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

Namiko Matsumoto Arya Mazumdar

July 8, 2022

Abstract

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

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

1 Introduction

One-bit compressed sensing (1bCS) is a basic nonlinear sampling method for high-dimensional sparse signals, introduced first in Boufounos and Baraniuk (2008). Consider an unknown sparse signal $\mathbf{x} \in \mathbb{R}^n$ with sparsity (number of nonzero coordinates) $||\mathbf{x}||_0 \leq k, k \ll n$. In the 1bCS framework, measurements of $\mathbf{x}$ are obtained with a sensing matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ via the observations of signs:

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

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

Compressed sensing, the method of obtaining signals by taking few linear projections Donoho (2006); Candès et al. (2006) has seen a lot of success in the past two decades. 1bCS is an extremely-quantized version of compressed sensing where only 1 bit per sample of the signal is observed. In terms of nonlinearity, this is one of the simplest example of a single-index model Plan and Vershynin (2016): $y_i = f(\langle \mathbf{a}_i, \mathbf{x}\rangle), i = 1, \ldots, m$, where f is a coordinate-wise nonlinear operation. Both as a practical case-study, and for being aesthetically appealing, 1bCS has been studied with interest in the last few years, for example, in Haupt and Baraniuk (2011); Gopi et al. (2013); Acharya et al. (2017); Plan and Vershynin (2013a); Li (2016).

*The authors are with the University of California San Diego. This work is supported in part by NSF awards 2133484 and 2127929. Emails: {nmatsumo,arya}@ucsd.edu.

Notably, it was shown in Jacques et al. (2013b) that $m = \tilde{\Theta}(k/\epsilon)$ measurements are necessary and sufficient (up to logarithmic factors) to approximate $\mathbf{x}$ within an ϵ-ball. But the reconstruction method used to obtain this measurement complexity is via exhaustive search, which is practically infeasible. A linear programming based solution (which runs in polynomial time) that has measurement complexity $O(\frac{k}{\epsilon^5}\log^2\frac{n}{k})$ was provided in Plan and Vershynin (2013b). Note the suboptimal dependence on ϵ.

An incredibly well-performing algorithm turned out to the *binary iterative hard thresholding* (BIHT) algorithm, proposed in the former work Jacques et al. (2013b). BIHT is a simple iterative algorithm that converges to the correct solution quickly in practice. However, until later, the reason of its good performance was somewhat unexplained, barring the fact that it is actually a proximal gradient descent algorithm on a certain loss function (provided in Eq. (7)). In the algorithm, the projection is taken onto a nonconvex set (namely, selecting the "top-k" coordinates and then normalizing), which usually makes a theoretical analysis unwieldy. Since the work of Jacques et al. (2013b) there have been some progress explaining the empirical success of the BIHT algorithm. In particular, it was shown in (Boufounos et al. 2015, Sec. 3.4.2) that after only the first iteration of BIHT algorithm an approximation error ϵ is achievable with $\tilde{O}(\frac{k}{\epsilon^4})$ measurements, though the same result is shown in (Jacques et al. 2013a, Sec. 5) with $\tilde{O}(\frac{k}{\epsilon^2})$ measurements, so the former result might just be a typo. Similar results also appear in (Plan et al. 2017, Sec. 3.5). In all these results, the dependence on ϵ, which is also referred to as error-rate, is suboptimal. Furthermore, these works also do not show convergence as the algorithm iterates further. Indeed, according to these works $O(\frac{k}{\epsilon^2}\log\frac{n}{k})$ measurements are sufficient to bring the error down to ϵ after just the first iteration of BIHT. Beyond the first iteration, it was shown in Liu et al. (2019) that the iterates of BIHT remain bounded maintaining the same order of accuracy for the subsequent iterations. This, however, does not imply a reduction in the approximation error after the first iteration. This issue have been partially mitigated in Friedlander et al. (2021), which uses a *normalized* version of the BIHT algorithm. While Friedlander et al. (2021) manage to show that the normalized BIHT algorithm can achieve optimal dependence on the error-rate as the number of iterations of BIHT tends to infinity, i.e., $m \sim \frac{1}{\epsilon}$, their result is only valid when $m > \max\{ck^{10}\log^{10}\frac{n}{k}, 24^{48}, \frac{c'}{\epsilon}(k\log\frac{n}{k})^{7/2}\}$. This clearly is very sub-optimal in terms of dependence on k, and do not explain the empirical performance of the algorithm. This has been left as the main open problem in this area as per Friedlander et al. (2021).

1.1 Our Contribution and Techniques

In this paper, we show that normalized BIHT converges with sample complexity having optimal dependence on both sparsity k and error ϵ (see, Theorem 3.1 below). And as such, we also show the convergence rate with respect to iterations for this algorithm. In particular, we show that the approximation error of BIHT decays as $O(\epsilon^{1-2^{-t}})$ with the number of iteration t. This encapsulates the very fast convergence of BIHT to the ϵ-ball of the actual signal. Furthermore, this also shows that after just 1 iteration of BIHT, an approximation error of $\sqrt{\epsilon}$ is achievable, with $O(\frac{k}{\epsilon}\log\frac{n}{k})$ measurements, which matches the observations of Jacques et al. (2013a); Plan et al. (2017) regarding the performance of BIHT with just one iteration. Due to the aforementioned fast rate, the approximation error quickly converges to ϵ resulting in a polynomial time algorithm for recovery in 1bCS with only $\tilde{O}(\frac{k}{\epsilon})$ measurements, the optimal possible.

There are several difficulties in analyzing BIHT that were pointed out in the past, for example in Friedlander et al. (2021). First of all, the loss function is not differentiable, therefore one has to rely on (sub)-gradients, which prohibits an easier analysis of convergence. Secondly, the algorithm projects onto nonconvex sets, so it is not apparent that in each iteration a better approximation is achieved. To tackle these hurdles, the key idea is to use some structural property of the measurement or sampling matrix. Our result relies on such a property of the sampling matrix $\mathbf{A}$, called the restricted invertibility condition. A somewhat different invertibility property of a matrix also appears in Friedlander et al. (2021). However, our definition, which looks more natural, allows for a significantly different analysis - and results in the improved sample complexity. Thereafter, we show that random matrices with i.i.d. Gaussian entries, satisfy the invertibility condition with overwhelmingly large probability.

The invertibility condition that is essential for our proof intuitively states that treating the signed measurements as some "scaled linear" measurements should lead to good enough estimates, which is an overarching theme of recovery in generalized linear models. However, our condition also quantifies "goodness" of this estimate in a way that allows us to show contraction in the BIHT iterations. This contraction of approximation error comes naturally from our definition. While similar idea appear in Friedlander et al. (2021), showing the contraction of approximate error is a much involved exercise therein. Also, it is empirically observed in (Jacques et al. 2013b, Sec. 4.2) that in normalized BIHT, the step-size of the gradient descent algorithm must be carefully chosen, or the algorithm will not converge. Our definition of the invertibility condition gives some intuitive justification on why the algorithm is so sensitive to step-size. Our analysis relies on the step-size being set exactly to $\eta = \sqrt{2\pi}$. More generally, if η were to deviate too far from $\sqrt{2\pi}$, the contraction would be lost.

So the technical burden of our main result turns out to be to show Gaussian matrices do satisfy the invertibility condition (Definition 3.1 below). We need to show that for every pair of sparse unit vectors the condition holds. We resort to constructing a cover, an "epsilon-net," of the unit sphere, and then decompose the invertibility conditions for any two vectors in the sphere into two components. First, we show that it is satisfied for two vectors in the epsilon-net whose distance is sufficiently large, then we show that only small error is added when instead of the net points, vectors close to them is considered. This leads to a "large-distance" and "small-distance" analysis. For these two parts, we require differently curated concentration inequalities, which form the bulk of the techniques used in this paper. Notably, we cannot just extend the invertibility condition to points outside the net by simply using, e.g., the triangle inequality, due to the sign operation. But at the same time, the sign operation significantly reduces the number of matrix-vector products we need to union bound over. It turns out that, because we condition on the rotational uniformity of the measurements, this number is not "too large," and will not increase the sample complexity beyond the optimal.

One important aspect of BIHT's convergence is that as the approximation error in t^{th} iteration improves, it makes possible an even smaller error for the $(t+1)^{\text{th}}$ approximation. This can again be intuitively explained by the rotational symmetry of the measurements, as well as the sign operation. Each iteration of BIHT involves fewer and fewer measurements, and we can track the number of measurements involved by tracking the number of measurements that *mismatch* between the vector $\mathbf{x}$ and its approximation at the t^{th} iteration. This is used in the "large-distance" regime, where the pairs of points must be at least distance τ from each other (note that that qualifier is necessary). On the other hand, once the distance is smaller than τ, the Chernoff bound that is used to track the mismatch is no longer sufficient (using that we would end up needing a suboptimal sample complexity). That is why the "small-distance" analysis is needed separately. However, as mentioned above, because of the rotational uniformity, the number of different ways that we have to include in union bound in this small distance regime is not that many. In some sense, what prevents us from simply extending the argument made for "large-distance" regime to the "small-distance" regime is also what enables the approach taken in the "small-distance" regime. Reconciling these two regimes was necessary for our approach.

1.2 Other Related Works

A generalization of 1bCS is the noisy version of the problem, where the binary observations $y_i \in \{+1, -1\}$ are random (noisy): i.e., $y_i = 1$ with probability $f(\langle \mathbf{a}_i, \mathbf{x} \rangle), i = 1, \ldots, m$, where f is a potentially nonlinear function, such as the sigmoid function. Recovery guarantee for such models were studied in Plan and Vershynin (2013a). Another model of observational noise can appear before the quantization, i.e., $y_i = \mathsf{sign}\left(\langle \mathbf{a}_i, \mathbf{x} \rangle + \eta_i\right), i = 1, \ldots, m$, where η_i is random noise. As observed in Plan and Vershynin (2016); Friedlander et al. (2021), the noiseless setting (the one we consider) is actually more difficult to handle, because the randomness of noise allows for a maximum likelihood analysis. Indeed, having some control-over η_is (or just assuming them to be i.i.d. Gaussian), helps estimate the norm of $\mathbf{x}$ Knudson et al. (2016), which is otherwise impossible with just sign measurements, as in our model (this is called introducing *dither*, a well-known paradigm in signal processing). In a related line of work, one bit measurements are taken by adaptively varying the threshold (in our case the threshold is 0 all the time), which can lead to much improved error-rate, for example see Baraniuk et al. (2017) and Saab et al. (2018), the later being on application of sigma-delta quantization methods.

Yet another line of works in 1bCS literature takes a more combinatorial avenue and looks at the support recovery problem and constructions of structured measurement matrices. Instances of these works are Gopi et al. (2013); Acharya et al. (2017); Flodin et al. (2019); Mazumdar and Pal (2021). However, the nature of these works are quite different from ours.

1.3 Organization

The rest of the paper is organized as follows. The required notations and definitions to state the main result appear in Section 2, where we also formally define the 1-bit compressed sensing problem and the reconstruction method, the normalized binary iterative hard thresholding algorithm (Algorithm 1). We provide our main result in Section 3, which establishes the convergence rate of BIHT (Theorem 3.1) and the asymptotic error rate (Corollary 3.2) with the optimal measurement complexity. In Section 3.2 we also provide an overview of how the result is derived. In Section 4 we provide the main proof of the BIHT convergence algorithm, assuming that a structural property if satisfied by the measurement matrix. Proof of this structural property for Gaussian matrices is the major technical contribution of this paper (Theorem 3.3), and it has been delegated to Appendix A. Proofs of all lemmas and intermediate results can be found in the appendix. We conclude with some future directions in Section 5.

2 Preliminaries

2.1 Notations and Definitions

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

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

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

Note that these are related by $\theta_{\mathbf{u},\mathbf{v}} = \arccos\left(1 - \frac{d^2_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{2}\right)$, equivalently, $d_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v}) = \sqrt{2(1 - \cos(\theta_{\mathbf{u},\mathbf{v}}))}$.

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

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

The function can be extended to vectors, i.e., $\mathsf{sign} : \mathbb{R}^n \to \{+1, -1\}^n$ by just applying the it on each coordinate.

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

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

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

Definition 2.1 (Hard thresholding operation (k-top))**.** *For $k \in \mathbb{Z}_+$, the hard thresholding operation $\mathcal{T}_k : \mathbb{R}^n \to \Sigma_k^n$ projects a real-valued vector $\mathbf{u} \in \mathbb{R}^n$ into the space of k-sparse real-valued vector by setting all but the k largest (in absolute value) entries in $\mathbf{u}$ to 0 (with ties broken arbitrarily).*

Definition 2.2 (Hard thresholding operation (by a set))**.** *For a subset of coordinates $J \subseteq [n]$, the hard thresholding operation $\mathcal{T}_J : \mathbb{R}^n \to \Sigma_k^n$ associated with J projects a real-valued vector $\mathbf{u} \in \mathbb{R}^n$ into the space of real-valued, k-sparse vectors by*

$$\mathcal{T}_J(\mathbf{u})_j = \begin{cases} u_j, & \text{if } j \in J, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

This operation $\mathcal{T}_J$ is a linear transformation (see Lemma D.10, Appendix D) associated with a diagonal $n \times n$ matrix denoted $\mathbf{T}_J = \mathsf{diag}(T_{1;J}, \ldots, T_{n;J})$, where

$$T_{j;J} = \begin{cases} 1, & \text{if } j \in J, \\ 0, & \text{if } j \notin J. \end{cases} \tag{5}$$

2.2 1-Bit Compressed Sensing and the BIHT Algorithm

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

$$\mathbf{b} = \mathsf{sign}(\mathbf{A}\mathbf{x}) \tag{6}$$

Throughout this work, the unknown signals, $\mathbf{x} \in \Sigma_k^n$, are assume to have unit norm since information about the norm is lost due to the binarization of the responses. (For interested readers, see Knudson et al. (2016) for techniques, e.g., dithering, to reconstruct the signal's norm in 1-bit compressed sensing.) Given $\mathbf{A}$ and $\mathbf{b}$, the goal of 1-bit compressed sensing is to recover $\mathbf{x}$ as accurately as possible. We measure the accuracy of reconstruction by the metric $d_{\mathcal{S}^{n-1}}(\cdot, \cdot)$.

The binary iterative hard thresholding (BIHT) reconstruction algorithm, proposed by Jacques et al. (2013c), comprises two iterative steps: (i) a gradient descent step, which finds a non-sparse approximation, $\tilde{\mathbf{x}} \in \mathbb{R}^n$, followed by (ii) a projection by $\tilde{\mathbf{x}} \mapsto \hat{\mathbf{x}} = \mathcal{T}_k(\tilde{\mathbf{x}})$ into the space of k-sparse, real-valued vectors. As shown by Jacques et al. (2013c), the gradient step, (i), aims to minimize the objective function

$$\mathcal{J}(\hat{\mathbf{x}}; \mathbf{x}) = \left\| [\, \mathsf{sign}(\mathbf{A}\mathbf{x}) \odot \mathsf{sign}(\mathbf{A}\hat{\mathbf{x}}) \,]_- \right\|_1, \tag{7}$$

where $\mathbf{u} \odot \mathbf{v} = (u_1 v_1, \ldots, u_n v_n)$ and $(\, [\, \mathbf{u} \,]_- \,)_j = u_j \cdot \mathbf{1}(u_j < 0)$. While several variants of the BIHT algorithm have been proposed, Jacques et al. (2013c), this work focuses on the normalized BIHT algorithm, where the projection step, (ii), is modified to project the approximation onto the k-sparse, ℓ_2-unit sphere, $\mathcal{S}^{n-1} \cap \Sigma_k^n$. Algorithm 1 provides the version of the BIHT algorithm studied in this work.

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

```
Set η = √(2π)
x̂^(0) ~ S^{n-1} ∩ Σ_k^n
for t = 1, 2, 3, ... do
    x̃^(t) ← x̂^(t-1) + (η / 2m) Aᵀ ( sign(Ax) − sign(A x̂^(t-1)) )
    x̂^(t) ← T_k(x̃^(t)) / ‖T_k(x̃^(t))‖_2
```

3 Main Results and Techniques

3.1 BIHT Convergence Theorem

Our main results is presented below. Informally, it states that with $m = O(\frac{k}{\epsilon} \log \frac{n}{k\sqrt{\epsilon}})$ 1-bit (sign) measurements, it is possible to recover any k-sparse vector within an ϵ-ball, by means of the normalized BIHT algorithm.

Theorem 3.1. *Let $a, b, c > 0$ be universal constants as in Eq. (3). Fix $\epsilon, \rho \in (0,1)$ and $k, m, n \in \mathbb{Z}_+$ where*

$$m \geq \frac{4bck}{\epsilon} \log\left(\frac{en}{k}\right) + \frac{2bck}{\epsilon} \log\left(\frac{12bc}{\epsilon}\right) + \frac{bc}{\epsilon} \log\left(\frac{a}{\rho}\right). \tag{8}$$

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

$$d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\right) \leq 2^{2^{-t}} \epsilon^{1-2^{-t}} \tag{9}$$

for all $t \in \mathbb{Z}_{>0}$.

Corollary 3.2. *Under the conditions stated in Theorem 3.1, uniformly with probability at least $1-\rho$, for all unknown k-sparse, real-valued unit vectors $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, the sequence of BIHT approximations, $\{\hat{\mathbf{x}}^{(t)}\}_{t \in \mathbb{Z}_{\geq 0}}$, converges asymptotically into the ϵ-ball around the unknown vector $\mathbf{x}$. Formally,*

$$\lim_{t \to \infty} d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\right) \leq \epsilon. \tag{10}$$

3.2 Technical Overview

The analysis in this work is divided into two components: (I) the proofs of Theorem 3.1 and Corollary 3.2, which show universal convergence of the BIHT approximations by using the *restricted approximate invertibility condition* (RAIC) for an i.i.d. standard normal measurement matrix (defined below), and (II) the proof of the main technical theorem, Theorem 3.3 (also below), which derives the RAIC for such a measurement matrix.

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

3.2.1 The Restricted Approximate Invertibility Condition

The main technical contribution is an improved sample complexity for the restricted approximate invertibility condition (RAIC). A different invertibilty condition was proposed by Friedlander et al. (2021). We have included the definition of Friedlander et al. (2021) in Appendix E, for comparison, and to emphasize the major differences. The definition of RAIC considered in this work is formalized in Definition 3.1, which uses the following notations. For $m, n \in \mathbb{Z}_+$, let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a measurement matrix with rows $\mathbf{A}^{(i)} \in \mathbb{R}^n$, $i \in [m]$. Then, define the functions $h_{\mathbf{A}}, h_{\mathbf{A};J} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$h_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) = \frac{\eta}{m} \mathbf{A}^\mathsf{T} \cdot \frac{1}{2} (\mathsf{sign}(\mathbf{A}\mathbf{x}) - \mathsf{sign}(\mathbf{A}\mathbf{y})) \tag{11}$$

$$h_{\mathbf{A};J}(\mathbf{x}, \mathbf{y}) = \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\mathbf{y}) \cup J}(h_{\mathbf{A}}(\mathbf{x}, \mathbf{y})) \tag{12}$$

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

Definition 3.1 (restricted approximate invertibility condition (RAIC)). *Fix $\delta, a_1, a_2 > 0$ and $k, m, n \in \mathbb{Z}_+$ such that $0 < k < n$. The (k, n, δ, a_1, a_2)-RAIC is satisfied by a measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ if*

$$\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 \leq a_1 \sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{x},\mathbf{y})} + a_2\delta \tag{13}$$

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

Theorem 3.3 below is the primary technical result in this analysis and establishes that m-many i.i.d. standard normal measurements satisfy the (k, n, δ, c_1, c_2)-RAIC, where the sample complexity for m matches the lower bound of (Jacques et al. 2013c, Lemma 1). The proof of the theorem is deferred to Appendix A, and a overview of the proof is given below in Section 3.2.3.

Theorem 3.3. *Let $a, b, c_1, c_2 > 0$ be universal constants as defined in Eq. (3). Fix $\delta, \rho \in (0,1)$ and $k, m, n \in \mathbb{Z}_+$ such that $0 < k < n$ and*

$$m = \frac{b}{\delta} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12b}{\delta}\right)^{2k} \left(\frac{a}{\rho}\right)\right) \in O\left(\frac{k}{\delta}\log\left(\frac{n}{\delta k}\right) + \frac{1}{\delta}\log\left(\frac{1}{\rho}\right)\right). \tag{14}$$

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a measurement matrix whose rows $\mathbf{A}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n \times n})$, $i \in [m]$, have i.i.d. standard normal entries. Then, the measurement matrix $\mathbf{A}$ satisfies the (k, n, δ, c_1, c_2)-RAIC with probability at least $1-\rho$. Formally, uniformly with probability at least $1-\rho$, for all $\mathbf{x}, \mathbf{y} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$ and all $J \subseteq [n]$, $|J| \leq k$,

$$\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 \leq c_1 \sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{x},\mathbf{y})} + c_2\delta. \tag{15}$$

3.2.2 The Uniform Convergence of BIHT Approximations

Assuming the desired RAIC property (i.e., correctness of Theorem 3.3), the uniform convergence of BIHT approximations is shown as follows. (a) The 0^{th} BIHT approximation, $\hat{\mathbf{x}}^{(0)}$, which is simply initialized by drawing a point uniformly at random from $\mathcal{S}^{n-1} \cap \Sigma_k^n$, i.e., $\hat{\mathbf{x}}^{(0)} \sim \mathcal{S}^{n-1} \cap \Sigma_k^n$, can be seen to have error at most 2. Then, the following argument handles each subsequent t^{th} BIHT approximation, $t \in \mathbb{Z}_+$. (b) Using standard techniques, the error of any t^{th} BIHT approximation, $t \in \mathbb{Z}_+$, can be shown to be (deterministically) upper bounded by

$$d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\right) \leq 4 \left\| \left(\mathbf{x} - \hat{\mathbf{x}}^{(t-1)}\right) - h_{\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^t)}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}\right) \right\|_2. \tag{16}$$

(c) Subsequently, observing the correspondence between Eq. (16) and the RAIC, Theorem 3.3 is applied to further bound the t^{th} approximation error in (16), $t \in \mathbb{Z}_+$, from above by

$$d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\right) \leq 4\left(c_1\sqrt{\frac{\epsilon}{c} d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}\right)} + c_2\frac{\epsilon}{c}\right) = 4c_1\sqrt{\frac{\epsilon}{c} d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}\right)} + 4c_2\frac{\epsilon}{c}. \tag{17}$$

(d) Then, the recurrence relation corresponding to the right-hand-side of Eq. (17),

$$\varepsilon(0) = 2 \tag{18}$$

$$\varepsilon(t) = 4c_1\sqrt{\frac{\epsilon}{c}\varepsilon(t-1)} + 4c_2\frac{\epsilon}{c}, \quad t \in \mathbb{Z}_+, \tag{19}$$

can be shown to be monotonically decreasing with respect to t, asymptotically converging to the ϵ-ball around the unknown vector $\mathbf{x}$, and pointwise upper bounded by $\varepsilon(t) \leq 2^{2^{-t}} \epsilon^{1-2^{-t}}$ for each $t \in \mathbb{Z}_{\geq 0}$. This will complete the analysis for the universal convergence of the BIHT algorithm.

3.2.3 The RAIC for an i.i.d. Standard Normal Measurement Matrix

Fixing $\delta, \rho \in (0,1)$ and letting $c_1, c_2 > 0$ be the universal constants as specified in Eq. (3), Theorem 3.3 establishes that the measurement matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ with i.i.d. standard normal entries satisfies the (k, n, δ, c_1, c_2)-RAIC with high probability (at least $1-\rho$) when the number of measurements m is at least what is given in Eq. (14). The proof of the theorem is outlined as follows.

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

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

$$\left\|(\mathbf{u} - \mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u}, \mathbf{v})\right\|_2 \leq b_1 \sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v})} \tag{20}$$

where $b_1 > 0$ is a small universal constant (see, Eq. (3)).

(c) The second property, which holds on a "small scale," requires that with probability at least $1 - \rho_2$, for each member of the τ-net $\mathbf{u} \in \mathcal{C}_\tau$, each $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$, and every $J \subseteq [n]$, $|J| \leq 2k$,

$$\left\|(\mathbf{x} - \mathbf{u}) - h_{\mathbf{A};J}(\mathbf{x}, \mathbf{u})\right\|_2 \leq b_2 \delta \tag{21}$$

where $b_2 > 0$ is a small universal constant (again see, Eq. (3)).

(d) Requiring $\rho_1 + \rho_2 = \rho$, the last step of the proof derives the RAIC claimed in the theorem by using the results from steps (b) and (c), such that the condition holds with probability at least $1 - \rho$. We provide an overview of these two steps subsequently.

3.2.4 The Large and Small-Scale Regimes, Steps (b) and (c)

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

$$h_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) = \sqrt{2\pi} \frac{1}{m} \mathbf{A}^\mathsf{T} \cdot \frac{1}{2} \left(\mathsf{sign}\left(\mathbf{A}\mathbf{x}\right) - \mathsf{sign}\left(\mathbf{A}\mathbf{y}\right)\right) \tag{22}$$

$$= \sqrt{2\pi} \frac{1}{m} \sum_{i=1}^{m} \mathbf{A}^{(i)} \cdot \frac{1}{2} \left(\mathsf{sign}\left(\langle \mathbf{A}^{(i)}, \mathbf{x} \rangle\right) - \mathsf{sign}\left(\langle \mathbf{A}^{(i)}, \mathbf{y} \rangle\right)\right) \tag{23}$$

$$= \sqrt{2\pi} \frac{1}{m} \sum_{i=1}^{m} \mathbf{A}^{(i)} \cdot \mathsf{sign}\left(\langle \mathbf{A}^{(i)}, \mathbf{x} \rangle\right) \cdot \mathbf{1}\left(\mathsf{sign}\left(\langle \mathbf{A}^{(i)}, \mathbf{x} \rangle\right) \neq \mathsf{sign}\left(\langle \mathbf{A}^{(i)}, \mathbf{y} \rangle\right)\right) \tag{24}$$

Hence, given the random vector $\mathbf{R}_{\mathbf{x},\mathbf{y}} = \frac{1}{2}\left(\mathsf{sign}\left(\mathbf{A}\mathbf{x}\right) - \mathsf{sign}\left(\mathbf{A}\mathbf{y}\right)\right)$, which takes values in $\{-1, 0, 1\}^m$, $(h_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) \mid \mathbf{R}_{\mathbf{x},\mathbf{y}})$ becomes a function of only $L_{\mathbf{x},\mathbf{y}} = \|\mathbf{R}_{\mathbf{x},\mathbf{y}}\|_0 \leq m$-many random vectors. Such conditioning on $\mathbf{R}$ will allow for tighter concentration inequalities related to (an orthogonal decomposition of) $(h_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) \mid \mathbf{R})$. Note that these concentration inequalities, stated in Appendix A, provide the same inequality for any $\mathbf{R}_{\mathbf{x},\mathbf{y}}, \mathbf{R}_{\mathbf{x}',\mathbf{y}'}$ such that $L_{\mathbf{x},\mathbf{y}} = L'_{\mathbf{x}',\mathbf{y}'}$, where $L_{\mathbf{x},\mathbf{y}} = \|\mathbf{R}_{\mathbf{x},\mathbf{y}}\|_0$, $L_{\mathbf{x}',\mathbf{y}'} = \left\|\mathbf{R}_{\mathbf{x}',\mathbf{y}'}\right\|_0$, $\mathbf{x}, \mathbf{y}, \mathbf{x}', \mathbf{y}' \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, and thus it suffices to have a handle on (an appropriate subset of) the random variables $\{L_{\mathbf{x},\mathbf{y}} : \mathbf{x}, \mathbf{y} \in \mathcal{S}^{n-1} \cap \Sigma_k^n\}$.

With this intuition, the large- and small-scale results in steps (b) and (c) are derived using two primary arguments. First, for a given $\mathbf{u}, \mathbf{v} \in \mathcal{C}_\tau$, the associated random variable L is bounded. Then, conditioning on L, the desired properties in steps (b) and (c) follow from the appropriate concentration inequalities related to the decomposition of $h_{\mathbf{A};J}(\mathbf{x}, \mathbf{y})$ into three orthogonal components.

Specifically, step (b) is achieved as follows. (i) Consider any $(\mathbf{u}, \mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau$ such that $d_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v}) \geq \tau$, and fix $J' \subseteq [n]$, $|J'| \leq 2k$ arbitrarily. Note that, by a known construction of a τ-net, all pairs of distinct points do satisfy $d_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v}) \geq \tau$. (ii) It can be shown that for a small $s \in (0, 1)$, the number $L_{\mathbf{u},\mathbf{v}}$ of points $\mathbf{A}^{(i)} \in \mathcal{A}$, $i \in [m]$, such that $\mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{u} \rangle) \neq \mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{v} \rangle)$ is bounded in the range $L_{\mathbf{u},\mathbf{v}} \in [(1-s)\frac{\theta_{\mathbf{u},\mathbf{v}} m}{\pi}, (1+s)\frac{\theta_{\mathbf{u},\mathbf{v}} m}{\pi}]$ uniformly for all $(\mathbf{u}, \mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau$ with high probability. (iii) Define $g_{\mathbf{A}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ by

$$g_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) = h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) - \left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) \right\rangle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2} - \left\langle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2}, h_{\mathbf{A}}(\mathbf{u}, \mathbf{v}) \right\rangle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2} \tag{25}$$

which imply,

$$h_{\mathbf{A}}(\mathbf{u},\mathbf{v}) = \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, h_{\mathbf{A}}(\mathbf{u},\mathbf{v}) \right\rangle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2} + \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, h_{\mathbf{A}}(\mathbf{u},\mathbf{v}) \right\rangle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2} + g_{\mathbf{A}}(\mathbf{u},\mathbf{v}) \tag{26}$$

and

$$\begin{aligned} h_{\mathbf{A};J'}(\mathbf{u},\mathbf{v}) &= \mathcal{T}_{\mathsf{supp}(\mathbf{u})\cup\mathsf{supp}(\mathbf{v})\cup J'}\left(h_{\mathbf{A}}(\mathbf{u},\mathbf{v})\right) \\ &= \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, h_{\mathbf{A}}(\mathbf{u},\mathbf{v}) \right\rangle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2} + \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, h_{\mathbf{A}}(\mathbf{u},\mathbf{v}) \right\rangle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2} + g_{\mathbf{A};J'}(\mathbf{u},\mathbf{v}) \end{aligned} \tag{27}$$

where $g_{\mathbf{A};J'}(\mathbf{u},\mathbf{v}) = \mathcal{T}_{\mathsf{supp}(\mathbf{u})\cup\mathsf{supp}(\mathbf{v})\cup J'}\left(g_{\mathbf{A}}(\mathbf{u},\mathbf{v})\right)$. Note that Friedlander et al. (2021) similarly use such a decomposition to show their RAIC. (iv) Then, conditioned on $L_{\mathbf{u},\mathbf{v}} \in [(1-s)\frac{\theta_{\mathbf{u},\mathbf{v}} m}{\pi}, (1+s)\frac{\theta_{\mathbf{u},\mathbf{v}} m}{\pi}]$, the desired property in Eq. (20) is derived from Eq. (26) using the concentration inequalities provided by Lemma A.1 in Appendix A together with standard techniques, e.g., the triangle inequality. (v) A union bound extends Eq. (20) to hold uniformly over $\mathcal{C}_\tau \times \mathcal{C}_\tau$ and all $J' \subseteq [n]$, $|J'| \le 2k$, with high probability.

