Binary Iterative Hard Thresholding Converges with Optimal Number of Measurements for 1-Bit Compressed Sensing

Namiko Matsumoto
Department of Computer Science and Engineering
University of California, San Diego
La Jolla, California, USA
nmatsumo@ucsd.edu

Arya Mazumdar
Halıcıoğlu Data Science Institute
University of California, San Diego
La Jolla, California, USA
arya@ucsd.edu

***Abstract*—Compressed sensing has been a very successful high-dimensional signal acquisition and recovery technique that relies on linear operations. However, the actual measurements of signals have to be quantized before storing or processing them. 1(One)-bit compressed sensing is a heavily quantized version of compressed sensing, where each linear measurement of a signal is reduced to just one bit: the sign of the measurement. Once enough of such measurements are collected, the recovery problem in 1-bit compressed sensing aims to find the original signal with as much accuracy as possible. The recovery problem is related to the traditional "halfspace-learning" problem in learning theory.**

For recovery of sparse vectors, a popular reconstruction method from one-bit measurements is the *binary iterative hard thresholding (BIHT)* algorithm. The algorithm is a simple projected subgradient descent method, and is known to converge well empirically, despite the nonconvexity of the problem. The convergence property of BIHT was not theoretically justified, except with an exorbitantly large number of measurements (i.e., a number of measurement greater than $\max\{k^{10}, 24^{48}, k^{3.5}/\epsilon\}$, where k is the sparsity and ϵ denotes the approximation error, and even this expression hides other factors). In this paper we show that the BIHT estimates converge to the original signal with only $\tilde{O}(\frac{k}{\epsilon})$ measurements. Note that, this dependence on k and ϵ is optimal for any recovery method in 1-bit compressed sensing. With this result, to the best of our knowledge, BIHT is the only practical and efficient (polynomial time) algorithm that requires the optimal number of measurements in all parameters (both k and ϵ). This is also an example of a gradient descent algorithm converging to the correct solution for a nonconvex problem, under suitable structural conditions.

***Index Terms*—compressed sensing, quantization, gradient descent, sparsity**

I. Introduction

One-bit compressed sensing (1bCS) is a basic nonlinear sampling method for high-dimensional sparse signals, introduced first in [2]. Consider an unknown sparse signal $\mathbf{x} \in \mathbb{R}^n$ with sparsity (number of nonzero coordinates) $||\mathbf{x}||_0 \leq k$, where $k \ll n$. In the 1bCS framework, measurements of $\mathbf{x}$ are obtained with a sensing matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ via the observations of signs:

$$\mathbf{b} = \mathsf{sign}(\mathbf{A}\mathbf{x}).$$

The sign function (formally defined later) is simply the $\pm$ signs of the coordinates.

Compressed sensing, the method of obtaining signals by taking few linear projections [3], [4] has seen a lot of success in the past two decades. 1bCS is an extremely quantized version of compressed sensing where only one bit per sample of the signal is observed. In terms of nonlinearity, this is one of the simplest examples of a single-index model [5]: $y_i = f(\langle \mathbf{a}_i, \mathbf{x}\rangle), i = 1, \ldots, m$, where f is a coordinate-wise nonlinear operation. As a practical case study and for its aesthetic appeal, 1bCS has been studied with interest in the last few years, for example, in [6]–[10].

Notably, it was shown in [11] that $m = \tilde{\Theta}(k/\epsilon)$ measurements are necessary and sufficient (up to logarithmic factors) to approximate $\mathbf{x}$ within an ϵ-ball. But the reconstruction method used to obtain this measurement complexity is via exhaustive search, which is practically infeasible. A linear programming based solution (which runs in polynomial time) that has measurement complexity $O(\frac{k}{\epsilon^5}\log^2\frac{n}{k})$ was provided in [12]. Note the suboptimal dependence on ϵ.

An incredibly well-performing algorithm turned out to the *binary iterative hard thresholding* (BIHT) algorithm, proposed in the former work [11]. BIHT is a simple iterative algorithm that converges to the correct solution quickly in practice. However, until later, the reason of its good performance was somewhat unexplained, barring the fact that it is actually a proximal gradient descent algorithm on a certain loss function (provided in Eq. (8)). In the algorithm, the projection is taken onto a nonconvex set (namely, selecting the "top-k" coordinates and then normalizing), which usually makes a theoretical analysis unwieldy. Since the work of [11] there has been some progress explaining the empirical success of the BIHT algorithm. In particular, it was shown in [13, Sec. 3.4.2] that after only the first iteration of the BIHT algorithm, an

This work is supported in part by NSF awards 2133484 and 2127929. A full version of this paper with detailed proofs is available online [1].

2575-8454/22/$31.00 ©2022 IEEE
DOI 10.1109/FOCS54457.2022.00082

approximation error ϵ is achievable with $\tilde{O}(\frac{k}{\epsilon^4})$ measurements, though the same result is shown in [14, Sec. 5] with $\tilde{O}(\frac{k}{\epsilon^2})$ measurements, so the former result might just be a typo. Similar results also appear in [15, Sec. 3.5]. In all these results, the dependence on ϵ, which is also referred to as the error-rate, is suboptimal. Furthermore, these works also do not show convergence as the algorithm iterates further. Indeed, according to these works, $O(\frac{k}{\epsilon^2}\log\frac{n}{k})$ measurements are sufficient to bring the error down to ϵ after just the first iteration of BIHT. Beyond the first iteration, it was shown in [16] that the iterates of BIHT remain bounded, maintaining the same order of accuracy for the subsequent iterations. This, however, does not imply a reduction in the approximation error after the first iteration. This issue has been partially mitigated in [17], which uses a *normalized* version of the BIHT algorithm. While [17] manage to show that the normalized BIHT algorithm can achieve optimal dependence on the error-rate as the number of iterations of BIHT tends to infinity, i.e., $m \sim \frac{1}{\epsilon}$, their result is only valid when $m > \max\{ck^{10}\log^{10}\frac{n}{k}, 24^{48}, \frac{c'}{\epsilon}(k\log\frac{n}{k})^{7/2}\}$. This clearly is highly sub-optimal in terms of dependence on k, and does not explain the empirical performance of the algorithm. This has been left as the main open problem in this area as per [17].

A. Our Contribution and Techniques

In this paper, we show that the normalized BIHT algorithm converges with a sample complexity having optimal dependence on both the sparsity k and error ϵ (see, Theorem III.1 below). As such, we further show the convergence rate with respect to iterations for this algorithm. In particular, we show that the approximation error of BIHT decays as $O(\epsilon^{1-2^{-t}})$ with the number of iteration t. This encapsulates the very fast convergence of BIHT to the ϵ-ball of the actual signal. Furthermore, this also shows that after just one iteration of BIHT, an approximation error of $\sqrt{\epsilon}$ is achievable, with $O(\frac{k}{\epsilon}\log\frac{n}{k})$ measurements, which matches the observations of [14], [15] regarding the performance of BIHT with just one iteration. Due to the aforementioned fast rate, the approximation error quickly converges to ϵ resulting in a polynomial time algorithm for recovery in 1bCS with only $\tilde{O}(\frac{k}{\epsilon})$ measurements, the optimal.

There are several difficulties in analyzing BIHT that were pointed out in the past, for example in [17]. First of all, the loss function is not differentiable, and therefore one has to rely on (sub)gradients, which prohibits an easier analysis of convergence. Secondly, the algorithm projects onto nonconvex sets, so the improvement of the approximation in each iteration is not immediately apparent. To tackle these hurdles, the key idea is to use some structural property of the measurement or sampling matrix. Our result relies on such a property of the sampling matrix $\mathbf{A}$, called the restricted approximate invertibility condition (RAIC). A somewhat different invertibility property of a matrix also appears in [17]. However, our definition, which looks more natural, allows for a significantly different analysis that yields the improved sample complexity. Thereafter, we show that random matrices with i.i.d. Gaussian entries satisfy the invertibility condition with overwhelmingly large probability.

