High-dimensional Asymptotics for Phase Retrieval with Structured Sensing Matrices

Rishabh Dudeja

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2021

Abstract

High-dimensional Asymptotics for Phase Retrieval with Structured Sensing Matrices

Rishabh Dudeja

Phase Retrieval is an inference problem where one seeks to recover an unknown complex-valued n-dimensional signal vector from the magnitudes of m linear measurements. The linear measurements are specified using a $m \times n$ sensing matrix. This problem is a mathematical model for imaging systems arising in X-ray crystallography and other applications where it is infeasible to acquire the phase of the measurements. This dissertation presents some results regarding the analysis of this problem in the high-dimensional asymptotic regime where the number of measurements and the signal dimension diverge proportionally so that their ratio remains fixed. A limitation of existing high-dimensional analyses of this problem is that they model the sensing matrix as a random matrix with independent and identically (i.i.d.) distributed Gaussian entries. In practice, this matrix is highly structured with limited randomness. This work studies a correction to the i.i.d. Gaussian sensing model, known as the sub-sampled Haar sensing model which faithfully captures a crucial orthogonality property of realistic sensing matrices. The first result of this thesis provides a precise asymptotic characterization of the performance of commonly used spectral estimators for phase retrieval in the sub-sampled Haar sensing model. This result can be leveraged to tune certain parameters involved in the spectral estimator optimally. The second part of this dissertation studies the information-theoretic limits for better-than-random (or weak) recovery in the sub-sampled Haar sensing model. The main result in this part shows that appropriately tuned spectral methods achieve weak recovery with the information-theoretically optimal

number of measurements. Simulations indicate that the performance curves derived for the sub-sampled Haar sensing model accurately describe the empirical performance curves for realistic sensing matrices such as randomly sub-sampled Fourier sensing matrices and Coded Diffraction Pattern (CDP) sensing matrices. The final part of this dissertation tries to provide a mathematical understanding of this empirical universality phenomenon: For the real-valued version of the phase retrieval problem, the main result of the final part proves that the dynamics of a class of iterative algorithms, called Linearized Approximate Message Passing schemes, are asymptotically identical in the sub-sampled Haar sensing model and a real-valued analog of the sub-sampled Fourier sensing model.

Table of Contents

Acknowledgements

This dissertation would not have been possible without the support and mentorship of my advisors Prof. Arian Maleki and Prof. Daniel Hsu. Over the past five years, they have been generous with their time, ideas and feedback. They have taught me so much, including mathematics, evaluating research problems, writing papers and research statements, and giving talks. In research, I often found myself lost in the weeds. Arian and Daniel have always been there to motivate me, show me a different path, or gently nudge me to cut my losses and move on. I couldn't have asked for better mentors. I am so grateful that they took a chance on me even when I don't think I deserved it.

I am grateful to my other committee members: Prof. Cynthia Rush, Prof. Sumit Mukherjee, and Prof. Yue Lu, for asking many thought-provoking questions and giving me feedback on my thesis. I am grateful to Prof. Sumit Mukherjee and Prof. Yue Lu for supporting my job applications. I cannot thank Prof. Cynthia Rush enough for giving me the opportunity to be a part of the Fall 2020 program on Probability, Geometry, and Computation in High Dimensions at the Simons Institute. I learned a lot during this program from my interactions with Prof. Cynthia Rush and Prof. Song Mei and the many reading group talks and seminars.

I also had the good fortune of working with several fantastic student and postdoc collaborators: Ji Xu, Junjie Ma, Kiran Vodrahalli, and Milad Bakhshizadeh. I benefitted not only from their ideas in our projects but also from their friendship. I am particularly indebted to Junjie for introducing me to the phase retrieval problem. I would not have started working in this direction if it were not for him.

Over the past six years, I attended courses taught by some of the most outstanding teachers I have met: Professors Alex Andoni, Arian Maleki, Bodhi Sen, Daniel Hsu, Michael Weinstein, Ming Yuan, Omri Weinstein, Sumit Mukherjee, and Vineet Goyal. I am thankful for their contribution to my education. I also learned so much from Arian and Daniel's other students and postdocs during the many reading groups at Columbia: Chris Tosh, Clayton Sanford, Deb Mandal, Geelon So, Giannis Karamanolakis, Haolei Weng, Ji Xu, Junjie Ma, Kevin Shi, Kiran Vodrahalli, Milad Bakhshizadeh, Morgane Austern, Shuiawen Wang, and Wenda Zhou. I thank them for sharing their knowledge and perspective.

I am also grateful to Anthony Cruz and Dood Kalicharan for their constant support; thanks to them, I didn't have to worry about anything other than my work since they took care of everything else. I don't think the statistics department can function without them.

Finally, I had my share of ups and downs during this journey. I am grateful to my sister Aditi, my friends Piyush and Tim, and my mom and dad for always being there with a sympathetic and patient ear. I owe everything to them.

Chapter 1: Introduction

1.1 The Phase Retrieval Problem

Phase retrieval is a statistical inference problem that arises in various imaging applications like electron microscopy, crystallography, astronomy, and optical imaging [1]. This problem originated in the field of X-ray crystallography [2], and we use this application to describe the physical considerations giving rise to the phase retrieval problem.

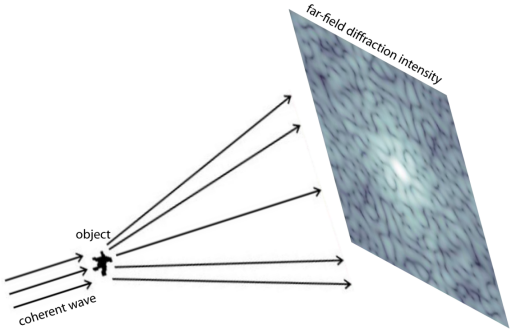

Figure 1.1: A schematic diagram of a typical X-ray crystallography setup. Source: Shechtman, Eldar, Cohen, Chapman, Miao, and Segev [1]

X-ray crystallography: The goal of X-ray crystallography is to infer the structure of a molecule of a compound from its crystalline sample. The structure of a molecule is captured by its electron density function which describes the probability of observing an electron in any given spatial location. In this imaging technology, the crystalline sample is irradiated with an X-ray beam. As the X-rays pass through the sample, they interact with the electron density of the sample and diffract. The intensity (or magnitude) of the diffraction pattern at various spatial locations is captured by a photographic plate. Due to physical limitations, it is infeasible to capture the phase of the diffraction pattern. The relationship between the spatial intensity of the diffraction pattern

and the electron density function of the sample is described by Fraunhofer (or far-field) diffraction principle. According to this principle, the intensity of the diffraction pattern at a specific spatial location is proportional to the magnitude of the Fourier transform of the electron density function at a suitable frequency (see [1, Page 4] for a precise formula). Hence, in the phase retrieval problem, one seeks to infer the unknown electron density function of the molecule from the magnitude of its Fourier transform. Since the Fourier transform is invertible, recovering the unknown electron density function is equivalent to recovering the Fourier transform of the electron density function. Since the magnitude of the Fourier transform is already observed, one simply needs to recover the unobserved phase information. This is why this problem is called "phase retrieval". A schematic diagram of a typical X-ray crystallography setup (reproduced from [1]) is shown in Figure 1.1.

Mathematical Formulation: A common mathematical formulation of the phase retrieval problem is to recover an unknown n-dimensional, complex-valued signal vector $\boldsymbol{x}_\star \in \mathbb{C}^n$ from the magnitudes of m linear measurements. The measurements are denoted by a m-dimensional vector $\boldsymbol{y} \in \mathbb{R}^m$. The relationship between the signal $\boldsymbol{x}_\star$ and the observed measurements is given by:

$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}_\star|^2. \tag{1.1}$$

In the above equation, $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ is a $m \times n$ matrix, known as the sensing matrix. The operation $|\cdot|^2$ is understood to act entry-wise on the vector $\boldsymbol{A}\boldsymbol{x}_\star \in \mathbb{C}^m$. The sensing matrix $\boldsymbol{A}$ is assumed to be known. This general mathematical formulation can be specialized to the setup of X-ray crystallography as follows:

- The signal vector $\boldsymbol{x}_\star$ encodes the unknown electron density function of the molecule of interest. It is constructed by sampling (or discretizing) the electron density function on a 2D grid of size $d \times d$ and encoding the resulting $d \times d$ matrix as a vector of dimension $n = d^2$. If $\boldsymbol{X} \in \mathbb{C}^{d\times d}$ denotes the sampled (or discretized) electron density function, then one such

encoding is given by:

$$(x_\star)_{(i-1)d+j} = X_{ij},\ i,j \in \{1,2,3\ldots,d\}. \tag{1.2}$$

- The sensing matrix $\boldsymbol{A} = \boldsymbol{F}_n$, the $n \times n$ linear operator which maps $\boldsymbol{x}_\star \in \mathbb{C}^n$ to the 2D Discrete Fourier Transform (DFT) of $\boldsymbol{X}_\star$ (encoded as a vector). For the encoding specified in (1.2), the sensing matrix is given by:

$$(F_n)_{(i_1-1)d+j_1,(i_2-1)d+j_2} = \frac{1}{\sqrt{n}} \exp\left(\frac{2\pi \mathbf{i}}{\sqrt{n}} \cdot (i_1-1)(i_2-1) + \frac{2\pi \mathbf{i}}{\sqrt{n}} \cdot (j_1-1)(j_2-1)\right), \tag{1.3}$$

  where $\mathbf{i} = \sqrt{-1}$.

Redundant Measurements: Note that in the Fourier phase retrieval problem discussed so far (1.3), the number of measurements equals the signal dimension, i.e., $m = n$. However, the mathematical formulation in (1.1) allows for the acquisition of $m > n$ redundant measurements. Acquiring $m > n$ redundant measurements is desirable for two reasons:

- For arbitrary signal vectors, the magnitude of the Fourier transform does not uniquely determine the signal [3]. Hence, acquiring redundant measurements can help ensure that the signal is uniquely determined (up to some trivial ambiguities) by magnitude-only measurements.

- Even for signal classes that are uniquely determined by the magnitude of their Fourier transform, acquiring redundant measurements can improve the stability properties of the inverse problem and provide robustness to some amount of noise in the measurements.

In this dissertation, we will be particularly interested in the following two approaches for obtaining redundant measurements.

Masks: In this scheme, proposed by Candès, Eldar, Strohmer, and Voroninski [4], a mask or a phase plate is placed between the sample and the photographic plate. By modulating the sample with several different masks, redundant measurements are obtained obtained. A schematic diagram of this setup is shown in Figure 1.2. The sensing matrix in this scheme is called the Coded Diffraction Pattern (CDP) ensemble and is given by:

$$\boldsymbol{A}_{\mathsf{CDP}} = \begin{bmatrix} \boldsymbol{F}_n \boldsymbol{D}_1 \\ \boldsymbol{F}_n \boldsymbol{D}_2 \\ \vdots \\ \boldsymbol{F}_n \boldsymbol{D}_\delta \end{bmatrix}. \tag{1.4a}$$

where $\boldsymbol{F}_n$ denotes the $n \times n$ 2D-DFT matrix (defined in (1.3)), $\delta \in \mathbb{N}$ is the number of masks used and $\boldsymbol{D}_{1:\delta}$ are diagonal matrices representing phase masks used to modulate the signal:

$$\boldsymbol{D}_\ell = \mathrm{Diag}\left(e^{\mathbf{i}\theta_{1,\ell}}, e^{\mathbf{i}\theta_{2,\ell}}, \cdots, e^{\mathbf{i}\theta_{n,\ell}}\right). \tag{1.4b}$$

A popular proposal for designing the phase masks is to sample them randomly [4, 5], for e.g.

$$\theta_{i,\ell} \overset{\text{i.i.d.}}{\sim} \mathrm{Unif}\left(\left\{-\frac{\pi}{2}, 0, \frac{\pi}{2}, \pi\right\}\right). \tag{1.4c}$$

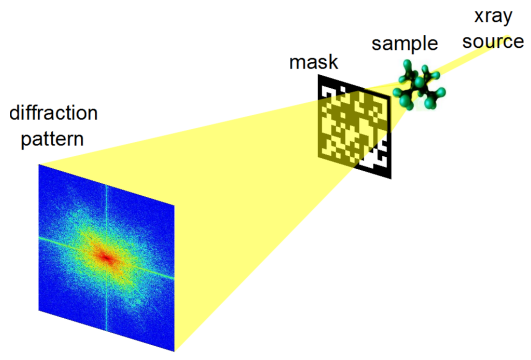

Figure 1.2: A schematic diagram of a typical setup for diffraction imaging with phase masks. Source: Candès, Eldar, Strohmer, and Voroninski [4]

Oversampling: Another strategy to obtain redundant measurements is to oversample the diffraction pattern on a grid of resolution finer than the Nyquist frequency [6]. This requires surrounding the sample with a background of known transmission properties [7, 8]. Mathematically, this is formulated as zero padding the n-dimensional signal vector with $m - n$ zeros [1]. The measurements are given by the magnitude of m-point 2D-DFT of the zero-padded signal. Consequently, in the oversampled Phase retrieval problem, the sensing matrix is given by sub-sampling the first n columns of the $m \times m$ 2D-DFT matrix $\boldsymbol{F}_m$ (as defined in (1.3)). In this dissertation, we will be interested in a semi-random model for oversampled phase retrieval, where the n columns are chosen uniformly at random (without replacement). Formally, the sensing matrix will be given by:

$$\boldsymbol{A}_{\mathsf{PDFT}} = \boldsymbol{F}_m \cdot \boldsymbol{P} \cdot \boldsymbol{S}_{m,n}, \tag{1.5a}$$

where, $\boldsymbol{F}_m$ is the $m \times m$ 2D-DFT matrix defined in (1.3) and,

$$\boldsymbol{P} \sim \text{Uniformly Random } m \times m \text{ Permutation Matrix}, \tag{1.5b}$$

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{I}_n \\ \boldsymbol{0}_{(m-n)\times n} \end{bmatrix}. \tag{1.5c}$$

The subscript PDFT in $\boldsymbol{A}_{\mathsf{PDFT}}$ stands for Partial DFT. We call this sensing ensemble the (randomly) sub-sampled Fourier ensemble.

1.2 A Statistical Perspective on Phase Retrieval

Modern statistical analyses of the phase retrieval problem seek to design computationally efficient estimators for recovering the signal $\boldsymbol{x}_\star$ using the minimum number of measurements.

A commonly used performance measure to quantify the quality of an estimator $\hat{\boldsymbol{x}}$ is the squared cosine similarity:

$$\cos^2(\angle(\boldsymbol{x}_\star, \hat{\boldsymbol{x}})) \stackrel{\text{def}}{=} \frac{|\langle \boldsymbol{x}_\star, \hat{\boldsymbol{x}} \rangle|^2}{\|\boldsymbol{x}_\star\|^2 \|\hat{\boldsymbol{x}}\|^2}. \tag{1.6}$$

This performance measure accounts for the inherent phase ambiguity in the phase retrieval problem: Since the signal vectors $\boldsymbol{x}_\star$ and $\boldsymbol{x}_\star e^{\mathbf{i}\phi}$ result in identical measurement vectors $\boldsymbol{y}$ for any $\phi \in \mathbb{R}$, it is possible to determine $\boldsymbol{x}_\star$ only upto a global phase. An estimator $\hat{\boldsymbol{x}}$ has good performance when $\cos^2(\angle(\boldsymbol{x}_\star, \hat{\boldsymbol{x}})) \approx 1$. In this case, the estimator provides an accurate estimate of the direction of the signal vector. On the other hand, when $\cos^2(\angle(\boldsymbol{x}_\star, \hat{\boldsymbol{x}})) \approx 0$, the estimator is nearly orthogonal to the signal vector, and hence uninformative.

Existing statistical analyses of the phase retrieval problem fall into roughly two categories:

Order-of-Magnitude Analyses: A number of recent statistical analyses of the phase retrieval problem design computationally efficient estimators which recover $\boldsymbol{x}_\star$ with information-theoretically rate-optimal $m = O(n)$ (or nearly optimal $m = O(n \operatorname{polylog}(n))$) measurements. A representative, but necessarily incomplete, list of such works includes the analysis of convex relaxations like PhaseLift [9, 10], PhaseMax [11, 12], and analysis of non-convex optimization-based methods [13, 5, 14]. The number of measurements required if the underlying signal has a low dimensional structure has also been investigated [15, 16, 17]. Though a number of these works study a physical unrealizable and stylized model of the sensing matrix, the order of magnitude of measurements required to solve the phase retrieval problem with certain sensing matrices that are close to practice such as the CDP ensemble (see (1.4)) is also understood: the works by Candès, Li, and Soltanolkotabi [18, 5] exhibit computationally efficient estimators for solving phase retrieval with CDP sensing ensembles with $m = O(n \operatorname{polylog}(n))$ measurements.

High-dimensional Asymptotic Analysis: The previously mentioned order-of-magnitude analyses show that a variety of different methods succeed in solving the phase retrieval problem with the optimal or nearly optimal order of magnitude of measurements. However, in practice, these meth-

ods can have a vast difference in performance, which is not captured by the order-of-magnitude analyses. Consequently, efforts have been made to complement these results with sharp high dimensional asymptotic analyses which shed light on the performance of different estimators and information-theoretic lower bounds in the high dimensional limit:

$$m, n \to \infty,\ m/n \to \delta. \tag{1.7}$$

The parameter δ is called the sampling ratio. This provides a high-resolution framework to compare different estimators based on the critical value of δ at which they achieve non-trivial performance (i.e. better than a random guess) or exact recovery of $\boldsymbol{x}_\star$. Comparing this to the critical value of δ required information-theoretically allows us to reason about the optimality of known estimators. This dissertation focuses on understanding the phase retrieval problem in the high-dimensional asymptotic regime.

A key challenge in analyzing the phase retrieval problem in the high-dimensional asymptotic regime (1.7) is that current techniques are unable to handle the highly-structured semi-random sensing matrices like the CDP ensemble (1.4) and the sub-sampled Fourier ensemble (1.5) that arise in practice. Consequently, various mathematically tractable, approximate models for sensing matrices have been proposed, which we introduce next. We refer to such a model as an ansatz, to emphasize that such a model is physically unrealizable, and has been chosen for mathematical convenience with the hope that it is a good approximation to sensing matrices that are closer to practice.

I.I.D. Gaussian Ansatz: In this ansatz, the entries of the sensing matrix are assumed to be i.i.d. Gaussian (real or complex). This is the most well-studied ensemble in the high dimensional asymptotic limit. For this ansatz, the precise performance curves for various estimators such as spectral methods [19, 20, 21], convex relaxation methods like PhaseLift [22] and PhaseMax [23], and a class of iterative algorithms called Approximate Message Passing [24] are now well understood. The precise asymptotic limit of the Bayes risk [25] for Bayesian phase

retrieval is also known. However, this ansatz does not accurately predict the performance of estimators on sensing ensembles closer to practice such as the CDP ensemble (1.4) and the sub-sampled Fourier ensemble (1.5).

Sub-sampled Haar Ansatz: In the sub-sampled Haar sensing ansatz, the sensing matrix is generated by picking the first n columns of a uniformly random $m \times m$ unitary matrix:

$$\boldsymbol{A} = \boldsymbol{H}_m \cdot \boldsymbol{S}_{m,n}, \tag{1.8a}$$

$$\boldsymbol{H}_m \sim \text{Unif}\left(\mathbb{U}_m\right), \ \boldsymbol{S}_{m,n} = \begin{bmatrix} \boldsymbol{I}_n \\ \boldsymbol{0}_{(m-n)\times n} \end{bmatrix}. \tag{1.8b}$$

The sub-sampled Haar ansatz captures a crucial aspect of sensing matrices that arise in practice: namely they have orthogonal columns (note that for both the CDP and the sub-sampled Fourier ensembles we have $\boldsymbol{A}_{\mathsf{PDFT}}^{\mathsf{H}}\boldsymbol{A}_{\mathsf{PDFT}} = \boldsymbol{A}_{\mathsf{CDP}}^{\mathsf{H}}\boldsymbol{A}_{\mathsf{CDP}} = \boldsymbol{I}_n$).

Rotationally Invariant Ansatz: This is a broad class of unstructured sensing ensembles that include the i.i.d. Gaussian ansatz and the sub-sampled Haar ansatz as special cases. Here, it is assumed that the SVD of the sensing matrix is given by:

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathsf{H}}, \tag{1.9a}$$

where $\boldsymbol{U}, \boldsymbol{V}$ are independent and uniformly random orthogonal matrices (or unitary in the complex case): $\boldsymbol{U} \sim \text{Unif}\left(\mathbb{U}(m)\right)$, $\boldsymbol{V} \sim \text{Unif}\left(\mathbb{U}(n)\right)$ and $\boldsymbol{S}$ is a deterministic matrix such that the empirical spectral distribution of $\boldsymbol{S}^{\mathsf{T}}\boldsymbol{S}$ converges to a limiting measure μ_S. This ansatz is able to exactly model the spectrum of sensing matrices of interest, but treats the singular vectors of the sensing matrix as generic. In comparision to the i.i.d. Gaussian ansatz, the subsampled Haar ansatz and the rotationally invariant ansatz are significantly less studied.

Universality Phenomena: Even though the sub-sampled Haar ansatz (1.8) is faithful only to a relatively coarse feature (column orthogonality) of the practically relevant sensing models, numerical simulations reveal an intriguing universality phenomenon: It has been observed that the performance curves derived theoretically for sub-sampled Haar ansatz provide a nearly perfect fit to the empirical performance on practical sensing ensembles like $\boldsymbol{A}_{\mathsf{CDP}}, \boldsymbol{A}_{\mathsf{PDFT}}$. This has been observed by several authors in the context of various signal processing problems. It was first pointed out by Donoho and Tanner [26] in the context of ℓ_1 norm minimization for noiseless compressed sensing and then again by Monajemi, Jafarpour, Gavish, and Donoho [27] for the same setup, but for many more structured sensing ensembles. More recently, Abbara, Baker, Krzakala, and Zdeborová [28] have observed this universality phenomenon in the context of approximate message passing algorithms for noiseless compressed sensing. For noiseless compressed sensing both the Gaussian ansatz and the sub-sampled Haar ansatz lead to identical predictions (and hence the simulations with structured sensing matrices match both of them). However, in noisy compressed sensing and non-linear inverse problems like phase retrieval, the predictions from the sub-sampled Haar ansatz and the Gaussian ansatz are different. The predictions from the sub-sampled Haar ansatz seem to be correct in simulations. Oymak and Hassibi [29] pointed out that structured ensembles generated by sub-sampling deterministic orthogonal matrices empirically behave like Sub-sampled Haar sensing matrices for noisy compressed sensing. In the context of phase retrieval, this phenomenon was reported by Ma, Dudeja, Xu, Maleki, and Wang [30] for the performance of the spectral method. The current theoretical understanding of this universality phenomenon is limited.

1.3 Overview of Contributions

The goal of this dissertation is to present some results that further our understanding of the phase retrieval problem in the high-dimensional asymptotic regime (1.7) for semi-random sensing matrices such as the CDP ensemble (1.4) and sub-sampled Fourier ensemble (1.5).

Towards this goal, we focus on understanding the performance of the spectral estimator for

phase retrieval. The spectral estimator is given by the largest eigenvector of a matrix $\boldsymbol{M}$ constructed using the measurements $\boldsymbol{y}$ and the sensing matrix $\boldsymbol{A}$ as follows:

$$\hat{\boldsymbol{x}} \stackrel{\text{def}}{=} \arg\max_{\|\boldsymbol{u}\|=1} \boldsymbol{u}^{\mathsf{H}} \boldsymbol{M} \boldsymbol{u}, \tag{1.10a}$$

$$\boldsymbol{M} \stackrel{\text{def}}{=} \boldsymbol{A}^{\mathsf{H}} \boldsymbol{T} \boldsymbol{A}, \tag{1.10b}$$

$$\boldsymbol{T} \stackrel{\text{def}}{=} \mathrm{Diag}(\mathcal{T}(y_1), \mathcal{T}(y_2), \ldots, \mathcal{T}(y_m)). \tag{1.10c}$$

In the above equation the function $\mathcal{T} : [0, \infty) \to \mathbb{R}$ is a suitable trimming function. This is a tuning parameter that can be chosen to optimize the performance of the spectral estimator. The spectral estimator is a widely used pilot estimator for phase retrieval. It is often used to initialize iterative algorithms which seek to solve the phase retrieval problem by optimizing a non-convex loss [13, 5, 31].

We study the performance of the spectral estimator under the sub-sampled Haar ansatz for the sensing matrix (1.8). Our choice of this ansatz is inspired by the empirical evidence for universality provided by previously mentioned prior works [26, 27, 28, 29, 30] which suggest that the sub-sampled Haar ansatz accurately describes the empirical performance of various estimators on practical sensing ensembles like $\boldsymbol{A}_{\mathsf{CDP}}, \boldsymbol{A}_{\mathsf{PDFT}}$. The main results obtained are summarized below:

1. In Chapter 3, we provide an expression for the limiting value of squared cosine similarity between the spectral estimator and the true signal for a broad class of trimming functions. Our analysis builds on the techniques introduced by Lu and Li [19] who analyzed the performance of the spectral estimator for the i.i.d. Gaussian ansatz. The precise expression for the limiting value had been previously conjectured by Ma, Dudeja, Xu, Maleki, and Wang [30], and the results of this chapter provide a proof for this conjecture. Figure 1.3 compares the theoretical performance curves for the sub-sampled Haar ansatz (obtained in Chapter 3) and the i.i.d. Gaussian ansatz (obtained by Lu and Li) with the empirical performance curves for the CDP ensemble. The figure suggests that the sub-sampled Haar ansatz accurately describes the empirical performance of spectral estimators on practical sensing ensembles like

the CDP ensemble, whereas the i.i.d. Gaussian ansatz does not.

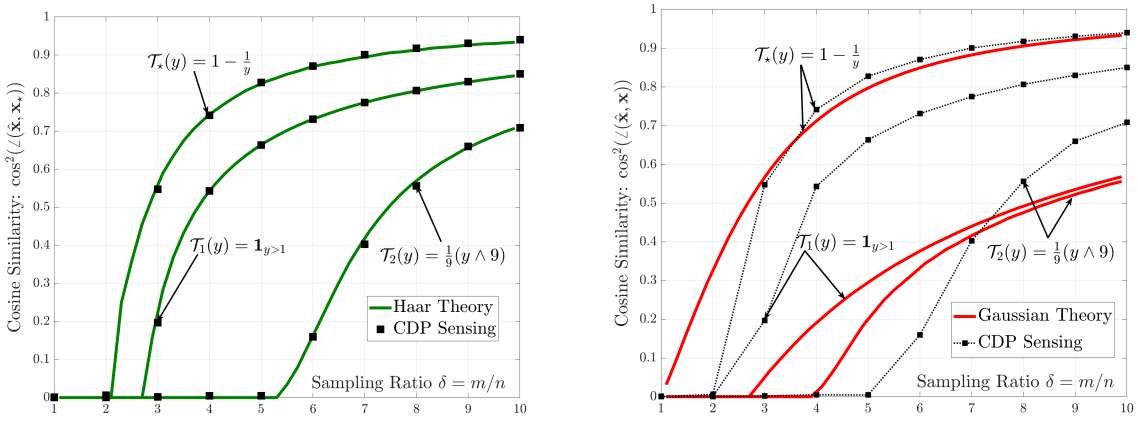

Figure 1.3: Comparison of the theoretical performance curves for the sub-sampled Haar ansatz (obtained in Chapter 3) and the i.i.d. Gaussian ansatz (obtained by Lu and Li) with the empirical performance curves for the CDP ensemble for three different trimming functions. $\mathcal{T}_\star$ is the optimal trimming function for the Gaussian [21] and the sub-sampled Haar sensing models [30]

2. Based on the conjectured formula for the limiting value of squared cosine similarity, Ma, Dudeja, Xu, Maleki, and Wang [30] derived the optimal choice of the trimming function $\mathcal{T}$. When $\delta > 2$, the optimal trimming functions achieves a non-trivial (or weak) recovery, that is,

$$\lim_{\substack{m,n\to\infty \\ m=\delta n}} \mathbb{E}\left[\cos^2(\angle(\boldsymbol{x}_\star, \hat{\boldsymbol{x}}))\right] > 0.$$

   In Chapter 4, we show that the threshold $\delta = 2$ is information-theoretically optimal: When $\delta < 2$, no estimator can achieve non-trivial (or weak) recovery. Our analysis in this chapter builds on the techniques used by Mondelli and Montanari [20], who proved the analogous result for the i.i.d. Gaussian ansatz.

3. In Chapter 5, we present some partial progress towards a mathematical understanding of the empirically observed universality. For the real-valued version of the phase retrieval problem, we show that the dynamics of a class of iterative algorithms that can match the performance

of any spectral estimator are asymptotically identical in the sub-sampled Haar ansatz (1.8) and a real-valued analog of the sub-sampled Fourier ensemble (1.5).

1.4 Notations

Notations for common sets

We use $\mathbb{N}, \mathbb{N}_0, \mathbb{R}, \mathbb{C}$ to denote the sets of natural numbers, non-negative integers, real numbers, and complex numbers, respectively.

$\mathbb{R}^n$ and $\mathbb{C}^n$ denote the n dimensional real and complex vector spaces respectively. $\mathbb{S}^{n-1} \subset \mathbb{C}^n$ is the set of complex n-dimensional vectors with unit norm.

The set of $m \times n$ real matrices is denoted by $\mathbb{R}^{m\times n}$ and the set of $m \times n$ complex matrices is denoted by $\mathbb{C}^{m\times n}$. $\mathbb{O}(m)$ refers to the set of all $m \times m$ orthogonal matrices and $\mathbb{U}(m)$ refers to the set of all $m \times m$ unitary matrices.

$[k]$ denotes the set $\{1, 2, \cdots, k\}$ and $[i : j]$ denotes the set $\{i, i + 1, i + 2 \cdots, j - 1, j\}$.

For Linear Algebraic Aspects

For a matrix $\boldsymbol{A}$, $\boldsymbol{A}^{\mathsf{H}}$ refers to the conjugate transpose of $\boldsymbol{A}$ and $\mathsf{Tr}(\cdot)$ denotes the trace of a square matrix.

For a matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, with real eigenvalues, we use $\lambda_1(\boldsymbol{A}) \geq \lambda_2(\boldsymbol{A}) \cdots \geq \lambda_n(\boldsymbol{A})$ to denote the eigenvalues arranged in descending order. We use $\sigma(\boldsymbol{A})$ to refer to the spectrum of $\boldsymbol{A}$ which is simply the set of eigenvalues $\{\lambda_1(\boldsymbol{A}), \lambda_2(\boldsymbol{A}) \ldots \lambda_n(\boldsymbol{A})\}$. We denote the largest and smallest eigenvalue of $\boldsymbol{A}$ by $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$. Finally we define the spectral measure of $\boldsymbol{A}$, denoted by $\mu_{\boldsymbol{A}}$ as,

$$\mu_{\boldsymbol{A}} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(\boldsymbol{A})}.$$

For $m, n \in \mathbb{N}$, we denote the $m \times m$ identity matrix by $\boldsymbol{I}_m$ and a $m \times n$ matrix of all zero

entries by $\mathbf{0}_{m,n}$. For $m \geq n$, We also define the special matrix $\boldsymbol{S}_{m,n}$ as:

$$\boldsymbol{S}_{m,n} \stackrel{\text{def}}{=} \begin{bmatrix} \boldsymbol{I}_n \\ \mathbf{0}_{m-n,n} \end{bmatrix}. \tag{1.11}$$

We use $\boldsymbol{e}_1, \boldsymbol{e}_2 \ldots, \boldsymbol{e}_n$ to denote the standard basis vectors in $\mathbb{R}^n$.

For vectors and matrices $\|\cdot\|$ denotes the ℓ_2 and the Frobenius norm respectively. For complex matrices $\|\cdot\|_{op}$ denotes the operator norm. For vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{C}^n$, the inner product $\langle \boldsymbol{a}, \boldsymbol{b} \rangle$ is defined as $\boldsymbol{a}^{\mathsf{H}}\boldsymbol{b}$. For matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{m \times n}$ the inner product $\langle \boldsymbol{A}, \boldsymbol{B} \rangle$ is defined as $\mathsf{Tr}(\boldsymbol{A}^{\mathsf{H}}\boldsymbol{B})$.

For Complex Analytic Aspects

For a complex number $z \in \mathbb{C}$, $\mathrm{Re}(z), \mathrm{Im}(z), \mathsf{Arg}(z), |z|, \overline{z}$ refer to the real part, imaginary part, argument, modulus and conjugate of z. We denote the complex upper half plane and lower half planes by

$$\mathbb{C}^+ \stackrel{\text{def}}{=} \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\} \text{ and } \mathbb{C}^- \stackrel{\text{def}}{=} \{z \in \mathbb{C} : \mathrm{Im}(z) < 0\}.$$

Notation for Asymptotic Analysis

We say a sequence $f(n)$ is $o(n)$ if $f(n)/n \to 0$ as $n \to \infty$. We use the generic constant C to refer to a positive finite constant that does not depend on m, n. This constant may change from line to line and may depend on the noise level σ (introduced in Chapter 4) and the sampling ratio δ unless stated otherwise. If this constant depends on any other parameters we will make this dependence explicit: For example, $C(\epsilon)$ denotes a positive, finite constant depending on some parameter ϵ, the noise level σ and possibly the sampling ratio δ but independent of m, n.

For Probabilistic Aspects

We denote almost sure convergence, convergence in probability and convergence in distribution by $\stackrel{\text{a.s.}}{\to}, \stackrel{\text{P}}{\to}$ and $\stackrel{\text{d}}{\to}$ respectively. If for a sequence of random variables we have $X_n \stackrel{\text{P}}{\to} c$ for a deterministic c, we say $\text{p-lim}\, X_n = c$. Two random variables X, Y are equal in distribution,

denoted by $X \overset{\mathrm{d}}{=} Y$ if they have the same distribution. For an event $\mathcal{E}$, $\mathbf{1}_{\mathcal{E}}$ denotes the indicator function of $\mathcal{E}$. For a probability measure μ, we use $\mathrm{Supp}(\mu)$ to denote the support of μ.

Some Special Distributions

The (real) multivariate Gaussian distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\Sigma}$ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. We say a complex random variable Z is standard complex Gaussian distributed, denoted by $Z \sim \mathcal{CN}(0,1)$ if $\mathrm{Re}(Z)$ and $\mathrm{Im}(Z)$ are i.i.d. $\mathcal{N}\left(0, \frac{1}{2}\right)$. We say a complex n-dimensional random vector $\boldsymbol{Z} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_n)$ if each entry $Z_i \overset{\text{i.i.d.}}{\sim} \mathcal{CN}(0,1)$. $\mathrm{Unif}(\mathbb{U}_m)$ denotes the Haar measure on the unitary group.

Chapter 2: Related Work

There are a large number of results on the phase retrieval problem with varying assumptions on the sensing matrix, studying different classes of estimators under different analysis frameworks. In this chapter, we summarize a few important and representative results. We organize our discussion as follows:

1. In Section 2.1, we summarize results that study the order-of-magnitude of measurements required to solve the phase retrieval problem.

2. In Section 2.2 we summarize results about the phase retrieval problem in the high-dimensional asymptotic regime (1.7).

3. In Section 2.3, we discuss empirical and theoretical studies of universality phenomena relevant to our work.

2.1 Order-of-Magnitude Analyses

A large number of estimators are known to solve the phase retrieval problem with the rate-optimal number of measurements $m = O(n)$ or the nearly optimal order-of-magnitude of measurements $m = O(n \cdot \text{poly}(\log(n)))$. The earliest such estimator is PhaseLift SDP relaxation proposed by Candès, Strohmer, and Voroninski [9]. A linear programming based relaxation called PhaseMax has also been proposed and analyzed by Goldstein and Studer [12] and Bahmani and Romberg [11]. More recently, approaches based on non-convex optimization have been analyzed. This includes an alternating minimization approach due to Netrapalli, Jain, and Sanghavi [13] and a gradient descent-based algorithm due to Candès, Li, and Soltanolkotabi [5]. Though a number of these works study a physical unrealizable and stylized model of the sensing matrix, order-of-magnitude analyses are flexible enough to extend to CDP sensing matrices. We refer the reader to

[18] for the analysis of PhaseLift for CDP matrices and to Candès, Li, and Soltanolkotabi [5] and Qu, Zhang, Eldar, and Wright [32] for the analysis of non-convex optimization approach for CDP matrices and random circulant sensing matrices respectively.

2.2 High-dimensional Asymptotic Analyses

Results for Gaussian Sensing Matrices: Order-of-magnitude analyses, though flexible, lack the resolution to compare the performance of various estimators which achieve the optimal sample complexity of $O(n)$ measurements. Consequently, recent years have seen a number of works that provide an analysis in the high dimensional asymptotic framework where $m, n \to \infty$ and $m/n = \delta$. Lu and Li [19] analyzed a class of spectral estimators in this asymptotic framework for Gaussian sensing matrices. Their analyses was leveraged by Mondelli and Montanari [20] and Luo, Alghamdi, and Lu [21] to design spectral estimators with optimal performance. Convex relaxation-based approaches, such as PhaseLift and PhaseMax have also been analyzed in this framework for Gaussian sensing matrices [23, 22]. Bayati and Montanari [24] have analyzed the dynamics of a broad class of iterative algorithms called Approximate Message Passing schemes, which seem to be capable of computing many estimators for a broad range of inference problems, including phase retrieval.

Information Theoretic Lower Bounds for Gaussian Sensing Matrices Mondelli and Montanari [20] showed that the weak recovery threshold for Gaussian sensing matrices was $\delta_{\mathsf{weak}} = 1$. Barbier, Krzakala, Macris, Miolane, and Zdeborová [25] have used interpolation methods to obtain expressions for the asymptotic Bayes risk for estimating generalized linear models. This includes real-valued phase retrieval with Gaussian sensing matrices as a special case. In particular, their results recover the results of Mondelli and Montanari [20] as a special case and also shed light on the minimum mean square error achievable above the weak recovery threshold. This work also shows that the expression of the Bayes risk for any sensing matrix with i.i.d. entries with some mild moment assumptions is the same as the Bayes risk for Gaussian sensing matrices.

Sharp Asymptotic Analyses for Non-i.i.d. Sensing Matrices Rigorous results for non-i.i.d. sensing matrices in the high dimensional asymptotic framework are limited. Thrampoulidis and Hassibi [33] provide an analysis of the generalized Lasso estimator for compressed sensing using uniformly random row orthogonal matrices using the Convex Gaussian Minmax Theorem (CGMT) framework. The analysis of Approximate Message Passing algorithms has been extended to the rotationally invariant ansatz (1.9) by Schniter, Rangan, and Fletcher [34], Rangan, Schniter, and Fletcher [35], and Takeuchi [36]. We note that the non-rigorous replica method can be used to derive conjectures for the asymptotic Bayes risk for the large class of rotationally invariant sensing ansatz (1.9) which includes sub-sampled Haar sensing ansatz (1.8) as a special case. The application of the replica method to rotationally invariant ensembles was pioneered in a sequence of papers by Takeda, Uda, and Kabashima [37], Takeda, Hatabu, and Kabashima [38] and Kabashima [39]. We refer the reader to Reeves [40] for a recent derivation of these conjectures. To the best of our knowledge, these conjectures have not been rigorously proved except in a few special cases, none of which cover the sub-sampled Haar sensing matrix. The only rigorous result about sharp information-theoretic lower bounds for non-i.i.d. sensing matrices is due to Barbier, Macris, Maillard, and Krzakala [41] who provide the expression for the limiting Bayes risk for a certain class of sensing matrices. The class of sensing matrices they consider are formed by a product of independent matrices each consisting of i.i.d. entries. This is significantly different from the sub-sampled Haar sensing model which we consider here. Moreover, the sensing problem they study is the real linear sensing problem and not the phase retrieval problem that we study here. Lastly, we note that the non-rigorous replica method has also been used to analyze convex relaxation methods like LASSO [42, 43] for rotationally invariant sensing matrices.

2.3 Universality Results

Empirical Results: It has been observed that the performance curves derived theoretically for sub-sampled Haar sensing provide a nearly perfect fit to the empirical performance of estimators on practical sensing ensembles like $\boldsymbol{A}_{\mathsf{CDP}}, \boldsymbol{A}_{\mathsf{DFT}}$. This has been observed by a number of authors

in the context of various signal processing problems. It was first pointed out by Donoho and Tanner [26] in the context of ℓ_1 norm minimization for noiseless compressed sensing and then again by Monajemi, Jafarpour, Gavish, and Donoho [27] for the same setup but for many more structured sensing ensembles. For noiseless compressed sensing both the Gaussian ensemble and the sub-sampled Haar ensemble lead to identical predictions (and hence the simulations with structured sensing matrices match both of them). However, in noisy compressed sensing, the predictions from the sub-sampled Haar model and the Gaussian model are different. Oymak and Hassibi [29] pointed out that structured ensembles generated by sub-sampling deterministic orthogonal matrices empirically behave like Sub-sampled Haar sensing matrices. More recently, Abbara, Baker, Krzakala, and Zdeborová [28] have observed this universality phenomenon in the context of approximate message passing algorithms for noiseless compressed sensing. In the context of phase retrieval, this phenomenon was reported by Ma, Dudeja, Xu, Maleki, and Wang [30] for the performance of the spectral method.

Gaussian Universality: A number of papers have tried to explain the observations of Donoho and Tanner [26] regarding the universality in performance of ℓ_1 minimization for noiseless linear sensing. For noiseless linear sensing, the Gaussian sensing ensemble, sub-sampled Haar sensing ensemble, and structured sensing ensembles like sub-sampled Fourier sensing ensemble behave identically. Consequently, a number of papers have tried to identify the class of sensing matrices which behave like Gaussian sensing matrices. It has been shown that sensing matrices with i.i.d. entries under mild moment assumptions behave like Gaussian sensing matrices in the context of the performance of general (non-linear) Approximate Message Passing schemes [24, 44], the limiting Bayes risk [41], and the performance of estimators based on convex optimization [45, 46]. The assumption that the sensing matrix has i.i.d. entries has been relaxed to the assumption that it has i.i.d. rows (with possible dependence within a row) [22]. Finally, we emphasize that in the presence of noise or when the measurements are non-linear, the structured ensembles that we consider here, obtained by sub-sampling a deterministic orthogonal matrix like the DFT matrix or

the Hadamard-Walsh matrix, no longer behave like Gaussian matrices, but rather like sub-sampled Haar matrices.

A result for highly structured ensembles: While the results mentioned above move beyond i.i.d. Gaussian sensing, the sensing matrices they consider are still largely unstructured and highly random. In particular, they do not apply to the sub-sampled Fourier or CDP ensembles. A notable exception is the work of Donoho and Tanner [47] which considers a random undetermined system of linear equations (in $\boldsymbol{x}$) of the form $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}_0$ for a random matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and a k-sparse non-negative vector $\boldsymbol{x}_0 \in \mathbb{R}^n_{\geq 0}$. Donoho and Tanner show that as $m, n, k \to \infty$ such that $n/m \to \kappa_1, k/m \to \kappa_2$, the probability that $\boldsymbol{x}_0$ is the unique non-negative solution to the system sharply transitions from 0 to 1 depending on the values κ_1, κ_2. Moreover, this transition is universal across a wide range of random $\boldsymbol{A}$, including Gaussian ensembles, random matrices with i.i.d. entries sampled from a symmetric distribution, and highly structured ensembles whose null space is given by a random matrix $\boldsymbol{B} \in \mathbb{R}^{n-m\times n}$ generated by multiplying the columns of a fixed matrix $\boldsymbol{B}_0$ whose columns are in general position by i.i.d. random signs. The proof technique of Donoho and Tanner uses results from the theory of random polytopes and it is not obvious how to extend their techniques beyond the case of solving underdetermined linear equations.

Universality Results in Random Matrix Theory: The phenomena that structured orthogonal matrices, such as Hadamard and Fourier matrices, behave like random Haar matrices in some aspects has been studied in the context of random matrix theory [48] and in particular free probability [49]. A well known result in free probability (see the book of Mingo and Speicher [49] for a textbook treatment) is that if $\boldsymbol{U} \sim \text{Unif}\big(\mathbb{U}(m)\big)$ and $\boldsymbol{D}_1, \boldsymbol{D}_2$ are deterministic $m\times m$ diagonal matrices then $\boldsymbol{U}\boldsymbol{D}_1\boldsymbol{U}^{\mathsf{H}}$ and $\boldsymbol{D}_2$ are asymptotically free and consequently the limiting spectral distribution of matrix polynomials in $\boldsymbol{D}_2$ and $\boldsymbol{U}\boldsymbol{D}_1\boldsymbol{U}^{\mathsf{H}}$ can be described in terms of the limiting spectral distribution of $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$. Tulino, Caire, Shamai, and Verdu [50] and Farrell [51] have obtained an extension of this result where a Haar unitary matrix is replaced by $m \times m$ Fourier matrix $\boldsymbol{F}_m$: If $\boldsymbol{D}_1, \boldsymbol{D}_2$ are independent diagonal matrices then $\boldsymbol{F}_m\boldsymbol{D}_1\boldsymbol{F}_m^{\mathsf{H}}$ is asymptotically free from $\boldsymbol{D}_2$. The

result of these authors has been extended to other deterministic orthogonal/unitary matrices (such as the Hadamard-Walsh matrix) conjugated by random signed permutation matrices by Anderson and Farrell [52].

Non-rigorous Results from Statistical Physics: In the statistical physics literature Cakmak, Opper, Winther, and Fleury [53, 54, 55, 56, 57] have developed an analysis of message passing algorithms for rotationally invariant ensembles via a non-rigorous technique called the dynamical functional theory. These works are interesting because they do not heavily rely on rotational invariance, but instead rely on results from Free probability. Since some of the free probability results have been extended to Fourier and Hadamard matrices [50, 51, 52], there is hope to generalize their analysis beyond rotationally invariant ensembles. However, currently, their results are non-rigorous due to two reasons: 1) due to the use of dynamical field theory, and 2) their application of Free probability results neglects dependence between matrices. In our work in Chapter 5, we avoid the use of dynamical functional theory since we analyze linearized AMP algorithms, and furthermore, we properly account for dependence that is heuristically neglected in their work.

The Hidden Manifold Model: Lastly, we discuss the recent works of Goldt, Mézard, Krzakala, and Zdeborová [58], Gerace, Loureiro, Krzakala, Mézard, and Zdeborová [59], and Goldt, Reeves, Mézard, Krzakala, and Zdeborová [60], where they study statistical learning problems where the feature matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ (the analogue of the sensing matrix in statistical learning) is generated as:

$$\boldsymbol{A} = \sigma(\boldsymbol{Z}\boldsymbol{F}),$$

where $\boldsymbol{F} \in \mathbb{R}^{d \times n}$ is a generic (possibly structured) deterministic weight matrix and $\boldsymbol{Z} \in \mathbb{R}^{m \times d}$ is an i.i.d. Gaussian matrix. The function $\sigma : \mathbb{R} \to \mathbb{R}$ acts entry-wise on the matrix $\boldsymbol{Z}\boldsymbol{F}$. For this model, the authors have analyzed the dynamics of online (one-pass) stochastic gradient descent (first non-rigorously [58] and then rigorously [60]) and the performance of regularized empirical

risk minimization with convex losses (non-rigorously) via the replica method [59] in the high dimensional asymptotic $m, n, d \to \infty$, $n/m \to \kappa_1, d/m \to \kappa_2$. Their results show that in this case the feature matrix behaves like a certain correlated Gaussian feature matrix. We note that the feature matrix $\boldsymbol{A}$ here is quite different from the sub-sampled Fourier ensemble (1.5) or the CDP ensemble (1.4) since it uses $O(m^2)$ i.i.d. random variables ($\boldsymbol{Z}$) where as the sub-sampled Fourier ensemble only uses m random variables (to specify the permutation matrix $\boldsymbol{P}$). However, a technical result proved by the authors (Lemma A.2 of [58]) appears to be a special case of a classical result of Mehler [61] and Slepian [62] which we find useful in our analysis in Chapter 5.

Chapter 3: Analysis of Spectral Estimators

In this chapter[1], we provide an analysis of the performance of spectral estimators for the sub-sampled Haar sensing ansatz.

3.1 Problem Formulation

3.1.1 Measurement Model and Spectral Estimator

In the phase retrieval problem we are given m observations $\boldsymbol{y} \in \mathbb{R}^m$ generated as:

$$\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}_\star|^2$$

where $\boldsymbol{x}_\star \in \mathbb{C}^n$ is the unknown signal vector and $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ is the sensing matrix. We assume that $\|\boldsymbol{x}_\star\| = \sqrt{m}$ and that the matrix $\boldsymbol{A}$ is generated according to the following process: Sample $\boldsymbol{H}_m \in \mathbb{U}(m)$ from the Haar measure on the unitary group $\mathbb{U}(m)$ and set $\boldsymbol{A}$ to be the matrix formed by picking the first n columns of $\boldsymbol{H}_m$. More formally,

$$\boldsymbol{A} = \boldsymbol{H}\boldsymbol{S}_{m,n},\ \boldsymbol{H} \sim \text{Unif}(\mathbb{U}(m)),$$

and $\boldsymbol{S}$ is defined in (1.11). An important parameter for our analysis will be the sampling ratio, denoted by $\delta \stackrel{\text{def}}{=} m/n$. Let $\mathcal{T} : \mathbb{R}_{\geq 0} \to \mathbb{R}$ be a trimming function. We study spectral estimators $\hat{\boldsymbol{x}}$ constructed as the leading eigenvector of the matrix $\boldsymbol{M}$, defined below:

$$\hat{\boldsymbol{x}} = \arg\max_{\|\boldsymbol{u}\|=1} \boldsymbol{u}^{\mathsf{H}}\boldsymbol{M}\boldsymbol{u},$$

[1]The results obtained in this chapter have been published in the paper R. Dudeja, M. Bakhshizadeh, J. Ma, and A. Maleki, "Analysis of spectral methods for phase retrieval with random orthogonal matrices," IEEE Transactions on Information Theory, 2020

where $\boldsymbol{M} = \boldsymbol{A}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}$ and $\boldsymbol{T} = \text{Diag}(\mathcal{T}(y_1), \mathcal{T}(y_2) \dots \mathcal{T}(y_m))$.

3.1.2 Assumptions & Asymptotic Framework

We analyze the performance of the spectral estimator in an asymptotic setup where $n, m \to \infty, m/n = \delta > 1$. In particular, we consider a sequence of independent phase retrieval problems realized on the same probability space with increasing n, m. We assume some regularity assumptions on the trimming function $\mathcal{T}$ which are stated below.

Assumption 1. *The trimming function $\mathcal{T}$ satisfies the following conditions:*

1. *$\mathcal{T}$ is Lipschitz continuous.*
2. *$\sup_{y\geq 0} \mathcal{T}(y) = 1,\ \inf_{y\geq 0} \mathcal{T}(y) = 0$.*
3. *The random variable T, defined by $Z \sim \mathcal{CN}(0,1)$ and $T = \mathcal{T}(|Z|^2)$ has a density with respect to the Lebesgue measure on $\mathbb{R}$.*

In the following remarks, we discuss why each of these assumptions are required and whether they can be relaxed.

Remark 1. *We need the trimming function $\mathcal{T}$ to be Lipschitz continuous so that the trimmed measurements $\mathcal{T}(y_i)$ can be approximated in distribution by $\mathcal{T}(|Z|^2), Z \sim \mathcal{CN}(0,1)$. We expect this approximation to hold under weaker smoothness hypothesis on $\mathcal{T}$ than Lipschitz continuity.*

Remark 2. *The assumptions:*

$$\sup_{y\geq 0} \mathcal{T}(y) = 1,\ \inf_{y\geq 0} \mathcal{T}(y) = 0$$

are no stronger than the assumption that $\mathcal{T}$ is a bounded trimming function. In fact, given any arbitary bounded trimming function with $\inf_{y\geq 0} \mathcal{T}(y) = a$ and $\sup_{y\geq 0} \mathcal{T}(y) = b$, the spectral estimator constructed using $\mathcal{T}$ has the same performance as the spectral measure constructed

using

$$\tilde{\mathcal{T}}(y) \stackrel{def}{=} (\mathcal{T}(y) - a)/(b-a).$$

This is because,

$$\begin{aligned}\widetilde{\boldsymbol{M}} \stackrel{def}{=} \boldsymbol{A}^{\mathsf{H}}\widetilde{\boldsymbol{T}}\boldsymbol{A} &= \frac{1}{b-a}\boldsymbol{A}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A} - \frac{a}{b-a}\boldsymbol{I}_n \\ &= \frac{1}{b-a}\boldsymbol{M} - \frac{a}{b-a}\boldsymbol{I}_n.\end{aligned}$$

In particular $\boldsymbol{M}$ and $\widetilde{\boldsymbol{M}}$ have the same leading eigenvector. We require the assumption that the trimming function is bounded since a number of results in free probability theory that we rely on assume this.

Remark 3. *We need (3) in Assumption 1 to ensure that the limiting spectral measure of the matrix $\boldsymbol{M}$ has no discrete component. We expect that this assumption can be completely removed by a careful analysis since the location of point masses in the limiting spectral measure of $\boldsymbol{M}$ is well understood.*

3.2 Main Result

In order to state our main result about the performance of the spectral estimator, we need to introduce the following four functions:

$$\begin{aligned}\Lambda(\tau) \triangleq \tau - \frac{(1-1/\delta)}{\mathbb{E}\left[\frac{1}{\tau - T}\right]},\ \psi_1(\tau) \triangleq \frac{\mathbb{E}\left[\frac{|Z|^2}{\tau-T}\right]}{\mathbb{E}\left[\frac{1}{\tau-T}\right]}, \\ \psi_2(\tau) \triangleq \frac{\mathbb{E}\left[\frac{1}{(\tau-T)^2}\right]}{\left(\mathbb{E}\left[\frac{1}{\tau-T}\right]\right)^2},\ \psi_3^2(\tau) \stackrel{\text{def}}{=} \frac{\mathbb{E}\left[\frac{|Z|^2}{(\tau-T)^2}\right]}{\left(\mathbb{E}\left[\frac{1}{\tau-T}\right]\right)^2}.\end{aligned} \tag{3.1}$$

In the above display, the random variables Z, T have the joint distribution given by $Z \sim \mathcal{CN}(0,1)$, $T = \mathcal{T}(|Z|^2)$. The functions Λ, ψ_1 are defined on $[1, \infty)$ and the functions ψ_2, ψ_3 are defined on $(1, \infty)$.

Remark 4. *Under Assumption 1, the support of the random variable T is the interval $[0,1]$. Hence the definition of these functions at $\tau = 1$ needs some clarification. First, note that the random variable $(1-T)^{-1} \geq 0$. Hence, the $\mathbb{E}[(1-T)^{-1}]$ is well-defined, but maybe ∞. If it is finite, each of the above functions are well-defined at $\tau = 1$. If $\mathbb{E}[(1-T)^{-1}] = \infty$, we define, $\Lambda(1) = 1, \psi_1(1) = 1$. This corresponds to interpreting $1/\infty = 0$ and $\infty/\infty = 1$ in the definition of these functions.*

Theorem 1. *Define $\tau_r \triangleq \arg\min_{\tau \in [1,\infty)} \Lambda(\tau)$. Also, let $\theta_\star$ denote the unique value of $\theta > \tau_r$ that satisfies $\psi_1(\theta) = \frac{\delta}{\delta - 1}$. Then, under Assumption 1, we have*

$$\lambda_1(\boldsymbol{M}) \overset{a.s.}{\to} \begin{cases} \Lambda(\tau_r), & \psi_1(\tau_r) \leq \frac{\delta}{\delta-1}, \\ \Lambda(\theta_\star), & \psi_1(\tau_r) > \frac{\delta}{\delta-1}. \end{cases}$$

Furthermore,

$$\frac{|\boldsymbol{x}_\star^{\mathsf{H}} \hat{\boldsymbol{x}}|^2}{\|\boldsymbol{x}_\star\|^2} \overset{a.s.}{\to} \begin{cases} 0, & \psi_1(\tau_r) < \frac{\delta}{\delta-1}, \\ \frac{\left(\frac{\delta}{\delta-1}\right)^2 - \frac{\delta}{\delta-1} \cdot \psi_2(\theta_\star)}{\psi_3(\theta_\star)^2 - \frac{\delta}{\delta-1} \cdot \psi_2(\theta_\star)}, & \psi_1(\tau_r) > \frac{\delta}{\delta-1}. \end{cases}$$

Remark 5. *The proof of Theorem 1 shows that if $\psi_1(\tau_r) > \delta/(\delta-1)$, there exists exactly one solution to the equation $\psi_1(\theta) = \delta/(\delta-1),\ \theta \in (\tau_r, \infty)$. Hence, $\theta_\star$ is well-defined.*

The proof of this result is postponed until Section 3.5. Before we proceed to the proof of this theorem, let us clarify some of its interesting features. First, note that similar to the Gaussian sensing matrices, even in the case of partial orthogonal matrices, the maximum eigenvector exhibits a phase transition behavior. For certain values of $\delta > 1$, the inequality $\psi_1(\tau_r) < \frac{\delta}{\delta-1}$ holds, and hence the maximum eigenvector does not carry information about $\boldsymbol{x}_*$. For other values of δ, the inequality $\psi_1(\tau_r) > \frac{\delta}{\delta-1}$ holds and hence, the direction of the maximum eigenvector starts to offer information about the direction of $\boldsymbol{x}_*$. For typical choices of the trimming function $\mathcal{T}$, there exists

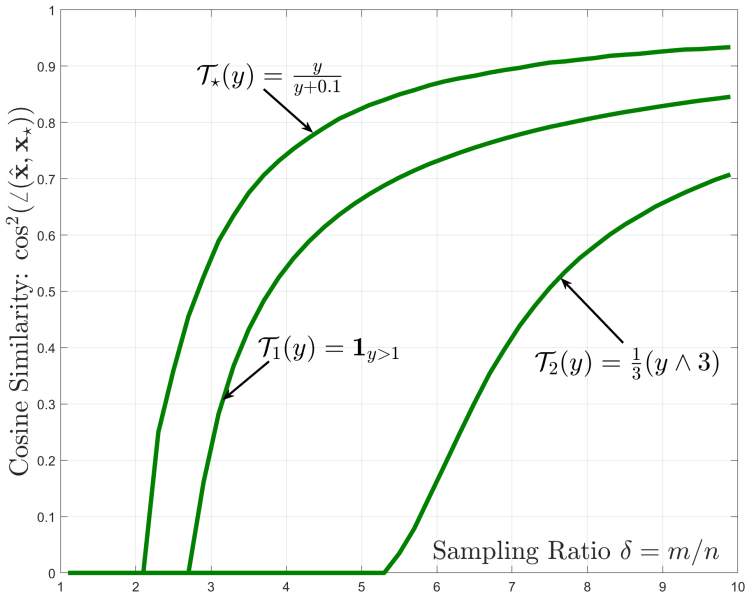

Figure 3.1: Plot of the asymptotic cosine similarity between $\hat{\boldsymbol{x}}$ and $\boldsymbol{x}_\star$ for three different choices of the trimming function.

a critical value of δ, denoted by $\delta_{\mathcal{T}}$ such that, when $\delta < \delta_{\mathcal{T}}$, the spectral estimator is asymptotically orthogonal to the signal vector. When $\delta > \delta_{\mathcal{T}}$, the spectral estimator makes a non-trivial angle with the signal vector. This phase transition phenomena is illustrated in Figure 3.1 for 3 different choices of $\mathcal{T}$.

Remark 6 (Choice of Trimming function)**.** *In Figure 3.1, we plot the asymptotic cosine similarity given by Theorem 1 for various values of the sampling ratio δ and 3 different trimming functions. The trimming function $\mathcal{T}_\star(y) = y/(y + 0.1)$ is a regularized version of the optimal trimming function for the i.i.d. Gaussian sensing model computed by Luo, Alghamdi, and Lu [21].*

Remark 7 (Extensions to generalized linear measurements)**.** *While we focus on the phase retrieval problem in this dissertation, our results extend straightforwardly to the generalized linear estimation, where the measurements y_i are generated as follows:*

$$y_i \sim f(\cdot|(\boldsymbol{A}\boldsymbol{x}_\star)_i),$$

where $f(\cdot|\cdot)$ denotes a conditional distribution modelling a possibly randomized output channel. Under suitable regularity assumptions on f, Theorem 1 holds with the change that the joint distribution of the random variables T, Z is now given by:

$$Z \sim \mathcal{CN}(0,1),\ Y \sim f(\cdot|Z),\ T = \mathcal{T}(Y).$$

3.3 Optimal Trimming Functions

Theorem 1 can used to design the trimming function $\mathcal{T}$ optimally in order to obtain the best possible value of $|\boldsymbol{x}_\star^{\mathsf{H}}\hat{\boldsymbol{x}}|^2$. Most of the work towards this goal was already done in [30] where the result in Theorem 1 was stated as a conjecture and was used to design the optimal trimming function. In particular, [30] showed the following impossibility result.

Proposition 1 ([30])**.** *Let $\mathcal{T}$ be any trimming function for which Theorem 1 holds. Then,*

$$\limsup_{\substack{m,n\to\infty \\ m=n\delta}} \frac{|\boldsymbol{x}_\star^{\mathsf{H}}\hat{\boldsymbol{x}}|^2}{\|\boldsymbol{x}_\star\|^2} \overset{a.s.}{\leq} \rho^2_{\mathsf{opt}}(\delta),$$

where,

$$\rho^2_{\mathsf{opt}}(\delta) \overset{def}{=} \begin{cases} 0, & \delta \leq 2 \\ \frac{\theta_\star^{\mathsf{opt}}-1}{\theta_\star^{\mathsf{opt}}-\frac{1}{\delta}}, & \delta > 2 \end{cases},$$

where $\theta_\star^{\mathsf{opt}}$ is the solution to the equation (in τ):

$$\psi_1^{\mathsf{opt}}(\tau) = \frac{\delta}{\delta-1},\ \psi_1^{\mathsf{opt}}(\tau) \overset{def}{=} \frac{\mathbb{E}\left[\frac{|Z|^2}{\tau - T_{\mathsf{opt}}}\right]}{\mathbb{E}\left[\frac{1}{\tau - T_{\mathsf{opt}}}\right]},\ \tau \in (1,\infty),$$

which exists uniquely when $\delta > 2$ and, the random variable T_{opt} is distributed as:

$$Z \sim \mathcal{CN}(0,1),\ T_{\mathsf{opt}} = 1 - \frac{1}{|Z|^2}.$$

The work [30] also provided a candidate for the optimal trimming function:

$$\mathcal{T}_{\mathsf{opt}}(y) = 1 - \frac{1}{y}.$$

They showed that if the characterization given in Theorem 1 holds for $\mathcal{T}_{\mathsf{opt}}$, then it achieves the asymptotic squared correlation $\rho^2_{\mathsf{opt}}(\delta)$. Unfortunately, since $\mathcal{T}_{\mathsf{opt}}$ is unbounded, Theorem 1 does not apply to it. Extending Theorem 1 to unbounded trimming functions would likely require extending previously known results in free probability to unbounded measures, and we don't pursue this approach in our work. Instead, we suitably modify the arguments of [30] to show that the family of bounded trimming functions:

$$\mathcal{T}_{\mathsf{opt},\epsilon}(y) = 1 - \frac{1}{y+\epsilon},\ \epsilon > 0,$$

attains an asymptotic squared correlation that can be made arbitrarily close to $\rho^2(\delta)$ as $\epsilon \downarrow 0$.

Proposition 2. *Let $\hat{\boldsymbol{x}}_\epsilon$ denote the spectral estimator for $\boldsymbol{x}_\star$ obtained by using $\mathcal{T}_{\mathsf{opt},\epsilon}$ as the trimming function. We have, almost surely,*

$$\lim_{\epsilon \downarrow 0} \lim_{\substack{m,n \to \infty \\ m = n\delta}} \frac{|\boldsymbol{x}_\star^{\mathsf{H}} \hat{\boldsymbol{x}}_\epsilon|^2}{\|\boldsymbol{x}_\star\|^2} = \rho^2_{\mathsf{opt}}(\delta).$$

We provide a proof of this result in Appendix A.2.

The regularized trimming functions $\mathcal{T}_{\mathsf{opt},\epsilon}$ are not only useful from a theoretical point of view to prove an achievability result, but also from a computational stand point: In simulations we have observed that the power iterations are slow to converge when $\mathcal{T}_{\mathsf{opt}}$ is used as the trimming function due to presence of large negative eigenvalues and this problem is mitigated by using $\mathcal{T}_{\mathsf{opt},\epsilon}$ with a small value of ϵ (such as 0.1 or 0.01) with a negligible degradation in performance.

3.4 Some Additional Notation

In this section, we introduce some additional notation we will find useful in this chapter.

The random variables Z, T: Throughout this chapter, the random variables Z, T refer to the pair of random variables with the joint distribution given by $Z \sim \mathcal{CN}(0,1), T = \mathcal{T}(|Z|^2)$.

Notation for topological aspects: Let A be a subset of $\mathbb{R}$ or $\mathbb{C}$. $\overline{A}$ denotes the closure of A. The distance from a point $x \in \mathbb{R}$ to A is defined by $\text{dist}(x, A) = \inf_{y \in A} |x - y|$. We define the ϵ neighborhood of A, denoted by A_ϵ as

$$A_\epsilon \stackrel{\text{def}}{=} \{x : \text{dist}(x, A) < \epsilon\}.$$

The symbol $\emptyset$ is used to denote the empty set.

3.5 Proof of Theorem 1

3.5.1 Roadmap

Our proof follows the general strategy taken by Lu and Li [19]. In this subsection, we state several key lemmas and show how they fit together in the proof of Theorem 5. First we note that without loss of generality, for the purpose of analysis of the spectral estimator, we can assume $\boldsymbol{x}_\star = \sqrt{m}\boldsymbol{e}_1$. The following lemma supports this claim.

Lemma 1. *The distribution of the cosine similarity, $\rho^2 = |\boldsymbol{x}_\star^{\mathsf{H}} \hat{\boldsymbol{x}}|^2 / \|\boldsymbol{x}_\star\|^2$ is independent of $\boldsymbol{x}_\star$.*

Proof. Let $\boldsymbol{x}_\star$ be an arbitrary signal vector with $\|\boldsymbol{x}_\star\| = \sqrt{m}$. Let $\boldsymbol{y}, \boldsymbol{T}, \hat{\boldsymbol{x}}$ denote the measurements, trimmed measurements and spectral estimate generated when the sensing matrix was $\boldsymbol{A}$ and the signal vector was $\boldsymbol{x}_\star$. Note that the cosine similarity ρ^2 is a (deterministic) function of $\boldsymbol{A}, \boldsymbol{x}_\star$ and hence we use the notation $\rho^2(\boldsymbol{A}, \boldsymbol{x}_\star)$ to denote the cosine similarity when the sensing matrix is $\boldsymbol{A}$ and the signal vector is $\boldsymbol{x}_\star$.

Let $\boldsymbol{\Gamma} \in \mathbb{U}(n)$ be such that $\sqrt{m}\boldsymbol{\Gamma}\boldsymbol{e}_1 = \boldsymbol{x}_\star$. We have $\boldsymbol{x}_\star^{\mathsf{H}} \hat{\boldsymbol{x}} = \sqrt{m}\boldsymbol{e}_1^{\mathsf{H}} \boldsymbol{\Gamma}^{\mathsf{H}} \hat{\boldsymbol{x}}$. Next we note that $\hat{\boldsymbol{x}}' \stackrel{\text{def}}{=} \boldsymbol{\Gamma}^{\mathsf{H}} \hat{\boldsymbol{x}}$ is the leading eigenvector of the matrix $\boldsymbol{M}' \stackrel{\text{def}}{=} \boldsymbol{\Gamma}^{\mathsf{H}} \boldsymbol{M} \boldsymbol{\Gamma} = (\boldsymbol{A}\boldsymbol{\Gamma})^{\mathsf{H}} \boldsymbol{T} \boldsymbol{A}\boldsymbol{\Gamma} = \boldsymbol{A}'^{\mathsf{H}} \boldsymbol{T} \boldsymbol{A}'$, where we defined $\boldsymbol{A}' \stackrel{\text{def}}{=} \boldsymbol{A}\boldsymbol{\Gamma}$. Noting that $\boldsymbol{T}$ is a diagonal matrix consisting of the trimmed observations $\boldsymbol{y} = |\boldsymbol{A}\boldsymbol{x}_\star|^2 = \sqrt{m}|\boldsymbol{A}'\boldsymbol{e}_1|$, we conclude that $\hat{\boldsymbol{x}}'$ is the spectral estimate generated when the

sensing matrix was $\boldsymbol{A}'$ and the signal vector was $\sqrt{m}\boldsymbol{e}_1$. Hence, we have concluded that

$$\rho^2(\boldsymbol{A}, \boldsymbol{x}_\star) = \rho^2(\boldsymbol{A}', \sqrt{m}\boldsymbol{e}_1).$$

Next we note that $\boldsymbol{A}$ was generated from the sub-sampled Haar model, that is $\boldsymbol{A} = \boldsymbol{H}_m\boldsymbol{S}_{m,n}$ where $\boldsymbol{H}_m \sim \text{Unif}(\mathbb{U}(m))$. Since the Haar measure on $\mathbb{U}(n)$ is invariant to right multiplication by unitary matrices, we have

$$\boldsymbol{H}_m \stackrel{\text{d}}{=} \boldsymbol{H}_m \cdot \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I}_{m-n} \end{bmatrix},$$

where the notation $\stackrel{\text{d}}{=}$ means that two random vectors have the same distributions. Consequently $\boldsymbol{A} = \boldsymbol{H}_m\boldsymbol{S}_{m,n} \stackrel{\text{d}}{=} \boldsymbol{A}\boldsymbol{\Gamma} = \boldsymbol{A}'$. Therefore, $\rho^2(\boldsymbol{A}, \boldsymbol{x}_\star) = \rho^2(\boldsymbol{A}', \sqrt{m}\boldsymbol{e}_1) \stackrel{\text{d}}{=} \rho^2(\boldsymbol{A}, \sqrt{m}\boldsymbol{e}_1)$, and the distribution of ρ^2 is independent of $\boldsymbol{x}_\star$. □

In the light of the above lemma, in the rest of the chapter, we will assume $\boldsymbol{x}_\star = \sqrt{m}\boldsymbol{e}_1$. Next, we partition $\boldsymbol{A}$ by separating the first column

$$\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_{-1}],$$

where $\boldsymbol{A}_{-1}$ denotes all the remaining columns of $\boldsymbol{A}$ (except $\boldsymbol{A}_1$). Hence we can partition $\boldsymbol{A}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}$ in the following way:

$$\boldsymbol{A}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}_1 & \boldsymbol{A}_1^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}_{-1} \\ \boldsymbol{A}_{-1}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}_1 & \boldsymbol{A}_{-1}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}_{-1} \end{bmatrix}. \tag{3.2}$$

Our strategy will be to reduce questions about the spectrum of the matrix $\boldsymbol{M}$ to questions about the spectrum of a matrix of the form $\boldsymbol{X} = \boldsymbol{E}\boldsymbol{U}\boldsymbol{F}\boldsymbol{U}^{\mathsf{H}}$, where $\boldsymbol{U}$ is a uniformly random unitary matrix, $\boldsymbol{E}$ is a random matrix independent of $\boldsymbol{U}$ and $\boldsymbol{F}$ is deterministic. This matrix model has been well studied in Free Probability [64]. The starting point of our reduction is Proposition 2 from Lu and

Li [19], stated below.

Proposition 3 (Lu and Li [19])**.** *Let $\boldsymbol{D}$ be an arbitrary deterministic symmetric matrix partitioned as:*

$$\boldsymbol{D} = \begin{bmatrix} a & \boldsymbol{q}^{\mathsf{H}} \\ \boldsymbol{q} & \boldsymbol{P} \end{bmatrix}.$$

Then, we have

$$\lambda_1(D) = L(\vartheta_\star),$$

where $L(\vartheta) = \lambda_1(\boldsymbol{P} + \vartheta \boldsymbol{q}\boldsymbol{q}^{\mathsf{H}})$, and $\vartheta_\star > 0$ is the unique solution to the fixed point equation $L(\vartheta) = \frac{1}{\vartheta} + a$. Furthermore, let $\boldsymbol{v}_1$ be the eigenvector corresponding to the largest eigenvalue of $\boldsymbol{D}$. Then,

$$|\boldsymbol{e}_1^{\mathsf{H}} \boldsymbol{v}_1|^2 \in \left[\frac{\partial_- L(\vartheta_\star)}{\partial_- L(\vartheta_\star) + (1/\vartheta_\star)^2}, \frac{\partial_+ L(\vartheta_\star)}{\partial_+ L(\vartheta_\star) + (1/\vartheta_\star)^2} \right],$$

where ∂_- and ∂_+ denote the left and right derivatives respectively. In particular, if $L(\vartheta)$ is differentiable at $\vartheta_\star$, then

$$|\boldsymbol{e}_1^{\mathsf{H}} \boldsymbol{v}_1|^2 = \frac{L'(\vartheta_\star)}{L'(\vartheta_\star) + (1/\vartheta_\star)^2}.$$

A straightforward corollary of the above proposition to our problem is given below. Define the function

$$L_m(\vartheta) \stackrel{\text{def}}{=} \lambda_1 \left(\boldsymbol{A}_{-1}^{\mathsf{H}} (\boldsymbol{T} + \vartheta \boldsymbol{T}\boldsymbol{A}_1 (\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}}) \boldsymbol{A}_{-1} \right).$$

Corollary 1. *Let $\vartheta_m > 0$ be the unique solution of $L_m(\vartheta) = 1/\vartheta + \boldsymbol{A}_1^{\mathsf{H}} \boldsymbol{T} \boldsymbol{A}_1$. Then, $\lambda_1(\boldsymbol{A}^{\mathsf{H}} \boldsymbol{T} \boldsymbol{A}) =$*

$L_m(\vartheta_m)$ *and*

$$|\boldsymbol{e}_1^{\mathsf{H}}\hat{\boldsymbol{x}}|^2 \in \left[\frac{\partial_- L_m(\vartheta_m)}{\partial_- L_m(\vartheta_m)+(1/\vartheta_m)^2}, \frac{\partial_+ L_m(\vartheta_m)}{\partial_+ L(\vartheta_m)+(1/\vartheta_m)^2}\right].$$

In particular, if $L_m(\vartheta)$ *is differentiable at* ϑ_m*, then*

$$|\boldsymbol{e}_1^{\mathsf{H}}\hat{\boldsymbol{x}}|^2 = \frac{L_m'(\vartheta_m)}{L_m'(\vartheta_m)+(1/\vartheta_m)^2}.$$

Hence, we shift our focus to characterizing the function $L_m(\vartheta)$. Recall the decomposition of the matrix $\boldsymbol{M}$ given in (3.2). Recall that since $\boldsymbol{x}_\star = \sqrt{m}\boldsymbol{e}_1$, the diagonal matrix $\boldsymbol{T}$ is a deterministic function of $\boldsymbol{A}_1$. If the sensing matrix $\boldsymbol{A}$ consisted of independent Gaussian entries, then $\boldsymbol{T}, \boldsymbol{A}_1$ would have been independent of $\boldsymbol{A}_{-1}$. This is no longer true when $\boldsymbol{A}$ is a partial unitary matrix. In order to take care of this, the following lemma leverages a conditioning trick to get rid of the dependence. The following lemma also establishes the link between the function $L_m(\vartheta)$ and the study of the spectrum of a matrix of the form $\boldsymbol{X} = \boldsymbol{E}\boldsymbol{U}\boldsymbol{F}\boldsymbol{U}^{\mathsf{H}}$, where $\boldsymbol{U}$ is a uniformly random unitary matrix, $\boldsymbol{E}$ is a random matrix independent of $\boldsymbol{U}$ and $\boldsymbol{F}$ is deterministic.

Lemma 2. *We have*

$$L_m(\vartheta) = \lambda_1\left(\boldsymbol{B}^{\mathsf{H}}(\boldsymbol{T}+\vartheta\boldsymbol{T}\boldsymbol{A}_1(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}})\boldsymbol{B}\boldsymbol{H}_{m-1}\boldsymbol{R}\boldsymbol{H}_{m-1}^{\mathsf{H}}\right), \tag{3.3}$$

where

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{I}_{n-1} & \boldsymbol{0}_{n-1,m-n} \\ \boldsymbol{0}_{m-n,n-1} & \boldsymbol{0}_{m-n,m-n} \end{bmatrix},$$

$\boldsymbol{B} \in \mathbb{C}^{m\times m-1}$ *is an arbitrary basis matrix for* $\boldsymbol{A}_1^{\perp}$*, which denotes the subspace orthogonal to* $\boldsymbol{A}_1$*, and* $\boldsymbol{H}_{m-1} \sim \mathrm{Unif}(\mathbb{U}(m-1))$ *is independent of* $\boldsymbol{A}_1$.

Proof. We condition on $\boldsymbol{A}_1$. Conditioned on $\boldsymbol{A}_1$, we can realize $\boldsymbol{A}_{-1}$ as:

$$\boldsymbol{A}_{-1} = \boldsymbol{B}\boldsymbol{H}_{m-1}\boldsymbol{S}_{m-1,n-1}.$$

In the above equation, $\boldsymbol{B} \in \mathbb{C}^{m\times m-1}$ is matrix whose columns form an orthonormal basis of the orthogonal complement of $\boldsymbol{A}_1$ and $\boldsymbol{H}_{m-1}$ is a Haar Unitary of size $m-1$ independent of $\boldsymbol{A}_1$. Hence, we obtain

$$\begin{aligned} L_m(\vartheta) &= \lambda_1\left(\boldsymbol{A}_{-1}^{\mathsf{H}}(\boldsymbol{T} + \vartheta\boldsymbol{T}\boldsymbol{A}_1(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}})\boldsymbol{A}_{-1}\right) \\ &\overset{\text{a}}{=} \lambda_1\left(\boldsymbol{B}^{\mathsf{H}}(\boldsymbol{T} + \vartheta\boldsymbol{T}\boldsymbol{A}_1(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}})\boldsymbol{B}\cdot\boldsymbol{H}_{m-1}\boldsymbol{R}\boldsymbol{H}_{m-1}^{\mathsf{H}}\right). \end{aligned}$$

In the step marked (a), We used the fact that for any two matrices $\boldsymbol{\Lambda}, \boldsymbol{\Gamma}$ (of appropriate dimensions), $\boldsymbol{\Lambda}\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}\boldsymbol{\Lambda}$ have the same non-zero eigenvalues. In particular, we used this fact with:

$$\begin{aligned} \boldsymbol{\Lambda} &= \boldsymbol{S}_{m-1,n-1}^{\mathsf{H}}\boldsymbol{H}_{m-1}^{\mathsf{H}} \\ \boldsymbol{\Gamma} &= \boldsymbol{B}^{\mathsf{H}}(\boldsymbol{T} + \vartheta\boldsymbol{T}\boldsymbol{A}_1(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}})\boldsymbol{B}\boldsymbol{H}_{m-1}\boldsymbol{S}_{m-1,n-1}. \end{aligned}$$

□

Define the matrix,

$$\boldsymbol{E}(\vartheta) \overset{\text{def}}{=} \boldsymbol{B}^{\mathsf{H}}(\boldsymbol{T} + \vartheta\boldsymbol{T}\boldsymbol{A}_1(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}})\boldsymbol{B}. \tag{3.4}$$

The following lemma characterizes the asymptotic limit of the function $L_m(\vartheta)$. Define $\Lambda_+(\tau)$ as

$$\Lambda_+(\tau) = \begin{cases} \tau - \frac{(1-1/\delta)}{\mathbb{E}\left[\frac{1}{\tau - T}\right]} & \text{if } \tau > \tau_r, \\ \min_{\tau\geq 1}\left(\tau - \frac{(1-1/\delta)}{\mathbb{E}\left[\frac{1}{\tau - T}\right]}\right) & \text{if } \tau \leq \tau_r, \end{cases}$$

where $T = \mathcal{T}(|Z|^2)$ and $Z \sim \mathcal{CN}(0,1)$, and

$$\tau_r \triangleq \arg\min_{\tau \geq 1} \left(\tau - \frac{(1-1/\delta)}{\mathbb{E}\left[\frac{1}{\tau - T}\right]} \right).$$

Lemma 3. *Let* $\vartheta_c \stackrel{def}{=} \left(1 - \left(\mathbb{E}\left[\frac{|Z|^2}{1-T}\right]\right)^{-1} - \mathbb{E}[|Z|^2 T]\right)^{-1}$. *Define the function* $\theta(\vartheta)$ *as:*

- *When* $\vartheta > \vartheta_c$*: Let* $\theta(\vartheta)$ *be the unique value of* λ *that satisfies the equation:*

$$\lambda - \mathbb{E}[|Z|^2 T] - 1/\vartheta = \left(\mathbb{E}\left[\frac{|Z|^2}{\lambda - T}\right]\right)^{-1},$$

  in the interval:

$$\lambda \in \left(\max(1, \mathbb{E}[|Z|^2 T] + 1/\vartheta), \infty\right).$$

- *When* $\vartheta \leq \vartheta_c$*:* $\theta(\vartheta) \stackrel{def}{=} 1$.

