Robust 1-bit Compressed Sensing with Iterative Hard Thresholding

Namiko Matsumoto* Arya Mazumdar†

Abstract

In 1-bit compressed sensing, the aim is to estimate a k-sparse unit vector $x \in S^{n-1}$ within an ϵ error (in ℓ_2) from minimal number of linear measurements that are quantized to just their signs, i.e., from measurements of the form $y = \mathsf{sign}(\langle a, x\rangle)$. In this paper, we study a noisy version where a fraction of the measurements can be flipped, potentially by an adversary. In particular, we analyze the Binary Iterative Hard Thresholding (BIHT) algorithm, a proximal gradient descent on a properly defined loss function used for 1-bit compressed sensing, in this noisy setting. It is known from recent results that, with $\tilde{O}(\frac{k}{\epsilon})$ noiseless measurements, BIHT provides an estimate within ϵ error. This result is optimal and universal, meaning one set of measurements work for all sparse vectors. In this paper, we show that BIHT also provides better results than all known methods for the noisy setting. We show that when up to τ-fraction of the sign measurements are incorrect (adversarial error), with the same number of measurements as before, BIHT agnostically provides an estimate of x within an $\tilde{O}(\epsilon + \tau)$ error, maintaining the universality of measurements. This establishes stability of iterative hard thresholding in the presence of measurement error. To obtain the result, we use the restricted approximate invertibility of Gaussian matrices, as well as a tight analysis of the high-dimensional geometry of the adversarially corrupted measurements.

1 Introduction

Compressed sensing is a framework in signal processing that exploits the inherent sparsity or compressibility of signals to efficiently acquire and reconstruct them with a sampling rate that is significantly lower than the dimensionality of the signal [8, 16]. By using a small number of non-adaptive measurements, often obtained through random projections, compressed sensing enables the recovery of the original signal with high accuracy.

In real-world signal acquisitions and storage, signals are often digitized. This led to introduction to 1-bit compressed sensing (1bCS) by [6]. In this model, a unit-norm sparse signal $\mathbf{x} \in S^{n-1}, \|\mathbf{x}\|_0 \le k$, is acquired through the operation $\mathbf{y} = \mathsf{sign}(\mathbf{Ax})$, where $\mathbf{A}$ is an $m \times n$ real matrix and $\mathbf{y} \in \{1, -1\}^m$ is a binary vector containing the coordinate-wise signs of $\mathbf{Ax}$. The primary objective is to design a measurement matrix $\mathbf{A}$ with minimal number of rows m, such that for any $\mathbf{x} \in S^{n-1}, \|\mathbf{x}\|_0 \le k$, an estimate $\hat{\mathbf{x}}$ from $\mathbf{y}$ and $\mathbf{A}$ via an efficient algorithm can be provided such that $\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \le \epsilon$, for a given $0 \le \epsilon \le 1$. We will refer to ϵ as the *parameter error*.

It is known that $m = \Omega(\frac{k}{\epsilon})$ measurements are necessary [25] for this. Also, if the entries of the matrix $\mathbf{A}$ are chosen to be standard normal random random variables then recovery is possible with high probability for all k-sparse unit norm vectors with $m = O(\frac{k}{\epsilon}\log\frac{n}{\epsilon})$ measurements [25]. Hereafter, there has been a series of work that tries to achieve this baseline number of measurements $\tilde{O}(\frac{k}{\epsilon})$ with a computationally tractable algorithm, such as convex relaxations [32, 7, 33]. In particular, the *linear estimator* of [33] shows that $\tilde{O}(\frac{k}{\epsilon^2})$ measurements are sufficient, which is suboptimal in its dependency on the parameter error.

In this paper, we study a very natural iterative estimation method proposed in [25], called *binary iterative hard thresholding* (BIHT). Iterative hard thresholding is a well-known algorithm for compressed sensing, where estimations of $\mathbf{x}$ are projected back to the "top-k" coordinates in each step to maintain sparsity of the solution [5]. The description of BIHT is provided in Algorithm 1, and will be formally discussed later. In short, it is a proximal gradient descent algorithm where an estimate of $\mathbf{x}$ is updated iteratively followed by the aforementioned projection. BIHT was empirically observed to have excellent performance which was analyzed in several papers such as [24, 7, 27, 20]. Ultimately, in [28], it was shown that $\tilde{O}(\frac{k}{\epsilon})$ measurements are sufficient for BIHT to produce an estimate with at most ϵ error, i.e., the optimal dependence on sparsity and error.

In this paper, we show that iterative hard thresholding is in fact even more powerful: it is robust to adversarial noise. Noisy one-bit compressed sensing has also been quite well-studied in the last few years [2, 11, 3, 22, 10].

*University of California San Diego.
†University of California San Diego.

In particular, we assume that any up to $\tau m, 0 \le \tau \le 1$, coordinates of the measurement vector $\mathsf{sign}(\mathbf{A}\mathbf{x})$ are flipped by an adversary. In this model, [31] showed that their *linear estimator* can provide an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ such that $\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2 \le \tilde{O}(\epsilon^2 + \tau)$ with $O(\frac{k}{\epsilon^{12}} \log \frac{2n}{k})$ measurements (Thm. 1.3 in [31]). In the same model, [2] provided an algorithm that returns an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ such that $\|\hat{\mathbf{x}} - \mathbf{x}\|_2 \le \epsilon + c\tau$, $c > 0$ being a constant, with $\tilde{O}(\frac{k}{\epsilon^3})$ measurements.

To mitigate such sign-flips in measurements, an algorithm called adaptive outlier pursuit was proposed in [36] that shows superior performance over BIHT empirically. However the algorithm requires precise knowledge of τ, and performance deteriorates rapidly without this knowledge. On the other hand, BIHT is agnostic to the number of sign-flips. Another algorithm based on MAP estimation was proposed in [11], relying on a stable embedding property of the measurement matrix which is known to take at least $\Omega(\frac{k}{\epsilon^2})$ rows. With $\Omega(\frac{k}{\epsilon^2})$ measurements a least-square decoding algorithm was also shown to be effective in [22] in the presence of sign-flips. More recently, the noisy 1-bit compressed sensing problem was also studied from both the perspective of parameter error, and prediction error; in particular the performance of the AdaBoost ([19]) algorithm was analyzed in [10]. The number of required measurements here scales as $\tilde{O}(\frac{k}{\epsilon^3})$. We have omitted the dependence on τ in the last few results for the sake of clarity, and also to point out suboptimal dependence on parameter error even in the absence of adversarial sign-flips.

1.1 Our Contributions

Under the adversarial sign-flip model described above, we show that BIHT still produces a good estimate of the sparse vector $\mathbf{x}$ with the same number of $\tilde{O}(\frac{k}{\epsilon})$ measurements. BIHT is also agnostic to the number of sign-flips: indeed, as long as there is sufficient number of measurements, a good estimate with small parameter error is produced. To be precise, we show that with $m = O(\frac{k}{\epsilon} \log \frac{n}{k\epsilon})$ measurements of which up to τm can be corrupted, BIHT converges to an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$, such that

$$\|\hat{\mathbf{x}} - \mathbf{x}\| \le \epsilon + O(\sqrt{\epsilon\tau} + \tau\sqrt{\log\frac{1}{\tau}}) \asymp \max\{\epsilon, \tau\sqrt{\log\frac{1}{\tau}}\}.$$

With only $\tilde{O}(\frac{k}{\epsilon})$ measurements, this result provides the best sample complexity guarantee, i.e., a number of measurements with better dependence on parameter error than [31, 2, 11, 22, 10] mentioned above.

While our work builds on [28], our analysis requires new insights as well as new technical tools. One of the key steps in [28] is to establish a property of Gaussian matrices called *restricted approximate invertibility condition* (RAIC). This condition ensures that the estimation error remains controlled throughout the iterations of BIHT by approximately preserving the discrepancy between two vectors and the average of the measurements (rows of matrix $\mathbf{A}$) that yield distinct outcomes for those vectors.

In this paper, we aim to prove a similar condition but account for the possibility of flipped measurements. To achieve this, we introduce a new definition of RAIC with measurement error. Our main technical achievement is demonstrating that Gaussian matrices possess this property. For this, in addition to validating the results obtained by [28] for Gaussian measurements, we also need to establish a (roughly) linear relationship between the expected norm of the sum of up to τm-many measurements and the expected error resulting from adversarial corruption of up to τm-many responses. Consequently, given that the norm of the sum of any set of up to τm-many measurements can be consistently bounded and not exceed a certain threshold with a high probability, we can establish an upper bound on the error introduced by adversarial noise with a high probability. With the goal of upper bounding the norm of the sum of the up to τm-many measurements, the vector sum is orthogonally decomposed into two components: (a) its projection onto a particular vector $\mathbf{u}$ (determined later), and (b) its projection into the kernel of $\mathbf{u}$, each of which will be bounded separately. The norm of the two components can be recombined via triangle inequality. Repeating this over a collection of vectors, $\mathbf{u}$, leads to a uniform result. Crucially, it turns out that the number of vectors, $\mathbf{u}$, which need to be considered in this collection is finite and quantifiable: it does not exceed the number of ways to choose up to τm-many responses to corrupt. This is related to the tracking of "mismatches," which was a key element in previous analysis.

1.2 Other Related Works

Without the sparsity constraint, the problem we consider is closely related to the noisy half-space learning problem. However, most of the time the focus of such works is to provide guarantee on prediction error, rather than parameter error [18, 26]. The objective of this line of work is to come up with distribution-agnostic efficient algorithms, and then to provide guarantees on their zero-one loss (probability of mismatch). This problem is also studied with different noise models, for example, Massart noise [12], instance-dependent noise [30, 9], random sign-flip noise [13]. In particular, the latter work shows that with standard Gaussian covariates, and with probability of sign-flip being η, one can come up with a classifier to guarantee probability of mismatch $O(\eta)$, with $\frac{n}{\eta^2}$ samples. Since for Gaussian covariates, parameter error and prediction error could be related - this will lead to a suboptimal sample complexity with respect to the error rate, if a "sparse" version could be made available. Active learning under this model was also considered in [37].

Interestingly, learning k-sparse half-spaces where labels can be corrupted has been considered in [38, 34], with guarantee on the prediction error. Note that these papers study the problem in an "active PAC learning" setting, which is different from even the adaptive version of 1-bit compressed sensing. We point the reader to [38] for a detailed discussion on this difference. Furthermore, if the covariates/measurements were Gaussian, prediction error could be related to parameter error, but that is not the case in general. While in the active learning set up the number of label queries are small, the total sample complexity in [34] scales quadratically with k, which is suboptimal in 1-bit compressed sensing.

In [31] a more general sparse signal recovery problem was studied where the binary observations $y_i \in \{+1, -1\}$ are random: i.e., $y_i = 1$ with probability $f(\langle a_i, x\rangle), i = 1, \ldots, m$, where f is a potentially nonlinear function, such as the logistic function.

The support recovery problem in 1-bit compressed sensing and constructions of structured measurement matrices are well-studied, though not directly related to this work, e.g., [21, 1, 17, 29]. Other lines of work on 1-bit compressed sensing include adaptive sensing for faster error-decay [4, 23], dithering to allow for magnitude recovery and robustness [14, 15], and 1-bit sensing with generative priors [27].

Organization. The rest of the paper is organized as follows. In Section 2 we introduced the notations used in the paper, and also the BIHT algorithm. Section 3 contains the main result (Theorem 3.1) and a technical overview of the proofs. Subsequently, proofs of the main results appear in Section 4 and 5. Intermediate, and longer proofs are delegated to the appendix.

2 Preliminaries

2.1 Notations

Throughout this work, the parameters $k, n \in \mathbb{Z}_+$ are taken to satisfy $n \geq 2k$, where k denotes sparsity (i.e., the maximum number of nonzero entries in a vector), and where n is the dimension of the signal vectors and measurements. The number of measurements (and rows in the measurement matrix) is denoted by $m \in \mathbb{Z}_+$. For notational simplicity, the parameter $\tau \in (0, 1]$—the largest allowable fraction of responses that can be corrupted—is assumed to satisfy $\tau m \in \mathbb{Z}_+$. This does not forgo generality since τm can be replaced by $\lceil \tau m \rceil$ throughout the analysis in this manuscript. Note that this work does not consider $\tau = 0$ since [28] already established the result under noiseless conditions.

For the purposes of this discussion, let $\ell, d \in \mathbb{Z}_+$, where $d \in \mathbb{Z}_+$ specifies an arbitrary dimension.

Let $\mathcal{D}$ be an arbitrary distribution. Then, $X \sim \mathcal{D}$ denotes a random variable which follows the distribution $\mathcal{D}$. Similarly, let $\mathcal{S}$ be an arbitrary set. Then, $X \sim \mathcal{S}$ denotes a random variable which follows the uniform distribution over $\mathcal{S}$. The univariate normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}_{\geq 0}$ is denoted by $\mathcal{N}(\mu, \sigma^2)$, while the d-variate normal distribution with mean vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

The d-dimensional identity matrix is denoted by $\mathbf{I}_d \in \mathbb{R}^{d\times d}$. More generally, matrices are written as capital letters in boldface, upright typeface, e.g., $\mathbf{A} \in \mathbb{R}^{\ell \times d}$, with the i^{th} rows denoted by, e.g., $\mathbf{A}_i \in \mathbb{R}^d$, $i \in [\ell]$, such that $\mathbf{A} = (\mathbf{A}_1 \cdots \mathbf{A}_\ell)^\mathsf{T}$, and with the (i, j)-entries written in italic typeface, e.g., $A_{i,j} \in \mathbb{R}$. Nonrandom vectors are written as lowercase letters in boldface, upright typeface, e.g., $\mathbf{u} \in \mathbb{R}^d$ with the j^{th} entries, $j \in [d]$, written in italic typeface, e.g., $u_j \in \mathbb{R}$, such that $\mathbf{u} = (u_1, \ldots, u_d)$. Random vectors follow the same convention as nonrandom vectors but with uppercase letters, e.g., $\mathbf{Z} = (Z_1, \ldots, Z_d) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$. For $J \subseteq [d]$, the restriction of $\mathbf{u} \in \mathbb{R}^d$ to the entries indexed by J is denoted by $\mathbf{u}|_J \in \mathbb{R}^{|J|}$. The support of a vector, $\mathbf{u} \in \mathbb{R}^d$, is denoted by $\mathsf{supp}(\mathbf{u}) \triangleq \{j \in [d] : u_j \neq 0\} \subseteq [d]$, and the number of nonzero entries in $\mathbf{u}$—the ℓ_0-"norm" of $\mathbf{u}$—is written as $\|\mathbf{u}\|_0 \triangleq |\mathsf{supp}(\mathbf{u})|$.

The ℓ_2-unit sphere in $\mathbb{R}^d$ is denoted by $S^{d-1} \triangleq \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\|_2 = 1\}$, and the set of k-sparse d-dimensional vectors is written as $\Sigma_k^d \triangleq \{\mathbf{u} \in \mathbb{R}^d : \|\mathbf{u}\|_0 \leq k\}$. Hence, the set of all d-dimensional, k-sparse, real-valued unit vectors is denote by $S^{d-1} \cap \Sigma_k^d \triangleq \{\mathbf{u} \in S^{d-1} : \|\mathbf{u}\|_0 \leq k\}$. The distance between two points projected onto the

ℓ_2-unit sphere is specified by the function $d_{S^{d-1}} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_{\geq 0}$, where

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

for $\mathbf{u},\mathbf{v} \in \mathbb{R}^d$. The Hamming distance between a pair of vectors, $\mathbf{u},\mathbf{v} \in \mathbb{R}^d$, is denoted by $d_H(\mathbf{u},\mathbf{v}) = |\{j \in [d] : u_j \neq v_j\}|$. The sign function, $\mathsf{sign} : \mathbb{R} \to \{-1,1\}$, follows the convention:

$$\mathsf{sign}(a) = \begin{cases} -1, & \text{if } a < 0, \\ \ \ 1, & \text{if } a \geq 0, \end{cases}$$

where $a \in \mathbb{R}$. This notation extends to vectors as $\mathsf{sign} : \mathbb{R}^d \to \{-1,1\}^d$ by taking the $\pm$-signs of each entry of a d-dimensional vector.

2.2 Hard Thresholding and the BIHT Algorithm This work considers two notions of hard thresholding as means to project points into the subspace of ℓ-sparse vectors, Σ^d_ℓ: top-ℓ and subset hard thresholding. These are formalized in the following definitions.

DEFINITION 2.1. (TOP-ℓ HARD THRESHOLDING) *The* top-ℓ hard thresholding operation*, denoted by $T_\ell : \mathbb{R}^d \to \mathbb{R}^d$, projects a vector $\mathbf{u} \in \mathbb{R}^d$ into Σ^d_ℓ by retaining only the ℓ largest (in absolute value) entries in $\mathbf{u}$ and setting all other entries to 0. Note that "ties" can be broken arbitrarily. More formally, writing $\mathcal{U}_\ell = \{\mathbf{u'} \in \mathbb{R}^d : \|\mathbf{u'}\|_0 = \ell,\ u'_j \in \{u_j, 0\}\ \forall j \in [d]\}$, the top-$\ell$ hard thresholding operation maps: $\mathbf{u} \mapsto T_\ell(\mathbf{u}) \in \arg\max_{\mathbf{u'} \in \mathcal{U}_\ell} \|\mathbf{u'}\|_1$.*

DEFINITION 2.2. (SUBSET HARD THRESHOLDING) *The* subset hard thresholding operation *associated with a coordinate subset $J \subseteq [d]$, denoted by $T_J : \mathbb{R}^d \to \mathbb{R}^d$, takes a vector, $\mathbf{u} \in \mathbb{R}^d$, into $\Sigma^d_{|J|}$ by setting all entries in $\mathbf{u}$ indexed by $[d] \setminus J$ to 0. More formally, $T_J(\mathbf{u})$ is the vector whose j^{th} entries, $j \in [d]$, are given by $T_J(\mathbf{u})_j = u_j \cdot \mathbb{I}(j \in J)$.*

The measurement matrix is denoted by $\mathbf{A} \in \mathbb{R}^{m\times n}$, with the measurements, i.e., its rows, written as $\mathbf{A}_1, \dots, \mathbf{A}_m \in \mathbb{R}^n$, such that $\mathbf{A} = (\mathbf{A}_1\ \cdots\ \mathbf{A}_m)^\mathsf{T}$. For a given unknown k-sparse unit signal vector, $\mathbf{x} \in S^{n-1} \cap \Sigma^n_k$, we call $\mathsf{sign}(\mathbf{A}\mathbf{x})$ the true responses. Now, suppose, $\mathbf{y} \in \{-1,1\}^m$ denotes an arbitrary vector that satisfies $d_H(\mathbf{y},\ \mathsf{sign}(\mathbf{A}\mathbf{x})) \leq \tau m$. This vector, $\mathbf{y}$, can be viewed as introducing adversarial noise into the true responses. At this point we can formally define the normalized BIHT algorithm. It is given as Algorithm 1 below.

Algorithm 1 Binary iterative hard thresholding (BIHT) algorithm: Input $\mathbf{y}, \mathbf{A}$

Set $\eta = \sqrt{2\pi}$
$\hat{\mathbf{x}}^{(0)} \sim S^{n-1} \cap \Sigma^n_k$
for $t = 1,2,3,\dots$ **do**
  $\tilde{\mathbf{x}}^{(t)} \leftarrow \hat{\mathbf{x}}^{(t-1)} + \frac{\eta}{m}\mathbf{A}^\mathsf{T} \cdot \frac{1}{2}\Big(\mathbf{y} - \mathsf{sign}\big(\mathbf{A}\hat{\mathbf{x}}^{(t-1)}\big)\Big)$
  $\hat{\mathbf{x}}^{(t)} \leftarrow \frac{T_k(\tilde{\mathbf{x}}^{(t)})}{\|T_k(\tilde{\mathbf{x}}^{(t)})\|_2}$
end for

2.3 Some Universal Constants None of the universal constants appearing in this work are very large. These constants, $a, b, c_1, c_2, c_3, c_4, c > 0$, appear throughout the results and analysis in this work. These universal constants are fixed as follows:

(2.1a) $a = 16,$ $\qquad b \lesssim 379.1038,$

(2.1b) $c_1 = \sqrt{\frac{3\pi}{b}}\left(1 + \frac{16\sqrt{2}}{3}\right) \in (1.3469, 1.3470),$ $\qquad c_2 = \frac{3}{b}\left(1 + \frac{4\pi}{3} + \frac{8\sqrt{3\pi}}{3} + 8\sqrt{6\pi}\right) \in (0.3806, 0.3807),$

$$\text{(2.1c)} \qquad c_3 = \frac{(12+\sqrt{3})\sqrt{\pi}}{\sqrt{b}} \in (1.2500, 1.2501), \qquad c_4 = 2 + 4\sqrt{\pi} \in (9.0898, 9.0899),$$

$$\text{(2.1d)} \qquad c = 4\left(c_1 + \sqrt{c_1^2 + c_2}\right)^2 \in (31.9999, 32).$$

3 Main Result and Technical Overview

Theorem 3.1, below, states the main result of this work, which establishes the convergence of BIHT when an arbitrary but bounded number of responses are corrupted. Note that it is a universal result in the sense that the measurement matrix, $\mathbf{A}$, is fixed across the recovery of all k-sparse, real-valued unit vectors.