The invertibility condition that is essential for our proof intuitively states that treating the signed measurements as some "scaled linear" measurements should lead to adequate estimates, which is an overarching theme of recovery in generalized linear models. Further, our condition quantifies the "goodness" of these estimates in a way that allows us to show a contraction in the BIHT iterations. This contraction of approximation error comes naturally from our definition. In contrast, while a similar idea appears in [17], showing the contraction of approximate error is a highly involved exercise therein. As another point of interest, [11, Sec. 4.2] empirically observed that in normalized BIHT, the step-size of the gradient descent algorithm must be carefully chosen, or else the algorithm will not converge. Our definition of the invertibility condition gives some intuitive justification on why the algorithm is so sensitive to step-size. Our analysis relies on the step-size being set exactly to $\eta = \sqrt{2\pi}$. More generally, if η were to deviate too far from $\sqrt{2\pi}$, the contraction would be lost.

So the technical burden of our main result turns out to be to show Gaussian matrices do satisfy the invertibility condition (Definition III.1 below). We need to show that for every pair of sparse unit vectors the condition holds. We resort to constructing a cover, an "epsilon-net," of the unit sphere, and then decompose the invertibility conditions for any two vectors in the sphere into two components. First, we show that it is satisfied for two vectors in the epsilon-net whose distance is sufficiently large, and then we show that only small error is added when instead of the net points, vectors close to them are considered. This leads to a "large-distance" and "small-distance" analysis. For these two parts, we require differently curated concentration inequalities, which form the bulk of the techniques used in this paper. Notably, we cannot just extend the invertibility condition to points outside the net by simply using, e.g., the triangle inequality, due to the sign operation. But at the same time, the sign operation significantly reduces the number of matrix-vector products we need to union bound over. It turns out that, because we condition on the rotational uniformity of the measurements, this number is not "too large," and will not increase the sample complexity beyond the optimal.

One important aspect of BIHT's convergence is that as the approximation error in t^{th} iteration improves, it makes possible an even smaller error for the $(t+1)^{\text{th}}$ approximation. This can again be intuitively explained by the rotational symmetry of the measurements, as well as the sign operation. Each iteration of BIHT involves fewer and fewer measurements, and we can track the number of measurements involved by tracking the number of measurements that are *mismatches* between the vector $\mathbf{x}$ and its approximation at the t^{th} iteration. This is used in the "large-distance" regime, where the pairs of points must be at least some distance τ from each other (note that this qualifier is necessary). On the other hand, once the distance is smaller than τ, the Chernoff bound that is used to track the

mismatch is no longer sufficient (using that we would end up needing a suboptimal sample complexity). That is why we need to use a separate analysis for the "small-distance" regime. In this regime, we instead try to keep a count of the number of distinct vectors obtained by $\mathsf{sign}(\mathbf{Ax})$ for all k-sparse unit norm $\mathbf{x}$ within a "small distance" from a fixed net point. Because of the rotational uniformity, this count can also be tightly quantified, and it turns out to be small enough to give us the optimal sample complexity.

B. Other Related Works

A generalization of 1bCS is the noisy version of the problem, where the binary observations $y_i \in \{+1, -1\}$ are random (noisy): i.e., $y_i = 1$ with probability $f(\langle \mathbf{a}_i, \mathbf{x} \rangle), i = 1, \ldots, m$, where f is a potentially nonlinear function, such as the sigmoid function. Recovery guarantees for such models were studied in [9]. In another model, observational noise can appear before the quantization, i.e., $y_i = \mathsf{sign}\left(\langle \mathbf{a}_i, \mathbf{x} \rangle + \xi_i\right), i = 1, \ldots, m$, where ξ_i is random noise. As observed in [5], [17], the noiseless setting (also considered in this work) is actually more difficult to handle because the randomness of noise allows for a maximum likelihood analysis. Indeed, having some control-over ξ_is (or just assuming them to be i.i.d. Gaussian), helps estimate the norm of $\mathbf{x}$ [18], which is otherwise impossible with just sign measurements, as in our model (this is called introducing *dither*, a well-known paradigm in signal processing). In a related line of work, one-bit measurements are taken by adaptively varying the threshold (in our case the threshold is always 0), which can significantly reduce the error-rate, for example see [19] and [20], the latter being an application of sigma-delta quantization methods.

Yet another line of work in 1bCS literature takes a more combinatorial avenue and looks at the support recovery problem and constructions of structured measurement matrices. Instances of these works are [7], [8], [21], [22]. However, the nature of these works is quite different from ours.

C. Organization

The rest of the paper is organized as follows. The required notations and definitions to state the main result appear in Section II, where we also formally define the 1-bit compressed sensing problem and the reconstruction method, the normalized binary iterative hard thresholding algorithm (Algorithm 1). We provide our main result in Section III, which establishes the convergence rate of BIHT (Theorem III.1) and the asymptotic error-rate (Corollary III.2) with the optimal measurement complexity. In Section III-B we also overview the derivation of the result, including our invertibility condition for Gaussian matrices. In Section IV we provide the main proof of the BIHT convergence algorithm, assuming that a structural property is satisfied by the measurement matrix. Proof of this structural property for Gaussian matrices is the major technical contribution of this paper (Theorem III.3). However, due to the space limitation, we are unable to give the full proof here. It can be found in the full version of this paper which is available online [1]. Proofs of all lemmas and intermediate results that are omitted here can also be found in the the full version. We conclude with some future directions in Section V.

II. Preliminaries

A. Notations and Definitions

The set of all k-sparse real-valued vectors in n dimension is denoted by Σ_k^n. The ℓ_2-sphere in $\mathbb{R}^n$ is written $\mathcal{S}^{n-1} \subset \mathbb{R}^n$, and hence, $\mathcal{S}^{n-1} \cap \Sigma_k^n \subset \Sigma_k^n$ is the subset of k-sparse real-valued vectors with unit norm. The Euclidean ball of radius $\tau \geq 0$ and center $\mathbf{u} \in \mathbb{R}^n$ is defined as $\mathcal{B}_\tau(\mathbf{u}) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{u} - \mathbf{x}\|_2 \leq \tau\}$. Matrices are denoted in uppercase, boldface text, e.g., $\mathbf{M} \in \mathbb{R}^{m \times n}$, with its (i, j)-entries written $M_{i,j}$. The $n \times n$ identity matrix written as $\mathbf{I}_{n \times n}$. Vectors are likewise indicated by boldface font, using lowercase and uppercase lettering for nonrandom and random vectors, respectively, e.g., $\mathbf{u} \in \mathbb{R}^n$ and $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n \times n})$, with entries denoted such that, e.g., $\mathbf{u} = (u_1, \ldots, u_n)$. As customary, $\mathcal{N}(\mathbf{0}, \mathbf{I}_{n \times n})$ denotes the i.i.d. n-variate standard normal distribution (with the univariate case, $\mathcal{N}(0, 1)$). Moreover, random sampling from a distribution $\mathcal{D}$ is denoted by $X \sim \mathcal{D}$, and likewise, drawing uniformly at random from a set $\mathcal{X}$ is written as $X \sim \mathcal{X}$. For any pair of real-valued vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, write $d_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v}) \in \mathbb{R}_{\geq 0}$ for the distance between their projections onto the ℓ_2-sphere, as well as $\theta_{\mathbf{u},\mathbf{v}} \in [0, \pi]$ and $\boldsymbol{\theta}_{\mathbf{u},\mathbf{v}} \in [-\pi, \pi]$ for, respectively, the angular distance and signed angular distance (for a given convention of positive and negative directions of rotation) between them. Formally,

$$d_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v}) = \begin{cases} \left\| \frac{\mathbf{u}}{\|\mathbf{u}\|_2} - \frac{\mathbf{v}}{\|\mathbf{v}\|_2} \right\|_2, & \text{if } \mathbf{u}, \mathbf{v} \neq \mathbf{0}, \\ 0, & \text{if } \mathbf{u} = \mathbf{v} = \mathbf{0}, \\ 1, & \text{otherwise}, \end{cases} \tag{1}$$

$$\theta_{\mathbf{u},\mathbf{v}} = \arccos\left(\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2}\right). \tag{2}$$

Note that these are related by

$$\theta_{\mathbf{u},\mathbf{v}} = \arccos\left(1 - \frac{d^2_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v})}{2}\right) \tag{3}$$

or equivalently,

$$d_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v}) = \sqrt{2(1 - \cos(\theta_{\mathbf{u},\mathbf{v}}))}. \tag{4}$$

The sign function, $\mathsf{sign} : \mathbb{R} \to \{+1, -1\}$, is defined in the following way:

$$\mathsf{sign}(x) = \begin{cases} 1, & x \geq 0, \\ -1, & x < 0. \end{cases}$$

The function can be extended to vectors, i.e., $\mathsf{sign} : \mathbb{R}^n \to \{+1, -1\}^n$, by just applying it on each coordinate. Additionally, for a condition $C \in \{\mathsf{true}, \mathsf{false}\}$, define the indicator function $\mathbb{I} : \{\mathsf{true}, \mathsf{false}\} \to \{0, 1\}$ by

$$\mathbb{I}(C) = \begin{cases} 0, & \text{if } C = \mathsf{false}, \\ 1, & \text{if } C = \mathsf{true}. \end{cases} \tag{5}$$

We are going use the following universal constants $a, b, c, c_1, c_2 > 0$ in the statement of our results. Their values are

$$\begin{aligned} a = 16, \quad b \gtrsim 379.1038, \quad c = 32, \\ c_1 = \sqrt{\frac{3\pi}{b}}\left(1 + \frac{16\sqrt{2}}{3}\right), \\ c_2 = \frac{3}{b}\left(1 + \frac{4\pi}{3} + \frac{8\sqrt{3\pi}}{3} + 8\sqrt{6\pi}\right). \end{aligned} \tag{6}$$

Additionally, in the BIHT algorithm, the step-size $\eta > 0$ is fixed as $\eta = \sqrt{2\pi}$.