Then, we have $L_m(\vartheta) \stackrel{a.s.}{\to} \Lambda_+(\theta(\vartheta))$*, where* $L_m(\vartheta)$ *is defined in* (3.3)*.*

The proof of Lemma 3 can be found in Section 3.5.5.

From Corollary 1, we know that $\lambda_1(\boldsymbol{M})$ solves the fixed point equation (in ϑ): $L_m(\vartheta) = 1/\vartheta + \boldsymbol{A}_1^{\mathsf{H}} \boldsymbol{T} \boldsymbol{A}_1$. Simple concentration arguments (see Lemma 7, Section 3.5.3) show that asymptotically:

$$\boldsymbol{A}_1^{\mathsf{H}} \boldsymbol{T} \boldsymbol{A}_1 \approx \mathbb{E}|Z|^2 T.$$

Combining this with Lemma 3 suggests that asymptotically $\lambda_1(\boldsymbol{M})$ behaves like the solution to the following fixed point equation (in ϑ):

$$\Lambda_+(\theta(\vartheta)) = 1/\vartheta + \mathbb{E}|Z|^2 T.$$

The following lemma analyzes the behavior of this asymptotic fixed point equation. The proof of this lemma can be found in Section 3.5.5.

Lemma 4. *The following hold for the equation:*

$$\Lambda_+(\theta(\vartheta)) = 1/\vartheta + \mathbb{E}[|Z|^2 T],\ \vartheta > 0.$$

1. *This equation has a unique solution.*

2. *Let $\vartheta_\star$ denote the solution of the above equation. Then:*

Case 1 *If $\psi_1(\tau_r) \leq \frac{\delta}{\delta-1}$, we have*

$$\Lambda_+(\theta(\vartheta_\star)) = \Lambda(\tau_r).$$

Furthermore,if the inequality is strict that is, $\psi_1(\tau_r) < \delta/(\delta-1)$ then,

$$\left.\frac{\mathrm{d}\Lambda_+(\theta(\vartheta))}{\mathrm{d}\vartheta}\right|_{\vartheta=\vartheta_\star} = 0,$$

Case 2 *If $\psi_1(\tau_r) > \frac{\delta}{\delta-1}$, we have*

$$\Lambda_+(\theta(\vartheta_\star)) = \Lambda(\theta_\star),$$

and,

$$\left.\frac{\mathrm{d}\Lambda_+(\theta(\vartheta))}{\mathrm{d}\vartheta}\right|_{\vartheta=\vartheta_\star} = \frac{1}{\vartheta_\star^2}\frac{\delta}{\delta-1} \cdot \left(\frac{\delta}{\delta-1} - \psi_2(\theta_\star)\right) \cdot \frac{1}{\psi_3^2(\theta_\star) - \frac{\delta^2}{(\delta-1)^2}}.$$

where $\theta_\star > 1$ is the unique $\theta \geq \tau_r$ that satisfies $\psi_1(\theta) = \frac{\delta}{\delta-1}$.

We are now in the position to prove our main result (restated below for convenience). Recall the definitions of the functions $\Lambda(\tau), \psi_1(\tau), \psi_2(\tau), \psi_3(\tau)$ from (3.1).

Theorem 1 *Define* $\tau_r \triangleq \arg\min_{\tau\in[1,\infty)} \Lambda(\tau)$. *Also, let* $\theta_\star$ *denote the unique value of* $\theta > \tau_r$ *that satisfies* $\psi_1(\theta) = \frac{\delta}{\delta-1}$. *Then, we have*

$$\lambda_1(\boldsymbol{M}) \xrightarrow{a.s.} \begin{cases} \Lambda(\tau_r), & \text{if } \psi_1(\tau_r) \leq \frac{\delta}{\delta-1}, \\ \Lambda(\theta_\star), & \text{if } \psi_1(\tau_r) > \frac{\delta}{\delta-1}. \end{cases}$$

Furthermore,

$$|\boldsymbol{e}_1^{\mathsf{H}}\hat{\boldsymbol{x}}|^2 \xrightarrow{a.s.} \begin{cases} 0, & \text{if } \psi_1(\tau_r) < \frac{\delta}{\delta-1}, \\ \frac{\left(\frac{\delta}{\delta-1}\right)^2 - \frac{\delta}{\delta-1}\cdot\psi_2(\theta_\star)}{\psi_3(\theta_\star)^2 - \frac{\delta}{\delta-1}\cdot\psi_2(\theta_\star)}, & \text{if } \psi_1(\tau_r) > \frac{\delta}{\delta-1}. \end{cases}$$

Proof. We start with the analysis of the largest eigenvalue. We recall the claim of Corollary 1, which tells us that $\lambda_1(\boldsymbol{M})$ is given by $L_m(\vartheta_m)$ where ϑ_m denotes the solution of $L_m(\vartheta) = 1/\vartheta + a_m$ and $a_m = \boldsymbol{A}_1^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}_1$.

We also know that there exists a probability 1 event $\mathcal{E}$, on which, $L_m(\vartheta) \xrightarrow{\text{a.s.}} \Lambda_+(\theta(\vartheta))$ (Lemma 3) and $a_m \xrightarrow{\text{a.s.}} \mathbb{E}[|Z|^2 T]$ (see Lemma 7 in Section 3.5.3).

We claim that on $\mathcal{E}$, $\vartheta_m \to \vartheta_\star$, where $\vartheta_\star$ is the solution of the limiting fixed point equation $\Lambda_+(\theta(\vartheta)) = 1/\vartheta + \mathbb{E}[|Z|^2 T]$ (which was analyzed in Lemma 4). To see this let $\overline{\vartheta} = \limsup \vartheta_m$. Consider a subsequence $\vartheta_{m_k} \to \overline{\vartheta}$. Then applying Lemma 3 (in Appendix E) of Lu and Li [19], we obtain,

$$\begin{aligned} 0 &= \lim_{k\to\infty} \left(L_{m_k}(\vartheta_{m_k}) - \frac{1}{\vartheta_{m_k}} - a_{m_k} \right) \\ &= \Lambda_+(\theta(\overline{\vartheta})) - \frac{1}{\overline{\vartheta}} - \mathbb{E}|Z|^2 T. \end{aligned}$$

That is, $\overline{\vartheta}$ is also a solution to the limiting fixed point equation $\Lambda_+(\theta(\vartheta)) = 1/\vartheta + \mathbb{E}[|Z|^2 T]$. But since this equation has a unique solution (Lemma 4), we have $\limsup \vartheta_m = \overline{\vartheta} = \vartheta_\star$. Likewise, an

analogous argument shows $\liminf \vartheta_m = \vartheta_\star$.

Now for any realization in the event $\mathcal{E}$, we have,

$$\lambda_1(\boldsymbol{M}) = L_m(\vartheta_m) \overset{\text{(a)}}{\to} \Lambda_+(\theta(\vartheta_\star)).$$

In the above display, in the step marked (a), we again appealed to Lemma 3 (Appendix E) of Lu and Li [19] and the fact that $\vartheta_m \to \vartheta_\star$. Finally, appealing to the alternative characterization of $\Lambda_+(\theta(\vartheta_\star))$ given in Lemma 4 gives us the claim of the theorem.

We now discuss our result about the cosine similarity. We recall that from Corollary 1, we have

$$|\boldsymbol{e}_1^{\mathsf{H}} \hat{\boldsymbol{x}}|^2 \in \left[\frac{\partial_- L_m(\vartheta_m)}{\partial_- L_m(\vartheta_m) + (1/\vartheta_m)^2}, \frac{\partial_+ L_m(\vartheta_n)}{\partial_+ L(\vartheta_m) + (1/\vartheta_m)^2} \right].$$

Appealing to Lemma 4 in Appendix E of Lu and Li [19], we have,

$$\partial_- L_m(\vartheta_m) \to \partial_- \Lambda_+(\theta(\vartheta_\star)),\ \partial_+ L_m(\vartheta_m) \to \partial_+ \Lambda_+(\theta(\vartheta_\star)).$$

The derivative of $\Lambda_+(\theta(\vartheta))$ at $\vartheta = \vartheta_\star$ was calculated in Lemma 4. Plugging this in the above expression gives the statement of the theorem. □

The remainder of this section is dedicated to the proof of Lemmas 3 and 4, and is organized as follows:

- Recall that (cf. 3.3)

  $$L_m(\vartheta) = \lambda_1 \left(\boldsymbol{E}(\vartheta) \boldsymbol{H}_{m-1} \boldsymbol{R} \boldsymbol{H}_{m-1}^{\mathsf{H}} \right),$$

  where

  $$\boldsymbol{E}(\vartheta) \overset{\text{def}}{=} \boldsymbol{B}^{\mathsf{H}} (\boldsymbol{T} + \vartheta \boldsymbol{T} \boldsymbol{A}_1 (\boldsymbol{T} \boldsymbol{A}_1)^{\mathsf{H}}) \boldsymbol{B}.$$

  Note that $\boldsymbol{E}(\vartheta)$ is independent of $\boldsymbol{H}_{m-1}$. The spectrum of such a matrix product has been studied in free probability theory, and we collect some results regarding this in Section 3.5.2.

- In order to apply the free probability results, we need to understand the spectrum of $\boldsymbol{E}(\vartheta)$. This is done in Section 3.5.3.

- It turns out that the limiting spectrum measure of $\boldsymbol{E}(\vartheta)\boldsymbol{H}_{m-1}\boldsymbol{R}\boldsymbol{H}_{m-1}^{\mathsf{H}}$ is given by the free convolution (defined in Section 3.5.2) of the measures γ and $\mathcal{L}_T$, where $\gamma \stackrel{\text{def}}{=} \frac{1}{\delta}\delta_1 + \left(1 - \frac{1}{\delta}\right)\delta_0$ and $\mathcal{L}_T$ is the law of the random variable $T = \mathcal{T}(|Z|/\sqrt{\delta})$. Section 3.5.4 is devoted to understanding the support of the free convolution.

- Finally, Section 3.5.5 proves lemmas 3 and 4.

3.5.2 Free Probability Background

Our analysis of the spectral estimators relies on a well-studied model in the theory of free probability; We will reduce the problem to the problem of understanding the spectrum of matrices of the form $\boldsymbol{X} = \boldsymbol{E}\boldsymbol{U}\boldsymbol{F}\boldsymbol{U}^{\mathsf{H}}$, where $\boldsymbol{E}$ and $\boldsymbol{F}$ are deterministic matrices and $\boldsymbol{U}$ is a Haar-distributed unitary matrix. Then, the limiting spectral distribution of $\boldsymbol{X}$ is the free multiplicative convolution of the limiting spectral distributions of $\boldsymbol{E}$ and $\boldsymbol{F}$. This section is a collection of the results and definitions regarding these aspects. Here is the organization of this section. Section 3.5.2 collects various facts from free harmonic analysis. Section 3.5.2 describes the two fundamental results about the model $\boldsymbol{X} = \boldsymbol{E}\boldsymbol{U}\boldsymbol{F}\boldsymbol{U}^{\mathsf{H}}$ that will be useful for our analysis. Section 3.5.2 reviews some results about the support of singular part of the free convolution of two measures. Throughout this section, we assume that γ and ν are two arbitrary compactly supported probability measures on $[0, \infty)$ and that neither of the two measures is completely concentrated at a single point.

Facts from Free Harmonic Analysis

In this section, we collect some facts from the field of free harmonic analysis. All these results can be found in Chapter 3 of Mingo and Speicher [49] or the papers by Belinschi, Bercovici, Capitaine, and Fevrier [64] and Belinschi [65].

Definition 1. *The Cauchy transform G_γ of γ at z is defined as follows:*

$$G_\gamma(z) = \int \frac{\gamma(\mathrm{d}t)}{z-t},\ z \in \mathbb{C}\backslash[0,\infty).$$

Definition 2. *The moment generating function of γ, ψ_γ at z is defined as follows:*

$$\psi_\gamma(z) = \int \frac{zt}{1-zt}\gamma(\mathrm{d}t),\ z \in \mathbb{C}\backslash[0,\infty).$$

The Cauchy transform and the moment generating function are related via the relation

$$G_\gamma(z) = \frac{1}{z} \cdot \left(\psi_\gamma\left(\frac{1}{z}\right) + 1\right).$$

Definition 3. *The η-transform of a measure is defined as,*

$$\eta_\gamma(z) = \frac{\psi_\gamma(z)}{1+\psi_\gamma(z)}.$$

The Cauchy transform (and hence the Moment Generating function) uniquely characterizes a measure. The measure can be obtained by the following inversion formula. The particular version we state is taken from Section 3.1 of Belinschi, Bercovici, Capitaine, and Fevrier [64].

Theorem 2. *For $a < b \in [0,\infty)$, we have*

$$\gamma((a,b)) + \frac{1}{2}\gamma(\{a,b\}) = \frac{1}{\pi}\lim_{\epsilon\to 0^+}\int_a^b \mathrm{Im}(G_\gamma(x-i\epsilon))\,\mathrm{d}x.$$

Furthermore, if γ satisfies $\gamma = \gamma_{ac} + \gamma_s$, where γ_{ac} and γ_s denote the absolutely continuous and the singular part of the measure with respect to the Lebesgue measure, then the density of the

absolutely continuous part is given by

$$\frac{\mathrm{d}\gamma_{ac}}{\mathrm{d}x}(x) = \lim_{\epsilon \to 0^+} \frac{1}{\pi} \mathrm{Im}(G_\gamma(x - i\epsilon)).$$

Next we recall the definition of the free convolution based on the subordination functions from Belinschi and Bercovici [66]. The statement we provide below appears in a more general form as Proposition 2.6 in Belinschi, Speicher, Treilhard, and Vargas [67].

Definition 4. *Let (γ, ν) be a pair of probability measures. There exist analytic functions w_γ, w_ν defined on $\mathbb{C}\backslash[0, \infty)$ such that, for all $z \in \mathbb{C}^+$ we have*

1. *$w_\gamma(z), w_\nu(z) \in \mathbb{C}^+$; $w_\gamma(\overline{z}) = \overline{w_\gamma(z)}, w_\nu(\overline{z}) = \overline{w_\nu(z)}$ and $\mathsf{Arg}(w_\gamma(z)) \geq \mathsf{Arg}(z), \mathsf{Arg}(w_\nu(z)) \geq \mathsf{Arg}(z)$.*

2. *For any $z \in \mathbb{C}^+$, $w_\nu(z)$ is the unique solution in $\mathbb{C}^+$ of the fixed point equation $Q_z(w) = w$, where Q_z is given by*

$$Q_z(w) = \frac{w}{\eta_\nu(w)} \eta_\gamma \left(\frac{z\eta_\nu(w)}{w} \right).$$

   An analogous characterization holds for w_γ with the role of γ and ν changed.

The free convolution of the measures γ and ν denoted by $\gamma \boxtimes \nu$ is the measure whose moment generating function satisfies

$$\psi_{\gamma \boxtimes \nu}(z) = \psi_\gamma(w_\gamma(z)) = \psi_\nu(w_\nu(z)) = \frac{w_\gamma(z) w_\nu(z)}{z - w_\gamma(z) w_\nu(z)}.$$

Remark 8. *We emphasize that each of the subordination functions w_γ, w_ν depend on both the measures γ, ν. This is clear since the function $Q_z(w)$ defining w_ν depends on both ν, γ.*

Note that the above definition defines w_ν and w_γ on $\mathbb{C}\backslash[0,\infty)$. However these functions can be continously extended to $\overline{\mathbb{C}^+} \cup \{\infty\}$ (Lemma 3.2 in [64]). These extensions to the real line will be important for Theorem 3.5.2.

Lemma 5. *The restrictions of subordination functions w_γ, w_ν on $\mathbb{C}^+$ have extensions to $\overline{\mathbb{C}^+} \cup \{\infty\}$ with the following properties:*

1. *$w_\gamma, w_\nu : \overline{\mathbb{C}^+} \cup \{\infty\} \to \overline{\mathbb{C}^+} \cup \{\infty\}$ are continuous.*

2. *If $1/x \in [0,\infty)\backslash Supp(\gamma \boxtimes \nu)$, then the functions w_γ, w_ν continue analytically to a neighborhood of x and*

$$\frac{1}{w_\gamma(x)} = \frac{w_\nu(x)}{x} \cdot \frac{1+\psi_\nu(w_\nu(x))}{\psi_\nu(w_\nu(x))} \in \mathbb{R}\backslash Supp(\gamma),$$
$$\frac{1}{w_\nu(x)} = \frac{w_\gamma(x)}{x} \cdot \frac{1+\psi_\gamma(w_\gamma(x))}{\psi_\gamma(w_\gamma(x))} \in \mathbb{R}\backslash Supp(\nu).$$

Spectrum of $\mathbf{X} = \mathbf{EUFU}^{\mathsf{H}}$

As we discussed before, we will convert the problem of analyzing the spectrum of $\boldsymbol{M}$ to problems involving the spectrum of matrices of the form $\mathbf{X}_N = \mathbf{E}_N\mathbf{U}_N\mathbf{F}_N\mathbf{U}_N^{\mathsf{H}}$, where $\boldsymbol{U}_N$ is a sequence of Haar distributed $N \times N$ random matrices, and $\boldsymbol{E}_N$ and $\boldsymbol{F}_N$ are sequences of deterministic positive semidefinite matrices. In this section, we review two important results from the field of free probability regarding such matrices.

Suppose that $\boldsymbol{E}_N$ and $\boldsymbol{F}_N$ satisfy the following hypotheses:

(i) $\mu_{\boldsymbol{E}_N} \xrightarrow{\text{d}} \mu_e$ and $\mu_{\boldsymbol{F}_N} \xrightarrow{\text{d}} \mu_f$, where μ_e, μ_f are compactly supported measures on $[0,\infty)$.

(ii) $\boldsymbol{E}_N$ has a single outlying eigenvalue θ not contained in Supp(μ_e). $\boldsymbol{F}_N$ has no eigenvalues outside Supp(μ_f).

(iii) The set of eigenvalues of $\boldsymbol{E}_N$ not equal to θ converge uniformly to $\text{Supp}(\mu_e)$ in the sense,

$$\lim_{N\to\infty} \max_{i:\lambda_i(\boldsymbol{E}_N)\neq\theta} \text{dist}(\lambda_i(\boldsymbol{E}_N), \text{Supp}(\mu_e)) = 0.$$

Our next theorem characterizes the bulk distribution of $\boldsymbol{X}_N$. The first part of this theorem is due to Voiculescu [68] and the second and third parts are due to Belinschi, Bercovici, Capitaine, and Fevrier [64] (Theorem 2.3).

Theorem 3. *Let w_e and w_f denote the subordination functions for the free multiplicative convolution of μ_e and μ_f. Define*

$$\tau_e(1/z) = \frac{1}{w_e(1/z)},\ K = \textit{Supp}(\mu_e \boxtimes \nu_f) \cup \tau_e^{-1}(\theta).$$

Then we have, almost surely for large enough N,

1. $\mu_{\boldsymbol{X}_N} \xrightarrow{d} \mu_e \boxtimes \mu_f$.
2. *Given $\epsilon > 0$, we have $\sigma(\boldsymbol{X}_N) \subset K_\epsilon$, where K_ϵ is the ϵ-neighborhood of K and $\sigma(\boldsymbol{X}_N)$ denotes the set of eigenvalues of $\boldsymbol{X}_N$.*
3. *For any $\rho \in \tau_e^{-1}(\theta)$ such that $\exists \epsilon > 0$ with $(\rho - 2\epsilon, \rho + 2\epsilon) \cap K = \{\rho\}$, we have $|\sigma(\boldsymbol{X}_N) \cap (\rho - \epsilon, \rho + \epsilon)| = 1$.*

Remark 9. *The hypothesis in the above theorem can be relaxed (as mentioned in Remark 5.11 of [64]) in the following two ways: 1) $\boldsymbol{E}_N$ is random, independent of $\boldsymbol{U}_N$ and $\boldsymbol{F}_N$ is deterministic, provided $\mu_{\boldsymbol{E}_N} \xrightarrow{d} \mu_e$ occurs almost surely, 2) The spike locations depend on N, θ_N provided $\theta_N \to \theta$ almost surely.*

Remark 10. *The above theorem is a simplified version of Theorem 2.3 in [64] which allows for multiple spikes in both $\boldsymbol{E}_N$ and $\boldsymbol{F}_N$.*

Remark 11. *The function τ might not be invertible. In such cases, $\tau^{-1}(\theta)$ can be a non-singleton set, and hence a single spike in $\boldsymbol{E}_N$ can create multiple spikes in $\boldsymbol{X}_N$. But we will see that this doesn't happen in our problem.*

Singular Part of Free Convolution

In the last section we discussed the bulk distribution of $\boldsymbol{X}_N = \boldsymbol{E}_N\boldsymbol{U}_N\boldsymbol{F}_N\boldsymbol{U}_N$. The main objective of this section is to mention a result regarding the largest eigenvalue of $\boldsymbol{X}_N$. We state regularity results for the singular part of $\gamma \boxtimes \nu$ from Belinschi [69] (Corollary 3.4) and Belinschi [65] (Theorem 4.1).

Theorem 4 (Singular Part of $\gamma\boxtimes\nu$)**.** *Decompose the singular part of $\gamma\boxtimes\nu$ as $(\gamma\boxtimes\nu)_s = (\gamma\boxtimes\nu)_d + (\gamma \boxtimes \nu)_{sc}$ where $(\gamma \boxtimes \nu)_d$ denotes the discrete part and $(\gamma \boxtimes \nu)_{sc}$ denotes the singular continous part. Then we have,*

1. *There can be at most two atoms. The possible locations of the atoms are:*
   - *(a) 0, with $\gamma \boxtimes \nu(\{0\}) = \max(\gamma(\{0\}), \nu(\{0\}))$.*
   - *(b) Any $a \in (0,\infty)$ such that there exist $u, v \in (0,\infty)$ with $uv = a$ and $\gamma(\{u\})+\nu(\{v\}) > 1$ and we have, $\gamma \boxtimes \nu(\{a\}) = \gamma(\{u\}) + \nu(\{v\}) - 1$. Note that there can be atmost one such a.*
2. *Suppose neither of γ, ν is completely concentrated at a single point. We have, Supp$((\gamma \boxtimes \nu)_{sc}) \subset$ Supp$((\gamma \boxtimes \nu)_{ac})$. Hence,*

$$\textit{Supp}(\gamma \boxtimes \nu) = \textit{Supp}((\gamma \boxtimes \nu)_{ac}) \cup \textit{Supp}((\gamma \boxtimes \nu)_d).$$

3.5.3 Analysis of the Spectrum of $\mathbf{E}(\vartheta)$

In order to apply Theorem 3, we need to understand the spectrum of $\boldsymbol{B}^{\mathsf{H}}(\boldsymbol{T}+\vartheta\boldsymbol{T}\boldsymbol{A}_1(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}})\boldsymbol{B}$. This is done in the following lemma.

Lemma 6. *Let*

$$T_{(1)} \geq T_{(2)} \cdots \geq T_{(m)}$$

denote the sorted trimmed measurements. Let $\boldsymbol{E}(\vartheta) \stackrel{def}{=} \boldsymbol{B}^{\mathsf{H}}(\boldsymbol{T} + \vartheta \boldsymbol{T}\boldsymbol{A}_1(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}})\boldsymbol{B}$*. Then,*

1. *The eigenvalues of* $\boldsymbol{E}(\vartheta)$ *interlace with* $T_{(1)}, T_{(2)} \ldots T_{(m)}$ *in the sense,*

$$\lambda_i(\boldsymbol{E}(\vartheta)) \leq T_{(i-1)} \; \forall \; i = 2, 3, \ldots m, \; \&$$
$$\lambda_i(\boldsymbol{E}(\vartheta)) \geq T_{(i+1)} \; \forall \; i = 1, 3, \ldots m-1.$$

2. $\boldsymbol{E}(\vartheta)$ *can have at most one eigenvalue bigger than* $T_{(1)}$*, which (if it exists) is given by the root of the following equation:*

$$Q_m(\lambda) = \frac{1}{\lambda - a_m - 1/\vartheta}, \; \lambda > \max(a_m + 1/\vartheta, T_{(1)}),$$

where $Q_m(\lambda)$ *is defined as*

$$Q_m(\lambda) \stackrel{def}{=} \sum_{i=1}^{m} \frac{|A_{1i}|^2}{\lambda - T_i}.$$

3. *Furthermore,* $\lambda_1(\boldsymbol{E}(\vartheta)) \leq 1 + \vartheta$ *and* $\lambda_{m-1}(\boldsymbol{E}(\vartheta)) \geq 0$*.*

Proof. Define the matrix $\boldsymbol{E}(\vartheta) = \boldsymbol{B}^{\mathsf{H}}(\boldsymbol{T} + \vartheta \boldsymbol{T}\boldsymbol{A}_1(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}})\boldsymbol{B}$. The main trick will be to choose the orthonormal basis matrix $\boldsymbol{B}$ conveniently, which will make our calculations easier. Recall that the columns of matrix $\boldsymbol{B}$, i.e. $\boldsymbol{B}_1, \boldsymbol{B}_2 \ldots \boldsymbol{B}_{m-1}$, span the subspace $\boldsymbol{A}_1^{\perp}$. Any basis for subspace $\boldsymbol{A}_1^{\perp}$ can serve as matrix $\boldsymbol{B}$. Hence, we chose the following specific construction of $\boldsymbol{B}$:

$$\boldsymbol{B}_1 = \frac{\boldsymbol{T}\boldsymbol{A}_1 - a_m \boldsymbol{A}_1}{\sqrt{b_m - a_m^2}},$$

where $a_m = \boldsymbol{A}_1^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}_1$ and $b_m = \boldsymbol{A}_1^{\mathsf{H}}\boldsymbol{T}^2\boldsymbol{A}_1$. With this choice, we note that

$$\begin{aligned}\boldsymbol{B}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}_1 &= [\boldsymbol{B}_1^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}_1, \boldsymbol{B}_2^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}_1 \ldots \boldsymbol{B}_{m-1}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}_1]^{\mathsf{H}} \\ &= \sqrt{b_m - a_m^2}\boldsymbol{e}_1.\end{aligned}$$

Hence $\boldsymbol{E}(\vartheta) = \boldsymbol{B}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{B} + \vartheta(b_m - a_m^2)\boldsymbol{e}_1\boldsymbol{e}_1^{\mathsf{H}}$. To obtain the eigenvalues of $\boldsymbol{E}(\vartheta)$ we use its characteristic polynomial. To evaluate the characteristic polynomial of $\boldsymbol{E}(\vartheta)$, we connect it to the characteristic polynomial of $\boldsymbol{O}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{O}$, where $\boldsymbol{O} = [\boldsymbol{A}_1, \boldsymbol{B}]$. Note that $\boldsymbol{O}$ is a unitary matrix. First, we have

$$\begin{aligned}\boldsymbol{O}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{O} &= \begin{bmatrix}\boldsymbol{A}_1^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}_1 & \boldsymbol{A}_1^{\mathsf{H}}\boldsymbol{T}\boldsymbol{B} \\ \boldsymbol{B}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{A}_1 & \boldsymbol{B}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{B}\end{bmatrix} \\ &= \begin{bmatrix} a_m & \sqrt{b_m - a_m^2}\boldsymbol{e}_1^{\mathsf{H}} \\ \sqrt{b_m - a_m^2}\boldsymbol{e}_1 & \boldsymbol{B}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{B}\end{bmatrix}.\end{aligned}$$

Consider the following matrix equation:

$$\begin{aligned}&\begin{bmatrix} a_m + \frac{1}{\vartheta} & \boldsymbol{0}^{\mathsf{H}} \\ \boldsymbol{0} & \boldsymbol{E}(\vartheta)\end{bmatrix} = \begin{bmatrix} a_m + \frac{1}{\vartheta} & \boldsymbol{0}^{\mathsf{H}} \\ \boldsymbol{0} & \boldsymbol{B}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{B}\end{bmatrix} + \vartheta(b_m - a_m^2)\boldsymbol{e}_2\boldsymbol{e}_2^{\mathsf{H}} \\ &= \begin{bmatrix} a_m & \sqrt{b_m - a_m^2}\boldsymbol{e}_1^{\mathsf{H}} \\ \sqrt{b_m - a_m^2}\boldsymbol{e}_1 & \boldsymbol{B}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{B}\end{bmatrix} + \begin{bmatrix} 1/\vartheta & -\sqrt{b_m - a_m^2} & \boldsymbol{0}_{m-2,1}^{\mathsf{H}} \\ -\sqrt{b_m - a_m^2} & \vartheta(b_m - a_m^2) & \boldsymbol{0}_{m-2,1}^{\mathsf{H}} \\ \boldsymbol{0}_{m-2,1} & \boldsymbol{0}_{m-2,1} & \boldsymbol{0}_{m-2,m-2}\end{bmatrix} \\ &= \boldsymbol{O}^{\mathsf{H}}\boldsymbol{T}\boldsymbol{O} + \begin{bmatrix} 1/\sqrt{\vartheta} \\ -\sqrt{\vartheta(b_m - a_m^2)} \\ \boldsymbol{0}_{m-2,1}\end{bmatrix}\begin{bmatrix} 1/\sqrt{\vartheta} \\ -\sqrt{\vartheta(b_m - a_m^2)} \\ \boldsymbol{0}_{m-2,1}\end{bmatrix}^{\mathsf{H}} \\ &= \boldsymbol{O}^{\mathsf{H}}(\boldsymbol{T} + \boldsymbol{u}\boldsymbol{u}^{\mathsf{H}})\boldsymbol{O}, \qquad (3.5)\end{aligned}$$

where

$$\boldsymbol{u} = \boldsymbol{O} \cdot \begin{bmatrix} 1/\sqrt{\vartheta} \\ -\sqrt{\vartheta(b_m - a_m^2)} \\ \mathbf{0}_{m-2,1} \end{bmatrix} = \frac{1}{\sqrt{\vartheta}} \boldsymbol{A}_1 - \sqrt{\vartheta(b_m - a_m^2)} \boldsymbol{B}_1$$
$$= \left(\frac{1}{\sqrt{\vartheta}} + a_m \sqrt{\vartheta} \right) \boldsymbol{A}_1 - \sqrt{\vartheta} \boldsymbol{T} \boldsymbol{A}_1$$

Therefore,

$$|u_i|^2 = \frac{(1 + a_m\vartheta - \vartheta T_i)^2 |A_{1i}|^2}{\vartheta}.$$

Now, we can compute the characteristic polynomial of $\boldsymbol{E}(\vartheta)$. We have

$$\begin{aligned} &\det(\lambda \boldsymbol{I} - \boldsymbol{E}(\vartheta)) \\ &\quad = \frac{1}{\lambda - a_m - \frac{1}{\vartheta}} \det \left(\lambda \boldsymbol{I} - \begin{bmatrix} a_m + \frac{1}{\vartheta} & \mathbf{0}^{\mathsf{H}} \\ \mathbf{0} & \boldsymbol{E}(\vartheta) \end{bmatrix} \right) \\ &\quad = \frac{1}{\lambda - a_m - 1/\vartheta} \cdot \det(\lambda \boldsymbol{I} - \boldsymbol{T} - \boldsymbol{u}\boldsymbol{u}^{\mathsf{H}}) \\ &\quad = \frac{\det(\lambda I - \boldsymbol{T})}{\lambda - a_m - 1/\vartheta} \cdot (1 - \boldsymbol{u}^{\mathsf{H}} (\lambda \boldsymbol{I} - \boldsymbol{T})^{-1} \boldsymbol{u}). \end{aligned}$$

Note that

$$\begin{aligned} &1 - \boldsymbol{u}^{\mathsf{H}} (\lambda \boldsymbol{I} - \boldsymbol{T})^{-1} \boldsymbol{u} = 1 - \sum_{i=1}^{m} \frac{|u_i|^2}{\lambda - T_i} \\ &= 1 - \frac{1}{\vartheta} \sum_{i=1}^{m} \frac{(1 + a_m\vartheta - \lambda\vartheta + (\lambda - T_i)\vartheta)^2 |A_{1i}|^2}{\lambda - T_i} \\ &= 1 - \frac{(1 + a_m\vartheta - \lambda\vartheta)^2}{\vartheta} \cdot \left(\sum_{i=1}^{m} \frac{|A_{1i}|^2}{\lambda - T_i} \right) - \vartheta \cdot \left(\sum_{i=1}^{m} (\lambda - T_i) \cdot |A_{1i}|^2 \right) - 2(1 + a_m\vartheta - \lambda\vartheta). \end{aligned}$$

Defining $Q_m(\lambda)$ in the following way:

$$Q_m(\lambda) \stackrel{\text{def}}{=} \sum_{i=1}^{m} \frac{|A_{1i}|^2}{\lambda - T_i},$$

we obtain,

$$\det(\lambda \boldsymbol{I} - \boldsymbol{E}(\vartheta)) = \det(\lambda \boldsymbol{I} - \boldsymbol{T})(\vartheta + (1 - \lambda\vartheta + a_m\vartheta)Q_m(\lambda)). \tag{3.6}$$

We emphasize that the above equation does not imply that $T_1, T_2, \ldots, T_m$ are the eigenvalues of $\boldsymbol{E}(\vartheta)$. This is because while $\det(\lambda \boldsymbol{I} - \boldsymbol{T})$ has zeros at T_i, the function $Q_m(\lambda)$ has poles at T_i. This prevents us from concluding that $\det(\lambda \boldsymbol{I} - \boldsymbol{E}(\vartheta)) = 0$ when $\lambda = T_i$. However, we can make the following observations:

1. By Cauchy's interlacing theorem, we have

$$\begin{aligned} \lambda_1(\boldsymbol{T} + \vartheta(\boldsymbol{T}\boldsymbol{A}_1)(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}}) &\geq T_{(1)} \\ &\geq \lambda_2(\boldsymbol{T} + \vartheta(\boldsymbol{T}\boldsymbol{A}_1)(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}}) \\ &\geq T_{(2)}. \end{aligned} \tag{3.7}$$

   The above is also true for the eigenvalues of:

$$\boldsymbol{O}^{\mathsf{H}}(\boldsymbol{T} + \vartheta(\boldsymbol{T}\boldsymbol{A}_1)(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}})\boldsymbol{O},$$

   since $\boldsymbol{O}$ is a unitary matrix.

2. (3.5) shows that $\boldsymbol{E}(\vartheta)$ is a principal submatrix of

$$\boldsymbol{O}^{\mathsf{H}}(\boldsymbol{T} + \vartheta(\boldsymbol{T}\boldsymbol{A}_1)(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}})\boldsymbol{O}.$$

Hence, the eigenvalues of $\boldsymbol{E}(\vartheta)$ will interlace the eigenvalues of $\boldsymbol{O}^{\mathsf{H}}(\boldsymbol{T}+\vartheta(\boldsymbol{T}\boldsymbol{A}_1)(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}})\boldsymbol{O}$:

$$\begin{aligned}\lambda_1(\boldsymbol{T}+\vartheta(\boldsymbol{T}\boldsymbol{A}_1)(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}} &\geq \lambda_1(\boldsymbol{E}(\vartheta)) \\ &\geq \lambda_2(\boldsymbol{T}+\vartheta(\boldsymbol{T}\boldsymbol{A}_1)(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}} \\ &\geq \lambda_2(\boldsymbol{E}(\vartheta)). \end{aligned} \tag{3.8}$$

Combining (3.7) and (3.8), one obtains

$$\lambda_2(\boldsymbol{E}(\vartheta)) \leq T_{(1)},\ \lambda_1(\boldsymbol{E}(\vartheta)) \geq T_{(2)}.$$

This proves statement (1) in the lemma. This means that $\boldsymbol{E}(\vartheta)$ has atmost one eigenvalue bigger than $T_{(1)}$. If $\lambda_1(\boldsymbol{E}(\vartheta)) \leq T_{(1)}$, then it has no outlying eigenvalue, if $\lambda_1(\boldsymbol{E}(\vartheta)) > T_{(1)}$, it has exactly one. We call this eigenvalue an outlying eigenvalue for reasons that will be clear later.

3. The outlying eigenvalue of $\boldsymbol{E}(\vartheta)$ (if it exists) is a root of the characteristic polynomial:

$$\begin{aligned}&\det(\lambda\boldsymbol{I}-\boldsymbol{E}(\vartheta)) = \\ &\qquad \det(\lambda I-\boldsymbol{T})\cdot(\vartheta+(1-\lambda\vartheta+a_m\vartheta)Q_m(\lambda)).\end{aligned}$$

Since this root lies in $(T_{(1)},\infty)$, it must be a root of:

$$Q_m(\lambda) = \frac{1}{\lambda - a_m - 1/\vartheta},\ \lambda > T_{(1)}. \tag{3.9}$$

Observing that:

$$\begin{aligned}\lambda > T_{(1)} &\implies Q_m(\lambda) > 0, \\ \lambda > a_m + 1/\vartheta &\implies (\lambda - a_m - 1/\vartheta)^{-1} > 0,\end{aligned}$$

we conclude the outlying eigenvalue is the unique solution (if it exists) to:

$$Q_m(\lambda) = \frac{1}{\lambda - a_m - 1/\vartheta},\ \lambda > \max(a_m + 1/\vartheta, T_{(1)}).$$

This proves statement (2).

4. Finally, we observe that $\boldsymbol{E}(\vartheta)$ is a positive semidefinite matrix for all $\vartheta \geq 0$, which shows $\lambda_{m-1}(\boldsymbol{E}(\vartheta)) \geq 0$. Also, we have $\lambda_1(\boldsymbol{E}(\vartheta)) \leq \|\boldsymbol{E}(\vartheta)\| \leq \|\boldsymbol{B}\|^2\|\boldsymbol{T} + \vartheta \boldsymbol{T}\boldsymbol{A}_1(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}}\|$. Note that $\|\boldsymbol{B}\| \leq 1$ and $\|\boldsymbol{T}\| \leq 1$ and $\|\boldsymbol{T}\boldsymbol{A}_1(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}}\| = \boldsymbol{A}_1^{\mathsf{H}}\boldsymbol{T}^2\boldsymbol{A}_1 \leq T_{(1)}^2 \leq 1$. Hence, by the triangle inequality we have $\lambda_1(\boldsymbol{E}(\vartheta)) \leq 1 + \vartheta$. This proves statement (3) of the lemma.

□

The following lemma analyzes the concentration of the function $Q_m(\lambda)$ to the deterministic function $Q(\lambda)$.

Lemma 7. *Suppose $\frac{m}{n} = \delta$. For a Lipschitz function $\mathcal{T}$ whose range is in $[0, 1]$, there exists an event of probability 1, on which the following three statements hold:*

1. $\frac{1}{m}\sum_{i=1}^{m} \delta_{T_i} \xrightarrow{d} \mathcal{L}_T$,
2. $Q_m(\lambda) \to Q(\lambda) \quad \forall\ \lambda \in (1, \infty)$,
3. $a_m \to \mathbb{E}|Z|^2 T$.

In the above equations, $Z \sim \mathcal{CN}(0, 1)$, and $T = \mathcal{T}(|Z|^2)$. Furthermore, $\mathcal{L}_T$ denotes the law of the random variable T, and

$$Q(\lambda) = \mathbb{E}\left[\frac{|Z|^2}{\lambda - T}\right].$$

The proof of the above result is provided in Appendix A.1

The next lemma analyzes the properties of the limiting fixed point equation $Q(\lambda) = (\lambda - \mathbb{E}|Z|^2T - 1/\vartheta)^{-1}$. Define the critical value ϑ_c as:

$$\vartheta_c \stackrel{\text{def}}{=} \left(1 - \left(\mathbb{E}\left[\frac{|Z|^2}{1-T}\right]\right)^{-1} - \mathbb{E}[|Z|^2T]\right)^{-1} \geq 0.$$

Lemma 8. *Consider the fixed point equation (in λ)*

$$\lambda - \mathbb{E}[|Z|^2T] - 1/\vartheta = \frac{1}{\mathbb{E}\left[\frac{|Z|^2}{\lambda - T}\right]}, \tag{3.10}$$

on the domain:

$$\lambda > \max(1, \mathbb{E}[|Z|^2T] + 1/\vartheta).$$

We have

1. *If $\vartheta > \vartheta_c$, then the above equation has exactly 1 solution, denoted by $\lambda = \theta(\vartheta)$. Furthermore,*

$$\lambda - \mathbb{E}[|Z|^2T] - 1/\vartheta > \frac{1}{\mathbb{E}\left[\frac{|Z|^2}{\lambda - T}\right]}$$
$$\forall\ \lambda \in \left(\max(1, \mathbb{E}[|Z|^2T] + 1/\vartheta), \theta(\vartheta)\right),$$
$$\lambda - \mathbb{E}[|Z|^2T] - 1/\vartheta < \frac{1}{\mathbb{E}\left[\frac{|Z|^2}{\lambda - T}\right]}\ \forall\ \lambda \in \left(\theta(\vartheta), \infty\right).$$

   Furthermore, we have $\theta(\vartheta)$ is an increasing function of ϑ and $\lim_{\vartheta\to\infty}\theta(\vartheta) = \infty$.

2. *If $\vartheta \leq \vartheta_c$, then the equation has no solutions. For any $\vartheta \leq \vartheta_c$, we define $\theta(\vartheta) = 1$.*

Proof. The following change of measure simplifies some of the proofs:

$$p(z) \stackrel{\text{def}}{=} \frac{|z|^2}{\pi} \exp(-|z|^2),$$
$$\tilde{\mathbb{E}}[f(Z)] \stackrel{\text{def}}{=} \int f(z)p(z)\,\mathrm{d}z.$$

Note that $p(z)$ is a proper probability density function since $\int p(z)\,\mathrm{d}z = \mathbb{E}[|Z|^2] = 1$. With this notation, (3.10) can be written as

$$\lambda - \tilde{\mathbb{E}}[T] - 1/\vartheta = \frac{1}{\tilde{\mathbb{E}}\left[\frac{1}{\lambda - T}\right]},\ \lambda > \max(1, \tilde{\mathbb{E}}[T] + 1/\vartheta).$$

Define the random variable $G(\lambda) = (\lambda - T)^{-1}$. Note that $G'(\lambda) = -G^2(\lambda)$. Further, define

$$f(\lambda) \stackrel{\text{def}}{=} \frac{1}{\tilde{\mathbb{E}}\left[G(\lambda)\right]};\ \lambda \in [1, \infty).$$

The first two derivatives of $f(\lambda)$ are

$$f'(\lambda) = \frac{\tilde{\mathbb{E}}[G^2]}{\tilde{\mathbb{E}}[G]^2},$$
$$f''(\lambda) = -2 \cdot \frac{\tilde{\mathbb{E}}[G^3]\tilde{\mathbb{E}}[G] - \tilde{\mathbb{E}}[G^2]^2}{\tilde{\mathbb{E}}[G]^3}.$$

First, since $f'(\lambda) \geq 0$, the function $f(\lambda)$ is increasing. By Jensen's Inequality $f'(\lambda) \geq 1$. Since the equality holds if and only if G is deterministic, and we have assumed that the support of T is $[0, 1]$, we conclude that $f(\lambda) > 1$. Noting that $G \geq 0$ and applying Chebychev's association inequality (See Fact 1, Appendix A.3) with $B = A = G$ and $f(a) = g(a) = a$ gives $f''(\lambda) \leq 0$. Hence $f(\lambda)$ is an increasing, concave function and $f'(\lambda) > 1$.

Next, we claim that $f(\lambda) = \lambda - \tilde{\mathbb{E}}[T] - 1/\vartheta$ can have atmost one solution in $(1, \infty)$. To see this, let λ_1 be the first point at which the two curves intersect. Hence $f(\lambda_1) = \lambda_1 - \tilde{\mathbb{E}}[T] - 1/\vartheta$.

Furthermore

$$f'(\lambda) > 1 = \frac{\mathrm{d}(\lambda - \tilde{\mathbb{E}}[T] - 1/\vartheta)}{\mathrm{d}\lambda}.$$

Hence there can be no other intersection point of the two curves after λ_1.

Now consider the following two cases:

Case 1: $\vartheta > \vartheta_c$. First note that since $(1-x)^{-1}$ is a convex function on $(-\infty, 1]$, according to Jensen's Inequality

$$\tilde{\mathbb{E}}\left[\frac{1}{1-T}\right] \geq \frac{1}{1-\tilde{\mathbb{E}}[T]} \geq 0.$$

Hence,

$$\frac{1}{\vartheta_c} = 1 - \left(\tilde{\mathbb{E}}\left[\frac{1}{1-T}\right]\right)^{-1} - \tilde{\mathbb{E}}[T] \geq 0.$$

This shows that $\vartheta_c \geq 0$. Furthermore,

$$\vartheta > \vartheta_c \Longleftrightarrow (\lambda - \tilde{\mathbb{E}}[T] - 1/\vartheta)_{\lambda=1} > f(1).$$

On the other hand, we can also compare the limiting behavior of $\lambda - \tilde{\mathbb{E}}[T] - 1/\vartheta$ and $f(\lambda)$ as $\lambda \to \infty$. We have

$$\frac{\lambda - \tilde{\mathbb{E}}[T] - 1/\vartheta}{\lambda} = 1 - \frac{\tilde{\mathbb{E}}[T] + 1/\vartheta}{\lambda},$$

and

$$\begin{aligned}
\frac{f(\lambda)}{\lambda} = \frac{1}{\tilde{\mathbb{E}}\left[\frac{1}{1-T/\lambda}\right]} &= \left(\tilde{\mathbb{E}}\left[\sum_{n=0}^{\infty}\left(\frac{T}{\lambda}\right)^n\right]\right)^{-1} \\
&= \Big(1 + \tilde{\mathbb{E}}[T]/\lambda + o(1/\lambda)\Big)^{-1} \\
&= 1 - \frac{\tilde{\mathbb{E}}[T]}{\lambda} + o(\lambda^{-1}).
\end{aligned}$$

Hence, $f(\lambda) > \lambda - \tilde{\mathbb{E}}[T] - 1/\vartheta$ for λ large enough and $f(1) < 1 - \tilde{\mathbb{E}}[T] - 1/\vartheta$. Hence the functions $f(\lambda)$ and $1 - \tilde{\mathbb{E}}[T] - 1/\vartheta$ intersect once in $(1, \infty)$. Finally note that,

$$\begin{aligned}\frac{1}{\vartheta} + \tilde{\mathbb{E}}[T] < \frac{1}{\vartheta_c} + \tilde{\mathbb{E}}[T] &= 1 - \left(\tilde{\mathbb{E}}\left[\frac{1}{1-T}\right]\right)^{-1} \\ &\leq 1.\end{aligned}$$

Hence $f(\lambda) = \lambda - \tilde{\mathbb{E}}[T] - 1/\vartheta$ has exactly one solution in $\lambda \geq \max(1, \tilde{\mathbb{E}}[T] + 1/\vartheta)$ as claimed. By the Implicit Function Theorem, we can compute

$$\theta'(\vartheta) = \frac{1/\vartheta^2}{f'(\theta(\vartheta)) - 1} \geq 0. \tag{3.11}$$

Hence $\theta(\vartheta)$ is an increasing function of ϑ. Finally, we verify that $\lim_{\vartheta \to \infty} \theta(\vartheta) = \infty$. Suppose that this is not the case, i.e. $\theta(\vartheta) \to \theta_\infty < \infty$ as $\vartheta \to \infty$. Recalling the fixed point characterization of $\theta(\vartheta)$, we obtain that θ_∞ satisfies the fixed point equation

$$\theta_\infty - \tilde{\mathbb{E}}[T] = \frac{1}{\tilde{\mathbb{E}}\left[\frac{1}{\theta_\infty - T}\right]}.$$

This means that Jensen's Inequality applied to the strictly convex function $(\theta_\infty - t)^{-1}$ should be tight. This means under the tilted measure ($\tilde{\mathbb{E}}$), T is deterministic. This is not possible since we have assumed that T is supported on $[0, 1]$.

Case 2: $\vartheta \leq \vartheta_c$ As in Case 1 we argue (this time with the opposite conclusion) that

$$\vartheta \leq \vartheta_c \implies f(1) \geq (\lambda - \tilde{\mathbb{E}}[T] - 1/\vartheta)_{\lambda=1}$$

Furthermore, since $f'(\lambda) > \frac{\mathrm{d}(\lambda - \tilde{\mathbb{E}}[T] - 1/\vartheta)}{\mathrm{d}\lambda} = 1$, $f(\lambda) = \lambda - \tilde{\mathbb{E}}[T] - 1/\vartheta$ has no solution in $(1, \infty)$.

□

Combining the above sequence of lemmas, we obtain the following proposition about the spec-

trum of the matrix $\boldsymbol{E}(\vartheta)$.

Proposition 4. *Let $\boldsymbol{E}(\vartheta) = \boldsymbol{B}^{\mathsf{H}}(\boldsymbol{T} + \vartheta \boldsymbol{T}\boldsymbol{A}_1(\boldsymbol{T}\boldsymbol{A}_1)^{\mathsf{H}}))\boldsymbol{B}$. Then, there exists an event of probability 1, on which we have,*

1. $\mu_{\boldsymbol{E}(\vartheta)} \xrightarrow{d} \mathcal{L}_T$.
2. *If $\vartheta \leq \vartheta_c$, $\sigma(\boldsymbol{E}(\vartheta)) \subset [0,1]$.*
3. *If $\vartheta > \vartheta_c$, then $\lambda_i(\boldsymbol{E}(\vartheta)) \in [0,1]\ \forall\ i \geq 2$, and,*

$$\lambda_1(\boldsymbol{E}(\vartheta)) \xrightarrow{a.s.} \theta(\vartheta),$$

where $\theta(\vartheta)$ is the unique solution to the equation (in λ):

$$\lambda - \mathbb{E}[|Z|^2 T] - 1/\vartheta = \frac{1}{\mathbb{E}\left[\frac{|Z|^2}{\lambda - T}\right]},$$

in the domain:

$$\lambda > \max(1, \mathbb{E}[|Z|^2 T] + 1/\vartheta).$$

Proof. We restrict ourselves to the event guaranteed by Lemma 7, on which,

1. $a_m \to \mathbb{E}|Z|^2 T$
2. $\frac{1}{m}\sum_{i=1}^{m} \delta_{T_i} \xrightarrow{\mathrm{d}} \mathcal{L}_T$
3. $Q_m(\lambda) \to Q(\lambda)\ \forall\ \lambda \in (1, \infty)$.

Let us denote this event by $\mathcal{E}$. Define the sequence of (random) functions $f_m(\lambda)$ as:

$$f_m(\lambda) = \lambda - a_m - 1/\vartheta - \left(\sum_{i=1}^{m} \frac{|A_{1i}|^2}{\lambda - T_i}\right)^{-1},$$

with the domain:

$$\lambda > \max(1, a_m + 1/\vartheta).$$

Define the (deterministic) function $f(\lambda)$:

$$f(\lambda) = \lambda - \mathbb{E}[|Z|^2 T] - 1/\vartheta - \left(\mathbb{E}\left[\frac{|Z|^2}{\lambda - T} \right] \right)^{-1},$$

with the domain:

$$\lambda > \max(1, \mathbb{E}[|Z|^2 T] + 1/\vartheta).$$

Note that on $\mathcal{E}$, we have $f_m(\lambda) \to f(\lambda) \; \forall \; \lambda > 1$.

1. By Lemma 6, we know that the eigenvalues of $\boldsymbol{E}(\vartheta)$ interlace with the eigenvalues of the diagonal matrix $\boldsymbol{T}$. On the event $\mathcal{E}$, $\mu_{\boldsymbol{T}} \to \mathcal{L}_T$. Hence indeed $\mu_{\boldsymbol{E}(\vartheta)} \xrightarrow{\text{d}} \mathcal{L}_T$. This proves statement (1) of the proposition.

2. Consider the case $\vartheta \leq \vartheta_c$. By Lemma 6, we already know that $\lambda_2(\boldsymbol{E}(\vartheta)) \leq T_{(1)} \leq 1$ and $\lambda_{m-1}(\boldsymbol{E}(\vartheta)) \geq 0$. Hence to prove (2), it is sufficient to show that

$$\overline{\lambda}_1 \stackrel{\text{def}}{=} \limsup_{m \to \infty} \lambda_1(\boldsymbol{E}(\vartheta)) \leq 1, \text{ on } \mathcal{E}.$$

   For the sake of contradiction, suppose that there is a realization in $\mathcal{E}$ such that $\overline{\lambda}_1 > 1$. On this realization we consider a subsequence such that $\lambda_1(\boldsymbol{E}(\vartheta)) \to \overline{\lambda}_1$. All the analysis henceforth is along this subsequence. Since for all m large enough $\lambda_1(\boldsymbol{E}(\vartheta)) > 1$, by Lemma 6, we must have $f_m(\lambda_1(\boldsymbol{E}(\vartheta)) = 0$. Applying Lemma 3 from Lu and Li [19] (Appendix E), we

obtain

$$0 = f_m(\lambda_1(\boldsymbol{E}(\vartheta)) \to f(\overline{\lambda}_1).$$

Since $\vartheta \leq \vartheta_c$, we know by Lemma 8 that $f(\lambda) = 0$ does not have any solution in $\lambda > \max(1, \mathbb{E}[|Z|^2 T] + 1/\vartheta)$. Hence,

$$1 < \overline{\lambda}_1 \leq \mathbb{E}[|Z|^2 T] + 1/\vartheta.$$

However,

$$f(\overline{\lambda}_1) = \underbrace{\overline{\lambda}_1 - \mathbb{E}[|Z|^2 T] - 1/\vartheta}_{\leq 0} - \Bigg(\underbrace{\mathbb{E}\left[\frac{|Z|^2}{\lambda - T}\right]}_{>0} \Bigg)^{-1}$$

$$< 0.$$

This contradicts $f(\overline{\lambda}_1) = 0$. Hence, $\limsup_{m \to \infty} \lambda_1(\boldsymbol{E}(\vartheta)) \leq 1,$ on $\mathcal{E}$. This concludes the proof of statement (2).

3. Now consider the case $\vartheta > \vartheta_c$. Again by Lemma 6, we know $\lambda_i(\boldsymbol{E}(\vartheta)) \in [0, 1]$ for all $i \geq 2$. By Lemma 8, we know that $f(\lambda) = 0$ has a unique solution in $\lambda > \max(1, \mathbb{E}|Z|^2 T + 1/\vartheta)$ denoted by $\theta(\vartheta)$. Fix an ϵ small enough such that $[\theta(\vartheta) - \epsilon, \theta(\vartheta) + \epsilon]$ lies in the domain of $f(\lambda)$. Note that $f(\theta(\vartheta)) = 0$, while $f(\theta(\vartheta) - \epsilon) > 0$ and $f(\theta(\vartheta) + \epsilon) < 0$ (by Lemma 8).

   Since $a_m \to \mathbb{E}|Z|^2 T$, for all m large enough, $[\theta(\vartheta) - \epsilon, \theta(\vartheta) + \epsilon]$ also lies in the domain of $f_m(\lambda)$. By Lemma 7, we have $f_m(\lambda) \to f(\lambda)$ for all $\lambda \in [\theta(\vartheta) - \epsilon, \theta(\vartheta) + \epsilon]$. In particular, we have, for all n large enough $f_m(\theta(\vartheta) - \epsilon) > 0$ while $f_m(\theta(\vartheta) + \epsilon) < 0$. Hence, by Lemma 6, we have $\lambda_1(\boldsymbol{E}(\vartheta)) \in [\theta(\vartheta) - \epsilon, \theta(\vartheta) + \epsilon]$ for all n large enough. Hence indeed, $\lambda_1(\boldsymbol{E}(\vartheta)) \overset{\text{a.s.}}{\to} \theta(\vartheta)$. This proves (3).

□

3.5.4 Analysis of the Support of $\gamma \boxtimes \mathcal{L}_T$

We recall that $\mathcal{L}_T$ is the law of the random variable $T = \mathcal{T}(|Z|^2)$, and $\gamma = \frac{1}{\delta}\delta_1 + \left(1 - \frac{1}{\delta}\right)\delta_0$. To keep the notation clean, we will refer to the analytic transforms corresponding to the measure $\mathcal{L}_T$ with the subscript T, for example the Cauchy transform for the measure $\mathcal{L}_T$ will be referred to as G_T.We begin by computing the Cauchy transform of $\gamma \boxtimes T$.

Lemma 9. *Let $z \in \mathbb{C}^-$. Then, we have,*

$$G_{\gamma \boxtimes T}(z) = \frac{1}{z} \cdot \frac{1 - 1/\delta}{1 - z w_T(1/z)}.$$

In the above display, the subordination function, $w_T(1/z)$, is the unique solution in $\mathbb{C}^+$ to the equation $\Lambda(1/w) = z$, where the function Λ is defined as:

$$\Lambda(\tau) \stackrel{def}{=} \tau - \frac{(1 - 1/\delta)}{\mathbb{E}\left[\frac{1}{\tau - T}\right]}.$$

Proof. First we can compute the moment generating functions:

$$\psi_\gamma(z) = \frac{1}{\delta} \cdot \frac{z}{1 - z},$$
$$\psi_T(z) = -1 + \mathbb{E}\left[\frac{1}{1 - zT}\right].$$

The η-transforms of the two measures are given by,

$$\eta_\gamma(z) = \frac{z/\delta}{z/\delta - z + 1},$$
$$\eta_T(z) = \frac{\mathbb{E}\left[\frac{zT}{1 - zT}\right]}{\mathbb{E}\left[\frac{1}{1 - zT}\right]}.$$

Hence, we can compute the function Q_z, given in Definition 4,

$$Q_z(w) = \frac{1/\delta}{(1/\delta - 1)\frac{\mathbb{E}\left[\frac{T}{1-wT}\right]}{\mathbb{E}\left[\frac{1}{1-wT}\right]} + 1/z}.$$

Hence w_T is the unique solution in $\mathbb{C}^+$ of the equation $Q_z(w) = w$. This equation can be simplified to

$$\frac{1}{z} = \Lambda(1/w),$$

where the function Λ is defined as $\Lambda(\tau) \stackrel{\text{def}}{=} \tau - \frac{(1-1/\delta)}{\mathbb{E}\left[\frac{1}{\tau - T}\right]}$. Hence, we can compute the moment generating function of $\gamma \boxtimes T$ in the following way:

$$\begin{aligned}
\psi_{\gamma \boxtimes T}(z) &= \psi_T(w_T(z)) \\
&= -1 + \mathbb{E}\left[\frac{1}{1 - w_T(z)T}\right] \\
&\stackrel{\text{(a)}}{=} -1 + \frac{1 - 1/\delta}{1 - w_T(z)/z}.
\end{aligned}$$

In the above display, in the step marked (a), we used the fact that w_T solves $\Lambda(1/w) = 1/z$. Finally, the Cauchy transform of $\gamma \boxtimes T$ is given by

$$\begin{aligned}
G_{\gamma \boxtimes T}(z) &= \frac{1}{z}\left(\psi_{\gamma \boxtimes T}\left(\frac{1}{z}\right) + 1\right) \\
&= \frac{1}{z} \cdot \frac{1 - 1/\delta}{1 - z w_T(1/z)}.
\end{aligned}$$

□

Our next goal is to characterize Supp$(\gamma \boxtimes T)$. Theorem 4 gives a complete characterization of the support of the singular part of $\gamma \boxtimes T$. Hence, we now need to understand the support of the absolutely continuous part of $\gamma \boxtimes T$. According to the Stieltjes Inversion theorem, (Theorem 2) the

density of the continuous part is given by

$$\frac{\mathrm{d}(\gamma \boxtimes T)_{ac}}{\mathrm{d}x}(x) = \frac{1}{\pi} \lim_{\epsilon \to 0^+} \mathrm{Im}\, G_{\gamma \boxtimes T}(x - i\epsilon)$$
$$= \frac{1}{\pi x} \mathrm{Im} \left(\frac{1 - \frac{1}{\delta}}{1 - x \lim_{\epsilon \to 0^+} w_T(1/(x - i\epsilon))} \right).$$

Since $\tau_T(x - i\epsilon) \stackrel{\text{def}}{=} 1/w_T(1/(x - i\epsilon))$ uniquely solves $\Lambda(\tau) = x - i\epsilon$ in $\mathbb{C}^-$, our interest will be to study the solutions of this equation for $\epsilon \approx 0$. Hence, we begin by studying the solutions of $\Lambda(\tau) = x$. Before doing so, we clarify the definition of $\Lambda(\tau)$ at $\tau = 1$ which is a subtle case because $1 \in \text{Supp}(T)$. We note that the random variable $(1 - T)^{-1}$ is non-negative and hence the expectation $\mathbb{E}[(1 - T)^{-1}]$ is well defined but might be ∞. If it is finite, then $\Lambda(\tau)$ is well defined at $\tau = 1$. If the expectation is ∞, we define $\Lambda(1) = 1$ which is consistent with intepreting $1/\infty = 0$. $\Lambda(\tau)$ is defined at $\tau = 0$ analogously. This definition ensures $\Lambda(\tau)$ is a continuous function on $(-\infty, 0] \cup [1, \infty)$. Next we discuss the solutions of $\Lambda(\tau) = x$. Figure 3.2 shows a typical plot $\Lambda(\tau)$. As is clear from this figure we expect the following two quantities to play major roles in determining the existence of a solution of $\Lambda(\tau) = x$: Define

$$\lambda_l = \max_{\tau \in (-\infty, 0]} \Lambda(\tau), \ \tau_l = \arg\max_{\tau \in (-\infty, 0]} \Lambda(\tau)$$
$$\lambda_r = \min_{\tau \in [1, \infty)} \Lambda(\tau), \ \tau_r = \arg\min_{\tau \in [1, \infty)} \Lambda(\tau).$$

Our next lemma proves the properties of $\Lambda(\tau)$ suggested by Figure 3.2.

Lemma 10. *The following statements are true about* $\Lambda(\tau)$*:*

1. $\Lambda(\tau)$ *is a convex function on* $[1, \infty)$ *and a concave function on* $(-\infty, 0]$.
2. $\lim_{\tau \to \infty} \Lambda(\tau) = \infty, \ \lim_{\tau \to -\infty} \Lambda(\tau) = -\infty$.
3. $\lambda_r > \lambda_l \geq 0$.
4. *Consider the 3 mutually exclusive and exhaustive cases:*

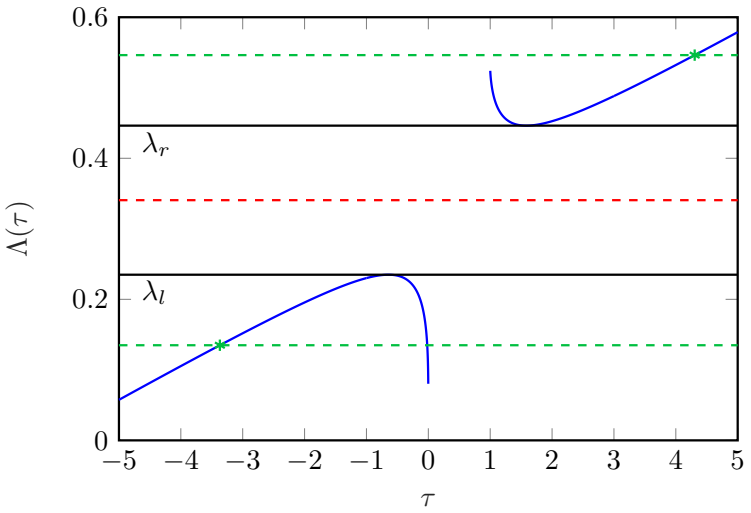

Figure 3.2: An Illustrative plot of the function $\Lambda(\tau)$: When $\lambda_l < x < \lambda_r$, the equation $\Lambda(\tau) = x$ has no solutions. When $x \geq \lambda_r$, the equation $\Lambda(\tau) = x, \Lambda'(\tau) > 0$ has a unique solution in $[1, \infty)$. When $x < \lambda_l$, then $\Lambda(\tau) = x, \Lambda'(\tau) > 0$ has a unique solution in $(-\infty, 0]$.

Case A: $x \leq \lambda_l$. *There is at least one and at most two solutions to* $\Lambda(\tau) = x$. *All solutions lie in* $(-\infty, 0]$. *Furthermore, when* $x < \lambda_l$, *there is exactly one solution for the equation* $\Lambda(\tau) = x, \Lambda'(\tau) > 0$. *This unique solution additionally satisfies* $\tau < \tau_l \leq 0$.

Case B: $\lambda_l < x < \lambda_r$. *There are no solutions of the equation* $\Lambda(\tau) = x,\ \tau \in (-\infty, 0] \cup [1, \infty)$.

Case C: $x \geq \lambda_r$. *There is at least one and at most two solutions to* $\Lambda(\tau) = x$. *All solutions lie in* $[1, \infty)$. *Furthermore, when,* $x > \lambda_r$, *there is a unique solution to* $\Lambda(\tau) = x, \Lambda'(\tau) > 0$. *This solution additionally satisfies* $\tau > \tau_r \geq 1$.

Proof. 1. We define the random variable $G(\tau)$,

$$G(\tau) \stackrel{\text{def}}{=} \frac{1}{\tau - T}.$$

We observe that for any $\tau \in [1, \infty)$, $G(\tau) \geq 0$ where as for $\tau \in (-\infty, 0]$, $G(\tau) \leq 0$. It is straightforward to see that $G'(\tau) = -G^2(\tau) \leq 0$. For notational simplicity, we will often

short hand $G(\tau)$ as G. We have

$$\Lambda'(\tau) = 1 - \left(1 - \frac{1}{\delta}\right) \cdot \frac{\mathbb{E}G^2}{(\mathbb{E}G)^2},$$
$$\Lambda''(\tau) = 2\left(1 - \frac{1}{\delta}\right) \cdot \frac{(\mathbb{E}G^3) \cdot (\mathbb{E}G) - (\mathbb{E}G^2)^2}{(\mathbb{E}G)^3}.$$

Consider the following two cases,

Case 1: $\tau \in [1, \infty)$**.** Applying Chebychev's Association Inequality (Fact 1) with $A = B = G$ and $f(a) = g(a) = a$ gives us that $\Lambda''(\tau) \geq 0$. In fact, an inspection of the proof of the Chebychev's Association Inequality from [70] allows us to rule out the equality case under the assumptions imposed on $\mathcal{T}$, and we have $\Lambda''(\tau) > 0$. Hence, Λ is strictly convex in $(1, \infty)$. Since $\Lambda(\tau)$ is continuous on $[1, \infty)$, we have Λ is convex on $[1, \infty)$

Case 2: $\tau \in (-\infty, 0]$**.** Again, applying Chebychev's Association Inequality with $A = B = -G$ and $f(a) = f(b) = a$ gives us $\Lambda''(\tau) \leq 0$, Hence Λ is concave in this region. As before, an inspection of the proof of Chebychev's Association inequality allows us to rule out the equality case under the assumptions imposed on $\mathcal{T}$, and we have $\Lambda''(\tau) < 0$. Hence, Λ is strictly concave in $(-\infty, 0)$. Since $\Lambda(\tau)$ is continuous on $(-\infty, 0)$, we have Λ is concave on $(-\infty, 0]$. This concludes the proof of statement (1) in the lemma.

2. Note that,

$$\lim_{\tau \to \infty} \tau - \frac{(1 - 1/\delta)}{\mathbb{E}\left[\frac{1}{\tau - T}\right]} = \tau \left(1 - \frac{(1 - 1/\delta)}{\mathbb{E}\left[\frac{\tau}{\tau - T}\right]}\right) = \infty.$$

This shows $\lim_{\tau \to \infty} \Lambda(\tau) = \infty$. The claim about the limit as $\tau \to -\infty$ can be analogously obtained. This proves item (2) in the statement of the lemma.

3. The infimum in the definition of λ_r is attained due to item (2) in the statement of the lemma.

Analogously, the supremum in the definition of λ_l is attained. Next consider any $\tau_+ \in (1, \infty)$ and any $\tau_- \in (-\infty, 0)$. Since the function $f(t) = (\tau_+ - t)^{-1}$ is convex on $[0, 1]$, according to Jensen's Inequality, we have

$$\begin{aligned}\Lambda(\tau_+) &\geq \tau_+ - \left(1 - \frac{1}{\delta}\right) \cdot (\tau_+ - \mathbb{E}[T]) \\ &= \frac{\tau_+}{\delta} + \left(1 - \frac{1}{\delta}\right) \cdot \mathbb{E}[T].\end{aligned}$$

On the other hand, since the function $f(t) = (\tau_- - t)^{-1}$ is concave on $[0, 1]$, we have

$$\begin{aligned}\Lambda(\tau_-) &\leq \tau_- - \left(1 - \frac{1}{\delta}\right) \cdot (\tau_- - \mathbb{E}[T]) \\ &= \frac{\tau_-}{\delta} + \left(1 - \frac{1}{\delta}\right) \cdot \mathbb{E}[T].\end{aligned}$$

Hence,

$$\begin{aligned}\Lambda(\tau_+) &\geq \frac{1}{\delta} + \left(1 - \frac{1}{\delta}\right) \cdot \mathbb{E}[T] \\ &> \left(1 - \frac{1}{\delta}\right) \cdot \mathbb{E}[T] \\ &\geq \Lambda(\tau_-).\end{aligned}$$

Taking the minimum over τ_+ and maximum of τ_- gives us $\lambda_r > \lambda_l$. Furthermore we note that $\Lambda(0^-) \geq 0$. Hence $\lambda_l \geq 0$. This concludes the proof of item (3) in the statement of the lemma.

4. For any $x \in (\lambda_l, \lambda_r)$, $\Lambda(\tau) = x$ doesn't have a solution in $(-\infty, 0] \cup [1, \infty)$ since $\Lambda(\tau) \leq \lambda_l \; \forall \; \tau \leq 0$ and $\Lambda(\tau) \geq \lambda_r \; \forall \; \tau \geq 1$. Now consider any $x \geq \lambda_r$. Since $\lambda(\tau) \leq \lambda_l < \lambda_r \; \forall \; \tau \leq 0$, we know that all solutions of $\Lambda(\tau) = x$ lie in $[1, \infty)$. Since Λ is strictly convex in $(1, \infty)$, there can be atmost 2 solutions. Now consider any $x > \lambda_r$. Let $\tau_r = \arg\min_{\tau \geq 1} \Lambda(\tau)$. Due to strict convexity of $\Lambda(\tau)$, we have $\Lambda'(\tau) > 0$ for any $\tau \in (\tau_r, \infty)$. Hence $\Lambda(\tau)$ is strictly

increasing on $[\tau_r, \infty)$. Since $\lambda_r = \Lambda(\tau_r) < x < \Lambda(\infty) = \infty$, we are guaranteed to have exactly one solution to $\Lambda(\tau) = x$ on (τ_r, ∞) which indeed satisfies $\Lambda'(\tau) > 0$. The analysis for the case when $x \leq \lambda_l$ can be done in a similar way. This concludes the proof of item (4) in the statement of the lemma.

□

We are now in the position to characterize the support of $\gamma \boxtimes T$ which is the content of the following proposition.

Proposition 5. *The support of $\gamma \boxtimes T$ is given by*

$$\textit{Supp}(\gamma \boxtimes T) = [\lambda_l, \lambda_r] \cup \textit{Supp}((\gamma \boxtimes T)_d),$$

where $(\gamma \boxtimes T)_d$ denotes the discrete part of the measure $\gamma \boxtimes T$. If the random variable T has a density with respect to the Lebesgue measure, then,

$$\textit{Supp}(\gamma \boxtimes T) = [\lambda_l, \lambda_r].$$

Proof. We first claim that $(\lambda_l, \lambda_r) \subset \text{Supp}(\gamma \boxtimes T)$. Since the support of a measure is closed, this means that $[\lambda_l, \lambda_r] \subset \text{Supp}(\gamma \boxtimes T)$. We prove this claim by contradiction. Suppose that $\exists \lambda \in (\lambda_l, \lambda_r)$ such that $\lambda \notin \text{Supp}(\gamma \boxtimes T)$. To simplify notation, for $z \in \mathbb{C}^-$, we introduce the following reciprocal subordination function $\tau_T(z)$

$$\tau_T(z) \stackrel{\text{def}}{=} \frac{1}{w_T(1/z)}.$$

According to Lemma 5, we have

$$\tau_T(\lambda) \stackrel{\text{def}}{=} \lim_{\epsilon \to 0^+} \tau_T(\lambda - i\epsilon) \in (-\infty, 0) \cup (1, \infty).$$

By Lemma 9, $\tau_T(\lambda - i\epsilon)$ uniquely solves the equation $\Lambda(\tau) = \lambda - i\epsilon$ in $\mathbb{C}^-$. Taking $\epsilon \to 0$, we

obtain,

$$\begin{aligned}\lambda &= \lim_{\epsilon\to 0^+} \Lambda(\tau_T(\lambda - i\epsilon)) \\ &= \lim_{\epsilon\to 0^+} \left(\tau_T(\lambda - i\epsilon) - \frac{1-1/\delta}{\mathbb{E}\left[\frac{1}{\tau_T(\lambda - i\epsilon)-T}\right]}\right) \\ &\overset{\text{(a)}}{=} \tau_T(\lambda) - \frac{1-1/\delta}{\mathbb{E}\left[\frac{1}{\tau_T(\lambda)-T}\right]}.\end{aligned}$$

In the step marked (a), we used the fact that since $\lim_{\epsilon\to 0^+} \tau_T(\lambda - i\epsilon) \notin \text{Supp}(T)$, we have $\exists c > 0$, such that for any ϵ small enough $\text{dist}(\tau_T(\lambda - i\epsilon), \text{Supp}(T)) \geq c$. This gives us a dominating function for an application of the dominated convergence theorem. Hence, we have found a solution for the equation $\lambda = \Lambda(\tau), \tau \in (-\infty, 0) \cup (1, \infty)$. But this contradicts Lemma 10. Hence, we have, $(\lambda_l, \lambda_r) \subset \text{Supp}(\gamma \boxtimes T)$.

Next, we claim that any $x \in [0, \lambda_l) \cup (\lambda_r, \infty)$ is not in the support of the absolutely continuous part of $\gamma \boxtimes T$. To show this, we first compute a first order asymptotic expansion of $\tau_T(x - i\epsilon)$ for $\epsilon \approx 0$. From Lemma 10, we know there exists a unique solution for the equation $\Lambda(\tau) = x, \tau \in (-\infty, 0) \cup (1, \infty)$ and $\Lambda'(\tau) > 0$. We denote this solution by $\tau_\star$. Since $\tau_\star \notin \text{Supp}(T)$, the function $\Lambda(\tau)$ is analytic in the neighborhood (in $\mathbb{C}$) of $\tau_\star$. The implicit function theorem guarantees us a solution $\tau(\epsilon) = \tau_R(\epsilon) + i\tau_I(\epsilon)$ of the equation $\Lambda(\tau) = x - i\epsilon$. However, this $\tau(\epsilon)$ may not be the reciprocal subordination function $\tau_T(x - i\epsilon)$ since we still need to verify it is in $\mathbb{C}^-$. To take care of this, again by the implicit function theorem we have

$$\Lambda'(\tau_\star) \cdot \frac{\mathrm{d}\tau}{\mathrm{d}\epsilon}(0) = -i.$$

This gives us

$$\frac{\mathrm{d}\tau_I}{\mathrm{d}\epsilon}(0) = -\frac{1}{\Lambda'(\tau_\star)} < 0, \ \frac{\mathrm{d}\tau_R}{\mathrm{d}\epsilon}(0) = 0.$$

Hence, we have

$$\tau(\epsilon) = \tau_\star - i\frac{\epsilon}{\Lambda'(\tau_\star)} + o(\epsilon).$$

This verifies that $\tau(\epsilon) \in \mathbb{C}^-$ for ϵ small enough. Finally since $\tau_T(x - i\epsilon)$ is the unique solution to the equation $\Lambda(\tau) = x - i\epsilon$ in $\mathbb{C}^-$, we have

$$\tau_T(x - i\epsilon) = \tau_\star - i\frac{\epsilon}{\Lambda'(\tau_\star)} + o(\epsilon).$$

According to the Stieltjes Inversion Formula, Theorem 2, we obtain

$$\begin{aligned}\frac{\mathrm{d}(\gamma \boxtimes T)_{ac}}{\mathrm{d}x}(x) &= \frac{1}{\pi x} \cdot \mathrm{Im}\left(\frac{1 - \frac{1}{\delta}}{1 - x \cdot \lim_{\epsilon \to 0^+} w_T\left(\frac{1}{x - i\epsilon}\right)}\right) \\ &\overset{\text{(b)}}{=} \frac{1}{\pi x} \cdot \mathrm{Im}\left(\frac{(1 - 1/\delta) \cdot \tau_\star}{\tau_\star - x}\right) = 0.\end{aligned}$$

In the step marked (b), we are relying on the assumption that $\tau_\star \neq x$. To verify this, we recall that $\tau_\star$ solves, $\Lambda(\tau_\star) = x$ and $\tau_\star \notin [0, 1]$. This means that

$$\begin{aligned}|\tau_\star - x| &= \frac{1 - 1/\delta}{\left|\mathbb{E}\left[\frac{1}{\tau_\star - T}\right]\right|} \\ &\geq \frac{1 - 1/\delta}{\mathbb{E}\left[\left|\frac{1}{\tau_\star - T}\right|\right]} \\ &\geq (1 - 1/\delta) \cdot \mathrm{dist}(\tau_\star, [0, 1]) > 0.\end{aligned}$$

Hence, we have shown

$$\frac{\mathrm{d}(\gamma \boxtimes T)_{ac}}{\mathrm{d}x}(x) \overset{\text{a.s.}}{=} 0, \forall x \in [0, \lambda_l) \cup (\lambda_r, \infty).$$

This implies,

$$[0, \lambda_l) \cup (\lambda_r, \infty) \subset \mathbb{R} \backslash \text{Supp}((\gamma \boxtimes T)_{ac}).$$

Taking complements, we have $\text{Supp}((\gamma \boxtimes T)_{ac}) \subset [\lambda_l, \lambda_r]$. Hence, we have shown that

$$\begin{aligned}
[\lambda_l, \lambda_r] \cup \text{Supp}((\gamma \boxtimes T)_d) &\subset \text{Supp}(\gamma \boxtimes T) \\
&= \text{Supp}((\gamma \boxtimes T)_{ac}) \cup \text{Supp}((\gamma \boxtimes T)_d) \\
&\subset [\lambda_l, \lambda_r] \cup \text{Supp}((\gamma \boxtimes T)_d).
\end{aligned}$$

Therefore, $\text{Supp}(\gamma \boxtimes T) = [\lambda_l, \lambda_r] \cup \text{Supp}((\gamma \boxtimes T)_d)$ which proves the claim of the proposition. Finally, when T has a density with respect to Lebesgue measure, Theorem 4 gives us $\text{Supp}((\gamma \boxtimes T)_d) = \emptyset$ which yields the second claim in the proposition. □

Finally we note that in order to apply Theorem 3, it is necessary to understand the set:

$$\tau_T^{-1}(\{\theta\}) \cap (\mathbb{R} \backslash \text{Supp}(\gamma \boxtimes T)),\ \theta \in \mathbb{R}$$

(see Theorem 3 to recall the definition of τ_T). This is done in the following lemma.