THEOREM 3.1. *Let $\epsilon, \epsilon_0, \tau, \rho \in (0,1]$, $r > 0$, $k, m, n \in \mathbb{Z}_+$, where*

$$r \triangleq \frac{c}{c_2}\left(c_3\sqrt{\epsilon\tau} + c_4\tau\sqrt{\log\left(\frac{2e}{\tau}\right)}\right), \tag{3.2}$$

$$\epsilon_0 \triangleq \epsilon + r, \tag{3.3}$$

and where

$$m \geq \frac{4bc}{\epsilon}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12b}{\epsilon}\right)^{2k}\left(\frac{3a}{\rho}\right)\right) = O\left(\frac{k}{\epsilon}\log\left(\frac{n}{\epsilon k}\right) + \frac{1}{\epsilon}\log\left(\frac{1}{\rho}\right)\right). \tag{3.4}$$

Fix an $m \times n$ measurement matrix, $\mathbf{A} \in \mathbb{R}^{m\times n}$, whose rows, $\mathbf{A}_1, \ldots, \mathbf{A}_m \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, are i.i.d. Gaussian random vectors. Uniformly with probability at least $1-\rho$, for all k-sparse, real-valued unit vectors, $\mathbf{x} \in S^{n-1} \cap \Sigma_k^n$, when given m noisy responses, $\mathbf{y} \in \{-1,1\}^m$ (i.e., with any choice of up to τm corrupted),

$$d_H(\mathbf{y},\ \mathsf{sign}(\mathbf{A}\mathbf{x})) \leq \tau m, \tag{3.5}$$

the sequence of approximations, $\{\hat{\mathbf{x}}^{(t)} \in S^{n-1} \cap \Sigma_k^n\}_{t\in\mathbb{Z}_{\geq 0}}$, produced by the normalized BIHT algorithm converges as

$$d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t)}) \leq 2^{2^{-t}}\epsilon_0^{1-2^{-t}} \tag{3.6a}$$

with an approximation error asymptotically bounded from above by

$$\lim_{t\to\infty} d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t)}) \leq \epsilon_0. \tag{3.6b}$$

3.1 The Restricted Approximate Invertibility Condition (RAIC) under Adversarial Noise

As we have discussed in the introduction, the key step of proving our result is to establish a property of Gaussian matrices called restricted approximate invertibility in the presence of adversarial sign-flips. Before we give a technical overview of our proof, here we define the notion of RAIC and formally present the result regarding Gaussian matrices.

Fixing the measurement matrix, $\mathbf{A} \in \mathbb{R}^{m\times n}$, let $f : \mathbb{R}^n \to \{-1,1\}^m$ denote an arbitrary map that satisfies $d_H(f(\mathbf{u}),\ \mathsf{sign}(\mathbf{A}\mathbf{u})) \leq \tau m$ for all $\mathbf{u} \in \mathbb{R}^n$. This map, f, can be viewed as introducing one particular adversarial error pattern into the true responses, $\mathsf{sign}(\mathbf{A}\mathbf{u}) \mapsto f(\mathbf{u})$. The set of all such functions, f, is denoted by $\mathcal{F}_{\mathbf{A}} \triangleq \{f : \mathbb{R}^n \to \{-1,1\}^m\ :\ d_H(f(\mathbf{u}),\ \mathsf{sign}(\mathbf{A}\mathbf{u})) \leq \tau m\ \forall \mathbf{u} \in \mathbb{R}^n\}$. This is essentially the set of all possible ways to adversarially corrupt the true responses.

Additionally, define the functions $h_{\mathbf{A}}, h_{f;\mathbf{A}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, at arbitrary ordered pairs points, $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^n \times \mathbb{R}^n$, by

$$h_{\mathbf{A}}(\mathbf{u},\mathbf{v}) = \frac{\sqrt{2\pi}}{m}\mathbf{A}^{\mathsf{T}} \cdot \frac{1}{2}\big(\mathsf{sign}(\mathbf{A}\mathbf{u}) - \mathsf{sign}(\mathbf{A}\mathbf{v})\big) = \frac{\sqrt{2\pi}}{m}\sum_{i=1}^m \mathbf{A}_i \cdot \frac{1}{2}\big(\mathsf{sign}(\langle\mathbf{A}_i,\mathbf{u}\rangle) - \mathsf{sign}(\langle\mathbf{A}_i,\mathbf{v}\rangle)\big), \tag{3.7a}$$

$$h_{f;\mathbf{A}}(\mathbf{u},\mathbf{v}) = \frac{\sqrt{2\pi}}{m}\mathbf{A}^{\mathsf{T}} \cdot \frac{1}{2}\big(f(\mathbf{u}) - \mathsf{sign}(\mathbf{A}\mathbf{v})\big) = \frac{\sqrt{2\pi}}{m}\sum_{i=1}^m \mathbf{A}_i \cdot \frac{1}{2}\big(f(\mathbf{u})_i - \mathsf{sign}(\langle\mathbf{A}_i,\mathbf{v}\rangle)\big), \tag{3.7b}$$

and for $J \subseteq [n]$, let $h_{\mathbf{A};J}, h_{f;\mathbf{A};J} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ denote the functions given at $(\mathbf{u}, \mathbf{v}) \in \mathbb{R}^n \times \mathbb{R}^n$ by

$$
\begin{aligned}
h_{\mathbf{A};J}(\mathbf{u}, \mathbf{v}) &= T_{\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v}) \cup J}(h_{\mathbf{A}}(\mathbf{u}, \mathbf{v})) \\
&= \frac{\sqrt{2\pi}}{m} \sum_{i=1}^{m} T_{\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v}) \cup J}\left(\mathbf{A}_i \cdot \frac{1}{2}\big(\mathsf{sign}(\langle \mathbf{A}_i, \mathbf{u} \rangle) - \mathsf{sign}(\langle \mathbf{A}_i, \mathbf{v} \rangle)\big)\right) \qquad (3.8a)
\end{aligned}
$$

and

$$
\begin{aligned}
h_{f;\mathbf{A};J}(\mathbf{u}, \mathbf{v}) &= T_{\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v}) \cup J}(h_{f;\mathbf{A}}(\mathbf{u}, \mathbf{v})) \\
&= \frac{\sqrt{2\pi}}{m} \sum_{i=1}^{m} T_{\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v}) \cup J}\left(\mathbf{A}_i \cdot \frac{1}{2}\big(f(\mathbf{u})_i - \mathsf{sign}(\langle \mathbf{A}_i, \mathbf{v} \rangle)\big)\right). \qquad (3.8b)
\end{aligned}
$$

Note that

$$
\frac{1}{2}\big(f(\mathbf{u})_i - \mathsf{sign}(\langle \mathbf{A}_i, \mathbf{v} \rangle)\big) = -\,\mathsf{sign}(\langle \mathbf{A}_i, \mathbf{v} \rangle) \cdot \mathbb{I}(f(\mathbf{u})_i \neq \mathsf{sign}(\langle \mathbf{A}_i, \mathbf{v} \rangle)).
$$

and hence, $h_{f;\mathbf{A}}, h_{f;\mathbf{A};J}$ are equivalently given by

$$
h_{f;\mathbf{A}}(\mathbf{u}, \mathbf{v}) = -\frac{\sqrt{2\pi}}{m} \sum_{i=1}^{m} \mathbf{A}_i \, \mathsf{sign}(\langle \mathbf{A}_i, \mathbf{v} \rangle) \cdot \mathbb{I}(f(\mathbf{u})_i \neq \mathsf{sign}(\langle \mathbf{A}_i, \mathbf{v} \rangle)), \qquad (3.9a)
$$

$$
h_{f;\mathbf{A};J}(\mathbf{u}, \mathbf{v}) = -\frac{\sqrt{2\pi}}{m} \sum_{i=1}^{m} T_{\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v}) \cup J}(\mathbf{A}_i \, \mathsf{sign}(\langle \mathbf{A}_i, \mathbf{v} \rangle)) \cdot \mathbb{I}(f(\mathbf{u})_i \neq \mathsf{sign}(\langle \mathbf{A}_i, \mathbf{v} \rangle)). \qquad (3.9b)
$$

For any two sparse vectors $\mathbf{u}$ and $\mathbf{v}$, $h_{f;\mathbf{A};J}(\mathbf{u}, \mathbf{v})$ is a "distance-vector" measuring their closeness through the the one-bit measurements. When $\mathbf{u} = \mathbf{v}$, it is simply the weighted sum of the measurements where the adversarial flips have occurred, restricted to the coordinates $\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v}) \cup J$.

The main technical theorem is stated next. Its proof is deferred to Section 5.

THEOREM 3.2. (RAIC UNDER ADVERSARIAL NOISE FOR GAUSSIAN MEASUREMENTS) *Fix $\rho \in (0, 1]$, $\delta, \tau \in (0, 1]$, $k, m, n \in \mathbb{Z}_+$, where*

$$
m \geq \frac{b}{\delta} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12b}{\delta}\right)^{2k} \left(\frac{3a}{\rho}\right)\right) = O\left(\frac{k}{\delta} \log\left(\frac{n}{\delta k}\right) + \frac{1}{\delta} \log\left(\frac{1}{\rho}\right)\right). \qquad (3.10)
$$

Let $\mathcal{A} = \{\mathbf{A}_1, \dots, \mathbf{A}_m \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)\}$ be a set of m i.i.d. standard multivariate normal random vectors, and define the matrix, $\mathbf{A} \in \mathbb{R}^{m \times n}$, which stacks them up: $\mathbf{A} = (\mathbf{A}_1 \ \cdots \ \mathbf{A}_m)^{\mathsf{T}}$. Then, with probability at least $1 - \rho$, uniformly for all $f \in \mathcal{F}_{\mathbf{A}}$, $\mathbf{x}, \mathbf{y} \in S^{n-1} \cap \Sigma_k^n$, $J \subseteq [n]$, $|J| \leq k$,

$$
\left\| (\mathbf{x} - \mathbf{y}) - h_{f;\mathbf{A};J}(\mathbf{x}, \mathbf{y}) \right\|_2 \leq c_1 \sqrt{\delta d_{S^{n-1}}(\mathbf{x}, \mathbf{y})} + c_2 \delta + c_3 \sqrt{\delta \tau} + c_4 \tau \sqrt{\log\left(\frac{2e}{\tau}\right)}. \qquad (3.11)
$$

3.2 Technical Overview The proof of the main theorem, Theorem 3.1, is broadly divided into three steps, each considered under (bounded) adversarial noise: (3.I) establish a stochastic result for Gaussian measurements, (3.II) establish a deterministic result for the iterative approximation errors of BIHT with arbitrary measurements, and (3.III) combine (3.I) and (3.II) to characterize the convergence of BIHT under adversarial noise. The result obtained in Step (3.I) establishes the RAIC for Gaussian measurements under adversarial noise (see, Theorem 3.2) by upper bounding:

$$
\left\| (\mathbf{x} - \mathbf{y}) - h_{f;\mathbf{A};J}(\mathbf{x}, \mathbf{y}) \right\|_2 \leq \tilde{O}\left(\sqrt{\delta d_{S^{n-1}}(\mathbf{x}, \mathbf{y})} + \delta + \tau\right) \qquad (3.12)
$$

uniformly with high probability for all $\mathbf{x}, \mathbf{y} \in S^{n-1} \cap \Sigma_k^n$ and all $J \subseteq [n]$, $|J| \leq k$. The result derived in Step (3.II) upper bounds the error of the t^{th} BIHT approximations deterministically by an expression similar to that in the definition of the RAIC (see, Lemma 4.1):

$$
d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t)}) \leq O\left(\left\| (\mathbf{x} - \hat{\mathbf{x}}^{(t-1)}) - h_{f;\mathbf{A};J}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}) \right\|_2\right). \qquad (3.13)
$$

Lastly, for Step (3.III), upon the establishment of Equations (3.12) and (3.13), the two equations taken together will bound the t^{th} approximation errors by:

$$d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t)}) \leq \begin{cases} 2, & \text{if } t = 0, \\ \tilde{O}\left(\sqrt{\delta d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})} + \delta + \tau\right), & \text{if } t > 0. \end{cases} \tag{3.14}$$

Note, however, that the above expression states the upper bound on the approximation error as a recurrence relation, rather than a closed-form result. Hence, Step (3.III) will also derive a closed-form expression for Equation (3.14) (see, Lemma 4.2), where much of the technical work here has already been accomplished by [28].

The majority of the analysis focuses on the stochastic result in Step (3.I), which is the main technical contribution of this work, while the analyses for the deterministic bound in Step (3.II) and the final step, Step (3.III), are less involved but allow the RAIC established in Step (3.I) to be related to the error of the approximations produced by the BIHT algorithm with corrupted responses. The arguments for Step (3.I) are briefly outlined below. On the other hand, Steps (3.II) and (3.III) are less technically demanding and hence omitted from this overview (see, Lemmas 4.1 and 4.2 and their proofs).

Overview of the Argument for Step (3.I). The idea behind the approach to Theorem 3.2 is the following. There is a (roughly) linear relationship between the expected norm of the sum of up to τm-many measurements and the expected error from adversarially corrupting up to τm-many responses. Hence, since the norm of the sum of every choice of up to τm-many measurements can be uniformly bounded as not "too large" with high probability, the error induced by the adversarial noise is similarly upper bounded with high probability.

More precisely, the argument for Theorem 3.2 is broken down into a few steps: (a) First, applying the triangle inequality, it can be shown that

$$\left\|(\mathbf{x}-\mathbf{y}) - h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y})\right\|_2 \leq \left\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\right\|_2 + \left\|h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{x},\mathbf{x})\right\|_2. \tag{3.15}$$

Then, the focus of the subsequent two steps is upper bounding the two terms on the right-hand-side of the above inequality. Note that Equation (3.15) gives a roughly linear dependence of the approximation error on the amount of adversarial noise. (b) The first term on the right-hand-side of Equation (3.15) can be upper bounded by directly applying [28, Theorem 3.3]. (c) On the other hand, the rightmost term in Equation (3.15)—which (roughly) quantifies the amount of error caused by adversarial noise—requires new analysis. As the first step towards bounding this term, it will be argued that it suffices to bound each element in the image of $h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}$, where $h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}[\{(\mathbf{w},\mathbf{w}) : \mathbf{w} \in S^{n-1} \cap \Sigma_k^n\}]$ has a finite and easily quantifiable size. Note that this approach will lead to a uniform bound on the norm of the image of $h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}$ at every real-valued point, $(\mathbf{x},\mathbf{x})$, $\mathbf{x} \in S^{n-1} \cap \Sigma_k^n$. (d) Finally, such a uniform bound is obtained by bounding the norm of the image of $h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}$ at an arbitrary point $(\mathbf{u},\mathbf{u})$, and subsequently union bounding over a specifically constructed set of such points. Worth noting, this step will orthogonally decompose $h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u})$ into two components, $\langle \mathbf{u}, h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}$ and $h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}$, such that

$$h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u}) = \langle \mathbf{u}, h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u})\rangle \mathbf{u} + \Big(h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}\Big).$$

The norm of each of the two components will be individually upper bounded using concentration inequalities for functions of Gaussians, and subsequently, these bounds will be combined via the triangle inequality,

$$\left\|h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u})\right\|_2 \leq \left|\langle \mathbf{u}, h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u})\rangle\right| + \left\|h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}\right\|_2, \tag{3.16}$$

and a union bound.

4 Proof of Theorem 3.1

As discussed in the technical overview (see, Section 3.2), the proof of the main theorem, Theorem 3.1, follows largely from three intermediate results, which are formalized as Theorem 3.2 —the stochastic result sought in Step (3.I)—and as Lemmas 4.1 and 4.2 in Section 4.1—the deterministic results sought in Steps (3.II) and (3.III), respectively. Recall that Theorem 3.2, the main technical contribution, establishes that with high probability Gaussian measurements satisfy the RAIC under adversarial noise, while Lemmas 4.1 and 4.2 provide a means to relate the RAIC under adversarial noise to a contraction inequality for the sequence of BIHT approximation errors, first as a recurrence relation and subsequently in closed-form.

4.1 Intermediate Results

As already mentioned, the following lemmas will facilitate the proof of Theorem 3.1. The proof of Lemmas 4.1 and Lemma 4.2, can be found in Section D. respectively.

LEMMA 4.1. *For all* $\mathbf{x} \in S^{n-1} \cap \Sigma_k^n$ *and* $t \in \mathbb{Z}_+$, *the error of the* t^{th} *BIHT approximation,* $\hat{\mathbf{x}}^{(t)} \in S^{n-1} \cap \Sigma_k^n$, *is bounded from above by*

$$d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t)}) \leq 4\left\|(\mathbf{x} - \hat{\mathbf{x}}^{(t-1)}) - h_{f;\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})\right\|_2. \tag{4.17}$$

LEMMA 4.2. (CF. [28, LEMMA 4.2]) *Let* $c, c_1, c_2, c_3 > 0$ *be defined as in Equation* (2.1), *and fix* $\tau \in [0, 1]$. *Let* $\gamma \in (0, 1]$, *and define the function* $\varepsilon : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ *by the recurrence relation*

$$\varepsilon(0) = 2, \tag{4.18a}$$

$$\varepsilon(t) = 4c_1\sqrt{\frac{\gamma}{c}\varepsilon(t-1)} + \frac{4c_2\gamma}{c}, \qquad t \in \mathbb{Z}_+. \tag{4.18b}$$

Then,

$$\lim_{t\to\infty} \varepsilon(t) \leq \gamma. \tag{4.19}$$

Moreover, the sequence $\{\varepsilon(t)\}_{t\in\mathbb{Z}_{\geq 0}}$ *is point-wise upper bounded by the sequence*

$$\left\{2^{2^{-t}}\gamma^{1-2^{-t}}\right\}_{t\in\mathbb{Z}_{\geq 0}}. \tag{4.20}$$

4.2 Proof of Theorem 3.1

Proof. (Theorem 3.1). The theorem will follow from an argument analogous to that which appeared in [28, proof of Theorem 3.1 and Corollary 3.2]. By Lemma 4.1, followed by Theorem 3.2, if $m \geq \frac{4bc}{\epsilon}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12b}{\epsilon}\right)^{2k}\left(\frac{3a}{\rho}\right)\right)$, then with probability at least $1-\rho$, for each $\mathbf{x} \in S^{n-1} \cap \Sigma_k^n$ and $t \in \mathbb{Z}_{\geq 0}$, the error of the t^{th} BIHT approximation of $\mathbf{x}$ is bounded from above by:

$$\begin{aligned}
&d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t)}) \\
&\leq 4\left\|(\mathbf{x} - \hat{\mathbf{x}}^{(t-1)}) - h_{f;\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})\right\|_2 \\
&\qquad \blacktriangleright \text{ by Lemma 4.1} \\
&\leq 4\left(c_1\sqrt{\frac{\epsilon}{c}d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})} + \frac{c_2\epsilon}{c} + c_3\sqrt{\epsilon\tau} + c_4\tau\sqrt{\log\left(\frac{2e}{\tau}\right)}\right) \\
&\qquad \blacktriangleright \text{ by Theorem 3.2} \\
&= 4\left(c_1\sqrt{\frac{\epsilon}{c}d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})} + \frac{c_2\epsilon}{c} + \frac{c_2 r}{c}\right) \\
&\qquad \blacktriangleright \text{ by the choice of } r = \frac{c}{c_2}\left(c_3\sqrt{\epsilon\tau} + c_4\tau\sqrt{\log\left(\frac{2e}{\tau}\right)}\right) \\
&= 4\left(c_1\sqrt{\frac{\epsilon}{c}d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})} + \frac{c_2(\epsilon + r - r)}{c} + \frac{c_2 r}{c}\right) \\
&= 4\left(c_1\sqrt{\frac{\epsilon}{c}d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})} + \frac{c_2(\epsilon_0 - r)}{c} + \frac{c_2 r}{c}\right) \\
&= 4\left(c_1\sqrt{\frac{\epsilon}{c}d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})} + \frac{c_2\epsilon_0}{c} - \frac{c_2 r}{c} + \frac{c_2 r}{c}\right) \\
&= 4\left(c_1\sqrt{\frac{\epsilon}{c}d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})} + \frac{c_2\epsilon_0}{c}\right)
\end{aligned}$$

$$\leq 4\left(c_1\sqrt{\frac{\epsilon_0}{c}d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t-1)})}+\frac{c_2\epsilon_0}{c}\right)$$
$$\leq 4c_1\sqrt{\frac{\epsilon_0}{c}d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t-1)})}+\frac{4c_2\epsilon_0}{c}.$$

In summary, with probability at least $1-\rho$, uniformly for all k-sparse, real-valued unit vectors, $\mathbf{x}\in S^{n-1}\cap\Sigma_k^n$, the following holds for all $t\in\mathbb{Z}_+$:

$$d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t)})\leq 4c_1\sqrt{\frac{\epsilon_0}{c}d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t-1)})}+\frac{4c_2\epsilon_0}{c} \tag{4.21}$$

Additionally, trivially, $d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(0)})\leq 2$ since

$$d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(0)})\leq d_{S^{n-1}}(\mathbf{x},-\mathbf{x})=2. \tag{4.22}$$

Next, arbitrarily fixing $\mathbf{x}\in S^{n-1}\cap\Sigma_k^n$, it will be shown by induction that whenever Equations (4.21) and (4.22) hold, the sequence of values $\{\varepsilon(t)\}_{t\in\mathbb{Z}_{\geq 0}}$ point-wise upper bounds the sequence $\{d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t)})\}_{t\in\mathbb{Z}_{\geq 0}}$. where $\varepsilon:\mathbb{Z}_{\geq 0}\to\mathbb{R}$ is defined as in Lemma 4.2.

