NBIHT: An Efficient Algorithm for 1-bit Compressed Sensing with Optimal Error Decay Rate

Michael P. Friedlander, Halyun Jeong, Yaniv Plan, and Özgür Yılmaz

***Abstract*—The *Binary Iterative Hard Thresholding* (BIHT) algorithm is a popular reconstruction method for one-bit compressed sensing due to its simplicity and fast empirical convergence. Despite considerable research on this algorithm, a theoretical understanding of the corresponding approximation error and convergence rate still remains an open problem.**

This paper shows that the normalized version of BIHT (NBIHT) achieves an approximation error rate optimal up to logarithmic factors. More precisely, using m one-bit measurements of an s-sparse vector x, we prove that the approximation error of NBIHT is of order $O\left(\frac{1}{m}\right)$ up to logarithmic factors, which matches the information-theoretic lower bound $\Omega\left(\frac{1}{m}\right)$ proved by Jacques, Laska, Boufounos, and Baraniuk in 2013. To our knowledge, this is the first theoretical analysis of a BIHT-type algorithm that explains the optimal rate of error decay empirically observed in the literature. This also makes NBIHT the first provable computationally-efficient one-bit compressed sensing algorithm that breaks the inverse square-root error decay rate $O\left(\frac{1}{m^{1/2}}\right)$.

***Index Terms*—One-bit compressed sensing, iterative reconstructions, binary iterative hard thresholding, optimal reconstruction error decay, quantization, sparse signals.**

I. Introduction

Manuscript received December 22, 2020; revised September 23, 2021.

M. P. Friedlander is with the Department of Computer Science and the Department of Mathematics, The University of British Columbia, Vancouver, BC, Canada (email:mpf@cs.ubc.ca).

H. Jeong is with the Department of Mathematics, University of California, Los Angeles, USA (email: hajeong@math.ucla.edu).

Y. Plan, and Ö Yılmaz are with the Department of Mathematics, The University of British Columbia, Vancouver, BC, Canada (email: hajeong@math.ubc.ca; yaniv@math.ubc.ca; oyilmaz@math.ubc.ca).

M. P. Friedlander was partially supported by the Office of Naval Research through grant N00014-17-1-2009. H. Jeong was funded in part by the University of British Columbia Data Science Institute (UBC DSI) and by the Pacific Institute of Mathematical Sciences (PIMS). Y. Plan was partially supported by an NSERC Discovery Grant (22R23068), a PIMS CRG 33: High-Dimensional Data Analysis, and a Tier II Canada Research Chair in Data Science. Ö. Yılmaz was partially supported by an NSERC Discovery Grant (22R82411) and PIMS CRG 33: High-Dimensional Data Analysis.

This work has been done when H. Jeong was at the University of British Columbia.

COMPRESSED sensing is an efficient signal acquisition and recovery paradigm that has received considerable attention from both researchers and practitioners. This novel paradigm has led to satisfactory solutions to many interesting problems in modern signal processing and machine learning that were once considered to be intractable by traditional approaches. In the most basic setup of compressed sensing, one aims to recover an unknown s-sparse signal $x \in \mathbb{R}^N$, i.e., a vector with at most s nonzero entries, from its linear measurements Ax where A is a known $m \times N$ matrix with $m \ll N$.

For measurement matrices with certain conditions, e.g., the *restricted isometry property (RIP)*, which is satisfied for a large class of matrices including Gaussian random matrices (random matrices whose entries are i.i.d. standard Gaussian random variables), the typical results in compressed sensing guarantee that signal recovery is possible if the number of measurements $m \gtrsim s\log(2N/s)$ [1], [2].

On the other hand, the digital nature of modern computing requires quantizing measurements to store and process them. The discrete nature of quantization presents exciting challenges in efficient signal recovery. One of the central challenges in this area is *quantized compressed sensing*, the problem of recovering a sparse vector from its quantized measurements $Q(Ax)$ where $Q(\cdot)$ is a *quantizer* that maps vectors in $\mathbb{R}^m$ to elements of a finite set. There are two main approaches to quantization in the setting of compressed sensing.

The first approach is to use noise-shaping quantizers which aim to design vector quantizers such that the quantization error in the measurement space lies in directions away from the set of all possible unquantized measurements. Thus, the quantization error can be mostly filtered out. Examples of noise-shaping quantizers in compressed sensing, which can be one-bit or multi-bit per measurement include $\Sigma\Delta$ quantization [3], [4], β-condensation [5], [6], and adaptive quatization [7]. These quantizers can reach approximation error rates that decay exponentially in m directly [8] or after additional encoding [9]. This near-optimal compression rate, however, comes at the expense of introducing memory components or sophisticated subroutine [7] in the feedback quantization (encoding) process — often not desired or sometimes not practical.

The second approach to quantization is memoryless scalar quantization (MSQ), which quantizes each measurement separately. In other words, the quantizer Q acts on Ax elementwise, so it can be implemented without the need for analog memory. Because of its non-adaptive nature in the encoding process, MSQ is more suitable to parallel or distributed-computing environments. One important case of quantized compressed sensing is when the quantizer only has two levels, i.e., it is one-bit. One-bit quantizers are easily implementable in hardware because of their simplicity, low-storage requirements, and robustness to scaling errors.

Boufounos and Baraniuk [10] first introduced *one-bit compressed sensing*. The goal is to recover an s-sparse signal x from its signed linear measurements, that is, from $b = \operatorname{sign}(Ax)$ for a Gaussian random matrix A. Here $\operatorname{sign}(\cdot)$ is applied elementwise on Ax where the scalar function $\operatorname{sign}(w)$ is equal to 1 if $w > 0$ and -1 otherwise. This is a quantized compressed sensing problem where the $\operatorname{sign}(\cdot)$ function is the one-bit memoryless scalar quantizer. Because one-bit measurements $\operatorname{sign}(Ax)$ are scale invariant, one cannot recover the magnitude of x. Consequently, in the one-bit compressed sensing problem, we may assume without loss of generality that x is a unit vector and restrict the signal set of interest to sparse vectors on the unit sphere.

Another motivation for one-bit compressed sensing is the sparse binary response model which aims to recover a sparse signal x from binary observations $\operatorname{sign}(Ax+\varepsilon)$, where ε is a noise vector [11]–[14]. Note that in these applications, we do not have access to Ax but only $\frac{1}{2}(\operatorname{sign}(Ax+\varepsilon)+1)$, its binary labels, which excludes the vector quantization idea. Here, the notation 1 indicates a vector of all-ones. Interestingly, the noiseless setting ($\varepsilon = 0$), which is identical to the one-bit compressed sensing model, has been thought difficult to analyze since a certain amount of noise is required to make the traditional maximum likelihood analysis feasible [12], [13]. Our theory provides a near-optimal theoretical guarantee in the number of observations in this case.

Jacques et al. [15] proposed *binary iterative hard thresholding (BIHT)* as a reconstruction method for one-bit compressed sensing, which is our main subject of study. BIHT and its variants are state-of-the-art reconstruction algorithms for one-bit compressed sensing because of their low computational cost and superior empirical recovery accuracy in low-noise or noiseless settings [15]–[17]. Formally, BIHT is a subgradient method coupled with hard thresolding to the set of s-sprase vectors to solve the following nonconvex optimization problem [15], [17]:

$$\min_{z\in\mathbb{R}^N} \|Az\|_1 - \langle b, Az\rangle \quad \text{subject to} \quad \|z\|_0 \le s,\ \|z\|_2 = 1. \tag{1}$$

Other BIHT-type algorithms includes *normalized BIHT (NBIHT)*, which features an addtional nonconvex projection step onto the unit sphere and the *hinge loss function BIHT* which replaces the objective in (1) with the hinge-loss function [15].

There are various aspects in the formulation (1) that make the convergence analysis of the BIHT challenging. First, the two constraints in (1) are nonconvex, so BIHT-type algorithms perform a nonconvex projection after the gradient step. Moreover, the objective function is not strictly convex and is not differentiable. As a consequence, the subgradients in BIHT-type algorithms have discontinuities, so traditional approaches based on the RIP are not directly applicable.

However, thanks to the randomness of measurement matrix A, we can use a multicale analysis to show that NBIHT converges to within a small radius of x. At large scales, i.e., for z far from the ground truth x, an RIP-type property holds. As z approaches x, the RIP constants become larger until the near isometry breaks down when z gets within the radius of the approximation error. This allows us to show that the iterates of NBIHT contract until they land within the radius of approximation error. We call this multiscale RIP-type guarantee the *restricted approximate invertibility condition (RAIC)*, and believe it may be useful in other problems involving binary measurements. This forms the foundation of our analysis for NBIHT.

Our main result shows that the NBIHT iterates x_k obey the bound

$$\|x_k - x\|_2 \le C\frac{(s\log(N/s))^{7/2}(\log m)^{12}}{m^{1-\frac{1}{2}\left(\frac{5}{6}\right)^{\lfloor k/25\rfloor -1}}}$$

for some absolute constants $C > 0$ with probability at least $1 - O\left(\frac{1}{m^3}\right)$.

To the best of our knowledge, there is no previous theoretical analysis for NBIHT despite much numerical evidence. There are some results on the convergence of BIHT (unnormalized), including [11], [16], [18], which showed that the first iteration of BIHT achieves an approximation error of order of $\frac{\sqrt{s\log(N/s)}}{m^{1/2}}$. Theorem V.1 in [17] states that the iterates of BIHT remain bounded after the first iteration and maintain the same order of accuracy for the rest of iterations. This analysis, however, does not imply that the approximation error may decrease after the first iteration. In contrast, our results show that the approximation error of NBIHT decays as $\frac{(s\log(N/s))^{7/2}(\log m)^{12}}{m}$ after sufficient iterates. On the other hand, Theorem 1 in [15] says that any one-bit compressed sensing recovery algorithms should exhibit error rate of at least order of $\frac{s}{m}$.

Since the lower bound of any reconstruction algorithm error is $\Omega(1/m)$ by that theorem, we give the first optimal dependence on m by a BIHT-type algorithm, and to our knowledge, the first provable polynomial-time one-bit compressed sensing algorithm achieving the optimal dependence. Furthermore, the error bound only depends poly-logarithmically on the ambient dimension, N, as is expected in the compressed sensing setup. We believe that the dependence on s and logarithmic factors can be improved and we conjecture that the linear dependence on s of the lower bound is achievable by NBIHT.

Lastly, we make a distinction between uniform and nonuniform recovery guarantees in compressed sensing. A nonuniform recovery problem considers the setting that aims to recover a fixed but unknown sparse vector using a random draw of the measurement matrix. On the other hand, in a uniform recovery problem, a set of sparse vectors are recovered using a same random matrix. In the development of compressed sensing, theoretical results for the nonuniform setting have often been a precursor to the uniform recovery guarantees. This includes many recent works in standard compressed sensing and a few in 1-bit compressed sensing, notably Theorem 2 in [15]. Our theory is nonuniform, partly because the theoretical guarantees for the approximation error of the first iteration of BIHT are nonuniform [11]. We leave extending our theory to the uniform recovery setting as a future research topic.

A. Related Literature

There is a vast literature about quantized compressed sensing, e.g., see [6], [16], [19]. As for one-bit compressed sensing, Plan and Vershynin [20], [21] have proposed recovery algorithms based on convex programs. These are computationally efficient and also cover more general classes of signals, but their recovery accuracy is only guaranteed of order $\sqrt{\frac{s\log(N/s)}{m}}$. Knudson et al. [22] studied one-bit compressed sensing when the observations are one-bit measurements with dither, i.e., $b = \mathrm{sign}(Ax+\varepsilon)$ where ε is a dither noise. This allows recovery of the magnitude of x under certain conditions, but their measurement model is different from ours and the recovery guarantee is still at most of order $\sqrt{\frac{s\log(N/s)}{m}}$.

In the overcomplete setting ($m \geq N$) without assuming sparsity of the signal, *consistent reconstruction* methods produce a signal w whose memoryless scalar quantized measurements $Q(Aw)$ agree with those of the ground truth x. Such methods can be implemented, for example, using linear programming with constraints that enforces the condition $Q(Aw) = Q(Ax)$. These methods offers recovery error decay of order $\frac{N}{m}$. Further considerations of computational efficiency have led to iterative signal recovery methods such as the *Rangan-Goyal algorithm* [23] or *frame permutation quantization* [24] in which only one quantized measurement is enforced at each iteration. But their approximation error is still of order $\frac{N}{m}$ [23]–[25], which would not provide a meaningful recovery when $N \gtrsim m$ (the typical compressed sensing setting); Because these algorithms are not specially designed to take into account sparsity, they assume the overcomplete setting and their error dependency on the signal dimension is $O(N)$, whereas ours is only $O(\log^{7/2} N)$.

B. Notation

Throughout this paper, we use the standard notation for Big-O and Big-Omega: If $f(n) = O(g(n))$, there exists a positive constant $C > 0$ such that $|f(n)| \leq C|g(n)|$ when n is sufficiently large, and if $f(n) = \Omega(g(n))$, there exists a positive constant $C > 0$ such that $|f(n)| \geq C|g(n)|$ for sufficiently large n.

We denote the unit sphere and the Euclidean unit ball in $\mathbb{R}^N$ respectively by $\mathbb{S}^{N-1}$ and B_2^N. The Euclidean ball centered at z with radius r is denoted by $B(z,r)$. The set of all s-sparse vectors in $\mathbb{R}^N$ is denoted by K. We write T_s for the hard thresholding operator onto K that keeps the s-largest magnitude elements of a vector and zeros out the rest of its entries. As usual, $\mathcal{N}(0,1)$ denotes the standard Gaussian distribution and $\mathcal{N}(0,I_N)$ denotes the N-dimensional standard Gaussian distribution. We assume that the measurement vectors a_j or the rows of A are the standard Gaussian random vectors in $\mathbb{R}^N$. As for norms, $\|\cdot\|$ and $\|\cdot\|_2$ denote the ℓ_2 norm for a vector, and $\|\cdot\|_1$ is the ℓ_1 norm. All other norms will be defined in later sections.

Given two sets $A, B \subset \mathbb{R}^N$, $A+B$ is the Minkowski sum of A and B, i.e., $A+B = \{v+w | v \in A, w \in B\}$. The normalized geodesic distance between two vectors x and y is $d_g(x,y) := \frac{\arccos\langle x,y\rangle}{\pi}$.

II. Global convergence of normalized BIHT

A. The normalized BIHT algorithm

We briefly describe the normalized binary iterative hard-thresholding (NBIHT) algorithm. The description of the NBIHT algorithm [15] is given in Algorithm 1.

An important property of NBIHT is that z_1 is an unbiased estimator of x for $\tau = \sqrt{\pi/2}$ (this is true for any initialization x_0 that does not depend on A). This follows from Proposition 17, which states that for any fixed unit vector $y \in \mathbb{R}^N$, $\sqrt{\frac{\pi}{2}}\mathbb{E}\left[\frac{1}{m}A^T \mathrm{sign}(Ay)\right] = y$. This shows that it is indeed unbiased, i.e., $\mathbb{E}z_1 = x$. Further iterations are

```
Algorithm 1 Normalized BIHT
Inputs: $m \times N$ matrix $A$, measurements $b = \text{sign}(Ax)$, sparsity level $s$, step size $\tau$
Initialize: Pick an $x_0$ with $||x_0|| = 1$ and sparsity level $s$.
Iterate: Repeat until a stopping criteria on $x_k$ is met
    Compute $z_{k+1}$
        $z_{k+1} := x_k + \frac{\tau}{m} A^T(\text{sign}(Ax) - \text{sign}(Ax_k))$
    Project to $s$-sparse vector sets
        $\tilde{x}_{k+1} := T_s(z_{k+1})$
    Normalize
        $x_{k+1} := \tilde{x}_{k+1}/||\tilde{x}_{k+1}||$
Outputs: The $s$-sparse approximate solution to $\text{sign}(Ax) = b$.
```

no longer unbiased estimators of x. Typically, analyses in this scenario proceed by giving uniform deviations from expectation over the whole set of values that the iterations can take (e.g., RIP). However, due to the discontinuity of the sign function, uniform bounds on the deviation (which scale according to the distance to x) are impossible. Instead, we give a multiscale approach that shows uniform deviations over a sequence of annuli of decreasing radii. See Section III for further details.

B. Main global convergence theorem

We are now ready to state our main theorem for the convergence of the NBIHT algorithm.

Theorem 1. *Let x be an s-sparse unit vector in $\mathbb{R}^N$. Assume that each entry of the measurement matrix $A \in \mathbb{R}^{m\times N}$ is drawn i.i.d. from $\mathcal{N}(0,1)$. Then, there exist positive universal constants c, C such that the iterates of NBIHT x_k with step size $\tau = \sqrt{\pi/2}$ satisfy*

$$\|x_k - x\|_2 \le C \frac{(s\log(N/s))^{7/2}(\log m)^{12}}{m^{\left(1-\frac{1}{2}\left(\frac{5}{6}\right)^{\lfloor k/25 \rfloor -1}\right)}}$$

for all $m > \max\{c\cdot s^{10}\log^{10}(N/s), 24^{48}\}$ with probability at least $1 - O\left(\frac{1}{m^3}\right)$.

Proof. See Theorem 25 and its proof. □

By letting the iteration number $k \to \infty$, Theorem 1 imples the following Corollary about the approximation error of NBIHT.

Corollary 2. *Under the same assumptions in Theorem 1, the NBIHT iterates x_k converge to x up to approximation error $C\frac{(s\log(N/s))^{7/2}(\log m)^{12}}{m}$ with probability at least $1 - O\left(\frac{1}{m^3}\right)$ provided $m > \max\{c\cdot s^{10}\log^{10}(N/s), 24^{48}\}$.*

Remark 1. *Jacques et.al. [15] provided a lower bound on the best achievable reconstruction error from one-bit (sign) measurements. Specifically, Theorem 1 in [15] shows that the approximation error decay rate cannot be faster than $\Omega(\frac{s}{m})$ no matter what algorithm we use as long as $m = \Omega(s^{3/2})$. Thus, Corollary 2 reveals that the NBIHT algorithm achieves the best possible approximation error decay rate in m.*

Remark 2. *The number of measurements required by Theorem 1 and Corollary 2 to achieve the optimal error decay rate is substantially large, in part because we have focused on proving the optimal error decay rate rather than optimizing the constant.*

III. Multiscale analysis based on restricted approximate invertibility condition

This section illustrates the idea behind the proof of our main convergence result. First, we introduce the dual norm which is used in the rest of the paper.

