Sublinear-Time Algorithms for Compressive Phase Retrieval

Yi Li, Vasileios Nakos

Abstract

In the problem of compressed phase retrieval, the goal is to reconstruct a sparse or approximately k-sparse vector $x \in \mathbb{C}^n$ given access to $y = |\Phi x|$, where $|v|$ denotes the vector obtained from taking the absolute value of $v \in \mathbb{C}^n$ coordinate-wise. In this paper we present sublinear-time algorithms for a few for-each variants of the compressive phase retrieval problem which are akin to the variants considered for the classical compressive sensing problem in theoretical computer science. Our algorithms use pure combinatorial techniques and near-optimal number of measurements.

Sublinear-Time Algorithms for Compressive Phase Retrieval[1]

I. INTRODUCTION

In the past decade, the sparse recovery problem, or compressive sensing, has attracted considerable research interest with extensive applications and fruitful results. The problem asks to recover a signal $x \in \mathbb{R}^n$ (or $\mathbb{C}^n$) from linear measurements $y = \Phi x$ for some matrix $\Phi \in \mathbb{R}^{m \times n}$ (or $\mathbb{C}^{m \times n}$), assuming that x is k-sparse (i.e. containing only k non-zero coordinates) or can be well-approximated by a k-sparse signal (intuitively x contains k large coordinates while the rest of the other coordinates are small). Often post-measurement noise is present, that is, $y = \Phi x + \nu$ for some noise vector ν. The primary goal is to use as few measurements as possible. The algorithms are largely divided into two categories: geometric algorithms and combinatorial algorithms. Geometric algorithms usually use fewer measurements but run in $\operatorname{poly}(n)$ time while combinatorial algorithms run in sublinear time, usually $\mathcal{O}(k \operatorname{poly}(\log n))$ or $\mathcal{O}(k^2 \operatorname{poly}(\log n))$, at the cost of slightly more measurements.

In recent years a closely related problem, called *compressive phase retrieval*, has become an active topic, which seeks to recover a sparse signal $x \in \mathbb{R}^n$ (or $\mathbb{C}^n$) from the *phaseless measurements* $y = |\Phi x|$ (or $y = |\Phi x| + \nu$ with post-measurement noise), where $|z|$ denotes a vector formed by taking the absolute value of every coordinate of z. The primary goal remains the same, i.e. to use as fewer measurements as possible. Such types of measurements arises in various fields such as optical imaging [37] and speech signal processing [36]. There has been rich research in geometric algorithms for recovering the whole vector from phaseless measurements (see, e.g. [7], [6], [5], [10], [19], [18], [14], [39]), as well as algorithms for the sparse phase retrieval problem that run in at least polynomial time (e.g. [23], [22], [11], [38]), mostly based on semidefinite programming. On the other side, there have been relatively few sublinear time algorithms – [4], [20], [32], [29] are the only algorithms to the best of our knowledge. This comes in contrast to the standard sparse recovery problem with linear measurements, which sees a long history of sublinear-time algorithms (e.g. [17], [33], [15], [13], [16], [24], [12], [25], [28], [30], [31]). For the case of phaseless measurements, most existing algorithms consider sparse signals, and thus such sublinear time algorithms have a flavour of code design, akin to Prony's method. Among the sublinear-time algorithms, [4] considers sparse signals only, [32] considers sparse signals with random post-measurement noise, [20] allows adversarial post-measurement noise but has poor recovery guarantee, [29] considers near-sparse real signals (signals of real coordinates) with no post-measurement noise but achieves constant-factor approximation and thus outperforms all other sublinear-time algorithms for real signals. The approach in [29] employs combinatorial techniques more widely used in the theoretical computer science literature for the classical sparse recovery problem. In this paper, we aim to improve on [29] for complex near-sparse signals using similar combinatorial techniques.

More quantitatively, suppose that the decoding algorithm $\mathcal{R}$, given input $y = |\Phi x|$, outputs an approximation $\hat{x}$ to x, with the guarantee that the approximation error $d(x, \hat{x})$ is bounded from above. For $x \in \mathbb{C}^n$, the approximation error $d(x, \hat{x}) = \min_{\theta \in [0, 2\pi)} \|x - e^{i\theta}\hat{x}\|$. Specifically we consider the following two types of error guarantee:

- $(\ell_\infty/\ell_2) \min_{\theta \in [0,2\pi)} \|x - e^{i\theta}\hat{x}\|_\infty \leq \frac{1}{\sqrt{k}}\|x_{-k}\|_2$;
- $(\ell_2/\ell_2) \min_{\theta \in [0,2\pi)} \|x - e^{i\theta}\hat{x}\|_2 \leq (1+\epsilon)\|x_{-k}\|_2$,

where x_{-k} denotes the vector formed by zeroing out the largest k coordinates (in magnitude) of x. Note that when x is noiseless, that is, when $x_{-k} = 0$, both guarantees mean exact recovery of x, i.e., $\hat{x} = x$.

Besides the error guarantees, the notions of for-all and for-each in the sparse recovery problems also extend to the compressive phase retrieval problem. In a for-all problem, the measurement matrix Φ is chosen in advance and will work for all input signals x, while in a for-each problem, the measurement matrix Φ is usually random such that for each input x, a random choice of Φ works with a good probability.

In the next subsection we shall give an overview of our sublinear-time results. Our results are all for-each results. As the problem has received little attention in the community of theoretical computer science, our results are the first and preliminary along the combinatorial approach though require mild assumptions on the possible phases of the coordinates of x. Those assumptions are automatically satisfied when the possible phases, as commonly seen in applications (see, e.g., [32]), belong to a set $P \subseteq \mathbb{S}^1$ which is equidistant with a constant gap, that is, up to a rotation, $P \subseteq \{e^{2\pi i \frac{j}{m}}\}_{j=0,\dots,m-1}$ for some constant m. We leave the case of general complex signals as an open problem. A major difficulty is that for heavy hitters of large magnitude, a small error in the phase estimate could incur a lot of error in the overall approximation.

A. Results

In this section we give an overview of the sublinear-time results which we have obtained for the sparse recovery problem with phaseless measurements.

---

[1] A preliminary version of this paper appeared in the Proceedings of *International Symposium on Information (ISIT)*, 2018.

First, we consider the case of noiseless signals. Similar to the classical sparse recovery where $\mathcal{O}(k)$ measurements suffice for noiseless signals by Prony's method [35], it is known that $\mathcal{O}(k)$ phaseless measurements also suffices for exact recovery (up to rotation) and the decoding algorithm runs in time $\mathcal{O}(k \log k)$ [4]. Their algorithm is based on a multi-phase traversal of a bipartite random graph in a way such that all magnitudes and all phases are recovered by resolving multi-tons. We prove a result with the same guarantee, but our algorithm takes a different route using more basic tools and being less technically demanding. Apart from being significantly simpler, it also can be modified so that it trades the decoding time with the failure probability; see Remark 1.

Theorem 1 (noiseless signals). *There exists a randomized construction of $\Phi \in \mathbb{C}^{m \times n}$ and a deterministic decoding procedure $\mathcal{R}$ such that for any signal $x \in \mathbb{C}^n$ with $|\operatorname{supp}(x)| \leq k$, the recovered signal $\hat{x} = \mathcal{R}(\Phi, |\Phi x|)$ satisfies that $\hat{x} = e^{i\theta}x$ for some $\theta \in [0, 2\pi)$ with probability at least $1 - 1/\operatorname{poly}(k)$, where Φ has $m = \mathcal{O}(k)$ measurements and $\mathcal{R}$ runs in time $\mathcal{O}(k \log k)$.*

The next results refer to approximately sparse signals and improve upon the previous ones with various degrees. For the ℓ_∞/ℓ_2 problem our result, stated below, improves upon [29] in terms of the error guarantee and the decoding time. It requires a modest assumption on the pattern of the valid phases of the heavy hitters as defined below, which is often satisfied in applications where the valid phases lie in a set of equidistant points on $\mathbb{S}^1$. Throughout this paper we identify $\mathbb{S}^1$ with $[0, 2\pi)$ or $[-\pi, \pi)$ and assume both the non-oriented distance $d(\cdot, \cdot)$ and the oriented distance $\vec{d}(\cdot, \cdot)$ on $\mathbb{S}^1$ are circular. We shall also use $[m]$ to denote the set $\{1, \ldots, m\}$ for any positive integer m, a conventional notation in computer science literature.