Lemma 11. *Let (w_γ, w_T) denote the subordination functions corresponding to the free multiplicative convolution of $\gamma, \mathcal{L}_T$. Define*

$$\tau_T(z) = \frac{1}{w_T(1/z)}.$$

Then, we have

$$\tau_T^{-1}(\{\theta\}) \cap (\mathbb{R} \backslash \textit{Supp}(\gamma \boxtimes T)) = \begin{cases} \theta \in [\tau_l, \tau_r] : & \emptyset \\ \theta \notin [\tau_l, \tau_r] : & \{\Lambda(\theta)\} \end{cases},$$

where where, $\tau_l \triangleq \arg\max_{\tau \le 0} \Lambda(\tau)$, $\tau_r \triangleq \arg\min_{\tau \ge 1} \Lambda(\tau)$.

Proof. From Proposition 5, we know that $\text{Supp}(\gamma \boxtimes T) = [\lambda_l, \lambda_r]$, where $\lambda_l \stackrel{\text{def}}{=} \max_{\tau \le 0} \Lambda(\tau)$ and $\lambda_r \stackrel{\text{def}}{=} \min_{\tau \ge 1} \Lambda(\tau)$. Furthermore, we showed that for any $x \notin [\lambda_l, \lambda_r]$, the reciprocal subordination function $\tau_T(x)$ is the unique solution to the equations: $\Lambda(\tau) = x, \Lambda'(\tau) > 0,\ \tau \notin [0, 1]$. From Lemma 10, we know that when $x > \lambda_r$, the unique solution to $\Lambda(\tau) = x, \Lambda'(x) > 0$ satisfies $\tau > \tau_r$ and when $x < \lambda_l$, the unique solution satisfies $\tau < \tau_l$. These considerations immediately yield the claim of the lemma. □

3.5.5 Proof of Lemmas 3 and 4

Recall we defined $\Lambda_+(\tau)$ as

$$\Lambda_+(\tau) = \begin{cases} \tau - \frac{(1-1/\delta)}{\mathbb{E}\left[\frac{1}{\tau - T}\right]} & \text{if } \tau > \tau_r, \\ \min_{\tau \ge 1} \left(\tau - \frac{(1-1/\delta)}{\mathbb{E}\left[\frac{1}{\tau - T}\right]} \right) & \text{if } \tau \le \tau_r, \end{cases}$$

where $T = \mathcal{T}(|Z|/\sqrt{\delta})$ and $Z \sim \mathcal{CN}(0, 1)$, and

$$\tau_r \triangleq \arg\min_{\tau \ge 1} \left(\tau - \frac{(1 - 1/\delta)}{\mathbb{E}\left[\frac{1}{\tau - T}\right]} \right).$$

We first prove Lemma 3, which we restated below for convenience.

Lemma 3. *Let* $\vartheta_c \stackrel{\text{def}}{=} \left(1 - \left(\mathbb{E}\left[\frac{|Z|^2}{1-T}\right]\right)^{-1} - \mathbb{E}[|Z|^2 T]\right)^{-1}$. *Define the function* $\theta(\vartheta)$ *as:*

- *When* $\vartheta > \vartheta_c$*: Let* $\theta(\vartheta)$ *be the unique value of* λ *that satisfies the equation:*

$$\lambda - \mathbb{E}[|Z|^2 T] - 1/\vartheta = \left(\mathbb{E}\left[\frac{|Z|^2}{\lambda - T}\right]\right)^{-1},$$

in the interval:

$$\lambda \in \left(\max(1, \mathbb{E}[|Z|^2 T] + 1/\vartheta), \infty\right).$$

- *When* $\vartheta \leq \vartheta_c$*:* $\theta(\vartheta) \overset{def}{=} 1$.

Then, we have $L_m(\vartheta) \overset{a.s.}{\to} \Lambda_+(\theta(\vartheta))$*, where* $L_m(\vartheta)$ *is defined in* (3.3)*.*

Proof. In Proposition 6, we obtained an asymptotic characterization of the spectrum of $\boldsymbol{E}(\vartheta)$. More specifically, we proved that

$$\mu_{\boldsymbol{E}(\vartheta)} \overset{\text{d}}{\to} \mathcal{L}_T, \ \lambda_1(\boldsymbol{E}(\vartheta)) \to \theta(\vartheta).$$

We recall the matrix $\boldsymbol{R}$ was defined as

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{I}_{n-1} & \boldsymbol{0}_{n-1,m-1} \\ \boldsymbol{0}_{m-n,n-1} & \boldsymbol{0}_{m-1,m-1} \end{bmatrix}.$$

In particular, $\mu_{\boldsymbol{R}} \overset{\text{d}}{\to} \gamma$, where the measure γ is given by

$$\gamma = \frac{1}{\delta}\delta_1 + \left(1 - \frac{1}{\delta}\right)\delta_0.$$

Applying Theorem 3, we obtain:

1. The spectral measure of $\boldsymbol{E}(\vartheta)\boldsymbol{H}_{m-1}\boldsymbol{R}\boldsymbol{H}_{m-1}^{\mathsf{H}}$ converges to:

$$\mu_{\boldsymbol{E}(\vartheta)\boldsymbol{H}_{m-1}\boldsymbol{R}\boldsymbol{H}_{m-1}^{\mathsf{H}}} \overset{\text{d}}{\to} \gamma \boxtimes \mathcal{L}_T.$$

2. For any $\epsilon > 0$, we have, almost surely, for m large enough that, $\sigma(\boldsymbol{E}(\vartheta)\boldsymbol{H}_{m-1}\boldsymbol{R}\boldsymbol{H}_{m-1}^{\mathsf{H}}) \subset K_\epsilon$, where K_ϵ is the ϵ-neighborhood of the set $K = \text{Supp}(\gamma \boxtimes \mathcal{L}_T) \cup \tau_T^{-1}(\{\theta(\vartheta)\})$.

3. For any $\lambda \in \tau_T^{-1}(\{\theta(\vartheta)\}) \cap (\mathbb{R} \backslash \text{Supp}(\gamma \boxtimes \mathcal{L}_T))$, we have almost surely exactly one eigenvalue of $\boldsymbol{E}(\vartheta)\boldsymbol{H}_{m-1}\boldsymbol{R}\boldsymbol{H}_{m-1}^{\mathsf{H}}$ in a small enough neighborhood of λ for large enough n.

In Proposition 5, we characterized $\text{Supp}(\gamma \boxtimes \mathcal{L}_T)$ as $[\lambda_l, \lambda_r]$, where $\lambda_l = \max_{\tau \leq 0} \Lambda(\tau)$, $\lambda_r = \min_{\tau \geq 1} \Lambda(\tau)$ and the function $\Lambda(\tau)$ is given by:

$$\Lambda(\tau) = \tau - \frac{(1 - 1/\delta)}{\mathbb{E}\left[\frac{1}{\tau - T}\right]}.$$

In Lemma 11, we characterized the set:

$$\tau_T^{-1}(\{\theta\}) \cap (\mathbb{R} \backslash \text{Supp}(\gamma \boxtimes T)) = \begin{cases} \emptyset & \theta \in [\tau_l, \tau_r], \\ \{\Lambda(\theta)\} & \theta \notin [\tau_l, \tau_r], \end{cases}$$

where, $\tau_l \triangleq \arg\max_{\tau \leq 0} \Lambda(\tau)$, $\tau_r \triangleq \arg\min_{\tau \geq 1} \Lambda(\tau)$. Putting these together, one obtains the following two cases:

Case 1: $\theta(\vartheta) \leq \tau_r$. In this case, the set $\tau_T^{-1}(\{\theta\}) \cap (\mathbb{R} \backslash \text{Supp}(\gamma \boxtimes T)) = \emptyset$. The matrix $\boldsymbol{E}(\vartheta)\boldsymbol{H}_{m-1}\boldsymbol{R}\boldsymbol{H}_{m-1}^{\mathsf{H}}$ has no eigenvalues outside the support of the bulk distribution, and

$$L_m(\vartheta) \overset{\text{a.s.}}{\to} \lambda_r = \Lambda(\tau_r).$$

Case 2: $\theta(\vartheta) > \tau_r$. In this case, the set

$$\tau_T^{-1}(\{\theta\}) \cap (\mathbb{R} \backslash \text{Supp}(\gamma \boxtimes T)) = \{\Lambda(\theta(\vartheta))\}.$$

Hence, there is an eigenvalue in the neighborhood of $\Lambda(\theta(\vartheta)))$. Since $\theta(\vartheta) > \tau_r$, and Λ is a strictly increasing function on $[\tau_r, \infty)$ (Lemma 10), we have $\Lambda(\theta(\vartheta)) > \lambda_r$. Hence the eigenvalue in the neighborhood of $\Lambda(\theta(\vartheta))$ is the largest one, and we have

$$L_m(\vartheta) \overset{\text{a.s.}}{\to} \Lambda(\theta(\vartheta)).$$

It is now straightforward to check that the above two cases can be combined into a concise form stated in the claim of the lemma. □

We end this section by proving Lemma 4, restated below for convenience.

Lemma 4. *The following hold for the equation:*

$$\Lambda_+(\theta(\vartheta)) = 1/\vartheta + \mathbb{E}[|Z|^2 T],\ \vartheta > 0.$$

1. *This equation has a unique solution.*

2. *Let $\vartheta_\star$ denote the solution of the above equation. Then:*

Case 1 *If $\psi_1(\tau_r) \leq \frac{\delta}{\delta-1}$, we have*

$$\Lambda_+(\theta(\vartheta_\star)) = \Lambda(\tau_r).$$

Furthermore if $\psi_1(\tau_r) < \delta/(\delta-1)$, then,

$$\left.\frac{\mathrm{d}\Lambda_+(\theta(\vartheta))}{\mathrm{d}\vartheta}\right|_{\vartheta=\vartheta_\star} = 0,$$

Case 2 *If $\psi_1(\tau_r) > \frac{\delta}{\delta-1}$, we have*

$$\Lambda_+(\theta(\vartheta_\star)) = \Lambda(\theta_\star),$$

and,

$$\left.\frac{\mathrm{d}\Lambda_+(\theta(\vartheta))}{\mathrm{d}\vartheta}\right|_{\vartheta=\vartheta_\star} = \frac{1}{\vartheta_\star^2} \cdot \frac{\delta}{\delta-1} \cdot \left(\frac{\delta}{\delta-1} - \psi_2(\theta_\star)\right) \cdot \frac{1}{\psi_3^2(\theta_\star) - \frac{\delta^2}{(\delta-1)^2}}.$$

where $\theta_\star > 1$ is the unique $\theta \geq \tau_r$ that satisfies $\psi_1(\theta) = \frac{\delta}{\delta-1}$.

Proof. Before we begin the proof of this lemma, it is helpful to list the conclusions of some of the previous lemmas.

Lemma 8: In this lemma, for $\vartheta > \vartheta_c$ we defined the function $\theta(\vartheta)$ as the unique value of $\lambda > \max(1, \mathbb{E}[|Z|^2 T] + 1/\vartheta)$ that satisfies

$$\lambda - \mathbb{E}[|Z|^2 T] - 1/\vartheta = \frac{1}{\mathbb{E}\left[\frac{|Z|^2}{\lambda - T}\right]}.$$

We also set $\theta(\vartheta) = 1$ when $\vartheta \leq \vartheta_c$. We also showed that $\theta(\vartheta)$ is strictly increasing on $[\vartheta_c, \infty)$ and $\theta(\infty) = \infty$. In particular $\theta(\vartheta)$ has a well defined inverse defined on the domain $[1, \infty)$ given by:

$$\theta^{-1}(\lambda) = \left(\lambda - \mathbb{E}[|Z|^2 T] - \frac{1}{\mathbb{E}\left[\frac{|Z|^2}{\lambda - T}\right]}\right)^{-1}. \tag{3.12}$$

Lemma 10: We defined the function $\Lambda(\tau)$ as

$$\Lambda(\tau) \triangleq \tau - \frac{(1 - 1/\delta)}{\mathbb{E}\left[\frac{1}{\tau - T}\right]}. \tag{3.13}$$

We showed the that $\Lambda(\tau)$ is strictly convex on $[1, \infty)$. We defined (τ_r, λ_r) to be the minimizing argument and the minimum value of $\Lambda(\tau)$ in $[1, \infty)$. In particular $\tau_r \geq 1$. We also showed that $\Lambda(\infty) = \infty$. We further defined $\Lambda_+(\tau)$ in the following way:

$$\Lambda_+(\tau) = \begin{cases} \lambda_r, & \tau \leq \tau_r. \\ \Lambda(\tau), & \tau > \tau_r. \end{cases}$$

Some simple implications of the above assertions are: First, since $\theta(\vartheta)$ and Λ_+ are both non-decreasing continuous functions $\Lambda_+(\theta(\vartheta))$ is non-decreasing and continuous. Second, since $\Lambda(\tau) = \lambda_r$ for $\tau \leq \tau_r$, we have, for all $\vartheta \leq \theta^{-1}(\tau_r)$, $\Lambda_+(\theta(\vartheta)) = \lambda_r$. Third since $\theta(\infty) = \infty$ and

$\Lambda(\infty) = \infty$, we have, $\Lambda_+(\theta(\vartheta)) \to \infty$ as $\vartheta \to \infty$. The only possible point of non-differentiability of $\Lambda_+(\theta(\vartheta))$ is at $\vartheta = \theta^{-1}(\tau_r)$. It is straightforward to compute the derivative of $\Lambda(\theta(\vartheta))$ at all other points using implicit function theorem and obtain

$$\frac{\mathrm{d}\Lambda_+(\theta(\vartheta))}{\mathrm{d}\vartheta} = \begin{cases} 0 & \vartheta < \theta^{-1}(\tau_r), \\ \Lambda'(\theta(\vartheta)) \cdot \theta'(\vartheta) & \vartheta > \theta^{-1}(\tau_r). \end{cases} \tag{3.14}$$

The derivatives of Λ, θ can be calculated as,

$$\Lambda'(\tau) = \frac{\delta - 1}{\delta}\left(\frac{\delta}{\delta - 1} - \psi_2(\tau)\right). \tag{3.15}$$

$$\theta'(\vartheta) = \frac{1}{\vartheta^2}\left(\frac{\left(\mathbb{E}\left[\frac{|Z|^2}{\theta(\vartheta)-T}\right]\right)^2}{\mathbb{E}\left[\frac{|Z|^2}{(\theta(\vartheta)-T)^2}\right] - \left(\mathbb{E}\left[\frac{|Z|^2}{\theta(\vartheta)-T}\right]\right)^2}\right). \tag{3.16}$$

A representative plot of the function $\Lambda_+(\theta(\vartheta))$ is shown in Figure 3.3.

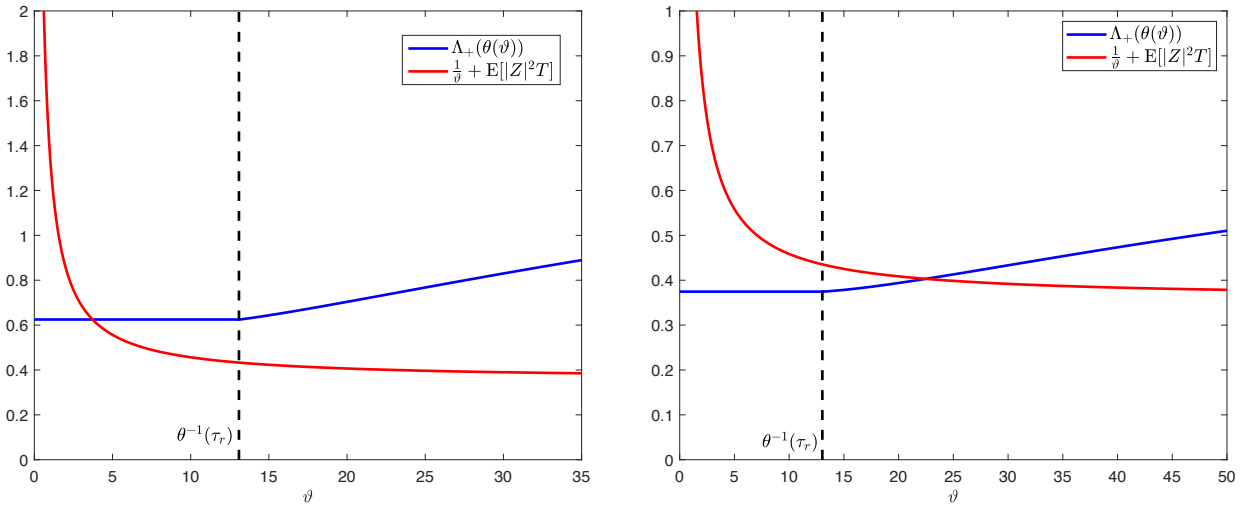

Figure 3.3: Typical Plots of the functions $\Lambda_+(\theta(\vartheta))$ (Blue) and $\mathbb{E}[|Z|^2 T] + \frac{1}{\vartheta}$ (Red). Case 1 (Left): The two functions intersect at the constant part of $\Lambda_+(\theta(\vartheta))$, Case 2 (Right): The The two functions intersect at the increasing part of $\Lambda_+(\theta(\vartheta))$

We are now in a position to prove the claims of the lemma.

1. Since $\Lambda_+(\theta(\vartheta))$ is continuous and non-decreasing and $1/\vartheta + \mathbb{E}[|Z|^2T]$ is continuous and strictly decreasing, the fixed point equation can have at most one solution. On the other hand comparing the values of the two sides of the fixed point equation at $\vartheta \to 0$ and $\vartheta \to \infty$ shows that there is at least one solution.

2. Let $\vartheta_\star$ be denote the solution of the fixed point equation $\Lambda_+(\theta(\vartheta)) = 1/\vartheta + \mathbb{E}[|Z|^2T]$. A typical plot of these two functions is shown in Figure 3.3. The figure shows two possible cases for the intersection of the two curves: *Case 1:* The curves intersect at a point $\vartheta_\star \leq \theta^{-1}(\tau_r)$ (or on the flat part of $\Lambda_+(\theta(\alpha))$. In this case we have, $\Lambda_+(\theta(\vartheta_\star)) = \lambda_r$.

   Case 2: The curves intersect at a point $\vartheta_\star > \theta^{-1}(\tau_r)$ or the rising part of $\Lambda_+(\theta(\alpha)$. We have $\Lambda_+(\theta(\vartheta_\star)) > \lambda_r$. We can distinguish between the two cases by comparing the value of the function $1/\vartheta + \mathbb{E}[|Z|^2T]$ at $\vartheta = \theta^{-1}(\tau_r)$ with λ_r. In particular, we have,

   Case 1:

$$\Lambda_+(\theta(\vartheta_\star)) = \lambda_r \Leftrightarrow 1/\theta^{-1}(\tau_r) + \mathbb{E}[|Z|^2T] \leq \lambda_r,$$

   Case 2:

$$\Lambda_+(\theta(\vartheta_\star)) > \lambda_r \Leftrightarrow 1/\theta^{-1}(\tau_r) + \mathbb{E}[|Z|^2T] > \lambda_r.$$

   Substituting the formula for $\theta^{-1}(\tau_r)$, mentioned in (3.12), and $\lambda_r = \Lambda(\tau_r)$ and the formula for Λ from (3.13), the 2 cases can be simplified slightly more.

   Case 1: This case occurs when

$$\frac{1}{\theta^{-1}(\tau_r)} + \mathbb{E}[|Z|^2T] \leq \lambda_r \Leftrightarrow \frac{\mathbb{E}\left[\frac{|Z|^2}{\tau_r - T}\right]}{\mathbb{E}\left[\frac{1}{\tau_r - T}\right]} \leq \frac{\delta}{\delta - 1}.$$

   In this situation, we have, $\Lambda_+(\theta(\vartheta_\star)) = \lambda_r$. Furthermore, if we additionally have

$$\frac{\mathbb{E}\left[\frac{|Z|^2}{\tau_r - T}\right]}{\mathbb{E}\left[\frac{1}{\tau_r - T}\right]} < \frac{\delta}{\delta - 1}$$

Then $\Lambda_+(\theta(\vartheta))$ is differentiable at $\vartheta_\star$ and, from (3.14), we have

$$\left.\frac{\mathrm{d}\Lambda_+(\theta(\vartheta))}{\mathrm{d}\vartheta}\right|_{\vartheta=\vartheta_\star} = 0.$$

Case 2: This case occurs when

$$\begin{aligned}\frac{1}{\theta^{-1}(\tau_r)} + \mathbb{E}[|Z|^2 T] > \lambda_r \\ \Leftrightarrow \frac{\mathbb{E}\left[\frac{|Z|^2}{\tau_r - T}\right]}{\mathbb{E}\left[\frac{1}{\tau_r - T}\right]} > \frac{\delta}{\delta - 1}.\end{aligned}$$

In this situation, we have, $\Lambda_+(\theta(\vartheta_\star)) > \lambda_r$. It turns out that we can give a simpler expression for $\Lambda_+(\theta(\vartheta_\star))$. In this case, $\vartheta_\star \geq \theta^{-1}(\tau_r)$ solves,

$$\Lambda(\theta(\vartheta_\star)) = \frac{1}{\vartheta_\star} + \mathbb{E}[|Z|^2 T], \tag{3.17}$$

and $\theta(\vartheta_\star) \geq 1$ is the solution of the equation

$$\mathbb{E}[|Z|^2 T] + \frac{1}{\vartheta_\star} = \theta(\vartheta_\star) - \frac{1}{\mathbb{E}\left[\frac{|Z|^2}{\theta(\vartheta_\star) - T}\right]}. \tag{3.18}$$

By definition the function $\Lambda(\tau(\alpha))$ is

$$\Lambda(\theta(\vartheta_\star)) = \theta(\vartheta_\star) - \frac{(1 - 1/\delta)}{\mathbb{E}\left[\frac{1}{\theta(\vartheta_\star) - T}\right]}. \tag{3.19}$$

We first eliminate $\vartheta_\star$ from Equations (3.17)-(3.19) and conclude that $\theta_\star \stackrel{\text{def}}{=} \theta(\vartheta_\star)$ solves

$$\frac{\mathbb{E}\left[\frac{|Z|^2}{\theta_\star - T}\right]}{\mathbb{E}\left[\frac{1}{\theta_\star - T}\right]} = \frac{\delta}{\delta - 1},\ \theta_\star \geq \tau_r, \tag{3.20}$$

and $\vartheta_\star$ is given by

$$\vartheta_\star = \left(\theta_\star - \frac{1}{\mathbb{E}\left[\frac{|Z|^2}{\theta_\star - T}\right]} - \mathbb{E}[|Z|^2 T] \right)^{-1}.$$

Since the solution to Equations (3.17)-(3.19) was guaranteed to be unique, the solution to (3.20) is guaranteed to be unique. Finally we can compute the derivative of $\Lambda_+(\theta(\vartheta))$ at $\vartheta = \vartheta_\star$. It will be convenient to introduce the random variable $G = (\theta_\star - T)^{-1}$ to write the equations in a compact form. From (3.14)-(3.16), we have

$$\begin{aligned}
&\left.\frac{\mathrm{d}\Lambda_+(\theta(\vartheta))}{\mathrm{d}\vartheta}\right|_{\vartheta=\vartheta_\star} = \Lambda'(\theta_\star)\cdot\theta'(\vartheta_\star) \\
&= \frac{\delta-1}{\delta\vartheta_\star^2}\left(\frac{\delta}{\delta-1} - \psi_2(\theta_\star)\right)\frac{\mathbb{E}[|Z|^2G]^2}{\mathbb{E}[|Z|^2G^2] - \mathbb{E}[|Z|^2G]^2} \\
&\overset{\text{(a)}}{=} \frac{\delta\cdot\left(\frac{\delta}{\delta-1} - \psi_2(\theta_\star)\right)}{\vartheta_\star^2\cdot(\delta-1)\cdot\psi_1^2(\theta_\star)}\cdot\frac{\mathbb{E}[|Z|^2G]^2}{\mathbb{E}[|Z|^2G^2] - \mathbb{E}[|Z|^2G]^2} \\
&= \frac{\delta\cdot\left(\frac{\delta}{\delta-1} - \psi_2(\theta_\star)\right)}{\vartheta_\star^2\cdot(\delta-1)}\cdot\frac{\mathbb{E}[G]^2}{\mathbb{E}[|Z|^2G^2] - \mathbb{E}[|Z|^2G]^2} \\
&= \frac{\delta}{\vartheta_\star^2(\delta-1)}\left(\frac{\delta}{\delta-1} - \psi_2(\theta_\star)\right)\frac{1}{\psi_3^2(\theta_\star) - \frac{\delta^2}{(\delta-1)^2}}.
\end{aligned}$$

In the above display, in the step marked (a) we used the fact that $\theta_\star$ satisfies $\psi_1(\theta_\star) = \delta/(\delta-1)$. This concludes the proof of the characterization (2) given in the statement of the lemma.

□

3.6 Conclusion

We analyzed the asymptotic performance of a spectral method for phase retrieval under a random column orthogonal matrix model. Our results provides a rigorous justification for the conjectures in [30], which were obtained by analyzing an expectation propagation algorithm.

Chapter 4: Information Theoretic Limits

4.1 Problem Formulation

In this chapter[1], we study information theoretic lower bounds for Phase Retrieval problem in the presence of (arbitrarily small) Gaussian measurement noise:

$$y_i = m|(\boldsymbol{A}\boldsymbol{x}_\star)_i|^2 + \sigma\epsilon_i,\ i = 1, 2 \dots m, \tag{4.1a}$$

$$\epsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)\,. \tag{4.1b}$$

We study this problem under the sub-sampled Haar ansatz for the sensing matrix:

$$\boldsymbol{A} = \boldsymbol{H}_m \cdot \boldsymbol{S}_{m,n}, \tag{4.2a}$$

$$\boldsymbol{H}_m \sim \text{Unif}\left(\mathbb{U}_m\right),\ \boldsymbol{S}_{m,n} = \begin{bmatrix} \boldsymbol{I}_n \\ \boldsymbol{0}_{(m-n)\times n} \end{bmatrix}. \tag{4.2b}$$

We assume that the signal vector is a uniformly random unit vector: $\boldsymbol{x}_\star \sim \text{Unif}\left(\mathbb{S}_{n-1}\right)$. This is intended to model situations where we don't have apriori knowledge regarding the structure of the signal (for example it is not known if it is sparse). Moreover, as we will clarify in a moment, this is the least favorable prior for this problem. The particular choice of scaling in (4.1) has been made so that the rescaled noiseless measurement $m \cdot (|\boldsymbol{A}\boldsymbol{x}_\star|)_i^2$ satisfies $m \cdot \mathbb{E}|\boldsymbol{A}\boldsymbol{x}_\star|_i^2 = 1$. We adopt the sharp high-dimensional asymptotic framework for our analysis and study a sequence of phase retrieval problems with $m, n \to \infty$, such that the oversampling ratio $\delta \overset{\text{def}}{=} m/n$ remains fixed.

[1]The results obtained in this chapter have been published in the paper R. Dudeja, J. Ma, and A. Maleki, "Information theoretic limits for phase retrieval with sub-sampled Haar sensing matrices," IEEE Transactions on Information Theory, vol. 66, no. 12, pp. 8002–8045, 2020

4.2 Main Result

Our main result is summarized in the following theorem:

Theorem 5. *For any $\delta < 2$ and for any noise level $\sigma > 0$, the Bayes risk satisfies:*

$$\lim_{\substack{m,n\to\infty \\ \frac{m}{n}=\delta}} \mathbb{E}\left\|\boldsymbol{x}_\star\boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}\left[\boldsymbol{x}_\star\boldsymbol{x}_\star^{\mathsf{H}}\middle|\boldsymbol{y},\boldsymbol{A}\right]\right\|^2 \to 1,$$

where $\|\cdot\|$ denotes the Frobenius norm.

We interpret the above result in two ways. First note that according to this theorem, for $\delta < 2$, the Bayes risk is the same as the risk of the estimator $\hat{\boldsymbol{x}} = \boldsymbol{0}$. Hence it is information theoretically impossible for any estimator to have a better performance than the trivial estimator $\hat{\boldsymbol{x}} = \boldsymbol{0}$. Second, we can make the above point more explicit as follows: Let $\hat{\boldsymbol{x}}(\boldsymbol{A},\boldsymbol{y})$ be any estimator for $\boldsymbol{x}_\star$ and let $r \geq 0$ be an arbitrary constant. By the optimality of the Bayes estimator, we have,

$$\min_{r\geq 0} \mathbb{E}\left\|\boldsymbol{x}_\star\boldsymbol{x}_\star^{\mathsf{H}} - r\frac{\hat{\boldsymbol{x}}(\boldsymbol{A},\boldsymbol{y})\hat{\boldsymbol{x}}(\boldsymbol{A},\boldsymbol{y})^{\mathsf{H}}}{\|\hat{\boldsymbol{x}}(\boldsymbol{A},\boldsymbol{y})\|^2}\right\|^2 \geq \mathbb{E}\left\|\boldsymbol{x}_\star\boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}\left[\boldsymbol{x}_\star\boldsymbol{x}_\star^{\mathsf{H}}\middle|\boldsymbol{y},\boldsymbol{A}\right]\right\|^2.$$

Taking $m, n \to \infty$ and some simple algebraic manipulations give us the following conclusion. When $\delta < 2$, then for any estimator $\hat{\boldsymbol{x}}(\boldsymbol{A},\boldsymbol{y})$ we have,

$$\lim_{\substack{m,n\to\infty \\ \frac{m}{n}=\delta}} \mathbb{E}\left[\frac{|\boldsymbol{x}_\star^{\mathsf{H}}\hat{\boldsymbol{x}}(\boldsymbol{A},\boldsymbol{y})|^2}{\|\hat{\boldsymbol{x}}(\boldsymbol{A},\boldsymbol{y})\|^2}\right] = 0.$$

That is, when $\delta < 2$, Theorem 5 provides an impossibility result: any estimator is asymptotically orthogonal to the signal vector $\boldsymbol{x}_\star$. This result complements our previous results [30, 63] which showed that the optimally designed spectral estimator is orthogonal to the signal vector in this regime. Moreover, these papers also provide the achievability result and exhibit estimators which achieve a strictly positive correlation with the signal vector when $\delta > 2$ and $\sigma = 0$. Hence, the sharp threshold for achieving a non-trivial correlation with the signal vector (called the weak recovery threshold in the literature) is $\delta_{\mathsf{weak}} = 2$ for phase retrieval with subsampled Haar sensing

matrix and vanishing measurement noise. This also shows that the uniform prior on $\boldsymbol{x}_\star$ as the least favorable prior in the following sense: The achievability results of these papers actually hold for an arbitrary signal vector (not necessarily drawn from a prior distribution). Consequently, when $\delta > 2$, for any prior on the signal vector, the Bayes risk for noiseless phase retrieval is non-trivial (< 1). Hence the uniform prior maximizes the δ threshold below which the Bayes risk is trivial and hence is least favorable.

Proof Techniques Our proof of Theorem 5 builds on the techniques of Mondelli and Montanari [20]: namely relating the Bayes risk to the Mutual Information and bounding the Mutual Information by the χ^2 divergence. However, unlike in the case of Gaussian sensing matrices, the evaluation of χ^2 divergence for our model is non trivial due to the dependence in the entries of the subsampled Haar sensing matrix. In our model, understanding the asymptotics of the χ^2 divergence reduces to understanding the asymptotics of a pair of high dimensional integrals defined on $\mathbb{S}^{m-1}$ and $\mathbb{S}^{m-1} \times \mathbb{S}^{m-1}$ (see Lemma 13) which we accomplish using Large Deviation techniques. These integrals are related to low rank Harish-Chandra-Itsker-Zuber (HCIZ) integrals studied by Guionnet and Maida [72] and our analysis is inspired by their approach. More specifically, our analysis of these integrals is based on the classical approach of Chaganty and Sethuraman [73] for obtaining strong large deviation results (i.e. results characterizing the leading exponential order as well as the second order polynomial factors in large deviation quantities of interest) using change of measure and local central limit theorems.

4.3 Some Additional Notation

In this section, we introduce some additional notations which we will find useful in this chapter.

Notations for special distributions $\mathcal{N}\left(\mu, \sigma^2\right)$ denotes the (real) Gaussian distribution with mean μ and variance σ^2. ψ_σ denotes the probability density function of $\mathcal{N}\left(0, \sigma^2\right)$:

$$\psi_\sigma(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}.$$

A random matrix $\boldsymbol{W}$ is a $\mathsf{GUE}(n)$ random matrix if it is a Hermitian $n \times n$ random matrix whose entries are sampled as follows:

$$W_{ii} \sim \mathcal{N}(0,1) \ \forall\, i \in [n],\ W_{ij} \sim \mathcal{CN}(0,1) \ \forall\, j < i,\ W_{ji} = \overline{W}_{ij} \ \forall\, j > i.$$

$\mathsf{Exp}\,(\lambda)$ denotes the exponential distribution with parameter λ which has the pdf:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & : x \geq 0 \\ 0 & : x < 0 \end{cases}.$$

$\mathsf{Gamma}\,(\alpha, \beta)$ denotes the Gamma distribution with shape parameter α and rate parameter β and has the pdf:

$$f(x; \alpha, \beta) = \begin{cases} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & : x \geq 0 \\ 0 & : x < 0 \end{cases}.$$

$\mathsf{Beta}\,(\alpha, \beta)$ denotes the Beta distribution with shape parameters $\alpha, \beta \geq 0$ which has the pdf:

$$f(x; \alpha, \beta) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1} & : x \in [0,1] \\ 0 & : x \notin [0,1] \end{cases}.$$

Let $\boldsymbol{g}_1, \boldsymbol{g}_2 \ldots, \boldsymbol{g}_n \overset{\text{i.i.d.}}{\sim} \mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{I}_p\right)$. Then the matrix $\boldsymbol{S} = \sum_{i=1}^n \boldsymbol{g}_i \boldsymbol{g}_i^{\mathsf{H}}$ has a complex Wishart distribution with parameters n, p denoted by $\mathsf{Wis}\,(n, p)$. The complex Wishart distribution is supported

on positive definite Hermitian matrices and has the pdf:

$$f(\boldsymbol{S}; n, p) = \frac{\det(\boldsymbol{S})^{n-p} \cdot e^{-\mathsf{Tr}(\boldsymbol{S})}}{\pi^{\frac{p(p-1)}{2}} \cdot \prod_{j=1}^{p}(n-j)!}.$$

The distribution Unif $(\mathbb{U}_m)$ denotes the uniform (Haar) probability measure on $\mathbb{U}(m)$.

Notation for other probabilistic aspects We will use $p(\boldsymbol{y})$ to denote the density of the measurements $\boldsymbol{y}$ with respect to the Lebesgue measure. Likewise $p(\boldsymbol{y}|\boldsymbol{A})$ and $p(\boldsymbol{y}|\boldsymbol{A}, \boldsymbol{x})$ denote the conditional density of the measurements $\boldsymbol{y}$ given the measurement matrix $\boldsymbol{A}$ and the conditional density of the $\boldsymbol{y}$ given the measurement matrix $\boldsymbol{A}$ and the signal vector $\boldsymbol{x}$ respectively.

Notation for Information Theoretic Aspects For random variables $A_1, A_2, \ldots A_k$, we denote the entropy of $(A_1 \ldots, A_k)$ by $\mathbf{H}(A_1, A_2 \ldots, A_k)$. If $(A_1, A_2 \ldots A_k)$ have a joint density $p(a_1, a_2 \ldots, a_k)$ with respect to the Lebesgue measure, this is defined as:

$$\mathbf{H}(A_{1:k}) = -\int_{\mathbb{R}^k} p(a_{1:k}) \ln p(a_{1:k}) \, \mathrm{d}a_{1:k}.$$

Let $B_1, B_2 \ldots B_l$ be another collection of random variables. We denote the conditional entropy of $(A_1, \ldots A_k)$ given $(B_1, \ldots, B_l)$ by $\mathbf{H}\left(A_1, A_2 \ldots, A_k \mid B_1, B_2 \ldots, B_l\right)$. When the conditional distribution of $(A_1, A_2 \ldots, A_k)$ given $(B_1, B_2 \ldots, B_l)$ has a density $p(a_1, a_2 \ldots, a_k|b_1, b_2 \ldots, b_l)$ (with respect to Lebesgue measure) and $(B_1, B_2 \ldots, B_l)$ has a marginal density $p(b_1, b_2 \ldots, b_l)$ (with respect to Lebesgue measure), then $\mathbf{H}\left(A_{1:k} \mid B_{1:l}\right)$ is given by:

$$\mathbf{H}\left(A_1, \ldots A_k \mid B_1 \ldots, B_l\right) = \int_{\mathbb{R}^l} p(b_{1:l}) \int_{\mathbb{R}^k} p(a_{1:k}|b_1 \ldots, b_l) \ln p(a_1, \ldots a_k|b_1, \ldots b_l) \, \mathrm{d}a_{1:k} \, \mathrm{d}b_{1:l}.$$

The mutual information between $A_1, A_2, \ldots A_k$ and $B_1, B_2 \ldots B_l$ is denoted $\mathbf{I}(A_1, \ldots A_k; B_1, \ldots B_l)$ and is defined by the following equivalent formulae:

$$\begin{aligned}\mathbf{I}(A_1, \ldots A_k; B_1, \ldots B_l) &\stackrel{\text{def}}{=} \mathbf{H}(A_1, \ldots A_k) - \mathbf{H}\left(A_1, \ldots A_k \mid B_1, \ldots B_l\right) \\ &= \mathbf{H}(B_1, \ldots B_l) - \mathbf{H}\left(B_1, \ldots B_l \mid A_1, \ldots A_k\right).\end{aligned}$$

The random variable Y We reserve the random variable Y to denote the random variable with one of the following two special distributions:

1. Y can be sampled from the empirical distribution of the phase retrieval measurements:

$$Y \sim \frac{1}{m}\sum_{i=1}^{m} \delta_{y_i}.$$

   For any $f : \mathbb{R} \mapsto \mathbb{R}$, we define $\hat{\mathbb{E}} f(Y)$ to be the expectation of $f(Y)$ with respect to the empirical measure of the measurements:

$$\hat{\mathbb{E}} f(Y) \stackrel{\text{def}}{=} \frac{1}{m}\sum_{i=1}^{m} f(y_i). \tag{4.3}$$

2. Alternatively the distribution of Y can be given by $Y = |Z|^2 + \sigma\epsilon$ where $Z \sim \mathcal{CN}(0, 1)$, $\epsilon \sim \mathcal{N}(0, 1)$. For any $f : \mathbb{R} \mapsto \mathbb{R}$, we define $\mathbb{E} f(Y)$ denotes the expectation of $f(Y)$ with respect to this measure, that is, $\mathbb{E} f(Y) = \mathbb{E} f(|Z|^2 + \sigma\epsilon)$. This special distribution is important to us because we will see that for a large class of test functions f, $\hat{\mathbb{E}} f(Y) \to \mathbb{E} f(Y)$ as $m \to \infty$.

Notations for linear algebraic aspects: For a 2×2 Hermitian matrix $\boldsymbol{A}$ we define the $\mathsf{Vec}\,(\cdot)$ operation by:

$$\mathsf{Vec}\,(\boldsymbol{A}) = \begin{bmatrix} A_{11} \\ A_{22} \\ \mathrm{Re}(A_{12}) \\ \mathrm{Im}(A_{12}) \end{bmatrix}.$$

4.4 Organization of the Proof

The remainder of this chapter is dedicated to proving Theorem 5. The proof consists of different steps which are split into various sections as follows:

In Section 4.5, we relate the Bayes risk to the Mutual Information for the phase retrieval problem with a small amount of side information and show that if the mutual information is $o(m)$, then the asymptotic Bayes risk is trivial. Hence, our focus shifts to showing that when $\delta < 2$, the Mutual information is $o(m)$. We then bound the mutual information by the χ^2 divergence. Understanding the χ^2 divergence in the Phase retrieval model requires us to understand the asymptotics of two high dimensional integrals denoted by $\mathscr{L}$ and $\mathscr{U}$ on $\mathbb{S}^{m-1}$ and $\mathbb{S}^{m-1} \times \mathbb{S}^{m-1}$ respectively.

In Section 4.6, we study the asymptotics of the integrals $\mathscr{U}, \mathscr{L}$ by change of measure techniques and local central limit theorems.

In Section 4.7, we use a stochastic version of the Laplace Principle along with the asymptotics of $\mathscr{U}, \mathscr{L}$ to understand the asymptotics of the χ^2 divergence. This results in a explicit condition on the sampling ratio δ and the noise level σ which guarantees that the mutual information is $o(m)$ and hence the Bayes risk is trivial.

In Section 4.8, we simplify the condition on δ, σ obtained previously in the low noise limit $\sigma \to 0$.

4.5 Mutual Information and Bayes Risk

We first relate the Bayes risk to the mutual information in the phase retrieval problem where one observes a small amount of side information about the signal vector $\boldsymbol{x}_\star$. The amount of side information we observe will be controlled by a parameter $\Delta > 0$ which will be a constant independent of n, m. The side information we observe will be linear gaussian measurements of the matrix $\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}}$. More precisely, for $i = 1, 2 \dots \lfloor \Delta \cdot m \rfloor$ we observe a measurement pair $(\boldsymbol{w}_i, z_i)$ drawn from the following model:

$$\boldsymbol{w}_i \overset{\text{i.i.d.}}{\sim} \mathsf{GUE}(n),\ z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\langle \boldsymbol{w}_i, \boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} \rangle, 1\right)\ \forall\, i = 1, 2, \dots \lfloor \Delta \cdot m \rfloor. \tag{4.4}$$

We collect all the side information measurements z_i's in a vector $\boldsymbol{z} \in \mathbb{R}^{\lfloor \Delta m \rfloor}$. We denote the collection of the GUE sensing matrices by $\boldsymbol{W} \overset{\text{def}}{=} \{\boldsymbol{w}_1, \boldsymbol{w}_2 \dots \boldsymbol{w}_{\lfloor \Delta m \rfloor}\}$. The following proposition establishes the connection between the Bayes Risk and $\mathbf{I}\left(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}\right)$.

Proposition 6. *Suppose that there exists a constant $\Delta > 0$ (independent of m, n) such that the mutual information $\mathbf{I}\left(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}\right) = o(m)$. Then we have,*

$$\lim_{\substack{m,n \to \infty \\ m = n\delta}} \mathbb{E}_{\boldsymbol{x}_\star, \boldsymbol{y}, \boldsymbol{A}} \| \boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}[\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} | \boldsymbol{y}, \boldsymbol{A}] \|^2 = 1.$$

In light of Proposition 6, in order to show that the Bayes risk is trivial, it is sufficient to show that an upper bound on the mutual information is $o(m)$. We will use the second moment upper bound (or the χ^2-divergence uppper bound) on mutual information. This upper bound was utilized by Mondelli and Montanari [20] for determining the weak recovery threshold for Gaussian sensing matrices. In our setup, the result of these authors can be stated as:

$$\mathbf{I}\left(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}\right) \leq \mathbb{E}_{\boldsymbol{y}, \boldsymbol{z}} \left[\frac{\mathbb{E}_{\boldsymbol{A}, \boldsymbol{W}} p^2(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W})}{p^2(\boldsymbol{y}, \boldsymbol{z})} \right] - 1.$$

It is also well known that the second moment upper bound is sensitive to bad but rare events that can

cause the upper bound to blow up. In order to exclude these bad events we will use a conditional version of the above bound which is stated below. A similar result was used by Reeves, Xu, and Zadik [74] in the context of a linear regression problem. The proof of this result is given in Appendix B.1.2.

Lemma 12. *Let $\mathcal{E}_m$ be any sequence of events depending only $\boldsymbol{y}$. We have,*

$$\mathbf{I}\left(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}\right) \leq \left(\int_{\mathcal{E}_m} \frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W})}{p(\boldsymbol{y}, \boldsymbol{z})} \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} - 1 \right) + C \cdot m \cdot \sqrt{\mathbb{P}(\mathcal{E}_m^c)}.$$

In the above display, $C \geq 0$ denotes a finite constant depending only on δ, Δ, σ^2.

The following lemma simplifies the upper bound on $\mathbf{I}\left(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}\right)$. For any $\boldsymbol{y} \in \mathbb{R}^m$ and any positive semidefinite 2×2 Hermitian matrix $\boldsymbol{Q}$, introduce the functions:

$$\mathscr{U}\left(\boldsymbol{y}, \boldsymbol{Q}\right) \stackrel{\text{def}}{=} \mathbb{E}\left[\prod_{i=1}^{m} \psi_\sigma(y_i - |G_{1i}|^2) \psi_\sigma(y_i - |G_{2i}|^2) \middle| \boldsymbol{G}^{\mathsf{H}} \boldsymbol{G} = m \boldsymbol{Q} \right], \tag{4.5}$$

$$\mathscr{L}\left(\boldsymbol{y}\right) \stackrel{\text{def}}{=} \mathbb{E}\left[\prod_{i=1}^{m} \psi_\sigma(y_i - |G_{1i}|^2) \middle| \|\boldsymbol{G}_1\|^2 = m \right], \tag{4.6}$$

where $\boldsymbol{G}_1, \boldsymbol{G}_2 \overset{\text{i.i.d.}}{\sim} \mathcal{CN}\left(\boldsymbol{0}, \boldsymbol{I}_m\right)$ and the matrix $\boldsymbol{G} = [\boldsymbol{G}_1 \ \boldsymbol{G}_2]$. We emphasize that in the definitions of $\mathscr{U}\left(\boldsymbol{y}, \boldsymbol{Q}\right)$ and $\mathscr{L}\left(\boldsymbol{y}\right)$, the measurements $\boldsymbol{y}$ are fixed, and the expectation is only with respect to the Gaussian matrix $\boldsymbol{G}$.

Lemma 13. *We have,*

$$\int_{\mathcal{E}_m} \frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W})}{p(\boldsymbol{y}, \boldsymbol{z})} \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} = \frac{2}{n-1} \mathbb{E}_{\boldsymbol{y}} \left[\frac{\int_0^1 \mathscr{U}\left(\boldsymbol{y}, \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix} \right) \cdot \frac{q \cdot (1-q^2)^{n-2}}{(1-q^2/2)^{\lfloor \Delta m \rfloor}} \, \mathrm{d}q}{\mathscr{L}^2(\boldsymbol{y})} \cdot \mathbf{1}_{\mathcal{E}_m} \right]$$

Proof. We have,

$$\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{A},\boldsymbol{W}) = \mathbb{E}_{\boldsymbol{A},\boldsymbol{W},\boldsymbol{x},\boldsymbol{x}'} p(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{A},\boldsymbol{W},\boldsymbol{x}) p(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{A},\boldsymbol{W},\boldsymbol{x}')$$
$$= \mathbb{E}\left[\prod_{i=1}^{m} \psi_\sigma(y_i - m|\langle \boldsymbol{a}_i,\boldsymbol{x}\rangle|^2)\psi_\sigma(y_i - m|\langle \boldsymbol{a}_i,\boldsymbol{x}'\rangle|^2) \prod_{i=1}^{\lfloor \Delta m \rfloor} \psi_1(z_i - \langle \boldsymbol{w}_i, \boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}\rangle)\psi_1(z_i - \langle \boldsymbol{w}_i, \boldsymbol{x}'\boldsymbol{x}'^{\mathsf{H}}\rangle)\right]$$

Define the scalar random variable:

$$q = \boldsymbol{x}^{\mathsf{H}}\boldsymbol{x}',$$

and the associated random matrices:

$$\boldsymbol{Q} = \begin{bmatrix} 1 & q \\ \overline{q} & 1 \end{bmatrix}, \ \boldsymbol{C} = \begin{bmatrix} 1 & q \\ 0 & \sqrt{1-|q|^2} \end{bmatrix}$$

Note that we have $\boldsymbol{C}^{\mathsf{H}}\boldsymbol{C} = \boldsymbol{Q}$. It is easy to see that, conditioned on $\boldsymbol{x}, \boldsymbol{x}'$:

$$\begin{bmatrix} \langle \boldsymbol{w}_i, \boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}\rangle \\ \langle \boldsymbol{w}_i, \boldsymbol{x}'\boldsymbol{x}'^{\mathsf{H}}\rangle \end{bmatrix} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} 1 & |q|^2 \\ |q|^2 & 1 \end{bmatrix}\right),$$
$$\boldsymbol{A}\boldsymbol{x}, \boldsymbol{A}\boldsymbol{x}' \overset{\text{d}}{=} \boldsymbol{U}_1, q\boldsymbol{U}_1 + \sqrt{1-|q|^2}\boldsymbol{U}_2.$$

In the above display, $\boldsymbol{U} = [\boldsymbol{U}_1 \ \boldsymbol{U}_2]$ is a uniformly random $m \times 2$ partial unitary matrix. By the rotational invariance of $\boldsymbol{U}$, we have,

$$\sqrt{m} \cdot \begin{bmatrix} \boldsymbol{A}\boldsymbol{x} & \boldsymbol{A}\boldsymbol{x}' \end{bmatrix} \overset{\text{d}}{=} \sqrt{m} \cdot \boldsymbol{U}\boldsymbol{C}$$
$$= \boldsymbol{U}\boldsymbol{C}\boldsymbol{Q}^{-1/2}(m \cdot \boldsymbol{Q})^{1/2}$$
$$\overset{\text{d},(1)}{=} \boldsymbol{U}(m\boldsymbol{Q})^{1/2}.$$

In the step marked (1), we used the fact that $\boldsymbol{C}\boldsymbol{Q}^{-1/2}$ is unitary consequently $\boldsymbol{U}\boldsymbol{C}\boldsymbol{Q}^{-1/2} \overset{\text{d}}{=} \boldsymbol{U}$. Let

$\boldsymbol{G}$ be a $m \times 2$ matrix consisting of $\mathcal{CN}(0,1)$ entries. Then we have,

$$\begin{aligned}\sqrt{m} \cdot \begin{bmatrix} \boldsymbol{A}\boldsymbol{x} & \boldsymbol{A}\boldsymbol{x}' \end{bmatrix} &\stackrel{\text{d}}{=} \boldsymbol{U}(m\boldsymbol{Q})^{1/2} \\ &\stackrel{\text{d}_{(2)}}{=} \boldsymbol{G}(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G})^{-1/2}(m \cdot \boldsymbol{Q})^{1/2} \\ &\stackrel{\text{d}_{(3)}}{=} \boldsymbol{G}(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G})^{-1/2}(m \cdot \boldsymbol{Q})^{1/2} | \boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} = m\boldsymbol{Q} \\ &= \boldsymbol{G} | \boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} = m\boldsymbol{Q}.\end{aligned}$$

In step (2) we used the well known fact that a uniformly random partial unitary matrix can be realized as $\boldsymbol{U} \stackrel{\text{d}}{=} \boldsymbol{G}(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G})^{-1/2}$. In the step marked (3) we used the fact that $\boldsymbol{G}(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G})^{-1/2}$ is independent of $\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G}$, and hence conditioning on the event $\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G}$ does not change the distribution of $\boldsymbol{G}(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G})^{-1/2}$. Hence we have shown that, conditioned on $\boldsymbol{x}, \boldsymbol{x}'$, the matrix $\sqrt{m} \cdot [\boldsymbol{A}\boldsymbol{x}\ \boldsymbol{A}\boldsymbol{x}']$ has the same distribution as a Gaussian matrix $\boldsymbol{G}$ conditioned on the event $\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} = m\boldsymbol{Q}$:

$$\sqrt{m} \cdot \begin{bmatrix} \boldsymbol{A}\boldsymbol{x} & \boldsymbol{A}\boldsymbol{x}' \end{bmatrix} \stackrel{\text{d}}{=} \boldsymbol{G} | \boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} = m\boldsymbol{Q}.$$

Hence we have,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}) &= \mathbb{E}_{\boldsymbol{A},\boldsymbol{W},\boldsymbol{x},\boldsymbol{x}'} p(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}, \boldsymbol{x}) p(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}, \boldsymbol{x}') \\ &= \mathbb{E}_q \left[\Psi_1(q; \boldsymbol{y}) \cdot \Psi_2(q; \boldsymbol{z}) \right].\end{aligned}$$

In the above display, we defined,

$$\Psi_1(q; \boldsymbol{y}) \stackrel{\text{def}}{=} \mathbb{E}\left[\prod_{i=1}^{m} \psi_\sigma(y_i - |G_{1i}|^2) \psi_\sigma(y_i - |G_{2i}|^2) \middle| \boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} = m\boldsymbol{Q} \right],$$

and,

$$\Psi_2(q; \boldsymbol{z}) \stackrel{\text{def}}{=} \mathbb{E}_{Z,Z'} \psi_1(z_i - Z) \psi_1(z_i - |q|^2 Z - \sqrt{1 - |q|^4} Z'),$$

where $Z, Z' \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. We observe that the conditional expectation:

$$\mathbb{E}\left[\prod_{i=1}^{m} \psi_\sigma(y_i - |G_{1i}|^2)\psi_\sigma(y_i - |G_{2i}|^2)\middle| \boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} = m\boldsymbol{Q}\right],$$

depends on q only via $|q|$. Consequently, we redefine the matrix $\boldsymbol{Q}$ as:

$$\boldsymbol{Q} = \begin{bmatrix} 1 & |q| \\ |q| & 1 \end{bmatrix}.$$

The following integral has been evaluated in Lemma 46 in Appendix B.9.

$$\mathbb{E}_{Z,Z'}\psi_1(z - Z)\psi_1(z - |q|^2 Z - \sqrt{1-|q|^4}Z') = \frac{1}{4\pi\sqrt{1-|q|^4/4}} \exp\left(-\frac{z^2}{2(1+|q|^2/2)}\right).$$

Hence,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}) \\
&\quad = \mathbb{E}_q\left[\mathbb{E}\left[\prod_{i=1}^{m} \psi_\sigma(y_i - |G_{1i}|^2)\psi_\sigma(y_i - |G_{2i}|^2)\middle| \boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} = m\boldsymbol{Q}\right] \cdot \frac{e^{-\frac{1}{2(1+|q|^2/2)}\cdot\sum_{i=1}^{\lfloor\Delta m\rfloor} z_i^2}}{(4\pi\sqrt{1-|q|^4/4})^{\lfloor\Delta m\rfloor}}\right].
\end{aligned}$$

Next we compute $p(\boldsymbol{y}, \boldsymbol{z})$. Since $\boldsymbol{y}$ and $\boldsymbol{z}$ are independent, $p(\boldsymbol{y}, \boldsymbol{z}) = p(\boldsymbol{y})p(\boldsymbol{z})$. $p(\boldsymbol{y})$ can be computed by following similar steps as before:

$$p(\boldsymbol{y}) = \mathbb{E}\left[\prod_{i=1}^{m} \psi_\sigma(y_i - |G_{1i}|^2)\middle| \|\boldsymbol{G}_1\|^2 = m\right].$$

It is also easy to check that $z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,2)$. Hence:

$$p(\boldsymbol{y}, \boldsymbol{z}) = \mathbb{E}\left[\prod_{i=1}^{m} \psi_\sigma(y_i - |G_{1i}|^2)\middle| \|\boldsymbol{G}_1\|^2 = m\right] \cdot \prod_{i=1}^{\lfloor\Delta m\rfloor} \psi_{\sqrt{2}}(z_i).$$

Consequently, introducing the functions:

$$\mathscr{U}(\boldsymbol{y},\boldsymbol{R}) \stackrel{\text{def}}{=} \mathbb{E}\left[\prod_{i=1}^{m}\psi_\sigma(y_i-|G_{1i}|^2)\psi_\sigma(y_i-|G_{2i}|^2)\middle|\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G}=m\boldsymbol{R}\right],$$
$$\mathscr{L}(\boldsymbol{y}) \stackrel{\text{def}}{=} \mathbb{E}\left[\prod_{i=1}^{m}\psi_\sigma(y_i-|G_{1i}|^2)\middle|\|\boldsymbol{G}_1\|^2=m\right].$$

we obtain,

$$\mathbb{E}_{\boldsymbol{z}}\frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}}p^2(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{A},\boldsymbol{W})}{p^2(\boldsymbol{y},\boldsymbol{z})} = \mathbb{E}_q\left[\frac{\mathscr{U}\left(\boldsymbol{y},\begin{bmatrix}1 & |q|\\ |q| & 1\end{bmatrix}\right)}{\mathscr{L}^2(\boldsymbol{y})}\cdot\left(\frac{\mathbb{E}_{Z\sim\mathcal{N}(0,2)}\exp\left(\frac{\frac{|q|^2}{2}}{1+\frac{|q|^2}{2}}\cdot\frac{Z^2}{2}\right)}{\sqrt{1-\frac{|q|^4}{4}}}\right)^{\lfloor\Delta m\rfloor}\right]$$
$$\stackrel{\text{(a)}}{=} \mathbb{E}_q\left[\frac{\mathscr{U}\left(\boldsymbol{y},\begin{bmatrix}1 & |q|\\ |q| & 1\end{bmatrix}\right)}{\mathscr{L}^2(\boldsymbol{y})}\cdot\frac{1}{(1-|q|^2/2)^{\Delta m}}\right].$$

In the step marked (a), we used the MGF of χ^2 distribution to compute:

$$\mathbb{E}_{Z\sim\mathcal{N}(0,2)}\exp\left(\frac{\frac{|q|^2}{2}}{1+\frac{|q|^2}{2}}\cdot\frac{Z^2}{2}\right)=\sqrt{\frac{1+|q|^2/2}{1-|q|^2/2}}$$

Hence we have,

$$\int_{\mathcal{E}_m} \frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{A},\boldsymbol{W})}{p(\boldsymbol{y},\boldsymbol{z})}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{z} = \mathbb{E}_{\boldsymbol{y},|q|}\left[\frac{\mathscr{U}\left(\boldsymbol{y},\begin{bmatrix}1 & |q|\\ |q| & 1\end{bmatrix}\right)}{\mathscr{L}^2(\boldsymbol{y})}\cdot\frac{1}{(1-|q|^2/2)^{\lfloor\Delta m\rfloor}}\cdot\mathbf{1}_{\mathcal{E}_m}\right]$$

Next we observe that,

$$|q|^2 \sim \mathsf{Beta}\,(1, n-1)\,.$$

Utilizing the formula for the pdf of Beta random variables we have,

$$\int_{\mathcal{E}_m} \frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{A},\boldsymbol{W})}{p(\boldsymbol{y},\boldsymbol{z})}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{z} = \frac{1}{n-1}\mathbb{E}_{\boldsymbol{y}}\left[\frac{\int_0^1 \mathscr{U}\left(\boldsymbol{y},\begin{bmatrix}1 & \sqrt{b}\\ \sqrt{b} & 1\end{bmatrix}\right)\cdot\frac{(1-b)^{n-2}}{(1-b/2)^{\lfloor\Delta m\rfloor}}\,\mathrm{d}b}{\mathscr{L}^2(\boldsymbol{y})}\cdot\mathbf{1}_{\mathcal{E}_m}\right].$$

Finally making the change of variable $b = q^2$ gives us:

$$\int_{\mathcal{E}_m} \frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{A},\boldsymbol{W})}{p(\boldsymbol{y},\boldsymbol{z})}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{z} \le \frac{2}{n-1}\mathbb{E}_{\boldsymbol{y}}\left[\frac{\int_0^1 \mathscr{U}\left(\boldsymbol{y},\begin{bmatrix}1 & q\\ q & 1\end{bmatrix}\right)\cdot\frac{q\cdot(1-q^2)^{n-2}}{(1-q^2/2)^{\Delta m}}\,\mathrm{d}q}{\mathscr{L}^2(\boldsymbol{y})}\cdot\mathbf{1}_{\mathcal{E}_m}\right].$$

□

Remark 12. *At this point, it is instructive to compare the claim of Lemma 13 to its counterpart from [20]. If $\boldsymbol{A}$ were Gaussian, then, Mondelli and Montanari [20] have shown that,*

$$\int \frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{A},\boldsymbol{W})}{p(\boldsymbol{y},\boldsymbol{z})}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{z} = \frac{2}{n-1}\mathbb{E}_{\boldsymbol{y}}\left[\frac{\int_0^1 \mathscr{U}_{\mathsf{Gauss}}\left(\boldsymbol{y},\begin{bmatrix}1 & q\\ q & 1\end{bmatrix}\right)\cdot\frac{q\cdot(1-q^2)^{n-2}}{(1-q^2/2)^{\lfloor\Delta m\rfloor}}\,\mathrm{d}q}{\mathscr{L}^2_{\mathsf{Gauss}}(\boldsymbol{y})}\cdot\right], \tag{4.7}$$

where the functions $\mathscr{U}_{\mathsf{Gauss}}$ and $\mathscr{L}_{\mathsf{Gauss}}$ are defined as follows:

$$\mathscr{U}_{\mathsf{Gauss}}\left(\boldsymbol{y},\begin{bmatrix}1 & q\\ q & 1\end{bmatrix}\right) \stackrel{def}{=} \mathbb{E}\left[\prod_{i=1}^m \psi_\sigma(y_i - |G_{1i}|^2)\psi_\sigma(y_i - |qG_{1i} + \sqrt{1-q^2}G_{2i}|^2)\right],$$

$$\mathscr{L}_{\mathsf{Gauss}}(\boldsymbol{y}) \stackrel{def}{=} \mathbb{E}\left[\prod_{i=1}^m \psi_\sigma(y_i - |G_{1i}|^2)\right]$$

Because the conditioning is absent in the definitions of $\mathscr{U}_{\mathsf{Gauss}}$ and $\mathscr{L}_{\mathsf{Gauss}}$, one can leverage the independence in $\boldsymbol{G}_1, \boldsymbol{G}_2$ and obtain straightforwardly:

$$\frac{\mathscr{U}_{\mathsf{Gauss}}\left(\boldsymbol{y},\begin{bmatrix}1 & q\\ q & 1\end{bmatrix}\right)}{\mathscr{L}^2_{\mathsf{Gauss}}(\boldsymbol{y})} = \prod_{i=1}^m \frac{\mathbb{E}_{G_1,G_2}\left[\psi_\sigma(y_i - |G_1|^2)\psi_\sigma(y_i - |qG_1 + \sqrt{1-q^2}G_2|^2)\right]}{\mathbb{E}^2_G\left[\psi_\sigma(y_i - |G|^2)\right]}.$$

Furthermore when the sensing matrix is Gaussian, the observations $y_1, y_2 \ldots y_m$ are i.i.d. Let Y be a random variable with the same distribution as y_i. The expression in (4.7) *simplifies significantly:*

$$\int \frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{A},\boldsymbol{W})}{p(\boldsymbol{y},\boldsymbol{z})}\,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{z} = \frac{2}{n-1}\int_0^1 F_{\mathsf{Gauss}}(q)^m \cdot \frac{q\cdot(1-q^2)^{n-2}}{(1-q^2/2)^{\lfloor\Delta m\rfloor}}\,\mathrm{d}q, \tag{4.8}$$

where,

$$F_{\mathsf{Gauss}}(q) \stackrel{def}{=} \mathbb{E}_Y \left[\frac{\mathbb{E}_{G_1,G_2}\left[\psi_\sigma(Y-|G_1|^2)\psi_\sigma(Y-|qG_1+\sqrt{1-q^2}G_2|^2)\right]}{\mathbb{E}_G^2\left[\psi_\sigma(Y-|G|^2)\right]} \right].$$

Mondelli and Montanari [20] analyze the integral in 4.8 by a straightforward application of the Laplace Principle. Note that this whole approach breaks down in our case because the conditioning in the definition of $\mathscr{U}$, $\mathscr{L}$ introduces dependence between the Gaussian random vectors $\boldsymbol{G}_1, \boldsymbol{G}_2$ and their entries. This dependence is a manifestation of the dependence present in a subsampled Haar unitary matrix.

4.6 Asymptotic Analysis of $\mathscr{L}$ and $\mathscr{U}$

In order to evaluate the upper bound on the mutual information that is given in Lemma 13, one needs to understand the asymptotic behaviour of the functions $\mathscr{L}$ and $\mathscr{U}$ introduced in Lemma 13.

4.6.1 Analysis of $\mathscr{L}$

Recall that $\mathscr{L}(\boldsymbol{y})$ was defined as:

$$\mathscr{L}(\boldsymbol{y}) \stackrel{\text{def}}{=} \mathbb{E}\left[\prod_{i=1}^{m} \psi_\sigma(y_i - |G_{1i}|^2) \middle| \|\boldsymbol{G}_1\|^2 = m \right].$$

We can rewrite $\mathscr{L}(\boldsymbol{y})$ as follows:

$$\mathscr{L}(\boldsymbol{y}) \stackrel{\text{def}}{=} \mathbb{E}\left[\exp\left(\sum_{i=1}^{m} \ln \psi_\sigma(y_i - |G_{1i}|^2) \right) \middle| \frac{1}{m}\|\boldsymbol{G}_1\|^2 = 1 \right].$$

The above equation suggests that the asymptotics of $\mathscr{L}$ are determined by the large deviation properties of the random variables:

$$\left(\frac{1}{m}\sum_{i=1}^{m}\ln\psi_\sigma(y_i-|G_{1i}|^2),\frac{1}{m}\|\boldsymbol{G}_1\|^2\right). \tag{4.9}$$

Note that the random variables in the display above are a sum of independent random variables. In our analysis we treat $\boldsymbol{y}$ as a fixed vector in $\mathbb{R}^m$ and only leverage the randomness in $\boldsymbol{G}_1$. Consequently, the two random variables in (4.9) are sums of independent, but not identically distributed random variables. This makes our analysis a bit delicate. Large deviation theory tells us that the Cramer Transform plays a crucial role in understanding the large deviations of sums of independent random variables. Hence, we define the Tilted Exponential distribution which is the Cramer Transform (or the exponential tilting) of the pair of random variables $(\ln\psi_\sigma(y-|G|^2),|G|^2)$ where $G\sim\mathcal{CN}(0,1)$ and $y\in\mathbb{R}$ is a fixed scalar below.

Definition 5 (The Tilted Exponential Distribution)**.** *The Tilted Exponential distribution with parameters* (λ,y) *denoted by* $\mathsf{TExp}(\lambda,y)$ *is the distribution on* $[0,\infty)$ *with the pdf:*

$$f(u)=\frac{e^{-(1-\lambda)u}\psi_\sigma(u-y)}{Z_{\mathsf{TExp}}(\lambda,y)},$$

where, $Z_{\mathsf{TExp}}(\lambda,y)$ *denotes the normalizing constant:*

$$Z_{\mathsf{TExp}}(\lambda,y)\overset{def}{=}\int_0^\infty e^{-(1-\lambda)u}\psi_\sigma(u-y)\,\mathrm{d}u=\mathbb{E}_{E\sim\mathsf{Exp}(1)}e^{\lambda E}\psi_\sigma(E-y).$$

We also denote the variance of $\mathsf{TExp}(\lambda,y)$ *by* $\sigma^2_{\mathsf{TExp}}(\lambda,y)$.

In Appendix B.6.1 we prove some essential properties of the Tilted Exponential distribution which will be useful in our analysis.

The analysis of $\mathscr{L}(\boldsymbol{y})$ uses two standard techniques from large deviation theory: performing an exponential change of measure and then applying a central limit theorem under the tilted measure.

The following lemma is a change of measure result that we use in our analysis. In order to state it we first introduce some notation. Fix any $\boldsymbol{y} \in \mathbb{R}^m$ and any $\lambda \in \mathbb{R}$. Let $u_1, u_2 \ldots u_m$ be independent non-negative random variables with $u_i \sim \mathsf{TExp}(\lambda, y_i)$. Let $F_{\lambda,\boldsymbol{y}}$ be the density of the random variable $\sum_{i=1}^m u_i$.

Lemma 14. *For any $\lambda \in \mathbb{R}, \boldsymbol{y} \in \mathbb{R}^m$ we have,*

$$\mathscr{L}(\boldsymbol{y}) = \frac{(m-1)! \cdot e^{m(1-\lambda)} \cdot F_{\lambda,\boldsymbol{y}}(m)}{m^{m-1}} \cdot \prod_{i=1}^m Z_{\mathsf{TExp}}(\lambda, y_i).$$

In the above display, $F_{\lambda,\boldsymbol{y}}$ is the density of the random variable $\sum_{i=1}^m u_i$ where the random variables u_i are sampled independently with marginal distribution $u_i \sim \mathsf{TExp}(\lambda, y_i)$.

Proof. Define the random variables:

$$U = \sum_{i=1}^m u_i, \; T = \sum_{i=1}^m \ln \psi_\sigma(y_i - u_i).$$

Consider two possible probability distributions for U and T:

1. u_i are i.i.d. $\mathsf{Exp}(1)$. Let $G(u,t)$ be the joint pdf of U and T in this setup.

2. u_i are sampled independently from $\mathsf{TExp}(\lambda, y_i)$ defined in the statement of the lemma. Let $F_{\lambda,\boldsymbol{y}}(u,t)$ denote the joint pdf of U, T in this setup.

We can compute $F_{\lambda,\boldsymbol{y}}(u,t)$ in terms of $G(u,t)$ in the following way:

$$F_{\lambda,\boldsymbol{y}}(u,t) = \frac{\exp(t+\lambda u)}{\prod_{i=1}^m Z_{\mathsf{TExp}}(\lambda, \boldsymbol{y}_i)} \cdot G(u,t). \tag{4.10}$$

Let $G(t|u)$ denote the conditional density of T given $U = u$ and $G(u)$ denote the marginal density of U under Setup 1. Analogously define $F_{\lambda,\boldsymbol{y}}(t|u)$ and $F_{\lambda,\boldsymbol{y}}(u)$. We can then compute $\mathscr{L}(\boldsymbol{y})$ as

follows:

$$\begin{aligned}
\mathscr{L}(\boldsymbol{y}) &= \mathbb{E}_{\boldsymbol{g}\sim\mathcal{CN}(\mathbf{0},\boldsymbol{I}_m)}\left[\exp\left(\sum_{i=1}^{m}\ln\phi_\sigma(y_i-|g_i|^2)\right)\Big|\|\boldsymbol{g}\|^2=m\right] \\
&= e^{-m\lambda}\mathbb{E}\left[\exp\left(\sum_{i=1}^{m}\ln\phi_\sigma(y_i-|g_i|^2)+\lambda m\right)\Big|\|\boldsymbol{g}\|^2=m\right] \\
&\overset{\text{(a)}}{=} e^{-m\lambda}\int e^{t+m\lambda}G(t|m)\,\mathrm{d}t \\
&= \frac{e^{-m\lambda}}{G(m)}\int e^{t+m\lambda}G(m,t)\,\mathrm{d}t.
\end{aligned}$$

In the step marked (a), we used the fact that if $G \sim \mathcal{CN}(0,1)$, then $|G|^2 \sim \mathsf{Exp}(1)$. Next, appealing to (4.10), we obtain:

$$\begin{aligned}
\mathscr{L}(\boldsymbol{y}) &\overset{\text{(b)}}{=} \frac{e^{-m\lambda}\prod_{i=1}^{m}Z_{\mathsf{TExp}}(\lambda,y_i)}{G(m)}\int F_{\lambda,\boldsymbol{y}}(m,t)\,\mathrm{d}t \\
&= \frac{F_{\lambda,\boldsymbol{y}}(m)e^{-m\lambda}}{G(m)}\cdot\prod_{i=1}^{m}Z_{\mathsf{TExp}}(\lambda,y_i) \\
&\overset{\text{(c)}}{=} \frac{(m-1)!\cdot e^{m(1-\lambda)}\cdot F_{\lambda,\boldsymbol{y}}(m)}{(m)^{m-1}}\cdot\prod_{i=1}^{m}Z_{\mathsf{TExp}}(\lambda,y_i)\,.
\end{aligned}$$

The equality marked (b) follows from (4.10). In the step (c), we used the fact that under Setup 1, U is a sum of exponential random variables and hence $U \sim \mathsf{Gamma}(m,1)$. Therefore the density of the Gamma distribution can be used to evaluate $G(m)$. This proves the claim of the lemma. □

Our next step will be to develop the asymptotics of $F_{\lambda,\boldsymbol{y}}$ by means of a local CLT. Note that in Lemma 14, $\lambda \in \mathbb{R}$ was arbitrary. We will set $\lambda = \hat{\lambda}_1(\sigma)$, where

$$\hat{\lambda}_1(\sigma) \overset{\text{def}}{=} \arg\max_{\lambda\in\mathbb{R}}\left(\lambda - \hat{\mathbb{E}}_Y \ln \mathbb{E}_{E\sim\mathsf{Exp}(1)}e^{\lambda E}\psi_\sigma(E-Y)\right). \tag{4.11}$$

We also define,

$$\hat{\Xi}_1(\sigma) \stackrel{\text{def}}{=} \max_{\lambda \in \mathbb{R}} \left(\lambda - \hat{\mathbb{E}}_Y \ln \mathbb{E}_{E \sim \mathsf{Exp}(1)} e^{\lambda E} \psi_\sigma(E - Y) \right). \tag{4.12}$$

The notation $\hat{\mathbb{E}}$ in the above display, has been introduced in (4.3). Note that the above quantities depend on the vector $\boldsymbol{y}$, but we have not made the dependence explicit in the notation. The intuition for setting λ in this way is that the first order stationarity condition applied to the concave variational problem in (4.11) and (4.12) give us:

$$\frac{1}{m} \sum_{i=1}^{m} \frac{\mathbb{E} E e^{\hat{\lambda}_1(\sigma) E} \psi_\sigma(E - y_i)}{Z_{\mathsf{TExp}}\left(\hat{\lambda}_1(\sigma), y_i\right)} = 1 \implies \mathbb{E}\left[\sum_{i=1}^{m} u_i \right] = m.$$

Consequently, by the central limit theorem, we expect that, $m^{-\frac{1}{2}} \cdot ((\sum_i u_i) - m)$ is close to a Gaussian distribution with variance:

$$\hat{v}(\sigma) \stackrel{\text{def}}{=} \frac{1}{m} \sum_{i=1}^{m} \sigma^2_{\mathsf{TExp}}\left(\hat{\lambda}_1(\sigma), y_i\right) = \hat{\mathbb{E}}_Y \sigma^2_{\mathsf{TExp}}\left(\hat{\lambda}_1(\sigma), Y\right). \tag{4.13}$$

Hence, $F_{\hat{\lambda}_1(\sigma), \boldsymbol{y}}$, which is the density of $\sum_{i=1}^{m} u_i$ can be approximated by the density of $\mathcal{N}\left(m, m\hat{v}(\sigma)\right)$.

$$F_{\hat{\lambda}_1(\sigma), \boldsymbol{y}}(m) \approx \psi_{m \cdot \hat{v}(\sigma)}(0) = \frac{1}{\sqrt{2\pi \hat{v}(\sigma) \cdot m}}.$$

This intuition is made rigorous in the following proposition.

Proposition 7 (A Local Central Limit Theorem)**.** *Suppose that there exists a constant $0 < K < \infty$, such that,*

$$|\hat{\lambda}_1(\sigma)| \leq K,\ \hat{\mathbb{E}}_Y(|Y| + |Y|^2 + |Y|^3) \leq K,\ \frac{1}{K} \leq \hat{v}(\sigma) \leq K.$$

Then, there exists a constant $C(K)$, depending only on K such that we have the following asymp-

totic expansion for $F_{\hat{\lambda}_1(\sigma),\boldsymbol{y}}(m)$*:*

$$\left| F_{\hat{\lambda}_1(\sigma),\boldsymbol{y}}(m) - \frac{1}{\sqrt{2\pi\hat{v}(\sigma)\cdot m}} \right| \leq \frac{C(K)\ln(m)}{m},$$

where $\hat{\lambda}_1(\sigma)$ *and* $\hat{v}(\sigma)$ *have been defined in* (4.11) *and* (4.13)*.*

There is a large literature on local central limit theorems. We refer the reader to Bhattacharya and Rao [75] for a textbook treatment of these results. We are unable to use the statements of local central limit theorems already available in the literature because we require a local central limit theorem for sums of independent but not identically distributed random variables and we further require some control on the error of normal approximation. The proof of Proposition 7 can be found in Appendix B.2.1. It closely follows the classical proofs of local central limit theorems based on characteristic functions (see for example Feller [76, Chapter 16]).

We conclude our analysis of $\mathscr{L}$ with the following result which is a straightforward corollary of the change of measure result given in Lemma 14 and the local central limit theorem in Proposition 7.