Definition 3. *Let G be a symmetric subset in $\mathbb{R}^N$. The dual norm of G is given by*

$$\|x\|_{G^\circ} := \sup_{u\in G}\langle x, u\rangle, \quad x \in \mathbb{R}^N.$$

We also denote $(G-G)\cap dB_2^N$ by G_d. Moreover, when G is a cone, we see that $\|x\|_{G_t^\circ} = t\|x\|_{G^\circ}$ for any $t > 0$.

Recall that in standard compressed sensing, one wishes to reconstruct an s-sparse vector x from its linear measurements Ax. There are numerous algorithms available for this problem, including *iterative hard thresholding (IHT)*, which is commonly used due to its simplicity and computational efficiency. We present IHT in Algorithm 2 because it has motivated BIHT-type algorithms and because our NBIHT convergence proof hinges, in part, on the analysis of IHT.

The proofs for the linear convergence of IHT typically start by showing $\|z_{k+1} - x\|_2 \le c\|x_k - x\|_2$ for some $0 < c < 1$. Since we want to convey the idea, let us assume $c = \frac{1}{8}$, i.e., we have $\|z_{k+1} - x\|_2 \le \frac{1}{8}\|x_k - x\|_2$. Then, a standard argument implies a contraction inequality $\|x_{k+1} - x\|_2 \le \frac{1}{2}\|x_k - x\|_2$ which leads to the linear convergence of IHT. To prove the inequality $\|z_{k+1} - x\|_2 \le \frac{1}{8}\|x_k - x\|_2$, define $\tilde{A} := \frac{1}{\sqrt{m}}A$, which satisfies the RIP with high probability for sufficiently large m [1]. Since $z_{k+1} - x = \frac{1}{m}A^T(Ax - Ax_k) - (x - x_k) = \tilde{A}^T(\tilde{A}x - \tilde{A}x_k) - (x - x_k)$, it is evident that controlling the term $\tilde{A}^T(\tilde{A}x - \tilde{A}x_k) - (x - x_k)$ plays a critical role in the convergence analysis. This term is controlled with the RIP which implies that $\tilde{A}^T\tilde{A}$ acts as approximately

Algorithm 2 IHT

```
Inputs: $m \times N$ standard Gaussian random matrix $A$, measurements $b = Ax$, sparsity level $s$
Initialize: Pick a $x_0$ with sparsity level $s$.
Iterate: Repeat until a stopping criteria on $x_k$ is met
    Compute $z_{k+1}$
        $z_{k+1} := x_k + \frac{1}{m} A^T(Ax - Ax_k)$
    Project to $s$-sparse vector sets
        $x_{k+1} := T_s(z_{k+1})$
Outputs: The approximate $s$-sparse solution to $Ax = b$.
```

the identity when restricted to sparse vectors. The RIP is equivalent to the following equation [1], which will naturally connect to the RAIC:

$$\begin{aligned}\sup_{y \in K} & \left\| \frac{1}{m} \cdot A^T(Ax - Ay) - (x-y) \right\|_{K_1^\circ} \\ &= \sup_{y \in K} \left\| \tilde{A}^T(\tilde{A}x - \tilde{A}y) - (x-y) \right\|_{K_1^\circ} \\ &\leq \frac{1}{8} ||x-y||_2.\end{aligned}$$

This is the key inequality in the proof for the linear convergence of IHT.

Unfortunately, due to the discontinuity of $\text{sign}(A\cdot)$, the measurement operator in one-bit compressed sensing does not have such handy properties; Two arbitrarily close vectors can be mapped to points which are "far" under the operator $\text{sign}(A\cdot)$. However, it turns out that the randomness of $\text{sign}(A\cdot)$ still allows us to derive a "restricted approximate invertibility condition (RAIC)" that makes it possible to replace the RIP to facilitate the analysis and we think it is interesting on its own. Informally, the RAIC is similar to RIP but holds only on an annulus at x ($r_{i+1} \leq ||x-y||_2 \leq r_i$) with a small additive error. More precisely, the RAIC with the associated parameters r_i and r_{i+1} holds if

$$\begin{aligned}\sup_{\substack{y \in K \cap \mathbb{S}^{N-1} \\ r_{i+1} \leq ||x-y||_2 \leq r_i}} & \left\| \frac{\tau}{m} \cdot A^T(\text{sign}(Ax) - \text{sign}(Ay)) - (x-y) \right\|_{K_1^\circ} \\ &\leq \frac{1}{8} ||x-y||_2 + \frac{3}{50} r_{i+1}.\end{aligned}$$

Here, $\{r_j\}_{j=1}^L$ is a finite decreasing sequence of real numbers with $r_1 \approx \frac{\sqrt{s \log(N/s)} + \sqrt{\log m}}{m^{1/2}}$ and $r_L \lesssim \frac{(s \log(N/s))^{7/2} (\log m)^{12}}{m}$, whose exact specification will be given later. It is well-known that the first iteration of the NBIHT, say x_1, satisfies $||x_1 - x|| \leq r_1$ with high probability.

Then, the rest of analysis is based on a contraction-type inequality for the approximation error (up to a small additive term) that is obtained by applying the RAIC with the current error scale. For example, if the k-th NBIHT iterate, x_k, satisfies $r_{i+1} \leq ||x - x_k||_2 \leq r_i$ for some $1 \leq i \leq L$, then we apply the RAIC with parameters r_i and r_{i+1}. Repeated applications of the RAIC combined with an elementary argument imply that x_k converges to x until it reaches the approximation error scale r_L. Lastly, once x_k reaches the error scale r_L, we show that the rest of the NIBHT iterates stay close to x with the same error scale as r_L.

IV. Outline of the proof of main technical results

In this section, we present the proof outline of the main technical results including the RAIC.

A. Sketch of the proof of our main convergence result based on the RAIC

1) Let K be the set of s-sparse vectors. First, from the gradient step iteration for the NBIHT, $z_{k+1} = x_k + \frac{\tau}{m} A^T(\text{sign}(Ax) - \text{sign}(Ax_k))$. Using a standard argument for the projections to $K \cap \mathbb{S}^{N-1}$, we will see that

$$\begin{aligned}& ||x_{k+1} - x||_2 \\ &\leq 4 \left\| \frac{\tau}{m} \cdot A^T(\text{sign}(Ax) - \text{sign}(Ax_k)) - (x - x_k) \right\|_{K_1^\circ}.\end{aligned} \tag{2}$$

2) Let L be the largest positive integer satisfying $m^{\frac{1}{40}(\frac{5}{6})^L} > 24$. Such L always exists from the assumption $m > 24^{48} = 24^{40\left(\frac{6}{5}\right)}$. Suppose i is a postive integer less than L. We define a sequence of sets, $K^{(i)} := \{y \mid \text{y is an } s\text{-sparse unit vector}, r_{i+1} \leq \|y - x\| \leq r_i\}$ where $\{r_i\}$ is a decreasing sequence of real numbers satisfying

$$r_i \lesssim m^{-\left(1 - \frac{1}{2}\left(\frac{5}{6}\right)^{i-1}\right)} (\log m)^{10} (s \log(N/s))^3.$$

By the choice of L and the sequence $\{r_i\}$, it turns out that $r_L \lesssim \frac{(s \log(N/s))^{7/2} (\log m)^{12}}{m}$. The exact specifications of r_i will be given in Section V.

3) The RAIC says that for all $y \in K^{(i)}$,

$$\begin{aligned}& \left\| \frac{\tau}{m} \cdot A^T(\text{sign}(Ax) - \text{sign}(Ay)) - (x-y) \right\|_{K_1^\circ} \\ &\leq \frac{3}{m^{\frac{1}{40}(\frac{5}{6})^i}} ||x-y||_2 + \frac{3}{50} r_{i+1}\end{aligned} \tag{3}$$

with high probability. This bound and (2) imply that

$$||x_{k+1} - x||_2 \leq \frac{12}{m^{\frac{1}{40}(\frac{5}{6})^i}} ||x - x_k||_2 + \frac{6}{25} r_{i+1}$$

for all x_k with $r_{i+1} \leq \|x - x_k\| \leq r_i$.
Since $m^{\frac{1}{40}(\frac{5}{6})^L} > 24$ and $i < L$, $m^{\frac{1}{40}(\frac{5}{6})^i} > 24$. Thus,

$$\|x_{k+1} - x\|_2 \leq \frac{1}{2}\|x_k - x\|_2 + \frac{6}{25} r_{i+1}.$$

Note that $\frac{1}{2}\|x_k - x\|_2 + \frac{6}{25} r_{i+1} \leq \frac{1}{2} r_i + \frac{1}{2} r_i \leq r_i$, so $\|x_{k+1} - x\|_2 \leq r_i$. Then, a simple induction argument yields the following: for any $t \geq 1$, either there exists p with $1 \leq p \leq t$ such that $\|x_{k+p} - x\| \leq r_{i+1}$ or

$$\|x_{k+t} - x\|_2 \leq \left(\frac{1}{2}\right)^t \|x - x_k\|_2 + \frac{12}{25} r_{i+1}$$

with high probability.
From the relation between r_i and r_{i+1}, and the previous inequality, one can easily show that $\|x_{k+p} - x\| \leq r_{i+1}$ for some p with $1 \leq p \leq 25$. In other words, 25 iterations are enough for the BIHT iterates to reach the next level $K^{(i+1)}$. Repeating this argument yields that x_k approaches x until it reaches the level set $K^{(L)}$, guaranteeing the approximation error $\approx r_L$ up to some logarithmic factors. This is the main idea of Theorem 1.

B. Sketch of the proof of the RAIC

It remains to present the idea of the proof the RAIC, which turns out to be the main challenge in our approach. Here are the key ingredients of the proof.

- Local Binary Embedding (LBE) [26]: The LBE is a refined version of the Binary ε-Stable Embedding (BεSE) by Jacques et.al. [15]. The BεSE states that the distance of any two s-sparse unit vectors is close to the normalized Hamming distance of their one-bit measurements up to additive error ε. It is also proved that the BεSE holds for an $m \times N$ Gaussian random measurement matrix with ε of order of $\sqrt{s/M}$ with high probability, but we were not able to apply BεSE to achieve our decay rate essentially because $\sqrt{\frac{1}{M}} \gg \frac{1}{M}$.
On the other hand, although the LBE resembles the BεSE, the key difference is that its additive error in the embedding becomes smaller as two vectors are sufficiently close to each other. In other words, we have more accurate control of the additive error if one vector belongs to a small neighborhood of the other.
- The standard ε-net argument is used for a uniform approximation of Lipschitz continuous parts in our analysis, whereas the LBE is applied to the discontinuous part involving the $\operatorname{sign}(\cdot)$ operation.
- We also use a concentration argument based on the bounded Bernstein's inequality in several places.

1) The first step of our proof is the orthogonal decomposition of a Gaussian random vector into three components along the directions of $x - x_k$, $x + x_k$ and their orthogonal complement to facilitate the analysis. More precisely, given a unit vector y, consider the following decomposition of each measurement vectors a_i:

$$a_i = \left\langle a_i, \frac{x-y}{\|x-y\|} \right\rangle \frac{x-y}{\|x-y\|} + \left\langle a_i, \frac{x+y}{\|x+y\|} \right\rangle \frac{x+y}{\|x+y\|} + b_i(x,y),$$

where $b_i(x,y)$ is the component of a_i orthogonal to $x-y$ and $x+y$. This decomposition allows us to decouple the right hand side of (3) into three parts, each of which has a similar structure. As will be made more precise in later sections, the analysis of each part essentially boils down to controlling the following form of a sum

$$\frac{\tau}{m}\sum_{i=1}^{m}[(\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, y\rangle))g(a_i, y)] - \tau\mathbb{E}_a[(\operatorname{sign}(\langle a, x\rangle) - \operatorname{sign}(\langle a, y\rangle))g(a, y)] \quad (4)$$

for all s-sparse unit vectors y. Here g is a jointly 1-Lipschitz continuous function bounded by $C(\sqrt{s\log(2N/s)} + \sqrt{\log m})$ for some fixed constant $C \geq 1$. For the sake of illustration, we further assume $\mathbb{E}_a[(\operatorname{sign}(\langle a, x\rangle) - \operatorname{sign}(\langle a, y\rangle))g(a, y)] = 0$ for all unit vector y.

2) We are now ready to present the idea on how to control the sum (4) by giving strings of inequalities with technical details followed by justifications. First, let $\mathscr{N}_{K^{(i)}}$ be an ε-net of $K^{(i)}$ with $\varepsilon := \delta_i$ which will be defined in Section V. So, for any $y \in K^{(i)}$, there exists $\hat{y} \in \mathscr{N}_{K^{(i)}}$ with $\|y - \hat{y}\|_2 \leq \delta_i$. Then,

$$\begin{aligned}
&\left|\frac{\tau}{m}\sum_{i=1}^{m}[(\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, y\rangle))g(a_i, y)\right| \\
&\overset{(i)}{\leq} \left|\frac{\tau}{m}\sum_{i=1}^{m}[(\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, y\rangle)g(a_i, y)] \right. \\
&\qquad \left. - \frac{\tau}{m}\sum_{i=1}^{m}[(\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle))g(a_i, y)]\right| \\
&\quad + \left|\frac{\tau}{m}\sum_{i=1}^{m}[(\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle))g(a_i, y)\right| \\
&\overset{(ii)}{\leq} \frac{\tau}{m}\sum_{i=1}^{m}|\operatorname{sign}(\langle a_i, y\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle)|\,|g(a_i, y)| \\
&\quad + \left|\frac{\tau}{m}\sum_{i=1}^{m}(\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle))g(a_i, y)\right| \\
&\overset{(iii)}{\leq} 2C\tau \cdot (\sqrt{s\log(2N/s)} + \sqrt{\log m})\log m \cdot \delta_i
\end{aligned}$$

$$
\begin{aligned}
&+\left|\frac{\tau}{m}\sum_{i=1}^{m}(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle))g(a_i,y)\right| \\
&\overset{(iv)}{\leq} 2C\tau\cdot(\sqrt{s\log(2N/s)}+\sqrt{\log m})\log m\cdot\delta_i \\
&\qquad+\frac{\tau}{m}\sum_{i=1}^{m}|\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle)| \\
&\qquad\qquad\times|g(a_i,y)-g(a_i,\hat{y})| \\
&\qquad+\left|\frac{\tau}{m}\sum_{i=1}^{m}(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle))g(a_i,\hat{y})\right| \\
&\overset{(v)}{\leq} 4C\tau\cdot(\sqrt{s\log(2N/s)}+\sqrt{\log m})\log m\cdot\delta_i \\
&\qquad+\left|\frac{\tau}{m}\sum_{i=1}^{m}(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle))g(a_i,\hat{y})\right| \\
&\overset{(vi)}{\leq} 4C\tau\cdot(\sqrt{s\log(2N/s)}+\sqrt{\log m})\log m\cdot\delta_i \\
&\qquad+\frac{2}{m^{\frac{1}{40}(\frac{5}{6})^i}}\|x-\hat{y}\|_2+\frac{1}{600}r_{i+1} \\
&\overset{(vii)}{\leq} \frac{2}{m^{\frac{1}{40}(\frac{5}{6})^i}}\|x-\hat{y}\|_2+\frac{4\tau}{600}r_{i+1} \\
&\overset{(viii)}{\leq} \frac{2}{m^{\frac{1}{40}(\frac{5}{6})^i}}\|x-y\|_2+2\|y-\hat{y}\|+\frac{4\tau}{600}r_{i+1} \\
&\overset{(ix)}{\leq} \frac{2}{m^{\frac{1}{40}(\frac{5}{6})^i}}\|x-y\|_2+\frac{3}{200}r_{i+1}.
\end{aligned}
$$

Here are justifications for the above chain of inequalities: (i), (ii), (iv), and (viii) follow from the triangle inequality. (iii) is due to the local binary embedding, the ε-net $\mathscr{N}_{K(i)}$, and the fact that g is bounded by $C(\sqrt{s\log(2N/s)}+\sqrt{\log m})$. For (v), we used a standard ε-net argument since g is 1-Lipschitz. Next, (vi) follows from the Bernstein's inequality for mean-zero bounded random variables. This step is important since we exploit the cancellation of terms in the sum based on a concentration inequality. Simple arguments without considering the cancellation effect in the sum would yield $C(\sqrt{s\log(2N/s)}+\sqrt{\log m})\log m\cdot r_{i+1}>r_{i+1}$, which is not good enough to show the RAIC used in (3). Lastly, (vii) and (ix) are from the relation between δ_i and r_{i+1}.

V. Toward the proof of NBIHT convergence

Several previous works in one-bit compressed sensing are based on the binary stable embedding property (BSE) type of inequalities [15], [18], [21]. Although this property is interesting on its own, it is not as strong as the RIP in the standard compressed sensing when we have full linear measurements for sparse signals, so we were not able to use the BSE to achieve our goal. Instead, our proof for the main theorem relies on various different tools and this section provides them before the proof appears.

We begin with the following proposition relating the approximation error of NBIHT iterates x_k to $T_s(z_k)$ in Algorithm 1.

Proposition 4. *If T_s is the hard thresholding operator and $x_k=T_s(z_k)/\|T_s(z_k)\|$, then*

$$
\begin{aligned}
\|x_k-x\|_2 &\leq \|T_s(z_k)-x_k\|_2+\|T_s(z_k)-x\|_2 \\
&\leq 2\|T_s(z_k)-x\|_2. \qquad (5)
\end{aligned}
$$

Proof. The first inequality is by the triangle inequality. The second inequality follows from the facts that $\|x\|=\|x_k\|=1$ and x_k is the closest point on the unit sphere to $T_s(z_k)$. □

A. The first iteration of normalized BIHT

The first iteration of the BIHT is investigated by several authors [11], [18]. In particular, Proposition 1 and 2 of Jacques et al. [18] state that $\|T_s(z_1)-x\|=O\left(\frac{\sqrt{s\log(N/s)}+\sqrt{\log m}}{m^{1/2}}\right)$ with high probability. The following statement makes this precise.