Step (c) takes a similar approach. (i) Let $\mathbf{u} \in \mathcal{C}_\tau$ be an arbitrary vector in the τ-net, and fix any $J' \subseteq [n]$, $|J'| \le 2k$. Recall that the desired property in Eq. (21) should hold for all $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma^n_k$. (ii) To ensure this uniform result over $\mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma^n_k$, construct a second net $\mathcal{D}_\tau(\mathbf{u}) \subseteq \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma^n_k$ such that for each $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma^n_k$, there exits a point $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$ such that $\mathsf{sign}(\mathbf{A}\mathbf{w}) = \mathsf{sign}(\mathbf{A}\mathbf{x})$. (iii) Let $\beta = \arccos(1-\frac{\tau^2}{2})$ be the angle associated with the distance τ, and define the random variable $M_{\beta,\mathbf{u}} = |\{\mathbf{A}^{(i)}, i \in [m] : \theta_{\mathbf{w},\mathbf{A}^{(i)}} \in$ Notice that the size of $\mathcal{D}_\tau(\mathbf{u})$ need not exceed $2^{M_{\beta,\mathbf{u}}}$. Moreover, for any $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma^n_k$, $\theta_{\mathbf{x},\mathbf{u}} \in [0,\beta]$, and the number $L_{\mathbf{x},\mathbf{u}}$ of points $\mathbf{A}^{(i)} \in \mathcal{A}$, $i \in [m]$, on which $\mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{x} \rangle)$ and $\mathsf{sign}(\langle \mathbf{A}^{(i)}, \mathbf{u} \rangle)$ disagree is upper bounded by $M_{\beta,\mathbf{u}}$, formally, $L_{\mathbf{x},\mathbf{u}} \le M_{\beta,\mathbf{u}}$ for every $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma^n_k$. (iv) By a Chernoff and union bound, the random variable $M_{\beta,\mathbf{u}}$ can be shown to be bounded from above by $M_{\beta,\mathbf{u}} \le \frac{4}{3}\tau m$ with high probability for every $\mathbf{u} \in \mathcal{C}_\tau$, and due to the above argument, this further implies $L_{\mathbf{x},\mathbf{u}} \le \frac{4}{3}\tau m$ for each $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma^n_k$. (v) Taking any $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$ and conditioning on $L_{\mathbf{x},\mathbf{u}}$, the norm of $h_{\mathbf{A};J'}(\mathbf{w},\mathbf{u})$ is bounded using an orthogonal decomposition analogous to that in step (b) and again applying the concentration inequalities in Lemma A.1 and standard techniques, such that $\|h_{\mathbf{A};J'}(\mathbf{w},\mathbf{u})\|_2 \le O(\tau)$. (vi) This bound is then extended to hold uniformly for all $\mathbf{u} \in \mathcal{C}_\tau$, $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$, and $J' \subseteq [n]$, $|J'| \le 2k$, by union bounding. (vii) Step (c) concludes by arguing that the uniform result from step (vi) suffices to ensure Eq. (21) holds uniformly for all $\mathbf{u} \in \mathcal{C}_\tau$, all $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u})$, and all $J' \subseteq [n]$, $|J'| \le 2k$, by observing that for each $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u})$, there exists $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$ such that $\|h_{\mathbf{A};J'}(\mathbf{x},\mathbf{u})\|_2 = \|h_{\mathbf{A};J'}(\mathbf{w},\mathbf{u})\|_2 \le O(\tau)$ due to the construction of the net $\mathcal{D}_\tau(\mathbf{u})$, and additionally, by the triangle inequality, $\|(\mathbf{x}-\mathbf{u}) - h_{\mathbf{A};J'}(\mathbf{x},\mathbf{u})\|_2 \le \|\mathbf{x}-\mathbf{u}\|_2 + \|h_{\mathbf{A};J'}(\mathbf{x},\mathbf{u})\|_2 \le O(\tau)$.

4 Proof of the Main Result—BIHT Convergence

4.1 Intermediate Results

Before proving the main theorems, Theorem 3.1 and 3.2, three intermediate results, in Lemmas 4.1-4.3, are presented to facilitate the analysis for the convergence of BIHT approximations. The proofs for these intermediate results are in Section 4.3.

Lemma 4.1. *Consider any $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma^n_k$ and any $t \in \mathbb{Z}_+$. The error of the t^{th} approximation produced by the BIHT algorithm satisfies*

$$d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\right) \le 4 \left\| \left(\mathbf{x} - \hat{\mathbf{x}}^{(t-1)}\right) - h_{\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}\right) \right\|_2. \tag{28}$$

Note that Lemma 4.1 is a deterministic result, arsing from the equation by which the BIHT algorithm computes its t^{th} approximations, $t \in \mathbb{Z}_+$. Hence, it hold for all $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma^n_k$ and all iterations $t \in \mathbb{Z}_+$.

Lemma 4.2. *Let $\varepsilon : \mathbb{Z}_{\ge 0} \to \mathbb{R}$ be a function given by the recurrence relation*

$$\varepsilon(0) = 2 \tag{29}$$

$$\varepsilon(t) = 4c_1\sqrt{\frac{\epsilon}{c}\varepsilon(t-1)} + 4c_2\frac{\epsilon}{c}, \quad t \in \mathbb{Z}_+ \tag{30}$$

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

$$\lim_{t\to\infty} \varepsilon(t) = \left(2c_1\left(c_1 + \sqrt{c_1^2 + c_2}\right) + c_2\right)\frac{4\epsilon}{c} < \epsilon \tag{31}$$

Lemma 4.3. *Let $\varepsilon : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ be the function as defined in Lemma 4.2. Then, the sequence $\{\varepsilon(t)\}_{t\in\mathbb{Z}_{\geq 0}}$ is bound from above by the sequence $\{2^{2^{-t}}\epsilon^{1-2^{-t}}\}_{t\in\mathbb{Z}_{\geq 0}}$.*

4.2 Proofs of Theorems 3.2 and 3.1

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

Theorem (restatement) (Theorem 3.1)**.** *Let $a, b, c > 0$ be universal constants as in Eq. (3). Fix $\epsilon, \rho \in (0, 1)$ and $k, m, n \in \mathbb{Z}_+$ where*

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

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

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

for all $t \in \mathbb{Z}_{>0}$.

Corollary (restatement) (Corollary 3.2)**.** *Under the conditions stated in Theorem 3.1, uniformly with probability at least $1-\rho$, for all unknown k-sparse, real-valued unit vectors $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, the sequence of BIHT approximations, $\{\hat{\mathbf{x}}^{(t)}\}_{t\in\mathbb{Z}_{\geq 0}}$, converges asymptotically into the ϵ-ball around the unknown vector $\mathbf{x}$. Formally,*

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

Proof (Theorem 3.1 and Corollary 3.2). The BIHT approximations for an arbitrary unknown, k-sparse unit vector, $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, will be shown to converge as claimed in the theorems by applying the main technical theorem, Theorem 3.3, and the intermediate lemmas, Lemmas 4.1-4.3. Recalling that Theorem 3.3 and Lemma 4.1 hold uniformly over $\mathcal{S}^{n-1} \cap \Sigma_k^n$, the argument then implies uniform convergence for all $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$.

Consider any unknown, k-sparse unit vector $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$ with an associated sequence of BIHT approximations $\{\hat{\mathbf{x}}^{(t)} \in \mathcal{S}^{n-1} \cap \Sigma_k^n\}_{t\in\mathbb{Z}_{\geq 0}}$. For each $t \in \mathbb{Z}_+$, Lemma 4.1 bounds the error of the t^{th} approximation from above by

$$d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\right) \leq 4\left\|\left(\mathbf{x} - \hat{\mathbf{x}}^{(t-1)}\right) - h_{\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t-1)})}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}\right)\right\|_2 \tag{32}$$

which is further bounded by Theorem 3.3—by setting $\delta = \frac{\epsilon}{c} = \frac{\epsilon}{32}$ in the theorem—as

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

$$\leq 4\left(c_1\sqrt{\frac{\epsilon}{c}d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}\right)} + c_2\frac{\epsilon}{c}\right) \tag{33b}$$

$$= 4c_1 \sqrt{\frac{\epsilon}{c} d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)} \right)} + 4c_2 \frac{\epsilon}{c} \tag{33c}$$

where in the case of $t = 1$, (33c),

$$d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(1)} \right) \leq 4c_1 \sqrt{\frac{\epsilon}{c} d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(0)} \right)} + 4c_2 \frac{\epsilon}{c} \leq 4c_1 \sqrt{\frac{\epsilon}{c} d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, -\mathbf{x} \right)} + 4c_2 \frac{\epsilon}{c} = c_1 \sqrt{2\epsilon} + \frac{c_2}{4} \epsilon. \tag{34}$$

Recall that Lemma 4.2 defines a function $\varepsilon : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ by the recurrence relation

$$\varepsilon(0) = 2 \tag{35}$$

$$\varepsilon(t) = 4c_1 \sqrt{\frac{\epsilon}{c} \varepsilon(t-1)} + 4c_2 \frac{\epsilon}{c}, \quad t \in \mathbb{Z}_+ \tag{36}$$

whose form is similar to (33c). It can be argued inductively that for every $t \in \mathbb{Z}_{\geq 0}$, the function $\varepsilon(t)$ upper bounds the error of the t^{th} BIHT approximation, $d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t)} \right)$, as discussed next. The base case, $t = 0$, is trivial since

$$d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(0)} \right) \leq d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, -\mathbf{x} \right) = 2 = \varepsilon(0). \tag{37}$$

On the other hand, supposing that for each $t' \in [t-1]$, $t \in \mathbb{Z}_+$, the error is upper bounded by

$$d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t')} \right) \leq \varepsilon(t'), \tag{38}$$

the t^{th} approximation satisfies

$$d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t)} \right) \leq 4c_1 \sqrt{\frac{\epsilon}{c} d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)} \right)} + 4c_2 \frac{\epsilon}{c} \leq 4c_1 \sqrt{\frac{\epsilon}{c} \varepsilon(t-1)} + 4c_2 \frac{\epsilon}{c} = \varepsilon(t). \tag{39}$$

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

$$d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t)} \right) \leq \varepsilon(t), \quad \forall\, t \in \mathbb{Z}_{\geq 0}. \tag{40}$$

Then, Lemmas 4.2 and 4.3 immediately imply the desired result since asymptotically (Lemma 4.2),

$$\lim_{t \to \infty} d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t)} \right) \leq \lim_{t \to \infty} \varepsilon(t) = \left(2c_1 \left(c_1 + \sqrt{c_1^2 + c_2} \right) + c_2 \right) \frac{4\epsilon}{c} < \epsilon \tag{41}$$

and pointwise (Lemma 4.3),

$$d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t)} \right) \leq \varepsilon(t) \leq 2^{2^{-t}} \epsilon^{1-2^{-t}}. \tag{42}$$

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

Notice that in the argument laid out above, Lemma 4.1 and Theorem 3.3 hold uniformly for every $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, where Lemma 4.1 is deterministic while Theorem 3.3 ensures the bound with probability at least $1 - \rho$. Thus, for every $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, the t^{th} BIHT approximation has error upper bounded by

$$d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t)} \right) \leq 4c_1 \sqrt{\frac{\epsilon}{c} d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)} \right)} + 4c_2 \frac{\epsilon}{c} \tag{43}$$

uniformly with probability at least $1 - \rho$. Furthermore, because Lemmas 4.2 and 4.3 are deterministic, the rate of decay and asymptotic behavior stated in the theorems also hold uniformly, such that for all $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$,

$$\lim_{t \to \infty} d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t)} \right) \leq \lim_{t \to \infty} \varepsilon(t) = \left(2c_1 \left(c_1 + \sqrt{c_1^2 + c_2} \right) + c_2 \right) \frac{4\epsilon}{c} < \epsilon \tag{44}$$

$$d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t)} \right) \leq \varepsilon(t) \leq 2^{2^{-t}} \epsilon^{1-2^{-t}}, \quad \forall\, t \in \mathbb{Z}_{\geq 0} \tag{45}$$

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

4.3 Proof of the Intermediate Lemmas (Lemmas 4.1-4.3)

4.3.1 Proof of Lemma 4.1

Proof (Lemma 4.1). Let $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$ be an arbitrary unknown, k-spare vector of unit norm, and consider any t^{th} BIHT approximation, $\hat{\mathbf{x}}^{(t)} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, $t \in \mathbb{Z}_+$. Recall that the BIHT algorithm computes its t^{th} approximation by

$$\tilde{\mathbf{x}}^{(t)} = \hat{\mathbf{x}}^{(t-1)} + \frac{\eta}{m}\mathbf{A}^{\mathsf{T}} \cdot \frac{1}{2}\left(\mathsf{sign}\left(\mathbf{A}\mathbf{x}\right) - \mathsf{sign}\left(\mathbf{A}\hat{\mathbf{x}}^{(t-1)}\right)\right) \tag{46}$$

$$\hat{\mathbf{x}}^{(t)} = \frac{\mathcal{T}_k\left(\tilde{\mathbf{x}}^{(t)}\right)}{\left\|\mathcal{T}_k\left(\tilde{\mathbf{x}}^{(t)}\right)\right\|_2} \tag{47}$$

and notice that

$$\tilde{\mathbf{x}}^{(t)} = \hat{\mathbf{x}}^{(t-1)} + h_{\mathbf{A}}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}) \tag{48}$$

$$\mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t-1)}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) = \hat{\mathbf{x}}^{(t-1)} + h_{\mathbf{A};\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}). \tag{49}$$

Applying the triangle inequality, the error of the t^{th} BIHT approximation, $\hat{\mathbf{x}}^{(t)}$, can be bounded from above.

$$d_{\mathcal{S}^{n-1}}\left(\mathbf{x}, \hat{\mathbf{x}}^{(t)}\right) \tag{50a}$$

$$= \left\|\mathbf{x} - \hat{\mathbf{x}}^{(t)}\right\|_2 \tag{50b}$$

$$= \left\|\left(\mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right) + \left(\mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) - \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right) + \left(\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) - \hat{\mathbf{x}}^{(t)}\right)\right\|_2 \tag{50c}$$

$$\leq \left\|\mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right\|_2 + \left\|\mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) - \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right\|_2 + \left\|\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) - \hat{\mathbf{x}}^{(t)}\right\|_2$$

$$\blacktriangleright \text{ by the triangle inequality} \tag{50d}$$

$$= \left\|\mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right\|_2 + \left\|\mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) - \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right\|_2 \tag{50e}$$
$$+ \left\|\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) - \frac{\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)}{\left\|\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right\|_2}\right\|_2$$

The rightmost term in the last line can be upper bounded as follows.

$$\left\|\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) - \frac{\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)}{\left\|\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right\|_2}\right\|_2 \tag{51a}$$

$$= \left|\left\|\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right\|_2 - 1\right| \left\|\frac{\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)}{\left\|\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right\|_2}\right\|_2 \tag{51b}$$

$$= \left|\left\|\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right\|_2 - 1\right| \tag{51c}$$

$$= \left|\left\|\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right\|_2 - \|\mathbf{x}\|_2\right| \tag{51d}$$

$$\leq \left\|\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) - \mathbf{x}\right\|_2 \qquad \blacktriangleright \text{ by the triangle inequality} \tag{51e}$$

$$= \left\|\left(\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right) + \left(\mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) - \mathbf{x}\right)\right\|_2 \tag{51f}$$

$$\leq \left\|\mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right)\right\|_2 + \left\|\mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})}\left(\tilde{\mathbf{x}}^{(t)}\right) - \mathbf{x}\right\|_2 \tag{51g}$$

► by the triangle inequality

$$= \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 + \left\| \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) - \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 \tag{51h}$$

Combing (50e) and (51h) yields

$$d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t)} \right) = 2 \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 + 2 \left\| \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) - \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 . \tag{52}$$

Taking a closer look at the last term in (52),

$$\left\| \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) - \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 = \left\| \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \setminus \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 \leq \left\| \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)}) \setminus \mathsf{supp}(\mathbf{x})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 \tag{53}$$

where the rightmost inequality follows from the definition of the thresholding operation $\mathcal{T}_k$, which ensures that for each $j \in \mathsf{supp}(\mathbf{x}) \setminus \mathsf{supp}(\hat{\mathbf{x}}^{(t)})$, the j^{th} entry of $\tilde{\mathbf{x}}^{(t)}$ satisfies $|\tilde{x}_j^{(t)}| \leq \min_{j' \in \mathsf{supp}(\hat{\mathbf{x}}^{(t)}) \setminus \mathsf{supp}(\mathbf{x})} |\tilde{x}_{j'}^{(t)}|$. Then, observe

$$\left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 = \sum_{j \in \mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(x_j - \tilde{x}_j^{(t)} \right)^2 \tag{54a}$$

$$= \sum_{j \in \mathsf{supp}(\hat{\mathbf{x}}^{(t)}) \setminus \mathsf{supp}(\mathbf{x})} \left(x_j - \tilde{x}_j^{(t)} \right)^2 + \sum_{j \in \mathsf{supp}(\mathbf{x})} \left(x_j - \tilde{x}_j^{(t)} \right)^2 \tag{54b}$$

$$= \sum_{j \in \mathsf{supp}(\hat{\mathbf{x}}^{(t)}) \setminus \mathsf{supp}(\mathbf{x})} \left(0 - \tilde{x}_j^{(t)} \right)^2 + \sum_{j \in \mathsf{supp}(\mathbf{x})} \left(x_j - \tilde{x}_j^{(t)} \right)^2 \tag{54c}$$

$$= \sum_{j \in \mathsf{supp}(\hat{\mathbf{x}}^{(t)}) \setminus \mathsf{supp}(\mathbf{x})} \left(\tilde{x}_j^{(t)} \right)^2 + \sum_{j \in \mathsf{supp}(\mathbf{x})} \left(x_j - \tilde{x}_j^{(t)} \right)^2 \tag{54d}$$

$$= \left\| \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)}) \setminus \mathsf{supp}(\mathbf{x})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 + \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 \tag{54e}$$

It follows that

$$\left\| \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)}) \setminus \mathsf{supp}(\mathbf{x})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 + \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 = \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 \tag{55a}$$

$$\longrightarrow \quad \left\| \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)}) \setminus \mathsf{supp}(\mathbf{x})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 = \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 - \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 \tag{55b}$$

$$\longrightarrow \quad \left\| \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)}) \setminus \mathsf{supp}(\mathbf{x})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 \leq \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 \tag{55c}$$

$$\longrightarrow \quad \left\| \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)}) \setminus \mathsf{supp}(\mathbf{x})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 \leq \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 \tag{55d}$$

Likewise,

$$\left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t-1)}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 \tag{56a}$$

$$= \sum_{j \in \mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t-1)}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(x_j - \tilde{x}_j^{(t)} \right)^2 \tag{56b}$$

$$= \sum_{j \in \mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(x_j - \tilde{x}_j^{(t)} \right)^2 + \sum_{j \in \mathsf{supp}(\hat{\mathbf{x}}^{(t-1)}) \setminus (\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)}))} \left(x_j - \tilde{x}_j^{(t)} \right)^2 \tag{56c}$$

$$= \left\| \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\mathbf{x} - \tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 + \left\| \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t-1)}) \setminus (\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)}))} \left(\mathbf{x} - \tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 \tag{56d}$$

$$\geq \left\| \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\mathbf{x} - \tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 \tag{56e}$$

$$= \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2^2 \tag{56f}$$

$$\longrightarrow \quad \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 \leq \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t-1)}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 \tag{56g}$$

Continuing from (52),

$$d_{\mathcal{S}^{n-1}} \left(\mathbf{x}, \hat{\mathbf{x}}^{(t)} \right) \tag{57a}$$

$$= 2 \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 + 2 \left\| \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) - \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 \tag{57b}$$

$$= 2 \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 + 2 \left\| \mathcal{T}_{\mathsf{supp}(\hat{\mathbf{x}}^{(t)}) \setminus \mathsf{supp}(\mathbf{x})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 \quad \blacktriangleright \text{ by Eq. (53)} \tag{57c}$$

$$\leq 4 \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 \quad \blacktriangleright \text{ by Eq. (55d)} \tag{57d}$$

$$\leq 4 \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t-1)}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\tilde{\mathbf{x}}^{(t)} \right) \right\|_2 \quad \blacktriangleright \text{ by Eq. (56g)} \tag{57e}$$

$$= 4 \left\| \mathbf{x} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t-1)}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(\hat{\mathbf{x}}^{(t-1)} + h_{\mathbf{A}}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}) \right) \right\|_2 \tag{57f}$$

$$= 4 \left\| \mathbf{x} - \hat{\mathbf{x}}^{(t-1)} - \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t-1)}) \cup \mathsf{supp}(\hat{\mathbf{x}}^{(t)})} \left(h_{\mathbf{A}}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}) \right) \right\|_2 \tag{57g}$$

$$= 4 \left\| \left(\mathbf{x} - \hat{\mathbf{x}}^{(t-1)} \right) - h_{\mathbf{A}; \mathsf{supp}(\hat{\mathbf{x}}^{(t)})}(\mathbf{x}, \hat{\mathbf{x}}^{(t-1)}) \right\|_2 \tag{57h}$$

as desired. ■

4.3.2 Proof of Lemmas 4.2 and 4.3

Lemmas 4.2 and 4.3, will be verified in tandem. Fact 4.1, stated below and proved in Section C, will facilitate the proof.

Fact 4.1. *Let $u, v, w, w_0 \in \mathbb{R}_+$ such that $u = \frac{1}{2}\left(1 + \sqrt{1 + 4w}\right)$, and $1 \leq u \leq \frac{2}{\sqrt{v}}$. Define the functions $f_1, f_2 : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ by*

$$f_1(0) = 2 \tag{58}$$

$$f_1(t) = vw + \sqrt{vg(t-1)}, \quad t \in \mathbb{Z}_+ \tag{59}$$

$$f_2(t) = 2^{2^{-t}} (u^2 v)^{1-2^{-t}}, \quad t \in \mathbb{Z}_{\geq 0}. \tag{60}$$

Then, f_1 and f_2 are strictly monotonically decreasing and asymptotically converges to $u^2 v$. Moreover, f_2 pointwise upper bounds f_1. Formally,

$$f_1(t) \leq f_2(t), \quad \forall\, t \in \mathbb{Z}_{\geq 0} \tag{61}$$

$$\lim_{t \to \infty} f_2(t) = \lim_{t \to \infty} f_1(t) = u^2 v. \tag{62}$$

Lemma (restatement) (Lemma 4.2)**.** *Let $\varepsilon : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ be a function given by the recurrence relation*

$$\varepsilon(0) = 2$$

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

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

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

Lemma (restatement) (Lemma 4.3)**.** *Let $\varepsilon : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ be the function as defined in Lemma 4.2. Then, the sequence $\{\varepsilon(t)\}_{t \in \mathbb{Z}_{\geq 0}}$ is bound from above by the sequence $\{2^{2^{-t}} \epsilon^{1-2^{-t}}\}_{t \in \mathbb{Z}_{\geq 0}}$.*

Proof (Lemmas 4.2 and 4.3). The lemmas are corollaries to Fact 4.1. All that is necessary is writing ε in the form of f_1 in Fact 4.1 and verifying that it satisfies the conditions of the fact. For $t = 0$, $\varepsilon(0) = 2 = f_1(0)$. Otherwise, for $t > 0$, observe

$$\varepsilon(t) = 4c_1\sqrt{\frac{\epsilon}{c}\varepsilon(t-1)} + 4c_2\frac{\epsilon}{c} = \left(\frac{16c_1^2\epsilon}{c}\right)\left(\frac{16c_1^2\epsilon}{c}\right)^{-1} 4c_2\frac{\epsilon}{c} + \sqrt{\left(\frac{16c_1^2\epsilon}{c}\right)\varepsilon(t-1)} \tag{63a}$$

$$= \left(\frac{16c_1^2\epsilon}{c}\right)\left(\frac{c_2}{4c_1^2}\right) + \sqrt{\left(\frac{16c_1^2\epsilon}{c}\right)\varepsilon(t-1)} \tag{63b}$$

$$= vw + \sqrt{v\varepsilon(t-1)} \tag{63c}$$

where $v = \frac{16c_1^2\epsilon}{c}$, $w = \frac{c_2}{4c_1^2}$, and $u = \frac{1}{2}(1 + \sqrt{1 + 4\cdot\frac{c_2}{4c_1^2}}) = \frac{1}{2}(1 + \sqrt{1+\frac{c_2}{c_1^2}}) = \frac{1}{2c_1}(c_1 + \sqrt{c_1^2 + c_2})$. Recall that the universal constants are fixed as $c_1 = \sqrt{\frac{3\pi}{b}}\left(1 + \frac{16\sqrt{2}}{3}\right), c_2 = \frac{3}{b}\left(1 + \frac{4\pi}{3} + \frac{8\sqrt{3\pi}}{3} + 8\sqrt{6\pi}\right), c = 32$. By numerical calculations, it can be shown that $u\sqrt{v} < \sqrt{2}$ whenever $b \gtrsim 379.1038$, and hence $u < \sqrt{\frac{2}{v}}$, as required by Fact 4.1. It then follows that ε monotonically decreases with $t \in \mathbb{Z}_{\geq 0}$ and

$$\lim_{t\to\infty} \varepsilon(t) = u^2 v = \left(2c_1\left(c_1 + \sqrt{c_1^2 + c_2}\right) + c_2\right)\frac{4\epsilon}{c} < \frac{32\epsilon}{c} = \epsilon, \tag{64}$$

where the last inequality follows from a numerical calculation. Moreover, Fact 4.1 further implies

$$\varepsilon(t) \leq 2^{2^{-t}}(u^2v)^{1-2^{-t}} < 2^{2^{-t}}\epsilon^{1-2^{-t}}. \tag{65}$$

■

5 Outlook

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

References

Jayadev Acharya, Arnab Bhattacharyya, and Pritish Kamath. Improved bounds for universal one-bit compressive sensing. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 2353–2357. IEEE, 2017.

Richard G Baraniuk, Simon Foucart, Deanna Needell, Yaniv Plan, and Mary Wootters. Exponential decay of reconstruction error from binary measurements of sparse signals. *IEEE Transactions on Information Theory*, 63(6):3368–3385, 2017.

LE Blumenson. A derivation of n-dimensional spherical coordinates. *The American Mathematical Monthly*, 67(1):63–66, 1960.

Petros Boufounos and Richard G. Baraniuk. 1-bit compressive sensing. In *42nd Annual Conference on Information Sciences and Systems, CISS 2008, Princeton, NJ, USA, 19-21 March 2008*, pages 16–21. IEEE, 2008. ISBN 978-1-4244-2246-3. doi: 10.1109/CISS.2008.4558487. URL https://doi.org/10.1109/CISS.2008.4558487.

Petros T Boufounos, Laurent Jacques, Felix Krahmer, and Rayan Saab. Quantization and compressive sensing. In *Compressed sensing and its applications*, pages 193–237. Springer, 2015.

Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52 (2):489–509, 2006.

David L. Donoho. Compressed sensing. *IEEE Trans. Information Theory*, 52(4):1289–1306, 2006.

Larkin Flodin, Venkata Gandikota, and Arya Mazumdar. Superset technique for approximate recovery in one-bit compressed sensing. In *Advances in Neural Information Processing Systems*, pages 10387–10396, 2019.

Michael P Friedlander, Halyun Jeong, Yaniv Plan, and Özgür Yılmaz. Nbiht: An efficient algorithm for 1-bit compressed sensing with optimal error decay rate. *IEEE Transactions on Information Theory*, 68 (2):1157–1177, 2021.

Sivakant Gopi, Praneeth Netrapalli, Prateek Jain, and Aditya Nori. One-bit compressed sensing: Provable support and vector recovery. In *International Conference on Machine Learning*, pages 154–162, 2013.

Jarvis D. Haupt and Richard G. Baraniuk. Robust support recovery using sparse compressive sensing matrices. In *45st Annual Conference on Information Sciences and Systems, CISS 2011, The John Hopkins University, Baltimore, MD, USA, 23-25 March 2011*, pages 1–6. IEEE, 2011. ISBN 978-1-4244-9846-8. doi: 10.1109/CISS.2011.5766202. URL https://doi.org/10.1109/CISS.2011.5766202.

Laurent Jacques, Kévin Degraux, and Christophe De Vleeschouwer. Quantized iterative hard thresholding: Bridging 1-bit and high-resolution quantized compressed sensing. *arXiv preprint arXiv:1305.1786*, 2013a.

Laurent Jacques, Jason N Laska, Petros T Boufounos, and Richard G Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Transactions on Information Theory*, 59 (4):2082–2102, 2013b.

Laurent Jacques, Jason N Laska, Petros T Boufounos, and Richard G Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE Transactions on Information Theory*, 59 (4):2082–2102, 2013c.

Karin Knudson, Rayan Saab, and Rachel Ward. One-bit compressive sensing with norm estimation. *IEEE Transactions on Information Theory*, 62(5):2748–2758, 2016.

Ping Li. One scan 1-bit compressed sensing. In Arthur Gretton and Christian C. Robert, editors, *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016, Cadiz, Spain, May 9-11, 2016*, volume 51 of *JMLR Workshop and Conference Proceedings*, pages 1515–1523. JMLR.org, 2016. URL http://jmlr.org/proceedings/papers/v51/li16g.html.

Dekai Liu, Song Li, and Yi Shen. One-bit compressive sensing with projected subgradient method under sparsity constraints. *IEEE Transactions on Information Theory*, 65(10):6650–6663, 2019.

Arya Mazumdar and Soumyabrata Pal. Support recovery in universal one-bit compressed sensing. *arXiv preprint arXiv:2107.09091*, 2021.

Yaniv Plan and Roman Vershynin. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Trans. Information Theory*, 59(1):482–494, 2013a.

Yaniv Plan and Roman Vershynin. One-bit compressed sensing by linear programming. *Communications on Pure and Applied Mathematics*, 66(8):1275–1297, 2013b.

Yaniv Plan and Roman Vershynin. The generalized lasso with non-linear observations. *IEEE Transactions on information theory*, 62(3):1528–1537, 2016.

Yaniv Plan, Roman Vershynin, and Elena Yudovina. High-dimensional estimation with geometric constraints. *Information and Inference: A Journal of the IMA*, 6(1):1–40, 2017.

Rayan Saab, Rongrong Wang, and Özgür Yılmaz. Quantization of compressive samples with stable and robust recovery. *Applied and Computational Harmonic Analysis*, 44(1):123–143, 2018.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

A Proof of Theorem 3.3

This section proves the main technical theorem, Theorem 3.3, which is restated for convenience.

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

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

Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a measurement matrix whose rows $\mathbf{A}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n \times n})$, $i \in [m]$, have i.i.d. standard normal entries. Then, the measurement matrix $\mathbf{A}$ satisfies the (k, n, δ, c_1, c_2)-RAIC. Formally, uniformly with probability at least $1 - \rho$, for all $\mathbf{x}, \mathbf{y} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$ and all $J \subseteq [n]$, $|J| \leq k$,

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

The proof of the theorem will consider two regimes—the first, in Section A.1, looks at points which are at least distance $\frac{\delta}{b}$ apart, while the second, in Section A.2, handles points which are very close (less than distance $\frac{\delta}{b}$). Section A.3 then combines the two regimes to establish the theorem.