The base case, when $t=0$, is trivial since $\sup_{\mathbf{u},\mathbf{v}\in\mathbb{R}^n} d_{S^{n-1}}(\mathbf{u},\mathbf{v})=2=\varepsilon(0)$. Now, arbitrarily fixing $t\in\mathbb{Z}_+$, suppose each t'^{th} BIHT approximation, $t'<t$, satisfies $d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t')})\leq\varepsilon(t')$. Then, the aim is to show that $d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t)})\leq\varepsilon(t)$, where $\varepsilon(t)=4c_1\sqrt{\frac{\gamma}{c}\varepsilon(t-1)}+\frac{4c_2\gamma}{c}$ with the fixing of $\gamma=\epsilon_0$. Observe:

$$\begin{aligned}
d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t)}) &\leq 4c_1\sqrt{\frac{\epsilon_0}{c}d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t-1)})}+\frac{4c_2\epsilon_0}{c}\\
&\qquad \blacktriangleright \text{ by Equation (4.21)}\\
&\leq 4c_1\sqrt{\frac{\epsilon_0}{c}\varepsilon(t-1)}+\frac{4c_2\epsilon_0}{c}\\
&\qquad \blacktriangleright \text{ by the inductive assumption}\\
&=\varepsilon(t)
\end{aligned}$$

Said briefly, $d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t)})\leq\varepsilon(t)$, as claimed. By induction, it follows that for all $t\in\mathbb{Z}_{\geq 0}$, the error of the t^{th} BIHT approximation is bounded from above by $d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t)})\leq\varepsilon(t)$. Extending this to all other $\mathbf{x}\in S^{n-1}\cap\Sigma_k^n$ via the earlier discussion, every such sequence of BIHT approximations is point-wise upper bounded by ε with high probability.

Having verified the above, the theorems follow immediately from Lemma 4.2:

$$d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t)})\leq\varepsilon(t)\leq 2^{2^{-t}}\epsilon_0^{1-2^{-t}},$$
$$\lim_{t\to\infty} d_{S^{n-1}}(\mathbf{x},\hat{\mathbf{x}}^{(t)})\leq\epsilon_0.$$

This completes the proof of the main theorem, Theorem 3.1. □

5 Proof of the Main Technical Theorem (Theorem 3.2)

5.1 Discussion and Preliminaries We begin by introducing and verifying some results that will set us up for proving the main technical theorem, Theorem 3.2. Throughout Section 5, the function $f:\mathbb{R}^n\to\{-1,1\}^m$ is taken to be any function which upholds: $d_H(f(\mathbf{w}),\,\mathsf{sign}(\mathbf{A}\mathbf{w}))\leq\tau m$ at every point, $\mathbf{w}\in\mathbb{R}^n$. Specification of this condition will be henceforth omitted to avoid redundancy.

First off, the left-hand-side of Equation (3.11) in Theorem 3.2, $\|(\mathbf{x}-\mathbf{y})-h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2$, is split into two components (with bounding).

Claim 5.1. *For all $\mathbf{x},\mathbf{y}\in\mathbb{R}^n$ and $J\subseteq[n]$, the following inequality (deterministically) holds:*

$$\|(\mathbf{x}-\mathbf{y})-h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2\leq\|(\mathbf{x}-\mathbf{y})-h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2+\left\|h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{x},\mathbf{x})\right\|_2. \tag{5.23}$$

Proof. (Claim 5.1). Fix $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $J \subseteq [n]$, arbitrarily. Let $J' \triangleq \mathsf{supp}(\mathbf{y}) \cup J$. The (random) vector $(\mathbf{x}-\mathbf{y}) - h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y})$ can be rewritten as follows:

$$(\mathbf{x}-\mathbf{y}) - h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y}) = \big((\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\big) - \big(h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\big),$$

where the second and fourth terms on the right-hand-side cancel. Additionally, observe:

$$\begin{aligned}
&h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\\
&= \frac{\sqrt{2\pi}}{m}\sum_{i=1}^{m} T_{\mathsf{supp}(\mathbf{x})\cup\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{A}_i)\cdot\frac{1}{2}\big(f(\mathbf{x})_i - \mathsf{sign}(\langle \mathbf{y}, \mathbf{A}_i\rangle)\big)\\
&\quad - \frac{\sqrt{2\pi}}{m}\sum_{i=1}^{m} T_{\mathsf{supp}(\mathbf{x})\cup\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{A}_i)\cdot\frac{1}{2}\big(\mathsf{sign}(\langle \mathbf{x}, \mathbf{A}_i\rangle) - \mathsf{sign}(\langle \mathbf{y}, \mathbf{A}_i\rangle)\big)\\
&= \frac{\sqrt{2\pi}}{m}\sum_{i=1}^{m} T_{\mathsf{supp}(\mathbf{x})\cup\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{A}_i)\cdot\frac{1}{2}\Big(\big(f(\mathbf{x})_i - \mathsf{sign}(\langle \mathbf{y}, \mathbf{A}_i\rangle)\big) - \big(\mathsf{sign}(\langle \mathbf{x}, \mathbf{A}_i\rangle) - \mathsf{sign}(\langle \mathbf{y}, \mathbf{A}_i\rangle)\big)\Big)\\
&= \frac{\sqrt{2\pi}}{m}\sum_{i=1}^{m} T_{\mathsf{supp}(\mathbf{x})\cup\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{A}_i)\cdot\frac{1}{2}\big(f(\mathbf{x})_i - \mathsf{sign}(\langle \mathbf{x}, \mathbf{A}_i\rangle)\big)\\
&= h_{f;\mathbf{A};J'}(\mathbf{x},\mathbf{x})
\end{aligned}$$

Thus, combining above work:

$$\begin{aligned}
(\mathbf{x}-\mathbf{y}) - h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y}) &= \big((\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\big) - \big(h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\big)\\
&= \big((\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\big) - h_{f;\mathbf{A};J'}(\mathbf{x},\mathbf{x}).
\end{aligned}$$

Then, the norm is upper bounded as follows:

$$\begin{aligned}
\|(\mathbf{x}-\mathbf{y}) - h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 &= \big\|\big((\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\big) - h_{f;\mathbf{A};J'}(\mathbf{x},\mathbf{x})\big\|_2\\
&\le \|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 + \|h_{f;\mathbf{A};J'}(\mathbf{x},\mathbf{x})\|_2
\end{aligned}$$

where the bottom line applies the triangle inequality. ∎

As in the the proof of Claim 5.1, write $J' \triangleq \mathsf{supp}(\mathbf{y}) \cup J$. For convenience, this notation will be used throughout the remainder of this manuscript. Note that $|J'| \le 2k$ by design. Equation (5.23) of Claim 5.1 decomposes (with bounding) the random variable of interest, $\|(\mathbf{x}-\mathbf{y}) - h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2$, into two terms which are individually easier to control. The majority of the argument for Theorem 3.2 is towards a uniform upper bound on the latter term of this decomposition, $D_{2;J'}(\mathbf{x},\mathbf{x}) \triangleq \|h_{f;\mathbf{A};J'}(\mathbf{x},\mathbf{x})\|_2$. This second term, $D_{2;J'}(\mathbf{x},\mathbf{x})$, requires new analysis, which takes up Section 5.1.1 and Appendices A-B. On the other hand, the first term, $D_{1;J}(\mathbf{x},\mathbf{x}) \triangleq \|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2$, is immediately upper bounded with high probability for all $\mathbf{x},\mathbf{y} \in S^{n-1} \cap \Sigma_k^n$ and for each $J \subseteq [n]$, $|J| \le k$, via [28, Theorem 3.3], stated below.

Lemma 5.1. ([28, Theorem 3.3]) *Fix $\epsilon', \rho' \in (0,1)$, $k, m, n \in \mathbb{Z}_+$, such that*

$$m \ge \frac{b}{\epsilon'}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12b}{\epsilon'}\right)^{2k}\left(\frac{a}{\rho'}\right)\right). \tag{5.24}$$

Then, uniformly with probability at least $1-\rho'$, the Gaussian measurement matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ satisfies the $(k, n, \epsilon', c_1, c_2)$-RAIC:

$$\big\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\big\|_2 \le c_1\sqrt{\epsilon' d_{S^{n-1}}(\mathbf{x},\mathbf{y})} + c_2\epsilon' \tag{5.25}$$

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

Next, we turn our attention to the random variable $D_{2;J'}(\mathbf{x},\mathbf{x})$.

5.1.1 Discussion Regarding $D_{2;J'}(\mathbf{x},\mathbf{x})$ The derivation of an upper bound on the random variable, $D_{2;J'}(\mathbf{x},\mathbf{x}) = \|h_{f;\mathbf{A};J'}(\mathbf{x},\mathbf{x})\|_2$, will entirely ignore the specification of the vector $\mathbf{x}$. Rather, by the definition of $h_{f;\mathbf{A};J'}$,

$$\begin{aligned} h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) &= \frac{\sqrt{2\pi}}{m}\sum_{i=1}^{m} T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i)\cdot\frac{1}{2}\Big(f(\mathbf{u})_i - \mathsf{sign}(\langle \mathbf{u},\mathbf{A}_i\rangle)\Big) \\ &= \frac{\sqrt{2\pi}}{m}\sum_{i=1}^{m} T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i)\cdot\frac{1}{2}\Big(f(\mathbf{u}) - \mathsf{sign}(\mathbf{A}\mathbf{u})\Big)_i \end{aligned}$$

where $\mathbf{u}\in\mathbb{R}^n$. The only dependence of $h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})$ on the preimage, $(\mathbf{u},\mathbf{u})$, is captured in the expression: $\frac{1}{2}(f(\mathbf{u}) - \mathsf{sign}(\mathbf{A}\mathbf{u}))$. In particular, taking $f\in\mathcal{F}_{\mathbf{A}}$, note that $\frac{1}{2}(f(\mathbf{u}) - \mathsf{sign}(\mathbf{A}\mathbf{u}))\in\{\mathbf{z}\in\{-1,0,1\}^m : 1\leq \|\mathbf{z}\|_0\leq \tau m\}$ for all $\mathbf{u}\in\mathbb{R}^n$. All other terms in the definition of $h_{f;\mathbf{A};J'}$ are fixed across all $\mathbf{u}\in\mathbb{R}^n$. Thus, upon fixing the set of Gaussian vectors, $\mathbf{A}_1,\dots,\mathbf{A}_m$, for each $J'\subseteq[n]$, $|J'|\leq 2k$, the image of the function $h_{f;\mathbf{A};J'}$ has finite cardinality no more than:

$$\left|h_{f;\mathbf{A};J'}\left[\{(\mathbf{w},\mathbf{w}) : \mathbf{w}\in S^{n-1}\cap\Sigma_k^n\}\right]\right| \leq \sum_{\ell=1}^{\tau m}\binom{m}{\ell}2^{\ell}.$$

Additionally, recall that earlier we took $J' = \mathsf{supp}(\mathbf{y})\cup J$, where in the case of $D_{2;J'}(\mathbf{x},\mathbf{x})$, the only way that $\mathbf{y}$ comes into play is in regard to its support. Later on, we will union bound over all possible coordinate subsets which can form $\mathsf{supp}(\mathbf{y})\cup J$—i.e., all $J'\subseteq[n]$, $|J'|\leq 2k$. Hence, in effect, we can consider J' to be an arbitrary coordinate subset of cardinality at most $2k$, as we will proceed with throughout this discussion and in the formal proof of the bound on $D_{2;J'}(\mathbf{x},\mathbf{x})$ (i.e., Section 5.1.1 and Appendices A-C). It therefore suffices to enumerate each of the up to $\sum_{\ell=1}^{\tau m}\binom{m}{\ell}2^{\ell}$-many vectors comprising $h_{f;\mathbf{A};J'}\left[\{(\mathbf{w},\mathbf{w}) : \mathbf{w}\in S^{n-1}\cap\Sigma_k^n\}\right]$ and bound their norms for each choice of $J'\subseteq[n]$, $|J'|\leq 2k$. With this motivation, construct a collection of sets, $\mathcal{U}_{J'}\subseteq S^{n-1}\cap\Sigma_k^n$, $J'\subseteq[n]$, by inserting precisely one vector, $\mathbf{u}\in S^{n-1}\cap\Sigma_k^n$, into $\mathcal{U}_{J'}$ for each vector $\mathbf{z}\in h_{f;\mathbf{A};J'}\left[\{(\mathbf{w},\mathbf{w}) : \mathbf{w}\in S^{n-1}\cap\Sigma_k^n\}\right]$ such that $h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) = \mathbf{z}$. Note that $|\mathcal{U}_{J'}| = \sum_{\ell=1}^{\tau m}\binom{m}{\ell}2^{\ell}$ by design. With this construction, the above discussion is formalized and verified in the following claim and its proof.

CLAIM 5.2. *Fix $\gamma\geq 0$. Suppose for all $J'\subseteq[n]$, $|J'|\leq 2k$, and $\mathbf{u}\in\mathcal{U}_{J'}$, the norm of $h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})$ is bounded from above by: $\|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\|_2\leq\gamma$. Then, uniformly for all $\mathbf{x}\in S^{n-1}\cap\Sigma_k^n$ and for all $J'\subseteq[n]$, $|J'|\leq 2k$, the same bound holds at $(\mathbf{x},\mathbf{x})$: $\|h_{f;\mathbf{A};J'}(\mathbf{x},\mathbf{x})\|_2\leq\gamma$.*

Proof. (Claim 5.2). Suppose for the sake of contradiction that $\|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\|_2\leq\gamma$ for each $J'\subseteq[n]$, $|J'|\leq 2k$, and $\mathbf{u}\in\mathcal{U}_{J'}$, but there exists $J_0\subseteq[n]$, $|J_0|\leq 2k$, and $\mathbf{x}\in S^{n-1}\cap\Sigma_k^n$, for which $\|h_{f;\mathbf{A};J_0}(\mathbf{x},\mathbf{x})\|_2 > \gamma$. Denote the image of $(\mathbf{x},\mathbf{x})$ under $h_{f;\mathbf{A};J_0}$ by $\mathbf{z} = h_{f;\mathbf{A};J_0}(\mathbf{x},\mathbf{x})$, where by assumption, $\|\mathbf{z}\|_2>\gamma$, and let $\mathcal{W} = \{\mathbf{w}\in S^{n-1}\cap\Sigma_k^n : h_{f;\mathbf{A};J_0}(\mathbf{w},\mathbf{w}) = \mathbf{z}\}$. Then, by the construction of the set $\mathcal{U}_{J_0}$, it must be that $|\mathcal{U}_{J_0}\cap\mathcal{W}| = 1$, which implies that there exists $\mathbf{u}\in\mathcal{U}_{J_0}$ for which $\|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\|_2 = \|\mathbf{z}\|_2>\gamma$—a contradiction. By this contradiction, the claim holds. ∎

Due Claim 5.2, the proof of Theorem 3.2 will seek to bound $\|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\|_2$ from above for all $J'\subseteq[n]$, $|J'|\leq 2k$, and $\mathbf{u}\in\mathcal{U}_{J'}$. Specifically, Lemma 5.2 controls this random variable, $D_{2;J'}(\mathbf{u},\mathbf{u}) = \|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\|_2$, uniformly for every $J'\subseteq[n]$, $|J'|\leq 2k$, and $\mathbf{u}\in\mathcal{U}_{J'}$.

LEMMA 5.2. *Let $m\in\mathbb{Z}_+$ satisfy*

$$m\geq\frac{b}{\delta}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12b}{\delta}\right)^{2k}\left(\frac{3a}{\rho}\right)\right) = O\left(\frac{k}{\delta}\log\left(\frac{n}{\delta k}\right)+\frac{1}{\delta}\log\left(\frac{1}{\rho}\right)\right).$$

With probability at least $1-\frac{2\rho}{3}$, uniformly for all $J'\subseteq[n]$, $|J'|\leq 2k$, and $\mathbf{u}\in\mathcal{U}_{J'}$,

$$D_{2;J'}(\mathbf{u},\mathbf{u})\leq(12+\sqrt{3})\sqrt{\frac{\pi\delta\tau}{b}}+\left(2+4\sqrt{\pi}\right)\tau\sqrt{\log\left(\frac{2e}{\tau}\right)} = c_3\sqrt{\delta\tau}+c_4\tau\sqrt{\log\left(\frac{2e}{\tau}\right)}. \tag{5.26}$$

The proof of the lemma is deferred to Appendices A-C. Next, Theorem 3.2 is proved, contingent on the proof of Lemma 5.2.

Proof. (Theorem 3.2). Fix $m = \frac{b}{\delta}\log\Big(\binom{n}{k}^2\binom{n}{2k}(\frac{12b}{\delta})^{2k}(\frac{3a}{\rho})\Big)$. Due to Claim 5.1, for every $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and every $J \subseteq [n]$,

$$\|(\mathbf{x}-\mathbf{y}) - h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 \leq \|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 + \left\|h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{x},\mathbf{x})\right\|_2$$

By Lemma 5.1 ([28, Theorem 3.3]), with probability at least $1-\frac{\rho}{3}$, uniformly for all $\mathbf{x}, \mathbf{y} \in S^{n-1} \cap \Sigma_k^n$ and $J \subseteq [n]$, $|J| \leq k$,

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

where $c_1, c_2 > 0$ are universal constants as defined in Equation (2.1). Additionally, by Lemma 5.2, with probability at least $1-\frac{2\rho}{3}$, uniformly for all $J \subseteq [n]$, $|J| \leq k$, and $\mathbf{u} \in \mathcal{U}_J$,

$$\left\|h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{u},\mathbf{u})\right\|_2 \leq (12+\sqrt{3})\sqrt{\frac{\pi\delta\tau}{b}} + \big(2+4\sqrt{\pi}\big)\tau\sqrt{\log\left(\frac{2e}{\tau}\right)} = c_3\sqrt{\delta\tau} + c_4\tau\sqrt{\log\left(\frac{2e}{\tau}\right)}.$$

Recalling Claim 5.2, it follows that with the same probability, the same bound on $\left\|h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{x},\mathbf{x})\right\|_2$ holds uniformly over all $\mathbf{x}, \mathbf{y} \in S^{n-1} \cap \Sigma_k^n$ and all choices of $J \subseteq [n]$, such that $|\mathsf{supp}(\mathbf{y}) \cup J| \leq 2k$. Combining the above bounds on $\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2$ and $\left\|h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{x},\mathbf{x})\right\|_2$ via a union bound, and applying Claim 5.1, the desired upper bound follows: with probability at least $1-\frac{\rho}{3}-\frac{2\rho}{3} = 1-\rho$, uniformly for all $\mathbf{x}, \mathbf{y} \in S^{n-1} \cap \Sigma_k^n$, $J \subseteq [n]$, $|J| \leq k$:

$$\begin{aligned}\|(\mathbf{x}-\mathbf{y}) - h_{f;\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 &\leq \|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 + \left\|h_{f;\mathbf{A};\mathsf{supp}(\mathbf{y})\cup J}(\mathbf{x},\mathbf{x})\right\|_2 \\ &\leq c_1\sqrt{\delta d_{S^{n-1}}(\mathbf{x},\mathbf{y})} + c_2\delta + (12+\sqrt{3})\sqrt{\frac{\pi\delta\tau}{b}} + \big(2+4\sqrt{\pi}\big)\tau\sqrt{\log\left(\frac{2e}{\tau}\right)} \\ &= c_1\sqrt{\delta d_{S^{n-1}}(\mathbf{x},\mathbf{y})} + c_2\delta + c_3\sqrt{\delta\tau} + c_4\tau\sqrt{\log\left(\frac{2e}{\tau}\right)}\end{aligned}$$

where $c_1, c_2, c_3, c_4 > 0$ are universal constants specified in Equation (2.1). ∎

Acknowledgement. This research is based upon work supported by the National Science Foundation under Grant No. 2133484 and 2217058, and by the NSF Graduate Research Fellowship Program under Grant No. DGE-2038238.

References

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

[2] P. Awasthi, M.-F. Balcan, N. Haghtalab, and H. Zhang, *Learning and 1-bit compressed sensing under asymmetric noise*, in Conference on Learning Theory, PMLR, 2016, pp. 152–192.

[3] P. Awasthi, M. F. Balcan, and P. M. Long, *The power of localization for efficiently learning linear separators with noise*, Journal of the ACM (JACM), 63 (2017), pp. 1–27.

[4] 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, 63 (2017), pp. 3368–3385.

[5] T. Blumensath and M. E. Davies, *Iterative hard thresholding for compressed sensing*, Appl. Comput. Harmon. Anal, 27 (2009), pp. 265–274.

[6] P. Boufounos and R. G. Baraniuk, *1-bit compressive sensing*, in 42nd Annual Conference on Information Sciences and Systems, CISS 2008, Princeton, NJ, USA, 19-21 March 2008, 2008, pp. 16–21.

[7] 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.