Corollary 2 (Lower Bound on $\mathscr{L}$)**.** *Under the assumptions of Proposition 7, there exists* $M(K) \in \mathbb{N}$ *depending only on* K *such that,*

$$\mathscr{L}(\boldsymbol{y}) \geq \frac{1}{2\sqrt{K}} \exp\left(-m\cdot\hat{\Xi}_1(\sigma)\right),\ \forall m \geq M(K),$$

where the function $\hat{\Xi}_1(\sigma)$ *has been defined in* (4.12)*.*

Proof. Applying Lemma 14 with $\hat{\lambda} = \hat{\lambda}_1(\sigma)$, we have,

$$\mathscr{L}(\boldsymbol{y}) = \frac{(m-1)!\cdot e^{m(1-\hat{\lambda})}\cdot F_{\hat{\lambda},\boldsymbol{y}}(m)}{(m)^{m-1}} \cdot \prod_{i=1}^{m} Z_{\mathsf{TExp}}\left(\hat{\lambda}, y_i\right).$$

Note by Stirling's Approximation, we have:

$$\begin{aligned}\frac{(m-1)!}{m^{m-1}} &\geq \sqrt{2\pi(m-1)} \cdot e^{-(m-1)} \cdot \left(1-\frac{1}{m}\right)^{m-1} \\ &\overset{\text{(a)}}{\geq} \sqrt{2\pi(m-1)} \cdot e^{-m}\end{aligned}$$

In the step marked (a), we used the bound $1-x \geq e^{-\frac{x}{1-x}},\ x \in (0,1)$. From Proposition 7, we conclude that there exists a constant $M(K)$, depending only on K, such that,

$$F_{\hat{\lambda},\boldsymbol{y}}(m) \geq \frac{1}{\sqrt{2\pi\hat{v}(\sigma)m}} - \frac{C(K)\ln(m)}{m}.$$

In particular, this means that there exists $M(K)$ depending only on K such that,

$$F_{\hat{\lambda},\boldsymbol{y}}(m) \geq \frac{1}{2\sqrt{2\pi K m}}\ \forall m \geq M(K).$$

This gives us the lower bound:

$$\begin{aligned}\mathscr{L}(\boldsymbol{y}) &\geq \frac{1}{2\sqrt{2K}} \cdot e^{-m\hat{\lambda}} \cdot \prod_{i=1}^{m} Z_{\mathsf{TExp}}\left(\hat{\lambda}, y_i\right),\ \forall m \geq M(K) \\ &= \frac{1}{2\sqrt{K}} \exp\left(-m \max_{\lambda\in\mathbb{R}}\left(\lambda r - \frac{1}{m}\sum_{i=1}^{m} \ln \mathbb{E}_{E\sim\mathsf{Exp}(1)} e^{\lambda E}\psi_\sigma(E-y_i)\right)\right) \\ &= \frac{1}{2\sqrt{K}} \exp\left(-m \cdot \hat{\Xi}_1(\sigma)\right).\end{aligned}$$

In the last step, we used (4.11) and (4.12). □

4.6.2 Analysis of $\mathscr{U}$

We recall the function $\mathscr{U}$ was defined as follows:

$$\mathscr{U}(\boldsymbol{y},\boldsymbol{Q}) \overset{\text{def}}{=} \mathbb{E}\left[\prod_{i=1}^{m} \psi_\sigma(y_i - |G_{1i}|^2)\psi_\sigma(y_i - |G_{2i}|^2)\middle| \boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} = m\boldsymbol{Q}\right],$$

where the matrix $\boldsymbol{Q}$ is of the form:

$$\boldsymbol{Q} = \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix}, \; q \in (0,1). \tag{4.14}$$

We observe that $\mathscr{U}$ can be rewritten as:

$$\mathscr{U}(\boldsymbol{y}, \boldsymbol{Q}) = \mathbb{E}\left[\exp\left(\sum_{i=1}^{m} \ln \psi_\sigma(y_i - |G_{1i}|^2) + \ln \psi_\sigma(y_i - |G_{2i}|^2) \right) \Big| \frac{1}{m}\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} = \boldsymbol{Q} \right].$$

The asymptotics of $\mathscr{U}$ are determined by the large deviation properties of the pair of random variables:

$$\left(\frac{1}{m} \sum_{i=1}^{m} \ln \psi_\sigma(y_i - |G_{1i}|^2) + \ln \psi_\sigma(y_i - |G_{2i}|^2), \frac{1}{m}\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} \right).$$

Both of these random variables are a sum of independent random variables. The Tilted Wishart distribution which is defined below will play a key role in our analysis. This distribution is the Cramer transform (or the exponential tilting) of the random variables defined above.

Definition 6 (The Tilted Wishart Distribution with Parameters (λ, ϕ, y))**.** *A 2×2 Hermitian matrix $\boldsymbol{S}$ is said to be* $\mathsf{TWis}\,(\lambda, \phi, y)$ *if*

$$\boldsymbol{S} = \begin{bmatrix} s & \sqrt{ss'}e^{\mathbf{i}\theta} \\ \sqrt{ss'}e^{-i\theta} & s' \end{bmatrix},$$

and the random variables $s \in [0, \infty), s' \in [0, \infty), \theta \in (-\pi, \pi]$ are sampled from the pdf:

$$h(s, s', \theta) \stackrel{def}{=} \frac{1}{2 \cdot \pi \cdot Z_{\mathsf{TWis}}(\lambda, \phi, y)} \cdot e^{-(1-\lambda)(s+s') + \phi\sqrt{ss'}\cos(\theta)} \cdot \psi_\sigma(s - y) \cdot \psi_\sigma(s' - y).$$

In the above display, the normalizing constant $Z_{\mathsf{TWis}}(\lambda, \phi, y)$ is defined as:

$$Z_{\mathsf{TWis}}(\lambda, \phi, y) \stackrel{def}{=} \frac{1}{2\pi} \int_0^\infty \int_0^\infty \int_{-\pi}^{\pi} e^{-(1-\lambda)(s+s')+\phi\sqrt{ss'}\cos(\theta)} \cdot \psi_\sigma(s-y) \cdot \psi_\sigma(s'-y)\, \mathrm{d}\theta\, \mathrm{d}s\, \mathrm{d}s'.$$

We denote the covariance matrix of the tilted Wishart distribution by $\mathbf{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, y)$, that is:

$$\mathbf{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, y) = \mathbb{E}\left[\mathsf{Vec}\left(\boldsymbol{S} - \mathbb{E}\boldsymbol{S}\right) \mathsf{Vec}\left(\boldsymbol{S} - \mathbb{E}\boldsymbol{S}\right)^{\mathsf{H}}\right].$$

Similar to the analysis of $\mathscr{L}$, the analysis of $\mathscr{U}$ consists of two steps: First, a change of measure step which is given in Lemma 15 and second, an application of the local central limit theorem which is given in Proposition 8.

We begin with the change of measure result. Let $\lambda, \phi \in \mathbb{R}$ be arbitrary. Let $\boldsymbol{S}_1, \boldsymbol{S}_2 \ldots \boldsymbol{S}_m$ be independent Hermitian random matrices with

$$\boldsymbol{S}_i \sim \mathsf{TWis}(\lambda, \phi, y_i), \ \forall\ i\ \in\ [m].$$

Define the random variable $\boldsymbol{S}$ as:

$$\boldsymbol{S} = \sum_{i=1}^{m} \boldsymbol{S}_i.$$

Let $H_{\lambda,\phi,\boldsymbol{y}}$ be the density of the random matrix $\boldsymbol{S}$.

Lemma 15. *For any $\boldsymbol{y} \in \mathbb{R}^m$ and any 2×2 positive definite Hermitian matrix $\boldsymbol{Q}$, we have,*

$$\mathscr{U}(\boldsymbol{y}, \boldsymbol{Q}) = \frac{\pi(m-1)!(m-2)!}{m^{2m-2} \cdot \det(\boldsymbol{Q})^{m-2}} \cdot e^{m(1-\lambda)\mathsf{Tr}(\boldsymbol{Q}) - m\phi\mathrm{Re}(Q_{12})} \cdot \left(\prod_{i=1}^{m} Z_{\mathsf{TWis}}(\lambda, \phi, y_i)\right) \cdot H_{\lambda,\phi,\boldsymbol{y}}(m\boldsymbol{Q}).$$

Proof. Let us index the entries of $\boldsymbol{S}_k,\ k \in [m]$ as follows:

$$\boldsymbol{S}_k = \begin{bmatrix} s_k & \sqrt{s_k s_k'} e^{\mathbf{i}\theta_k} \\ \sqrt{s_k s_k'} e^{-\mathbf{i}\theta_k} & r_k' \end{bmatrix}$$

Define the random variables:

$$\boldsymbol{S} = \sum_{k=1}^{m} \boldsymbol{S}_k,\ T = \sum_{k=1}^{m} \ln \psi_\sigma(y_k - r_k) + \ln \psi_\sigma(y_k - r_k').$$

Consider two possible probability distributions for $\boldsymbol{S}, T$:

Setup 1: $\boldsymbol{S}_k = \boldsymbol{g}_k \boldsymbol{g}_k^{\mathsf{H}}$ where $\boldsymbol{g}_k \sim \mathcal{CN}\left(\mathbf{0}, \boldsymbol{I}_2\right)$. Equivalently, s_i and s_i' are i.i.d. $\mathsf{Exp}\,(1)$ and θ_i are i.i.d. Unif$(-\pi, \pi]$. Let $H(\cdot, \cdot)$ be the joint pdf of $\boldsymbol{S}, T$ in this setup.

Setup 2: $\boldsymbol{S}_k$ are independent and distributed as $\boldsymbol{S}_k \sim \mathsf{TWis}\,(\lambda, \phi, y_k)$. Let $H_{\lambda,\phi,\boldsymbol{y}}(\cdot, \cdot)$ denote the joint pdf of $\boldsymbol{S}, T$ in this setup.

We can compute $H_{\lambda,\phi,\boldsymbol{y}}$ in terms of G as follows:

$$H_{\lambda,\phi,\boldsymbol{y}}(\boldsymbol{S}, T) = \frac{\exp(T + \lambda \cdot \mathsf{Tr}(\boldsymbol{S}) + \phi \cdot \mathrm{Re}(S_{12}))}{\prod_{i=1}^{m} Z_{\mathsf{TWis}}\,(\lambda, \phi, y_i)} \cdot H(\boldsymbol{S}, T).$$

Let $H(\cdot|\boldsymbol{S})$ denote the conditional density of T given $\boldsymbol{S}$ and $H(\boldsymbol{S})$ denote the marginal density of $\boldsymbol{S}$ under Setup 1. Analogously define $H_{\lambda,\phi,\boldsymbol{y}}(\cdot|\boldsymbol{S})$ and $H_{\lambda,\phi,\boldsymbol{y}}(\boldsymbol{S})$ under Setup 2. We can then compute $\mathscr{U}\,(\boldsymbol{y}, \boldsymbol{Q})$ as follows:

$$\begin{aligned}
\mathscr{U}\,(\boldsymbol{y}, \boldsymbol{Q}) &\overset{\text{def}}{=} \mathbb{E}\left[\prod_{i=1}^{m} \phi_\sigma(y_i - |G_{1i}|^2)\phi_\sigma(y_i - |G_{2i}|^2) \middle| \boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} = m\boldsymbol{Q}\right] \\
&= \mathbb{E}\left[\exp\left(\sum_{i=1}^{m} \ln \phi_\sigma(y_i - |g_{1i}|^2) + \ln \phi_\sigma(y_i - |g_{2i}|^2)\right) \middle| \boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} = m\boldsymbol{Q}\right] \\
&\overset{\text{(a)}}{=} \int e^t H(t|m\boldsymbol{Q})\,\mathrm{d}t.
\end{aligned}$$

In the step marked (a), we used the fact that under Setup 1, we have

$$(\boldsymbol{S}, T) \stackrel{\text{d}}{=} \left(\boldsymbol{G}^{\mathsf{H}}\boldsymbol{G},\ \sum_{i=1}^{m} \ln \phi_\sigma(y_i - |g_{1i}|^2) + \ln \phi_\sigma(y_i - |g_{2i}|^2) \right).$$

Hence,

$$\begin{aligned}
\mathscr{U}(\boldsymbol{y}, \boldsymbol{Q}) &= \int e^t H(t|m\boldsymbol{Q})\,\mathrm{d}t \\
&= \frac{e^{-m\lambda \mathsf{Tr}(\boldsymbol{Q}) - m\phi \mathrm{Re}(Q_{12})}}{H(m\boldsymbol{Q})} \int e^{t + m\lambda \mathsf{Tr}(\boldsymbol{Q}) + m\phi \mathrm{Re}(Q_{12})} H(m\boldsymbol{Q}, t)\,\mathrm{d}t \\
&= \frac{e^{-m\lambda \mathsf{Tr}(\boldsymbol{Q}) - m\phi \mathrm{Re}(Q_{12})}}{H(m\boldsymbol{Q})} \cdot \prod_{i=1}^{m} Z_{\mathsf{TWis}}(\lambda, \phi, y_i) \cdot \int H_{\lambda,\phi,\boldsymbol{y}}(m\boldsymbol{Q}, t)\,\mathrm{d}t \\
&= \frac{e^{-m\lambda \mathsf{Tr}(\boldsymbol{Q}) - m\phi \mathrm{Re}(Q_{12})}}{H(m\boldsymbol{Q})} \cdot \prod_{i=1}^{m} Z_{\mathsf{TWis}}(\lambda, \phi, y_i) \cdot H_{\lambda,\phi,\boldsymbol{y}}(m\boldsymbol{Q}) \\
&\stackrel{\text{(b)}}{=} \frac{\pi (m-1)!(m-2)!}{m^{2m-2} \cdot \det(\boldsymbol{Q})^{m-2}} \cdot e^{m(1-\lambda)\mathsf{Tr}(\boldsymbol{Q}) - m\phi \mathrm{Re}(Q_{12})} \cdot \left(\prod_{i=1}^{m} Z_{\mathsf{TWis}}(\lambda, \phi, y_i) \right) \cdot H_{\lambda,\phi,\boldsymbol{y}}(m\boldsymbol{Q}).
\end{aligned}$$

In the step marked (b), we used the fact that under Setup 1, $\boldsymbol{S}$ is distributed as a complex Wishart random matrix and hence,

$$H(m\boldsymbol{Q}) = \frac{1}{\pi} \cdot \frac{m^{2m-2}}{(m-1)!(m-2)!} \cdot \exp(-m\mathsf{Tr}(\boldsymbol{Q})) \cdot \det(\boldsymbol{Q})^{m-2}.$$

This concludes the proof of the lemma. □

Next, we will use a local central limit theorem to characterize the asymptotics of $H_{\lambda,\phi,\boldsymbol{y}}(m\boldsymbol{Q})$. Note that Lemma 15 holds for any $\lambda, \phi \in \mathbb{R}$. We will set $\lambda = \hat{\lambda}_2(q;\sigma), \phi = \hat{\phi}(q;\sigma)$, where

$$(\hat{\lambda}_2(q;\sigma), \hat{\phi}(q;\sigma)) \stackrel{\text{def}}{=} \arg \max_{(\lambda,\phi)\in\mathbb{R}} \left(2\lambda + q\phi - \hat{\mathbb{E}}_Y \ln Z_{\mathsf{TWis}}(\lambda, \phi, Y) \right). \tag{4.15}$$

We also define

$$\hat{\Xi}_2(q;\sigma) \stackrel{\text{def}}{=} \max_{(\lambda,\phi)\in\mathbb{R}} \left(2\lambda + q\phi - \hat{\mathbb{E}}_Y \ln Z_{\mathsf{TWis}}(\lambda,\phi,Y)\right). \tag{4.16}$$

The rational behind this choice of λ, ϕ is that the first order optimality conditions for the above concave variational problem give us:

$$\begin{aligned} 2 &= \frac{1}{m}\sum_{i=1}^{m} \frac{\partial_\lambda Z_{\mathsf{TWis}}\left(\hat{\lambda}_2(q;\sigma), \hat{\phi}(q;\sigma), y_i\right)}{Z_{\mathsf{TWis}}\left(\hat{\lambda}_2(q;\sigma), \hat{\phi}(q;\sigma), y_i\right)} \stackrel{\text{(a)}}{=} \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}(s_i + s_i') \\ q &= \frac{1}{m}\sum_{i=1}^{m} \frac{\partial_\phi Z_{\mathsf{TWis}}\left(\hat{\lambda}_2(q;\sigma), \hat{\phi}(q;\sigma), y_i\right)}{Z_{\mathsf{TWis}}\left(\hat{\lambda}_2(q;\sigma), \hat{\phi}(q;\sigma), y_i\right)} \stackrel{\text{(a)}}{=} \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}\sqrt{s_i s_i'}\cos(\theta_i). \end{aligned}$$

In the steps marked (a), we used the formula for the normalizing constant $Z_{\mathsf{TWis}}(\lambda,\phi,y)$, given in Definition 6, to compute the partial derivatives. It is also clear by the symmetry of Definition 6 that:

$$\mathbb{E}s_i = \mathbb{E}s_i',\ \mathbb{E}\sqrt{s_i s_i'}\sin(\theta) = 0.$$

Hence, the first order optimality conditions imply:

$$\mathbb{E}\boldsymbol{S} = \sum_{i=1}^{m} \mathbb{E}\boldsymbol{S}_i = m\boldsymbol{Q}.$$

By the Multivariate Central Limit Theorem, we expect that $m^{-\frac{1}{2}} \cdot (\boldsymbol{S} - m\boldsymbol{Q})$ to be asymptotically Gaussian. We also define the covariance matrix of $m^{-\frac{1}{2}} \cdot (\boldsymbol{S} - m\boldsymbol{Q})$ as $\hat{\boldsymbol{V}}(q;\sigma)$:

$$\hat{\boldsymbol{V}}(q;\sigma) \stackrel{\text{def}}{=} \frac{\mathbb{E}\mathsf{Vec}\left(\boldsymbol{S} - \mathbb{E}\boldsymbol{S}\right)\mathsf{Vec}\left(\boldsymbol{S} - \mathbb{E}\boldsymbol{S}\right)^{\mathsf{H}}}{m} = \hat{\mathbb{E}}\boldsymbol{\Sigma}_{\mathsf{TWis}}\left(\hat{\lambda}_2(\boldsymbol{Q};\sigma), \hat{\phi}(\boldsymbol{Q};\sigma), Y\right). \tag{4.17}$$

By the CLT, we expect

$$m^{-\frac{1}{2}} \cdot \mathsf{Vec}\left(\boldsymbol{S} - m\boldsymbol{Q}\right) \approx \mathcal{N}\left(\mathbf{0}, \hat{\boldsymbol{V}}(q;\sigma)\right).$$

Hence,

$$H_{\hat{\lambda}_2(q;\sigma),\hat{\phi}(q;\sigma),\boldsymbol{y}}(m\boldsymbol{Q}) \approx \frac{1}{\sqrt{(2\pi m)^4 \det(\hat{\boldsymbol{V}}(q;\sigma))}}.$$

The following proposition makes this argument rigorous.

Proposition 8 (A Local Central Limit Theorem)**.** *Suppose that there exists a constant* $0 < K < \infty$ *such that:*

$$|\hat{\lambda}_2(q;\sigma)| + |\hat{\phi}(q;\sigma)| \le K,\ \hat{\mathbb{E}}_Y|Y|^{40} \le K,\ \frac{1}{K} \le \lambda_{\min}\left(\hat{\boldsymbol{V}}(q;\sigma)\right) \le \lambda_{\max}\left(\hat{\boldsymbol{V}}(q;\sigma)\right) \le K.$$

Then, there exists a constant $C(K)$*, depending only on* K *such that we have the following asymptotic expansion for* $H_{\hat{\lambda}_2(q;\sigma),\hat{\phi}(q;\sigma),\boldsymbol{y}}$*:*

$$\left| H_{\hat{\lambda}_2(q;\sigma),\hat{\phi}(q;\sigma),\boldsymbol{y}} - \frac{1}{\sqrt{(2\pi m)^4 \det(\hat{\boldsymbol{V}}(q;\sigma))}} \right| \le \frac{C(K)\ln^5(m)}{m^2\sqrt{m}}.$$

The proof of this proposition appears in Appendix B.2.2 and closely follows classical proofs of local central limit theorems based on characteristic functions (see for example, Feller [76, Chapter 16]). We conclude our analysis of $\mathscr{U}$ with the following upper bound on $\mathscr{U}$ which is a straightforward corollary of Lemma 15 and Proposition 8.

Corollary 3. *Under the assumptions of Proposition 2, there exists* $M(K) \in \mathbb{N}$ *depending only on* K *such that*

$$\mathscr{U}(\boldsymbol{y}, \boldsymbol{Q}) \le \frac{C(K)}{m^2 \cdot (1-q^2)^{m-2}} \cdot \exp\left(-m \cdot \hat{\Xi}_2(q;\sigma)\right),$$

for all $m \geq M(K)$.

Proof. From Lemma 5, we know that

$$\mathscr{U}(\boldsymbol{y}, \boldsymbol{Q}) =$$
$$\frac{\pi(m-1)!(m-2)!}{m^{2m-2}} \cdot \det(\boldsymbol{Q})^{m-2} \cdot e^{m(1-\lambda)\mathsf{Tr}(\boldsymbol{Q})-m\phi\mathrm{Re}(Q_{12})} \cdot \left(\prod_{i=1}^{m} Z_{\mathsf{TWis}}(\lambda, \phi, y_i)\right) \cdot H_{\lambda,\phi,\boldsymbol{y}}(m\boldsymbol{Q}).$$

In Proposition 8, we obtained the bound

$$\left| H_{\hat{\lambda},\hat{\phi},\boldsymbol{y}}(m\boldsymbol{Q}) - \frac{1}{\sqrt{(2\pi m)^4 \det(\hat{\boldsymbol{V}}(q;\sigma))}} \right| \leq \frac{C(K)\ln^5(m)}{m^2\sqrt{m}}.$$

Note that under the assumptions of Proposition 8, we have

$$\det(\hat{\boldsymbol{V}}(q;\sigma)) \geq \lambda_{\min}^4(\hat{\boldsymbol{V}}(q;\sigma)) \geq \frac{1}{K^4}.$$

This tells us, that there is a $M(K) \in \mathbb{N}$ depending only on K, such that,

$$H_{\hat{\lambda},\hat{\phi},\boldsymbol{y}}(m\boldsymbol{Q}) \leq \frac{C(K)}{m^2}, \ \forall m \geq M(K).$$

By Stirling's approximation, we have

$$\frac{\pi(m-1)!(m-2)!}{m^{2m-2}} \leq \frac{\pi e^5}{e^{2m}}.$$

These estimates give us the upper bound:

$$\mathscr{U}(\boldsymbol{y}, \boldsymbol{Q}) \leq \frac{C(K)e^{m(\mathsf{Tr}(\boldsymbol{Q})-2)}}{m^2 \cdot \det(\boldsymbol{Q})^{m-2}} \cdot e^{-m\max_{(\lambda,\phi)\in\mathbb{R}}\left(\lambda\mathsf{Tr}(\boldsymbol{Q})+\phi\mathrm{Re}(Q_{12})-\frac{1}{m}\sum_{i=1}^{m}\ln Z_{\mathsf{TWis}}(\lambda,\phi,y_i)\right)},$$

for all $m \geq M(K)$. Recalling the definition of $\hat{\Xi}_2(q;\sigma)$ (See (4.15)) and the form of the matrix $\boldsymbol{Q}$ (see (4.14)) gives us the claim of the corollary. □

4.7 The Stochastic Laplace Method

Recall that in Lemmas 12 and 13 we have shown the following upper bound on $\mathbf{I}\left(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}\right)$:

$$\mathbf{I}\left(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}\right) \leq \frac{2}{n-1}\mathbb{E}_{\boldsymbol{y}}\left[\frac{\int_0^1 \mathscr{U}\left(\boldsymbol{y}, \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix}\right) \cdot \frac{q\cdot(1-q^2)^{n-2}}{(1-q^2/2)^{\Delta m}}\,\mathrm{d}q}{\mathscr{L}^2(\boldsymbol{y})} \cdot \mathbf{1}_{\mathcal{E}_m}\right] - 1 + Cm\sqrt{\mathbb{P}(\mathcal{E}_m^c)},$$

where $\mathcal{E}_m$ is an arbitrary event depending on $\boldsymbol{y}$ and the functions $\mathscr{U}, \mathscr{L}$ were defined in (4.5) and (4.6). Let us for the moment, also assume that the conditions required for Corollary 2 and 3 are met. Then, tracking only the exponential order terms, we obtain,

$$\mathbf{I}\left(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}\right) \lessapprox \mathbb{E}_{\boldsymbol{y}}\left[\int_0^1 e^{-m\cdot\hat{\mathcal{F}}(q;\delta,\Delta,\sigma)}\,\mathrm{d}q \cdot \mathbf{1}_{\mathcal{E}_m}\right], \tag{4.18}$$

where,

$$\hat{\mathcal{F}}(q;\delta,\Delta,\sigma) = \hat{\Xi}_2(q;\sigma) - 2\hat{\Xi}_1(\sigma) + \left(1-\frac{1}{\delta}\right)\ln(1-q^2) + \Delta\ln\left(1-\frac{q^2}{2}\right). \tag{4.19}$$

Our goal will be to evaluate the integral in (4.18) via the Laplace Method. However, we observe that the function $\hat{\mathcal{F}}(q;\delta,\Delta,\sigma)$ is stochastic since it depends on the empirical distribution of the phase retrieval observations $\boldsymbol{y}$. It turns out that $\hat{\Xi}_2(q;\sigma)$, defined in (4.15), and $\hat{\Xi}_1(\sigma)$, defined in (4.12), and hence $\hat{\mathcal{F}}(q;\delta,\Delta,\sigma)$ concentrate around deterministic functions $\Xi_2(q;\sigma)$, $\Xi_1(\sigma)$, $\mathcal{F}(q;\delta,\Delta,\sigma)$ defined below:

$$\Xi_1(\sigma) \stackrel{\text{def}}{=} \max_{\lambda\in\mathbb{R}}\left(\lambda - \mathbb{E}_Y \ln \mathbb{E}_{E\sim\mathsf{Exp}(1)} e^{\lambda E}\psi_\sigma(E-Y)\right), \tag{4.20}$$

$$\Xi_2(q;\sigma) \stackrel{\text{def}}{=} \max_{(\lambda,\phi)\in\mathbb{R}}\left(2\lambda + q\phi - \mathbb{E}_Y \ln Z_{\mathsf{TWis}}\left(\lambda,\phi,Y\right)\right), \tag{4.21}$$

$$\mathcal{F}(q;\delta,\Delta,\sigma) \stackrel{\text{def}}{=} \Xi_2(q;\sigma) - 2\Xi_1(\sigma) + \left(1-\frac{1}{\delta}\right)\ln(1-q^2) + \Delta\ln\left(1-\frac{q^2}{2}\right). \tag{4.22}$$

In the above display, the random variable $Y = |Z|^2 + \sigma\epsilon$ where $Z \sim \mathcal{CN}(0,1)$, $\epsilon \sim \mathcal{N}(0,1)$. We also define the deterministic counterparts to $\hat{\lambda}_1(\sigma)$, defined in (4.11) and $\hat{\lambda}_2(q;\sigma), \hat{\phi}(q;\sigma)$, defined in (4.15):

$$\lambda_1(\sigma) \stackrel{\text{def}}{=} \arg\max_{\lambda\in\mathbb{R}} \left(\lambda - \mathbb{E}_Y \ln \mathbb{E}_{E\sim\mathsf{Exp}(1)} e^{\lambda E}\psi_\sigma(E-Y)\right), \tag{4.23}$$

$$(\lambda_2(q;\sigma),\phi(q;\sigma)) \stackrel{\text{def}}{=} \max_{(\lambda,\phi)\in\mathbb{R}} \left(2\lambda + q\phi - \mathbb{E}_Y \ln Z_{\mathsf{TWis}}(\lambda,\phi,Y)\right). \tag{4.24}$$

The convergence to these deterministic functions allows to design a high probability event $\mathcal{E}_m$ on which applying Laplace method to the stochastic function $\hat{\mathcal{F}}(q;\delta,\Delta,\sigma)$ is essentially the same as applying it to the deterministic function $\mathcal{F}(q;\delta,\Delta,\sigma)$. We state our concentration result in the proposition below.

Proposition 9. *For any fixed $\sigma > 0$, we have the following convergence results:*

1. *Convergence of Moments: $\hat{\mathbb{E}}Y^k \xrightarrow{P} \mathbb{E}Y^k$ for any $k \in \mathbb{N}$, where $Y = |Z|^2 + \sigma\epsilon$, $Z \sim \mathcal{CN}(0,1)$ and $\epsilon \sim \mathcal{N}(0,1)$.*

2. *For any $R \in (0,\infty)$, we have the uniform convergence of the functions:*

$$\sup_{|\lambda|\le R} |\hat{\mathbb{E}}\sigma^2_{\mathsf{TExp}}(\lambda,Y) - \mathbb{E}\sigma^2_{\mathsf{TExp}}(\lambda,Y)| \xrightarrow{P} 0,$$

$$\sup_{|\lambda|+|\phi|\le R} \|\hat{\mathbb{E}}\boldsymbol{\Sigma}_{\mathsf{TWis}}(\lambda,\phi,Y) - \mathbb{E}\boldsymbol{\Sigma}_{\mathsf{TWis}}(\lambda,\phi,Y)\| \xrightarrow{P} 0.$$

3. *$\hat{\lambda}_1(\sigma)$ is tight in the sense that, there exists a constant $R \in (0,\infty)$, depending only on σ such that,*

$$\mathbb{P}\left(|\hat{\lambda}_1(\sigma)| > R\right) \to 0.$$

4. $\hat{\Xi}_1(\sigma) \xrightarrow{P} \Xi_1(\sigma)$

5. *For any* $\eta \in (0,1)$*, there exists* $R_\eta \in (0,\infty)$ *(depending only on* η, σ*) such that:*

$$\mathbb{P}\left(\max_{0\le q\le 1-\eta} |\hat{\lambda}_2(q;\sigma)| + |\hat{\phi}(q;\sigma)| > R_\eta\right) \to 0.$$

6. *For any* $\eta \in (0,1)$*, we have,*

$$\sup_{q\in[0,1-\eta]} |\hat{\Xi}_2(q;\sigma) - \Xi_2(q;\sigma)| \xrightarrow{P} 0.$$

7. *For any* $\eta \in (0,1)$*, we have,*

$$\sup_{q\in[0,1-\eta]} |\hat{\lambda}_2(q;\sigma) - \lambda_2(q;\sigma)| \xrightarrow{P} 0, \quad \sup_{q\in[0,1-\eta]} |\hat{\phi}(q;\sigma) - \phi(q;\sigma)| \xrightarrow{P} 0.$$

8. *For any* $\eta \in (0,1)$*, we have,*

$$\sup_{q\in[0,1-\eta]} \left|\frac{\mathrm{d}^2}{\mathrm{d}q^2}\hat{\Xi}_2(q;\sigma) - \frac{\mathrm{d}^2}{\mathrm{d}q^2}\Xi_2(q;\sigma)\right| \xrightarrow{P} 0.$$

The proof of this Proposition appears in Appendix B.3. It uses standard empirical process theory results from Van Der Vaart and Wellner [77] with some modification to account for the fact that the observations $y_1, y_2 \ldots, y_m$ are not independent. With the above concentration result, we suitably design an event $\mathcal{E}_m$ with $\mathbb{P}(\mathcal{E}_m) \to 1$ such that on the event $\mathcal{E}_m$, we are able to adapt the usual proof of Laplace Method to obtain the following conclusion.

Proposition 10. *Suppose that* δ, Δ, σ *are such that* $\mathcal{F}(q;\delta,\Delta,\sigma) > \mathcal{F}(0;\delta,\Delta,\sigma) = 0 \;\forall\; q \in (0,1)$ *and* $\frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(0;\delta,\Delta,\sigma) > 0$*. Then,* $\mathbf{I}\left(\boldsymbol{y},\boldsymbol{z};\boldsymbol{A},\boldsymbol{W}\right) = o(m)$.

The proof of this proposition can be found in Appendix B.4. The claim of this Proposition is very intuitive: It says that due to the concentration of $\hat{\mathcal{F}}(q;\delta,\Delta,\sigma)$ to $\mathcal{F}(q;\delta,\Delta,\sigma)$, the stochastic

and the deterministic integrals:

$$\int_0^1 e^{-m\hat{\mathcal{F}}(q;\delta,\Delta,\sigma)}\,\mathrm{d}q \approx \int_0^1 e^{-m\mathcal{F}(q;\delta,\Delta,\sigma)}\,\mathrm{d}q,$$

behave very similarly. According to the standard Laplace method, the condition $\mathcal{F}(q;\delta,\Delta,\sigma) > \mathcal{F}(0;\delta,\Delta,\sigma) = 0$ ensures that,

$$\frac{1}{m}\ln\left(\int_0^1 e^{-m\mathcal{F}(q;\delta,\Delta,\sigma)}\,\mathrm{d}q\right) \to 0,$$

whereas the positivity requirement on the second derivative ensures that the second order, subexponential factors in the Laplace integral are sufficiently well controlled to obtain $\mathbf{I}\left(\boldsymbol{y},\boldsymbol{z};\boldsymbol{A},\boldsymbol{W}\right) = o(m)$.

4.8 Low Noise Asymptotics

Proposition 10 and Proposition 6 tell us that if for some δ, σ, we can find $\Delta > 0$ such that:

$$\mathcal{F}(q;\delta,\Delta,\sigma) > \mathcal{F}(0;\delta,\Delta,\sigma)\ \forall\ q \in (0,1),\ \frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(0;\delta,\Delta,\sigma) > 0, \tag{4.25}$$

then,

$$\lim_{\substack{m,n\to\infty \\ m=n\delta}} \mathbb{E}_{\boldsymbol{x}_\star,\boldsymbol{y},\boldsymbol{A}} \|\boldsymbol{x}_\star\boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}[\boldsymbol{x}_\star\boldsymbol{x}_\star^{\mathsf{H}}|\boldsymbol{y},\boldsymbol{A}]\|^2 = 1.$$

Note that the Bayes risk increases monotonically with the noise level σ (that is, the phase retrieval problem is harder for larger noise levels). Furthermore, the Bayes risk is atmost the risk of the trivial estimator $\hat{\boldsymbol{x}} = \mathbf{0}$:

$$\limsup_{\substack{m,n\to\infty \\ m=n\delta}} \mathbb{E}_{\boldsymbol{x}_\star,\boldsymbol{y},\boldsymbol{A}} \|\boldsymbol{x}_\star\boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}[\boldsymbol{x}_\star\boldsymbol{x}_\star^{\mathsf{H}}|\boldsymbol{y},\boldsymbol{A}]\|^2 \le \mathbb{E}_{\boldsymbol{x}_\star,\boldsymbol{y},\boldsymbol{A}} \|\boldsymbol{x}_\star\boldsymbol{x}_\star^{\mathsf{H}} - \mathbf{0}\|^2 = 1.$$

Hence if show that the asymptotic Bayes risk is trivial (that is, equal to 1) for an arbitrarily small $\sigma > 0$, it automatically implies the Bayes risk is trivial for larger values of noise. Consequently we will focus on verifying condition (4.25) for small values of noise, where the analysis of the variational problems involved simplifies considerably. We show the following result:

Proposition 11. *Recall that $\mathcal{F}(q; \delta, \Delta, \sigma)$ was defined as:*

$$\mathcal{F}(q; \delta, \Delta, \sigma) = \Xi_2(q; \sigma) - 2\Xi_1(\sigma) + \left(1 - \frac{1}{\delta}\right) \ln(1 - q^2) + \Delta \ln\left(1 - \frac{q^2}{2}\right).$$

For any δ and Δ that satisfy

$$1 \leq \delta < 2,\ 0 < \Delta < \frac{2 - \delta}{\delta},$$

there exists a critical value of the noise level $\sigma_c(\delta, \Delta) > 0$ such that, for any $0 < \sigma < \sigma_c(\delta, \Delta)$, we have

1. *The function $\mathcal{F}(q; \delta, \Delta, \sigma)$ has a unique minimum at $q = 0$ and $\mathcal{F}(q; \delta, \Delta, \sigma) > \mathcal{F}(0; \delta, \Delta, \sigma)$ for any $q \in (0, 1)$.*
2. $\frac{\mathrm{d}^2 \mathcal{F}}{\mathrm{d}q^2}(q; \delta, \Delta, \sigma)\Big|_{q=0} > 0.$

Combined with Proposition 10 and Proposition 6 it immediately gives us Theorem 5 as a corollary.

Corollary 4. *Theorem 5 holds.*

Proof. When $\delta < 2$, we can set:

$$\Delta = \frac{2 - \delta}{2\delta} > 0.$$

Proposition 11 guarantees that (4.25) holds for all values of $0 < \sigma \leq \sigma_c(\delta, \Delta)$. Proposition 10 lets us conclude that for all $0 < \sigma \leq \sigma_c(\delta, \Delta)$, $\mathbf{I}(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}) = o(m)$. Consequently, by Proposition

6, for any $0 < \sigma \leq \sigma_c(\delta, \Delta)$ we have,

$$\lim_{\substack{m,n\to\infty \\ m=n\delta}} \mathbb{E}_{\boldsymbol{x}_\star,\boldsymbol{y},\boldsymbol{A}} \|\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}[\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} | \boldsymbol{y}, \boldsymbol{A}]\|^2 = 1.$$

Since the Bayes risk is atmost 1 and increases monotonically with σ, this means for any $\sigma > 0$:

$$\lim_{\substack{m,n\to\infty \\ m=n\delta}} \mathbb{E}_{\boldsymbol{x}_\star,\boldsymbol{y},\boldsymbol{A}} \|\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}[\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} | \boldsymbol{y}, \boldsymbol{A}]\|^2 = 1.$$

□

The proof of Proposition 11 can be found in Appendix B.5. The main idea of the proof is that in the limit $\sigma \to 0$, the analysis of the function $\mathcal{F}(q; \delta, \Delta, \sigma)$ simplifies considerably.

4.9 Conclusion

In this chapter, we studied the Phase Retrieval problem with subsampled Haar sensing matrices with non-zero but vanishing measurement noise in the high dimensional asymptotic where the signal dimension (n) and the number of measurements (m) diverge such that the sampling ratio $\delta = m/n$ remains fixed. We showed that when the sampling ratio $\delta = m/n < 2$, then it is information theoretically impossible for any estimator to obtain an asymptotically non-trivial performance: any estimator is asymptotically uncorrelated with the signal vector. Since previous work [30, 63] has designed estimators which achieve a non trivial correlation with the planted vector when $\delta > 2$, this shows that the weak recovery threshold for this model is $\delta_{\mathsf{weak}} = 2$.

Chapter 5: Universality in Dynamics of Linearized Message Passing

In this chapter[1], we present some partial progress towards a mathematical understanding of the empirically observed universality. We study the real-valued analog of the phase retrieval problem where the sensing matrix is generated by sub-sampling n columns of the $m \times m$ Hadamard-Walsh matrix. Under an average case assumption on the signal vector, our main result (Theorem 6) shows that the dynamics of a class of linearized Approximate message passing schemes for this structured ensemble are asymptotically identical to the dynamics of the same algorithm in the sub-sampled Haar sensing model in the high-dimensional limit.

5.1 Problem Formulation

In the real-valued analog of the phase retrieval problem (also called sign-retrieval), one observes magnitudes of m linear measurements (denoted by $y_{1:m}$) of an unknown n dimensional signal vector $\boldsymbol{x}_\star \in \mathbb{R}^n$:

$$y_i = |(\boldsymbol{A}\boldsymbol{x}_\star)_i|^2,$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a $m \times n$ sensing matrix.

We also define $\boldsymbol{z} \stackrel{\text{def}}{=} \boldsymbol{A}\boldsymbol{x}_\star$ which we refer to as the signed measurements (which are not observed).

We will study this model in the high-dimensional asymptotic regimen $m, n \to \infty, m = n\delta$. In this chapter, we will find it convenient to state the results in terms of the inverse sampling-ratio:

[1]The results obtained in this chapter have been submitted for possible publication in a journal and appear in the preprint R. Dudeja and M. Bakhshizadeh, "Universality of linearized message passing for phase retrieval with structured sensing matrices," arXiv preprint arXiv:2008.10503, 2020

$$\kappa \stackrel{\text{def}}{=} \frac{1}{\delta} = \frac{n}{m} \tag{5.1}$$

5.1.1 Sensing Models

Next, we introduce 3 different models for the sensing matrix $\boldsymbol{A}$. In all the equations below, $\boldsymbol{P}$ is a uniformly random $m \times m$ permutation matrix and $\boldsymbol{S}$ is the column-selection matrix:

$$\boldsymbol{P} \sim \text{Uniformly Random } m \times m \text{ Permutation Matrix}, \tag{5.2a}$$

$$\boldsymbol{S} = \begin{bmatrix} \boldsymbol{I}_n \\ \boldsymbol{0}_{(m-n)\times n} \end{bmatrix}. \tag{5.2b}$$

Sub-sampled Hadamard Sensing Model: Assume that $m = 2^\ell$ for some $\ell \in \mathbb{N}$. In the sub-sampled Hadamard sensing model the sensing matrix is generated by sub-sampling n columns of a $m \times m$ Hadamard-Walsh matrix $\boldsymbol{H}$ uniformly at random:

$$\boldsymbol{A} = \boldsymbol{H}\boldsymbol{P}\boldsymbol{S}, \tag{5.3}$$

Recall that the Hadamard-Walsh matrix as a closed form formula: For any $i, j \in [m]$, let $\boldsymbol{i}, \boldsymbol{j}$ denote the binary representations of $i-1, j-1$. Hence, $\boldsymbol{i}, \boldsymbol{j} \in \{0,1\}^\ell$. Then the (i,j)-th entry of $\boldsymbol{H}$ is given by:

$$H_{ij} = \frac{(-1)^{\langle \boldsymbol{i}, \boldsymbol{j} \rangle}}{\sqrt{m}}, \tag{5.4}$$

where $\langle \boldsymbol{i}, \boldsymbol{j} \rangle = \sum_{k=1}^{\ell} i_k j_k$. It is well known that $\boldsymbol{H}$ is orthogonal, i.e. $\boldsymbol{H}^\mathsf{T}\boldsymbol{H} = \boldsymbol{I}_m$. This sensing model can be thought of as a real analogue of the sub-sampled Fourier sensing model. Our primary goal is to develop a theory for this sensing model which is not covered by existing results.

We believe that our analysis can be extended to the Fourier case without much effort as well as some other deterministic orthogonal matrices like the discrete cosine transform matrix.

Remark 13. *Some authors refer to any orthogonal matrix with* ± 1 *entries as a Hadamard matrix. We emphasize that we claim results only about the Hadamard-Walsh construction given in* (5.4) *and not arbitrary Hadamard matrices.*

Sub-sampled Haar Sensing Model: In this model the sensing matrix is generated by sub-sampling n columns, chosen uniformly at random, of a $m \times m$ uniformly random orthogonal matrix:

$$\boldsymbol{A} = \boldsymbol{OPS}, \tag{5.5}$$

where $\boldsymbol{O} \sim \text{Unif}\big(\mathbb{O}(m)\big)$. Existing theory applies to this sensing model and our goal will be to transfer these results to the sub-sampled Hadamard model.

Sub-sampled Orthogonal Model: This model includes both sub-sampled Hadamard and Haar models as special cases. In this model the sensing matrix is generated by sub-sampling n columns chosen uniformly at random of a $m \times m$ orthogonal matrix $\boldsymbol{U}$:

$$\boldsymbol{A} = \boldsymbol{UPS}, \tag{5.6}$$

where $\boldsymbol{U}$ is a fixed or random orthogonal matrix. Setting $\boldsymbol{U} = \boldsymbol{O}$ gives the sub-sampled Haar model and setting $\boldsymbol{U} = \boldsymbol{H}$ gives the sub-sampled Hadamard model. Our primary purpose for introducing this general model is that it allows us to handle both the sub-sampled Haar and Hadamard models in a unified way. Additionally, some of our intermediate results hold for any orthogonal matrix $\boldsymbol{U}$ whose entries are delocalized, and we wish to record that when possible.

In addition, we introduce the following matrices which will play an important role in our analysis:

1. We define $\boldsymbol{B} \stackrel{\text{def}}{=} \boldsymbol{P}\boldsymbol{S}\boldsymbol{S}^{\mathsf{T}}\boldsymbol{P}^{\mathsf{T}}$. Observe that $\boldsymbol{B}$ is a random diagonal matrix with $\{0,1\}$ entries. It is easy to check that the distribution of $\boldsymbol{B}$ is described as follows: pick a uniformly random subset $S \subset [m]$ with $|S| = n$ and set:

$$B_{ii} = \begin{cases} 1: & i \in S \\ 0: & i \notin S \end{cases}.$$

2. Note that $\mathbb{E}\boldsymbol{B} = \kappa \boldsymbol{I}_m$. We define the zero mean random diagonal matrix $\overline{\boldsymbol{B}} \stackrel{\text{def}}{=} \boldsymbol{B} - \kappa \boldsymbol{I}_m$.

3. We define the matrix $\boldsymbol{\Psi} \stackrel{\text{def}}{=} \boldsymbol{U}\overline{\boldsymbol{B}}\boldsymbol{U}^{\mathsf{T}} = \boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} - \kappa \boldsymbol{I}_m$.

Finally, note that all the sensing ensembles introduced in this section make sense only when $n \leq m$ or equivalently $\kappa \in [0,1]$. We will additionally assume that κ lies in the open interval $(0,1)$.

Linearized Approximate Message Passing (AMP) Algorithms

We study a class of linearized message passing algorithms. This is a class of iterative schemes which execute the following updates:

$$\hat{\boldsymbol{z}}^{(t+1)} := \left(\frac{1}{\kappa}\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} - \boldsymbol{I}\right) \cdot \left(\eta_t(\boldsymbol{Y}) - \frac{\mathbb{E}\mathsf{Tr}(\eta_t(\boldsymbol{Y}))}{m}\boldsymbol{I}\right) \cdot \hat{\boldsymbol{z}}^{(t)}, \tag{5.7a}$$

$$\hat{\boldsymbol{x}}^{(t+1)} := \boldsymbol{A}^{\mathsf{T}}\hat{\boldsymbol{z}}^{(t+1)}, \tag{5.7b}$$

where

$$\boldsymbol{Y} = \text{Diag}\left(y_1, y_2 \dots y_m\right),$$

and $\eta_t : \mathbb{R} \to \mathbb{R}$ are bounded Lipchitz functions that act entry-wise on the diagonal matrix $\boldsymbol{Y}$. The iterates $(\hat{\boldsymbol{z}}^{(t)})_{t\geq 0}$ should be thought as estimates of the signed measurements $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}_\star$. We now provide further context regarding the iteration in (5.7).

Interpretation as Linearized AMP: The iteration (5.7) can be thought of as a linearization of a broad class of non-linear approximate message passing algorithms. These algorithms execute the iteration:

$$\hat{\boldsymbol{z}}^{(t+1)} := \left(\frac{1}{\kappa}\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} - \boldsymbol{I}\right) \cdot H_t(\boldsymbol{y}, \hat{\boldsymbol{z}}^{(t)}), \tag{5.8a}$$

$$\hat{\boldsymbol{x}}^{(t+1)} := \boldsymbol{A}^{\mathsf{T}}\hat{\boldsymbol{z}}^{(t+1)}. \tag{5.8b}$$

where $H_t : \mathbb{R}^2 \to \mathbb{R}$ is a bounded Lipchitz function which satisfies the divergence-free property:

$$\frac{1}{m}\sum_{i=1}^{m} \mathbb{E}\partial_z H_t(y_i, \hat{z}_i^{(t)}) = 0.$$

Indeed, if H_t was linear in the second (z) argument (or was approximated by its linearization) one obtains the iteration in (5.7). By choosing the function H_t in the iteration appropriately, one can obtain the state-of-the-art performance for phase retrieval with sub-sampled Haar sensing. This algorithm achieves non-trivial (better than random) performance when $\kappa < 2/3$, and exact recovery when $\kappa < 0.63$ [79]. While our analysis currently does not cover the non-linear iteration (5.8), we hope our techniques can be extended to analyze (5.8).

Connection to Spectral Methods: Given that the algorithm we analyze (5.7) does not cover the state-of-the-art algorithm, one can reasonably ask what performance can one achieve with the linearized iteration (5.7). It turns out that the iteration in (5.7) can implement a popular class of spectral methods which estimates the signal vector $\boldsymbol{x}_\star$ as proportional to the leading eigenvector of the matrix:

$$\boldsymbol{M} = \frac{1}{m}\sum_{i=1}^{m} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{T}},$$

where $\boldsymbol{a}_{1:m}$ denote the columns of $\boldsymbol{A}$ and $\mathcal{T} : \mathbb{R}_{\geq 0} \to (-\infty, 1)$ is a trimming function. The performance of these spectral estimators have been analyzed in the high dimensional limit [30, 63]

for the sub-sampled Haar model and they are known to have a non-trivial (better than random) performance when $\kappa < 2/3$. Furthermore, simulations show that the same result holds for sub-sampled Hadamard sensing. In order to connect the iteration (5.7) to the spectral estimator, Ma, Dudeja, Xu, Maleki, and Wang [30] proposed setting the functions η_t in the following way:

$$\eta_t(y) = \left(\frac{1}{\mu} - \mathcal{T}(y)\right)^{-1}, \tag{5.9}$$

where $\mu \in (0, 1)$ is a tuning parameter. Ma, Dudeja, Xu, Maleki, and Wang shows that with this choice of η_t, every fixed point of the iteration (5.7) denoted by $\boldsymbol{z}^\infty$, $\boldsymbol{A}^\mathsf{T}\boldsymbol{z}^\infty$ is an eigenvector of the matrix $\boldsymbol{M}$. Furthermore, suppose μ is set to be the solution to the equation:

$$\psi_1(\mu) = \frac{1}{1-\kappa},\ \psi_1(\mu) \stackrel{\text{def}}{=} \frac{\mathbb{E}|Z|^2 G}{\mathbb{E}G}, \tag{5.10}$$

where the joint distribution of (Z, G) is given by:

$$Z \sim \mathcal{N}(0, 1),\ G = \left(\frac{1}{\mu} - \mathcal{T}(|Z|^2)\right)^{-1}.$$

Then, Ma, Dudeja, Xu, Maleki, and Wang have shown that the linearized message passing iterations (5.7) achieve the same performance as the spectral method for the sub-sampled Haar model as $t \to \infty$.

The State Evolution Formalism: An important property of the AMP algorithms of (5.7) and (5.8) is that for the sub-sampled Haar model, the dynamics of the algorithm can be tracked by a deterministic scalar recursion known as the state evolution. This was first shown for Gaussian sensing matrices by Bayati and Montanari [24] and subsequently for rotationally invariant ensembles by Rangan, Schniter, and Fletcher [35]. We instantiate their result for our problem in the following proposition.

Proposition 12 (State Evolution [35])**.** *Suppose that the sensing matrix is generated from the sub-*

sampled Haar model and the signal vector is normalized such that $\|_\star\|_2^2/m \xrightarrow{P} 1$ *and the iteration* (5.7) *is initialized as:*

$$\hat{\boldsymbol{z}}^{(0)} = \alpha_0 \boldsymbol{z} + \sigma_0 \boldsymbol{w},$$

where $\alpha_0 \in \mathbb{R}, \sigma_0 \in \mathbb{R}_+$ *are fixed and* $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_m)$*. Then for any fixed* $t \in \mathbb{N}$*, as* $m, n \to \infty$*,* $n/m \to \kappa$*, we have,*

$$\begin{aligned} \frac{\langle \hat{\boldsymbol{z}}^{(t)}, \boldsymbol{z} \rangle}{m} &\xrightarrow{P} \alpha_t, \ \frac{\|\hat{\boldsymbol{z}}^{(t)}\|_2^2}{m} \xrightarrow{P} \alpha_t^2 + \sigma_t^2, \\ \frac{\langle \hat{\boldsymbol{x}}^{(t)}, \boldsymbol{x}_\star \rangle}{m} &\xrightarrow{P} \alpha_t, \ \frac{\|\hat{\boldsymbol{x}}^{(t)}\|_2^2}{m} \xrightarrow{P} \alpha_t^2 + (1-\kappa)\sigma_t^2, \end{aligned}$$

where (α_t, σ_t^2) *are given by the recursion:*

$$\alpha_{t+1} = (\delta - 1) \cdot \alpha_t \cdot \mathbb{E} Z^2 \overline{\eta}_t(|Z|), \tag{5.11a}$$

$$\sigma_{t+1}^2 = \left(\frac{1}{\kappa} - 1\right) \cdot \left(\alpha_t^2 \cdot \left\{\mathbb{E} Z^2 \overline{\eta}_t^2(|Z|) - (\mathbb{E} Z^2 \overline{\eta}_t(|Z|))^2\right\} + \sigma_t^2 \mathbb{E} \overline{\eta}_t^2(|Z|)\right). \tag{5.11b}$$

In the above display, $Z \sim \mathcal{N}(0,1)$ *and* $\overline{\eta}_t(z) = \eta_t(|z|^2) - \mathbb{E}\eta_t(|Z|^2)$*.*

The above proposition lets us track the evolution of some performance metrics like the mean squared error (MSE) and the cosine similarity of the iterates. The proof of Proposition 12 crucially relies on the rotational invariance of the sub-sampled Haar ensemble via Bolthausen's conditioning technique [80] and does not extend to structured sensing ensembles like the sub-sampled Hadamard sensing matrix. However, empirically, the state evolution accurately describes the dynamics of the Linearized AMP algorithm even for the sub-sampled Hadamard ensemble. In this chapter, we seek to understand this universality phenomenon.

A Demonstration of the Universality Phenomena: For the sake of completeness, we provide a self contained demonstration of the universality phenomena that we seek to study in Figure 5.1. In order to generate this figure:

1. We used a 1024×256 image (after vectorization, shown as inset in Figure 5.1) as the signal vector. Each of the red, blue, green channels were centered so that that their mean was zero and standard deviation was 1.

2. We set $m = 1024 \times 256$.

3. In order to generate problems with different κ we down-sampled the original image to obtain a new signal with $n \approx m\kappa$ (upto rounding errors).

4. We used a randomly sub-sampled Hadamard matrix for sensing. This was used to construct a phase retrieval problem for each of the red, blue and green channels.

5. We used the linearized message passing configured to implement the spectral estimator (c.f. (5.9) and (5.10)) with the optimal trimming function [21, 30]:

$$\mathcal{T}_\star(y) = 1 - \frac{1}{y}.$$

   We ran the algorithm for 20 iterations and tracked the squared cosine similarity:

$$\cos^2(\angle(\hat{\boldsymbol{x}}^{(t)}, \boldsymbol{x}_\star)) \stackrel{\text{def}}{=} \frac{\left|\langle \hat{\boldsymbol{x}}^{(t)}, \boldsymbol{x}_\star \rangle\right|^2}{\|\hat{\boldsymbol{x}}^{(t)}\|_2^2 \|\boldsymbol{x}_\star\|_2^2}.$$

   We averaged the squared cosine similarity across the RGB channels.

6. We repeated this for 10 different random sensing matrices. The average cosine similarity is represented by $+$ markers in Figure 5.1 and the error bars represent the standard error across 10 repetitions. The solid curves represent the predictions derived from State Evolution (see Proposition 12). We can observe that the State Evolution closely tracks the empirical dynamics.

Assumption on the signal: It is easy to see that, unlike in the sub-sampled Haar case, the state evolution cannot hold for arbitrary worst case signal vectors for the sub-sampled Hadamard sensing

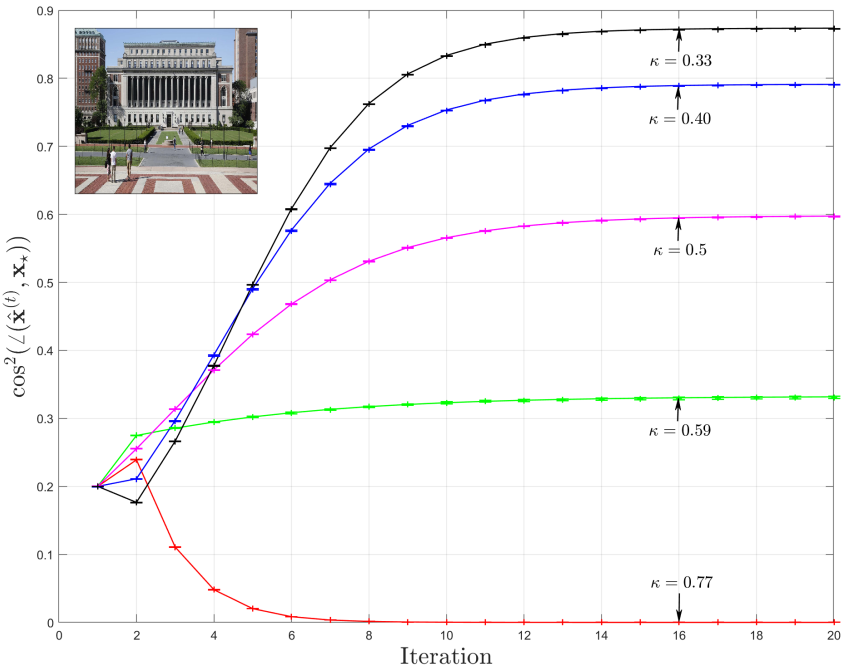

Figure 5.1: Solid Lines: Predicted Dynamics derived using State Evolution (Prop. 12 developed for sub-sampled Haar sensing, + markers: Dynamics of Linearized Message Passing averaged over 10 repetitions with sub-sampled Hadamard sensing and a real image (shown in inset) used as the signal vector. The error bars represent the standard error across repetitions.

models since the orthogonal signal vectors $\sqrt{m}\boldsymbol{e}_1$ and $\sqrt{m}\boldsymbol{e}_2$ generate the same measurement vector $\boldsymbol{y} = (1, 1 \cdots, 1)^\mathsf{T}$. This is a folklore argument for non-indentifiability of the phase retrieval problem for ± 1 sensing matrices [81]. Hence we study the universality phenomena under the simplest average case assumption on the signal, namely $\boldsymbol{x}_\star \sim \mathcal{N}\left(\mathbf{0}, \boldsymbol{I}_n/\kappa\right)$.

5.2 Main Result

Now, we are ready to state our main result.

Theorem 6. *Consider the linear message passing iterations* (5.7)*. Suppose that:*

1. *The functions η_t are bounded and Lipchitz.*

2. *The signal is generated from the Gaussian prior: $\boldsymbol{x}_\star \sim \mathcal{N}\left(\mathbf{0}, \frac{1}{\kappa}\boldsymbol{I}_n\right)$.*

3. *The sensing matrix is generated from the sub-sampled Hadamard ensemble.*

4. *the iteration* (5.7) *is initialized as:*

$$\hat{\boldsymbol{z}}^{(0)} = \alpha_0 \boldsymbol{z} + \sigma_0 \boldsymbol{w},$$

where $\alpha_0 \in \mathbb{R}, \sigma_0 \in \mathbb{R}_+$ *are fixed and* $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_m)$.

Then for any fixed $t \in \mathbb{N}$, *as* $m, n \to \infty$, $n = \kappa m$, *we have,*

$$\begin{aligned} \frac{\langle \hat{\boldsymbol{z}}^{(t)}, \boldsymbol{z} \rangle}{m} &\xrightarrow{P} \alpha_t, \ \frac{\|\hat{\boldsymbol{z}}^{(t)}\|_2^2}{m} \xrightarrow{P} \alpha_t^2 + \sigma_t^2, \\ \frac{\langle \hat{\boldsymbol{x}}^{(t)}, \boldsymbol{x}_\star \rangle}{m} &\xrightarrow{P} \alpha_t, \ \frac{\|\hat{\boldsymbol{x}}^{(t)}\|_2^2}{m} \xrightarrow{P} \alpha_t^2 + (1-\kappa)\sigma_t^2, \end{aligned}$$

where (α_t, σ_t^2) *are given by the recursion in* (5.11).

Theorem 6 simply states that the dynamics of linearized message passing in the sub-sampled Hadamard model are asymptotically indistinguishable from the dynamics in the sub-sampled Haar model. This provides a theoretical justification for the universality depicted in Figure 5.1.

Remarks on Proof Techniques: The proof of Theorem 6 is inspired by certain universality results in random matrix theory [48] and in particular free probability [49]. A well known result in free probability (see the book of Mingo and Speicher [49] for a textbook treatment) is that if $\boldsymbol{U} \sim \text{Unif}\big(\mathbb{U}(m)\big)$ and $\boldsymbol{D}_1, \boldsymbol{D}_2$ are deterministic $m \times m$ diagonal matrices then $\boldsymbol{U}\boldsymbol{D}_1\boldsymbol{U}^{\mathsf{H}}$ and $\boldsymbol{D}_2$ are asymptotically free and consequently the limiting spectral distribution of matrix polynomials in $\boldsymbol{D}_2$ and $\boldsymbol{U}\boldsymbol{D}_1\boldsymbol{U}^{\mathsf{H}}$ can be described in terms of the limiting spectral distribution of $\boldsymbol{D}_1$ and $\boldsymbol{D}_2$. Tulino, Caire, Shamai, and Verdu [50] and Farrell [51] have obtained an extension of this result where a Haar unitary matrix is replaced by $m \times m$ Fourier matrix: If $\boldsymbol{D}_1, \boldsymbol{D}_2$ are independent diagonal matrices then $\boldsymbol{F}_m \boldsymbol{D}_1 \boldsymbol{F}_m^{\mathsf{H}}$ is asymptotically free from $\boldsymbol{D}_2$. The result of these authors has been extended to other deterministic orthogonal/unitary matrices (such as the Hadamard-Walsh matrix) conjugated by random signed permutation matrices by Anderson and Farrell [52]. In order

to see how the result of Tulino, Caire, Shamai, and Verdu connects with ours note that the linearized AMP iterations (5.7) involve 2 random matrices: $\boldsymbol{H}\overline{\boldsymbol{B}}\boldsymbol{H}^{\mathsf{T}}$ and $q(\boldsymbol{Y})$. Note that if $\boldsymbol{B}$ and the diagonal matrix $q(\boldsymbol{Y})$ were independent, then the result of Tulino, Caire, Shamai, and Verdu would imply that $\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\mathsf{T}}$ and $q(\boldsymbol{Y})$ are asymptotically free and this could potentially be used to analyze the linearized AMP algorithm. However, the key difficulty is that the measurements $\boldsymbol{y}$ depend on which columns of the Hadamard-Walsh matrix were selected (specified by $\boldsymbol{B}$). Infact, this dependence is precisely what allows the linearized AMP algorithm to recover the signal. However, we still find some of the techniques introduced by Tulino, Caire, Shamai, and Verdu useful in our analysis. We also emphasize that asymptotic freeness of $\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\mathsf{T}},\ q(\boldsymbol{Y})$ alone seems to be insufficient to characterize the behavior of Linearized AMP algorithms. Asymptotic freeness implies that the expected normalized trace of certain matrix products involving $\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\mathsf{T}},\ q(\boldsymbol{Y})$ vanish in the limit $m \to \infty$. On the other hand, our proof also requires the analysis of certain quadratic forms involving $\boldsymbol{H}\boldsymbol{B}\boldsymbol{H}^{\mathsf{T}},\ q(\boldsymbol{Y})$ (see Proposition 14) which do not appear to have been studied in the free probability literature.

5.3 Additional Notation

In this section, we introduce some additional notations we rely on in this chapter.

Linear Algebraic Aspects: We will use bold face letters to refer to vectors and matrices. For a matrix $\boldsymbol{V} \in \mathbb{R}^{m\times n}$, we adopt the convention of referring to the columns of $\boldsymbol{V}$ by $\boldsymbol{V}_1, \boldsymbol{V}_2 \cdots \boldsymbol{V}_n \in \mathbb{R}^m$ and to the rows by $\boldsymbol{v}_1, \boldsymbol{v}_2 \cdots \boldsymbol{v}_m \in \mathbb{R}^n$. For a vector $\boldsymbol{v}$, $\|\boldsymbol{v}\|_1, \|\boldsymbol{v}\|_2, \|\boldsymbol{v}\|_\infty$ denote the ℓ_1, ℓ_2, and ℓ_∞ norms, respectively. By default, $\|\boldsymbol{v}\|$ denotes the ℓ_2 norm. For a matrix $\boldsymbol{V}$, $\|\boldsymbol{V}\|_{\mathsf{op}}, \|\boldsymbol{V}\|_{\mathsf{Fr}}, \|\boldsymbol{V}\|_\infty$ denote the operator norm, Frobenius norm, and the entry-wise ∞-norm, respectively.

Important distributions: $\mathsf{Bern}(p)$ denotes Bernoulli distribution with bias p. $\mathsf{Binom}(n, p)$ denotes the Binomial distribution with n trials and bias p. For an arbitrary set S, $\mathsf{Unif}\,(S)$ denotes the uniform distribution on the elements of S.

Order Notation and Constants: We use the standard $O(\cdot)$ notation. C will be used to refer to a universal constant independent of all parameters. When the constant C depends on a parameter k we will make this explicit by using the notation C_k or $C(k)$. We say a sequence $a_n = O(\text{polylog}(n))$ if there exists a fixed, finite constant K such that $a_n \leq O(\log^K(n))$.

5.4 Proof Overview

Our basic strategy to prove Theorem 6 will be as follows: Throughout the chapter, we will assume that Assumptions 1, 2, and 4 of Theorem 6 hold. We will seek to only show that the observables:

$$\frac{\langle \hat{\boldsymbol{z}}^{(t)}, \boldsymbol{z} \rangle}{m}, \ \frac{\|\hat{\boldsymbol{z}}^{(t)}\|_2^2}{m}, \frac{\langle \hat{\boldsymbol{x}}^{(t)}, \boldsymbol{x}_\star \rangle}{m}, \ \frac{\|\hat{\boldsymbol{x}}^{(t)}\|_2^2}{m}, \tag{5.12}$$

have the same limit in probability under both the sub-sampled Haar and the sub-sampled Hadamard sensing models. We will not need to explicitly identify their limits since Proposition 12 already identifies the limit for us, and hence, Theorem 6 will follow.

It turns out the limits of the observables (5.12) depends only on normalized traces and quadratic forms of certain alternating products of the matrices $\boldsymbol{\Psi}$ and $\boldsymbol{Z}$. Hence, we introduce the following definition.

Definition 7 (Alternating Product)**.** *A matrix $\mathcal{A}$ is said to be a alternating product of matrices $\boldsymbol{\Psi}, \boldsymbol{Z}$ if there exist polynomials $p_i : \mathbb{R} \to \mathbb{R}, \ i \in 1, 2 \ldots, k$, and bounded, Lipchitz functions $q_i : \mathbb{R} \to \mathbb{R}, \ i \in \{1, 2 \ldots k\}$ such that:*

1. *If $B \sim \mathsf{Bern}(\kappa)$, $\mathbb{E} p_i(B - \kappa) = 0$.*
2. *q_i are even functions i.e. $q_i(\xi) = q_i(-\xi)$ and if $\xi \sim \mathcal{N}(0, 1)$, then, $\mathbb{E} q_i(\xi) = 0$,*

and, $\mathcal{A}$ is one of the following:

1. *Type 1: $\mathcal{A} = p_1(\boldsymbol{\Psi}) q_1(\boldsymbol{Z}) p_2(\boldsymbol{\Psi}) \cdots q_{k-1}(\boldsymbol{Z}) p_k(\boldsymbol{\Psi})$*

2. *Type 2:* $\mathcal{A} = p_1(\boldsymbol{\Psi})q_1(\boldsymbol{Z})p_2(\boldsymbol{\Psi})q_2(\boldsymbol{Z})\cdots p_k(\boldsymbol{\Psi})q_k(\boldsymbol{Z})$

3. *Type 3:* $\mathcal{A} = q_1(\boldsymbol{Z})p_2(\boldsymbol{\Psi})q_2(\boldsymbol{Z})\cdots p_k(\boldsymbol{\Psi})q_k(\boldsymbol{Z})$.

4. *Type 4:* $\mathcal{A} = q_1(\boldsymbol{Z})p_2(\boldsymbol{\Psi})q_2(\boldsymbol{Z})p_3(\boldsymbol{\Psi})\cdots q_{k-1}(\boldsymbol{Z})p_k(\boldsymbol{\Psi})$.

In the above definitions:

1. *The scalar polynomial p_i is evaluated at the matrix $\boldsymbol{\Psi}$ in the usual sense, for example if $p(\psi) = \psi^2$, then, $p(\boldsymbol{\Psi}) = \boldsymbol{\Psi}^2$.*

2. *The functions q_i are evaluated entry-wise on the diagonal matrix $\boldsymbol{Z}$, i.e.*

$$q_i(\boldsymbol{Z}) = \textit{Diag}\left(q_i(z_1), q_i(z_2)\ldots q_i(z_m)\right).$$

We note that alternating products are a central notion in free probability [49]. The difference here is that we have additionally constrained the functions p_i, q_i in Definition 7.

Theorem 6 is a consequence of two properties of alternating products which may be of independent interest. These are stated in the following propositions.

Proposition 13. *Let $\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})$ be an alternating product of matrices $\boldsymbol{\Psi}, \boldsymbol{Z}$. Suppose the sensing matrix $\boldsymbol{A}$ is generated from the sub-sampled Haar sensing model, or the sub-sampled Hadamard sensing model, or by sub-sampling a deterministic orthogonal matrix $\boldsymbol{U}$ with the property:*

$$\|\boldsymbol{U}\|_\infty \leq \sqrt{\frac{K_1 \log^{K_2}(m)}{m}},\ \forall m\ \geq K_3,$$

for some fixed constants K_1, K_2, K_3. Then,

$$\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}))/m \xrightarrow{P} 0.$$

Proposition 14. *Let $\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})$ be an alternating product of matrices $\boldsymbol{\Psi}, \boldsymbol{Z}$. Then for the sub-sampled Haar sensing model and for sub-sampled Hadamard ($\boldsymbol{U} = \boldsymbol{H}$) sensing model, we have,*

$$\text{p-lim } \frac{\langle \boldsymbol{z}, \mathcal{A}\boldsymbol{z} \rangle}{m}$$

exists and is identical for the two models.

Outline of the Remaining Chapter: The remainder of the chapter is organized as follows:

1. In Section 5.5 we provide a proof of Theorem 6 assuming Propositions 13 and 14.
2. In Section 5.6 we introduce some key tools required for the proof of Propositions 13 and 14.
3. The proof of Proposition 13 can be found in Section 5.7.
4. The proof of Proposition 14 can be found in Section 5.8.

5.5 Proof of Theorem 6

In this section we will show the analysis of the observables (5.12) reduces to the analysis of the normalized traces and quadratic forms of alternating products. In particular, we will prove Theorem 6 using Propositions 13 and 14.

Proof of Theorem 6. For simplicity, we will assume the functions η_t do not change with t, i.e. $\eta_t = \eta \; \forall \; t \geq 0$. This is just to simplify notations, and the proof of time varying η_t is exactly the same. Define the function:

$$q(z) = \eta(|z|^2) - \mathbb{E}_{Z \sim \mathcal{N}(0,1)}[\eta(|Z|^2)].$$

Note that the linearized message passing iterations (5.7) can be expressed as:

$$\hat{\boldsymbol{z}}^{(t+1)} = \frac{1}{\kappa} \cdot \boldsymbol{\Psi} \cdot q(\boldsymbol{Z}) \cdot \hat{\boldsymbol{z}}^{(t)}.$$

Unrolling the iterations we obtain:

$$\hat{\boldsymbol{z}}^{(t)} = \frac{1}{\kappa^t} \cdot (\boldsymbol{\Psi} \cdot q(\boldsymbol{Z}))^t \cdot \hat{\boldsymbol{z}}^{(0)}.$$

Note that the initialization is assumed to be of the form: $\hat{\boldsymbol{z}}^{(0)} = \alpha_0 \boldsymbol{z} + \sigma_0 \boldsymbol{w}$, where $\boldsymbol{w} \sim \mathcal{N}(0, \boldsymbol{I})$. Hence:

$$\begin{aligned}
\hat{\boldsymbol{z}}^{(t)} &= \alpha_0 \frac{1}{\kappa^t} \cdot (\boldsymbol{\Psi} \cdot q(\boldsymbol{Z}))^t \cdot \boldsymbol{z} + \sigma_0 \cdot \frac{1}{\kappa^t} \cdot (\boldsymbol{\Psi} \cdot q(\boldsymbol{Z}))^t \cdot \boldsymbol{w}, \\
\hat{\boldsymbol{x}}^{(t)} &= \boldsymbol{A}^\mathsf{T} \hat{\boldsymbol{z}}^{(t)}.
\end{aligned}$$

We will focus on showing that the limits:

$$\text{p-lim} \, \frac{\langle \boldsymbol{x}_\star, \hat{\boldsymbol{x}}^{(t)} \rangle}{m}, \ \text{p-lim} \, \frac{\|\hat{\boldsymbol{x}}^{(t)}\|_2^2}{m}, \tag{5.13}$$

exist and are identical for the two models. The claim for the limits corresponding to $\hat{\boldsymbol{z}}^{(t)}$ are exactly analogous and omitted. Hence, the remainder of the proof is devoted to analyzing the above limits.

Analysis of $\langle \boldsymbol{x}_\star, \hat{\boldsymbol{x}}^{(t)} \rangle$: Observe that:

$$\begin{aligned}
\langle \boldsymbol{x}_\star, \hat{\boldsymbol{x}}^{(t)} \rangle &= \langle \boldsymbol{A}^\mathsf{T} \boldsymbol{z}, \boldsymbol{A}^\mathsf{T} \hat{\boldsymbol{z}}^{(t)} \rangle \\
&= \alpha_0 \frac{1}{\kappa^t} \cdot \underbrace{\langle \boldsymbol{A}^\mathsf{T} \boldsymbol{z}, \boldsymbol{A}^\mathsf{T} (\boldsymbol{\Psi} \cdot q(\boldsymbol{Z}))^t \cdot \boldsymbol{z} \rangle}_{(T_1)} + \sigma_0 \cdot \frac{1}{\kappa^t} \cdot \underbrace{\langle \boldsymbol{A}^\mathsf{T} \boldsymbol{z}, \boldsymbol{A}^\mathsf{T} \cdot (\boldsymbol{\Psi} \cdot q(\boldsymbol{Z}))^t \cdot \boldsymbol{w} \rangle}_{(T_2)}.
\end{aligned}$$

We first analyze term (T_1). Observe that:

$$
\begin{aligned}
(T_1) &= \boldsymbol{z}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t\boldsymbol{z} \\
&= \boldsymbol{z}^{\mathsf{T}}\boldsymbol{\Psi}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t\boldsymbol{z} + \kappa\boldsymbol{z}^{\mathsf{T}}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t\boldsymbol{z} \\
&= \boldsymbol{z}^{\mathsf{T}}\boldsymbol{\Psi}^2(q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1}q(\boldsymbol{Z})\boldsymbol{z} + \kappa\boldsymbol{z}^{\mathsf{T}}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t\boldsymbol{z} \\
&\overset{\text{(a)}}{=} \boldsymbol{z}^{\mathsf{T}}p(\boldsymbol{\Psi})(q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1}q(\boldsymbol{Z})\boldsymbol{z} + \kappa(1-\kappa)\boldsymbol{z}^{\mathsf{T}}(q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1}q(\boldsymbol{Z})\boldsymbol{z} + \kappa\boldsymbol{z}^{\mathsf{T}}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t\boldsymbol{z}.
\end{aligned}
$$

In the step marked (a) we defined the polynomial $p(\psi) = \psi^2 - \kappa(1-\kappa)$ which has the property $\mathbb{E}p(B-\kappa) = 0$ when $B \sim \mathsf{Bern}(\kappa)$. One can check that $Z \sim \mathcal{N}(0,1)$, $\mathbb{E}q(Z) = 0$, and q is a bounded, Lipchitz, even function. Hence, each of the terms appearing in step (a) are of the form $\boldsymbol{z}^{\mathsf{T}}\mathcal{A}\boldsymbol{z}$ for some alternating product $\mathcal{A}$ (Definition 7) of matrices $\boldsymbol{\Psi}, \boldsymbol{Z}$. Consequently, by Proposition 14 we obtain that term (1) divided by m converges to the same limit in probability under both the sub-sampled Haar sensing and the sub-sampled Hadamard sensing model. Next, we analyze (T_2). Note that:

$$
\begin{aligned}
\frac{\langle \boldsymbol{A}^{\mathsf{T}}\boldsymbol{z}, \boldsymbol{A}^{\mathsf{T}}\cdot(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t\cdot\boldsymbol{w}\rangle}{m} &= \boldsymbol{z}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t\boldsymbol{w}/m \\
&\overset{\text{d}}{=} \frac{\|(q(\boldsymbol{Z})\boldsymbol{\Psi})^t\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{z}\|_2}{m}\cdot W,\ W \sim \mathcal{N}(0,1)\,,
\end{aligned}
$$

where $\overset{\text{d}}{=}$ means both sides have a same distribution. Observe that:

$$
\begin{aligned}
\frac{\|(q(\boldsymbol{Z})\boldsymbol{\Psi})^t\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{z}\|_2}{m} &= \frac{\|(q(\boldsymbol{Z})\boldsymbol{\Psi})^t\boldsymbol{A}\boldsymbol{x}_\star\|_2}{m} \\
&\le \|(q(\boldsymbol{Z})\boldsymbol{\Psi})^t\boldsymbol{A}\|_{\mathsf{op}}\cdot\frac{\|\boldsymbol{x}_\star\|_2}{m} \\
&\le \|q(\boldsymbol{Z})\|_{\mathsf{op}}^t\|\boldsymbol{\Psi}\|_{\mathsf{op}}^t\|\boldsymbol{A}\|_{\mathsf{op}}\cdot\frac{\|\boldsymbol{x}_\star\|_2}{m}.
\end{aligned}
$$

It is easy to check that: $\|q(\boldsymbol{Z})\|_{\mathsf{op}} \le 2\|\eta\|_\infty < \infty$. Similarly, $\|\boldsymbol{\Psi}\|_{\mathsf{op}} \le 1,\ \|\boldsymbol{A}\|_{\mathsf{op}} = 1$.

Hence,

$$\frac{\|(q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{A}\boldsymbol{A}^\mathsf{T}\boldsymbol{z}\|_2}{m} \leq 2^t \|\eta\|_\infty^t \cdot \sqrt{\frac{\|\boldsymbol{x}_\star\|^2}{m}} \cdot \frac{1}{\sqrt{m}}$$

Observing that $\|\boldsymbol{x}_\star\|^2/m \xrightarrow{\text{P}} 1$ we obtain:

$$\left| \frac{\langle \boldsymbol{A}^\mathsf{T}\boldsymbol{z}, \boldsymbol{A}^\mathsf{T} \cdot (\boldsymbol{\Psi} \cdot q(\boldsymbol{Z}))^t \cdot \boldsymbol{w} \rangle}{m} \right| \leq 2^t \|\eta\|_\infty^t \cdot \sqrt{\frac{\|\boldsymbol{x}_\star\|^2}{m}} \cdot \frac{|W|}{\sqrt{m}} \xrightarrow{\text{P}} 0.$$

Note the above result holds for both subsampled Haar sensing and subsampled Hadamard sensing. This proves that the limit

$$\text{p-lim} \, \frac{\langle \boldsymbol{x}_\star, \hat{\boldsymbol{x}}^{(t)} \rangle}{m}$$

exists and is identical for the two models.

