(Nearly) Sample-Optimal Sparse Fourier Transform

Piotr Indyk
MIT

Michael Kapralov
MIT*

Eric Price
MIT

November 14, 2013

Abstract

We consider the problem of computing a k-sparse approximation to the discrete Fourier transform of an n-dimensional signal. Our main result is a randomized algorithm that computes such an approximation using $O(k \log n (\log \log n)^{O(1)})$ signal samples in time $O(k \log^2 n (\log \log n)^{O(1)})$, assuming that the entries of the signal are polynomially bounded. The sampling complexity improves over the recent bound of $O(k \log n \log(n/k))$ given in [HIKP12b], and matches the lower bound of $\Omega(k \log(n/k)/\log \log n)$ from the same paper up to poly$(\log \log n)$ factors when $k = O(n^{1-\delta})$ for a constant $\delta > 0$.

*We acknowledge financial support from grant #FA9550-12-1-0411 from the U.S. Air Force Office of Scientific Research (AFOSR) and the Defense Advanced Research Projects Agency (DARPA).

1 Introduction

The discrete Fourier transform (DFT) is a ubiquitous computational problem. Its applications are broad and include signal processing, communications, and audio/image/video compression. The fastest algorithm for this problem is the Fast Fourier Transform (FFT), which computes DFT of an n-dimensional signal in $O(n \log n)$ time. The existence of DFT algorithms that are faster than FFT is among the central algorithmic questions that still remain open.

A general algorithm for computing the exact DFT must take at least linear time. In many applications, however, most of the Fourier coefficients of a signal are small or equal to zero, i.e., the output of the DFT is (approximately) *sparse*. This includes audio, image and video compression, where the sparsity provides the rationale underlying compression schemes such as MPEG and JPEG. Other applications involving sparse signals include MRI [Nis10], NMR [MEH09] and ultrasound imaging [KS01]. For sparse signals, the $\Omega(n)$ lower bound for the complexity of DFT no longer applies.

The goal of designing efficient DFT algorithms for (approximately) sparse signals has been a subject of a large body of research [Man92, GGI$^+$02, AGS03, GMS05, Iwe10, Aka10, HIKP12a, HIKP12b, LWC12, BCG$^+$12, HAKI12, PR13, HKPV13]. These works show that, for a wide range of signals, both the time complexity and the number of signal samples taken can be significantly sub-linear in n. From a different perspective, minimizing the sampling complexity for signals that are approximate sparse in the Fourier domain was also a focus of an extensive research in the area of *compressive sensing* [Don06, CT06].

The best known results obtained in those areas are summarized in the following table. For the sake of uniformity, we focus on results for signals that are not necessarily exactly sparse, and provide the so-called ℓ_2/ℓ_2 approximation guarantee[1]. In this case, the goal of an algorithm is as follows: given the signal x and the sparsity parameter k, output $\widehat{x}'$ satisfying

$$\|\widehat{x} - \widehat{x}'\|_2 \leq C \min_{k\text{-sparse } y} \|\widehat{x} - y\|_2, \tag{1}$$

where $\widehat{x}$ denotes the complex DFT of x. The algorithms are randomized and succeed with constant probability.

Reference	Time	Samples	Approximation	Signal model
[CT06, RV08, CGV12]	$\Omega(n)$	$O(k \log^3(k) \log(n))$	$C = O(1)$	worst case
[CP10]	$\Omega(n)$	$O(k \log n)$	$C = (\log n)^{O(1)}$	worst case
[HIKP12b]	$O(k \log(n) \log(n/k))$	$O(k \log(n) \log(n/k))$	any $C > 1$	worst case
[GHI$^+$13]	$O(k \log^2 n)$	$O(k \log n)$	$C = O(1)$	average case, $k = \Theta(\sqrt{n})$
This paper	$O^*(k \log^2 n)$	$O^*(k \log n)$	any $C > 1$	worst case

Figure 1: Bounds for the algorithms that recover k-sparse Fourier approximations . All algorithms produce an output satisfying Equation 1 with probability of success that is at least constant. We use $O^*(f(n))$ to denote a function bounded by $f(n)(\log \log n)^{O(1)}$.

In summary, it has been known how to either perform the sparse recovery in sub-linear time, or achieve $O(k \log n)$ sampling complexity, or obtain a constant-factor approximation guarantee, or make the algorithm

[1]Some of the algorithms [CT06, RV08, CGV12] can in fact be made deterministic, but at the cost of satisfying a somewhat weaker ℓ_2/ℓ_1 guarantee. Also, additional results that hold for exactly sparse signals are known, see e.g., [BCG$^+$12] and references therein.

work for arbitrary (i.e., worst case) signals. However, it was not known how to obtain all of these guarantees simultaneously, or even how to satisfy various subsets of those guarantees.

1.1 Our results

Our main result is an algorithm that, given an arbitrary signal x, computes a k-sparse approximation for any factor $C > 1$ using $O^*(k \log n)$ signal samples in time $O^*(k \log^2 n)$. This assumes that n is a power of 2 and has additional additive error $\|x\|_2/n^{\Theta(1)}$; both restrictions are common in previous work. Thus, the algorithm essentially provides the "best of all worlds" guarantees, modulo the poly$(\log \log n)$ factors and assuming $k < n^{1-\delta}$ for any constant $\delta > 0$. The sampling complexity of the algorithm essentially matches the lower bound of $\Omega(k \log(n/k)/\log \log n)$ from [HIKP12b], again up to the same restrictions.

1.2 Our Techniques

For the rest of this paper, we will consider the *inverse* discrete Fourier transform problem of estimating a sparse x from samples of $\widehat{x}$. This is an equivalent problem modulo some conjugation, and lets the notation be simpler because the analysis almost always works with the sparse vector.

Our algorithm follows a similar approach to [GMS05, HIKP12b], which try to adapt the methods of [CCF02, GLPS10] from arbitrary linear measurements to Fourier ones. We use a "filter" that lets us "hash" the k large frequencies to $B = O(k)$ buckets. This lets us "locate" – i.e., find the indices of – many of the large frequencies. We then "estimate" the value of x at these frequencies, giving a sparse estimate χ of x. To improve this estimate, we can repeat the process on $x - \chi$ by subtracting the influence of χ during hashing. This repetition will yield a good sparse approximation χ of x.

The methods of [CCF02, GLPS10] will, multiple times, take a set of B linear measurements of the form

$$\tilde{u}_j = \sum_{i:h(i)=j} s_i x_i$$

for random hash functions $h : [n] \to [B]$ and random sign changes s_i with $|s_i| = 1$. This denotes *hashing* to B *buckets*. With such ideal linear measurements, $O(\log(n/k))$ hashes suffice for sparse recovery, giving an $O(k \log(n/k))$ sample complexity.

To perform sparse Fourier transforms, [GMS05] and [HIKP12b] approximate $\tilde{u}$ using linear combinations of Fourier samples. They use *filters* to compute $u \approx \tilde{u}$ using somewhat more than B *Fourier* measurements. Choosing a filter involves a tradeoff between the approximation quality and increase in number of samples. As described in Section 3, for any parameter $R > 2$, using $O(B \log R)$ Fourier measurements we can get (very roughly) that $\|u - \tilde{u}\|_2 \leq \|x\|_2/R$. We refer to this error $(u - \tilde{u})$, which is mostly caused by elements x_i contributing to buckets other than $h(i)$, as "leakage."

The difference between [GMS05] and [HIKP12b] is largely driven by a different choice of filters. [GMS05] uses a filter with $R = O(1)$, which gives efficient sample complexity per hashing but involves lots of leakage. Dealing with this leakage requires multiple logarithmic factors of overhead in the number of hashes. By contrast, [HIKP12b] uses a filter with $R = n^{O(1)}$. This filter loses one logarithmic factor in sample complexity, but makes leakage negligible for polynomially bounded inputs. The rest of the algorithm then can proceed somewhat similarly to [GLPS10] and be optimal, giving $O(k \log n \log(n/k))$ sample complexity.

In this paper we observe that setting $R = n^{O(1)}$ is often overkill: in many cases the post-filtering parts of [HIKP12b] can tolerate a larger amount of leakage (and hence use a filter that performs fewer

measurements). Moreover, the situations where R must be large are precisely the situations where the post-filtering parts of [HIKP12b] can be made more efficient and use $o(\log(n/k))$ hashings. We give a broad outline of our analysis, starting with a special case.

Similar magnitude heavy hitters. Even with the "ideal" hashing $\tilde{u}$, we expect an average of around $\mu^2 = \mathrm{Err}_k^2(x)/B$ "noise" from the tail in each of the $B = O(k)$ buckets, where $\mathrm{Err}_k(x)$ denotes $\min_{k\text{-sparse } y} \|x - y\|_2$. This means that the post-filtering steps of the algorithm must already tolerate average noise of order μ^2.

For intuition, it is useful to consider recovery of a signal where the largest k coordinates are all between $\sqrt{R}\mu$ and $R\mu$ for a parameter $R \geq 2$. Then choosing the filter with $O(\log R)$ overhead, i.e. performing $O(B \log R)$ Fourier measurements, the average leakage will be

$$\frac{1}{B}\|\tilde{u} - u\|_2^2 \leq \frac{1}{R^2 B}\|x\|_2^2 \leq \frac{k \cdot (R\mu)^2 + \mathrm{Err}_k^2(x)}{R^2 B} < \mu^2.$$

This means that the post-filtering steps of the algorithm will succeed, giving a sample complexity of $O(k \log R \log(n/k))$. This is a great improvement over the $O(k \log n \log(n/k))$ sampling complexity of [HIKP12b] when R is small, but if R is polynomially large we have not gained anything.

The next insight is that we can use fewer than $\log(n/k)$ hashings if the smallest heavy hitter has value $\sqrt{R}\mu^2 \gg \mu^2$. Indeed, the bottleneck in [GMS05, HIKP12b] is the location phase, where we need to recover $\log(n/k)$ bits about each large frequency (in order to identify it among the n/k different frequencies in the bucket). While [GMS05, HIKP12b] recover these bits one at a time, their methods can actually recover $\Omega(\log R)$ bits per hashing in this case because the expected signal to noise ratio in each bucket is $\Omega(R)$. This gives a sample complexity of $O(k \log R \log_R(n/k)) = O(k \log(Rn/k))$.

Our algorithm uses the approach we just outlined, but also needs to cope with additional difficulties that we ignored in the sketch above. First, in the general case we cannot expect all heavy hitters to be in the range $[\sqrt{R}\mu^2, R\mu^2]$, and the argument above does not give any guarantees on recovery of smaller elements. Additionally, the sketch above ignores collisions during hashing, which cause us to only recover a constant fraction of the heavy hitters in each round. We now give an outline of our approach to the general problem.

General vectors. The above algorithm finds most of the large frequencies if they all have value between $\sqrt{R}\mu^2$ and $R\mu^2$ for a known R. More generally, if $\|x\|_2^2 \leq Rk\mu^2$, the same techniques can recover most of the frequencies of magnitude larger than $R^\delta \mu^2$ with sample complexity $O(\frac{1}{\delta} k \log(Rn/k))$, for a parameter $\delta > 0$: we perform $O(\log_{R^\delta}(n/k))$ hashings that each take $O(k \log R)$ samples. Call this algorithm $\mathcal{A}(R, \delta)$.

Our algorithm will repeat $\mathcal{A}(R, \delta)$ multiple times for some δ. After enough repetitions, we will recover almost every coordinate larger than $\sqrt{R}\mu^2$. The residual will then have norm bounded by $O(\sqrt{R}k\mu^2)$. Our algorithm takes the following form: we repeat $\mathcal{A}(\sqrt{R}, \delta)$ multiple times, then $\mathcal{A}(R^{1/4}, \delta)$, and so on. After $\log \log R$ rounds of this, the residual will have norm $O(k\mu^2)$ and we can perform recovery directly. For this technique to work with $(\log \log(Rn))^c$ overhead, we will show that $\log \log R$ repetitions of $\mathcal{A}(R, \delta)$ suffice to reduce the residual norm to $\sqrt{R}k\mu^2$, for some $\delta = \Omega(1/\log \log R)$.

A first attempt might be to set $\delta = 1/2$, thus recovering most of the coordinates larger than $\sqrt{R}k\mu^2$ in each stage. This leads to problems if, for example, the vector has $k/2$ elements of value $R^{.4}\mu^2$ and $k/2$ elements of value $R^{.6}\mu^2$. Then $\mathcal{A}(R, 1/2)$ will never recover the first $k/2$ coordinates, and collisions with those coordinates mean it will only recover a constant fraction of the second $k/2$ coordinates. So it takes

$\Omega(\log R) \gg \log\log R$ repetitions to reduce the residual from $R^{.6}k\mu^2$ to $\sqrt{R}k\mu^2$. This is too slow; we need to make the number of elements above $\sqrt{R}\mu^2$ decay doubly exponentially.

This suggests that we need a more delicate characterization of $\mathcal{A}(r,\delta)$. We show in our analysis that coordinates are recovered with high probability if they are "well-hashed," meaning that the total noise in the bucket is R^δ smaller than the value of the coordinate. Coordinates of magnitude $R^\delta\mu^2$ have a constant chance of being well-hashed (leading to singly exponential decay), and coordinates that are much larger than $R^\delta\mu^2$ have a higher chance of being well-hashed (ultimately yielding the required doubly exponential decay). Our analysis follows this outline, but has to handle further complications that arise from imperfect estimation phase. For simplicity, we first present the analysis assuming perfect estimation, and then give the proof without any assumptions.

General vectors: perfect estimation. We classify the elements of the signal into $1/\delta$ "levels" of elements between $[R^{\delta j}\mu^2, R^{\delta(j+1)}\mu^2]$ for $j = 0,\dots,1/\delta - 1$, as opposed to a single range like $[\sqrt{R}\mu^2, R\mu^2]$. We then bound the success rate of recovery at each level in terms of the number of elements in various levels above and below it.

To first approximation, coordinates are recovered and eliminated from the residual if they are well-hashed, and are not recovered if they are not well-hashed. And in most cases the probability that a large coordinate j is not well-hashed is dominated by the probability that it collides with a coordinate of magnitude at least $R^{-\delta}|x_j|^2$. In this approximation, if we set $m_\ell(t)$ to be the number of $|x_j|^2$ larger than $R^{\ell\delta}\mu^2$ after t rounds of the algorithm, then $\mathbb{E}[m_\ell(t+1)] \le m_\ell(t)m_{\ell-1}(t)/B$. Then m_0 doesn't decay—coordinates less than $R^\delta\mu^2$ will not be recovered by $\mathcal{A}(R,\delta)$—but m_1 decays exponentially, m_2 will then decay as 2^{-t^2}, and in general m_ℓ will decay as $2^{-\binom{t}{\ell}}$. With $\delta = 1/\log\log R$, we find that $m_{1/\delta-1}$ (which contains all coordinates larger than $\sqrt{R}\mu^2$) will decay to $1/R^c$ in $O(\log\log R)$ rounds. As a result, the squared norm of the residual will be at most $O(\sqrt{R}\mu^2)$. The details of this part of the analysis are presented in Section 6.

General vectors: actual behavior. In the actual algorithm, coordinates do not just disappear if they are located, but are estimated with some error. This means large components can appear in the residual where no component was before, if lots of small components were hashed to a certain bucket. This causes the m_ℓ to not obey the nice recurrence in the previous paragraph. To deal with this, we introduce the notion of *splittings* of the residual. For analysis purposes, we split each component of the residual into multiple terms whose total magnitude is the same. We define the m_ℓ in terms of the number of components in the splitting, not the actual residual.

The intuition is that the residual error when estimating an element x_i is approximately $\|x_C\|_2$, where $C \subset [n]$ is the set that "collides" with i. Rather than thinking of the residual as a single coordinate with value $\|x_C\|_2$, we "split" it and imagine duplicating x_j for each $j \in C$. Because $j \in C$ was not recovered from the bucket, j was (most likely) not well-hashed. So the contribution of the duplicated x_j to m_ℓ is comparable to the contribution of the x_j that remain after not being well-hashed. Hence the m_ℓ obey almost the same recurrence as in the perfect estimation setting above.

As a result, $O(\log\log R)$ repetitions of $\mathcal{A}(R, 1/\log\log R)$ reduce the residual norm to $\sqrt{R}k\mu^2$. Repeating for $\log\log n$ rounds decreases R from n^c to $O(1)$, and we can finish off by accepting a $\log R$ loss. The details of this part of the analysis are presented in Section 7.

2 Notation and definitions

We use the notation $[n] = \{0, 1, \ldots, n-1\}$. For $x \in \mathbb{R}^n$ the Fourier transform of x is given by

$$\widehat{x_j} = \frac{1}{\sqrt{n}} \sum_{i \in [n]} \omega^{ij} x_i, \tag{2}$$

where ω is a root of unity of order n. We may also denote the Fourier transform $\widehat{x}$ of x by $\mathcal{F}(x)$. The inverse transform is given by

$$x_j = \frac{1}{\sqrt{n}} \sum_{i \in [n]} \omega^{-ij} \widehat{x_i}. \tag{3}$$

By the convolution theorem, $\widehat{x * y} = \sqrt{n}\widehat{x} \cdot \widehat{y}$.

We assume that n is a power of 2.

2.1 Notation

We use $f \gtrsim g$ to denote $f = \Omega(g)$ and $f \lesssim g$ to denote $f = O(g)$.

We will use (pseudorandom) spectrum permutations [HIKP12b, GMS05], which we now define.

Definition 2.1. *Suppose that σ^{-1} exists $\mod n$. We define the permutation $P_{\sigma,a,b}$ by $(P_{\sigma,a,b}\hat{x})_i = \hat{x}_{\sigma(i+a)}\omega^{-\sigma bi}$. We also define $\pi_{\sigma,b}(i) = \sigma(i-b) \mod n$.*

This has the following useful property, proven in Section 11.

Claim 2.2. (Claim 2.2 of [HIKP12b]). *Let $\mathcal{F}^{-1}(x)$ denote the inverse Fourier transform of x. Then*

$$(\mathcal{F}^{-1}(P_{\sigma,a,b}\widehat{x}))_{\pi(i)} = x_i \omega^{a\sigma i}.$$

Also, define

- $h_{\sigma,b}(i) = \text{round}(\pi_{\sigma,b}(i)n/B)$ to be an $[n] \to [B]$ "hash function" .
- $o_i(j) = \pi(j) - (n/B)h(i)$ to be the "offset" of j relative to i.

This "hashing" h is approximately pairwise independent in the following sense:

Lemma 2.3 (Lemma 3.6 of [HIKP12a])**.** *If $j \neq 0$, n is a power of two, and σ is a uniformly random odd number in $[n]$, then $\Pr[\sigma j \in [-C, C] \pmod n] \leq 4C/n$ for all C.*

In much of the paper, we use $|i|$ for $i \in [n]$ to denote $\min_{z \in \mathbb{Z}} |i + zn|$; this is the "absolute value modulo n." So the above lemma, for example, bounds $\Pr[|\sigma j| \leq C]$.

Define

$$\text{Err}_k(x) = \min_{k\text{-sparse } y} ||x - y||_2.$$

Our algorithm will start with an input $\widehat{x}^*$ and find progressively better approximations χ to x^*. Most of the analysis will depend only on $x := x^* - \chi$. Our algorithm will involve decreasing the "signal to noise ratio" $R \approx \|x\|_2^2 / \text{Err}_k^2(x^*)$.