Definition 1 (η-distinctness). Let $P = \{p_1, \ldots, p_m\}$ be a finite set on $\mathbb{S}^1$. We say P is η-distinct if the following conditions hold:

  (i) $d(p_i, p_j) \geq \eta$ for all distinct $i, j \in [m]$;
  (ii) it holds for every pair of distinct $i, j \in [m]$ that

$$\max_{\ell \in [m]} d(p_\ell + p_j - p_i, P) \in \{0\} \cup [\eta, \pi].$$

Intuitively, (i) means that the phases are at least η apart from each other, and (ii) means that if we rotate the set P of the valid phases to another set P' such that some valid phase coincides with another one (in the expression above p_i is rotated to the position of p_j), then either $P = P'$ or there exists an additive gap of at least η around some phase. This precludes the case where P is approximately, but not exactly, equidistant.

Definition 2 (head). Let $x \in \mathbb{C}^n$. Define $H_k(x)$ to be (a fixed choice of) the index set of the k largest coordinates of x in magnitude, breaking ties arbitrarily.

Definition 3 (ϵ-heavy hitters). Let $x \in \mathbb{C}^n$. We say x_i is an ϵ-heavy hitter if $|x_i|^2 \geq \epsilon \|x_{-1/\epsilon}\|_2^2$.

Definition 4 (phase-compliant signals). Let $x \in \mathbb{C}^n$. Let $P \subseteq \mathbb{S}^1$ be a set of possible phases and T be the set of all $(1/k)$-heavy hitters in T. We say that x is (k, P)-compliant if $\{\arg x_i : i \in T\} \subseteq P$.

Theorem 2 (ℓ_∞/ℓ_2 with optimal measurements). *There exists a randomized construction of $\Phi \in \mathbb{C}^{m \times n}$ and a deterministic decoding procedure $\mathcal{R}$ such that for $x \in \mathbb{C}^n$ which is $(\mathcal{O}(k), P)$-compliant for some η-distinct $P \subset \mathbb{S}^1$, the recovered signal $\hat{x} = \mathcal{R}(\Phi, |\Phi x|, P)$ satisfies the ℓ_∞/ℓ_2 error guarantee with probability at least $1 - 1/\operatorname{poly}(n)$, and Φ has $m = \mathcal{O}((k/\eta) \log n)$ rows and $\mathcal{R}$ runs in time $\mathcal{O}(k/\eta + k \operatorname{poly}(\log n))$.*

It is clear that the lower bound for the traditional compressive sensing problem is also a lower bound for the compressive phase retrieval problem, and it is known that the ℓ_∞/ℓ_2 compressive sensing problem requires $\Omega(k \log n)$ measurements [2]. Therefore, the theorem above achieves the optimal measurements up to a constant factor when η is a constant.

An immediate corollary of the ℓ_∞/ℓ_2 sparse recovery algorithm is an ℓ_2/ℓ_2 sparse recovery algorithm, stated below, which improves upon [29] in approximation ratio (from a constant factor to $1 + \epsilon$).

Corollary 1 (ℓ_2/ℓ_2 with near-optimal measurements). *There exists a randomized construction of $\Phi \in \mathbb{R}^{m \times n}$ and a deterministic decoding procedure $\mathcal{R}$ such that for $x \in \mathbb{C}^n$ which is $(\mathcal{O}(k), P)$-compliant for some η-distinct $P \subset \mathbb{S}^1$, the recovered signal $\hat{x} = \mathcal{R}(\Phi, |\Phi x|, P)$ satisfies the ℓ_2/ℓ_2 error guarantee with probability at least $1 - 1/\operatorname{poly}(n)$, and Φ has $m = \mathcal{O}((k/\min\{\eta, \epsilon\}) \log n)$ rows and $\mathcal{R}$ runs in time $\mathcal{O}((k/\min\{\eta, \epsilon\}) \operatorname{poly}(\log n))$.*

It is also known that the classical compressive sensing problem with for-each ℓ_2/ℓ_2 error guarantee and constant failure probability requires $\Omega((k/\epsilon) \log(n/k))$ measurements [34]. Our result above achieves the optimal number of measurements up to a logarithmic factor.

II. TOOLKIT

The first two results concern heavy hitters, one for estimating the value of a heavy hitter and the other for finding the positions of the heavy hitters.

Theorem 3 (COUNT-SKETCH, [8])**.** *There exist a randomized construction of a matrix $\Phi \in \mathbb{R}^{m \times n}$ with $m = \mathcal{O}(K \log n)$ and a deterministic algorithm $\mathcal{R}$ such that given $y = |\Phi x|$ for $x \in \mathbb{C}^n$, with probability at least $1 - 1/\operatorname{poly}(n)$, for every $i \in [n]$, the algorithm $\mathcal{R}$ returns in time $\mathcal{O}(\log n)$ an estimate $|\hat{x}_i|$ such that*

$$||x_i| - |\hat{x}_i||^2 \leq \frac{1}{K}\|x_{-K}\|_2^2.$$

Theorem 4 (Heavy hitters, [27])**.** *There exist a randomized construction of a matrix $\Phi \in \mathbb{R}^{m \times n}$ with $m = \mathcal{O}(K \log n)$ and a deterministic algorithm $\mathcal{R}$ such that given $y = |\Phi x|$ for $x \in \mathbb{C}^n$, with probability at least $1 - 1/\operatorname{poly}(n)$ the algorithm $\mathcal{R}$ returns in time $\mathcal{O}(K \cdot \operatorname{poly}(\log n))$ a set S of size $\mathcal{O}(K)$ containing all $(1/K)$-heavy hitters of x.*

We remark that the paper [27] does not consider complex signals but the extension to complex signals is straightforward. The algorithm is not designed for the phaseless sparse recovery either, the identification algorithm nevertheless works when the measurements are phaseless because it only relies on the magnitudes of the bucket measurements; see Theorem 2 and Section B in [27]. Estimating the values of the candidate coordinates requires knowing the phases of the measurements but our theorem above does not concern this part.

Theorem 5 ([9])**.** *Let V be a set of n vertices. There exists an absolute constant κ such that $\kappa n \log n$ uniform samples of pairs of distinct vertices in V induce a connected graph with probability at least $1 - 1/\operatorname{poly}(n)$.*

The following lemmata will be crucial in the analysis of our algorithms.

Lemma 1. *Let $\theta_0 \in (0, \pi/12)$ be a constant. There exist constants $\epsilon_0, c_0, c > 0$ such that the following holds. Suppose that $x, y, n_1, n_2, n_3 \in \mathbb{C}$ such that $|n_1|, |n_2|, |n_3| \leq \epsilon \min\{|x|, |y|\}$ for some $\epsilon \leq \epsilon_0$. Denote by θ be the phase difference between x and y. Then given the norms*

$$|x + n_1|, |y + n_2|, |x + y + n_1 + n_2 + n_3|,$$

we can recover $|\theta|$ up to an additive error of $c_0 \sqrt{\epsilon}$. Furthermore, we can recover $|\theta|$ up to an additive error of $c\epsilon$ whenever $|\theta| \in (\theta_0, \pi - \theta_0)$.

Proof. If we know $|x|$, $|y|$ and $|x + y|$, it follows from the Law of Cosines that

$$\cos(\pi - |\theta|) = \frac{|x|^2 + |y|^2 - |x+y|^2}{2|x| \cdot |y|} = \frac{-\Re x \bar{y}}{|x| \cdot |y|}.$$

Let $x' = x + n_1$ and $y' = y + n_2$ then $x + y + n_1 + n_2 + n_3 = x' + y' + n_3$. Suppose the phase difference between x' and y' is θ', then we would pretend $x' + y' + n_3$ to be $x' + y'$. We compute

$$\xi = \frac{|x'|^2 + |y'|^2 - |x' + y' + n_3|^2}{2|x'| \cdot |y'|} = \frac{-\Re x'\overline{y'} - \Re x'\overline{n_3} - \Re y\overline{n_3} - |n_3|^2}{|x'| \cdot |y'|}.$$

Hence

$$|\xi - \cos(\pi - |\theta'|)| \leq \frac{|x'||n_3| + |y'||n_3| + |n_3|^2}{|x'| \cdot |y'|} \leq \frac{\epsilon}{1 - \epsilon} + \frac{\epsilon}{1 - \epsilon} + \frac{\epsilon^2}{1 - \epsilon^2} \leq 3\epsilon,$$

provided that $\epsilon \leq 1/4$.