Analysis of $\|\hat{\boldsymbol{x}}^{(t)}\|^2$: Recalling that:

$$\begin{aligned}
\hat{\boldsymbol{z}}^{(t)} &= \alpha_0 \frac{1}{\kappa^t} \cdot (\boldsymbol{\Psi} \cdot q(\boldsymbol{Z}))^t \cdot \boldsymbol{z} + \sigma_0 \frac{1}{\kappa^t} \cdot (\boldsymbol{\Psi} \cdot q(\boldsymbol{Z}))^t \cdot \boldsymbol{w}, \\
\hat{\boldsymbol{x}}^{(t)} &= \boldsymbol{A}^\mathsf{T} \hat{\boldsymbol{z}}^{(t)},
\end{aligned}$$

we can compute:

$$\frac{1}{m}\|\hat{\boldsymbol{x}}^{(t)}\|_2^2 = \frac{1}{\kappa^{2t}} \cdot \left(\alpha_0^2 \cdot (T_3) + 2\alpha_0\sigma_0 (T_4) + \sigma_0^2 \cdot (T_5) \right),$$

where the terms $(T_3 - T_5)$ are defined as:

$$\begin{aligned}
(T_3) &= \frac{\boldsymbol{z}^\mathsf{T}(q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{A}\boldsymbol{A}^\mathsf{T}(\boldsymbol{\Psi} \cdot q(\boldsymbol{Z}))^t \cdot \boldsymbol{z}}{m}, \\
(T_4) &= \frac{\boldsymbol{z}^\mathsf{T}(q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{A}\boldsymbol{A}^\mathsf{T}(\boldsymbol{\Psi} \cdot q(\boldsymbol{Z}))^t \cdot \boldsymbol{w}}{m}, \\
(T_5) &= \frac{\boldsymbol{w}^\mathsf{T}(q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{A}\boldsymbol{A}^\mathsf{T}(\boldsymbol{\Psi} \cdot q(\boldsymbol{Z}))^t \cdot \boldsymbol{w}}{m}.
\end{aligned}$$

We analyze each of these terms separately. First, consider (T_3). Our goal will be to decompose the matrix $(q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{A}\boldsymbol{A}^\mathsf{T}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t$ as:

$$(q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{A}\boldsymbol{A}^\mathsf{T}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t = c_0\boldsymbol{I} + \sum_{i=1}^{N_t} c_i \mathcal{A}_i,$$

where $\mathcal{A}_i$ are alternating products of the matrices $\boldsymbol{\Psi}, \boldsymbol{Z}$ (see Definition 7) and c_i are some scalar constants. This decomposition has the following properties: 1) It is independent of the choice of the orthogonal matrix $\boldsymbol{U}$ used to generate the sensing matrix. 2) The number of terms in the decomposition N_t depends only on t and not on m, n. In order to see why such a decomposition exists: first recall that $\boldsymbol{A}\boldsymbol{A}^\mathsf{T} = \boldsymbol{\Psi} + \kappa \boldsymbol{I}_m$. Hence, we can write:

$$\begin{aligned}(q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{A}\boldsymbol{A}^\mathsf{T}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t &= (q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{\Psi}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t + \kappa \boldsymbol{z}^\mathsf{T}(q(\boldsymbol{Z})\boldsymbol{\Psi})^t(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t \\ &= (q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1} q(\boldsymbol{Z})\Psi^3 q(\boldsymbol{Z})(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^{t-1} + \kappa \boldsymbol{z}^\mathsf{T}(q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1} q(\boldsymbol{Z})\boldsymbol{\Psi}^2 q(\boldsymbol{Z})(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^{t-1}.\end{aligned}$$

For any $i \in \mathbb{N}$, we write $\Psi^i = p_i(\boldsymbol{\Psi}) + \mu_i \boldsymbol{I}$, where $\mu_i = \mathbb{E}(B-\kappa)^i,\ B \sim \mathsf{Bern}(\kappa)$, and $p_i(\psi) = \psi^i - \mu_i$. This polynomial satisfies $\mathbb{E}p_i(B-\kappa) = 0$. This gives us:

$$\begin{aligned}(q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{A}\boldsymbol{A}^\mathsf{T}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t = (q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1} q(\boldsymbol{Z}) p_3(\boldsymbol{\Psi}) q(\boldsymbol{Z})(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^{t-1} \\ + \kappa \boldsymbol{z}^\mathsf{T}(q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1} q(\boldsymbol{Z}) \boldsymbol{p}_2(\Psi) q(\boldsymbol{Z})(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^{t-1} \\ + (\mu_3 + \kappa\mu_2)\cdot (q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1} q(\boldsymbol{Z})^2 (\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^{t-1}.\end{aligned}$$

In the above display, the first two terms on the RHS are in the desired alternating product form. We center the last term. For any $i \in \mathbb{N}$ we define $q_i(z) = q^i(z) - \nu_i,\ \ \nu_i = \mathbb{E}q(\xi)^i,\ \xi \sim$

$\mathcal{N}(0,1)$. Hence, $q^i(\boldsymbol{Z}) = q_i(\boldsymbol{Z}) + \nu_i \boldsymbol{I}_m$. Hence:

$$\begin{aligned}(q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{A}\boldsymbol{A}^\mathsf{T}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t = (q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1} q(\boldsymbol{Z}) p_3(\boldsymbol{\Psi}) q(\boldsymbol{Z})(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^{t-1} \\ + \kappa \boldsymbol{z}^\mathsf{T}(q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1} q(\boldsymbol{Z}) \boldsymbol{p}_2(\Psi) q(\boldsymbol{Z})(\boldsymbol{\Psi} q(\boldsymbol{Z}))^{t-1} \\ + (\mu_3 + \kappa\mu_2)(q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1} q_2(\boldsymbol{Z})(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^{t-1} \\ + \nu_2\,(\mu_3 + \kappa\mu_2)(q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^{t-1}.\end{aligned}$$

In the above display, each of the terms in the right hand side is an alternating product except $(\mu_3 + \kappa\mu_2)\cdot(q(\boldsymbol{Z})\boldsymbol{\Psi})^{t-1}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^{t-1}$. We inductively center this term. Note that this centering procedure does not depend on the choice of the orthogonal matrix $\boldsymbol{U}$ used to generate the sensing matrix. Furthermore, the number of terms is bounded by $N_t \leq N_{t-1} + 3$, so $N_t \leq 1 + 3t$. Hence, we have obtained the desired decomposition:

$$(q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{A}\boldsymbol{A}^\mathsf{T}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t = c_0 \boldsymbol{I} + \sum_{i=1}^{N_t} c_i \mathcal{A}_i. \tag{5.14}$$

Therefore, we can write (T_3) as:

$$(T_3) = c_0 \frac{\|\boldsymbol{z}\|^2}{m} + \frac{1}{m}\sum_{i=1}^{N_t} c_i\ \boldsymbol{z}^\mathsf{T}\mathcal{A}_i\boldsymbol{z} = c_0 \frac{\|\boldsymbol{x}_\star\|^2}{m} + \frac{1}{m}\sum_{i=1}^{N_t} c_i\ \boldsymbol{z}^\mathsf{T}\mathcal{A}_i\boldsymbol{z}.$$

Observe that $\|\boldsymbol{x}_\star\|^2/m \xrightarrow{\text{P}} 1$, and Proposition 14 guarantees $\boldsymbol{z}^\mathsf{T}\mathcal{A}_i\boldsymbol{z}/m$ converges in probability to the same limit irrespective of whether $\boldsymbol{U} = \boldsymbol{O}$ or $\boldsymbol{U} = \boldsymbol{H}$. Hence, term (T_3) converges in probability to the same limit for both the subsampled Haar sensing and the subsampled Hadamard sensing model.

Next, we analyze term (T_4). Repeating the arguments we made for the analysis of the term

(T_2) we find:

$$(T_4) = \frac{\boldsymbol{z}^{\mathsf{T}}(q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t \cdot \boldsymbol{w}}{m}$$
$$\stackrel{\mathrm{d}}{=} \frac{\|(q(\boldsymbol{Z})\boldsymbol{\Psi})^t \boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}(\boldsymbol{\Psi}\cdot q(\boldsymbol{Z}))^t \boldsymbol{z}\|_2}{m} \cdot W \stackrel{\mathrm{P}}{\to} 0,$$

where $W \sim \mathcal{N}(0,1)$. Finally, we analyze the term (T_5). Using the decomposition (5.14) we have:

$$(T_5) = c_0 \frac{\|\boldsymbol{w}\|_2^2}{m} + \frac{1}{m}\sum_{i=1}^{N_t} c_i\ \boldsymbol{w}^{\mathsf{T}}\mathcal{A}_i\boldsymbol{w}.$$

We know that $\|\boldsymbol{w}\|_2^2/m \stackrel{\mathrm{P}}{\to} 1$. Hence, we focus on analyzing $\boldsymbol{w}^{\mathsf{T}}\mathcal{A}_i\boldsymbol{w}/m$. We decompose this as:

$$\frac{\boldsymbol{w}^{\mathsf{T}}\mathcal{A}_i\boldsymbol{w}}{m} = \frac{w^{\mathsf{T}}\mathcal{A}_i\boldsymbol{w} - \mathbb{E}[w^{\mathsf{T}}\mathcal{A}_i\boldsymbol{w}|\mathcal{A}_i]}{m} + \frac{\mathbb{E}[w^{\mathsf{T}}\mathcal{A}_i\boldsymbol{w}|\mathcal{A}_i]}{m}.$$

Observe that:

$$\frac{\mathbb{E}[\boldsymbol{w}^{\mathsf{T}}\mathcal{A}_i\boldsymbol{w}|\mathcal{A}_i]}{m} = \frac{\kappa\cdot\mathsf{Tr}(\mathcal{A}_i)}{m} \stackrel{\mathrm{P}}{\to} 0 \quad \text{(By Proposition 13)}.$$

On the other hand, using the Hanson-Wright Inequality (Fact 4) together with the estimates

$$\|\mathcal{A}_i\|_{\mathsf{op}} \le C(\mathcal{A}_i),\ \|\mathcal{A}_i\|_{\mathsf{Fr}} \le \sqrt{m}\cdot C(\mathcal{A}_i),$$

for a fixed constant $C(\mathcal{A}_i)$ (independent of m, n) depending only on the formula for $\mathcal{A}_i$, we obtain:

$$\mathbb{P}\left(\left|\boldsymbol{w}^{\mathsf{T}}\mathcal{A}_i\boldsymbol{w} - \mathbb{E}[w^{\mathsf{T}}\mathcal{A}_i\boldsymbol{w}|\mathcal{A}_i]\right| > mt \;\middle|\; \mathcal{A}_i\right) \le 2\exp\left(-\frac{c}{C(\mathcal{A}_i)}\cdot m\cdot\min(t,t^2)\right) \to 0.$$

Hence,

$$\frac{\boldsymbol{w}^{\mathsf{T}}\mathcal{A}_i\boldsymbol{w} - \mathbb{E}[w^{\mathsf{T}}\mathcal{A}_i\boldsymbol{w}|\mathcal{A}_i]}{m} \xrightarrow{\text{P}} 0.$$

This implies $(T_5) \xrightarrow{\text{P}} c_0$ for both the models. This proves the limit :

$$\text{p-lim}\, \frac{\|\hat{\boldsymbol{x}}^{(t)}\|_2^2}{m}$$

exists and is identical for the two sensing models, which concludes the proof of Theorem 6.

□

5.6 Key Ideas for the Proof of Propositions 13 and 14

In this section, we introduce some key ideas that are important in the proof of Propositions 13 and 14. Recall that we wish to analyze the limit in probability of the normalized trace and the quadratic form. A natural candidate for this limit is the limiting value of their expectation:

$$\text{p-lim}\, \frac{1}{m}\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}) \stackrel{?}{=} \lim_{m\to\infty} \frac{1}{m}\mathbb{E}\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}),$$
$$\text{p-lim}\, \frac{\langle \boldsymbol{z}, \mathcal{A}\boldsymbol{z}\rangle}{m} \stackrel{?}{=} \lim_{m\to\infty} \frac{\mathbb{E}\langle \boldsymbol{z}, \mathcal{A}\boldsymbol{z}\rangle}{m}.$$

In order to show this, one needs to show that the variance of the normalized trace and the normalized quadratic form converge to 0, which involves analyzing the second moment of these quantities. However, since the analysis of the second moment uses very similar ideas as the analysis of the expectation, we focus on outlining the main ideas in the context of the analysis of expectation.

First, we observe that alternating products can be simplified significantly due to the following property of polynomials of centered Bernoulli random variables.

Lemma 16. *For any polynomial p such that if $B \sim \mathsf{Bern}(\kappa)$, $\mathbb{E}\, p(B-\kappa) = 0$ we have,*

$$p(\mathbf{\Psi}) = (p(1-\kappa) - p(-\kappa)) \cdot \mathbf{\Psi}.$$

Proof. Observe that since $\mathbf{\Psi} = \boldsymbol{U}\overline{\boldsymbol{B}}\boldsymbol{U}^{\mathsf{T}}$, and $\boldsymbol{U}$ is orthogonal, we have $p(\mathbf{\Psi}) = \boldsymbol{U}p(\overline{\boldsymbol{B}})\boldsymbol{U}^{\mathsf{T}}$. Next, observe that:

$$\begin{aligned} p(\overline{B}_{ii}) &= p(1-\kappa)B_{ii} + p(-\kappa)(1-B_{ii}) \\ &= (p(1-\kappa) - p(-\kappa)) \cdot \overline{B}_{ii} + \underbrace{\kappa p(1-\kappa) + (1-\kappa)p(-\kappa)}_{=0}, \end{aligned}$$

where the last step follows from the assumption $\mathbb{E}\, p(B-\kappa) = 0$. Hence, $p(\overline{\boldsymbol{B}}) = (p(1-\kappa) - p(-\kappa))\overline{\boldsymbol{B}}$ and $p(\mathbf{\Psi}) = (p(1-\kappa) - p(-\kappa))\mathbf{\Psi}$. □

Hence, without loss of generality we can assume that each of the p_i in an alternating product satisfy $p_i(\xi) = \xi$.

5.6.1 Partitions

Note that the expected normalized trace and the expected quadratic form in Propositions 13 and 14 can be expanded as follows:

$$\begin{aligned} \frac{1}{m}\mathbb{E}\mathsf{Tr}\mathcal{A}(\mathbf{\Psi}, \boldsymbol{Z}) &= \frac{1}{m} \sum_{a_1, a_2, \ldots a_k = 1}^{m} \mathbb{E}[(\mathbf{\Psi})_{a_1,a_2} q_1(z_{a_2}) \cdots q_{k-1}(z_{a_k})(\mathbf{\Psi})_{a_k,a_1}], \\ \frac{\mathbb{E}\langle \boldsymbol{z}, \mathcal{A}\boldsymbol{z}\rangle}{m} &= \frac{1}{m} \sum_{a_{1:k+1} \in [m]} \mathbb{E}[z_{a_1}(\mathbf{\Psi})_{a_1,a_2} q_1(z_{a_2})(\mathbf{\Psi})_{a_2,a_3} \cdots q_{k-1}(z_{a_k})(\mathbf{\Psi})_{a_k,a_{k+1}} z_{a_{k+1}}]. \end{aligned}$$

Some Notation: Let $\mathcal{P}([k])$ denotes the set of all partitions of a discrete set $[k]$. We use $|\pi|$ to denote the number of blocks in π. Recall that a partition $\pi \in \mathcal{P}([k])$ is simply a collection of disjoint subsets of $[k]$ whose union is $[k]$ i.e.

$$\pi = \{\mathcal{V}_1, \mathcal{V}_2 \ldots \mathcal{V}_{|\pi|}\},\ \sqcup_{t=1}^{|\pi|} \mathcal{V}_t = [k].$$

The symbol $\sqcup$ is exclusively reserved for representing a set as a union of disjoint sets. For any element $s \in [k]$, we use the notation $\pi(s)$ to refer to the block that s lies in. That is, $\pi(s) = \mathcal{V}_i$ iff $s \in \mathcal{V}_i$. For any $\pi \in \mathcal{P}([k])$, define the set $\mathcal{C}(\pi)$ the set of all vectors $\boldsymbol{a} \in [m]^k$ which are constant exactly on the blocks of π:

$$\mathcal{C}(\pi) \stackrel{\text{def}}{=} \{\boldsymbol{a} \in [m]^k : a_s = a_t \Leftrightarrow \pi(s) = \pi(t)\}.$$

Consider any $\boldsymbol{a} \in \mathcal{C}(\pi)$. If $\mathcal{V}_i$ is a block in π, we use $a_{\mathcal{V}_i}$ to denote the unique value the vector $\boldsymbol{a}$ assigns to the all the elements of $\mathcal{V}_i$.

The rationale for introducing this notation is the observation that:

$$[m]^k = \bigsqcup_{\pi \in \mathcal{P}([k])} \mathcal{C}(\pi),$$

and hence we can write the normalized trace and quadratic forms as:

$$\frac{\mathbb{E}\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})}{m} = \frac{1}{m} \sum_{\pi \in \mathcal{P}([k])} \sum_{\boldsymbol{a} \in \mathcal{C}(\pi)} \mathbb{E}[(\boldsymbol{\Psi})_{a_1,a_2} q_1(z_{a_2}) \cdots q_{k-1}(z_{a_k})(\boldsymbol{\Psi})_{a_k,a_1}], \tag{5.15a}$$

$$\frac{\mathbb{E}\langle \boldsymbol{z}, \mathcal{A}\boldsymbol{z}\rangle}{m} = \frac{1}{m} \sum_{\pi \in \mathcal{P}([k+1])} \sum_{\boldsymbol{a} \in \mathcal{C}(\pi)} \mathbb{E}[z_{a_1}(\boldsymbol{\Psi})_{a_1,a_2} q_1(z_{a_2}) \cdots q_{k-1}(z_{a_k})(\boldsymbol{\Psi})_{a_k,a_{k+1}} z_{a_{k+1}}]. \tag{5.15b}$$

This idea of organizing the combinatorial calculations is due to Tulino, Caire, Shamai, and Verdú [82] and the rationale for doing so will be clear in a moment.

5.6.2 Concentration

Lemma 17. *Let the sensing matrix $\boldsymbol{A}$ be generated by sub-sampling an orthogonal matrix $\boldsymbol{U}$. We have, for any $a, b \in [m]$:*

$$\mathbb{P}\left(|\Psi_{ab}| \geq \epsilon | \boldsymbol{U}\right) \leq 4 \exp\left(-\frac{\epsilon^2}{8m\|\boldsymbol{U}\|_\infty^4}\right).$$

Proof. Recall that $\boldsymbol{\Psi} = \boldsymbol{U}(\boldsymbol{B} - \kappa \boldsymbol{I}_m)\boldsymbol{U}^{\mathsf{T}}$, where the distribution of the diagonal matrix

$$\boldsymbol{B} = \mathsf{Diag}\left(B_{11}, B_{22} \dots B_{mm}\right)$$

is described as follows: First draw a uniformly random subset $S \subset [m]$ with $|S| = n$ and set:

$$B_{ii} = \begin{cases} 0 & : i \notin S \\ & \\ 1 & : i \in S \end{cases}.$$

Due to the constraint that $\sum_{i=1}^{m} B_{ii} = n$, these random variables are not independent. In order to address this issue we couple $\boldsymbol{B}$ with another random diagonal matrix $\tilde{\boldsymbol{B}}$ generated as follows:

1. First sample $N \sim \mathsf{Binom}(m, \kappa)$.
2. Sample a subset $\tilde{S} \subset [m]$ with $|\tilde{S}| = N$ as follows:
   - If $N \leq n$, then set $\tilde{S}$ to be a uniformly random subset of S of size N.
   - If $N > n$ first sample a uniformly random subset A of S^c of size $N - n$ and set $\tilde{S} = S \cup A$.
3. Set $\tilde{\boldsymbol{B}}$ as follows:

$$\tilde{B}_{ii} = \begin{cases} 0 & : i \notin \tilde{S} \\ & \\ 1 & : i \in \tilde{S}. \end{cases}.$$

It is easy to check that conditional on N, $\tilde{S}$ is a uniformly random subset of $[m]$ with cardinality N. Since $N \sim \mathsf{Binom}(m, \kappa)$, we have $\tilde{B}_{ii} \overset{\text{i.i.d.}}{\sim} \mathsf{Bern}(\kappa)$. Define:

$$T \overset{\text{def}}{=} \Psi_{ab} = \boldsymbol{u}_a^{\mathsf{T}}(\boldsymbol{B} - \kappa \boldsymbol{I}_m)\boldsymbol{u}_b = \sum_{i=1}^{m} u_{ai} u_{bi} (B_{ii} - \mathbb{E}B_{ii}),$$

$$\tilde{T} \overset{\text{def}}{=} \boldsymbol{u}_a^{\mathsf{T}}(\tilde{\boldsymbol{B}} - \kappa \boldsymbol{I}_m)\boldsymbol{u}_b = \sum_{i=1}^{m} u_{ai} u_{bi} (\tilde{B}_{ii} - \mathbb{E}\tilde{B}_{ii}).$$

Observe that $|T - \tilde{T}| \leq |N - n| \|\boldsymbol{U}\|_\infty^2$. Hence,

$$\begin{aligned}
\mathbb{P}\left(|T| \geq \epsilon\right) &\leq \mathbb{P}\left(|\tilde{T}| \geq \frac{\epsilon}{2}\right) + \mathbb{P}\left(|T - \tilde{T}| \geq \frac{\epsilon}{2}\right) \\
&= \mathbb{P}\left(|\tilde{T}| \geq \frac{\epsilon}{2}\right) + \mathbb{P}\left(|N - \mathbb{E}N| \geq \frac{\epsilon}{2\|\boldsymbol{U}\|_\infty^2}\right) \\
&\overset{\text{(a)}}{\leq} 4 \exp\left(-\frac{\epsilon^2}{8m\|\boldsymbol{U}\|_\infty^4}\right).
\end{aligned}$$

In the step marked (a), we used Hoeffding's Inequality. □

Hence the above lemma shows that,

$$\|\boldsymbol{\Psi}\|_\infty \leq O\left(\sqrt{m}\|\boldsymbol{U}\|_\infty^2 \operatorname{polylog}(m)\right),$$

with high probability. Recall that in the subsampled Hadamard model $\boldsymbol{U} = \boldsymbol{H}$ and $\|\boldsymbol{H}\|_\infty = 1/\sqrt{m}$. Similarly, in the subsampled Haar model $\boldsymbol{U} = \boldsymbol{O}$ and $\|\boldsymbol{O}\|_\infty \leq O(\operatorname{polylog}(m)/\sqrt{m})$. Hence, we expect:

$$\|\boldsymbol{\Psi}\|_\infty \leq O\left(\frac{\operatorname{polylog}(m)}{\sqrt{m}}\right), \quad \text{with high probability.} \tag{5.16}$$

5.6.3 Mehler's Formula

Note that in order to compute the expected normalized trace and quadratic form as given in (5.15), we need to compute:

$$\begin{aligned}
&\mathbb{E}[(\boldsymbol{\Psi})_{a_1,a_2} q_1(z_{a_2}) \cdots q_{k-1}(z_{a_k})(\boldsymbol{\Psi})_{a_k,a_1}], \\
&\mathbb{E}[z_{a_1}(\boldsymbol{\Psi})_{a_1,a_2} q_1(z_{a_2})(\boldsymbol{\Psi})_{a_2,a_3} \cdots q_{k-1}(z_{a_k})(\boldsymbol{\Psi})_{a_k,a_{k+1}} z_{a_{k+1}}].
\end{aligned}$$

Note that by the Tower property:

$$\mathbb{E}[(\boldsymbol{\Psi})_{a_1,a_2} q_1(z_{a_2}) \cdots q_{k-1}(z_{a_k})(\boldsymbol{\Psi})_{a_k,a_1}] = \mathbb{E}\left[(\boldsymbol{\Psi})_{a_1,a_2} \cdots (\boldsymbol{\Psi})_{a_k,a_1} \mathbb{E}[q_1(z_{a_2}) \cdots q_{k-1}(z_{a_k})|\boldsymbol{A}]\right],$$

and analogously for $\mathbb{E}[z_{a_1}(\boldsymbol{\Psi})_{a_1,a_2}q_1(z_{a_2})(\boldsymbol{\Psi})_{a_2,a_3}\cdots q_{k-1}(z_{a_k})(\boldsymbol{\Psi})_{a_k,a_{k+1}}z_{a_{k+1}}]$. Suppose that $\boldsymbol{a} \in \mathcal{C}(\pi)$ for some $\pi \in \mathcal{P}([k])$. Let $\pi = \mathcal{V}_1 \sqcup \mathcal{V}_2 \cdots \sqcup \mathcal{V}_{|\pi|}$. Define:

$$F_{\mathcal{V}_i}(\xi) = \prod_{\substack{j \in \mathcal{V}_i \\ j \neq 1}} q_{j-1}(\xi).$$

Then, we have:

$$\mathbb{E}[q_1(z_{a_2})\cdots q_{k-1}(z_{a_k})|\boldsymbol{A}] = \mathbb{E}\left[\prod_{i=1}^{|\pi|} F_{\mathcal{V}_i}(z_{a_{\mathcal{V}_i}})\middle|\boldsymbol{A}\right].$$

In order to compute the conditional expectation we observe that conditional on $\boldsymbol{A}$, $\boldsymbol{z}$ is a zero mean Gaussian vector with covariance:

$$\mathbb{E}[\boldsymbol{z}\boldsymbol{z}^{\mathsf{T}}|\boldsymbol{A}] = \frac{1}{\kappa}\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} = \frac{1}{\kappa}\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^{\mathsf{T}} = \boldsymbol{I} + \frac{\boldsymbol{\Psi}}{\kappa}.$$

Note that since $a_{\mathcal{V}_i} \neq a_{\mathcal{V}_j}$ for $i \neq j$, we have as a consequence of (5.16), $\{z_{a_{\mathcal{V}_i}}\}_{i=1}^{|\pi|}$ are weakly correlated Gaussians. Hence we expect,

$$\mathbb{E}[q_1(z_{a_2})\cdots q_{k-1}(z_{a_k})|\boldsymbol{A}] = \prod_{i=1}^{|\pi|}\mathbb{E}_{Z\sim\mathcal{N}(0,1)}F_{\mathcal{V}_i}(Z) + \text{ A small error term,}$$

where the error term is a term that goes to zero as $m \to \infty$. Mehler's formula given in the proposition below provides an explicit formula for the error term. Observe that in (5.15):

1. the sum over $\pi \in \mathcal{P}([k])$ cannot cause the error terms to add up since $|\mathcal{P}([k])|$ is a constant depending on k but independent of m.

2. On the other hand, the sum over $\boldsymbol{a} \in \mathcal{C}(\pi)$ can cause the errors to add up since:

$$|\mathcal{C}(\pi)| = m \cdot (m-1)\cdots(m-|\pi|+1).$$

It is not obvious right away how accurately the error must be estimated, but it turns out that for the proof of Proposition 13 it suffices to estimate the order of magnitude of the error term. For the proof of Proposition 14 we need to be more accurate and the leading order term in the error needs to be tracked precisely.

Before we state Mehler's formula we recall some preliminaries regarding Fourier analysis on the Gaussian space. Let $Z \sim \mathcal{N}(0,1)$. Let $f : \mathbb{R} \to \mathbb{R}$ be such that $\mathbb{E}f^2(Z) < \infty$, i.e. $f \in L^2(\mathcal{N}(0,1))$. The Hermite polynomials $\{H_j : j \in \mathbb{N}_0\}$ form an orthogonal polynomial basis for $L^2(\mathcal{N}(0,1))$. The polynomial H_j is a degree j polynomial. They satisfy the orthogonality property:

$$\mathbb{E}H_i(Z)H_j(Z) = i! \cdot \delta_{ij}.$$

The first few Hermite polynomials are given by:

$$H_0(z) = 1,\ H_1(z) = z,\ H_2(z) = z^2 - 1.$$

Proposition 15 (Mehler [61] and Slepian [62])**.** *Consider a k dimensional Gaussian vector $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, such that $\Sigma_{ii} = 1$ for all $i \in [k]$. Let $f_1, f_2, \ldots, f_k : \mathbb{R} \to \mathbb{R}$ be k arbitrary functions whose absolute value can be upper bounded by a polynomial. Then,*

$$\left| \mathbb{E}\left[\prod_{i=1}^{k} f_i(z_i)\right] - \sum_{\substack{\boldsymbol{w} \in \mathcal{G}(k) \\ \|\boldsymbol{w}\| \leq t}} \left(\prod_{i=1}^{k} \hat{f}_i(\mathsf{d}_i(\boldsymbol{w}))\right) \cdot \frac{\boldsymbol{\Sigma}^{\boldsymbol{w}}}{\boldsymbol{w}!} \right| \leq C\left(1 + \frac{1}{\lambda_{\min}^{4t+4}(\boldsymbol{\Sigma})}\right)\left(\max_{i \neq j} |\Sigma_{ij}|\right)^{t+1},$$

where:

1. *$\mathcal{G}(k)$ denotes the set of undirected weighted graphs with non-negative integer weights on k nodes with no self loops.*

2. *An element $\boldsymbol{w} \in \mathcal{G}(k)$ is represented by a $k \times k$ symmetric matrix $\boldsymbol{w}$ with $w_{ij} = w_{ji} \in$*

$\mathbb{N} \cup \{0\}$, *and* $w_{ii} = 0$.

3. $\mathsf{d}_i(\boldsymbol{w})$ *denotes the degree of node* i*:* $\mathsf{d}_i(\boldsymbol{w}) = \sum_{j=1}^{k} w_{ij}$.

4. $\|\boldsymbol{w}\|$ *denotes the total weight of the graph defined as:*

$$\|\boldsymbol{w}\| \stackrel{def}{=} \sum_{i<j} w_{ij} = \frac{1}{2} \sum_{i=1}^{k} \mathsf{d}_i(\boldsymbol{w}).$$

5. *The coefficients* $\hat{f}_i(j)$ *are defined as:* $\hat{f}_i(j) = \mathbb{E} f_i(Z) H_j(Z)$ *where* $Z \sim \mathcal{N}(0, 1)$.

6. $\boldsymbol{\Sigma}^{\boldsymbol{w}}, \boldsymbol{w}!$ *denote the entry-wise powering and factorial:*

$$\boldsymbol{\Sigma}^{\boldsymbol{w}} = \prod_{i<j} \Sigma_{ij}^{w_{ij}}, \ \boldsymbol{w}! = \prod_{i<j} w_{ij}!$$

7. $C = C_{t,k,f_{1:k}}$ *is a finite constant depending only on the* t, k*, and the functions* $f_{1:k}$ *but is independent of* $\boldsymbol{\Sigma}$.

This result is essentially due to Mehler [61] in the case $k = 2$, and the result for general k was obtained by Slepian [62]. Actually the results of these authors show that the pdf of $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ denoted by $\psi(\boldsymbol{z}; \boldsymbol{\Sigma})$ has the following Taylor expansion around $\boldsymbol{\Sigma} = \boldsymbol{I}_k$:

$$\psi(\boldsymbol{z}; \boldsymbol{\Sigma}) = \psi(\boldsymbol{z}; \boldsymbol{I}_k) \cdot \left(\sum_{\boldsymbol{w} \in \mathcal{G}(k)} \frac{\boldsymbol{\Sigma}^{\boldsymbol{w}}}{\boldsymbol{w}!} \cdot \prod_{i=1}^{k} H_{\mathsf{d}_i(\boldsymbol{w})}(z_i) \right).$$

In Appendix C.5 of the supplementary materials we check that this Taylor's expansion can be integrated, and estimate the truncation error to obtain Proposition 15.

At this point, we have introduced all the tools used in the proof of Proposition 13 and we refer the reader to Section 5.7 for the proof of Proposition 13.

5.6.4 Central Limit Theorem

We introduce the following definition.

Definition 8 (Matrix Moment)**.** *Let $\boldsymbol{M}$ be a symmetric matrix. Given:*

1. *A partition $\pi \in \mathcal{P}([k])$ with blocks $\pi = \{\mathcal{V}_1, \mathcal{V}_2, \cdots, \mathcal{V}_{|\pi|}\}$.*
2. *A $k \times k$ symmetric weight matrix $\boldsymbol{w} \in \mathcal{G}(k)$ with non-negative valued entries and $w_{ii} = 0 \; \forall \; i \in [k]$.*
3. *A vector $\boldsymbol{a} \in \mathcal{C}(\pi)$.*

Define the $(\boldsymbol{w}, \pi, \boldsymbol{a})$ - matrix moment of the matrix $\boldsymbol{M}$ as:

$$\mathcal{M}(\boldsymbol{M}, \boldsymbol{w}, \pi, \boldsymbol{a}) \stackrel{def}{=} \prod_{i,j \in [k], i<j} M_{a_i, a_j}^{w_{ij}}.$$

By defining:

$$W_{st}(\boldsymbol{w}, \pi) \stackrel{def}{=} \sum_{\substack{i,j \in [k], i<j \\ \{\pi(i), \pi(j)\} = \{\mathcal{V}_s, \mathcal{V}_t\}}} w_{ij},$$

we can write $\mathcal{M}(\boldsymbol{M}, \boldsymbol{w}, \pi, \boldsymbol{a})$ in the form:

$$\mathcal{M}(\boldsymbol{M}, \boldsymbol{w}, \pi, \boldsymbol{a}) = \prod_{\substack{s,t \in [|\pi|] \\ s \leq t}} M_{a_{\mathcal{V}_s}, a_{\mathcal{V}_t}}^{W_{st}(\boldsymbol{w}, \boldsymbol{\pi})}.$$

Remark 14 (Graph Interpretation)**.** *It is often useful to interpret the tuple $(\boldsymbol{w}, \pi, \boldsymbol{a})$ in terms of graphs:*

1. *$\boldsymbol{w}$ represents the adjacency matrix of an undirected weighted graph on the vertex set $[k]$ with no self-edges $(w_{ii} = 0)$. We say an edge exists between nodes $i, j \in [k]$ if $w_{ij} \geq 1$ and the weight of the edge is given by w_{ij}.*
2. *The partition π of the vertex set $[k]$ represents a community structure on the graph. Two vertices $i, j \in [k]$ are in the same community iff $\pi(i) = \pi(j)$.*

3. *$\boldsymbol{a}$ represents a labelling of the vertices $[k]$ with labels in the set $[m]$ which respects the community structure.*

4. *The weights $W_{st}(\boldsymbol{w}, \pi)$ simply denote the total weight of edges between communities s, t.*

The rationale for introducing this definition is as follows: When we use Mehler's formula to compute $\mathbb{E}[q_1(z_{a_2}) \cdots q_{k-1}(z_{a_k})|\boldsymbol{A}]$ and $\mathbb{E}[z_{a_1} q_1(z_{a_2}) \cdots q_{k-1}(z_{a_k}) z_{a_{k+1}}|\boldsymbol{A}]$, and substitute the resulting expression in (5.15), it expresses:

$$\frac{\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})}{m}, \ \frac{\mathbb{E}\langle \boldsymbol{z}, \mathcal{A}\boldsymbol{z}\rangle}{m},$$

in terms of the matrix moments $\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})$.

For the proof of Proposition 13 it suffices to upper bound $|\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})|$. We do so in the following lemma.

Lemma 18. *Consider an arbitrary matrix moment $\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})$ of $\boldsymbol{\Psi}$. There exists a universal constant C (independent of $m, \boldsymbol{a}, \pi, \boldsymbol{w}$) such that,*

$$\mathbb{E}|\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})| \leq \left(\sqrt{\frac{C\|\boldsymbol{w}\| \log^2(m)}{m}} \right)^{\|\boldsymbol{w}\|},$$

for both the sub-sampled Haar and the sub-sampled Hadamard sensing model.

The claim of the lemma is not surprising in light of (5.16). The complete proof follows from the concentration inequality in Lemma 17, which can be found in Appendix C.3.1.

On the other hand, to prove Proposition 14 we need a more refined analysis and we need to estimate the leading order term in $\mathbb{E}\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})$. In order to do so, we first consider any fixed entry of $\sqrt{m}\boldsymbol{\Psi}$:

$$\sqrt{m}\Psi_{ab} = \sqrt{m}(\boldsymbol{U}\overline{\boldsymbol{B}}\boldsymbol{U}^{\mathsf{T}})_{ab} = \sum_{i=1}^{m} \sqrt{m} \cdot u_{ai} \cdot u_{bi}(B_{ii} - \kappa).$$

Observe that:

1. $B_{ii} - \kappa$ are centered and weakly dependent.
2. $\sqrt{m} u_{ai} u_{bi} = O(m^{-\frac{1}{2}})$ under the sub-sampled Haar model and the sub-sampled Hadamard model.

Consequently, we expect $\sqrt{m}\Psi_{ab}$ to converge to a Gaussian random variable and hence, we expect that:

$$\mathbb{E}\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})$$

to converge to a suitable Gaussian moment. In order to show that the normalized quadratic form $\mathbb{E}\langle \boldsymbol{z}, \mathcal{A}\boldsymbol{z}\rangle/m$ converges to the same limit under both the sensing models, we need to understand what is the limiting value of $\mathbb{E}\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})$ under both the models. Understanding this uses the following simple but important property of Hadamard matrices.

Lemma 19. *For any $i, j \in [m]$, we have:*

$$\sqrt{m}\boldsymbol{h}_i \odot \boldsymbol{h}_j = \boldsymbol{h}_{i \oplus j},$$

where $\odot$ denotes the entry-wise multiplication of vectors, and $i \oplus j \in [m]$ denotes the result of the following computation:

Step 1: *Compute $\boldsymbol{i}, \boldsymbol{j} \in \{0,1\}^m$ which are the binary representations of $(i-1)$ and $(j-1)$ respectively.*

Step 2: *Compute $\boldsymbol{i} + \boldsymbol{j}$ by adding $\boldsymbol{i}, \boldsymbol{j}$ bit-wise (modulo 2).*

Step 3: *Compute the number in $[0 : m-1]$ whose binary representation is given by $\boldsymbol{i} + \boldsymbol{j}$.*

Step 4: *Add one to the number obtained in Step 3 to obtain $i \oplus j \in [m]$.*

Proof. Recall by the definition of the Hadamard matrix, we have,

$$h_{ik} = \frac{1}{\sqrt{m}}(-1)^{\langle \boldsymbol{i},\boldsymbol{k}\rangle},\ h_{jk} = \frac{1}{\sqrt{m}}(-1)^{\langle \boldsymbol{j},\boldsymbol{k}\rangle}.$$

Hence,

$$\sqrt{m}(\boldsymbol{h}_i \odot \boldsymbol{h}_j)_k = \frac{(-1)^{\langle \boldsymbol{i}+\boldsymbol{j},\boldsymbol{k}\rangle}}{\sqrt{m}} = (\boldsymbol{h}_{i\oplus j})_k,$$

as claimed. □

Due to the structure in Hadamard matrices, $\mathbb{E}\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})$ might not always converge to the same limit under the subsampled Haar and the Hadamard models. There are two kinds of exceptions:

Exception 1: Note that for the subsampled Hadamard Model,

$$\sqrt{m}\Psi_{aa} = \sqrt{m}\sum_{i=1}^{m}\overline{B}_{ii}|h_{ai}|^2 = \frac{1}{\sqrt{m}}\sum_{i=1}^{m}\overline{B}_{ii} = 0.$$

In contrast, under the subsampled Haar model, it can be shown that $\sqrt{m}\Psi_{aa}$ converges to a non-degenerate Gaussian. These exceptions are ruled out by requiring the weight matrix $\boldsymbol{w}$ to be dissassortative with respect to π (See definition below).

Exception 2: Define $\overline{\boldsymbol{b}} \in \mathbb{R}^m$ to be the vector formed by the diagonal entries of $\overline{\boldsymbol{B}}$. Observe that for the subsampled Hadamard model:

$$\sqrt{m}\Psi_{ab} = \langle \overline{\boldsymbol{b}}, \sqrt{m}\boldsymbol{h}_a \odot \boldsymbol{h}_b\rangle = \langle \overline{\boldsymbol{b}}, \boldsymbol{h}_{a\oplus b}\rangle.$$

Consequently, if two distinct pairs (a_1, b_1) and (a_2, b_2) are such that $a_1 \oplus b_1 = a_2 \oplus b_2$, then $\sqrt{m}\Psi_{a_1,b_1}$ and $\sqrt{m}\Psi_{a_2,b_2}$ are perfectly correlated in the subsampled Hadamard model. In contrast, unless $(a_1, b_1) = (a_2, b_2)$, it can be shown they are asymptotically uncorrelated in

the subsampled Haar model. This exception is ruled out by requiring the labelling $\boldsymbol{a}$ to be conflict free with respect to $(\boldsymbol{w}, \pi)$ (defined below).

Definition 9 (Disassortative Graphs)**.** *We say the weight matrix $\boldsymbol{w}$ is disassortative with respect to the partition π if: $\forall\ i, j \in [k],\ i < j$ such that $\pi(i) = \pi(j)$, we have $w_{ij} = 0$. This is equivalent to $W_{ss}(\boldsymbol{w}, \pi) = 0$ for all $s \in [|\pi|]$. In terms of the graph interpretation, this means that there are no intra-community edges in the graph. For any $\pi \in \mathcal{P}([k])$,we denote the set of all weight matrices dissortive with respect to π by $\mathcal{G}_{\mathsf{DA}}(\pi)$:*

$$\mathcal{G}_{\mathsf{DA}}(\pi) \stackrel{def}{=} \{\boldsymbol{w} \in \mathcal{G}(k) : W_{ss}(\boldsymbol{w}, \pi) = 0\ \forall\ s\ \in [|\pi|]\}.$$

Definition 10 (Conflict Freeness)**.** *Let $\pi \in \mathcal{P}([k])$ be a partition and let $\boldsymbol{w} \in \mathcal{G}_{\mathsf{DA}}(\pi)$ be a weight matrix disassortative with respect to π. Let $s_1 < t_1$ and $s_2 < t_2$ be distinct pairs of communities: $s_1, s_2, t_1, t_2 \in [|\pi|]$, $(s_1, t_1) \neq (s_2, t_2)$. We say a labelling $\boldsymbol{a} \in \mathcal{C}(\pi)$ has a conflict between distinct community pairs (s_1, t_1) and (s_2, t_2) if:*

1. $W_{s_1,t_1}(\boldsymbol{w}, \pi) \geq 1,\ W_{s_2,t_2}(\boldsymbol{w}, \pi) \geq 1.$
2. $a_{\mathcal{V}_{s_1}} \oplus a_{\mathcal{V}_{t_1}} = a_{\mathcal{V}_{s_2}} \oplus a_{\mathcal{V}_{t_2}}.$

We say a labelling $\boldsymbol{a}$ is conflict-free if it has no conflicting community pairs. The set of all conflict free labellings of $(\boldsymbol{w}, \pi)$ is denoted by $\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}, \pi)$.

The following two propositions show that if Exception 1 and Exception 2 are ruled out, then indeed $\mathbb{E}\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})$ converges to the same Gaussian moment under both the subsampled Haar and the Hadamard models.

Proposition 16. *Consider the sub-sampled Haar model ($\boldsymbol{\Psi} = \boldsymbol{O}\overline{\boldsymbol{B}}\boldsymbol{O}^{\mathsf{T}}$). Fix a partition $\pi \in \mathcal{P}(k)$ and a weight matrix $\boldsymbol{w} \in \mathcal{G}(k)$. Then, there exist constants $K_1, K_2, K_3 > 0$ depending only on*

$\|\boldsymbol{w}\|$ (independent of m), such that for any $\boldsymbol{a} \in \mathcal{C}(\pi)$ we have:

$$\left| \mathbb{E}\, \mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) - \prod_{\substack{s,t \in [|\pi|] \\ s \leq t}} \mathbb{E}\left[Z_{st}^{W_{st}(\boldsymbol{w},\pi)} \right] \right| \leq \frac{K_1 \log^{K_2}(m)}{m^{\frac{1}{4}}},\ \forall\, m \geq K_3.$$

In the above display, $Z_{st},\ s \leq t,\ s,t\ \in\ [|\pi|]$ are independent Gaussian random variables with the distribution:

$$Z_{st} \sim \begin{cases} s < t: & \mathcal{N}\left(0, \kappa(1-\kappa)\right) \\ s = t: & \mathcal{N}\left(0, 2\kappa(1-\kappa)\right) \end{cases}.$$

Proposition 17. *Consider the sub-sampled Hadamard model ($\boldsymbol{\Psi} = \boldsymbol{H}\overline{\boldsymbol{B}}\boldsymbol{H}^{\mathsf{T}}$). Fix a partition $\pi \in \mathcal{P}(k)$ and a weight matrix $\boldsymbol{w} \in \mathbb{N}_0^{k \times k}$. Then,*

1. *Suppose that $\boldsymbol{w} \notin \mathcal{G}_{\mathsf{DA}}(\pi)$, then,*

$$\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) = 0.$$

2. *Suppose that $\boldsymbol{w} \in \mathcal{G}_{\mathsf{DA}}(\pi)$. Then, there exist constants $K_1, K_2, K_3 > 0$ depending only on $\|\boldsymbol{w}\|$ (independent of m), such that for any conflict free labelling $\boldsymbol{a} \in \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}, \pi)$, we have:*

$$\left| \mathbb{E}\, \mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) - \prod_{\substack{s,t \in [|\pi|] \\ s < t}} \mathbb{E}\left[Z_{\kappa}^{W_{st}(\boldsymbol{w},\pi)} \right] \right| \leq \frac{K_1 \log^{K_2}(m)}{m^{\frac{1}{4}}},\ \forall\, m \geq K_3.$$

In the above display, $Z_\kappa \sim \mathcal{N}\left(0, \kappa(1-\kappa)\right)$.

The proof of these Propositions can be found in Appendix C.3.2 in the supplementary materials. The proofs use a coupling argument to replace the weakly dependent diagonal matrix $\overline{\boldsymbol{B}}$ with a i.i.d. diagonal entries (as in the proof of Lemma 17) along with a classical Berry Eseen inequality due to Bhattacharya [83].

Finally, in order to finish the proof of Proposition 14 regarding the universality of the normalized quadratic form we need to argue the exceptional labellings for which $\mathbb{E}\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})$ doesn't converge to the same Gaussian moment under the sub-sampled Hadamard and Haar models are an asymptotically negligible fraction of the total labellings.

Lemma 20. *Let $\pi \in \mathcal{P}([k])$ be a partition and $\boldsymbol{w} \in \mathcal{G}_{\mathsf{DA}}(\pi)$ be a weight matrix disassortative with respect to π. We have, $|\mathcal{C}(\pi)\backslash\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}, \pi)| \leq |\pi|^4 \cdot m^{|\pi|-1}$, and*

$$\lim_{m\to\infty} \frac{\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}, \pi)}{m^{|\pi|}} = 1.$$

Proof. Let $(s_1, t_1) \neq (s_2, t_2)$ be two distinct community pairs such that:

$$W_{s_1,t_1}(\boldsymbol{w}, \pi) \geq 1,\ W_{s_2,t_2}(\boldsymbol{w}, \pi) \geq 1.$$

Let $\mathcal{L}_{(\mathsf{s}_1,\mathsf{t}_1;\mathsf{s}_2,\mathsf{t}_2)}(\boldsymbol{w}, \pi)$ denote the set of all labellings $\boldsymbol{a} \in \mathcal{C}(\pi)$ that have a conflict between distinct community pairs (s_1, t_1) and (s_2, t_2):

$$\mathcal{L}_{(\mathsf{s}_1,\mathsf{t}_1;\mathsf{s}_2,\mathsf{t}_2)}(\boldsymbol{w}, \pi) \stackrel{\text{def}}{=} \{\boldsymbol{a} \in \mathcal{C}(\pi) : a_{\mathcal{V}_{s_1}} \oplus a_{\mathcal{V}_{t_1}} = a_{\mathcal{V}_{s_2}} \oplus a_{\mathcal{V}_{t_2}}\}.$$

Then, we note that

$$\mathcal{C}(\pi)\backslash\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}, \pi) = \bigcup_{s_1,t_1,s_2,t_2} \mathcal{L}_{(\mathsf{s}_1,\mathsf{t}_1;\mathsf{s}_2,\mathsf{t}_2)}(\boldsymbol{w}, \pi),$$

where the union ranges over s_1, t_1, s_2, t_2 such that $1 \leq s_1 < t_1 \leq |\pi|, 1 \leq s_2 < t_2 \leq |\pi|$ and $(s_1, t_1) \neq (s_2, t_2)$ and $W_{s_1,t_1}(\boldsymbol{w}, \pi) \geq 1, W_{s_2,t_2}(\boldsymbol{w}, \pi) \geq 1$. Next, we bound $|\mathcal{L}_{(\mathsf{s}_1,\mathsf{t}_1;\mathsf{s}_2,\mathsf{t}_2)}(\boldsymbol{w}, \pi)|$. Since we know that $(s_1, t_1) \neq (s_2, t_2)$ and $s_1 < t_1$ and $s_2 < t_2$ out of the 4 indices s_1, t_1, s_2, t_2, there must be one index which is different from all the others. Let us assume that this index is t_2 (the remaining cases are analogous). To count $|\mathcal{L}_{(\mathsf{s}_1,\mathsf{t}_1;\mathsf{s}_2,\mathsf{t}_2)}(\boldsymbol{w}, \pi)|$ we assign labels to all blocks of π except t_2. The number of ways of doing so is at most $m^{|\pi|-1}$. After we do so, we note that $a_{\mathcal{V}_{t_2}}$

is uniquely determined by the constraint:

$$a_{\mathcal{V}_{s_1}} \oplus a_{\mathcal{V}_{t_1}} = a_{\mathcal{V}_{s_2}} \oplus a_{\mathcal{V}_{t_2}}.$$

Hence, $|\mathcal{L}_{(\mathsf{s}_1,\mathsf{t}_1;\mathsf{s}_2,\mathsf{t}_2)}(\boldsymbol{w},\pi)| \leq m^{|\pi|-1}$. Therefore,

$$|\mathcal{C}(\pi)\backslash\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w},\pi)| = \sum_{s_1,t_1,s_2,t_2} |\mathcal{L}_{(\mathsf{s}_1,\mathsf{t}_1;\mathsf{s}_2,\mathsf{t}_2)}(\boldsymbol{w},\pi)| \leq |\pi|^4 m^{|\pi|-1}.$$

Finally, we note that,

$$|\mathcal{C}(\pi)| - |\mathcal{C}(\pi)\backslash\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w},\pi)| = |\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w},\pi)| \leq |\mathcal{C}(\pi)|.$$

$|\mathcal{C}(\pi)|$ is given by:

$$|\mathcal{C}(\pi)| = m(m-1)\cdots(m-|\pi|+1) = m^{|\pi|}\cdot(1+o_m(1)).$$

Combining this with the already obtained upper bound $|\mathcal{C}(\pi)\backslash\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w},\pi)| \leq |\pi|^4\cdot m^{|\pi|-1}$, we obtain the second claim of the lemma. □

We now have all the tools required to finish the proof of Proposition 14 and we refer the reader to Section 5.8 for the proof of this result.

5.7 Proof of Proposition 13

In this Section we prove Proposition 13.

Let us consider a fixed alternating product $\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})$ as given in Definition 7. As a consequence of Lemma 16 we can assume that all the polynomials $p_i(\xi) = \xi$. We begin by stating a few intermediate lemmas which will be used to prove Proposition 13.

Lemma 21 (A high probability event)**.** *Let $\boldsymbol{U}$ denote the $m\times m$ orthogonal matrix used to generate*

the sensing matrix . Define the event:

$$\mathcal{E} = \left\{ \max_{i \neq j} |(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}|)_{ij} \leq \sqrt{32 \cdot m \cdot \|\boldsymbol{U}\|_{\infty}^{4} \cdot \log(m)}, \right. \\ \left. \max_{i \in [m]} |(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})_{ii} - \kappa| \leq \sqrt{32 \cdot m \cdot \|\boldsymbol{U}\|_{\infty}^{4} \cdot \log(m)} \right\}. \tag{5.17}$$

Then,

$$\mathbb{P}(\mathcal{E}|\boldsymbol{U}) \geq 1 - 4/m^2.$$

Furthermore, for the subsampled Haar model, when $\boldsymbol{U} = \boldsymbol{O} \sim$ Unif $\big(\mathbb{O}(m)\big)$, we have:

$$\mathbb{P}\left(\left\{ \|\boldsymbol{O}\|_{\infty} \leq \sqrt{\frac{8\log(m)}{m}} \right\} \cap \mathcal{E} \right) \geq 1 - 6/m^2.$$

The above Lemma follows from the concentration result in Lemma 17 and a union bound. Complete details are provided in Appendix C.1 in the supplementary materials.

Lemma 22 (A Continuity Estimate)**.** *Let $\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})$ be an alternating product of the matrices $\boldsymbol{\Psi}, \boldsymbol{Z}$ (see Definition 7). Then the map $\boldsymbol{Z} \mapsto \mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})/m$ is Lipchitz in Z, i.e. for any two diagonal matrices $\boldsymbol{Z} =$ Diag $(z_1, z_2 \ldots, z_m)$, $\boldsymbol{Z}' =$ Diag $(z_1', z_2' \ldots, z_m')$ we have:*

$$\left| \frac{\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})}{m} - \frac{\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}')}{m} \right| \leq \frac{C(\mathcal{A})}{\sqrt{m}} \cdot \|\boldsymbol{Z} - \boldsymbol{Z}'\|_{\mathsf{Fr}},$$

where $C(\mathcal{A})$ denotes a constant depending only on the formula for the alternating product $\mathcal{A}$ (independent of m, n).

This lemma follows from a straightforward computation provided in C.1 in the supplementary materals.

Lemma 23 (Analysis of Expectation)**.** *Let the sensing matrix $\boldsymbol{A}$ be drawn either from the subsam-*

pled Haar model or be generated using a deterministic orthogonal matrix $\boldsymbol{U}$ with the property:

$$\|\boldsymbol{U}\|_\infty \leq \sqrt{\frac{K_1 \log^{K_2}(m)}{m}},$$

for some universal constants $K_1, K_2 \geq 0$, then, we have:

$$\frac{1}{m}\mathbb{E}[\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}))|\boldsymbol{A}] \xrightarrow{P} 0.$$

Lemma 24 (Analysis of Variance)**.** *Let $\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})$ be any alternating product of the matrices $\boldsymbol{\Psi}, \boldsymbol{Z}$. Then,*

$$\mathrm{Var}\left(\frac{\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})}{m}\middle|\boldsymbol{A}\right) \leq \frac{C(\mathcal{A})}{n},$$

where $C(\mathcal{A})$ denotes a constant depending only on the formula for the alternating product $\mathcal{A}$ (independent of m, n).

Proofs of Lemmas 23 and 24 can be found at Section 5.7.1. Before moving forward to the proofs of these lemmas, let us conclude the proof of Proposition 13 assuming Lemmas 23 and 24 are true.

Proof of Proposition 13. We write $\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}))/m$ as:

$$\frac{\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}))}{m} = \mathbb{E}\left[\frac{\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}))}{m}\middle|\boldsymbol{A}\right] + \left(\frac{\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}))}{m} - \mathbb{E}\left[\frac{\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}))}{m}\middle|\boldsymbol{A}\right]\right).$$

We will show each of the two terms on the right hand side converge to zero in probability. Lemma 23 already gives:

$$\mathbb{E}\left[\frac{\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}))}{m}\middle|\boldsymbol{A}\right] \xrightarrow{\mathrm{P}} 0.$$

On the other hand, by Chebychev's Inequality and Lemma 24 we have:

$$\mathbb{P}\left[\left|\frac{\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z}))-\mathbb{E}[\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z}))|\boldsymbol{A}]}{m}\right|>\epsilon\middle|\boldsymbol{A}\right]\leq\frac{1}{\epsilon^2}\cdot\mathrm{Var}\left(\frac{\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})}{m}\middle|\boldsymbol{A}\right)\leq\frac{C(\mathcal{A})}{n\epsilon^2}.$$

Hence,

$$\mathbb{P}\left[\left|\frac{\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z}))-\mathbb{E}[\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z}))|\boldsymbol{A}]}{m}\right|>\epsilon\right]\to 0.$$

This concludes the proof of the proposition. □

5.7.1 Proof of Lemmas 23 and 24

Proof of Lemma 23. Recall the notation regarding partitions introduced in Section 5.6.1. We will organize the proof into various steps.

Step 1: Restricting to a Good Event. We first observe that $\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z}))/m$ is uniformly bounded:

$$\frac{\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})}{m}\leq\|\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\|_{\mathsf{op}}\leq\prod_{i=1}^{k}\|q_i\|_\infty=C(\mathcal{A})<\infty,$$

where $\|q_i\|_\infty=\sup_{\xi\in\mathbb{R}}|q_i(\xi)|$, and $C(\mathcal{A})$ denotes a finite constant independent of m,n. Recall the definition of $\mathcal{E}$ in (5.17). If the sensing matrix $\boldsymbol{A}$ was generated by subsampling a deterministic orthogonal matrix $\boldsymbol{U}$ with the property

$$\|\boldsymbol{U}\|_\infty\leq\sqrt{\frac{K_1\log^{K_2}(m)}{m}},$$

then Lemma 21 gives $\mathbb{P}(\mathcal{E}^c)\leq 4/m^2$. On the other hand, if $\boldsymbol{A}$ was generated by subsampling a uniformly random column orthogonal matrix $\boldsymbol{O}$ then we set $K_1=8,K_2=1$ and Lemma 21 gives $\mathbb{P}(\mathcal{E}^c)\leq 6/m^2$. Using this event, we decompose $\mathbb{E}[\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})|\boldsymbol{A}]/m$ as:

$$\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})|\boldsymbol{A}]}{m}=\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})|\boldsymbol{A}]}{m}\cdot\mathbb{I}\left(\mathcal{E}\right)+\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})|\boldsymbol{A}]}{m}\cdot\mathbb{I}\left(\mathcal{E}^c\right).$$

Since $\mathbb{P}(\mathcal{E}^c) \to 0$ and $\mathbb{E}[\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})|\boldsymbol{A}]/m < C(\mathcal{A}) < \infty$ is uniformly bounded, we immediately obtain $\mathbb{E}[\mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})|\boldsymbol{A}] \cdot \mathbb{I}\left(\mathcal{E}^c\right)/m \xrightarrow{\text{P}} 0$. Hence, we simply need to show:

$$\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})|\boldsymbol{A}]}{m} \cdot \mathbb{I}\left(\mathcal{E}\right) \xrightarrow{\text{P}} 0.$$

Step 2: Variance Normalization. Recall that $\boldsymbol{Z} = \text{Diag}\left(\boldsymbol{z}\right),\ \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}_\star \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{A}\boldsymbol{A}^\mathsf{T}/\kappa\right)$. We define the normalized random vector $\tilde{\boldsymbol{z}}$ as:

$$\tilde{z}_i = \frac{z_i}{\sigma_i},\ \sigma_i^2 = \frac{(\boldsymbol{A}\boldsymbol{A}^\mathsf{T})_{ii}}{\kappa}. \tag{5.18}$$

Note that conditional on $\boldsymbol{A}$, $\tilde{\boldsymbol{z}}$ is a zero mean Gaussian vector with:

$$\mathbb{E}[\tilde{z}_i{}^2|\boldsymbol{A}] = 1,\ \mathbb{E}[\tilde{z}_i\tilde{z}_j|\boldsymbol{A}] = \frac{(\boldsymbol{A}\boldsymbol{A}^\mathsf{T})_{ij}/\kappa}{\sigma_i\sigma_j}.$$

We define the diagonal matrix $\tilde{\boldsymbol{Z}} = \text{Diag}\left(\tilde{\boldsymbol{z}}\right)$. Using the continuity estimate from Lemma 22 we have,

$$\begin{aligned}
\left|\frac{\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})}{m} - \frac{\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \tilde{\boldsymbol{Z}})}{m}\right| &\leq \frac{C(\mathcal{A})}{\sqrt{m}}\|\boldsymbol{z} - \tilde{\boldsymbol{z}}\|_2 \\
&\leq C(\mathcal{A}) \cdot \left(\frac{1}{m}\sum_{i=1}^m z_i^2\right)^{\frac{1}{2}} \cdot \left(\max_{i\in[m]}\left|\frac{1}{\sigma_i} - 1\right|\right) \\
&\leq C(\mathcal{A}) \cdot \left(\frac{1}{m}\sum_{i=1}^m x_i^2\right)^{\frac{1}{2}} \cdot \left(\max_{i\in[m]}\left|\frac{1}{\sigma_i} - 1\right|\right).
\end{aligned}$$

We observe that $\|\boldsymbol{x}_\star\|^2/m \xrightarrow{\text{P}} \kappa^{-1}$, and on the event $\mathcal{E}$,

$$\max_{i\in[m]}\left|\frac{1}{\sigma_i} - 1\right| \to 0.$$

Hence,

$$\left| \frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})|\boldsymbol{A}]}{m} - \frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \tilde{\boldsymbol{Z}})|\boldsymbol{A}]}{m} \right| \cdot \mathbb{I}\left(\mathcal{E}\right) \xrightarrow{\text{P}} 0,$$

and hence, to conclude the proof of the lemma we simply need to show:

$$\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \tilde{\boldsymbol{Z}})|\boldsymbol{A}]}{m} \cdot \mathbb{I}\left(\mathcal{E}\right) \xrightarrow{\text{P}} 0.$$

Step 3: Mehler's Formula. Supposing that alternating product is of the Type 2 form (recall Definition 7):

$$\mathcal{A}(\boldsymbol{\Psi}, \tilde{\boldsymbol{Z}}) = (\boldsymbol{\Psi})q_1(\tilde{\boldsymbol{Z}})(\boldsymbol{\Psi})q_2(\tilde{\boldsymbol{Z}}) \cdots (\boldsymbol{\Psi})q_k(\tilde{\boldsymbol{Z}}).$$

The argument for the other types is very similar and we will sketch it in the end. We expand $\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \tilde{\boldsymbol{Z}})$ as follows:

$$\frac{1}{m}\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \tilde{\boldsymbol{Z}}) = \frac{1}{m} \sum_{a_1, a_2, \ldots a_k = 1}^{m} (\boldsymbol{\Psi})_{a_1, a_2} q_1(\tilde{\boldsymbol{Z}})_{a_2, a_2} \cdots (\boldsymbol{\Psi})_{a_k, a_1} q_k(\tilde{\boldsymbol{Z}})_{a_1, a_1}.$$

Next, we observe that:

$$[m]^k = \bigsqcup_{\pi \in \mathcal{P}([k])} \mathcal{C}(\pi).$$

Hence we can decompose the above sum as:

$$\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \tilde{\boldsymbol{Z}}) \ |\boldsymbol{A}]}{m} = \sum_{\pi \in \mathcal{P}([k])} \frac{1}{m} \sum_{a \in \mathcal{C}(\pi)} (\boldsymbol{\Psi})_{a_1, a_2} \cdots (\boldsymbol{\Psi})_{a_k, a_1} \mathbb{E}[\ q_1(\tilde{z}_{a_2}) \cdots q_k(\tilde{z}_{a_{k+1}})|\boldsymbol{A}].$$

By the triangle inequality,

$$\left|\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi},\tilde{\boldsymbol{Z}})\,|\boldsymbol{A}]}{m}\right| \leq \sum_{\pi\in\mathcal{P}([k])}\frac{1}{m}\sum_{a\in\mathcal{C}(\pi)}|(\boldsymbol{\Psi})_{a_1,a_2}\cdots(\boldsymbol{\Psi})_{a_k,a_1}||\mathbb{E}[\,q_1(\tilde{z}_{a_2})\cdots q_k(\tilde{z}_{a_1})|\boldsymbol{A}]|. \tag{5.19}$$

We first bound $|\mathbb{E}[\,q_1(\tilde{z}_{a_2})q_2(\tilde{z}_{a_3})\cdots q_k(\tilde{z}_{a_1})|\boldsymbol{A}]|$. Observe that if we denote the blocks of $\pi=\{\mathcal{V}_1,\mathcal{V}_2\ldots\mathcal{V}_{|\pi|}\}$, we can write:

$$\left|\mathbb{E}[\,q_1(\tilde{z}_{a_2})q_2(\tilde{z}_{a_3})\cdots q_k(\tilde{z}_{a_1})|\boldsymbol{A}]\right| = \left|\mathbb{E}\left[\prod_{i=1}^{|\pi|}\prod_{j\in\mathcal{V}_i}q_{j-1}(\tilde{z}_{a_{\mathcal{V}_i}})\middle|\boldsymbol{A}\right]\right|.$$

In the above display, we have defined $q_0\stackrel{\text{def}}{=}q_k$. Define the functions $\overline{q}_1,\overline{q}_2\ldots\overline{q}_{|\pi|}$ as:

$$\overline{q}_i(\xi)=\prod_{j\in\mathcal{V}_i}q_{j-1}(\xi)-\nu_i,\ \nu_i=\mathbb{E}_{\xi\sim\mathcal{N}(0,1)}\left[\prod_{j\in\mathcal{V}_i}q_{j-1}(\xi)\right].$$

Hence, we obtain:

$$\begin{aligned}\left|\mathbb{E}[\,q_1(\tilde{z}_{a_2})q_2(\tilde{z}_{a_3})\cdots q_k(\tilde{z}_{a_1})|\boldsymbol{A}]\right| &= \left|\mathbb{E}\left[\prod_{i=1}^{|\pi|}(\overline{q}_i(z_{a_{\mathcal{V}_i}})+\nu_i)\middle|\boldsymbol{A}\right]\right| \\ &\leq \sum_{V\subset[|\pi|]}\left(\prod_{i\notin V}|\nu_i|\right)\cdot\left|\mathbb{E}\left[\prod_{i\in V}\overline{q}_i(\tilde{z}_{a_{\mathcal{V}_i}})\middle|\boldsymbol{A}\right]\right|. \end{aligned}\tag{5.20}$$

Let $\mathscr{S}(\pi)$ denote the singleton blocks of the partition π: $\mathscr{S}(\pi)=\{i\in[|\pi|]:|\mathcal{V}_i|=1\}$. Note that for any $i\in\mathscr{S}(\pi)$, $\nu_i=0$ since the functions q_i satisfy $\mathbb{E}q_i(\xi)=0$ when $\xi\sim\mathcal{N}(0,1)$ (Definition 7). Hence,

$$\left|\mathbb{E}[\,q_1(\tilde{z}_{a_2})q_2(\tilde{z}_{a_3})\cdots q_k(\tilde{z}_{a_1})|\boldsymbol{A}]\right| \leq \sum_{V\subset[|\pi|]:\mathscr{S}(\pi)\subset V}\left(\prod_{i\notin V}|\nu_i|\right)\cdot\left|\mathbb{E}\left[\prod_{i\in V}\overline{q}_i(\tilde{z}_{a_{\mathcal{V}_i}})\middle|\boldsymbol{A}\right]\right|.$$

Next, we apply Mehler's Formula (Proposition 15) to bound:

$$\left| \mathbb{E}\left[\prod_{i \in V} \overline{q}_i(\tilde{z}_{a_{\mathcal{V}_i}}) \middle| \boldsymbol{A} \right] \right| \mathbb{I}\left(\mathcal{E}\right).$$

We make the following observations:

1. Recall the distribution of $\tilde{\boldsymbol{z}}$ given in (5.18) and the definition of the event $\mathcal{E}$ in (5.17), we obtain:

$$\max_{i \neq j} |\mathbb{E}[\tilde{z}_i \tilde{z}_j | \boldsymbol{A}]| \leq \left(\max_{i \neq j} \frac{1}{\kappa \sigma_i \sigma_j} \sqrt{\frac{32 \cdot K_1^2 \cdot \log^{2K_2+1}(m)}{m}} \right).$$

   Note that for large enough m, event $\mathcal{E}$ guarantees $\min_i \sigma_i \geq 1/2$. Hence,

$$\max_{i \neq j} |\mathbb{E}[\tilde{z}_i \tilde{z}_j | \boldsymbol{A}]| \leq \left(\frac{4}{\kappa} \sqrt{\frac{32 \cdot K_1^2 \cdot \log^{2K_2+1}(m)}{m}} \right).$$

   For any $S \subset [m]$ with $|S| \leq k$, let $\mathbb{E}[\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}^{\mathsf{T}} | \boldsymbol{A}]_{S,S}$ be the principal submatrix of the covariance matrix $\mathbb{E}[\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}^{\mathsf{T}} | \boldsymbol{A}]$. By Gershgorin's Circle Theorem we have.

$$\lambda_{\min}\left(\mathbb{E}[\tilde{\boldsymbol{z}}\tilde{\boldsymbol{z}}^{\mathsf{T}} | \boldsymbol{A}]_{S,S} \right) \geq 1 - k \max_{i \neq j} |\mathbb{E}[\tilde{z}_i \tilde{z}_j | \boldsymbol{A}]| \geq \frac{1}{2} \text{ (for } m \text{ large enough).}$$

2. We note that $\overline{q}_i$ satisfy $\mathbb{E}\overline{q}_i(\xi) = 0$ and $\mathbb{E}\xi\overline{q}_i(\xi) = 0$ (since $\overline{q}_i$ are even functions) when $\xi \sim \mathcal{N}(0,1)$. Hence, the first non-zero term in Mehler's expansion corresponds to $\boldsymbol{w}$ such that:

$$\mathsf{d}_i(\boldsymbol{w}) \geq 2, \quad \forall\, i \in V,$$

   thus,

$$\|\boldsymbol{w}\| \geq |V|.$$

Hence, by Mehler's Formula (Proposition 15), we obtain:

$$\left|\mathbb{E}\left[\prod_{i\in V}\overline{q}_i(\tilde{z}_{a_{\mathcal{V}_i}})\middle|\boldsymbol{A}\right]\right|\mathbb{I}\left(\mathcal{E}\right) \leq C\cdot\left(\max_{i\neq j}\mathbb{E}[\tilde{z}_i\tilde{z}_j|\boldsymbol{A}]\right)^{|V|}$$
$$\leq C\cdot\left(\frac{4}{\kappa}\sqrt{\frac{32\cdot K_1^2\cdot\log^{2K_2+1}(m)}{m}}\right)^{|V|},$$

for some finite constant C depending only on k and the functions $q_{1:k}$. Substituting this bound in (5.20) we obtain:

$$\left|\mathbb{E}[\ q_1(\tilde{z}_{a_2})q_2(\tilde{z}_{a_3})\cdots q_k(\tilde{z}_{a_1})|\boldsymbol{A}]\right|\cdot\mathbb{I}\left(\mathcal{E}\right) \leq \sum_{V\subset[|\pi|]}\left(\prod_{i\notin V}|\nu_i|\right)\cdot\left|\mathbb{E}\left[\prod_{i\in V}\overline{q}_i(\tilde{z}_{a_{\mathcal{V}_i}})\middle|\boldsymbol{A}\right]\right|$$
$$\leq C\sum_{V\subset[|\pi|]}\left(\prod_{i\notin V}|\nu_i|\right)\cdot\left(\frac{4}{\kappa}\sqrt{\frac{32\cdot K_1^2\cdot\log^{2K_2+1}(m)}{m}}\right)^{|V|}$$
$$\leq C(\mathcal{A})\cdot\left(\frac{4}{\kappa}\sqrt{\frac{32\cdot K_1^2\cdot\log^{2K_2+1}(m)}{m}}\right)^{|\mathscr{S}(\pi)|}.$$

In the above display, $C(\mathcal{A})$ denotes a finite constant depending only on k and the functions appearing in the definition of $\mathcal{A}$. Substituting this in (5.19):

$$\left|\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi},\tilde{\boldsymbol{Z}})\ |\boldsymbol{A}]}{m}\right|\mathbb{I}\left(\mathcal{E}\right)$$
$$\leq\sum_{\pi\in\mathcal{P}([k])}\frac{C(\mathcal{A})}{m}\sum_{a\in\mathcal{C}(\pi)}|(\boldsymbol{\Psi})_{a_1,a_2}\cdots(\boldsymbol{\Psi})_{a_k,a_1}|\left(\frac{4}{\kappa}\sqrt{\frac{32\cdot K_1^2\cdot\log^{2K_2+1}(m)}{m}}\right)^{|\mathscr{S}(\pi)|}.$$

Again, recalling the definition of $\mathcal{E}$ in (5.17), we can upper bound $|(\mathbf{\Psi})_{a_1,a_2}\cdots(\mathbf{\Psi})_{a_k,a_1}|$:

$$\left|\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\mathbf{\Psi},\tilde{\boldsymbol{Z}})\,|\boldsymbol{A}]}{m}\right|\cdot\mathbb{I}\left(\mathcal{E}\right)\leq\sum_{\pi\in\mathcal{P}([k])}\frac{C(\mathcal{A})}{m}\sum_{a\in\mathcal{C}(\pi)}\cdot\left(\sqrt{\frac{\cdot K_1^2\cdot\log^{2K_2+1}(m)}{m}}\right)^{|\mathscr{S}(\pi)|+k}$$
$$=\frac{C(\mathcal{A})}{m}\sum_{\pi\in\mathcal{P}([k])}|\mathcal{C}(\pi)|\cdot\left(\sqrt{\frac{\cdot K_1^2\cdot\log^{2K_2+1}(m)}{m}}\right)^{|\mathscr{S}(\pi)|+k}. \tag{5.21}$$

Step 4: Conclusion. Observe that: $|\mathcal{C}(\pi)|\leq m^{|\pi|}$. Recall that π has $|\mathscr{S}(\pi)|$ singleton blocks. All remaining blocks of π have at least 2 elements. Hence, we can upper bound $|\pi|$ as follows:

$$|\pi|\leq\frac{k-|\mathscr{S}(\pi)|}{2}+|\mathscr{S}(\pi)|=\frac{k+|\mathscr{S}(\pi)|}{2}.$$

Substituting this in (5.21) along with the trivial bounds $|\mathscr{S}(\pi)|\leq k,\ |\mathcal{P}([k])\leq k^k$, we obtain:

$$\left|\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\mathbf{\Psi},\tilde{\boldsymbol{Z}})\,|\boldsymbol{A}]}{m}\right|\cdot\mathbb{I}\left(\mathcal{E}\right)\leq\frac{C(\mathcal{A})\cdot k^k\cdot(K_1^2\log^{2K_2+1}(m))^k}{m}\to 0,$$

as desired.

Step 5: Other Cases. Recall that we had assumed that the alternating product was of Type 2:

$$\mathcal{A}(\mathbf{\Psi},\tilde{\boldsymbol{Z}})=(\mathbf{\Psi})q_1(\tilde{\boldsymbol{Z}})(\mathbf{\Psi})q_2(\tilde{\boldsymbol{Z}})\cdots(\mathbf{\Psi})q_k(\tilde{\boldsymbol{Z}}).$$

The analysis for the other types is analogous, and we briefly sketch these cases:

Type 1: $\mathcal{A}(\mathbf{\Psi},\tilde{\boldsymbol{Z}})=(\mathbf{\Psi})q_1(\tilde{\boldsymbol{Z}})(\mathbf{\Psi})q_2(\tilde{\boldsymbol{Z}})\cdots(\mathbf{\Psi})q_k(\tilde{\boldsymbol{Z}})(\mathbf{\Psi})$**.** In this case, we can expand the

normalized trace as:

$$\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi},\tilde{\boldsymbol{Z}})\,|\boldsymbol{A}]}{m} = \frac{1}{m}\sum_{a_0,a_1,\ldots a_k=1}^{m}\mathbb{E}[(\boldsymbol{\Psi})_{a_0,a_1}q_1(\tilde{\boldsymbol{Z}})_{a_1,a_1}\cdots q_k(\tilde{\boldsymbol{Z}})_{a_k,a_k}(\boldsymbol{\Psi})_{a_k,a_0}|\boldsymbol{A}]$$
$$= \frac{1}{m}\sum_{a_0=1}^{m}\sum_{\pi\in\mathcal{P}([k])}\sum_{a\in\mathcal{C}(\pi)}(\boldsymbol{\Psi})_{a_0,a_1}(\boldsymbol{\Psi})_{a_1,a_2}\cdots(\boldsymbol{\Psi})_{a_k,a_0}\mathbb{E}[q_1(\tilde{z}_{a_1})\cdots q_k(\tilde{z}_{a_k})|\boldsymbol{A}].$$

As before, we can argue on the event $\mathcal{E}$, for any $a_{0:k}$:

$$|\mathbb{E}[q_1(\tilde{z}_{a_1})\cdots q_k(\tilde{z}_{a_k})|\boldsymbol{A}]| \le O\left(\left(\frac{\mathrm{polylog}(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|}{2}}\right),$$
$$|(\boldsymbol{\Psi})_{a_0,a_1}(\boldsymbol{\Psi})_{a_1,a_2}\cdots(\boldsymbol{\Psi})_{a_k,a_0}| \le O\left(\left(\frac{\mathrm{polylog}(m)}{m}\right)^{\frac{k+1}{2}}\right),$$
$$|\mathcal{C}(\pi)| \le m^{\frac{k+|\mathscr{S}(\pi)|}{2}},$$
$$|\mathcal{P}([k])| \le k^k.$$

This gives us:

$$\left|\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi},\tilde{\boldsymbol{Z}})\,|\boldsymbol{A}]}{m}\right|\mathbb{I}(\mathcal{E}) \le \frac{1}{m}\cdot\overbrace{m}^{\text{choices for } a_0}\cdot\overbrace{|\mathcal{P}([k])|}^{\text{choices for } \pi}\cdot\overbrace{|\mathcal{C}(k)|}^{\text{choices for } a_{1:k}}\cdot O\left(\frac{\mathrm{polylog}(m)}{m^{\frac{k+|\mathscr{S}(\pi)|+1}{2}}}\right)$$
$$= O\left(\frac{\mathrm{polylog}(m)}{\sqrt{m}}\right)\to 0.$$

Type 3: $\mathcal{A} = q_0(\boldsymbol{Z})(\boldsymbol{\Psi})q_1(\boldsymbol{Z})\cdots(\boldsymbol{\Psi})q_k(\boldsymbol{Z})$**.** This case can be reduced to Type 1 and Type 2. Define $\tilde{q}_k(\xi) = q_0(\xi)q_k(\xi) - \nu,\ \nu = \mathbb{E}_{\xi\sim\mathcal{N}(0,1)}\ q_0(\xi)q_k(\xi)$. Then:

$$\frac{\mathbb{E}[\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})|\boldsymbol{A}]}{m} = \frac{\mathbb{E}[\mathsf{Tr}(q_0(\boldsymbol{Z})(\boldsymbol{\Psi})q_1(\boldsymbol{Z})\cdots(\boldsymbol{\Psi})q_k(\boldsymbol{Z}))|\boldsymbol{A}]}{m}$$
$$= \frac{\mathbb{E}[\mathsf{Tr}((\boldsymbol{\Psi})q_1(\boldsymbol{Z})\cdots(\boldsymbol{\Psi})q_k(\boldsymbol{Z})q_0(\boldsymbol{Z}))|\boldsymbol{A}]}{m}$$
$$= \underbrace{\frac{\mathbb{E}[\mathsf{Tr}((\boldsymbol{\Psi})q_1(\boldsymbol{Z})\cdots(\boldsymbol{\Psi})\tilde{q}_k(\boldsymbol{Z}))|\boldsymbol{A}]}{m}}_{\text{Type 2}} + \nu\underbrace{\frac{\mathbb{E}[\mathsf{Tr}((\boldsymbol{\Psi})q_1(\boldsymbol{Z})\cdots(\boldsymbol{\Psi}))|\boldsymbol{A}]}{m}}_{\text{Type 1}}.$$

Type 4: $\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}) = q_1(\boldsymbol{Z})(\boldsymbol{\Psi})q_2(\boldsymbol{Z})(\boldsymbol{\Psi})\cdots q_k(\boldsymbol{Z})(\boldsymbol{\Psi})$**.** This case is exactly the same as Type 2, and exactly the same bounds hold.