We give the pseudocode for our algorithm below (the pseudocode for the function LOCATESIGNAL, which follows the location function in [HIKP12b] quite closely, appears in Section 10 together with its analysis).

The main algorithm is Algorithm 1. Its performance analysis, which builds upon the analysis of primitives REDUCESNR and RECOVERATCONSTANTSNR, is provided in Section 8.

Algorithm 1 Overall algorithm: perform Sparse Fourier Transform

procedure SPARSEFFT($\widehat{x}, k, \epsilon, R, p$)
  $\chi^{(0)} \leftarrow 0$ ▷ in $\mathbb{C}^n$.
  $R_0 \leftarrow R$
  $r \leftarrow \Theta(\log \log R)$
  for $i = 0, 1, \ldots, r-1$ **do**
    $\chi' \leftarrow$ REDUCESNR($\widehat{x}, \chi^{(i)}, 3k, R_i, p/(2r)$)
    $\chi^{(i+1)} \leftarrow$ SPARSIFY($\chi^{(i)} + \chi', 2k$) ▷ Zero out all but top $2k$ entries
    $R_{i+1} \leftarrow c\sqrt{R_i}$ ▷ For some constant c
  end for
  $\chi' \leftarrow$ RECOVERATCONSTANTSNR($\widehat{x}, \chi^{(r)}, 3k, \epsilon, p/2$)
  return $\chi^{(r)} + \chi'$
end procedure

Our SNR reduction primitive is given by Algorithm 2. Its analysis is the most technical part of the paper and is given in Sections 6 and Section 7.

Algorithm 2 Reduce the SNR $\|x\|_2^2/\xi^2$ from R to $O(\sqrt{R})$

procedure REDUCESNR($\widehat{x}, \chi, k, R, p$)
  $B \leftarrow \frac{1}{\alpha}k$ for sufficiently small $\alpha > 0$.
  $\chi^{(1)} \leftarrow \chi$
  $N \leftarrow \Theta(\log_2 \log_2 R)$
  for $t = 0, 1, \ldots, N-1$ **do**
    $k_t \leftarrow O(k4^{-t})$
    $L \leftarrow$ LOCATESIGNAL($\widehat{x}, \chi^{(i)}, B, \sigma, b, R, \alpha R^{-20}$)
    $\tilde{x} \leftarrow$ ESTIMATEVALUES($\widehat{x}, \chi^{(i)}, L, B, 3k_t, 13, R$)
    $\chi^{(i+1)} \leftarrow \chi^{(i)} + \tilde{x}$
  end for
  return $\chi^N - \chi$
end procedure

The SNR reduction primitive given by Algorithm 2 allows us to reduce the problem to the constant SNR case. The recovery primitive for that case is given by Algorithm 3. Its analysis is presented in Section 5.

Algorithm 3 Recovery when $\|x - \chi\|_2 \lesssim \mathrm{Err}_k^2(x)$

procedure RECOVERATCONSTANTSNR($\hat{x}, \chi, k, \epsilon, p$)
  $R \leftarrow 20$
  $B \leftarrow Rk/(\epsilon\alpha p)$ for a sufficiently small constant $\alpha > 0$
  Choose σ, b uniformly at random in $[n]$, σ odd.
  $L \leftarrow$ LOCATESIGNAL$(\hat{x}, \chi, B, \sigma, b, R, \alpha\epsilon p)$
  $\chi' \leftarrow$ ESTIMATEVALUES$(\hat{x}, \chi, L, B, 3k, \log(B/(4k)), R)$
  return χ'
end procedure

The ESTIMATEVALUES and HASHTOBINS primitives are similar to the ones in [HIKP12b], but differ in that in HASHTOBINS the signal recovered so far is subtracted from the samples as opposed to buckets:

Algorithm 4 Estimation: estimate $(x - \chi)_L$ using T rounds of B-bucket, contrast R hashing.

procedure ESTIMATEVALUES($\widehat{x}, \chi, L, B, k, T, R$)
  for $t = 1$ to T **do**
    Choose $\sigma, b, a \in [n]$ uniformly at random, σ odd
    $u \leftarrow$ HASHTOBINS$(x, \chi, P_{\sigma,a,b}, B, R)$
    $\tilde{x}_i^{(t)} \leftarrow G^{-1}_{o_i(i)} u_{h_{\sigma,b}(i)} \omega^{-a\sigma i}$ for all $i \in L$. ▷ Note that $G_{o_i(i)}$ depends on σ, b, a
  end for
  $\tilde{x}_i \leftarrow \mathrm{median}_t(\tilde{x}_i^{(t)})$ for all $i \in L$. ▷ Median in real and imaginary axis separately
  return SPARSIFY$(\tilde{x}, k)$.
end procedure

Algorithm 5 Hashing using Fourier samples (analyzed in Lemma 11.3)

procedure HASHTOBINS($\widehat{x}, \chi, P_{\sigma,a,b}, B, R$)
  $G \leftarrow$ flat window function with B buckets and contrast R.
  Compute $y' = \hat{G} \cdot P_{\sigma,a,b}(\hat{x} - \hat{\chi}')$, for some χ' with $\|\widehat{\chi} - \widehat{\chi}'\|_\infty < \frac{\|\chi\|_2}{R^* n^{13}}$ ▷ Have $\|y'\|_0 \lesssim B \log R$
  Compute $u_j = \sqrt{n}\mathcal{F}^{-1}(y')_{jn/B}$ for $j \in [B]$
  return u
end procedure

2.2 Glossary of Terms in REDUCESNR and RECOVERATCONSTANTSNR

In REDUCESNR and RECOVERATCONSTANTSNR, there are a lot of variables with common names and similar purposes. This section provides a glossary, which may be useful for reference. We have globally:

- $\widehat{x}^* \in \mathbb{C}^n$ is the original input, where we want to recover an approximation to x^*.
- k^* is the original value of k, for which we expect $\widehat{x}^*$ to be approximately k^*-sparse.
- $R^* \geq \|x^*\|_2^2 / \mathrm{Err}_k^2(x^*)$ is an upper bound on the SNR for the original signal. We have $R^* = O(\mathrm{poly}(n))$ by the input assumption.

and for each different call to REDUCESNR and RECOVERATCONSTANTSNR, we have

- $\chi \in \mathbb{C}^n$ is our current best guess at x, which will (in all calls) be $2k^*$-sparse.
- $x = x^* - \chi$ is the "residual" signal that we want to recover in this call. (When analyzing REDUCESNR, we set $x = x^* - \chi^{(i)}$ in each inner loop.)
- $k = 3k^*$ is the approximate sparsity of x.
- $\alpha > 0$ is sufficiently small, but at least $1/(\log\log n)^c$.
- $B > k/\alpha$ to be the number of buckets in each hashing.
- T to be the number of hashings done inside ESTIMATEVALUES.
- $\tilde{x}^{(t)} \in \mathbb{C}^n$ is the estimation of x in ESTIMATEVALUES for each $t \in [T]$.
- $\tilde{x} \in \mathbb{C}^n$ is the median of $\tilde{x}^{(t)}$ over t.
- R will is a parameter (in REDUCESNR) and sufficiently large constant (in RECOVERATCONSTANTSNR). It roughly represents the "signal-to-noise ratio".
- $\delta = 1/(40\log_2\log_2 R)$.
- $\gamma = R^{-\delta}$ to be the "contrast" our LOCATESIGNAL requires.

3 Properties of the bucketing scheme

Our algorithm uses filters and various choices of σ, b, a to "hash" the coordinates of x into buckets. For each (σ, b, a) and each bucket $j \in [B]$ we recover an estimate i^* of the heavy coordinate in that bucket. Also, for each $i \in [n]$ we can recover an estimate $\tilde{x}_i$ of x_i.

3.1 Filter properties

Our main tool is a generalization of filters from [HIKP12b] that allows the noise in a single bucket depends on both the energy of the signal as well as the number of elements that hashed into it.

Definition 3.1 (Flat Window Functions)**.** *A flat window function G over $\mathbb{C}^n$ has B* buckets *and* contrast *R if, for $|i| \le n/2$, we have*

- *$G_i \ge 1/3$ for $|i| \le n/(2B)$.*
- *$0 \le G_i \le 1$ for all i.*
- *$G_i \le (\frac{cn}{|i|B})^{\log R}$ for all i for some constant c*

The filters of [HIKP12b] were essentially the case of $R = n^{O(1)}$. We will prove in Section 11 that

Lemma 3.2. *There exist flat window functions where $|\operatorname{supp}(\widehat{G})| \lesssim B\log R$. Moreover, $\operatorname{supp}(\widehat{G}) \subset [-O(B\log R), O(B\log R)]$.*

The analysis in this paper will assume we have precomputed $\widehat{G}$ and G and may access them with unit cost. This is unnecessary: in Section 12.1 we describe how to compute them on the fly to $1/n^c$ precision without affecting our overall running time. This precision is sufficient for our purposes.

Lemma 3.3. *Let $(\sigma, a, b) \in [n]$ be uniform subject to σ being odd. Let $u \in \mathbb{C}^B$ denote the result of* HASHTOBINS*$(\widehat{x}^*, \chi, P_{\sigma,a,b}, B, R)$. Fix a coordinate $i \in [n]$ and define $x = x^* - \chi$. For each (σ, b), we can define variables $C \subset [n]$ and $w > 0$ (and in particular, $C = \{j \ne i : |\sigma(i-j) \bmod n| \le cn/B\}$ for some constant c,) so that*

- *For all j, as a distribution over (σ, b),*

$$\Pr[j \in C] \lesssim 1/B.$$

- *As a distribution over (σ, b),*

$$\mathbb{E}[w^2] \lesssim \frac{\|x\|_2^2}{R^2 B} + \frac{\|x^*\|_2^2}{R^* n^{11}}$$

- *Conditioned on (σ, b) and as a distribution over a,*

$$\mathop{\mathbb{E}}_{a}[|G^{-1}_{o_i(i)}\omega^{-a\sigma i}u_{h(i)} - x_i|^2] \lesssim w^2 + \|x_C\|_2^2.$$

Intuitively, C denotes the elements of x that collide with i, and w denotes the rest of the noise. The two terms of w correspond to leakage of x from other hash locations and to errors in the subtraction of χ, respectively. This latter term should be thought of as negligible.

We also define the notion of being "well-hashed," which depends on another parameter $\gamma = R^\delta$ from the glossary:

Definition 3.4. *Let $\sigma, b \in [n], \sigma$ odd. An element i is well-hashed for a particular σ, b and filter G if over uniformly random $a \in [n]$,*

$$\mathop{\mathbb{E}}_{a}[|G^{-1}_{o_i(i)}\omega^{-a\sigma i}u_{h(i)} - x'_i|^2] \le \gamma^{1/2}|x'_i|^2.$$

Intuitively, a well-hashed element contains little noise in the bucket that it hashed to, relative to its own energy, and will hence be likely to be recovered in LOCATESIGNAL. This is formalized in Lemma 10.2.

4 Proof Overview

This section gives the key lemmas that are proven in later sections. Our procedures try to reduce the ℓ_2 norm of the residual to the "noise level" $\xi^2 := \mathrm{Err}^2_{k^*}(x^*) + \|x^*\|_2^2/(R^* n^{10})$. The polynomial n^{10} can be arbitrary, and only affects the running time of the algorithm; we choose a specific constant for simplicity of notation. The $\|x^*\|_2^2/(R^* n^{10})$ term is essentially irrelevant to the behavior of the algorithm, and will be ignored when discussing intuition.

First, we give an algorithm RECOVERATCONSTANTSNR that is efficient when $\|x\|_2^2 \lesssim \mathrm{Err}^2_k(x)$.

Lemma 5.1. *For $x^*, \chi \in \mathbb{C}^n$ define $x = x^* - \chi$. Then* RECOVERATCONSTANTSNR$(\widehat{x}, \chi, k, \epsilon, p)$ *returns χ' such that*

$$\|x - \chi'\|_2^2 \le \mathrm{Err}^2_k(x) + \epsilon\|x\|_2^2 + \frac{\|x^*\|_2^2}{n^{10}}$$

with probability $1-p$, using $O(\frac{1}{p\epsilon}k\log(n/k)\log\log(n/k)\log(1/(\epsilon p)))$ measurements and (assuming $\|\chi\|_0 \lesssim k$) a $\log n$ factor more time.

This is relatively straightforward. To see why it is useful, for $k = 3k^*$ we have $\mathrm{Err}^2_k(x) \le \mathrm{Err}_{k^*}(x^*)$. Therefore, once χ is close enough to x^* that $x = x^* - \chi$ has $\|x\|_2^2 \lesssim \mathrm{Err}^2_{k^*}(x^*)$, this lemma gives that $\chi + \chi'$ is within $(1+\epsilon)\,\mathrm{Err}^2_{k^*}(x^*)$ of x^* using only $O^*(\frac{1}{p\epsilon}k\log(n/k)\log(1/(\epsilon p)))$ measurements. (As stated above, for intuition we are ignoring the negligible $\frac{\|x^*\|_2^2}{n^{10}}$ term.)

We then show how to quickly reduce $\|x\|_2^2$ to $O(\mathrm{Err}^2_{k^*}(x^*))$:

Lemma 7.11. *For $x^*, \chi \in \mathbb{C}^n$ define $x = x^* - \chi$ and*

$$\xi^2 = \mathrm{Err}_k^2(x) + \frac{\|x\|_2^2}{R} + \frac{\|x^*\|_2^2}{R^* n^{10}}.$$

Suppose $\|\chi\|_0 \le k$. Then REDUCESNR$(\widehat{x}^*, \chi, k, R, p)$ *returns χ' such that*

$$\|x - \chi'\|_2^2 \lesssim \sqrt{R}\xi^2$$

with probability $1-p$, using $O(\frac{1}{p^2} k \log(Rn/k)(\log\log(Rn/k))^c)$ measurements and a $\log(Rn)$ factor more time.

This is the most technical part of our paper. By iterating it $\log\log R$ times and finishing off with Lemma 5.1, we get the final result:

Theorem 8.1. *Let $x \in \mathbb{C}^n$ satisfy $\|x\|_2^2 \le R\,\mathrm{Err}_k^2(x)$. Then* SPARSEFFT$(\widehat{x}, k, R, p)$ *returns a χ' such that*

$$\|x - \chi'\|_2^2 \le (1+\epsilon)\,\mathrm{Err}_k^2(x) + \|x\|_2^2/(R^* n^{10})$$

with probability $1-p$ and using $O(\frac{1}{p^2\epsilon} k \log(Rn/k)(\log\log(Rn/k))^c \log(1/\epsilon))$ measurements and a $\log(Rn)$ factor more time.

We now summarize the proof of Lemma 7.11. Let $x = x^* - \chi$, and define

$$\mu^2 = \frac{1}{k}\xi^2 \ge \left(\mathrm{Err}_k^2(x) + \frac{\|x\|_2^2}{R}\right)/k.$$

If we hash to $B = k/\alpha$ buckets with flat window functions of contrast R, then the expected magnitude of the contribution of the tail of x to any bucket is $O(\alpha\mu^2)$.

REDUCESNR involves $O(\log\log R)$ stages. In each stage, we hash to B bins and call LOCATESIGNAL to get a set L of candidate locations for heavy hitters. We then estimate x_i for each $i \in L$ as the median $\tilde{x}_i$ of $O(1)$ random hashings to B bins. We then subtract off $\tilde{x}_{L'}$, where L' contains the largest k' coordinatse of $\tilde{x}$ and k' starts out $\Theta(k)$ in the first stage and decreases exponentially. So the recurrence in each stage is $x \to x - \tilde{x}_{L'}$.

This process is somewhat complicated, so we start by analyzing a simpler process in each stage. Let S denote the set of "well-hashed" coordinates $i \in [n]$, i.e. coordinates that are hashed to bins with noise less than $\gamma^{1/2}|x_i|^2$. In Section 6 we analyze the recurrence $x \to x - x_S$. Generally, we expect larger elements to be more likely to be well-hashed, and so the number of them to decay more quickly. We analyze the number $m_\ell(t)$ of i with $|x_i| > \mu^2\gamma^{-\ell}$ that remain at each stage t, for each level ℓ. We show that these quantities obey a nice system of equations, causing the $m_\ell(t)$ to decay doubly exponentially for the first ℓ rounds. Then after $t = O(\log\log R)$ rounds, an R^{-10} fraction of the coordinates larger than $\mu^2\sqrt{R}$ remain. This means that the recurrence $x \to x - x_S$ would leave a remainder of norm $O(k\mu^2\sqrt{R})$ as desired.

In Section 7, we relate this to the true recurrence $x \to x - \tilde{x}_{L'}$. We study recurrences that are *admissible*, meaning that they satisfy a similar system of equations to that in Section 6. Admissible recurrences satisfy composition rules that let us find them sequentially, and using Section 6 we can show the remainder after $\log\log R$ iterations of any admissible recurrence has small norm. In a series of results, we show that $x \to x - \tilde{x}_S$, $x \to x - x_{L'}$, and finally $x \to x - \tilde{x}_{L'}$ are admissible. This then proves Lemma 7.11.

5 Constant SNR

Our procedure for recovery at constant SNR is given by Algorithm 3. In this section we prove

Lemma 5.1. *For $x^*, \chi \in \mathbb{C}^n$ define $x = x^* - \chi$. Then* RECOVERATCONSTANTSNR$(\widehat{x}, \chi, k, \epsilon, p)$ *returns χ' such that*

$$\|x - \chi'\|_2^2 \leq \mathrm{Err}_k^2(x) + \epsilon\|x\|_2^2 + \frac{\|x^*\|_2^2}{n^{10}}$$

with probability $1-p$, using $O(\frac{1}{p\epsilon}k\log(n/k)\log\log(n/k)\log(1/(\epsilon p)))$ measurements and (assuming $\|\chi\|_0 \lesssim k$) a $\log n$ factor more time.

In what follows we define

$$\xi^2 = \|x\|_2^2 + \|x^*\|_2^2/(R^*n^{11}).$$

and $\mu^2 = \xi^2/k$. By definition of the algorithm, $B = Rk/(\epsilon\alpha p)$ for some constants R, α. We will show that, if R is a sufficiently large constant, then with probability $1-p$,

$$\|x - \chi'\|_2^2 - \mathrm{Err}_k^2(x) \lesssim \alpha\epsilon\xi^2.$$

For sufficiently small α this gives the result.

A simple consequence of Lemma 3.3 is that for each i, and for random (σ, a, b), we have

$$\mathop{\mathbb{E}}_{a,\sigma,b}[|G^{-1}_{o_i(i)}\omega^{-a\sigma i}u_{h(i)} - x_i|^2] \lesssim \|x\|_2^2/B + \|x^*\|_2^2/(R^*n^{11}) \leq \xi^2/B = \epsilon\alpha p\mu^2/R. \tag{4}$$

There are two sources of error in RECOVERATCONSTANTSNR, coming from location and estimation respectively. The proof of Lemma 5.1 proceeds in two stages, bounding the error introduced in both steps.