Proposition 5. *Let A be a $m\times N$ Gaussian random matrix. Then, there exists constant $\bar{c}\geq 1$ such that*

$$
\|T_s(z_1)-x\|\leq\frac{1}{2}\cdot\frac{\sqrt{s\log(N/s)}+\sqrt{\log m}}{m^{1/2}}
$$

with probability at least $1-\bar{c}\left((\frac{s}{N})^{5s}\cdot\frac{1}{m^5}\right)$.

Propositions 4 and 5 together imply that

$$
\|x_1-x\|\leq\frac{\sqrt{s\log(N/s)}+\sqrt{\log m}}{m^{1/2}} \qquad (6)
$$

with probability at least $1-\bar{c}\left((\frac{s}{N})^{5s}\cdot\frac{1}{m^5}\right)$. Also note that x_1 is a unit s-sparse vector since it is an iterate of NBIHT.

B. Metric projection

In this section, we introduce the notion of the restricted approximate invertibility condition (RAIC) and its connection to the approximation error $\|x_k-x\|$. We start with the definition of Gaussian width of a set.

Definition 6. *The Gaussian width of a set $G\in\mathbb{R}^N$ is defined by*

$$
w(G):=\mathbb{E}\sup_{u\in G}\langle h,u\rangle
$$

where $h\sim\mathscr{N}(0,I_N)$. Note that $w(G)=\mathbb{E}\|h\|_{G^\circ}$.

The following Corollary 8.3. in [11] allows us to control $\|x_{k+1}-x\|_2$.

Corollary 7. *Let P_G be the orthogonal projection on a star-shaped set G. Then, for $z \in G$, we have*

$$\|P_G(w)-z\|_2 \le \max\left(t, \frac{2}{t}\|w-z\|_{G_t^\circ}\right) \quad \text{for any } t>0.$$

Since T_s is the hard thresholding operator to the s-sparse vectors, by Corollary 7, we have

$$\begin{aligned}\|T_s(z_{k+1})-x\|_2 &\le \max\left(t, \frac{2}{t}\,\|z_{k+1}-x\|_{K_t^\circ}\right)\\ &\le \max\left(t, \frac{2}{t}\cdot t\,\|z_{k+1}-x\|_{K_1^\circ}\right)\\ &\le 2\,\|z_{k+1}-x\|_{K_1^\circ},\end{aligned}$$

where the second inequality follows from the fact that K is a symmetric cone and the third one is by taking $t = 2\,\|z_{k+1}-x\|_{K_1^\circ}$.

Hence, combining the above inequality, Proposition 4, and $z_{k+1} = x_k + \frac{\tau}{m}\cdot A^T(\mathrm{sign}(Ax)-\mathrm{sign}(Ax_k))$ yields

$$\begin{aligned}&\|x_{k+1}-x\|_2\\ &\le 4\left\|\frac{\tau}{m}\cdot A^T(\mathrm{sign}(Ax)-\mathrm{sign}(Ax_k))-(x-x_k)\right\|_{K_1^\circ}. \quad (7)\end{aligned}$$

Remark 3. *Note that if we had linear measurements of x, not the one-bit measurements, the restricted isometry property (RIP) would be sufficient to establish a contraction, that is, $\|x_{k+1}-x\|_2 \le \rho\|x_k-x\|_2$ for some $0<\rho<1$. Indeed, the δ_{2s}-RIP for matrix $\frac{1}{\sqrt{m}}A$ in can be recast as*

$$\max_{S\subset[N],card(S)\le 2s}\left\|\frac{1}{m}A^T|_S A|_S - I\right\|_{2\to 2} = \delta_{2s}$$

which implies that

$$\left\|\frac{1}{m}A^T(Ax-Ay)-(x-y)\right\|_{K_1^\circ} \le \delta_{2s}\|x-y\|_2$$

for all s-sparse vectors x and y (See [1] for more details).

Hence, if we had $\left\|\frac{1}{m}A^T(Ax-Ax_k)-(x-x_k)\right\|_{K_1^\circ}$ instead of $\left\|\frac{\tau}{m}\cdot A^T(sign(Ax)-sign(Ax_k))-(x-x_k)\right\|_{K_1^\circ}$ in the right hand side of (7), the RIP would give a contraction for sufficiently small δ_{2s}, which leads to a linear convergence.

Unfortunately, because of a severe discontinuity of the operator $\mathrm{sign}(A\cdot)$, we don't have the RIP. However we will show that the following property still holds, which provides "local approximate version of RIP".

Definition 8 (Restricted Approximate Invertibility Condition)**.**
We say that the matrix M satisfies the $(\nu,\delta_{2s},\eta_{2s},r_{lb},r_{ub})$-restricted approximate invertibility condition (RAIC) at x (the ground truth s-sparse unit vector) if the following inequality holds for all s-sparse unit vectors y with $r_{lb} \le \|x-y\|_2 \le r_{ub}$,

$$\begin{aligned}&\left\|\nu\cdot M^T(sign(Mx)-sign(My))-(x-y)\right\|_{K_1^\circ}\\ &\le \delta_{2s}\|x-y\|_2+\eta_{2s}.\end{aligned}$$

Note that the RAIC is similar to the RIP except it holds for a certain region at x with an additive error η_{2s}. As will be more clear in Proposition 18, we have the RAIC for the matrix A with high probability within certain regions around x. At this point, one would notice that the RAIC can be applied to (7) under appropriate conditions.

C. Orthogonal decomposition of measurement vectors

We will use the following orthogonal decomposition of Gaussian measurement vectors a_i, which is inspired by the decomposition technique in Plan et. al [11], as the first step in the proof of Theorem 1 in the next section.

Lemma 9. *Suppose that a_i's are the standard Gaussian random vectors. Let x, y be unit vectors in $\mathbb{R}^N$. Since $x-y$ and $x+y$ are orthogonal to each other, then we have*

$$\begin{aligned}a_i = &\left\langle a_i, \frac{x-y}{\|x-y\|}\right\rangle\frac{x-y}{\|x-y\|}\\ &+\left\langle a_i, \frac{x+y}{\|x+y\|}\right\rangle\frac{x+y}{\|x+y\|}+b_i(x,y)\end{aligned}$$

where $b_i(x,y)$ is the component of a_i orthogonal to $x-y$ and $x+y$.

Proof. Since x,y are unit vectors, one can easily check that $\left\langle a_i, \frac{x-y}{\|x-y\|}\right\rangle\frac{x-y}{\|x-y\|}$ and $\left\langle a_i, \frac{x+y}{\|x+y\|}\right\rangle\frac{x+y}{\|x+y\|}$ are orthogonal by a direct calculation. □

D. Uniform bounds

Since we will use the Bernstein's inequality for bounded random variables later, we need to show that $\left|\left\langle a_i, \frac{x+y}{\|x+y\|}\right\rangle\right|$, $\left|\left\langle a_i, \frac{x-y}{\|x-y\|}\right\rangle\right|$, $\|b_i(x,y)\|_{K_1^\circ}$ are bounded with high probability. First, the following lemma provides an upper bound of $\|a\|_{K_1^\circ}$ for the standard Gaussian random vector a.

Proposition 10. *Suppose $a\sim\mathcal{N}(0,I_N)$. Then there exist absolute constants $C_b \ge 1$ and $0<c_b\le 1$ such that with probability at least $1-\frac{2}{m^5}$, we have*

$$\|a\|_{K_1^\circ} \le C_b\sqrt{s\log(N/s)}+\sqrt{\frac{5\log m}{c_b}}.$$

Proof. From the definition of the dual norm $\|a\|_{K_1^\circ}$,

$$\mathop{\mathbb{E}}_a\|a\|_{K_1^\circ} = \mathop{\mathbb{E}}_a \sup_{\substack{v-w\in K-K\\ \|v-w\|_2\le 1}}\langle a, v-w\rangle$$

$$= w(K_1) \le C_b\sqrt{s\log(N/s)},$$

where the inequality is by a well-known bound of Gaussian width of the set of $2s$-sparse unit vectors for some constant $C_b \ge 1$.

Since $a \to \|a\|_{K_1^\circ}$ is a Lipschitz continuous function with Lipschitz constant at most 1, by the Gaussian concentration inequality [27], [28],

$$\|a\|_{K_1^\circ} \le C_b\sqrt{s\log(N/s)} + u \tag{8}$$

with probability at least $1-2\exp(-c_b u^2)$ for some constant $0 < c_b \le 1$. Finally, set $u = \sqrt{\frac{5\log m}{c_b}}$ to obtain the proposition. □

Define the quantity

$$C(N,s,m) := 3\left(C_b\sqrt{s\log(N/s)} + \sqrt{\frac{5\log m}{c_b}}\right). \tag{9}$$

Note that $C(N,s,m) \le 9C_b\sqrt{s\log(N/s)} \cdot \sqrt{\frac{5\log m}{c_b}}$.

Let E_i be the event that the bound in Proposition 10 holds for a_i. Note that $\mathbb{P}(E_i) \ge 1 - \frac{2}{m^5}$ and also define E_u such that

$$E_u = \cap_{i=1}^m E_i.$$

Then, by the union bound, $\mathbb{P}(E_u) \ge 1 - \frac{2}{m^4}$ which implies that $\max\limits_{1\le j\le m} \|a_j\|_{K_1^\circ} \le C(N,s,m)$ with probability at least $1-\frac{2}{m^4}$.

Lemma 11. *Under the event E_u, for all i with $1 \le i \le m$, we have*

$$\sup_{y\in K\cap\mathbb{S}^{N-1}} \left\{\left|\left\langle a_i, \frac{x+y}{\|x+y\|}\right\rangle\right|, \left|\left\langle a_i, \frac{x-y}{\|x-y\|}\right\rangle\right|, \|b_i(x,y)\|_{K_1^\circ}\right\}$$
$$\le C(N,s,m)$$

and

$$\max\left\{|\langle a_i,z\rangle - \langle a_i,\hat{z}\rangle|, \|\langle a_i,z\rangle z - \langle a_i,\hat{z}\rangle\hat{z}\|_{K_1^\circ}\right\}$$
$$\le 2\delta\cdot C(N,s,m)$$

for any s-sparse unit vectors $z,\hat{z}$ with $\|z-\hat{z}\|_2 \le \delta$. Moreover the event E_u occurs with probability at least $1-\frac{2}{m^4}$.

Proof. See Appendix A for the proof of this lemma. □

E. Local binary embedding for small regions at x

One of key ingredients of our proof of the RAIC is the local binary embedding (local sensitivity hashing) property by Oymak and Recht [26]. Bilyk and Lacey [29] also reported a similar property.

Theorem 12 (Local δ-binary embedding [26], [29])**.** *Let $A \in \mathbb{R}^{m\times N}$ be a standard Gaussian random matrix. Then, there exists universal constants $C_l, C_L \ge 1$, and $\tilde{c} > 0$ such that given a set $G \in \mathbb{S}^{N-1}$ and a constant $\delta \in (0,1)$, we have*

1) *For all $x,y \in G$ with $\|x-y\|_2 \le \delta/\sqrt{\log\delta^{-1}}$, $\left|\frac{1}{2m}\|sign(Ax)-sign(Ay)\|_1 - d_g(x,y)\right| \le C_L\delta$,*
2) *Conversely for all $x,y \in G$ with $\left|\frac{1}{2m}\|sign(Ax)-sign(Ay)\|_1 - d_g(x,y)\right| \le C_L\delta$, $\|x-y\|_2 \le \delta$,*

with probability $1-\exp(-\tilde{c}\delta m)$ whenever $m \ge C_l\delta^{-3}\log\delta^{-1}w^2(G)$. Here $w(G)$ is the Gaussian width of G.

Define a constant C_{10} as $C_{10} := \max\{C_l, C_L, 2C_b^2\} + \pi$. We are looking for two sequences $\{r_i\}_{i=1}$ and $\{\delta_i\}_{i=1}$ satisfying the following properties.

1) $r_1 = m^{-1/2}C(N,s,m)$.
2) $\delta_i = C(N,s,m)\left(\frac{r_i^2\log m}{m}\right)^{1/3}\log m$.
3) $r_{i+1}^2 = 600C_{10}\log m \cdot r_i\delta_i C(N,s,m)$ or $r_{i+1} = 10\sqrt{6}C_{10}^{1/2}r_i^{5/6}m^{-1/6}(\log m)^{7/6}C(N,s,m)$.

Proposition 13. *It is easy to see that $\delta_i \ge \frac{1}{m}$ and $r_i \ge \frac{1}{m}$ for all i, so $\log m \ge \log\delta_i^{-1}$ and $\log m \ge \log r_i^{-1}$. Also note that $\frac{r_{i+1}}{600} = C_{10}\frac{r_i\delta_i\log m\cdot C(N,s,m)}{r_{i+1}}$ and $\frac{r_{i+1}}{10\sqrt{6}} = (C_{10}r_i\delta_i\log m\cdot C(N,s,m))^{1/2}$.*

The following two sequences $\{r_i\}$ and $\{\delta_i\}$ are constructed by induction based on above three requirements for the sequences above.

Definition 14. *For $i \ge 0$,*

$$\begin{aligned} r_{i+1} &= (600C_{10})^{3\left(1-\left(\frac{5}{6}\right)^i\right)} r_1^{\left(\frac{5}{6}\right)^i} m^{-\left(1-\left(\frac{5}{6}\right)^i\right)} \\ &\quad \times (\log m)^{7\left(1-\left(\frac{5}{6}\right)^i\right)} C(N,s,m)^{6\left(1-\left(\frac{5}{6}\right)^i\right)}, \\ &= (600C_{10})^{3\left(1-\left(\frac{5}{6}\right)^i\right)} m^{-\left(1-\frac{1}{2}\left(\frac{5}{6}\right)^i\right)} \\ &\quad \times (\log m)^{7\left(1-\left(\frac{5}{6}\right)^i\right)} C(N,s,m)^{6-5\left(\frac{5}{6}\right)^i} \end{aligned}$$

and

$$\begin{aligned} \delta_{i+1} &= (600C_{10})^{2\left(1-\left(\frac{5}{6}\right)^i\right)} m^{-\left(1-\frac{1}{3}\left(\frac{5}{6}\right)^i\right)} \\ &\quad \times (\log m)^{\frac{14}{3}\left(1-\left(\frac{5}{6}\right)^i\right)+\frac{4}{3}} C(N,s,m)^{5-\frac{10}{3}\left(\frac{5}{6}\right)^i}. \end{aligned}$$

Proposition 15. *From the definitions of r_i, r_{i+1}, and $C(N,s,m)$, it is straightforward to check that there exists $c > 0$ such that $r_{i+1} \le r_i$, $\frac{r_{i+1}}{600} \le C_{10}\delta_i\log m\cdot C(N,s,m)$, and $\delta_i \le \frac{r_{i+1}}{600}$ as long as $m > cs^{10}\log^{10}(N/s)$.*

Let $K^{(i)} := B(x,r_i)\cap K$.

Then, for $y \in K^{(i)}$, $y-x \in B(0,r_i)\cap(K-K)$, so $y \in B(0,r_i)\cap(K-K)+x$, i.e., $K^{(i)} \subset B(0,r_i)\cap(K-K)+x$. Hence, we have

$$\begin{aligned} w(K^{(i)}) &= w(B(x,r_i)\cap K) \\ &\le w(B(0,r_i)\cap(K-K)+x) \end{aligned}$$

$$
\begin{aligned}
&\quad (\text{ for any sets } S,T \text{ with } S\subseteq T, w(S)\le w(T) \text{)}\\
&= w(B(0,r_i)\cap(K-K))\\
&\quad (\text{ for any set } S, w(S+x)=w(S) \text{)}\\
&\le r_i\cdot C_b\sqrt{2s\log(N/s)}.
\end{aligned}
$$

Next, we apply the local binary embedding to the set $K^{(i)}$ as follows.

Corollary 16 (Corollary of local δ-binary embedding)**.** *Under the same notations in Theorem 12, we have*

1) *For all $y\in K^{(i)}$ with $\|x-y\|_2\le\delta_i$, $\left|\frac{1}{2m}\|sign(Ax)-sign(Ay)\|_1\right|\le C_{10}\delta_i\log m$,*
2) *For all $y\in K^{(i)}$ with $\|x-y\|_2\le r_i$, $\left|\frac{1}{2m}\|sign(Ax)-sign(Ay)\|_1\right|\le C_{10}r_i\log m$,*

with probability $1-\frac{c_2}{m^5}$ for some universal constant $c_2>0$.