This concludes the proof of Lemma 23. □

Proof of Lemma 24. We observe that since $\boldsymbol{\Psi} = \boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} - \kappa \boldsymbol{I}_m$, conditioning on $\boldsymbol{A}$ fixes $\boldsymbol{\Psi}$. Hence, the only source of randomness in $\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})$ is $\boldsymbol{Z} = \mathrm{Diag}\left(\boldsymbol{z}\right), \boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}_\star, \boldsymbol{x}_\star \sim \mathcal{N}\left(0, 1/\kappa\right)$. Define the map $f(\boldsymbol{x}_\star) \stackrel{\text{def}}{=} \mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi}, \mathrm{Diag}\left(\boldsymbol{A}\boldsymbol{x}_\star\right))/m$. By Lemma 22, we have:

$$|f(\boldsymbol{x}) - f(\boldsymbol{x}')| \leq \frac{C(\mathcal{A})}{\sqrt{m}} \cdot \|\boldsymbol{A}(\boldsymbol{x} - \boldsymbol{x}')\|_2 \leq \frac{C(\mathcal{A})\|\boldsymbol{A}\|_{\mathsf{op}}}{\sqrt{m}} \cdot \|\boldsymbol{x} - \boldsymbol{x}'\|_2 = \frac{C(\mathcal{A})}{\sqrt{m}} \cdot \|\boldsymbol{x} - \boldsymbol{x}'\|_2.$$

Hence, f is $C(\mathcal{A})/\sqrt{n}$-Lipchitz. The claim of Lemma follows from the Gaussian Poincare Inequality (see Fact 5). □

5.8 Proof of Proposition 14

In this section, we provide a proof of Proposition 14. The proof follows from the following three results.

Lemma 25 (Continuity Estimates)**.** *We have:*

$$\left|\frac{\boldsymbol{z}^{\mathsf{T}}\mathcal{A}(\boldsymbol{U}\overline{\boldsymbol{B}}\boldsymbol{U}^{\mathsf{T}}, \mathit{Diag}\,(\boldsymbol{z}))\boldsymbol{z}}{m} - \frac{\widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}(\boldsymbol{U}\overline{\boldsymbol{B}}\boldsymbol{U}^{\mathsf{T}}, \mathit{Diag}\,(\widetilde{\boldsymbol{z}}))\widetilde{\boldsymbol{z}}}{m}\right| \leq \frac{C(\mathcal{A})}{m} \cdot \left(\|\boldsymbol{z}\|_2^2 \cdot \|\boldsymbol{z} - \widetilde{\boldsymbol{z}}\|_\infty + \|\boldsymbol{z} - \widetilde{\boldsymbol{z}}\|_2 \cdot (\|\boldsymbol{z}\|_2 + \|\widetilde{\boldsymbol{z}}\|_2)\right),$$

where $C(\mathcal{A})$ depends only on k, the $\|\|_\infty$-norms, and Lipchitz constants of the functions appearing in $\mathcal{A}$.

We have relegated the proof of the above continuity estimate to Appendix C.4.1 in the supplementary materials.

Proposition 18 (Universality of the first moment of the quadratic form)**.** *For both the subsampled Haar sensing model and the subsampled Hadamard sensing model, we have:*

$$\lim_{m\to\infty} \frac{\mathbb{E}\boldsymbol{z}^{\mathsf{T}}\mathcal{A}\boldsymbol{z}}{m} = (1-\kappa)^k \cdot \left(\prod_i \hat{q}_i(2)\right) \cdot \left(\prod_i (p_i(1-\kappa) - p_i(-\kappa))\right),$$

where the index i in the product ranges over all the p_i, q_i functions appearing in $\mathcal{A}$. In the above display:

$$\hat{q}_i(2) = \mathbb{E}q_i(\xi)H_2(\xi),\ \xi \sim \mathcal{N}(0,1), \tag{5.22}$$

where $H_2(\xi) = \xi^2 - 1$ is the degree 2 Hermite polynomial.

Proposition 19 (Universality of the second moment of the quadratic form)**.** *For both the subsampled Haar sensing model and the subsampled Hadamard sensing model we have:*

$$\lim_{m\to\infty} \frac{\mathbb{E}(\boldsymbol{z}^{\mathsf{T}}\mathcal{A}\boldsymbol{z})^2}{m^2} = (1-\kappa)^{2k} \cdot \left(\prod_i \hat{q}_i^2(2)\right) \cdot \left(\prod_i (p_i(1-\kappa) - p_i(-\kappa))^2\right).$$

In the above expression, $\hat{q}_i(2)$ are as defined in (5.22)*.*

We now provide a proof of Proposition 14 using the above results.

Proof of Proposition 14. Note that Propositions 18, 19 together imply that,

$$\mathrm{Var}\left(\frac{\boldsymbol{z}^{\mathsf{T}}\mathcal{A}\boldsymbol{z}}{m}\right) \to 0,$$

for both the sensing models. Hence, by Chebychev's inequality and Proposition 18, we have, for both the sensing models,

$$\text{p-lim}\, \frac{\boldsymbol{z}^{\mathsf{T}}\mathcal{A}\boldsymbol{z}}{m} = (1-\kappa)^k \cdot \left(\prod_i \hat{q}_i(2)\right) \cdot \left(\prod_i (p_i(1-\kappa) - p_i(-\kappa))\right).$$

This proves the claim of Proposition 14. □

The remainder of the section is dedicated to the proof of Proposition 18. The proof of Proposition 19 is very similar and can be found in Appendix C.2 in the supplementary materials.

5.8.1 Proof of Proposition 18

We provide a proof of Proposition 18 assuming that alternating form is of Type 1.

$$\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}) = p_1(\boldsymbol{\Psi})q_1(\boldsymbol{Z})p_2(\boldsymbol{\Psi}) \cdots q_{k-1}(\boldsymbol{Z})p_k(\boldsymbol{\Psi}).$$

We will outline how to handle the other types at the end of the proof (see Remark 15). Furthermore, in light of Lemma 16 we can further assume that all polynomials $p_i(\psi) = \psi$. Hence, we assume that $\mathcal{A}$ is of the form:

$$\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}) = \boldsymbol{\Psi}q_1(\boldsymbol{Z})\boldsymbol{\Psi} \cdots q_{k-1}(\boldsymbol{Z})\boldsymbol{\Psi}.$$

The proof of Proposition 18 consists of various steps which will be organized as separate lemmas. We begin by recall that

$$\boldsymbol{z} \sim \mathcal{N}\left(0, \frac{\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}}{\kappa}\right).$$

Define the event:

$$\mathcal{E} = \left\{ \max_{i \neq j} |(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}|)_{ij} \leq \sqrt{\frac{2048 \cdot \log^3(m)}{m}},\ \max_{i \in [m]} |(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})_{ii} - \kappa| \leq \sqrt{\frac{2048 \cdot \log^3(m)}{m}} \right\}. \tag{5.23}$$

By Lemma 21, we know that $\mathbb{P}(\mathcal{E}^c) \to 0$ for both the subsampled Haar sensing and the subsampled Hadamard model. We define the normalized random vector $\widetilde{\boldsymbol{z}}$ as:

$$\widetilde{z}_i = \frac{z_i}{\sigma_i},\ \sigma_i^2 = \frac{(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})_{ii}}{\kappa}.$$

Note that conditional on $\boldsymbol{A}$, $\widetilde{\boldsymbol{z}}$ is a zero mean Gaussian vector with:

$$\mathbb{E}[\widetilde{z}_i^2|\boldsymbol{A}] = 1,\ \mathbb{E}[\widetilde{z}_i\widetilde{z}_j|\boldsymbol{A}] = \frac{(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})_{ij}/\kappa}{\sigma_i\sigma_j}.$$

We define the diagonal matrix $\widetilde{\boldsymbol{Z}} = \mathrm{Diag}\,(\widetilde{\boldsymbol{z}})$.

Lemma 26. *We have,*

$$\lim_{m\to\infty} \frac{\mathbb{E}\boldsymbol{z}^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\boldsymbol{z}}{m} = \lim_{m\to\infty} \frac{\mathbb{E}\widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}})\widetilde{\boldsymbol{z}}}{m}\mathbb{I}\left(\mathcal{E}\right),$$

provided the latter limit exists.

The proof of the lemma uses the fact that $\mathbb{P}(\mathcal{E}^c) \to 0$, and that on the event $\mathcal{E}$ since $\sigma_i^2 \approx 1$, we have $\boldsymbol{z} \approx \widetilde{\boldsymbol{z}}$ and hence, the continuity estimates of Lemma 25 give the claim of this result. Complete details have been provided in Appendix C.4.2 in the supplementary materials.

The advantage of Lemma 26 is that $\widetilde{z}_i \sim \mathcal{N}\left(0,1\right)$, and on the event $\mathcal{E}$ the coordinates of $\widetilde{\boldsymbol{z}}$ have weak correlations. Consequently, Mehler's Formula (Proposition 15) can be used to analyze the leading order term in $\mathbb{E}[\widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}})\widetilde{\boldsymbol{z}}\,\mathbb{I}\left(\mathcal{E}\right)]$. Before we do so, we do one additional preprocessing step.

Lemma 27. *We have:*

$$\lim_{m\to\infty} \frac{\mathbb{E}\widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}})\widetilde{\boldsymbol{z}}}{m}\mathbb{I}\left(\mathcal{E}\right) = \lim_{m\to\infty} \frac{\mathbb{E}\langle\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}}),\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2\rangle\mathbb{I}\left(\mathcal{E}\right)}{m},$$

provided the latter limit exists.

Proof Sketch. Observe that we can write:

$$\begin{aligned}\widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}\widetilde{\boldsymbol{z}} &= \langle\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}}),\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}\rangle \\ &\overset{\text{(a)}}{=} \langle\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}}),\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2\rangle + \mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}})\cdot q(\widetilde{\boldsymbol{Z}})) + \mathsf{Tr}(\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}})).\end{aligned}$$

In the step marked (a), we defined $q(\xi) = \xi^2 - 1$ which is an even function. Note that we know $|\mathsf{Tr}(\mathcal{A})|/m \leq \|\mathcal{A}\|_{\mathsf{op}} \leq C(\mathcal{A}) < \infty$. Furthermore, by Proposition 13, we know $\mathsf{Tr}(\mathcal{A})/m \xrightarrow{\text{P}} 0$, and hence by Dominated Convergence Theorem $\mathbb{E}\mathsf{Tr}(\mathcal{A})\mathbb{I}\left(\mathcal{E}\right)/m \to 0$. Additionally, note that $\mathsf{Tr}(\mathcal{A}q(\widetilde{\boldsymbol{Z}}))$ is also an alternating form except for minor issue that $q(\xi)$ is not uniformly bounded and Lipchitz. However, the combinatorial calculations in Proposition 13 can be repeated to show that $\mathbb{E}\mathsf{Tr}(\mathcal{A} \cdot q(\widetilde{\boldsymbol{Z}}))/m \to 0$. Since we will see a more complicated version of these arguments in the remainder of the proof, we omit the details of this step. □

Note that, so far, Lemmas 26 and 27 show that:

$$\lim_{m\to\infty} \frac{\mathbb{E}\boldsymbol{z}^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})\boldsymbol{z}}{m} = \lim_{m\to\infty} \frac{\mathbb{E}\langle \mathcal{A}(\boldsymbol{\Psi}, \widetilde{\boldsymbol{Z}}), \widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2\rangle \mathbb{I}\left(\mathcal{E}\right)}{m},$$

provided the latter limit exists. We now focus on analyzing the RHS. We expand

$$\frac{\langle \mathcal{A}(\boldsymbol{\Psi}, \widetilde{\boldsymbol{Z}}), \widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2\rangle}{m} = \frac{1}{m} \sum_{\substack{a_{1:k+1}\in[m] \\ a_1 \neq a_{k+1}}} \widetilde{z}_{a_1}(\boldsymbol{\Psi})_{a_1,a_2} q_1(\widetilde{z}_{a_2}) \cdots q_{k-1}(\widetilde{z}_{a_k})(\boldsymbol{\Psi})_{a_k,a_{k+1}} \widetilde{z}_{a_{k+1}}.$$

Recall the notation for partitions introduced in Section 5.6.1. Observe that:

$$\{(a_1 \dots a_{k+1}) \in [m]^{k+1} : \ a_1 \neq a_{k+1}\} = \bigsqcup_{\substack{\pi\in\mathcal{P}([k+1]) \\ \pi(1)\neq\pi(k+1)}} \mathcal{C}(\pi).$$

Hence,

$$\frac{\mathbb{E}\langle \mathcal{A}(\boldsymbol{\Psi}, \widetilde{\boldsymbol{Z}}), \widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2\rangle \cdot \mathbb{I}\left(\mathcal{E}\right)}{m} =$$
$$\frac{1}{m} \sum_{\substack{\pi\in\mathcal{P}([1:k+1]) \\ \pi(1)\neq\pi(k+1)}} \sum_{a\in\mathcal{C}(\pi)} \mathbb{E}\, \widetilde{z}_{a_1}(\boldsymbol{\Psi})_{a_1,a_2} q_1(\widetilde{z}_{a_2})(\boldsymbol{\Psi})_{a_2,a_3} \cdots q_{k-1}(\widetilde{z}_{a_k})(\boldsymbol{\Psi})_{a_k,a_{k+1}} \widetilde{z}_{a_{k+1}} \cdot \mathbb{I}\left(\mathcal{E}\right).$$

Fix a $\pi \in \mathcal{P}([k+1])$ such that $\pi(1) \neq \pi(k+1)$, and consider a labelling $\boldsymbol{a} \in \mathcal{C}(\pi)$. By the tower

property,

$$\mathbb{E}\widetilde{z}_{a_1}(\boldsymbol{\Psi})_{a_1,a_2}q_1(\widetilde{z}_{a_2})(\boldsymbol{\Psi})_{a_2,a_3}\cdots q_{k-1}(\widetilde{z}_{a_k})(\boldsymbol{\Psi})_{a_k,a_{k+1}}\widetilde{z}_{a_{k+1}}\mathbb{I}\left(\mathcal{E}\right) =$$
$$\mathbb{E}\left[(\boldsymbol{\Psi})_{a_1,a_2}(\boldsymbol{\Psi})_{a_2,a_3}\cdots(\boldsymbol{\Psi})_{a_k,a_{k+1}}\cdot\mathbb{E}[\widetilde{z}_{a_1}q_1(\widetilde{z}_{a_2})q_2(\widetilde{z}_{a_3})\cdots q_{k-1}(\widetilde{z}_{a_k})\widetilde{z}_{a_{k+1}}|\boldsymbol{A}]\mathbb{I}\left(\mathcal{E}\right)\right].$$

We will now use Mehler's formula (Proposition 15) to evaluate the conditional expectation upto leading order. Note that some of the random variables $\widetilde{z}_{a_{1:k+1}}$ are equal (as given by the partition π). Hence, we group them together and recenter the resulting functions. The blocks corresponding to a_1, a_{k+1} need to be treated specially due to the presence of $\widetilde{z}_{a_1}, \widetilde{z}_{a_{k+1}}$ in the above expectations. Hence, we introduce the following notations:

$$\mathscr{F}(\pi) = \pi(1),\ \mathscr{L}(\pi) = \pi(k+1),\ \mathscr{S}(\pi) = \{i \in [2:k] : |\pi(i)| = 1\}.$$

We label all the remaining blocks of π as $\mathcal{V}_1, \mathcal{V}_2 \ldots \mathcal{V}_{|\pi|-|\mathscr{S}(\pi)|-2}$. Hence, the partition π is given by:

$$\pi = \mathscr{F}(\pi) \sqcup \mathscr{L}(\pi) \sqcup \left(\bigsqcup_{i\in\mathscr{S}(\pi)}\{i\}\right) \sqcup \left(\bigsqcup_{t=1}^{|\pi|-|\mathscr{S}(\pi)|-2}\mathcal{V}_i\right).$$

Note that:

$$\widetilde{z}_{a_1}\widetilde{z}_{a_{k+1}}\prod_{i=2}^{k}q_{i-1}(\widetilde{z}_{a_i}) = Q_{\mathscr{F}}(\widetilde{z}_{a_1})Q_{\mathscr{L}}(\widetilde{z}_{a_{k+1}})\left(\prod_{i\in\mathscr{S}(\pi)}q_{i-1}(\widetilde{z}_{a_i})\right)\cdot\prod_{i=1}^{|\pi|-|\mathscr{S}(\pi)|-2}(Q_{\mathcal{V}_i}(z_{a_{\mathcal{V}_i}}) + \mu_{\mathcal{V}_i}),$$

where:

$$Q_{\mathscr{F}}(\xi) = \xi\cdot\prod_{i\in\mathscr{F}(\pi),i\neq 1}q_{i-1}(\xi), \tag{5.24}$$

(5.25)

$$Q_{\mathscr{L}}(\xi) = \xi \cdot \prod_{i \in \mathscr{L}(\pi), i \neq k+1} q_{i-1}(\xi), \tag{5.26}$$

$$\mu_{\mathcal{V}_i} = \mathbb{E}_{\xi \sim \mathcal{N}(0,1)} \left[\prod_{j \in \mathcal{V}_i} q_{j-1}(\xi) \right], \tag{5.27}$$

$$Q_{\mathcal{V}_i}(\xi) = \prod_{j \in \mathcal{V}_i} q_{j-1}(\xi) - \mu_{\mathcal{V}_i}. \tag{5.28}$$

With this notation in place, we can apply Mehler's formula. The result is summarized in the following lemma.

Lemma 28. *For any $\pi \in \mathcal{P}([k+1])$ such that $\pi(1) \neq \pi(k+1)$, and any labelling $\boldsymbol{a} \in \mathcal{C}(\pi)$ we have:*

$$\mathbb{I}(\mathcal{E}) \cdot \left| \mathbb{E}[\widetilde{z}_{a_1} q_1(\widetilde{z}_{a_2}) q_2(\widetilde{z}_{a_3}) \cdots q_{k-1}(\widetilde{z}_{a_k}) \widetilde{z}_{a_{k+1}} | \boldsymbol{A}] - \sum_{\boldsymbol{w} \in \mathcal{G}_1(\pi)} g(\boldsymbol{w}, \pi) \cdot \mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) \right| \leq C(\mathcal{A}) \cdot \left(\frac{\log^3(m)}{m\kappa^2} \right)^{\frac{2+|\mathscr{S}(\pi)|}{2}}, \tag{5.29a}$$

where $\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})$ is the matrix moment as defined in Definition 8. The coefficients $g(\boldsymbol{w}, \pi)$ are given by:

$$g(\boldsymbol{w}, \pi) = \frac{1}{\kappa^{\|\boldsymbol{w}\|} \boldsymbol{w}!} \cdot \left(\hat{Q}_{\mathscr{F}}(1) \hat{Q}_{\mathscr{L}}(1) \prod_{i \in \mathscr{S}(\pi)} \hat{q}_{i-1}(2) \right) \cdot \left(\prod_{i \in [|\pi| - |\mathscr{S}(\pi)| - 2]} \mu_{\mathcal{V}_i} \right), \tag{5.29b}$$

and, the set $\mathcal{G}_1(\pi)$ is defined as:

$$\mathcal{G}_1(\pi) \stackrel{def}{=} \big\{ \boldsymbol{w} \in \mathcal{G}(k+1) : \mathsf{d}_1(\boldsymbol{w}) = 1,\ \mathsf{d}_{k+1}(\boldsymbol{w}) = 1,\ \mathsf{d}_i(\boldsymbol{w}) = 2\ \forall\ i\ \in\ \mathscr{S}(\pi), \mathsf{d}_i(\boldsymbol{w}) = 0\ \forall\ i\ \notin\ \{1, k+1\} \cup \mathscr{S}(\pi) \big\}. \tag{5.29c}$$

The proof of the lemma is obtained by instantiating Mehler's formula for this situation and

identifying the leading order term. Additional details for this step are provided in Appendix C.4.3 in the supplementary materials.

With this, we return to our analysis of:

$$\frac{\mathbb{E}\langle \mathcal{A}(\boldsymbol{\Psi}, \widetilde{\boldsymbol{Z}}), \widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2 \rangle \cdot \mathbb{I}(\mathcal{E})}{m} = \frac{1}{m} \sum_{\substack{\pi \in \mathcal{P}([1:k+1]) \\ \pi(1) \neq \pi(k+1)}} \sum_{a \in \mathcal{C}(\pi)} \mathbb{E}\, \widetilde{z}_{a_1} (\boldsymbol{\Psi})_{a_1,a_2} q_1(\widetilde{z}_{a_2}) (\boldsymbol{\Psi})_{a_2,a_3} \cdots q_{k-1}(\widetilde{z}_{a_k}) (\boldsymbol{\Psi})_{a_k,a_{k+1}} \widetilde{z}_{a_{k+1}} \cdot \mathbb{I}(\mathcal{E}) .$$

We define the following subsets of $\mathcal{P}(k+1)$ as:

$$\mathcal{P}_1([k+1]) \stackrel{\text{def}}{=} \{\pi \in \mathcal{P}(k+1) : \ \pi(1) \neq \pi(k+1), \ |\pi(1)| = 1, |\pi(k+1)| = 1, |\pi(j)| \leq 2 \ \forall \ j \ \in \ [k+1]\}, \tag{5.30a}$$

$$\mathcal{P}_2([k+1]) \stackrel{\text{def}}{=} \{\pi \in \mathcal{P}(k+1) : \ \pi(1) \neq \pi(k+1)\} \backslash \mathcal{P}_1([k+1]), \tag{5.30b}$$

and the error term which was controlled in Lemma 28:

$$\epsilon(\boldsymbol{\Psi}, \boldsymbol{a}) \stackrel{\text{def}}{=} \mathbb{I}(\mathcal{E}) \cdot \left(\mathbb{E}[\widetilde{z}_{a_1} q_1(\widetilde{z}_{a_2}) \cdots q_{k-1}(\widetilde{z}_{a_k}) \widetilde{z}_{a_{k+1}} | \boldsymbol{A}] - \sum_{\boldsymbol{w} \in \mathcal{G}_1(\pi)} g(\boldsymbol{w}, \pi) \cdot \mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) \right).$$

With these definitions we consider the decomposition:

$$\frac{\mathbb{E}\langle \mathcal{A}(\boldsymbol{\Psi}, \widetilde{\boldsymbol{Z}}), \widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2 \rangle \cdot \mathbb{I}(\mathcal{E})}{m} = \frac{1}{m} \sum_{\pi \in \mathcal{P}_1([k+1])} \sum_{a \in \mathcal{C}(\pi)} \sum_{\boldsymbol{w} \in \mathcal{G}_1(\pi)} g(\boldsymbol{w}, \pi) \mathbb{E}\left[(\boldsymbol{\Psi})_{a_1,a_2} \cdots (\boldsymbol{\Psi})_{a_k,a_{k+1}} \mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) \right] - \mathsf{I} + \mathsf{II} + \mathsf{III},$$

where:

$$\mathsf{I} \stackrel{\text{def}}{=} \frac{1}{m} \sum_{\substack{\pi \in \mathcal{P}([k+1]) \\ \pi(1) \neq \pi(k+1)}} \sum_{a \in \mathcal{C}(\pi)} \sum_{\boldsymbol{w} \in \mathcal{G}_1(\pi)} g(\boldsymbol{w}, \pi) \mathbb{E}\left[(\boldsymbol{\Psi})_{a_1, a_2} \cdots (\boldsymbol{\Psi})_{a_k, a_{k+1}} \mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) \mathbb{I}\left(\mathcal{E}^c\right) \right],$$

$$\mathsf{II} \stackrel{\text{def}}{=} \frac{1}{m} \sum_{\substack{\pi \in \mathcal{P}([k+1]) \\ \pi(1) \neq \pi(k+1)}} \sum_{a \in \mathcal{C}(\pi)} \mathbb{E}\left[(\boldsymbol{\Psi})_{a_1, a_2} \cdots (\boldsymbol{\Psi})_{a_k, a_{k+1}} \epsilon(\boldsymbol{\Psi}, \boldsymbol{a}) \mathbb{I}\left(\mathcal{E}\right) \right],$$

$$\mathsf{III} \stackrel{\text{def}}{=} \frac{1}{m} \sum_{\pi \in \mathcal{P}_2([k+1])} \sum_{a \in \mathcal{C}(\pi)} \sum_{\boldsymbol{w} \in \mathcal{G}_1(\pi)} g(\boldsymbol{w}, \pi) \mathbb{E}\left[(\boldsymbol{\Psi})_{a_1, a_2} \cdots (\boldsymbol{\Psi})_{a_k, a_{k+1}} \mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) \right].$$

Define $\boldsymbol{\ell}_{k+1} \in \mathcal{G}(k+1)$ to be the weight matrix of a simple line graph, i.e.

$$(\boldsymbol{\ell}_{k+1})_{ij} = \begin{cases} 1: & |j-i| = 1 \\ 0: & \text{otherwise} \end{cases}.$$

This decomposition can be written compactly as:

$$\mathsf{I} = \frac{1}{m} \sum_{\substack{\pi \in \mathcal{P}([1:k+1]) \\ \pi(1) \neq \pi(k+1)}} \sum_{a \in \mathcal{C}(\pi)} \sum_{\boldsymbol{w} \in \mathcal{G}_1(\pi)} g(\boldsymbol{w}, \pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w} + \boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a}) \mathbb{I}\left(\mathcal{E}^c\right) \right],$$

$$\mathsf{II} = \frac{1}{m} \sum_{\substack{\pi \in \mathcal{P}([1:k+1]) \\ \pi(1) \neq \pi(k+1)}} \sum_{a \in \mathcal{C}(\pi)} \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a}) \epsilon(\boldsymbol{\Psi}, \boldsymbol{a}) \mathbb{I}\left(\mathcal{E}\right) \right],$$

$$\mathsf{III} = \frac{1}{m} \sum_{\pi \in \mathcal{P}_2([1:k+1])} \sum_{a \in \mathcal{C}(\pi)} \sum_{\boldsymbol{w} \in \mathcal{G}_1(\pi)} g(\boldsymbol{w}, \pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w} + \boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a}) \right].$$

We will show that $\mathsf{I}, \mathsf{II}, \mathsf{III} \to 0$. Showing this involves the following components:

1. Bounds on matrix moments $\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w} + \boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a}) \right]$, which have been developed in Lemma 18.

2. Controlling the size of the set $|\mathcal{C}(\pi)|$ (since we sum over $\boldsymbol{a} \in \mathcal{C}(\pi)$ in the above terms).

Since,

$$|\mathcal{C}(\pi)| = m(m-1)\cdots(m-|\pi|+1) \asymp m^{|\pi|},$$

we need to develop bounds on $|\pi|$. This is done in the following lemma. In contrast, the sums over $\pi \in \mathcal{P}([k+1])$ and $\boldsymbol{w} \in \mathcal{G}_1(\pi)$ are not a cause of concern since $|\mathcal{P}([k+1])|, |\mathcal{G}_1(\pi)|$ depend only on k (which is held fixed), and not on m.

Lemma 29. *For any $\pi \in \mathcal{P}_1([k+1])$, we have:*

$$|\pi| = \frac{k+3+|\mathscr{S}(\pi)|}{2} \implies |\mathcal{C}(\pi)| \leq m^{\frac{k+3+|\mathscr{S}(\pi)|}{2}}.$$

For any $\pi \in \mathcal{P}_2([k+1])$, we have:

$$|\pi| \leq \frac{k+2+|\mathscr{S}(\pi)|}{2} \implies |\mathcal{C}(\pi)| \leq m^{\frac{k+2+|\mathscr{S}(\pi)|}{2}}.$$

Proof. Consider any $\pi \in \mathcal{P}([k+1])$ such that $\pi(1) \neq \pi(k+1)$. Recall that the disjoint blocks of $|\pi|$ were given by:

$$\pi = \mathscr{F}(\pi) \sqcup \mathscr{L}(\pi) \sqcup \left(\bigsqcup_{i \in \mathscr{S}(\pi)} \{i\} \right) \sqcup \left(\bigsqcup_{t=1}^{|\pi|-|\mathscr{S}(\pi)|-2} \mathcal{V}_i \right).$$

Hence,

$$k+1 = |\mathscr{F}(\pi)| + |\mathscr{L}(\pi)| + |\mathscr{S}(\pi)| + \sum_{t=1}^{|\pi|-|\mathscr{S}(\pi)|-2} |\mathcal{V}_i|.$$

Note that:

$$|\mathscr{F}(\pi)| \geq 1 \qquad \text{(Since } 1 \in \mathscr{F}(\pi)\text{)}, \tag{5.31a}$$

$$|\mathscr{L}(\pi)| \geq 1 \qquad \text{(Since } k+1 \in \mathscr{L}(\pi)\text{)}, \tag{5.31b}$$

$$|\mathcal{V}_i| \geq 2 \qquad \text{(Since } \mathcal{V}_i \text{ are not singletons)}. \tag{5.31c}$$

Hence,

$$k+1 \geq |\mathscr{F}(\pi)| + |\mathscr{L}(\pi)| + |\mathscr{S}(\pi)| + 2|\pi| - 2|\mathscr{S}(\pi)| - 4,$$

which implies:

$$\begin{aligned} |\pi| &\leq \frac{k+5+|\mathscr{S}(\pi)| - |\mathscr{F}(\pi)| - |\mathscr{L}(\pi)|}{2} \\ &\leq \frac{k+3+|\mathscr{S}(\pi)|}{2}, \end{aligned} \tag{5.32}$$

and hence,

$$|\mathcal{C}(\pi)| \leq m^{|\pi|} \leq m^{\frac{k+3+|\mathscr{S}(\pi)|}{2}}.$$

Finally, observe that:

1. For any $\pi \in \mathcal{P}_1([k+1])$ each of the inequalities in (5.31) are exactly tight by the definition of $\mathcal{P}_1([k+1])$ in (5.30), and hence:

$$|\pi| = \frac{k+3+|\mathscr{S}(\pi)|}{2}.$$

2. For any $\pi \in \mathcal{P}_2([k+1])$, one of the inequalities in (5.31) must be strict (see (5.30)). Hence,

when $\pi \in \mathcal{P}_2([k+1])$, we have the improved bound:

$$|\pi| \leq \frac{k+2+|\mathscr{S}(\pi)|}{2}.$$

This proves the claims of the lemma. □

We will now show that $\mathsf{I}, \mathsf{II}, \mathsf{III} \to 0$.

Lemma 30. *We have,*

$$\mathsf{I} \to 0,\ \mathsf{II} \to 0,\ \mathsf{III} \to 0 \ \text{ as } m \to \infty,$$

and hence:

$$\lim_{m\to\infty} \frac{\mathbb{E}\boldsymbol{z}^{\mathsf{T}}\mathcal{A}\boldsymbol{z}}{m} = \lim_{m\to\infty} \frac{1}{m} \sum_{\pi\in\mathcal{P}_1([k+1])} \sum_{a\in\mathcal{C}(\pi)} \sum_{\boldsymbol{w}\in\mathcal{G}_1(\pi)} g(\boldsymbol{w},\pi)\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right],$$

provided the latter limit exists.

Proof. First, note that for any $\boldsymbol{w} \in \mathcal{G}_1(\pi)$, we have:

$$\|\boldsymbol{w}\| = \frac{1}{2}\sum_{i=1}^{k+1} \mathsf{d}_i(\boldsymbol{w}) = \frac{1+1+2|\mathscr{S}(\pi)|}{2} = 1+|\mathscr{S}(\pi)| \ \text{ (See (5.29)).}$$

Furthermore, recalling that $\boldsymbol{\ell}_{k+1}$ is the weight matrix of a simple line graph, $\|\boldsymbol{\ell}_{k+1}\| = k$. Now, we apply Lemma 18 to obtain:

$$\begin{aligned}
|\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\mathbb{I}\left(\mathcal{E}^c\right)\right]| &\leq \sqrt{\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, 2\boldsymbol{w}+2\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right]}\sqrt{\mathbb{P}(\mathcal{E}^c)} \\
&\overset{(a)}{\leq} \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+1+k}{2}} \cdot \sqrt{\mathbb{P}(\mathcal{E}^c)} \\
&\leq \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+1+k}{2}} \cdot \frac{C_k}{m}.
\end{aligned}$$

Analogously we can obtain:

$$\mathbb{E}|\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})| \leq \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{k}{2}},$$
$$\mathbb{E}\left[|\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w} + \boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})|\right] \leq \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+1+k}{2}}$$

Further, recall that by Lemma 28 we have:

$$|\epsilon(\boldsymbol{\Psi}, \boldsymbol{a})| \leq C(\mathcal{A}) \cdot \left(\frac{\log^3(m)}{m\kappa^2}\right)^{\frac{2+|\mathscr{S}(\pi)|}{2}}.$$

Using these estimates, we obtain:

$$\begin{aligned}
|\mathsf{I}| &\leq \frac{C(\mathcal{A})}{m} \cdot \sum_{\substack{\pi:\mathcal{P}([k+1]) \\ \pi(0)\neq\pi(k+1)}} |\mathcal{C}(\pi)| \cdot \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+1+k}{2}} \cdot \frac{C_k}{m} \\
&\overset{\text{(a)}}{\leq} \frac{C(\mathcal{A})}{m} \cdot \sum_{\substack{\pi:\mathcal{P}([k+1]) \\ \pi(0)\neq\pi(k+1)}} m^{\frac{k+3+|\mathscr{S}(\pi)|}{2}} \cdot \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+1+k}{2}} \cdot \frac{C_k}{m} \\
&= O\left(\frac{\mathrm{polylog}(m)}{m}\right).
\end{aligned}$$

In addition:

$$\begin{aligned}
|\mathsf{II}| &\leq \frac{C(\mathcal{A})}{m} \cdot \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{k}{2}} \cdot \sum_{\substack{\pi:\mathcal{P}([k+1]) \\ \pi(0)\neq\pi(k+1)}} |\mathcal{C}(\pi)| \cdot \left(\frac{\log^3(m)}{m\kappa^2}\right)^{\frac{2+|\mathscr{S}(\pi)|}{2}} \\
&\overset{\text{(a)}}{\leq} \frac{C(\mathcal{A})}{m} \cdot \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{k}{2}} \cdot \sum_{\substack{\pi:\mathcal{P}([k+1]) \\ \pi(0)\neq\pi(k+1)}} m^{\frac{k+3+|\mathscr{S}(\pi)|}{2}} \cdot \left(\frac{\log^3(m)}{m\kappa^2}\right)^{\frac{2+|\mathscr{S}(\pi)|}{2}} \\
&= O\left(\frac{\mathrm{polylog}(m)}{\sqrt{m}}\right).
\end{aligned}$$

Furthermore:

$$
\begin{aligned}
|\mathsf{III}| &\leq \frac{C(\mathcal{A})}{m} \cdot \sum_{\pi:\mathcal{P}_2([k+1])} |\mathcal{C}(\pi)| \cdot \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+1+k}{2}} \\
&\overset{\text{(a)}}{\leq} \frac{C(\mathcal{A})}{m} \cdot \sum_{\pi:\mathcal{P}_2([k+1])} m^{\frac{k+2+|\mathcal{C}(\pi)|}{2}} \cdot \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+1+k}{2}} \\
&= O\left(\frac{\mathrm{polylog}(m)}{\sqrt{m}}\right).
\end{aligned}
$$

In each of the above displays, in the steps marked (a), we used the bounds on $|\mathcal{C}(\pi)|$ from Lemma 29. C_k denotes a constant depending only on k and $C(\mathcal{A})$ denotes a constant depending only on k and the functions appearing in $\mathcal{A}$. This concludes the proof of this lemma. □

So far we have shown that:

$$
\lim_{m\to\infty} \frac{\mathbb{E}\boldsymbol{z}^{\mathsf{T}}\mathcal{A}\boldsymbol{z}}{m} = \lim_{m\to\infty} \frac{1}{m} \sum_{\pi\in\mathcal{P}_1([k+1])} \sum_{a\in\mathcal{C}(\pi)} \sum_{\boldsymbol{w}\in\mathcal{G}_1(\pi)} g(\boldsymbol{w},\pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right],
$$

provided the latter limit exists. Our goal is to show that the limit on the LHS exists and is universal across the subsampled Haar and Hadamard models. In order to do so, we will leverage the fact that the first order term in the expansion of $\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right]$ is the same for the two models if $\boldsymbol{w}+\boldsymbol{\ell}_{k+1}$ is dissortive with respect to π and if $\boldsymbol{a}$ is a conflict-free labelling (Propositions 16 and 17). Hence, we need to argue that the contribution of terms corresponding to $\boldsymbol{w} : \boldsymbol{w}+\boldsymbol{\ell}_{k+1} \notin \mathcal{G}_{\mathsf{DA}}(\pi)$ and $\boldsymbol{a} \notin \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)$ are negligible. Towards this end, we consider the decomposition:

$$
\begin{aligned}
&\frac{1}{m} \sum_{\pi\in\mathcal{P}_1([k+1])} \sum_{a\in\mathcal{C}(\pi)} \sum_{\boldsymbol{w}\in\mathcal{G}_1(\pi)} g(\boldsymbol{w},\pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right] = \\
&\frac{1}{m} \sum_{\pi\in\mathcal{P}_1([k+1])} \sum_{\substack{\boldsymbol{w}\in\mathcal{G}_1(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}\in\mathcal{G}_{\mathsf{DA}}(\pi)}} \sum_{a\in\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi)} g(\boldsymbol{w},\pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right] + \mathsf{IV} + \mathsf{V},
\end{aligned}
$$

where:

$$\mathsf{IV} \stackrel{\text{def}}{=} \frac{1}{m} \sum_{\pi \in \mathcal{P}_1([k+1])} \sum_{a \in \mathcal{C}(\pi)} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_1(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1} \notin \mathcal{G}_{\mathsf{DA}}(\pi)}} g(\boldsymbol{w}, \pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right],$$

$$\mathsf{V} \stackrel{\text{def}}{=} \frac{1}{m} \sum_{\pi \in \mathcal{P}_1([k+1])} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_1(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1} \in \mathcal{G}_{\mathsf{DA}}(\pi)}} \sum_{a \in \mathcal{C}(\pi) \setminus \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)} g(\boldsymbol{w}, \pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right].$$

Lemma 31. *We have* $\mathsf{IV} \to 0, \mathsf{V} \to 0$*, as* $m \to \infty$*, and hence:*

$$\lim_{m \to \infty} \frac{\mathbb{E}\boldsymbol{z}^{\mathsf{T}} \mathcal{A} \boldsymbol{z}}{m} =$$
$$\lim_{m \to \infty} \frac{1}{m} \sum_{\pi \in \mathcal{P}_1([k+1])} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_1(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1} \in \mathcal{G}_{\mathsf{DA}}(\pi)}} \sum_{a \in \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)} g(\boldsymbol{w}, \pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right],$$

provided the latter limit exists.

Proof. We will prove this in two steps.

Step 1: $\mathsf{IV} \to 0$**.** We consider the two sensing models separately:

1. Subsampled Hadamard Sensing: In this case, Proposition 17 tells us that if $\boldsymbol{w} + \boldsymbol{\ell}_{k+1} \notin \mathcal{G}_{\mathsf{DA}}(\pi)$, then:

   $$\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right] = 0,$$

   and hence, $\mathsf{IV} = 0$.

2. Subsampled Haar Sensing: Observe that, since $\|\boldsymbol{w}\| + \|\boldsymbol{\ell}_{k+1}\| = 1 + |\mathscr{S}(\pi)| + k$, we have:

   $$\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right] = \frac{\mathbb{E}\left[\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right]}{m^{\frac{1+|\mathscr{S}(\pi)|+k}{2}}}.$$

By Proposition 16, we know that:

$$\left| \mathbb{E}\left[\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right] - \prod_{\substack{s,t\in[|\pi|]\\ s\leq t}} \mathbb{E}\left[Z_{st}^{W_{st}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi)}\right] \right| \leq \frac{K_1 \log^{K_2}(m)}{m^{\frac{1}{4}}},$$

where K_1, K_2, K_3 are universal constants depending only on k. Note that since $\boldsymbol{w} + \boldsymbol{\ell}_{k+1} \notin \mathcal{G}_{\mathsf{DA}}(\pi)$, we must have some $s \in [|\pi|]$ such that:

$$W_{ss}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi) \geq 1.$$

Recall that $\mathsf{d}_i(\boldsymbol{w}) = 0$ for any $i \notin \{1, k+1\} \cup \mathscr{S}(\pi)$ (since $\boldsymbol{w} \in \mathcal{G}_{\mathsf{1}}(\pi)$), and furthermore, $|\pi(i)| = 1 \ \forall \ i \ \in \ \{1, k+1\} \cup \mathscr{S}(\pi)$ (since $\pi \in \mathcal{P}_1(k+1)$). Hence, we have $\boldsymbol{w} \in \mathcal{G}_{\mathsf{DA}}(\pi)$ and in particular, $W_{ss}(\boldsymbol{w}, \pi) = 0$. Consequently, we must have $W_{ss}(\boldsymbol{\ell}_{k+1}, \pi) \geq 1$. Recall that $\boldsymbol{\ell}_{k+1}$ is the weight matrix of a line graph:

$$(\boldsymbol{\ell}_{k+1})_{ij} = \begin{cases} 1: & |i-j| = 1 \\ 0: & \text{otherwise} \end{cases}.$$

Consequently, since $W_{ss}(\boldsymbol{\ell}_{k+1}, \pi) \geq 1$, we must have for some $i \in [k]$, $\pi(i) = \pi(i+1) = \mathcal{V}_s$. However, since $\pi \in \mathcal{P}_1(k+1)$, $|\mathcal{V}_s| \leq 2$, and hence, $\mathcal{V}_s = \{i, i+1\}$. This means that $W_{ss}(\boldsymbol{\ell}_{k+1}, \pi) = 1 = W_{ss}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)$. Consequently, since $\mathbb{E}Z_{ss} = 0$, we have:

$$\prod_{\substack{s,t\in[|\pi|]\\ s\leq t}} \mathbb{E}\left[Z_{st}^{W_{st}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi)}\right] = 0,$$

or

$$|\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right]| = \frac{C_k \log^{K}(m)}{m^{\frac{1+|\mathscr{S}(\pi)|+k}{2}+\frac{1}{4}}},$$

where C_k, K are constants that depend only on k. Recalling Lemma 29,

$$|\mathcal{C}(\pi)| \leq m^{|\pi|} \leq m^{\frac{k+3+|\mathscr{S}(\pi)|}{2}},$$

we obtain:

$$|\mathsf{IV}| \leq \frac{C(\mathcal{A})}{m} \sum_{\pi \in \mathcal{P}_1([k+1])} |\mathcal{C}(\pi)| \cdot \frac{C_k \log^K(m)}{m^{\frac{1+|\mathscr{S}(\pi)|+k}{2}+\frac{1}{4}}} = O\left(\frac{\text{polylog}(m)}{m^{\frac{1}{4}}}\right) \to 0.$$

Step 2: $\mathsf{V} \to 0$**.** Using Lemma 20, we know that

$$|\mathcal{C}(\pi) \backslash \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w} + \boldsymbol{\ell}_{k+1}, \pi)| \leq (k+1)^4 m^{|\pi|-1}.$$

In Lemma 29, we showed that for any $\pi \in \mathcal{P}_1([k+1])$,

$$|\pi| = \frac{k+3+|\mathscr{S}(\pi)|}{2}.$$

Hence,

$$|\mathcal{C}(\pi) \backslash \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w} + \boldsymbol{\ell}_{k+1}, \pi)| \leq (k+1)^4 \cdot m^{\frac{k+1+|\mathscr{S}(\pi)|}{2}}.$$

We already know from Lemma 18 that:

$$|\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w} + \boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right]| \leq \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{\|\boldsymbol{w}\|+\|\boldsymbol{\ell}_{k+1}\|}{2}} \leq \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+1+k}{2}}.$$

This gives us:

$$|\mathsf{V}| \le \frac{C}{m} \sum_{\pi \in \mathcal{P}_1([k+1])} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_1(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1} \in \mathcal{G}_{\mathsf{DA}}(\pi)}} |\mathcal{C}(\pi) \backslash \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)| \cdot \left(\frac{C_k \log^2(m)}{m} \right)^{\frac{|\mathscr{S}(\pi)|+1+k}{2}}$$
$$= O\left(\frac{\mathrm{polylog}(m)}{m} \right)$$

which goes to zero as claimed.

□

To conclude, we have shown that:

$$\lim_{m\to\infty} \frac{\mathbb{E}\boldsymbol{z}^{\mathsf{T}} \mathcal{A} \boldsymbol{z}}{m} =$$
$$\lim_{m\to\infty} \frac{1}{m} \sum_{\pi \in \mathcal{P}_1([k+1])} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_1(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1} \in \mathcal{G}_{\mathsf{DA}}(\pi)}} \sum_{a \in \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)} g(\boldsymbol{w}, \pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a}) \right],$$

provided the limit on the RHS exists. In the following lemma we explicitly evaluate the limit on the RHS, and in particular, show it exists and is identical for the two sensing models.

Lemma 32. *For both the subsampled Haar sensing and Hadamard sensing model, we have:*

$$\lim_{m\to\infty} \frac{\mathbb{E}\boldsymbol{z}^{\mathsf{T}} \mathcal{A} \boldsymbol{z}}{m} = \sum_{\pi \in \mathcal{P}_1([k+1])} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_1(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1} \in \mathcal{G}_{\mathsf{DA}}(\pi)}} g(\boldsymbol{w}, \pi) \cdot \mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi),$$

where,

$$\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi) \stackrel{def}{=} \prod_{\substack{s,t \in [|\pi|] \\ s<t}} \mathbb{E}\left[Z^{W_{st}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)} \right],\ Z \sim \mathcal{N}\left(0, \kappa(1-\kappa)\right).$$

Proof. By Propositions 17 (for the subsampled Hadamard model) and 16 (for the subsampled Haar

model) we know that, if $\boldsymbol{w}+\boldsymbol{\ell}_{k+1} \in \mathcal{G}_{\mathsf{DA}}(\pi)$ and $\boldsymbol{a} \in \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)$, we have:

$$\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a}) = \mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi) + \epsilon(\boldsymbol{w}, \pi, \boldsymbol{a}),$$

where

$$|\epsilon(\boldsymbol{w}, \pi, \boldsymbol{a})| \leq \frac{K_1 \log^{K_2}(m)}{m^{\frac{1}{4}}}, \ \forall\, m \geq K_3,$$

for some constants K_1, K_2, K_3 depending only on k. Hence, we can consider the decomposition:

$$\frac{1}{m} \sum_{\pi \in \mathcal{P}_1([k+1])} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_1(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1} \in \mathcal{G}_{\mathsf{DA}}(\pi)}} \sum_{a \in \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)} g(\boldsymbol{w}, \pi) \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})\right] = \mathsf{VI} + \mathsf{VII},$$

where:

$$\mathsf{VI} \stackrel{\text{def}}{=} \frac{1}{m} \sum_{\pi \in \mathcal{P}_1([k+1])} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_1(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1} \in \mathcal{G}_{\mathsf{DA}}(\pi)}} \sum_{a \in \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)} g(\boldsymbol{w}, \pi) \cdot \frac{\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)}{m^{\frac{1+\mathscr{S}(\pi)+k}{2}}},$$

$$\mathsf{VII} \stackrel{\text{def}}{=} \frac{1}{m} \sum_{\pi \in \mathcal{P}_1([k+1])} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_1(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1} \in \mathcal{G}_{\mathsf{DA}}(\pi)}} \sum_{a \in \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)} g(\boldsymbol{w}, \pi) \cdot \frac{\epsilon(\boldsymbol{w}, \pi, \boldsymbol{a})}{m^{\frac{1+\mathscr{S}(\pi)+k}{2}}}.$$

We can upper bound $|\mathsf{VII}|$ as follows:

$$|\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)| \leq |\mathcal{C}(\pi)| \leq m^{\frac{k+3+|\mathscr{S}(\pi)|}{2}}.$$

Thus:

$$\begin{aligned} |\mathsf{VII}| &\leq \frac{C(\mathcal{A})}{m} \cdot C_k \cdot |\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}, \pi)| \cdot \frac{1}{m^{\frac{1+|\mathscr{S}(\pi)|+k}{2}}} \cdot \frac{K_1 \log^{K_2}(m)}{m^{\frac{1}{4}}} \\ &= O\left(\frac{\text{polylog}(m)}{m^{\frac{1}{4}}}\right) \to 0. \end{aligned}$$

Moreover, can compute:

$$\lim_{m\to\infty}(\mathsf{VI}) = \lim_{m\to\infty}\frac{1}{m}\sum_{\pi\in\mathcal{P}_1([k+1])}\sum_{\substack{\boldsymbol{w}\in\mathcal{G}_1(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}\in\mathcal{G}_{\mathsf{DA}}(\pi)}}\sum_{a\in\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi)} g(\boldsymbol{w},\pi)\cdot\frac{\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi)}{m^{\frac{1+\mathscr{S}(\pi)+k}{2}}}$$

$$= \lim_{m\to\infty}\frac{1}{m}\sum_{\pi\in\mathcal{P}_1([k+1])}\sum_{\substack{\boldsymbol{w}\in\mathcal{G}_1(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}\in\mathcal{G}_{\mathsf{DA}}(\pi)}} g(\boldsymbol{w},\pi)\cdot\frac{\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi)}{m^{\frac{1+|\mathscr{S}(\pi)|+k}{2}}}\cdot|\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi)|$$

$$\overset{\text{(a)}}{=} \lim_{m\to\infty}\sum_{\pi\in\mathcal{P}_1([k+1])}\sum_{\substack{\boldsymbol{w}\in\mathcal{G}_1(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}\in\mathcal{G}_{\mathsf{DA}}(\pi)}} g(\boldsymbol{w},\pi)\cdot\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi)\cdot\frac{|\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi)|}{m^{|\pi|}}$$

$$\overset{\text{(b)}}{=} \sum_{\pi\in\mathcal{P}_1([k+1])}\sum_{\substack{\boldsymbol{w}\in\mathcal{G}_1(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}\in\mathcal{G}_{\mathsf{DA}}(\pi)}} g(\boldsymbol{w},\pi)\cdot\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi).$$

In the step marked (a) we used the fact that $|\pi| = (3+|\mathscr{S}(\pi)|+k)/2$ for any $\pi\in\mathcal{P}_1([k+1])$ (Lemma 29), and in step (b) we used Lemma 20 ($|\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi)|/m^{|\pi|}\to 1$). This proves the claim of the lemma. □

In the following lemma, we show that the combinatorial sum obtained in Lemma 32 can be significantly simplified.

Lemma 33. *For both the subsampled Haar sensing and Hadamard sensing models, we have:*

$$\lim_{m\to\infty}\frac{\mathbb{E}\boldsymbol{z}^{\mathsf{T}}\mathcal{A}\boldsymbol{z}}{m} = (1-\kappa)^k\cdot\prod_{i=1}^{k-1}\hat{q}_i(2).$$

In particular, Proposition 18 holds.

Proof. We claim that the only partition with a non-zero contribution is:

$$\pi = \bigsqcup_{i=1}^{k+1}\{i\}.$$

In order to see this, suppose π is not entirely composed of singleton blocks. Define:

$$i_\star \overset{\text{def}}{=} \min\{i\in[k+1] : |\pi(i)|>1\}.$$

Note that $i_\star > 1$ since we know that $|\pi(1)| = |\mathscr{F}(\pi)| = 1$ for any $\pi \in \mathcal{P}_1(k+1)$. Since $\pi \in \mathcal{P}_1([k+1])$, we must have $|\pi(i_\star)| = 2$, hence, denote:

$$\pi(i_\star) = \{i_\star, j_\star\},$$

for some $j_\star > i_\star + 1$ ($i_\star \leq j_\star$ since it is the first index which is not in a singleton block, and $j_\star \neq i_\star + 1$ since otherwise $\boldsymbol{w} + \boldsymbol{\ell}_{k+1}$ will not be disassortative). Let us label the first few blocks of π as:

$$\mathcal{V}_1 = \{1\},\ \mathcal{V}_2 = \{2\}, \ldots, \mathcal{V}_{i_\star - 1} = \{i_\star - 1\},\ \mathcal{V}_{i_\star} = \{i_\star, j_\star\}.$$

Next, we compute:

$$\begin{aligned}
W_{i_\star-1,i_\star}(\boldsymbol{w} + \boldsymbol{\ell}_{k+1}, \pi) &= W_{i_\star-1,i_\star}(\boldsymbol{\ell}_{k+1}, \pi) + W_{i_\star-1,i_\star}(\boldsymbol{w}, \pi) \\
&\overset{\text{(a)}}{=} W_{i_\star-1,i_\star}(\boldsymbol{\ell}_{k+1}, \pi) \\
&\overset{\text{(b)}}{=} \mathbf{1}_{i_\star-1 \in \mathcal{V}_{i_\star-1}} + \mathbf{1}_{i_\star+1 \in \mathcal{V}_{i_\star-1}} + \mathbf{1}_{j_\star-1 \in \mathcal{V}_{i_\star-1}} + \mathbf{1}_{j_\star+1 \in \mathcal{V}_{i_\star-1}} \\
&\overset{\text{(c)}}{=} \mathbf{1}_{i_\star-1 = i_\star-1} + \mathbf{1}_{i_\star+1 = i_\star-1} + \mathbf{1}_{j_\star-1 = i_\star-1} + \mathbf{1}_{j_\star+1 = i_\star-1} \\
&\overset{\text{(d)}}{=} 1.
\end{aligned}$$

In the step marked (a), we used the fact that since $\boldsymbol{w} \in \mathcal{G}_1(\pi)$ and $|\pi(i_\star)| = |\pi(j_\star)| = 2$, we must have $d_{i_\star}(\boldsymbol{w}) = d_{j_\star}(\boldsymbol{w}) = 0$ and $W_{i_\star-1,i_\star}(\boldsymbol{w}, \pi) = 0$. In the step marked (b), we used the definition of $\boldsymbol{\ell}_{k+1}$ (that it is the line graph). In the step marked (c), we used the fact that $\mathcal{V}_{i_\star-1} = \{i_\star-1\}$. In the step marked (d), we used the fact that $j_\star > i_\star + 1$.

Hence, we have shown that for any $\pi \neq \sqcup_{i=1}^{k+1}\{i\}$, we have:

$$\mu(\boldsymbol{w}, \pi) = 0 \ \forall\ \boldsymbol{w} \text{ such that } \boldsymbol{w} \in \mathcal{G}_1(\pi),\ \boldsymbol{w} + \boldsymbol{\ell}_{k+1} \in \mathcal{G}_{\mathsf{DA}}(\pi).$$

Next, let $\pi = \sqcup_{i=1}^{k+1}\{i\}$. We observe for any $\boldsymbol{w}$ such that $\boldsymbol{w} \in \mathcal{G}_1(\pi),\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1} \in \mathcal{G}_{\mathsf{DA}}(\pi)$, we have:

$$\begin{aligned}\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi) &= \prod_{\substack{s,t\in[|\pi|]\\ s<t}} \mathbb{E}\left[Z^{W_{st}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi)}\right],\ Z\sim\mathcal{N}\left(0,\kappa(1-\kappa)\right)\\ &= \prod_{\substack{i,j\in[k+1]\\ i<j}} \mathbb{E}\left[Z^{w_{ij}+(\ell_{k+1})_{ij},\pi)}\right],\ Z\sim\mathcal{N}\left(0,\kappa(1-\kappa)\right).\end{aligned}$$

Note that since $\mathbb{E}Z = 0$, for $\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1},\pi) \neq 0$, we must have:

$$w_{ij} \geq (\ell_{k+1})_{ij},\ \forall\, i,j\ \in\ [k].$$

However, since $\boldsymbol{w} \in \mathcal{G}_1(\pi)$ we have:

$$\mathsf{d}_1(\boldsymbol{w}) = \mathsf{d}_{k+1}(\boldsymbol{w}) = 1,\ \mathsf{d}_i(\boldsymbol{w}) = 2\ \forall\ i\ \in\ [2:k],$$

so, $\boldsymbol{w} = \boldsymbol{\ell}_{k+1}$. Hence, recalling the formula for $g(\boldsymbol{w},\pi)$ from Lemma 28, we obtain:

$$\lim_{m\to\infty}\frac{\mathbb{E}\boldsymbol{z}^{\mathsf{T}}\mathcal{A}\boldsymbol{z}}{m} = (1-\kappa)^k\cdot\prod_{i=1}^{k-1}\hat{q}_i(2).$$

This proves the statement of the lemma and also Proposition 18 (see Remark 15 regarding how the analysis extends to other types). □

Throughout this section, we assumed that the alternating product $\mathcal{A}$ was of Type I. The following remark outlines how the analysis of this section extends to other types.

Remark 15. *The analysis of the other cases can be reduced to Type 1 as follows: Consider an alternating form $\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})$ of Type 1:*

$$\mathcal{A} = p_1(\boldsymbol{\Psi})q_1(\boldsymbol{Z})p_1(\boldsymbol{\Psi})\cdots q_{k-1}(\boldsymbol{Z})p_k(\boldsymbol{\Psi}),$$

but the more general quadratic form:

$$\frac{1}{m}\mathbb{E}\alpha(\boldsymbol{z})^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\beta(\boldsymbol{z}), \tag{5.33}$$

where $\alpha, \beta : \mathbb{R} \to \mathbb{R}$ are odd functions whose absolute values can be upper bounded by a polynomial. They act on the vector $\boldsymbol{z}$ entry-wise. This covers all the types in a unified way:

1. *For Type 1 case: We take $\alpha(z) = \beta(z) = z$.*

2. *For the Type 2 case, we write:*

$$\boldsymbol{z}^{\mathsf{T}}p_1(\boldsymbol{\Psi})q_1(\boldsymbol{Z})p_1(\boldsymbol{\Psi})\cdots q_k(\boldsymbol{Z})p_k(\boldsymbol{\Psi})q_k(\boldsymbol{Z})\boldsymbol{z} = \alpha(\boldsymbol{z})^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\beta(\boldsymbol{z}),$$

   where $\alpha(z) = z, \beta(z) = zq_k(z)$.

3. *For the Type 3 case:*

$$\boldsymbol{z}^{\mathsf{T}}q_0(\boldsymbol{Z})p_1(\boldsymbol{\Psi})q_1(\boldsymbol{Z})p_1(\boldsymbol{\Psi})\cdots q_{k-1}(\boldsymbol{Z})p_k(\boldsymbol{\Psi})q_k(\boldsymbol{Z})\boldsymbol{z} = \alpha(\boldsymbol{z})^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\beta(\boldsymbol{z}),$$

   where $\alpha(z) = zq_0(z), \beta(z) = zq_k(z)$.

4. *For the Type 4 case:*

$$\boldsymbol{z}^{\mathsf{T}}q_0(\boldsymbol{Z})p_1(\boldsymbol{\Psi})q_1(\boldsymbol{Z})p_2(\boldsymbol{\Psi})\cdots q_{k-1}(\boldsymbol{Z})p_k(\boldsymbol{\Psi})\boldsymbol{z} = \alpha(\boldsymbol{z})^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\beta(\boldsymbol{z}),$$

   where $\alpha(z) = zq_0(z), \beta(z) = z$.

The analysis of the more general quadratic form in (5.33) *is analogous to the analysis outlined in this section. Lemmas 26 and 27 extend straightforwardly. Inspecting the proof of Lemma 28 shows that the same error bound continues to hold (after suitably redefining $c(\boldsymbol{w}, \pi)$), since α, β are odd (as in the case $\alpha(z) = \beta(z) = z$). The subsequent lemmas after that hold verbatim for the more general quadratic form* (5.33)*.*

5.9 Conclusion

In this chapter, we analyzed the dynamics of linearized Approximate message passing algorithms for phase retrieval when the sensing matrix is generated by sub-sampling n columns of a $m \times m$ Hadamard-Walsh matrix under an average-case Gaussian prior assumption on the signal. We showed that the dynamics of linearized AMP algorithms for these sensing matrices are asymptotically indistinguishable from the dynamics in the case when the sensing matrix is generated by sampling n columns of a uniformly random $m \times m$ orthogonal matrix. This provides a theoretical justification for an empirically observed universality phenomena in a particular case.

Chapter 6: Conclusion and Future Directions

We end this dissertation by mentioning some interesting directions for future work.

6.1 Beyond Spectral Estimators for Phase Retrieval

In Chapter 3, we provided an analysis of the performance of spectral methods under the sub-sampled Haar ansatz for the sensing matrix. However, spectral estimators are not the state-of-the-art estimators for the Phase retrieval problem. It would be interesting to analyze the following estimators for the phase retrieval problem with sub-sampled Haar sensing matrices:

Analysis of Maximum Likelihood Estimator: The maximum likelihood estimator for the (noiseless) phase retrieval problem is any solution to the following feasibility problem:

$$\text{Find } \boldsymbol{u} \in \mathbb{C}^n : |\boldsymbol{A}\boldsymbol{u}|^2 = \boldsymbol{y}. \tag{6.1}$$

It would be interesting to understand at what value of δ does this feasibility problem have $\boldsymbol{u} = \boldsymbol{x}$ as the unique solution. Maillard, Loureiro, Krzakala, and Zdeborová [79] have analyzed the conjectured replica-symmetric prediction for the Bayes risk for this problem. Their analysis suggests that exact recovery is possible as soon as $\delta > 2$. This leads to the conjecture that the feasibility problem (6.1) has a unique solution when $\delta > 2$. It would be interesting to prove this conjecture. Combined with the results of Chapter 2 of this dissertation, such a result would show that this problem exhibits a "all-or-nothing" phase transition [74]: When $\delta < 2$ any estimator is asymptotically orthogonal to the signal and when $\delta > 2$, there is an estimator which recovers $\boldsymbol{x}$ exactly.

Analysis of Bayes Optimal Approximate Message Passing with Spectral Initialization: It is not clear if the maximum likelihood estimator in (6.1) can be computed efficiently. Maillard, Loureiro, Krzakala, and Zdeborová have also studied the performance of a computationally efficient Bayes-optimal Approximate Message Passing algorithm. Their analysis suggests that this algorithm achieves exact recovery of the unknown signal vector when $\delta > 2.265$. Unfortunately, since the state evolution of the Bayes-optimal AMP algorithm for this problem has an uninformative fixed point, it requires an arbitrarily small amount of side information to recover the signal. Consequently, it does not yield a valid estimator. It would be interesting to provide an analysis of the Bayes optimal AMP algorithm initialized with the spectral initialization similar to the work of Montanari and Venkataramanan [84] and Mondelli and Venkataramanan [85] for Gaussian sensing matrices.

6.2 Understanding Bayes risk above the Weak Recovery Threshold

In Chapter 4, we studied the Phase Retrieval problem with sub-sampled Haar sensing matrices with non-zero but vanishing measurement noise in the high dimensional asymptotic regime. We showed that when the sampling ratio $\delta = m/n < 2$, then it is information-theoretically impossible for any estimator to obtain an asymptotically non-trivial performance: any estimator is asymptotically uncorrelated with the signal vector. Since Chapter 3 exhibits an estimator which achieves a nontrivial correlation with the signal vector when $\delta > 2$, this shows that the weak recovery threshold for this model is $\delta_{\mathsf{weak}} = 2$.

Our proof techniques in this chapter do not offer any information about the behavior of Bayes risk above the weak recovery threshold, particularly in the presence of measurement noise. For Gaussian sensing matrices Barbier, Krzakala, Macris, Miolane, and Zdeborová [25] have developed interpolation-based methods to compute the exact expression of Bayes risk. Furthermore, this technique appears to be general enough to handle interesting models of measurement noise and prior information about the signal (e.g. sparsity). It would be interesting to see if this technique can be extended beyond i.i.d. sensing matrices to sub-sampled Haar sensing matrices. Re-

cent works by Barbier, Macris, Maillard, and Krzakala [41] and Maillard, Loureiro, Krzakala, and Zdeborová [79] take a step in this direction and study generalized linear models where the sensing matrix is a Gaussian matrix whose rows are sampled i.i.d. from a correlated multivariate Gaussian distribution.

6.3 Further exploration of Universality Phenomenon

In Chapter 5, we analyzed the dynamics of linearized Approximate message passing algorithms for phase retrieval when the sensing matrix is generated by sub-sampling n columns of a $m \times m$ Hadamard-Walsh matrix under an average-case Gaussian prior assumption on the signal. We showed that the dynamics of linearized AMP algorithms for these sensing matrices are asymptotically indistinguishable from the dynamics in the case when the sensing matrix is generated by sampling n columns of a uniformly random $m \times m$ orthogonal matrix. This provides a theoretical justification for an empirically observed universality phenomenon in a particular case. It would be interesting to extend our results in the following ways:

Other structured ensembles: While we focused on the sub-sampled Hadamard sensing model in Chapter 5, we believe our results should extend to other popular structured matrices with orthogonal columns such as randomly sub-sampled Fourier, Discrete Cosine Transform matrices, and CDP matrices. For these ensembles, there exist analogs of Lemma 19 which would make it possible to prove counterparts of Proposition 17.

Non-linear AMP Algorithms: Our results hold for linearized AMP algorithms which are not the state-of-the-art message-passing algorithms for phase retrieval. It would be interesting to extend our results to include general non-linear AMP algorithms. This could provide a unified approach to understanding universality in a broad class of estimators.

Non-Gaussian Priors: Simulations show that the universality of the dynamics of linearized AMP algorithms continues to hold even if the signal is not drawn from a Gaussian prior, but is

an actual image. Hence it would be interesting to extend our results to general i.i.d. priors and more realistic models for signals.

References

[1] Y. Shechtman, Y. C. Eldar, O. Cohen, H. N. Chapman, J. Miao, and M. Segev, "Phase retrieval with application to optical imaging: A contemporary overview," IEEE signal processing magazine, vol. 32, no. 3, pp. 87–109, 2015.

[2] V. Elser, T.-Y. Lan, and T. Bendory, "Benchmark problems for phase retrieval," SIAM Journal on Imaging Sciences, vol. 11, no. 4, pp. 2429–2455, 2018.

[3] A. Walther, "The question of phase retrieval in optics," Optica Acta: International Journal of Optics, vol. 10, no. 1, pp. 41–49, 1963.

[4] E. J. Candès, Y. C. Eldar, T. Strohmer, and V. Voroninski, "Phase retrieval via matrix completion," SIAM review, vol. 57, no. 2, pp. 225–251, 2015.

[5] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval via Wirtinger flow: Theory and algorithms," IEEE Transactions on Information Theory, vol. 61, no. 4, pp. 1985–2007, 2015.

[6] J. Miao, J Kirz, and D Sayre, "The oversampling phasing method," Acta Crystallographica Section D: Biological Crystallography, vol. 56, no. 10, pp. 1312–1315, 2000.

[7] J. Miao, D. Sayre, and H. Chapman, "Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects," JOSA A, vol. 15, no. 6, pp. 1662–1669, 1998.

[8] A. Fannjiang and T. Strohmer, "The numerics of phase retrieval," arXiv preprint arXiv:2004.05788, 2020.

[9] E. J. Candès, T. Strohmer, and V. Voroninski, "Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming," Communications on Pure and Applied Mathematics, vol. 66, no. 8, pp. 1241–1274, 2013.

[10] E. J. Candès and X. Li, "Solving quadratic equations via phaselift when there are about as many equations as unknowns," Foundations of Computational Mathematics, vol. 14, no. 5, pp. 1017–1026, 2014.

[11] S. Bahmani and J. Romberg, "Phase retrieval meets statistical learning theory: A flexible convex relaxation," in Artificial Intelligence and Statistics, 2017, pp. 252–260.

[12] T. Goldstein and C. Studer, "Phasemax: Convex phase retrieval via basis pursuit," IEEE Transactions on Information Theory, vol. 64, no. 4, pp. 2675–2689, 2018.

[13] P. Netrapalli, P. Jain, and S. Sanghavi, "Phase retrieval using alternating minimization," in Advances in Neural Information Processing Systems, 2013, pp. 2796–2804.

[14] J. Sun, Q. Qu, and J. Wright, "A geometric analysis of phase retrieval," Foundations of Computational Mathematics, vol. 18, no. 5, pp. 1131–1198, 2018.

[15] T. T. Cai, X. Li, and Z. Ma, "Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow," The Annals of Statistics, vol. 44, no. 5, pp. 2221–2251, 2016.

[16] M. Bakhshizadeh, A. Maleki, and S. Jalali, "Using black-box compression algorithms for phase retrieval," IEEE Transactions on Information Theory, vol. 66, no. 12, pp. 7978–8001, 2020.

[17] P. Hand, O. Leong, and V. Voroninski, "Phase retrieval under a generative prior," in Advances in Neural Information Processing Systems, 2018, pp. 9136–9146.

[18] E. J. Candès, X. Li, and M. Soltanolkotabi, "Phase retrieval from coded diffraction patterns," Applied and Computational Harmonic Analysis, vol. 39, no. 2, pp. 277–299, 2015.

[19] Y. M. Lu and G. Li, "Phase transitions of spectral initialization for high-dimensional nonconvex estimation," Information and Inference, to appear, 2019.

[20] M. Mondelli and A. Montanari, "Fundamental limits of weak recovery with applications to phase retrieval," Foundations of Computational Mathematics, vol. 19, no. 3, pp. 703–773, 2019.

[21] W. Luo, W. Alghamdi, and Y. M. Lu, "Optimal spectral initialization for signal recovery with applications to phase retrieval," IEEE Transactions on Signal Processing, vol. 67, no. 9, pp. 2347–2356, 2019.

[22] A. Abbara, A. Baker, F. Krzakala, and L. Zdeborová, "On the universality of noiseless linear estimation with respect to the measurement matrix," Journal of Physics A: Mathematical and Theoretical, vol. 53, no. 16, p. 164 001, 2020.

[23] O. Dhifallah, C. Thrampoulidis, and Y. M. Lu, "Phase retrieval via polytope optimization: Geometry, phase transitions, and new algorithms," arXiv preprint arXiv:1805.09555, 2018.

[24] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," IEEE Transactions on Information Theory, vol. 57, no. 2, pp. 764–785, 2011.

[25] J. Barbier, F. Krzakala, N. Macris, L. Miolane, and L. Zdeborová, "Optimal errors and phase transitions in high-dimensional generalized linear models," Proceedings of the National Academy of Sciences, vol. 116, no. 12, pp. 5451–5460, 2019.

[26] D. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 367, no. 1906, pp. 4273–4293, 2009.

[27] H. Monajemi, S. Jafarpour, M. Gavish, and D. L. Donoho, "Deterministic matrices matching the compressed sensing phase transitions of Gaussian random matrices," Proceedings of the National Academy of Sciences, vol. 110, no. 4, pp. 1181–1186, 2013. eprint: https://www.pnas.org/content/110/4/1181.full.pdf.

[28] A. Abbara, A. Baker, F. Krzakala, and L. Zdeborová, "On the universality of noiseless linear estimation with respect to the measurement matrix," arXiv preprint arXiv:1906.04735, 2019.

[29] S. Oymak and B. Hassibi, "A case for orthogonal measurements in linear inverse problems," in 2014 IEEE International Symposium on Information Theory, IEEE, 2014, pp. 3175–3179.

[30] J. Ma, R. Dudeja, J. Xu, A. Maleki, and X. Wang, "Spectral method for phase retrieval: An expectation propagation perspective," IEEE Transactions on Information Theory, vol. 67, no. 2, pp. 1332–1355, 2021.

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

[32] Q. Qu, Y. Zhang, Y. Eldar, and J. Wright, "Convolutional phase retrieval," in Advances in Neural Information Processing Systems, 2017, pp. 6086–6096.

[33] C. Thrampoulidis and B. Hassibi, "Isotropically random orthogonal matrices: Performance of lasso and minimum conic singular values," in 2015 IEEE International Symposium on Information Theory (ISIT), IEEE, 2015, pp. 556–560.

[34] P. Schniter, S. Rangan, and A. K. Fletcher, "Vector approximate message passing for the generalized linear model," in 2016 50th Asilomar Conference on Signals, Systems and Computers, IEEE, 2016, pp. 1525–1529.

[35] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," IEEE Transactions on Information Theory, vol. 65, no. 10, pp. 6664–6684, 2019.

[36] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," in 2017 IEEE International Symposium on Information Theory (ISIT), IEEE, 2017, pp. 501–505.

[37] K. Takeda, S. Uda, and Y. Kabashima, "Analysis of CDMA systems that are characterized by eigenvalue spectrum," EPL (Europhysics Letters), vol. 76, no. 6, p. 1193, 2006.

[38] K. Takeda, A. Hatabu, and Y. Kabashima, "Statistical mechanical analysis of the linear vector channel in digital communication," Journal of Physics A: Mathematical and Theoretical, vol. 40, no. 47, p. 14 085, 2007.

[39] Y. Kabashima, "Inference from correlated patterns: A unified theory for perceptron learning and linear vector channels," in Journal of Physics: Conference Series, IOP Publishing, vol. 95, 2008, p. 012 001.

[40] G. Reeves, "Additivity of information in multilayer networks via additive Gaussian noise transforms," in 2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton), IEEE, 2017, pp. 1064–1070.

[41] J. Barbier, N. Macris, A. Maillard, and F. Krzakala, "The mutual information in random linear estimation beyond iid matrices," in 2018 IEEE International Symposium on Information Theory (ISIT), IEEE, 2018, pp. 1390–1394.

[42] M. Vehkaperä, Y. Kabashima, and S. Chatterjee, "Analysis of regularized LS reconstruction and random matrix ensembles in compressed sensing," IEEE Transactions on Information Theory, vol. 62, no. 4, pp. 2100–2124, 2016.

[43] C.-K. Wen, J. Zhang, K.-K. Wong, J.-C. Chen, and C. Yuen, "On sparse vector recovery performance in structurally orthogonal matrices via lasso," IEEE Transactions on Signal Processing, vol. 64, no. 17, pp. 4519–4533, 2016.

[44] W.-K. Chen and W.-K. Lam, Universality of approximate message passing algorithms, 2020. arXiv: 2003.10431 [math.PR].

[45] S. B. Korada and A. Montanari, "Applications of the Lindeberg principle in communications and statistical learning," IEEE transactions on information theory, vol. 57, no. 4, pp. 2440–2450, 2011.

[46] S. Oymak and J. A. Tropp, "Universality laws for randomized dimension reduction, with applications," Information and Inference: A Journal of the IMA, vol. 7, no. 3, pp. 337–446, 2018.

[47] D. L. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," Discrete & computational geometry, vol. 43, no. 3, pp. 522–541, 2010.

[48] G. W. Anderson, A. Guionnet, and O. Zeitouni, An introduction to random matrices. Cambridge university press, 2010, vol. 118.

[49] J. A. Mingo and R. Speicher, Free probability and random matrices. Springer, 2017, vol. 35.

[50] A. M. Tulino, G. Caire, S. Shamai, and S. Verdu, "Capacity of channels with frequency-selective and time-selective fading," IEEE Transactions on Information Theory, vol. 56, no. 3, pp. 1187–1215, 2010.

[51] B. Farrell, "Limiting empirical singular value distribution of restrictions of discrete Fourier transform matrices," Journal of Fourier Analysis and Applications, vol. 17, no. 4, pp. 733–753, 2011.

[52] G. W. Anderson and B. Farrell, "Asymptotically liberating sequences of random unitary matrices," Advances in Mathematics, vol. 255, pp. 381–413, 2014.

[53] B. Cakmak, M. Opper, O. Winther, and B. H. Fleury, "Dynamical functional theory for compressed sensing," in 2017 IEEE International Symposium on Information Theory (ISIT), IEEE, 2017, pp. 2143–2147.

[54] B. Çakmak and M. Opper, "Memory-free dynamics for the TAP equations of ising models with arbitrary rotation invariant ensembles of random coupling matrices," arXiv preprint arXiv:1901.08583, 2019.

[55] B. Cakmak and M. Opper, "Analysis of Bayesian inference algorithms by the dynamical functional approach," Journal of Physics A: Mathematical and Theoretical, 2020.

[56] B. Çakmak and M. Opper, "A dynamical mean-field theory for learning in restricted boltzmann machines," arXiv preprint arXiv:2005.01560, 2020.

[57] M. Opper and B. Çakmak, "Understanding the dynamics of message passing algorithms: A free probability heuristics," arXiv preprint arXiv:2002.02533, 2020.

[58] S. Goldt, M. Mézard, F. Krzakala, and L. Zdeborová, "Modeling the influence of data structure on learning in neural networks: The hidden manifold model," Physical Review X, vol. 10, no. 4, p. 041 044, 2020.

[59] F. Gerace, B. Loureiro, F. Krzakala, M. Mézard, and L. Zdeborová, "Generalisation error in learning with random features and the hidden manifold model," in International Conference on Machine Learning, PMLR, 2020, pp. 3452–3462.

[60] S. Goldt, G. Reeves, M. Mézard, F. Krzakala, and L. Zdeborová, "The Gaussian equivalence of generative models for learning with two-layer neural networks," arXiv preprint arXiv:2006.14709, 2020.

[61] F. G. Mehler, "Ueber die entwicklung einer function von beliebig vielen variablen nach laplaceschen functionen höherer ordnung.," Journal für die reine und angewandte Mathematik, vol. 1866, no. 66, pp. 161–176, 1866.

[62] D. Slepian, "On the symmetrized Kronecker power of a matrix and extensions of Mehler's formula for hermite polynomials," SIAM Journal on Mathematical Analysis, vol. 3, no. 4, pp. 606–616, 1972.

[63] R. Dudeja, M. Bakhshizadeh, J. Ma, and A. Maleki, "Analysis of spectral methods for phase retrieval with random orthogonal matrices," IEEE Transactions on Information Theory, 2020.

[64] S. T. Belinschi, H. Bercovici, M. Capitaine, and M. Fevrier, "Outliers in the spectrum of large deformed unitarily invariant models," The Annals of Probability, vol. 45, no. 6A, pp. 3571–3625, 2017.

[65] S. T. Belinschi, "The atoms of the free multiplicative convolution of two probability distributions," Integral Equations and Operator Theory, vol. 46, no. 4, pp. 377–386, 2003.

[66] S. T. Belinschi and H. Bercovici, "A new approach to subordination results in free probability," Journal d'Analyse Mathématique, vol. 101, no. 1, pp. 357–365, 2007.

[67] S. T. Belinschi, R. Speicher, J. Treilhard, and C. Vargas, "Operator-valued free multiplicative convolution: Analytic subordination theory and applications to random matrix theory," International Mathematics Research Notices, vol. 2015, no. 14, pp. 5933–5958, 2014.

[68] D. Voiculescu, "Limit laws for random matrices and free products," Inventiones Mathematicae, vol. 104, no. 1, pp. 201–220, 1991.

[69] S. T. Belinschi, "A note on regularity for free convolutions," in Annales de l'Institut Henri Poincare (B) Probability and Statistics, vol. 42, 2006, pp. 635–648.

[70] S. Boucheron, G. Lugosi, and P. Massart, Concentration inequalities: A nonasymptotic theory of independence. Oxford university press, 2013.