[8] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Transactions on Information Theory, 52 (2006), pp. 489–509.

[9] J. Cheng, T. Liu, K. Ramamohanarao, and D. Tao, *Learning with bounded instance and label-dependent label noise*, in International Conference on Machine Learning, PMLR, 2020, pp. 1789–1799.

[10] G. Chinot, F. Kuchelmeister, M. Löffler, and S. van de Geer, *Adaboost and robust one-bit compressed sensing*, Mathematical Statistics and Learning, 5 (2022), pp. 117–158.

[11] D.-Q. Dai, L. Shen, Y. Xu, and N. Zhang, *Noisy 1-bit compressive sensing: models and algorithms*, Applied and Computational Harmonic Analysis, 40 (2016), pp. 1–32.

[12] I. Diakonikolas, T. Gouleakis, and C. Tzamos, *Distribution-independent pac learning of halfspaces with massart noise*, Advances in Neural Information Processing Systems, 32 (2019).

[13] I. Diakonikolas, V. Kontonis, C. Tzamos, and N. Zarifis, *Learning general halfspaces with adversarial label noise via online gradient descent*, in International Conference on Machine Learning, PMLR, 2022, pp. 5118–5141.

[14] S. Dirksen and S. Mendelson, *Non-gaussian hyperplane tessellations and robust one-bit compressed sensing*, Journal of the European Mathematical Society, 23 (2021), pp. 2913–2947.

[15] ——, *Robust one-bit compressed sensing with partial circulant matrices*, The Annals of Applied Probability, 33 (2023), pp. 1874–1903.

[16] D. L. Donoho, *Compressed sensing*, IEEE Transactions on information theory, 52 (2006), pp. 1289–1306.

[17] L. Flodin, V. Gandikota, and A. Mazumdar, *Superset technique for approximate recovery in one-bit compressed sensing*, in Advances in Neural Information Processing Systems, 2019, pp. 10387–10396.

[18] S. Frei, Y. Cao, and Q. Gu, *Agnostic learning of halfspaces with gradient descent via soft margins*, in International Conference on Machine Learning, PMLR, 2021, pp. 3417–3426.

[19] Y. Freund and R. E. Schapire, *A decision-theoretic generalization of on-line learning and an application to boosting*, Journal of computer and system sciences, 55 (1997), pp. 119–139.

[20] M. P. Friedlander, H. Jeong, Y. Plan, and Ö. Yilmaz, *Nbiht: An efficient algorithm for 1-bit compressed sensing with optimal error decay rate*, IEEE Transactions on Information Theory, 68 (2021), pp. 1157–1177.

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

[22] J. Huang, Y. Jiao, X. Lu, and L. Zhu, *Robust decoding from 1-bit compressive sampling with ordinary and regularized least squares*, SIAM Journal on Scientific Computing, 40 (2018), pp. A2062–A2086.

[23] T. Huynh and R. Saab, *Fast binary embeddings and quantized compressed sensing with structured matrices*, Communications on Pure and Applied Mathematics, 73 (2020), pp. 110–149.

[24] 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).

[25] 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, 59 (2013), pp. 2082–2102.

[26] Z. Ji, K. Ahn, P. Awasthi, S. Kale, and S. Karp, *Agnostic learnability of halfspaces via logistic loss*, in International Conference on Machine Learning, PMLR, 2022, pp. 10068–10103.

[27] D. Liu, S. Li, and Y. Shen, *One-bit compressive sensing with projected subgradient method under sparsity constraints*, IEEE Transactions on Information Theory, 65 (2019), pp. 6650–6663.

[28] N. Matsumoto and A. Mazumdar, *Binary iterative hard thresholding converges with optimal number of measurements for 1-bit compressed sensing*, in 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2022, pp. 813–822.

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

[30] A. K. Menon, B. Van Rooyen, and N. Natarajan, *Learning from binary labels with instance-dependent noise*, Machine Learning, 107 (2018), pp. 1561–1595.

[31] Y. Plan and R. Vershynin, *Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach*, IEEE Transactions on Information Theory, 59 (2012), pp. 482–494.

[32] ———, *One-bit compressed sensing by linear programming*, Communications on Pure and Applied Mathematics, 66 (2013), pp. 1275–1297.

[33] Y. Plan, R. Vershynin, and E. Yudovina, *High-dimensional estimation with geometric constraints*, Information and Inference: A Journal of the IMA, 6 (2017), pp. 1–40.

[34] J. Shen and C. Zhang, *Attribute-efficient learning of halfspaces with malicious noise: Near-optimal label complexity and noise tolerance*, in Algorithmic Learning Theory, PMLR, 2021, pp. 1072–1113.

[35] M. J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*, vol. 48, Cambridge university press, 2019.

[36] M. Yan, Y. Yang, and S. Osher, *Robust 1-bit compressive sensing using adaptive outlier pursuit*, IEEE Transactions on Signal Processing, 60 (2012), pp. 3868–3875.

[37] S. Yan and C. Zhang, *Revisiting perceptron: Efficient and label-optimal learning of halfspaces*, Advances in Neural Information Processing Systems, 30 (2017).

[38] C. Zhang, *Efficient active learning of sparse halfspaces*, in Conference on Learning Theory, PMLR, 2018, pp. 1856–1880.

A Analysis for $D_{2;J'}(\mathbf{u},\mathbf{u})$

Appendices A-C contain the analysis to bound the random variable $D_{2;J'}(\mathbf{u},\mathbf{u}) \triangleq \|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\|_2$ from above, as per Lemma 5.2. Appendix A orthogonally decomposes the random vector $h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})$ into two components and upper bounds the norm of each. The derivations of these bounds largely rely on tailored concentration inequalities, whose proofs constitute Appendix C. The lemma that we ultimately seek, Lemma 5.2, is established in Appendix B by combining the results proved in Appendix A.

A.1 An Orthogonal Decomposition To control the random variable $D_{2;J'}(\mathbf{u},\mathbf{u}) \triangleq \|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\|_2$, we begin by orthogonally decomposing the random vector $h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})$ into two components: $\langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}$ and $h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}$, where

$$h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) = \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u} + \big(h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}\big). \tag{A.1}$$

As such, define the random variables, $D'_{1;J'}(\mathbf{u},\mathbf{u})$ and $D'_{2;J'}(\mathbf{u},\mathbf{u})$, by

$$D'_{1;J'}(\mathbf{u},\mathbf{u}) = |\langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle|, \tag{A.2}$$

$$D'_{2;J'}(\mathbf{u},\mathbf{u}) = \|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}\|_2. \tag{A.3}$$

We make use of this decomposition and these random variables in the following claim.

CLAIM A.1. *For any $J' \subseteq [n]$ and $\mathbf{u} \in \mathbb{R}^n$,*

$$D_{2;J'}(\mathbf{u},\mathbf{u}) \le D'_{1;J'}(\mathbf{u},\mathbf{u}) + D'_{2;J'}(\mathbf{u},\mathbf{u}). \tag{A.4}$$

Proof. (Claim A.1). The claim directly follows from the orthogonal decomposition discussed above and the triangle inequality:

$$\begin{aligned}
D_{2;J'}(\mathbf{u},\mathbf{u}) &= \|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\|_2 \\
&\qquad \blacktriangleright \text{ by the definition of the random variable } D_{2;J'}(\mathbf{u},\mathbf{u}) \\
&= \big\|\langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u} + \big(h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}\big)\big\|_2 \\
&\qquad \blacktriangleright \text{ by Equation (A.1)} \\
&\le \|\langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}\|_2 + \|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}\|_2 \\
&\qquad \blacktriangleright \text{ by the triangle inequality} \\
&\le |\langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle| \|\mathbf{u}\|_2 + \|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}\|_2 \\
&\qquad \blacktriangleright \text{ due to the homogeneity of norms} \\
&\le |\langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle| + \|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}\|_2 \\
&\qquad \blacktriangleright \because \|\mathbf{u}\|_2 = 1 \\
&= D'_{1;J'}(\mathbf{u},\mathbf{u}) + D'_{2;J'}(\mathbf{u},\mathbf{u}) \\
&\qquad \blacktriangleright \text{ by the definitions of the random variables } D'_{1;J'}(\mathbf{u},\mathbf{u}), D'_{2;J'}(\mathbf{u},\mathbf{u})
\end{aligned}$$

Therefore, $D_{2;J'}(\mathbf{u},\mathbf{u}) \le D'_{1;J'}(\mathbf{u},\mathbf{u}) + D'_{2;J'}(\mathbf{u},\mathbf{u})$, completing the proof of the claim. ∎

Due to Claim A.1, above, $D_{2;J'}(\mathbf{u},\mathbf{u})$ can be upper bounded by individually bounding $D'_{1;J'}(\mathbf{u},\mathbf{u})$ and $D'_{2;J'}(\mathbf{u},\mathbf{u})$, which are simpler to handle than directly characterizing $D_{2;J'}(\mathbf{u},\mathbf{u})$. Such bounds are obtained in Appendices A.3 and A.4, respectively. These results will lead to the proof of Lemma 5.2 in Appendix B, which formally upper bounds $D_{2;J'}(\mathbf{u},\mathbf{u})$.

A.2 Concentration Inequalities for the Orthogonal Decomposition Before the random variables, $D'_{1;J'}(\mathbf{u},\mathbf{u})$ and $D'_{2;J'}(\mathbf{u},\mathbf{u})$, are bounded, two concentration inequalities are stated below as Lemmas A.1 and A.2 to facilitate the analysis. The proofs of these lemmas are deferred to Appendix C.

LEMMA A.1. *Fix $t > 0$, $\ell \in \mathbb{Z}_+$. Let $\mathcal{Z} = \{\mathbf{Z}_1, \dots, \mathbf{Z}_\ell \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)\}$ be a collection of ℓ i.i.d. Gaussian vectors, and fix a k-sparse, real-valued unit vector, $\mathbf{u} \in S^{n-1} \cap \Sigma^n_k$, and a coordinate subset, $J'' \subseteq [n]$, Define the random variables*

$$X_i \triangleq \big\langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i\rangle))\big\rangle,$$

for $i \in [\ell]$, and write their sum as

$$\bar{X} \triangleq \sum_{i=1}^{\ell} X_i = \left\langle \mathbf{u}, \sum_{i=1}^{\ell} T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i\rangle)) \right\rangle.$$

The concentration and mean of the random variable $\bar{X}$ are such that

$$\Pr\left(\bar{X} > \left(\sqrt{\frac{2}{\pi}} + t\right)\ell\right) \le e^{-\frac{1}{2}\ell t^2}. \tag{A.5}$$

Additionally, for each $i \in [\ell]$, there is an equivalence: $|X_i| = X_i$.

LEMMA A.2. *Fix $t > 0$, $\ell \in \mathbb{Z}_+$. Let $\mathcal{Z} = \{\mathbf{Z}_1, \ldots, \mathbf{Z}_\ell \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)\}$ be a collection of ℓ i.i.d. Gaussian vectors, and fix a k-sparse, real-valued unit vector, $\mathbf{u} \in S^{n-1} \cap \Sigma_k^n$, and a coordinate subset, $J'' \subseteq [n]$. Write $k' \triangleq |\mathsf{supp}(\mathbf{u}) \cup J''|$. Define the random vector*

$$\bar{\mathbf{Y}}_{\mathbf{u}} = \sum_{i=1}^{\ell} \Big(T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)) - \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)) \rangle \mathbf{u} \Big).$$

The concentration and mean of the random variable representing its norm, $\|\bar{\mathbf{Y}}_{\mathbf{u}}\|_2$, are such that

$$\Pr\left(\left\| \bar{\mathbf{Y}}_{\mathbf{u}} \right\|_2 > \sqrt{\frac{(k'-1)\ell}{2}} + \ell t \right) \leq e^{-\frac{1}{2}\ell t^2}. \tag{A.6}$$

A.3 Bounding $D'_{1;J'}(\mathbf{u}, \mathbf{u}) \triangleq |\langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u}, \mathbf{u}) \rangle|$ Having introduced the concentration inequalities in Appendix A.2, we are ready to bound the random variables $D'_{1;J'}(\mathbf{u}, \mathbf{u}) \triangleq |\langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u}, \mathbf{u}) \rangle|$ and $D'_{2;J'}(\mathbf{u}, \mathbf{u}) \triangleq \|h_{f;\mathbf{A};J'}(\mathbf{u}, \mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u}, \mathbf{u}) \rangle \mathbf{u}\|_2$. To start off, the random variable $D'_{1;J'}(\mathbf{u}, \mathbf{u})$ is bounded from above per the following lemma.

LEMMA A.3. *Let $a, b > 0$ be the universal constants specified in Equation (2.1), and fix $\rho \in (0, 1]$. Suppose*

$$m \geq \frac{b}{\delta} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12b}{\delta} \right)^{2k} \left(\frac{3a}{\rho} \right) \right) = O\left(\frac{k}{\delta} \log\left(\frac{n}{\delta k} \right) + \frac{1}{\delta} \log\left(\frac{1}{\rho} \right) \right).$$

Then, with probability at least $1 - \frac{\rho}{3}$, uniformly for all $J' \subseteq [n]$, $|J'| \leq 2k$, and $\mathbf{u} \in \mathcal{U}_{J'}$,

$$D'_{1;J'}(\mathbf{u}, \mathbf{u}) \leq \frac{2\ell}{m} + \frac{\sqrt{2\pi}\ell t}{m} \tag{A.7}$$

where

$$\ell \leq \tau m, \tag{A.8}$$

$$t = \sqrt{\frac{2}{\ell} \log\left(2 \cdot 2^{\ell} \binom{m}{\ell} \binom{n}{2k} \frac{3\tau m}{\rho} \right)}. \tag{A.9}$$

Proof. (Lemma A.3). First, expanding out and rewriting the expression for $h_{f;\mathbf{A};J'}$ yields:

$$\begin{aligned} h_{f;\mathbf{A};J'}(\mathbf{u}, \mathbf{u}) &= -\frac{\sqrt{2\pi}}{m} \sum_{i=1}^{m} T_{\mathsf{supp}(\mathbf{u}) \cup J'}(\mathbf{A}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i \rangle)) \cdot \mathbb{I}(f(\mathbf{u})_i \neq \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i \rangle)) \\ & \qquad \blacktriangleright \text{ by Equation (3.9a) in Section 3.1} \\ &= -\frac{\sqrt{2\pi}}{m} \sum_{i \in I} T_{\mathsf{supp}(\mathbf{u}) \cup J'}(\mathbf{A}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i \rangle)) \\ & \qquad \blacktriangleright \text{ per the remark below} \end{aligned}$$

where $I \subseteq [m]$ indexes the sign-mismatches:

$$I \triangleq \{i \in [m] : f(\mathbf{u})_i \neq \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i \rangle)\} \equiv \{i \in [m] : \mathbb{I}(f(\mathbf{u})_i \neq \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i \rangle)) \neq 0\}.$$

Note that the assumption on f stated at the beginning of Appendix 5—that the number of corrupted responses is bounded—ensures that $|I| \leq \tau m$. Fix $\ell = |I| \leq \tau m$, and without loss of generality, assume $I = [\ell]$. Under this assumption, the above derivation implies:

$$h_{f;\mathbf{A};J'}(\mathbf{u}, \mathbf{u}) = -\frac{\sqrt{2\pi}}{m} \sum_{i=1}^{\ell} T_{\mathsf{supp}(\mathbf{u}) \cup J'}(\mathbf{A}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i \rangle)),$$

or equivalently,

$$-h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) = \frac{\sqrt{2\pi}}{m}\sum_{i=1}^{\ell} T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle)).$$

Now, define the random variables

$$X_i \triangleq \big\langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle))\big\rangle, \qquad i \in [\ell],$$

and let $\bar{X} \triangleq \sum_{i=1}^{\ell} X_i$. Note that by Lemma A.1, $|X_i| = X_i$ for each $i \in [\ell]$, and thus, $|\bar{X}| = \bar{X}$, as shown in the following brief derivation:

$$|\bar{X}| = \left|\sum_{i=1}^{\ell} X_i\right| = \left|\sum_{i=1}^{\ell} |X_i|\right| = \sum_{i=1}^{\ell} |X_i| = \sum_{i=1}^{\ell} X_i = \bar{X}.$$

Since $|\bar{X}| = \bar{X}$, any bound that holds for $\bar{X}$ must also hold for $|\bar{X}|$. Hence, this proof will focus on upper bounding the value taken by $\bar{X}$, rather that directly characterizing $|\bar{X}|$. Using the notation of these random variables, observe:

$$\begin{aligned}
\langle \mathbf{u}, -h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle &= \left\langle \mathbf{u}, \frac{\sqrt{2\pi}}{m}\sum_{i=1}^{\ell} T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle))\right\rangle \\
&= \frac{\sqrt{2\pi}}{m}\sum_{i=1}^{\ell} \big\langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle))\big\rangle \\
&= \frac{\sqrt{2\pi}}{m}\sum_{i=1}^{\ell} X_i \\
&= \frac{\sqrt{2\pi}}{m}\bar{X}.
\end{aligned}$$

Note that by the above observations,

$$\langle \mathbf{u}, -h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle = \frac{\sqrt{2\pi}}{m}\bar{X} = \frac{\sqrt{2\pi}}{m}\big|\bar{X}\big| = \left|\frac{\sqrt{2\pi}}{m}\bar{X}\right| \geq 0,$$

and therefore, $|\langle \mathbf{u}, -h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle| = \langle \mathbf{u}, -h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle$. Due to Lemma A.1, the random variable $\bar{X}$ is bounded from above by

$$\bar{X} \leq \mathbb{E}[\bar{X}] + \ell t \leq \sqrt{\frac{2}{\pi}}\ell + \ell t = \left(\sqrt{\frac{2}{\pi}} + t\right)\ell$$

with probability at least $1 - e^{-\frac{1}{2}\ell t^2}$. Take

$$t = \sqrt{\frac{2}{\ell}\log\left(2\cdot 2^{\ell}\binom{m}{\ell}\binom{n}{2k}\frac{3\tau m}{\rho}\right)}.$$

Then, the desired bound follows:

$$\begin{aligned}
|\langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle| &= |-\langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle| \\
&= |\langle \mathbf{u}, -h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle| \\
&= \langle \mathbf{u}, -h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \\
&= \frac{\sqrt{2\pi}}{m}\bar{X}
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{\sqrt{2\pi}}{m}\left(\sqrt{\frac{2}{\pi}} + t\right)\ell \\
&= \frac{2\ell}{m} + \frac{\sqrt{2\pi}\ell t}{m}.
\end{aligned}
$$

This inequality holds for any *single* choice of $J' \subseteq [n]$, $|J'| \le 2k$, and $\mathbf{u} \in \mathcal{U}_{J'}$ with probability at least

$$1 - e^{-\frac{1}{2}\ell t^2} = 1 - \frac{\frac{\rho}{3\tau m}}{2 \cdot 2^\ell \binom{m}{\ell}\binom{n}{2k}},$$

and by a union bound, the above inequality holds uniformly for *every* $J' \subseteq [n]$, $|J'| \le 2k$, $\mathbf{u} \in \mathcal{U}_{J'}$, and $\ell \in [\tau m]$ with probability at least

$$
\begin{aligned}
1 - \sum_{\ell=1}^{\tau m} 2 \cdot 2^\ell \binom{m}{\ell}\binom{n}{2k} e^{-\frac{1}{2}\ell t^2} &= 1 - \sum_{\ell=1}^{\tau m} 2 \cdot 2^\ell \binom{m}{\ell}\binom{n}{2k} \frac{\frac{\rho}{3\tau m}}{2 \cdot 2^\ell \binom{m}{\ell}\binom{n}{2k}} \\
&= 1 - \sum_{\ell=1}^{\tau m} \frac{\rho}{3\tau m} \\
&= 1 - \tau m \cdot \frac{\rho}{3\tau m} \\
&= 1 - \frac{\rho}{3}.
\end{aligned}
$$

Thus, Lemma A.3 is proved. ∎

A.4 Bounding $D'_{2;J'}(\mathbf{u},\mathbf{u}) \triangleq \|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}\|_2$ Next, the second random variable in the orthogonal decomposition, $D'_{2;J'}(\mathbf{u},\mathbf{u}) \triangleq \|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}\|_2$, is upper bounded in Lemma A.4, laid out below.