We define two hard thresholding operations: the *top-k hard thresholding operation* and the *subset hard thresholding operation*, defined below in Definitions II.1 and II.2. When clear from context, we will omit the distinction simply refer to a *hard thresholding operation*.

Definition II.1 (Top-k hard thresholding operation)**.** *For $k \in \mathbb{Z}_+$, $k \leq n$, the* top-k hard thresholding operation*, $\mathcal{T}_k : \mathbb{R}^n \to \Sigma_k^n$, projects a real-valued vector $\mathbf{u} \in \mathbb{R}^n$ into the space of k-sparse real-valued vectors by setting all but the k largest (in absolute value) entries in $\mathbf{u}$ to 0 (with ties broken arbitrarily).*

Definition II.2 (Subset hard thresholding operation)**.** *For a k-subset of coordinates $J \subseteq [n]$, the* subset hard thresholding operation *associated with J, $\mathcal{T}_J : \mathbb{R}^n \to \Sigma_k^n$, projects a real-valued vector $\mathbf{u} \in \mathbb{R}^n$ into the space of k-sparse real-valued vectors by $\mathcal{T}_J(\mathbf{u})_j = u_j \cdot \mathbb{I}(j \in J)$ for each $j \in [n]$.*

B. 1-Bit Compressed Sensing and the BIHT Algorithm

A measurement matrix is denoted by $\mathbf{A} \in \mathbb{R}^{m \times n}$ and has rows, $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(m)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n \times n})$, with i.i.d. Gaussian entries. The one-bit measurements of an unknown signal, $\mathbf{x} \in \Sigma_k^n$, are performed by:

$$\mathbf{b} = \mathsf{sign}(\mathbf{A}\mathbf{x}) \tag{7}$$

Throughout this work, the unknown signals, $\mathbf{x} \in \Sigma_k^n$, are assume to have unit norm since information about the norm is lost due to the one-bit quanitization of the measurements. (For interested readers, see [18] for techniques, e.g., dithering, to reconstruct the signal's norm in 1-bit compressed sensing.) Given $\mathbf{A}$ and $\mathbf{b}$, the goal of 1-bit compressed sensing is to recover $\mathbf{x}$ as accurately as possible. We measure the accuracy of the reconstruction, $\hat{\mathbf{x}} \in \Sigma_k^n$, by the metric $d_{\mathcal{S}^{n-1}}(\mathbf{x}, \hat{\mathbf{x}})$.

The BIHT reconstruction algorithm, proposed by [11], comprises two iterative steps: (i) a gradient descent step, which finds a dense approximation, $\tilde{\mathbf{x}} \in \mathbb{R}^n$, followed by (ii) a projection by $\tilde{\mathbf{x}} \mapsto \hat{\mathbf{x}} = \mathcal{T}_k(\tilde{\mathbf{x}})$ into the space of k-sparse real-valued vectors. As shown by [11], the gradient step, (i), aims to minimize the objective function

$$\mathcal{J}(\hat{\mathbf{x}}; \mathbf{x}) = \left\| [\, \mathsf{sign}(\mathbf{A}\mathbf{x}) \odot \mathsf{sign}(\mathbf{A}\hat{\mathbf{x}}) \,]_- \right\|_1, \tag{8}$$

where $\mathbf{u} \odot \mathbf{v} = (u_1 v_1, \ldots, u_n v_n)$ and $([\mathbf{u}]_-)_j = u_j \cdot \mathbb{I}(u_j < 0)$. While several variants of the BIHT algorithm have been proposed, see, [11, Section 4], this work focuses on the normalized BIHT algorithm, where the projection step, (ii), is modified to project the approximation onto the k-sparse, ℓ_2-unit sphere, $\mathcal{S}^{n-1} \cap \Sigma_k^n$. Algorithm 1 provides the version of the BIHT algorithm studied in this work.

Algorithm 1: Binary iterative hard thresholding with normalized projections (normalized BIHT)

Set $\eta = \sqrt{2\pi}$
$\hat{\mathbf{x}}^{(0)} \sim \mathcal{S}^{n-1} \cap \Sigma_k^n$
for $t = 1, 2, 3, \ldots$ **do**
  $\tilde{\mathbf{x}}^{(t)} \leftarrow \hat{\mathbf{x}}^{(t-1)} + \frac{\eta}{m}\mathbf{A}^T \cdot \frac{1}{2}\big(\mathsf{sign}(\mathbf{A}\mathbf{x}) - \mathsf{sign}(\mathbf{A}\hat{\mathbf{x}}^{(t-1)})\big)$
  $\hat{\mathbf{x}}^{(t)} \leftarrow \frac{\mathcal{T}_k(\tilde{\mathbf{x}}^{(t)})}{\|\mathcal{T}_k(\tilde{\mathbf{x}}^{(t)})\|_2}$

III. Main Results and Techniques

A. BIHT Convergence Theorem

Our main result is presented below. Informally, it states that with $m = O(\frac{k}{\epsilon}\log\frac{n}{k\sqrt{\epsilon}})$ one-bit (sign) measurements, it is possible to recover any k-sparse unit vector within an ϵ-ball, by means of the normalized BIHT algorithm.

Theorem III.1. *Let $a, b, c > 0$ be universal constants as in Eq. (6). Fix $\epsilon, \rho \in (0, 1)$ and $k, m, n \in \mathbb{Z}_+$, where*

$$m \geq \frac{4bck}{\epsilon}\log\left(\frac{en}{k}\right) + \frac{2bck}{\epsilon}\log\left(\frac{12bc}{\epsilon}\right) + \frac{bc}{\epsilon}\log\left(\frac{a}{\rho}\right). \tag{9}$$

Let the measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ have rows with i.i.d. Gaussian entries. Then, uniformly with probability at least $1 - \rho$, for every unknown k-sparse real-valued unit vector, $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, the normalized BIHT algorithm produces a sequence of approximations, $\{\hat{\mathbf{x}}^{(t)} \in \mathcal{S}^{n-1} \cap \Sigma_k^n\}_{t \in \mathbb{Z}_{\geq 0}}$, which converges to the ϵ-ball around the unknown vector $\mathbf{x}$ at a rate upper bounded by

$$d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\big) \leq 2^{2^{-t}}\epsilon^{1-2^{-t}} \tag{10}$$

for each $t \in \mathbb{Z}_{\geq 0}$.

Corollary III.2. *Under the conditions stated in Theorem III.1, uniformly with probability at least $1 - \rho$, for every unknown k-sparse real-valued unit vector, $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, the sequence of BIHT approximations, $\{\hat{\mathbf{x}}^{(t)}\}_{t \in \mathbb{Z}_{\geq 0}}$, converges asymptotically to the ϵ-ball around the unknown vector $\mathbf{x}$. Formally,*

$$\lim_{t \to \infty} d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\big) \leq \epsilon. \tag{11}$$

B. Technical Overview

The analysis in this work is divided into two components: (I) the proofs of Theorem III.1 and Corollary III.2, which show the universal convergence of the BIHT approximations by using the *restricted approximate invertibility condition*

(RAIC) for Gaussian measurement matries (defined below), and (II) the proof of the main technical theorem, Theorem III.3 (also below), which derives the RAIC for such a measurement matrix. As already mentioned, the second piece of analysis is only outlined this version but can be found in its entirety in the full version [1].