Before beginning the proof, let us introduce some notation and intermediate results. Recall the definition of $h_{\mathbf{A}} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,

$$h_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) = \sqrt{2\pi} \frac{1}{m} \mathbf{A}^\top \cdot \frac{1}{2} (\mathsf{sign}(\mathbf{A}\mathbf{x}) - \mathsf{sign}(\mathbf{A}\mathbf{y})) \tag{66}$$

$$h_{\mathbf{A};J}(\mathbf{x}, \mathbf{y}) = \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\mathbf{y}) \cup J}(h_{\mathbf{A}}(\mathbf{x}, \mathbf{y})) \tag{67}$$

and further define

$$g_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) = h_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) - \left\langle \frac{\frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{y}}{\|\mathbf{y}\|_2}}{\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{y}}{\|\mathbf{y}\|_2} \right\|_2}, h_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) \right\rangle \frac{\frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{y}}{\|\mathbf{y}\|_2}}{\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} - \frac{\mathbf{y}}{\|\mathbf{y}\|_2} \right\|_2} - \left\langle \frac{\frac{\mathbf{x}}{\|\mathbf{x}\|_2} + \frac{\mathbf{y}}{\|\mathbf{y}\|_2}}{\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} + \frac{\mathbf{y}}{\|\mathbf{y}\|_2} \right\|_2}, h_{\mathbf{A}}(\mathbf{x}, \mathbf{y}) \right\rangle \frac{\frac{\mathbf{x}}{\|\mathbf{x}\|_2} + \frac{\mathbf{y}}{\|\mathbf{y}\|_2}}{\left\| \frac{\mathbf{x}}{\|\mathbf{x}\|_2} + \frac{\mathbf{y}}{\|\mathbf{y}\|_2} \right\|_2} \tag{68}$$

$$g_{\mathbf{A};J}(\mathbf{x}, \mathbf{y}) = \mathcal{T}_{\mathsf{supp}(\mathbf{x}) \cup \mathsf{supp}(\mathbf{y}) \cup J}(g_{\mathbf{A}}(\mathbf{x}, \mathbf{y})) \tag{69}$$

for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $J \subseteq [n]$. The first of the two following lemmas provides concentration inequalities related to these functions $h_{\mathbf{A}}$ and $g_{\mathbf{A}}$. The latter lemma characterizes the number of measurements which lie in an angularly defined subset of $\mathbb{R}^n$. Both lemmas are verified in Appendix B.

Lemma A.1. *Fix $\ell, t > 0$, $\mathbf{r} \in \{-1,0,1\}^m$, and $J \subseteq [n]$, such that $\|\mathbf{r}\|_0 = \ell > 0$ and $|J| \leq 2k$. Let $(\mathbf{u},\mathbf{v}) \in \mathcal{S}^{n-1} \cap \Sigma_k^n \times \mathcal{S}^{n-1} \cap \Sigma_k^n$ be an ordered pair of real-valued unit vectors, and define the random variables $\mathbf{R}_{\mathbf{u},\mathbf{v}} = (R_{1;\mathbf{u},\mathbf{v}},\ldots,R_{m;\mathbf{u},\mathbf{v}}) = \frac{1}{2}(\mathsf{sign}(\mathbf{Au}) - \mathsf{sign}(\mathbf{Av}))$ and $L_{\mathbf{u},\mathbf{v}} = \|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0$, and suppose $\mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}$ and $L_{\mathbf{u},\mathbf{v}} = \ell$. Then, conditioned on $\mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}$ and $L_{\mathbf{u},\mathbf{v}} = \ell$, the following concentration inequalities hold.*

$$\Pr\left(\left|\left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\right\rangle - \sqrt{\frac{\pi}{2}}\frac{\ell}{m}\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{\theta_{\mathbf{u},\mathbf{v}}}\right| \geq \frac{\ell t}{m}\middle| \mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell\right) \leq 2e^{-\frac{1}{2}\ell t^2} \tag{70}$$

$$\Pr\left(\left|\left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\right\rangle\right| \geq \frac{\ell t}{m}\middle| \mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell\right) \leq 2e^{-\frac{1}{2}\ell t^2} \tag{71}$$

$$\Pr\left(\left\|\frac{1}{\eta} g_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\right\|_2 \geq \frac{2\sqrt{2k\ell}}{m} + \frac{\ell t}{m}\middle| \mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell\right) \leq 2e^{-\frac{1}{8}\ell t^2} \tag{72}$$

Lemma A.2. *Fix $t \in (0,1)$, $\beta \in [0, \frac{\pi}{2}]$. Let $\mathbf{u} \in \mathbb{R}^n$, and define the random variable $M_{\beta,\mathbf{u}} = |\{\mathbf{A}^{(i)} \in \mathcal{A} : \theta_{\mathbf{u},\mathbf{A}^{(i)}} \in \left[\frac{\pi}{2} - \beta, \frac{\pi}{2} + \beta\right]\}|$. Then,*

$$\mu_{M_{\beta,\mathbf{u}}} = \mathbb{E}\left[M_{\beta,\mathbf{u}}\right] = \frac{2}{\pi}\beta m \tag{73}$$

and

$$\Pr\left(M_{\beta,\mathbf{u}} \notin \left[(1-t)\mu_{M_{\beta,\mathbf{u}}}, (1+t)\mu_{M_{\beta,\mathbf{u}}}\right]\right) \leq 2e^{-\frac{1}{3}\mu_{M_{\beta,\mathbf{u}}} t^2}. \tag{74}$$

Lastly, for the purposes of the proof, a τ-net $\mathcal{C}_\tau \subset \mathcal{S}^{n-1} \cap \Sigma_k^n$ over the set of k-sparse, real-valued unit vectors is designed as follows, where $\tau = \frac{\delta}{b}$ is defined to lighten the notation. For each $J \subseteq [n]$, $|J| \leq k$, let $\mathcal{C}_{\tau;J} \subset \mathcal{S}^{n-1} \cap \Sigma_k^n$ be a τ-net over the set $\{\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n : \mathsf{supp}(\mathbf{x}) = J\}$. Then, construct the τ-net $\mathcal{C}_\tau \subset \mathcal{S}^{n-1} \cap \Sigma_k^n$ as their union, $\mathcal{C}_\tau = \bigcup_{J \subseteq [n]:|J|\leq k} \mathcal{C}_{\tau;J}$. Note that $|\mathcal{C}_\tau| \leq \binom{n}{k}\left(\frac{3}{\tau}\right)^k 2^k = \binom{n}{k}\left(\frac{6}{\tau}\right)^k$ and $|\mathcal{C}_\tau \times \mathcal{C}_\tau| \leq \binom{n}{k}^2\left(\frac{3}{\tau}\right)^{2k} 2^{2k} = \binom{n}{k}^2\left(\frac{6}{\tau}\right)^{2k}$. This construction is consistent throughout Sections A.1-A.3.

A.1 "Large distances" regime

The first regime considers the RAIC for ordered pairs of points in the τ-net which are at least distance τ from each other. Lemma A.3 formalizes a uniform result in this regime.

Lemma A.3. *Let $b_1 > 0$ be a universal constant. Fix $\delta, \rho_1 \in (0,1)$, and let $\tau = \frac{\delta}{b}$. Uniformly with probability at least $1 - \rho_1$,*

$$\|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 \leq b_1\sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})} \tag{75}$$

for all $(\mathbf{u},\mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau$ satisfying $d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) \geq \tau$, and $J \subseteq [n]$, $|J| \leq 2k$.

Proof (Lemma A.3). Let $(\mathbf{u},\mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau$ be an arbitrary ordered pair of points in the τ-net whose distance is at least $d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) \geq \tau$. Similar to the approach by Friedlander et al. (2021) and seen in Plan and Vershynin (2016), the function $h_{\mathbf{A};J}$ can be orthogonally decomposed as

$$h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) = \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\right\rangle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2} + \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\right\rangle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2} + g_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \tag{76}$$

Combining (76) with the triangle inequality yields

$$\|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 \tag{77a}$$

$$= \left\|(\mathbf{u}-\mathbf{v}) - \left(\left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\right\rangle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2} + \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\right\rangle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2} + g_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\right)\right\|_2 \tag{77b}$$

$$\leq \left\|(\mathbf{u}-\mathbf{v}) - \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\right\rangle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}\right\|_2 + \left\|\left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\right\rangle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}\right\|_2 + \|g_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 \tag{77c}$$

▶ by the triangle inequality

$$= \left| \|\mathbf{u}-\mathbf{v}\|_2 - \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \right| \left\| \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2} \right\|_2 + \left| \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \right| \left\| \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2} \right\|_2 + \|g_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 \tag{77d}$$

$$= \left| \|\mathbf{u}-\mathbf{v}\|_2 - \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \right| + \left| \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \right| + \|g_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 \tag{77e}$$

Lemma A.1 provides the following concentration inequalities.

$$\Pr\left(\left| \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle - \sqrt{\frac{\pi}{2}} \frac{\ell_{\mathbf{u},\mathbf{v}}}{m} \frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{\theta_{\mathbf{u},\mathbf{v}}} \right| > \frac{\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell_{\mathbf{u},\mathbf{v}} \right) \leq 2e^{-\frac{1}{2}\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}^2} \tag{78}$$

$$\Pr\left(\left| \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \right| > \frac{\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell_{\mathbf{u},\mathbf{v}} \right) \leq 2e^{-\frac{1}{2}\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}^2} \tag{79}$$

$$\Pr\left(\left\| \frac{1}{\eta} g_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\|_2 > \frac{2\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m} + \frac{\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell_{\mathbf{u},\mathbf{v}} \right) \leq 2e^{-\frac{1}{8}\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}^2} \tag{80}$$

where $\mathbf{R}_{\mathbf{u},\mathbf{v}}$ and $L_{\mathbf{u},\mathbf{v}}$ are random variables defined as $\mathbf{R}_{\mathbf{u},\mathbf{v}} = (R_{1;\mathbf{u},\mathbf{v}}, \dots, R_{m;\mathbf{u},\mathbf{v}}) = \frac{1}{2}(\mathsf{sign}(\mathbf{Au}) - \mathsf{sign}(\mathbf{Av}))$ and $L_{\mathbf{u},\mathbf{v}} = \|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0$, and $\mathbf{r} \in \{-1,0,1\}^m$, $\ell_{\mathbf{u},\mathbf{v}} \in [m]$. Eq. (78) further implies

$$\Pr\left(\left| \left(\|\mathbf{u}-\mathbf{v}\|_2 - \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \right) - \left(\|\mathbf{u}-\mathbf{v}\|_2 - \sqrt{\frac{\pi}{2}} \frac{\eta\ell_{\mathbf{u},\mathbf{v}}}{m} \frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{\theta_{\mathbf{u},\mathbf{v}}} \right) \right| > \frac{\eta\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell_{\mathbf{u},\mathbf{v}} \right)$$
$$\leq 2e^{-\frac{1}{2}\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}^2} \tag{81}$$

while Eqs. (79) and (80) can be written

$$\Pr\left(\left| \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \right| > \frac{\eta\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell_{\mathbf{u},\mathbf{v}} \right) \leq 2e^{-\frac{1}{2}\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}^2} \tag{82}$$

$$\Pr\left(\|g_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 > \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m} + \frac{\eta\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell_{\mathbf{u},\mathbf{v}} \right) \leq 2e^{-\frac{1}{8}\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}^2} \tag{83}$$

It follows that given $L_{\mathbf{u},\mathbf{v}} = \ell_{\mathbf{u},\mathbf{v}}$, with probability at least $1 - 6e^{-\frac{1}{8}\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}^2}$, the following holds

$$\|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 \tag{84a}$$

$$\leq \left| \|\mathbf{u}-\mathbf{v}\|_2 - \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \right| + \left| \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \right| + \|g_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 \tag{84b}$$

$$\leq \left| \|\mathbf{u}-\mathbf{v}\|_2 - \sqrt{\frac{\pi}{2}} \frac{\eta\ell_{\mathbf{u},\mathbf{v}}}{m} \frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{\theta_{\mathbf{u},\mathbf{v}}} \right| + \frac{\eta\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} + \frac{\eta\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m} + \frac{\eta\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} \tag{84c}$$

$$= \left| d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) - \sqrt{\frac{\pi}{2}} \frac{\eta\ell_{\mathbf{u},\mathbf{v}}}{m} \frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{\theta_{\mathbf{u},\mathbf{v}}} \right| + \frac{3\eta\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m} \tag{84d}$$

$$= \left| 1 - \sqrt{\frac{\pi}{2}} \frac{\eta\ell_{\mathbf{u},\mathbf{v}}}{m} \frac{1}{\theta_{\mathbf{u},\mathbf{v}}} \right| d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) + \frac{3\eta\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m} \tag{84e}$$

Let us next get a handle on the random variable $L_{\mathbf{u},\mathbf{v}}$, which tallies up the number of sign differences between $\mathsf{sign}(\mathbf{Au})$ and $\mathsf{sign}(\mathbf{Av})$. Note that this is precisely the number of i^{th} measurements $\mathbf{A}^{(i)} \in \mathcal{A}$, $i \in [m]$, such that $\theta_{\mathbf{w},\mathbf{A}^{(i)}} \in [\frac{\pi}{2} - \frac{\theta_{\mathbf{u},\mathbf{v}}}{2}, \frac{\pi}{2} + \frac{\theta_{\mathbf{u},\mathbf{v}}}{2}]$, where $\mathbf{w} = \mathbf{u} - \mathbf{v}$. By Lemma A.2, the random variable $L_{\mathbf{u},\mathbf{v}}$ can be characterized expectation

$$\mathbb{E}[L_{\mathbf{u},\mathbf{v}}] = \frac{\theta_{\mathbf{u},\mathbf{v}} m}{\pi} \tag{85}$$

and the concentration inequality

$$\Pr\left(L_{\mathbf{u},\mathbf{v}} \notin \left[(1-s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi}, (1+s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi}\right]\right) \leq 2e^{-\frac{1}{3\pi}\theta_{\mathbf{u},\mathbf{v}}ms_{\mathbf{u},\mathbf{v}}}. \tag{86}$$

Thus far, it has been shown that for a given pair $(\mathbf{u},\mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau$, where $d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) \geq \tau$, with probability at least $1 - 6e^{-\frac{1}{8}\ell_{\mathbf{u},\mathbf{v}}t^2_{\mathbf{u},\mathbf{v}}} - 2e^{-\frac{1}{3\pi}\theta_{\mathbf{u},\mathbf{v}}ms_{\mathbf{u},\mathbf{v}}}$,

$$\|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 \leq \left|1 - \sqrt{\frac{\pi}{2}}\frac{\eta\ell_{\mathbf{u},\mathbf{v}}}{m}\frac{1}{\theta_{\mathbf{u},\mathbf{v}}}\right| d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) + \frac{3\eta\ell_{\mathbf{u},\mathbf{v}}t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m} \tag{87}$$

where $\ell_{\mathbf{u},\mathbf{v}} \in [(1-s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi}, (1+s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi}]$. Next, this result will be extended—via union bounding—to hold uniformly for over all pairs $(\mathbf{u},\mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau$ with $d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) \geq \tau$ and each $J \subseteq [n]$, $|J| \leq 2k$. Let $\rho'_1, \rho''_1 \in (0,1)$ such that $\rho'_1 + \rho''_1 = \rho_1$. For each pair $\mathbf{u},\mathbf{v} \in \mathcal{C}_\tau$ and every $J \subseteq [n]$, $|J| = 2k$, the parameters $s_{\mathbf{u},\mathbf{v}}$ and $t_{\mathbf{u},\mathbf{v}}$ should ensure

$$\Pr\left(\exists\, \mathbf{u},\mathbf{v} \in \mathcal{C}_\tau,\ d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) \geq \tau,\quad L_{\mathbf{u},\mathbf{v}} \notin \left[(1-s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi}, (1+s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi}\right]\right) \leq \rho'_1 \tag{88}$$

and

$$\Pr\left(\begin{array}{l} \exists\, (\mathbf{u},\mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau,\ d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) \geq \tau, \\ \exists\, J \subseteq [n],\ |J| \leq 2k, \\ \|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 \\ \quad > \left|1 - \sqrt{\frac{\pi}{2}}\frac{\eta\ell_{\mathbf{u},\mathbf{v}}}{m}\frac{1}{\theta_{\mathbf{u},\mathbf{v}}}\right| d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) + \frac{3\eta\ell_{\mathbf{u},\mathbf{v}}t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m} \end{array} \middle| L_{\mathbf{u},\mathbf{v}} = \ell_{\mathbf{u},\mathbf{v}} \in \left[(1 \pm s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi}\right]\right) \leq \rho''_1 \tag{89}$$

For the former, (88), observe,

$$\Pr\left(\exists\, \mathbf{u},\mathbf{v} \in \mathcal{C}_\tau,\ d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) \geq \tau,\quad L_{\mathbf{u},\mathbf{v}} \notin \left[(1-s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi}, (1+s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi}\right]\right) \leq \rho'_1 \tag{90a}$$

$$\longrightarrow \quad \binom{n}{k}^2 \left(\frac{6}{\tau}\right)^{2k} \Pr\left(L_{\mathbf{u},\mathbf{v}} \notin \left[(1-s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi}, (1+s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi}\right]\right) \leq \rho'_1 \tag{90b}$$

$$\longrightarrow \quad \binom{n}{k}^2 \left(\frac{6}{\tau}\right)^{2k} 2e^{-\frac{1}{3\pi}\theta_{\mathbf{u},\mathbf{v}}ms_{\mathbf{u},\mathbf{v}}} \leq \rho'_1 \tag{90c}$$

$$\longrightarrow \quad s_{\mathbf{u},\mathbf{v}} \geq \sqrt{\frac{3\pi}{\theta_{\mathbf{u},\mathbf{v}}m}\log\left(\binom{n}{k}^2\left(\frac{6}{\tau}\right)^{2k}\left(\frac{2}{\rho'_1}\right)\right)} \tag{90d}$$

Hence, the parameter is set as

$$s_{\mathbf{u},\mathbf{v}} = \sqrt{\frac{3\pi}{\theta_{\mathbf{u},\mathbf{v}}m}\log\left(\binom{n}{k}^2\left(\frac{6}{\tau}\right)^{2k}\left(\frac{2}{\rho'_1}\right)\right)} \in (0,1) \tag{91}$$

Then,

$$\ell_{\mathbf{u},\mathbf{v}} \leq (1+s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi} \leq \left(1 + \sqrt{\frac{3\pi}{\theta_{\mathbf{u},\mathbf{v}}m}\log\left(\binom{n}{k}^2\left(\frac{6}{\tau}\right)^{2k}\left(\frac{2}{\rho'_1}\right)\right)}\right)\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi} \leq \frac{2}{\pi}\theta_{\mathbf{u},\mathbf{v}}m. \tag{92}$$

On the other hand, using (89), $t_{\mathbf{u},\mathbf{v}}$ is determined as follows. Note that the number subsets $J \subseteq [n]$, $|J| \leq 2k$, is at most $\binom{n}{2k}2^{2k}$ (which will be used momentarily in a union bound), and then observe,

$$\Pr\left(\begin{array}{l} \exists\, (\mathbf{u},\mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau,\ d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) \geq \tau, \\ \exists\, J \subseteq [n],\ |J| \leq 2k, \\ \|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 \\ \quad > \left|1 - \sqrt{\frac{\pi}{2}}\frac{\eta\ell_{\mathbf{u},\mathbf{v}}}{m}\frac{1}{\theta_{\mathbf{u},\mathbf{v}}}\right| d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) + \frac{3\eta\ell_{\mathbf{u},\mathbf{v}}t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m} \end{array} \middle| L_{\mathbf{u},\mathbf{v}} = \ell_{\mathbf{u},\mathbf{v}} \in \left[(1 \pm s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}}m}{\pi}\right]\right) \leq \rho''_1 \tag{93a}$$

$$\longrightarrow \quad \binom{n}{k}^2 \left(\frac{6}{\tau}\right)^{2k} 2^{2k}\binom{n}{2k} 6e^{-\frac{1}{8}\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}^2} \leq \rho_1'' \tag{93b}$$

$$\longrightarrow \quad \binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} 6e^{-\frac{1}{8}\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}^2} \leq \rho_1'' \tag{93c}$$

$$\longrightarrow \quad t_{\mathbf{u},\mathbf{v}} \geq \sqrt{\frac{8}{\ell_{\mathbf{u},\mathbf{v}}} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{6}{\rho_1''}\right)\right)} \tag{93d}$$

Thus, the parameter can be set as

$$t_{\mathbf{u},\mathbf{v}} = \sqrt{\frac{8}{\ell_{\mathbf{u},\mathbf{v}}} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{6}{\rho_1''}\right)\right)}. \tag{94}$$

Note that

$$\frac{\ell_{\mathbf{u},\mathbf{v}}}{m} \leq (1+s_{\mathbf{u},\mathbf{v}})\frac{\theta_{\mathbf{u},\mathbf{v}} m}{\pi} \cdot \frac{1}{m} = \frac{(1+s_{\mathbf{u},\mathbf{v}})}{\pi}\theta_{\mathbf{u},\mathbf{v}} \leq \frac{2}{\pi}\theta_{\mathbf{u},\mathbf{v}} \tag{95}$$

and

$$\frac{\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} \leq \frac{\ell_{\mathbf{u},\mathbf{v}}}{m}\sqrt{\frac{8}{\ell_{\mathbf{u},\mathbf{v}}} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{6}{\rho_1''}\right)\right)} = \frac{1}{m}\sqrt{8\ell_{\mathbf{u},\mathbf{v}} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{6}{\rho_1''}\right)\right)} \tag{96a}$$

$$\leq \frac{1}{m}\sqrt{8 \cdot \frac{2}{\pi}\theta_{\mathbf{u},\mathbf{v}} m \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{6}{\rho_1''}\right)\right)} \tag{96b}$$

$$= \sqrt{\frac{16}{\pi}\frac{\theta_{\mathbf{u},\mathbf{v}}}{m} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{6}{\rho_1''}\right)\right)} \tag{96c}$$

$$\leq \sqrt{\frac{16}{3}\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{6}{\rho_1''}\right)\right)} \tag{96d}$$

$$\leq \frac{4}{\sqrt{3}} \cdot \sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{6}{\rho_1''}\right)\right)} \tag{96e}$$

In regard to the parameter $s_{\mathbf{u},\mathbf{v}}$, observe

$$s_{\mathbf{u},\mathbf{v}} d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) = d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})\sqrt{\frac{3\pi}{\theta_{\mathbf{u},\mathbf{v}} m} \log\left(\binom{n}{k}^2 \left(\frac{6}{\tau}\right)^{2k} \left(\frac{2}{\rho_1'}\right)\right)} \tag{97a}$$

$$\leq \sqrt{\frac{3\pi d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m} \log\left(\binom{n}{k}^2 \left(\frac{6}{\tau}\right)^{2k} \left(\frac{2}{\rho_1'}\right)\right)} \tag{97b}$$

Then, from the above discussion, with high probability, $\|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2$ is upper bounded as follows.

$$\|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2$$
$$\leq \left|1 - \sqrt{\frac{\pi}{2}}\frac{\eta\ell_{\mathbf{u},\mathbf{v}}}{m}\frac{1}{\theta_{\mathbf{u},\mathbf{v}}}\right| d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) + \frac{3\eta\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m}$$
$$\leq \left|1 - \sqrt{\frac{\pi}{2}}\eta\frac{(1+s_{\mathbf{u},\mathbf{v}})\theta_{\mathbf{u},\mathbf{v}}}{\pi}\frac{1}{\theta_{\mathbf{u},\mathbf{v}}}\right| d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) + \frac{3\eta\ell_{\mathbf{u},\mathbf{v}} t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m}$$

$$= \left|1-(1+s_{\mathbf{u},\mathbf{v}})\sqrt{\frac{\pi}{2}}\frac{\eta}{\pi}\right| d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) + \frac{3\eta\ell_{\mathbf{u},\mathbf{v}}t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m}$$

$$= \left|1-(1+s_{\mathbf{u},\mathbf{v}})\sqrt{\frac{\pi}{2}}\frac{\sqrt{2\pi}}{\pi}\right| d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) + \frac{3\eta\ell_{\mathbf{u},\mathbf{v}}t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m}$$

$$= |1-(1+s_{\mathbf{u},\mathbf{v}})|\, d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) + \frac{3\eta\ell_{\mathbf{u},\mathbf{v}}t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m}$$

$$= s_{\mathbf{u},\mathbf{v}} d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) + \frac{3\eta\ell_{\mathbf{u},\mathbf{v}}t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m}$$

$$\le \sqrt{\frac{3\pi d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\left(\frac{6}{\tau}\right)^{2k}\left(\frac{2}{\rho_1'}\right)\right)} + \frac{3\eta\ell_{\mathbf{u},\mathbf{v}}t_{\mathbf{u},\mathbf{v}}}{m} + \frac{2\eta\sqrt{2k\ell_{\mathbf{u},\mathbf{v}}}}{m}$$

$$\le \sqrt{\frac{3\pi d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\left(\frac{6}{\tau}\right)^{2k}\left(\frac{2}{\rho_1'}\right)\right)} + \frac{12\eta}{\sqrt{3}}\cdot\sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12}{\tau}\right)^{2k}\left(\frac{6}{\rho_1''}\right)\right)}$$

$$+ 2\eta\sqrt{\frac{2k}{m}\cdot\frac{2}{\pi}\theta_{\mathbf{u},\mathbf{v}}}$$

$$= \sqrt{\frac{3\pi d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\left(\frac{6}{\tau}\right)^{2k}\left(\frac{2}{\rho_1'}\right)\right)} + 4\sqrt{3}\eta\cdot\sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12}{\tau}\right)^{2k}\left(\frac{6}{\rho_1''}\right)\right)}$$

$$+ 4\eta\sqrt{\frac{k}{m}\cdot\frac{1}{\pi}\theta_{\mathbf{u},\mathbf{v}}}$$

$$\le \sqrt{\frac{3\pi d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\left(\frac{6}{\tau}\right)^{2k}\left(\frac{2}{\rho_1'}\right)\right)} + 4\sqrt{3}\eta\cdot\sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12}{\tau}\right)^{2k}\left(\frac{6}{\rho_1''}\right)\right)}$$

$$+ \frac{4\eta}{\sqrt{3}}\sqrt{\frac{k d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}}$$

$$= \sqrt{3\pi}\cdot\sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\left(\frac{6}{\tau}\right)^{2k}\left(\frac{2}{\rho_1'}\right)\right)} + 4\sqrt{6\pi}\cdot\sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12}{\tau}\right)^{2k}\left(\frac{6}{\rho_1''}\right)\right)}$$

$$+ \frac{4\sqrt{3}\eta}{3}\cdot\sqrt{\frac{k d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}}$$

$$\le \sqrt{3\pi}\cdot\sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12}{\tau}\right)^{2k}\left(\frac{2}{\rho_1'}\right)\right)} + 4\sqrt{6\pi}\cdot\sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12}{\tau}\right)^{2k}\left(\frac{6}{\rho_1''}\right)\right)}$$

$$+ \frac{4\sqrt{3}\eta}{3}\cdot\sqrt{\frac{k d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}}$$

$$= \sqrt{3\pi}\cdot\sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12}{\tau}\right)^{2k}\left(\frac{8}{\rho_1}\right)\right)} + 4\sqrt{6\pi}\cdot\sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12}{\tau}\right)^{2k}\left(\frac{8}{\rho_1}\right)\right)}$$

$$+ \frac{4\sqrt{3}\eta}{3}\cdot\sqrt{\frac{k d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}}$$

► set $\rho_1' = \frac{1}{4}\rho_1,\ \rho_1'' = \frac{3}{4}\rho_1$

$$\le \sqrt{3\pi}\cdot\sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12}{\tau}\right)^{2k}\left(\frac{8}{\rho_1}\right)\right)} + 4\sqrt{6\pi}\cdot\sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m}\log\left(\binom{n}{k}^2\binom{n}{2k}\left(\frac{12}{\tau}\right)^{2k}\left(\frac{8}{\rho_1}\right)\right)}$$

$$
\begin{aligned}
&+ \frac{4\sqrt{6\pi}}{3} \cdot \sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{8}{\rho_1}\right)\right)} \\
&= \left(\sqrt{3\pi} + 4\sqrt{6\pi} + \frac{4\sqrt{6\pi}}{3}\right) \sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{8}{\rho_1}\right)\right)} \\
&= \left(\sqrt{3\pi} + \frac{16\sqrt{6\pi}}{3}\right) \sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{8}{\rho_1}\right)\right)} \\
&= \sqrt{3\pi}\left(1 + \frac{16\sqrt{2}}{3}\right) \sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{8}{\rho_1}\right)\right)} \\
&\leq \sqrt{3\pi}\left(1 + \frac{16\sqrt{2}}{3}\right) \sqrt{\frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{m} \log\left(\binom{n}{k}^2 \binom{n}{2k} \left(\frac{12}{\tau}\right)^{2k} \left(\frac{a}{\rho}\right)\right)} \\
&\leq \sqrt{3\pi}\left(1 + \frac{16\sqrt{2}}{3}\right) \sqrt{\tau d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})} \\
&= \sqrt{3\pi}\left(1 + \frac{16\sqrt{2}}{3}\right) \sqrt{\frac{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{b}} \\
&= \sqrt{\frac{3\pi}{b}}\left(1 + \frac{16\sqrt{2}}{3}\right) \sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}
\end{aligned}
$$

In short, the above step yields

$$
\left\|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\right\|_2 \leq b_1 \sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})} \tag{99}
$$

where the universal constant is set as

$$
b_1 = \sqrt{\frac{3\pi}{b}}\left(1 + \frac{16\sqrt{2}}{3}\right). \tag{100}
$$

Then, the lemma's universal result follows—with probability at least $1 - \rho_1$,

$$
\left\|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\right\|_2 \leq b_1 \sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})} \tag{101}
$$

uniformly for all $(\mathbf{u},\mathbf{v}) \in \mathcal{C}_\tau \times \mathcal{C}_\tau$, $d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) \geq \tau$, and all $J \subseteq [n]$, $|J| \leq 2k$. ■

A.2 "Small distances" regime

In contrast to the regime in Section A.1, the regime under consideration in this section looks at points in the τ-ball around each point in the τ-net, $\mathcal{C}_\tau$. Lemma A.4 states the formal result.

Lemma A.4. *Let $b_2 > 0$ be a universal constant. Fix $\delta, \rho_2 \in (0,1)$, and let $\tau = \frac{\delta}{b}$. Uniformly with probability at least $1 - \rho_2$,*

$$
\left\|(\mathbf{x}-\mathbf{u}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{u})\right\|_2 \leq b_2 \delta \tag{102}
$$

for all $\mathbf{u} \in \mathcal{C}_\tau$, for all $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$, and for all $J \subseteq [n]$, $|J| \leq 2k$.

Proof (Lemma A.4). To motivate the approach taken in this proof, consider an arbitrary point $\mathbf{u} \in \mathcal{C}_\tau$ of the τ-net and any $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$. Using the triangle inequality and then the membership of $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$—which implies $\|\mathbf{x}-\mathbf{u}\|_2 \leq \tau$—it follows that

$$
\left\|(\mathbf{x}-\mathbf{u}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{u})\right\|_2 \leq \|\mathbf{x}-\mathbf{u}\|_2 + \|h_{\mathbf{A};J}(\mathbf{x},\mathbf{u})\|_2 \leq \tau + \|h_{\mathbf{A};J}(\mathbf{x},\mathbf{u})\|_2 \tag{103}
$$

Hence, the primary task in proving the lemma is controlling the rightmost term in (103), $\|h_{\mathbf{A};J}(\mathbf{x},\mathbf{u})\|_2$.