Proof. First, choose δ such that $\delta_i=\delta/\sqrt{\log\delta^{-1}}$. Since $y\in K^{(i)}$ with $\|x-y\|_2\le\delta_i$ and $\log m\ge\log\delta_i^{-1}\ge\log\delta^{-1}$ in Proposition 13, the first part of Theorem 12 implies that

$$\left|\frac{1}{2m}\|\mathrm{sign}(Ax)-\mathrm{sign}(Ay)\|_1-d_g(x,y)\right|\le C_L\delta_i\log m.$$

Next, note that $d_g(x,y)\le\pi\|x-y\|_2$ since x, y are unit vectors, so $d_g(x,y)\le\pi\delta_i$ from the relation between the arc and chord lengths in the unit sphere. Thus, from the fact that $C_{10}\ge C_L+\pi$, we have $\left|\frac{1}{2m}\|\mathrm{sign}(Ax)-\mathrm{sign}(Ay)\|_1\right|\le C_{10}\delta_i\log m$ with probability at least

$$
\begin{aligned}
&1-\exp(-\tilde{c}\delta m)\\
&\ge 1-\exp(-\tilde{c}\delta_i m)\\
&\ge 1-\exp\Bigg(-\tilde{c}(600C_{10})^{2\left(1-\left(\frac{5}{6}\right)^{i-1}\right)}m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}\\
&\qquad\qquad\times(\log m)^{\frac{14}{3}\left(1-\left(\frac{5}{6}\right)^{i-1}\right)+\frac{4}{3}}C(N,s,m)^{5-\frac{10}{3}\left(\frac{5}{6}\right)^{i-1}}\Bigg)\\
&\ge 1-\exp\left(-\tilde{c}(\log m)^2\right)\ge 1-\frac{c_2}{m^5}
\end{aligned}
$$

for some absolute constant $c_2>0$, if the condition $m\ge C_l\delta^{-3}\log\delta^{-1}w^2(K^{(i)})$ is met. By the construction of δ_i and $C(N,s,m)$, we have $m\ge\left(\frac{\delta_i}{\log m}\right)^{-3}r_i^2\log m\cdot C^3(N,s,m)\ge C_l\delta^{-3}\log\delta^{-1}w^2(K^{(i)})$, so this condition is satisfied. This proves the first part of the corollary and the second part follows from the same arguments. □

F. ε-net for $K^{(i)}$.

Consider an ε-net $\mathscr{N}_{K^{(i)}}$ for $K^{(i)}$ with $\varepsilon=\delta_i$. Then, by the Sudakov minorization inequality [Theorem 7.4.1 in Vershynin [28]], there exist constants $C',C^{''}>0$ such that for all $i\ge1$,

$$
\begin{aligned}
&\log|\mathscr{N}_{K^{(i)}}|\\
&\le\frac{C'w^2(K^{(i)})}{\delta_i^2}\\
&\le\frac{C'r_i^2w^2(K)}{\delta_i^2}\\
&\le C'(600C_{10})^{2\left(1-\left(\frac{5}{6}\right)^{i-1}\right)}m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}(\log m)^{2\left(1-\frac{7}{3}\left(\frac{5}{6}\right)^{i-1}\right)}\\
&\qquad\times C(N,s,m)^{2-\frac{10}{3}\left(\frac{5}{6}\right)^{i-1}}w^2(K)\\
&\le C^{''}(600C_{10})^{2\left(1-\left(\frac{5}{6}\right)^{i-1}\right)}m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}(\log m)^{2\left(1-\frac{7}{3}\left(\frac{5}{6}\right)^{i-1}\right)}\\
&\qquad\times C(N,s,m)^{4-\frac{10}{3}\left(\frac{5}{6}\right)^{i-1}}.
\end{aligned}
$$

G. *Correlation between sign$\langle a,x\rangle$ and $\langle a,y\rangle$*

Proposition 17. *Let u,v be unit vectors and a be a Gaussian random vector in $\mathbb{R}^N$. Then,*

$$\sqrt{\frac{\pi}{2}}\mathbb{E}[(sign\langle a,u\rangle)\langle a,v\rangle]=\langle u,v\rangle$$

and

$$\sqrt{\frac{\pi}{2}}\mathbb{E}\Big[\frac{1}{m}A^T sign(Au)\Big]=u.$$

Proof. See the proof of Lemma 4.1 in [21] for the first equality. For the second equality, consider the standard basis vectors $e_1,\dots,e_N$ for $\mathbb{R}^N$. For each $1\le i\le N$, using the first equality for unit vectors e_i and u gives us

$$\sqrt{\frac{\pi}{2}}\mathbb{E}[(\mathrm{sign}\langle a,u\rangle)\langle a,e_i\rangle]=\langle u,e_i\rangle=u_i,$$

where u_i is the i-th component of u. After rewriting these N equalities in a vector form, we have

$$\sqrt{\frac{\pi}{2}}\mathbb{E}[(\mathrm{sign}\langle a,u\rangle)a]=u.$$

Then, the second equality follows because

$$
\begin{aligned}
\sqrt{\frac{\pi}{2}}\mathbb{E}\Big[\frac{1}{m}A^T\mathrm{sign}(Au)\Big]&=\sqrt{\frac{\pi}{2}}\mathbb{E}\Big[\frac{1}{m}\sum_{i=1}^m a_i\mathrm{sign}(\langle a_i,u\rangle)\Big]\\
&=\frac{1}{m}\sum_{i=1}^m\sqrt{\frac{\pi}{2}}\mathbb{E}\Big[\mathrm{sign}(\langle a_i,u\rangle)a_i\Big]\\
&=u.
\end{aligned}
$$

□

VI. Main technical propositions

This section proves our main theorem and technical propositions. Our goal is to establish the RAIC for a certain small region around x which is crucial in our analysis. With all the facts gathered in the previous section, we are prepared to state our main technical proposition for RAIC precisely.

Proposition 18. *Let x be an s-sparse unit vector in $\mathbb{R}^N$. Then, the following bound holds uniformly for any s-sparse unit vector y with $r_{i+1} \leq \|x-y\| \leq r_i$.*

$$\left\| \frac{\tau}{m} \cdot A^T(sign(Ax) - sign(Ay)) - (x-y) \right\|_{K_1^\circ} \leq \frac{3}{m^{\frac{1}{40}(\frac{5}{6})^i}} \|x-y\|_2 + \frac{3}{50} r_{i+1}$$

with probability exceeding

$$1 - 3\bar{c}\left(\frac{s}{N}\right)^{5s} \cdot \frac{1}{m^5} - \frac{c_9}{m^4}$$

for some universal constants $\bar{c}, c_9 > 0$.

In other words, the measurement matrix A satisfies $\left(\sqrt{\frac{2}{\pi}} \cdot \frac{1}{m}, \frac{3}{m^{\frac{1}{40}(\frac{5}{6})^i}}, \frac{3}{50} r_{i+1}, r_{i+1}, r_i\right)$-RAIC with high probability.

A. Orthogonal decomposition of measurement vectors

The first step of the proof of Proposition 18 is the decomposition of $\frac{\tau}{m}A^T(\text{sign}(Ax) - \text{sign}(Ay)) - (x-y)$. Essentially, it decomposes into three parts: the components along the direction $x-y$, $x+y$, and their orthogonal part. This decomposition is based on Lemma 9.

Proof of Proposition 18 . Using the expansion of a_i in Lemma 9 and the triangle inequality, we have the chain of inequalities starting with (10). The chain of inequalities brings to control the three terms $(I), (II)$, and (III) in (11) and (12), which is the main technical challenge of our work. The subsequent three lemmas provide upper bounds for these terms, which will be proven in the next section.

Lemma 19. *There exist universal constants $c, \bar{c} > 0$ such that for all s-sparse unit vector y with $r_i \leq \|x-y\|_2 \leq r_{i+1}$, we have*

$$(I) \leq \frac{1}{m^{\frac{1}{40}(\frac{5}{6})^i}} \|x-y\|_2 + \frac{r_{i+1}}{50}$$

for all $m > cs^{10}\log^{10}(N/s)$ with probability at least $1 - \bar{c}\left(\frac{s}{N}\right)^{5s} \cdot \frac{1}{m^5} - \frac{c_6}{m^4}$

Lemma 20. *There exist universal constants $c, \bar{c} > 0$ such that for all s-sparse unit vector y with $r_i \leq \|x-y\|_2 \leq r_{i+1}$,*

$$(II) \leq \frac{2}{m^{\frac{1}{40}(\frac{5}{6})^i}} \|x-y\|_2 + \frac{r_{i+1}}{50}$$

for all $m > cs^{10}\log^{10}(N/s)$ with probability at least $1 - \bar{c}\left(\frac{s}{N}\right)^{5s} \cdot \frac{1}{m^5} - \frac{c_6}{m^4}$.

Lemma 21. *There exist universal constants $c, \bar{c} > 0$ such that for all s-sparse unit vector y with $r_i \leq \|x-y\|_2 \leq r_{i+1}$,*

$$(III) \leq \frac{r_{i+1}}{50}$$

for all $m > cs^{10}\log^{10}(N/s)$ with probability at least $1 - \bar{c}\left(\frac{s}{N}\right)^{5s} \cdot \frac{1}{m^5} - \frac{c_8}{m^4}$.

Applying Lemma 19, 20, and 21 to terms (I), (II), (III) and setting the constant $c_9 := 2c_6 + c_8$ complete the proof of Proposition 18.

□

Proposition 18 leads to our main global convergence theorems.

Theorem 22. *If x_k, the k-th iterate of NBIHT satisfies $r_{i+1} \leq \|x_k - x\| \leq r_i$ for some i, then we have*

$$\|x_{k+1} - x\|_2 \leq \frac{12}{m^{\frac{1}{40}(\frac{5}{6})^i}} \|x_k - x\|_2 + \frac{6}{25} r_{i+1}$$

with probability at least

$$1 - 3\bar{c}\left(\frac{s}{N}\right)^{5s} \cdot \frac{1}{m^5} - \frac{c_9}{m^4}.$$

Proof of Theorem 22.

Applying Proposition 18 to x_k and using the relation $z_{k+1} = x_k + \frac{\tau}{m}A^T(\text{sign}(Ax) - \text{sign}(Ax_k))$ yield

$$\|z_{k+1} - x\|_{K_1^\circ} \leq \frac{3}{m^{\frac{1}{40}(\frac{5}{6})^i}} \|x-y\|_2 + \frac{3}{50} r_{i+1}.$$

Then, from (7),

$$\begin{aligned} \|x_{k+1} - x\|_2 &\leq 4\left(\frac{3}{m^{\frac{1}{40}(\frac{5}{6})^i}} \|x-y\|_2 + \frac{3}{50} r_{i+1}\right) \\ &\leq \frac{12}{m^{\frac{1}{40}(\frac{5}{6})^i}} \|x_k - x\|_2 + \frac{6}{25} r_{i+1} \end{aligned}$$

□

Corollary 23. *Suppose $m^{\frac{1}{40}(\frac{5}{6})^L} > 24$ for some integer $L > 0$. Let i be an integer with $1 \leq i \leq L$ and x_k satisfy $r_{i+1} \leq \|x_k - x\| \leq r_i$. Then, we have for any $t \geq 1$, either there exists p with $1 \leq p \leq t$ such that $\|x_{k+p} - x\| \leq r_{i+1}$ or*

$$\|x_{k+t} - x\|_2 \leq \frac{1}{2^t} \cdot \frac{1}{m^{\frac{1}{40}((\frac{5}{6})^i - (\frac{5}{6})^L)}} \|x_k - x\|_2 + \frac{12}{25} r_{i+1}.$$

Proof. Since $m^{\frac{1}{40}(\frac{5}{6})^L} > 24$, $\frac{12}{m^{\frac{1}{40}(\frac{5}{6})^i}} < \frac{1}{2} \cdot \frac{1}{m^{\frac{1}{40}((\frac{5}{6})^i - (\frac{5}{6})^L)}}$. Thus, from Theorem 22

$$\|x_{k+1} - x\|_2 \leq \frac{1}{2} \cdot \frac{1}{m^{\frac{1}{40}((\frac{5}{6})^i - (\frac{5}{6})^L)}} \|x_k - x\|_2 + \frac{6}{25} r_{i+1}.$$

Note that $\frac{1}{2}\|x_k - x\|_2 + \frac{6}{25} r_{i+1} \leq \frac{1}{2} r_i + \frac{1}{2} r_i \leq r_i$, so $\|x_{k+1} - x\|_2 \leq r_i$.

$$\left\|\frac{\tau}{m}\cdot A^T(\mathrm{sign}(Ax)-\mathrm{sign}(Ay))-(x-y)\right\|_{K_1^\circ} \tag{10}$$

$$\begin{aligned}
&=\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)a_i-(x-y)\right\|_{K_1^\circ}\\
&\le\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle))\left(\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle\frac{x-y}{\|x-y\|}+\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\frac{x+y}{\|x+y\|}\right)-(x-y)\right\|_{K_1^\circ}\\
&\quad+\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)b_i(x,y)\right\|_{K_1^\circ}\\
&\le\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle))\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle\frac{x-y}{\|x-y\|}-(x-y)\right\|_{K_1^\circ}\\
&\quad+\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle))\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\frac{x+y}{\|x+y\|}\right\|_{K_1^\circ}+\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)b_i(x,y)\right\|_{K_1^\circ}\\
&\le\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle))\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle-\|x-y\|\right|\left\|\frac{x-y}{\|x-y\|}\right\|_{K_1^\circ}\\
&\quad+\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle))\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right|\left\|\frac{x+y}{\|x+y\|}\right\|_{K_1^\circ}+\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)b_i(x,y)\right\|_{K_1^\circ}
\end{aligned}$$

$$=\underbrace{\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle))\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle-\|x-y\|\right|}_{(I)} \tag{11}$$

$$+\underbrace{\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle))\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right|}_{(II)}+\underbrace{\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)b_i(x,y)\right\|_{K_1^\circ}}_{(III)}. \tag{12}$$

After noticing $\frac{1}{2}\cdot\frac{1}{m^{\frac{1}{40}\left((\frac{5}{6})^i-(\frac{5}{6})^L\right)}}\le\frac{1}{2}$, one can show that the following by induction: For any $t\ge 1$, either there exists p with $1\le p\le t$ such that $\|x_{k+p}-x\|\le r_{i+1}$ or

$$\|x_{k+t}-x\|_2\le\frac{1}{2^t}\cdot\frac{1}{m^{\frac{t}{40}\left((\frac{5}{6})^i-(\frac{5}{6})^L\right)}}\|x_k-x\|_2+\frac{12}{25}r_{i+1}$$

with high probability.

□

From the relation between r_i and r_{i+1}, it turns out that $\|x_{k+p}-x\|\le r_{i+1}$ for some p with $1\le p\le 25$ whenever $x_k\in K^i$. In other words, 25 iterations are enough for the BIHT iterates to reach next level $K^{(i+1)}$. This is basically the idea of Theorem 1, which we present in the following corollary.

Corollary 24. *Suppose $m^{\frac{1}{40}(\frac{5}{6})^L}>24$ for some positive integer L. Then, for any integer i with $0\le i<L$, if $\|x_k-x\|\le r_i$, there exists $t\le 25$ such that*

$$\|x_{k+j}-x\|_2\le r_i\quad\textit{for all } 0\le j\le t$$

and

$$\|x_{k+t}-x\|_2\le r_{i+1}.$$

Proof. Suppose the claim in the Corollary is not true. Then, there exists $t>25$ such that $\|x_{k+p}-x\|>r_{i+1}$ for all p with $1\le p\le t$. Also, $\|x_{k+p}-x\|\le r_i$ as in the proof in Corollary 23. On the other hand, it is easy to check that

$$\begin{aligned}
&t+t\left(1-\left(\frac{5}{6}\right)^{L-i}\right)\cdot\frac{1}{40}\left(\frac{5}{6}\right)^{i}\log_2 m\\
&\ge\log_2 25+\frac{1}{10}\left(\frac{5}{6}\right)^{i}\log_2 m\quad\text{for } t>25.
\end{aligned}$$

Then, from Definition 14 for r_i,r_{i+1}, this implies that $\frac{1}{2^t}\cdot\frac{1}{m^{\frac{t}{40}\left((\frac{5}{6})^i-(\frac{5}{6})^L\right)}}r_i\le\frac{1}{25}r_{i+1}$ by taking the logarithm to the base 2 on this equality. However, then we have $\|x_{k+25}-x\|\le\frac{1}{25}r_{i+1}+\frac{12}{25}r_{i+1}\le r_{i+1}$ by Corollary 23, which contradicts to the assumption that $\|x_{k+25}-x\|>r_{i+1}$. □

We restate Theorem 1 for the convenience of readers.

Theorem 25. *Suppose* $m > \max\{cs^{10}\log^{10}(N/s), 24^{48}\}$ *for a universal constant* $c > 0$*. Then, there exists a universal constant* $C_{12} > 0$ *such that the NBIHT iterates obey*

$$\|x_k - x\|_2 \leq C_{12}\frac{(s\log(N/s))^{7/2}(\log m)^{12}}{m^{\left(1-\frac{1}{2}\left(\frac{5}{6}\right)^{\lfloor k/25\rfloor-1}\right)}}$$

for $k \geq 1$ *with probability exceeding* $1 - O\left(\frac{1}{m^3}\right)$*.*

Proof of Theorem 25 . Our proof is based on Corollary 24 and a similar type of argument used to show the stability of the Truncated Wirtinger Flow [30].

Let L be the largest positive integer satisfying $m^{\frac{1}{40}(\frac{5}{6})^L} > 24$. Note that such L always exists from the assumption $m > 24^{48} = 24^{40\left(\frac{6}{5}\right)}$. We shall call $\{y : \|y-x\| > r_L\}$ Regime A and $\{y : \|y-x\| \leq r_L\}$ Regime B.

First, for x_k belongs to Regime A, applying Corollary 24 repeatedly shows that there exists $k \leq 25i$ such that $\|x_k - x\| \leq r_i$ as long as $i \leq L$.

Next, we will prove that if x_k is in Regime B ($\|x_k - x\| \leq r_L$), then $\|x_{k+1} - x\| \leq 8C_{10}r_L\log m C(N,s,m)$. In other words, for the approximation error of x_{k+1} is still well-controlled for x_k belonging to Regime B. To see this, we start from (7),

$$\begin{aligned}
&\|x_{k+1} - x\|_2 \\
&\leq 4\left\|\frac{\tau}{m}\cdot A^T(\mathrm{sign}(Ax) - \mathrm{sign}(Ax_k)) - (x - x_k)\right\|_{K_1^\circ} \\
&\leq 4\left\|\frac{\tau}{m}\cdot A^T(\mathrm{sign}(Ax) - \mathrm{sign}(Ax_k))\right\|_{K_1^\circ} \\
&\qquad + 4\|x - x_k\|_{K_1^\circ} \\
&\leq 4\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\mathrm{sign}(\langle a_i, x\rangle) - \mathrm{sign}(\langle a_i, x_k\rangle)a_i\right\|_{K_1^\circ} \\
&\qquad + 4\|x - x_k\|_{K_1^\circ} \\
&\leq 4\cdot\frac{\tau}{m}\sum_{i=1}^{m}|\mathrm{sign}(\langle a_i, x\rangle) - \mathrm{sign}(\langle a_i, x_k\rangle)|\,\|a_i\|_{K_1^\circ} \\
&\qquad + 4\|x - x_k\|_{K_1^\circ} \\
&\leq 4\max_{1\leq j\leq m}\|a_j\|_{K_1^\circ}\cdot\frac{\tau}{m}\sum_{i=1}^{m}|\mathrm{sign}(\langle a_i, x\rangle) - \mathrm{sign}(\langle a_i, x_k\rangle)| \\
&\qquad + 4\|x - x_k\|_{K_1^\circ} \\
&\leq 4C_{10}r_L\log m\cdot C(N,s,m) + 4\|x - x_k\|_2 \\
&\leq 4C_{10}r_L\log m\cdot C(N,s,m) + 4r_L \\
&\leq 8C_{10}r_L\log m\cdot C(N,s,m)
\end{aligned}$$

where the first five inequalities follow from the triangle inequality. The sixth inequality arises from Corollary 16 and the fact about the event E_u in Section V-D.