[71] R. Dudeja, J. Ma, and A. Maleki, "Information theoretic limits for phase retrieval with sub-sampled Haar sensing matrices," IEEE Transactions on Information Theory, vol. 66, no. 12, pp. 8002–8045, 2020.

[72] A. Guionnet and M. Maida, "A Fourier view on the R-transform and related asymptotics of spherical integrals," Journal of functional analysis, vol. 222, no. 2, pp. 435–490, 2005.

[73] N. R. Chaganty and J. Sethuraman, "Strong large deviation and local limit theorems," The Annals of Probability, pp. 1671–1690, 1993.

[74] G. Reeves, J. Xu, and I. Zadik, "The all-or-nothing phenomenon in sparse linear regression," in Conference on Learning Theory, PMLR, 2019, pp. 2652–2663.

[75] R. N. Bhattacharya and R. R. Rao, Normal approximation and asymptotic expansions. SIAM, 1986, vol. 64.

[76] W. Feller, An introduction to probability theory and its applications. John Wiley & Sons, 2008, vol. 2.

[77] A. W. Van Der Vaart and J. A. Wellner, Weak convergence and empirical processes, 1996.

[78] R. Dudeja and M. Bakhshizadeh, "Universality of linearized message passing for phase retrieval with structured sensing matrices," arXiv preprint arXiv:2008.10503, 2020.

[79] A. Maillard, B. Loureiro, F. Krzakala, and L. Zdeborová, "Phase retrieval in high dimensions: Statistical and computational phase transitions," arXiv preprint arXiv:2006.05228, 2020.

[80] E. Bolthausen, "On the high-temperature phase of the Sherrington-Kirkpatrick model," in Seminar at EURANDOM, Eindhoven, 2009.

[81] F. Krahmer and H. Rauhut, "Structured random measurements in signal processing," GAMM-Mitteilungen, vol. 37, no. 2, pp. 217–238, 2014.

[82] A. M. Tulino, G. Caire, S. Shamai, and S. Verdú, "Capacity of channels with frequency-selective and time-selective fading," IEEE Transactions on Information Theory, vol. 56, no. 3, pp. 1187–1215, 2010.

[83] R. N. Bhattacharya, "On errors of normal approximation," The Annals of Probability, vol. 3, no. 5, pp. 815–828, 1975.

[84] A. Montanari and R. Venkataramanan, "Estimation of low-rank matrices via approximate message passing," The Annals of Statistics, vol. 49, no. 1, pp. 321–345, 2021.

[85] M. Mondelli and R. Venkataramanan, "Approximate message passing with spectral initialization for generalized linear models," in International Conference on Artificial Intelligence and Statistics, PMLR, 2021, pp. 397–405.

[86] M. Spruill, "Asymptotic distribution of coordinates on high dimensional spheres," Electronic communications in probability, vol. 12, pp. 234–247, 2007.

[87] P.-Å. Wedin, "Perturbation theory for pseudo-inverses," BIT Numerical Mathematics, vol. 13, no. 2, pp. 217–232, 1973.

[88] V. V. Petrov, Sums of independent random variables. Springer Science & Business Media, 2012, vol. 82.

[89] N. G. Ushakov, Selected topics in characteristic functions. Walter de Gruyter, 2011.

[90] M. Abramowitz and I. A. Stegun, Handbook of mathematical functions: with formulas, graphs, and mathematical tables. Courier Corporation, 1965, vol. 55.

[91] G. N. Watson, A treatise on the theory of Bessel functions. Cambridge university press, 1995.

[92] G. S. Watson, "Statistics on spheres," 1983.

[93] E. S. Meckes, The random matrix theory of the classical compact groups. Cambridge University Press, 2019, vol. 218.

[94] M. Rudelson and R. Vershynin, "Hanson-wright inequality and sub-Gaussian concentration," Electronic Communications in Probability, vol. 18, 2013.

[95] K. Ball, "An elementary introduction to modern convex geometry," Flavors of geometry, vol. 31, pp. 1–58, 1997.

[96] M. Gromov and V. D. Milman, "A topological application of the isoperimetric inequality," American Journal of Mathematics, vol. 105, no. 4, pp. 843–854, 1983.

[97] R. van Handel, "Probability in high dimension," PRINCETON UNIV NJ, Tech. Rep., 2014.

[98] B. A. Schmitt, "Perturbation bounds for matrix square roots and pythagorean sums," Linear algebra and its applications, vol. 174, pp. 215–227, 1992.

Appendix A: Omitted Proofs from Chapter 3

A.1 Proof of Lemma 7

This section is dedicated to the proof of Lemma 7.

Proof of Lemma 7. It is sufficient to show each item holds almost surely.

1. The argument for this part is a minor modification of the argument sketched in [86]. To prove statement (1) it suffices to show that

$$\frac{1}{m}\sum_{i=1}^{n}\delta_{\sqrt{m}|A_{i1}|} \xrightarrow{d} |Z|, \tag{A.1}$$

almost surely. Because if we have (A.1), then for every bounded continuous function f,

$$f\left(\mathcal{T}\left(m|A_{i1}|^2\right)\right) = g\left(\sqrt{m}|A_{1i}|\right),$$

where $g(x) = f(\mathcal{T}(x^2))$ is a bounded continuous function as well. Hence by (A.1),

$$\frac{1}{m}\sum_{i=1}^{m} f(T_i) \to \mathbb{E}\left[g(Z)\right] = \mathbb{E}\left[f\left(\mathcal{T}\left(|Z|^2\right)\right)\right],$$

which implies $\frac{1}{m}\sum_{i=1}^{m}\delta_{T_i} \xrightarrow{d} \mathcal{L}_T$.

To show (A.1), note that $\mathbf{A_1}$ has the same distribution as $\frac{\mathbf{z}}{\|\mathbf{z}\|}$, where $\mathbf{z} = (\mathbf{z_1}, ..., \mathbf{z_m})$, and

$z_i \overset{i.i.d.}{\sim} \mathcal{CN}(0,1)$. Let Φ denote the cumulative distribution function of $|Z|$ and define

$$\begin{aligned} F_m(t) &\overset{\text{def}}{=} \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}\left(\sqrt{m}|A_{1i}| \leq t\right), \\ G_m(t) &\overset{\text{def}}{=} \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}\left(|z_i| \leq t\right). \end{aligned}$$

Then, we have

$$F_m(t) \overset{d}{=} G_m\left(t\frac{\|\mathbf{z}\|}{\sqrt{m}}\right). \tag{A.2}$$

Moreover,

$$\begin{aligned} &G_m\left(t\frac{\|\mathbf{z}\|}{\sqrt{m}}\right) - \Phi(t) = \\ &\quad G_m\left(t\frac{\|\mathbf{z}\|}{\sqrt{m}}\right) - \Phi\left(t\frac{\|\mathbf{z}\|}{\sqrt{m}}\right) + \Phi\left(t\frac{\|\mathbf{z}\|}{\sqrt{m}}\right) - \Phi(t) \\ &\quad \xrightarrow{a.s.} 0 + 0. \end{aligned}$$

$G_m(t\|\mathbf{z}\|) - \Phi(t\|\mathbf{z}\|)$ goes to 0 almost surely by Glivenko-Cantelli lemma. Furthermore, since

$$\frac{\|\mathbf{z}\|}{\sqrt{m}} \xrightarrow{a.s.} 1,$$

and Φ is a continuous function we conclude that

$$\Phi\left(t\frac{\|\mathbf{z}\|}{\sqrt{m}}\right) - \Phi(t) \overset{a.s.}{\to} 0.$$

Hence,

$$F_m(t) \to \Phi(t),$$

almost surely which yields (A.1).

2. We now focus on the proof of statement (2). Let

$$\mathcal{C}_k \stackrel{\text{def}}{=} \left[1 + \frac{1}{k}, k\right], \quad k \in \mathbb{N}.$$

We will show that

$$Q_m(\lambda) \to Q(\lambda) \quad \forall\, \lambda \,\in\, \mathcal{C}_k, \tag{A.3}$$

almost surely. This means there is a set $\mathcal{C}'_k$, with measure 0, out of which we have the convergence for all $\lambda \in \mathcal{C}_k$. If we define $\mathcal{C}' \stackrel{\text{def}}{=} \bigcup_{k=1}^{\infty} \mathcal{C}'_k$, then $Q_m(\lambda) \to Q(\lambda) \quad \forall \lambda \in (1, \infty)$ out of $\mathcal{C}'$ and clearly $\mathbb{P}(\mathcal{C}') = 0$.

First note that $\mathbf{A_1} \stackrel{\mathbf{d}}{=} \frac{\mathbf{z}}{\|\mathbf{z}\|}$, where

$$\mathbf{z} = (\mathbf{z_1}, ..., \mathbf{z_m}), \quad \mathbf{z_i} \overset{\mathbf{i.i.d.}}{\sim} \mathcal{CN}(\mathbf{0}, \mathbf{1}).$$

Define

$$\tilde{Q}_m(\lambda) \stackrel{\text{def}}{=} \frac{1}{m} \sum_{i=1}^{m} \frac{|z_i|^2}{\lambda - \mathcal{T}\left(|z_i|^2\right)}. \tag{A.4}$$

Note that for a fixed λ we have $\tilde{Q}_m(\lambda) \to Q(\lambda)$ almost surely by the strong law of large numbers. Since $\tilde{Q}_m(\lambda)$ is a decreasing function in λ and we have $\tilde{Q}_m(\lambda) \to Q(\lambda) \quad \forall \lambda \in \mathcal{C}_k \cap \mathbb{Q}$ almost surely, we obtain $\tilde{Q}_m(\lambda) \to Q(\lambda)$ for all $\lambda \in \mathcal{C}_k$ with probability 1. Hence, it suffices to show under an event that holds with probability 1,

$$Q_m(\lambda) - \tilde{Q}_m(\lambda) \to 0 \quad \forall \lambda \in \mathcal{C}_k. \tag{A.5}$$

To prove (A.5), we will find a sequence τ_m such that $\tau_m \to 0$ as $m \to \infty$, and,

$$\sum_{m \geq 1} \mathbb{P}\left(\sup_{\lambda \in \mathcal{C}_k} \left|Q_m(\lambda) - \tilde{Q}_m(\lambda)\right| > \tau_m\right) < \infty.$$

With this, Borel-Cantelli lemma yields that event

$$E = \left\{ \sup_{\lambda \in \mathcal{C}_k} \left| Q_m(\lambda) - \tilde{Q}_m(\lambda) \right| > \tau_m \text{ infinitely often} \right\}$$

has measure 0. Out of the event E we have (A.5) as it was desired.

Define the events:

$$E_1 \triangleq \left\{ \sup_{i \leq m} |z_i| \leq \sqrt{6 \log m} \right\},$$
$$E_{2,\epsilon} \triangleq \left\{ \left| \frac{\|\mathbf{z}\|^2}{m} - 1 \right| \leq \epsilon \right\},$$

where ϵ is parameter we will set later. Note that,

$$\begin{aligned} \left| Q_m(\lambda) - \tilde{Q}_m(\lambda) \right| \leq & \\ & \sum_{i=1}^{m} \frac{|z_i|^2}{\|\mathbf{z}\|^2} \left| \frac{\frac{\|\mathbf{z}\|^2}{m}}{\lambda - \mathcal{T}\left(|z_i|^2\right)} - \frac{1}{\lambda - \mathcal{T}\left(\frac{m}{\|\mathbf{z}\|^2}|z_i|^2\right)} \right| \\ & \leq \mathsf{I} + \mathsf{II}, \end{aligned}$$

where we defined the terms I, II as:

$$\mathsf{I} = \left| \frac{\|\boldsymbol{z}\|^2}{m} - 1 \right| \cdot \sum_{i=1}^{m} \frac{|z_i|^2}{\|\boldsymbol{z}\|^2} \cdot \left| \frac{1}{\lambda - \mathcal{T}\big(|z_i|^2\big)} \right|$$
$$\mathsf{II} = \sum_{i=1}^{m} \frac{|z_i|^2}{\|\boldsymbol{z}\|^2} \cdot \frac{\left| \mathcal{T}\big(|z_i|^2\big) - \mathcal{T}\big(\frac{m|z_i|^2}{\|\boldsymbol{z}\|^2}\big) \right|}{\left| \lambda - \mathcal{T}\big(|z_i|^2\big) \right| \cdot \left| \lambda - \mathcal{T}\big(\frac{m|z_i^2|}{\|\boldsymbol{z}\|^2}\big) \right|}.$$

Using the fact that $\boldsymbol{z} \in E_1 \cap E_{2,\epsilon}$ and $\lambda \in \mathcal{C}_k$, we have,

$$\begin{aligned}
\mathsf{I} &\leq k\epsilon, \\
\mathsf{II} &\leq k^2 \cdot \max_{i \leq n} \left| \mathcal{T}\left(|z_i|^2\right) - \mathcal{T}\left(\frac{m|z_i|^2}{\|\boldsymbol{z}\|^2}\right) \right|.
\end{aligned}$$

Observe that, on the event $E_1 \cap E_{2,\epsilon}$,

$$\begin{aligned}
\left| |z_i|^2 - \frac{m}{\|\mathbf{z}\|^2}|z_i|^2 \right| &\leq |z_i|^2 \left| 1 - \frac{m}{\|\mathbf{z}\|^2} \right| \\
&\leq 6\log(m) \cdot \frac{\epsilon}{1-\epsilon}.
\end{aligned}$$

Since $\mathcal{T}$ was assumed to be Lipchitz,

$$\mathsf{II} \leq k^2 \cdot \|\mathcal{T}\|_{\mathsf{Lip}} \cdot 6\log(m) \cdot \frac{\epsilon}{1-\epsilon},$$

where $\|\mathcal{T}\|_{\mathsf{Lip}}$ denotes the Lipchitz constant of $\mathcal{T}$. Hence, when $m \geq e^2$, setting $\epsilon = \frac{1}{\log^2(m)} \leq 0.5$, we obtain, on the event $E_1 \cap E_{2,\epsilon}$

$$\left| Q_m(\lambda) - \tilde{Q}_m(\lambda) \right| \leq \tau_m, \ \forall \ \lambda \ \in \ \mathcal{C}_k. \tag{A.6}$$

where

$$\tau_m = \frac{k}{\log^2(m)} + \frac{2k^2 \cdot \|\mathcal{T}\|_{\mathsf{Lip}}}{\log(m)}.$$

Note that $\tau_m \to 0$ as $m \to \infty$ as required. And,

$$\begin{aligned}\mathbb{P}\left(\sup_{\lambda\in\mathcal{C}_k}\left|Q_m(\lambda)-\tilde{Q}_m(\lambda)\right|>\tau_m\right)&\\ &\leq \mathbb{P}\left(E_1^c\right)+\mathbb{P}\left(E_{2,\epsilon}^c\right)\\ &\leq 2\cdot m^{-2}+2e^{-\frac{m}{8\log^4(m)}},\end{aligned}$$

where the last step follows from standard bounds on the tail Gaussian random variables and χ^2 random variables. In particular, we have,

$$\sum_{m\geq 1}\mathbb{P}\left(\sup_{\lambda\in\mathcal{C}_k}\left|Q_m(\lambda)-\tilde{Q}_m(\lambda)\right|>\tau_m\right)<\infty,$$

as required.

3. The proof is similar to the proof of the second statement. Hence, we skip the details.

□

A.2 Proof of Proposition 2

This section is devoted to the proof of Proposition 2. We denote the functions $\Lambda, \psi_1, \psi_2, \psi_3$ (recall (3.1)) with $\mathcal{T} = \mathcal{T}_{\mathsf{opt}}$ as $\Lambda_{\mathsf{opt}}, \psi_1^{\mathsf{opt}}, \psi_2^{\mathsf{opt}}, \psi_3^{\mathsf{opt}}$ and those with $\mathcal{T} = \mathcal{T}_{\mathsf{opt},\epsilon}$ as $\Lambda_\epsilon, \psi_1^\epsilon, \psi_2^\epsilon, \psi_3^\epsilon$. Define the random variables:

$$Z \sim \mathcal{CN}(0,1),\ T_{\mathsf{opt}} = \mathcal{T}_{\mathsf{opt}}(|Z|^2),\ T_\epsilon = \mathcal{T}_{\mathsf{opt},\epsilon}(|Z|^2).$$

Next we observe that the function $\mathcal{T}_{\mathsf{opt},\epsilon}$ is a bounded, strictly increasing, Lipchitz function and consequently T_ϵ has a density with respect to the Lebesgue measure. Hence by the rescale and shift argument outlined in Remark 2, Theorem 1 applies to a equivalent modification of $\mathcal{T}_{\mathsf{opt},\epsilon}$ which can used to infer the corresponding result for $\mathcal{T}_{\mathsf{opt},\epsilon}$ (after another rescale and shift argument). This gives us the result:

$$\frac{|\boldsymbol{x}_\star^{\mathsf{H}}\hat{\boldsymbol{x}}_\epsilon|^2}{\|\boldsymbol{x}\|^2} \overset{\text{a.s.}}{\to} \begin{cases} 0, & \psi_1^\epsilon(\tau_r^\epsilon) < \frac{\delta}{\delta-1}, \\ \frac{\left(\frac{\delta}{\delta-1}\right)^2 - \frac{\delta}{\delta-1}\cdot\psi_2^\epsilon(\theta_\star^\epsilon)}{\psi_3^\epsilon(\theta_\star^\epsilon)^2 - \frac{\delta}{\delta-1}\cdot\psi_2^\epsilon(\theta_\star^\epsilon)}, & \psi_1(\tau_r^\epsilon) > \frac{\delta}{\delta-1}. \end{cases}, \tag{A.7}$$

where $\tau_r^\epsilon \overset{\text{def}}{=} \arg\min_{\tau\in[1,\infty)} \Lambda_\epsilon(\tau)$ and $\theta_\star^\epsilon$ is the solution to the fixed point equation (in τ): $\psi_1^\epsilon(\tau) = \delta/(\delta-1)$ which is guaranteed to exist uniquely provided $\psi_1(\tau_r^\epsilon) > \delta/(\delta-1)$. First we observe that,

$$\Lambda_\epsilon'(\tau) = 1 - \left(1 - \frac{1}{\delta}\right)\cdot\frac{\mathbb{E}G_\epsilon^2(\tau)}{(\mathbb{E}G_\epsilon(\tau))^2},\ G_\epsilon(\tau) = (\tau - T_\epsilon)^{-1}.$$

In particular, at $\tau = 1$, we have,

$$\begin{aligned}\Lambda_\epsilon'(1) &= 1 - \left(1 - \frac{1}{\delta}\right)\cdot\frac{(1+\epsilon)^2+1}{(1+\epsilon)^2} \\ &\implies \lim_{\epsilon\downarrow 0}\Lambda'(1) = \frac{2-\delta}{\delta},\end{aligned}$$

and,

$$\psi_1^\epsilon(1) = 2 + \epsilon.$$

We consider the following two cases.

Case 1: $1 < \delta < 2$. Lemma 10 shows that $\Lambda_\epsilon(\tau)$ is convex on $[1,\infty)$. When $\delta < 2$, $\Lambda_\epsilon'(1) > 0$ for ϵ small enough, and hence Λ_ϵ is strictly increasing and $\tau_r^\epsilon = 1$. Moreover, in this case, for ϵ small enough,

$$\frac{\delta}{\delta-1} = 2 + \frac{2-\delta}{\delta-1} > 2 + \epsilon = \psi_1^\epsilon(1).$$

Hence, using (A.7),

$$\lim_{\epsilon \downarrow 0} \lim_{\substack{m,n \to \infty \\ m = \delta n}} \frac{|\boldsymbol{x}_\star^{\mathsf{H}} \hat{\boldsymbol{x}}_\epsilon|^2}{n} = 0.$$

Case 2: $\delta > 2$ In this case, for small enough ϵ, $\Lambda_\epsilon'(1) < 0$. Hence the τ_r^ϵ, the minimizer of the convex function Λ_ϵ occurs in the region $(1, \infty)$. This means it satisfies the optimality condition:

$$\Lambda_\epsilon'(\tau_r^\epsilon) = 0 \Leftrightarrow \psi_2(\tau_r^\epsilon) = \frac{\delta}{\delta - 1}.$$

Next we claim that, $\forall \tau \in [1, \infty)$,

$$\psi_1^\epsilon(\tau) > \psi_2^\epsilon(\tau) \Leftrightarrow \mathbb{E}[G_\epsilon(\tau)] \cdot \mathbb{E}[|Z|^2 G_\epsilon(\tau)] > \mathbb{E}[G_\epsilon^2(\tau)],$$

which is a consequence of Chebychev's association inequality (Fact 1) with the choice:

$$\begin{aligned} B = G_\epsilon(\tau),\ A = |Z|, \\ f(a) = a^2 \left(\tau - \mathcal{T}_\epsilon\left(a^2\right)\right),\ g(a) = \left(\tau - \mathcal{T}_\epsilon\left(a^2\right)\right)^{-1}. \end{aligned}$$

In particular we have $\psi_1^\epsilon(\tau_r^\epsilon) > \delta/(\delta - 1)$, and hence Theorem 1 gives us:

1. There exists a unique solution $\theta_\star^\epsilon \in (\tau_r^\epsilon, \infty)$ such that $\psi_1^\epsilon(\theta_\star^\epsilon) = \delta/(\delta - 1)$,
2. and,

$$\frac{|\boldsymbol{x}_\star^{\mathsf{H}} \hat{\boldsymbol{x}}_\epsilon|^2}{\|\boldsymbol{x}\|^2} \xrightarrow{\text{a.s.}} \frac{\left(\frac{\delta}{\delta-1}\right)^2 - \frac{\delta}{\delta-1} \cdot \psi_2^\epsilon(\theta_\star^\epsilon)}{\psi_3^\epsilon(\theta_\star^\epsilon)^2 - \frac{\delta}{\delta-1} \cdot \psi_2^\epsilon(\theta_\star^\epsilon)}.$$

Next we claim that,

$$1 < \liminf_{\epsilon \downarrow 0} \theta_\star^\epsilon \leq \limsup_{\epsilon \downarrow 0} \theta_\star^\epsilon < \infty.$$

To see this, observe

$$\psi_1^\epsilon(\theta_\star^\epsilon) = \frac{\mathbb{E}\frac{|Z|^2(|Z|^2+\epsilon)}{(\theta_\star^\epsilon-1)(|Z|^2+\epsilon)+1}}{\mathbb{E}\frac{(|Z|^2+\epsilon)}{(\theta_\star^\epsilon-1)(|Z|^2+\epsilon)+1}}.$$

If $\liminf_{\epsilon\downarrow 0}\theta_\star^\epsilon = 1$, one can select a subsequence along which $\psi_1^\epsilon(\theta_\star^\epsilon) \to \mathbb{E}|Z|^4 = 2$ by dominated convergence which contradicts: $\psi_2^\epsilon(\theta_\star^\epsilon) = \delta/(\delta-1) < 2$. Likewise if $\limsup_{\epsilon\downarrow 0}\theta_\star^\epsilon = \infty$, one can find a subsequence along which $\theta_\star^\epsilon \to \infty$ and, by dominated convergence,

$$\psi_1^\epsilon(\theta_\star^\epsilon) = \frac{\mathbb{E}\frac{|Z|^2(|Z|^2+\epsilon)(\theta_\star^\epsilon-1)}{(\theta_\star^\epsilon-1)(|Z|^2+\epsilon)+1}}{\mathbb{E}\frac{(|Z|^2+\epsilon)(\theta_\star^\epsilon-1)}{(\theta_\star^\epsilon-1)(|Z|^2+\epsilon)+1}} \to 1,$$

which contradicts $\psi_1^\epsilon(\theta_\star^\epsilon) = \delta/(\delta-1) < 1 \;\forall\; \delta \in (2,\infty)$. We can now conclude that,

$$\liminf_{\epsilon\downarrow 0}\theta_\star^\epsilon = \limsup_{\epsilon\downarrow 0}\theta_\star^\epsilon = \theta_\star^{\mathsf{opt}},$$

where $\theta_\star^{\mathsf{opt}}$ is the unique solution to $\psi_1^{\mathsf{opt}}(\tau) = \delta/(\delta-1)$ in $\tau \in (1,\infty)$ guaranteed by Proposition 1 (due to [30]). This is because, by selecting a subsequence along with $\theta_\star^\epsilon \to \liminf_{\epsilon\downarrow 0}\theta_\star^\epsilon$, we can conclude that, along that subsequence,

$$\frac{\delta}{\delta-1} = \psi_1^\epsilon(\theta_\star^\epsilon) \to \psi_1^{\mathsf{opt}}\left(\liminf_{\epsilon\downarrow 0}\theta_\star^\epsilon\right).$$

This implies,

$$\psi_1^{\mathsf{opt}}\left(\liminf_{\epsilon\downarrow 0}\theta_\star^\epsilon\right) = \frac{\delta}{\delta-1},$$

and analogously,

$$\psi_1^{\mathsf{opt}}\left(\limsup_{\epsilon\downarrow 0}\theta_\star^\epsilon\right) = \frac{\delta}{\delta-1}.$$

Since Proposition 1 guarantees that the equation $\psi_1^{\mathsf{opt}}(\tau) = \delta/(\delta-1)$ has a unique solution in

$(1, \infty)$ we get,

$$\liminf_{\epsilon \downarrow 0} \theta_\star^\epsilon = \limsup_{\epsilon \downarrow 0} \theta_\star^\epsilon = \theta_\star^{\mathsf{opt}}.$$

Dominated convergence now yields,

$$\psi_i^\epsilon(\theta_\star^\epsilon) \to \psi_i^{\mathsf{opt}}(\theta_\star^{\mathsf{opt}}), \text{ as } \epsilon \downarrow 0 \; \forall \; i = 1, 2, 3,$$

and consequently, almost surely,

$$\lim_{\epsilon \downarrow 0} \lim_{\substack{m,n \to \infty, \\ m = n\delta}} \frac{|\boldsymbol{x}_\star^{\mathsf{H}} \hat{\boldsymbol{x}}_\epsilon|^2}{n} \overset{\text{a.s.}}{=} \frac{\left(\frac{\delta}{\delta - 1}\right)^2 - \frac{\delta}{\delta - 1} \cdot \psi_2^{\mathsf{opt}}(\theta_\star^{\mathsf{opt}})}{\psi_3^{\mathsf{opt}}(\theta_\star^{\mathsf{opt}})^2 - \frac{\delta}{\delta - 1} \cdot \psi_2^{\mathsf{opt}}(\theta_\star^{\mathsf{opt}})}.$$

The right hand side of the above display can be simplified to:

$$\frac{\left(\frac{\delta}{\delta - 1}\right)^2 - \frac{\delta}{\delta - 1} \cdot \psi_2^{\mathsf{opt}}(\theta_\star^{\mathsf{opt}})}{\psi_3^{\mathsf{opt}}(\theta_\star^{\mathsf{opt}})^2 - \frac{\delta}{\delta - 1} \cdot \psi_2^{\mathsf{opt}}(\theta_\star^{\mathsf{opt}})} = \frac{\theta_\star^{\mathsf{opt}} - 1}{\theta_\star^{\mathsf{opt}} - \frac{1}{\delta}}.$$

This clean formula is due to [30] and we refer the reader to Appendix B in [30] for a proof.

A.3 Miscellaneous results

Fact 1 (Chebychev Association Inequality, [70])**.** *Let A, B be r.v.s and $B \geq 0$. Suppose f, g are two non-decreasing functions. Then,*

$$\mathbb{E}[B]\mathbb{E}[Bf(A)g(A)] \geq \mathbb{E}[f(A)B]\mathbb{E}[g(A)B].$$

Furthermore, if, $\mathbb{P}(B = 0) = 0$ and,

$$\mathbb{P}\left(f(A) = x\right) = 0, \; \mathbb{P}\left(g(A) = x\right) = 0, \; \forall \; x \; \in \; \mathbb{R},$$

then, the above inequality is strict.

Proof. The proof of the inequality appears in [70]. Inspecting the proof we can derive a sufficient condition for the inequality to be strict. The proof in [70] shows,

$$2 \cdot (\mathbb{E}[B]\mathbb{E}[Bf(A)g(A)] - \mathbb{E}[f(A)B]\mathbb{E}[g(A)B]) = \\ \mathbb{E}BB'(f(A) - f(A')) \cdot (g(A) - g(A')).$$

where (B', A') is an independent sample of the random variables (B, A). Since, f, g are increasing $(f(A) - f(A')) \cdot (g(A) - g(A')) \geq 0$ and $B \geq 0, B' \geq 0$. Hence the equality is tight iff:

$$BB'(f(A) - f(A')) \cdot (g(A) - g(A')) \overset{\text{a.s.}}{=} 0,$$

which is ruled out by the assumptions of the claim. □

Appendix B: Omitted Proofs from Chapter 4

B.1 Proofs from Section 4.5

In this section, we collect the missing proofs from Section 4.5. We begin with a lemma describing the joint distribution of the phase retrieval measurements $\boldsymbol{y}$ and the side information $\boldsymbol{z}$.

Lemma 34. *Let $\boldsymbol{x}_\star, \boldsymbol{y}$ and $\boldsymbol{z}$ denote the signal vector, the measurements and side information sampled from the phase retrieval with side information model (see* (4.2)*,* (4.1) *and* (4.4)*). Then conditioned on $\boldsymbol{x}_\star$, $\boldsymbol{y}$ and $\boldsymbol{z}$ are independent with marginal distributions:*

$$\begin{aligned}
&\boldsymbol{y} \overset{d}{=} m|\boldsymbol{U}|^2 + \sigma\boldsymbol{\epsilon},\ \boldsymbol{U} \sim \mathit{Unif}\left(\mathbb{S}^{m-1}\right),\ \boldsymbol{\epsilon} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}_m\right), \\
&\boldsymbol{z} \sim \mathcal{N}\left(\boldsymbol{0}, 2 \cdot \boldsymbol{I}_{\lfloor \Delta \cdot m \rfloor}\right).
\end{aligned}$$

Furthermore, since the above distributions do not depend on $\boldsymbol{x}_\star$, this result holds even without conditioning on $\boldsymbol{x}_\star$.

Proof. From (4.2), (4.1) and (4.4), we know that,

$$\boldsymbol{y} = m|\boldsymbol{A}\boldsymbol{x}_\star|^2 + \sigma\boldsymbol{\epsilon},\ z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(\langle \boldsymbol{w}_i, \boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} \rangle, 1\right),\ i \in \{1, 2 \ldots, \lfloor \Delta m \rfloor\},$$

where $\boldsymbol{A}$ is a uniformly random $m \times n$ partial unitary matrix and the matrices $\boldsymbol{w}_i \overset{\text{i.i.d.}}{\sim} \mathsf{GUE}(n)$. Since $\boldsymbol{A}$ is independent of $\boldsymbol{w}_1, \boldsymbol{w}_2 \ldots, \boldsymbol{w}_{\lfloor \Delta m \rfloor}$, we have $\boldsymbol{y}, \boldsymbol{z}$ are conditionally independent given $\boldsymbol{x}_\star$. Let $\boldsymbol{B}$ be the $n \times n$ unitary matrix whose first column $\boldsymbol{B}_1 = \boldsymbol{x}_\star$ (and the remaining columns

can be arbitrary). Then note that, conditioned on $\boldsymbol{x}_\star$,

$$\begin{aligned}\boldsymbol{A}\boldsymbol{x}_\star &\overset{(1)}{=} \boldsymbol{A}\boldsymbol{B}\boldsymbol{B}^{\mathsf{H}}\boldsymbol{x}_\star \\ &= \boldsymbol{A}\boldsymbol{B}\boldsymbol{e}_1 \\ &\overset{\text{d},(2)}{=} \boldsymbol{A}\boldsymbol{e}_1.\end{aligned}$$

In the above display, the step marked (1) used the fact that $\boldsymbol{B}\boldsymbol{B}^{\mathsf{H}} = \boldsymbol{I}_n$, the distribution inequality (2) used the fact that since $\boldsymbol{A}$ is a uniformly random partial unitary matrix, its distribution is invariant to left multiplication by a unitary matrix. Finally note that the first column of a partial unitary matrix $\boldsymbol{A}\boldsymbol{e}_1 \sim \text{Unif}\left(\mathbb{S}^{m-1}\right)$. This gives us:

$$\boldsymbol{y} \overset{\text{d}}{=} m|\boldsymbol{U}|^2 + \sigma\boldsymbol{\epsilon},\ \boldsymbol{U} \sim \text{Unif}\left(\mathbb{S}^{m-1}\right),\ \boldsymbol{\epsilon} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}_m\right). \tag{B.1}$$

Next observe that since $\boldsymbol{w}_i \sim \mathsf{GUE}(n)$, conditioned on $\boldsymbol{x}_\star$,

$$\langle \boldsymbol{w}_i, \boldsymbol{x}_\star\boldsymbol{x}_\star^{\mathsf{H}}\rangle \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0,1\right),\ i \in \{1, 2\ldots, \lfloor \Delta m \rfloor\}.$$

Hence, conditioned on $\boldsymbol{x}_\star$,

$$z_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0,2\right).$$

This proves the claim of the lemma. Note that since the conditional distributions do not depend on $\boldsymbol{x}_\star$, this result holds even without conditioning on $\boldsymbol{x}_\star$ □

The remainder of this section is organized as follows:

1. Section B.1.1 is devoted to the proof of Proposition 6.
2. Section B.1.2 is devoted to the proof of Lemma 12.

B.1.1 Proof of Proposition 6

Proof. Let $\Delta > 0$ be fixed to any value that guarantees:

$$\mathbf{I}\left(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}\right) = o(m).$$

By the chain rule for mutual information,

$$\begin{aligned}
\mathbf{I}\left(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}\right) &= \mathbf{I}\left(\boldsymbol{y}; \boldsymbol{A}, \boldsymbol{W}\right) + \mathbf{I}\left(\boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W} | \boldsymbol{y}\right) \\
&= \mathbf{I}\left(\boldsymbol{y}; \boldsymbol{A}\right) + \mathbf{I}\left(\boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W} | \boldsymbol{y}\right) \\
&\geq \mathbf{I}\left(\boldsymbol{y}; \boldsymbol{A}\right).
\end{aligned}$$

Consequently $\mathbf{I}\left(\boldsymbol{y}; \boldsymbol{A}\right) = o(m)$. This means,

$$\mathbf{H}\left(\boldsymbol{y} \mid \boldsymbol{A}\right) = \mathbf{H}\left(\boldsymbol{y}\right) - o(m) \tag{B.2}$$

$$\mathbf{H}\left(\boldsymbol{y}, \boldsymbol{z} \mid \boldsymbol{A}, \boldsymbol{W}\right) = \mathbf{H}\left(\boldsymbol{y}, \boldsymbol{z}\right) - o(m). \tag{B.3}$$

In order to prove the claim of the proposition, we will costruct an upper bound and a lower bound on the quantity $\mathbf{H}\left(\boldsymbol{z} \mid \boldsymbol{y}, \boldsymbol{A}, \boldsymbol{W}\right)$. Comparing the upper and lower bound will give us the claim of the proposition.

$$\begin{aligned}
\mathbf{H}\left(\boldsymbol{z} \mid \boldsymbol{y}, \boldsymbol{A}, \boldsymbol{W}\right) &= \mathbf{H}\left(\boldsymbol{y}, \boldsymbol{z} \mid \boldsymbol{A}, \boldsymbol{W}\right) - \mathbf{H}\left(\boldsymbol{y} \mid \boldsymbol{A}, \boldsymbol{W}\right) \\
&= \mathbf{H}\left(\boldsymbol{y}, \boldsymbol{z} \mid \boldsymbol{A}, \boldsymbol{W}\right) - \mathbf{H}\left(\boldsymbol{y} \mid \boldsymbol{A}\right) \\
&\overset{\text{(a)}}{=} \mathbf{H}\left(\boldsymbol{y}, \boldsymbol{z}\right) - \mathbf{H}\left(\boldsymbol{y}\right) - o(m) \\
&\overset{\text{(b)}}{=} \mathbf{H}\left(\boldsymbol{z}\right) - o(m) \\
&\overset{\text{(c)}}{=} \lfloor \Delta m \rfloor \cdot h(2) \cdot (1 - o(1)) \qquad \text{(B.4)} \\
&= \Delta m \cdot h(2) \cdot (1 - o(1)). \qquad \text{(B.5)}
\end{aligned}$$

In the equality marked (a), we used the conclusions derived in (B.2) and (B.3). In the step marked (b), we used the fact that $\boldsymbol{y}, \boldsymbol{z}$ are independent (see Lemma 34). In step (c) we defined $h(v) \stackrel{\text{def}}{=} \frac{1}{2}\ln(2\pi v)$, which is the entropy of $\mathcal{N}(0, v)$ and recalled the claim of Lemma 34: $z_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. On the other hand we can upper bound $\mathbf{H}\left(\boldsymbol{z} \mid \boldsymbol{y}, \boldsymbol{A}, \boldsymbol{W}\right)$ as follows:

$$\begin{aligned}
\mathbf{H}\left(\boldsymbol{z} \mid \boldsymbol{y}, \boldsymbol{A}, \boldsymbol{W}\right) &\leq \sum_{i=1}^{\lfloor \Delta m \rfloor} \mathbf{H}\left(z_i \mid \boldsymbol{y}, \boldsymbol{A}, \boldsymbol{W}\right) \\
&\stackrel{\text{(a)}}{\leq} \sum_{i=1}^{\lfloor \Delta m \rfloor} \mathbb{E} h\left(\text{Var}(z_i | \boldsymbol{y}, \boldsymbol{A}, \boldsymbol{W})\right) \\
&\stackrel{\text{(b)}}{\leq} \sum_{i=1}^{\lfloor \Delta m \rfloor} h(\mathbb{E}\text{Var}(z_i | \boldsymbol{y}, \boldsymbol{A}, \boldsymbol{W})) \qquad \text{(B.6)}
\end{aligned}$$

In the step marked (a) we used the fact that the Gaussian Distribution has the maximal entropy for a fixed variance and in step (b) we used the concavity of h. Next we compute $\mathbb{E}\text{Var}(Z_i | \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{W}_{1:\Delta m})$. We have,

$$\begin{aligned}
\mathbb{E}\text{Var}(z_i | \boldsymbol{y}, \boldsymbol{A}, \boldsymbol{W}) &= \mathbb{E}(z_i - \mathbb{E}[z_i | \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{W}])^2 \\
&\stackrel{\text{(a)}}{=} \mathbb{E}\langle \boldsymbol{w}_i, \boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}[\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} | \boldsymbol{y}, \boldsymbol{A}, \boldsymbol{W}] \rangle^2 + 1 \\
&\stackrel{\text{(b)}}{=} \mathbb{E}\langle \boldsymbol{w}_i, \boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}[\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} | \boldsymbol{y}, \boldsymbol{A}] \rangle^2 + 1 \\
&\stackrel{\text{(c)}}{=} \mathbb{E}\| \boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}[\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} | \boldsymbol{y}, \boldsymbol{A}] \|^2 + 1. \qquad \text{(B.7)}
\end{aligned}$$

In the above display, the equality (a) follows from the fact that $z_i \sim \mathcal{N}\left(\langle \boldsymbol{w}_i, \boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} \rangle, 1\right)$ and equality (b) used the fact that $\boldsymbol{W}$ is independent of $\boldsymbol{x}_\star, \boldsymbol{y}, \boldsymbol{A}$. In the step (c), we used the following property of a GUE matrix: for a deterministic Hermitian matrix $\boldsymbol{M}$, $\langle \boldsymbol{w}_i, \boldsymbol{M} \rangle \sim \mathcal{N}\left(0, \|\boldsymbol{M}\|^2\right)$. (B.5), (B.6) and (B.7) give us the conclusion:

$$\Delta m \cdot h(\mathbb{E}\| \boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}[\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} | \boldsymbol{Y}, \boldsymbol{A}] \|^2 + 1) \geq \Delta m \cdot h(2)(1 - o(1)).$$

Since h is an increasing function this gives us:

$$\liminf_{\substack{m,n\to\infty \\ m=\delta n}} \mathbb{E}_{\boldsymbol{x}_\star,\boldsymbol{y},\boldsymbol{A}} \|\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}[\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} | \boldsymbol{y}, \boldsymbol{A}]\|^2 \geq 1.$$

On the other hand, by the optimality of the Bayes estimator, we have: $\mathbb{E}_{\boldsymbol{x}_\star,\boldsymbol{y},\boldsymbol{A}} \|\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}[\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} | \boldsymbol{y}, \boldsymbol{A}]\|^2 \leq \mathbb{E}_{\boldsymbol{x}_\star,\boldsymbol{y},\boldsymbol{A}} \|\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} - \mathbf{0}\|^2 = 1$. Hence,

$$\lim_{\substack{m,n\to\infty \\ m=n\delta}} \mathbb{E}_{\boldsymbol{x}_\star,\boldsymbol{y},\boldsymbol{A}} \|\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} - \mathbb{E}[\boldsymbol{x}_\star \boldsymbol{x}_\star^{\mathsf{H}} | \boldsymbol{y}, \boldsymbol{A}]\|^2 = 1.$$

This concludes the proof of the proposition. □

B.1.2 Proof of Lemma 12

Proof. Through out this proof C refers to a finite non-negative constant independent of m, n that can possibly depend on δ, σ^2, Δ. This constant may change from line to line. Recall that,

$$\mathbf{I}(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}) = \mathbf{H}(\boldsymbol{y}, \boldsymbol{z}) - \mathbf{H}\left(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}\right).$$

We can split $\mathbf{H}\left(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}\right)$ as follows:

$$\begin{aligned}
\mathbf{H}\left(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}\right) &= -\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} \left(\int p(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}) \ln p(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}) \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} \right) \\
&= -\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} \left(\int_{\mathcal{E}_m} p(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}) \ln p(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}) \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} \right) \\
&\qquad - \mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} \left(\int_{\mathcal{E}_m^c} p(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}) \ln p(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}) \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} \right) \\
&\overset{\text{(a)}}{=} -\int_{\mathcal{E}_m} \mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} \left[p(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}) \ln p(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}) \right] \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} \\
&\qquad - \int_{\mathcal{E}_m^c} \mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}) \ln p(\boldsymbol{y}, \boldsymbol{z} | \boldsymbol{A}, \boldsymbol{W}) \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z}.
\end{aligned}$$

In the step marked (a) we used Fubini's Theorem. Likewise we can split $\mathbf{H}(\boldsymbol{y}, \boldsymbol{z})$ as follows:

$$\mathbf{H}(\boldsymbol{y}, \boldsymbol{z}) = -\int_{\mathcal{E}_m} p(\boldsymbol{y}, \boldsymbol{z}) \ln p(\boldsymbol{y}, \boldsymbol{z}) \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} - \int_{\mathcal{E}_m^c} p(\boldsymbol{y}, \boldsymbol{z}) \ln p(\boldsymbol{y}, \boldsymbol{z}) \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z}.$$

Hence,

$$\mathbf{I}(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}) = \mathsf{I} + \mathsf{II} + \mathsf{III},$$

where the terms $\mathsf{I}, \mathsf{II}, \mathsf{III}$ are defined as:

$$\begin{aligned}
\mathsf{I} &\stackrel{\text{def}}{=} -\int_{\mathcal{E}_m} p(\boldsymbol{y}, \boldsymbol{z}) \ln p(\boldsymbol{y}, \boldsymbol{z}) \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} + \int_{\mathcal{E}_m} \mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} \left[p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}) \ln p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}) \right] \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z}, \\
\mathsf{II} &\stackrel{\text{def}}{=} -\int_{\mathcal{E}_m^c} p(\boldsymbol{y}, \boldsymbol{z}) \ln p(\boldsymbol{y}, \boldsymbol{z}) \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z}, \\
\mathsf{III} &\stackrel{\text{def}}{=} \int_{\mathcal{E}_m^c} \mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}) \ln p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}) \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z}.
\end{aligned}$$

Analysis of I : Consider the following inequality:

$$\ln(x) \leq (x-1) \implies x \ln(x) \leq x(x-1), \ \forall x \geq 0.$$

Applying this to $\frac{p(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{A},\boldsymbol{W})}{p(\boldsymbol{y},\boldsymbol{z})}$, we obtain,

$$p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}) \ln p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}) \leq p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}) \ln(p(\boldsymbol{y}, \boldsymbol{z})) - p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}) + \frac{p^2(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W})}{p(\boldsymbol{y}, \boldsymbol{z})}.$$

Substituting this in the expression for I we obtain,

$$\begin{aligned}
\mathsf{I} &\leq -\int_{\mathcal{E}_m} p(\boldsymbol{y}, \boldsymbol{z}) \ln p(\boldsymbol{y}, \boldsymbol{z}) \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} + \int_{\mathcal{E}_m} p(\boldsymbol{y}, \boldsymbol{z}) \ln p(\boldsymbol{y}, \boldsymbol{z}) \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} - \Pr(\mathcal{E}_m) \\
&\qquad + \int_{\mathcal{E}_m} \frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W})}{p(\boldsymbol{y}, \boldsymbol{z})} \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} \\
&= \left(\int_{\mathcal{E}_m} \frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W})}{p(\boldsymbol{y}, \boldsymbol{z})} \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} - 1 \right) + \mathbb{P}(\mathcal{E}_m^c).
\end{aligned}$$

Hence we have,

$$\mathsf{I} \leq \left(\int_{\mathcal{E}_m} \frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{A},\boldsymbol{W})}{p(\boldsymbol{y},\boldsymbol{z})} \,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{z} - 1 \right) + \mathbb{P}(\mathcal{E}_m^c). \tag{B.8}$$

Analysis of II : We can handle II as follows:

$$\begin{aligned}
\mathsf{II} &\overset{\text{def}}{=} -\int_{\mathcal{E}_m^c} p(\boldsymbol{y},\boldsymbol{z}) \ln p(\boldsymbol{y},\boldsymbol{z}) \,\mathrm{d}\boldsymbol{y}\,\mathrm{d}\boldsymbol{z} \\
&\overset{\text{(a)}}{=} -\int_{\mathcal{E}_m^c} p(\boldsymbol{y}) \ln p(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y} + \mathbf{H}\left(\boldsymbol{z}\right) \mathbb{P}(\mathcal{E}_m^c) \\
&\overset{\text{(b)}}{\leq} -\int_{\mathcal{E}_m^c} p(\boldsymbol{y}) \ln p(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y} + C \cdot m \cdot \mathbb{P}(\mathcal{E}_m^c) \\
&= -\int_{\mathcal{E}_m^c} p(\boldsymbol{y}) \ln \mathbb{E}_{\boldsymbol{x},\boldsymbol{A}} p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{A}) \,\mathrm{d}\boldsymbol{y} + C \cdot m \cdot \mathbb{P}(\mathcal{E}_m^c) \\
&\overset{\text{(c)}}{\leq} -\mathbb{E}_{\boldsymbol{x},\boldsymbol{A},\boldsymbol{y}} \mathbf{1}_{\mathcal{E}_m^c} \ln p(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{A}) + C \cdot m \cdot \mathbb{P}(\mathcal{E}_m^c) \\
&= \frac{1}{2\sigma^2} \mathbb{E}\|\boldsymbol{y} - m|\boldsymbol{A}\boldsymbol{x}|^2\|^2 \mathbf{1}_{\mathcal{E}_m^c} + \frac{m \ln(2\pi\sigma^2)}{2} \mathbb{P}(\mathcal{E}_m^c) + C \cdot m \cdot \mathbb{P}(\mathcal{E}_m^c) \\
&\leq C \cdot \left(m^2 \mathbb{P}(\mathcal{E}_m^c) \mathbb{E}\|\boldsymbol{A}\boldsymbol{x}\|_4^4 + \mathbb{E}\|\boldsymbol{y}\|^2 \mathbf{1}_{\mathcal{E}_m^c} \right) + C \cdot m \cdot \mathbb{P}(\mathcal{E}_m^c).
\end{aligned}$$

In the step marked (a) we used the fact that $\boldsymbol{y}, \boldsymbol{z}$ are marginally independent. In the step marked (b) we used the fact that $\mathbf{H}\left(\boldsymbol{z}\right) \leq Cm$ for a suitable C. In the step marked (c) we applied Jensen's Inequality and note that the random variables $\boldsymbol{x}$, $\boldsymbol{A}$ and $\boldsymbol{y}$ are independent. Note that by Cauchy Schwartz Inequality, we have,

$$\mathbb{E}\|\boldsymbol{y}\|^2 \mathbf{1}_{\mathcal{E}_m^c} \leq \sqrt{\mathbb{E}\|\boldsymbol{y}\|^4 \cdot \mathbb{P}(\mathcal{E}_m^c)}.$$

It is also straightforward to obtain the following estimates by simple moment computations:

$$\mathbb{E}\|\boldsymbol{A}\boldsymbol{x}\|_4^4 = \sum_{i=1}^m \mathbb{E}|\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|^4 = \sum_{i=1}^m \mathbb{E}\|\boldsymbol{a}_i\|^4 |x_1|^4 \leq m\mathbb{E}|x_1|^4 \leq \frac{C}{m},\ \mathbb{E}\|\boldsymbol{y}\|^4 \leq Cm^2.$$

for some $0 \leq C < \infty$. This gives us:

$$\mathsf{II} \leq Cm \left(\mathbb{P}(\mathcal{E}_m^c) + \sqrt{\mathbb{P}(\mathcal{E}_m^c)} \right). \tag{B.9}$$

Analysis of III : Next we analyze the term III:

$$\mathsf{III} = \int_{\mathcal{E}_m^c} \mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}) \ln p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}) \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z}$$

Noting that:

$$\begin{aligned}
&\ln p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}) = \ln \mathbb{E}_{\boldsymbol{x}} p(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W}, \boldsymbol{x}) \\
&= \ln \mathbb{E}_{\boldsymbol{x}} e^{-\|\boldsymbol{y} - m|\boldsymbol{A}\boldsymbol{x}|^2\|^2/2\sigma^2} + \ln \mathbb{E}_{\boldsymbol{x}} \left[\prod_{i=1}^{\lfloor \Delta m \rfloor} e^{-(z_i - \langle \boldsymbol{x}\boldsymbol{x}^{\mathsf{H}}, \boldsymbol{w}_i \rangle)^2/2} \right] - \frac{m \ln(2\pi\sigma^2) + \lfloor \Delta m \rfloor \ln(2\pi)}{2} \\
&\leq -\frac{m \ln(2\pi\sigma^2) + \lfloor \Delta m \rfloor \ln(2\pi)}{2} \\
&\leq Cm.
\end{aligned}$$

Hence we obtain,

$$\mathsf{III} \leq Cm\mathbb{P}(\mathcal{E}_m^c).$$

Combining the estimates on $\mathsf{I}, \mathsf{II}, \mathsf{III}$ we obtain,

$$\mathbf{I}(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}) \leq \left(\int_{\mathcal{E}_m} \frac{\mathbb{E}_{\boldsymbol{A},\boldsymbol{W}} p^2(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{A}, \boldsymbol{W})}{p(\boldsymbol{y}, \boldsymbol{z})} \, \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{z} - 1 \right) + C \cdot m \cdot \sqrt{\mathbb{P}(\mathcal{E}_m^c)}.$$

□

B.2 Proofs of Local Central Limit Theorems

The proofs of the Local central limit theorems are based on the classical approach using characteristic functions. Section B.2.1 contains the proof of the local CLT in Proposition 7 and Section B.2.2 contains the proof of the local CLT in Proposition 8. The proofs use some standard properties of characteristic functions which have been collected in Appendix B.8 for reference. We will also rely on some analytic properties of the Tilted Exponential distribution and Tilted Wishart distribution given in Appedices B.6.1 and B.6.2.

B.2.1 Proof of Proposition 7

Proof. Recall the random variable U was defined as:

$$U = \sum_{i=1}^{m} u_i,\ u_i \sim \mathsf{TExp}\left(\hat{\lambda}_1(\sigma), y_i\right),\ i \in [\Delta m],$$

where,

$$\hat{\lambda}_1(\sigma) \stackrel{\text{def}}{=} \arg\max_{\lambda \in \mathbb{R}} \left(\lambda - \hat{\mathbb{E}}_Y \ln \mathbb{E}_{E \sim \mathsf{Exp}(1)} e^{\lambda E} \psi_\sigma(E - Y)\right).$$

Note that $\lambda = \hat{\lambda}_1(\sigma)$ satisfies the first order stationarity condition:

$$1 = \frac{1}{m} \sum_{i=1}^{m} \frac{\mathbb{E}_{E \sim \mathsf{Exp}(1)} E \cdot e^{\hat{\lambda}_1(\sigma) E} \psi_\sigma(E - y_i)}{\mathbb{E}_{E \sim \mathsf{Exp}(1)} e^{\hat{\lambda}_1(\sigma) E} \psi_\sigma(E - y_i)} \Leftrightarrow \sum_{i=1}^{m} \mathbb{E} u_i = m.$$

From here on, throughout this proof, we will shorthand $\hat{\lambda}_1(\sigma)$ as simply $\hat{\lambda}_1$. Define the centered random variables: $\check{u}_i = u_i - \mathbb{E} u_i$ and centered and normalized random variable:

$$\check{U} = \frac{U - m}{\sqrt{m}} = \frac{\sum_{i=1}^{m} \check{u}_i}{\sqrt{m}}.$$

Let $\hat{v}(\sigma)$ denote the variance of $\check{U}$:

$$\hat{v}(\sigma) \stackrel{\text{def}}{=} \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}\check{u}_i^2 = \frac{1}{m}\sum_{i=1}^{m} \sigma^2_{\mathsf{TExp}}\left(\hat{\lambda}_1, y_i\right).$$

Again for ease of notation we will short hand $\hat{v}(\sigma)$ as $\hat{v}$. Let $\check{F}$ denote the density of $\check{U}$. Let $\check{\psi}(t) = \mathbb{E}e^{\mathbf{i}t\check{U}}$ denote the characteristic function of $\check{U}$. By the change of variable formula, we have,

$$F_{\hat{\lambda}_1,\boldsymbol{y}}(m) = \frac{\check{F}(0)}{\sqrt{m}}$$

Hence we focus on computing $\check{F}(0)$. By the Fourier Inversion formula (Lemma 7, Appendix B.8) we have,

$$\begin{aligned}
|\check{F}(u) - \phi_{\sqrt{\hat{v}}}(u)| &= \frac{1}{2\pi}\left|\int_{\mathbb{R}} e^{-\mathbf{i}tu}\left(\check{\psi}(t) - e^{-\frac{\hat{v}t^2}{2}}\right)\mathrm{d}t\right| \\
&\stackrel{(a)}{\leq} \frac{1}{2\pi}\left(\int_{|t|\leq t_1}\left|\check{\psi}(t) - e^{-\frac{\hat{v}t^2}{2}}\right|\mathrm{d}t + \int_{t_1\leq|t|\leq t_2\sqrt{m}}|\check{\psi}(t)|\,\mathrm{d}t + \int_{|t|\geq t_2\sqrt{m}}|\check{\psi}(t)|\,\mathrm{d}t + \int_{|t|\geq t_1} e^{-\frac{\hat{v}t^2}{2}}\,\mathrm{d}t\right) \\
&\stackrel{(b)}{\leq} \frac{1}{2\pi}\left(\underbrace{\int_{|t|\leq t_1}\left|\check{\psi}(t) - e^{-\frac{\hat{v}t^2}{2}}\right|\mathrm{d}t}_{(1)} + \underbrace{\int_{t_1\leq|t|\leq t_2\sqrt{m}}|\check{\psi}(t)|\,\mathrm{d}t}_{(2)} + \underbrace{\int_{|t|\geq t_2\sqrt{m}}|\check{\psi}(t)|\,\mathrm{d}t}_{(3)} + \frac{2}{\hat{v}}e^{-\frac{\hat{v}t_1^2}{2}}\right)
\end{aligned}$$

In the step marked (a), the cutoff parameters t_1, t_2 are arbitrary and will be fixed later. In the step marked (b), we used standard bounds on the tail of a gaussian integral (see Lemma 47, Appendix B.9). In the following sequence of steps, we upper bound each of the error terms $(1), (2)$ and (3). We will be able to show, for a suitable selection of t_1, t_2, that,

$$(1) + (2) + (3) + \frac{2}{\hat{v}}e^{-\frac{\hat{v}t_1^2}{2}} \leq \frac{C(K)\cdot\ln(m)}{\sqrt{m}}.$$

This gives us,

$$\left|\check{F}(0) - \frac{1}{\sqrt{2\pi\hat{v}}}\right| \leq \frac{C(K)\ln(m)}{\sqrt{m}} \implies \left|F_{\hat{\lambda}_1,\boldsymbol{y}}(m) - \frac{1}{\sqrt{2\pi\hat{v}\cdot m}}\right| \leq \frac{C(K)\ln(m)}{m},$$

which is the claim of this proposition. The remaining proof is devoted to the analysis of (1), (2) and (3).

Analysis of (1): Recall $\check{\psi}(t) = \mathbb{E}e^{\mathbf{i}t\check{U}}$ and $f(x) = e^{\mathbf{i}tx}$ is bounded, t-Lipchitz function of x. Applying the Berry-Eseen Inequality (Theorem 9, Appendix B.8), we have,

$$\left|\check{\psi}(t) - e^{-\frac{\hat{v}t^2}{2}}\right| \leq \frac{C\cdot(1+\sqrt{\hat{v}}|t|)\cdot\rho_3}{\sqrt{m\cdot\hat{v}^3}}.$$

In the above display, C is a universal constant and ρ_3 is given by:

$$\begin{aligned}\rho_3 &= \frac{1}{m}\sum_{i=1}^{m}\mathbb{E}|u_i - \mathbb{E}u_i|^3 \\ &\leq \frac{8}{m}\sum_{i=1}^{m}\mathbb{E}|u_i|^3 \\ &\overset{(c)}{\leq} C\left(1 + |\hat{\lambda}_1|^3 + \frac{1}{m}\sum_{i=1}^{m}|y_i|^3\right).\end{aligned}$$

In the step marked (c) we used the estimate on $\mathbb{E}|u_i|^3$ proved in Lemma 43. Integrating the pointwise bound above we obtain:

$$(1) \leq C\cdot\left(1 + |\hat{\lambda}_1|^3 + \frac{1}{m}\sum_{i=1}^{m}|y_i|^3\right)\cdot\frac{t_1(1+\sqrt{\hat{v}}t_1)}{\sqrt{m\cdot\hat{v}^3}}.$$

We set:

$$t_1 = \sqrt{\frac{2\ln(m)}{\hat{v}}}.$$

This gives us:

$$(1) \leq \frac{C}{\hat{v}^2} \cdot \left(1 + |\hat{\lambda}_1|^3 + \frac{1}{m}\sum_{i=1}^{m} |y_i|^3\right) \cdot \frac{\ln(m)}{\sqrt{m}} \leq \frac{C(K) \cdot \ln(m)}{\sqrt{m}}.$$

Analysis of (2): Let $(u'_1, u'_2 \dots u'_m)$ be independent and identically distributed as $(u_1, u_2 \dots u_m)$.

Note that,

$$\left|\mathbb{E}e^{\mathbf{i}t\check{u}_i}\right|^2 = \left|\mathbb{E}e^{\mathbf{i}tu_i}\right|^2 = \mathbb{E}e^{\mathbf{i}t(u_i - u'_i)}.$$

Hence,

$$\left|\check{\psi}(t)\right|^2 = \prod_{i=1}^{m} \left|\mathbb{E}\exp\left(\frac{\mathbf{i}t\check{u}_i}{\sqrt{m}}\right)\right|^2 = \prod_{i=1}^{m} \mathbb{E}\exp\left(\frac{\mathbf{i}t(u_i - u'_i)}{\sqrt{m}}\right).$$

By the Taylor's theorem for CF (Theorem 8, Appendix B.8), we have,

$$\mathbb{E}\exp\left(\frac{\mathbf{i}t(u_i - u'_i)}{\sqrt{m}}\right) = 1 - \frac{\mathbb{E}(u_i - u'_i)^2 \cdot t^2}{2m} + E_i,\ |E_i| \leq \frac{\mathbb{E}|u_i - u'_i|^3 \cdot |t|^3}{6m\sqrt{m}}.$$

Now consider any $t \leq t_2\sqrt{m}$:

$$\begin{aligned}
|\check{\psi}(t)|^2 &= \prod_{i=1}^{m}\left(1 - \frac{\mathbb{E}(u_i - u'_i)^2 \cdot t^2}{2m} + E_i\right) \\
&\leq \prod_{i=1}^{m}\left(1 - \frac{\mathbb{E}(u_i - u'_i)^2 \cdot t^2}{2m} + \frac{\mathbb{E}|u_i - u'_i|^3 \cdot |t|^3}{6m\sqrt{m}}\right) \\
&\leq \exp\left(-\frac{t^2}{2m}\sum_{i=1}^{m}\mathbb{E}(u_i - u'_i)^2 + \frac{|t|^3}{6m\sqrt{m}}\sum_{i=1}^{m}\mathbb{E}|u_i - u'_i|^3\right).
\end{aligned}$$

Next we observe that,

$$\frac{1}{2m}\sum_{i=1}^{m}\mathbb{E}(u_i - u'_i)^2 = \hat{v}.$$

We set:

$$t_2 = \frac{\hat{v}}{2} \cdot \left(\frac{1}{m} \sum_{i=1}^{m} \mathbb{E}|u_i - u_i'|^3 \right)^{-1}.$$

This ensures, for any $|t| \leq t_2\sqrt{m}$, we have,

$$|\check{\psi}(t)|^2 \leq \exp\left(-\frac{t^2}{2m} \sum_{i=1}^{m} \mathbb{E}(u_i - u_i')^2 + \frac{t^3}{6m\sqrt{m}} \sum_{i=1}^{m} \mathbb{E}|u_i - u_i'|^3 \right) \leq \exp\left(-\frac{\hat{v}t^2}{2} \right).$$

Consequently,

$$\begin{aligned}
(2) = \int_{t_1 \leq |t| \leq t_2\sqrt{m}} |\check{\psi}(t)| \, \mathrm{d}t \leq \int_{t_1 \leq |t| \leq t_2\sqrt{m}} e^{-\hat{v}t^2/4} \, \mathrm{d}t &\leq \int_{t_1 \leq |t|} e^{-\hat{v}t^2/4} \, \mathrm{d}t \\
&\overset{\text{(d)}}{\leq} \frac{4}{\hat{v}} \exp\left(-\frac{\hat{v}t_1^2}{4} \right) \\
&\overset{\text{(e)}}{=} \frac{4}{\hat{v}\sqrt{m}} \leq \frac{C(K)}{\sqrt{m}}.
\end{aligned}$$

In the step marked (d), we used the standard bound on gaussian tail integrals (Lemma 47) and in the step marked (e) we substituted the value of t_1 fixed in the analysis of (1). Finally, to wrap up this step, we note that there exists a finite positive constant $C(K)$ such that,

$$t_2 \geq \frac{1}{C(K)}.$$

Indeed,

$$\frac{1}{m} \sum_{i=1}^{m} \mathbb{E}|u_i - u_i'|^3 \leq \frac{8}{m} \sum_{i=1}^{m} \mathbb{E}|u_i|^3 \leq C \left(1 + |\hat{\lambda}_1|^3 + \frac{1}{m} \sum_{i=1}^{m} |y_i|^3 \right) \leq C(K),$$

and,

$$t_2 = \frac{\hat{v}}{2} \cdot \left(\frac{1}{m} \sum_{i=1}^{m} \mathbb{E}|u_i - u_i'|^3 \right)^{-1} \geq \frac{1}{C(K)}.$$

Analysis of (3): Recall the term (3) was given by:

$$(3) = \int_{|t| \geq t_2 \sqrt{m}} |\check{\psi}(t)| \, \mathrm{d}t = \sqrt{m} \int_{|t| \geq t_2} |\check{\psi}(t\sqrt{m})| \, \mathrm{d}t.$$

By AM-GM for non-negative real numbers we have,

$$\left| \check{\psi}(t\sqrt{m}) \right|^2 = \prod_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i} t u_i} \right|^2 \leq \left(\frac{1}{m} \sum_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i} t u_i} \right|^2 \right)^m.$$

We use two different strategies to further control the above bound:

1. Applying Lemma 43, we obtain,

$$\frac{1}{m} \sum_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i} t u_i} \right|^2 \leq \frac{C}{|t|^2} \cdot \frac{1}{m} \sum_{i=1}^{m} (1 + |\hat{\lambda}_1| + |y|_i)^2 \leq \frac{C(K)}{|t|^2}.$$

2. The above bound tells us that for $|t| \geq \sqrt{2C(K)}$, we have,

$$\frac{1}{m} \sum_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i} t u_i} \right|^2 \leq \frac{1}{2}.$$

Applying Lemma B.51 in Appendix B.8, we can find a constant $0 < \eta(K) < 1$ depending only on K such that,

$$\frac{1}{m} \sum_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i} t u_i} \right|^2 \leq (1 - \eta(K)), \ \forall |t| \geq t_2.$$

We can combine the above to bounds to control (3) as follows:

$$
\begin{aligned}
(3) &= \sqrt{m} \int_{|t| \geq t_2} |\check{\psi}(t\sqrt{m})| \, \mathrm{d}t \\
&\leq \sqrt{m} \int_{|t| \geq t_2} \left(\frac{1}{m} \sum_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i} t u_i} \right|^2 \right)^{\frac{m}{2}} \mathrm{d}t \\
&\leq \sqrt{m} \cdot C(K) \int_{|t| \geq t_2} \left(\frac{1}{m} \sum_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i} t u_i} \right|^2 \right)^{\frac{m}{2}-1} \cdot \frac{1}{|t|^2} \, \mathrm{d}t \\
&\leq C(K) \cdot \sqrt{m} \cdot (1 - \eta(K))^{\frac{m}{2}-1} \cdot \int_{|t| \geq t_2} \frac{1}{|t|^2} \, \mathrm{d}t \\
&\leq \frac{C(K)}{\sqrt{m}}.
\end{aligned}
$$

This concludes the proof of the proposition. □

B.2.2 Proof of Proposition 8

Proof. Recall that the random variable $\boldsymbol{S}$ was defined as:

$$
\boldsymbol{S} = \sum_{k=1}^{m} \boldsymbol{S}_k, \;\; \boldsymbol{S}_k = \begin{bmatrix} s_k & \sqrt{s_k s_k'} e^{\mathbf{i}\theta_k} \\ \sqrt{s_k s_k'} e^{-\mathbf{i}\theta_k} & r_k' \end{bmatrix} \sim \mathsf{TWis}\left(\hat{\lambda}_2(q;\sigma), \hat{\phi}_2(q;\sigma), y_i \right),
$$

where $(\hat{\lambda}_2(q;\sigma), \hat{\phi}(q;\sigma))$ solved the concave variational problem:

$$
(\hat{\lambda}_2(q;\sigma), \hat{\phi}(q;\sigma)) \stackrel{\text{def}}{=} \arg \max_{(\lambda,\phi) \in \mathbb{R}} \left(2\lambda + q\phi - \hat{\mathbb{E}}_Y \ln Z_{\mathsf{TWis}}(\lambda, \phi, Y) \right).
$$

Throughout this proof for easy of notation we will omit the dependence of quantities like $\hat{\lambda}_2(q;\sigma)$, $\hat{\phi}(q;\sigma)$ and $\hat{\boldsymbol{V}}(q;\sigma)$ on q, σ and denote them by $\hat{\lambda}_2, \hat{\phi}, \hat{\boldsymbol{V}}$. Since the optimizer of the variational

problem lies in a compact set, we know that $\hat{\lambda}_2, \hat{\phi}$ satisfy the first order optimality conditions:

$$2 = \frac{1}{m}\sum_{i=1}^{m} \frac{\partial_\lambda Z_{\mathsf{TWis}}\left(\hat{\lambda}_2, \hat{\phi}, y_i\right)}{Z_{\mathsf{TWis}}\left(\hat{\lambda}_2, \hat{\phi}, y_i\right)} \overset{\text{(a)}}{=} \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}(s_i + s_i')$$
$$q = \frac{1}{m}\sum_{i=1}^{m} \frac{\partial_\phi Z_{\mathsf{TWis}}\left(\hat{\lambda}_2, \hat{\phi}, y_i\right)}{Z_{\mathsf{TWis}}\left(\hat{\lambda}_2, \hat{\phi}, y_i\right)} \overset{\text{(a)}}{=} \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}\sqrt{s_i s_i'}\cos(\theta_i).$$

In the steps marked (a), we used the formula for the normalizing constant $Z_{\mathsf{TWis}}(\lambda, \phi, y)$ given in Definition 6 to compute the partial derivatives. It is also clear from Definition 6 that:

$$\mathbb{E}s_i = \mathbb{E}s_i', \ \mathbb{E}\sqrt{s_i s_i'}\sin(\theta) = 0.$$

Hence the first order optimality conditions imply:

$$\mathbb{E}\boldsymbol{S} = m\boldsymbol{Q}.$$

Next we define the centered random variables:

$$\check{\boldsymbol{S}}_i = \boldsymbol{S}_i - \mathbb{E}\boldsymbol{S}_i, \ \check{\boldsymbol{S}} = \frac{\boldsymbol{S} - \mathbb{E}\boldsymbol{S}}{\sqrt{m}} = \frac{1}{\sqrt{m}}\sum_{i=1}^{m} \check{\boldsymbol{S}}_i.$$

Note that,

$$\mathbb{E}\mathsf{Vec}\left(\check{\boldsymbol{S}}\right)\mathsf{Vec}\left(\hat{\boldsymbol{S}}\right)^{\mathsf{H}} = \frac{1}{m}\sum_{i=1}^{m} \boldsymbol{\Sigma}_{\mathsf{TWis}}\left(\hat{\lambda}_2, \hat{\phi}, y_i\right) = \hat{\boldsymbol{V}}.$$

Let $\check{H}$ denote the density of $\check{\boldsymbol{S}}$. We note that it is sufficient to study the asymptotics of $\check{H}(\boldsymbol{0})$ since

by the change of variable formula we have:

$$H_{\hat{\lambda}_2,\hat{\phi},\boldsymbol{y}}(m\boldsymbol{Q}) = \frac{\check{H}(\mathbf{0})}{m^2}.$$

In the remainder of the proof we focus on developing asymptotic expansions for $\check{H}$. We define the characteristic function of $\check{\boldsymbol{S}}$:

$$\check{\Psi}(\boldsymbol{t}) = \mathbb{E}\exp\left(\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}\left(\check{\boldsymbol{S}}\right)\rangle\right).$$

By the Fourier Inversion formula (Lemma 7) we have,

$$\check{H}(\boldsymbol{U}) = \frac{1}{(2\pi)^4}\int_{\mathbb{R}^4} e^{-\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}(\boldsymbol{U})\rangle}\check{\Psi}(\boldsymbol{t})\,\mathrm{d}\boldsymbol{t}.$$

Applying the inversion formula to $\mathcal{N}\left(\mathbf{0}, \hat{\boldsymbol{V}}\right)$ gives us:

$$\frac{1}{\sqrt{(2\pi)^4\det(\hat{\boldsymbol{V}})}} e^{-\frac{1}{2}\mathsf{Vec}(\boldsymbol{U})^{\mathsf{H}}\hat{\boldsymbol{V}}^{-1}\mathsf{Vec}(\boldsymbol{U})} = \frac{1}{(2\pi)^4}\int_{\mathbb{R}^4} e^{-\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}(\boldsymbol{U})\rangle} e^{-\frac{1}{2}\boldsymbol{t}^{\mathsf{H}}\hat{\boldsymbol{V}}\boldsymbol{t}}\,\mathrm{d}\boldsymbol{t}.$$

Setting $\boldsymbol{U} = 0$ and computing the error between the above two displays we obtain: Appendix B.8) we have,

$$\begin{aligned}(2\pi)^4\left|\check{H}(\mathbf{0}) - \frac{1}{\sqrt{(2\pi)^4\det(\hat{\boldsymbol{V}})}}\right| &= \left|\int_{\mathbb{R}^4} e^{-\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}(\boldsymbol{U})\rangle}\left(\check{\Psi}(\boldsymbol{t}) - e^{-\frac{1}{2}\boldsymbol{t}^{\mathsf{H}}\hat{\boldsymbol{V}}\boldsymbol{t}}\right)\mathrm{d}\boldsymbol{t}\right| \\ &\overset{(a)}{\leq} \big((1)+(2)+(3)+(4)\big),\end{aligned}$$

where,

$$
\begin{aligned}
(1) &\stackrel{\text{def}}{=} \int_{\|\boldsymbol{t}\| \le t_1} \left| \check{\Psi}(\boldsymbol{t}) - e^{-\frac{1}{2}\boldsymbol{t}^{\mathsf{H}} \hat{\boldsymbol{V}} \boldsymbol{t}} \right| \mathrm{d}t, \\
(2) &\stackrel{\text{def}}{=} \int_{t_1 \le \|\boldsymbol{t}\| \le t_2\sqrt{m}} |\check{\Psi}(\boldsymbol{t})| \, \mathrm{d}t, \\
(3) &\stackrel{\text{def}}{=} \int_{\|\boldsymbol{t}\| \ge t_2\sqrt{m}} |\check{\Psi}(\boldsymbol{t})| \, \mathrm{d}t, \\
(4) &\stackrel{\text{def}}{=} \int_{\|\boldsymbol{t}\| \ge t_1} e^{-\frac{1}{2}\boldsymbol{t}^{\mathsf{H}} \hat{\boldsymbol{V}} \boldsymbol{t}} \, \mathrm{d}t.
\end{aligned}
$$

In the step marked (a), the cutoff parameters t_1, t_2 are arbitrary and will be fixed later. We will be able to choose t_1, t_2 such that the following bound holds:

$$
(1) + (2) + (3) + (4) \le \frac{C(K) \cdot \ln^5(m)}{\sqrt{m}}.
$$

This gives us,

$$
\left| \check{H}(\mathbf{0}) - \frac{1}{\sqrt{(2\pi)^4 \det(\hat{\boldsymbol{V}})}} \right| \le \frac{C(K) \ln^5(m)}{\sqrt{m}},
$$

and hence,

$$
\left| H_{\hat{\lambda}_2, \hat{\phi}, \boldsymbol{y}}(m\boldsymbol{Q}) - \frac{1}{\sqrt{(2\pi m)^4 \det(\hat{\boldsymbol{V}})}} \right| \le \frac{C(K) \ln^5(m)}{m^2\sqrt{m}},
$$

which is the claim of this proposition. The remaining proof is devoted to the analysis of (1), (2), (3) and (4).

Analysis of (1): Recall $\check{\Psi}(\boldsymbol{t}) = \mathbb{E} e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}(\check{\boldsymbol{S}}) \rangle}$ and $f(\boldsymbol{x}) = e^{\mathbf{i}\langle \boldsymbol{t}, \boldsymbol{x} \rangle}$ is bounded, $\|\boldsymbol{t}\|$-Lipchitz function

of $\boldsymbol{x}$. Applying the Berry-Eseen Inequality (Theorem 9, Appendix B.8), we have,

$$\left|\check{\Psi}(\boldsymbol{t}) - e^{-\frac{1}{2}\boldsymbol{t}^{\mathsf{H}}\hat{\boldsymbol{V}}\boldsymbol{t}}\right| \leq \frac{C \cdot (1 + \|\hat{\boldsymbol{V}}\|^{1/2}\|\boldsymbol{t}\|) \cdot \rho_3}{\sqrt{m \cdot \lambda^3_{\min}(\hat{\boldsymbol{V}})}}.$$

In the above display, C is a universal constant and ρ_3 is given by:

$$\begin{aligned}
\rho_3 &= \frac{1}{m}\sum_{i=1}^{m} \mathbb{E}\|\mathsf{Vec}\left(\boldsymbol{S}_i\right) - \mathbb{E}\mathsf{Vec}\left(\boldsymbol{S}_i\right)\|^3 \\
&\leq \frac{8}{m}\sum_{i=1}^{m} \mathbb{E}\|\mathsf{Vec}\left(\boldsymbol{S}_i\right)\|^3 \\
&\leq \frac{16}{m}\sum_{i=1}^{m} \mathbb{E}(s_i^3 + \mathbb{E}s_i'^3) \\
&\overset{\text{(a)}}{\leq} C\left(1 + |\hat{\lambda}_2|^3 + |\hat{\phi}|^3 + \frac{1}{m}\sum_{i=1}^{m}|y_i|^3\right).
\end{aligned}$$

In the step marked (a) we used the estimate on $\mathbb{E}s_i^3$ proved in Lemma 44. Recalling the assumptions

$$K^{-1} \leq \lambda_{\min}(\hat{\boldsymbol{V}}) \leq \lambda_{\max}(\hat{\boldsymbol{V}}) \leq K,\ |\hat{\phi}| + |\hat{\lambda}_2| < K,\ \frac{1}{m}\sum_{i=1}^{m}|y_i|^3 \leq K,$$

we obtain,

$$\left|\check{\Psi}(\boldsymbol{t}) - e^{-\frac{1}{2}\boldsymbol{t}^{\mathsf{H}}\hat{\boldsymbol{V}}\boldsymbol{t}}\right| \leq \frac{C(K) \cdot (1 + \|\boldsymbol{t}\|)}{\sqrt{m}}$$

Integrating the pointwise bound above we obtain:

$$(1) \leq \frac{C(K) \cdot (1 + t_1) \cdot t_1^4}{\sqrt{m}}$$

We set:

$$t_1^2 = \frac{4\ln(m)}{\lambda_{\min}(\hat{\boldsymbol{V}})}$$

This gives us:

$$(1) \leq \frac{C(K) \cdot \ln^5(m)}{\sqrt{m}}.$$

Analysis of (2): Let $(\tilde{\boldsymbol{S}}_1, \tilde{\boldsymbol{S}}_2 \ldots \tilde{\boldsymbol{S}}_m)$ be independent and identically distributed as $(\boldsymbol{S}_1, \boldsymbol{S}_2 \ldots \boldsymbol{S}_m)$.