Similarly we have

$$\cos(\pi - |\theta'|) = \cos(\pi - |\theta|) \cdot \frac{|x|}{|x + n_1|} \cdot \frac{|y|}{|y + n_2|} + \nu, \quad |\nu| \leq 3\epsilon,$$

and thus

$$\cos(\pi - |\theta'|) - \cos(\pi - |\theta|) = \cos\theta \left(\frac{|x|}{|x + n_1|} \cdot \frac{|y|}{|y + n_2|} - 1 \right) + \nu.$$

Note that $\frac{|x|}{|x+n_1|}, \frac{|y|}{|y+n_2|} \in [\frac{1}{1+\epsilon}, \frac{1}{1-\epsilon}]$, it follows that

$$|\cos(\pi - |\theta'|) - \cos(\pi - |\theta|)| \leq c_1 \epsilon$$

for some absolute constant $c_1 > 0$, and thus

$$|\xi - \cos(\pi - |\theta|)| \leq (c_1 + 3)\epsilon.$$

Therefore we can estimate $|\theta|$ up to an additive error of $c_0 \sqrt{\epsilon}$ for some absolute constant $c_0 > 0$; and furthermore, there exists an absolute constant $c > 0$ such that when $\theta_0 = \pi/9$, ϵ is small enough and $|\theta| \in (\theta_0, \pi - \theta_0)$, it holds that

$$|\arccos\xi - |\theta|| \leq c\epsilon. \qquad \square$$

Lemma 2. *Let $x, y, n_1, n_3, \epsilon, \theta_0, c$ be as in Lemma 1 and further assume that $\epsilon < \theta_0/(2c)$. Suppose that $\arg y = \arg x + \theta$ for some $\theta \in [0, 2\pi)$, where addition is modulo 2π. Given the norms*

$$|x + n_1|, |y|, |x + y + n_1 + n_3|, |x + \beta y + n_1 + n_3|, \quad \beta = e^{i\theta_0},$$

we can recover θ up to an additive error of $c\epsilon$, provided that $\theta \in (2\theta_0, \pi - 2\theta_0) \cup (\pi + 2\theta_0, 2\pi - 2\theta_0)$.

Proof. By Lemma 1, we can recover $|\theta|$ up to an additive error of $c\epsilon$ when $|\theta| \in (\theta_0, \pi - \theta_0)$. To determine the sign, we rotate y by angle θ_0 and test the angle between x and this rotated y again by Lemma 1. Suppose that the angle between x and βy is ϕ and we have an estimate of $|\phi|$ up to an additive error of $c\epsilon$, provided that $|\phi| \in (\theta_0, \pi - \theta_0)$, which is satisfied when $|\theta| \in (2\theta_0, \pi - 2\theta_0)$. It holds in this case that

$$|\phi| - |\theta| = \begin{cases} \theta_0, & \theta > 0; \\ -\theta_0, & \theta < 0. \end{cases}$$

The left-hand side is approximated up to an additive error of $2c\epsilon$ and thus we can distinguish the two cases. □

Lemma 3 (relative phase estimate). *Let $x, y, n_1, n_3, \epsilon, c_0, \theta_0, c$ be as in Lemma 2 and further assume that $\epsilon < \min\{\theta_0/(2c), (\theta_0/(2c_0))^2\}$. Suppose that $\arg y = \arg x + \theta$ for some $\theta \in [0, 2\pi)$, where addition is modulo 2π. Given the norms*

$$|x + n_1|, |y|, |x + e^{i(\theta_0 j + \frac{\pi}{2}\ell)} y + n_1 + n_3|, \quad j, \ell = 0, 1$$

we can recover θ up to an additive error of $c\epsilon$.

Proof. From Lemma 2, we know that we can recover θ up to an additive error of $c\epsilon$ when $\theta \in I$, where $I = (2\theta_0, \pi - 2\theta_0) \cup (\pi + 2\theta_0, 2\pi - 2\theta_0)$. We accept the estimate if the estimate is in the range of $I' := (3\theta_0, \pi - 3\theta_0) \cup (\pi + 3\theta_0, 2\pi - 3\theta_0)$. It follows from the assumption of ϵ that if we ever accept the estimate, $|\theta|$ must be contained in $(\frac{5}{2}\theta_0, \pi - \frac{5}{2}\theta_0)$ and we have therefore an $c\epsilon$ additive error.

Consider the phase difference between x and $e^{i\pi/2} y$ and suppose that $\arg(e^{i\pi/2} y) = \arg x + \phi$, then we can recover ϕ up to an additive error of $c\epsilon$ for $\phi \in I$, that is, for $\theta \in J := (\pi/2 + 2\theta_0, 3\pi/2 - 2\theta_0) \cup (-\pi/2 + 2\theta_0, \pi/2 - 2\theta_0)$, which is I rotated by $\pi/2$. We accept the estimate when it is in the range of $J' := (\pi/2 + 3\theta_0, 3\pi/2 - 3\theta_0) \cup (-\pi/2 + 3\theta_0, \pi/2 - 3\theta_0)$.

Note that $I' \cup J'$ covers the whole $\mathbb{S}^1$ when $\theta_0 < \pi/12$. □

III. NOISELESS SIGNALS

We shall need the following theorem from [29], which shows that one can recover an exactly K-sparse signal up to a global phase using $\mathcal{O}(K)$ measurements and in time $\mathcal{O}(K^2)$.[2]

Theorem 6 ([29]). *There exists a matrix $M \in \mathbb{C}^{(6k-2) \times n}$, such that given given $y = |Mx|$ for $x \in \mathbb{C}^n$ such that $\|x\|_0 \leq K$ we can recover x up to a rotation in time $\mathcal{O}(K^2)$.*

We are now ready to prove Theorem 1, which we restate below.

Theorem 1 (noiseless signals). *There exists a randomized construction of $\Phi \in \mathbb{C}^{m \times n}$ and a deterministic decoding procedure $\mathcal{R}$ such that for any signal $x \in \mathbb{C}^n$ with $|\mathrm{supp}(x)| \leq k$, the recovered signal $\hat{x} = \mathcal{R}(\Phi, |\Phi x|)$ satisfies that $\hat{x} = e^{i\theta} x$ for some $\theta \in [0, 2\pi)$ with probability at least $1 - 1/\mathrm{poly}(k)$, where Φ has $m = \mathcal{O}(k)$ measurements and $\mathcal{R}$ runs in time $\mathcal{O}(k \log k)$.*

Let us first give some intuition on the algorithm and the proof. We hash all n coordinates to $\Theta(k/\log k)$ buckets, and in each bucket we apply the algorithm in Theorem 6 with $K = O(\log k)$. By standard concetration bounds there exist at most $O(\log k)$ coordinates of $\mathrm{supp}(x)$ in each bucket, and hence, conditioned on this event, we can find them deterministically along with their relative phases. Next we find all the relative phases by finding the relative phases "between buckets", which means finding the relative phases between representatives from each bucket. To that end, we build a random graph with $O(k/\log k)$ buckets and $O(k)$ edges, where each vertex represents a bucket and each edge has a label that indicates the relative phase between its two endpoints. If the graph is connected, a standard traversal will give the desired result. We show how to sample random edges from this graph by performing a careful subsampling on the vector x. The number of measurements and the running time follows by standard results in random graph theory. Below is the formal proof.

Proof. Let $B = k/(c \log k)$ and $h : [n] \to [B]$ a random hash function, where c is a constant. We hash all n coordinates into B buckets using h. It is a typical application of Chernoff bound that the buckets have small size, more specifically, for some constant $\bar{c} > c$,

$$\Pr\left\{\exists j \in [B] : |h^{-1}(j) \cap \mathrm{supp}(x)| > \bar{c} \log k\right\} \leq \frac{1}{\mathrm{poly}(k)}. \tag{1}$$

---

[2]The runtime is dominated by Prony's method, which can be implemented in $O(K^2)$ time using a Lanczos process (see, e.g. [3]) and solving a linear Vandermonde system (see, e.g. [1]). A detailed discussion can be found in the arXiv version of [29].