LEMMA A.4. *Fix $\rho \in (0,1]$. Suppose*

$$m \ge \frac{b}{\delta}\log\left(\binom{n}{k}^2 \binom{n}{2k}\left(\frac{12b}{\delta}\right)^{2k}\left(\frac{3a}{\rho}\right)\right) = O\left(\frac{k}{\delta}\log\left(\frac{n}{\delta k}\right) + \frac{1}{\delta}\log\left(\frac{1}{\rho}\right)\right).$$

Then, with probability at least $1 - \frac{\rho}{3}$, uniformly for all $J' \subseteq [n]$, $|J'| \le 2k$, and $\mathbf{u} \in \mathcal{U}_{J'}$,

$$D'_{2;J'}(\mathbf{u},\mathbf{u}) \le \frac{\sqrt{2\pi(3k-1)\ell}}{m} + \frac{\sqrt{2\pi}\ell t}{m} \tag{A.10}$$

where

$$\ell \le \tau m, \tag{A.11}$$

$$t = \sqrt{\frac{2}{\ell}\log\left(2 \cdot 2^\ell \binom{m}{\ell}\binom{n}{2k}\frac{3\tau m}{\rho}\right)}. \tag{A.12}$$

Proof. (Lemma A.4). Define the random variables X_i, $i \in [\ell]$, and $\bar{X} = \sum_{i=1}^{\ell} X_i$ as in the proof of Lemma A.3. As before, write $I \subseteq [m]$, $I \triangleq \{i \in [m] : f(\mathbf{u})_i \neq \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle)\}$, and let $\ell \triangleq |I|$, where $\ell = |I| \le \tau m$. Without loss of generality, once again take $I = [\ell]$. As in the proof of Lemma A.3, we have:

$$
\begin{aligned}
h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) &= -\frac{\sqrt{2\pi}}{m}\sum_{i \in I} T_{\mathsf{supp}(\mathbf{u}) \cup J'}(\mathbf{A}_i\,\mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle)) \\
&= -\frac{\sqrt{2\pi}}{m}\sum_{i=1}^{\ell} T_{\mathsf{supp}(\mathbf{u}) \cup J'}(\mathbf{A}_i\,\mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle)) \qquad \text{(A.13)}
\end{aligned}
$$

where the second equality uses the assumption that $I = [\ell]$. Recall that due to the linearity of inner products,

$$\langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle = \left\langle \mathbf{u}, -\frac{\sqrt{2\pi}}{m}\sum_{i=1}^{\ell} T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i\ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle))\right\rangle$$

$$= -\frac{\sqrt{2\pi}}{m}\sum_{i=1}^{\ell}\Big\langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i\ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle))\Big\rangle. \tag{A.14}$$

Thus, the difference of Equations (A.13) and (A.14) is given by:

$$h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}$$

$$= -\frac{\sqrt{2\pi}}{m}\sum_{i=1}^{\ell} T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i\ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle)) + \frac{\sqrt{2\pi}}{m}\sum_{i=1}^{\ell}\Big\langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i\ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle))\Big\rangle \mathbf{u}$$

▶ by expanding the terms via Equations (A.13) and (A.14)

$$= -\frac{\sqrt{2\pi}}{m}\sum_{i=1}^{\ell}\Big(T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i\ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle)) - \big\langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i\ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle))\big\rangle \mathbf{u}\Big) \tag{A.15}$$

▶ by combining the summations and factoring out the $\frac{\sqrt{2\pi}}{m}$ term.

Taking the norm of the above expression, Equation (A.15), yields the following:

$$\|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u}) - \langle \mathbf{u}, h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\rangle \mathbf{u}\|_2$$

$$= \left\| -\frac{\sqrt{2\pi}}{m}\sum_{i=1}^{\ell}\Big(T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i\ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle)) - \big\langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i\ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle))\big\rangle \mathbf{u}\Big)\right\|_2$$

▶ by Equation (A.15)

$$= \frac{\sqrt{2\pi}}{m}\left\| \sum_{i=1}^{\ell}\Big(T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i\ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle)) - \big\langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{A}_i\ \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}_i\rangle))\big\rangle \mathbf{u}\Big)\right\|_2$$

▶ due to the homogeneity of norms

$$\leq \frac{\sqrt{2\pi}}{m}\left(\sqrt{\frac{(3k-1)\ell}{2}} + \ell t\right)$$

▶ by Lemma A.2, setting $k' = |\mathsf{supp}(\mathbf{u}) \cup J'| \leq 3k$

$$= \frac{\sqrt{\pi(3k-1)\ell}}{m} + \frac{\sqrt{2\pi}\ell t}{m}$$

▶ by distributing the $\frac{\sqrt{2\pi}}{m}$ term

where the second to last expression (the inequality) holds with probability at least $1 - e^{-\frac{1}{2}\ell t^2} \geq 1 - \frac{\rho}{3}$. ∎

B Controlling $D_{2;J'}(\mathbf{u},\mathbf{u}) \triangleq \|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\|_2$

The analysis in Appendices A.3 and A.4 makes possible a straightforward derivation of the upper bound on $D_{2;J'}(\mathbf{u},\mathbf{u}) \triangleq \|h_{f;\mathbf{A};J'}(\mathbf{u},\mathbf{u})\|_2$ as stated in Lemma 5.2, which is proved next.

Proof. (Lemma 5.2). Due to Lemmas A.3 and A.4 and by a union bound over their results, the following inequalities hold simultaneously for all $J' \subseteq [n]$, $|J'| \leq 2k$, and $\mathbf{u} \in \mathcal{U}_{J'}$ with probability at least $1 - \frac{2\rho}{3}$ when $m \geq \frac{b}{\delta}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12b}{\delta}\right)^{2k}\left(\frac{3a}{\rho}\right)\right)$:

$$D'_{1;J'}(\mathbf{u},\mathbf{u}) \leq \frac{2\ell}{m} + \frac{\sqrt{2\pi}\ell t}{m},$$

$$D'_{2;J'}(\mathbf{u},\mathbf{u}) \le \frac{\sqrt{2\pi}\ell t}{m} + \frac{\sqrt{\pi(3k-1)\ell}}{m},$$

where

$$\ell \le \tau m,$$
$$t = \sqrt{\frac{2}{\ell}\log\left(2\cdot 2^{\ell}\binom{m}{\ell}\binom{n}{2k}\frac{3\tau m}{\rho}\right)}.$$

Observe:

$$\begin{aligned}
\ell t &= \ell\sqrt{\frac{2}{\ell}\log\left(2\cdot 2^{\ell}\binom{m}{\ell}\binom{n}{2k}\frac{3\tau m}{\rho}\right)}\\
&= \sqrt{2\ell\log\left(2\cdot 2^{\ell}\binom{m}{\ell}\binom{n}{2k}\frac{3\tau m}{\rho}\right)}\\
&\le \sqrt{2\tau m\log\left(2\cdot 2^{\tau m}\binom{m}{\tau m}\binom{n}{2k}\frac{3\tau m}{\rho}\right)}\\
&\qquad \blacktriangleright\ \because \ell \le \tau m\\
&= \sqrt{2\tau m\log\left(2^{\tau m}\binom{m}{\tau m}\right) + 2\tau m\log\binom{n}{2k} + 2\tau m\log\left(\frac{3\tau m}{\rho}\right) + 2\tau m\log(2)}\\
&\le \sqrt{2(\tau m)^2\log\left(\frac{2e}{\tau}\right) + 2\tau m\log\binom{n}{2k} + 2\tau m\log\left(\frac{3\tau m}{\rho}\right) + 2\tau m\log(2)}\\
&\qquad \blacktriangleright\ \because 2^{\tau m}\binom{m}{\tau m} \le 2^{\tau m}\left(\frac{em}{\tau m}\right)^{\tau m} = 2^{\tau m}\left(\frac{e}{\tau}\right)^{\tau m} = \left(\frac{2e}{\tau}\right)^{\tau m}\\
&\le \sqrt{2(\tau m)^2\log\left(\frac{2e}{\tau}\right)} + \sqrt{2\tau m\log\binom{n}{2k}} + \sqrt{2\tau m\log\left(\frac{3\tau m}{\rho}\right)} + \sqrt{2\tau m\log(2)}\\
&\qquad \blacktriangleright\ \text{by the triangle inequality}\\
&= \sqrt{2(\tau m)^2\log\left(\frac{2e}{\tau}\right)} + \sqrt{\frac{2\tau m^2\log\binom{n}{2k}}{m}} + \sqrt{\frac{2\tau m^2\log(\frac{3\tau m}{\rho})}{m}} + \sqrt{\frac{2\tau m^2\log(2)}{m}}\\
&\qquad \blacktriangleright\ \text{by multiplying each of the last three terms by } \sqrt{\frac{m}{m}}\\
&= \tau m\sqrt{2\log\left(\frac{2e}{\tau}\right)} + m\sqrt{\frac{2\tau\log\binom{n}{2k}}{m}} + m\sqrt{\frac{2\tau\log(\frac{3\tau m}{\rho})}{m}} + m\sqrt{\frac{2\tau\log(2)}{m}}\\
&= \tau m\sqrt{2\log\left(\frac{2e}{\tau}\right)} + m\sqrt{\frac{2\delta\tau\log\binom{n}{2k}}{\delta m}} + m\sqrt{\frac{2\delta\tau\log(\frac{3\tau m}{\rho})}{\delta m}} + m\sqrt{\frac{2\delta\tau\log(2)}{\delta m}}\\
&\qquad \blacktriangleright\ \text{by multiplying each of the last three terms by } \sqrt{\frac{\delta}{\delta}}\\
&\le \tau m\sqrt{2\log\left(\frac{2e}{\tau}\right)} + m\sqrt{\frac{2\delta\tau}{b}} + m\sqrt{\frac{2\delta\tau}{b}} + m\sqrt{\frac{2\delta\tau}{b}}\\
&\qquad \blacktriangleright\ \because \delta m \ge b\max\left\{\log\binom{n}{2k}, \log\left(\frac{3\tau m}{\rho}\right), \log(2)\right\},\\
&\qquad \text{where } b > 0 \text{ is a universal constant specified in Equation (2.1)}
\end{aligned}$$

$$= \tau m \sqrt{2 \log\left(\frac{2e}{\tau}\right)} + 3m \sqrt{\frac{2\delta\tau}{b}}$$
$$= m \left(\tau \sqrt{2 \log\left(\frac{2e}{\tau}\right)} + 3 \sqrt{\frac{2\delta\tau}{b}} \right).$$

Then, dividing the above expressions by m, it follows that

$$\frac{\ell t}{m} \leq \tau \sqrt{2 \log\left(\frac{2e}{\tau}\right)} + 3 \sqrt{\frac{2\delta\tau}{b}}. \tag{B.16}$$

Additionally, note that

$$\frac{\ell}{m} \leq \frac{\tau m}{m} = \tau, \tag{B.17}$$

$$\frac{\sqrt{\pi(3k-1)\ell}}{m} \leq \frac{\sqrt{\pi(3k-1)\tau m}}{m} = \sqrt{\frac{\pi(3k-1)\tau}{m}} \leq \sqrt{\frac{\pi(3k-1)\delta\tau}{bk}} \leq \sqrt{\frac{3\pi\delta\tau}{b}}. \tag{B.18}$$

Combining the above results yields the following upper bound:

$$\begin{aligned}
\frac{2\ell}{m} + \frac{2\sqrt{2\pi}\ell t}{m} + \frac{\sqrt{\pi(3k-1)\ell}}{m} &\leq 2\tau + 2\sqrt{2\pi} \cdot \tau \sqrt{2 \log\left(\frac{2e}{\tau}\right)} + 2\sqrt{2\pi} \cdot 3 \sqrt{\frac{2\delta\tau}{b}} + \sqrt{\frac{3\pi\delta\tau}{b}} \\
&\qquad \blacktriangleright \text{ due to Equations (B.16)-(B.18)} \\
&= 2\tau + 4\sqrt{\pi} \cdot \tau \sqrt{\log\left(\frac{2e}{\tau}\right)} + 12 \sqrt{\frac{\pi\delta\tau}{b}} + \sqrt{3} \cdot \sqrt{\frac{\pi\delta\tau}{b}} \\
&= 2\tau + 4\sqrt{\pi} \cdot \tau \sqrt{\log\left(\frac{2e}{\tau}\right)} + 13 \sqrt{\frac{\pi\delta\tau}{b}} \\
&\leq 2\tau \sqrt{\log\left(\frac{2e}{\tau}\right)} + 4\sqrt{\pi} \cdot \tau \sqrt{\log\left(\frac{2e}{\tau}\right)} + (12 + \sqrt{3}) \sqrt{\frac{\pi\delta\tau}{b}} \\
&\qquad \blacktriangleright \because \tau \leq \tau \sqrt{\log\left(\frac{2e}{\tau}\right)} \text{ for } \tau \in (0, 1] \\
&= \left(2 + 4\sqrt{\pi}\right) \tau \sqrt{\log\left(\frac{2e}{\tau}\right)} + (12 + \sqrt{3}) \sqrt{\frac{\pi\delta\tau}{b}} \\
&= (12 + \sqrt{3}) \sqrt{\frac{\pi\delta\tau}{b}} + \left(2 + 4\sqrt{\pi}\right) \tau \sqrt{\log\left(\frac{2e}{\tau}\right)}. \qquad \text{(B.19)}
\end{aligned}$$

Therefore, by Claim A.1 and an earlier remark, with probability at least $1 - \frac{2\rho}{3}$, uniformly for all $J' \subseteq [n]$, $|J'| \leq k$, and $\mathbf{u} \in \mathcal{U}_{J'}$, $D_{2;J'}(\mathbf{u}, \mathbf{u})$ is bounded from above as follows:

$$\begin{aligned}
D_{2;J'}(\mathbf{u}, \mathbf{u}) &\leq D'_{1;J'}(\mathbf{u}, \mathbf{u}) + D'_{2;J'}(\mathbf{u}, \mathbf{u}) \\
&\qquad \blacktriangleright \text{ by Claim A.1} \\
&\leq \left(\frac{2\ell}{m} + \frac{\sqrt{2\pi}\ell t}{m} \right) + \left(\frac{\sqrt{2\pi}\ell t}{m} + \frac{\sqrt{\pi(3k-1)\ell}}{m} \right) \\
&\qquad \blacktriangleright \text{ by Lemmas A.3 and A.4} \\
&= \frac{2\ell}{m} + \frac{2\sqrt{2\pi}\ell t}{m} + \frac{\sqrt{\pi(3k-1)\ell}}{m}
\end{aligned}$$

$$\leq (12+\sqrt{3})\sqrt{\frac{\pi\delta\tau}{b}} + \big(2+4\sqrt{\pi}\big)\tau\sqrt{\log\left(\frac{2e}{\tau}\right)}$$

▶ by Equation (B.19)

$$= c_3\sqrt{\delta\tau} + c_4\tau\sqrt{\log\left(\frac{2e}{\tau}\right)}$$

▶ due to an appropriate choice of the universal constants, $c_3, c_4 > 0$, as defined in Equation (2.1)

which completes the proof. ∎

C Proofs of the Concentration Inequalities – Lemmas A.1 and A.2

Before tackling Lemmas A.1 and A.2, the following four intermediate results—Lemmas C.1-C.4—are stated and derived to facilitate their proofs.

Lemma C.1. *Fix a k-sparse, real-valued unit vector,* $\mathbf{u} \in S^{n-1} \cap \Sigma_k^n$*, and let* $J'' \subseteq [n]$*. Let* $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ *be a Gaussian vector with i.i.d. entries. Define the random variable* $X_\mathbf{u}$ *by*

$$X_\mathbf{u} = \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J''}(\mathbf{Z}\,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle))\rangle.$$

Then, $X_\mathbf{u} = |\langle \mathbf{u}, \mathbf{Z}\rangle| = |X_\mathbf{u}|$*, and* $\mathbb{E}[X_\mathbf{u}] = \sqrt{\frac{2}{\pi}}$*.*

Lemma C.2. *Fix* $\mathbf{u}, \mathbf{v} \in S^{n-1} \cap \Sigma_k^n$ *and* $J'' \subseteq [n]$*, where* $\mathsf{supp}(\mathbf{v}) \subseteq \mathsf{supp}(\mathbf{u}) \cup J''$*. Let*

$$Y_{\mathbf{u},\mathbf{v}} = \big\langle \mathbf{v}, T_{\mathsf{supp}(\mathbf{u})\cup J''}(\mathbf{Z}\,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle)) - \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J''}(\mathbf{Z}\,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle))\rangle \mathbf{u}\big\rangle.$$

Then, $Y_{\mathbf{u},\mathbf{v}} \sim \mathcal{N}(0,1)$*.*

Lemma C.3. *Fix* $\ell \in \mathbb{Z}_+$*,* $\mathbf{u} \in S^{n-1} \cap \Sigma_k^n$*, and* $J'' \subseteq [n]$*, and write* $k' \triangleq |\mathsf{supp}(\mathbf{u}) \cup J''|$*. Let*

$$\bar{\mathbf{Y}}_\mathbf{u} = \sum_{i=1}^{\ell} \Big(T_{\mathsf{supp}(\mathbf{u})\cup J''}(\mathbf{Z}\,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle)) - \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J''}(\mathbf{Z}\,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle))\rangle \mathbf{u} \Big)$$

and let $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \ell\mathbf{I}_{k'-1})$*. Then,* $\|\mathbf{Y}_\mathbf{u}\|_2 \sim \|\mathbf{W}\|_2$*.*

Lemma C.4. *Fix* $d \in \mathbb{Z}_+$*, and let* $\mathbf{W} \sim \mathcal{N}(0, \sigma^2\mathbf{I}_d)$ *be a Gaussian random vector with i.i.d. entries. Then,*

$$\Pr\left(\|\mathbf{W}\|_2 > \sigma\sqrt{\frac{d}{2}} + \sigma^2 t \right) \leq e^{-\frac{1}{2}\sigma^2 t^2}. \tag{C.20}$$

Proof. (Lemma C.1). Fix $\ell \in \mathbb{Z}_+$, $\mathbf{u} \in S^{n-1} \cap \Sigma_k^n$, and $J'' \subseteq [n]$ arbitrarily, and taking $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, let $X_\mathbf{u} = \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J''}(\mathbf{Z}\,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle))\rangle$. To start off, the relation, $\mathsf{supp}(\mathbf{u}) \subseteq \mathsf{supp}(\mathbf{u}) \cup J''$, motivates the following claim.

Claim C.1. *Let* $\mathbf{u}, \mathbf{v} \in S^{n-1} \cap \Sigma_k^n$ *(not necessarily distinct) and* $J'' \subseteq [n]$*, such that* $\mathsf{supp}(\mathbf{v}) \subseteq \mathsf{supp}(\mathbf{u}) \cup J''$*. Then, for any* $\mathbf{z} \in \mathbb{R}^n$ *and* $J'' \subseteq [n]$*, there is an equality:*

$$\langle \mathbf{v}, \mathbf{z}\rangle = \langle \mathbf{v}, T_{\mathsf{supp}(\mathbf{u})\cup J''}(\mathbf{z})\rangle = \langle T_{\mathsf{supp}(\mathbf{u})\cup J''}(\mathbf{v}), T_{\mathsf{supp}(\mathbf{u})\cup J''}(\mathbf{z})\rangle.$$

Proof. (Claim C.1). The first equality in this claim is simple to verify:

$$\langle \mathbf{v}, T_{\mathsf{supp}(\mathbf{u})\cup J''}(\mathbf{z})\rangle$$

$$= \sum_{j=1}^{n} v_j T_{\mathsf{supp}(\mathbf{u})\cup J''}(\mathbf{z})_j$$

$$
\begin{aligned}
&= \sum_{j \in \mathsf{supp}(\mathbf{v}) \cap \mathsf{supp}(T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z}))} v_j T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z})_j + \sum_{j \in [n] \setminus (\mathsf{supp}(\mathbf{v}) \cap \mathsf{supp}(T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z})))} v_j T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z})_j \\
&= \sum_{j \in \mathsf{supp}(\mathbf{v}) \cap (\mathsf{supp}(\mathbf{u}) \cup J'')} v_j T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z})_j + \sum_{j \in [n] \setminus (\mathsf{supp}(\mathbf{v}) \cap (\mathsf{supp}(\mathbf{u}) \cup J''))} v_j T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z})_j \\
&= \sum_{j \in \mathsf{supp}(\mathbf{v})} v_j T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z})_j + \sum_{j \in [n] \setminus \mathsf{supp}(\mathbf{v})} v_j T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z})_j \\
&= \sum_{j \in \mathsf{supp}(\mathbf{v})} v_j T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z})_j + \sum_{j \in [n] \setminus \mathsf{supp}(\mathbf{v})} 0 \cdot T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z})_j \\
&= \sum_{j \in \mathsf{supp}(\mathbf{v})} v_j z_j + \sum_{j \in [n] \setminus \mathsf{supp}(\mathbf{v})} 0 \cdot T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z})_j \\
&= \sum_{j \in \mathsf{supp}(\mathbf{v})} v_j z_j + \sum_{j \in [n] \setminus \mathsf{supp}(\mathbf{v})} 0 \cdot z_j \\
&= \sum_{j \in \mathsf{supp}(\mathbf{v})} v_j z_j + \sum_{j \in [n] \setminus \mathsf{supp}(\mathbf{v})} v_j z_j \\
&= \sum_{j=1}^{n} v_j z_j \\
&= \langle \mathbf{v}, \mathbf{z} \rangle.
\end{aligned}
$$

Thus, the claim's first equality holds: $\langle \mathbf{v}, \mathbf{z} \rangle = \langle \mathbf{v}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z}) \rangle$. For the second equality of the claim, note that $\mathsf{supp}(\mathbf{v}) \subseteq \mathsf{supp}(\mathbf{u}) \cup J''$ implies that $T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{v}) = \mathbf{v}$, and therefore, $\langle \mathbf{v}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z}) \rangle = \langle T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{v}), T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{z}) \rangle$, ∎

Thus, the random variable $X_{\mathbf{u}}$ is equivalently given by

$$
\begin{aligned}
X_{\mathbf{u}} &= \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z} \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)) \rangle \\
&= \langle \mathbf{u}, \mathbf{Z} \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) \rangle \\
&= \langle \mathbf{u}, \mathbf{Z} \rangle \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)
\end{aligned}
$$