Informally speaking, we show that the approximation error, $\varepsilon(t)$, of the BIHT algorithm at step $t > 0$ satisfies a recurrence relation of the form $\varepsilon(t) = a_1\sqrt{\epsilon\varepsilon(t-1)} + a_2\epsilon$. It is not a difficult exercise to see that we get the desired convergence rate from this recursion, starting from a constant error. The recursion itself is a result of the RAIC property, which tries to capture the fact that the difference between two vectors $\mathbf{x}$ and $\mathbf{y}$ can be reconstructed by applying $\mathbf{A}^T$ on the difference of the corresponding one-bit measurements. Next we explain the technicalities of these different components of the proof.

1) The Restricted Approximate Invertibility Condition

The main technical contribution is an improved sample complexity for the restricted approximate invertibility condition (RAIC). A different invertibilty condition was proposed by [17]. A comparison of the two definitions can be found in the full version of this paper [1]. The definition of RAIC considered in this work is formalized in Definition III.1, which uses the following notations. For $m, n \in \mathbb{Z}_+$, let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be a measurement matrix with rows $\mathbf{A}^{(i)} \in \mathbb{R}^n$, $i \in [m]$. Then, define the functions $h_{\mathbf{A}}, h_{\mathbf{A};J} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$h_{\mathbf{A}}(\mathbf{x},\mathbf{y}) = \frac{\eta}{m}\mathbf{A}^T \cdot \frac{1}{2}\left(\mathsf{sign}(\mathbf{A}\mathbf{x}) - \mathsf{sign}(\mathbf{A}\mathbf{y})\right) \tag{12}$$

and

$$h_{\mathbf{A};J}(\mathbf{x},\mathbf{y}) = \mathcal{T}_{\mathsf{supp}(\mathbf{x})\cup\mathsf{supp}(\mathbf{y})\cup J}\left(h_{\mathbf{A}}(\mathbf{x},\mathbf{y})\right) \tag{13}$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $J \subseteq [n]$, and where $\eta = \sqrt{2\pi}$.

Definition III.1 (Restricted approximate invertibility condition (RAIC))**.** *Fix $\delta, a_1, a_2 > 0$ and $k, m, n \in \mathbb{Z}_+$ such that $0 < k < n$. The (k, n, δ, a_1, a_2)-RAIC is satisfied by a measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ if*

$$\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 \le a_1\sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{x},\mathbf{y})} + a_2\delta \tag{14}$$

uniformly for all $\mathbf{x}, \mathbf{y} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$ and all $J \subseteq [n]$, $|J| \le k$.

Theorem III.3 below is the primary technical result in this analysis and establishes that m-many i.i.d. Gaussian measurements satisfy the (k, n, δ, c_1, c_2)-RAIC, where the sample complexity for m matches the lower bound of [11, Lemma 1]. The proof of the theorem is deferred to the full version [1], while an overview of the proof is given below in Section III-B3.

Theorem III.3. *Let $a, b, c_1, c_2 > 0$ be universal constants as defined in Eq.* (6)*. Fix $\delta, \rho \in (0, 1)$ and $k, m, n \in \mathbb{Z}_+$ such that $0 < k < n$ and*

$$\begin{aligned} m &= \frac{b}{\delta}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12b}{\delta}\right)^{2k}\left(\frac{a}{\rho}\right)\right) \\ &= O\left(\frac{k}{\delta}\log\left(\frac{n}{\delta k}\right) + \frac{1}{\delta}\log\left(\frac{1}{\rho}\right)\right). \end{aligned} \tag{15}$$

Let $\mathbf{A} \in \mathbb{R}^{m\times n}$ be a measurement matrix whose rows have i.i.d. Gaussian entries. Then, $\mathbf{A}$ satisfies the (k, n, δ, c_1, c_2)-RAIC with probability at least $1 - \rho$. To state this explicitly, uniformly with probability at least $1-\rho$, for all $\mathbf{x}, \mathbf{y} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$ and all $J \subseteq [n]$, $|J| \le k$,

$$\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 \le c_1\sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{x},\mathbf{y})} + c_2\delta. \tag{16}$$

2) The Uniform Convergence of BIHT Approximations

Assuming the desired RAIC property (i.e., the correctness of Theorem III.3), the uniform convergence of BIHT approximations is shown as follows.

(a) The 0^{th} BIHT approximation, which is simply drawn uniformly at random, $\hat{\mathbf{x}}^{(0)} \sim \mathcal{S}^{n-1} \cap \Sigma_k^n$, can be seen to have an error of at most 2 (the diameter of the unit sphere). Then, the following argument handles each subsequent t^{th} BIHT approximation, $t \in \mathbb{Z}_+$.

(b) Using standard techniques, the error of any t^{th} BIHT approximation, $t \in \mathbb{Z}_+$, can be shown to be (deterministically) upper bounded by

$$\begin{aligned} &d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\right) \\ &\quad\le 4\left\|\left(\mathbf{x} - \hat{\mathbf{x}}^{(t-1)}\right) - h_{\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^t)}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}\right)\right\|_2. \end{aligned} \tag{17}$$

(c) Subsequently, observing the correspondence between Eq. (17) and the RAIC, Theorem III.3 is applied to further bound the t^{th} approximation error in (17) from above by

$$\begin{aligned} d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\right) &\le 4\left(c_1\sqrt{\frac{\epsilon}{c}d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}\right)} + c_2\frac{\epsilon}{c}\right) \\ &= 4c_1\sqrt{\frac{\epsilon}{c}d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}\right)} + 4c_2\frac{\epsilon}{c}. \end{aligned} \tag{18}$$

(d) Then, the recurrence relation corresponding to the right-hand-side of Eq. (18),

$$\varepsilon(0) = 2, \tag{19}$$

$$\varepsilon(t) = 4c_1\sqrt{\frac{\epsilon}{c}\varepsilon(t-1)} + 4c_2\frac{\epsilon}{c}, \quad t \in \mathbb{Z}_+, \tag{20}$$

can be shown to monotonically decrease with t, asymptotically converging as $\varepsilon(t) \sim \epsilon$, and pointwise upper bounded by $\varepsilon(t) \le 2^{2^{-t}}\epsilon^{1-2^{-t}}$ for each $t \in \mathbb{Z}_{\ge 0}$. The asymptotic convergence and convergence rate of the BIHT apprximations to the ϵ-ball around the unknown vector $\mathbf{x}$ directly follow. This will complete the analysis for the universal convergence of the BIHT algorithm.

3) The RAIC for an i.i.d. Gaussian Matrix

Fixing $\delta, \rho \in (0, 1)$ and letting $c_1, c_2 > 0$ be the universal constants specified in Eq. (6), Theorem III.3 establishes that the measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with i.i.d. Gaussian entries satisfies the (k, n, δ, c_1, c_2)-RAIC with high probability (at least $1-\rho$) when the number of measurements m is at least

what is stated in Eq. (15). The proof of the theorem is outlined as follows.

(a) Writing $\tau = \frac{\delta}{b}$, suppose $\{\mathcal{C}_{\tau;J} \subseteq \mathcal{S}^{n-1} \cap \Sigma_k^n : J \subseteq [n], |J| \leq k\}$ are τ-nets over the subset of vectors in $\mathcal{S}^{n-1} \cap \Sigma_k^n$ whose support sets are precisely J. Then, a τ-net over the entire set of k-sparse real-valued vectors, $\mathcal{S}^{n-1} \cap \Sigma_k^n$, is constructed by the union $\mathcal{C}_\tau = \bigcup_{J\subseteq[n]:|J|\leq k} \mathcal{C}_{\tau;J}$. The goal will be to show that with high probability certain properties hold for (almost) every ordered pair $(\mathbf{u}, \mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau$, or for every vector $\mathbf{u} \in \mathcal{C}_\tau$. The desired RAIC will then follow from extending the properties to every pair $\mathbf{x}, \mathbf{y} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$.

(b) The first property, corresponding with the "large distance" regime (recall the discussion in Section I-A), requires that with probability at least $1 - \rho_1$, for every ordered pair, $(\mathbf{u}, \mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau$, in the τ-net with distance at least $d_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v}) \geq \tau$ and for every $J \subseteq [n]$, $|J| \leq 2k$,

$$\|(\mathbf{u} - \mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u}, \mathbf{v})\|_2 \leq b_1 \sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v})}, \quad (21)$$

where $b_1 > 0$ is a small universal constant (see, Eq. (6)).