If x_{k+1} lands on the region A (i.e., $\|x_{k+1} - x\| > r_L$), then we can again apply Corollary 24. So the rest of iterations are guaranteed to satisfy

$$\left\|x_{k+j} - x\right\|_2 \leq 8C_{10}\log m\cdot C(N,s,m)r_L$$

for all $j \geq 1$. Hence, above argument, Definition 14, and (9) yield

$$\begin{aligned}
\|x_k - x\|_2 &\leq 8C_{10}\log m\cdot C(N,s,m)\cdot\max\{r_{\lfloor k/25\rfloor}, r_L\} \\
&\leq C_{11}\frac{(s\log(N/s))^{7/2}(\log m)^{12}}{m^{\left(1-\frac{1}{2}\left(\frac{5}{6}\right)^{\min\{\lfloor k/25\rfloor, L\}-1}\right)}}
\end{aligned}$$

for some universal constant C_{11}.

Now, note that the above expression for the error bound can be rewritten as

$$\begin{aligned}
&C_{11}\frac{(s\log(N/s))^{7/2}(\log m)^{12}}{m^{\left(1-\frac{1}{2}\left(\frac{5}{6}\right)^{\min\{\lfloor k/25\rfloor, L\}-1}\right)}} \qquad (13) \\
&= C_{11}\frac{(s\log(N/s))^{7/2}(\log m)^{12}m^{\frac{1}{2}\left(\frac{5}{6}\right)^{\min\{\lfloor k/25\rfloor, L\}-1}}}{m} \\
&= C_{11}\frac{(s\log(N/s))^{7/2}(\log m)^{12}}{m} \qquad (14) \\
&\qquad \times\max\left\{m^{\frac{1}{2}\left(\frac{5}{6}\right)^{\lfloor k/25\rfloor-1}}, m^{\frac{1}{2}\left(\frac{5}{6}\right)^{L-1}}\right\} \\
&\leq C_{11}\frac{(s\log(N/s))^{7/2}(\log m)^{12}\cdot m^{\frac{1}{2}\left(\frac{5}{6}\right)^{\lfloor k/25\rfloor-1}}\cdot m^{\frac{1}{2}\left(\frac{5}{6}\right)^{L-1}}}{m}. \qquad (15)
\end{aligned}$$

From the definition of L, we have $m^{\frac{1}{40}(\frac{5}{6})^{L+1}} \leq 24$ or $\frac{1}{40}(\frac{5}{6})^{L+1}\log m \leq \log 24$. This implies that $\frac{1}{2}\left(\frac{5}{6}\right)^{L-1} \lesssim \frac{1}{\log m}$, i.e., there exists a universal constant $\hat{c} > 0$ such that $\frac{1}{2}\left(\frac{5}{6}\right)^{L-1} \leq \frac{\hat{c}}{\log m}$. Since $m^{\frac{1}{2}\left(\frac{5}{6}\right)^{L-1}} \leq m^{\frac{\hat{c}}{\log m}} = \exp\left(\frac{\hat{c}}{\log m}\cdot\log m\right) = \exp(\hat{c})$, which is another constant, so applying this fact to (15) implies that

$$C_{11}\frac{(s\log(N/s))^{7/2}(\log m)^{12}}{m^{\left(1-\frac{1}{2}\left(\frac{5}{6}\right)^{\min\{\lfloor k/25\rfloor, L\}-1}\right)}} \leq C_{12}\frac{(s\log(N/s))^{7/2}(\log m)^{12}}{m^{\left(1-\frac{1}{2}\left(\frac{5}{6}\right)^{\lfloor k/25\rfloor-1}\right)}}$$

for some universal constant $C_{12} := C_{11}\cdot\exp(\hat{c}) > 0$. Hence, we establish that

$$\|x_k - x\|_2 \leq C_{12}\frac{(s\log(N/s))^{7/2}(\log m)^{12}}{m^{\left(1-\frac{1}{2}\left(\frac{5}{6}\right)^{\lfloor k/25\rfloor-1}\right)}}.$$

As for the probability of success, we apply the union bound over all the levels $K_1, K_2, \ldots, K_L$ where each holds with probability at least

$$1 - \left(3\bar{c}\left(\frac{s}{N}\right)^{5s}\cdot\frac{1}{m^5} - \frac{c_9}{m^4}\right),$$

from Theorem 22. Thus, the success probability should be at least

$$1 - L\left(3\bar{c}\left(\frac{s}{N}\right)^{5s}\cdot\frac{1}{m^5} - \frac{c_9}{m^4}\right) \geq 1 - O\left(\frac{1}{m^3}\right),$$

where the last inequality is by observing $L = O(\log\log m)$, which can obtained from $m^{\frac{1}{40}(\frac{5}{6})^L} > 24$. This proves the theorem.

□

VII. Proofs for Section VI

In this section, we prove Lemma 19, 20, and 21. The ideas for their proofs are similar and we start with the proof of Lemma 20 because we believe it is simpler than the other two.

A. Proof of Lemma 20

For $y \in K^{(i)}$, choose $\hat{y}$ in $\mathscr{N}_{K^{(i)}}$ with $\|y-\hat{y}\| \le \delta_i$. Note that since $y \in K^{(i)}$, $\|x-y\| \le r_i$.

a) Step 1. Approximate the term (II) by ε-net $\mathscr{N}_{K^{(i)}}$: Continuing from (II), we obtain

$$\begin{aligned}
&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right| \\
\overset{(i)}{\le}&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right| \\
&+\left|\frac{1}{m}\sum_{i=1}^{m}\left[\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right]\right. \\
&\qquad\left.-\frac{1}{m}\sum_{i=1}^{m}\left[\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right]\right| \\
\overset{(ii)}{\le}&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right| \\
&+\frac{1}{m}\sum_{i=1}^{m}\tau|\mathrm{sign}(\langle a_i,y\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle)|\left|\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right| \\
\overset{(iii)}{\le}&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle)\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right| \\
&+C(N,s,m)\cdot\frac{\tau}{m}\sum_{i=1}^{m}|\mathrm{sign}(\langle a_i,y\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle)| \\
\overset{(iv)}{\le}&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right| \\
&+2C_{10}\tau\cdot\log m\cdot C(N,s,m)\delta_i,
\end{aligned}$$

as long as we have Lemma 11 (uniform bound lemma) and Corollary 16 (local binary embedding). Here, in (i) and (ii) follow by the triangle inequality. To have (iii), we used the uniform bound for $\left|\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right|$ in Lemma 11. The inequality (iv) is due to Corollary 16.

Now consider the first term in the right side of (iv):

$$\begin{aligned}
&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right| \\
\overset{(v)}{\le}&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+\hat{y}}{\|x+y\|}\right\rangle\right| \\
&+\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\right. \\
&\qquad\left.\times\left(\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle-\left\langle a_i,\frac{x+\hat{y}}{\|x+y\|}\right\rangle\right)\right| \\
\overset{(vi)}{\le}&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+\hat{y}}{\|x+y\|}\right\rangle\right| \\
&+\frac{1}{m}\sum_{i=1}^{m}\tau\,|(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))| \\
&\qquad\times\left|\left(\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle-\left\langle a_i,\frac{x+\hat{y}}{\|x+y\|}\right\rangle\right)\right| \\
\overset{(vii)}{\le}&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+\hat{y}}{\|x+y\|}\right\rangle\right| \\
&+\frac{2C(N,s,m)\delta_i}{\|x+y\|}\cdot\frac{1}{m}\sum_{i=1}^{m}\tau\,|(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))| \\
\overset{(viii)}{\le}&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+\hat{y}}{\|x+y\|}\right\rangle\right| \\
&+2C_{10}\tau\frac{r_i\log m\cdot C(N,s,m)\delta_i}{\|x+y\|_2} \\
\overset{(ix)}{\le}&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+\hat{y}}{\|x+y\|}\right\rangle\right| \\
&+4C_{10}\tau\cdot r_i\log m\cdot C(N,s,m)\delta_i \\
=&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+\hat{y}}{\|x+\hat{y}\|}\right\rangle\right|\cdot\frac{\|x+\hat{y}\|}{\|x+y\|} \\
&+4C_{10}\tau\cdot r_i\log m\cdot C(N,s,m)\delta_i.
\end{aligned}$$

In the chain of inequalities above, (v) and (vi) follow from the triangle inequality. We applied Lemma 11 to have (vii). To get (viii), first note that $\|\hat{y}-x\| \le r_i$ since $\hat{y} \in K^{(i)}$ and apply Corollary 16. (ix) is easily follows from $\|x+y\| \ge 1$ since $y \in K^{(i)}$ implying $\|x-y\| \le r_1 \ll 1$.

Hence, we obtain

$$\begin{aligned}
&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right| \\
\le&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+\hat{y}}{\|x+\hat{y}\|}\right\rangle\right|\cdot\frac{\|x+\hat{y}\|}{\|x+y\|} \\
&+2C_{10}\tau\cdot\log m\cdot C(N,s,m)\delta_i+4C_{10}\tau\cdot r_i\log m\cdot C(N,s,m)\delta_i \\
\le&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+\hat{y}}{\|x+\hat{y}\|}\right\rangle\right|\cdot\frac{\|x+\hat{y}\|}{\|x+y\|} \\
&+\frac{2\tau r_{i+1}}{600}+\frac{4\tau r_{i+1}}{600} \\
\le&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i,\frac{x+\hat{y}}{\|x+\hat{y}\|}\right\rangle\right|\cdot\frac{\|x+\hat{y}\|}{\|x+y\|} \\
&+\frac{8r_{i+1}}{600},
\end{aligned}$$

where we used Proposition 15 in the second inequality.

b) Step 2. Bounding the mean and variance of truncated terms: To keep notation light, let us define a random variable $\mathbf{X_2}(a_i,x,y)$ by

$$\mathbf{X_2}(a_i,x,y)=(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,y\rangle))\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle. \tag{16}$$

First note that $\mathbf{X_2}(a_i,x,y)$ is mean-zero which can be verified by direct expansion of the right hand side of (16) and Proposition 17. This implies

$$\begin{aligned}0&=\mathbb{E}\left[\mathbf{X_2}(a_i,x,\hat{y})\right]\\&=\mathbb{E}\left[\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i}\right]+\mathbb{E}\left[\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i^c}\right]\end{aligned}$$

where the event E_i is defined in Section V-D.

Hence, we obtain

$$\begin{aligned}&|\mathbb{E}\left[\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i}\right]|\\&\leq\mathbb{E}\left[|\mathbf{X_2}(a_i,x,\hat{y})|\,\mathbb{1}_{E_i^c}\right]\\&\leq\mathbb{E}\left[|\mathbf{X_2}(a_i,x,\hat{y})|^2\right]^{1/2}\cdot[\mathbb{E}\mathbb{1}_{E_i^c}]^{1/2}\\&\leq\mathbb{E}\left[|\mathbf{X_2}(a_i,x,\hat{y})|^2\right]^{1/2}\cdot\mathbb{P}(E_i^c)^{1/2}\\&=\mathbb{E}\left[(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle))^2\left|\left\langle a_i,\frac{x+\hat{y}}{\|x+\hat{y}\|}\right\rangle\right|^2\right]^{1/2}\\&\qquad\times\mathbb{P}(E_i^c)^{1/2}\\&\leq\mathbb{E}\left[(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle))^4\right]^{1/4}\mathbb{E}\left[\left\langle a_i,\frac{x+\hat{y}}{\|x+\hat{y}\|}\right\rangle^4\right]^{1/4}\\&\qquad\times\mathbb{P}(E_i^c)^{1/2}\\&\leq 2\cdot 3^{1/4}\cdot\frac{\sqrt{2}}{m^2}\\&\leq 4\cdot\frac{\sqrt{2}}{m^2}.\end{aligned}$$

Here, the second and fourth inequalities are by the Cauchy-Schwartz inequality. The second last line is due to the facts that $|\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle)|\leq 2$, the fourth moment of the standard Gaussian random variable is 3, and $\mathbb{P}(E_i)\geq 1-\frac{2}{m^5}$.

By the construction of the event E_i and Lemma 11, $\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i}$ is bounded by

$$|\mathbf{X_2}(a,x,\hat{y})\mathbb{1}_{E_i}|\leq 2C(N,s,m)$$

and its second moment is bounded (so is its variance) by

$$\begin{aligned}&\mathbb{E}_{a_i}\left[\mathbf{X_2}(a,x,\hat{y})^2\mathbb{1}_{E_i}\right]\\&=\mathbb{E}_{a_i}\left[(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle))^2\left\langle a_i,\frac{x+\hat{y}}{\|x+\hat{y}\|}\right\rangle^2\mathbb{1}_{E_i}\right]\\&\leq C(N,s,m)^2\mathbb{E}_{a_i}\left[(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle))^2\mathbb{1}_{E_i}\right]\\&\leq C(N,s,m)^2\mathbb{E}_{a_i}\left[(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle))^2\right]\\&\leq 4C(N,s,m)^2\cdot d_g(x,\hat{y})\\&\leq 4\pi C(N,s,m)^2\|x-\hat{y}\|_2,\end{aligned}$$

where $d_g(x,\hat{y})$ is the normalized geodesic distance between x and $\hat{y}$. Here, we applied Lemma 11 to get the first inequality and used

$$\begin{aligned}&\mathbb{E}_{a_i}\left[(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle))^2\right]\\&\quad=4\mathbb{P}_{a_i}(\operatorname{sign}(\langle a_i,x\rangle)\neq\operatorname{sign}(\langle a_i,\hat{y}\rangle))=4d_g(x,\hat{y}),\end{aligned}$$

which is due to the rotation invariance of the standard Gaussian vector.

c) Step 3. Bounding the sum of truncated terms by the Bernstein's inequality: Next, we apply the Bernstein's inequality for mean-zero bounded random variables to have

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i}-\mathbb{E}[\mathbf{X_2}(a,x,\hat{y})\mathbb{1}_{E_i}])\right|\geq t\right)\\&\qquad\leq 2\exp\left(-\frac{m^2t^2/2}{\sigma^2+Kmt/3}\right),\end{aligned}$$

where σ^2 is the sum of the variances and K is the bound of the random variables $\tau\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i}$. Because $\sigma^2\leq\tau^2\cdot m\mathbb{E}_{a_i}\left[\mathbf{X_2}(a,x,\hat{y})^2\mathbb{1}_{E_i}\right]$ and $K\leq 2\tau C(N,s,m)$ from Step 2, we obtain

$$\begin{aligned}&\mathbb{P}\left(\left|\frac{1}{m}\sum_{i=1}^{m}\tau\left(\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i}-\mathbb{E}\left[\mathbf{X_2}(a,x,\hat{y})\mathbb{1}_{E_i}\right]\right)\right|\geq t\right)\\&\leq 2\exp\left(-\frac{mt^2/2}{\tau^2\,\mathbb{E}_{a_i}\left[\mathbf{X_2}(a,x,\hat{y})^2\mathbb{1}_{E_i}\right]+2\tau C(N,s,m)t/3}\right)\\&\leq 2\exp\left(-\frac{mt^2/2}{4\pi\tau^2C(N,s,m)^2\cdot\|x-\hat{y}\|_2+2\tau C(N,s,m)t/3}\right)\\&\leq 2\exp\left(-\frac{mt^2/2}{4\pi\tau^2C(N,s,m)^2\cdot\|x-\hat{y}\|_2+4\pi\tau^2C(N,s,m)^2t/3}\right)\end{aligned}$$

where we used the upper estimates of the first and second moments of $\mathbf{X_2}(a,x,\hat{y})\mathbb{1}_{E_i}$ in Step 2.

Choose $t=\frac{1}{m^{1/2-\xi_i}}\|x-\hat{y}\|_2$ in which $\xi_i\in[0,1/2]$ will be determined later. Then, we have

$$\begin{aligned}&\frac{mt^2/2}{4\pi\tau^2C(N,s,m)^2\cdot\|x-\hat{y}\|_2+4\pi\tau^2C(N,s,m)^2t/3}\\&=\frac{m^{2\xi_i}\|x-\hat{y}\|_2^2}{8\pi\tau^2C(N,s,m)^2\cdot[\|x-\hat{y}\|_2+\frac{1}{m^{1/2-\xi_i}}\|x-\hat{y}\|_2/3]}\\&\geq\frac{m^{2\xi_i}\|x-\hat{y}\|_2^2}{16\pi\tau^2C(N,s,m)^2\cdot[\max\{\|x-\hat{y}\|_2,\frac{1}{m^{1/2-\xi_i}}\|x-\hat{y}\|_2/3\}]}\\&\geq\frac{\min\{m^{2\xi_i}\|x-\hat{y}\|_2,3m^{1/2+\xi_i}\|x-\hat{y}\|_2\}}{16\pi\tau^2C(N,s,m)^2}.\end{aligned}$$

Using the fact that $r_{i+1}\leq\|x-\hat{y}\|\leq r_i$ since $\hat{y}\in\mathcal{N}_{K^{(i)}}\subset K^{(i)}$, above further reduces to

$$\begin{aligned}&\frac{1}{16\pi\tau^2C(N,s,m)^2}\cdot\min\{m^{2\xi_i}\|x-\hat{y}\|_2,3m^{1/2+\xi_i}\|x-\hat{y}\|_2\}\\&\quad\geq\frac{1}{16\pi\tau^2C(N,s,m)^2}\|x-\hat{y}\|_2\cdot\min\{m^{2\xi_i},3m^{1/2+\xi_i}\}\end{aligned}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3124598, IEEE Transactions on Information Theory

$$\geq \frac{1}{16\pi\tau^2 C(N,s,m)^2} r_{i+1} \cdot \min\{m^{2\xi_i}, 3m^{1/2+\xi_i}\}$$
$$\geq \frac{1}{80C(N,s,m)^2} r_{i+1} m^{2\xi_i},$$

where the last inequality is from the fact that $0 \leq \xi_i \leq 1/2$.

Thus, with a probability at least

$$1 - 2\exp\left(-\frac{r_{i+1}m^{2\xi_i}}{80C(N,s,m)^2}\right),$$

we have

$$\left|\frac{1}{m}\sum_{i=1}^{m} \tau\left(\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i} - \mathbb{E}\left[\mathbf{X_2}(a,x,\hat{y})\mathbb{1}_{E_i}\right]\right)\right|$$
$$\leq \frac{1}{m^{1/2-\xi_i}}\|x-\hat{y}\|_2. \tag{17}$$

By the union bound over all $\hat{y}$ in $\mathcal{N}_{K(i)}$, above bound holds uniformly for all $\hat{y} \in \mathcal{N}_{K(i)}$ with probability at least (18), where the bound for $|\mathcal{N}_{K(i)}|$ is from Subsection V-F and r_{i+1} is from Definition 14.