Note that for any $a \in \mathbb{R}$,

$$a \,\mathsf{sign}(a) = |a|.$$

Therefore, $X_{\mathbf{u}} = \langle \mathbf{u}, \mathbf{Z} \rangle \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) = |\langle \mathbf{u}, \mathbf{Z} \rangle| = |X_{\mathbf{u}}|$, as claimed. By a well-known property of Gaussians, $\langle \mathbf{u}, \mathbf{Z} \rangle \sim \mathcal{N}(0,1)$, and hence, $X_{\mathbf{u}} = |\langle \mathbf{u}, \mathbf{Z} \rangle| \sim |U|$, where $U \sim \mathcal{N}(0,1)$ is a half-normal random variable. Since $X_{\mathbf{u}} \sim |U|$, these two random variables are equal in expectation: $\mathbb{E}[X_{\mathbf{u}}] = \mathbb{E}\big[|U|\big] = \sqrt{\frac{2}{\pi}}$. ∎

Proof. (Lemma C.2). Fix a pair of orthonormal vectors, $\mathbf{u}, \mathbf{v} \in S^{n-1} \cap \Sigma_k^n$, arbitrarily, and let

$$Y_{\mathbf{u},\mathbf{v}} = \Big\langle \mathbf{v}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z} \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)) - \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z} \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)) \rangle \mathbf{u} \Big\rangle$$

Observe:

$$
\begin{aligned}
Y_{\mathbf{u},\mathbf{v}} &= \Big\langle \mathbf{v}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z} \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)) - \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z} \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)) \rangle \mathbf{u} \Big\rangle \\
&= \Big\langle \mathbf{v}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z} \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)) \Big\rangle - \Big\langle \mathbf{v}, \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z} \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)) \rangle \mathbf{u} \Big\rangle \\
&\qquad \blacktriangleright \text{ by the linearity of inner products} \\
&= \Big\langle \mathbf{v}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z} \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)) \Big\rangle - \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z} \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)) \rangle \langle \mathbf{v}, \mathbf{u} \rangle \\
&\qquad \blacktriangleright \text{ by the linearity of inner products}
\end{aligned}
$$

$$= \left\langle \mathbf{v}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z} \, \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)) \right\rangle - 0$$

▶ due to the orthogonality of $\mathbf{u}$ and $\mathbf{v}$

$$= \left\langle \mathbf{v}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z} \, \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)) \right\rangle$$

$$= \left\langle \mathbf{v}, \mathbf{Z} \, \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) \right\rangle$$

▶ due to Claim C.1, wherein $\mathbf{z} = \mathbf{Z} \, \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)$, and since $\mathsf{supp}(\mathbf{v}) \subseteq \mathsf{supp}(\mathbf{u}) \cup J''$

$$= \langle \mathbf{v}, \mathbf{Z} \rangle \, \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)$$

▶ by the linearity of inner products

The remaining step is to show that $\langle \mathbf{v}, \mathbf{Z} \rangle \, \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) \sim \mathcal{N}(0,1)$. This can be achieved by a two-step argument: (a) First, we will argue that $\langle \mathbf{v}, \mathbf{Z} \rangle$ and $\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)$ are independent. (b) Then, by standard facts about Gaussians and due to the independence shown in Step (a), the claim will follow. Starting with Step (a), note that if $\langle \mathbf{u}, \mathbf{Z} \rangle$ and $\langle \mathbf{v}, \mathbf{Z} \rangle$ are independent, so are $\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)$ and $\langle \mathbf{v}, \mathbf{Z} \rangle$. Therefore, it suffices to establish the independence of $\langle \mathbf{u}, \mathbf{Z} \rangle$ and $\langle \mathbf{v}, \mathbf{Z} \rangle$. Write $U_1 = \langle \mathbf{u}, \mathbf{Z} \rangle$ and $U_2 = \langle \mathbf{v}, \mathbf{Z} \rangle$. By a well-known fact about Gaussians, $U_1, U_2 \sim \mathcal{N}(0,1)$. Now, consider the joint distribution of

$$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma} \right)$$

which is a 0-mean bivariate Gaussian with covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{2 \times 2}$. The goal is to show that $\boldsymbol{\Sigma} = \mathbf{I}_2$. Each i^{th} diagonal entry, $i \in \{1,2\}$, is given by:

$$\begin{aligned} \Sigma_{i,i} &= \mathsf{Cov}(U_i, U_i) \\ &= \mathsf{Var}(U_i) \\ &= 1 \end{aligned}$$

where the last line follows from the earlier observation that $U_1, U_2 \sim \mathcal{N}(0,1)$. On the other hand, each off-diagonal (i,j)-entry, $i, j \in \{1,2\}$, $i \neq j$, is obtained as follows. Assuming without loss of generality that $i = 1$ and $j = 2$, the corresponding covariance is 0-valued due to the next derivation:

$$\begin{aligned} \Sigma_{j,i} = \Sigma_{i,j} &= \mathsf{Cov}(U_i, U_j) \\ &= \mathbb{E}[U_i U_j] - \mathbb{E}[U_i] \, \mathbb{E}[U_j] \\ &= \mathbb{E}[U_i U_j] \\ &= \mathbb{E}[U_1 U_2] \\ &= \mathbb{E}[\mathbf{u}^{\mathsf{T}} \mathbf{Z} \mathbf{Z}^{\mathsf{T}} \mathbf{v}] \\ &= \mathbf{u}^{\mathsf{T}} \, \mathbb{E}[\mathbf{Z} \mathbf{Z}^{\mathsf{T}}] \mathbf{v} \\ &= \mathbf{u}^{\mathsf{T}} \, \mathsf{Cov}(\mathbf{Z}, \mathbf{Z}) \mathbf{v} \\ &= \mathbf{u}^{\mathsf{T}} \mathbf{I}_2 \mathbf{v} = \mathbf{u}^{\mathsf{T}} \mathbf{v} = \cos(\theta_{\mathbf{u},\mathbf{v}}) = \cos\left(\frac{\pi}{2}\right) = 0 \end{aligned}$$

as said. From the above work, it follows that $\boldsymbol{\Sigma} = \mathbf{I}_2$ and $(U_1, U_2) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$. This therefore establishes the desired independence of $U_1 = \langle \mathbf{u}, \mathbf{Z} \rangle$ and $U_2 = \langle \mathbf{v}, \mathbf{Z} \rangle$. The independence of $\langle \mathbf{v}, \mathbf{Z} \rangle$ and $\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle)$ follows, completing Step (a).

Proceeding to Step (b), recall that the goal of this step is to show that $\langle \mathbf{v}, \mathbf{Z} \rangle \, \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) \sim \mathcal{N}(0,1)$. Note that because $\langle \mathbf{u}, \mathbf{Z} \rangle \sim \mathcal{N}(0,1)$, the symmetry of the distribution $\mathcal{N}(0,1)$ around its mean (0) leads to $\Pr(\langle \mathbf{u}, \mathbf{Z} \rangle < 0) = \Pr(\langle \mathbf{u}, \mathbf{Z} \rangle \geq 0) = \frac{1}{2}$, and hence $\Pr(\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) = -1) = \Pr(\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) = 1) = \frac{1}{2}$. This in turn implies that $\langle \mathbf{v}, \mathbf{Z} \rangle \, \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) \sim WS$, where $W \sim \mathcal{N}(0,1)$ and $S \sim \{-1,1\}$ are independent. The density function of W (a univariate Gaussian) is given at $w \in \mathbb{R}$ by $f_W(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}}$, while the mass function of S is given at $s \in \{-1,1\}$ by $f_S(s) = \frac{1}{2}$ and is otherwise 0-valued. Additionally, $f_{-W}(w) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(-w)^2}{2}}$. Due to the independence of W and S, their joint density function is simply the product of their individual densities: $f_{W,S}(w,s) = f_W(w) f_S(s)$. Notice that

$$(WS \mid S = s) = (sW \mid S = s) = sW$$

where the last equality uses the independence discussed above. The density of WS is then given at $z \in \mathbb{R}$ by:

$$\begin{aligned}
f_{WS}(z) &= f_{WS|S=-1}(z|-1) f_S(-1) + f_{WS|S=1}(z|1) f_S(1) \\
&\quad \blacktriangleright \text{ by the law of total probability} \\
&= \frac{1}{2} f_{WS|S=-1}(z|-1) + \frac{1}{2} f_{WS|S=1}(z|1) \\
&\quad \blacktriangleright \text{ by the definition of } f_S \\
&= \frac{1}{2} f_{-W}(z) + \frac{1}{2} f_W(z) \\
&\quad \blacktriangleright \text{ by an earlier remark} \\
&= \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(-z)^2}{2}} + \frac{1}{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \\
&\quad \blacktriangleright \text{ by the definitions of } f_{-W}, f_W \\
&= \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} + \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \\
&\quad \blacktriangleright \text{ by squaring the negative term in the first exponent} \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \\
&\quad \blacktriangleright \text{ by simplification}
\end{aligned}$$

In short, $f_{WS}(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$, which is precisely the density function of a standard normal random variable, $\mathcal{N}(0,1)$. Therefore, $\langle \mathbf{v}, \mathbf{Z} \rangle \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) \sim WS \sim \mathcal{N}(0,1)$. This completes Step (b). Moreover, combined with an earlier argument, it follows that $Y_{\mathbf{u},\mathbf{v}} = \langle \mathbf{v}, \mathbf{Z} \rangle \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) \sim \mathcal{N}(0,1)$, as the lemma claimed. ◻

Proof. (Lemma C.3). Fix $\mathbf{u} \in S^{n-1} \cap \Sigma_k^n$ and $J'' \subseteq [n]$ arbitrarily, where $k' = |\mathsf{supp}(\mathbf{u}) \cup J''|$, and let $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \ell \mathbf{I}_{k'-1})$. Define

$$\bar{\mathbf{Y}}_{\mathbf{u}} = \sum_{i=1}^{\ell} \Big(T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)) - \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)) \rangle \mathbf{u} \Big)$$

and note that since $\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \in \{-1, 1\}$ is a scalar,

$$\begin{aligned}
\bar{\mathbf{Y}}_{\mathbf{u}} &= \sum_{i=1}^{\ell} \Big(T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)) - \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)) \rangle \mathbf{u} \Big) \\
&= \sum_{i=1}^{\ell} \Big(T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i) \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) - \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i) \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \rangle \mathbf{u} \Big) \\
&= \sum_{i=1}^{\ell} \Big(T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i) \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) - \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i) \rangle \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \mathbf{u} \Big)
\end{aligned}$$

Notice that $T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{u}) = \mathbf{u}$ since $\mathsf{supp}(\mathbf{u}) \subseteq \mathsf{supp}(\mathbf{u}) \cup J''$. Additionally, taking $\mathbf{v} = \mathbf{u} = T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{u})$ and $\mathbf{z} = \mathbf{Z}_i$ in Claim C.1 (see, the proof of Lemma C.1), it follows that

$$\langle \mathbf{u}, \mathbf{Z}_i \rangle = \langle T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{u}), T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)) \rangle,$$

and thus,

$$\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) = \mathsf{sign}\big(\langle T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{u}), T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i) \rangle\big).$$

Combining the above arguments now yields:

$$\begin{aligned}
\bar{\mathbf{Y}}_{\mathbf{u}} &= \sum_{i=1}^{\ell} \Big(T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i) \, \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) - \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i) \rangle \, \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \mathbf{u} \Big) \\
&= \sum_{i=1}^{\ell} \Big(T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i) \, \mathsf{sign}\big(\langle T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{u}), T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i) \rangle\big) \\
&\qquad - \langle T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{u}), T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i) \rangle \, \mathsf{sign}\big(\langle T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{u}), T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i) \rangle\big) \, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{u}) \Big) \\
&= \sum_{i=1}^{\ell} \Big(T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{Z}_i) \, \mathsf{sign}\Big(\big\langle \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \big\rangle \Big) \\
&\qquad - \big\langle \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \big\rangle \, \mathsf{sign}\Big(\big\langle \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \big\rangle \Big) \, T_{\mathsf{supp}(\mathbf{u}) \cup J''}(\mathbf{u}) \Big) \\
&= \sum_{i=1}^{\ell} T_{\mathsf{supp}(\mathbf{u}) \cup J''} \Bigg(\mathbf{Z}_i \, \mathsf{sign}\Big(\big\langle \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \big\rangle \Big) \\
&\qquad - \big\langle \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \big\rangle \, \mathsf{sign}\Big(\big\langle \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \big\rangle \Big) \, \mathbf{u} \Bigg) \\
&= T_{\mathsf{supp}(\mathbf{u}) \cup J''} \Bigg(\sum_{i=1}^{\ell} \mathbf{Z}_i \, \mathsf{sign}\Big(\big\langle \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \big\rangle \Big) \\
&\qquad\qquad - \big\langle \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \big\rangle \, \mathsf{sign}\Big(\big\langle \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \big\rangle \Big) \, \mathbf{u} \Bigg)
\end{aligned}$$

The random variable of interest is $\|\bar{\mathbf{Y}}_{\mathbf{u}}\|_2$, where, as can be seen from the last expression above,

$$\begin{aligned}
\big\|\bar{\mathbf{Y}}_{\mathbf{u}}\big\|_2 &= \sqrt{\sum_{j=1}^{n} \bar{Y}_{\mathbf{u};j}^2} \\
&= \sqrt{\sum_{j \in \mathsf{supp}(\mathbf{u}) \cup J''} \bar{Y}_{\mathbf{u};j}^2} \\
&= \Big\| \bar{\mathbf{Y}}_{\mathbf{u}}|_{\mathsf{supp}(\mathbf{u}) \cup J''} \Big\|_2 \\
&= \Bigg\| \sum_{i=1}^{\ell} \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \, \mathsf{sign}\Big(\big\langle \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \big\rangle \Big) \\
&\qquad - \big\langle \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \big\rangle \, \mathsf{sign}\Big(\big\langle \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \big\rangle \Big) \, \mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''} \Bigg\|_2
\end{aligned}$$

Writing $k' \triangleq |\mathsf{supp}(\mathbf{u}) \cup J''|$, note that $\mathbf{u}|_{\mathsf{supp}(\mathbf{u}) \cup J''}, \mathbf{Z}_i|_{\mathsf{supp}(\mathbf{u}) \cup J''} \in \mathbb{R}^{k'}$ are the restriction of the vectors $\mathbf{u}, \mathbf{Z}_i \in \mathbb{R}^n$, respectively, onto the coordinates indexed by $\mathsf{supp}(\mathbf{u}) \cup J''$. As a result, the last line above does not depend on the choice of n, subject to $n \geq k'$. Hence, in order to simplify notations in this proof, assume without loss of generality that $n = k' = |\mathsf{supp}(\mathbf{u}) \cup J''|$. Then, $\mathbf{u}$, $\mathbf{Z}_i$, $i \in [\ell]$, and $\bar{\mathbf{Y}}_{\mathbf{u}}$ are all k'-dimensional, and the definition of $\bar{\mathbf{Y}}_{\mathbf{u}}$ can be written as

$$\bar{\mathbf{Y}}_{\mathbf{u}} = \sum_{i=1}^{\ell} \Big(\mathbf{Z}_i \, \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) - \langle \mathbf{u}, \mathbf{Z}_i \rangle \, \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \mathbf{u} \Big).$$

Let $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_{k'}\} \subset \mathbb{R}^{k'}$ be an orthonormal basis for $\mathbb{R}^{k'}$, where $\mathbf{v}_{k'} = \mathbf{u}$. Then,

$$\bar{\mathbf{Y}}_{\mathbf{u}} = \sum_{i=1}^{\ell} \Big(\mathbf{Z}_i \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) - \langle \mathbf{u}, \mathbf{Z}_i \rangle \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \mathbf{u} \Big)$$

$$= \sum_{i=1}^{\ell} \Big(\mathbf{Z}_i - \langle \mathbf{u}, \mathbf{Z}_i \rangle \mathbf{u} \Big) \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)$$

▶ by distributivity

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{k'} \Big\langle \mathbf{v}_j, \big(\mathbf{Z}_i - \langle \mathbf{u}, \mathbf{Z}_i \rangle \mathbf{u} \big) \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \Big\rangle \mathbf{v}_j$$

▶ orthogonal decomposition via the basis $\mathcal{V}$

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{k'} \Big\langle \mathbf{v}_j, \mathbf{Z}_i - \langle \mathbf{u}, \mathbf{Z}_i \rangle \mathbf{u} \Big\rangle \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)$$

▶ by the linearity of inner products

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{k'} \Big(\langle \mathbf{v}_j, \mathbf{Z}_i \rangle - \langle \mathbf{v}_j, \langle \mathbf{u}, \mathbf{Z}_i \rangle \mathbf{u} \rangle \Big) \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)$$

▶ by the linearity of inner products

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{k'-1} \Big(\langle \mathbf{v}_j, \mathbf{Z}_i \rangle - \langle \mathbf{v}_j, \langle \mathbf{u}, \mathbf{Z}_i \rangle \mathbf{u} \rangle \Big) \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) + \Big(\langle \mathbf{u}, \mathbf{Z}_i \rangle - \langle \mathbf{u}, \langle \mathbf{u}, \mathbf{Z}_i \rangle \mathbf{u} \rangle \Big) \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)$$

▶ by distributivity

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{k'-1} \Big(\langle \mathbf{v}_j, \mathbf{Z}_i \rangle - \langle \mathbf{v}_j, \langle \mathbf{u}, \mathbf{Z}_i \rangle \mathbf{u} \rangle \Big) \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) + \Big(\langle \mathbf{u}, \mathbf{Z}_i \rangle - \langle \mathbf{u}, \mathbf{u} \rangle \langle \mathbf{u}, \mathbf{Z}_i \rangle \Big) \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)$$

▶ by the linearity of inner products

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{k'-1} \Big(\langle \mathbf{v}_j, \mathbf{Z}_i \rangle - \langle \mathbf{v}_j, \langle \mathbf{u}, \mathbf{Z}_i \rangle \mathbf{u} \rangle \Big) \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) + \Big(\langle \mathbf{u}, a\mathbf{Z}_i \rangle - \langle \mathbf{u}, \mathbf{Z}_i \rangle \Big) \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)$$

▶ $\because \langle \mathbf{u}, \mathbf{u} \rangle = \|\mathbf{u}\|_2^2 = 1$

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{k'-1} \Big(\langle \mathbf{v}_j, \mathbf{Z}_i \rangle - \langle \mathbf{v}_j, \langle \mathbf{u}, \mathbf{Z}_i \rangle \mathbf{u} \rangle \Big) \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) + 0 \cdot \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)$$

▶ $\because \langle \mathbf{u}, \mathbf{Z}_i \rangle - \langle \mathbf{u}, \mathbf{Z}_i \rangle = 0$

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{k'-1} \Big(\langle \mathbf{v}_j, \mathbf{Z}_i \rangle - \langle \mathbf{v}_j, \langle \mathbf{u}, \mathbf{Z}_i \rangle \mathbf{u} \rangle \Big) \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)$$

▶ via simplification

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{k'-1} \Big(\langle \mathbf{v}_j, \mathbf{Z}_i \rangle - \langle \mathbf{v}_j, \mathbf{u} \rangle \langle \mathbf{u}, \mathbf{Z}_i \rangle \Big) \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)$$

▶ by the linearity of inner products

$$= \sum_{i=1}^{\ell} \sum_{j=1}^{k'-1} \langle \mathbf{v}_j, \mathbf{Z}_i \rangle \mathbf{v}_j \,\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)$$

▶ $\because \mathbf{v}_j \perp \mathbf{v}_{k'} = \mathbf{u}$ when $j \neq k'$

$$= \sum_{j=1}^{k'-1} \sum_{i=1}^{\ell} \langle \mathbf{v}_j, \mathbf{Z}_i \rangle \mathbf{v}_j \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)$$

▶ the summations can be reordered since they do not have dependencies

$$= \sum_{j=1}^{k'-1} \sum_{i=1}^{\ell} \langle \mathbf{v}_j, \mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \rangle \mathbf{v}_j$$

▶ by the linearity of inner products

$$= \sum_{j=1}^{k'-1} \mathbf{v}_j \left\langle \mathbf{v}_j, \sum_{i=1}^{\ell} \mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \right\rangle$$

▶ by the linearity of inner products

Let $S_1, \ldots, S_\ell \sim \{-1, 1\}$ be i.i.d. Rademacher random variables which are also independent of $\mathbf{Z}_1, \ldots, \mathbf{Z}_\ell$. Due to an argument in the proof of Lemma C.2, $\mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \sim \mathbf{Z}_i S_i \sim \mathbf{Z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{k'})$, and additionally, the random vectors, $\{\mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle)\}_{i \in [\ell]}$, are mutually independent. Hence,

$$\sum_{i=1}^{\ell} \mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \sim \mathcal{N}(\mathbf{0}, \ell \mathbf{I}_{k'-1}).$$