(c) The second property, corresponding with the "small distance" regime, requires that with probability at least $1 - \rho_2$, for each $\mathbf{u} \in \mathcal{C}_\tau$, each $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$, and each $J \subseteq [n]$, $|J| \leq 2k$,

$$\|(\mathbf{x} - \mathbf{u}) - h_{\mathbf{A};J}(\mathbf{x}, \mathbf{u})\|_2 \leq b_2 \delta, \quad (22)$$

where $b_2 > 0$ is a small universal constant (again see, Eq. (6)).

(d) Requiring $\rho_1 + \rho_2 = \rho$, the last step of the proof derives the RAIC claimed in the theorem by using the results from Steps (b) and (c), such that the condition holds with probability at least $1 - \rho$ uniformly in all possible cases.

We provide a more thorough overview of Steps (b) and (c) next in Section III-B4, and do likewise for Step (d) in Section III-B5.

4) Large- and Small-Distance Regimes – Steps (b) and (c)

Before discussing the approach to Steps (b) and (c), let us first motivate the argument. Let $\mathbf{x}, \mathbf{y} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$. Notice that the function $h_{\mathbf{A}}(\mathbf{x}, \mathbf{y})$ can be written as

$$\begin{aligned} &h_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) \\ &= \frac{\sqrt{2\pi}}{m} \mathbf{A}^T \cdot \frac{1}{2}\big(\mathsf{sign}(\mathbf{A}\mathbf{x}) - \mathsf{sign}(\mathbf{A}\mathbf{y})\big) \\ &= \frac{\sqrt{2\pi}}{m} \sum_{i=1}^m \mathbf{A}^{(i)} \cdot \frac{1}{2}\big(\mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{x}\rangle) - \mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{y}\rangle)\big) \\ &= \frac{\sqrt{2\pi}}{m} \sum_{i=1}^m \mathbf{A}^{(i)} \cdot \mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{x}\rangle) \\ &\qquad \cdot \mathbb{I}(\mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{x}\rangle) \neq \mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{y}\rangle)). \end{aligned}$$

Hence, given the random vector

$$\mathbf{R}_{\mathbf{x},\mathbf{y}} = \frac{1}{2}\left(\mathsf{sign}\left(\mathbf{A}\mathbf{x}\right) - \mathsf{sign}\left(\mathbf{A}\mathbf{y}\right)\right),$$

which takes values in $\{-1, 0, 1\}^m$, and defining the random variable

$$L_{\mathbf{x},\mathbf{y}} = \|\mathbf{R}_{\mathbf{x},\mathbf{y}}\|_0 = \sum_{i=1}^m \mathbb{I}(\mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{x}\rangle) \neq \mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{y}\rangle)),$$

which tracks number of *mismatches* (again, recall the discussion in Section I-A), the random vector $(h_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) \mid \mathbf{R}_{\mathbf{x},\mathbf{y}})$ becomes a function of only $L_{\mathbf{x},\mathbf{y}}$-many random vectors, where $L_{\mathbf{x},\mathbf{y}} \leq m$. Such conditioning on $\mathbf{R}_{\mathbf{x},\mathbf{y}}$ will allow for tighter concentration inequalities related to (an orthogonal decomposition of) the random vector $(h_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) \mid \mathbf{R}_{\mathbf{x},\mathbf{y}})$. Note that these concentration inequalities (detailed in the full version [1, Lemma A.1]) provide the same inequality for any $L_{\mathbf{x},\mathbf{y}} = \|\mathbf{R}_{\mathbf{x},\mathbf{y}}\|_0$ and $L_{\mathbf{x}',\mathbf{y}'} = \|\mathbf{R}_{\mathbf{x}',\mathbf{y}'}\|_0$, whenever $L_{\mathbf{x},\mathbf{y}} = L'_{\mathbf{x}',\mathbf{y}'}$, where $\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}' \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, and thus it suffices to have a handle on (an appropriate subset of) the random variables $\{L_{\mathbf{x},\mathbf{y}} : \mathbf{x}, \mathbf{y} \in \mathcal{S}^{n-1} \cap \Sigma_k^n\}$.

With this intuition in mind, we will now lay down the specifics of deriving the results achieved by Steps (b) and (c) for the "large-" and "small-distance" regimes. Each follows from two primary arguments. First, for a given $\mathbf{u}, \mathbf{v} \in \mathcal{C}_\tau$, the associated random variable $L_{\mathbf{u},\mathbf{v}}$ is bounded. Then, conditioning on $L_{\mathbf{u},\mathbf{v}}$, the desired properties in Steps (b) and (c) follow from the appropriate concentration inequalities related to the decomposition of $h_{\mathbf{A};J}(\mathbf{x}, \mathbf{y})$ into three orthogonal components.

Specifically, Step (b) is achieved as follows.

(i) Consider any $(\mathbf{u}, \mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau$ such that $d_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v}) \geq \tau$, and fix $J' \subseteq [n]$, $|J'| \leq 2k$, arbitrarily.

(ii) It can be shown that for a small $s \in (0, 1)$, the number, $L_{\mathbf{u},\mathbf{v}}$, of points among $\mathbf{A}^{(i)}, i \in [m]$, for which a mismatch occurs, i.e., $\mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{u}\rangle) \neq \mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{v}\rangle)$, is bounded in the range

$$L_{\mathbf{u},\mathbf{v}} \in \left[(1-s)\frac{\theta_{\mathbf{u},\mathbf{v}} m}{\pi}, (1+s)\frac{\theta_{\mathbf{u},\mathbf{v}} m}{\pi}\right] \quad (23)$$

uniformly with high probability for all $(\mathbf{u}, \mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau$.

(iii) Define $g_{\mathbf{A}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$\begin{aligned} g_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) =& h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) \\ &- \left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) \right\rangle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2} \\ &- \left\langle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2}, h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) \right\rangle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2} \end{aligned} \quad (24)$$

where $g_{\mathbf{A};J'}(\mathbf{u}, \mathbf{v}) = \mathcal{T}_{\mathsf{supp}(\mathbf{u})\cup\mathsf{supp}(\mathbf{v})\cup J'}(g_{\mathbf{A}}(\mathbf{u}, \mathbf{v}))$. Note that $h_{\mathbf{A}}$ and $h_{\mathbf{A};J'}$ can then be orthogonally decomposed into

$$\begin{aligned} h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) =& \left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) \right\rangle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2} \\ &+ \left\langle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2}, h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) \right\rangle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2} \end{aligned}$$

$$+ g_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) \tag{25}$$

and

$$\begin{aligned} h_{\mathbf{A};J'}(\mathbf{u}, \mathbf{v}) =& \mathcal{T}_{\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v}) \cup J'} \left(h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) \right) \\ =& \left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) \right\rangle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2} \\ &+ \left\langle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2}, h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) \right\rangle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2} \\ &+ g_{\mathbf{A};J'}(\mathbf{u}, \mathbf{v}). \end{aligned} \tag{26}$$

Note that [17] similarly uses such a decomposition to show their RAIC, and this decomposition technique appears earlier in [15].

(iv) Conditioned on $L_{\mathbf{u},\mathbf{v}} \in [(1-s)\frac{\theta_{\mathbf{u},\mathbf{v}} m}{\pi}, (1+s)\frac{\theta_{\mathbf{u},\mathbf{v}} m}{\pi}]$, the desired property in Eq. (21) is derived from Eq. (26) using a concentration inequality proved in the full version [1] together with standard techniques, e.g., the triangle inequality.

(v) A union bound extends Eq. (21) to hold uniformly over $\mathcal{C}_\tau \times \mathcal{C}_\tau$ and all $J' \subseteq [n]$, $|J'| \leq 2k$, with high probability, completely Step (b).

While Step (c) takes a similar approach, it requires a somewhat different argument involving an additional construction, as detailed next.

(i) Let $\mathbf{u} \in \mathcal{C}_\tau$ be an arbitrary vector in the τ-net, and fix any $J' \subseteq [n]$, $|J'| \leq 2k$. Recall that the desired property in Eq. (22) should hold for all $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$.

(ii) To ensure this uniform result over $\mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$, construct a second net $\mathcal{D}_\tau(\mathbf{u}) \subseteq \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$ such that for each $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$, there exits a point $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$ such that $\mathsf{sign}(\mathbf{A}\mathbf{w}) = \mathsf{sign}(\mathbf{A}\mathbf{x})$. The next step will upper bound the size of $\mathcal{D}_\tau(\mathbf{u})$.