To see this, for $i \in [n]$ and $j \in [B]$, let $X_{i,j}$ be the indicator variable of the event $h(i) = j$. Then $\mathbb{E}X_{i,j} = 1/B$, and thus $\mathbb{E}\sum_{i \in \mathrm{supp}(x)} X_{i,j} = k/B = \bar{c} \log k$. Note that $X_{i,j}$ are negatively associated, thus the Chernoff bound can be applied, which, with appropriate constants, yields that for each $j \in [B]$,

$$\Pr\left\{|h^{-1}(j) \cap \mathrm{supp}(x)| > 5 \log k\right\} = \Pr\left\{\sum_{i \in \mathrm{supp}(x)} X_{i,j} > \bar{c} \log k\right\} \leq \frac{1}{\mathrm{poly}(k)}.$$

Take a union bound over all $j \in [B]$ gives (1).

In each bucket we run the algorithm of Theorem 6 with $K = \bar{c} \log k$. The number of measurements used for each bucket is $\Theta(\log k)$. The failure probability is $2^{-\Theta(\log k)}$, and this allows us to take a union-bound to conclude correct execution over all B buckets. For each $j \in [B]$, we can find $x_{h^{-1}(j)}$ up to a global phase. We set $\mathrm{supp}(x) = \bigcup_{j=1}^{B} \mathrm{supp}(x_{h^{-1}(j)})$ and, for notational convenience, let $s = |\mathrm{supp}(x)|$ and ℓ be such that $2^{\ell-1} < s \leq 2^{\ell}$.

Now, let $B' = 2^{\ell}/(c\ell)$ and let $h' : [n] \to [B']$ be a random hash function. We hash all n coordinates into B' coordinates. Similar to the above, with probability $\geq 1 - 2^{-\Omega(\ell)}$ every bucket contains at most $\bar{c}\ell$ nonzero coordinates of x. In each bucket, we run the algorithm of Theorem 6 with $K = \bar{c}\ell$, so that in $\mathcal{O}(\ell)$ measurements we can find $x_{(h')^{-1}(j)}$ up to a global phase for each $j \in [B']$, so it remains to find the relative phases across different $x_{(h')^{-1}(j)}$.

Let F_{ℓ} be a matrix of n columns and $2\alpha\kappa 2^{\ell}$ rows, where α is a sufficiently large constants and κ is the same constant in Theorem 5. The rows are split into $\alpha\kappa 2^{\ell}$ groups, each has 2 rows. For each group j, the first row $(F_{\ell})_{2j-1}$ has independent random $\{0,1\}$-entries such that $\mathbb{E}[(F_{\ell})_{2j-1,t}] = 2^{-\ell}$. The second row has the form $(F_{\ell})_{2j,t} = \rho_{j,t} \cdot (F_{\ell})_{2j-1,t}$, where $\{\rho_{j,t}\}$ are independent uniform random variables in $\{1, i\}$.

Next we describe how to recover the relative phases across different $x_{(h')^{-1}(j)}$. Define a row index set $\hat{J}$ to be

$$\hat{J} = \left\{j : \left|\mathrm{supp}((F_{\ell})_{5(j-1)+1}) \cap \mathrm{supp}(x)\right| = 2\right\}.$$

Observe that for each j,

$$\Pr\{j \in \hat{J}\} = \binom{s}{2} \frac{1}{2^{2\ell}} \left(1 - \frac{1}{2^{\ell}}\right)^{s-2} \geq \frac{1}{2} \left(\frac{s-1}{2^{\ell}}\right)^2 \left(1 - \frac{1}{s}\right)^{s-2} \geq \frac{1}{2} \left(\frac{1}{2}\right)^2 \frac{1}{e},$$

that is, each j is contained in $\hat{J}$ with probability at least an absolute constant. For each such j, let $\{u_j, v_j\} = \mathrm{supp}((F_{\ell})_{2j-1}) \cap \mathrm{supp}(x)$. It is clear that $\rho_{j,u_j} \neq \rho_{j,v_j}$ with probability $1/2$. When this event happens, we say j is good. We have shown that each j is good with probability at least an absolute constant. Let $J \subseteq \hat{J}$ denote the set of good j's.

We shall focus on the measurements corresponding to the groups in J. From each $j \in J$ we can obtain a random pair $\{u_j, v_j\} \subseteq \mathrm{supp}(x)$ and, moreover, $(h'(u_j), h'(v_j))$ is uniformly random on $[B'] \times [B'] \setminus \{(b,b), b \in [B']\}$ (in other words, such a measurement corresponds to a random pair of buckets). Since j is good, we obtain both $|x_{u_j} + x_{v_j}|$ and $|x_{u_j} + ix_{v_j}|$. Since we also know $|x_{u_j}|$ and $|x_{v_j}|$, we can infer the relative phase between x_u, x_v. The relative phases we obtain are always correct since the signal is noiseless. Let $\mathcal{M}$ be the ordered set of such pairs (u_j, v_j) along with the label that we obtain about the relative phase between u_j and v_j. By taking α to be big enough, we have $|\mathcal{M}| = |J| \geq \kappa B' \log B'$ (since $B' \log B' = \Theta(2^{\ell})$) with probability at least $1 - 2^{-\Omega(|J|)}$. We run a depth-first search on $\mathcal{M}$ to infer the relative phases. We infer all the relative phases correctly when the $\mathcal{M}$ is connected, which, by Theorem 5, happens with probability at least $1 - 1/\mathrm{poly}(B') \geq 1 - 2^{-\Omega(\ell)}$.

So far we have shown that with probability at least $1 - 2^{-\Omega(\ell)}$ we can recover $\hat{x}$ up to a global phase difference. To improve the success probability to $1 - 1/\mathrm{poly}(k)$, we can repeat the procedure above by $\Theta(1)$ times for $\ell > \lceil \frac{1}{2} \log k \rceil$ and $\Theta(\log k)$ times for $\ell \leq \lceil \frac{1}{2} \log k \rceil$ and take the relative phase pattern that appears most often.

Number of Measurements: The total number of measurements is therefore (taking a union over all possible values of $\ell = 1, 2, \ldots, \lceil \log k \rceil$)

$$\frac{k}{c \log k} \cdot \Theta(\log k) + \sum_{\ell > \lceil \frac{1}{2} \log k \rceil} \left(\frac{2^{\ell}}{c\ell} \cdot \Theta(\ell) + \alpha\kappa 2^{\ell}\right) \cdot \Theta(1) + \sum_{\ell \leq \lceil \frac{1}{2} \log k \rceil} \left(\frac{2^{\ell}}{c\ell} \cdot \Theta(\ell) + \alpha\kappa 2^{\ell}\right) \cdot \Theta(\log k) = \mathcal{O}(k)$$

as desired.

Runtime: Recovering the in-bucket signals takes time $\mathcal{O}(B \log^2 k(\cdot \mathrm{poly}(\log \log k))) = \mathcal{O}(k \log k \cdot \mathrm{poly}(\log \log k))$. Finding the set J takes time $\mathcal{O}(k)$ (see Remark 2). Thus, it takes time $\mathcal{O}(k + B' \log B') = \mathcal{O}(k)$ to build a pattern of relative phase differences and $\mathcal{O}(k \log k)$ time to build all patterns[3]. Finding the most frequent pattern can be implemented with a sorting followed by a linear scan. There are $\Theta(\log k)$ patterns and the comparison of two patterns takes time $\mathcal{O}(B)$. Hence finding the most frequent pattern takes time $\mathcal{O}(\log k \log \log k \cdot B + B \log k) = \mathcal{O}(k \log \log k)$. The overall runtime is therefore $\mathcal{O}(k \log k)$. $\square$

---

[3]In fact, building the patterns can be done in $O(k)$ time, since the $\log k$ factor due to the number of repetitions occurs only for "small" ℓ, that is, $\ell \leq \lceil \frac{1}{2} \log k \rceil$. Nevertheless, this will not make any difference in our final running time.

Remark 1. Note that if we hash to $k^{1-\alpha}$ buckets, solve in each bucket and then combine the buckets, we can obtain a failure probability at most $\exp(-k^\alpha)$ and a running time of $\mathcal{O}(k^{1+\alpha})$. This is a trade-off between decoding time and failure probability that the previous algorithms did not achieve.