5.1 Energy lost from LOCATESIGNAL

Let S contain the largest k coordinates of x and L be the list of locations output by LOCATESIGNAL. In this section we bound the energy of the vector $x_{S\setminus L}$. Define

$$A_{large} = \{i \in S : |x_i|^2 \geq \alpha\epsilon\mu^2/R\}$$
$$A_{small} = \{i \in S : |x_i|^2 \leq \alpha\epsilon\mu^2/R\},$$

so that

$$||x_{A_{small}}||^2 \leq \alpha\epsilon\mu^2 k/R \leq \alpha\epsilon\xi^2. \tag{5}$$

For each $i \in [n]$ by (4) we have

$$\mathop{\mathbb{E}}_{a,\sigma,b}[|G^{-1}_{o_i(i)}\omega^{-a\sigma i}u_{h(i)} - x_i|^2] \lesssim \alpha\epsilon p\mu^2/R.$$

Consider $i \in A_{large}$, and recall Definition 3.4 of being well-hashed. By Markov's inequality applied to (4) and $R > \gamma^{-1/2}$, the probability that $i \in A_{large}$ is not well-hashed is bounded by

$$\frac{\epsilon\alpha p\mu^2/R}{\gamma^{1/2}|x_i|^2} \leq \frac{\epsilon\alpha p\mu^2}{|x_i|^2}. \tag{6}$$

Each well-hashed element is then located in LOCATESIGNAL with probability at least $1 - \alpha\epsilon p$ by our choice of parameters. Thus, for $i \in A_{large}$ one has

$$\Pr[i \notin L] \leq \frac{\epsilon\alpha p\mu^2}{|x_i|^2} + O(\alpha\epsilon p).$$

It then follows that

$$\begin{aligned}\mathbb{E}[||x_{S\setminus L}||^2 - ||x_{A_{small}}||^2] &= \mathbb{E}[||x_{A_{large}\setminus L}||^2] \\ &\leq \sum_{i\in A_{large}} \frac{\epsilon\alpha p\mu^2}{|x_i|^2}|x_i|^2 + \alpha\epsilon p\|x\|_2^2 \leq \alpha\epsilon p\xi^2 \\ &\leq 2\alpha\epsilon p\xi^2.\end{aligned} \tag{7}$$

Combined with (5) one has

$$||x_{S\setminus L}||^2 \lesssim \alpha\epsilon\xi^2 \tag{8}$$

with probability at least $1 - p/2$ by Markov's inequality. It remains to consider the effect of pruning in ESTIMATEVALUES.

5.2 Energy of $x - \chi'$

We now analyze the errors introduced in the estimation step. These errors come from two sources: estimation noise and the pruning step in ESTIMATEVALUES. Let $\tilde{x}^{(t)}$ denote the estimate in each hashing in ESTIMATEVALUES (defined to be zero outside L), and $\tilde{x}$ denote the coordinate-wise median over t of $\tilde{x}^{(t)}$. By definition, $\chi' = \tilde{x}_{L'}$ where L' denotes the largest $3k$ elements of $\tilde{x}$. By (4), for each $i \in L$ and $t \in [T]$ during estimation we have

$$\mathbb{E}[|\tilde{x}_i^{(t)} - x_i|^2] \lesssim \epsilon\alpha p\mu^2/R \leq \epsilon\alpha p\mu^2,$$

and so by properties of the median (Lemma 9.5),

$$\mathbb{E}[|\tilde{x}_i - x_i|^2] \leq 4\,\mathbb{E}[|\tilde{x}_i^{(t)} - x_i|^2] \lesssim \epsilon\alpha p\mu^2 \tag{9}$$

for all i. Now, by Lemma 9.1,

$$\|x - \chi'\|_2^2 = \|x - \tilde{x}_{L'}\|_2^2 \leq \mathrm{Err}_k^2(x) + 4\|(x - \tilde{x})_{S\cup L'}\|_2^2. \tag{10}$$

The first term appears in our output, so it is sufficient to upper bound the last term by $O(\alpha\epsilon\xi^2)$ with probability $1 - p$. We write

$$\|(x - \tilde{x})_{S\cup L'}\|_2^2 \leq \|(x - \tilde{x})_{S\setminus L}\|_2^2 + \|(x - \tilde{x})_{(S\cap L)\cup L'}\|_2^2. \tag{11}$$

The first term is bounded by (8). It remains to bound this last bit, which is entirely the effect of estimation error since $(S \cap L) \cup L' \subseteq L$. By the fact that $|(S \cap L) \cup L'| \leq 4k$, Lemma 9.4 with $T = O(\log(B/4k))$, and (9),

$$\begin{aligned}\mathbb{E}[\|(x - \tilde{x})_{(S\cap L)\cup L'}\|_2^2] &\leq \max_{A\subseteq L, |A|=4k} \|(x - \tilde{x})_A\|_2^2 \\ &\lesssim 4k \cdot (B/4k)^{\Theta(1/T)} \cdot \max_i \mathbb{E}[|x_i - \tilde{x}_i^{(t)}|^2] \\ &\lesssim k \cdot 1 \cdot \epsilon\alpha p\mu^2 \\ &= \epsilon\alpha p\xi^2.\end{aligned}$$

Hence by Markov's inequality, with probability at least $1 - p/2$ one has $\|(x - \tilde{x})_{(S\cap L)\cup L'}\|_2^2 \lesssim \alpha\epsilon\xi^2$, and putting this together with (10) and (11), we get

$$\begin{aligned} \|x - \chi'\|_2^2 &\leq \mathrm{Err}_k^2(x) + O(\alpha\epsilon)\xi^2 \\ &\leq \mathrm{Err}_k^2(x) + \epsilon\xi^2 \end{aligned} \tag{12}$$

with probability at least $1 - p$, for sufficiently small constant α.

Proof of Lemma 5.1. The guarantee on the residual error is provided by (12), so it remains to verify sampling complexity. The call to LOCATESIGNAL takes order

$$B\log(Rn/B)\log\log R\log\log(n/B)\log(1/(\alpha\epsilon p)) \lesssim \frac{1}{p\epsilon}k\log(n/k)\log\log(n/k)\log(1/(\epsilon p))$$

samples by Lemma 10.2. The call to ESTIMATEVALUES takes order

$$\log(B/4k)B\log R \lesssim \frac{1}{\epsilon p}k\log(1/(\epsilon p))$$

samples, giving the desired total sample complexity. □

6 Reducing SNR: idealized analysis

6.1 Dynamics of the process with simplifying assumptions

The goal of this section and the next is to prove the following lemma:

Lemma 7.11. *For $x^*, \chi \in \mathbb{C}^n$ define $x = x^* - \chi$ and*

$$\xi^2 = \mathrm{Err}_k^2(x) + \frac{\|x\|_2^2}{R} + \frac{\|x^*\|_2^2}{R^* n^{10}}.$$

Suppose $\|\chi\|_0 \leq k$. Then REDUCESNR$(\widehat{x}^*, \chi, k, R, p)$ *returns χ' such that*

$$\|x - \chi'\|_2^2 \lesssim \sqrt{R}\xi^2$$

with probability $1-p$, using $O(\frac{1}{p^2}k\log(Rn/k)(\log\log(Rn/k))^c)$ measurements and a $\log(Rn)$ factor more time.

In this section we give a description of iterative process in REDUCESNR under simplifying assumptions, demonstrating the basic dynamics of the process. We will later give a general analysis.

Define $\mu^2 = \xi^2/k$, and from the glossary (Section 2.2) recall the definitions

$$\delta = \frac{1}{40\log_2\log_2 R}, \quad \gamma = R^{-\delta}.$$

We define the following *energy levels*. For each $j = 1, \ldots, 1/\delta - 1$ let

$$L_j = [\mu^2 \cdot \gamma^{-j}, \mu^2\gamma^{-(j+1)}],$$

and let $L_0 := [0, \mu^2\gamma^{-(j+1)}]$ and $L_{1/\delta} := [\mu^2\gamma^{-1/\delta}, \infty) = [\mu^2 R, \infty)$.

Simplifying assumptions. Recall the notion of well-hashedness (Definition 3.4). The crucial property of well-hashed elements is that if $i \in [n]$ is well-hashed, then an invocation of LOCATESIGNAL locates it with probability at least $1 - 1/\text{poly}(R)$. This property is proved in Lemma 10.2. In this section we make the following simplifying assumption: we assume that each well-hashed element $i \in [n]$ is estimated with zero error and removed from the signal. The elements that are not well-hashed, on the other hand, we assume simply remain in the system untouched. Let H denote the set of well-hashed elements (which is close to the list of locations output by LOCATESIGNAL). In this section, therefore, we analyze the recursion $x \to x - x_H$.

For each $x_i \in L_j$ and each $t \geq 1$ let $\mathbf{1}_{i,t}$ denote the indicator variable equal to 1 if x_i survived up to the t-th round of the process and 0 otherwise. For each $j \in [1 : 1/\delta]$ and $t \geq 1$ let

$$m_j(t) = \frac{1}{k} \sum_{j' \geq j} \sum_{i \in [k] : (x_i)^2 \in L_{j'}} \mathbf{1}_{i,t}.$$

Recall that by Definition 3.4 an element $i \in [n]$ is well-hashed for a particular choice of $\sigma, b \in [n], \sigma$ odd, and filter G if over uniformly random $a \in [n]$,

$$\mathop{\mathbb{E}}_{a}[|G^{-1}_{o_i(i)} \omega^{-a\sigma i} u_{h(i)} - x_i|^2] \leq \gamma^{1/2} |x_i|^2.$$

Lemma 6.1. *Let $\sigma, b \in [n]$ be chosen uniformly at random, σ odd. Let $i \in [n]$ denote an index such that $|x_i|^2 \in L_j$. Then the probability that i is not well-hashed at time t is at most of order*

$$\alpha \left(\gamma^{j-1/2} + \sum_{j' < j-1} \gamma^{j-j'-3/2} m_{j'}(t) \right) + \alpha m_{j-1}(t),$$

where the number of buckets B satisfies $B \geq k/\alpha$.

Proof. Let $x^{(h)}$ denote all elements of x in levels L_{j-1} and above. Denote the set of such elements by S^+. Let $x^{(t)}$ denote all elements of x in $L_{j'}, j' < j - 1$. Since $|x_i|^2 \in L_j$, we have $|x_i|^2 \geq \gamma^{-j} \mu^2$.

Define C to be the indices that "collide with" i as in Lemma 3.3. We have that

$$\Pr[C \cap S^+ \neq \{\}] \lesssim |S^+|/B = \alpha m_{j-1}(t).$$

Condition on the event that $C \cap S^+ = \{\}$; since this happens with more than $1/2$ probability, the conditioning only loses a constant factor in expectations and we may neglect this influence. We have by Lemma 3.3 that

$$\mathop{\mathbb{E}}_{\sigma,b,a}[|G^{-1}_{o_i(i)} \omega^{-a\sigma i} u_{h(i)} - x_i|^2] \lesssim \|x\|_2^2/(R^2 B) + \mathop{\mathbb{E}}_{\sigma,b}[\|x_C\|_2^2] + \frac{1}{R^* n^{11}} \|x^*\|_2^2. \tag{13}$$

$$\tag{14}$$

Recall that by the definition of $\mu^2 = \xi^2/k$,

$$\|x\|_2^2/(R^2 B) + \frac{1}{R^* n^{11}} \|x^*\|_2^2 \leq \|x\|_2^2/(RB) + \frac{1}{BR^* n^{10}} \|x^*\|_2^2 \leq \alpha \mu^2.$$

Furthermore, recall that by Lemma 3.3, (1) any given element belongs to C with probability $O(1/B)$. Since the energy of an element in $L_{j'}$ is bounded above by $\gamma^{-(j'+1)} \mu^2$ by definition of $L_{j'}$, we get that

$$\mathop{\mathbb{E}}_{\sigma,b}[\|x_C\|_2^2 | C \cap S^+ = \{\}] \leq \alpha \mu^2 \sum_{j' < j-1} \gamma^{-(j'+1)} m_{j'}(t).$$

Putting these two estimates together, we get that the rhs of (13) is bounded by

$$\alpha\mu^2 + \alpha\mu^2 \sum_{j'<j-1} \gamma^{-(j'+1)} m_{j'}(t),$$

Therefore, conditioned on $C \cap S^+ = \{\}$, we have

$$\begin{aligned}
\Pr[i \text{ not well-hashed}] &= \Pr_{\sigma,b}[\mathop{\mathbb{E}}_{a}[|G^{-1}_{o_i(i)}\omega^{-a\sigma i}u_{h(i)} - x_i|^2] \geq \gamma^{1/2}|x_i|^2] \\
&\leq \frac{\mathbb{E}_{\sigma,b,a}[|G^{-1}_{o_i(i)}\omega^{-a\sigma i}u_{h(i)} - x_i|^2]}{\gamma^{1/2}|x_i|^2} \\
&\lesssim \frac{1}{\gamma^{1/2-j}\mu^2}\alpha\mu^2\left(1 + \mu^2 \sum_{j'<j-1} \gamma^{-(j'+1)} m_{j'}(t)\right) \\
&= \alpha\left(\gamma^{j-1/2} + \sum_{j'<j-1} \gamma^{j-j'-3/2} m_{j'}(t)\right).
\end{aligned}$$

Adding the $\alpha m_{j-1}(1)$ chance that $C \cap S^+ \neq \{\}$ in a union bound gives the result. □

Let S_t denote the state of the system at time t. By Lemma 6.1 at each time step t we have

$$\begin{aligned}
\mathbb{E}[m_1(t+1)|S_t] &\leq \alpha m_1(t) \cdot (m_0(t)) \\
\mathbb{E}[m_2(t+1)|S_t] &\leq \alpha m_2(t) \cdot (\gamma^{1/2} m_0(t) + m_1(t) + R^{-20}) \\
\mathbb{E}[m_3(t+1)|S_t] &\leq \alpha m_3(t) \cdot (\gamma^{3/2} m_0(t) + \gamma^{1/2} m_1(t) + m_2(t) + R^{-20}) \\
&\vdots \\
\mathbb{E}[m_j(t+1)|S_t] &\leq \alpha m_j(t) \cdot (\gamma^{j-3/2} m_0(t) + \ldots + \gamma^{1/2} m_{j-2}(t) + m_{j-1}(t) + R^{-20}).
\end{aligned} \tag{15}$$

Note that Lemma 6.1 in fact yields the bound without the additive term of R^{-20}. We analyze the weaker recurrence (15) in what follows, since the additional term of R^{-20} will be useful later in section 7 for handling location errors. Lemma 6.1 does not provide any guarantees on the evolution of $m_0(t)$. It will be convenient to assume that $m_0(t)$ is chosen arbitrarily from the range $[0, C]$ for a constant $C > 0$ and all $t \geq 1$. Note that the contribution of μ^2 to the rhs in Lemma 6.1 disappeared since it is dominated by the contribution of $m_0(t)$.

In what follows we first analyze a related deterministic process, and then show that the randomized process closely follows its deterministic version with high probability.

6.2 Deterministic process

Let $m_j(1) \in [0, C]$ for a constant $C > 0$, and let $m_0^{det}(t) \in [0, C]$ be chosen arbitrarily for every t. Further, let for each $t \geq 1$ and $j \in [1 : 1/\delta]$

$$\begin{aligned}
m_1^{det}(t+1) &= \alpha m_1^{det}(t) \cdot (m_0^{det}(t)) \\
m_2^{det}(t+1) &= \alpha m_2^{det}(t) \cdot (\gamma^{1/2} m_0^{det}(t) + m_1^{det}(t) + R^{-20}) \\
m_3^{det}(t+1) &= \alpha m_3^{det}(t) \cdot (\gamma^{3/2} m_0^{det}(t) + \gamma^{1/2} m_1^{det}(t) + m_2^{det}(t) + R^{-20}) \\
&\vdots \\
m_j^{det}(t+1) &= \alpha m_j^{det}(t) \cdot (\gamma^{j-1/2} m_0^{det}(t) + \ldots + \gamma^{1/2} m_{j-2}^{det}(t) + m_{j-1}^{det}(t) + R^{-20}).
\end{aligned} \tag{16}$$

We now analyze the evolution of solutions to (16):

Lemma 6.2. *For each $j = 1, \ldots, 1/\delta$ and $t \le j$ one has either $m_j^{det}(t) \le 2^{-2^t}$ or $m_{j-1}^{det}(t-1) = O(\gamma^{1/2})$.*

The same conclusion holds if the equations for $m_j^{det}(t)$ are modified to include a R^{-20} additive term to obtain

$$m_j^{det}(t+1) = \alpha m_j^{det}(t) \cdot (\gamma^{j-1/2} m_0^{det}(t) + \ldots + \gamma^{1/2} m_{j-2}^{det}(t) + m_{j-1}^{det}(t) + R^{-20}).$$

for $j = 1, \ldots, 1/\delta$.

Proof. Induction on j and t.

Base: $j = 1$**,** $t = 1$ Trivial by appropriate choice of α.

Inductive step:(j, t) Suppose that $m_{j'}^{det}(t') \le 2^{-2^{t'}}$ for all $j' < j$ and $t' \le j'$. Then we have

$$\begin{aligned} m_j^{det}(t+1) &\le \alpha m_j^{det}(t) \cdot (\gamma^{j-1/2} m_0^{det}(t) + \ldots + \gamma^{1/2} m_{j-2}^{det}(t) + m_{j-1}^{det}(t)) \\ &\le \alpha m_j^{det}(t) \cdot (O(\gamma^{1/2}) + m_{j-1}^{det}(t) + R^{-20}) \\ &\le \alpha m_j^{det}(t) \cdot (O(\gamma^{1/2}) + m_{j-1}^{det}(t)), \end{aligned}$$

where we used the fact that $R^{-20} = O(\gamma^{1/2})$.

Thus, if t is the first index such that $m_{j-1}^{det}(t) = O(\gamma^{1/2})$, we are done since $m_{j-1}^{det}(t)$ is non-increasing in t; Otherwise by the inductive hypothesis $m_j^{det}(t) \le 2^{-2^t}$, so

$$\begin{aligned} m_j^{det}(t+1) &\le \alpha m_j^{det}(t) \cdot (\gamma^{j-1/2} m_0^{det}(t) + \ldots + \gamma^{1/2} m_{j-2}^{det}(t) + m_{j-1}^{det}(t)) \\ &\le \alpha m_j^{det}(t) \cdot (O(\gamma^{1/2}) + m_{j-1}^{det}(t)) \\ &\le m_j^{det}(t) \cdot 2^{-2^t} \le 2^{-2^{t+1}} \end{aligned}$$

as long as α is smaller than an appropriate constant.

□

Thus, we obtain

Lemma 6.3. *One has for all $j \ge 1/(4\delta)$ and any $t \ge c \log\log R$*

$$m_j^{det}(t) \le R^{-10},$$

where $c > 0$ is a sufficiently large constant.

Proof. We use Lemma 6.2. First note that for $t \ge 1/(4\delta) \ge 10 \log_2 \log_2 R$ one has $2^{-2^t} < 2^{-2^{10 \log_2 \log_2 R}} < 2^{-\log_2^{10} R} < R^{-10}$. Thus, if the first case in Lemma 6.2 holds for $m_j(t)$, $j = t = (1/(2\delta))$, we are done. Otherwise if the second case holds, we have

$$\begin{aligned} m_j^{det}(t+1) &\le \alpha m_j^{det}(t) \cdot (\gamma^{j-1/2} m_0^{det}(t) + \ldots + \gamma^{1/2} m_{j-2}^{det}(t) + m_{j-1}^{det}(t)) \\ &\le \alpha m_j^{det}(t) \cdot (O(\gamma^{1/2}) + m_{j-1}^{det}(t)) = \alpha m_j^{det}(t) \cdot (O(\gamma^{1/2}) + m_{j-1}^{det}(t)), \end{aligned}$$

and then $m_j^{det}(t+t') = \gamma^{O(t')} = R^{O(t'/\log_2 \log_2 R)}$, and hence $m_j^{det}(t+t') \le R^{-10}$ for $t' = O(\log_2 \log_2 R)$.
□

We have proved

Lemma 6.4. *Let $\gamma = R^{-\delta}$ for some parameters $R > 1$ and $\delta = \Theta(1/\log\log R)$. Let $m_\ell(t) \in [0, C]$ be defined for some constant C, integer $\ell \in [0, 1/\delta - 1]$, and integer $t > 0$. Suppose it satisfies*

$$m_{\ell+1}(t+1) \le \alpha m_{\ell+1}(t)\left(m_\ell(t) + \sum_{i=1}^{\ell} \gamma^{i-1/2} m_{\ell-i}(t) + R^{-20}\right) \text{ for } \ell \ge 0.$$

for some sufficiently small constant α. Then there exists a universal constant c such that for all $t > c\log\log R$ and $\ell \ge 1/(4\delta)$,

$$m_\ell(t) \le R^{-10}.$$

6.3 Bound in Expectation

In this section we show similar convergence if the decay is only in expectation, and using a continuous version of the recurrence.