Towards this, consider the set of k-sparse, real-valued unit vectors, $\mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n \subseteq \mathcal{S}^{n-1} \cap \Sigma_k^n$, within distance-$\tau$ of $\mathbf{u}$, and let $\mathcal{Y}_\tau(\mathbf{u}) = \{\mathsf{sign}(\mathbf{Aw}) : \mathbf{w} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n\}$. Construct a net $\mathcal{D}_\tau(\mathbf{u}) \subseteq \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$ over the set of points $\mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$ such that for every distinct $\mathbf{y} \in \mathcal{Y}_\tau(\mathbf{u})$, the net $\mathcal{D}_\tau(\mathbf{u})$ contains exactly one point $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$ such that $\mathsf{sign}(\mathbf{Aw}) = \mathbf{y}$. By this construction of $\mathcal{D}_\tau(\mathbf{u})$, for any $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$, there exits $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$ such that $\mathsf{sign}(\mathbf{Aw}) = \mathsf{sign}(\mathbf{Ax})$. Noticing that the dependency of $h_{\mathbf{A};J}(\mathbf{x},\mathbf{u})$ on $\mathbf{x}$ is limited to the its dependence on $\mathsf{sign}(\mathbf{Ax})$, it follows that $h_{\mathbf{A};J}(\mathbf{x},\mathbf{u}) = h_{\mathbf{A};J}(\mathbf{w},\mathbf{u})$. Hence, it suffices to upper bound $\|h_{\mathbf{A};J}(\mathbf{w},\mathbf{u})\|_2$ uniformly over every $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$. Next, such a uniform bound is derived.

Fix any $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$. As in the proof of Lemma A.3, the function $h_{\mathbf{A};J}$ can be expressed using orthogonal projections as

$$h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) = \left\langle \frac{\mathbf{w}-\mathbf{u}}{\|\mathbf{w}-\mathbf{u}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \frac{\mathbf{w}-\mathbf{u}}{\|\mathbf{w}-\mathbf{u}\|_2} + \left\langle \frac{\mathbf{w}+\mathbf{u}}{\|\mathbf{w}+\mathbf{u}\|_2}, h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\rangle \frac{\mathbf{w}+\mathbf{v}}{\|\mathbf{w}+\mathbf{u}\|_2} + g_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \tag{104}$$

and by the triangle inequality

$$\|h_{\mathbf{A};J}(\mathbf{w},\mathbf{u})\|_2 \tag{105a}$$
$$= \left\| \left\langle \frac{\mathbf{w}-\mathbf{u}}{\|\mathbf{w}-\mathbf{u}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \frac{\mathbf{w}-\mathbf{u}}{\|\mathbf{w}-\mathbf{u}\|_2} + \left\langle \frac{\mathbf{x}+\mathbf{u}}{\|\mathbf{w}+\mathbf{u}\|_2}, h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\rangle \frac{\mathbf{w}+\mathbf{v}}{\|\mathbf{w}+\mathbf{u}\|_2} + g_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\|_2 \tag{105b}$$
$$\leq \left\| \left\langle \frac{\mathbf{w}-\mathbf{u}}{\|\mathbf{w}-\mathbf{u}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \frac{\mathbf{w}-\mathbf{u}}{\|\mathbf{w}-\mathbf{u}\|_2} \right\|_2 + \left\| \left\langle \frac{\mathbf{w}+\mathbf{u}}{\|\mathbf{w}+\mathbf{u}\|_2}, h_{\mathbf{A};J}(\mathbf{x},\mathbf{u}) \right\rangle \frac{\mathbf{w}+\mathbf{v}}{\|\mathbf{w}+\mathbf{u}\|_2} \right\|_2 + \|g_{\mathbf{A};J}(\mathbf{w},\mathbf{u})\|_2 \tag{105c}$$
$$= \left| \left\langle \frac{\mathbf{w}-\mathbf{u}}{\|\mathbf{w}-\mathbf{u}\|_2}, h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) \right\rangle \right| + \left| \left\langle \frac{\mathbf{w}+\mathbf{u}}{\|\mathbf{w}+\mathbf{u}\|_2}, h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\rangle \right| + \|g_{\mathbf{A};J}(\mathbf{w},\mathbf{u})\|_2 \tag{105d}$$

Recall the concentration inequalities provided in Lemma A.1.

$$\Pr\left(\left| \left\langle \frac{\mathbf{w}-\mathbf{u}}{\|\mathbf{w}-\mathbf{u}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\rangle - \sqrt{\frac{\pi}{2}} \frac{\ell_{\mathbf{w},\mathbf{u}}}{m} \frac{d_{\mathcal{S}^{n-1}}(\mathbf{w},\mathbf{u})}{\theta_{\mathbf{w},\mathbf{u}}} \right| \geq \frac{\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}}{m} \middle| \mathbf{R}_{\mathbf{w},\mathbf{u}} = \mathbf{r}, L_{\mathbf{w},\mathbf{u}} = \ell_{\mathbf{w},\mathbf{u}} \right) \leq 2e^{-\frac{1}{2}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2} \tag{106}$$

$$\Pr\left(\left| \left\langle \frac{\mathbf{w}+\mathbf{u}}{\|\mathbf{w}+\mathbf{u}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\rangle \right| \geq \frac{\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}}{m} \middle| \mathbf{R}_{\mathbf{w},\mathbf{u}} = \mathbf{r}, L_{\mathbf{w},\mathbf{u}} = \ell_{\mathbf{w},\mathbf{u}} \right) \leq 2e^{-\frac{1}{2}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2} \tag{107}$$

$$\Pr\left(\left\| \frac{1}{\eta} g_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\|_2 \geq \frac{2\sqrt{2k\ell_{\mathbf{w},\mathbf{u}}}}{m} + \frac{\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}}{m} \middle| \mathbf{R}_{\mathbf{w},\mathbf{u}} = \mathbf{r}, L_{\mathbf{w},\mathbf{u}} = \ell_{\mathbf{w},\mathbf{u}} \right) \leq 2e^{-\frac{1}{8}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2} \tag{108}$$

where $\mathbf{R}_{\mathbf{w},\mathbf{u}}$ and $L_{\mathbf{w},\mathbf{u}}$ are random variables defined as $\mathbf{R}_{\mathbf{w},\mathbf{u}} = (R_{1;\mathbf{w},\mathbf{u}}, \dots, R_{m;\mathbf{w},\mathbf{u}}) = \frac{1}{2}(\mathsf{sign}(\mathbf{Aw}) - \mathsf{sign}(\mathbf{Au}))$ and $L_{\mathbf{w},\mathbf{u}} = \|\mathbf{R}_{\mathbf{w},\mathbf{u}}\|_0$, and $\mathbf{r} \in \{-1,0,1\}^m$, $\ell_{\mathbf{w},\mathbf{u}} \in [m]$. Eq. (106) can be replaced by

$$\Pr\left(\left\langle \frac{\mathbf{w}-\mathbf{u}}{\|\mathbf{w}-\mathbf{u}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\rangle \geq \sqrt{\frac{\pi}{2}} \frac{\ell_{\mathbf{w},\mathbf{u}}}{m} \frac{d_{\mathcal{S}^{n-1}}(\mathbf{w},\mathbf{u})}{\theta_{\mathbf{w},\mathbf{u}}} + \frac{\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}}{m} \middle| \mathbf{R}_{\mathbf{w},\mathbf{u}} = \mathbf{r}, L_{\mathbf{w},\mathbf{u}} = \ell_{\mathbf{w},\mathbf{u}} \right) \leq e^{-\frac{1}{2}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2} \tag{109a}$$

$$\longrightarrow \quad \Pr\left(\left\langle \frac{\mathbf{w}-\mathbf{u}}{\|\mathbf{w}-\mathbf{u}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\rangle \geq \sqrt{\frac{\pi}{2}} \frac{\ell_{\mathbf{w},\mathbf{u}}}{m} + \frac{\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}}{m} \middle| \mathbf{R}_{\mathbf{w},\mathbf{u}} = \mathbf{r}, L_{\mathbf{w},\mathbf{u}} = \ell_{\mathbf{w},\mathbf{u}} \right) \leq e^{-\frac{1}{2}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2} \tag{109b}$$

$$\longrightarrow \quad \Pr\left(\left\langle \frac{\mathbf{w}-\mathbf{u}}{\|\mathbf{w}-\mathbf{u}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\rangle \geq \left(\sqrt{\frac{\pi}{2}} + t_{\mathbf{w},\mathbf{u}} \right) \frac{\ell_{\mathbf{w},\mathbf{u}}}{m} \middle| \mathbf{R}_{\mathbf{w},\mathbf{u}} = \mathbf{r}, L_{\mathbf{w},\mathbf{u}} = \ell_{\mathbf{w},\mathbf{u}} \right) \leq e^{-\frac{1}{2}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2} \tag{109c}$$

Due to the conditioning in the above concentration bounds, we will need to have a handle on the random variable $L_{\mathbf{w},\mathbf{u}} = \|\mathbf{R}_{\mathbf{w},\mathbf{u}}\|_0$. Note that the random variable $R_{i;\mathbf{w},\mathbf{u}}$, $i \in [m]$, takes a nonzero value precisely when $\mathsf{sign}(\langle \mathbf{w}, \mathbf{A}^{(i)} \rangle) \neq \mathsf{sign}(\langle \mathbf{u}, \mathbf{A}^{(i)} \rangle)$. However, because $d_{\mathcal{S}^{n-1}}(\mathbf{w}, \mathbf{u}) \leq \tau$, this sign difference can only occur for the i^{th} points $\mathbf{A}^{(i)} \in \mathcal{A}$, $i \in [m]$, whose angular distance from $\mathbf{u}$ is in the range $[\frac{\pi}{2} - \beta, \frac{\pi}{2} + \beta]$, where $\beta = \arccos\left(1 - \frac{\tau^2}{2}\right)$ is the angular distance associated with the distance τ. In light of this, define the random variable $M_{\beta,\mathbf{u}} = |\{\mathbf{A}^{(i)} \in \mathcal{A} : \theta_{\mathbf{u},\mathbf{A}^{(i)}} \in \left[\frac{\pi}{2} - \beta, \frac{\pi}{2} + \beta\right]\}|$. By Lemma A.2,

$$\mathbb{E}\left[M_{\beta,\mathbf{u}}\right] = \frac{2}{\pi}\beta m \tag{110}$$

and for $s \in (0, 1)$,

$$\Pr\left(M_{\beta,\mathbf{u}} > (1+s)\frac{2}{\pi}\beta m\right) \leq e^{-\frac{2}{3\pi}\beta m s^2} \tag{111a}$$

$$\longrightarrow \quad \Pr\left(M_{\beta,\mathbf{u}} > \frac{4}{\pi}\beta m\right) \leq e^{-\frac{8}{3\pi}\beta m} \tag{111b}$$

$$\longrightarrow \quad \Pr\left(M_{\beta,\mathbf{u}} > \frac{4}{\pi}\beta m\right) \leq e^{-\frac{8}{3\pi}\tau m} \tag{111c}$$

$$\longrightarrow \quad \Pr\left(M_{\beta,\mathbf{u}} > \frac{4}{3}\tau m\right) \leq e^{-\frac{8}{3\pi}\tau m}. \tag{111d}$$

It follows that with probability at least $1 - e^{-\frac{8}{3\pi}\tau m}$, the size of the net $\mathcal{D}_\tau(\mathbf{u})$ does not exceed $|\mathcal{D}_\tau(\mathbf{u})| \leq 2^{\frac{4}{3}\tau m}$. Later, this observation will be used to union bound over $\mathcal{D}_\tau(\mathbf{u})$.

This completes the necessary preparation for deriving a uniform bound on $\|h_{\mathbf{A};J}(\mathbf{w}, \mathbf{u})\|_2$. Let us first summarize the relevant concentration inequalities, which are

$$\Pr\left(\left\langle \frac{\mathbf{w} - \mathbf{u}}{\|\mathbf{w} - \mathbf{u}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{w}, \mathbf{u}) \right\rangle \geq \left(\sqrt{\frac{\pi}{2}} + t_{\mathbf{w},\mathbf{u}}\right) \frac{\ell_{\mathbf{w},\mathbf{u}}}{m} \middle| \mathbf{R}_{\mathbf{w},\mathbf{u}} = \mathbf{r}, L_{\mathbf{w},\mathbf{u}} = \ell_{\mathbf{w},\mathbf{u}}\right) \leq e^{-\frac{1}{2}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2} \tag{112}$$

$$\Pr\left(\left|\left\langle \frac{\mathbf{w} + \mathbf{u}}{\|\mathbf{w} + \mathbf{u}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{w}, \mathbf{u}) \right\rangle\right| \geq \frac{\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}}{m} \middle| \mathbf{R}_{\mathbf{w},\mathbf{u}} = \mathbf{r}, L_{\mathbf{w},\mathbf{u}} = \ell_{\mathbf{w},\mathbf{u}}\right) \leq 2e^{-\frac{1}{2}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2} \tag{113}$$

$$\Pr\left(\left\|\frac{1}{\eta} g_{\mathbf{A};J}(\mathbf{w}, \mathbf{u})\right\|_2 \geq \frac{2\sqrt{2k\ell_{\mathbf{w},\mathbf{u}}}}{m} + \frac{\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}}{m} \middle| \mathbf{R}_{\mathbf{w},\mathbf{u}} = \mathbf{r}, L_{\mathbf{w},\mathbf{u}} = \ell_{\mathbf{w},\mathbf{u}}\right) \leq 2e^{-\frac{1}{8}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2} \tag{114}$$

$$\Pr\left(M_{\beta,\mathbf{u}} > \frac{4}{3}\tau m\right) \leq e^{-\frac{8}{3\pi}\tau m} \tag{115}$$

To obtain a uniform result, a union bound can be taken over all $\mathbf{u} \in \mathcal{C}_\tau$, all $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$, and all $J \subseteq [n]$, $|J| \leq 2k$, to upper bound the probability that the uniform result fails to occur. For each pair $\mathbf{w}, \mathbf{u}$, the parameter $t_{\mathbf{w},\mathbf{u}}$ should be selected with consideration for this probability of failure so that it does not exceed ρ_2. Let

$$\rho_2' = \rho_2 - \binom{n}{k}\left(\frac{6}{\tau}\right)^k \binom{n}{2k} 2^{2k} e^{-\frac{8}{3\pi}\tau m} = \rho_2 - \binom{n}{k}\binom{n}{2k}\left(\frac{24}{\tau}\right)^k e^{-\frac{8}{3\pi}\tau m}. \tag{116}$$

Then, the lemma's uniform result fails to hold with probability not exceeding ρ_2 as long as

$$\binom{n}{k}\binom{n}{2k}\left(\frac{24}{\tau}\right)^k 2^{\frac{4}{3}\tau m} \cdot 5e^{-\frac{1}{8}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2} \leq \rho_2' \tag{117a}$$

$$\longrightarrow \quad e^{\frac{1}{8}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2} \geq \binom{n}{k}\binom{n}{2k}\left(\frac{24}{\tau}\right)^k e^{\frac{4}{3}\tau m} \cdot \frac{5}{\rho_2'} \tag{117b}$$

$$\longrightarrow \quad e^{\frac{1}{8}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2} \geq e^{\frac{8}{3}\tau m} \tag{117c}$$

$$\longrightarrow \quad \frac{1}{8}\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}^2 \geq \frac{8}{3}\tau m \tag{117d}$$

$$\longrightarrow \quad t_{\mathbf{w},\mathbf{u}} \geq \sqrt{\frac{64}{3}\frac{\tau m}{\ell_{\mathbf{w},\mathbf{u}}}} = \frac{8}{\sqrt{3}} \cdot \sqrt{\frac{\tau m}{\ell_{\mathbf{w},\mathbf{u}}}} \tag{117e}$$

Hence, $t_{\mathbf{w},\mathbf{u}}$ can be set as small as

$$t_{\mathbf{w},\mathbf{u}} = \frac{8}{\sqrt{3}} \cdot \sqrt{\frac{\tau m}{\ell_{\mathbf{w},\mathbf{u}}}} \tag{118}$$

Thus, with probability at least $1 - \rho_2$, the following holds.

$$\begin{aligned}
\frac{1}{\eta} \left\| h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\|_2 &\leq \left| \left\langle \frac{\mathbf{w}-\mathbf{u}}{\|\mathbf{w}-\mathbf{u}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\rangle \right| + \left| \left\langle \frac{\mathbf{w}+\mathbf{u}}{\|\mathbf{w}+\mathbf{u}\|_2}, \frac{1}{\eta} h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\rangle \right| + \left\| g_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\|_2 && \text{(119a)} \\
&\leq \left(\sqrt{\frac{\pi}{2}} + t_{\mathbf{w},\mathbf{u}} \right) \frac{\ell_{\mathbf{w},\mathbf{u}}}{m} + \frac{\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}}{m} + \frac{2\sqrt{2k\ell_{\mathbf{w},\mathbf{u}}}}{m} + \frac{\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}}{m} && \text{(119b)} \\
&= \sqrt{\frac{\pi}{2}} \frac{\ell_{\mathbf{w},\mathbf{u}}}{m} + \frac{2\sqrt{2k\ell_{\mathbf{w},\mathbf{u}}}}{m} + 3 \frac{\ell_{\mathbf{w},\mathbf{u}} t_{\mathbf{w},\mathbf{u}}}{m} && \text{(119c)} \\
&= \sqrt{\frac{\pi}{2}} \frac{\ell_{\mathbf{w},\mathbf{u}}}{m} + \frac{2\sqrt{2k\ell_{\mathbf{w},\mathbf{u}}}}{m} + 3 \frac{\ell_{\mathbf{w},\mathbf{u}}}{m} \frac{8}{\sqrt{3}} \cdot \sqrt{\frac{\tau m}{\ell_{\mathbf{w},\mathbf{u}}}} && \text{(119d)} \\
&= \sqrt{\frac{\pi}{2}} \frac{\ell_{\mathbf{w},\mathbf{u}}}{m} + \frac{2\sqrt{2k\ell_{\mathbf{w},\mathbf{u}}}}{m} + 8\sqrt{3} \cdot \sqrt{\frac{\tau \ell_{\mathbf{w},\mathbf{u}}}{m}} && \text{(119e)}
\end{aligned}$$

Note that the random variable $L_{\mathbf{w},\mathbf{u}}$ counts the number of sign differences between $\mathsf{sign}(\mathbf{A}\mathbf{w})$ and $\mathsf{sign}(\mathbf{A}\mathbf{u})$, which cannot exceed $M_{\beta,\mathbf{u}}$ because $\theta_{\mathbf{w},\mathbf{u}} \leq \beta$. Earlier, it was argued that with probability at least $1 - e^{-\frac{8}{3\pi}\tau m}$, the random variable $M_{\beta,\mathbf{u}}$ takes a value no larger than $M_{\beta,\mathbf{u}} \leq \frac{4}{3}\tau m$, and therefore, the value taken by the random variable $L_{\mathbf{w},\mathbf{u}}$ is bounded by $L_{\mathbf{w},\mathbf{u}} \leq M_{\beta,\mathbf{u}} \leq \frac{4}{3}\tau m$, implying that

$$\frac{\ell_{\mathbf{w},\mathbf{u}}}{m} \leq \frac{4}{3}\tau. \tag{120}$$

Applying (120) to (119e), the bound becomes

$$\begin{aligned}
\frac{1}{\eta} \left\| h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\|_2 &\leq \sqrt{\frac{\pi}{2}} \frac{\ell_{\mathbf{w},\mathbf{u}}}{m} + \frac{2\sqrt{2k\ell_{\mathbf{w},\mathbf{u}}}}{m} + 8\sqrt{3} \cdot \sqrt{\frac{\tau \ell_{\mathbf{w},\mathbf{u}}}{m}} && \text{(121a)} \\
&\leq \sqrt{\frac{\pi}{2}} \cdot \frac{4}{3}\tau + \frac{4\sqrt{6}}{3}\tau + 8\sqrt{3} \cdot \tau && \text{(121b)} \\
&\leq \sqrt{\frac{\pi}{2}} \cdot \frac{4}{3}\tau + \frac{4\sqrt{6}}{3}\tau + 8\sqrt{3}\tau && \text{(121c)} \\
&= \frac{2\sqrt{2\pi}}{3}\tau + \frac{4\sqrt{6}}{3}\tau + 8\sqrt{3}\tau && \text{(121d)} \\
&= \left(\frac{2\sqrt{2\pi}}{3} + \frac{4\sqrt{6}}{3} + 8\sqrt{3} \right) \tau && \text{(121e)}
\end{aligned}$$

Then, substituting $\eta = \sqrt{2\pi}$ yields

$$\left\| h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\|_2 \leq \left(\frac{2\sqrt{2\pi}}{3} + \frac{4\sqrt{6}}{3} + 8\sqrt{3} \right) \eta\tau = \left(\frac{2\sqrt{2\pi}}{3} + \frac{4\sqrt{6}}{3} + 8\sqrt{3} \right) \sqrt{2\pi}\tau = \left(\frac{4\pi}{3} + \frac{8\sqrt{3\pi}}{3} + 8\sqrt{6\pi} \right) \tau \tag{122}$$

Therefore, with probability at least $1 - \rho_2$, uniformly for every $\mathbf{u} \in \mathcal{C}_\tau$ and each $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$,

$$\left\| h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\|_2 \leq \left(\frac{4\pi}{3} + \frac{8\sqrt{3\pi}}{3} + 8\sqrt{6\pi} \right) \tau = \left(\frac{4\pi}{3} + \frac{8\sqrt{3\pi}}{3} + 8\sqrt{6\pi} \right) \frac{\delta}{b} \tag{123}$$

As previously discussed, for any given $\mathbf{u} \in \mathcal{C}_\tau$ and $\mathbf{x} \in \mathcal{B}_\tau(\mathbf{u}) \cap \mathcal{S}^{n-1} \cap \Sigma_k^n$, there exists $\mathbf{w} \in \mathcal{D}_\tau(\mathbf{u})$ such that $\mathsf{sign}(\mathbf{A}\mathbf{w}) = \mathsf{sign}(\mathbf{A}\mathbf{x})$, which implies that $h_{\mathbf{A};J}(\mathbf{x},\mathbf{u}) = h_{\mathbf{A};J}(\mathbf{w},\mathbf{u})$. It follows that

$$\left\| h_{\mathbf{A};J}(\mathbf{x},\mathbf{u}) \right\|_2 = \left\| h_{\mathbf{A};J}(\mathbf{w},\mathbf{u}) \right\|_2 \leq \left(\frac{4\pi}{3} + \frac{8\sqrt{3\pi}}{3} + 8\sqrt{6\pi} \right) \frac{\delta}{b}. \tag{124}$$

Combining (124) with Eq. (103),

$$\|(\mathbf{x}-\mathbf{u})-h_{\mathbf{A};J}(\mathbf{x},\mathbf{u})\|_2 \leq \|\mathbf{x}-\mathbf{u}\|_2 + \|h_{\mathbf{A};J}(\mathbf{x},\mathbf{u})\|_2 \leq \tau + \|h_{\mathbf{A};J}(\mathbf{w},\mathbf{u})\|_2 \tag{125a}$$

$$\leq \frac{\delta}{b} + \left(\frac{4\pi}{3} + \frac{8\sqrt{3\pi}}{3} + 8\sqrt{6\pi}\right)\frac{\delta}{b} = \left(1 + \frac{4\pi}{3} + \frac{8\sqrt{3\pi}}{3} + 8\sqrt{6\pi}\right)\frac{\delta}{b} \tag{125b}$$

In short, uniformly with probability at least $1-\rho_2$, for all $\mathbf{u}\in\mathcal{C}_\tau$ and each $\mathbf{x}\in\mathcal{B}_\tau(\mathbf{u})\cap\mathcal{S}^{n-1}\cap\Sigma_k^n$.

$$\|(\mathbf{x}-\mathbf{u})-h_{\mathbf{A};J}(\mathbf{x},\mathbf{u})\|_2 \leq b_2\delta \tag{126}$$

where the universal constant $b_2>0$ is set to

$$b_2 = \frac{1}{b}\left(1 + \frac{4\pi}{3} + \frac{8\sqrt{3\pi}}{3} + 8\sqrt{6\pi}\right). \tag{127}$$

■

A.3 Combining the regimes to prove Theorem 3.3

Using Lemmas A.3 and A.4, Theorem 3.3 can now be established with a direct argument.

Proof (Theorem 3.3). Fix $\rho_1,\rho_2\in(0,1)$ such that $\rho_1+\rho_2=\rho$. With the universal constant $a=16$, setting $\rho_1=\rho_2=\frac{\rho}{2}$ suffices. Let $\mathbf{x},\mathbf{y}\in\mathcal{S}^{n-1}\cap\Sigma_k^n$ be an arbitrary pair of k-sparse unit vectors. Suppose $\mathbf{u},\mathbf{v}\in\mathcal{C}_\tau$ are the closest points to $\mathbf{x},\mathbf{y}$, respectively, subject to $\mathsf{supp}(\mathbf{u})=\mathsf{supp}(\mathbf{x})$ and $\mathsf{supp}(\mathbf{v})=\mathsf{supp}(\mathbf{y})$. Formally,

$$\mathbf{u} = \arg\min_{\substack{\mathbf{u}'\in\mathcal{C}_\tau:\\ \mathsf{supp}(\mathbf{u}')=\mathsf{supp}(\mathbf{x})}} \|\mathbf{x}-\mathbf{u}'\|_2 \tag{128}$$

$$\mathbf{v} = \arg\min_{\substack{\mathbf{v}'\in\mathcal{C}_\tau:\\ \mathsf{supp}(\mathbf{v}')=\mathsf{supp}(\mathbf{y})}} \|\mathbf{y}-\mathbf{v}'\|_2 \tag{129}$$

Note that the requirement $\mathsf{supp}(\mathbf{u})=\mathsf{supp}(\mathbf{x})$ and $\mathsf{supp}(\mathbf{v})=\mathsf{supp}(\mathbf{y})$ is possible due to the design of the τ-net $\mathcal{C}_\tau$ as specified at the beginning of Section A. Observe

$$(\mathbf{x}-\mathbf{y})-h_{\mathbf{A}}(\mathbf{x},\mathbf{y}) \tag{130a}$$

$$= (\mathbf{x}-\mathbf{y}) - \sqrt{2\pi}\frac{1}{m}\mathbf{A}^\mathsf{T}\cdot\frac{1}{2}\left(\mathsf{sign}\left(\mathbf{A}\mathbf{x}\right)-\mathsf{sign}\left(\mathbf{A}\mathbf{y}\right)\right) \tag{130b}$$

$$= (\mathbf{u}-\mathbf{v}) + (\mathbf{x}-\mathbf{u}) + (\mathbf{v}-\mathbf{y}) - \sqrt{2\pi}\frac{1}{m}\mathbf{A}^\mathsf{T}\cdot\frac{1}{2}\left(\mathsf{sign}\left(\mathbf{A}\mathbf{u}\right)-\mathsf{sign}\left(\mathbf{A}\mathbf{v}\right)\right) \tag{130c}$$

$$- \sqrt{2\pi}\frac{1}{m}\mathbf{A}^\mathsf{T}\cdot\frac{1}{2}\left(\mathsf{sign}\left(\mathbf{A}\mathbf{x}\right)-\mathsf{sign}\left(\mathbf{A}\mathbf{u}\right)\right) - \sqrt{2\pi}\frac{1}{m}\mathbf{A}^\mathsf{T}\cdot\frac{1}{2}\left(\mathsf{sign}\left(\mathbf{A}\mathbf{v}\right)-\mathsf{sign}\left(\mathbf{A}\mathbf{y}\right)\right)$$

$$= (\mathbf{u}-\mathbf{v}) - \sqrt{2\pi}\frac{1}{m}\mathbf{A}^\mathsf{T}\cdot\frac{1}{2}\left(\mathsf{sign}\left(\mathbf{A}\mathbf{x}\right)-\mathsf{sign}\left(\mathbf{A}\mathbf{u}\right)\right) \tag{130d}$$

$$+ (\mathbf{x}-\mathbf{u}) - \sqrt{2\pi}\frac{1}{m}\mathbf{A}^\mathsf{T}\cdot\frac{1}{2}\left(\mathsf{sign}\left(\mathbf{A}\mathbf{u}\right)-\mathsf{sign}\left(\mathbf{A}\mathbf{v}\right)\right)$$

$$+ (\mathbf{v}-\mathbf{y}) - \sqrt{2\pi}\frac{1}{m}\mathbf{A}^\mathsf{T}\cdot\frac{1}{2}\left(\mathsf{sign}\left(\mathbf{A}\mathbf{v}\right)-\mathsf{sign}\left(\mathbf{A}\mathbf{y}\right)\right)$$

$$= (\mathbf{u}-\mathbf{v}) - h_{\mathbf{A}}(\mathbf{u},\mathbf{v}) + (\mathbf{x}-\mathbf{u}) - h_{\mathbf{A}}(\mathbf{x},\mathbf{u}) + (\mathbf{v}-\mathbf{y}) - h_{\mathbf{A}}(\mathbf{v},\mathbf{y}) \tag{130e}$$

Write $J_\mathbf{x} = J\cup\mathsf{supp}(\mathbf{x})$ and $J_\mathbf{y} = J\cup\mathsf{supp}(\mathbf{y})$, where $|J_\mathbf{x}|,|J_\mathbf{y}|\leq 2k$. Then,

$$(\mathbf{x}-\mathbf{y})-h_{\mathbf{A};J}(\mathbf{x},\mathbf{y}) = (\mathbf{x}-\mathbf{y}) - \mathcal{T}_{\mathsf{supp}(\mathbf{x})\cup\mathsf{supp}(\mathbf{y})\cup J}\left(h_{\mathbf{A}}(\mathbf{x},\mathbf{y})\right) \tag{131a}$$

$$= (\mathbf{u}-\mathbf{v}) - \mathcal{T}_{\mathsf{supp}(\mathbf{x})\cup\mathsf{supp}(\mathbf{y})\cup J}\left(h_{\mathbf{A}}(\mathbf{u},\mathbf{v})\right) \tag{131b}$$

$$+ (\mathbf{x}-\mathbf{u}) - \mathcal{T}_{\mathsf{supp}(\mathbf{x})\cup\mathsf{supp}(\mathbf{y})\cup J}\left(h_{\mathbf{A}}(\mathbf{x},\mathbf{u})\right)$$

$$
\begin{aligned}
&+ (\mathbf{v}-\mathbf{y}) - \mathcal{T}_{\mathsf{supp}(\mathbf{x})\cup\mathsf{supp}(\mathbf{y})\cup J}\left(h_{\mathbf{A}}(\mathbf{v},\mathbf{y})\right) \\
&= (\mathbf{u}-\mathbf{v}) - \mathcal{T}_{\mathsf{supp}(\mathbf{u})\cup\mathsf{supp}(\mathbf{v})\cup J}\left(h_{\mathbf{A}}(\mathbf{u},\mathbf{v})\right) && (131c) \\
&+ (\mathbf{x}-\mathbf{u}) - \mathcal{T}_{\mathsf{supp}(\mathbf{x})\cup\mathsf{supp}(\mathbf{u})\cup J_{\mathbf{y}}}\left(h_{\mathbf{A}}(\mathbf{x},\mathbf{u})\right) \\
&+ (\mathbf{v}-\mathbf{y}) - \mathcal{T}_{\mathsf{supp}(\mathbf{v})\cup\mathsf{supp}(\mathbf{y})\cup J_{\mathbf{x}}}\left(h_{\mathbf{A}}(\mathbf{v},\mathbf{y})\right) \\
&= (\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) + (\mathbf{x}-\mathbf{u}) - h_{\mathbf{A};J_{\mathbf{y}}}(\mathbf{x},\mathbf{u}) + (\mathbf{v}-\mathbf{y}) - h_{\mathbf{A};J_{\mathbf{x}}}(\mathbf{v},\mathbf{y}) && (131d)
\end{aligned}
$$

The norm of (131) is then bounded by the triangle inequality.

$$
\begin{aligned}
&\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 && (132a) \\
&= \left\|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v}) + (\mathbf{x}-\mathbf{u}) - h_{\mathbf{A};J_{\mathbf{y}}}(\mathbf{x},\mathbf{u}) + (\mathbf{v}-\mathbf{y}) - h_{\mathbf{A};J_{\mathbf{x}}}(\mathbf{v},\mathbf{y})\right\|_2 && (132b) \\
&\le \|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 + \left\|(\mathbf{x}-\mathbf{u}) - h_{\mathbf{A};J_{\mathbf{y}}}(\mathbf{x},\mathbf{u})\right\|_2 + \|(\mathbf{v}-\mathbf{y}) - h_{\mathbf{A};J_{\mathbf{x}}}(\mathbf{v},\mathbf{y})\|_2 && (132c)
\end{aligned}
$$

Suppose $d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) < \tau$. Then, by Lemma A.4,

$$
\begin{aligned}
&\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 && (133a) \\
&\le \|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 + \left\|(\mathbf{x}-\mathbf{u}) - h_{\mathbf{A};J_{\mathbf{y}}}(\mathbf{x},\mathbf{u})\right\|_2 + \|(\mathbf{v}-\mathbf{y}) - h_{\mathbf{A};J_{\mathbf{x}}}(\mathbf{v},\mathbf{y})\|_2 && (133b) \\
&\le 3b_2\delta && (133c) \\
&\le b_1\sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})} + 3b_2\delta && (133d)
\end{aligned}
$$

uniformly with probability at least $1-\rho_2 > 1-\rho$. On the other hand, if $d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v}) \ge \tau$, then by Lemmas A.3 and A.4,

$$
\begin{aligned}
&\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 && (134a) \\
&\le \|(\mathbf{u}-\mathbf{v}) - h_{\mathbf{A};J}(\mathbf{u},\mathbf{v})\|_2 + \left\|(\mathbf{x}-\mathbf{u}) - h_{\mathbf{A};J_{\mathbf{y}}}(\mathbf{x},\mathbf{u})\right\|_2 + \|(\mathbf{v}-\mathbf{y}) - h_{\mathbf{A};J_{\mathbf{x}}}(\mathbf{v},\mathbf{y})\|_2 && (134b) \\
&\le b_1\sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})} + b_2\delta + b_2\delta && (134c) \\
&= b_1\sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})} + 2b_2\delta && (134d) \\
&\le b_1\sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})} + 3b_2\delta && (134e)
\end{aligned}
$$

uniformly with probability at least $1-\rho_1-\rho_2 = 1-\rho$. Therefore, with probability at least $1-\rho$, for all $\mathbf{x},\mathbf{y} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$ and all $J \subseteq [n]$, $|J| \le k$,

$$
\|(\mathbf{x}-\mathbf{y}) - h_{\mathbf{A};J}(\mathbf{x},\mathbf{y})\|_2 \le c_1\sqrt{\delta d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})} + c_2\delta \tag{135}
$$

where $c_1 = b_1 = \sqrt{\frac{3\pi}{b}}\left(1+\frac{16\sqrt{2}}{3}\right)$, $c_2 = 3b_2 = \frac{3}{b}\left(1+\frac{4\pi}{3}+\frac{8\sqrt{3\pi}}{3}+8\sqrt{6\pi}\right)$, and $b \gtrsim 379.1038$, as specified in Eq. (3). Succinctly, the measurement matrix $\mathbf{A}$ satisfies the (k,n,δ,c_1,c_2)-RAIC with probability at least $1-\rho$. ∎

B Proofs of the concentration inequalities, Lemmas A.1 and A.2

B.1 Orthogonal projections: proof of Lemma A.1

This section proves a slightly more general form of the three concentration inequalities in Lemma A.1, stated in Lemmas B.1-B.3. It is easy to see that Lemma A.1 is a direct corollary.