Remark 2. We show how to implement efficiently the routine which finds the set of rows of F whose support intersects $\operatorname{supp}(x)$ at exactly two coordinates. For $\ell > \lceil \frac{1}{2} \log k \rceil$, in each repetition, the expected number of rows of F_ℓ containing an index $i \in \operatorname{supp}(x)$ is $2^{-\ell}\alpha\kappa 2^\ell = \alpha\kappa$. So the probability that there are more than $2\alpha\kappa 2^\ell$ pairs (i, q) such that $i \in \operatorname{supp}(x) \cap \operatorname{supp}(F_\ell)_q$ is $\exp(-\Omega(2^\ell)) < 1/\operatorname{poly}(k)$. A similar result can be obtained for $\ell \leq \lceil \frac{1}{2} \log k \rceil$. Suppose that F_ℓ is stored using n lists of nonzero coordinates in each column, we can afford to iterate over all such pairs (i, q), keep an array $C[q]$ that holds the cardinality of $\operatorname{supp}(x) \cap \operatorname{supp}((F_\ell)_q)$. At the end, we find the values of q with $C[q] = 2$. This implementation makes the algorithm run in $\mathcal{O}(k \log k)$ time.

IV. ℓ_∞/ℓ_2 ALGORITHM

We first give a high-level overview of our algorithm. Invoking Theorems 3 and 4, we can find a set S of size $O(k)$ that contains all heavy hitters. It remains to find their relative phases. Our approach is very different from the previous section; we shall exploit the presence of the rest $n - O(k)$ coordinates to find the relative phases among the heavy hitters. We downsample the signal at the rate of $\Theta(1/k)$, and combine with gaussians to form $L = \sum_{i \in T} g_i x_i$, where T is the support of the downsampled signal. With constant probability, $T \cap S = \emptyset$ (note that we do not know S when designing the measurements, we only know T). We then hash every coordinate in $[n] \setminus T$ to $\Theta(k)$ buckets, combine with random rotations and add L to each bucket. We repeat with fresh randomness $\Theta(\log n)$ times. Then, for every $i \in S$ we can find the buckets in which it is isolated from $S \setminus \{i\}$ and its relative phase with respect to L (note that L is present in every bucket). This allows us to find the relative phase among all $i \in S$. The risk is that L might be 0, in which case the argument above would not work. In this case, however, T would be an empty set, which in turn implies that x would be $O(k)$-sparse, and hence we could run the algorithm from the previous section.

The remaining of the section is devoted to the details of our algorithm and its analysis.

We set ϵ_0, c_0, c to be the constants in Lemma 3 and $\epsilon = \min\{\epsilon_0, \eta/(5c), \pi/(25c), \pi^2/(145c_0^2)\}$. Let $P \subseteq \mathbb{S}^1$ be η-distinct and suppose that it contains the phases of all $1/(\tilde{C}k)$-heavy hitters for some (large) constant $\tilde{C}$.

We first describe our construction of the measurement matrix Φ and then present the analysis and the recovery algorithm. Let $R = c_R \log n$ for some constant c_R to be determined. The overall sensing matrix Φ is a layered one as

$$\Phi = \begin{bmatrix} \Phi_{\mathrm{HH}} \\ \Phi_{\mathrm{CS}} \\ \Phi_{\mathrm{L}} \\ \Phi_1 \\ \vdots \\ \Phi_R \end{bmatrix}.$$

Here
- Φ_{HH} is the sensing matrix in Theorem 4 with $K = k$.
- Φ_{CS} is the sensing matrix of COUNT-SKETCH with $K = Ck/\epsilon$.
- Φ_{L} is an $R \times N$ matrix, the r-th row ($r \in [R]$) of which is

$$\rho_r = \begin{pmatrix} \eta_{r,1}g_{r,1} & \eta_{r,2}g_{r,2} & \cdots & \eta_{r,n}g_{r,n} \end{pmatrix},$$

  where $\eta_{r,i}$ are i.i.d. Bernoulli variables with $\mathbb{E}\,\eta_{r,i} = 1/(C_0 k)$ and $g_{r,i}$ are i.i.d. $\mathcal{N}(0,1)$ variables.
- Each Φ_r ($r \in [R]$) is a matrix of $4B$ rows defined as follows, where $B = C_B k/\epsilon$. Let $h_r : [n] \to [B]$ be a random hash function and $\{\sigma_{r,i}\}_{i=1}^n$ be random signs. Define a $B \times n$ hashing matrix H_r as

$$(H_r)_{j,i} = \begin{cases} (1 - \eta_{r,i})\sigma_{r,i}, & i \in h_r^{-1}(j); \\ 0, & \text{otherwise.} \end{cases}$$

  The $4B$ rows of Φ_r are defined to be

$$e^{i(\theta_0\ell_1 + \frac{\pi}{2}\ell_2)}\rho_r + (H_r)_{b,\cdot}, \qquad \ell_1, \ell_2 = 0, 1, \quad b = 1, \ldots, B.$$

We present the recovery algorithm in Algorithm 1, where we assume that the set P of valid phases have been sorted. Next we analyse the algorithm in four steps as follows.

Algorithm 1 Algorithm for the ℓ_∞/ℓ_2 phaseless sparse recovery. Assume that the elements in P are sorted.

1: $S \leftarrow$ the set returned by the algorithm in Theorem 4 with $K = k$
2: Run a COUNT-SKETCH algorithm with $K = Ck/\epsilon$ to obtain an approximation $|\hat{x}_i|$ to $|x_i|$ for all $i \in S$
3: $L_r \leftarrow |\sum_{i=1}^n \eta_{r,i} g_{r,i} x_i|$ for all $r \in [R]$
4: **if** $L_1 = 0$ **then**
5:     Run the algorithm for the noiseless case with sparsity $C_2 k$
6: **else**
7:     **for** each $r \in [R]$ **do**
8:         $S'_r \leftarrow \{i \in S : |\hat{x}_i| \geq L_r\}$
9:         $b_{r,i} \leftarrow h_r(i)$ for all $i \in S'_r$
10:         $I_r \leftarrow \{i \in S'_r : b_{r,i} \neq b_{r,j} \text{ for all } j \in S'_r \setminus \{i\}\}$
11:         **for** each $i \in I_r$ **do**
12:             $\tilde{\theta}_{r,i} \leftarrow$ estimate of phase difference between x_i and $\langle \rho_r, x \rangle$ using Lemma 3
13:     $S'' \leftarrow \bigcap_r S'_r$
14:     Choose an arbitrary $i_0 \in S''$
15:     **for** each $i \in S'' \setminus \{i_0\}$ **do**
16:         $\theta'_{i_0,i} = \text{median}_{r:i,i_0 \in I_r}(\tilde{\theta}_{r,i} - \tilde{\theta}_{r,i_0})$
17:     **for** each $p \in P$ **do**
18:         $\theta'_{i_0} \leftarrow p$
19:         $\theta'_i \leftarrow \theta'_{i_0} + \theta'_{i_0,i}$ for all $i \in S'' \setminus \{i_0\}$
20:         **if** $d(\theta'_i, P) \leq \eta/2$ for all $i \in S''$ **then**
21:             **return** $\hat{x}$ supported on S'' with $\arg \hat{x}_i = \theta_i$, where θ_i is the rounded value of θ'_i to P
22:         **end if**
23: **end if**

---

    a) Step 1: By Theorem 4, the set S has size $\mathcal{O}(k)$ and, with probability $1 - 1/\text{poly}(n)$, contains all $(1/k)$-heavy hitters. The COUNT-SKETCH (Theorem 3) guarantees that

$$\left||x_i| - |\hat{x}_i|\right|^2 \leq \frac{\epsilon}{Ck}\|x_{-Ck/\epsilon}\|_2^2 \tag{2}$$

for all $i \in S$ with probability at least $1 - 1/\text{poly}(n)$.

    b) Step 2: Fix a repetition $r \in [R]$. We shall see that L_r, calculated in Line 3, 'approximates' the desirable tail $\frac{1}{k}\|x_{-k}\|_2^2$. For notational simplicity, we omit r in the subscript and write $L_{r,i}$ as L, $\eta_{r,i}$ as η_i, etc., since we have fixed the repetition index r.

    First we upper bound L. Decompose x into real and imaginary parts as $x = y + iz$ with $y, z \in \mathbb{R}^n$ and consider $L_1 = \sum_i \eta_i g_i y_i$ and $L_2 = \sum_i \eta_i g_i z_i$. Note that $L^2 = L_1^2 + L_2^2$.