For all $\eta \ge 0$, define the function $f_\eta : \mathbb{C}^n \to [0, \infty)$ by

$$f_\eta(x) = \frac{1}{k}|\{i : |x^i|^2 \ge \eta\}|$$

to be roughly the "fraction" of heavy hitters that remain above η.

Lemma 6.5. *Let k, R, μ^2 be arbitrary with $\delta = \Theta(\log\log R)$ and $\gamma = R^\delta$. Consider a recursion $x \to x'$ of vectors $x \in \mathbb{C}^n$ that is repeated $N = \Theta(\log\log R)$ times as $x^0 \to x^1 \to \cdots \to x^N$, and for all $\ell \ge 0$ and all inputs x satisfies*

$$\mathbb{E}[f_\eta(x')] \lesssim \alpha f_\eta(x)\left(R^{-20} + \frac{\mu^2}{\gamma\eta} + \frac{1}{\gamma\eta}\int_0^{\gamma\eta} f_t(x)dt\right) \tag{17}$$

for some sufficiently small parameter α. Suppose that $\|x^0\|_2^2 \lesssim Rk\mu^2$ and we know for all $i \in [0, N]$ that $f_0(x^i) \lesssim 1$. Then

$$\|x^N\|_2^2 \lesssim \sqrt{R}k\mu^2$$

with probability $1 - O(\alpha N^2)$. Furthermore, with the same probability we also have for all $i \le N$ that

$$\|x^i\|_2^2 \lesssim Rk\mu^2$$
$$f_{\mu^2/\gamma}(x^i) \lesssim 1/4^i.$$

Proof. For simplicity of notation, we will prove the result about x^{N+1} rather than x^N; adjusting N gives the lemma statement.

The only properties of f that we use are (17), $f_a(x) \ge f_b(x)$ for $a \le b$, and that

$$\|x\|_2^2 = k\int_0^\infty f_\eta(x)d\eta.$$

The desired claims are made more difficult by increasing the $f_\eta(x)$. Since we know that $f_\eta(x) \le f_0(x) \le C$ for some constant C, we may set

$$f_\eta(x) = C \quad \text{for } \eta < \mu^2/\gamma$$

for all x without loss of generality.

Then the μ^2 term in (17) may be absorbed by the integral, giving for each $x \to x'$ that:

$$\text{for any } \eta \geq \mu^2/\gamma, \quad \mathbb{E}[f_\eta(x')] \lesssim \alpha f_\eta(x) \left(R^{-20} + \frac{1}{\gamma\eta} \int_0^{\gamma\eta} f_t(x) dt \right) \tag{18}$$

$$\lesssim \alpha f_\eta(x) \tag{19}$$

where the last step uses that $f_t(x) \leq C \lesssim 1$.

For $i \geq 1$ we have

$$\mathbb{E}[f_{\mu^2/\gamma}(x^i)] \leq (O(\alpha))^i f_{\mu^2/\gamma}(x^0) \lesssim \alpha/4^i.$$

for sufficiently small α, so by Markov's inequality and a union bound, with $1 - O(\alpha N)$ probability we have $f_{\mu^2/\gamma}(x^i) \leq 1/4^i$ for all $i \leq N+1$. This gives the last claim in the lemma statement.

Part 1: We know prove that $\|x^i\|_2^2 \lesssim Rk\mu^2$ for all i. We have that

$$\frac{1}{k}\|x'\|_2^2 = \int_0^\infty f_\eta(x') d\eta \lesssim \mu^2/\gamma + \int_{\mu^2/\gamma}^\infty f_\eta(x') d\eta$$

and by (19),

$$\begin{aligned}
\mathbb{E}[\int_{\mu^2/\gamma}^\infty f_\eta(x') d\eta] &= \int_{\mu^2/\gamma}^\infty \mathbb{E}[f_\eta(x')] d\eta \\
&\lesssim \alpha \int_{\mu^2/\gamma}^\infty f_\eta(x) d\eta \\
&\leq \alpha \|x\|_2^2/k.
\end{aligned}$$

Hence with probability $1 - O(\alpha N^2)$, in all N stages this is at most $\|x\|_2^2/(Nk)$ and we have

$$\frac{1}{k}\|x'\|_2^2 \lesssim \mu^2/\gamma + \|x\|_2^2/(Nk).$$

Hence for all $i \leq N$,

$$\|x^i\|_2^2 \lesssim k\mu^2/\gamma + \|x^0\|_2^2 \lesssim Rk\mu^2. \tag{20}$$

Part 2: We now prove that

$$f_{R^{1/4}\mu^2}(x^N) \lesssim R^{-10} \tag{21}$$

with the desired probability.

Define the functions $m_\ell : \mathbb{C}^n \to [0, C]$ for integer ℓ by

$$\begin{aligned}
m_0(x) &= f_0(x) \\
m_\ell(x) &= f_{\gamma^{-2\ell}\mu^2}(x) \quad \text{for } \ell > 0
\end{aligned}$$

We will show that they satisfy the recurrence in Lemma 6.4 with γ^2 replacing γ. By (18), for $\ell \geq 1$ we have

$$\mathbb{E}[m_\ell(x)] \lesssim \alpha m_\ell(x) \left(R^{-20} + \frac{\gamma^{2\ell-1}}{\mu^2} \int_0^{\gamma^{1-2\ell}\mu^2} f_t(x) dt \right)$$

and we know that

$$\begin{aligned}\int_0^{\gamma^{1-2\ell}\mu^2} f_t(x)dt &\le C\mu^2/\gamma^2 + \sum_{i=1}^{\ell-2}\int_{\mu^2\gamma^{-2i}}^{\mu^2\gamma^{-2i-2}} f_t(x)dt + \int_{\mu^2\gamma^{2-2\ell}}^{\mu^2\gamma^{1-2\ell}} f_t(x)dt \\ &\le C\mu^2/\gamma^2 + \sum_{i=1}^{\ell-2}\mu^2\gamma^{-2i-2}m_i(x) + \mu^2\gamma^{1-2\ell}m_{\ell-1}(x) \\ &= \mu^2\gamma^{1-2\ell}m_{\ell-1}(x) + \sum_{i=0}^{\ell-2}\gamma^{-2i-2}\mu^2 m_i(x)\end{aligned}$$

so

$$\begin{aligned}\mathbb{E}[m_\ell(x)] &\lesssim \alpha m_\ell(x)(R^{-20} + m_{\ell-1}(x) + \sum_{i=0}^{\ell-2}\gamma^{2\ell-2i-3}m_i(x)) \\ &= \alpha m_\ell(x)(R^{-20} + m_{\ell-1}(x) + \sum_{i=2}^{\ell}(\gamma^2)^{i-3/2}m_{\ell-i}(x))\end{aligned}$$

for $\ell \ge 1$. But for the expectation, this is precisely the recurrence of Lemma 6.4 after substituting γ^2 for γ. Since Lemma 6.4 only considers $N/\delta \lesssim N^2$ different ℓ and x^i, by Markov's inequality the recurrence will hold in all instances for a sufficiently small constant α' with probability $1 - O(\alpha N^2)$. Assume this happens. Since Lemma 6.4 is applied with $\gamma \to \gamma^2$, $\delta \to \delta/2$, this implies

$$m_{1/8\delta}(x^N) \lesssim R^{-10}.$$

This gives (21), because

$$f_{R^{1/4}\mu^2}(x^N) = f_{\gamma^{2\cdot 1/(8\delta)}\mu^2}(x^N) = m_{1/8\delta}(x^N) \lesssim R^{-10}.$$

Part 3: We now prove that $\|x^{N+1}\|_2^2 \lesssim \sqrt{R}k\mu^2$ with $1 - O(\alpha)$ probability conditioned on the above.

We have

$$\frac{1}{k}\|x^{N+1}\|_2^2 = \int_0^\infty f_\eta(x^{N+1})d\eta \lesssim R^{1/4}\mu^2 + \int_{R^{1/4}\mu^2}^\infty f_\eta(x^{N+1})d\eta$$

Define V to be the latter term. We have by (18) and (21) that

$$\begin{aligned}\mathbb{E}[V] &= \int_{R^{1/4}\mu^2}^\infty \mathbb{E}[f_\eta(x^{N+1})]d\eta \\ &\lesssim \int_{R^{1/4}\mu^2}^\infty \alpha f_\eta(x^N)(R^{-20} + \frac{1}{\gamma\eta}\int_0^{\gamma\eta} f_t(x^N)dt)d\eta \\ &= \int_{R^{1/4}\mu^2}^\infty \alpha f_\eta(x^N)(R^{-20} + \frac{1}{\gamma\eta}(\int_0^{R^{1/4}\mu^2} f_t(x^N)dt + \int_{R^{1/4}\mu^2}^{\gamma\eta} f_t(x^N)dt))d\eta \\ &\lesssim \int_{R^{1/4}\mu^2}^\infty \alpha f_\eta(x^N)(R^{-20} + \frac{1}{\gamma\eta}(R^{1/4}\mu^2 + \gamma\eta R^{-10}))d\eta \\ &\lesssim \alpha R^{-10}\|x^N\|_2^2/k + \int_{R^{1/4}\mu^2}^\infty \alpha f_\eta(x^N)\frac{R^{1/4}\mu^2}{\gamma\eta}d\eta\end{aligned}$$

In the latter term, for fixed $\int_0^\infty f_\eta(x^N)d\eta = \|x^N\|_2^2/k$ this is maximized when the mass of f_η is pushed towards smaller η. Hence

$$\begin{aligned}\int_{R^{1/4}\mu^2}^{\infty} \alpha f_\eta(x^N)\frac{R^{1/4}\mu^2}{\gamma\eta}d\eta &\le \int_{R^{1/4}\mu^2}^{R^{1/4}\mu^2+\|x^N\|_2^2/k} \alpha\cdot C\cdot\frac{R^{1/4}\mu^2}{\gamma\eta}d\eta \\ &= C\alpha R^{1/4}\mu^2\gamma^{-1}\log(1+\frac{\|x^N\|_2^2/k}{R^{1/4}\mu^2}) \\ &\le C\alpha R^{1/4+\delta}\mu^2\log R \\ &\lesssim \alpha\sqrt{R}\mu^2.\end{aligned}$$

by (20). But then $\mathbb{E}[V]\lesssim\alpha\sqrt{R}\mu^2$, so with $1-O(\alpha)$ probability $V\lesssim\sqrt{R}\mu^2$ and

$$\|x^{N+1}\|_2^2\lesssim\sqrt{R}k\mu^2$$

as desired. □

7 Reducing SNR: general analysis

Recall our goal:

Lemma 7.11. *For $x^*,\chi\in\mathbb{C}^n$ define $x=x^*-\chi$ and*

$$\xi^2=\mathrm{Err}_k^2(x)+\frac{\|x\|_2^2}{R}+\frac{\|x^*\|_2^2}{R^*n^{10}}.$$

Suppose $\|\chi\|_0\le k$. Then REDUCESNR$(\widehat{x}^*,\chi,k,R,p)$ *returns χ' such that*

$$\|x-\chi'\|_2^2\lesssim\sqrt{R}\xi^2$$

with probability $1-p$, using $O(\frac{1}{p^2}k\log(Rn/k)(\log\log(Rn/k))^c)$ measurements and a $\log(Rn)$ factor more time.

We will show that each inner loop of REDUCESNR satisfies some nice properties (simiar to those of Lemma 6.4) that cause the residual to reduce from signal-to-noise ratio R to $\sqrt{R}$. As in REDUCESNR and 2.2, we define

- $B=k/\alpha$ to be the size of each hash table, where $\alpha=O(1/\log\log^c n)$
- $T=O(1)$ to be the number of hashings done in each ESTIMATEVALUES
- $\xi^2=\mathrm{Err}_k^2(x^*-\chi)+\frac{\|x^*-\chi\|_2^2}{R}+\|x^*\|_2^2/(R^*n^{10})$.
- $\mu^2=\xi^2/k\ge\frac{1}{k}(\mathrm{Err}_k^2(x^*-\chi)+\|x^*-\chi\|_2^2/R)$ to be the "noise level."
- $\delta=1/(40\log_2\log_2 R)$.
- $\gamma=R^{-\delta}$ to be the "contrast" our LOCATESIGNAL requires.

In round t of the inner loop, we define the following variables:

- $\chi^{(t)}\in\mathbb{C}^n$: the estimate of x^* recovered so far.
- $x=x^*-\chi^{(t)}\in\mathbb{C}^n$: The vector we want to recover.

- $k' = k_t = \Theta(k4^{-t})$: The number of coordinates to update this round.
- $L \subset [n]$: Indices located by LOCATESIGNAL (with $|L| \leq B$)
- $\tilde{x}^{(t)} \in \mathbb{C}^n$ for $t \in T$: The estimations of x in each inner loop of ESTIMATEVALUES.
- $\tilde{x} \in \mathbb{C}^n$: $\tilde{x}_i = \text{median}_t\, \tilde{x}_i^{(t)}$ is the estimation of x that would result from ESTIMATEVALUES (although the algorithm only computes $\tilde{x}_L$).
- $S \subseteq L$ contains the largest $k'/4$ coordinates of x_L.
- $L' \subseteq L$: The indices of the largest k' coordinates of $\tilde{x}_L$

In the algorithm REDUCESNR, the inner loop replaces x with $x - \tilde{x}_{L'}$. This is then repeated $N = O(\log\log R)$ times. We say that this is a "recurrence" $x \to x - \tilde{x}_{L'}$, and will prove that the final result x^N has $\|x^N\|_2^2 \lesssim \sqrt{R}\xi^2$.

We will split our analysis of REDUCESNR into stages, where the earlier stages analyze the algorithm with the inner loop giving a simpler recurrence. In subsequent sections, we will consider the following different recurrences:

1. $x \to x - x_S$
2. $x \to x - \tilde{x}_S$
3. $x \to x - x_{L'}$
4. $x \to x - \tilde{x}_{L'}$

and show that each would reduce the noise level after $O(\log\log R)$ repetitions.

7.1 Splittings and Admissibility

We introduce the notion of *splittings*. These allow us to show that the error introduced by the estimation phase is of the same order as the error from coordinates that are not well hashed. Since that level of error is tolerable according to Section 6, we get that the total error is also tolerable.

Definition 7.1. *For $x \in \mathbb{C}^n$, (z, ν) is a* splitting *of x if, for all $i \in [n]$, z^i is a vector and $\nu_i \in \mathbb{R}$ with*

$$\|z^i\|_2^2 + \nu_i^2 \geq |x_i|^2.$$

Analogously to the previous section, we can measure the number of elements of z above any value $\eta \geq 0$:

$$f_\eta(z) = \frac{1}{k}|\{(i,j) : |z_j^i|^2 \geq \eta\}|$$

We will want to deal with "nice" splittings that satisfy additional properties, as described below.

Definition 7.2. *We say (z, ν) is a* concise *splitting of x if (z, ν) is a splitting of x and also*

$$\begin{aligned}
\|z^i\|_2^2 + \nu_i^2 &= |x_i|^2 \text{ for all } i \\
f_0(z) &\lesssim 1 \\
\|\nu\|_2^2 &\lesssim k\mu^2 \\
f_{\mu^2/\gamma}(z) &\leq k'/(4k) \\
\sum_i \|z^i\|_2^2 &\lesssim R^2 k\mu^2
\end{aligned}$$

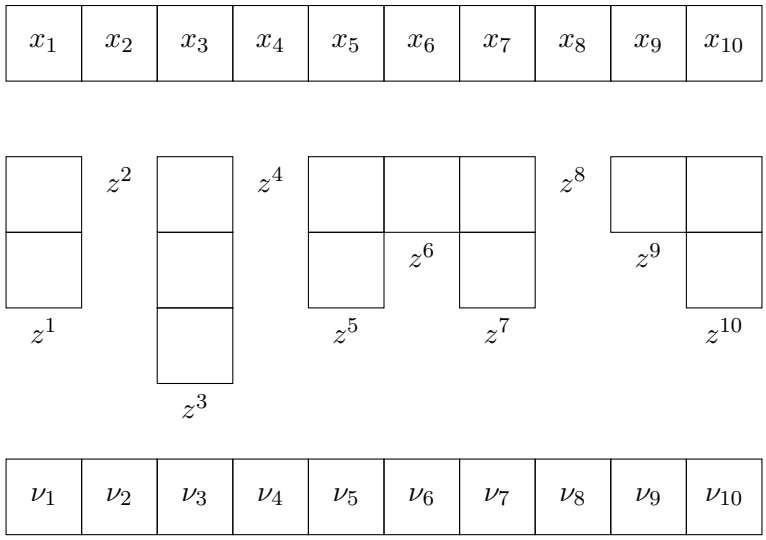

Figure 2: A representation of a splitting of x. In each column, $|x_i|^2 \leq \|z^i\|_2^2 + \nu_i^2$.

For various recurrences $x \to x'$ and any concise splitting (z, ν) of x, we would like to find methods of assigning splittings (z', ν') to x' that satisfy some nice properties. In particular, we will want decay as in Lemma 6.5:

$$\mathbb{E}[f_\eta(z')] \lesssim \alpha f_\eta(z) \left(R^{-20} + \frac{\mu^2}{\gamma\eta} + \frac{1}{\gamma\eta} \int_0^{\gamma\eta} f_t(z) dt \right) \tag{D}$$

and we will want relatively slow growth in the size of the splitting:

$$\begin{aligned} \mathbb{E}[\max\left(0, (\sum_i \|(z')^i\|_0) - (\sum_i \|z^i\|_0)\right)] &\lesssim \sqrt{\alpha k' k} \\ \mathbb{E}[\max\left(0, \sum_i (\nu_i')^2 - \sum_i \nu_i^2.\right)] &\lesssim \sqrt{\alpha k' k}\mu^2 \end{aligned} \tag{G}$$

For some recurrences, we will get the stronger condition:

$$\begin{aligned} \mathbb{E}[\sum_i \|(z')^i\|_0] &\lesssim \sqrt{\alpha k' k} \\ \mathbb{E}[\sum_i (\nu_i')^2] &\lesssim \sqrt{\alpha k' k}\mu^2. \end{aligned} \tag{G'}$$

Definition 7.3. *A recurrence $x \to x'$ is*

- admissible *if for any concise splitting (z, ν) of x, we can assign splittings (z', ν') to x' that satisfy* (D) *and* (G)*.*
- fully admissible *if (z', ν') can also satisfy* (G')*.*

Note that we require the input (z, ν) to be a concise splitting, but the result (z', ν') may not be concise. Analyzing repeated applications of (D) gives the following lemma:

Lemma 7.4. *Suppose $x \to x'$ is admissible. Then consider $r = \log\log R$ repetitions of the recurrence, i.e. $x^0 \to x^1 \to \cdots \to x^r$, where the jth round is run with $k' = k_j = k/4^j$ and x^0 has a concise splitting and $\|x^0\|_2^2 \lesssim Rk\mu^2$. Then for any parameter p, as long as $\alpha = \Theta(p^2/(\log\log R)^2)$ is sufficiently small, we get*

$$\|x^r\|_2^2 \lesssim \sqrt{R}k\mu^2$$

with $1 - p$ probability.