Lemma B.1. *Let $\ell, t > 0$ and $\mathbf{r} \in \{-1,0,1\}^m$ such that $\|\mathbf{r}\|_0 = \ell$. Fix an ordered pair of real-valued unit vectors, $(\mathbf{u},\mathbf{v}) \in \mathcal{S}^{n-1}\times\mathcal{S}^{n-1}$. Define the random variable $L_{\mathbf{u},\mathbf{v}} = \|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0$, and suppose $\mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}$ and $L_{\mathbf{u},\mathbf{v}} = \ell$. Then, the random variable $X_{\mathbf{u},\mathbf{v}} = \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, \sum_{i=1}^m \mathbf{Z}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right\rangle$ conditioned on $\mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell$ is concentrated around its mean such that*

$$
\Pr\left(|X_{\mathbf{u},\mathbf{v}} - \mathbb{E}[X_{\mathbf{u},\mathbf{v}}|L_{\mathbf{u},\mathbf{v}}=\ell]| \ge \ell t \,\middle|\, \mathbf{R}_{\mathbf{u},\mathbf{v}}=\mathbf{r}, L_{\mathbf{u},\mathbf{v}}=\ell\right) \le 2e^{-\frac{1}{2}\ell t^2}. \tag{136}
$$

Lemma B.2. *Let $\ell, t > 0$ and $\mathbf{r} \in \{-1,0,1\}^m$ such that $\|\mathbf{r}\|_0 = \ell$. Fix an ordered pair of real-valued unit vectors, $(\mathbf{u},\mathbf{v}) \in \mathcal{S}^{n-1} \times \mathcal{S}^{n-1}$. Define the random variable $L_{\mathbf{u},\mathbf{v}} = \|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0$, and suppose $\mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}$ and $L_{\mathbf{u},\mathbf{v}} = \ell$. Then, the random variable $X_{\mathbf{u},\mathbf{v}} = \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, \sum_{i=1}^m \mathbf{Z}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right\rangle$ conditioned on $\mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell$ is concentrated around zero such that*

$$\Pr\left(|X_{\mathbf{u},\mathbf{v}}| \geq \ell t \,\middle|\, \mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell\right) \leq 2e^{-\frac{1}{2}\ell t^2}. \tag{137}$$

Lemma B.3. *Let $d, \ell, t > 0$. Fix an ordered pair of k-sparse, real-valued unit vectors, $(\mathbf{u},\mathbf{v}) \in (\mathcal{S}^{n-1} \cap \Sigma_k^n) \times (\mathcal{S}^{n-1} \cap \Sigma_k^n)$, and let $J \subseteq [n]$ with $|J| \leq d$. Define the random variables $\mathbf{Y}^{(i)}_{\mathbf{u},\mathbf{v}} = \mathbf{Z}^{(i)} - \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, \mathbf{Z}^{(i)} \right\rangle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2} - \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, \mathbf{Z}^{(i)} \right\rangle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}$, $X_{\mathbf{u},\mathbf{v}} = \left\| \mathcal{T}_J \left(\sum_{i=1}^m \mathbf{Y}^{(i)}_{\mathbf{u},\mathbf{v}} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2$, and $L_{\mathbf{u},\mathbf{v}} = \|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0$, and suppose $\mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}$ and $L_{\mathbf{u},\mathbf{v}} = \ell$. Then,*

$$\Pr\left(X_{\mathbf{u},\mathbf{v}} \geq \left(\sqrt{2k} + \sqrt{d}\right)\sqrt{\ell} + \ell t \,\middle|\, \mathbf{R}_{\mathbf{u},\mathbf{v}} = \mathbf{r}, L_{\mathbf{u},\mathbf{v}} = \ell\right) \leq 2e^{-\frac{1}{8}\ell t^2} \tag{138}$$

Before proving the lemma (see Appendix B.1.2), several intermediate results are stated and proved in Appendix B.1.1 to facilitate the proof.

B.1.1 The distributions of orthogonal projections of i.i.d. standard normal vectors

Lemma B.4. *Fix an ordered pair of real-valued vectors, $(\mathbf{u},\mathbf{v}) \in \mathcal{S}^{n-1} \times \mathcal{S}^{n-1}$, of unit norm. Let $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n\times n})$ be a standard normal random vector, and let R be the (discrete) random variable taking values in $\{-1,0,1\}$ and given by $R_{\mathbf{u},\mathbf{v}} = \frac{1}{2}(\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle) - \mathsf{sign}(\langle \mathbf{v}, \mathbf{Z}\rangle))$. Define the map $\alpha : \mathbb{R} \to \mathbb{R}$ by $\alpha(x) = x \tan\left(\frac{\theta_{\mathbf{u},\mathbf{v}}}{2}\right) = x\sqrt{\frac{d^2_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{4 - d^2_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}}$. Then, the density function $f_{X|R} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ for the random variable $X_{\mathbf{u},\mathbf{v}} = \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, \mathbf{Z} \right\rangle R_{\mathbf{u},\mathbf{v}}$ conditioned on $R \neq 0$ is given by*

$$f_{X_{\mathbf{u},\mathbf{v}}|R_{\mathbf{u},\mathbf{v}}}(x \mid r \neq 0) = \begin{cases} \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}}\sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dy, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases} \tag{139}$$

Moreover, in expectation,

$$\mathbb{E}(X_{\mathbf{u},\mathbf{v}} \mid R_{\mathbf{u},\mathbf{v}} \neq 0) = \sqrt{\frac{\pi}{2}} \frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{\theta_{\mathbf{u},\mathbf{v}}}. \tag{140}$$

Proof (Lemma B.4). Before deriving the density function of $X_{\mathbf{u},\mathbf{v}}$, $\mathbf{u},\mathbf{v} \in \mathcal{S}^{n-1}$, let us show that for $\mathbf{u},\mathbf{v},\mathbf{u}',\mathbf{v}' \in \mathcal{S}^{n-1}$, such that $\theta_{\mathbf{u},\mathbf{v}} = \theta_{\mathbf{u}',\mathbf{v}'}$, the pair of random variables $(X_{\mathbf{u},\mathbf{v}} \mid R_{\mathbf{u},\mathbf{v}} = 0)$ and $(X_{\mathbf{u}',\mathbf{v}'} \mid R_{\mathbf{u}',\mathbf{v}'} = 0)$ follow the same distribution, as do the pair $(X_{\mathbf{u},\mathbf{v}} \mid R_{\mathbf{u},\mathbf{v}} \neq 0)$ and $(X_{\mathbf{u}',\mathbf{v}'} \mid R_{\mathbf{u}',\mathbf{v}'} \neq 0)$. This will simplify the characterization of the distribution of $X_{\mathbf{u},\mathbf{v}}$ by allowing $\mathbf{u},\mathbf{v}$ to be chosen non-arbitrarily. Conditioned on $R_{\mathbf{u},\mathbf{v}} = R_{\mathbf{u}',\mathbf{v}'} = 0$, $X_{\mathbf{u},\mathbf{v}} = X_{\mathbf{u}',\mathbf{v}'} = 0$ with probability 1. Otherwise, when $R_{\mathbf{u},\mathbf{v}}, R_{\mathbf{u}',\mathbf{v}'} \neq 0$, write $q = \|\mathbf{u}-\mathbf{v}\|_2 = \|\mathbf{u}'-\mathbf{v}'\|_2$, and observe

$$X_{\mathbf{u},\mathbf{v}} = \left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, \mathbf{Z} \right\rangle R_{\mathbf{u},\mathbf{v}} \tag{141a}$$

$$= \frac{1}{q}\left(\langle \mathbf{u}, \mathbf{Z}\rangle R_{\mathbf{u},\mathbf{v}} - \langle \mathbf{v}, \mathbf{Z}\rangle R_{\mathbf{u},\mathbf{v}}\right) \tag{141b}$$

$$= \frac{1}{q}\left(\langle \mathbf{u}, \mathbf{Z}\rangle \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle) - \langle \mathbf{v}, \mathbf{Z}\rangle (-\mathsf{sign}(\langle \mathbf{v}, \mathbf{Z}\rangle))\right) \tag{141c}$$

$$= \frac{1}{q}\left(\langle \mathbf{u}, \mathbf{Z}\rangle \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle) + \langle \mathbf{v}, \mathbf{Z}\rangle \mathsf{sign}(\langle \mathbf{v}, \mathbf{Z}\rangle)\right) \tag{141d}$$

$$= \frac{1}{q}\left(|\langle \mathbf{u}, \mathbf{Z}\rangle| + |\langle \mathbf{v}, \mathbf{Z}\rangle|\right) \tag{141e}$$

Likewise,

$$X_{\mathbf{u}',\mathbf{v}'} = \frac{1}{q}\left(\langle \mathbf{u}', \mathbf{Z}\rangle \mathsf{sign}(\langle \mathbf{u}', \mathbf{Z}\rangle) + \langle \mathbf{v}', \mathbf{Z}\rangle \mathsf{sign}(\langle \mathbf{v}', \mathbf{Z}\rangle)\right) = \frac{1}{q}\left(|\langle \mathbf{u}', \mathbf{Z}\rangle| + |\langle \mathbf{v}', \mathbf{Z}\rangle|\right) \tag{142}$$

Then, letting

$$(Y, Y') \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \cos(\theta_{\mathbf{u},\mathbf{v}})\\ \cos(\theta_{\mathbf{u},\mathbf{v}}) & 1\end{pmatrix}\right) \equiv \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \cos(\theta_{\mathbf{u}',\mathbf{v}'})\\ \cos(\theta_{\mathbf{u}',\mathbf{v}'}) & 1\end{pmatrix}\right), \tag{143}$$

notice that $X_{\mathbf{u},\mathbf{v}}$ and $X_{\mathbf{u}',\mathbf{v}'}$, conditioned on $R_{\mathbf{u},\mathbf{v}}, R_{\mathbf{u}',\mathbf{v}'} \neq 0$, both follow the same distribution as $\frac{1}{q}(|Y| + |Y'|)$. Hence, the claim is proved.

We are ready to derive Lemma B.4. To simplify notation, we will drop the subscript of $\mathbf{u}, \mathbf{v}$ on the random variables, writing $X = X_{\mathbf{u},\mathbf{v}}, R = R_{\mathbf{u},\mathbf{v}}$. Let $\mathbf{Z} = (Z_1, \ldots, Z_n) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n\times n})$. For an arbitrary choice of $\theta \in [0, 2\pi)$, fix $\mathbf{u}, \mathbf{v} \in \mathcal{S}^{n-1}$ such that $\theta_{\mathbf{u},\mathbf{v}} = \theta$ and $\mathbf{u} = (u_1, u_2, \ldots, u_n)$, $\mathbf{v} = (-u_1, u_2, \ldots, u_n)$ with $u_1 > 0$, which is made possible by the claim argued above. This choice will now be shown to induce the distribution of $(|Z_1| \mid R \neq 0)$ on the random variable $(X \mid R \neq 0)$. First, observe that

$$\frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2} = (1, 0, \ldots, 0) \tag{144}$$

and thus

$$X = \left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, \mathbf{Z}\right\rangle R = Z_1 R. \tag{145}$$

Moreover, conditioned on $R \neq 0$, by its definition, R takes the value

$$R = \mathsf{sign}\left(\left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, \mathbf{Z}\right\rangle\right) = \mathsf{sign}\left(Z_1\right). \tag{146}$$

It follows that

$$(X \mid R \neq 0) = \left(\left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, \mathbf{Z}\right\rangle R \,\middle|\, R \neq 0\right) = (Z_1 R \mid R \neq 0) = (Z_1 \mathsf{sign}\left(Z_1\right) \mid R \neq 0) = (|Z_1| \mid R \neq 0), \tag{147}$$

as claimed.

Next, the density function $f_{X|R\neq 0} : \mathbb{R} \to \mathbb{R}_{\geq 0}$ of the conditioned random variable $(X \mid R \neq 0)$ is found by deriving the equivalent density function $f_{|Z_1||R\neq 0} : \mathbb{R} \to \mathbb{R}_{\geq 0}$. By Bayes' rule, this density function can be written as

$$f_{|Z_1||R}(x \mid r \neq 0) = \frac{f_{|Z_1|}(x) p_{R||Z_1|}(r \neq 0 \mid x)}{p_R(r \neq 0)}, \tag{148}$$

which expresses $f_{|Z_1||R\neq 0}$ using three more manageable density (mass) functions. Beginning with $p_R(r \neq 0)$, let the random variable I be the indicator of the event $R \neq 0$, formally, $I = \mathbf{1}(R \neq 0)$. Observing the following biconditionals

$$R \neq 0 \iff \frac{1}{2}\left(\mathsf{sign}\left(\langle \mathbf{u}, \mathbf{Z}\rangle\right) - \mathsf{sign}\left(\langle \mathbf{v}, \mathbf{Z}\rangle\right)\right) \neq 0 \iff \left(\mathsf{sign}\left(\langle \mathbf{u}, \mathbf{Z}\rangle\right) - \mathsf{sign}\left(\langle \mathbf{v}, \mathbf{Z}\rangle\right)\right) \neq 0, \tag{149}$$

it follows that

$$I = \mathbf{1}(R \neq 0) \tag{150a}$$

$$I = \mathbf{1}\left(\frac{1}{2}\mathsf{sign}\left(\langle \mathbf{u}, \mathbf{Z}\rangle\right) - \mathsf{sign}\left(\langle \mathbf{v}, \mathbf{Z}\rangle\right) \neq 0\right) \tag{150b}$$

$$I = \mathbf{1}(\mathsf{sign}\left(\langle \mathbf{u}, \mathbf{Z}\rangle\right) - \mathsf{sign}\left(\langle \mathbf{v}, \mathbf{Z}\rangle\right) \neq 0) \tag{150c}$$

are equivalent definitions for the random variable I. Then, the mass associated with $R \neq 0$ is $p_R(r \neq 0) = \Pr(I = 1) = \frac{\theta_{\mathbf{u},\mathbf{v}}}{\pi}$, where the last equality follows from Lemma B.5, stated below and proved in Appendix D. (see the proof of Lemma D.2).

Lemma B.5. *Fix any pair of real-valued vectors* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$*, and suppose* $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n\times n})$ *is a standard normal vector with i.i.d. entries. Define the indicator random variable* $I = \mathbf{1}(\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle) - \mathsf{sign}(\langle \mathbf{v}, \mathbf{Z}\rangle) \neq 0)$*. Then,*

$$\Pr(I = 1) = \frac{\theta_{\mathbf{u},\mathbf{v}}}{\pi}. \tag{151}$$

In short, the above argument yields $p_R(r \neq 0) = \Pr(I = 1) = \frac{\theta_{\mathbf{u},\mathbf{v}}}{\pi}$.

Next, the density function for the random variable $|Z_1|$, which is the absolute value of the standard normal random variable Z_1, is the well-known folded standard normal distribution and takes the form

$$f_{|Z_1|}(x) = \begin{cases} f_{Z_1}(-x) + f_{Z_1}(x), & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases} \tag{152}$$

$$= \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{(-x)^2}{2}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases} \tag{153}$$

$$= \begin{cases} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} + \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases} \tag{154}$$

$$= \begin{cases} 2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases} \tag{155}$$

$$= \begin{cases} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases} \tag{156}$$

In summary,

$$f_{|Z_1|}(x) = \begin{cases} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}}, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases} \tag{157}$$

Lastly, consider the mass function of $(R \mid |Z_1|)$, which need only be evaluated when $R \neq 0$. The next argument will show that

$$p_{R||Z_1|}(r \neq 0 \mid x) = \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}}\, dy \tag{158}$$

where $\alpha : \mathbb{R} \to \mathbb{R}$ is as defined in the lemma (and repeated here for convenience):

$$\alpha(x) = x \tan\left(\frac{\theta_{\mathbf{u},\mathbf{v}}}{2}\right) = x\sqrt{\frac{d^2_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v})}{4 - d^2_{\mathcal{S}^{n-1}}(\mathbf{u}, \mathbf{v})}}. \tag{159}$$

Notice that given $|Z_1| = x$, $x \geq 0$, the event $R \neq 0$ occurs precisely when

$$\left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, \mathbf{Z} \right\rangle \in \left[-x \tan\left(\frac{\theta_{\mathbf{u},\mathbf{v}}}{2}\right), x \tan\left(\frac{\theta_{\mathbf{u},\mathbf{v}}}{2}\right) \right] \tag{160}$$

where $\tan\left(\frac{\theta_{\mathbf{u},\mathbf{v}}}{2}\right)$ can be expressed as follows by using the half-angle trigonometric formula (applied in (161a)):

$$\tan\left(\frac{\theta_{\mathbf{u},\mathbf{v}}}{2}\right) = \sqrt{\frac{1 - \cos(\theta_{\mathbf{u},\mathbf{v}})}{1 + \cos(\theta_{\mathbf{u},\mathbf{v}})}} \tag{161a}$$

$$= \sqrt{\frac{1 - \cos\left(\arccos\left(1 - \frac{d^2_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{2}\right)\right)}{1 + \cos\left(\arccos\left(1 - \frac{d^2_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{2}\right)\right)}} \tag{161b}$$

$$= \sqrt{\frac{\frac{d^2_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{2}}{2 - \frac{d^2_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{2}}} \tag{161c}$$

$$= \sqrt{\frac{d^2_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{4 - d^2_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}} \tag{161d}$$

$$= \frac{\alpha(x)}{x} \tag{161e}$$

Thus,

$$p_R(r \neq 0) = \Pr\left(\left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, \mathbf{Z} \right\rangle \in \left[-x\tan\left(\frac{\theta_{\mathbf{u},\mathbf{v}}}{2}\right), x\tan\left(\frac{\theta_{\mathbf{u},\mathbf{v}}}{2}\right)\right]\right) \tag{162a}$$

$$= \Pr\left(\left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, \mathbf{Z} \right\rangle \in \left[-x\frac{\alpha(x)}{x}, x\frac{\alpha(x)}{x}\right]\right) \tag{162b}$$

$$= \Pr\left(\left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, \mathbf{Z} \right\rangle \in [-\alpha(x), \alpha(x)]\right) \tag{162c}$$

But $\mathbf{Z}$ is invariant under inner products with unit vectors, and hence, the distribution of $\left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, \mathbf{Z} \right\rangle$ follows that of $\left\langle \frac{\mathbf{u}-\mathbf{v}}{\|\mathbf{u}-\mathbf{v}\|_2}, \mathbf{Z} \right\rangle \sim \mathcal{N}(0,1)$. Therefore,

$$p_R(r \neq 0) = \Pr_{Y \sim \mathcal{N}(0,1)}(Y \in [-\alpha(x), \alpha(x)]) = \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dy, \tag{163}$$

as claimed.

Combining the above derivations, the density function of $|Z_1| \mid R \neq 0$ is obtained via (148):

$$f_{|Z_1||R}(x \mid r \neq 0) = \frac{f_{|Z_1|}(x) p_{R||Z_1|}(r \neq 0 \mid x)}{p_R(r \neq 0)} = \frac{\sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dy}{\frac{\theta_{\mathbf{u},\mathbf{v}}}{\pi}} \tag{164a}$$

$$= \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dy \tag{164b}$$

if $x \geq 0$, and $f_{|Z_1||R}(x \mid r \neq 0)$ if $x < 0$, where the support of $f_{|Z_1||R}$ is restricted to the interval $[0, \infty)$ due the the latter case in (152).

The remaining task is finding the expectation of $(X \mid R \neq 0)$ to verify (140), which is done by a direct calculation using the density function, (139), that was just proved:

$$\mathbb{E}(X \mid R \neq 0) = \int_{-\infty}^{\infty} x f_{|Z_1||R}(x \mid r \neq 0) dx \tag{165a}$$

$$= \lim_{t \to \infty} \int_{x=0}^{x=t} \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} x e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dy dx \tag{165b}$$

$$= \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} \frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{2} \tag{165c}$$

$$= \sqrt{\frac{\pi}{2}} \frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{\theta_{\mathbf{u},\mathbf{v}}} \tag{165d}$$

as claimed. ■

Lemma B.6. *Fix an ordered pair of real-valued vectors,* $(\mathbf{u},\mathbf{v}) \in \mathcal{S}^{n-1} \times \mathcal{S}^{n-1}$*, of unit norm. Let* $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n \times n})$ *be a standard normal random vector, and let* $R_{\mathbf{u},\mathbf{v}}$ *be a discrete random variable given by* $R_{\mathbf{u},\mathbf{v}} = \frac{1}{2}(\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) - \mathsf{sign}(\langle \mathbf{v}, \mathbf{Z} \rangle))$*, which takes values in* $\{-1, 0, 1\}$*. Then, the distribution of the random variable* $Y_{\mathbf{u},\mathbf{v}} = \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, \mathbf{Z} \right\rangle R_{\mathbf{u},\mathbf{v}}$ *conditioned on* $R_{\mathbf{u},\mathbf{v}} \neq 0$ *is standard normal, i.e.,* $(Y_{\mathbf{u},\mathbf{v}} \mid R_{\mathbf{u},\mathbf{v}} \neq 0) \sim \mathcal{N}(0,1)$.

Proof (Lemma B.6). Analogously to the claim in the proof of Lemma B.4, it can be shown that for $\mathbf{u}, \mathbf{v}, \mathbf{u'}, \mathbf{v'} \in \mathcal{S}^{n-1}$, such that $\theta_{\mathbf{u},\mathbf{v}} = \theta_{\mathbf{u'},\mathbf{v'}}$, the random variables $(Y_{\mathbf{u},\mathbf{v}} \mid R_{\mathbf{u},\mathbf{v}} = 0)$ and $(Y_{\mathbf{u'},\mathbf{v'}} \mid R_{\mathbf{u'},\mathbf{v'}} = 0)$ follow the same distribution, as do $(Y_{\mathbf{u},\mathbf{v}} \mid R_{\mathbf{u},\mathbf{v}} \neq 0)$ and $(Y_{\mathbf{u'},\mathbf{v'}} \mid R_{\mathbf{u'},\mathbf{v'}} \neq 0)$. We will omit the formal argument since it is nearly identical to that provided in the proof of Lemma B.4.

Fix any $\theta \in [0, 2\pi)$, and let $\mathbf{u} = (u_1, \ldots, u_n) \in \mathcal{S}^{n-1}$ and take $\mathbf{v} = (u_1, -u_2 \ldots, -u_n)$ such that $u_1 > 0$ and $\theta_{\mathbf{u},\mathbf{v}} = \theta$. This construction yields

$$\frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2} = (1, 0, \ldots, 0) \tag{166}$$

as well as

$$\mathbf{u} - \mathbf{v} \propto (0, u_2, \ldots u_n) \tag{167}$$

We will again drop the subscript $\mathbf{u}, \mathbf{v}$ from the random variables for simplicity and denote $Y = Y_{\mathbf{u},\mathbf{v}}, R = R_{\mathbf{u},\mathbf{v}}$. From (166), it follows that

$$X = \left\langle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2}, \mathbf{Z} \right\rangle = Z_1 \tag{168}$$

On the other hand, observe that the event $R \neq 0$ implies that $\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) \neq \mathsf{sign}(\langle \mathbf{v}, \mathbf{Z} \rangle)$ and hence that $\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) = -\mathsf{sign}(\langle \mathbf{v}, \mathbf{Z} \rangle)$. Then,

$$\begin{aligned} R &= \frac{1}{2} \left(\mathsf{sign}\left(\langle \mathbf{u}, \mathbf{Z} \rangle \right) - \mathsf{sign}\left(\langle \mathbf{v}, \mathbf{Z} \rangle \right) \right) && \text{(169a)} \\ &= \mathsf{sign}\left(\mathsf{sign}\left(\langle \mathbf{u}, \mathbf{Z} \rangle \right) - \mathsf{sign}\left(\langle \mathbf{v}, \mathbf{Z} \rangle \right) \right) && \text{(169b)} \\ &= \mathsf{sign}\left(\mathsf{sign}\left(\langle \mathbf{u}, \mathbf{Z} \rangle \right) + \mathsf{sign}\left(\langle -\mathbf{v}, \mathbf{Z} \rangle \right) \right) && \text{(169c)} \\ &= \mathsf{sign}\left(\langle \mathbf{u} - \mathbf{v}, \mathbf{Z} \rangle \right) && \text{(169d)} \end{aligned}$$

But recall from (167) that $\mathbf{u} - \mathbf{v} \propto (0, u_2, \ldots u_n)$, and thus, given $R \neq 0$,

$$R = \mathsf{sign}\left(\langle \mathbf{u} - \mathbf{v}, \mathbf{Z} \rangle \right) = \mathsf{sign}\left(\langle (0, u_2, \ldots, u_n), \mathbf{Z} \rangle \right) \tag{170}$$

which implies conditional independence of $(R \mid R \neq 0)$ and $(Z_1 \mid R \neq 0) = (X \mid R \neq 0)$. Then, $(Y \mid R \neq 0) = (XR \mid R \neq 0) = (Z_1 R \mid R \neq 0)$, and so $(Y \mid R \neq 0)$ $(Y \mid R \neq 0)$ follows the same distribution as either the random variable Z' or $-Z'$, where $Z' \sim \mathcal{N}(0, 1)$. But it is well-known that the standard normal random variable Z' and its negation $-Z'$ have the same distribution, implying that $(Y \mid R \neq 0) \sim \mathcal{N}(0, 1)$, as claimed. ∎

Lemma B.7. *Fix an ordered pair of real-valued unit vectors,* $(\mathbf{u}, \mathbf{v}) \in \mathcal{S}^{n-1} \times \mathcal{S}^{n-1}$*, and let* $\mathbf{w} \in \mathcal{S}^{n-1} \cap \mathsf{span}\left(\{ \mathbf{u} - \mathbf{v}, \mathbf{u} + \mathbf{v} \} \right)^{\perp}$ *be any real-valued unit vector in the orthogonal complement of* $\mathsf{span}\left(\{ \mathbf{u} - \mathbf{v}, \mathbf{u} + \mathbf{v} \} \right)$*. Let* $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n \times n})$ *be a standard normal random vector, let* $\mathbf{Y}$ *be the random vector given by*

$$\mathbf{Y} = \mathbf{Z} - \left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, \mathbf{Z} \right\rangle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2} - \left\langle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2}, \mathbf{Z} \right\rangle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2}, \tag{171}$$

and let R *be the (discrete) random variable taking values in* $\{-1, 0, 1\}$ *and given by* $R = \frac{1}{2}\left(\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) - \mathsf{sign}(\langle \mathbf{v}, \mathbf{Z} \rangle) \right)$*. Then, the random vector* $X = \langle \mathbf{w}, \mathbf{Y} \rangle R$ *conditioned on* $R \neq 0$ *is standard normal, i.e.,* $(X \mid R \neq 0) \sim \mathcal{N}(0, 1)$*.*

Proof (Lemma B.7). As in the previous two lemmas, the ordered pair of unit vectors $(\mathbf{u}, \mathbf{v}) \in \mathcal{S}^{n-1} \times \mathcal{S}^{n-1}$ can be chosen nonarbitrarily due to the rotational invariance of the standard normal distribution and the argument laid out in the proof of Lemma B.4. For the purposes of this proof, we will select $\mathbf{u}$ and $\mathbf{v}$ as follows. For any pair of constants p, q, subject to $p^2 + q^2 = 1$, set $\mathbf{u} = (p, q, 0, \ldots, 0)$ and $\mathbf{v} = (-p, q, 0, \ldots, 0)$. Note that

$$\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1 \tag{172}$$

$$\mathbf{u} - \mathbf{v} = (2p, 0, \ldots, 0), \quad \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2} = (1, 0, \ldots, 0) = \mathbf{e}_1 \tag{173}$$

$$\mathbf{u} + \mathbf{v} = (0, 2q, \ldots, 0), \quad \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2} = (0, 1, \ldots, 0) = \mathbf{e}_2 \tag{174}$$

where $\mathbf{e}_1 = (1, 0, \ldots, 0), \mathbf{e}_1 = (0, 1, \ldots, 0) \in \mathbb{R}^n$ are the first and second standard basis vectors or $\mathbb{R}^n$. Fix any $\mathbf{w} \in \mathcal{S}^{n-1} \cap \mathsf{span}\left(\{\mathbf{u} - \mathbf{v}, \mathbf{u} + \mathbf{v}\}\right)^{\perp}$. Then,

$$\mathbf{Y} = \mathbf{Z} - \left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, \mathbf{Z} \right\rangle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2} - \left\langle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2}, \mathbf{Z} \right\rangle \frac{\mathbf{u} + \mathbf{v}}{\|\mathbf{u} + \mathbf{v}\|_2} \tag{175}$$

$$= \mathbf{Z} - Z_1 \mathbf{e}_1 - Z_2 \mathbf{e}_2 \tag{176}$$

$$= (0, 0, Z_3, \ldots, Z_n) \tag{177}$$

Notice that $\mathsf{span}\left(\{\mathbf{u} - \mathbf{v}, \mathbf{u} + \mathbf{v}\}\right) = \mathsf{span}\left(\{\mathbf{e}_1, \mathbf{e}_2\}\right)$ and $\mathsf{span}\left(\{\mathbf{u} - \mathbf{v}, \mathbf{u} + \mathbf{v}\}\right)^{\perp} = \mathsf{span}\left(\{\mathbf{e}_1, \mathbf{e}_2\}\right)^{\perp} = \{\mathbf{x} \in \mathbb{R}^n : x_1 = x_2 = 0\}$. Then, writing $\tilde{\mathbf{Z}} = (Z_3, \ldots, Z_n)$ and $\tilde{\mathbf{w}} = (w_3, \ldots, w_n)$, the random variable $\langle \mathbf{w}, \mathbf{Y} \rangle$ follows the same distribution as $\langle \tilde{\mathbf{w}}, \tilde{\mathbf{Z}} \rangle = \langle \frac{\tilde{\mathbf{w}}}{\|\tilde{\mathbf{w}}\|_2}, \tilde{\mathbf{Z}} \rangle$ with $\|\tilde{\mathbf{w}}\|_2 = 1$. But it is well-known that $\langle \tilde{\mathbf{w}}, \tilde{\mathbf{Z}} \rangle \sim \mathcal{N}(0, 1)$.