Note that ξ_i should satisfy $2\xi_i - \left(1 - \frac{1}{2}(\frac{5}{6})^i\right) \geq \frac{3}{8}(\frac{5}{6})^{i-1}$ for this to hold with a high probability. Now choose $\xi := \frac{1}{2}\left(1 - \frac{1}{20}(\frac{5}{6})^i\right)$ which gives the chain of inequality starting with (19). for some small fixed constants $C^{(3)}, C^{(4)} > 0$. Here the inequality (20) follows from $m^{\frac{1}{20}} \geq C(N,s,m)$ if $m > cs^{10}\log^{10}(N/s)$ for a sufficiently large constant $c > 0$.

Hence, we have

$$1 - 2|\mathcal{N}_{K(i)}|\exp\left(-\frac{r_{i+1}m^{2\xi_i}}{80C(N,s,m)^2}\right)$$
$$\geq 1 - 2\exp\left(-\frac{C^{(4)}}{2}(\log m)^{\frac{7}{6}}\right)$$
$$\geq 1 - \frac{c_4}{m^4}$$

for some absolute constant $c_4 > 0$ and for all $m > cs^{10}\log^{10}(N/s)$ where c is a sufficiently large constant.

d) Step 4. Establishing the bound for (II): Continuing from Step 1, recall that (II) is now bounded by

$$\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\operatorname{sign}(\langle a_i,x\rangle) - \operatorname{sign}(\langle a_i,y\rangle))\left\langle a_i, \frac{x+y}{\|x+y\|}\right\rangle\right|$$
$$\leq \left|\frac{1}{m}\sum_{i=1}^{m}\tau(\operatorname{sign}(\langle a_i,x\rangle) - \operatorname{sign}(\langle a_i,\hat{y}\rangle))\left\langle a_i, \frac{x+\hat{y}}{\|x+\hat{y}\|}\right\rangle\right| \frac{\|x+\hat{y}\|}{\|x+y\|}$$
$$+ \frac{8r_{i+1}}{600}$$
$$\leq 2\left|\frac{1}{m}\sum_{i=1}^{m}\tau\mathbf{X_2}(a_i,x,\hat{y})\right| + \frac{8r_{i+1}}{600}.$$

Also by the definition of the event E_u, whenever the event E_u occurs, we have

$$\frac{1}{m}\sum_{i=1}^{m}\tau\mathbf{X_2}(a_i,x,\hat{y}) = \frac{1}{m}\sum_{i=1}^{m}\tau\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i}.$$

Hence, by combining all the results together, for all y with $r_{i+1} \leq \|x-y\| \leq r_i$, we have

$$\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\operatorname{sign}(\langle a_i,x\rangle) - \operatorname{sign}(\langle a_i,y\rangle))\left\langle a_i, \frac{x+y}{\|x+y\|}\right\rangle\right|$$
$$\overset{(a)}{\leq} 2\left|\frac{1}{m}\sum_{i=1}^{m}\tau\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i}\right| + \frac{8r_{i+1}}{600}$$
$$\overset{(b)}{\leq} 2\tau\left|\frac{1}{m}\sum_{i=1}^{m}\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_u} - \mathbb{E}\left[\mathbf{X_2}(a,x,\hat{y})\mathbb{1}_{E_u}\right]\right|$$
$$+ 2\tau\left|\mathbb{E}\left[\mathbf{X_2}(a,x,\hat{y})\mathbb{1}_{E_u}\right]\right| + \frac{8r_{i+1}}{600}$$
$$\overset{(c)}{\leq} \frac{2}{m^{\frac{1}{40}(\frac{5}{6})^i}}\|x-\hat{y}\|_2 + \frac{8\sqrt{2}\tau}{m^2} + \frac{8r_{i+1}}{600}$$
$$\overset{(d)}{\leq} \frac{2}{m^{\frac{1}{40}(\frac{5}{6})^i}}(\|x-y\|_2 + \delta_i) + \frac{8\sqrt{2}\tau}{m^2} + \frac{8r_{i+1}}{600}$$
$$\overset{(e)}{\leq} \frac{2}{m^{\frac{1}{40}(\frac{5}{6})^i}}\|x-y\|_2 + 2\delta_i + \frac{8\sqrt{2}\tau}{m^2} + \frac{8r_{i+1}}{600}$$
$$\overset{(f)}{\leq} \frac{2}{m^{\frac{1}{40}(\frac{5}{6})^i}}\|x-y\|_2 + \frac{2}{600}r_{i+1} + \frac{1}{600}r_{i+1} + \frac{8r_{i+1}}{600}$$
$$\leq \frac{2}{m^{\frac{1}{40}(\frac{5}{6})^i}}\|x-y\|_2 + \frac{12r_{i+1}}{600}$$

for all $m > cs^{10}\log^{10}(N/s)$. Here, the first inequality follows from the assumption that E_u occurs. We applied the triangle inequality to have (b), and (d). The inequality (c) results from (17), $\xi = \frac{1}{2}\left(1 - \frac{1}{20}(\frac{5}{6})^i\right)$, and the bound for $|\mathbb{E}[\mathbf{X_2}(a,x,\hat{y})\mathbb{1}_{E_u}]|$ in Step 2. The inequality (e) follows from $m^{\frac{1}{40}(\frac{5}{6})^i} > m^{\frac{1}{40}(\frac{5}{6})^L} > 24$. We applied Proposition 15 to get (f).

Combining the results we have so far, (II) is bounded by

$$(II) \leq \frac{2}{m^{\frac{1}{40}(\frac{5}{6})^i}}\|x-y\|_2 + \frac{r_{i+1}}{50}.$$

As for the success probability for this bound to hold, we need Lemma 11, Corollary 16, the event E_u occurred, the concentration bound in Step 3, and the error bound for the first iteration of NBIHT (6). Hence, applying the union bound gives us

$$1 - \frac{2}{m^4} - \frac{c_2}{m^5} - \frac{2}{m^4} - \frac{c_4}{m^4} - \bar{c}\left(\frac{s}{N}\right)^{5s}\cdot\frac{1}{m^5}$$
$$\geq 1 - \bar{c}\left(\frac{s}{N}\right)^{5s}\cdot\frac{1}{m^5} - \frac{c_6}{m^4}.$$

B. Proof of Lemma 19

This subsection is devoted to derive the bound for the term (I). The idea is quite similar to the proof of Lemma 20, so we omitted some parts of the proof to avoid repetitions.

$$\begin{aligned}
&1-2|\mathcal{N}_{K^{(i)}}|\exp\left(-\frac{r_{i+1}m^{2\xi_i}}{80C(N,s,m)^2}\right)\\
&\geq 1-2\exp\left(C''(600C_{10})^{2\left(1-\left(\frac{5}{6}\right)^i\right)}m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}(\log m)^{2\left(1-\frac{7}{3}\left(\frac{5}{6}\right)^{i-1}\right)}C(N,s,m)^{4-\frac{10}{3}\left(\frac{5}{6}\right)^{i-1}}\right)\\
&\qquad\qquad\times\exp\left(-(600C_{10})^{3\left(1-\left(\frac{5}{6}\right)^i\right)}\frac{m^{2\xi_i-\left(1-\frac{1}{2}\left(\frac{5}{6}\right)^i\right)}(\log m)^{7\left(1-\left(\frac{5}{6}\right)^i\right)}C(N,s,m)^{6-5\left(\frac{5}{6}\right)^i}}{80C(N,s,m)^2}\right) \qquad (18)
\end{aligned}$$

$$\begin{aligned}
&C''(600C_{10})^{2\left(1-\left(\frac{5}{6}\right)^{i-1}\right)}m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}(\log m)^{2\left(1-\frac{7}{3}\left(\frac{5}{6}\right)^{i-1}\right)}C(N,s,m)^{4-\frac{10}{3}\left(\frac{5}{6}\right)^{i-1}} \qquad (19)\\
&\quad-(600C_{10})^{3\left(1-\left(\frac{5}{6}\right)^i\right)}m^{\frac{3}{8}\left(\frac{5}{6}\right)^{i-1}}(\log m)^{7\left(1-\left(\frac{5}{6}\right)^i\right)}C(N,s,m)^{4-5\left(\frac{5}{6}\right)^i}\\
&\leq C^{(3)}m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}\left[(\log m)^{2\left(1-\frac{7}{3}\left(\frac{5}{6}\right)^i\right)}C(N,s,m)^{4-4\left(\frac{5}{6}\right)^i}-m^{\frac{1}{24}\left(\frac{5}{6}\right)^{i-1}}(\log m)^{7\left(1-\left(\frac{5}{6}\right)^i\right)}C(N,s,m)^{4-5\left(\frac{5}{6}\right)^i}\right]\\
&=C^{(3)}m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}C(N,s,m)^{4-4\left(\frac{5}{6}\right)^i}\left[(\log m)^{2\left(1-\frac{7}{3}\left(\frac{5}{6}\right)^i\right)}-(\log m)^{7\left(1-\left(\frac{5}{6}\right)^i\right)}m^{\frac{1}{20}\left(\frac{5}{6}\right)^i}C(N,s,m)^{-\left(\frac{5}{6}\right)^i}\right]\\
&\leq C^{(4)}m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}C(N,s,m)^{4-4\left(\frac{5}{6}\right)^i}\left[(\log m)^{2\left(1-\frac{7}{3}\left(\frac{5}{6}\right)^i\right)}-(\log m)^{7\left(1-\left(\frac{5}{6}\right)^i\right)}\right] \qquad (20)\\
&\leq-\frac{C^{(4)}}{2}m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}(\log m)^{\frac{7}{6}}\\
&\leq-\frac{C^{(4)}}{2}(\log m)^{\frac{7}{6}}
\end{aligned}$$

As before let $y\in K^{(i)}$, so $r_{i+1}\leq\|x-y\|\leq r_i$. We proceed with the similar arguments used for the bound of (II) in subsection VII-A.

a) Step 1. Approximate (I) *by* ε*-net* $\mathcal{N}_{K^{(i)}}$*:* We begin with approximating (I) with the ε-net $\mathcal{N}_{K^{(i)}}$. The following inequality holds

$$\begin{aligned}
&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle-\|x-y\|\right|\\
&\leq\left|\frac{\tau}{m}\sum_{i=1}^{m}(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle)\left\langle a_i,\frac{x-\hat{y}}{\|x-\hat{y}\|}\right\rangle-\|x-\hat{y}\|\right| \qquad (21)\\
&+\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle)\left\langle a_i,\frac{x-\hat{y}}{\|x-\hat{y}\|}\right\rangle\right.\\
&\quad\left.-\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle\right| \qquad (22)\\
&+|\|x-y\|-\|x-\hat{y}\||
\end{aligned}$$

provided we have Lemma 11 and Corollary 16.

The third term in the right hand side is bounded by the triangle inequality:

$$|\|x-y\|-\|x-\hat{y}\||\leq\|x-y-x+\hat{y}\|=\|y-\hat{y}\|\leq\delta_i.$$

b) Step 2. Bounding the mean and variance of truncated terms: Define a random variable $\mathbf{X_1}(a_i,x,y)$ as

$$\mathbf{X_1}(a_i,x,y)=(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle.$$

As in the previous subsection, we have

$$\tau\mathbb{E}\left[\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)\langle a_i,x-y\rangle\right]=2-2\langle x,y\rangle=\|x-y\|_2^2$$

by directly expanding terms in the expectation and applying Proposition 17.

Hence, we obtain

$$\begin{aligned}
\|x-\hat{y}\|_2&=\mathbb{E}\left[\tau\mathbf{X_1}(a_i,x,\hat{y})\right]\\
&=\mathbb{E}\left[\tau\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i}\right]+\mathbb{E}\left[\tau\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i^c}\right].
\end{aligned}$$

So, the same arguments in the previous section give us the following bound.

$$\begin{aligned}
&|\mathbb{E}\left[\tau\mathbf{X_1}(a_i,x,\hat{y})\mathbb{1}_{E_i}\right]-\|x-\hat{y}\|_2|\\
&\leq\mathbb{E}\left[|\tau\mathbf{X_1}(a_i,x,\hat{y})|\,\mathbb{1}_{E_i^c}\right]\\
&\leq\frac{4c_5}{m^2}.
\end{aligned}$$

By the construction of the event E_i, $\mathbf{X_1}(a_i,x,\hat{y})\mathbb{1}_{E_i}$ is bounded by

$$|\mathbf{X_1}(a,x,\hat{y})\mathbb{1}_{E_i}|\leq 2C(N,s,m)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIT.2021.3124598, IEEE Transactions on Information Theory

and the bound for the second moment for $\mathbf{X_1}(a_i,x,\hat{y})\mathbb{1}_{E_i}$ is given by

$$\mathop{\mathbb{E}}_{a_i}\left[\mathbf{X_1}(a,x,\hat{y})^2\mathbb{1}_{E_i}\right] \le 4\pi C(N,s,m)^2\|x-\hat{y}\|_2$$

by the same argument used for $\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i}$ in the previous subsection.

c) Step 3. Bounding the sum of truncated terms by the Bernstein's inequality: Step 2 shows that the magnitude and variance of $\mathbf{X_1}(a_i,x,\hat{y})\mathbb{1}_{E_i}$ can be bounded exactly by those of $\mathbf{X_2}(a_i,x,\hat{y})\mathbb{1}_{E_i}$ in Step 2 in Subsection VII-A. Hence, applying the bounded Bernstein inequality and using the ε-net covering argument for $K^{(i)}$ as in Subsection VII-A yield the following statement:

For all $\hat{y} \in \mathscr{N}_{K^{(i)}}$, we obtain

$$\begin{aligned}&\left|\frac{1}{m}\sum_{i=1}^{m}\tau\mathbf{X_1}(a_i,x,\hat{y})\mathbb{1}_{E_i} - \mathbb{E}\left[\tau\mathbf{X_1}(a,x,\hat{y})\mathbb{1}_{E_i}\right]\right|\\ &\qquad \le \frac{1}{m^{1/2-\xi_i}}\|x-\hat{y}\|_2\end{aligned}$$

with probability $1-\frac{c_4}{m^4}$.

Then, we apply the triangle inequality to above to get

$$\begin{aligned}&\left|\frac{1}{m}\sum_{i=1}^{m}\tau\mathbf{X_1}(a_i,x,\hat{y})\mathbb{1}_{E_i} - \|x-\hat{y}\|_2\right|\\ &\qquad \le \frac{1}{m^{1/2-\xi_i}}\|x-\hat{y}\|_2 + \frac{4c_5}{m^2}\\ &\qquad \le \frac{1}{m^{1/2-\xi_i}}\|x-y\|_2 + \|y-\hat{y}\| + \frac{r_{i+1}}{600}\\ &\qquad \le \frac{1}{m^{1/2-\xi_i}}\|x-y\|_2 + \delta_i + \frac{r_{i+1}}{600}\\ &\qquad \le \frac{1}{m^{\frac{1}{40}(\frac{5}{6})^i}}\|x-y\|_2 + \frac{2r_{i+1}}{600}\end{aligned}$$

for all $m > cs^{10}\log^{10}(N/s)$ with a probability at least $1-\frac{c_4}{m^4}$.

d) Step 4. Establishing the bound for (I): As we derived the bound for for $\tau\mathbf{X_2}(a_i,x,\hat{y})$ in Section VII-A, whenever the event E_u occurs, we have

$$\frac{1}{m}\sum_{i=1}^{m}\tau\mathbf{X_1}(a_i,x,\hat{y}) = \frac{1}{m}\sum_{i=1}^{m}\tau\mathbf{X_1}(a_i,x,\hat{y})\mathbb{1}_{E_i}.$$

This conditional equality and the bound for the truncated terms in (21) in Step 3 and allow us to control (21).

On the other hand, the term (22) is bounded as below:

$$\begin{aligned}&\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle)\left\langle a_i,\frac{x-\hat{y}}{\|x-\hat{y}\|}\right\rangle\right.\\ &\quad\left.-\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle\right|\\ &\overset{(a)}{\le}\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle)\right.\\ &\qquad\left.\times\left(\left\langle a_i,\frac{x-\hat{y}}{\|x-\hat{y}\|}\right\rangle-\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle\right)\right|\\ &\quad+\left|\frac{1}{m}\sum_{i=1}^{m}\tau(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle-\mathrm{sign}(\langle a_i,x\rangle)\right.\\ &\qquad\left.+\mathrm{sign}(\langle a_i,y\rangle)\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle\right|\\ &\overset{(b)}{\le}\frac{\tau}{m}\sum_{i=1}^{m}|\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle|\left|\left\langle a_i,\frac{x-\hat{y}}{\|x-\hat{y}\|}-\frac{x-y}{\|x-y\|}\right\rangle\right|\\ &\quad+\frac{1}{m}\sum_{i=1}^{m}\tau|\mathrm{sign}(\langle a_i,\hat{y}\rangle)-\mathrm{sign}(\langle a_i,y\rangle)|\left|\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle\right|\\ &\overset{(c)}{\le}\frac{\tau}{m}\sum_{i=1}^{m}|\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle|\left|\left\langle a_i,\frac{x-\hat{y}}{\|x-\hat{y}\|}-\frac{x-y}{\|x-y\|}\right\rangle\right|\\ &\quad+2\tau C_{10}C(N,s,m)\log m\cdot\delta_i\\ &\overset{(d)}{\le}\frac{1}{m}\sum_{i=1}^{m}\tau|\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,\hat{y}\rangle|\cdot C(N,s,m)\cdot\frac{2\delta_i}{r_{i+1}}\\ &\quad+2\tau C_{10}C(N,s,m)\log m\cdot\delta_i\\ &\overset{(e)}{\le}C(N,s,m)\log m\cdot\frac{4\tau r_i\delta_i}{r_{i+1}}+2\tau C_{10}C(N,s,m)\log m\cdot\delta_i\\ &\overset{(f)}{\le}\frac{4\tau r_{i+1}}{600}+2\tau C_{10}C(N,s,m)\log m\cdot\delta_i\\ &\overset{(g)}{\le}\frac{4\tau r_{i+1}}{600}+\frac{2\tau r_{i+1}}{600}\\ &\le\frac{8}{600}r_{i+1}.\end{aligned}$$

Here are the justifications for the above chain of inequalities: (a) and (b) are by the triangle inequality. We applied Lemma 11 and Corollary 16 to obtain (c). The inequality (d) results from Proposition 27 and Lemma 11. (e) arises from applying Corollary 16. In (f) and (g), we used Proposition 15.