Proof. Because $x \to x'$ is admissible, there is a corresponding recurrence

$$(z, \nu) \to (z', \nu')$$

of splittings of x and x' that satisfies (D) and (G) whenever (z, ν) is concise. For this analysis, we will suppose that it satisfies (D) and (G) unconditionally, and bound the probability that (z, ν) is ever not concise.

The conditions (D) and (G) are only made more true by decreasing z' and ν' in absolute value, so we may also assume the (z', ν') resulting from the recurrence satisfies the first requirement of concise splittings,

$$\|(z')^i\|_2^2 + (\nu_i')^2 = |x_i|^2.$$

At each stage, with probability $1 - O(\sqrt{\alpha})$ we have by Markov's inequality and a union bound that

$$\begin{aligned} \max\left(0, (\sum_i \|(z')^i\|_0) - (\sum_i \|z^i\|_0)\right) &\le \sqrt{k'k} \\ \max\left(0, \sum_i (\nu_i')^2 - \sum_i \nu_i^2.\right) &\le \sqrt{k'k}\mu^2 \end{aligned} \tag{22}$$

Hence with $1 - O((\log\log R)\sqrt{\alpha}) > 1 - p/2$ probability, equation set (22) holds for all r stages. Assume this happens.

Then at any stage j, the resulting (z', ν') has $f_0(z') = \frac{1}{k}\sum_i \|(z')^i\|_0 \le \frac{1}{k}(k + \sum_{t \le j}\sqrt{k_t k}) \le 3$ and $\|\nu'\|_2^2 \le k\mu^2 + \sum_{t \le j}\sqrt{k_t k}\mu^2 \le 3k\mu^2$. Therefore the second and third requirements for conciseness are satisfied in every stage.

Now, we apply Lemma 6.5 to observe that with $1 - O(\alpha N^2) > 1 - p/2$ probability, the remaining two requirements for conciseness are satisfied in all stages and the final splitting (z, ν) of x^r satisfies

$$\sum_i \|z^i\|_2^2 \lesssim \sqrt{R}k\mu^2.$$

Therefore with probability $1 - p$, our supposition of conciseness is correct in all stages and the final x^r satisfies

$$\|x^r\|_2^2 \le \sum_i \|z^i\|_2^2 + \nu_i^2 \lesssim (\sqrt{R} + 3)k\mu^2 \lesssim \sqrt{R}k\mu^2$$

which is our result. □

Given that admissibility is a sufficient condition, we construct tools to prove that recurrences are admissible.

Lemma 7.5. *If $x \to x'$ is admissible, $x \to x^\#$ is fully admissible, and $x'_{\mathrm{supp}(x^\#)}$ is identically zero then $x \to x' + x^\#$ is admissible.*

Proof. For any splitting (z, ν) of x, we have splittings (z', ν') and $(z^\#, \nu^\#)$ of x' and $x^\#$. We would like to combine them for a splitting of $x' + x^\#$.

Let $A = \text{supp}(x^\#)$. For $i \notin A$, we use $((z')^i, \nu'_i)$. For $i \in A$, we use $((z^\#)^i, \nu_i^\#)$. This is a valid splitting of $x' + x^\#$.

By linearity it satisfies (D) and (G) with a minor loss in the constants. □

Lemma 7.6. *If $x \to x'$ and $x \to x^\#$ are both fully admissible, then $x \to x' + x^\#$ is fully admissible.*

Proof. For any splitting (z, ν) of x, we have splittings (z', ν') and $(z^\#, \nu^\#)$ of x' and $x^\#$. We would like to combine them for a splitting of $x' + x^\#$.

For each coordinate i, let $u = (z')^i$ and $v = (z^\#)^i$, and $a = |x'_i + x_i^\#|$. We will find a vector w and scalar g such that

$$\begin{aligned} \|w\|_2^2 + g^2 &\geq a^2 \\ \|w\|_0 &\lesssim \|u\|_0 + \|v\|_0 \\ g^2 &\lesssim (\nu'_i)^2 + (\nu_i^\#)^2 \\ |\{i \mid w_i \geq \eta\}| &\lesssim |\{i \mid u_i \geq \eta\}| + |\{i \mid v_i \geq \eta\}|. \end{aligned}$$

for all thresholds η. This will only lose a constant factor in (G') and (D). In particular, we set w to be the concatenation of two copies of u and two copies of v, and $g^2 = 2(\nu'_i)^2 + 2(\nu_i^\#)^2$. Then

$$\|w\|_2^2 + g^2 = 2(\|u\|_2^2 + (\nu'_i)^2) + 2(\|v\|_2^2 + (\nu_i^\#)^2) \geq 2|x'_i|^2 + 2|x_i^\#|^2 \geq a^2,$$

so (w, g) is a valid splitting for each coordinate, giving us that $x' + x^\#$ is fully admissible. □

7.2 Recurrence $x \to x - x_S$

Lemma 7.7. *Let S contain the largest $k'/4$ coordinates of L. Then $x \to x - x_S$ is admissible.*

Proof. Consider any concise splitting (z, ν) of x. Let $S' = \{i \in L : \|z^i\|_\infty^2 \geq \mu^2 \gamma^{-1}\}$.

We have $|S'| \leq k f_{\mu^2/\gamma}(z) \leq k'/4$ because (z, ν) is concise. Since $x - x_{S'}$ can be permuted to be coordinate-wise dominated by $x - x_S$, it suffices to split $x - x_{S'}$.

For $i \in S'$, we set $(z')^i = \{\}$ and $\nu'_i = 0$; for $i \notin S'$, we set $((z')^i, \nu'_i) = (z^i, \nu_i)$. We must only show (D) holds, because (G) is trivial (the growth is zero). That is, we must show that if $|z_j^i|^2 \geq \eta$ then

$$\Pr[i \notin S'] \lesssim \alpha \left(R^{-20} + \frac{\mu^2}{\gamma\eta} + \frac{1}{\gamma\eta} \int_0^{\gamma\eta} f_t(x) dt \right) =: M. \tag{23}$$

Let M denote the right hand side of (23). For such an i, $|x_i|^2 \geq |z_j^i|^2 \geq \mu^2\gamma^{-1}$, and

$$\Pr[i \notin S'] = \Pr[i \notin L] \leq \Pr[i \text{ not well-hashed}] + \Pr[i \notin L | i \text{ well-hashed}]$$

Define $H = \{i : \|z^i\|_\infty^2 \geq \gamma\eta\}$. Then from Lemma 3.3 we get a subset $C \subset [n]$ and variable w so that i is well-hashed if

$$w^2 + \|x_C\|_2^2 \leq c\gamma^{1/2}|x_i|^2$$

for some constant c, which is implied by $w^2 + \|x_C\|_2^2 \leq \gamma\eta$. We have that

$$\Pr[H \cap C \neq \{\}] \lesssim |H|/B \leq k f_{\gamma\eta}(z)/B = \alpha f_{\gamma\eta}(z)$$

and that

$$\begin{aligned}\mathbb{E}[w^2 + \|x_{C\setminus H}\|_2^2] &\lesssim \frac{\|x\|_2^2}{R^2 B} + \frac{\|x^*\|_2^2}{R^* n^{11}} + \|x_{\overline{H}}\|_2^2/B \\ &\lesssim \alpha\mu^2 + \|x_{\overline{H}}\|_2^2/B\end{aligned}$$

by the definition of μ^2. We know that

$$\|x_{\overline{H}}\|_2^2 \leq \sum_{(i,j):|z_j^i|^2 \leq \gamma\eta} |z_j^i|^2 = k\int_0^{\gamma\eta} (f_t(z) - f_{\gamma\eta}(z))dt.$$

Therefore by Markov's inequality,

$$\begin{aligned}\Pr[i \text{ not well-hashed}] &\leq \Pr[C \cap H \neq \{\}] + \Pr[w^2 + \|x_{C\setminus H}\|_2^2 \geq \gamma\eta] \\ &\lesssim \alpha f_{\gamma\eta}(z) + \frac{1}{\gamma\eta}(\alpha\mu^2 + \alpha\int_0^{\gamma\eta} (f_t(z) - f_{\gamma\eta}(z))dt) \\ &= \frac{1}{\gamma\eta}(\alpha\mu^2 + \alpha\int_0^{\gamma\eta} f_t(z)dt) < M.\end{aligned}$$

Next, by Lemma 10.2, since we call LOCATESIGNAL with failure probability αR^{-20}, we have

$$\Pr[i \notin L | i \text{ well-hashed}] \lesssim \alpha R^{-20} < M.$$

giving $\Pr[i \notin S'] \lesssim M$ for each i, as desired. □

7.3 Recurrence $x \to x - \tilde{x}_S$

Lemma 7.8. *Let L be independent of the estimation phase with $|L| \leq B$, and $A \subseteq L$ be possibly dependent on the estimation phase with $|A| \lesssim k'$. Then $x \to x_A - \tilde{x}_A$ is fully admissible.*

Proof. Let (z, ν) be a concise splitting of x. For $i \in L$, we have

$$|\tilde{x}_i - x_i|^2 = |\operatorname*{median}_t \tilde{x}_i^{(t)} - x_i|^2 \leq 2\operatorname*{median}_t |\tilde{x}_i^{(t)} - x_i|^2 \tag{24}$$

because we take the median in real and imaginary components separately. We have by $\tilde{x}_i^{(t)} = G_{o_i(i)}^{-1}\omega^{-a\sigma i}u_{h(i)}$ and Lemma 3.3 that

$$\mathop{\mathbb{E}}_a[|\tilde{x}_i^{(t)} - x_i|^2] \lesssim w_i^2 + \|x_{C_i^t}\|_2^2$$

for some C with $\Pr[j \in C] \lesssim 1/B$ for all j, and some w with

$$\mathbb{E}[w_i^2] \lesssim \frac{\|x\|_2^2}{R^2 B} + \|x^*\|_2^2/(R^* n^{11}) \lesssim \alpha\mu^2, \tag{25}$$

where the last step uses that $\|x\|_2^2 \lesssim R^2 k\mu^2$ because a concise splitting (z, ν) of x exists. Then

$$\mathop{\mathbb{E}}_a[|\tilde{x}_i^{(t)} - x_i|^2] \lesssim w_i^2 + \sum_{j \in C_i^t} \|z^j\|_2^2 + \nu_j^2.$$

Define

$$(y_i^t)^2 := w_i^2 + \sum_{j \in C_i^t} \|z^j\|_2^2 + \nu_j^2$$
$$\tau_i^t := \lceil 2|\tilde{x}_i^{(t)} - x_i|^2/(y_i^t)^2 \rceil$$

so

$$\mathop{\mathbb{E}}_{a}[\tau_i^t] \lesssim 1 \tag{26}$$

even after conditioning on the hash function (σ, b).

For any $t \in [T]$ and $i \in L$, let $U^{(t),i}$ be the concatenation of τ_i^t copies of z^j for each $j \in C_i^t$ and $\nu_i^{(t)} = \sqrt{\tau_i^t}((w_i^t)^2 + \sum_{j \in C_i^t} \nu_j^2)$. Then we have that

$$\|U^{(t),i}\|_2^2 + (\nu_i^{(t)})^2 \geq \tau_i^t (y_i^t)^2 \geq 2|\tilde{x}_i^{(t)} - x_i|^2$$

and so by (24), for at least $1 + \lfloor T/2 \rfloor$ different $t \in [T]$ we have

$$|\tilde{x}_i - x_i| \leq \|U^{(t),i}\|_2^2 + (\nu_i^{(t)})^2. \tag{27}$$

For each $i \in A$, our $(\tilde{z}^i, \tilde{\nu}_i)$ will equal $(U^{(t^*),i}, \nu_i^{(t^*)})$ for a t^* satisfying (27) as well as

$$\begin{aligned} \|\tilde{z}^i\|_\infty &\leq \mathop{\mathrm{quant}}_{t}^{1/6} \|U^{(t),i}\|_\infty \\ \|\tilde{z}^i\|_0 &\leq \mathop{\mathrm{quant}}_{t}^{1/6} \|U^{(t),i}\|_0 \\ \tilde{\nu}_i^2 &\leq \mathop{\mathrm{quant}}_{t}^{1/6} (\nu_i^{(t)})^2 \end{aligned} \tag{28}$$

where $\mathrm{quant}^{1/6}$ is the "quantile" defined in Section 9.1. This is always possible, because the number of t excluded by these additional conditions is at most $3\lfloor T/6 \rfloor \leq \lfloor T/2 \rfloor$. Choosing such a t^* for each i gives us a splitting $(\tilde{z}, \tilde{\nu})$ of $x_A - \tilde{x}_A$.

To show (D), for any $i \in L$ and threshold η define

$$\begin{aligned} m &= |\{(\ell, j) : |z_j^\ell| \geq \eta\}| \\ \tilde{m} &= |\{(\ell, j) : |\tilde{z}_j^\ell| \geq \eta\}| \\ m_t^i &= |\{j : U_j^{(t),i} \geq \eta\}| \end{aligned}$$

We bound $\mathbb{E}[m_{t^*}^i]$ using Lemma 9.3. Since $\Pr[j \in C_i^t] \lesssim 1/B$ and $\mathbb{E}[\tau_i^t] \lesssim 1$ after conditioning on (σ, b) and so fixing C_i^t, for fixed i and t we have

$$\begin{aligned} \mathbb{E}[m_t^i] &= \mathbb{E}[(|\{(\ell, j) : |z_j^\ell| \geq \eta, \ell \in C_i^t\}|)\tau_i^t] \\ &= \mathbb{E}[(|\{(\ell, j) : |z_j^\ell| \geq \eta, \ell \in C_i^t\}|)\, \mathbb{E}[\tau_i^t \mid (\sigma, b)]] \\ &\lesssim \mathbb{E}[|\{(\ell, j) : |z_j^\ell| \geq \eta, \ell \in C_i^t\}|] \\ &= \sum_{(\ell,j):|z_j^\ell| \geq \eta} \Pr[\ell \in C_i^t] \\ &\lesssim \sum_{(\ell,j):|z_j^\ell| \geq \eta} 1/B \\ &= m/B \end{aligned}$$

We also have $m^i_{t^*} = 0$ if $\text{quant}^{1/6}{}_t\, m^i_t = 0$ and $m^i_{t^*} \le \sum_t m^i_t$ always; hence for each fixed index $i \in L$, by Lemma 9.3

$$\mathbb{E}[m^i_{t^*}] \lesssim (m/B)^{T/6}.$$

But then for $T \ge 12$ we have

$$\mathbb{E}[\tilde{m}] = \mathbb{E}[\sum_{i\in A} m^i_{t^*}] \le \sum_{i\in L} \mathbb{E}[m^i_{t^*}] \lesssim B(m/B)^2 = m^2/B$$
$$\mathbb{E}[\tilde{m}/k] \lesssim \alpha(m/k)^2$$

which says that

$$\mathbb{E}[f_\eta(z')] \lesssim \alpha(f_\eta(z))^2 \le \alpha f_\eta(\frac{1}{\gamma\eta}\int_0^{\gamma\eta} f_t(z)dt)$$

for each η, implying (D).

We now show (G'). For any nonnegative random variable $X_{i,t}$ (which will be either $\|U^{(t),i}\|_0$ or $(\nu^{(t)}_i)^2$) and $Y_i \le \text{quant}^{1/6}{}_t\, X_{i,t}$ (which will be $\|\tilde{z}^i\|_0$ or $\tilde{\nu}^2_i$), for sufficiently large constant T we have by Lemma 9.4 with $\delta = 1/2$ that

$$\mathbb{E}[\max_{|A'|\lesssim k'} \sum_{i\in A'} Y_i] \lesssim \sqrt{k'B}\max_{i,t} \mathbb{E}[X_{i,t}]$$

Plugging in that, for each fixed i, by (26) and conciseness we have

$$\mathop{\mathbb{E}}_{\sigma,b,a}[\|U^{(t),i}\|_0] = \mathop{\mathbb{E}}_{\sigma,b,a}\left[\sum_{j\in C^t_i} \|z^j\|_0 \tau^t_i\right] \le \mathop{\mathbb{E}}_{\sigma,b}\left[\sum_{j\in C^t_i} \|z^j\|_0 \max_{\sigma,b} \mathbb{E}[\tau^t_i \mid (\sigma,b)]\right] \lesssim \mathop{\mathbb{E}}_{\sigma,b}\left[\sum_{j\in C^t_i} \|z^j\|_0\right] \lesssim k/B = \alpha$$

gives

$$\mathbb{E}[\max_{|A'|\lesssim k'} \sum_{i\in A'} \|\tilde{z}^i\|_0] \lesssim \sqrt{k'B}\alpha = \sqrt{\alpha k' k},$$

which is the first part of (G'). Similarly, plugging in

$$\mathbb{E}[(\nu^{(t)}_i)^2] \lesssim (\mathop{\mathbb{E}}_{\sigma,b}[w^2_i] + \|\nu\|^2_2/B)\max_{\sigma,b} \mathbb{E}[\tau^t_i|(\sigma,b)] \lesssim \alpha\mu^2$$

gives

$$\mathbb{E}[\max_{|A'|\lesssim k'} \sum_{i\in A'} \tilde{\nu}^2_i] \lesssim \sqrt{k'B}\alpha\mu^2 = \sqrt{\alpha k' k}\mu^2,$$

which is the second part.

Therefore $(\tilde{z}, \tilde{\nu})$ is a splitting of $x_A - \tilde{x}_A$ that satisfies (D) and (G'), so $x \to x_A - \tilde{x}_A$ is fully admissible. □

7.4 Recurrence $x \to x - \tilde{x}_{L'}$

The following lemma has a similar proof structure to Lemma 9.1.

Lemma 7.9. *The recurrence $x \to x - x_{L'}$ is admissible.*

Proof. Recall the set S from Lemma 7.7, which has $|S| = k'/4 \le |L'|/4$ and for which $x - x_S$ is admissible. Let $A = L' \setminus S$ and $B = S \setminus L'$. We have $|A| \ge 4|B|$. Furthermore $\min_{i\in A} |\tilde{x}_i| \ge \max_{i\in B} |\tilde{x}_i|$ because ESTIMATEVALUES chose A over B.

By Lemma 7.7, $x \to x - x_S$ is admissible. Let $y = (x - \tilde{x})_A + 2(\tilde{x} - x)_B$. Using Lemma 7.8, $x \to y$ is admissible. Hence for every splitting (z, ν) of x there are splittings (z^S, ν^S) of $x - x_S$ that satisfies (D) and (G) and (z^{AB}, ν^{AB}) of y that satisfies (D) and (G').