Recall the definition of the random variable $R = \frac{1}{2}(\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) - \mathsf{sign}(\langle \mathbf{v}, \mathbf{Z} \rangle))$. Because $\mathbf{u}, \mathbf{v} \in \mathsf{span}\left(\{\mathbf{u} - \mathbf{v}, \mathbf{u} + \mathbf{v}\}\right)$, the random variable R is entirely dependent on the projection of $\mathbf{Z}$ onto $\mathsf{span}\left(\{\mathbf{u} - \mathbf{v}, \mathbf{u} + \mathbf{v}\}\right)$ and hence independent of its projection onto $\mathsf{span}\left(\{\mathbf{u} - \mathbf{v}, \mathbf{u} + \mathbf{v}\}\right)^{\perp}$. More formally,

$$R = \frac{1}{2}\left(\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z} \rangle) - \mathsf{sign}(\langle \mathbf{v}, \mathbf{Z} \rangle)\right) \tag{178a}$$

$$= \frac{1}{2}\left(\mathsf{sign}(pZ_1 + qZ_2) - \mathsf{sign}(-pZ_1 + qZ_2)\right) \tag{178b}$$

and thus, R depends only on the random variables Z_1 and Z_2. However, it was already noted that $\mathsf{span}\left(\{\mathbf{u} - \mathbf{v}, \mathbf{u} + \mathbf{v}\}\right)^{\perp} = \{\mathbf{x} \in \mathbb{R}^n : x_1 = x_2 = 0\}$, which implies that the projection Y depend only on a (possibly improper) subset of $\{Z_j\}_{j \in [n] \setminus \{1,2\}}$. The independence of Y and R follows. Then, the conditioned random variable $(X \mid R \neq 0) = (\langle \mathbf{w}, \mathbf{Y} \rangle R \mid R \neq 0)$ is equivalent to either $\langle \mathbf{w}, \mathbf{Y} \rangle R$ or $-\langle \mathbf{w}, \mathbf{Y} \rangle R$, both of which follow the standard normal distribution. Hence, $(X \mid R \neq 0) \sim \mathcal{N}(0, 1)$. ■

B.1.2 Concentration inequalities for orthogonal projections of normal vectors

We are ready to prove Lemmas B.1-B.3. Note that the subscripts $\mathbf{u}, \mathbf{v}$ are dropped from some random variables for ease of notation.

Proof (Lemma B.1). Using the linearity of inner products, the random variable X can be written as

$$X = \left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, \sum_{i=1}^{m} \mathbf{Z}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right\rangle = \sum_{i=1}^{m} \left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, \mathbf{Z}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right\rangle = \sum_{i=1}^{m} X_i, \tag{179}$$

where the random variables $X_i = \left\langle \frac{\mathbf{u} - \mathbf{v}}{\|\mathbf{u} - \mathbf{v}\|_2}, \mathbf{Z}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right\rangle$, $i \in [m]$, are i.i.d. and have (conditional) distributions formally defined in Lemma B.4. The concentration inequality will follow from (i) controlling the MGF, $\psi_{X_i - \mu \mid R_{i;\mathbf{u},\mathbf{v}} \neq 0}$, of each zero-mean i.i.d. random variable $(X_i - \mu \mid R_{i;\mathbf{u},\mathbf{v}} \neq 0)$, such that $\psi_{X_i - \mu \mid R_{i;\mathbf{u},\mathbf{v}} \neq 0}(s) \leq e^{\frac{s^2}{2}}$. The negation of this random variable, $(-X_i + \mu \mid R_{i;\mathbf{u},\mathbf{v}} \neq 0)$, is handled likewise. (ii) Then, the MGFs of $(X - \mathbb{E}[X] \mid \|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0)$ and $(-X + \mathbb{E}[X] \mid \|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0)$ follow from step (i) and the i.i.d. property of $\{X_i\}_{i \in [m]}$. (iii) Lastly, two Chernoff bounds using the MGFs found in step (ii) will yield the lemma's two-sided bound. in (136).

Beginning with the derivation of the MGF of the i.i.d. random variables, as outlined in step (i), fix any $i \in [m]$ such that $R_{i;\mathbf{u},\mathbf{v}} \neq 0$. Then, the density function of $(X_i \mid R_{i;\mathbf{u},\mathbf{v}} \neq 0)$ is given in Eq. (139) of Lemma B.4:

$$f_{X_i \mid R_{i;\mathbf{u},\mathbf{v}}}(x \mid r \neq 0) = \begin{cases} \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dy, & \text{if } x \geq 0, \\ 0, & \text{if } x < 0. \end{cases} \tag{180}$$

with

$$\mu \overset{\mathsf{def}}{=} \mathbb{E}(X_i \mid R_{i;\mathbf{u},\mathbf{v}} \neq 0) = \sqrt{\frac{\pi}{2}} \frac{d_{\mathcal{S}^{n-1}}(\mathbf{u},\mathbf{v})}{\theta_{\mathbf{u},\mathbf{v}}}, \tag{181}$$

as specified in (140) of Lemma B.4. The MGF of $(X_i \mid R_{i;\mathbf{u},\mathbf{v}} \neq 0)$ at $s \geq 0$ is then bounded from above by

$$\psi_{X_i-\mu|R_{i;\mathbf{u},\mathbf{v}}\neq 0}(s) \leq e^{\frac{s^2}{2}} \tag{182}$$

as derived next in (183).

$$\begin{aligned}
\psi_{X_i-\mu|R_{i;\mathbf{u},\mathbf{v}}\neq 0}(s) &= \mathbb{E}\left[e^{s(X_i-\mathbb{E}(X_i|R_{i;\mathbf{u},\mathbf{v}}\neq 0))} \middle| R_{i;\mathbf{u},\mathbf{v}} \neq 0 \right] && \text{(183a)}\\
&= \mathbb{E}\left[e^{s(X_i-\mu)} \middle| R_{i;\mathbf{u},\mathbf{v}} \neq 0 \right] && \text{(183b)}\\
&= e^{-s\mu}\, \mathbb{E}\left[e^{sX_i} \middle| R_{i;\mathbf{u},\mathbf{v}} \neq 0 \right] && \text{(183c)}\\
&= e^{-s\mu} \int_{x=-\infty}^{x=\infty} e^{sx} f_{X_i|R_{i;\mathbf{u},\mathbf{v}}}(x \mid r \neq 0) dx && \text{(183d)}\\
&= e^{-s\mu} \int_{x=0}^{x=\infty} e^{sx} \cdot \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dydx && \text{(183e)}\\
&= e^{-s\mu} \int_{x=0}^{x=\infty} e^{sx} e^{-\frac{x^2}{2}} \cdot \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dydx && \text{(183f)}\\
&= e^{-s\mu} \int_{x=0}^{x=\infty} e^{-\left(\frac{x^2}{2}-sx\right)} \cdot \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dydx && \text{(183g)}\\
&= e^{-s\mu} \int_{x=0}^{x=\infty} e^{-\frac{x^2-2sx}{2}} \cdot \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dydx && \text{(183h)}\\
&= e^{-s\mu} \int_{x=0}^{x=\infty} e^{-\frac{x^2-2sx+s^2-s^2}{2}} \cdot \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dydx && \text{(183i)}\\
&= e^{-s\mu} \int_{x=0}^{x=\infty} e^{\frac{s^2}{2}} e^{-\frac{x^2-2sx+s^2}{2}} \cdot \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dydx && \text{(183j)}\\
&= e^{-s\mu} \int_{x=0}^{x=\infty} e^{\frac{s^2}{2}} e^{-\frac{(x-s)^2}{2}} \cdot \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dydx && \text{(183k)}\\
&= e^{\frac{s^2}{2}} e^{-s\mu} \int_{x=0}^{x=\infty} e^{-\frac{(x-s)^2}{2}} \cdot \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dydx && \text{(183l)}
\end{aligned}$$

Note that the function

$$q(s) = e^{-s\mu} \int_{x=0}^{x=\infty} e^{-\frac{(x-s)^2}{2}} \cdot \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dydx = \mathbb{E}\left[e^{s(X-\mu)} e^{-\frac{s^2}{2}} \right] \tag{183m}$$

decreases monotonically w.r.t. s over the interval $s \in [0, \infty)$ (see Lemma B.8). Formally, this implies

$$\max_{s\in[0,\infty)} q(s) = q(0) = 1 \tag{183n}$$

where the last equality follows from the fact that $q(0)$ reduces to the evaluation of the density function $f_{X_i|R_{i;\mathbf{u},\mathbf{v}}}$ over its entire support. Then, continuing (183a)-(183l) arrives at the desired bound, (182):

$$\begin{aligned}
\psi_{X_i-\mu|R_{i;\mathbf{u},\mathbf{v}}\neq 0}(s) &= e^{\frac{s^2}{2}} e^{-s\mu} \int_{x=0}^{x=\infty} e^{-\frac{(x-s)^2}{2}} \cdot \frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}} \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sqrt{2\pi}} \int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}} dydx && \text{(183o)}\\
&\leq e^{\frac{s^2}{2}} \cdot 1 && \text{(183p)}
\end{aligned}$$

$$= e^{\frac{s^2}{2}} \tag{183q}$$

Next, the MGF of the negated random variable, $(-X_i + \mu \mid R_{i;\mathbf{u},\mathbf{v}} \neq 0)$ is upper bounded by

$$\psi_{-X_i+\mu|R_{i;\mathbf{u},\mathbf{v}}\neq 0}(s) \leq e^{\frac{s^2}{2}}. \tag{184}$$

The derivation of (184) is similar to that above.

$$\psi_{-X_i+\mu|R_{i;\mathbf{u},\mathbf{v}}\neq 0}(s) = \mathbb{E}\left[e^{s(-X_i+\mathbb{E}(X_i|R_{i;\mathbf{u},\mathbf{v}}\neq 0))}\middle| R_{i;\mathbf{u},\mathbf{v}} \neq 0\right] \tag{185a}$$
$$= \mathbb{E}\left[e^{-s(X_i-\mu)}\middle| R_{i;\mathbf{u},\mathbf{v}} \neq 0\right] \tag{185b}$$
$$= e^{s\mu}\,\mathbb{E}\left[e^{-sX_i}\middle| R_{i;\mathbf{u},\mathbf{v}} \neq 0\right] \tag{185c}$$
$$= e^{s\mu}\int_{x=-\infty}^{x=\infty} e^{-sx} f_{X_i|R_{i;\mathbf{u},\mathbf{v}}}(x \mid r \neq 0)dx \tag{185d}$$
$$= e^{s\mu}\int_{x=0}^{x=\infty} e^{-sx}\cdot\frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}}\sqrt{\frac{2}{\pi}}e^{-\frac{x^2}{2}}\cdot\frac{1}{\sqrt{2\pi}}\int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}}dydx \tag{185e}$$
$$= e^{s\mu}\int_{x=0}^{x=\infty} e^{-sx}e^{-\frac{x^2}{2}}\cdot\frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}}\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\sqrt{2\pi}}\int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}}dydx \tag{185f}$$
$$= e^{s\mu}\int_{x=0}^{x=\infty} e^{-\left(\frac{x^2}{2}+sx\right)}\cdot\frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}}\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\sqrt{2\pi}}\int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}}dydx \tag{185g}$$
$$= e^{s\mu}\int_{x=0}^{x=\infty} e^{-\frac{x^2+2sx}{2}}\cdot\frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}}\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\sqrt{2\pi}}\int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}}dydx \tag{185h}$$
$$= e^{s\mu}\int_{x=0}^{x=\infty} e^{-\frac{x^2+2sx+s^2-s^2}{2}}\cdot\frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}}\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\sqrt{2\pi}}\int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}}dydx \tag{185i}$$
$$= e^{s\mu}\int_{x=0}^{x=\infty} e^{\frac{s^2}{2}}e^{-\frac{x^2-2sx+s^2}{2}}\cdot\frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}}\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\sqrt{2\pi}}\int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}}dydx \tag{185j}$$
$$= e^{s\mu}\int_{x=0}^{x=\infty} e^{\frac{s^2}{2}}e^{-\frac{(x+s)^2}{2}}\cdot\frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}}\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\sqrt{2\pi}}\int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}}dydx \tag{185k}$$
$$= e^{\frac{s^2}{2}}e^{-s\mu}\int_{x=0}^{x=\infty} e^{-\frac{(x+s)^2}{2}}\cdot\frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}}\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\sqrt{2\pi}}\int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}}dydx \tag{185l}$$

Again, the function

$$r(s) = e^{s\mu}\int_{x=0}^{x=\infty} e^{-\frac{(x+s)^2}{2}}\cdot\frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}}\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\sqrt{2\pi}}\int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}}dydx = \mathbb{E}\left[e^{-s(X-\mu)}e^{-\frac{s^2}{2}}\right] \tag{185m}$$

decreases monotonically w.r.t. $s \in [0, \infty)$ (see, again, Lemma B.8), and thus

$$\max_{s\in[0,\infty)} r(s) = r(0) = 1 \tag{185n}$$

where, as before, the last equality holds because $r'(0)$ simply evaluates the density function $f_{X_i|R_{i;\mathbf{u},\mathbf{v}}}$ over its entire support. Then, the desired bound in (184) can now be established by continuing from (185a)-(185l) as follows.

$$\psi_{-X_i+\mu|R_{i;\mathbf{u},\mathbf{v}}\neq 0}(s) = e^{\frac{s^2}{2}}e^{s\mu}\int_{x=0}^{x=\infty} e^{-\frac{(x+s)^2}{2}}\cdot\frac{\pi}{\theta_{\mathbf{u},\mathbf{v}}}\sqrt{\frac{2}{\pi}}\cdot\frac{1}{\sqrt{2\pi}}\int_{y=-\alpha(x)}^{y=\alpha(x)} e^{-\frac{y^2}{2}}dydx \tag{185o}$$
$$\leq e^{\frac{s^2}{2}}\cdot 1 \tag{185p}$$
$$= e^{\frac{s^2}{2}} \tag{185q}$$

Note that (182) and (184) holds likewise for every $i \in [m]$. This completes the first outline step.

The second task, outlined in (ii), is controlling the MGFs of the sums of i.i.d. random variables, $(X - \mathbb{E}[X] \mid L = \ell)$ and $(-X + \mathbb{E}[X] \mid L = \ell)$. Writing $\mu_X = \mathbb{E}[X \mid L = \ell]$, the MGF for the sum of i.i.d. random variables, $X = \sum_{i=1}^{m}(X_i - \mu)$, conditioned on $L = \ell$ can then be bounded from above as follows.

$$\psi_{X-\mu_X \,|\, \|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0=\ell}(s) = \mathbb{E}\left[e^{s(X-\mu_X)}\middle| L = \ell\right] \tag{186a}$$
$$= \mathbb{E}\left[e^{s\sum_{i=1}^{m}(X_i-\mu)}\middle| L = \ell\right] \tag{186b}$$
$$= \mathbb{E}\left[e^{s\sum_{i\in\mathsf{supp}(\|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0)}(X_i-\mu)}\right] \tag{186c}$$
$$= \prod_{i\in\mathsf{supp}(\|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0)} \mathbb{E}\left[e^{s(X_i-\mu)}\right], \qquad \blacktriangleright \because \text{the random variables } X_i, i \in [m] \text{ are independent} \tag{186d}$$
$$= \mathbb{E}\left[e^{s(X_i-\mu)}\right]^{\ell}, \qquad \blacktriangleright \because \text{the random variables } X_i, i \in [m] \text{ are identically distributed} \tag{186e}$$
$$\leq e^{\frac{1}{2}\ell s^2}, \qquad \blacktriangleright \text{ by (182)} \tag{186f}$$

Moreover, by an analogous argument, the MGF of the negated random variable $(-X - \mathbb{E}[-X] \mid L = \ell) = (-X + \mathbb{E}[X] \mid L = \ell)$ can be upper bounded. Notice that $-X = -\sum_{i=1}^{m}(X_i - \mu) = \sum_{i=1}^{m}(-X_i + \mu)$, which allows the MGF of $-X + \mathbb{E}[X]$ conditioned on $L = \ell$ to be upper bounded by the following.

$$\psi_{-X+\mu_X \,|\, \|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0=\ell}(s) = \mathbb{E}\left[e^{s(-X+\mu_X)}\middle| L = \ell\right] \tag{187a}$$
$$= \mathbb{E}\left[e^{s\sum_{i=1}^{m}(-X_i+\mu)}\middle| L = \ell\right] \tag{187b}$$
$$= \mathbb{E}\left[e^{s\sum_{i\in\mathsf{supp}(\|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0)}(-X_i+\mu)}\right] \tag{187c}$$
$$= \prod_{i\in\mathsf{supp}(\|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0)} \mathbb{E}\left[e^{s(-X_i+\mu)}\right], \qquad \blacktriangleright \because \text{the random variables } X_i, i \in [m] \text{ are independent} \tag{187d}$$
$$= \mathbb{E}\left[e^{s(-X_i+\mu)}\right]^{\ell}, \qquad \blacktriangleright \because \text{the random variables } X_i, i \in [m] \text{ are identically distributed} \tag{187e}$$
$$\leq e^{\frac{1}{2}\ell s^2}, \qquad \blacktriangleright \text{ by (184)} \tag{187f}$$

To summarize, this step, (ii), has shown

$$\psi_{X-\mu_X \,|\, \|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0=\ell}(s) \leq e^{\frac{1}{2}\ell s^2} \tag{188}$$
$$\psi_{-X+\mu_X \,|\, \|\mathbf{R}_{\mathbf{u},\mathbf{v}}\|_0=\ell}(s) \leq e^{\frac{1}{2}\ell s^2}. \tag{189}$$

The aim in the final outlined step, (iii), is bounding X from each sides by a Chernoff bound and subsequently union bounding to obtain the lemma's two-sided result. The upper bound, derived first, will use the MGF of $(X - \mathbb{E}[X] \mid L = \ell)$, while the lower bound will use the MGF of $(-X + \mathbb{E}[X] \mid L = \ell)$. In both cases, the bounds will be shown to fail with probability not exceeding $e^{-\frac{1}{2}\ell t^2}$. For the upper bound,

$$\Pr\left(X - \mathbb{E}\left[X | L = \ell\right] \geq \ell t \,\middle| L = \ell\right) \tag{190a}$$
$$= \Pr\left(X - \mu_X \geq \ell t \,\middle| L = \ell\right) \tag{190b}$$
$$= \Pr\left(e^{X-\mathbb{E}[X|L=\ell]} \geq e^{\ell t} \,\middle| L = \ell\right) \tag{190c}$$
$$\leq \min_{s\geq 0} e^{-\ell s t} \cdot \psi_{X-\mu_X|L=\ell}(s), \qquad \blacktriangleright \text{ due to Bernstein (see, e.g., Vershynin (2018))} \tag{190d}$$

$$\leq \min_{s\geq 0} e^{-\ell st} e^{\frac{1}{2}\ell s^2}, \qquad \blacktriangleright \text{ by Eq. (188)} \tag{190e}$$

$$= \min_{s\geq 0} e^{-\ell\left(st-\frac{s^2}{2}\right)} \tag{190f}$$

A maximizer of $st - \frac{s^2}{2}$ a minimizer of $e^{-\ell(st-\frac{s^2}{2})}$. The unique zero of $\frac{\partial}{\partial s} st - \frac{s^2}{2}$ is at $s = t$ (moreover, $\frac{\partial^2}{\partial s^2} st - \frac{s^2}{2} < 0$ and hence this is indeed a (global) maximum). Note additionally that setting $s = t$ ensures that $s \in [0, 1]$, which was assumed in step (i). Then, continuing from above,

$$\Pr\left(X - \mathbb{E}\left[X|L=\ell\right] \geq \ell t \,\middle|\, L=\ell\right) \leq \min_{s\geq 0} e^{-\ell\left(st-\frac{s^2}{2}\right)} \tag{190g}$$

$$= e^{-\ell\left(t^2-\frac{t^2}{2}\right)}, \qquad \blacktriangleright \text{ as argued above} \tag{190h}$$

$$\leq e^{-\frac{1}{2}\ell t^2} \tag{190i}$$

as desired. The derivation of the lower bound is nearly identical, as seen next.

$$\Pr\left(X - \mathbb{E}\left[X|L=\ell\right] \leq -\ell t \,\middle|\, L=\ell\right) \tag{191a}$$

$$= \Pr\left(-X + \mathbb{E}\left[X|L=\ell\right] \geq \ell t \,\middle|\, L=\ell\right) \tag{191b}$$

$$= \Pr\left(-X + \mu_X \geq \ell t \,\middle|\, L=\ell\right) \tag{191c}$$

$$= \Pr\left(e^{-X+\mathbb{E}[X|L=\ell]} \geq e^{\ell t} \,\middle|\, L=\ell\right) \tag{191d}$$

$$\leq \min_{s\geq 0} e^{-\ell st} \cdot \psi_{-X+\mu_X|L=\ell}(s), \qquad \blacktriangleright \text{ due to Bernstein (see, e.g., Vershynin (2018))} \tag{191e}$$

$$\leq \min_{s\geq 0} e^{-\ell st} e^{\frac{1}{2}\ell s^2}, \qquad \blacktriangleright \text{ by Eq. (189)} \tag{191f}$$

$$= \min_{s\geq 0} e^{-\ell\left(st-\frac{s^2}{2}\right)} \tag{191g}$$

$$= e^{-\ell\left(t^2-\frac{t^2}{2}\right)}, \qquad \blacktriangleright \text{ the same minimization problem as (190f),} \tag{191h}$$

$$\text{whose solution is at } s = t \tag{191i}$$

$$= e^{-\frac{1}{2}\ell t^2} \tag{191j}$$

Thus far, it has been shown that

$$\Pr\left(X - \mathbb{E}\left[X|L=\ell\right] \geq \ell t \,\middle|\, L=\ell\right) \leq e^{-\frac{1}{2}\ell t^2}, \tag{192}$$

$$\Pr\left(X - \mathbb{E}\left[X|L=\ell\right] \leq -\ell t \,\middle|\, L=\ell\right) \leq e^{-\frac{1}{2}\ell t^2}. \tag{193}$$

To complete the proof, (192) and (193) are combined by a union bound, yielding the lemma's concentration inequality,

$$\Pr\left(\left|X - \mathbb{E}\left[X|L=\ell\right]\right| \geq \ell t \,\middle|\, L=\ell\right) \leq 2e^{-\frac{1}{2}\ell t^2}. \tag{194}$$

∎

Proof (Lemma B.2). As in the proof of Lemma B.4, let $X_i = \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, \mathbf{Z}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right\rangle$ for each $i \in [m]$, which are i.i.d. with (conditional) distributions described in Lemma B.6. Then the random variable X can be written as

$$X = \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, \sum_{i=1}^{m} \mathbf{Z}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right\rangle = \sum_{i=1}^{m} \left\langle \frac{\mathbf{u}+\mathbf{v}}{\|\mathbf{u}+\mathbf{v}\|_2}, \mathbf{Z}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right\rangle = \sum_{i=1}^{m} X_i. \tag{195}$$

Recall from Lemma B.6 that for each $i \in [m]$, the random variable $(X_i \mid A_{i;\mathbf{u},\mathbf{v}} \neq 0)$ is standard normal. It follows

$$X = \sum_{i=1}^{m} X_i = \sum_{i \in \mathsf{supp}(\mathbf{R}_{\mathbf{u},\mathbf{v}})} (X_i \mid A_{i;\mathbf{u},\mathbf{v}} \neq 0) \sim \mathcal{N}(0, \sigma^2 = \ell) \tag{196}$$

where the distribution of X depends only on the number ℓ of random variables summed up but not the exact subset $\mathsf{supp}(\mathbf{R}_{\mathbf{u},\mathbf{v}}) \subseteq [n]$ (since the random variables X_i, $i \in [m]$, are identically distributed). Therefore,

$$\Pr\left(|X| \geq t' | L = \ell\right) \leq 2e^{-\frac{t'^2}{2\ell}}. \tag{197}$$

Taking $t' = \ell t$, (197) implies

$$\Pr\left(|X| \geq \ell t | L = \ell\right) \leq 2e^{-\frac{\ell^2 t^2}{2\ell}} = 2e^{-\frac{1}{2}\ell t^2}. \tag{198}$$

Thus proved. ■

Proof (Lemma B.3). By Lemma B.7, for each $i \in [m]$, $\mathbf{Z}^{(i)} R_{i;\mathbf{u},\mathbf{v}}$ Write $J' = J \cap (\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v}))$ and $J'' = J \setminus (\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v}))$. By the triangle inequality,

$$\begin{aligned}
\left\| \mathcal{T}_J \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 &= \left\| \mathcal{T}_{J'} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) + \mathcal{T}_{J''} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 && \text{(199a)} \\
&\leq \left\| \mathcal{T}_{J'} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 + \left\| \mathcal{T}_{J''} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 && \text{(199b)} \\
&\leq \left\| \mathcal{T}_{\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 + \left\| \mathcal{T}_{J''} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 . && \text{(199c)}
\end{aligned}$$

Let $d' = |\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})|$ and $\mathbf{V}^{(i)} = V_1^{(i)}, \ldots, V_{d'-2}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{(d'-2)\times(d'-2)})$, $i \in [m]$, and suppose $\{\mathbf{b}_j \in \mathbb{R}^n\}_{j \in [d'-2]}$ is an orthonormal basis over $\mathsf{span}\left(\{\mathbf{u}-\mathbf{v}, \mathbf{u}+\mathbf{v}\}\right)^\perp \cap \{\mathbf{x} \in \mathbb{R}^n : \mathsf{supp}(\mathbf{x}) \subseteq \mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})\}$ with $\mathbf{Y}^{(i)} = \sum_{j=1}^{d'-2} \langle \mathbf{b}_j, \mathbf{Y}^{(i)} \rangle \mathbf{b}_j$. Due to Lemma B.7, $\langle \mathbf{b}_j, \mathbf{Y}^{(i)} \rangle \sim \mathcal{N}(0,1)$.

$$\begin{aligned}
\left\| \mathcal{T}_{\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 &= \left\| \sum_{i=1}^{m} \mathcal{T}_{\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})} \left(\mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 && \text{(200a)} \\
&= \left\| \sum_{\substack{i=1: \\ R_{i;\mathbf{u},\mathbf{v}} \neq 0}}^{m} \sum_{j=1}^{d'-2} \langle \mathbf{b}_j, \mathbf{Y}^{(i)} \rangle \mathbf{b}_j \right\|_2 && \text{(200b)} \\
&= \left\| \sum_{j=1}^{d'-2} \mathbf{b}_j \sum_{\substack{i=1: \\ R_{i;\mathbf{u},\mathbf{v}} \neq 0}}^{m} \langle \mathbf{b}_j, \mathbf{Y}^{(i)} \rangle \right\|_2 && \text{(200c)} \\
&= \left(\sum_{j=1}^{d'-2} \sum_{j'=1}^{d'-2} \langle \mathbf{b}_j, \mathbf{b}_{j'} \rangle \left(\sum_{\substack{i=1: \\ R_{i;\mathbf{u},\mathbf{v}} \neq 0}}^{m} \langle \mathbf{b}_j, \mathbf{Y}^{(i)} \rangle \right)^2 \right)^{\frac{1}{2}} && \text{(200d)} \\
&= \left(\sum_{j=1}^{d'-2} \left(\sum_{\substack{i=1: \\ R_{i;\mathbf{u},\mathbf{v}} \neq 0}}^{m} \langle \mathbf{b}_j, \mathbf{Y}^{(i)} \rangle \right)^2 \right)^{\frac{1}{2}} && \text{(200e)}
\end{aligned}$$

$$\sim \left(\sum_{j=1}^{d'-2} \left(\sum_{\substack{i=1:\\ R_{i;\mathbf{u},\mathbf{v}} \neq 0}}^{m} V_j^{(i)} \right)^2 \right)^{\frac{1}{2}} \tag{200f}$$

$$= \left\| \sum_{\substack{i=1:\\ R_{i;\mathbf{u},\mathbf{v}} \neq 0}}^{m} \mathbf{V}^{(i)} \right\|_2 \tag{200g}$$

$$\sim \left\| \sum_{i}^{\ell} \mathbf{V}^{(i)} \right\|_2 . \tag{200h}$$

Then, by a standard Chernoff bound for standard normal random vectors (see, e.g., Corollary D.8 later in the appendix),

$$\Pr\left(\left\| \mathcal{T}_{J'} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 > \sqrt{2k\ell} + \frac{1}{2}\ell t \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}}, L_{\mathbf{u},\mathbf{v}} = \ell \right) \tag{201}$$

$$\leq \Pr\left(\left\| \mathcal{T}_{\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 > \sqrt{2k\ell} + \frac{1}{2}\ell t \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}}, L_{\mathbf{u},\mathbf{v}} = \ell \right) \tag{202}$$

$$= \Pr\left(\left\| \sum_{i=1}^{\ell} \mathbf{V}^{(i)} \right\|_2 > \sqrt{2k\ell} + \frac{1}{2}\ell t \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}}, L_{\mathbf{u},\mathbf{v}} = \ell \right) \tag{203}$$

$$\leq \Pr\left(\left\| \sum_{i=1}^{\ell} \mathbf{V}^{(i)} \right\|_2 > \mathbb{E}\left[\left\| \sum_{i=1}^{\ell} \mathbf{V}^{(i)} \right\|_2 \right] + \frac{1}{2}\ell t \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}}, L_{\mathbf{u},\mathbf{v}} = \ell \right) \tag{204}$$