    Choosing $C_0 \geq 200$, we have

$$\Pr\{\eta_i = 0 \text{ for all } i \in H_k(y) \cup H_k(z)\} \geq 0.99 \tag{3}$$

Condition on this event below. Note that $L_1 \sim \mathcal{N}(0, |\sum_i \eta_i y_i|_2^2)$, and

$$\Pr\left\{ L_1^2 \geq 2.282^2 \left|\sum_i \eta_i y_i\right|_2^2 \right\} \leq 1 - 2F(2.282) \leq 0.0225,$$

where $F(t)$ denotes the cumulative distribution function of the standard normal distribution.

    On the other hand, $\mathbb{E}|\sum_i \eta_i y_i|^2 = \sum_i (\mathbb{E}\,\eta_i) y_i^2 \leq \|y_{-k}\|_2^2/(C_0 k)$ thus

$$\Pr\left\{ \left|\sum_i \eta_i y_i\right|^2 \geq \frac{20}{C_0 k}\|x_{-k}\|_2^2 \right\} \leq 0.05,$$

and hence

$$\Pr\left\{ L_1^2 \geq \frac{105}{C_0 k}\|y_{-k}\|_2^2 \right\} \leq 0.0725.$$

Similarly we have

$$\Pr\left\{ L_2^2 \geq \frac{105}{C_0 k}\|z_{-k}\|_2^2 \right\} \leq 0.0725.$$

Therefore, taking a union bound of both events above and noting that $\|y_{-k}\|_2^2 + \|z_{-k}\|_2^2 \le \|x_{-k}\|_2^2$, we have that

$$\Pr\left\{ L^2 \ge \frac{105}{C_0 k}\|x_{-k}\|_2^2 \right\} \le 0.145. \tag{4}$$

We therefore obtained an upper bound of L. The next lemma lower bounds L.

Lemma 4. *With probability at least* 0.8*, it holds that* $L^2 \ge \frac{1}{C_1 k}\|x_{-C_2 k}\|_2^2$*, where* C_1, C_2 *are absolute constants.*

Proof. Decompose x into real and imaginary parts as $x = y + iz$ with $y, z \in \mathbb{R}^n$. Consider $L_1 = \sum_i \eta_i g_i y_i$ and $L_2 = \sum_i \eta_i g_i z_i$, which are both real. Note that $L^2 = L_1^2 + L_2^2$.

First consider L_1. We sort coordinates $[n] \setminus H_k(y)$ by decreasing order of magnitude. Then, we split the sorted coordinates into continuous blocks of size $C_2' k$ and let S_j denote the j-th block. Let δ_j be the indicator variable of the event that there exists $i \in S_j$ such that $\eta_i = 1$, then δ_j's are i.i.d. Bernoulli variables with $\mathbb{E}\,\delta_j = 1 - (1 - 1/(C_0 k))^{C_2' k} \ge 1 - \exp(-C_2'/C_0)$, which can be made arbitrary close to 1 by choosing C_2' big enough. It is a standard result in random walks (see, e.g., [26, p67]) that when $\mathbb{E}\,\delta_j$ is big enough, with probability at least 0.95, every partial prefix of the 0/1 sequence $(\delta_1, \delta_2, \delta_3, \dots)$ will have more 1s than 0s. Call this event $\mathcal{E}$. In fact, one can directly calculate that $\Pr(\mathcal{E}) = 1 - (1-p)^2/p$ when $p := \mathbb{E}\,\delta_j \ge 1/2$, and thus one can take $C_2' = \lceil 1.61 C_0 \rceil$ such that $\Pr(\mathcal{E}) \ge 0.95$.

Condition on $\mathcal{E}$. We can then define an injective function π from $\{j : \delta_j = 0\}$ to $\{j : \delta_j = 1\}$. Specifically, we define $\pi(j) = \ell$, where δ_j is the k-th 0 in the sequence and ℓ is the k-th 1 in the sequence. Clearly π is injective, $\pi(j) < j$ and $\delta_{\pi(j)} = 1$. It follows that

$$
\begin{aligned}
\sum_{i \in [n]} \eta_i |y_i|^2 &\ge \sum_j \delta_j \|S_{j+1}\|_\infty^2 \ge \sum_{j : \delta_j = 1} \frac{1}{C_2' k}\|S_{j+1}\|_2^2 \\
&\ge \frac{1}{2}\sum_{j : \delta_j = 1} \frac{1}{C_2' k}\|S_{j+1}\|_2^2 + \frac{1}{2}\sum_{\substack{j : \delta_j = 1 \\ \pi^{-1}(j) \text{ exists}}} \frac{1}{C_2' k}\|S_{\pi^{-1}(j)}\|_2^2 \\
&\ge \frac{1}{2}\sum_{j : \delta_j = 1} \frac{1}{C_2' k}\|S_{j+1}\|_2^2 + \frac{1}{2}\sum_{j : \delta_j = 0} \frac{1}{C_2' k}\|S_j\|_2^2 \\
&\ge \frac{1}{2 C_2' k}\|y_{-C_2' k}\|_2^2.
\end{aligned}
$$

This implies that $L_1 = \sum_i \eta_i g_i y_i \sim \mathcal{N}(0, \sum_i \eta_i |y_i|^2)$ with probability at least 0.95 will stochastically dominate a gaussian variable $\mathcal{N}(0, \frac{1}{2 C_2 k}\|y_{-C_2 k}\|_2^2)$. Combining with the fact that $\Pr_{g \sim \mathcal{N}(0,1)}\{|g| \le \frac{1}{16}\} \le 0.05$, we see that

$$\Pr\left\{ L_1^2 \ge \frac{1}{16^2 \cdot 2 C_2 k}\|y_{-C_2' k}\|_2^2 \right\} \ge 0.9.$$

Similarly for the imaginary part z and L_2,

$$\Pr\left\{ L_2^2 \ge \frac{1}{16^2 \cdot 2 C_2 k}\|z_{-C_2' k}\|_2^2 \right\} \ge 0.9.$$

Condition on that both events above happen. For notational convenience, let $T_1 = H_{C_2' k}(y)$ and $T_2 = H_{C_2' k}(z)$, then

$$
\begin{aligned}
L^2 = L_1^2 + L_2^2 &\ge \frac{1}{16^2 \cdot 2 C_2' k}\left(\|y_{T_1^c}\|_2^2 + \|z_{T_2^c}\|_2^2\right) \\
&\ge \frac{1}{16^2 \cdot 2 C_2' k}\|y_{(T_1 \cup T_2)^c}\|_2^2 + \|z_{(T_1 \cup T_2)^c}\|_2^2 \\
&= \frac{1}{16^2 \cdot 2 C_2' k}\|x_{(T_1 \cup T_2)^c}\|_2^2 \\
&\ge \frac{1}{16^2 \cdot 2 C_2' k}\|x_{-2 C_2' k}\|_2^2.
\end{aligned}
$$

Therefore, we can take $C_2 = 2 C_2'$ above and $C_1 = 16^2 C_2$. $\square$

Combining (3), (4) and Lemma 4 and taking $C_0 = 210$, we conclude that with probability at least $1 - 0.365$,

$$\frac{1}{C_1 k}\|x_{-C_2 k}\|_2^2 \le L^2 \le \frac{1}{2k}\|x_{-k}\|_2^2. \tag{5}$$

c) Step 3: We keep the repetition r. We now show that the trimmed set S' is good in the sense that its elements are not too small and it contains all $(1/k)$-heavy hitters. This is formalized in the following lemma.

Lemma 5. *Suppose that event* (2) *happens. With probability at least* 0.635*, it holds that*

(i) $|x_i|^2 \geq \frac{1}{2C_1 k} \|x_{-C_2 k}\|_2^2$ *for all* $i \in S'$*; and*

(ii) S' *contains all coordinates* i *such that* $|x_i|^2 \geq \frac{1}{k} \|x_{-k}\|_2^2$*.*

Proof. The events (3) and (5) happen simultaneously with probability at least $1 - 0.365$. Condition on both events. Let $C = \frac{\sqrt{2}}{\sqrt{2}-1} C_1$. By our choice of constants, $C \geq C_1 \geq C_2$, then

(i) for $i \in S'$, it holds that $|x_i| \geq L - \frac{1}{\sqrt{Ck}} \|x_{-C_2 k}\|_2 \geq \frac{1}{\sqrt{2C_1 k}} \|x_{-C_2 k}\|_2$;

(ii) if $|x_i|^2 \geq \frac{1}{k} \|x_{-k}\|_2^2$, then $|\hat{x}_i| \geq \frac{1}{\sqrt{k}} \|x_{-k}\|_2 - \frac{1}{\sqrt{Ck}} \|x_{-k}\|_2 \geq L$. $\qquad\square$

d) Step 4: We now show that each $\tilde{\theta}_{r,i}$ is good for $i \in I_r$. Let $\theta_{r,i}$ be the (oriented) phase difference between x_i and $\sum_{j=1}^{n} \eta_{r,j} g_{r,j} x_j$.