For $i \notin A \cup B$, we set $((z')^i, \nu'_i) = ((z^S)^i, \nu^S_i)$. For $i \in A$, we set $((z')^i, \nu'_i) = (\{\}, 0)$. Finally, we want to fill $((z')^i, \nu'_i)$ for $i \in B$. To do this, pair up each $i \in B$ with a disjoint set P_i of four elements in A. We know that

$$\begin{aligned} 2|\tilde{x}_i| &\le \|\tilde{x}_{P_i}\|_2^2 \\ |2x_i + y_i| &\le \|x_{P_i} + y_{P_i}\|_2 \\ 2|x_i| &\le |y_i| + \|x_{P_i}\|_2 + \|y_{P_i}\|_2 \\ 4|x_i|^2 &\le 3(|y_i|^2 + \|x_{P_i}\|_2^2 + \|y_{P_i}\|_2^2) \\ |x_i|^2 &\le |y_i|^2 + \|x_{P_i}\|_2^2 + \|y_{P_i}\|_2^2 \end{aligned} \tag{29}$$

Set $(z')^i$ to the concatenation of $(z^{AB})^i$ and, for all $j \in P_i$, $(z^S)^j$ and $(z^{AB})^j$. Similarly, set $(\nu'_i)^2 = (\nu_i^{AB})^2 + \sum_{j\in P_i} (\nu_j^S)^2 + (\nu_j^{AB})^2$. By (29), this makes (z', ν') be a valid splitting of $x - x_{L'}$.

Then each element of z^S, z^{AB}, ν^S, and ν^{AB} appears exactly once in (z', ν'); hence (z', ν') satisfies (D) and (G) so $x \to x - x_{L'}$ is admissible. □

Lemma 7.10. *The recurrence $x \to x - \tilde{x}_{L'}$ is admissible.*

Proof. We have

$$x - \tilde{x}_{L'} = (x - x_{L'}) + (x_{L'} - \tilde{x}_{L'}).$$

The first term is admissible by Lemma 7.9 and zero over L'. The second is fully admissible by Lemma 7.8, with support inside L'. Hence $x \to x - \tilde{x}_{L'}$ is admissible by Lemma 7.5. □

Lemma 7.11. *For $x^*, \chi \in \mathbb{C}^n$ define $x = x^* - \chi$ and*

$$\xi^2 = \mathrm{Err}_k^2(x) + \frac{\|x\|_2^2}{R} + \frac{\|x^*\|_2^2}{R^* n^{10}}.$$

Suppose $\|\chi\|_0 \le k$. Then REDUCESNR$(\widehat{x}^*, \chi, k, R, p)$ *returns χ' such that*

$$\|x - \chi'\|_2^2 \lesssim \sqrt{R}\xi^2$$

with probability $1-p$, using $O(\frac{1}{p^2} k \log(Rn/k)(\log\log(Rn/k))^c)$ measurements and a $\log(Rn)$ factor more time.

Proof. The following is a concise splitting (z, ν) of $x = x^* - \chi^{(0)}$: place the largest k coordinates of x into z, and the rest into ν. By Lemma 7.10, $x^* - \chi^{(i)} \to x^* - \chi^{(i+1)}$ is admissible. Therefore, by Lemma 7.4, χ^N satisfies

$$\|x^* - \chi^{(N)}\|_2^2 \lesssim \sqrt{R}\xi^2.$$

as desired for correctness.

In each of $O(\log\log R)$ rounds we call LOCATESIGNAL and ESTIMATEVALUES with $B = k/\alpha = O(k(\log\log R)^2/p^2)$ and failure probability R^{-20}. The sampling complexity of each LOCATESIGNAL is

$$O(B\log(Rn/B)\log\log R\log\log(n/B)\max(1,\log_R(1/p))) \lesssim \frac{1}{p^2}k\log(Rn/B)(\log\log(Rn/B))^4$$

by Lemma 10.2. The complexity of ESTIMATEVALUES is bounded by $O(B\log R) = O(k\log R)$ since we perform $O(1)$ bucketings using a filter with contrast R. The overall sampling complexity over $O(\log\log R)$ rounds is hence bounded by

$$O\left(\frac{1}{p^2}k\log(Rn/B)(\log\log(Rn/B))^c\right)$$

for a constant $c > 0$. □

8 Final Result

Repeating Lemma 7.11 $\log\log R$ times and applying Lemma 5.1 gives the result:

Theorem 8.1. *Let $x \in \mathbb{C}^n$ satisfy $\|x\|_2^2 \le R\,\mathrm{Err}_k^2(x)$. Then* SPARSEFFT$(\widehat{x}, k, R, p)$ *returns a χ' such that*

$$\|x - \chi'\|_2^2 \le (1+\epsilon)\,\mathrm{Err}_k^2(x) + \|x\|_2^2/(R^*n^{10})$$

with probability $1-p$ and using $O(\frac{1}{p^2\epsilon}k\log(Rn/k)(\log\log(Rn/k))^c\log(1/\epsilon))$ measurements and a $\log(Rn)$ factor more time.

Proof. During this proof, we define $x^* := x$.

The algorithm performs $r = O(\log\log R)$ rounds of REDUCESNR. We may assume that all the calls succeed, as happens with $1 - p/2$ probability. We will show that at each stage i of the algorithm,

$$\|x^* - \chi^{(i)}\|_2^2 \le R_i\xi^2$$

for $\xi^2 = \mathrm{Err}_k^2(x^*) + \|x^*\|_2^2/(R^*n^{10})$. This is true by assumption at $i = 0$, and by Lemma 7.11, in each iteration REDUCESNR causes

$$\begin{aligned}\|x^* - \chi^{(i)} - \chi'\|_2^2 &\le c\sqrt{R_i}(\mathrm{Err}_{3k}^2(x^* - \chi^{(i)}) + \xi^2)\\ &\le c\sqrt{R_i}(\mathrm{Err}_k^2(x^*) + \xi^2)\\ &\le 2c\sqrt{R_i}\xi^2\end{aligned}$$

for some constant c. By Lemma 9.1,

$$\begin{aligned}\|x^* - \text{Sparsify}(\chi^{(i)} + \chi', 2k)\|_2^2 &\le \mathrm{Err}_k^2(x^*) + 4\|x^* - \chi^{(i)} - \chi'\|_2^2\\ &\le (1 + 8c\sqrt{R_i})\xi^2\\ &\le R_{i+1}\xi^2\end{aligned}$$

for sufficient constant in the recurrence for R. This proves the induction.

For some $r = O(\log\log R)$, we have $R_r = O(1)$ and so

$$\|x^* - \chi^{(r)}\|_2^2 \lesssim \xi^2.$$

Then Lemma 5.1 shows that the χ' given by RECOVERATCONSTANTSNR satisfies

$$\begin{aligned}\|x^* - \chi^{(r)} - \chi'\|_2^2 &\le \mathrm{Err}_{3k}^2(x^* - \chi^{(r)}) + \epsilon\|x - \chi^{(r)}\|_2^2 + \|x^*\|_2^2/n^{10} \\ &\le \mathrm{Err}_k^2(x^*) + O(\epsilon\xi^2) \\ &\le (1 + O(\epsilon))\,\mathrm{Err}_k^2(x^*) + \|x^*\|_2^2/n^9\end{aligned}$$

which is the desired bound after rescaling ϵ.

□

9 Utility Lemmas

This section proves a few standalone technical lemmas.

Lemma 9.1. *Let $x, z \in \mathbb{C}^n$ and $k \le n$. Let S contain the largest k terms of x and T contain the largest $2k$ terms of z. Then*

$$\|x - z_T\|_2^2 \le \|x - x_S\|_2^2 + 4\|(x - z)_{S\cup T}\|_2^2.$$

Proof. Note that each term in $\overline{S \cup T}$ and T appears exactly once on each side. Hence it suffices to show that

$$\|x_{S\setminus T}\|_2^2 \le \|x_{T\setminus S}\|_2^2 + 4\|(x - z)_{S\setminus T}\|_2^2 + 3\|(x - z)_T\|_2^2.$$

Consider any $i \in S \setminus T$ and $j \in T \setminus S$. Then $|z_j| \ge |z_i|$ by the choice of T, so by the triangle inequality

$$\begin{aligned}|x_i| &\le |x_i - z_i| + |z_i| \\ &\le |x_i - z_i| + |z_j| \\ &\le |x_i - z_i| + |x_j - z_j| + |x_j|\end{aligned}$$

and so by Cauchy-Schwarz inequality

$$|x_i|^2 \le 2|x_j|^2 + 4|x_i - z_i|^2 + 4|x_j - z_j|^2. \tag{30}$$

Since $|T| = 2|S|$, we can match up each element $i \in S \setminus T$ with a distinct pair P_i of two elements of $T \setminus S$. Summing up (30) for $j \in P_i$ and dividing by two,

$$|x_i|^2 \le \|x_{P_i}\|_2^2 + 4|x_i - z_i|^2 + 2\|(x - z)_{P_i}\|_2^2.$$

Summing up over $i \in S \setminus T$, we have

$$\|x_{S\setminus T}\|_2^2 \le \|x_{T\setminus S}\|_2^2 + 4\|(x - z)_{S\setminus T}\|_2^2 + 2\|(x - z)_{T\setminus S}\|_2^2$$

which gives the desired result. □

9.1 Lemmas on Quantiles

This section proves some elementary lemmas on quantiles of random variables, which are a generalization of the median.

Definition 9.2. *For $f \ge 0$, we define the $1 - f$ quantile quant^f of any list $x_1, \ldots, x_n \in R$ to be the $\lceil (1 - f)n \rceil$th smallest element in the list.*

Then $\text{median} = \text{quant}^{1/2}$ for lists of odd length.

Lemma 9.3. *Let $f > 0$ and T be constants. Let $X_1, \dots, X_T$ be independent nonnegative integer random variables with $\mathbb{E}[X_i] \le \epsilon < 1$ for all i. Let Y satisfy*

$$Y \le \begin{cases} 0 & \text{if } \text{quant}^f X_i = 0 \\ \sum X_i & \text{otherwise} \end{cases}$$

Then $\mathbb{E}[Y] \lesssim \epsilon^{fT}$.

Proof. For each i, we have $\Pr[X_i = 0] \ge 1 - \epsilon$. Let B_i be a $\{0,1\}$-valued random variable with $\Pr[B_i = 1] = \epsilon$ and jointly distributed with X_i such that $X_i = 0$ whenever $B_i = 0$. Then let X_i' be a random variable distributed according to $(X_i \mid B_i = 1)$ independent of B_i, so that $X_i = B_i X_i'$. Then $\mathbb{E}[X_i'] = \mathbb{E}[X_i]/\mathbb{E}[B_i] \le 1$, and we have

$$Y \le \begin{cases} 0 & \text{if } \text{quant}^f B_i = 0 \\ \sum X_i' & \text{otherwise.} \end{cases}$$

Therefore

$$\begin{aligned} \mathbb{E}[Y] &\le \mathbb{E}[\sum_i X_i'] \Pr[\text{quant}^f B_i \ne 0] \\ &\le T \Pr[\sum B_i \ge 1 + \lfloor fT \rfloor] \\ &\le T \binom{T}{1 + \lfloor fT \rfloor} \epsilon^{fT} \lesssim \epsilon^{fT}. \end{aligned}$$

□

Lemma 9.4. *Let $f, \delta > 0$ be constants and T be a sufficiently large constant (depending on f and δ). Let $X^1, \dots, X^T \in \mathbb{R}^n$ be independent random variables with nonnegative coordinates and $\mathbb{E}[X_i^t] \le \mu$ independent of i and t. Then for any $k \le n$,*

$$\mathbb{E}[\max_{|A|=k} \sum_{i \in A} \text{quant}^f{}_t \, X_i^t] \lesssim k\mu(n/k)^\delta$$

Proof. Let $Y_i = \text{quant}^f{}_t \, X_i^t$. We have for any threshold η that

$$\begin{aligned} \Pr[Y_i \ge \eta] &= \Pr[|\{t : X_i^t \ge \eta\}| \ge 1 + \lfloor fT \rfloor] \\ &\le \binom{T}{1 + \lfloor fT \rfloor} (\mu/\eta)^{fT} \\ &\lesssim (\mu/\eta)^{fT}. \end{aligned}$$

Therefore $\mathbb{E}[|\{i : Y_i \ge \eta\}|] \le n(\mu/\eta)^{fT}$. But then

$$\begin{aligned} \mathbb{E}[\max_{|A|=k} \sum_{i \in A} Y_i] &= \mathbb{E} \int_0^\infty \min(k, |\{i : Y_i \ge \eta\}|) d\eta \\ &\le \int_0^\infty \min(k, n(\mu/\eta)^{fT}) d\eta \\ &= k\mu(n/k)^{1/fT} \int_0^\infty \min(1, u^{-fT}) du \\ &= k\mu(n/k)^{1/fT} \left(1 + \frac{1}{fT - 1}\right). \end{aligned}$$

If $T > 1/(\delta f)$ and $T > 2/f$ this gives the result. □

Lemma 9.5. *For any $x_1, \dots, x_n \in \mathbb{C}$ with n odd we have*

$$\mathbb{E}[|\operatorname*{median}_t x_t|^2] \le 4 \max_t \mathbb{E}[|x_t|^2]$$

where the median is taken separately in the real and imaginary axes.

Proof. We will show that if $x_i \in \mathbb{R}$ then

$$\mathbb{E}[(\operatorname*{median}_t x_t)^2] \le 2 \max_t \mathbb{E}[x_t^2].$$

applying this separately to the real and imaginary axes gives the result.

Let S be jointly distributed with x as a set of $(n+1)/2$ coordinates i with $x_i^2 \ge \operatorname{median}_t x_t^2$. This must exist by choosing coordinates less than or greater than x_i. Then

$$\mathbb{E}[(\operatorname*{median}_t x_t)^2] \le \operatorname*{mean}_{i \in S} x_i^2 \le \frac{2}{n+1} \sum_i x_i^2 \le 2 \operatorname*{mean}_{i \in [n]} x_i^2 \le 2 \max_i x_i^2.$$

□

10 Location

Algorithm 6 Location

procedure LOCATESIGNAL($\widehat{x}, \chi, B, \sigma, b, R, p$)
  $n \leftarrow \text{DIM}(\widehat{x})$. ▷ Dimension of vector
  $\gamma \leftarrow R^{1/40 \log_2 \log_2 R}$
  $c \leftarrow O(\log\log(n/B)\log(1/p))$.
  $T \leftarrow$ LOCATE1SPARSESAMPLES($n, \gamma, c, n/B$).
  $u^a_{[B]} \leftarrow$ HASHTOBINS($\widehat{x}, \chi, P_{\sigma,a,b}, B, R$) for $a \in T$.
  $\widehat{v^j}_a := u^a_j$ for $a \in T$ and $j \in [B]$.
  $L \leftarrow \{\}$
  for $j \in [B]$ **do**
    $L \leftarrow L \cup \{\sigma^{-1}(\text{LOCATE1SPARSE}(\widehat{v^j}, T, \gamma, jn/B, n/B))\}$
  end for
  return L
end procedure
procedure LOCATE1SPARSESAMPLES(n, γ, c, w)
  $\delta \leftarrow \gamma^{1/10}$
  $t_{max} \leftarrow O(\log_{1/\delta} w)$.
  $g_{i,t} \in [n]$ uniformly for $i \in [c], t \in [t_{max}]$.
  $f_t \in [\delta^{1-t}/8, \delta^{1-t}/4]$ an arbitrary integer, for all $t \in [t_{max}]$.
  return $T = \cup_{t\in[t_{max}],i\in[c]}\{g_{i,t}, g_{i,t} + f_t\}$ for all i, t.
end procedure
procedure LOCATE1SPARSE($\widehat{v}_T$, T, γ, l, w)
  $\delta \leftarrow \gamma^{1/10}$
  w_t defined to be $w\delta^{t-1}$.
  f_t defined to be any integer in $[(n/w_t)/8, (n/w_t)/4]$.
  Expects $T = \cup_{t\in[t_{max}],i\in[c]}\{(g_{i,t}, g_{i,t} + f_t)\}$ for $t_{max} = O(\log_{1/\gamma} w)$
  Define $m^{(i)}_t = \phi(\widehat{v}_{g_{i,t}+f_t}/\widehat{v}_{g_{i,t}})$. ▷ Estimates of $f_t i^* 2\pi/n$
  Define $m_t = \text{median}_i\, m^{(i)}_t$.
  $l_1 \leftarrow l, w_1 \leftarrow w$. ▷ Location in $l_1 - w_1/2, l_1 + w_1/2$
  for $t = 1, \ldots, t_{max}$ **do**
    $o_t \leftarrow \frac{m_t n/(2\pi) - (f_t l_t \bmod n)}{f_t}$ ▷ Within $[-n/2f_t, n/2f_t]$
    $l_{t+1} \leftarrow l_t + o_t$
  end for
  return ROUND($l_{t_{max}+1}$).
end procedure

We first show that LOCATE1SPARSE solves the 1-sparse recovery problem. This result is independent of the rest of the machinery in this paper: if v has a single component with $1 - \gamma^{1/2}$ of the mass, we find it with $\widetilde{O}(\log_{1/\gamma} n)$ samples of $\widehat{v}$.

Lemma 10.1. *Let $1/\gamma, c$ be larger than a sufficiently large constant. Let $\widehat{v} \in \mathbb{C}^n$, and suppose that there*

exists an $i^* \in [l - w/2, l + w/2]$ *such that*

$$\gamma^{1/2}|v_{i^*}|^2 \geq \sum_{j \neq i^*} |v_j|^2.$$

Then LOCATE1SPARSE*(*$\widehat{v}_T$*,* T*,* γ*,* l_1*,* $l_1 + w_1$*) returns* i^* *with all but* $\gamma^{\Omega(c)} \log w$ *probability, where the set* T *is the output of* LOCATE1SPARSESAMPLES*(*n*,* γ*,* c*,* w*) and has size* $|T| = O(c(1 + \log_{1/\gamma} w))$*. The time taken is* $O(|T|) = O(c(1 + \log_{1/\gamma} w))$*.*

Proof. Note that for uniformly random $g \in [n]$, by Parseval's theorem

$$\mathbb{E}[|\sqrt{n}\widehat{v}_g - \omega^{gi^*} v_{i^*}|^2] = \sum_{j \neq i^*} |v_j|^2 \leq \gamma^{1/2}|v_{i^*}|^2$$

Set $b = \gamma^{1/20}$. By Markov's inequality, with $1 - b$ probability

$$|\sqrt{n}\widehat{v}_g - \omega^{gi^*} v_{i^*}| \leq \sqrt{\gamma^{1/2}/b}|v_{i^*}|$$

and so

$$\|\phi(\widehat{v}_g) - (\frac{2\pi}{n} g i^* + \phi(v_{i^*}))\|_{\bigcirc} = \|\phi(\sqrt{n}\widehat{v}_g) - \phi\omega^{gi^*} v_{i^*}\|_{\bigcirc} \leq \sin^{-1}(\sqrt{\gamma^{1/2}/b}) \leq 2\sqrt{\gamma^{1/2}/b}$$

where $\|a - b\|_{\bigcirc} = \min_{i \in \mathbb{Z}}(|a - b - 2\pi i|)$ denotes the "circular distance" between a and b. Hence for any $(g_{i,t}, g_{i,t} + f_t)$, we have that

$$m_t^{(i)} = \phi(\widehat{v}_{g_{i,t}+f_t} / \widehat{v}_{g_{i,t}})$$

satisfies

$$\|m_t^{(i)} - f_t i^* 2\pi/n\|_{\bigcirc} \leq 4\sqrt{\gamma^{1/2}/b} \tag{31}$$

with probability $1 - 2b$ as a distribution over $g_{i,t}$. Because this is independent for different i, for any t by a Chernoff bound we have that (31) holds for at least $3c/4$ of the $m_t^{(i)}$ with probability at least

$$1 - \binom{c}{c/4}(2b)^{c/4} \geq 1 - 2^c(2b)^{c/4} = 1 - (32b)^{c/4} = 1 - \gamma^{\Omega(c)}.$$

If so, their median satisfies the same property[2]

$$\|m_t - f_t i^* 2\pi/n\|_{\bigcirc} \leq 4\sqrt{\gamma^{1/2}/b} \leq 2\pi b\delta. \tag{32}$$

Since there are $\log_{1/\gamma^{1/2}} w < \log w$ different t, by a union bound (32) holds for all t with the desired probability

$$1 - \gamma^{-\Omega(c)} \log w.$$

We will show that this implies that i^* is recovered by the algorithm.