By an argument analogous to that which appeared in the proof of Lemma C.2, the random variables, $\{\langle \mathbf{v}_j, \sum_{i=1}^{\ell} \mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \rangle\}_{j \in [k'-1]}$, are mutually independent, and therefore,

$$\left\langle \mathbf{v}_j, \sum_{i=1}^{\ell} \mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \right\rangle \sim W_j \sim \mathcal{N}(0, \sigma^2 = \ell),$$

where the random variables, $\{W_j\}_{j \in [k'-1]}$, are likewise mutually independent. Using these random variables, the lemma's result is obtained as follows:

$$\begin{aligned} \left\| \bar{\mathbf{Y}}_{\mathbf{u}} \right\|_2 &= \left\| \sum_{j=1}^{k'-1} \mathbf{v}_j \left\langle \mathbf{v}_j, \sum_{i=1}^{\ell} \mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \right\rangle \right\|_2 \\ &= \sqrt{\sum_{j=1}^{k'-1} \left\langle \mathbf{v}_j, \sum_{i=1}^{\ell} \mathbf{Z}_i \ \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}_i \rangle) \right\rangle^2} \\ &\sim \sqrt{\sum_{j=1}^{k'-1} W_j^2} \\ &= \|\mathbf{W}\|_2 \end{aligned}$$

To summarize, we have now shown that $\|\bar{\mathbf{Y}}_{\mathbf{u}}\|_2 \sim \|\mathbf{W}\|_2$, where $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \ell \mathbf{I}_{k'-1})$, thus completing the proof of Lemma C.3. ∎

Proof. (Lemma C.4). Let $\mathbf{U} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_d)$ and $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_d)$. Note that $\mathbf{W} \sim \sigma \mathbf{U}$ and $\|\mathbf{W}\|_2 \sim \sigma \|\mathbf{U}\|_2$, and hence,

$$\mathbb{E}\big[\|\mathbf{W}\|_2\big] = \mathbb{E}\big[\|\sigma \mathbf{U}\|_2\big] = \mathbb{E}\big[\sigma \|\mathbf{U}\|_2\big] = \sigma \, \mathbb{E}\big[\|\mathbf{U}\|_2\big].$$

It is well-known that $\|\mathbf{U}\|_2 \sim \chi_d$, and therefore,

$$\mathbb{E}\big[\|\mathbf{U}\|_2\big] = \frac{\Gamma\left(\frac{d+1}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \leq \sqrt{\frac{d}{2}},$$

where the inequality on the right-hand-side can be derived from the Legendre duplication formula and Stirling's approximation. Plugging this into the expression for the expectation of $\|\mathbf{W}\|_2$ yields

$$\mathbb{E}\big[\|\mathbf{W}\|_2\big] = \sigma\,\mathbb{E}\big[\|\mathbf{U}\|_2\big] \leq \sigma\sqrt{\frac{d}{2}} = \sqrt{\frac{\sigma^2 d}{2}}.$$

By a standard concentration inequality for L-Lipschitz functions on Gaussian vectors, where here, $\|\cdot\|_2$ is $(L=1)$-Lipschitz (see, e.g., [35]),

$$\Pr\left(\|\mathbf{U}\|_2 > \sqrt{\frac{d}{2}} + t'\right) \leq \Pr\big(\|\mathbf{U}\|_2 > \mathbb{E}\big[\|\mathbf{U}\|_2\big] + t'\big) \leq e^{-\frac{t'^2}{2L^2}} = e^{-\frac{t'^2}{2}}.$$

Setting $t' = \sigma t$,

$$\Pr\left(\|\mathbf{U}\|_2 > \sqrt{\frac{d}{2}} + \sigma t\right) \leq e^{-\frac{\sigma^2 t^2}{2}}.$$

Finally, by the earlier observation that$\mathbf{W} \sim \sigma\mathbf{U}$ and $\|\mathbf{W}\|_2 \sim \sigma\|\mathbf{U}\|_2$ the lemma's concentration inequality follows:

$$\Pr\left(\|\mathbf{W}\|_2 > \sigma\sqrt{\frac{d}{2}} + \sigma^2 t\right) = \Pr\left(\sigma\|\mathbf{U}\|_2 > \sigma\sqrt{\frac{d}{2}} + \sigma^2 t\right) = \Pr\left(\|\mathbf{U}\|_2 > \sqrt{\frac{d}{2}} + \sigma t\right) \leq e^{-\frac{\sigma^2 t^2}{2}}.$$

□

Having proved Lemmas C.1-C.4, we are ready to apply their concentration inequalities in order to establish Lemmas A.1 and A.2. Let us begin with the former, Lemma A.1.

Proof. (Lemma A.1). Fix $\mathbf{u} \in S^{n-1} \cap \Sigma_k^n$ and $J' \subseteq [n]$, $|J'| \leq 2k$, arbitrarily. Taking $\mathbf{Z}_1, \ldots, \mathbf{Z}_\ell \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, let $X_{\mathbf{u};i} = \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J'}(\mathbf{Z}_i)\rangle\ \mathsf{sign}(\langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u}) \cup J'}(\mathbf{Z}_i)\rangle)$, $i \in [\ell]$, and write $\bar{X}_{\mathbf{u}} = \sum_{i=1}^{\ell} X_{\mathbf{u};i}$. Let $U_1, \ldots, U_\ell \sim \mathcal{N}(0,1)$ be independent standard normal random variables, and let $\bar{U} = \sum_{i=1}^{\ell} |U_i|$. By Lemma C.1, $X_{\mathbf{u};i} \sim |U_i|$ for each $i \in [\ell]$, and therefore, $\bar{X}_{\mathbf{u}} \sim \bar{U}$. Hence, since $\bar{X}_{\mathbf{u}}$ and $\bar{U}$ follow the same distribution, it suffices to bound the concentration the random variable $\bar{U}$. Define the random vector, $\mathbf{U} = (U_1, \ldots, U_\ell)$, and note that

$$\bar{U} = \sum_{i=1}^{\ell} |U_i| = \|\mathbf{U}\|_1.$$

Recall that for any $\mathbf{w} \in \mathbb{R}^\ell$, $\|\mathbf{w}\|_1 \leq \sqrt{\ell}\|\mathbf{w}\|_2$ since

$$\begin{aligned}
\|\mathbf{w}\|_1 &= \sum_{i=1}^{\ell} |w_i| \\
&= \langle \mathbf{1}, (|w_1|, \ldots, |w_\ell|)\rangle \\
&\leq \|\mathbf{1}\|_2 \|(|w_1|, \ldots, |w_\ell|)\|_2 \\
&\qquad \blacktriangleright \text{ by the Cauchy-Schwarz inequality} \\
&= \|\mathbf{1}\|_2 \sqrt{\sum_{j=1}^{\ell} |w_j|^2} \\
&= \|\mathbf{1}\|_2 \sqrt{\sum_{j=1}^{\ell} w_j^2} \\
&= \|\mathbf{1}\|_2 \|(w_1, \ldots, w_\ell)\|_2 \\
&= \|\mathbf{1}\|_2 \|\mathbf{w}\|_2
\end{aligned}$$

$$= \sqrt{\ell}\|\mathbf{w}\|_2$$

Additionally, observe:

$$\begin{aligned}
\|\mathbf{v}\|_1 &= \|(\mathbf{v}-\mathbf{w})+\mathbf{w}\|_1 \\
&\le \|\mathbf{v}-\mathbf{w}\|_1 + \|\mathbf{w}\|_1 \\
&\qquad \blacktriangleright \text{ by the triangle inequality} \\
&\longrightarrow \|\mathbf{v}\|_1 \le \|\mathbf{v}-\mathbf{w}\|_1 + \|\mathbf{w}\|_1 \\
&\longrightarrow \|\mathbf{v}\|_1 - \|\mathbf{w}\|_1 \le \|\mathbf{v}-\mathbf{w}\|_1 \\
&\qquad \blacktriangleright \text{ rearrangement of terms} \\
&\longrightarrow \|\mathbf{v}\|_1 - \|\mathbf{w}\|_1 \le \sqrt{\ell}\|\mathbf{v}-\mathbf{w}\|_2 \\
&\qquad \blacktriangleright \text{ as argued earlier}
\end{aligned}$$

and thus, $\|\cdot\|_1$ is L-Lipschitz, where $L = \sqrt{\ell}$. By a standard concentration for Gaussian random vectors under L-Lipschitz functions (see, e.g., [35]),

$$\Pr\Big(\|\mathbf{U}\|_1 \ge \mathbb{E}\big[\|\mathbf{U}\|_1\big] + \ell t\Big) \le e^{-\frac{\ell^2 t^2}{2L^2}} = e^{-\frac{\ell^2 t^2}{2\ell}} = e^{-\frac{1}{2}\ell t^2},$$

where

$$\begin{aligned}
\mathbb{E}\big[\|\mathbf{U}\|_1\big] &= \mathbb{E}\left[\sum_{i=1}^{\ell} |U_i|\right] \\
&= \sum_{i=1}^{\ell} \mathbb{E}\big[|U_i|\big] \\
&\qquad \blacktriangleright \text{ by the linearity of expectation} \\
&= \sum_{i=1}^{\ell} \sqrt{\frac{2}{\pi}} \\
&\qquad \blacktriangleright \text{ the mean of a half-normal random variable (well-known)} \\
&= \sqrt{\frac{2}{\pi}}\ell
\end{aligned}$$

Combining the last two derivations yields:

$$\Pr\left(\|\mathbf{U}\|_1 \ge \left(\sqrt{\frac{2}{\pi}} + t\right)\ell\right) \le e^{-\frac{1}{2}\ell t^2}.$$

From this and the earlier discussion, since $\bar{X}_{\mathbf{u}} \sim \bar{U} = \|\mathbf{U}\|_1$, it follows that

$$\Pr\left(\bar{X}_{\mathbf{u}} \ge \left(\sqrt{\frac{2}{\pi}} + t\right)\ell\right) \le e^{-\frac{1}{2}\ell t^2}$$

as desired. This completes the proof of Lemma A.1. ∎

Next, we will proceed to the proof of Lemma A.2.

Proof. (Lemma A.2). Let $\mathbf{u} \in S^{n-1} \cap \Sigma_k^n$ and $J' \subseteq [n]$, $|J'| \le 2k$, and let $\mathbf{Z}_1, \ldots, \mathbf{Z}_\ell \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$ and $\mathbf{W} \sim \mathcal{N}(\mathbf{0}, \ell\mathbf{I}_{2k-1})$ be independent Gaussian vectors, each with i.i.d. entries. Define the random variable

$$\bar{\mathbf{Y}}_{\mathbf{u}} = \sum_{i=1}^{\ell} \Big(T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{Z}_i) - \langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{Z}_i)\rangle \,\mathsf{sign}\big(\langle \mathbf{u}, T_{\mathsf{supp}(\mathbf{u})\cup J'}(\mathbf{Z}_i)\rangle\big)\mathbf{u}\Big)$$

where the random variable of interest is $\|\bar{\mathbf{Y}}_{\mathbf{u}}\|_2$. The majority of the necessary work has already been achieved in Lemmas C.2 and C.4. By Lemma C.2, $\|\bar{\mathbf{Y}}_{\mathbf{u}}\|_2 \sim \|\mathbf{W}\|_2$, and thus, by Lemma C.4,

$$\Pr\Big(\|\bar{\mathbf{Y}}_{\mathbf{u}}\|_2 > \sqrt{(2k-1)\ell} + \ell t\Big) \leq e^{-\frac{1}{2}\ell t^2}$$

as claimed. ∎

D Proof of the Deterministic Results, Lemmas 4.1 and 4.2

D.1 Proof of Lemma 4.1

Proof. (Lemma 4.1). The proof will focus on verifying a slight generalization of Lemma 4.1, which is formally stated as the following claim.

CLAIM D.1. *Let $\mathbf{u}, \mathbf{v}, \mathbf{z} \in S^{n-1} \cap \Sigma_k^n$, and $\mathbf{w} \in \mathbb{R}^n$, where*

$$\mathbf{u} = \frac{T_k(\mathbf{v}+\mathbf{w})}{\|T_k(\mathbf{v}+\mathbf{w})\|_2}, \tag{D.21}$$

and where $\|\mathbf{v}+\mathbf{w}\|_0 \geq k$. Then,

$$\|\mathbf{z}-\mathbf{u}\|_2 \leq 4\big\|(\mathbf{z}-\mathbf{v}) - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})\cup\mathsf{supp}(\mathbf{v})}(\mathbf{w})\big\|_2. \tag{D.22}$$

Note that $\|\hat{\mathbf{x}}^{(t-1)}\|_0 \leq 2k$, by design, and since the random vector $h_{f;\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})$ follows a continuous distribution, $\|h_{f;\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})\|_0 = n$. Moreover, due to the condition that $n \geq 2k$, $\|\hat{\mathbf{x}}^{(t-1)} + h_{f;\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})\|_0 \geq n-k \geq 2k-k = k$. Hence, by taking $\mathbf{z} = \mathbf{x}$, $\mathbf{u} = \hat{\mathbf{x}}^{(t)}$, $\mathbf{v} = \hat{\mathbf{x}}^{(t-1)}$, and $\mathbf{w} = h_{f;\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})$, where $\|\mathbf{v}+\mathbf{w}\|_0 \geq k$ due to the above discussion, Claim D.1 bounds the approximation error, $d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t)})$, as follows:

$$d_{S^{n-1}}(\mathbf{x}, \hat{\mathbf{x}}^{(t)}) = \|\mathbf{x}-\hat{\mathbf{x}}^{(t)}\|_2 \leq 4\|(\mathbf{x}-\hat{\mathbf{x}}^{(t-1)}) - h_{f;\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)})\|_2.$$

Hence, the proof of Lemma 4.1 amounts to verifying Claim D.1, as accomplished next.

Proof. (Claim D.1). The following work is nearly identical to the arguments in [28, proof of Lemma 4.1]. First, note that

$$\mathbf{u} = \frac{T_k(\mathbf{v}+\mathbf{w})}{\|T_k(\mathbf{v}+\mathbf{w})\|_2} = \frac{T_{\mathsf{supp}(T_k(\mathbf{v}+\mathbf{w}))}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(T_k(\mathbf{v}+\mathbf{w}))}(\mathbf{v}+\mathbf{w})\|_2} = \frac{T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2}.$$

This will be useful later on. Next, observe:

$$\begin{aligned}
&\mathbf{z}-\mathbf{u}\\
&= \mathbf{z} - \frac{T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2}\\
&= \big(\mathbf{z} - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\big) + \big(T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\big) + \left(T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - \frac{T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2}\right)
\end{aligned}$$

Then,

$$\begin{aligned}
&\|\mathbf{z}-\mathbf{u}\|_2\\
&= \left\|\big(\mathbf{z} - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\big) + \big(T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\big) + \left(T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - \frac{T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2}\right)\right\|_2\\
&\text{(D.23)}\\
&\leq \big\|\mathbf{z} - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\big\|_2 + \big\|T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\big\|_2 + \left\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - \frac{T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2}\right\|_2
\end{aligned}$$

where the last inequality applies the triangle inequality. For clarity, denote the three terms in the last line by $\alpha_1, \alpha_2, \alpha_3 \geq 0$, where

$$\alpha_1 \triangleq \left\|\mathbf{z} - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2,$$
$$\alpha_2 \triangleq \left\|T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2,$$
$$\alpha_3 \triangleq \left\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - \frac{T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2}\right\|_2.$$

The remainder of the proof is carried out in the following three steps. (a) First, (D.23) will be upper bounded by

$$\left\|\mathbf{z} - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2 + \left\|T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2 + \left\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - \frac{T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2}\right\|_2$$
$$\leq 2\alpha_1 + 2\alpha_2.$$

(b) Then, simple arguments yield upper bounds on α_1 and α_2. (c) Lastly, combining the preceding work will provide the desired upper bound on $\|\mathbf{z}-\mathbf{u}\|_2$.

Step (a). First, the following derivation establishes the bound: $\alpha_3 \leq \alpha_1 + \alpha_2$.

$$\begin{aligned}
\alpha_3 &= \left\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - \frac{T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2}\right\|_2 \\
&\qquad \blacktriangleright \text{ by the definition of } \alpha_3 \\
&= \left\|\left(\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2 - 1\right)\frac{T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2}\right\|_2 \\
&\qquad \blacktriangleright \text{ by distributivity} \\
&= \left|\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2 - 1\right|\left\|\frac{T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2}\right\|_2 \\
&\qquad \blacktriangleright \text{ by an axiom for metrics} \\
&= \left|\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2 - 1\right| \\
&\qquad \blacktriangleright \because \left\|\frac{T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2}\right\|_2 = 1 \\
&= \left|\left\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2 - \|\mathbf{z}\|_2\right| \\
&\qquad \blacktriangleright \because \|\mathbf{z}\|_2 = 1 \\
&\leq \left\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - \mathbf{z}\right\|_2 \\
&\qquad \blacktriangleright \text{ by the (reverse) triangle inequality (see, Remark D.1, below)} \\
&= \left\|\left(T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right) + \left(T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - \mathbf{z}\right)\right\|_2 \\
&\qquad \blacktriangleright \text{ the inserted } \pm T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) \text{ terms cancel out each other} \\
&\leq \left\|T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2 + \left\|T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - \mathbf{z}\right\|_2 \\
&\qquad \blacktriangleright \text{ by the triangle inequality} \\
&= \left\|\mathbf{z} - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2 + \left\|T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w}) - T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2 \\
&\qquad \blacktriangleright \text{ by rearrangement} \\
&= \alpha_1 + \alpha_2 \\
&\qquad \blacktriangleright \text{ by the definitions of } \alpha_1, \alpha_2
\end{aligned}$$

Remark D.1. *The reverse triangle inequality applied in the above derivation can be established from the triangle inequality. To formalize this, fix* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ *arbitrarily. The reverse triangle inequality claims that* $\big|\|\mathbf{a}\|_2 - \|\mathbf{b}\|_2\big| \leq \|\mathbf{a} - \mathbf{b}\|_2$*. To verify this, note that the triangle inequality implies that* $\|\mathbf{a}\|_2 = \|(\mathbf{a} - \mathbf{b}) + \mathbf{b}\|_2 \leq \|\mathbf{a} - \mathbf{b}\|_2 + \|\mathbf{b}\|_2$*. Thus, by rearrangement,* $\|\mathbf{a}\|_2 - \|\mathbf{b}\|_2 \leq \|\mathbf{a} - \mathbf{b}\|_2$*. Likewise, by swapping the roles of* $\mathbf{a}$ *and* $\mathbf{b}$ *in the above arguments, it follows that* $\|\mathbf{b}\|_2 - \|\mathbf{a}\|_2 \leq \|\mathbf{a} - \mathbf{b}\|_2$*. Combining the two bounds then yields the reverse triangle inequality:* $\big|\|\mathbf{a}\|_2 - \|\mathbf{b}\|_2\big| \leq \|\mathbf{a} - \mathbf{b}\|_2$*.*

Now, we have that

$$\|\mathbf{z} - \mathbf{u}\|_2 \leq \alpha_1 + \alpha_2 + \alpha_3 \leq \alpha_1 + \alpha_2 + \alpha_1 + \alpha_2 = 2\alpha_1 + 2\alpha_2$$

which completes Step (a).