▶ due to Lemma D.9

$$\leq e^{-\frac{1}{8}\ell t^2} \qquad \text{▶ due to Corollary D.8, with the parameter set as } \sigma^2 = \frac{\ell}{m^2} \tag{205}$$

On the other hand, observe,

$$\left\| \mathcal{T}_{J''} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 = \left\| \sum_{i=1}^{m} \sum_{j \in J''} \left\langle \mathbf{e}_j, \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right\rangle \mathbf{e}_j \right\|_2 \tag{206a}$$

$$= \left\| \sum_{i=1}^{m} \sum_{j \in J''} Y_j^{(i)} R_{i;\mathbf{u},\mathbf{v}} \mathbf{e}_j \right\|_2 \tag{206b}$$

Let $d'' = |J''|$ and $\mathbf{W}^{(i)} = (W_1, \ldots, W_{d''}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{d'' \times d''})$, $i \in [m]$. Due to Lemma B.7, $(\| \sum_{j \in J''} Y_j^{(i)} R_{i;\mathbf{u},\mathbf{v}} \mathbf{e}_j \|_2 \mid R_{i;\mathbf{u},\mathbf{v}} \neq 0)$ and $\|\mathbf{W}^{(i)}\|_2$, $i \in [m]$, share the same distribution. Then, by a standard Chernoff bound for standard normal random vectors,

$$\Pr\left(\left\| \mathcal{T}_{J''} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 > \sqrt{d\ell} + \frac{1}{2}\ell t \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}}, L_{\mathbf{u},\mathbf{v}} = \ell \right) \tag{207}$$

$$\leq \Pr\left(\left\| \mathcal{T}_{J''} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 > \sqrt{d''\ell} + \frac{1}{2}\ell t \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}}, L_{\mathbf{u},\mathbf{v}} = \ell \right) \tag{208}$$

$$= \Pr\left(\left\| \sum_{\substack{i=1:\\ R_{i;\mathbf{u},\mathbf{v}} \neq 0}}^{m} \mathbf{W}^{(i)} \right\|_2 > \sqrt{d''\ell} + \frac{1}{2}\ell t \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}}, L_{\mathbf{u},\mathbf{v}} = \ell \right) \tag{209}$$

$$\leq \Pr\left(\left\| \sum_{i=1}^{\ell} \mathbf{W}^{(i)} \right\|_2 > \mathbb{E}\left[\left\| \sum_{i=1}^{\ell} \mathbf{W}^{(i)} \right\|_2 \right] + \frac{1}{2}\ell t \middle| \mathbf{R}_{\mathbf{u},\mathbf{v}}, L_{\mathbf{u},\mathbf{v}} = \ell \right) \tag{210}$$

$$\blacktriangleright \text{ again due to Lemma D.9} \tag{211}$$

$$\leq e^{-\frac{1}{8}\ell t^2} \qquad \blacktriangleright \text{ again due to Corollary D.8, with the parameter set as } \sigma^2 = \frac{\ell}{m^2} \tag{212}$$

Then, since

$$\sqrt{2k\ell} + \frac{1}{2}\ell t + \sqrt{d\ell} + \frac{1}{2}\ell t = \left(\sqrt{2k} + \sqrt{d}\right)\sqrt{\ell} + \ell t \tag{213}$$

and

$$\left\| \mathcal{T}_J \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 \leq \left\| \mathcal{T}_{\mathsf{supp}(\mathbf{u}) \cup \mathsf{supp}(\mathbf{v})} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 + \left\| \mathcal{T}_{J''} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2, \tag{214}$$

it follows from a union bound that

$$\Pr\left(\left\| \mathcal{T}_J \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 \geq \left(\sqrt{2k} + \sqrt{d}\right)\sqrt{\ell} + \ell t \,\middle|\, \mathbf{R}_{\mathbf{u},\mathbf{v}}, L_{\mathbf{u},\mathbf{v}} = \ell \right) \tag{215a}$$

$$\leq \Pr\left(\left\| \mathcal{T}_{J'} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 \geq \sqrt{2k\ell} + \frac{1}{2}\ell t \,\middle|\, \mathbf{R}_{\mathbf{u},\mathbf{v}}, L_{\mathbf{u},\mathbf{v}} = \ell \right) \tag{215b}$$

$$+ \Pr\left(\left\| \mathcal{T}_{J''} \left(\sum_{i=1}^{m} \mathbf{Y}^{(i)} R_{i;\mathbf{u},\mathbf{v}} \right) \right\|_2 \geq \sqrt{d\ell} + \frac{1}{2}\ell t \,\middle|\, \mathbf{R}_{\mathbf{u},\mathbf{v}}, L_{\mathbf{u},\mathbf{v}} = \ell \right) \tag{215c}$$

$$\leq 2e^{-\frac{1}{8}\ell t^2} \tag{215d}$$

■

B.1.3 Proof of Lemma B.8

Lemma B.8. *Let X be a random variable with a finite, positive mean $\mu = \mathbb{E}[X]$ and a density function f of the form*

$$f(x) = \begin{cases} \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2}} p(x), & \text{if } x \geq 0, \\ 0, & \text{if } x < 0, \end{cases} \tag{216}$$

where the image of the function $p : \mathbb{R} \to \mathbb{R}$ is given by $p(x) = \frac{\pi}{\theta} \frac{1}{\sqrt{2\pi}} \int_{y=-x\tan(\frac{\theta}{2})}^{y=x\tan(\frac{\theta}{2})} e^{-\frac{y^2}{2}} dy$ for $x \in \mathbb{R}$. Define the functions $q, r : \mathbb{R} \to \mathbb{R}$ by

$$q(s) = \mathop{\mathbb{E}}_{X \sim f}\left[e^{s(X-\mu)} e^{-\frac{s^2}{2}} \right] \tag{217}$$

$$r(s) = \mathop{\mathbb{E}}_{X \sim f}\left[e^{-s(X-\mu)} e^{-\frac{s^2}{2}} \right] \tag{218}$$

for $s \in \mathbb{R}$. Then, $q(s)$ and $r(s)$ monotonically decrease with s over the interval $s \in [0, \infty)$.

Proof (Lemma B.8). Let $s \in \mathbb{R}$, $f, p, q, r : \mathbb{R} \to \mathbb{R}$ be satisfy the conditions of the lemma. Notice that q, r can be expressed as

$$q(s) = \int_{x=-\infty}^{x=\infty} e^{s(x-\mu)} e^{-\frac{s^2}{2}} f(x) dx = \int_{x=0}^{x=\infty} \sqrt{\frac{2}{\pi}} e^{-s\mu} e^{-\frac{(x-s)^2}{2}} p(x) dx \tag{219}$$

$$r(s) = \int_{x=-\infty}^{x=\infty} e^{-s(x-\mu)} e^{-\frac{s^2}{2}} f(x) dx = \int_{x=0}^{x=\infty} \sqrt{\frac{2}{\pi}} e^{s\mu} e^{-\frac{(x+s)^2}{2}} p(x) dx \tag{220}$$

The functions q, r can be shown to (non-strictly) monotonically decrease with s over the interval $s \in [0, \infty)$ by verifying that their partial derivatives w.r.t. s are non-positive on this interval, which will be argued by contradiction. First, suppose $q(s)$ is not monotonically decreasing with s over all $s \geq 0$, such that there exists $s' \geq 0$ for which $\frac{\partial}{\partial s} q(s)\big|_{s=s'} > 0$. Write $p'(a, b) = \frac{\pi}{\theta} \frac{1}{\sqrt{2\pi}} \int_{a \tan(\frac{\theta}{2})}^{b \tan(\frac{\theta}{2})} e^{-\frac{y^2}{2}} dy$, $a \leq b \in \mathbb{R}$, and notice that $p'(a, b) \leq p'(0, b - a)$. Then, observe

$$\left.\frac{\partial}{\partial s} q(s)\right|_{s=s'} \tag{221a}$$

$$= \left.\frac{\partial}{\partial s} \int_{x=0}^{x=\infty} \sqrt{\frac{2}{\pi}} e^{-s\mu} e^{-\frac{(x-s)^2}{2}} p(x) dx \right|_{s=s'} \tag{221b}$$

$$= \left.\int_{x=0}^{x=\infty} \frac{\partial}{\partial s} \sqrt{\frac{2}{\pi}} e^{-s\mu} e^{-\frac{(x-s)^2}{2}} p(x) dx \right|_{s=s'} \tag{221c}$$

$$= \left.\int_{x=0}^{x=\infty} (x - s - \mu) \sqrt{\frac{2}{\pi}} e^{-s\mu} e^{-\frac{(x-s)^2}{2}} p(x) dx \right|_{s=s'} \tag{221d}$$

$$= \int_{x=0}^{x=\infty} (x - s' - \mu) \sqrt{\frac{2}{\pi}} e^{-s\mu} e^{-\frac{(x-s')^2}{2}} p(x) dx \tag{221e}$$

$$= e^{-s'\mu} \int_{x=0}^{x=\infty} (x - s' - \mu) \sqrt{\frac{2}{\pi}} e^{-\frac{(x-s')^2}{2}} p(x) dx \tag{221f}$$

$$= e^{-s'\mu} \int_{u=-s'}^{u=\infty} (u - \mu) \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} p(u + s') du, \qquad \blacktriangleright\ u = x - s' \tag{221g}$$

$$= e^{-s'\mu} \int_{u=-s'}^{u=\infty} (u - \mu) \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} \left(p(u) + 2p'(u, u + s')\right) du \tag{221h}$$

$$= e^{-s'\mu} \left(\int_{u=-s'}^{u=\infty} u \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} \left(p(u) + 2p'(u, u + s')\right) du - \mu \int_{u=-s'}^{u=\infty} \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} \left(p(u) + 2p'(u, u + s')\right) du \right) \tag{221i}$$

$$= e^{-s'\mu} \Bigg(\int_{u=-s'}^{u=0} u \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} \left(p(u) + 2p'(u, u + s')\right) du + \int_{u=0}^{u=\infty} u \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} \left(p(u) + 2p'(u, u + s')\right) du \tag{221j}$$

$$-\mu \int_{u=-s'}^{u=0} \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} \left(p(u) + 2p'(u, u + s')\right) du - \mu \int_{u=0}^{u=\infty} \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} \left(p(u) + 2p'(u, u + s')\right) du \Bigg)$$

$$\leq e^{-s'\mu} \left(\int_{u=0}^{u=\infty} u \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} \left(p(u) + 2p'(u, u + s')\right) du - \mu \int_{u=0}^{u=\infty} \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} \left(p(u) + 2p'(u, u + s')\right) du \right), \tag{221k}$$

$\blacktriangleright$ the first integral in (221j) is nonpositive; the third is nonnegative

$$\leq e^{-s'\mu} \left(\int_{u=0}^{u=\infty} u \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} \left(p(u) + 2p'(0, s')\right) du - \mu \int_{u=0}^{u=\infty} \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} \left(p(u) + 2p'(0, s')\right) du \right) \tag{221l}$$

$\blacktriangleright$ at $s = s'$, $\frac{\partial}{\partial s} q(s) > 0$ by assumption

$$= e^{-s'\mu} \Bigg(\int_{u=0}^{u=\infty} u \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} p(u) du + 2p'(0, s') \int_{u=0}^{u=\infty} u \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} du \tag{221m}$$

$$-\mu \int_{u=0}^{u=\infty} \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} p(u) du - 2\mu p'(0, s') \int_{u=0}^{u=\infty} \sqrt{\frac{2}{\pi}} e^{-\frac{u^2}{2}} du \Bigg) \tag{221n}$$

$$= e^{-s'\mu}\left(\int_{u=0}^{u=\infty} uf(u)du + 2p'(0,s')\int_{u=0}^{u=\infty} uf_{|Z|}(u)du - \mu\int_{u=0}^{u=\infty} f(u)du - 2\mu p'(0,s')\int_{u=0}^{u=\infty} f_{|Z|}(u)du\right) \tag{221o}$$

$$= e^{-s'\mu}\left(\mu + 2\sqrt{\frac{2}{\pi}}p'(0,s') - \mu - 2\mu p'(0,s')\right) \tag{221p}$$

$$= e^{-s'\mu}\left((\mu-\mu) + 2p'(0,s')(\sqrt{\frac{2}{\pi}} - \mu)\right) \tag{221q}$$

$$\leq 0, \qquad \blacktriangleright \text{ equality only if } \theta = \pi \tag{221r}$$

But this shows that $\frac{\partial}{\partial s}q(s)\big|_{s=s'} \leq 0$ which is a contradiction. Hence, monotonicity of q holds.

Now consider $r(s)$, and again assume there exists $s' \geq 0$ such that $\frac{\partial}{\partial s}r(s)\big|_{s=s'} > 0$. The following will similarly arrive at a contradiction.

$$\frac{\partial}{\partial s}r(s)\bigg|_{s=s'} \tag{222a}$$

$$= \frac{\partial}{\partial s}\int_{x=0}^{x=\infty}\sqrt{\frac{2}{\pi}}e^{s\mu}e^{-\frac{(x+s)^2}{2}}p(x)dx\bigg|_{s=s'} \tag{222b}$$

$$= \int_{x=0}^{x=\infty}\frac{\partial}{\partial s}\sqrt{\frac{2}{\pi}}e^{s\mu}e^{-\frac{(x+s)^2}{2}}p(x)dx\bigg|_{s=s'} \tag{222c}$$

$$= \int_{x=0}^{x=\infty}(\mu - s - x)\sqrt{\frac{2}{\pi}}e^{s\mu}e^{-\frac{(x+s)^2}{2}}p(x)dx\bigg|_{s=s'} \tag{222d}$$

$$= \int_{x=0}^{x=\infty}(\mu - s - x)\sqrt{\frac{2}{\pi}}e^{s\mu}e^{-\frac{(x+s')^2}{2}}p(x)dx \tag{222e}$$

$$\leq \int_{x=0}^{x=\infty}(\mu - s - x)\sqrt{\frac{2}{\pi}}e^{s'\mu}e^{-\frac{(x+s')^2}{2}}p(x)dx, \qquad \blacktriangleright \text{ at } s = s',\ \frac{\partial}{\partial s}r(s) > 0 \text{ by assumption} \tag{222f}$$

$$= e^{s'\mu}\int_{x=0}^{x=\infty}(\mu - s - x)\sqrt{\frac{2}{\pi}}e^{-\frac{(x+s')^2}{2}}p(x)dx \tag{222g}$$

$$= e^{s'\mu}\int_{u=s'}^{u=\infty}(\mu - u)\sqrt{\frac{2}{\pi}}e^{-\frac{u^2}{2}}p(u-s')du, \qquad \blacktriangleright\ u = x + s' \tag{222h}$$

$$\leq e^{s'\mu}\int_{u=s'}^{u=\infty}(\mu - u)\sqrt{\frac{2}{\pi}}e^{-\frac{u^2}{2}}p(u)du, \qquad \blacktriangleright \text{ equality only if } s' = 0 \tag{222i}$$

$$= e^{s'\mu}\left(\int_{u=0}^{u=\infty}(\mu - u)\sqrt{\frac{2}{\pi}}e^{-\frac{u^2}{2}}p(u)du - \int_{u=0}^{u=s'}(\mu - u)\sqrt{\frac{2}{\pi}}e^{-\frac{u^2}{2}}p(u)du\right) \tag{222j}$$

$$\leq e^{s'\mu}\int_{u=0}^{u=\infty}(\mu - u)\sqrt{\frac{2}{\pi}}e^{-\frac{u^2}{2}}p(u)du, \qquad \blacktriangleright \text{ the right integral in (222j) is nonnegative} \tag{222k}$$

$$= e^{s'\mu}\left(\mu\int_{u=0}^{u=\infty}\sqrt{\frac{2}{\pi}}e^{-\frac{u^2}{2}}p(u)du - \int_{u=0}^{u=\infty}u\sqrt{\frac{2}{\pi}}e^{-\frac{u^2}{2}}p(u)du\right) \tag{222l}$$

$$= e^{s'\mu}\left(\mu\int_{u=0}^{u=\infty}f(u)du - \int_{u=0}^{u=\infty}uf(u)du\right) \tag{222m}$$

$$= e^{s'\mu}(\mu - \mu) \tag{222n}$$

$$= 0 \tag{222o}$$

Thus, $\frac{\partial}{\partial s}r(s)\big|_{s=s'} \leq 0$ implies $\frac{\partial}{\partial s}r(s)\big|_{s=s'} \leq 0$, a contradiction. Therefore, the monotonicity of r also holds. ■

B.1.4 Proof of Lemma A.2

Lemma (restatement) (Lemma A.2)**.** *Fix* $t \in (0,1)$, $\beta \in [0, \frac{\pi}{2}]$. *Let* $\mathbf{u} \in \mathbb{R}^n$, *and define the random variable* $M_{\beta,\mathbf{u}} = |\{\mathbf{A}^{(i)}, i \in [m] : \theta_{\mathbf{u},\mathbf{A}^{(i)}} \in \left[\frac{\pi}{2} - \beta, \frac{\pi}{2} + \beta\right]\}|$. *Then,*

$$\mu_{M_{\beta,\mathbf{u}}} = \mathbb{E}\left[M_{\beta,\mathbf{u}}\right] = \frac{2}{\pi}\beta m$$

and

$$\Pr\left(M_{\beta,\mathbf{u}} \notin \left[(1-t)\mu_{M_{\beta,\mathbf{u}}}, (1+t)\mu_{M_{\beta,\mathbf{u}}}\right]\right) \leq 2e^{-\frac{1}{3}\mu_{M_{\beta,\mathbf{u}}} t^2}.$$

Proof (Lemma A.2). Denote $\mathcal{H} = \{\mathbf{A}^{(i)}, i \in [m] : \theta_{\mathbf{u},\mathbf{A}^{(i)}} \in \left[\frac{\pi}{2} - \beta, \frac{\pi}{2} + \beta\right]\}$. It is well known that standard normal vectors $\mathbf{A}^{(i)} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n\times n})$, $i \in [m]$, with i.i.d. entries are rotationally uniform. Hence, each i^{th} indicator random variable $I_i = \mathbf{1}\Big(\mathbf{A}^{(i)} \in \mathcal{H}\Big)$, $i \in [m]$, has

$$\Pr\left(I_i = 1\right) = 2 \cdot \frac{2\beta}{2\pi} = \frac{2\beta}{\pi}. \tag{223}$$

Moreover, $M_{\beta,\mathbf{u}} = \sum_{i=1}^{m} I_i$, and by the linearity of expectation and the fact that the random variables $\{I_i\}_{i\in[m]}$ are i.i.d.,

$$\mu_{M_{\beta,\mathbf{u}}} = \mathbb{E}\left[M_{\beta,\mathbf{u}}\right] = \frac{2\beta m}{\pi} \tag{224}$$

as desired. Using standard Chernoff bounds, for any $t \in (0,1)$,

$$\Pr\left(M_{\beta,\mathbf{u}} < (1-t)\mu_{M_{\beta,\mathbf{u}}}\right) \leq e^{-\frac{1}{2}\mu_{M_{\beta,\mathbf{u}}} t^2} \tag{225}$$

$$\Pr\left(M_{\beta,\mathbf{u}} > (1+t)\mu_{M_{\beta,\mathbf{u}}}\right) \leq e^{-\frac{1}{3}\mu_{M_{\beta,\mathbf{u}}} t^2} \tag{226}$$

and via a union bound,

$$\Pr\left(M_{\beta,\mathbf{u}} \notin \left[(1-t)\mu_{M_{\beta,\mathbf{u}}}, (1+t)\mu_{M_{\beta,\mathbf{u}}}\right]\right) \leq e^{-\frac{1}{2}\mu_{M_{\beta,\mathbf{u}}} t^2} + e^{-\frac{1}{3}\mu_{M_{\beta,\mathbf{u}}} t^2} \leq 2e^{-\frac{1}{3}\mu_{M_{\beta,\mathbf{u}}} t^2}, \tag{227}$$

as claimed. ∎

C Proof of Fact 4.1

Recall Fact 4.1 from Section 4.3.2.

Fact (restatement) (Fact 4.1)**.** *Let* $u, v, w, w_0 \in \mathbb{R}_+$ *such that* $u = \frac{1}{2}\left(1 + \sqrt{1+4w}\right)$, *and* $1 \leq u \leq \sqrt{\frac{2}{v}}$. *Define the functions* $f_1, f_2 : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ *by*

$$\begin{aligned} f_1(0) &= 2 \\ f_1(t) &= vw + \sqrt{vf_1(t-1)}, \quad t \in \mathbb{Z}_+ \\ f_2(t) &= 2^{2^{-t}}(u^2v)^{1-2^{-t}}, \quad t \in \mathbb{Z}_{\geq 0}. \end{aligned}$$

Then, f_1 *and* f_2 *are strictly monotonically decreasing and asymptotically converges to* u^2v. *Moreover,* f_2 *pointwise upper bounds* f_1. *Formally,*

$$\begin{aligned} f_1(t) &\leq f_2(t), \quad \forall\, t \in \mathbb{Z}_{\geq 0} \\ \lim_{t\to\infty} f_2(t) &= \lim_{t\to\infty} f_1(t) = u^2v. \end{aligned}$$

The verification of the fact will use Fact C.1.

Fact C.1. *Let* $u, w, w_0 \in \mathbb{R}_+$ $u = \frac{1}{2}\left(1+\sqrt{1+4w}\right)$*. Define the function* $f : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ *by*

$$f(0) = w_0, \tag{228}$$

$$f(t) = \sqrt{w + f(t-1)}, \quad t \in \mathbb{Z}_+. \tag{229}$$

Then,

$$\lim_{t\to\infty} f(t) = u \tag{230}$$

Moreover, when $w_0 > u$ *(*$w_0 < u$*,* $w_0 = u$*),* f *strictly monotonically decreases (respectively, strictly monotonically increases, is constant) with respect to* t*.*

Proof (Fact C.1). Let us first show that f is monotone over $t \in \mathbb{Z}_+$. Write

$$\mathsf{sign}_0(a) = \begin{cases} -1, & \text{if } a < 0, \\ 0, & \text{if } a = 0, \\ 1, & \text{if } a > 0, \end{cases} \tag{231}$$

and note that $\mathsf{sign}_0(f^2(t) - f^2(t')) = \mathsf{sign}_0(f(t) - f(t'))$ for any $t, t' \geq 0$. Moreover, notice that $f^2(t) = (\sqrt{w + f(t-1)})^2 = w + f(t-1)$, $t \in \mathbb{Z}_{\geq 0}$. The goal will be to show that for each $t \in \mathbb{Z}_+$, the sign of $f(t) - f(t+1)$ and $f(t-1) - f(t)$ match. Fix $t \in \mathbb{Z}_+$ arbitrarily, and observe

$$f^2(t) - f^2(t+1) = w + f(t-1) - (w + f(t)) \tag{232}$$

$$= f(t-1) - f(t) \tag{233}$$

and thus

$$\mathsf{sign}_0(f(t) - f(t+1)) = \mathsf{sign}_0(f^2(t) - f^2(t+1)) = \mathsf{sign}_0(f(t-1) - f(t)) \tag{234}$$

as desired. The monotonicity of f over $\mathbb{Z}_{\geq 0}$ follows.

To find the direction of the monotonicity, it suffices to look at $\mathsf{sign}_0(f(1) - f(0))$ since the monotonicity has already been argued. This can be given by

$$\mathsf{sign}_0(f(1) - f(0)) = \mathsf{sign}_0(f^2(1) - f^2(0)) = \mathsf{sign}_0(w + f(0) - f^2(0)) = \mathsf{sign}_0(w + w_0 - w_0^2). \tag{235}$$

To determine from this the condition under which f is constant, observe,

$$\mathsf{sign}_0(w + w_0 - w_0^2) = 0 \tag{236a}$$

$$\longrightarrow \quad w + w_0 - w_0^2 = 0 \tag{236b}$$

$$\longrightarrow \quad w_0^2 - w_0 - w = 0 \tag{236c}$$

$$\longrightarrow \quad w_0 \in \left\{\frac{1}{2}(1 \pm \sqrt{1+4w})\right\} \tag{236d}$$

$$\longrightarrow \quad w_0 = \frac{1}{2}(1 + \sqrt{1+4w}) = u \tag{236e}$$

$$w + w_0 - w_0^2 \begin{cases} < 0, & \text{if } w_0 > \frac{1}{2}(1+\sqrt{1+4w}), \\ = 0, & \text{if } w_0 = \frac{1}{2}(1+\sqrt{1+4w}), \\ > 0, & \text{if } w_0 < \frac{1}{2}(1+\sqrt{1+4w}). \end{cases} \tag{237}$$

Hence, f is strictly monotonically decreasing when $w_0 > u$, constant when $w_0 = u$, and strictly monotonically increasing when $w_0 > u$, as claimed.

The final step is to determine the asymptotic behavior of f as $t \to \infty$. If $w_0 = u$, then f is constant, implying that $\lim_{t\to\infty} f(t) = f(0) = w_0 = u$. On the other hand, when $w_0 \neq u$ we would like to characterize some behavior such as

$$\lim_{t\to\infty} f^2(t+1) - f^2(t) = 0 \tag{238}$$

Observe,

$$\begin{aligned} & f^2(t+1) - f^2(t) = 0 && (239a)\\ \longrightarrow\quad & w + f(t) - f^2(t) = 0 && (239b)\\ \longrightarrow\quad & f(t) = \frac{1}{2}(1+\sqrt{1+4w}) = u && (239c)\end{aligned}$$

Hence, if $w_0 > u$, the strictly monotonically decreasing function is lower bounded by $\inf_{t\in\mathbb{Z}_{\geq 0}} f(t) = u$, while the strictly monotonically increasing function is upper bounded by $\sup_{t\in\mathbb{Z}_{\geq 0}} f(t) = u$ when $w_0 < u$. But in both cases, the function has strict monotonicity, and therefore it must happen that $\lim_{t\to\infty} f(t) = u$. ■

Proof (Fact 4.1). In addition to defining f_1 and f_2 as in Fact 4.1, let $f : \mathbb{Z}_{\geq 0} \to \mathbb{R}$ be the function as defined in Fact C.1, which is given by the recurrence relation

$$f(0) = w_0 \tag{240}$$

$$f(t) = \sqrt{w + f(t-1)} \tag{241}$$

where for the purposes of this proof, w_0 is fixed as $w_0 = \sqrt{\frac{2}{v}}$. Notice that the function f_1 can be written as

$$f_1(t) = vw + \sqrt{vf_1(t-1)} = v\left(w + \sqrt{\frac{f_1(t-1)}{v}}\right) = v\left(w + f(t-1)\right) = vf^2(t) \tag{242}$$

Then, the monotonicity and asymptotic behavior of the functions f_1 follow directly from Fact C.1.

$$\lim_{t\to\infty} f_1(t) = \lim_{t\to\infty} vf^2(t) = u^2 v \tag{243}$$

On the other hand, for f_2,

$$\lim_{t\to\infty} f_2(t) = \lim_{t\to\infty} 2^{2^{-t}}(u^2v)^{1-2^{-t}} = 1\cdot u^2v = u^2v \tag{244}$$

The function f_2 can be shown inductively to pointwise upper bound f_1. The base case, $t = 0$, is trivial since $f_2(0) = 2^{2^0}(u^2v)^{1-2^0} = 2\cdot 1 = 2 = f_1(0)$. Letting $t \in \mathbb{Z}_+$, suppose that for each $t' \in \{2,\ldots,t-1\}$, the bound $f_1(t') \leq f_2(t')$ holds. Then, the desired result will follow from induction if it is shown that $f_1(t) \leq f_2(t)$. To verify this, note that f_2 can be written as the following recurrence relation

$$f_2(0) = 2 \tag{245}$$

$$f_2(t) = \sqrt{u^2vf_2(t-1)} \tag{246}$$

since it was already argued that $f_2(0) = 2$ and otherwise for $t \in \mathbb{Z}_+$,

$$\begin{aligned}\sqrt{u^2vf_2(t-1)} &= \left(u^2v\right)^{\frac{1}{2}}\left(f_2(t-1)\right)^{\frac{1}{2}} && (247a)\\ &= \left(u^2v\right)^{\frac{1}{2}}\left(u^2v\right)^{\frac{1}{2^2}}\left(f_2(t-2)\right)^{\frac{1}{2^2}} = \left(u^2v\right)^{\frac{1}{2}+\frac{1}{2^2}}\left(f_2(t-2)\right)^{\frac{1}{2^2}} && (247b)\\ &= \left(u^2v\right)^{\frac{1}{2}+\frac{1}{2^2}+\frac{1}{2^3}}\left(f_2(t-3)\right)^{\frac{1}{2^3}} && (247c)\\ &\vdots && (247d)\\ &= \left(u^2v\right)^{\sum_{s=1}^{t'}2^{-s}}\left(f_2(t-t')\right)^{2^{-t'}} && (247e)\\ &\vdots && (247f)\\ &= \left(u^2v\right)^{\sum_{s=1}^{t}2^{-s}}\left(f_2(t-t)\right)^{2^{-t}} = \left(u^2v\right)^{\sum_{s=1}^{t}2^{-s}}\left(f_2(0)\right)^{2^{-t}} = 2^{2^{-t}}\left(u^2v\right)^{1-2^{-t}} && (247g)\\ &= f_2(t) && (247h)\end{aligned}$$

as desired. With the above argument, it suffices to show that $f_1(t) \leq \sqrt{u^2 v f_2(t-1)}$. Note that

$$u^2 = \frac{1}{4}\left(1+\sqrt{1+w}\right)^2 = u + w \tag{248a}$$

$$\longrightarrow \quad w = u^2 - u \tag{248b}$$

Then, observe,

$$\begin{aligned}
f_1(t) - \sqrt{u^2 v f_2(t-1)} &= vw + \sqrt{v f_1(t-1)} - \sqrt{u^2 v f_2(t-1)} && \text{(249a)}\\
&\leq vw + \sqrt{v f_2(t-1)} - \sqrt{u^2 v f_2(t-1)}, \quad \blacktriangleright \text{ by the inductive hypothesis} && \text{(249b)}\\
&= v\left(u^2 - u\right) + \sqrt{v f_2(t-1)} - \sqrt{u^2 v f_2(t-1)} && \text{(249c)}\\
&= vu^2 - vu + \sqrt{v f_2(t-1)} - u\sqrt{v f_2(t-1)} && \text{(249d)}\\
&= (u-1)uv - (u-1)\sqrt{v f_2(t-1)} && \text{(249e)}\\
&\leq (u-1)uv - (u-1)\sqrt{v(u^2 v)} && \text{(249f)}\\
&\leq (u-1)uv - (u-1)uv && \text{(249g)}\\
&= 0 && \text{(249h)}
\end{aligned}$$