Lemma 6. *Suppose that event in Lemma 5 happens for* S'_r*. Then it holds for each* $i \in I_r$ *that, with probability at least* 0.95*,* $|\tilde{\theta}_{r,i} - \theta_{r,i}| \leq c\epsilon$ *for some absolute constant* c*.*

Proof. We have from our construction the measurements

$$\left| e^{i(2c\epsilon\ell_1 + \frac{\pi}{2}\ell_2)} \sum_{i=1}^{n} \eta_{r,i} g_{r,i} x_i + \sum_{i \in h_r^{-1}(j)} (1 - \eta_{r,i}) \sigma_{r,i} x_i \right|, \quad \ell_1, \ell_2 = 0, 1, \; j \in [B], \; r \in [R].$$

We note that Line 12 in the algorithm is valid because we have access to

$$|\hat{x}_i|, \; \left| \sum_{i=1}^{n} \eta_{r,i} g_{r,i} x_{r,i} \right|, \; \left| \hat{x}_i + e^{i(2c\epsilon\ell_1 + \frac{\pi}{2}\ell_2)} \sum_{i=1}^{n} \eta_{r,i} \sigma_{r,i} x_i + \sum_{i' \in h_r^{-1}(h_r(i)) \setminus S'} (1 - \eta_{r,i'}) g_{r,i'} x_{i'} \right|.$$

Let

$$Z = \sum_{i' \in h_r^{-1}(h_r(i)) \setminus S'} (1 - \eta_{r,i'}) \sigma_{r,i'} x_{i'}.$$

Observe that

$$\Pr\left\{ \left| h_r^{-1}(h_r(i)) \cap H_{C_2 k}(x) \right| = 1 \right\} \geq 1 - \frac{C_2}{C_B}$$

and that

$$\mathbb{E} |Z|^2 = \frac{\epsilon}{C_B k} \|x_{-C_2 k}\|_2^2.$$

By Markov's inequality, we have that $|Z|^2 \leq \frac{40\epsilon}{C_B k} \|x_{-k}\|_2^2$ with probability at least 0.025. Choose C_B such that $\frac{40}{C_B} \leq \frac{1}{2C_1}$ and $\frac{C_2}{C_B} < \frac{1}{40}$, the assumptions on noise magnitude in Lemma 3 will be satisfied for x_i with probability at least 0.95. Then Lemma 3 yields an estimate $\tilde{\theta}_{r,i}$ which satisfies $|\tilde{\theta}_{r,i} - \theta_i| \leq c\epsilon$. $\qquad\square$

e) Step 5: We are now ready to prove Theorem 2.

Proof of Theorem 2. Let $i \in S''$. Observe that for each r, it is isolated from every other $i' \in S'_r$ with probability $\frac{1}{C_B}$. A simple Chernoff bound shows that with probability at least $1 - 1/\text{poly}(n)$, it is contained in at least 0.9 fraction of $\{I_r\}_{r \in [R]}$, provided that C_B is big enough.

Suppose that event (2) happens. It follows from Lemmata 5 and 6 that in each repetition $r \in [R]$, for each $i \in I_r \setminus \{i_0\}$ it holds with probability at least $1 - 0.465$ that

$$\left| (\tilde{\theta}_{r,i} - \tilde{\theta}_{r,i_0}) - (\theta_{r,i} - \theta_{r,i_0}) \right| \leq 2c\epsilon.$$

Next we take the median over valid repetitions. To do this, we first check if there are at least half of $\tilde{\theta}_{r,i} - \tilde{\theta}_{r,i_0}$, when interpreted as real numbers in $[-\pi, \pi)$, satisfy that $\tilde{\theta}_{r,i} - \tilde{\theta}_{r,i_0} \in [-\frac{3}{25}\pi, \frac{3}{25}\pi]$. If this happens, take $\theta'_{i_0,i}$ to be the median of those numbers in $[-\pi, \pi)$; otherwise, take $\theta'_{i_0,i}$ to be the median with the interpretation of all $\tilde{\theta}_{r,i} - \tilde{\theta}_{r,i_0}$ as real numbers in $[0, 2\pi)$. Since both i and i_0 appear in at least 0.8 fraction of $\{I_r\}_{r \in [R]}$, we can, by choosing the constant c_R large enough, guarantee that

$$\left| \theta'_{i_0,i} - (\theta_{r,i} - \theta_{r,i_0}) \right| \leq 2c\epsilon \tag{6}$$

with probability at least $1 - 1/\text{poly}(n)$. This allows for a union bound over all $i \in S'' \setminus \{i_0\}$. Therefore (6) holds for all $i \in S'' \setminus \{i_0\}$ simultaneously with probability $\geq 1 - 1/\text{poly}(n)$.

Next, assume that it happens that (6) holds for all $i \in S'' \setminus \{i_0\}$. Consider the for-loop from Line 17 to 22. It is clear that when θ'_{i_0} is exactly the phase of x_{i_0}, it will hold that θ'_i is an accurate estimate to the phase of x_i up to an additive error of $2c\epsilon < \eta/2$. The if-clause in Line 20 will be true and the algorithm will terminate with an $\hat{x}$. Since the phases of the entries are at least η apart, there will be no ambiguity in rounding and the phases in $\hat{x}$ are all correct, hence the error $\|x - \hat{x}\|_2$ only depends on the magnitude errors, which is exactly (2), obtained from applying COUNT-SKETCH. When θ'_{i_0} is not x_{i_0}, by the rotational (k, η)-distinctness, $\{\theta'_i\}$ will coincide with P exactly or the if-clause will not be true. This shows the correctness.

Removing the conditioning of (2) increases the failure probability $1/\operatorname{poly}(n)$ and the overall failure probability is $1/\operatorname{poly}(n)$.

Number of Measurements: The submatrix Φ_{HH} has $\mathcal{O}(k \log n)$ rows, the submatrix Φ_{CS} has $\mathcal{O}((k/\epsilon)\log n)$ rows, the submatrix Φ_{L} has $\mathcal{O}(\log n)$ rows, each Φ_r for $r \in [R]$ has $\mathcal{O}(k/\epsilon)$ rows. Hence the total number of rows is dominated by that of Φ_{CS} and the R independent copies of Φ_r's, that is, $\mathcal{O}((k/\epsilon)\log n + R(k/\epsilon)) = \mathcal{O}((k/\epsilon)\log n) = \mathcal{O}((k/\eta)\log n)$.

Runtime: Line 1 runs in time $\mathcal{O}(k\operatorname{poly}(\log n))$, Line 2 in time $\mathcal{O}(|S|\log n) = \mathcal{O}(k \log n)$, Line 3 in time $O(\log n)$. The runtime before the if-branch of Line 4 is thus $\mathcal{O}(k\operatorname{poly}(\log n))$.

The noiseless case in Line 5 runs in time $\mathcal{O}(k \log k)$.

For the for-loop from Line 7 to 12, Line 8 runs in time $\mathcal{O}(k)$, Line 9 in $\mathcal{O}(k)$, Line 10 in $\mathcal{O}(k \log k)$, Lines 11–12 in time $O(k)$. Hence this for-loop takes time $\mathcal{O}(Rk \log k) = \mathcal{O}(k\operatorname{poly}(\log n))$. Line 13 runs in time $O(Rk) = O(k \log n)$. Calculating each median takes time $O(R)$ and thus Lines 15–16 takes time $O(Rk) = O(k \log n)$. The for-loop from Line 17 to 22 takes time $\mathcal{O}(k/\eta)$, since $|P| = \mathcal{O}(1/\eta)$, the if-clause in Line 20 can be verified in time $\mathcal{O}(k)$ if the elements in P are sorted in advance. Hence the total runtime of Lines 7–22 is $\mathcal{O}(k/\eta + k\operatorname{poly}(\log n))$.