(iii) Let $\beta = \arccos(1 - \frac{\tau^2}{2})$ be the angle associated with the distance τ, and define the random variable $M_{\beta,\mathbf{u}} = |\{\mathbf{A}^{(i)}, i \in [m] : \theta_{\mathbf{w},\mathbf{A}^{(i)}} \in [\frac{\pi}{2} - \beta, \frac{\pi}{2} + \beta]\}|$. Notice that the size of $\mathcal{D}_\tau(\mathbf{u})$ need not exceed $2^{M_{\beta,\mathbf{u}}}$. Moreover, for any $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$ with $\theta_{\mathbf{x},\mathbf{u}} \in [0, \beta]$, the value taken by the random variable $M_{\beta,\mathbf{u}}$ upper bounds the number of points $\mathbf{A}^{(i)}$, $i \in [m]$, on which $\mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{x} \rangle)$ and $\mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{u} \rangle)$ mismatch—or more formally, $L_{\mathbf{x},\mathbf{u}} \leq M_{\beta,\mathbf{u}}$ for every $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$.

(iv) By a Chernoff and union bound, the random variable $M_{\beta,\mathbf{u}}$ can be shown to be bounded from above by $M_{\beta,\mathbf{u}} \leq \frac{4}{3}\tau m$ with high probability for every $\mathbf{u} \in \mathcal{C}_\tau$, and taken with the above argument, this further implies $L_{\mathbf{x},\mathbf{u}} \leq \frac{4}{3}\tau m$ for each $\mathbf{u} \in \mathcal{C}_\tau$ and each $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$.

(v) Taking any $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$ and conditioning on $L_{\mathbf{x},\mathbf{u}}$, the norm of $h_{\mathbf{A};J'}(\mathbf{w}, \mathbf{u})$ is bounded using an orthogonal decomposition analogous to that in Step (b), and again applying the concentration inequalities in the full version [1, Lemma A.1], along with standard techniques, to obtain $\|h_{\mathbf{A};J'}(\mathbf{w}, \mathbf{u})\|_2 \leq O(\tau)$.

(vi) This bound is then extended to hold uniformly for all $\mathbf{u} \in \mathcal{C}_\tau$, $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$, and $J' \subseteq [n]$, $|J'| \leq 2k$, by union bounding.

(vii) Step (c) concludes by arguing that the uniform result from step (vi) suffices to ensure Eq. (22) holds uniformly for all $\mathbf{u} \in \mathcal{C}_\tau$, $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u})$, and $J' \subseteq [n]$, $|J'| \leq 2k$, by observing that for each $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u})$, the construction of the net, $\mathcal{D}_\tau(\mathbf{u})$, ensures the existence of $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$ such that $\|h_{\mathbf{A};J'}(\mathbf{x}, \mathbf{u})\|_2 = \|h_{\mathbf{A};J'}(\mathbf{w}, \mathbf{u})\|_2 \leq O(\tau)$. The argument additionally applies the triangle inequality: $\|(\mathbf{x} - \mathbf{u}) - h_{\mathbf{A};J'}(\mathbf{x}, \mathbf{u})\|_2 \leq \|\mathbf{x} - \mathbf{u}\|_2 + \|h_{\mathbf{A};J'}(\mathbf{x}, \mathbf{u})\|_2 \leq O(\tau)$.

5) Combining the Intermediate Results to Complete the Proof – Step (d)

The final step, Step (d), combines the results obtained in Steps (b) and (c), i.e., Eqs. (21) and (22), to conclude that the i.i.d. Gaussian measurement matrix $\mathbf{A}$ satisfies the (k, n, δ, c_1, c_2)-RAIC with bounded probability.

(i) Fix an arbitrary pair of k-sparse unit vectors $\mathbf{x}, \mathbf{y} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, and let $\mathbf{u}, \mathbf{v} \in \mathcal{C}_\tau$ be the closest net points, respectively, subject to $\mathsf{supp}(\mathbf{u}) = \mathsf{supp}(\mathbf{x})$ and $\mathsf{supp}(\mathbf{v}) = \mathsf{supp}(\mathbf{y})$. Note that our specific construction of $\mathcal{C}_\tau$ ensures that there exist net points $\mathbf{u}$ and $\mathbf{v}$ which are at most τ-far from $\mathbf{x}$ and $\mathbf{y}$, respectively, and satisfy the condition on the support sets. Additionally, it is possible to have $\mathbf{u} = \mathbf{x}$ in the case when $\mathbf{x} \in \mathcal{C}_\tau$, and likewise for $\mathbf{v}$ when $\mathbf{y} \in \mathcal{C}_\tau$. Let $J \subseteq [n]$, $|J| \leq k$, be any k-subset of coordinates. Moreover, write $J_{\mathbf{x}} = J \cup \mathsf{supp}(\mathbf{x})$ and $J_{\mathbf{y}} = J \cup \mathsf{supp}(\mathbf{y})$, each having size no more than $2k$.

(ii) It is straightforward to show with algebraic manipulation that

$$\begin{aligned} (\mathbf{x} - \mathbf{y}) - h_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) =& (\mathbf{u} - \mathbf{v}) - h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) \\ &+ (\mathbf{x} - \mathbf{u}) - h_{\mathbf{A}}(\mathbf{x}, \mathbf{u}) \\ &+ (\mathbf{v} - \mathbf{y}) - h_{\mathbf{A}}(\mathbf{v}, \mathbf{y}), \end{aligned} \tag{27}$$

and similarly that

$$\begin{aligned} (\mathbf{x} - \mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x}, \mathbf{y}) =& (\mathbf{u} - \mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u}, \mathbf{v}) \\ &+ (\mathbf{x} - \mathbf{u}) - h_{\mathbf{A};J_{\mathbf{y}}}(\mathbf{x}, \mathbf{u}) \\ &+ (\mathbf{v} - \mathbf{y}) - h_{\mathbf{A};J_{\mathbf{x}}}(\mathbf{v}, \mathbf{y}). \end{aligned} \tag{28}$$

(iii) The ℓ_2-norm of the left-hand-side of Eq. (28) can be bounded by splitting it up into the sum of three terms via the triangle inequality. Specifically,

$$\begin{aligned} &\|(\mathbf{x} - \mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x}, \mathbf{y})\|_2 \\ &\leq \|(\mathbf{u} - \mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u}, \mathbf{v})\|_2 \\ &\quad + \|(\mathbf{x} - \mathbf{u}) - h_{\mathbf{A};J_{\mathbf{y}}}(\mathbf{x}, \mathbf{u})\|_2 \\ &\quad + \|(\mathbf{v} - \mathbf{y}) - h_{\mathbf{A};J_{\mathbf{x}}}(\mathbf{v}, \mathbf{y})\|_2. \end{aligned} \tag{29}$$

(iv) Now, we divide up the argument into two cases based on whether $d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})$ is above or below the threshold τ. If $d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) < \tau$, then using the result from Step (c), we can obtain

$$\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 \leq 3b_2\delta. \tag{30}$$

Otherwise, when $d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) \geq \tau$, by applying the results from both Steps (b) and (c), we can show

$$\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 \leq b_1\sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})} + 2b_2\delta. \tag{31}$$

Above, $b_1, b_2 > 0$ are universal constants specified in Eq. (6). Both Eqs. (30) and (31) are trivially upper bounded by

$$\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 \leq b_1\sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})} + 3b_2\delta, \tag{32}$$

where this will hold with probability at least $1-\rho$.

(v) Then, adjusting the constants with the universal constants defined in Eq. (6), the (k,n,δ,c_1,c_2)-RAIC for the Gaussian measurement matrix $\mathbf{A}$ will hold with probability at least $1-\rho$, completing the proof of Theorem III.3.

IV. Proof of the Main Result—BIHT Convergence

A. Intermediate Results

Before proving the main theorems, Theorem III.1 and III.2, three intermediate results, in Lemmas IV.1-IV.3, are presented to facilitate the analysis for the convergence of BIHT approximations. The proofs for these intermediate results can be found in the full version [1].