[2]To define a median over the circle, we need to cut the circle somewhere; we may do so at any position not within $4\sqrt{\gamma^{1/2}/b}$ of at least $c/4$ of the points.

We will have by induction that, for all t, $i^* \in [l_t - w_t/2, l_t + w_t/2]$. This certainly holds at $t = 1$. Recall that $4w_t \le n/f_t \le 8w_t$ by the construction of f_t.

For any t, by (32) we have that $o_t f_t$ lies within δbn of $(f_t i^* - f_t l_t)$ (modulo n). Hence $(i^* - l_t)$ lies within $\delta bn/f_t$ of $o_t + zn/f_t$ for $|o_t| < n/(2f_t)$ and some integer z. But since $|i^* - l_t| \le w_t/2 \le n/(8f_t)$ and $\delta bn/f_t < n/(4f_t)$, this means that $z = 0$ and we have that $(i^* - l_t)$ lies within $\delta bn/f_t$ of o_t. Since

$$\delta bn/f_t \le \delta b 8 w_t \le \delta w_t/2 \le w_{t+1}/2,$$

i^* lies within $w_{t+1}/2$ of $l_{t+1} = l_t + o_t$ and the inductive step holds.

In the end, therefore, i^* lies within $w_{t_{max}}/2 = w\delta^{t_{max}-1}/2 < 1/2$ of l, so it is returned by the algorithm. □

We now relate Lemma 10.1, which guarantees 1-sparse recovery, to k-sparse recovery of well-hashed signals.

Lemma 10.2. *Let x be a signal, and B and R larger than sufficiently large constants. An invocation of* LOCATESIGNAL *returns a list L of size B such that each well-hashed (per Definition 3.4) $i \in [n]$ is present in L with probability at least $1-p$. The sample complexity is $O(B \log(Rn/B) \log\log R \log\log(n/B) \max(1, \log_R(1/p))$, and the time complexity is $O(\log R)$ larger.*

Proof. Consider any well-hashed i and $j = h(i)$. We define the vector $y^j \in \mathbb{C}^n$ by

$$y^j_{\sigma\ell} = x_\ell G_{\pi(\ell) - jn/B} = x_\ell G_{o_i(\ell)}.$$

Then

$$u^a_j = \sum_\ell \omega^{a\ell} y^j_\ell = \sqrt{n}\widehat{y^j}_a,$$

i.e. $\widehat{v^j} = \widehat{y^j}/\sqrt{n}$, so $v^j_{\sigma\ell} = x_\ell G_{o_i(\ell)}/\sqrt{n}$.

By the definition 3.4 of well-hashedness, over uniformly random $a \in [n]$,

$$\gamma^{1/2} x_i^2 \ge \mathop{\mathbb{E}}_a[|G^{-1}_{o_i(i)} \omega^{-a\sigma i} v^j_a - x_i|^2]$$

If we define $v_{-\sigma i} = v_{[n]\setminus\{\sigma i\}}$, we have after multiplying by $G^2_{o_i(i)}$ that

$$\begin{aligned}
\gamma^{1/2}|v^j_{\sigma i}|^2/n = G^2_{o_i(i)}\gamma^{1/2}|x_i|^2 &\ge \mathop{\mathbb{E}}_a[|\widehat{v^j}_a - \omega^{a\sigma i} G_{o_i(i)} x_i|^2] \\
&= \mathop{\mathbb{E}}_a[|\widehat{v^j}_a - \frac{1}{\sqrt{n}}\omega^{a\sigma i} v^j_{\sigma i}|^2] \\
&= \mathop{\mathbb{E}}_a[|\widehat{(v^j_{-\sigma i})}_a|^2]
\end{aligned}$$

Therefore by Parseval's inequality,

$$\gamma^{1/2}|v^j_{\sigma i}|^2 \ge \|v^j_{-\sigma i}\|_2^2.$$

This is precisely the requirement of Lemma 10.1. Hence LOCATE1SPARSE will return σi with all but $\gamma^{\Omega(c)} \log(n/B)$ probability, in which case i will be in the output set L.

Recall that $\log_{1/\gamma} R \lesssim \log\log R$. Setting

$$\begin{aligned}
c &= \Theta(\max(1, \log_\gamma(\log(n/B)/p))) \\
&\lesssim \max(1, (\log_R \log(n/B) + \log_R(1/p)) \log\log R) \\
&\lesssim \log\log(n/B) \max(1, \log_R(1/p))
\end{aligned}$$

gives the desired probability $1 - p.$, the number of samples is

$$\begin{aligned}|T|B\log R &= cB\log R\max(1,\log_{1/\gamma}(n/B))\\ &\lesssim B\log(Rn/B)\log\log R\log\log(n/B)\max(1,\log_R(1/p)).\end{aligned}$$

The time taken is dominated by HASHTOBINS, which takes sample complexity times $\log R$ time. □

11 Filter construction

Claim 2.2. (Claim 2.2 of [HIKP12b]). *Let $\mathcal{F}^{-1}(x)$ denote the inverse Fourier transform of x. Then*

$$(\mathcal{F}^{-1}(P_{\sigma,a,b}\widehat{x}))_{\pi(i)} = x_i\omega^{a\sigma i}.$$

Proof.

$$\begin{aligned}\mathcal{F}^{-1}(P_{\sigma,a,b}\widehat{x})_{\sigma(i-b)} &= \frac{1}{\sqrt{n}}\sum_{j\in[n]}\omega^{-\sigma(i-b)j}(P_{\sigma,a,b}\widehat{x})_j\\ &= \frac{1}{\sqrt{n}}\sum_{j\in[n]}\omega^{-\sigma(i-b)j}\widehat{x}_{\sigma(j+a)}\omega^{-\sigma bj}\\ &= \omega^{a\sigma i}\frac{1}{\sqrt{n}}\sum_{j\in[n]}\omega^{-i\sigma(j+a)}\widehat{x}_{\sigma(j+a)}\\ &= x_i\omega^{a\sigma i}.\end{aligned}$$

□

Lemma 11.1. *If G is a flat window function with B buckets and contrast $R > 2$, then for some constant c,*

$$\sum_{|i|>cn/2B} G_i^2 \lesssim \frac{n}{R^2B}$$

Proof. Let c be the constant such that $G_i \le (\frac{cn}{|i|B})^t$ for $t = \log R$. Then

$$\begin{aligned}\sum_{|i|>2cn/B} G_i^2 &\le 2\sum_{i=2cn/B}^{\infty}(\frac{cn}{|i|B})^{2\log R}\\ &\le \frac{4cn}{B}\sum_{i=1}^{\infty}(\frac{1}{2i})^{2\log R}\\ &= \frac{4cn}{R^2B}\sum_{i=1}^{\infty}i^{-2\log R}\\ &\lesssim \frac{n}{R^2B}\end{aligned}$$

and rescaling c gives the result. □

Lemma 3.2. *There exist flat window functions where* $|\operatorname{supp}(\widehat{G})| \lesssim B \log R$. *Moreover,* $\operatorname{supp}(\widehat{G}) \subset [-O(B \log R), O(B \log R)]$.

Proof. Suppose B is an odd integer; otherwise, replace B with $B' = B - 1$. The properties for B' will imply the properties for B, albeit with a worse constant in the third property.

Define $\widehat{F}$ to be a rectangular filter of length B, scaled so F is the Dirichlet kernel

$$F_i = \frac{\sin(\pi B i/n)}{B \sin(\pi i/n)}.$$

Noting that $2|x| \le |\sin(\pi x)| \le \pi|x|$ for $|x| \le 1/2$, we have for all i that

$$|F_i| \le \frac{|\sin(\pi B i/n)|}{2Bi/n} \le \frac{n}{2B|i|} \tag{33}$$

and for $i \in [-n/2B, n/2B]$ that

$$|F_i| \ge |\frac{2Bi/n}{B\pi i/n}| = \frac{2}{\pi}. \tag{34}$$

Define $\widehat{F'}$ to be $\widehat{F}$ convolved with itself $t = \Theta(\log R)$ times for an even integer t, so $\|\widehat{F'}\|_0 \lesssim B \log R$ and $F'_i = F_i^t$, and by (33)

$$0 \le F'_i \le \left(\frac{n}{2Bi}\right)^t. \tag{35}$$

Now, define G to be F' convolved with a length $\ell = 2\lfloor n/(2B) \rfloor + 1$ rectangular filter, i.e.

$$G_i = \sum_{|j-i| \le n/(2B)} F'_j,$$

so $\widehat{G}$ is $\widehat{F'}$ multiplied by a scaled Dirichlet kernel. By the last equation, it follows that $\|\widehat{G}\|_0 \le \|\widehat{F'}\|_0 \lesssim B \log R$. We would just like to show that $G/\|G\|_\infty$ satisfies the flat window function requirements.

Since $F'_i \ge 0$ per (35), we have $0 \le G_i/\|G\|_\infty \le 1$ so $G/\|G\|_\infty$ passes the second property of a flat window function.

For the first property of flat window functions, let $a = \sum_{i=0}^{\lfloor n/(2B) \rfloor} F'_i$. We have that $G_i \ge a$ for $|i| \le n/(2B)$ because each of those terms (or their symmetries F'_{-i}) appear in the summation that forms G_i. So it suffices to show that $G_i \le 3a$ for all i.

Define $S_k = \mathbb{Z} \cap [kn/(2B), (k+1)n/(2B)]$ for $k \in \mathbb{Z}$, so $|S_k| \le \lceil n/(2B) \rceil$ for all k. For any i, $\{j : |j - i| \le n/(2B)\}$ has nonzero intersection with at most 3 different S_k. Hence it suffices to show for all k that

$$a \ge \sum_{j \in S_k} F'_j.$$

To do this, we extend the definition of F'_x to all $x \in \mathbb{R}$. By symmetry, it suffices to consider $k \ge 0$. We have that $\sin(\pi x/n)$ is increasing on $[0, n/2]$, so for $0 \le x \le n/2 - n/B$ we have

$$F'_{x+n/B}/F'_x = \left(\frac{\sin(\pi x/n)}{\sin(\pi(x + n/B)/n)}\right)^t < 1.$$

Therefore, for each $j \in S_k$,

$$F'_j \le F'_{j-\lceil k/2 \rceil (n/B)} = F'_{|j-\lceil k/2 \rceil (n/B)|}.$$

Let $T_k = \{|j - \lceil k/2 \rceil (n/B)| : j \in S_k\}$. By considering the even and odd cases for k, we conclude that $T_k \subset [0, n/(2B)]$ and that for some parameter $\theta \ge 0$ we have

$$T_k = \{\theta, \theta + 1, \ldots, \theta + |T| - 1\}.$$

Since F' is decreasing on $[0, n/(2B)]$ we have that

$$\sum_{j \in S_k} F'_j \le \sum_{j \in T_k} F'_j = \sum_{j=0}^{|T|-1} F'_{\theta+j} \le \sum_{j=0}^{|T|-1} F'_j \le \sum_{j=0}^{\lfloor n/(2B) \rfloor} F'_j = a.$$

Therefore $G/\|G\|_\infty$ satisfies the first property of a flat window function.

Lastly, the third property of flat window functions. Consider $i = \alpha n/2B$ with $\alpha \ge 2$ (for smaller i, $G_i \le 1$ suffices as a bound). We have by (35) that

$$G_i \le \ell \max_{|j-i| \le n/2B} F'_j \le \ell (\frac{n}{2B(|i| - n/(2B))})^t = \ell (\frac{1}{\alpha - 1})^t.$$

We also have by (34) that

$$\|G\|_\infty \ge G_0 \ge \ell \min_{|i| \le n/(2B)} F'_i \ge \ell (2/\pi)^t.$$

Hence

$$G_i/\|G\|_\infty \le (\frac{\pi}{2(\alpha - 1)})^t = (O(1/\alpha))^t = (O(\frac{n}{B|i|}))^t$$

which is the third property of flat window functions. Thus $G/\|G\|_\infty$ is the desired flat window function. □

For a bucketing (σ, b), each coordinate j is permuted to an index $\pi(j) = \sigma j - b$, with nearest multiple of (n/B) being $(n/B)h(j)$. Define the offset of j relative to i to be $o_i(j) = \pi(j) - (n/B)h(i)$.

Given a bucketing (σ, b), for each bucket $j \in [B]$ we define the associated "bucket vectors" $v^{(j)}$ given by

$$v^{(j)}_{\sigma i} := x_i G_{\pi(i) - (n/B)j}.$$

This has the property that running the algorithm with offset a yields $u \in \mathbb{R}^B$ given by

$$u_j = \sum_i v^{(j)}_i \omega^{ia} = \widehat{v^{(j)}}_a.$$

For any bucketing (σ, b), we say that a bucket j has *noise at most* μ^2 if $\|v^{(j)}\|_2^2 \le \mu^2$. We say that an index i is hashed with noise at most μ^2 if, for $j = h(i)$, we have

$$\|v^{(j)} - x_i G_{\pi(i) - (n/B)j}\|_2^2 \le \mu^2.$$

We show how to relate the pairwise independence property 2.3 to flat window functions:

Lemma 11.2. *Let G be a flat window function with B buckets and contrast R. Then for $i \ne j$, there exists a constant c such that*

$$\mathbb{E}[G^2_{o_i(j)} \cdot I[|o_i(j)| > cn/B]] \lesssim \frac{1}{R^2 B}.$$

where $I[a > b]$ is 1 when $a > b$ and 0 otherwise.

Proof. Note that $o_i(j) = \pi(j) - (n/B)h(i)$ is within $n/(2B)$ of $\pi(j) - \pi(i) = \sigma(j-i)$. Let $f \geq 1$ be the constant such that

$$G_{o_i(j)} \leq \left(\frac{f}{B|o_i(j)|/n}\right)^{\log R}.$$

Then

$$\begin{aligned} G_{o_i(j)} &\leq \max_{|a-\sigma(i-j)|<n/(2B)} G_a \\ &\leq \max_{|a-\sigma(i-j)|<n/(2B)} \left(\frac{f}{B|a|/n}\right)^{\log R} \\ &\leq \left|\frac{f}{B|\sigma(i-j)|/n - 1/2}\right|^{\log R} \end{aligned}$$

as well as $G_{o_i(j)} \leq 1$. Define

$$y_b = \min\left(1, \left|\frac{f}{B|b|/n - 1/2}\right|^{\log R}\right).$$

It suffices to show that, for any $a \neq 0$ and as a distribution over σ,

$$\mathbb{E}[y_{\sigma a}^2 \cdot I[|\sigma a| > cn/B]] \lesssim \frac{1}{R^2 B}.$$

Let $D = 3fn/B \lesssim n/B$. Note that, for $d \geq 1$ and $|b| \geq dD > (2df + 1/2)n/B$,

$$y_b \leq \left(\frac{1}{2d}\right)^{\log R} = \frac{1}{R}\frac{1}{R^{\log d}}.$$

Consider the "levels sets" $S_l : \{b \mid 2^l D \leq |b| < 2^{l+1}D\}$, for $l \geq 0$. Then by Lemma 2.3,

$$\Pr[\sigma a \in S_l] \leq 4 \cdot 2^{l+1} D/n \lesssim 2^l D/n$$

and

$$\max_{b \in S_l} y_b \leq \frac{1}{R^{l+1}}.$$

Hence

$$\begin{aligned} \mathbb{E}[y_{\sigma a}^2 \cdot I[|\sigma a| \geq D]] &\lesssim \sum_{l=0}^{\infty} (2^l D/n) R^{-2l-2} \\ &\lesssim D/(R^2 n) \lesssim 1/(R^2 B) \end{aligned}$$

because $R^2 > 2$. Since $D \lesssim n/B$, this gives the result. □

Lemma 11.3. HASHTOBINS*($\widehat{x}, \chi, P_{\sigma,a,b}, B, R$) computes u such that for any $i \in [n]$,*

$$u_{h(i)} = \Delta_{h(i)} + \sum_j G_{o_i(j)}(x - \chi)_j \omega^{a\sigma j}$$

where G is the flat window function with B buckets and contrast R from Lemma 3.2, and $\Delta_{h(i)}^2 \leq \|\chi\|_2^2/(R^* n^{11})$ is a negligible error term. It takes $O(B \log R)$ samples, and if $\|\chi\|_0 \lesssim B$, it takes and $O(B \log R \log(Rn)$ time.

Proof. Let $S = \mathrm{supp}(\widehat{G})$, so $|S| \lesssim B \log R$ and in fact $S \subset [-O(B \log R), O(B \log R)]$.

First, HASHTOBINS computes

$$y' = \widehat{G} \cdot P_{\sigma,a,b}\widehat{x - \chi'} = \widehat{G} \cdot P_{\sigma,a,b}\widehat{x - \chi} + \widehat{G} \cdot P_{\sigma,a,b}\widehat{\chi - \chi'},$$

for an approximation $\widehat{\chi}'$ to $\widehat{\chi}$. This is efficient because one can compute $(P_{\sigma,a,b}\widehat{x})_S$ with $O(|S|)$ time and samples, and $P_{\sigma,a,b}\widehat{\chi}'_S$ is easily computed from $\widehat{\chi}'_T$ for $T = \{\sigma(j-b) : j \in S\}$. Since T is an arithmetic sequence and χ is B-sparse, by Corollary 12.2, an approximation $\widehat{\chi}'$ to $\widehat{\chi}$ can be computed in $O(B \log R \log(Rn))$ time such that

$$|\widehat{\chi}_i - \widehat{\chi}'_i| < \frac{\|\chi\|_2}{R^* n^{13}}$$

for all $i \in T$. Since $\|\widehat{G}\|_1 \le \sqrt{n}\|\widehat{G}\|_2 = \sqrt{n}\|G\|_2 \le n\|G\|_\infty \le n$ and $\widehat{G}$ is 0 outside S, this implies that

$$\|\widehat{G} \cdot P_{\sigma,a,b}(\widehat{\chi - \chi'})\|_2 \le \|\widehat{G}\|_1 \max_{i \in S} |(P_{\sigma,a,b}(\widehat{\chi - \chi'}))_i| = \|\widehat{G}\|_1 \max_{i \in T} |(\widehat{\chi - \chi'})_i| \le \frac{\|\chi\|_2}{R^* n^{12}}. \tag{36}$$

Define Δ by $\widehat{\Delta} = \sqrt{n}\widehat{G} \cdot P_{\sigma,a,b}(\widehat{\chi - \chi'})$. Then HASHTOBINS computes $u \in \mathbb{C}^B$ such that for all i,

$$u_{h(i)} = \sqrt{n}\mathcal{F}^{-1}(y')_{h(i)n/B} = \sqrt{n}\mathcal{F}^{-1}(y)_{h(i)n/B} + \Delta_{h(i)n/B},$$

for $y = \widehat{G} \cdot P_{\sigma,a,b}\widehat{x - \chi}$. This computation takes $O(\|y'\|_0 + B \log B) \lesssim B \log(Rn)$ time. We have by the convolution theorem that

$$\begin{aligned}
u_{h(i)} &= \sqrt{n}\mathcal{F}^{-1}(\widehat{G} \cdot P_{\sigma,a,b}(\widehat{x - \chi}))_{h(i)n/B} + \Delta_{h(i)n/B} \\
&= (G * \mathcal{F}(P_{\sigma,a,b}(\widehat{x - \chi})))_{h(i)n/B} + \Delta_{h(i)n/B} \\
&= \sum_{\pi(j) \in [n]} G_{h(i)n/B - \pi(j)} \mathcal{F}(P_{\sigma,a,b}(\widehat{x - \chi}))_{\pi(j)} + \Delta_{h(i)n/B} \\
&= \sum_{i \in [n]} G_{o_i(j)} (x - \chi)_j \omega^{a\sigma j} + \Delta_{h(i)n/B}
\end{aligned}$$

where the last step is the definition of $o_i(j)$ and Claim 2.2 of [HIKP12b] (reproduced here as Claim 2.2).