Therefore, the overall runtime is $\mathcal{O}(k/\eta + k\operatorname{poly}(\log n))$. $\qquad\square$

V. FUTURE WORK

In this paper we obtained sublinear-time algorithms for different versions of the compressive phase retrieval problem, using purely combinatorial techniques. Our results include the first sublinear-time algorithm for the ℓ_∞/ℓ_2 problem, a popular problem in the literature. We suggest future directions and open questions that may be of interest to the computer science community.

- A natural question is to generalize our results to a more general set of valid phases. The difficulty lies in the fact that we can only estimate the phase up to an additive error of $O(\eta)$ using the Law of Cosines, and this would incur a large error if the corresponding heavy hitter has a huge magnitude.
- Is it possible to obtain a uniform guarantee for noiseless signals that runs in $\tilde{\mathcal{O}}(k)$ time and uses $\mathcal{O}(k)$ measurements? Even something that is substantially less than $\mathcal{O}(k^2)$ time would be interesting.
- Another question is whether one can achieve almost optimal bounds using structured measurements, one example being local correlation measurements as in [20]. This would pave the way of tackling the more constrained problem of phase retrieval, where the sensing matrix is the Discrete Fourier Matrix, a problem of great importance in engineering. There are numerous directions and open problems in this direction; interested readers can refer to [21], [37].

ACKNOWELEDGEMENT

REFERENCES

[1] Áke Björck and Victor Pereya. Solution of vandermonde systems of equations. *Mathematics of Computation*, 24(112):893–903, October 1970.

[2] Khanh Do Ba, Piotr Indyk, Eric Price, and David P. Woodruff. Lower bounds for sparse recovery. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 1190–1197, 2010.

[3] Daniel L. Boley, Franklin T. Luk, and David Vandevoorde. A fast method to diagonalize a hankel matrix. *Linear Algebra and its Applications*, 284(1):41–52, 1998. International Linear Algebra Society (ILAS) Symposium on Fast Algorithms for Control, Signals and Image Processing.

[4] S. Cai, M. Bakshi, S. Jaggi, and M. Chen. Super: Sparse signals with unknown phases efficiently recovered. In *2014 IEEE International Symposium on Information Theory*, pages 2007–2011, June 2014.

[5] E. J. Candès, X. Li, and M. Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, April 2015.

[6] Emmanuel J. Candès, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2):277 – 299, 2015.

[7] Emmanuel J. Candès, Thomas Strohmer, and Vladislav Voroninski. Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

[8] Moses Charikar, Kevin Chen, and Martin Farach-Colton. Finding frequent items in data streams. *Theoretical Computer Science*, 312(1):3 – 15, 2004. Automata, Languages and Programming.

[9] Paul Erdős and Alfréd Rényi. On random graphs I. *Publicationes Mathematicae (Debrecen)*, 6:290–297, 1959 1959.

[10] Bing Gao, Yang Wang, and Zhiqiang Xu. Stable signal recovery from phaseless measurements. *Journal of Fourier Analysis and Applications*, 22(4):787–808, Aug 2016.

[11] Bing Gao, Yang Wang, and Zhiqiang Xu. Stable signal recovery from phaseless measurements. *Journal of Fourier Analysis and Applications*, 22(4):787–808, 2016.

[12] Anna C. Gilbert, Yi Li, Ely Porat, and Martin J. Strauss. For-all sparse recovery in near-optimal time. *ACM Trans. Algorithms*, 13(3):32:1–32:26, March 2017.

[13] Anna C Gilbert, Hung Q Ngo, Ely Porat, Atri Rudra, and Martin J Strauss. ℓ_2/ℓ_2-foreach sparse recovery with low risk. In *International Colloquium on Automata, Languages, and Programming*, pages 461–472. Springer, 2013.

[14] Tom Goldstein and Christoph Studer. Phasemax: Convex phase retrieval via basis pursuit. *IEEE Transactions on Information Theory*, 64(4):2675–2689, 2018.

[15] Haitham Hassanieh, Piotr Indyk, Dina Katabi, and Eric Price. Nearly optimal sparse fourier transform. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 563–578. ACM, 2012.

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

[17] Piotr Indyk, Eric Price, and David P Woodruff. On the power of adaptivity in sparse recovery. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 285–294. IEEE, 2011.

[18] Mark Iwen, Aditya Viswanathan, and Yang Wang. Robust sparse phase retrieval made easy. *Applied and Computational Harmonic Analysis*, 42(1):135 – 142, 2017.

[19] Mark A. Iwen, Brian Preskitt, Rayan Saab, and Aditya Viswanathan. Phase retrieval from local measurements: Improved robustness via eigenvector-based angular synchronization. arXiv:1612.01182 [math.NA], 2016.

[20] Mark A. Iwen, Aditya Viswanathan, and Yang Wang. Fast phase retrieval from local correlation measurements. *SIAM Journal on Imaging Sciences*, 9(4):1655–1688, 2016.

[21] Kishore Jaganathan, Yonina C Eldar, and Babak Hassibi. Phase retrieval: An overview of recent developments. In *Optical Compressive Imaging.*, pages 263–296. CRC Press, 2016.

[22] Kishore Jaganathan, Yonina C Eldar, and Babak Hassibi. Stft phase retrieval: Uniqueness guarantees and recovery algorithms. *IEEE Journal of selected topics in signal processing*, 10(4):770–781, 2016.

[23] Kishore Jaganathan, Samet Oymak, and Babak Hassibi. Sparse phase retrieval: Convex algorithms and limitations. In *2013 IEEE International Symposium on Information Theory*, pages 1022–1026. IEEE, 2013.

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

[25] Michael Kapralov. Sample efficient estimation and recovery in sparse fft via isolation on average. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pages 651–662. Ieee, 2017.

[26] Samuel Karlin and Howard M. Taylor. *A First Course in Stochastic Processes*. Elsevier, 2nd edition, 1975.

[27] Kasper Green Larsen, Jelani Nelson, Huy L Nguyên, and Mikkel Thorup. Heavy hitters via cluster-preserving clustering. In *Foundations of Computer Science (FOCS), 2016 IEEE 57th Annual Symposium on*, pages 61–70. IEEE, 2016.

[28] Yi Li, Vasileios Nakos, and David P. Woodruff. On low-risk heavy hitters and sparse recovery schemes. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2018, August 20-22, 2018 - Princeton, NJ, USA*, pages 19:1–19:13, 2018.

[29] Vasileios Nakos. Almost optimal phaseless compressed sensing with sublinear decoding time. In *2017 IEEE International Symposium on Information Theory, ISIT 2017, Aachen, Germany, June 25-30, 2017*, pages 1142–1146, 2017.

[30] Vasileios Nakos, Xiaofei Shi, David P. Woodruff, and Hongyang Zhang. Improved algorithms for adaptive compressed sensing. In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic*, 2018.

[31] Vasileios Nakos and Zhao Song. Stronger l_2/l_2 compressed sensing; without iterating. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing, STOC 2019, Phoenix, AZ, USA, June 23-26, 2019.*, 2019.

[32] R. Pedarsani, D. Yin, K. Lee, and K. Ramchandran. Phasecode: Fast and efficient compressive phase retrieval based on sparse-graph codes. *IEEE Transactions on Information Theory*, 63(6):3663–3691, June 2017.

[33] Ely Porat and Martin J. Strauss. Sublinear time, measurement-optimal, sparse recovery for all. In *Proceedings of the Twenty-third Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '12, pages 1215–1227, Philadelphia, PA, USA, 2012. Society for Industrial and Applied Mathematics.

[34] E. Price and D. P. Woodruff. $(1 + \epsilon)$-approximate sparse recovery. In *2011 IEEE 52nd Annual Symposium on Foundations of Computer Science*, pages 295–304, Oct 2011.

[35] R. Prony. Essai éxperimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures. *Journal de l'École Polytechnique Floréal et Plairial*, (22):24–76, 1795.

[36] Lawrence Rabiner and Biing-Hwang Juang. *Fundamentals of Speech Recognition*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[37] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev. Phase retrieval with application to optical imaging: A contemporary overview. *IEEE Signal Processing Magazine*, 32(3):87–109, 2015.

[38] Mahdi Soltanolkotabi. Structured signal recovery from quadratic measurements: Breaking sample complexity barriers via nonconvex optimization. *IEEE Transactions on Information Theory*, 65(4):2374–2400, 2019.

[39] Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. *Foundations of Computational Mathematics*, 18(5):1131–1198, 2018.