Putting all the bounds we have so far together, we establish

$$\begin{aligned}(I) &\le \frac{1}{m^{\frac{1}{40}(\frac{5}{6})^i}}\|x-y\|_2 + \frac{2r_{i+1}}{600} + \delta_i + \frac{8}{600}r_{i+1}\\ &\le \frac{1}{m^{\frac{1}{40}(\frac{5}{6})^i}}\|x-y\|_2 + \frac{r_{i+1}}{50}\end{aligned}$$

with probability at least

$$\begin{aligned}&1-\frac{2}{m^4}-\frac{c_2}{m^5}-\frac{2}{m^4}-\frac{c_4}{m^4}-\bar{c}\left(\frac{s}{N}\right)^{5s}\cdot\frac{1}{m^5}\\ &\qquad\ge 1-\bar{c}\left(\frac{s}{N}\right)^{5s}\cdot\frac{1}{m^5}-\frac{c_6}{m^4}.\end{aligned}$$

C. Proof of Lemma 21

As before, let $y \in K^{(i)}$, so $r_{i+1} \le \|x-y\| \le r_i$ and $\hat{y} \in \mathscr{N}_{K^{(i)}}$ with $\|y-\hat{y}\| \le \delta_i$.

a) Step 1. Approximate (III) by ε-net $\mathscr{N}_{K^{(i)}}$: First, we apply the triangle inequality and Lemma 9 to have

$$\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\mathrm{sign}(\langle a_i,x\rangle)-\mathrm{sign}(\langle a_i,y\rangle)b_i(x,y)\right\|_{K_1^\circ}$$

$$\begin{aligned}
&\le \left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, y\rangle) b_i(x,\hat{y}) \right\|_{K_t^\circ} \\
&+ \left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, y\rangle)(b_i(x,y) - b_i(x,\hat{y})) \right\|_{K_1^\circ} \\
&\le \left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, y\rangle) b_i(x,\hat{y}) \right\|_{K_1^\circ} \\
&+ \left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, y\rangle) \right. \\
&\qquad \left. \times \left(\left\langle a_i, \frac{x-y}{\|x-y\|} \right\rangle \frac{x-y}{\|x-y\|} - \left\langle a_i, \frac{x-\hat{y}}{\|x-\hat{y}\|} \right\rangle \frac{x-\hat{y}}{\|x-\hat{y}\|} \right) \right\|_{K_1^\circ} \\
&+ \left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, y\rangle) \right. \\
&\qquad \left. \times \left(\left\langle a_i, \frac{x+y}{\|x+y\|} \right\rangle \frac{x+y}{\|x+y\|} - \left\langle a_i, \frac{x+\hat{y}}{\|x+\hat{y}\|} \right\rangle \frac{x+\hat{y}}{\|x+\hat{y}\|} \right) \right\|_{K_1^\circ}
\end{aligned}$$

We will bound the second and third term using Lemma 29.

By applying Lemma 29, Corollary 16, and Proposition 15 to the second term, we have

$$\begin{aligned}
&\left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, y\rangle) \right. \\
&\qquad \left. \times \left(\left\langle a_i, \frac{x-y}{\|x-y\|} \right\rangle \frac{x-y}{\|x-y\|} - \left\langle a_i, \frac{x-\hat{y}}{\|x-\hat{y}\|} \right\rangle \frac{x-\hat{y}}{\|x-\hat{y}\|} \right) \right\|_{K_1^\circ} \\
&\le \frac{\tau}{m} \sum_{i=1}^{m} |\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, y\rangle| \\
&\quad \times \left\| \left(\left\langle a_i, \frac{x-y}{\|x-y\|} \right\rangle \frac{x-y}{\|x-y\|} - \left\langle a_i, \frac{x-\hat{y}}{\|x-\hat{y}\|} \right\rangle \frac{x-\hat{y}}{\|x-\hat{y}\|} \right) \right\|_{K_1^\circ} \\
&\le r_i (\log m) \cdot \frac{4\tau C(N,s,m)\delta_i}{r_{i+1}} \\
&\le \frac{4\tau}{600} r_{i+1}.
\end{aligned}$$

Similarly, the third term is bounded by $\frac{4\tau}{600} r_{i+1}$.

It remains to show that the first term is well-controlled.

$$\begin{aligned}
&\left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, y\rangle) b_i(x,\hat{y}) \right\|_{K_1^\circ} \\
&\le \left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle) b_i(x,\hat{y}) \right\|_{K_1^\circ} \\
&\quad + \left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, y\rangle) - \operatorname{sign}(\langle a_i, x\rangle) \right. \\
&\qquad\qquad \left. + \operatorname{sign}(\langle a_i, \hat{y}\rangle) b_i(x,\hat{y}) \right\|_{K_1^\circ} \\
&\le \left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle) b_i(x,\hat{y}) \right\|_{K_1^\circ} \\
&\quad + \frac{\tau}{m} \sum_{i=1}^{m} |\operatorname{sign}(\langle a_i, y\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle)| \, \|b_i(x,\hat{y})\|_{K_1^\circ} \\
&\le \left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle) b_i(x,\hat{y}) \right\|_{K_1^\circ} \\
&\quad + 2\tau\delta_i \log m \cdot \|b_i(x,\hat{y})\|_{K_1^\circ} \\
&\le \left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle) b_i(x,\hat{y}) \right\|_{K_1^\circ} \\
&\quad + 2\tau C(N,s,m) \log m \cdot \delta_i \\
&\le \left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle) b_i(x,\hat{y}) \right\|_{K_1^\circ} + \frac{2\tau r_{i+1}}{600}
\end{aligned} \tag{23}$$

whenever Lemma 11 and Corollary 16 hold. Here, the second last inequality is from Lemma 11 and the last is by Proposition 15. Thus, it boils down to control the first term in the right hand side of (23), which is presented in the next step.

b) Step 2. Bounding the first term by decoupling: As for the first term $\left\| \frac{\tau}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle) b_i(x,\hat{y}) \right\|_{K_1^\circ}$, we apply a simple variant of Lemma 8.1 in [11]. This implies that $\langle a_i, x\rangle, \langle a_i, \hat{y}\rangle$ and $b_i(x,\hat{y})$ are independent, which consequently shows that $(\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle))$ and $b_i(x,\hat{y})$ are independent. This allows us to apply the concentration inequality conditioned on $((\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle))$ as we describe below.

Define

$$h(x,\hat{y}) := \frac{1}{m} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle) b_i(x,\hat{y}).$$

This object has a very similar structure as the function h in Section 8.4.3 in [11] and we are going to follow the arguments appeared in Section 8.4.3 and 9.1 in that paper.

Conditioned on $\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle$, we have

$$h \sim N(0, \lambda I_{(x,\hat{y})^\perp}),$$
$$\text{where} \quad \lambda^2 := \frac{1}{m^2} \sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle)^2.$$

By following the argument in Section 8.4.3 of [11], conditional on $\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle$, we have

$$\mathbb{E}\|h\|_{K_t^\circ} \le \lambda \cdot w(K_t).$$

Since

$$\begin{aligned}
\lambda &= \frac{1}{m} \sqrt{\sum_{i=1}^{m} (\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle)^2} \\
&= \frac{\sqrt{2}}{\sqrt{m}} \sqrt{\frac{1}{m} \sum_{i=1}^{m} |\operatorname{sign}(\langle a_i, x\rangle) - \operatorname{sign}(\langle a_i, \hat{y}\rangle|},
\end{aligned}$$

we have $\lambda \le \frac{2}{m^{1/2}} \sqrt{C_{10} r_i \log m}$ with probability $1 - \frac{c_2}{m^4}$ by Corollary 16.

Again, by the same consideration in [11] based on the Gaussian concentration inequality (Section 8.3 and 9.1

in [11] to control the term E_2) for any $\varepsilon \geq 2m^{-1/2}$, we have

$$\begin{aligned}&\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle)b_i(x,\hat{y})\right\|_{K_1^\circ}\\ &\leq 4\tau\left(\frac{C_{10}r_i\log m}{m}\right)^{1/2}\left(w(K_1)+\varepsilon m^{1/2}\right)\\ &\leq 4\tau\left(\frac{C_{10}r_i\log m}{m}\right)^{1/2}\left(C(N,s,m)^{1/2}+\varepsilon m^{1/2}\right)\\ &\leq 4\tau\left(\frac{C_{10}r_i\log m}{m}\right)^{1/2}\left(C(N,s,m)^{1/2}\cdot\varepsilon m^{1/2}\right)\end{aligned}$$

with probability at least $1-\exp(-c'\varepsilon^2 m)$ for some universal constant $c'>0$ (in the last inequality, we have used the simple fact that $ab \geq a+b$ if $a,b \geq 2$).

From Proposition 13,

$$\begin{aligned}&\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle)b_i(x,\hat{y})\right\|_{K_1^\circ}\\ &\leq 4\frac{r_{i+1}}{10\sqrt{6}}\cdot\frac{m^{-1/2}}{\delta_i^{1/2}}\cdot\varepsilon m^{1/2}\\ &\leq \frac{2r_{i+1}}{10}\cdot\frac{\varepsilon}{\delta_i^{1/2}}.\end{aligned}$$

Taking $\varepsilon := \frac{\delta_i^{1/2}}{\sqrt{60}}$ (note that this choice of $\varepsilon \gg 2m^{-1/2}$ from the construction of δ_i) yields

$$\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,\hat{y}\rangle)b_i(x,\hat{y})\right\|_{K_1^\circ} \leq \frac{2\tau r_{i+1}}{600} \tag{24}$$

with probability at least (25).

Since we want the bound (24) holds for all $\hat{y}$ in $\mathcal{N}_{K^{(i)}}$, by the union bound, we have (24) for all $\hat{y}$ in $\mathcal{N}_{K^{(i)}}$ with probability greater than (26). By the same argument in the previous subsection based on comparing the exponents of $\log m$ and $C(N,s,m)$, this probability is at least

$$1-\exp\left(-C^{(5)}m^{\frac{1}{3}\left(\frac{5}{6}\right)^i}(\log m)^{\frac{7}{6}}C(N,s,m)\right) \geq 1-\frac{c_7}{m^5}.$$

c) Step 3. Establishing the bound for (III) *:* Combining the previous bounds in this subsection yields

$$\begin{aligned}&\left\|\frac{\tau}{m}\sum_{i=1}^{m}(\operatorname{sign}(\langle a_i,x\rangle)-\operatorname{sign}(\langle a_i,y\rangle)b_i(x,y)\right\|_{K_1^\circ}\\ &\leq \frac{4\tau}{600}r_{i+1}+\frac{2\tau r_{i+1}}{600}+\frac{2\tau r_{i+1}}{600}\\ &\leq \left(\frac{6}{600}+\frac{3}{600}+\frac{3}{600}\right)r_{i+1}\\ &= \frac{r_{i+1}}{50}\end{aligned}$$

with probability exceeding

$$\begin{aligned}&1-\frac{2}{m^4}-\frac{c_2}{m^5}-\frac{2}{m^4}-\frac{c_7}{m^4}-\bar{c}\left(\frac{s}{N}\right)^{5s}\cdot\frac{1}{m^5}\\ &\geq 1-\bar{c}\left(\frac{s}{N}\right)^{5s}\cdot\frac{1}{m^5}-\frac{c_8}{m^4}.\end{aligned}$$

VIII. Discussion

We show that NBIHT enjoys the optimal approximation error decay in the number of measurements for the one-bit compressed sensing problem. While this demonstrates its efficiency, there are still several aspects worth further investigation:

(1) Throughout this paper, we develop our theory for NBIHT with the step size $\tau = \sqrt{\pi/2}$. This may sound restrictive but the numerical experiments in Figure 1 suggest that the choice of the step size $\sqrt{\pi/2}$ is possibly optimal since with other constant step sizes or diminishing ones, the algorithm either converges more slowly or doesn't converge at all. Also note that our step size is already normalized by m as one can see in the NBIHT algorithm 1. The need for a careful choice and normalization of the step size for NBIHT to work is also empirically observed in [15].

(2) It would be interesting to see how much the requirement for the minimum number of measurements in Theorem 1 can be relaxed. Although improving this requirement is not the main focus of this paper, the extensive numerical experiments in the literature indicate that BIHT-type algorithms perform much better than other algorithms even for a moderate number of measurements [15]–[17]. Our reconstruction error bound is optimal in its dependence on the number of measurements m in Theorem 1 and Corollary 2, but we believe it is highly suboptimal in its dependence on the sparsity level s since it scales as $(s\log(N/s))^{7/2}$. Indeed, simply due to this large factor and the fact that the reconstruction error can be at most 2, we found that the error bound in Corollary 2 is empirically satisfied for all m, which would not be informative to understand the true measurement requirement.

We instead include numerical results supporting the conjecture in the introduction — the actual dependence on s in the error bound is linear — to provide readers a better insight about the requirement.

The left plot in Figure 2 shows the reconstruction error of NBIHT versus the sparsity level and compares it against the graphs of the functions $2s\log(n/s)/m$, $3s\log(n/s)/m$, and $(s\log(n/s))^2/m$. Comparison between these curves implies that the error dependence on s is of order $O(s\log(n/s))$, which is linear in s up to a logarithmic factor as stated in our conjecture.

On the other hand, the right plot in Figure 2 illustrates how the reconstruction error of NBIHT decays as m

$$\begin{aligned}1-\exp(-c'\varepsilon^2 m) &= 1-\exp\left(-c'\frac{\delta_i m^1}{60}\right)\\ &= 1-\exp\left(-c'(600C)^{3\left(1-\left(\frac{5}{6}\right)^{i-1}\right)}\frac{m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}(\log m)^{\frac{14}{3}\left(1-\left(\frac{5}{6}\right)^{i-1}\right)+\frac{4}{3}}C(N,s,m)^{5-\frac{10}{3}\left(\frac{5}{6}\right)^{i-1}}}{60}\right) \qquad (25)\end{aligned}$$

$$\begin{aligned}&1-2|\mathscr{N}_{K^{(i)}}|\exp\left(-c'(600C)^{2\left(1-\left(\frac{5}{6}\right)^{i-1}\right)}\frac{m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}(\log m)^{\frac{14}{3}\left(1-\left(\frac{5}{6}\right)^{i-1}\right)+\frac{4}{3}}C(N,s,m)^{5-\frac{10}{3}\left(\frac{5}{6}\right)^{i-1}}}{60}\right)\\ &\geq 1-2\exp\left(C''(600C)^{2\left(1-\left(\frac{5}{6}\right)^{i-1}\right)}m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}(\log m)^{2\left(1-\frac{7}{3}\left(\frac{5}{6}\right)^{i-1}\right)}C(N,s,m)^{4-\frac{10}{3}\left(\frac{5}{6}\right)^{i-1}}\right)\\ &\qquad\times\exp\left(-c'(600C)^{3\left(1-\left(\frac{5}{6}\right)^{i-1}\right)}\frac{m^{\frac{1}{3}\left(\frac{5}{6}\right)^{i-1}}(\log m)^{\frac{14}{3}\left(1-\left(\frac{5}{6}\right)^{i-1}\right)+\frac{4}{3}}C(N,s,m)^{5-\frac{10}{3}\left(\frac{5}{6}\right)^{i-1}}}{60}\right) \qquad (26)\end{aligned}$$

increases for a fixed sparsity level $s = 25$. We observe the error curve is below than $3s\log(n/s)/m$ for $m > 300$ and than $2s\log(n/s)/m$ for $m > 800$. This plot also suggests that the error decay rate in m is not $O(1/\sqrt{m})$, the previously known rate, but actually $O(1/m)$ as our theory tells us for sufficiently large m. More precisely, we empirically observe that the actual error curve scales $O(s\log(n/s)/m)$ once m becomes larger than a relatively small number. Therefore, proving the conjecture — the error dependence on s is at most $O(s\log(n/s))$ — would be next important step to investigate the behavior of the NBIHT algorithm further such as the true requirement for optimal error-decay rate (both optimal in s and m). We leave proving this as a future work.

(3) Note that in general it is not possible to recover a sparse signal from 1-bit measurements with non-Gaussian vectors even if we have infinitely many measurements [31]. However, under some extra assumptions on the signal set, we can reconstruct the signal with a reasonable accuracy. It could be worth to explore whether BIHT-type algorithms still exhibit superior performance for non-Gaussian measurements under these assumptions.

(4) Another possible direction would be to extend and analyze BIHT-type algorithms for sparse signals with respect to a dictionary. It is easy to see that our results naturally extend to sparse signals with respect to any orthogonal basis. We expect that BIHT-type algorithms might offer a good approximation error decay for a certain type of dictionaries as well, assuming that the hard thresholding operator for the dictionary is well defined and can be implemented in a computationally-efficient way.

ACKNOWLEDGEMENT

The authors thank Xiaowei Li for reading this manuscript and giving us several valuable comments.

APPENDIX A
PROOF OF THE UNIFORM BOUND LEMMA

In this section, we prove Lemma 11.

Lemma 26. *With probability at least* $1-\frac{2}{m^4}$, *we have*

$$\sup_{y\in K\cap\mathbb{S}^{N-1}}\left\{\left|\left\langle a_i, \frac{x+y}{\|x+y\|}\right\rangle\right|, \left|\left\langle a_i, \frac{x-y}{\|x-y\|}\right\rangle\right|, \|b_i(x,y)\|_{K_1^\circ}\right\}$$
$$\leq 3\left(C_b\sqrt{s\log(N/s)}+\sqrt{\frac{5\log m}{c_b}}\right)$$

for all i *with* $1 \leq i \leq m$.