Finally, we note that

$$|\Delta_{h(i)n/B}| \le \|\Delta\|_2 = \|\widehat{\Delta}\|_2 = \sqrt{n}\|\widehat{G} \cdot P_{\sigma,a,b}(\widehat{\chi - \chi'})\|_2 \le \frac{\|\chi\|_2}{R^* n^{11}},$$

where we used (36) in the last step. This completes the proof. □

Lemma 3.3. *Let $(\sigma, a, b) \in [n]$ be uniform subject to σ being odd. Let $u \in \mathbb{C}^B$ denote the result of* HASHTOBINS$(\widehat{x}^*, \chi, P_{\sigma,a,b}, B, R)$. *Fix a coordinate $i \in [n]$ and define $x = x^* - \chi$. For each (σ, b), we can define variables $C \subset [n]$ and $w > 0$ (and in particular, $C = \{j \ne i : |\sigma(i-j) \bmod n| \le cn/B\}$ for some constant c,) so that*

- *For all j, as a distribution over (σ, b),*

$$\Pr[j \in C] \lesssim 1/B.$$

- *As a distribution over* (σ, b)*,*

$$\mathbb{E}[w^2] \lesssim \frac{\|x\|_2^2}{R^2 B} + \frac{\|x^*\|_2^2}{R^* n^{11}}$$

- *Conditioned on* (σ, b) *and as a distribution over* a*,*

$$\mathop{\mathbb{E}}_{a}[|G^{-1}_{o_i(i)}\omega^{-a\sigma i}u_{h(i)} - x_i|^2] \lesssim w^2 + \|x_C\|_2^2.$$

Proof. By Lemma 2.3, for any fixed i and j,

$$\Pr[j \in C] = \Pr[|\sigma(i-j)| \le cn/B] \lesssim 1/B$$

which gives the first part.

Define $x' = x - \chi$. Per Lemma 11.3, HASHTOBINS computes the vector $u \in \mathbb{C}^B$ given by

$$u_{h(i)} - \Delta_{h(i)} = \sum_j G_{o_i(j)} x'_j \omega^{a\sigma j}$$

for some Δ with $\|\Delta\|_\infty^2 \le \|x\|_2^2/(R^* n^{11})$. We define the vector $v \in \mathbb{C}^n$ by $v_{\sigma j} = x'_j G_{o_i(j)}$, so that

$$u_{h(i)} - \Delta_{h(i)} = \sum_j \omega^{aj} v_j = \sqrt{n}\widehat{v}_a$$

so

$$u_{h(i)} - \omega^{a\sigma i} G_{o_i(i)} x'_i - \Delta_{h(i)} = \sqrt{n}(\widehat{v_{\overline{\{\sigma i\}}}})_a.$$

By Parseval's theorem, therefore,

$$\begin{aligned}
\mathop{\mathbb{E}}_{a}[|G^{-1}_{o_i(i)}\omega^{-a\sigma i}u_{h(i)} - x'_i|^2] &\le 2G^{-2}_{o_i(i)}(\mathop{\mathbb{E}}_{a}[|u_{h(i)} - \omega^{a\sigma i}G_{o_i(i)}x'_i - \Delta_{h(i)}|^2] + \mathop{\mathbb{E}}_{a}[\Delta^2_{h(i)}]) \\
&= 2G^{-2}_{o_i(i)}(\|v_{\overline{\{\sigma i\}}}\|_2^2 + \Delta^2_{h(i)}) \\
&\lesssim \frac{\|\chi\|_2^2}{R^* n^{11}} + \sum_{j \ne i} |x'_j G_{o_i(j)}|^2 \\
&\le \frac{\|\chi\|_2^2}{R^* n^{11}} + \sum_{j \notin C \cup \{i\}} |x'_j G_{o_i(j)}|^2 + \sum_{j \in C} |x'_j|^2
\end{aligned}$$

If we define w^2 to be the first two terms, we satisfy the third part of the lemma statement. Next, we have that

$$\frac{\|\chi\|_2^2}{R^* n^{11}} \le 2(\frac{\|x\|_2^2 + \|x-\chi\|_2^2}{R^* n^{11}}) \lesssim \frac{\|x\|_2^2}{R^* n^{11}} + \frac{\|x-\chi\|_2^2}{R^2 B}.$$

From the other term, for $j \notin C \cup \{i\}$, $|\sigma(i-j)| > cn/B$ so $o_i(j) > (c-1)n/B$. Hence for sufficiently large c, by Lemma 11.2,

$$\mathbb{E}[\sum_{j \notin C \cup \{i\}} |x'_j G_{o_i(j)}|^2] \le \sum_{j \ne i} |x_j - \chi_j|^2 \, \mathbb{E}[G^2_{o_i(j)} \cdot I[o_i(j) > (c-1)n/B]] \le \frac{\|x-\chi\|_2^2}{R^2 B}.$$

Hence their sum has

$$\mathbb{E}[w^2] \lesssim \frac{\|x\|_2^2}{R^* n^{11}} + \frac{\|x-\chi\|_2^2}{R^2 B}.$$

This proves the second part of the lemma statement, completing the proof. □

12 Semi-equispaced Fourier Transform

Algorithm 7 Semi-equispaced Fourier Transform in $O(ck \log n)$ time

procedure SEMIEQUISPACEFFT(x, c) ▷ $x \in \mathbb{C}^n$ is k-sparse
  Round k up to a factor of n.
  $G, \widehat{G'} \leftarrow \text{FILTERS}(n, k, c)$.
  $y_i \leftarrow \frac{1}{\sqrt{n}}(x * G)_{in/2k}$ for $i \in [2k]$.
  $\widehat{y} \leftarrow \text{FFT}(y)$ ▷ $2k$ dimensional
  $\widehat{x}'_i \leftarrow \widehat{y}_i$ for $|i| \leq k/2$.
  return $\widehat{x}'$
end procedure

The following is similar to results of [DR93, PST01].

Lemma 12.1. *Let n be a power of two and $c \geq 1$. Suppose $x \in \mathbb{C}^n$ is k-sparse for some k. We can compute $\widehat{x}'_i$ for all $|i| \leq k/2$ in $O(ck \log n)$ time such that*

$$|\widehat{x}'_i - \widehat{x}_i| \leq \|x\|_2 / n^c.$$

Proof. Without loss of generality k is a power of two (round up), so $2k$ divides n.

Let $G, \widehat{G'}$ be the flat window functions of [HIKP12a], so that $G_i = 0$ for all $|i| \gtrsim (n/k)c \log n$, $\|G - G'\|_2 \leq n^{-c}$,

$$\widehat{G'}_i = \begin{cases} 1 & \text{if } |i| \leq k/2 \\ 0 & \text{if } |i| \geq k \end{cases},$$

and $\widehat{G'}_i \in [0, 1]$ everywhere. The construction is that G approximates a Gaussian convolved with a rectangular filter and G is a (truncated) Gaussian times a sinc function, and is efficiently computable.

Define

$$z = \frac{1}{\sqrt{n}} x * G.$$

We have that $\widehat{z}_i = \widehat{x}_i \widehat{G}_i$ for all i. Furthermore, because subsampling and aliasing are dual under the Fourier transform, since $y_i = z_{in/(2k)}$ is a subsampling of z we have for $|i| \leq k/2$ that

$$\begin{aligned}
\widehat{x}'_i = \widehat{y}_i &= \sum_{j=0}^{n/2k-1} \widehat{z}_{i+2kj} \\
&= \sum_{j=0}^{n/2k-1} \widehat{x}_{i+2kj} \widehat{G}_{i+2kj} \\
&= \sum_{j=0}^{n/2k-1} \widehat{x}_{i+2kj} \widehat{G'}_{i+2kj} + \sum_{j=0}^{n/2k-1} \widehat{x}_{i+2kj} (\widehat{G}_{i+2kj} - \widehat{G'}_{i+2kj}) \\
&= \widehat{x}_i + \sum_{j=0}^{n/2k-1} \widehat{x}_{i+2kj} (\widehat{G}_{i+2kj} - \widehat{G'}_{i+2kj})
\end{aligned}$$

and so

$$|\widehat{x}'_i - \widehat{x}_i| \le \|\widehat{x}\|_2 \|\widehat{G} - \widehat{G'}\|_2 \le \|x\|_2 n^{-c}$$

as desired.

The time complexity is $O(k \log k)$ for a $2k$-dimensional FFT, plus the time to construct y. Because G_i has a localized support, each nonzero coordinate i of x only contributes to $O(c \log n)$ entries of y. Hence the time to construct y is $O(ck \log n)$ times the time to evaluate G at an arbitrary position. Because G is a Gaussian times a sinc function, assuming we can evaluate exponentials in unit time this is $O(ck \log n)$ total. □

This can be easily generalized to arbitrary arithmetic sequences of length k:

Corollary 12.2. *Let n be a power of two, $c \ge 1$, and σ odd. Suppose $x \in \mathbb{C}^n$ is k-sparse for some k, and $S = \{\sigma(i - b) : i \in \mathbb{Z}, |i| \le k\}$. Then we can compute $\widehat{x}'_i$ for all $i \in S$ in $O(ck \log n)$ time such that*

$$|\widehat{x}'_i - \widehat{x}_i| \le \|x\|_2 / n^c.$$

Proof. Let σ^{-1} denote the inverse of σ modulo n. Define $x^*_j = \omega^{-bj} x_{\sigma^{-1} j}$. Then for all $i \in [n]$,

$$\begin{aligned}
\widehat{x}_{\sigma(i-b)} &= \frac{1}{\sqrt{n}} \sum_{j \in [n]} \omega^{\sigma(i-b)j} x_j \\
&= \frac{1}{\sqrt{n}} \sum_{j \in [n]} \omega^{i\sigma j} \omega^{-b\sigma j} x_j \\
&= \frac{1}{\sqrt{n}} \sum_{j' = \sigma j \in [n]} \omega^{ij'} \omega^{-bj'} x_{\sigma^{-1} j'} \\
&= \widehat{x}^*_i.
\end{aligned}$$

We can sample from $\widehat{x}^*_i$ with $O(1)$ overhead, so by Lemma 12.1 we can approximate $\widehat{x}_{\sigma(i-b)} = \widehat{x}^*_i$ for $|i| \le k$ in $O(ck \log n)$ time. □

To compute $G_{o_i(i)}$, we take the opposite semi-equispaced Fourier transform.

Algorithm 8 Converse semi-equispaced Fourier Transform in $O(k \log(n/\delta))$ time

procedure CONVERSESEMIEQUISPACEFFT($\widehat{x}$, S, c) ▷ $\mathrm{supp}(x) \in [-k/2, k/2]$
  Round k up to a factor of n.
  $G, \widehat{G'} \leftarrow$ FILTERS(n, k, c).
  $u \leftarrow$ INVFFT$(\widehat{x}_{[-k,k]})$ ▷ $2k$ dimensional
  $y_{in/(2k)} \leftarrow u_i$ for $i \in [2k]$.
  $x'_i \leftarrow \frac{1}{\sqrt{n}} \sum_{j \in \mathrm{supp}(G) : i + j \equiv 0 \bmod n/(2k)} G_j y_{i+j}$ for $i \in S$.
  return x'
end procedure

Lemma 12.3. *Let n be a power of two and $c \ge 1$. Suppose $\widehat{x} \in \mathbb{C}^n$ has $\mathrm{supp}(x) \in [-k/2, k/2]$, and let $S \subset [n]$ have $|S| = k$. We can compute x'_i for all $i \in S$ in $O(ck \log n)$ time such that*

$$|x'_i - x_i| \le \|x\|_2 / n^c.$$

Proof. Without loss of generality k is a power of two (round up), so $2k$ divides n.

Let $G, \widehat{G'}$ be the flat window functions of [HIKP12a], so that $G_i = 0$ for all $|i| \gtrsim (n/k)c\log n$, $\|G - G'\|_2 \leq n^{-c}$,

$$\widehat{G'}_i = \begin{cases} 1 & \text{if } |i| \leq k/2 \\ 0 & \text{if } |i| \geq k \end{cases},$$

and $\widehat{G'}_i \in [0,1]$ everywhere. The construction is that G approximates a Gaussian convolved with a rectangular filter and G is a (truncated) Gaussian times a sinc function, and is efficiently computable.

For the y defined in the algorithm, we have that $y_{in/(2k)} = x_{in/(2k)}\sqrt{n/(2k)}$ by the definition of the Fourier transform. Setting $y_j = 0$ elsewhere, y is a scaled subsampling x. Since subsampling and aliasing are dual under the Fourier transform, we have that $\widehat{y}_i = \sum_{j=-\infty}^{\infty} \widehat{x}_{i+2kj}$.

Therefore $\widehat{x} = \widehat{y} \cdot \widehat{G'}$, so $x = \frac{1}{\sqrt{n}} y * G'$. Then for any i,

$$\begin{aligned} |x'_i - x_i| &= \frac{1}{\sqrt{n}} |\sum_j (G_j - G'_j) y_{i+j}| \\ &\leq \frac{1}{\sqrt{n}} \|G - G'\|_2 \|y\|_2 \\ &\lesssim \frac{1}{\sqrt{n}} n^{-c} \sqrt{n/(2k)} \|x\|_2. \end{aligned}$$

Rescaling c gives the result.

The time complexity is $O(k \log k)$ for the Fourier transform and $O(ck \log n)$ for the summation to form x', giving $O(ck \log n)$ time total. □

12.1 Computing G, $\widehat{G}$

Our algorithm needs to know, for each R, both $\widehat{G}_i$ for $|i| \leq B \log R$ and $G_{o_i(j)}$ for various j. Here we show how to compute these up to $1/n^c$ precision for an arbitrary constant c with no additional time overhead beyond the already existing $\log(Rn)$ factor.

Computing $\widehat{G}_i$ for all i is possible in $O(B \log^2 R)$ time, because it is a sinc function times a degree $\log R$ polynomial at each position i. Since we only need this once for each R, the total time is at most a $\log R$ factor above the sample complexity.

For each hashing in estimation phases, we will need to compute $G_{o_i(j)}$ for the set L of $O(B)$ coordinates. We will already know $\widehat{G}$, which is $O(B \log R)$ sparse and dense around the origin. Hence Lemma 12.3 can compute $G_{o_i(j)}$ in $O(B \log R \log n)$ time, which is only $\log n$ more than the sample complexity to perform the hashing.

References

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

[Aka10] A. Akavia. Deterministic sparse Fourier approximation via fooling arithmetic progressions. *COLT*, pages 381–393, 2010.

[BCG+12] P. Boufounos, V. Cevher, A. C. Gilbert, Y. Li, and M. J. Strauss. What's the frequency, kenneth?: Sublinear fourier sampling off the grid. *RANDOM/APPROX*, 2012.

[CCF02] M. Charikar, K. Chen, and M. Farach-Colton. Finding frequent items in data streams. *ICALP*, 2002.

[CGV12] Mahdi Cheraghchi, Venkatesan Guruswami, and Ameya Velingker. Restricted isometry of fourier matrices and list decodability of random linear codes. *"SODA"*, 2012.

[CP10] E. Candes and Y. Plan. A probabilistic and ripless theory of compressed sensing. *IEEE Transactions on Information Theory*, 2010.

[CT06] E. Candes and T. Tao. Near optimal signal recovery from random projections: Universal encoding strategies. *IEEE Trans. on Info.Theory*, 2006.

[Don06] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[DR93] A. Dutt and V. Rokhlin. Fast fourier transforms for nonequispaced data. *SIAM J. Sci. Comput.*, 14(6):1368–1393, November 1993.

[GGI+02] A. Gilbert, S. Guha, P. Indyk, M. Muthukrishnan, and M. Strauss. Near-optimal sparse Fourier representations via sampling. *STOC*, 2002.

[GHI+13] Badih Ghazi, Haitham Hassanieh, Piotr Indyk, Dina Katabi, Eric Price, and Lixin Shi. Sample-optimal average-case sparse fourier transform in two dimensions. *arXiv preprint arXiv:1303.1209*, 2013.

[GLPS10] Anna C. Gilbert, Yi Li, Ely Porat, and Martin J. Strauss. Approximate sparse recovery: optimizing time and measurements. In *STOC*, pages 475–484, 2010.

[GMS05] A. Gilbert, M. Muthukrishnan, and M. Strauss. Improved time bounds for near-optimal space Fourier representations. *SPIE Conference, Wavelets*, 2005.

[HAKI12] Haitham Hassanieh, Fadel Adib, Dina Katabi, and Piotr Indyk. Faster gps via the sparse fourier transform. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 353–364. ACM, 2012.

[HIKP12a] H. Hassanieh, P. Indyk, D. Katabi, and E. Price. Simple and practical algorithm for sparse Fourier transform. *SODA*, 2012.

[HIKP12b] Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. Nearly optimal sparse fourier transform. In *STOC*, pages 563–578, 2012.

[HKPV13] Sabine Heider, Stefan Kunis, Daniel Potts, and Michael Veit. A sparse prony fft. *SAMPTA*, 2013.

[Iwe10] M. A. Iwen. Combinatorial sublinear-time Fourier algorithms. *Foundations of Computational Mathematics*, 10:303–338, 2010.

[KS01] A. Kak and M. Slaney. *Principles of Computerized Tomographic Imaging*. Society for Industrial and Applied Mathematics, 2001.

[LWC12] D. Lawlor, Y. Wang, and A. Christlieb. Adaptive sub-linear time fourier algorithms. *arXiv:1207.6368*, 2012.

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

[MEH09] Y. Matsuki, M. Eddy, and J. Herzfeld. Spectroscopy by integration of frequency and time domain information (sift) for fast acquisition of high resolution dark spectra. *J. Am. Chem. Soc.*, 2009.

[Nis10] D. Nishimura. *Principles of Magnetic Resonance Imaging*. Society for Industrial and, 2010.

[PR13] Sameer Pawar and Kannan Ramchandran. Computing a k-sparse n-length discrete fourier transform using at most 4k samples and o (k log k) complexity. *arXiv preprint arXiv:1305.0870*, 2013.

[PST01] Daniel Potts, Gabriele Steidl, and Manfred Tasche. Fast fourier transforms for nonequispaced data: A tutorial. In *Modern sampling theory*, pages 247–270. Springer, 2001.

[RV08] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *CPAM*, 61(8):1025–1171, 2008.