Hence,

$$f_1(t) - \sqrt{u^2 v f_2(t-1)} \leq 0 \Longrightarrow f_1(t) \leq \sqrt{u^2 v f_2(t-1)} = f_2(t) \tag{250}$$

By induction, $f_1(t) \leq f_2(t)$ for every $t \in \mathbb{Z}_{\geq 0}$. ■

D Miscellaneous results

Lemma D.1. *Let* $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n\times n})$ *be a standard normal vector with i.i.d. entries. Fix any unit vector* $\mathbf{u} \in \mathcal{S}^{n-1}$*. Then, the random variable* $X = \boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}$ *taking values in* $[-\pi, \pi]$ *follows the uniform distribution over* $[-\pi, \pi]$*.*

Proof (Lemma D.1). Let $Y \sim \mathsf{Unif}\left([-\pi, \pi]\right)$ and $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n\times n})$, and define $X = \boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}$. Lemma D.1 will follow from showing the equivalence of the MGFs of X and Y, where both are given by

$$\psi_X(s) = \psi_Y(s) = \psi(s) = \frac{e^{s\pi} - e^{-s\pi}}{2\pi s} \tag{251}$$

for any $s \geq 0$. Recall that the density function associated with the uniform distribution over $[-\pi, \pi]$ is

$$f_Y(y) = \begin{cases} \frac{1}{\pi - (-\pi)} = \frac{1}{2\pi}, & \text{if } y \in [-\pi, \pi], \\ 0, & \text{otherwise.} \end{cases} \tag{252}$$

The MGF of Y is then

$$\begin{aligned}
\psi_Y(s) &= \mathbb{E}\left[e^{sY}\right] && \text{(253a)}\\
&= \int_{y=-\infty}^{y=\infty} e^{sy} f_Y(y)\, dy && \text{(253b)}\\
&= \int_{y=-\pi}^{y=\pi} \frac{e^{sy}}{2\pi} dy && \text{(253c)}\\
&= \frac{1}{2\pi s}\int_{u=-s\pi}^{u=s\pi} e^{u} du, \quad \blacktriangleright\ u = sy,\ dy = \frac{du}{s} && \text{(253d)}\\
&= \frac{e^{s\pi} - e^{-s\pi}}{2\pi s} && \text{(253e)}
\end{aligned}$$

$$= \psi(s) \tag{253f}$$

as desired. On the other hand, the MGF of X is obtained as follows. For any $r \geq 0$, recall that the volume $V_n(r)$ of the n-ball with radius r is given in closed form by

$$V_n(r) = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2}+1)} = \frac{2 \cdot \pi^{\frac{n}{2}}}{n\Gamma(\frac{n}{2})} \tag{254}$$

and can also be represented in spherical coordinates as (see, e.g., Blumenson (1960))

$$V_n(r) = \int_{r'=0}^{r'=r} \int_{w_1=-\frac{\pi}{2}}^{w_1=\frac{\pi}{2}} \cdots \int_{w_{n-2}=-\frac{\pi}{2}}^{w_{n-2}=\frac{\pi}{2}} \int_{w_{n-1}=-\pi}^{w_{n-1}=\pi} r^{n-1} \sin^{n-2}(w_1) \cdots \sin(w_{n-2})\, dw_{n-1}\, dw_{n-2} \cdots dw_1\, dr' \tag{255a}$$

$$= \left(\int_{r'=0}^{r'=r} r^{n-1}\, dr'\right) \left(\int_{w_1=-\frac{\pi}{2}}^{w_1=\frac{\pi}{2}} \sin^{n-2}(w_1)\, dw_1\right) \cdots \left(\int_{w_{n-2}=-\frac{\pi}{2}}^{w_{n-2}=\frac{\pi}{2}} \sin(w_{n-2})\, dw_{n-2}\right) \left(\int_{w_{n-1}=-\pi}^{w_{n-1}=\pi} dw_1\right) \tag{255b}$$

$$= \frac{r^n}{n} \cdot 2\pi \left(\int_{w_1=-\frac{\pi}{2}}^{w_1=\frac{\pi}{2}} \sin^{n-2}(w_1)\, dw_1\right) \cdots \left(\int_{w_{n-2}=-\frac{\pi}{2}}^{w_{n-2}=\frac{\pi}{2}} \sin(w_{n-2})\, dw_{n-2}\right) \tag{255c}$$

It follows that

$$\left(\int_{w_1=-\frac{\pi}{2}}^{w_1=\frac{\pi}{2}} \sin^{n-2}(w_1)\, dw_1\right) \cdots \left(\int_{w_{n-2}=-\frac{\pi}{2}}^{w_{n-2}=\frac{\pi}{2}} \sin(w_{n-2})\, dw_{n-2}\right) = \frac{nV_n(r)}{2\pi r^n} = \frac{n2\pi^{\frac{n}{2}} r^n}{2\pi r^n n\Gamma(\frac{n}{2})} = \frac{\pi^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} \tag{256}$$

Then, again using spherical coordinates,

$$\psi_X(s) = \mathbb{E}\left[e^{sX}\right] \tag{257a}$$

$$= \int_{r=0}^{r=\infty} \int_{w_1=-\frac{\pi}{2}}^{w_1=\frac{\pi}{2}} \cdots \int_{w_{n-2}=-\frac{\pi}{2}}^{w_{n-2}=\frac{\pi}{2}} \int_{x=-\pi}^{x=\pi} \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{r^2}{2}} e^{sx} r^{n-1} \sin^{n-2}(w_1) \cdots \sin(w_{n-2})\, dx\, dw_{n-2} \cdots dw_1\, dr \tag{257b}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \left(\int_{r=0}^{r=\infty} r^{n-1} e^{-\frac{r^2}{2}}\, dr\right) \left(\int_{w_1=-\frac{\pi}{2}}^{w_1=\frac{\pi}{2}} \sin^{n-2}(w_1)\, dw_1\right) \cdots \left(\int_{w_{n-2}=-\frac{\pi}{2}}^{w_{n-2}=\frac{\pi}{2}} \sin(w_{n-2})\, dw_{n-2}\right) \left(\int_{x=-\pi}^{x=\pi} e^{sx}\, dx\right) \tag{257c}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \left(\int_{r=0}^{r=\infty} r^{n-1} e^{-\frac{r^2}{2}}\, dr\right) \frac{\pi^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} \left(\int_{x=-\pi}^{x=\pi} e^{sx}\, dx\right) \tag{257d}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot \frac{\pi^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} \left(\int_{r=0}^{r=\infty} r^{n-1} e^{-\frac{r^2}{2}}\, dr\right) \left(\int_{x=-\pi}^{x=\pi} e^{sx}\, dx\right) \tag{257e}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot \frac{\pi^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} \left(\int_{u=0}^{u=\infty} r^{n-1} e^{-u} \left(\frac{du}{r}\right)\right) \left(\int_{x=-\pi}^{x=\pi} e^{sx}\, dx\right) \tag{257f}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot \frac{\pi^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} \left(\int_{u=0}^{u=\infty} r^{n-2} e^{-u}\, du\right) \left(\int_{x=-\pi}^{x=\pi} e^{sx}\, dx\right) \tag{257g}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot \frac{\pi^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} \left(\int_{u=0}^{u=\infty} \left(\sqrt{2u}\right)^{n-2} e^{-u}\, du\right) \left(\int_{x=-\pi}^{x=\pi} e^{sx}\, dx\right) \tag{257h}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot \frac{\pi^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} \left(2^{\frac{n}{2}-1} \int_{u=0}^{u=\infty} u^{\frac{n}{2}-1} e^{-u}\, du\right) \left(\int_{x=-\pi}^{x=\pi} e^{sx}\, dx\right) \tag{257i}$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \cdot \frac{\pi^{\frac{n}{2}-1}}{\Gamma(\frac{n}{2})} \left(2^{\frac{n}{2}-1} \Gamma\left(\frac{n}{2}\right)\right) \left(\int_{x=-\pi}^{x=\pi} e^{sx}\, dx\right) \tag{257j}$$

$$= \frac{(2\pi)^{\frac{n}{2}-1}\Gamma(\frac{n}{2})}{(2\pi)^{\frac{n}{2}}\Gamma(\frac{n}{2})} \int_{x=-\pi}^{x=\pi} e^{sx}\, dx \tag{257k}$$

$$= \frac{1}{2\pi} \int_{x=-\pi}^{x=\pi} e^{sx}\, dx \tag{257l}$$

$$= \frac{e^{s\pi} - e^{-s\pi}}{2\pi s} \tag{257m}$$

$$= \psi(s) \tag{257n}$$

Therefore, $\psi_X = \psi = \psi_Y$, which immediately implies that the random variables X and Y follow the same distribution, as claimed. ∎

Lemma (restatement) D.2 (Lemma B.5)**.** *Fix any pair of real-valued vectors* $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$*, and suppose* $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n\times n})$ *is a standard normal vector with i.i.d. entries. Define the indicator random variable* $I = \mathbf{1}(\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle) - \mathsf{sign}(\langle \mathbf{v}, \mathbf{Z}\rangle) \neq 0)$*. Then,*

$$\Pr(I = 1) = \frac{\theta_{\mathbf{u},\mathbf{v}}}{\pi}. \tag{258}$$

Proof (Lemma D.2). The result will follow from showing that the random variable I is equivalently defined angularly as $\mathbf{1}(\cos(\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}) \cos(\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} - \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}) < 0)$. Subsequently, Lemma D.1 simplifies the derivation of (258). For the first step, observe the following equivalence:

$$I = 1 \iff (\mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle) - \mathsf{sign}(\langle \mathbf{v}, \mathbf{Z}\rangle)) = 1, \qquad \blacktriangleright \text{ by definition} \tag{259a}$$

$$\iff \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle) \neq \mathsf{sign}(\langle \mathbf{v}, \mathbf{Z}\rangle) \tag{259b}$$

$$\iff \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle)\,\mathsf{sign}(\langle \mathbf{v}, \mathbf{Z}\rangle) \neq \mathsf{sign}(\langle \mathbf{v}, \mathbf{Z}\rangle)\,\mathsf{sign}(\langle \mathbf{v}, \mathbf{Z}\rangle) = 1 \tag{259c}$$

$$\iff \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle)\,\mathsf{sign}(\langle \mathbf{v}, \mathbf{Z}\rangle) \neq 1 \tag{259d}$$

$$\iff \mathsf{sign}(\langle \mathbf{u}, \mathbf{Z}\rangle)\,\mathsf{sign}(\langle \mathbf{v}, \mathbf{Z}\rangle) = -1 \tag{259e}$$

$$\iff \mathsf{sign}\left(\left\langle \frac{\mathbf{u}}{\|\mathbf{u}\|_2}, \frac{\mathbf{Z}}{\|\mathbf{Z}\|_2}\right\rangle\right) \mathsf{sign}\left(\left\langle \frac{\mathbf{v}}{\|\mathbf{v}\|_2}, \frac{\mathbf{Z}}{\|\mathbf{Z}\|_2}\right\rangle\right) = -1 \tag{259f}$$

$$\iff \left\langle \frac{\mathbf{u}}{\|\mathbf{u}\|_2}, \frac{\mathbf{Z}}{\|\mathbf{Z}\|_2}\right\rangle \left\langle \frac{\mathbf{v}}{\|\mathbf{v}\|_2}, \frac{\mathbf{Z}}{\|\mathbf{Z}\|_2}\right\rangle < 0 \tag{259g}$$

$$\iff \cos(\theta_{\mathbf{u},\mathbf{Z}}) \cos(\theta_{\mathbf{v},\mathbf{Z}}) < 0 \tag{259h}$$

$$\iff \cos(\theta_{\mathbf{u},\mathbf{Z}}) \cos(\theta_{\mathbf{u},\mathbf{Z}}) < 0 \tag{259i}$$

$$\iff \cos(\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}) \cos(\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} + \boldsymbol{\theta}_{\mathbf{v},\mathbf{u}}) < 0 \tag{259j}$$

$$\iff \cos(\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}) \cos(\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} - \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}) < 0 \tag{259k}$$

$$\iff \left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} + \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}| > \frac{\pi}{2} \text{ and } |\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| < \frac{\pi}{2}\right) \text{ or } \left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} - \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}| < \frac{\pi}{2} \text{ and } |\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| > \frac{\pi}{2}\right) \tag{259l}$$

where the random variables $\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}, \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}, \boldsymbol{\theta}_{\mathbf{v},\mathbf{u}} \in [-\pi, \pi]$ are signed rotations under some convention for rotations in the origin-centered hyperplane containing $\mathbf{u}$ and $\mathbf{v}$. Recall from Lemma D.1 that the random variable $\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}$ follows the uniform distribution over $[-\pi, \pi]$. In light of this, suppose $Y \sim \mathsf{Unif}([-\pi, \pi])$ is a random variable under the uniform distribution. Note that for any fixed size $b \in [0, 2\pi]$, every interval $[a, a + b) \subseteq [-\pi, \pi]$ of size b is equally probable, formally

$$\Pr(Y \in [a, a + b)) = \Pr(Y \in [a', a' + b)) \tag{260}$$

for every choice of $a, a' \in [-\pi, \pi - b]$. Then,

$$\Pr\left(\left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} + \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}| > \frac{\pi}{2} \text{ and } |\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| < \frac{\pi}{2}\right) \text{ or } \left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} - \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}| < \frac{\pi}{2} \text{ and } |\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| > \frac{\pi}{2}\right)\right) \tag{261a}$$

$$= \Pr\left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} + \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}| > \frac{\pi}{2} \text{ and } |\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| < \frac{\pi}{2}\right) + \Pr\left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} - \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}| < \frac{\pi}{2} \text{ and } |\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| > \frac{\pi}{2}\right) \tag{261b}$$

▶ the two events are disjoint

$$= \Pr\left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} + \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}| > \frac{\pi}{2} \middle| |\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| < \frac{\pi}{2}\right) \Pr\left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| < \frac{\pi}{2}\right) + \Pr\left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} - \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}| < \frac{\pi}{2} \middle| |\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| > \frac{\pi}{2}\right) \Pr\left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| > \frac{\pi}{2}\right) \tag{261c}$$

$$= \frac{1}{2}\Pr\left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} + \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}| > \frac{\pi}{2} \middle| |\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| < \frac{\pi}{2}\right) + \frac{1}{2}\Pr\left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} - \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}| < \frac{\pi}{2} \middle| |\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| > \frac{\pi}{2}\right) \tag{261d}$$

$$= \frac{1}{2}\Pr\left(Y \in [0, \theta_{\mathbf{u},\mathbf{v}})\right) + \frac{1}{2}\Pr\left(Y \in [0, \theta_{\mathbf{u},\mathbf{v}})\right) \tag{261e}$$

$$= \Pr\left(Y \in [0, \theta_{\mathbf{u},\mathbf{v}})\right) \tag{261f}$$

$$= \frac{\theta_{\mathbf{u},\mathbf{v}}}{\pi} \tag{261g}$$

Therefore, combining (259a) and (261) yields the result:

$$\Pr(I = 1) = \Pr\left(\left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} + \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}| > \frac{\pi}{2} \text{ and } |\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| < \frac{\pi}{2}\right) \text{ or } \left(|\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}} - \boldsymbol{\theta}_{\mathbf{u},\mathbf{v}}| < \frac{\pi}{2} \text{ and } |\boldsymbol{\theta}_{\mathbf{u},\mathbf{Z}}| > \frac{\pi}{2}\right)\right) = \frac{\theta_{\mathbf{u},\mathbf{v}}}{\pi}. \tag{262}$$

■

Lemma D.3. *Fix $\sigma > 0$. Let $Z_1, \dots, Z_m \sim \mathcal{N}(0, \sigma^2)$ be m i.i.d. normal variables. Then, their sum $\sum_{i=1}^m Z_i$ follows the mean-zero, variance-$m\sigma^2$ normal distribution.*

Corollary D.4. *Let $\mathbf{Z}^{(1)}, \dots, \mathbf{Z}^{(m)} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{n\times n})$. Then, their sum $\sum_{i=1}^m \mathbf{Z}^{(i)}$ follows the normal distribution $\mathcal{N}(\mathbf{0}, m\sigma^2\mathbf{I}_{n\times n})$.*

Proof (Lemma D.4). This follows directly from applying Lemma D.3 to each of the n coordinates, which suffices since the entries are independent. ■

Lemma D.5. *Fix $\sigma, t > 0$. Let $Z \sim \mathcal{N}(0, \sigma^2)$ be a normal random variable. Then,*

$$\Pr(|Z| > t) \leq 2e^{-\frac{t^2}{2\sigma^2}}. \tag{263}$$

Proof (Lemma D.5). By Chernoff bounds,

$$\Pr(Z > t) \leq e^{-\frac{t^2}{2\sigma^2}} \tag{264}$$

$$\Pr(Z < t) \leq e^{-\frac{t^2}{2\sigma^2}}. \tag{265}$$

By a union bound, the result follows. ■

Corollary D.6. *Fix $\sigma, t > 0$. Let $Z_1, \dots, Z_m \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{n\times n})$ be m i.i.d. normal random variables, and write their sum as $Z = \sum_{i=1}^m Z_i$. Then,*

$$\Pr(|Z| > t) \leq 2e^{-\frac{t^2}{2m\sigma^2}}. \tag{266}$$

Proof (of Corollary D.6). The corollary directly follows from Lemmas D.5 and D.3. ■

Lemma D.7. *Fix $\sigma, t > 0$. Let $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{n\times n})$, be a normal vector with i.i.d. entries. Then,*

$$\Pr\left(\left|\|\mathbf{Z}\|_2 - \mathbb{E}\left[\|\mathbf{Z}\|_2\right]\right| \geq t\right) \leq 2e^{-\frac{t^2}{2\sigma^2}} \tag{267}$$

Moreover, for any coordinate subset, $\mathcal{J} \subseteq [n]$,

$$\Pr\left(\left|\|\mathbf{T}_{\mathcal{J}}\mathbf{Z}\|_2 - \mathbb{E}\left[\|\mathbf{T}_{\mathcal{J}}\mathbf{Z}\|_2\right]\right| \geq t\right) \leq 2e^{-\frac{t^2}{2\sigma^2}} \tag{268}$$

Proof (Lemma D.7). Fix any $\sigma, t > 0$, and suppose $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n\times n})$. The lemma will follow from deriving a Chernoff bound using the MGF of the random vector $\mathbf{Z}$. Recall that the zero-mean multivariate normal distribution has as its density function

$$f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\mathbf{z}^\top(\frac{1}{\sigma^2}\mathbf{I}_{n\times n})\mathbf{z}}{2}} = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} e^{-\frac{\|\mathbf{z}\|_2^2}{2\sigma^2}}. \tag{269}$$

The MGF of $\|\mathbf{Z}\|_2$ is then

$$\psi_{\mathbf{Z}}(s) = e^{\frac{\sigma^2 s^2}{2}} \tag{270}$$

which is obtained as follows.

$$\begin{aligned}
\psi_{\|\mathbf{Z}\|_2}(s) &= \mathbb{E}\left[e^{-s\|\mathbf{Z}\|_2}\right] && \text{(271a)}\\
&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \int_{\mathbf{z}\in\mathbb{R}^n} e^{s\|\mathbf{z}\|_2} e^{-\frac{\|\mathbf{z}\|_2^2}{2\sigma^2}} d\mathbf{z}, \quad \blacktriangleright \text{ by the law of lazy statistician} && \text{(271b)}\\
&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \int_{\mathbf{z}\in\mathbb{R}^n} e^{\frac{\sigma^2 s^2}{2}} e^{-\frac{\sigma^2 s^2}{2}} e^{s\|\mathbf{z}\|_2} e^{-\frac{\|\mathbf{z}\|_2^2}{2\sigma^2}} d\mathbf{z} && \text{(271c)}\\
&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \int_{\mathbf{z}\in\mathbb{R}^n} e^{\frac{\sigma^2 s^2}{2}} e^{-\frac{\sigma^4 s^2}{2\sigma^2}} e^{\frac{2\sigma^2 s\|\mathbf{z}\|_2}{2\sigma^2}} e^{-\frac{\|\mathbf{z}\|_2^2}{2\sigma^2}} d\mathbf{z} && \text{(271d)}\\
&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \int_{\mathbf{z}\in\mathbb{R}^n} e^{\frac{\sigma^2 s^2}{2}} e^{-\frac{\|\mathbf{z}\|_2^2 - 2\sigma^2 s\|\mathbf{z}\|_2 + \sigma^4 s^2}{2\sigma^2}} d\mathbf{z} && \text{(271e)}\\
&= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \int_{\mathbf{z}\in\mathbb{R}^n} e^{\frac{\sigma^2 s^2}{2}} e^{-\frac{(\|\mathbf{z}\|_2 - \sigma^2 s)^2}{2\sigma^2}} d\mathbf{z} && \text{(271f)}\\
&= e^{\frac{\sigma^2 s^2}{2}} \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \int_{\mathbf{z}\in\mathbb{R}^n} e^{-\frac{(\|\mathbf{z}\|_2 - \sigma^2 s)^2}{2\sigma^2}} d\mathbf{z} && \text{(271g)}\\
&= e^{\frac{\sigma^2 s^2}{2}} \cdot 1, \quad \blacktriangleright \text{ evaluating a density function over its entire support} && \text{(271h)}\\
&= e^{\frac{\sigma^2 s^2}{2}} && \text{(271i)}
\end{aligned}$$

The upper bound in (267) follows.

$$\begin{aligned}
\Pr\left(\|\mathbf{Z}\|_2 - \mathbb{E}\left[\|\mathbf{Z}\|_2\right] \geq t\right) &= \Pr\left(e^{\|\mathbf{Z}\|_2 - \mathbb{E}\left[\|\mathbf{Z}\|_2\right]} \geq e^t\right) && \text{(272a)}\\
&\leq \min_{s\geq 0} e^{-st}\psi_{\mathbf{Z}}(s), \quad \blacktriangleright \text{ due to Bernstein (see, e.g., Vershynin (2018))} && \text{(272b)}\\
&\leq \min_{s\geq 0} e^{-st} e^{\frac{\sigma^2 s^2}{2}}, \quad \blacktriangleright \text{ by Eq. (270)} && \text{(272c)}\\
&= e^{-\frac{t^2}{\sigma^2}} e^{\frac{t^2}{2\sigma^2}}, \quad \blacktriangleright\ s = \frac{t}{\sigma^2} \text{ minimizes (272c)} && \text{(272d)}\\
&= e^{-\frac{t^2}{2\sigma^2}} && \text{(272e)}
\end{aligned}$$

as desired. The lower bound in (267) follows from a similar argument. Combined with the upper bound by union bounding yields (267).

The lemma's second result, (268) follows immediately from (267), which does not depend on the dimension of the random vector $\mathbf{Z}$ (and thus the same concentration inequality holds after hard thresholding a subset of coordinates). More formally, this can be shown by contradiction. Suppose that

$$\Pr\left(\left|\|\mathbf{T}_{\mathcal{J}}\mathbf{Z}\|_2 - \mathbb{E}\,\|\mathbf{T}_{\mathcal{J}}\mathbf{Z}\|_2\right| \geq t\right) > 2e^{-\frac{t^2}{2\sigma^2}}. \tag{273}$$

Fix any $d \in [n]$, and with out loss of generality, suppose $\mathcal{J} = [d]$. Construct a second random vector $\mathbf{Y} = (Y_1, \ldots, Y_d)$ in d-dimensions by assigning $Y_j = Z_j$ for each $j \in [d]$, such that $\mathbf{Y} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{d\times d})$. By (267), the following holds

$$\Pr\left(\left|\|\mathbf{Y}\|_2 - \mathbb{E}\left[\|\mathbf{Y}\|_2\right]\right| \geq t\right) \leq 2e^{-\frac{t^2}{2\sigma^2}}. \tag{274}$$

Then, it must happen that for some random draw of $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n\times n})$, $\|\mathbf{T}_{\mathcal{J}}\mathbf{Z}\|_2 \neq \|\mathbf{Y}\|_2$. But similarly justified by the proof of Lemma D.10 (stated later in the appendix),

$$\|\mathbf{T}_{\mathcal{J}}\mathbf{Z}\|_2 = \left(\sum_{j=1}^{n} (T_{j;\mathcal{J}} Z_j)^2\right)^{\frac{1}{2}} = \left(\sum_{\substack{j\in[n]: \\ T_{j;\mathcal{J}}\neq 0}} Z_j^2\right)^{\frac{1}{2}} = \left(\sum_{j\in[d]} Z_j^2\right)^{\frac{1}{2}} = \|\mathbf{Y}\|_2 \tag{275a}$$

which is a contradiction. Hence, (268) holds. ■

Corollary D.8. *Fix $\sigma > 0$. Let $\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(m)} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n\times n})$, be m i.i.d. normal vectors, and write their sum as $\mathbf{Z} = \sum_{i=1}^{m} \mathbf{Z}^{(i)}$. Then,*

$$\Pr\left(\left|\|\mathbf{Z}\|_2 - \mathbb{E}\,\|\mathbf{Z}\|_2\right| \geq t\right) \leq 2e^{-\frac{t^2}{2m\sigma^2}} \tag{276}$$

Moreover, for any coordinate subset, $\mathcal{J} \subseteq [n]$,

$$\Pr\left(\left|\|\mathbf{T}_{\mathcal{J}}\mathbf{Z}\|_2 - \mathbb{E}\,\|\mathbf{T}_{\mathcal{J}}\mathbf{Z}\|_2\right| \geq t\right) \leq 2e^{-\frac{t^2}{2m\sigma^2}} \tag{277}$$

Proof (of Corollary D.8). Notice that $\mathbf{Z} = \sum_{i=1}^{m} \mathbf{Z}^{(i)}$ follows the multivariate normal distribution $\mathcal{N}(\mathbf{0}, m\sigma^2)$ due to Corollary D.4. Hence, the corollary is immediately realized from Lemma D.7. ■

Lemma D.9. *Fix $\sigma > 0$, and let $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n\times n})$. Then, its expected norm is at most*

$$\mathbb{E}\left[\|\mathbf{Z}\|_2\right] \leq \sqrt{n\sigma^2} = \sqrt{n}\sigma. \tag{278}$$

In the case when $n = 1$, the expected norm is precisely

$$\mathbb{E}\left[\|\mathbf{Z}\|_2\right] = \sqrt{\frac{2}{\pi\sigma^2}}. \tag{279}$$

Proof (Lemma D.9). Let $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{n\times n})$. Then,

$$\begin{aligned}
\mathbb{E}\left[\|\mathbf{Z}\|_2\right] &= \mathbb{E}\left[\sqrt{\sum_{j=1}^{n} Z_j^2}\right] && (280a)\\
&\leq \sqrt{\mathbb{E}\left[\sum_{j=1}^{n} Z_j^2\right]}, && \blacktriangleright \text{ Jensen's inequality (for concave functions)} \quad (280b)\\
&= \sqrt{\sum_{j=1}^{n} \mathbb{E}\left[Z_j^2\right]} && (280c)\\
&= \sqrt{\sum_{j=1}^{n} \left(\mathsf{Var}\left(Z_j^2\right) + \mathbb{E}\left[Z_j\right]^2\right)} && (280d)\\
&= \sqrt{\sum_{j=1}^{n} \mathsf{Var}\left(Z_j^2\right)} && (280e)\\
&= \sqrt{\sum_{j=1}^{n} \sigma^2} && (280f)\\
&= \sqrt{n\sigma^2} && (280g)\\
&= \sqrt{n}\sigma && (280h)
\end{aligned}$$

as claimed. ■

Lemma D.10. *Fix $\mathcal{J} \subset [n]$. Then, the map $\mathcal{T}_J : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation.*

Proof (Lemma D.10). Fix $\mathcal{J} \subset [n]$ arbitrarily, and construct a diagonal matrix $\mathbf{T}_{\mathcal{J}} \in \mathbb{R}^{n \times n}$ such that $\mathbf{T}_{\mathcal{J}} = \mathsf{diag}(T_{1;\mathcal{J}}, \dots, T_{n;\mathcal{J}})$ with $T_{j;\mathcal{J}} = \mathbf{1}(j \in \mathcal{J})$ for each $j \in [n]$, as in Definition 2.2 in Section 2. Then, clearly the map $T : \mathbb{R}^n \to \mathbb{R}^n$ by $\mathbf{x} \overset{T}{\mapsto} \mathbf{T}_{\mathcal{J}}\mathbf{x}$ is equivalent to the map $\mathcal{T}_{\mathcal{J}} : \mathbb{R}^n \to \mathbb{R}^n$ since, writing $\mathbf{y} = T(\mathbf{x}) = \mathbf{A}\mathbf{x}$,

$$\mathbf{y} = \begin{pmatrix} T_{1;\mathcal{J}} x_1 \\ \vdots \\ T_{n;\mathcal{J}} x_n \end{pmatrix} = \begin{pmatrix} x_1 \cdot \mathbf{1}(1 \in \mathcal{J}) \\ \vdots \\ x_n \cdot \mathbf{1}(n \in \mathcal{J}) \end{pmatrix} = \mathcal{T}_{\mathcal{J}}(\mathbf{x}) \tag{281a}$$

for all $\mathbf{x} \in \mathbb{R}^n$. It is well-known that a map from $\mathbb{R}^n$ to $\mathbb{R}^n$ is a linear transformation if can be specified by a matrix-vector product for some real-valued $n \times n$ matrix, hence completing the proof. ■

E A Different Invertibility Condition Friedlander et al. (2021)

Definition E.1 (restricted approximate invertibility condition as defined in (Friedlander et al. 2021, Def. 8))**.** *Fix $\nu, \delta, \eta, r, r' > 0$. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ be a measurement matrix, and let $\mathbf{x} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$. The $(\nu, \delta, \eta, r, r')$-RAIC holds for $\mathbf{A}$ at $\mathbf{x}$ if for every $\mathbf{y} \in \mathcal{S}^{n-1} \cap \Sigma_k^n$, $r \le d_{\mathcal{S}^{n-1}}(\mathbf{x}, \mathbf{y}) \le r'$,*

$$\left\| (\mathbf{x} - \mathbf{y}) - \nu \mathbf{A}^{\mathsf{T}} (\mathsf{sign}(\mathbf{A}\mathbf{x}) - \mathsf{sign}(\mathbf{A}\mathbf{y})) \right\|_{(\mathcal{S}^{n-1} \cap \Sigma_k^n)^\circ} \le \delta d_{\mathcal{S}^{n-1}}(\mathbf{x}, \mathbf{y}) + \eta \tag{282}$$

where $\|\cdot\|_{(\mathcal{S}^{n-1} \cap \Sigma_k^n)^\circ}$ denotes the dual norm given by $\|\mathbf{u}\|_{(\mathcal{S}^{n-1} \cap \Sigma_k^n)^\circ} = \sup_{\mathbf{u}' \in \mathcal{S}^{n-1} \cap \Sigma_k^n} \langle \mathbf{u}, \mathbf{u}' \rangle$ for $\mathbf{u} \in \mathbb{R}^n$.

Instead of the ℓ_2-norm as in our definition, this definition resorts to the dual norm. Furthermore, our definition of RAIC should hold for all pair of vectors uniformly; whereas in the above definition invertibitily condition is asked for vectors within distance $[r, r']$. Both of these two differences make our definition simpler to state and handle, and also allow us to do a precise analysis in the "small-distance" regime.