Lemma IV.1. *Consider any $\mathbf{x} \in \mathcal{S}^{n-1}\cap\Sigma_k^n$ and any $t \in \mathbb{Z}_+$. The error of the t^{th} approximation produced by the BIHT algorithm satisfies*

$$d_{\mathcal{S}^{n-1}}\big(\mathbf{x},\hat{\mathbf{x}}^{(t)}\big) \leq 4\Big\|\big(\mathbf{x}-\hat{\mathbf{x}}^{(t-1)}\big) - h_{\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\big(\mathbf{x},\hat{\mathbf{x}}^{(t-1)}\big)\Big\|_2. \tag{33}$$

Note that Lemma IV.1 is a deterministic result, arsing from the equation by which the BIHT algorithm computes its t^{th} approximations, $t \in \mathbb{Z}_+$. Hence, it hold for all $\mathbf{x} \in \mathcal{S}^{n-1}\cap\Sigma_k^n$ and all iterations $t \in \mathbb{Z}_+$.

Lemma IV.2. *Let $\varepsilon : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ be a function given by the recurrence relation*

$$\varepsilon(0) = 2, \tag{34}$$

$$\varepsilon(t) = 4c_1\sqrt{\frac{\epsilon}{c}\varepsilon(t-1)} + 4c_2\frac{\epsilon}{c}, \quad t \in \mathbb{Z}_+. \tag{35}$$

The function ε decreases monotonically with t and asymptotically tends to a value not exceeding ϵ—formally,

$$\lim_{t\to\infty}\varepsilon(t) = \left(2c_1\left(c_1+\sqrt{c_1^2+c_2}\right)+c_2\right)\frac{4\epsilon}{c} < \epsilon. \tag{36}$$

Lemma IV.3. *Let $\varepsilon : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ be the function as defined in Lemma IV.2. Then, the sequence $\{\varepsilon(t)\}_{t\in\mathbb{Z}_{\geq 0}}$ is bound from above by the sequence $\{2^{2^{-t}}\epsilon^{1-2^{-t}}\}_{t\in\mathbb{Z}_{\geq 0}}$.*

B. Proofs of Theorems III.2 and III.1

The main theorems for the analysis of the BIHT algorithm are restated below for convenience and will subsequently be proved in tandem.

Theorem (restatement) (Theorem III.1)**.** *Let $a,b,c > 0$ be universal constants as in Eq. (6). Fix $\epsilon,\rho \in (0,1)$ and $k,m,n \in \mathbb{Z}_+$ where*

$$m \geq \frac{4bck}{\epsilon}\log\left(\frac{en}{k}\right) + \frac{2bck}{\epsilon}\log\left(\frac{12bc}{\epsilon}\right) + \frac{bc}{\epsilon}\log\left(\frac{a}{\rho}\right).$$

Let the measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ have rows with i.i.d. Gaussian entries. Then, uniformly with probability at least $1-\rho$, for every unknown k-sparse real-valued unit vector, $\mathbf{x} \in \mathcal{S}^{n-1}\cap\Sigma_k^n$, the normalized BIHT algorithm produces a sequence of approximations, $\{\hat{\mathbf{x}}^{(t)} \in \mathcal{S}^{n-1}\cap\Sigma_k^n\}_{t\in\mathbb{Z}_{\geq 0}}$, which converges to the ϵ-ball around the unknown vector $\mathbf{x}$ at a rate upper bounded by

$$d_{\mathcal{S}^{n-1}}\big(\mathbf{x},\hat{\mathbf{x}}^{(t)}\big) \leq 2^{2^{-t}}\epsilon^{1-2^{-t}}$$

for each $t \in \mathbb{Z}_{\geq 0}$.

Corollary (restatement) (Corollary III.2)**.** *Under the conditions stated in Theorem III.1, uniformly with probability at least $1-\rho$, for every unknown k-sparse real-valued unit vector, $\mathbf{x} \in \mathcal{S}^{n-1}\cap\Sigma_k^n$, the sequence of BIHT approximations, $\{\hat{\mathbf{x}}^{(t)}\}_{t\in\mathbb{Z}_{\geq 0}}$, converges asymptotically to the ϵ-ball around the unknown vector $\mathbf{x}$. Formally,*

$$\lim_{t\to\infty} d_{\mathcal{S}^{n-1}}\big(\mathbf{x},\hat{\mathbf{x}}^{(t)}\big) \leq \epsilon.$$

Proof (Theorem III.1 and Corollary III.2). The convergence of BIHT approximations for an arbitrary unknown, k-sparse unit vector, $\mathbf{x} \in \mathcal{S}^{n-1}\cap\Sigma_k^n$, will follow from the main technical theorem, Theorem III.3, and the intermediate lemmas, Lemmas IV.1-IV.3. Recalling that Theorem III.3 and Lemma IV.1 hold uniformly over $\mathcal{S}^{n-1}\cap\Sigma_k^n$ (respectively, with bounded probability and deterministically), the argument then implies uniform convergence for all unknown k-sparse vectors, $\mathbf{x} \in \mathcal{S}^{n-1}\cap\Sigma_k^n$.

Consider any unknown, k-sparse unit vector $\mathbf{x} \in \mathcal{S}^{n-1}\cap\Sigma_k^n$ with an associated sequence of BIHT approximations, $\{\hat{\mathbf{x}}^{(t)} \in \mathcal{S}^{n-1}\cap\Sigma_k^n\}_{t\in\mathbb{Z}_{\geq 0}}$. For each $t \in \mathbb{Z}_+$, Lemma IV.1 bounds the error of the t^{th} approximation from above by

$$d_{\mathcal{S}^{n-1}}\big(\mathbf{x},\hat{\mathbf{x}}^{(t)}\big) \leq 4\Big\|\big(\mathbf{x}-\hat{\mathbf{x}}^{(t-1)}\big) - h_{\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t-1)})}\big(\mathbf{x},\hat{\mathbf{x}}^{(t-1)}\big)\Big\|_2 \tag{37}$$

which is further bounded by Theorem III.3 (by setting $\delta = \frac{\epsilon}{c} = \frac{\epsilon}{32}$ in the theorem) as

$$\begin{aligned} d_{\mathcal{S}^{n-1}}\big(\mathbf{x},\hat{\mathbf{x}}^{(t)}\big) &\leq 4\Big\|\big(\mathbf{x}-\hat{\mathbf{x}}^{(t-1)}\big) - h_{\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t-1)})}\big(\mathbf{x},\hat{\mathbf{x}}^{(t-1)}\big)\Big\|_2 && \text{(38a)}\\ &\leq 4\left(c_1\sqrt{\frac{\epsilon}{c}d_{\mathcal{S}^{n-1}}\big(\mathbf{x},\hat{\mathbf{x}}^{(t-1)}\big)} + c_2\frac{\epsilon}{c}\right) && \text{(38b)}\\ &= 4c_1\sqrt{\frac{\epsilon}{c}d_{\mathcal{S}^{n-1}}\big(\mathbf{x},\hat{\mathbf{x}}^{(t-1)}\big)} + 4c_2\frac{\epsilon}{c} && \text{(38c)} \end{aligned}$$

where in the case of $t = 1$, (38c),

$$\begin{aligned} d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(1)}\big) &\le 4c_1\sqrt{\frac{\epsilon}{c} d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(0)}\big)} + 4c_2\frac{\epsilon}{c} \\ &\le 4c_1\sqrt{\frac{\epsilon}{c} d_{\mathcal{S}^{n-1}}(\mathbf{x}, -\mathbf{x})} + 4c_2\frac{\epsilon}{c} \\ &= c_1\sqrt{\epsilon} + \frac{c_2}{8}\epsilon. \end{aligned} \tag{39}$$

Recall that Lemma IV.2 defines a function $\varepsilon : \mathbb{Z}_{\ge 0} \to \mathbb{R}$ by the recurrence relation

$$\varepsilon(0) = 2, \tag{40}$$

$$\varepsilon(t) = 4c_1\sqrt{\frac{\epsilon}{c}\varepsilon(t-1)} + 4c_2\frac{\epsilon}{c}, \quad t \in \mathbb{Z}_+, \tag{41}$$

whose form is similar to (38c). It can be argued inductively that for every $t \in \mathbb{Z}_{\ge 0}$, the function $\varepsilon(t)$ upper bounds the error of the t^{th} BIHT approximation, $d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\big)$, as discussed next. The base case, $t = 0$, is trivial since

$$d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(0)}\big) \le d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, -\mathbf{x}\big) = 2 = \varepsilon(0). \tag{42}$$