Step (b). Write

$$\alpha_1' \triangleq \left\|\mathbf{z} - T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})}(\mathbf{v} + \mathbf{w})\right\|_2,$$
$$\alpha_2' \triangleq \left\|T_{\mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})}(\mathbf{v} + \mathbf{w})\right\|_2.$$

The goal in this step will be to show that $\alpha_2 \leq \alpha_1 \leq \alpha_1'$. To bound α_2, observe:

$$\alpha_2 = \left\|T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})}(\mathbf{v} + \mathbf{w}) - T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v} + \mathbf{w})\right\|_2$$
$$= \sqrt{\sum_{j=1}^{n} (T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})}(\mathbf{v} + \mathbf{w}))_j^2 - \sum_{j=1}^{n} (T_{\mathsf{supp}(\mathbf{u})}(\mathbf{v} + \mathbf{w}))_j^2}$$

▶ by expanding out the definition of the ℓ_2-norm

$$= \sqrt{\sum_{j \in \mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})} (\mathbf{v} + \mathbf{w})_j^2 - \sum_{j \in \mathsf{supp}(\mathbf{u})} (\mathbf{v} + \mathbf{w})_j^2}$$

▶ due to the definition of the thresholding operation, T

$$= \sqrt{\sum_{j \in \mathsf{supp}(\mathbf{z}) \setminus \mathsf{supp}(\mathbf{u})} (\mathbf{v} + \mathbf{w})_j^2 + \sum_{j \in \mathsf{supp}(\mathbf{u})} (\mathbf{v} + \mathbf{w})_j^2 - \sum_{j \in \mathsf{supp}(\mathbf{u})} (\mathbf{v} + \mathbf{w})_j^2}$$

▶ $\because$ $(\mathsf{supp}(\mathbf{z}) \setminus \mathsf{supp}(\mathbf{u})) \sqcup \mathsf{supp}(\mathbf{u}) = \mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})$ is a disjoint partition

$$= \sqrt{\sum_{j \in \mathsf{supp}(\mathbf{z}) \setminus \mathsf{supp}(\mathbf{u})} (\mathbf{v} + \mathbf{w})_j^2}$$

▶ the leftmost pair of summations in the preceding line cancel out each other

$$= \sqrt{\sum_{j=1}^{n} T_{\mathsf{supp}(\mathbf{z}) \setminus \mathsf{supp}(\mathbf{u})}(\mathbf{v} + \mathbf{w})_j^2}$$

▶ due to the definition of the thresholding operation, T

$$= \left\|T_{\mathsf{supp}(\mathbf{z}) \setminus \mathsf{supp}(\mathbf{u})}(\mathbf{v} + \mathbf{w})\right\|_2$$

▶ by condensing notation via the definition of the ℓ_2-norm

$$\leq \left\|T_{\mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})}(\mathbf{v} + \mathbf{w})\right\|_2$$

▶ see, Remark D.2 below

$$= \alpha_2'$$

▶ by the definition of α_2'

Remark D.2. *The above derivation uses, where noted, the inequality:*

$$\left\|T_{\mathsf{supp}(\mathbf{z}) \setminus \mathsf{supp}(\mathbf{u})}(\mathbf{v} + \mathbf{w})\right\|_2 \leq \left\|T_{\mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})}(\mathbf{v} + \mathbf{w})\right\|_2.$$

This inequality is verified as follows. Recall that

$$\mathbf{u} = \frac{T_k(\mathbf{v}+\mathbf{w})}{\|T_k(\mathbf{v}+\mathbf{w})\|_2} = \frac{T_{\mathsf{supp}(T_k(\mathbf{v}+\mathbf{w}))}(\mathbf{v}+\mathbf{w})}{\|T_{\mathsf{supp}(T_k(\mathbf{v}+\mathbf{w}))}(\mathbf{v}+\mathbf{w})\|_2},$$

and hence, $\mathsf{supp}(\mathbf{u}) = \mathsf{supp}(T_k(\mathbf{v}+\mathbf{w}))$. *For any* $j \notin \mathsf{supp}(\mathbf{u})$, *the definition of the top-k hard thresholding operation,* T_k, *enforces:* $|v_j + w_j| \le \min_{j' \in \mathsf{supp}(\mathbf{u})} |v_{j'} + w_{j'}|$. *Additionally,* $\|\mathbf{u}\|_0 = k$ *since* $\|\mathbf{v}+\mathbf{w}\|_0 \ge k$ *ensures the top-k entries in* $\mathbf{v}+\mathbf{w}$ *are all nonzero. On the other hand, recall that* $\|\mathbf{z}\|_0 \le k$. *As a result,* $|\mathsf{supp}(\mathbf{z}) \setminus \mathsf{supp}(\mathbf{u})| \le |\mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})|$, *and at the same time, for any* $j \in \mathsf{supp}(\mathbf{z}) \setminus \mathsf{supp}(\mathbf{u})$ *and* $j' \in \mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})$, $|v_j + w_j| \le |v_{j'} + w_{j'}|$. *Write* $\ell \triangleq |\mathsf{supp}(\mathbf{z}) \setminus \mathsf{supp}(\mathbf{u})|$ *and* $\ell' \triangleq |\mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})|$, *and let* $\{j_1, \ldots, j_\ell\} = \mathsf{supp}(\mathbf{z}) \setminus \mathsf{supp}(\mathbf{u})$ *and* $\{j'_1, \ldots, j'_{\ell'}\} = \mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})$. *Note that by the above discussion,* $\ell \le \ell'$, *and* $|v_{j_s} + w_{j_s}| \le |v_{j'_s} + w_{j'_s}|$ *for each* $s \in [\ell]$. *Taken together, the desired inequality follows:*

$$\begin{aligned}
\left\|T_{\mathsf{supp}(\mathbf{z})\setminus\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2 &= \sum_{j=1}^{n} T_{\mathsf{supp}(\mathbf{z})\setminus\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})_j^2 \\
&\quad \blacktriangleright \textit{ by expanding out the definition of the } \ell_2\textit{-norm} \\
&= \sum_{j \in \mathsf{supp}(\mathbf{z})\setminus\mathsf{supp}(\mathbf{u})} (v_j + w_j)^2 \\
&\quad \blacktriangleright \textit{ due to the definition of the hard thresholding operation, } T \\
&= \sum_{s=1}^{\ell} (v_{j_s} + w_{j_s})^2 \\
&\quad \blacktriangleright \textit{ by reindexing with } \{j_1, \ldots, j_\ell\} = \mathsf{supp}(\mathbf{z}) \setminus \mathsf{supp}(\mathbf{u}) \\
&\le \sum_{s=1}^{\ell} (v_{j'_s} + w_{j'_s})^2 \\
&\quad \blacktriangleright \textit{ by the earlier observation that } |v_{j_s} + w_{j_s}| \le |v_{j'_s} + w_{j'_s}| \\
&\le \sum_{s=1}^{\ell'} (v_{j'_s} + w_{j'_s})^2 \\
&\quad \blacktriangleright \textit{ since all summands are nonnegative, and } \ell \le \ell' \\
&= \sum_{j \in \mathsf{supp}(\mathbf{u})\setminus\mathsf{supp}(\mathbf{z})} (v_j + w_j)^2 \\
&\quad \blacktriangleright \textit{ by reindexing with } \{j'_1, \ldots, j'_{\ell'}\} = \mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z}) \\
&= \sum_{j=1}^{n} T_{\mathsf{supp}(\mathbf{u})\setminus\mathsf{supp}(\mathbf{z})}(\mathbf{v}+\mathbf{w})_j^2 \\
&\quad \blacktriangleright \textit{ due to the definition of the hard thresholding operation, } T \\
&= \left\|T_{\mathsf{supp}(\mathbf{u})\setminus\mathsf{supp}(\mathbf{z})}(\mathbf{v}+\mathbf{w})\right\|_2 \\
&\quad \blacktriangleright \textit{ by condensing notation via the definition of the } \ell_2\textit{-norm}
\end{aligned}$$

Thus, the desired inequality, $\|T_{\mathsf{supp}(\mathbf{z})\setminus\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\|_2 \le \|T_{\mathsf{supp}(\mathbf{u})\setminus\mathsf{supp}(\mathbf{z})}(\mathbf{v}+\mathbf{w})\|_2$, *has been verified.*

The above work has shown that $\alpha_2 \le \alpha'_2$. Next, to bound α'_2, observe:

$$\begin{aligned}
\alpha_1^2 &= \left\|\mathbf{z} - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2^2 \\
&\quad \blacktriangleright \text{ by the definition of } \alpha_1 \\
&= \sum_{j=1}^{n} \Big(z_j - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})_j\Big)^2
\end{aligned}$$

▶ by expanding out the definition of the ℓ_2-norm

$$= \sum_{j \in \mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})} (z_j - v_j - w_j)^2$$

▶ since $\mathsf{supp}(\mathbf{z} - T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})}(\mathbf{v} + \mathbf{w})) \subseteq \mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})$

$$= \sum_{j \in \mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})} (z_j - v_j - w_j)^2 + \sum_{j \in \mathsf{supp}(\mathbf{z})} (z_j - v_j - w_j)^2$$

▶ $\because (\mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})) \sqcup \mathsf{supp}(\mathbf{z}) = \mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})$ is a disjoint partition

$$= \sum_{j \in \mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})} (-v_j - w_j)^2 + \sum_{j \in \mathsf{supp}(\mathbf{z})} (z_j - v_j - w_j)^2$$

▶ $\because z_j = 0$ for any $j \in [n] \setminus \mathsf{supp}(\mathbf{z})$

$$= \sum_{j \in \mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})} (v_j + w_j)^2 + \sum_{j \in \mathsf{supp}(\mathbf{z})} (z_j - v_j - w_j)^2$$

▶ by squaring the -1 factor

$$= \sum_{j=1}^{n} T_{\mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})}(\mathbf{v} + \mathbf{w})_j^2 + \sum_{j=1}^{n} T_{\mathsf{supp}(\mathbf{z})}(\mathbf{z} - (\mathbf{v} + \mathbf{w}))_j^2$$

▶ due to the definition of the hard thresholding operation, T

$$= \left\| T_{\mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})}(\mathbf{v} + \mathbf{w}) \right\|_2^2 + \left\| T_{\mathsf{supp}(\mathbf{z})}(\mathbf{z} - (\mathbf{v} + \mathbf{w})) \right\|_2^2$$

▶ by condensing notation via the definition of the ℓ_2-norm

In short,

$$\alpha_1^2 = \left\| \mathbf{z} - T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})}(\mathbf{v} + \mathbf{w}) \right\|_2^2 = \left\| T_{\mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})}(\mathbf{v} + \mathbf{w}) \right\|_2^2 + \left\| T_{\mathsf{supp}(\mathbf{z})}(\mathbf{z} - (\mathbf{v} + \mathbf{w})) \right\|_2^2.$$

Rearranging the terms obtains:

$$\begin{aligned} \left\| T_{\mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})}(\mathbf{v} + \mathbf{w}) \right\|_2^2 &= \left\| \mathbf{z} - T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})}(\mathbf{v} + \mathbf{w}) \right\|_2^2 - \left\| T_{\mathsf{supp}(\mathbf{z})}(\mathbf{z} - (\mathbf{v} + \mathbf{w})) \right\|_2^2 \\ &\leq \left\| \mathbf{z} - T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})}(\mathbf{v} + \mathbf{w}) \right\|_2^2. \end{aligned}$$

Hence, after taking a square root:

$$\alpha_2' = \left\| T_{\mathsf{supp}(\mathbf{u}) \setminus \mathsf{supp}(\mathbf{z})}(\mathbf{v} + \mathbf{w}) \right\|_2 \leq \left\| \mathbf{z} - T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})}(\mathbf{v} + \mathbf{w}) \right\|_2 = \alpha_1$$

The final task for Step (b) is bounding α_1. For this purpose, observe the following. (Note that the comments throughout the derivation below take $J \subseteq [n]$, $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$.)

$$\left\| \mathbf{z} - T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})}(\mathbf{v} + \mathbf{w}) \right\|_2^2$$

$$= \left\| T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})}(\mathbf{z}) - T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})}(\mathbf{v} + \mathbf{w}) \right\|_2^2$$

▶ $\because T_J(\mathbf{a}) = \mathbf{a}$ if $J \supseteq \mathsf{supp}(\mathbf{a})$

$$= \left\| T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})}(\mathbf{z} - \mathbf{v} - \mathbf{w}) \right\|_2^2$$

▶ $\because T_J(\mathbf{a}) + T_J(\mathbf{b}) = T_J(\mathbf{a} + \mathbf{b})$

$$= \left\| T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})}(\mathbf{z} - \mathbf{v} - \mathbf{w}) + T_{\mathsf{supp}(\mathbf{v}) \setminus (\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u}))}(\mathbf{z} - \mathbf{v} - \mathbf{w}) \right\|_2^2$$

▶ $(\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})) \sqcup (\mathsf{supp}(\mathbf{v}) \setminus (\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u}))) = \mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})$ is a (disjoint) partition

$$= \left\| T_{\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u})}(\mathbf{z} - \mathbf{v} - \mathbf{w}) \right\|_2^2 + \left\| T_{\mathsf{supp}(\mathbf{v}) \setminus (\mathsf{supp}(\mathbf{z}) \cup \mathsf{supp}(\mathbf{u}))}(\mathbf{z} - \mathbf{v} - \mathbf{w}) \right\|_2^2$$

▶ by the Pythagorean theorem

▶ note that $T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{z}-\mathbf{v}-\mathbf{w})$ and $T_{\mathsf{supp}(\mathbf{v})\setminus(\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u}))}(\mathbf{z}-\mathbf{v}-\mathbf{w})$ are orthogonal since their support sets are disjoint

$\geq \left\|T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{z}-\mathbf{v}-\mathbf{w})\right\|_2^2$

▶ since both terms in the preceding line are nonnegative, deleting one cannot increase the value of the expression

$= \left\|T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{z}) - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2^2$

▶ $\because T_J(\mathbf{a}) + T_J(\mathbf{b}) = T_J(\mathbf{a}+\mathbf{b})$

$= \left\|\mathbf{z} - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2^2$

▶ $\because T_J(\mathbf{a}) = \mathbf{a}$ if $J \supseteq \mathsf{supp}(\mathbf{a})$

$= \alpha_1^2$

▶ by the definition of α_1

Thus,

$$\alpha_1 = \left\|\mathbf{z} - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})}(\mathbf{v}+\mathbf{w})\right\|_2 \leq \left\|\mathbf{z} - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})\cup\mathsf{supp}(\mathbf{v})}(\mathbf{v}+\mathbf{w})\right\|_2 = \alpha_1'$$

as claimed. To summarize, this step has shown:

$$\alpha_2 \leq \alpha_2' \leq \alpha_1 \leq \alpha_1',$$

Step (c). By combining the arguments of Steps (a) and (b), Equation (D.22) follows:

$$\begin{aligned}\|\mathbf{z}-\mathbf{u}\|_2 \leq 2\alpha_1 + 2\alpha_2 \leq 4\alpha_1 \leq 4\alpha_1' &\leq 4\left\|\mathbf{z} - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})\cup\mathsf{supp}(\mathbf{v})}(\mathbf{v}+\mathbf{w})\right\|_2 \\ &= 4\left\|(\mathbf{z}-\mathbf{v}) - T_{\mathsf{supp}(\mathbf{z})\cup\mathsf{supp}(\mathbf{u})\cup\mathsf{supp}(\mathbf{v})}(\mathbf{w})\right\|_2.\end{aligned}$$

This completes the proof of Claim D.1. ∎

By the discussion at the beginning of this proof, due to the proof of Claim D.1, Lemma 4.1 also holds. ∎

D.2 Proof of Lemma 4.2 Before the proof of Lemma 4.2 is laid out, the following fact is stated to facilitate this. The proof of this fact can be found in [28].

FACT D.1. ([28, FACT 4.1]) *Let $u, v, w, w_0 \in \mathbb{R}_+$, where $u = \frac{1}{2}(1+\sqrt{1+4w})$ and $u \in [1, \sqrt{\frac{2}{v}}]$. Let $f_1, f_2 : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ be the functions given by*

(D.24a) $$f_1(0) = 2,$$

(D.24b) $$f_1(t) = vw + \sqrt{vf_1(t-1)}, \quad t \in \mathbb{Z}_+,$$

(D.25) $$f_2(t) = 2^{2^{-t}}(u^2v)^{1-2^{-t}}, \quad t \in \mathbb{Z}_{\geq 0}.$$

The functions, f_1 and f_2, are strictly decreasing and satisfy

(D.26) $$f_1(t) \leq f_2(t), \quad \forall t \in \mathbb{Z}_{\geq 0},$$

(D.27) $$\lim_{t\to\infty} f_2(t) = \lim_{t\to\infty} f_1(t) = u^2v.$$

Proof. (Lemma 4.2). The lemma's results follow from an argument nearly identical to the proofs of [28, Lemmas 4.2 and 4.3] with just a couple changes in constants. The (combined) proofs are reproduced below with the appropriate adjustments to constants. The results are derived simply via Fact D.1. Recall the definition of the function, $\varepsilon : \mathbb{Z}_{\geq 0} \to \mathbb{R}$, by the recurrence relation:

(D.28a) $$\varepsilon(0) = 2$$

$$\varepsilon(t) = 4c_1\sqrt{\frac{\gamma}{c}\varepsilon(t-1)} + \frac{4c_2\gamma}{c}, \qquad t \in \mathbb{Z}_+. \tag{D.28b}$$

The first task is writing Equations (D.28) in the form of Equations (D.24). When $t = 0$, then trivially, Equation (D.28a) matches Equation (D.24a), whereas for $t \in \mathbb{Z}_+$, Equation (D.28b) can match the form of Equation (D.24b) by simply writing:

$$\begin{aligned}
\varepsilon(t) &= 4c_1\sqrt{\frac{\gamma}{c}\varepsilon(t-1)} + \frac{4c_2\gamma}{c} \\
&= \sqrt{\frac{16c_1^2\gamma}{c}\varepsilon(t-1)} + \frac{4c_2\gamma}{c} \\
&= \sqrt{\frac{16c_1^2\gamma}{c}\varepsilon(t-1)} + \frac{4c_2}{16c_1^2}\frac{16c_1^2\gamma}{c} \\
&= \sqrt{\frac{16c_1^2\gamma}{c}\varepsilon(t-1)} + \frac{c_2}{4c_1^2}\frac{16c_1^2\gamma}{c} \\
&= \frac{16c_1^2\gamma}{c}\frac{c_2}{4c_1^2} + \sqrt{\frac{16c_1^2\gamma}{c}\varepsilon(t-1)} \\
&= vw + \sqrt{vf_1(t-1)}
\end{aligned}$$

where in the last line,

$$f_1 = \varepsilon, \tag{D.29a}$$

$$v = \frac{16c_1^2\gamma}{c}, \tag{D.29b}$$

$$w = \frac{c_2}{4c_1^2}. \tag{D.29c}$$

Now we have that Equation (D.28b),

$$\varepsilon(t) = 4c_1\sqrt{\frac{\gamma}{c}\varepsilon(t-1)} + \frac{4c_2\gamma}{c},$$

is equivalent to

$$f_1(t) = vw + \sqrt{vf_1(t-1)},$$

the latter of which is precisely the form of Equation (D.24b).

Before Fact D.1 can be applied, it is necessary to verify that the fact's conditions are satisfied when the parameters v, w are chosen as in (D.29). Specifically, writing $u = \frac{1}{2}(1 + \sqrt{1+4w})$, Fact D.1 requires that $1 \le u \le \sqrt{\frac{2}{v}}$. Clearly, $u \ge 1$ since $\frac{1}{2}(1+\sqrt{1+z}) \ge \frac{1}{2}(1+1) = 1$ for $z \ge 0$. Towards verifying the other side of the bound, $u \le \sqrt{\frac{2}{v}}$, expand out u and $\frac{1}{\sqrt{v}}$ as follows. For u, observe:

$$\begin{aligned}
u &= \frac{1}{2}\left(1 + \sqrt{1+4w}\right) \\
&= \frac{1}{2}\left(1 + \sqrt{1 + 4\cdot\frac{c_2}{4c_1^2}}\right) \\
&= \frac{1}{2}\left(1 + \sqrt{1 + \frac{c_2}{c_1^2}}\right) \\
&= \frac{1}{2}\left(1 + \frac{1}{c_1}\sqrt{c_1^2 + c_2}\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2}\left(\frac{1}{c_1}c_1 + \frac{1}{c_1}\sqrt{c_1^2 + c_2}\right) \\
&= \frac{1}{2c_1}\left(c_1 + \sqrt{c_1^2 + c_2}\right) \\
&= \frac{c_1 + \sqrt{c_1^2 + c_2}}{2c_1}
\end{aligned}
$$

Next, $\frac{1}{\sqrt{v}}$ is rewritten as:

$$
\begin{aligned}
\frac{1}{\sqrt{v}} &= \sqrt{\frac{c}{16c_1^2\gamma}} \\
&= \sqrt{\frac{c}{4c_1^2 \cdot 4\gamma}} \\
&= \frac{1}{2c_1}\sqrt{\frac{c}{4\gamma}} \\
&= \frac{1}{\sqrt{\gamma}} \cdot \frac{\sqrt{c/4}}{2c_1}
\end{aligned}
$$

Taking

$$
c = 4\left(c_1 + \sqrt{c_1^2 + c_2}\right)^2,
$$

it follows that $u \leq \sqrt{\frac{2}{v}}$, as required, since:

$$
\begin{aligned}
\sqrt{\frac{2}{v}} &\geq \sqrt{\frac{2}{\gamma}} \cdot \frac{\sqrt{c/4}}{2c_1} \\
&= \sqrt{\frac{2}{\gamma}} \cdot \frac{\sqrt{\frac{1}{4} \cdot 4\left(c_1 + \sqrt{c_1^2 + c_2}\right)^2}}{2c_1} \\
&= \sqrt{\frac{2}{\gamma}} \cdot \frac{\sqrt{\left(c_1 + \sqrt{c_1^2 + c_2}\right)^2}}{2c_1} \\
&= \sqrt{\frac{2}{\gamma}} \cdot \frac{c_1 + \sqrt{c_1^2 + c_2}}{2c_1} \\
&= \sqrt{\frac{2}{\gamma}} u \\
&\geq u
\end{aligned}
$$

Hence, the fact applies since $1 \leq u \leq \sqrt{\frac{2}{v}}$.

The lemma's results can now be obtained via Fact D.1. Note that $\sqrt{\frac{2}{v}} \geq \sqrt{\frac{2}{\gamma}} \cdot \frac{\sqrt{c/4}}{2c_1}$ implies $\sqrt{\frac{\gamma}{v}} \geq \frac{\sqrt{c/4}}{2c_1}$. Then, observe:

$$
\sqrt{\frac{\gamma}{v}} \geq \frac{\sqrt{\gamma}}{\sqrt{\gamma}} \cdot \frac{\sqrt{c/4}}{2c_1} = \frac{\sqrt{c/4}}{2c_1} = \frac{c_1 + \sqrt{c_1^2 + c_2}}{2c_1} = u
$$

To state the result briefly, we have established that $u \leq \sqrt{\frac{\gamma}{v}}$. Thus,

$$
u^2 v \leq \left(\sqrt{\frac{\gamma}{v}}\right)^2 v = \frac{\gamma}{v} \cdot v = \gamma.
$$

Lastly, applying Fact D.1 yields

$$\varepsilon(t) \leq 2^{2^{-t}} (u^2 v)^{1-2^{-t}} \leq 2^{2^{-t}} \gamma^{1-2^{-t}},$$
$$\lim_{t\to\infty} \varepsilon(t) = u^2 v \leq \gamma,$$

as desired. ∎