Proof. For any $y \in K\cap\mathbb{S}^{N-1}$, we observe that $\frac{x+y}{\|x+y\|} \in (K-K)\cap\mathbb{S}^{N-1}$ because K is symmetric.

$$\begin{aligned}&\sup_{y\in K\cap\mathbb{S}^{N-1}}\left|\left\langle a_i, \frac{x+y}{\|x+y\|}\right\rangle\right|\\ &\leq \sup_{w,y\in K\cap\mathbb{S}^{N-1}}\left|\left\langle a_i, \frac{w+y}{\|w+y\|}\right\rangle\right|\\ &\leq \sup_{u-v\in(K-K),\|u-v\|_2\leq 1}|\langle a_i, u-v\rangle|\\ &= \sup_{u-v\in(K-K),\|u-v\|_2\leq 1}\langle a_i, u-v\rangle\\ &= \|a_i\|_{K_1^\circ} \leq C_b\sqrt{s\log(N/s)}+\sqrt{\frac{5\log m}{c_b}}\end{aligned}$$

under the event E_u. Here the second equality holds since K is symmetric and the last inequality is from Proposition 10.

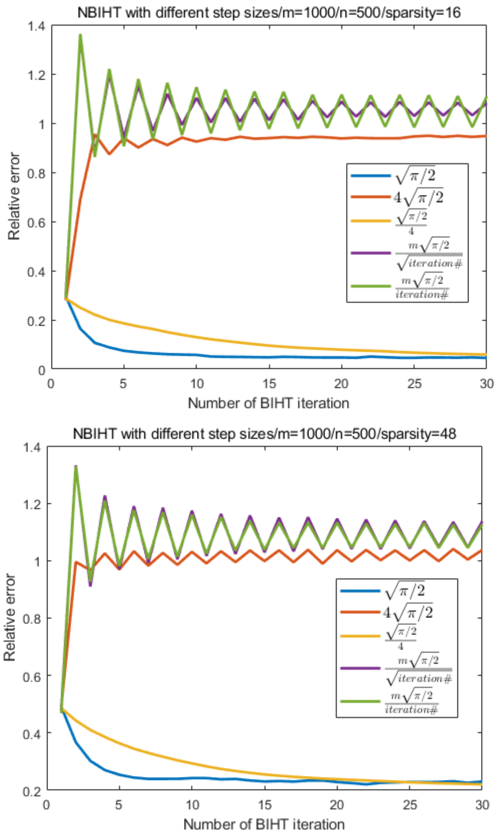

Fig. 1. Convergence behavior of NBIHT for different step sizes τ. We create 1000×500 Gaussian random matrix and draw vectors with sparsity levels 16 and 48 uniformly at random from the unit sphere. In each experiment, we run NBIHT with constant step sizes $\sqrt{\pi/2}, 4\sqrt{\pi/2}, \frac{1}{4}\sqrt{\pi/2}$, and diminishing (and unnormalized) step sizes $\frac{m\sqrt{\pi/2}}{\sqrt{k}}$ and $\frac{m\sqrt{\pi/2}}{k}$. We average the relative error over 30 trials and record the error versus the number of NBIHT iterations. Both plots indicate that with constant step size $4\sqrt{\pi/2}$ and diminishing step sizes without proper normalization, NBIHT doesn't converge at all, and with the step size $\frac{1}{4}\sqrt{\pi/2}$, it converges more slowly than with our choice of step size $\sqrt{\pi/2}$.

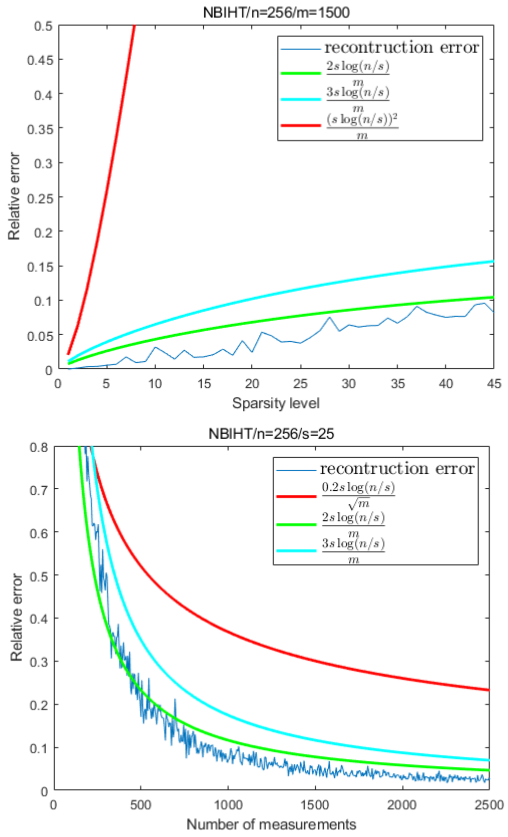

Fig. 2. NBIHT recovery error curve for fixed sparsity level and the number of measurements. A vector with specified sparsity level s is drawn uniformly at random from unit sphere and measured with $m\times 256$ standard Gaussian random matrix followed by the sign function. The reconstruction error is averaged over 30 trials. The left plot suggests that the error dependence on s is of order $O(s\log(n/s)/m)$. The right plot empirically shows that the error decay rate is actually strictly better than $O(1/\sqrt{m})$, the previously known rate, but $O(1/m)$ as our theory predict.

Again by the symmetry of K, this also implies that

$$\sup_{y\in K\cap\mathbb{S}^{N-1}}\left|\left\langle a_i, \frac{x-y}{\|x-y\|}\right\rangle\right| \le C_b\sqrt{s\log(N/s)}+\sqrt{\frac{5\log m}{c_b}}.$$

As for the bound for $\|b_i(x,y)\|_{K_1^\circ}$, we first start from the decomposition of a_i in Lemma 9.

$$b_i(x,y)=\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle\frac{x-y}{\|x-y\|}+\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\frac{x+y}{\|x+y\|}-a_i.$$

After taking the dual norm $\|\cdot\|_{K_1^\circ}$ on both sides, we have

$$\begin{aligned}&\|b_i(x,y)\|_{K_1^\circ}\\ &\le\left\|\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle\frac{x-y}{\|x-y\|}\right\|_{K_1^\circ}+\left\|\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\frac{x+y}{\|x+y\|}\right\|_{K_1^\circ}\\ &\quad+\|a_i\|_{K_1^\circ}\\ &\le\left|\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle\right|\left\|\frac{x-y}{\|x-y\|}\right\|_{K_1^\circ}+\left|\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right|\left\|\frac{x+y}{\|x+y\|}\right\|_{K_1^\circ}\\ &\quad+\|a_i\|_{K_1^\circ}\\ &=\left|\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle\right|+\left|\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right|+\|a_i\|_{K_1^\circ}\\ &\le 3\left(C_b\sqrt{s\log(N/s)}+\sqrt{\frac{5\log m}{c_b}}\right),\end{aligned}$$

where the first and second inequalities are by the triangle inequality, the equality is from the definition of the dual norm $\|\cdot\|_{K_1^\circ}$, and we applied the bounds for $\left|\left\langle a_i,\frac{x-y}{\|x-y\|}\right\rangle\right|, \left|\left\langle a_i,\frac{x+y}{\|x+y\|}\right\rangle\right|$, and $\|a_i\|_{K_1^\circ}$ in the last inequality. Thus we have the bound in the lemma under the event E_u which holds with probability exceeding $1-\frac{2}{m^4}$.

□

Proposition 27. *Let $y,\hat{y}\in\mathbb{S}^{N-1}$ with $r\le\|x-y\|$ and $r\le\|x-\hat{y}\|$ for some $r>0$. Also, assume that $\|y-\hat{y}\|\le\delta$.*

Then, we have

$$\left\| \frac{x-y}{\|x-y\|} - \frac{x-\hat{y}}{\|x-\hat{y}\|} \right\| \le \frac{2\delta}{r}.$$

Proof. Repeated applications of triangle inequality yield the following chain of inequalities.

$$\begin{aligned}
&\left\| \frac{x-y}{\|x-y\|} - \frac{x-\hat{y}}{\|x-\hat{y}\|} \right\| \\
&\le \left\| \frac{x-y}{\|x-y\|} - \frac{x-y}{\|x-\hat{y}\|} \right\| + \left\| \frac{x-y}{\|x-\hat{y}\|} - \frac{x-\hat{y}}{\|x-\hat{y}\|} \right\| \\
&\le \left| \frac{1}{\|x-y\|} - \frac{1}{\|x-\hat{y}\|} \right| \|x-y\| + \frac{1}{\|x-\hat{y}\|} \|x-y-(x-\hat{y})\| \\
&\le \left| \frac{\|x-\hat{y}\| - \|x-y\|}{\|x-y\|\|x-\hat{y}\|} \right| \|x-y\| + \frac{1}{\|x-\hat{y}\|} \|y-\hat{y}\| \\
&\le \left| \frac{\|y-\hat{y}\|}{\|x-\hat{y}\|} \right| + \frac{1}{\|x-\hat{y}\|} \|y-\hat{y}\| \\
&= \frac{2}{\|x-\hat{y}\|} \|y-\hat{y}\| \le \frac{2\delta}{r}.
\end{aligned}$$

□

Lemma 28. *Let $z, \hat{z}$ be s-sparse vectors with $\|z-\hat{z}\|_2 \le \delta$. Then, we have*

$$|\langle a_i, z\rangle - \langle a_i, \hat{z}\rangle| \le \delta \left(C_b\sqrt{s\log(N/s)} + \sqrt{\frac{5\log m}{c_b}} \right)$$

for all $1 \le i \le m$ with probability at least $1 - \frac{2}{m^4}$.

Proof. Let $h = z - \hat{z}$. By Proposition 10, for all $i \in [m]$, we have

$$\begin{aligned}
\sup_{h \in K-K\cap\delta\mathbb{S}^{N-1}} |\langle a_i, h\rangle| &= \|a_i\|_{K_\delta^\circ} = \delta\|a_i\|_{K_1^\circ} \\
&\le \delta \left(C_b\sqrt{s\log(N/s)} + \sqrt{\frac{5\log m}{c_b}} \right) \qquad (27)
\end{aligned}$$

with probability at least $1 - \frac{2}{m^5}$. Then the lemma follows from the union bound. □

Lemma 29. *Let $z, \hat{z}$ be unit s-sparse vectors with $\|z-\hat{z}\|_2 \le \delta$. Then, for $u > 0$, we have*

$$\|\langle a_i, z\rangle z - \langle a_i, \hat{z}\rangle \hat{z}\|_{K_1^\circ} \le 2\delta \left(C_b\sqrt{s\log(N/s)} + \sqrt{\frac{5\log m}{c_b}} \right)$$

for all $1 \le i \le m$ with probability at least $1 - \frac{2}{m^4}$.

Proof. Note that

$$\begin{aligned}
&\|\langle a_i, z\rangle z - \langle a_i, \hat{z}\rangle \hat{z}\|_{K_1^\circ} \\
&\quad \le \|\langle a_i, z\rangle z - \langle a_i, \hat{z}\rangle z\|_{K_1^\circ} + \|\langle a_i, \hat{z}\rangle z - \langle a_i, \hat{z}\rangle \hat{z}\|_{K_1^\circ} \\
&\quad \le |\langle a_i, z\rangle - \langle a_i, \hat{z}\rangle| \cdot \|z\|_{K_1^\circ} + |\langle a_i, \hat{z}\rangle| \cdot \|z-\hat{z}\|_{K_1^\circ} \\
&\quad \le |\langle a_i, z\rangle - \langle a_i, \hat{z}\rangle| \cdot \|z\|_2 + |\langle a_i, \hat{z}\rangle| \cdot \|z-\hat{z}\|_2 \\
&\le \delta \left(C_b\sqrt{s\log(N/s)} + \sqrt{\frac{5\log m}{c_b}} \right) \\
&\qquad + \left(C_b\sqrt{s\log(N/s)} + \sqrt{\frac{5\log m}{c_b}} \right) \cdot \delta \\
&\le 2\delta \left(C_b\sqrt{s\log(N/s)} + \sqrt{\frac{5\log m}{c_b}} \right).
\end{aligned}$$

The explanations to above inequalities are as follows: We used the triangle inequality in the first and second inequalities. The third inequalities is from the definition of the dual norm $\|\cdot\|_{K_1^\circ}$. We applied Lemma 28 to the fourth inequality. □

Finally, collecting the results in Lemma 26, 28, and 29 yield Lemma 11.

References

[1] S. Foucart and H. Rauhut, "An invitation to compressive sensing," in *A mathematical introduction to compressive sensing*. Springer, 2013, pp. 1–39.

[2] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge university press, 2012.

[3] C. S. Güntürk, M. Lammers, A. M. Powell, R. Saab, and Ö. Yılmaz, "Sobolev duals for random frames and $\sigma\delta$ quantization of compressed sensing measurements," *Foundations of Computational mathematics*, vol. 13, no. 1, pp. 1–36, 2013.

[4] 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.

[5] E. Chou, C. S. Güntürk, F. Krahmer, R. Saab, and Ö. Yılmaz, "Noise-shaping quantization methods for frame-based and compressive sampling systems," in *Sampling theory, a renaissance*. Springer, 2015, pp. 157–184.

[6] E. Chou and C. S. Güntürk, "Distributed noise-shaping quantization: I. beta duals of finite frames and near-optimal quantization of random measurements," *Constructive Approximation*, vol. 44, no. 1, pp. 1–22, 2016.

[7] 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.

[8] P. Deift, F. Krahmer, and C. S. Güntürk, "An optimal family of exponentially accurate one-bit sigma-delta quantization schemes," *Communications on Pure and Applied Mathematics*, vol. 64, no. 7, pp. 883–919, 2011.

[9] R. Saab, R. Wang, and Ö. Yılmaz, "From compressed sensing to compressed bit-streams: practical encoders, tractable decoders," *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6098–6114, 2017.

[10] P. T. Boufounos and R. G. Baraniuk, "1-bit compressive sensing," in *2008 42nd Annual Conference on Information Sciences and Systems*. IEEE, 2008, pp. 16–21.

[11] 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, 2016.

[12] 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.

[13] M. A. Davenport, Y. Plan, E. Van Den Berg, and M. Wootters, "1-bit matrix completion," *Information and Inference: A Journal of the IMA*, vol. 3, no. 3, pp. 189–223, 2014.

[14] J. L. Horowitz, *Semiparametric and nonparametric methods in econometrics*. Springer, 2009, vol. 12.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] A. M. Powell, R. Saab, and Ö. Yılmaz, “Quantization and finite frames,” in *Finite frames*. Springer, 2013, pp. 267–302.

[20] 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.

[21] ——, “Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach,” *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 482–494, 2012.

[22] 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.

[23] S. Rangan and V. K. Goyal, “Recursive consistent estimation with bounded noise,” *IEEE Transactions on Information Theory*, vol. 47, no. 1, pp. 457–464, 2001.

[24] H. Q. Nguyen, V. K. Goyal, and L. R. Varshney, “Frame permutation quantization,” *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 74–97, 2011.

[25] A. M. Powell, “Mean squared error bounds for the rangan–goyal soft thresholding algorithm,” *Applied and Computational Harmonic Analysis*, vol. 29, no. 3, pp. 251–271, 2010.

[26] S. Oymak and B. Recht, “Near-optimal bounds for binary embeddings of arbitrary sets,” *arXiv preprint arXiv:1512.04433*, 2015.

[27] M. Ledoux, *The concentration of measure phenomenon*. American Mathematical Soc., 2001, no. 89.

[28] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge University Press, 2018, vol. 47.

[29] D. Bilyk and M. T. Lacey, “Random tessellations, restricted isometric embeddings, and one bit sensing,” *arXiv preprint arXiv:1512.06697*, 2015.

[30] Y. Chen and E. J. Candès, “Solving random quadratic systems of equations is nearly as easy as solving linear systems,” *Communications on Pure and Applied Mathematics*, vol. 70, no. 5, pp. 822–883, 2017.

[31] A. Ai, A. Lapanowski, Y. Plan, and R. Vershynin, “One-bit compressed sensing with non-gaussian measurements,” *Linear Algebra and its Applications*, vol. 441, pp. 222–239, 2014.

Michael P. Friedlander Michael P. Friedlander is a professor of computer science and mathematics at the University of British Columbia. His research is primarily in developing and implementing numerical methods for large-scale optimization. Friedlander serves on the editorial boards of SIAM J. Optimization, SIAM J. Matrix Analysis and Applications, Mathematics of Operations Research, and Mathematical Programming. He received a Ph.D. in operations research from Stanford University.

Halyun Jeong Halyun Jeong is an assistant adjunct professor at the University of California, Los Angeles since July 2021. He was a Pacific Institute for the Mathematical Sciences (PIMS) postdoctoral fellow at the University of British Columbia from August 2017 to June 2021. He received his Ph.D. in Mathematics at Courant Institute of Mathematical Sciences, New York University in 2017. He was a recipient of the Kwanjeong scholarship foundation for his Ph.D. program. Prior to this, he obtained his master’s degree in Electrical and Computer Engineering from the University of California, San Diego and received his bachelor’s degrees in mathematics, computer science, and electrical engineering with the highest distinction from Pohang University of Science and Technology (POSTECH).

Yaniv Plan Yaniv Plan has been an Associate Professor of Math at the University of British Columbia (UBC) since July 2021. Previously, he was an assistant professor at UBC, starting in June 2014. He was awarded a Tier 2 Canada Research Chair for this position. In 2016, he won the UBC Mathematics and Pacific Institute for the Mathematical Sciences Faculty Award. Prior to this he was a Hildebrand Assistant Professor, and also an NSF postdoc, at University of Michigan in the Math department. He received his PhD in the Applied and Computational Mathematics program at Caltech and received the W.P. Carey and Co. Inc. prize for an outstanding doctoral dissertation. He also spent two years as a visiting researcher at Stanford University during a Sabbatical from his PhD. His research interests lie in applied probability, high-dimensional inference, random matrix theory, compressive sensing, matrix completion, and learning theory.

Özgür Yılmaz Özgür Yılmaz received the B.Sc. degrees in mathematics and electrical engineering from Bogazic¸i University, Istanbul, Turkey, 1997, and the Ph.D. degree in applied and computational mathematics from Princeton University in 2001. From 2002 to 2004, he was an Avron Douglis Lecturer at the University of Maryland, College Park. He joined the Mathematics Department at the University of British Columbia in 2004, where he is currently a professor. He is a member of IAM. His research interests include applied harmonic analysis, information theory, and signal processing.