Meanwhile, arbitrarily fixing $t \in \mathbb{Z}_+$, suppose that for each $t' \in [t-1]$, the error is upper bounded by

$$d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t')}\big) \le \varepsilon(t'). \tag{43}$$

Then, applying Eq. (38), the t^{th} approximation satisfies

$$\begin{aligned} d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\big) &\le 4c_1\sqrt{\frac{\epsilon}{c} d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}\big)} + 4c_2\frac{\epsilon}{c} \\ &\le 4c_1\sqrt{\frac{\epsilon}{c}\varepsilon(t-1)} + 4c_2\frac{\epsilon}{c} \\ &= \varepsilon(t) \end{aligned} \tag{44}$$

as desired. By induction, it follows that the sequence of BIHT approximations for the unknown vector $\mathbf{x}$ satisfies

$$d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\big) \le \varepsilon(t), \quad \forall\, t \in \mathbb{Z}_{\ge 0}. \tag{45}$$

Then, Lemmas IV.2 and IV.3 immediately imply the desired results since asymptotically (Lemma IV.2),

$$\begin{aligned} \lim_{t\to\infty} d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\big) &\le \lim_{t\to\infty} \varepsilon(t) \\ &= \left(2c_1\left(c_1 + \sqrt{c_1^2 + c_2}\right) + c_2\right)\frac{4\epsilon}{c} \\ &< \epsilon \end{aligned} \tag{46}$$

whereas pointwise (Lemma IV.3),

$$d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\big) \le \varepsilon(t) \le 2^{2^{-t}}\epsilon^{1-2^{-t}}. \tag{47}$$

This completes the first step of the proof. Next, the proof concludes by extending the argument to the uniform results claimed in the theorems.

As briefly mentioned at the beginning of the proof, in the argument laid out above, Lemma IV.1 and Theorem III.3 hold uniformly for every $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, where Lemma IV.1 is deterministic while Theorem III.3 ensures the bound with probability at least $1 - \rho$. Thus, for every $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, the t^{th} BIHT approximation has error upper bounded by

$$d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\big) \le 4c_1\sqrt{\frac{\epsilon}{c} d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}\big)} + 4c_2\frac{\epsilon}{c} \tag{48}$$

uniformly with probability at least $1-\rho$. Furthermore, because Lemmas IV.2 and IV.3 are deterministic, the rate of decay and asymptotic behavior stated in the theorems also hold uniformly—specifically, for all $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$,

$$\begin{aligned} \lim_{t\to\infty}\ & d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\big) \\ &\le \lim_{t\to\infty} \varepsilon(t) \\ &= \left(2c_1\left(c_1 + \sqrt{c_1^2 + c_2}\right) + c_2\right)\frac{4\epsilon}{c} \\ &< \epsilon \end{aligned} \tag{49}$$

and

$$d_{\mathcal{S}^{n-1}}\big(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\big) \le \varepsilon(t) \le 2^{2^{-t}}\epsilon^{1-2^{-t}}, \quad \forall\, t \in \mathbb{Z}_{\ge 0} \tag{50}$$

with probability at least $1 - \rho$. ■

V. Outlook

In this paper, we have shown that the binary iterative hard thresholding algorithm, an iterative (proximal) subgradient descent algorithm for a nonconvex optimization problem, converges under certain structural assumptions, with the optimal number of measurements. It is worth exploring how general this result can be: what other nonlinear measurements can be handled this way, and what type of measurement noise can be tolerated by such iterative algorithms? This direction is hopeful because the noiseless sign measurements are often thought to be the hardest to analyze. As another point of interest, our result is deterministic given a measurement matrix with a certain property. Incidentally, Gaussian measurements satisfy this property with high probability. However, the spherical symmetry of these measurements is crucial in the proof laid out in this work, and it is not clear whether other non-Gaussian (even sub-Gaussian) measurement matrices can have this property, or whether derandomized, explicit construction of measurement matrices is possible.

References

[1] N. Matsumoto and A. Mazumdar, "Binary iterative hard thresholding converges with optimal number of measurements for 1-bit compressed sensing," *arXiv preprint arXiv:2207.03427*, 2022.

[2] P. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *42nd Annual Conference on Information Sciences and Systems, CISS 2008, Princeton, NJ, USA, 19-21 March 2008*. IEEE, 2008, pp. 16–21. [Online]. Available: https://doi.org/10.1109/CISS.2008.4558487

[3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[5] Y. Plan and R. Vershynin, "The generalized lasso with non-linear observations," *IEEE Transactions on information theory*, vol. 62, no. 3, pp. 1528–1537, 2016.

[6] J. D. Haupt and R. G. Baraniuk, "Robust support recovery using sparse compressive sensing matrices," in *45st Annual Conference on Information Sciences and Systems, CISS 2011, The John Hopkins University, Baltimore, MD, USA, 23-25 March 2011*. IEEE, 2011, pp. 1–6. [Online]. Available: https://doi.org/10.1109/CISS.2011.5766202

[7] S. Gopi, P. Netrapalli, P. Jain, and A. Nori, "One-bit compressed sensing: Provable support and vector recovery," in *International Conference on Machine Learning*, 2013, pp. 154–162.

[8] J. Acharya, A. Bhattacharyya, and P. Kamath, "Improved bounds for universal one-bit compressive sensing," in *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2017, pp. 2353–2357.

[9] Y. Plan and R. Vershynin, "Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach," *IEEE Trans. Information Theory*, vol. 59, no. 1, pp. 482–494, 2013.

[10] P. Li, "One scan 1-bit compressed sensing," in *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016, Cadiz, Spain, May 9-11, 2016*, ser. JMLR Workshop and Conference Proceedings, A. Gretton and C. C. Robert, Eds., vol. 51. JMLR.org, 2016, pp. 1515–1523. [Online]. Available: http://jmlr.org/proceedings/papers/v51/li16g.html

[11] L. Jacques, J. N. Laska, P. T. Boufounos, and R. G. Baraniuk, "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2082–2102, 2013.

[12] Y. Plan and R. Vershynin, "One-bit compressed sensing by linear programming," *Communications on Pure and Applied Mathematics*, vol. 66, no. 8, pp. 1275–1297, 2013.

[13] P. T. Boufounos, L. Jacques, F. Krahmer, and R. Saab, "Quantization and compressive sensing," in *Compressed sensing and its applications*. Springer, 2015, pp. 193–237.

[14] L. Jacques, K. Degraux, and C. De Vleeschouwer, "Quantized iterative hard thresholding: Bridging 1-bit and high-resolution quantized compressed sensing," *arXiv preprint arXiv:1305.1786*, 2013.

[15] Y. Plan, R. Vershynin, and E. Yudovina, "High-dimensional estimation with geometric constraints," *Information and Inference: A Journal of the IMA*, vol. 6, no. 1, pp. 1–40, 2017.

[16] D. Liu, S. Li, and Y. Shen, "One-bit compressive sensing with projected subgradient method under sparsity constraints," *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6650–6663, 2019.

[17] M. P. Friedlander, H. Jeong, Y. Plan, and Ö. Yılmaz, "Nbiht: An efficient algorithm for 1-bit compressed sensing with optimal error decay rate," *IEEE Transactions on Information Theory*, vol. 68, no. 2, pp. 1157–1177, 2021.

[18] K. Knudson, R. Saab, and R. Ward, "One-bit compressive sensing with norm estimation," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2748–2758, 2016.

[19] R. G. Baraniuk, S. Foucart, D. Needell, Y. Plan, and M. Wootters, "Exponential decay of reconstruction error from binary measurements of sparse signals," *IEEE Transactions on Information Theory*, vol. 63, no. 6, pp. 3368–3385, 2017.

[20] R. Saab, R. Wang, and Ö. Yılmaz, "Quantization of compressive samples with stable and robust recovery," *Applied and Computational Harmonic Analysis*, vol. 44, no. 1, pp. 123–143, 2018.

[21] L. Flodin, V. Gandikota, and A. Mazumdar, "Superset technique for approximate recovery in one-bit compressed sensing," in *Advances in Neural Information Processing Systems*, 2019, pp. 10 387–10 396.

[22] A. Mazumdar and S. Pal, "Support recovery in universal one-bit compressed sensing," in *13th Innovations in Theoretical Computer Science Conference, ITCS 2022, January 31 - February 3, 2022, Berkeley, CA, USA*, ser. LIPIcs, M. Braverman, Ed., vol. 215. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022, pp. 106:1–106:20.