Note that,

$$\left|\mathbb{E}e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}\left(\tilde{\boldsymbol{S}}_i\right)\rangle}\right|^2 = \left|\mathbb{E}e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}(\boldsymbol{S}_i)\rangle}\right|^2 = \mathbb{E}e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\rangle}$$

Hence,

$$\left|\check{\Psi}(\boldsymbol{t})\right|^2 = \prod_{i=1}^{m} \left|\mathbb{E}e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}\left(\tilde{\boldsymbol{S}}_i\right)\rangle/\sqrt{m}}\right|^2 = \prod_{i=1}^{m} \mathbb{E}e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\rangle/\sqrt{m}}.$$

By the Taylor's theorem for CF (Theorem 8, Appendix B.8), we have,

$$\mathbb{E}\exp\left(\mathbf{i}\frac{\langle \boldsymbol{t}, \mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\rangle}{\sqrt{m}}\right) = 1 - \frac{\mathbb{E}\langle \boldsymbol{t}, \mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\rangle^2}{2m} + E_i,$$

where $|E_i|$ is controlled by:

$$|E_i| \leq \frac{\mathbb{E}|\langle \boldsymbol{t}, \mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\rangle|^3}{6m\sqrt{m}} \leq \frac{\|\boldsymbol{t}\|^3 \mathbb{E}\|\mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\|^3}{6m\sqrt{m}}.$$

Now consider any $\|\boldsymbol{t}\| \leq t_2\sqrt{m}$:

$$
\begin{aligned}
|\check{\Psi}(\boldsymbol{t})|^2 &= \prod_{i=1}^{m}\left(1 - \frac{\mathbb{E}\langle \boldsymbol{t}, \mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\rangle^2}{2m} + E_i\right) \\
&\leq \prod_{i=1}^{m}\left(1 - \frac{\mathbb{E}\langle \boldsymbol{t}, \mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\rangle^2}{2m} + \frac{\|\boldsymbol{t}\|^3\mathbb{E}\|\mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\|^3}{6m\sqrt{m}}\right) \\
&\leq \exp\left(-\sum_{i=1}^{m}\frac{\mathbb{E}\langle \boldsymbol{t}, \mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\rangle^2}{2m} + \frac{\|\boldsymbol{t}\|^3\mathbb{E}\|\mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\|^3}{6m\sqrt{m}}\right).
\end{aligned}
$$

Next we observe that,

$$
\frac{1}{2m}\sum_{i=1}^{m}\mathbb{E}\langle \boldsymbol{t}, \mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\rangle^2 = \boldsymbol{t}^{\mathsf{H}}\hat{\boldsymbol{V}}\boldsymbol{t}.
$$

We set:

$$
t_2 = 3\lambda_{\min}(\hat{\boldsymbol{V}}) \cdot \left(\frac{1}{m}\sum_{i=1}^{m}\mathbb{E}\|\mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\|^3\right)^{-1}.
$$

This ensures, for any $|t| \leq t_2\sqrt{m}$, we have,

$$
\begin{aligned}
|\check{\Psi}(\boldsymbol{t})|^2 &\leq \exp\left(-\sum_{i=1}^{m}\frac{\mathbb{E}\langle \boldsymbol{t}, \mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\rangle^2}{2m} + \frac{\|\boldsymbol{t}\|^3\mathbb{E}\|\mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\|^3}{6m\sqrt{m}}\right) \\
&\leq \exp\left(-\frac{\boldsymbol{t}^{\mathsf{H}}\hat{\boldsymbol{V}}\boldsymbol{t}}{2}\right) \leq \exp\left(-\frac{\lambda_{\min}(\hat{\boldsymbol{V}})\|\boldsymbol{t}\|^2}{2}\right).
\end{aligned}
$$

Consequently,

$$(2) = \int_{t_1 \leq \|\boldsymbol{t}\| \leq t_2\sqrt{m}} |\check{\Psi}(\boldsymbol{t})| \, \mathrm{d}\boldsymbol{t} \tag{B.10}$$

$$\leq \int_{t_1 \leq \|\boldsymbol{t}\| \leq t_2\sqrt{m}} \exp\left(-\frac{\lambda_{\min}(\hat{\boldsymbol{V}})\|\boldsymbol{t}\|^2}{4}\right) \mathrm{d}\boldsymbol{t} \tag{B.11}$$

$$\overset{\text{(a)}}{\leq} C \int_{t_1}^{\infty} \exp\left(-\frac{\lambda_{\min}(\hat{\boldsymbol{V}})}{4} l^2\right) l^3 \, \mathrm{d}l \tag{B.12}$$

$$\overset{\text{(b)}}{\leq} C(K) \cdot \left(\frac{\lambda_{\min}(\hat{\boldsymbol{V}}) t_1^2}{4} + 1\right) \cdot \exp\left(-\frac{\lambda_{\min}(\hat{\boldsymbol{V}}) t_1^2}{4}\right) \tag{B.13}$$

In the step marked (a), we converted the integral into polar coordinates from cartesian coordinates. In the step marked (b), we used Lemma 47 and used the assumption that $\lambda_{\min}(\hat{\boldsymbol{V}}) \geq K^{-1}$. Recalling that we set:

$$t_1^2 = \frac{4\ln(m)}{\lambda_{\min}(\hat{\boldsymbol{V}})},$$

we obtain,

$$(2) \leq \frac{C(K) \cdot \ln(m)}{m}.$$

Finally, to wrap up this step, we note that there exists a finite positive constant $C(K)$ such that,

$$t_2 \geq \frac{1}{C(K)}.$$

Indeed,

$$\frac{1}{m}\sum_{i=1}^{m} \mathbb{E}\|\mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i\right)\|^3 \leq \frac{C}{m}\sum_{i=1}^{m} \mathbb{E}|s_i|^3 \leq C\left(1 + |\lambda|^3 + |\phi|^3 + \frac{1}{m}\sum_{i=1}^{m}|y_i|^3\right) \leq C(K),$$

and,

$$t_2 = 3\lambda_{\min}(\hat{\boldsymbol{V}}) \cdot \left(\frac{1}{m} \sum_{i=1}^{m} \mathbb{E} \| \mathsf{Vec}\left(\boldsymbol{S}_i - \tilde{\boldsymbol{S}}_i \right) \|^3 \right)^{-1} \geq \frac{1}{C(K)}.$$

Analysis of (3): Recall the term (3) was given by:

$$(3) = \int_{\|\boldsymbol{t}\| \geq t_2 \sqrt{m}} |\check{\Psi}(\boldsymbol{t})| \, \mathrm{d}\boldsymbol{t} = m^2 \int_{\|\boldsymbol{t}\| \geq t_2} |\check{\Psi}(\boldsymbol{t}\sqrt{m})| \, \mathrm{d}\boldsymbol{t}.$$

By AM-GM for non-negative real numbers we have,

$$|\check{\Psi}(\boldsymbol{t}\sqrt{m})|^2 = \prod_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}(\boldsymbol{S}_i) \rangle} \right|^2 \leq \left(\frac{1}{m} \sum_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}(\boldsymbol{S}_i) \rangle} \right|^2 \right)^m.$$

We use two different strategies to further control the above bound:

1. Applying Lemma 44, we obtain,

$$\frac{1}{m} \sum_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}(\boldsymbol{S}_i) \rangle} \right|^2 \leq \frac{C}{\|\boldsymbol{t}\|^{\frac{2}{3}}} \cdot \frac{1}{m} \sum_{i=1}^{m} (1 + |\hat{\lambda}_2|^{20} + |\hat{\phi}|^{20} + |y|_i^{20})^2 \leq \frac{C(K)}{\|\boldsymbol{t}\|^{\frac{2}{3}}}. \tag{B.14}$$

2. The above bound tells us that for $\|\boldsymbol{t}\| \geq \sqrt{8C^3(K)}$, we have,

$$\frac{1}{m} \sum_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}(\boldsymbol{S}_i) \rangle} \right|^2 \leq \frac{1}{2}.$$

   Applying Lemma B.51 in Appendix B.8, we can find a constant $0 < \eta(K) < 1$ depending only on K such that,

$$\frac{1}{m} \sum_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}(\boldsymbol{S}_i) \rangle} \right|^2 \leq (1 - \eta(K)), \ \forall \ \|\boldsymbol{t}\| \geq t_2 \geq \frac{1}{C(K)}. \tag{B.15}$$

We can combine the above to bounds to control (3) as follows:

$$
\begin{aligned}
(3) &= m^2 \int_{\|\boldsymbol{t}\| \geq t_2} |\check{\Psi}(\boldsymbol{t}\sqrt{m})| \, \mathrm{d}\boldsymbol{t} \\
&\leq m^2 \int_{\|\boldsymbol{t}\| \geq t_2} \left(\frac{1}{m} \sum_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}(\boldsymbol{S}_i) \rangle} \right|^2 \right)^{\frac{m}{2}} \mathrm{d}\boldsymbol{t} \\
&\overset{\text{(a)}}{\leq} m^2 \cdot C(K) \cdot \int_{\|\boldsymbol{t}\| \geq t_2} \left(\frac{1}{m} \sum_{i=1}^{m} \left| \mathbb{E} e^{\mathbf{i}\langle \boldsymbol{t}, \mathsf{Vec}(\boldsymbol{S}_i) \rangle} \right|^2 \right)^{\frac{m}{2}-9} \cdot \frac{1}{\|\boldsymbol{t}\|^6} \, \mathrm{d}\boldsymbol{t} \\
&\overset{\text{(b)}}{\leq} C(K) \cdot m^2 \cdot (1 - \eta(K))^{\frac{m}{2}-9} \cdot \int_{\|\boldsymbol{t}\| \geq t_2} \frac{1}{\|\boldsymbol{t}\|^6} \, \mathrm{d}t \\
&\overset{\text{(c)}}{\leq} \frac{C(K)}{\sqrt{m}} \cdot \int_{t_2}^{\infty} \frac{1}{l^6} \cdot l^3 \, \mathrm{d}l \\
&\leq \frac{C(K)}{\sqrt{m}}
\end{aligned}
$$

In the above display, in step (a), we utilized the bound in (B.14). In the step marked (b) we utilized the bound in (B.15). In the equation marked (c) we converted the integral into polar coordinates and checked that the integral was finite.

Analysis of (4): We recall that:

$$
\begin{aligned}
(4) &= \int_{\|\boldsymbol{t}\| \geq t_1} e^{-\frac{1}{2}\boldsymbol{t}^{\mathsf{H}} \hat{\boldsymbol{V}} \boldsymbol{t}} \, \mathrm{d}t \\
&\leq \int_{\|\boldsymbol{t}\| \geq t_1} e^{-\frac{\lambda_{\min}(\hat{\boldsymbol{V}})}{2} \|\boldsymbol{t}\|^2} \, \mathrm{d}t.
\end{aligned}
$$

After this, we can exactly repeat the arguments following (B.10) and obtain,

$$
(4) \leq \frac{C(K) \ln(m)}{m}.
$$

This concludes the proof. □

B.3 Concentration Analysis

This section is devoted to proving the concentration result Proposition 9. Throughout this section, we will use Y to denote the random variable $|Z|^2 + \sigma\epsilon$, where $Z \sim \mathcal{CN}(0,1)$ and $\epsilon \sim \mathcal{N}(0,1)$. Hence, for any $f : \mathbb{R} \to \mathbb{R}$, $\mathbb{E}f(Y) = \mathbb{E}f(|Z|^2 + \sigma\epsilon)$. We also recall the $\hat{\mathbb{E}}$ notation, for any real valued function f on $\mathbb{R}$:

$$\hat{\mathbb{E}}f(Y) \stackrel{\text{def}}{=} \frac{1}{m}\sum_{i=1}^{m} f(y_i),$$

where $y_1, y_2 \ldots, y_m$ are the observations in the phase retrieval problem. The main intuition behind all of the results in this section is that the empirical measure of the measurements converges to the law of Y. Hence for a large class test functions f, $\hat{\mathbb{E}}f(Y) \approx \mathbb{E}f(Y)$. This intuition is made rigorous in terms of a general Weak Law of Large Numbers (WLLN) and a Uniform WLLN (ULLN) for the empirical measure of the measurements in Section B.3.1. We then use these general results to prove Proposition 9 in Section B.3.2.

B.3.1 A General Uniform Weak Law of Large Numbers

The following proposition establishes a weak law of large numbers (WLLN) for empirical averages of measurements $y_1, y_2 \ldots y_m$ in the phase retrieval model.

Proposition 20 (A WLLN)**.** *Let $y_1, y_2 \ldots y_m$ be the m measurements from the Phase Retrieval model. Let $f : \mathbb{R} \to \mathbb{R}$ satisfy the local Lipchitz assumption:*

$$|f(a) - f(b)| \leq L \cdot (1 + |a|^k + |b|^k) \cdot |a - b|,$$

for some $L > 0,\ k \in \mathbb{N}$. Then we have,

$$\frac{1}{m}\sum_{i=1}^{m} f(y_i) \xrightarrow{p} \mathbb{E}f(|Z|^2 + \sigma\epsilon).$$

In the above display, Z, ϵ are independent r.v.s with the distributions: $Z \sim \mathcal{CN}(0,1)$, $\epsilon \sim \mathcal{N}(0,1)$.

Proof. Recall that in the phase retrieval model, we have,

$$(y_1, y_2 \dots y_m) \stackrel{\text{d}}{=} \left(\frac{m|g_1|^2}{\|\boldsymbol{g}\|^2} + \sigma\epsilon_1, \frac{m|g_2|^2}{\|\boldsymbol{g}\|^2} + \sigma\epsilon_2 \dots \frac{m|g_m|^2}{\|\boldsymbol{g}\|^2} + \sigma\epsilon_m \right).$$

In the above display $\boldsymbol{g}$ and $\boldsymbol{\epsilon}$ are independent with $\boldsymbol{g} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_m)$ and $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_m)$. To obtain the claim of the proposition we write,

$$\begin{aligned} \frac{1}{m}\sum_{i=1}^{m} f(y_i) - \mathbb{E}f(|Z|^2 + \sigma\epsilon) &\stackrel{\text{d}}{=} \frac{1}{m}\sum_{i=1}^{m} f\left(\frac{m|g_i|^2}{\|\boldsymbol{g}\|^2} + \sigma\epsilon_i \right) - \mathbb{E}f(|Z|^2 + \sigma\epsilon) \\ &= (1) + (2). \end{aligned}$$

where the terms $(1), (2)$ are defined below:

$$\begin{aligned} (1) &\stackrel{\text{def}}{=} \left(\frac{1}{m}\sum_{i=1}^{m} f\left(\frac{m|g_i|^2}{\|\boldsymbol{g}\|^2} + \sigma\epsilon_i \right) - \frac{1}{m}\sum_{i=1}^{m} f\left(|g_i|^2 + \sigma\epsilon_i \right) \right), \\ (2) &\stackrel{\text{def}}{=} \left(\frac{1}{m}\sum_{i=1}^{m} f\left(|g_i|^2 + \sigma\epsilon_i \right) - \mathbb{E}f(|Z|^2 + \sigma\epsilon) \right). \end{aligned}$$

Note that,

$$(2) \stackrel{\text{P}}{\to} 0,$$

by WLLN for sums of i.i.d. random variables. On the other hand, by the local Lipchitz assumption

on f:

$$
\begin{aligned}
(1) &\leq L \cdot \left(\frac{m}{\|\boldsymbol{g}\|^2} - 1 \right) \cdot \frac{1}{m} \sum_{i=1}^{m} |g_i|^2 \left(1 + \frac{m^k |g_i|^{2k}}{\|\boldsymbol{g}\|^{2k}} + |g_i|^{2k} \right) \\
&= L \cdot \left(\frac{m}{\|\boldsymbol{g}\|^2} - 1 \right) \cdot \left(\frac{1}{m} \sum_{i=1}^{m} |g_i|^2 + \left(\left(\frac{m}{\|\boldsymbol{g}\|^2} \right)^k + 1 \right) \frac{1}{m} \sum_{i=1}^{m} |g_i|^{2k+2} \right) \qquad \text{(B.16)}
\end{aligned}
$$

By WLLN and continuous mapping theorem:

$$
\begin{gathered}
\frac{m}{\|\boldsymbol{g}\|^2} - 1 \xrightarrow{\text{P}} 0 \\
\frac{1}{m} \sum_{i=1}^{m} |g_i|^2 \xrightarrow{\text{P}} \mathbb{E}|Z|^2 < \infty, \\
\left(\left(\frac{m}{\|\boldsymbol{g}\|^2} \right)^k + 1 \right) \cdot \frac{1}{m} \sum_{i=1}^{m} |g_i|^{2k+2} \xrightarrow{\text{P}} \left(\frac{1}{(\mathbb{E}|Z|^2)^k} + 1 \right) \cdot \mathbb{E}|Z|^{2k+2} < \infty.
\end{gathered}
$$

Hence $(1) \xrightarrow{\text{P}} 0$. This proves the claim of the proposition. □

The following proposition proves a Uniform Law of Large Numbers (ULLN) for empirical averages of the measurements $y_1, y_2 \ldots y_m$ using some results from empirical process theory [77].

Proposition 21 (A Uniform Law of Large Numbers)**.** *Let $\mathcal{F}_T$ be a collection of functions $f_{\boldsymbol{t}} : \mathbb{R} \to \mathbb{R}$ indexed by a parameter $\boldsymbol{t}$ which takes values in the set T, a bounded subset of $\mathbb{R}^k$. Suppose that the collection $\mathcal{F}_T$ satisfies the following Lipchitz conditions:*

$$
\begin{aligned}
&\textit{Lipchitz in parameter: } |f_{\boldsymbol{t}}(y) - f_{\boldsymbol{s}}(y)| \leq L \cdot \|\boldsymbol{t} - \boldsymbol{s}\| \cdot (1 + |y|^l) \; \forall \; \boldsymbol{t}, \boldsymbol{s} \in T, \; y \in \mathbb{R}, \\
&\textit{Lipchitz in argument: } |f_{\boldsymbol{t}}(y) - f_{\boldsymbol{t}}(y')| \leq L \cdot |y - y'| \cdot (|y|^l + |y'|^l + 1) \; \forall \; \boldsymbol{t} \in T, \; y, y' \in \mathbb{R}.
\end{aligned}
$$

for some $L > 0, l \in \mathbb{N}$. Then we have,

$$
\sup_{\boldsymbol{t} \in T} \left(\frac{1}{m} \sum_{i=1}^{m} f_{\boldsymbol{t}}(y_i) - \mathbb{E} f_{\boldsymbol{t}}(|Z|^2 + \sigma\epsilon) \right) \xrightarrow{P} 0.
$$

Proof. As in the proof of Proposition 20, we have the decomposition:

$$\frac{1}{m}\sum_{i=1}^{m} f(y_i) - \mathbb{E}f(|Z|^2 + \sigma\epsilon) = (1) + (2).$$

where,

$$(1) \stackrel{\text{def}}{=} \left(\frac{1}{m}\sum_{i=1}^{m} f_{\boldsymbol{t}}\left(\frac{m|g_i|^2}{\|\boldsymbol{g}\|^2} + \sigma\epsilon_i \right) - \frac{1}{m}\sum_{i=1}^{m} f_{\boldsymbol{t}}\left(|g_i|^2 + \sigma\epsilon_i \right) \right),$$
$$(2) \stackrel{\text{def}}{=} \left(\frac{1}{m}\sum_{i=1}^{m} f_{\boldsymbol{t}}\left(|g_i|^2 + \sigma\epsilon_i \right) - \mathbb{E}f_{\boldsymbol{t}}(|Z|^2 + \sigma\epsilon) \right).$$

The analysis (1) is exactly the same as in Proposition 20. The upper bound in (B.16) holds uniformly over T and hence,

$$\sup_{\boldsymbol{t}\in T} (1) \xrightarrow{\text{P}} 0.$$

For the term (2), we appeal to standard empirical process theory results from Van Der Vaart and Wellner [77]. By Theorem 2.7.11 of Van Der Vaart and Wellner [77], the function class $\mathcal{F}_T$ has bounded bracketting number. Consequently, by Theorem 2.4.1 of Van Der Vaart and Wellner [77], $\mathcal{F}_T$ is Glivenko-Cantelli, that is,

$$\sup_{\boldsymbol{t}\in T} (2) \xrightarrow{\text{P}} 0.$$

This concludes the proof of the proposition. □

Next we will apply the ULLN of Proposition 21 to obtain uniform convergence of empirical averages of the log-normalizing constants and moments of the Tilted Exponential and Wishart

distributions. In particular, we recall the definitions:

$$\ln Z_{\mathsf{TExp}}(\lambda, y) \stackrel{\text{def}}{=} \ln \mathbb{E}_{E\sim\mathsf{Exp}(1)} e^{\lambda E}\psi_\sigma(y-E),$$
$$\ln Z_{\mathsf{TWis}}(\lambda, \phi, y) \stackrel{\text{def}}{=} \ln \mathbb{E}_{\boldsymbol{g}\sim\mathcal{CN}(\mathbf{0},\boldsymbol{I}_2)} e^{\lambda(|g_1|^2+|g_2|^2)+\phi\mathrm{Re}(g_1\overline{g}_2)}\psi_\sigma(y-|g_1|^2)\psi_\sigma(y-|g_2|^2).$$

For any $a, b, c, d \in \mathbb{N}$ we also define the moments of the tilted exponential and wishart distributions:

$$\mu^{(a)}_{\mathsf{TExp}}(\lambda, y) \stackrel{\text{def}}{=} \mathbb{E}T^j,\ T\sim\mathsf{TExp}(\lambda, y)$$
$$\mu^{(a,b,c,d)}_{\mathsf{TExp}}(\lambda, \phi, y) \stackrel{\text{def}}{=} \mathbb{E}S_{11}^a\mathrm{Re}(S_{12})^b\mathrm{Im}(S_{12})^c S_{22}^d,\ \boldsymbol{S}\sim\mathsf{TWis}(\lambda, \phi, y).$$

Recalling the Definitions 5 and 6, we have,

$$\mu^{(a)}_{\mathsf{TExp}}(\lambda, y) = \frac{\mathbb{E}E^a e^{\lambda E}\psi_\sigma(y-E)}{\mathbb{E}_{E\sim\mathsf{Exp}(1)} e^{\lambda E}\psi_\sigma(y-E)},$$
$$\mu^{(a,b,c,d)}_{\mathsf{TExp}}(\lambda, \phi, y) = \frac{\mathbb{E}|g_1|^{2a}\mathrm{Re}(g_1\overline{g}_2)^b\mathrm{Im}(g_1\overline{g}_2)^c|g_2|^{2d}e^{\lambda(|g_1|^2+|g_2|^2)+\phi\mathrm{Re}(g_1\overline{g}_2)}\psi_\sigma(y-|g_1|^2)\psi_\sigma(y-|g_2|^2)}{\mathbb{E}e^{\lambda(|g_1|^2+|g_2|^2)+\phi\mathrm{Re}(g_1\overline{g}_2)}\psi_\sigma(y-|g_1|^2)\psi_\sigma(y-|g_2|^2)}.$$

In the above display $E\sim\mathsf{Exp}(1)$, $\boldsymbol{g}\sim\mathcal{CN}(\mathbf{0}, \boldsymbol{I}_2)$. The following corollary applies the obtained ULLN to the above functions to obtain uniform convergence for these functions.

Corollary 5 (Uniform Convergence of Log-Normalizing Constants and Moments)**.** *For any $R > 0$ and $a, b, c, d \in \mathbb{N}$, we have,*

$$1)\ \sup_{|\lambda|\le R}\left(\frac{1}{m}\sum_{i=1}^m \ln Z_{\mathsf{TExp}}(\lambda, y_i) - \mathbb{E}_{Z,\epsilon}\ln Z_{\mathsf{TExp}}\left(\lambda, |Z|^2+\sigma\epsilon\right)\right) \xrightarrow{P} 0,$$
$$2)\ \sup_{|\lambda|+|\phi|\le R}\left(\frac{1}{m}\sum_{i=1}^m \ln Z_{\mathsf{TWis}}(\lambda, \phi, y_i) - \mathbb{E}_{Z,\epsilon}\ln Z_{\mathsf{TWis}}\left(\lambda, \phi, |Z|^2+\sigma\epsilon\right)\right) \xrightarrow{P} 0,$$
$$3)\ \sup_{|\lambda|\le R}\left(\frac{1}{m}\sum_{i=1}^m \mu^{(a)}_{\mathsf{TExp}}(\lambda, y_i) - \mathbb{E}_{Z,\epsilon}\mu^{(a)}_{\mathsf{TExp}}\left(\lambda, |Z|^2+\sigma\epsilon\right)\right) \xrightarrow{P} 0,$$
$$4)\ \sup_{|\lambda|+|\phi|\le R}\left(\frac{1}{m}\sum_{i=1}^m \mu^{(a,b,c,d)}_{\mathsf{TExp}}(\lambda, \phi, y_i) - \mathbb{E}_{Z,\epsilon}\mu^{((a,b,c,d))}_{\mathsf{TExp}}\left(\lambda, \phi, |Z|^2+\sigma\epsilon\right)\right) \xrightarrow{P} 0.$$

Proof. In order to prove the corollary, we just need to verify the Lipchitz conditions in Proposition 21. In order to do so, we observe that,

$$\frac{\partial}{\partial y}\ln Z_{\mathsf{TExp}}(\lambda,\phi)=\frac{\mu^{(1)}_{\mathsf{TExp}}(\lambda,y)-y}{\sigma^2},\ \frac{\partial}{\partial \lambda}\ln Z_{\mathsf{TExp}}(\lambda,\phi)=\mu^{(1)}_{\mathsf{TExp}}(\lambda,y).$$

The moments of the Tilted Exponential distribution are bounded in Lemma 43. Using this we obtain,

$$\max_{|\lambda|\le R}\left|\frac{\partial}{\partial y}\ln Z_{\mathsf{TExp}}(\lambda,y)\right|\le C(R+|y|),\ \max_{|\lambda|\le R}\left|\frac{\partial}{\partial \lambda}\ln Z_{\mathsf{TExp}}(\lambda,y)\right|\le C(R+|y|).$$

Integrating these derivative bounds gives us the following Lipchitz estimates:

$$\left|\ln Z_{\mathsf{TExp}}(\lambda,y)-\ln Z_{\mathsf{TExp}}\left(\lambda,y'\right)\right|\le C\cdot(R+|y|+|y'|)\cdot|y-y'|\ \forall\ |\lambda|\le R,\ y,y'\in\mathbb{R},$$
$$\left|\ln Z_{\mathsf{TExp}}(\lambda,y)-\ln Z_{\mathsf{TExp}}\left(\lambda',y\right)\right|\le C\cdot(R+|y|)\cdot|\lambda-\lambda'|\ \forall\ |\lambda|\le R,\ |\lambda'|\le R,\ y\in\mathbb{R},$$

which verifies the assumptions of Proposition 21 and hence (1) follows. Likewise the uniform convergence in (3) follows from the observation:

$$\frac{\partial}{\partial y}\mu^{(a)}_{\mathsf{TExp}}(\lambda,y)=\frac{\mu^{(a+1)}_{\mathsf{TExp}}(\lambda,y)-\mu^{(a)}_{\mathsf{TExp}}(\lambda,y)\,\mu^{(1)}_{\mathsf{TExp}}(\lambda,y)}{\sigma^2},$$
$$\frac{\partial}{\partial \lambda}\mu^{(a)}_{\mathsf{TExp}}(\lambda,y)=\mu^{(a+1)}_{\mathsf{TExp}}(\lambda,y)-\mu^{(a)}_{\mathsf{TExp}}(\lambda,y)\,\mu^{(1)}_{\mathsf{TExp}}(\lambda,y).$$

The proofs of (2) and (4) are analogous and rely on moment bounds for the tilted wishart distribution given in Lemma 44. □

B.3.2 Proof of Proposition 9

We now present the proof of Proposition 9.

Proof. Since polynomial functions are locally Lipchitz, the claim (1) follows from the WLLN proved in Proposition 20. Item (2) is a special case of Corollary 5. The proofs of items (3-4) is

very similar to (and easier) items (5-6) and is omitted. Hence we focus on proving claims 5-8. Define the concave (in λ, ϕ) potential functions:

$$V_2(\lambda, \phi; q) \stackrel{\text{def}}{=} 2\lambda + \phi q - \mathbb{E}_Y \ln Z_{\mathsf{TWis}}(\lambda, \phi, Y),$$
$$\hat{V}_2(\lambda, \phi; q) \stackrel{\text{def}}{=} 2\lambda + \phi q - \hat{\mathbb{E}}_Y \ln Z_{\mathsf{TWis}}(\lambda, \phi, Y).$$

The potential functions are important because:

$$(\lambda_2(q;\sigma), \phi(q;\sigma)) = \arg\max_{\lambda,\phi\in\mathbb{R}} V_2(\lambda, \phi; q),\ \Xi_2(q;\sigma) = \max_{\lambda,\phi\in\mathbb{R}} V_2(\lambda, \phi; q).$$

And likewise,

$$(\hat{\lambda}_2(q;\sigma), \hat{\phi}(q;\sigma)) = \arg\max_{\lambda,\phi\in\mathbb{R}} \hat{V}_2(\lambda, \phi; q),\ \hat{\Xi}_2(q;\sigma) = \max_{\lambda,\phi\in\mathbb{R}} \hat{V}_2(\lambda, \phi; q).$$

The proof of this proposition relies on coercivity estimates for the above potential functions which have been proved in Appendix B.7.

For the ease of notation, in this proof we will short hand $\Xi_2(q;\sigma)$, $\hat{\Xi}_2(q;\sigma)$, $\lambda_2(q;\sigma)$, $\hat{\lambda}_2(q;\sigma)$, $\phi(q;\sigma)$ and $\hat{\phi}(q;\sigma)$ as $\Xi_2(q)$, $\hat{\Xi}_2(q)$, $\lambda_2(q)$, $\hat{\lambda}_2(q)$, $\phi(q)$ and $\hat{\phi}(q)$, omitting the dependence on σ. We consider each of the claims (5-8) one by one:

5. In Proposition 23 (Appendix B.7), we have shown that the solutions to the variation problems lie in the compact intervals:

$$|\lambda_2(q)| + |\phi(q)| \le C\left(1 + q + \frac{1}{1-q}\right) \cdot (\mathbb{E}|Y|^2 + 1),$$
$$|\hat{\lambda}_2(q)| + |\hat{\phi}(q)| \le C\left(1 + q + \frac{1}{1-q}\right) \cdot (\hat{\mathbb{E}}|Y|^2 + 1)$$

   On the other hand we know from Proposition 20 that,

$$\hat{\mathbb{E}}Y^2 \stackrel{\text{P}}{\to} \mathbb{E}Y^2 < \infty.$$

Consequently, we can find constant R that depends only on η, σ such that,

$$\max_{0\le q\le 1-\eta} |\lambda_2(q)| + |\phi(q)| \le R,\ \mathbb{P}\left(\max_{0\le q\le 1-\eta} |\hat{\lambda}_2(q)| + |\hat{\phi}(q)| > R\right) \to 0.$$

For instance taking R as:

$$R = C\left(2 + \frac{1}{1-\eta}\right)(2 + \mathbb{E}Y^2),$$

is sufficient. This proves item (5) of the proposition.

6. We upper bound $\Xi_2(q) - \hat{\Xi}_2(q)$ and $\hat{\Xi}_2(q) - \Xi_2(q)$ separately:

$$\begin{aligned}
&\Xi_2(q) - \hat{\Xi}_2(q) = V_2(\lambda_2(q), \phi(q); q) - \hat{V}(\hat{\lambda}_2(q), \hat{\phi}(q); q) \\
&= V_2(\lambda_2(q), \phi(q); q) - \hat{V}_2(\lambda_2(q), \phi(q); q) + \underbrace{\hat{V}_2(\lambda_2(q), \phi(q); q) - \hat{V}_2(\hat{\lambda}_2(q), \hat{\phi}(q); q)}_{\le 0} \\
&\le V_2(\lambda_2(q), \phi(q); q) - \hat{V}_2(\lambda_2(q), \phi(q); q) \\
&\le \sup_{q\in[0,1-\eta], |\lambda|+|\phi|\le R} |V_2(\lambda, \phi; q) - \hat{V}_2(\lambda, \phi; q)|.
\end{aligned}$$

Analogously, we can obtain $\hat{\Xi}_2(q) - \Xi_2(q) \le \sup_{q\in[0,1-\eta], |\lambda|+|\phi|\le R} |V_2(\lambda, \phi; q) - \hat{V}_2(\lambda, \phi; q)|$. Consequently we have,

$$\begin{aligned}
\sup_{q\in[0,1-\eta]} |\Xi_2(q) - \hat{\Xi}_2(q)| &\le \sup_{q\in[0,1-\eta], \lambda, \phi\in\mathbb{R}} |V_2(\lambda, \phi; q) - \hat{V}_2(\lambda, \phi; q)| \\
&= \sup_{\lambda, \phi: |\lambda|+|\phi|\le R} \left|\mathbb{E}_Y \ln Z_{\mathsf{TWis}}(\lambda, \phi, Y) - \hat{\mathbb{E}}_Y \ln Z_{\mathsf{TWis}}(\lambda, \phi, Y)\right| \\
&\xrightarrow{\text{P}} 0.
\end{aligned}$$

In the last step we appealed to Corollary 5. This concludes the proof of item (6).

7. For the purpose of demonstrating convergence in probability it is sufficient to restrict our-

selves to the event:

$$\max_{0\le q\le 1-\eta} |\hat{\lambda}_2(q)| + |\hat{\phi}(q)| \le R,$$

since this event occurs with probability tending to 1. Proposition 23 shows that the function $\mathbb{E}_Y \ln Z_{\mathsf{TWis}}(\lambda, \phi, Y)$ is strongly convex on compact intervals. Hence for some universal constant $C < \infty$, we have, for any $\lambda, \phi: \ |\lambda| + |\phi| \le R, \ \forall q \in [0, 1-\eta]$,

$$V_2(\lambda, \phi; q) \le V_2(\lambda_2(q), \phi(q); q) - \frac{1}{C} \cdot (|\lambda - \lambda_2(q)|^2 + |\phi - \phi(q)|^2).$$

Applying the strong convexity estimate to $\lambda = \hat{\lambda}_2(q), \ \phi = \hat{\phi}(q)$ gives us:

$$\begin{aligned} |\hat{\lambda}_2(q) - \lambda_2(q)|^2 + |\hat{\phi}(q) - \phi(q)|^2 &\le C(V_2(\lambda_2(q), \phi(q); q) - V_2(\hat{\lambda}_2(q), \hat{\phi}(q); q)) \\ &= C \cdot \big((1) + (2) + (3)\big). \end{aligned}$$

In the above display, we defined the terms (1), (2) and (3) as:

$$\begin{aligned} (1) &= V_2(\lambda_2(q), \phi(q); q) - \hat{V}_2(\lambda_2(q), \phi(q); q), \\ (2) &= \hat{V}_2(\lambda_2(q), \phi(q); q) - \hat{V}_2(\hat{\lambda}_2(q), \hat{\phi}(q); q), \\ (3) &= \hat{V}_2(\hat{\lambda}_2(q), \hat{\phi}(q); q) - V_2(\hat{\lambda}_2(q), \hat{\phi}(q); q). \end{aligned}$$

Since $(\hat{\lambda}_2(q), \hat{\phi}(q))$ maximizes $\hat{V}_2(\lambda, \phi; q)$, we have,

$$(2) \le 0.$$

On the other hand, both (1) and (2) can be bounded by:

$$(1) \leq \sup_{\lambda,\phi:|\lambda|+|\phi|\leq R, q\in[0,1-\eta]} \left| V_2(\lambda,\phi;q) - \hat{V}_2(\lambda,\phi;q) \right|,$$
$$(2) \leq \sup_{\lambda,\phi:|\lambda|+|\phi|\leq R, q\in[0,1-\eta]} \left| V_2(\lambda,\phi;q) - \hat{V}_2(\lambda,\phi;q) \right|.$$

Hence we have obtained,

$$|\hat{\lambda}_2(q) - \lambda_2(q)|^2 + |\hat{\phi}(q) - \phi(q)|^2 \leq 2C \cdot \sup_{\lambda,\phi:|\lambda|+|\phi|\leq R, q\in[0,1-\eta]} \left| V_2(\lambda,\phi;q) - \hat{V}_2(\lambda,\phi;q) \right|.$$

Corollary 5 gives us the uniform convergence:

$$\sup_{\lambda,\phi:|\lambda|+|\phi|\leq R, q\in[0,1-\eta]} \left| V_2(\lambda,\phi;q) - \hat{V}_2(\lambda,\phi;q) \right| =$$
$$\sup_{\lambda,\phi:|\lambda|+|\phi|\leq R} \left| \mathbb{E}_Y \ln Z_{\mathsf{TWis}}(\lambda,\phi,Y) - \hat{\mathbb{E}}_Y \ln Z_{\mathsf{TWis}}(\lambda,\phi,Y) \right| \xrightarrow{\text{P}} 0.$$

Hence we obtain,

$$\sup_{q\in[0,1-\eta]} |\hat{\lambda}_2(q) - \lambda_2(q)|^2 + |\hat{\phi}(q) - \phi(q)|^2 \xrightarrow{\text{P}} 0.$$

This shows claim (7) of the proposition.

8. A simple computation shows that:

$$\frac{\mathrm{d}^2\Xi_2(q)}{\mathrm{d}q^2} - \frac{\mathrm{d}^2\hat{\Xi}_2(q)}{\mathrm{d}q^2} = \boldsymbol{e}_2^{\mathsf{H}} \left(\nabla^2_{\lambda,\phi} V_2(\lambda_2(q),\phi(q);q)^{-1} - \nabla^2_{\lambda,\phi} \hat{V}_2(\hat{\lambda}_2(q),\hat{\phi}(q);q)^{-1} \right) \boldsymbol{e}_2.$$

Hence,

$$\sup_{q\in[0,1-\eta]} \left| \frac{\mathrm{d}^2\Xi_2(q)}{\mathrm{d}q^2} - \frac{\mathrm{d}^2\hat{\Xi}_2(q)}{\mathrm{d}q^2} \right| \leq \sup_{q\in[0,1-\eta]} \left\| \nabla^2_{\lambda,\phi} V_2(\lambda_2(q),\phi(q);q)^{-1} - \nabla^2_{\lambda,\phi} \hat{V}_2(\hat{\lambda}_2(q),\hat{\phi}(q);q)^{-1} \right\|$$

Hence it is sufficient to show that,

$$\sup_{q\in[0,1-\eta]} \left\| \nabla^2_{\lambda,\phi} V_2(\lambda_2(q),\phi(q);q)^{-1} - \nabla^2_{\lambda,\phi}\hat{V}_2(\hat{\lambda}_2(q),\hat{\phi}(q);q)^{-1} \right\| \xrightarrow{\text{P}} 0.$$

By triangle inequality, we can write,

$$\sup_{q\in[0,1-\eta]} \left\| \nabla^2 \hat{V}_2(\hat{\lambda}_2(q),\hat{\phi}(q);q) - \nabla^2\, V_2(\lambda_2(q),\phi(q);q) \right\| \leq (1)+(2),$$

where we define the terms (1) and (2) as:

$$\begin{aligned}
(1) &\stackrel{\text{def}}{=} \sup_{q\in[0,1-\eta]} \left\| \nabla^2 \hat{V}_2(\hat{\lambda}_2(q),\hat{\phi}(q);q) - \nabla^2\, V_2(\hat{\lambda}_2(q),\hat{\phi}(q);q) \right\|, \\
(2) &\stackrel{\text{def}}{=} \sup_{q\in[0,1-\eta]} \left\| \nabla^2 V_2(\hat{\lambda}_2(q),\hat{\phi}(q);q) - \nabla^2\, V_2(\lambda_2(q),\phi(q);q) \right\|.
\end{aligned}$$

We control the first term as follows:

$$\begin{aligned}
(1) &\leq \sup_{q\in[0,1-\eta],\lambda,\phi:|\lambda|+|\phi|\leq R} \left\| \nabla^2 \hat{V}_2(\lambda,\phi;q) - \nabla^2\, V_2(\lambda,\phi;q) \right\| \\
&= \sup_{\lambda,\phi:|\lambda|+|\phi|\leq R} \| \nabla^2 \mathbb{E}_Y \ln Z_{\mathsf{TWis}}(\lambda,\phi,Y) - \nabla^2 \hat{\mathbb{E}}_Y \ln Z_{\mathsf{TWis}}(\lambda,\phi,Y) \|
\end{aligned}$$

Noting that the entries of matrix $\nabla^2_{\lambda,\phi} \ln Z_{\mathsf{TWis}}(\lambda,\phi,Y)$ are moments of the Tilted Wishart distribution and appealing to Corollary 5 gives us the uniform convergence:

$$(1) \leq \sup_{q\in[0,1-\eta],\lambda,\phi:|\lambda|+|\phi|\leq R} \left\| \nabla^2 \hat{V}_2(\lambda,\phi;q) - \nabla^2\, V_2(\lambda,\phi;q) \right\| \xrightarrow{\text{P}} 0. \tag{B.17}$$

To control the second term, we first note that $\nabla^2 V_2(\lambda,\phi;q)$ is independent of q. It is also easy to check that it is locally Lipchitz of λ,ϕ, consequently we have the estimate,

$$\left\| \nabla^2 V_2(\hat{\lambda}_2(q),\hat{\phi}(q);q) - \nabla^2\, V_2(\lambda_2(q),\phi(q);q) \right\| \leq C\left(|\lambda_2(q)-\hat{\lambda}_2(q)| + |\phi(q)-\hat{\phi}(q)| \right),$$

for some constant C depending only on R (in particular, C does not depend on q). Combining this with the conclusion obtained in item (4) of the lemma gives us:

$$(2) \xrightarrow{\text{P}} 0.$$

Hence we have,

$$\sup_{q\in[0,1-\eta]} \left\| \nabla^2 \hat{V}_2(\hat{\lambda}_2(q), \hat{\phi}(q); q) - \nabla^2 V_2(\lambda_2(q), \phi(q); q) \right\| \xrightarrow{\text{P}} 0. \tag{B.18}$$

In order to obtain the analogous result for the inverse-hessian, we note that by Proposition 23, $V_2(\lambda, \phi; q)$ is strongly concave on compact sets. Furthermore, $\nabla^2 V_2(\lambda, \phi; q)$ does not depend on q. Hence we have,

$$\lambda_{\max}(\nabla^2 V_2(\lambda, \phi; q)) \leq -\frac{1}{C}, \ \forall |\lambda| + |\phi| \leq R, \forall q,$$

for a large enough universal constant C. Recalling the uniform convergence in (B.17), we have,

$$\mathbb{P}\left(\max_{\lambda,\phi:|\lambda|+|\phi|\leq R} \lambda_{\max}(\nabla^2 \hat{V}_2(\lambda, \phi; q)) \leq -\frac{1}{2C} \right) \to 1.$$

Since both $V, \hat{V}$ are concave functions (c.f. Proposition 23), we have,

$$\sup_{q\in[0,1-\eta]} \|\nabla^2 V_2(\lambda_2(q), \phi(q); q)^{-1}\|_{op} = O(1), \quad \sup_{q\in[0,1-\eta]} \|\nabla^2 \hat{V}_2(\hat{\lambda}_2(q), \hat{\phi}(q); q)^{-1}\|_{op} = O_P(1). \tag{B.19}$$

Wedin [87] has shown the following perturbation bounds for matrix inverse for any two invertible matrices A, B:

$$\|A^{-1} - B^{-1}\| \leq \sqrt{2} \cdot \max(\|A^{-1}\|_{op}, \|B^{-1}\|_{op}) \cdot \|A - B\|.$$

Combining the tightness result in (B.19) and the uniform convergence of hessians (see (B.17)) gives us,

$$\sup_{q\in[0,1-\eta]} \|\nabla^2 V_2(\lambda_2(q),\phi(q);q)^{-1} - \nabla^2 \hat{V}_2(\hat{\lambda}_2(q),\hat{\phi}(q);q)^{-1}\| \xrightarrow{\text{P}} 0.$$

This concludes the proof of item (8).

□

B.4 Proof of Proposition 10

Recall that we had introduced the following functions:

$$\mathcal{F}(q;\delta,\Delta,\sigma) = \Xi_2(q;\sigma) - 2\Xi_1(\sigma) + \left(1-\frac{1}{\delta}\right)\ln(1-q^2) + \Delta\ln\left(1-\frac{q^2}{2}\right)$$
$$\hat{\mathcal{F}}(q;\delta,\Delta,\sigma) = \hat{\Xi}_2(q;\sigma) - 2\hat{\Xi}_1(\sigma) + \left(1-\frac{1}{\delta}\right)\ln(1-q^2) + \Delta\ln\left(1-\frac{q^2}{2}\right),$$

where,

$$\Xi_2(q;\sigma) \stackrel{\text{def}}{=} \max_{(\lambda,\phi)\in\mathbb{R}} \left(2\lambda + \phi q - \mathbb{E}_Y \ln Z_{\mathsf{TWis}}(\lambda,\phi,Y)\right),$$
$$\hat{\Xi}_2(q;\sigma) \stackrel{\text{def}}{=} \max_{(\lambda,\phi)\in\mathbb{R}} \left(2\lambda + \phi q - \hat{\mathbb{E}}_Y \ln Z_{\mathsf{TWis}}(\lambda,\phi,Y)\right),$$
$$\Xi_1(\sigma) \stackrel{\text{def}}{=} \max_{\lambda\in\mathbb{R}} \left(\lambda - \mathbb{E}_Y \ln \mathbb{E}_{E\sim\mathsf{Exp}(1)} e^{\lambda E}\psi_\sigma(E-Y)\right),$$
$$\hat{\Xi}_1(\sigma) \stackrel{\text{def}}{=} \max_{\lambda\in\mathbb{R}} \left(\lambda - \mathbb{E}_Y \ln \hat{\mathbb{E}}_{E\sim\mathsf{Exp}(1)} e^{\lambda E}\psi_\sigma(E-Y)\right).$$

Consider any δ that satisfies the assumptions of Proposition 10:

$$\mathcal{F}(0;\delta,\Delta,\sigma) < \mathcal{F}(q;\delta,\Delta,\sigma)\ \forall\ q\ \in\ (0,1), \tag{B.20}$$

and,

$$\frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(0;\delta,\Delta,\sigma) > 0. \tag{B.21}$$

In Lemmas 12 and 13, we showed that,

$$\mathbf{I}\left(\boldsymbol{y},\boldsymbol{z};\boldsymbol{A},\boldsymbol{W}\right) \leq \frac{2}{n-1}\mathbb{E}_{\boldsymbol{y}}\left[\frac{\int_0^1 \mathscr{U}\left(\boldsymbol{y},\begin{bmatrix}1 & q\\ q & 1\end{bmatrix}\right)\cdot\frac{q(1-q^2)^{n-2}}{(1-q^2/2)^{\Delta m}}\,\mathrm{d}q}{\mathscr{L}^2(\boldsymbol{y},1)}\cdot\mathbf{1}_{\mathcal{E}_m}\right] + C\cdot m\cdot\sqrt{\mathbb{P}(\mathcal{E}_m^c)}. \tag{B.22}$$

We will set $\mathcal{E}_m$ as:

$$\mathcal{E}_m = \mathcal{E}_m^{(1)}(L)\cap\mathcal{E}_m^{(2)}(R,\eta)\cap\mathcal{E}_m^{(3)}(R,\eta)\cap\mathcal{E}_m^{(4)}(\eta)\cap\mathcal{E}_m^{(5)}(\eta,\epsilon_2)\cap\mathcal{E}_m^{(6)}(R,\epsilon_2) \tag{B.23}$$

where:

$$\mathcal{E}_m^{(1)}(L) = \left\{\boldsymbol{y} : 1+\hat{\mathbb{E}}Y^{40}\leq L\right\}, \tag{B.24}$$

$$\mathcal{E}_m^{(2)}(R,\eta) = \left\{\boldsymbol{y} : \sup_{|\lambda|\leq R}\left|\hat{\mathbb{E}}\sigma^2_{\mathsf{TExp}}(\lambda,Y) - \mathbb{E}\sigma^2_{\mathsf{TExp}}(\lambda,Y)\right|\leq\eta\right\}, \tag{B.25}$$

$$\mathcal{E}_m^{(3)}(R,\eta) = \left\{\boldsymbol{y} : \sup_{|\lambda|+|\phi|\leq R}\left\|\hat{\mathbb{E}}\boldsymbol{\Sigma}_{\mathsf{TWis}}(\lambda,\phi,Y) - \mathbb{E}\boldsymbol{\Sigma}_{\mathsf{TWis}}(\lambda,\phi,Y)\right\|\leq\eta\right\}, \tag{B.26}$$

$$\mathcal{E}_m^{(4)}(\eta) = \left\{\boldsymbol{y} : \sup_{q\leq 1/2}\left|\frac{\mathrm{d}^2}{\mathrm{d}q^2}\mathcal{F}(q;\delta,\Delta,\sigma) - \frac{\mathrm{d}^2}{\mathrm{d}q^2}\hat{\mathcal{F}}(q;\delta,\Delta,\sigma)\right|\leq\eta\right\}, \tag{B.27}$$

$$\mathcal{E}_m^{(5)}(\eta,\epsilon_2) = \left\{\boldsymbol{y} : |\Xi_1(\sigma)-\hat{\Xi}_1(\sigma)|\leq\eta, \sup_{q\in[0,1-\epsilon_2]}|\Xi_2(q;\sigma)-\hat{\Xi}_2(q;\sigma)|\leq\eta\right\}, \tag{B.28}$$

$$\mathcal{E}_m^{(6)}(R,\epsilon_2) = \left\{\boldsymbol{y} : |\hat{\lambda}_1(\sigma)|\leq R, \sup_{q\in[0,1-\epsilon_2]}|\hat{\lambda}_2(q;\sigma)|+|\hat{\phi}(q;\sigma)|\leq R\right\}. \tag{B.29}$$

In the above display L, R, η, ϵ_2 are parameters which will be set appropriately later. Recall that the notation $\hat{\mathbb{E}}$ is used to denote empirical averages:

$$\hat{\mathbb{E}} f(Y) = \frac{1}{m} \sum_{i=1}^{m} f(y_i),$$

and the notation $\mathbb{E} f(Y) = \mathbb{E}_{Z,\epsilon} f(|Z|^2 + \sigma\epsilon)$ where $Z \sim \mathcal{CN}(0,1)$, $\epsilon \sim \mathcal{N}(0,1)$. Recall the upper bound in (B.22). Our goal in this section is to show $\mathbf{I}(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{A}, \boldsymbol{W}) = o(m)$. Towards this goal, the remainder of this section is organized as follows:

1. In Lemma 35 we show that $\mathbb{P}(\mathcal{E}_m^c) = o(1)$.
2. In Lemmas 36 and 37 we show that under the event $\mathcal{E}_m$, the assumptions of Corollary 2 and 3 are met, and hence we can use them to obtain an upper bound on $\mathscr{U}$ and a lower bound on $\mathscr{L}$.
3. Finally the proof of Proposition 10 is restated and proved.

Lemma 35 (Analysis of $\mathbb{P}(\mathcal{E}_m)$)**.** *For any $\epsilon_2 \in (0,1)$, there exists a critical value $R_c(\epsilon_2)$ such that, for any $L > 1 + \mathbb{E}Y^{40}$, any $R > R_c(\epsilon_2)$ and any $\eta > 0$, we have, for the event,*

$$\mathcal{E}_m = \mathcal{E}_m^{(1)}(L) \cap \mathcal{E}_m^{(2)}(R,\eta) \cap \mathcal{E}_m^{(3)}(R,\eta) \cap \mathcal{E}_m^{(4)}(\eta) \cap \mathcal{E}_m^{(5)}(\eta,\epsilon_2) \cap \mathcal{E}_m^{(6)}(R,\epsilon_2),$$
$$\mathbb{P}(\mathcal{E}_m) \to 1.$$

Proof. This lemma is essentially a consequence of the concentration analysis in Proposition 9. By claim (1) of Proposition 20 we know that,

$$\hat{\mathbb{E}} Y^{40} \xrightarrow{\text{P}} \mathbb{E}Y^{40} < \infty.$$

Consequently any $L > \mathbb{E}Y^{40}$ we have,

$$\mathbb{P}(\mathcal{E}_m^{(1)}(L)) \to 1.$$

For any $\epsilon_2 > 0$. Claims (3) and (5) of Proposition 9 guarantee the existence of $R_c(\epsilon_2)$ such that,

$$\mathbb{P}(\mathcal{E}_m^{(6)}(R, \epsilon_2)) \to 0, \ \forall\ R > R_c(\epsilon_2), \ \forall\ \epsilon_2 > 0.$$

Claim (2) of Proposition 9 gives for any $R \in (0, \infty), \ \eta > 0$,

$$\mathbb{P}(\mathcal{E}_m^{(2)}(R, \eta)) \to 1, \ \mathcal{E}_m^{(3)}(R, \eta) \to 1.$$

Like wise Claim (4) and (6) 9 guarantee for any $\epsilon_2 \in (0, 1)$ and in $\eta > 0$, we have, $\mathbb{P}(\mathcal{E}_m^{(5)}(\eta, \epsilon_2)) \to 1$. Finally we observe that:

$$\left| \frac{\mathrm{d}^2}{\mathrm{d}q^2} \mathcal{F}(q; \delta, \Delta, \sigma) - \frac{\mathrm{d}^2}{\mathrm{d}q^2} \hat{\mathcal{F}}(q; \delta, \Delta, \sigma) \right| = \left| \frac{\mathrm{d}^2}{\mathrm{d}q^2} \Xi_2(q; \sigma) - \frac{\mathrm{d}^2}{\mathrm{d}q^2} \Xi_2(q; \sigma) \right|,$$

Hence Claim (8) of Proposition 9 shows that for any $\eta > 0$, we have,$\mathbb{P}(\mathcal{E}_n^{(4)}(\eta)) \to 1$. Finally a union bound gives us the claim $\mathbb{P}(\mathcal{E}_m) \to 1$. □

Lemma 36 (A Lower Bound on $\mathscr{L}$)**.** *For any $R, L \in (0, \infty)$, there exists a critical value of η denoted by $\eta_1(R)$ depending only on R such that for any $\eta < \eta_1(R), \ \epsilon_2 > 0$ on the event $\mathcal{E}_m^{(1)}(L) \cap \mathcal{E}_m^{(6)}(R, \epsilon_2) \cap \mathcal{E}_m^{(2)}(R, \eta) \cap \mathcal{E}_m^{(6)}(R, \epsilon_2)$, we have the lower bound,*

$$\mathscr{L}(\boldsymbol{y}, 1) \geq \frac{1}{C(L, R)} e^{-m \cdot \hat{\Xi}_1}, \ \forall m \geq M(L, R). \tag{B.30}$$

where $C(L, R), M(L, R)$ are large enough, finite constants depending only on L, R.

Proof. Recall that from Corollary 2, we obtained the lower bound:

$$\mathscr{L}(\boldsymbol{y}, 1) \geq \frac{1}{2\sqrt{K}} \exp\left(-m \max_{\lambda \in \mathbb{R}} \left(\lambda - \hat{\mathbb{E}}_Y \ln \mathbb{E}_{E \sim \mathsf{Exp}(1)} e^{\lambda E} \psi_\sigma(E - Y) \right) \right) = \frac{1}{2\sqrt{K}} e^{-m\hat{\Xi}_1}, \ \forall m \geq M(K).$$

provided we can verify:

- $\hat{\mathbb{E}}(|Y| + |Y|^2 + |Y|^3) \leq K$: This can be ensured by taking $K \geq 3L$ and observing that under

event $\mathcal{E}_m^{(1)}(L)$ we have $1 + \hat{\mathbb{E}}Y^{40} \leq L$.

- $\hat{\lambda}$ which is the solution of the variational problem:

$$\hat{\lambda} = \arg\max_{\lambda \in \mathbb{R}} \left(\lambda - \hat{\mathbb{E}}_Y \ln \mathbb{E}_{E \sim \mathsf{Exp}(1)} e^{\lambda E} \psi_\sigma(E - Y) \right),$$

lies in a compact set $|\hat{\lambda}_1(\sigma)| \leq K$. Taking $K \geq R$ guarantees this under the event $\mathcal{E}_m^{(6)}(R, \epsilon_2)$.

- Finally we need to check:

$$\frac{1}{K} \leq \hat{\mathbb{E}}\sigma^2_{\mathsf{TExp}}\left(\hat{\lambda}, Y\right) \leq K, \tag{B.31}$$

for some value of K. Note that event $\mathcal{E}_m^{(6)}(R, \epsilon_2)$, guarantees $|\hat{\lambda}_1(\sigma)| \leq R$. The function $\lambda \mapsto \mathbb{E}\sigma^2_{\mathsf{TExp}}(\lambda, Y)$ is strictly positive and finite on compact sets, that is:

$$0 < \min_{|\lambda| \leq R} \mathbb{E}\sigma^2_{\mathsf{TExp}}(\lambda, Y) \leq \max_{|\lambda| \leq R} \mathbb{E}\sigma^2_{\mathsf{TExp}}(\lambda, Y) < \infty.$$

This can be checked by observing $\lambda \mapsto \mathbb{E}\sigma^2_{\mathsf{TExp}}(\lambda, Y)$ is continuous and if $\mathbb{E}\sigma^2_{\mathsf{TExp}}(\lambda, Y) = 0$ for some λ then, $\sigma^2_{\mathsf{TExp}}(\lambda, Y) \overset{\text{a.s.}}{=} 0$. This is clearly not possible since $\mathsf{TExp}(\lambda, y)$ is not deterministic for any finite λ, y. Hence there exists a constant depending only on R such that,

$$\frac{1}{C_1(R)} \leq \mathbb{E}\sigma^2_{\mathsf{TExp}}\left(\hat{\lambda}, Y\right) \leq C_1(R).$$

The event $\mathcal{E}_m^{(2)}(R, \eta)$ guarantees:

$$\sup_{|\lambda| \leq R} \left| \hat{\mathbb{E}}\sigma^2_{\mathsf{TExp}}(\lambda, Y) - \mathbb{E}\sigma^2_{\mathsf{TExp}}(\lambda, Y) \right| \leq \eta.$$

Since $|\hat{\lambda}_1(\sigma)| \leq R$, the above error bound holds for $\lambda = \hat{\lambda}_1(\sigma)$. Taking $\eta \leq (2C_1(R))^{-1}$

guarantees:

$$\frac{1}{2C_1(R)} \leq \hat{\mathbb{E}}\sigma^2_{\mathsf{TExp}}\left(\hat{\lambda}_1(\sigma), Y\right) \leq C_2(R) + \frac{1}{C_2(R)}.$$

This verifies (B.31) for a suitable K.

Hence, all the requirements of Proposition 22 are satisfied which gives us the claim of the lemma.

□

Lemma 37 (An Upper Bound on $\mathscr{U}$)**.** *We have the following upper bounds on $\mathscr{U}$:*

1. *Unconditional Upper Bound: For any $\boldsymbol{y} \in \mathbb{R}^m$, for any $q \in [0, 1)$, we have,*

$$\mathscr{U}\left(\boldsymbol{y}, \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix}\right) \leq e^{C_{\mathscr{U}} \cdot m},$$

for a universal constant $C_{\mathscr{U}}$ which depends only on the noise level σ.

2. *For any $R, L \in (0, \infty)$, there exists a critical value of η denoted by $\eta_2(R)$ depending only on R such that for any $\eta < \eta_2(R),\ \epsilon_2 > 0$, we have the upper bound,*

$$\mathscr{U}\left(\boldsymbol{y}, \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix}\right) \leq \frac{C(L,R) \cdot e^{-m\hat{\Xi}_2(q)}}{m^2 \cdot (1-q^2)^{m-2}} \ \forall\ q \in [0, 1-\epsilon_2],$$

for any $\boldsymbol{y} \in \mathcal{E}^{(1)}_m(L) \cap \mathcal{E}^{(3)}_m(R,\eta) \cap \mathcal{E}^{(6)}_m(R,\epsilon_2)$. In the above display, $C(L,R)$ is a constant depending only on the choice of L, R.

Proof. 1. We recall the definition of $\mathscr{U}$:

$$\mathscr{U}(\boldsymbol{y}, \boldsymbol{Q}) \stackrel{\text{def}}{=} \mathbb{E}\left[\prod_{i=1}^{m} \psi_\sigma(y_i - m|G_{1i}|^2)\psi_\sigma(y_i - m|G_{2i}|^2)\middle| \boldsymbol{G}^{\mathsf{H}}\boldsymbol{G} = \boldsymbol{Q}\right].$$

Observing that, $\psi_\sigma(x) \leq (2\pi\sigma^2)^{-1/2}$, we obtain, $\forall \boldsymbol{y} \in \mathbb{R}^m,\ \forall q \in (0,1)$,

$$\mathscr{U}\left(\boldsymbol{y}, \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix}\right) \leq e^{C_{\mathscr{U}} \cdot m},$$

for a universal constant $C_{\mathscr{U}} < \infty$ that depends only on σ.

2. Recall that Corollary 3 shows,

$$\mathscr{U}\left(\boldsymbol{y}, \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix}\right) \leq \frac{C(K)e^{-m\hat{\Xi}_2(q)}}{m^2 \cdot (1-q^2)^{m-2}}.$$

provided we can show:

- $\hat{\mathbb{E}}Y^{40} \leq K$. This is true under the event $\mathcal{E}_m^{(1)}(L)$ if we choose $K \geq L$.
- The minimizing arguments $(\hat{\lambda}_2(q;\sigma), \hat{\phi}(q;\sigma))$ satisfy $|\hat{\lambda}_2(q;\sigma)| + |\hat{\phi}(q;\sigma)| \leq K$ for any $q \in [0, 1-\epsilon_2]$. This is guaranteed by the event $\mathcal{E}_m^{(6)}(R, \epsilon_2)$ if $K \geq R$.
- Finally, we need to check:

$$\frac{1}{K} \leq \lambda_{\min}\left(\hat{\mathbb{E}}\boldsymbol{\Sigma}_{\mathsf{TWis}}\left(\hat{\lambda}, \hat{\phi}, Y\right)\right) \leq \lambda_{\max}\left(\hat{\mathbb{E}}\boldsymbol{\Sigma}_{\mathsf{TWis}}\left(\hat{\lambda}, \hat{\phi}, Y\right)\right) \leq K. \tag{B.32}$$

The event $\mathcal{E}_m^{(6)}(R, \epsilon_2)$ guarantees $|\hat{\lambda}_2(q;\sigma)| + |\hat{\phi}(q;\sigma)| \leq R,\ \forall\ q\ \in [0, 1-\epsilon_2]$. The matrix function $(\lambda, \phi) \mapsto \mathbb{E}\boldsymbol{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, Y)$ is:

(a) Bounded on the compact set $|\lambda| + |\phi| \leq R$. Indeed:

$$\begin{aligned} \|\mathbb{E}\boldsymbol{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, Y)\| &\leq \mathbb{E}\|\boldsymbol{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, Y)\| \\ &\overset{\text{(a)}}{\leq} C(1 + |\lambda|^2 + |\phi|^2 + \mathbb{E}Y^2) \leq C(1+R^2). \end{aligned}$$

In the inequality marked (a), we used the moment bounds for the tilted Wishart

distribution derived in Claim (4) of Lemma 44.

(b) Strictly positive definite on the compact set $|\lambda| + |\phi| \leq R$. To see this we note that if $\lambda_{\min}(\mathbb{E}\mathbf{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, Y)) = 0$ for some λ, ϕ then since

$$\lambda_{\min}(\mathbb{E}\mathbf{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, Y)) \geq \mathbb{E}\lambda_{\min}(\mathbf{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, Y)),$$

we have $\lambda_{\min}(\mathbf{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, Y)) = 0$ almost surely (with respect to the distribution of Y). This contradicts Claim (6) of Lemma 44.

Hence, there exists a positive and finite constant $C_2(R)$ depending only on R such that,

$$\frac{1}{C_2(R)} \leq \lambda_{\min}\left(\mathbb{E}\mathbf{\Sigma}_{\mathsf{TWis}}\left(\hat{\lambda}, \hat{\phi}, Y\right)\right) \leq \lambda_{\max}\left(\mathbb{E}\mathbf{\Sigma}_{\mathsf{TWis}}\left(\hat{\lambda}, \hat{\phi}, Y\right)\right) \leq C_2(R).$$

The event $\mathcal{E}_m^{(3)}(R, \eta)$ guarantees:

$$\left|\lambda_{\min}\left(\mathbb{E}\mathbf{\Sigma}_{\mathsf{TWis}}\left(\hat{\lambda}, \hat{\phi}, Y\right)\right) - \lambda_{\min}\left(\hat{\mathbb{E}}\mathbf{\Sigma}_{\mathsf{TWis}}\left(\hat{\lambda}, \phi, Y\right)\right)\right| \leq \eta,$$
$$\left|\lambda_{\max}\left(\mathbb{E}\mathbf{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, Y)\right) - \lambda_{\max}\left(\hat{\mathbb{E}}\mathbf{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, Y)\right)\right| \leq \eta.$$

Choosing $\eta \leq (2C_2(R))^{-1}$, we have,

$$\leq \lambda_{\min}\left(\hat{\mathbb{E}}\mathbf{\Sigma}_{\mathsf{TWis}}\left(\hat{\lambda}, \hat{\phi}, Y\right)\right) \geq \frac{1}{2C_2(R)},$$
$$\lambda_{\max}\left(\hat{\mathbb{E}}\mathbf{\Sigma}_{\mathsf{TWis}}\left(\hat{\lambda}, \hat{\phi}, Y\right)\right) \leq C_2(R) + \frac{1}{2C_2(R)},$$

which verifies (B.32) for a suitable K.

Hence all the assumptions of Corollary 3 have been verified, which gives us the claim in item (2) of the lemma.

□

Finally we restate and prove Proposition 10.

Proposition 10. *Suppose that δ, Δ, σ are such that $\mathcal{F}(q;\delta,\Delta,\sigma) > \mathcal{F}(0;\delta,\Delta,\sigma) = 0 \;\forall\; q \in (0,1)$ and $\frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(0;\delta,\Delta,\sigma) > 0$. Then, $\mathbf{I}(\boldsymbol{y},\boldsymbol{z};\boldsymbol{A},\boldsymbol{W}) = o(m)$.*

Proof. In Lemmas 12 and 13, we showed that,

$$\mathbf{I}(\boldsymbol{y},\boldsymbol{z};\boldsymbol{A},\boldsymbol{W}) \leq \frac{1}{n-1}\mathbb{E}_{\boldsymbol{y}}\left[\frac{\int_0^1 \mathscr{U}\left(\boldsymbol{y},\begin{bmatrix}1 & \sqrt{b}\\ \sqrt{b} & 1\end{bmatrix}\right)\cdot\frac{(1-b)^{n-2}}{(1-b/2)^{\Delta m}}\,\mathrm{d}b}{\mathscr{L}^2(\boldsymbol{y},1)}\cdot\mathbf{1}_{\mathcal{E}_m}\right] + C\cdot m\cdot\sqrt{\mathbb{P}(\mathcal{E}_m^c)}$$

$$= \frac{2}{n-1}\mathbb{E}_{\boldsymbol{y}}\left[\frac{\int_0^1 \mathscr{U}\left(\boldsymbol{y},\begin{bmatrix}1 & q\\ q & 1\end{bmatrix}\right)\cdot\frac{q(1-q^2)^{n-2}}{(1-q^2/2)^{\Delta m}}\,\mathrm{d}q}{\mathscr{L}^2(\boldsymbol{y},1)}\cdot\mathbf{1}_{\mathcal{E}_m}\right] + C\cdot m\cdot\sqrt{\mathbb{P}(\mathcal{E}_m^c)}$$

$$\overset{\text{(a)}}{=} (1)+(2)+(3)+C\cdot m\cdot\sqrt{\mathbb{P}(\mathcal{E}_m^c)}. \tag{B.33}$$

In the step marked (a), we split the integral into three parts:

$$(1) \stackrel{\text{def}}{=} \frac{2}{n-1}\mathbb{E}_{\boldsymbol{y}}\left[\frac{\int_{1-\epsilon_2}^{1}\mathscr{U}\left(\boldsymbol{y},\begin{bmatrix}1 & q\\ q & 1\end{bmatrix}\right)\cdot\frac{q(1-q^2)^{n-2}}{(1-q^2/2)^{\Delta m}}\,\mathrm{d}q}{\mathscr{L}^2(\boldsymbol{y},1)}\cdot\mathbf{1}_{\mathcal{E}_m}\right],$$

$$(2) \stackrel{\text{def}}{=} \frac{2}{n-1}\mathbb{E}_{\boldsymbol{y}}\left[\frac{\int_{\epsilon_1}^{1-\epsilon_2}\mathscr{U}\left(\boldsymbol{y},\begin{bmatrix}1 & q\\ q & 1\end{bmatrix}\right)\cdot\frac{q(1-q^2)^{n-2}}{(1-q^2/2)^{\Delta m}}\,\mathrm{d}q}{\mathscr{L}^2(\boldsymbol{y},1)}\cdot\mathbf{1}_{\mathcal{E}_m}\right],$$

$$(3) \stackrel{\text{def}}{=} \frac{2}{n-1}\mathbb{E}_{\boldsymbol{y}}\left[\frac{\int_{0}^{\epsilon_1}\mathscr{U}\left(\boldsymbol{y},\begin{bmatrix}1 & q\\ q & 1\end{bmatrix}\right)\cdot\frac{q(1-q^2)^{n-2}}{(1-q^2/2)^{\Delta m}}\,\mathrm{d}q}{\mathscr{L}^2(\boldsymbol{y},1)}\cdot\mathbf{1}_{\mathcal{E}_m}\right].$$

In the above display $\epsilon_1, \epsilon_2 \in (0,1)$ are parameters which will be set appropriately. We also recall that we had set:

$$\mathcal{E}_m = \mathcal{E}_m^{(1)}(L) \cap \mathcal{E}_m^{(2)}(R,\eta) \cap \mathcal{E}_m^{(3)}(R,\eta) \cap \mathcal{E}_m^{(4)}(\eta) \cap \mathcal{E}_m^{(5)}(\eta,\epsilon_2) \cap \mathcal{E}_m^{(6)}(R,\epsilon_2).$$

where the various events have been defined in Equations B.24. We now describe how to set the parameters $L, R, \eta, \epsilon_1, \epsilon_2$ so that each of the terms in (B.33) is $o(m)$. We also draw the readers attention to the point that the parameter ϵ_2 used to define the cutoff points for the integrals (1) and

(2) is same as the ϵ_2 in the definition of the event $\mathcal{E}_m^{(5)}(\eta, \epsilon_2), \mathcal{E}_m^{(6)}(R, \epsilon_2)$. Notice also the same parameter R is involved in the definitions of the events $\mathcal{E}_m^{(2)}(R, \eta), \mathcal{E}_m^{(3)}(R, \eta), \mathcal{E}_m^{(6)}(R, \epsilon_2)$.

Analysis of (1)**:** By Lemma 37, we know that,

$$\mathscr{U}\left(\boldsymbol{y}, \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix}\right) \leq e^{C_{\mathscr{U}} \cdot m}.$$

We next appeal to Lemma 36. We enforce the requirement

$$\eta < \eta_1(R) \tag{B.34}$$

and obtain,

$$\mathscr{L}(\boldsymbol{y}, 1) \geq \frac{1}{C(L, R)} \cdot e^{-m\hat{\Xi}_1}.$$

The event $\mathcal{E}_m^{(5)}(\eta, \epsilon_2)$ guarantees that $\hat{\Xi}_1(\sigma) \leq \Xi_1(\sigma) + \eta$. By enforcing:

$$\eta \leq 1, \tag{B.35}$$

we have $\hat{\Xi}_1(\sigma) \leq \Xi_1(\sigma) + 1$ which is an absolute constant (depending only on the noise level). Consequently, we have $\mathscr{L}(\boldsymbol{y}, 1) \geq C(L, R)^{-1} \cdot e^{-C_{\mathscr{L}} \cdot m}$ for some universal constant

$C_{\mathscr{L}} \in (0, \infty)$ depending only on the noise level. Hence we have, for $\min(n, m) \geq 4$

$$\begin{aligned}
(1) &\leq \frac{2 \cdot C^2(L, R) \cdot e^{(C_{\mathscr{U}} + 2C_{\mathscr{L}}) \cdot m}}{n-1} \cdot \int_{1-\epsilon_2}^{1} (1-q^2)^{n-2} \, \mathrm{d}q \\
&\leq \frac{2 \cdot C^2(L, R)}{n-1} \cdot e^{(C_{\mathscr{U}} + 2C_{\mathscr{L}}) \cdot m} \cdot (1 - (1-\epsilon_2)^2)^{\frac{n}{2}} \\
&\leq \frac{2 \cdot C^2(L, R)}{n-1} \cdot e^{(C_{\mathscr{U}} + 2C_{\mathscr{L}}) \cdot m} \cdot (2\epsilon_2)^{\frac{n}{2}} \\
&= \frac{2 \cdot C^2(L, R)}{n-1} \cdot \exp\left(m \cdot \left(C_{\mathscr{U}} + 2C_{\mathscr{L}} + \frac{\ln(2)}{2\delta} - \frac{\ln \frac{1}{\epsilon_2}}{2\delta} \right) \right)
\end{aligned}$$

We set:

$$\epsilon_2 = \frac{1}{2} \cdot e^{-2\delta(C_{\mathscr{U}} + 2C_{\mathscr{L}})} < 1, \tag{B.36}$$

which gives us $(1) = O(1/n) = o(1)$.

Analysis of $\mathbb{P}(\mathcal{E}_m^c)$: As suggested by Lemma 35, we set $L > \mathbb{E}|Y|^{40}$. For example, we can set $L = 1 + \mathbb{E}Y^{40}$. We will also enforce the constraint $R > R_c(\epsilon_2)$ for example by setting $R = R_c(\epsilon_2) + 1$ (note that ϵ_2 has been set in (B.36)). This ensures that $\mathbb{P}(\mathcal{E}_m^c) = o(1)$. At this set we have set R, ϵ_2, L and we are still free to set $\eta > 0, \epsilon_1 \in (0, 1)$ arbitrarily subject to the requirements in (B.34)-(B.35).

Analysis of (2): We enforce:

$$\eta < \min(\eta_1(R), \eta_2(R)) \tag{B.37}$$

which is enough to satisfy the assumptions of Lemma 36 and item (2) of Lemma 37, which

gives us,

$$\frac{\mathscr{U}\left(\boldsymbol{y}, \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix}\right)}{\mathscr{L}^2(\boldsymbol{y}, 1)} \leq \frac{C \cdot e^{-m(\hat{\Xi}_2(q) - 2\hat{\Xi}_1)}}{m^2 \cdot (1-q^2)^{m-2}} \; \forall \; q \in [0, 1-\epsilon_2]. \tag{B.38}$$

This allows us to upper bound the term (2) as follows:

$$\begin{aligned}
(2) &\overset{\text{def}}{=} \frac{2}{n-1} \mathbb{E}_{\boldsymbol{y}} \left[\frac{\int_{\epsilon_1}^{1-\epsilon_2} \mathscr{U}\left(\boldsymbol{y}, \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix}\right) \cdot \frac{q(1-q^2)^{n-2}}{(1-q^2/2)^{\Delta m}} \, \mathrm{d}q}{\mathscr{L}^2(\boldsymbol{y}, 1)} \cdot \mathbf{1}_{\mathcal{E}_m} \right] \\
&\leq \frac{C}{(n-1) \cdot m^2} \cdot \mathbb{E}_{\boldsymbol{y}} \left[\int_{\epsilon_1}^{1-\epsilon_2} e^{-m\hat{\mathcal{F}}(q;\delta,\Delta,\sigma)} \, \mathrm{d}q \cdot \mathbf{1}_{\mathcal{E}_m} \right]
\end{aligned}$$

Since event $\mathcal{E}_m^{(5)}(\eta, \epsilon_2)$ guarantees $|\hat{\Xi}_1(\sigma) - \Xi_1(\sigma)| \leq \eta, \; \sup_{q \in [0, 1-\epsilon_2]} |\hat{\Xi}_2(q;\sigma) - \Xi_2(q;\sigma)| \leq \eta$, we have,

$$|\hat{\mathcal{F}}(q;\delta,\Delta,\sigma) - \mathcal{F}(q;\delta,\Delta,\sigma)| \leq 3\eta \; \forall \; q \in [0, 1-\epsilon_2].$$

Since $\delta < \delta_c(\sigma^2, \Delta)$ and $\epsilon_1 > 0$, Recall that we have, $\inf_{q \in [\epsilon_1, 1]} \mathcal{F}(q;\delta,\Delta,\sigma) > 0 = \mathcal{F}(0;\delta,\Delta,\sigma)$ (see (B.20)). Hence, we can enforce that η, ϵ_1 satisfy:

$$\eta < \frac{1}{6} \inf_{q \in [\epsilon_1, 1]} \mathcal{F}(q;\delta,\Delta,\sigma) \tag{B.39}$$

This guarantees, for any $q \in [\epsilon_1, 1-\epsilon_2]$,

$$\begin{aligned}\hat{\mathcal{F}}(q;\delta,\Delta,\sigma) &\geq \mathcal{F}(q;\delta,\Delta,\sigma) - 3\eta \\ &\geq \inf_{q\in[\epsilon_1,1]} \mathcal{F}(q;\delta,\Delta,\sigma) - 3\eta \\ &\geq \frac{1}{2}\inf_{q\in[\epsilon_1,1]} \mathcal{F}(q;\delta,\Delta,\sigma) > 0.\end{aligned}$$

Hence,

$$(2) \leq \frac{C}{(n-1)\cdot m^2} \cdot \exp\left(-\frac{m}{2}\cdot \inf_{q\in[\epsilon_1,1]} \mathcal{F}(q;\delta,\Delta,\sigma)\right) = o(1).$$

Analysis of (3): We recall that Term (3) was given by:

$$(3) = \frac{2}{n-1}\mathbb{E}_{\boldsymbol{y}}\left[\frac{\int_0^{\epsilon_1} \mathscr{U}\left(\boldsymbol{y}, \begin{bmatrix}1 & q\\ q & 1\end{bmatrix}\right)\cdot \frac{q(1-q^2)^{n-2}}{(1-q^2/2)^{\Delta m}}\,\mathrm{d}q}{\mathscr{L}^2(\boldsymbol{y},1)}\cdot \mathbf{1}_{\mathcal{E}_m}\right].$$

The upper bound in (B.38) applies to $q \in [0, \epsilon_1]$. Hence we obtain,

$$(3) \leq \frac{C}{(n-1)\cdot m^2}\cdot \mathbb{E}_{\boldsymbol{y}}\left[\int_0^{\epsilon_1} e^{-m\hat{\mathcal{F}}(q;\delta,\Delta,\sigma)}\,\mathrm{d}q \cdot \mathbf{1}_{\mathcal{E}_m}\right]$$

Next we approximate $\hat{\mathcal{F}}$ by its taylors expansion at $q = 0$. First observe that, $\hat{\mathcal{F}}(0;\delta,\Delta,\sigma) = 0$ and,

$$\frac{\mathrm{d}\hat{\mathcal{F}}}{\mathrm{d}q}(q;\delta,\Delta,\sigma) = \hat{\phi}(q;\sigma) - 2\left(1-\frac{1}{\delta}\right)\frac{q}{1-q^2} - \frac{\Delta q}{1-q^2/2} \implies \frac{\mathrm{d}\hat{\mathcal{F}}}{\mathrm{d}q}(0;\delta,\Delta,\sigma) = 0.$$

We enforce the constraint

$$\eta < \frac{1}{4}\frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(0;\delta,\Delta,\sigma). \tag{B.40}$$

We set $\epsilon_1 \in (0, 1/2)$ which guarantees:

$$\left|\frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(q;\delta,\Delta,\sigma) - \frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(0;\delta,\Delta,\sigma)\right| \leq \frac{1}{2}\frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(0;\delta,\Delta,\sigma).$$

(B.21) and the fact that $\mathcal{F}(\cdot;\delta,\Delta,\sigma)$ has a continuous second derivative at $q = 0$ ensures this is possible. Hence we have,

$$\frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(q;\delta,\Delta,\sigma) > 2\eta,\ \forall\ q < \epsilon_1.$$

The event $\mathcal{E}_m^{(4)}(\eta)$ guarantees:

$$\sup_{q\in[0,1/2]}\left|\frac{\mathrm{d}^2\hat{\mathcal{F}}}{\mathrm{d}q^2}(q;\delta,\Delta,\sigma) - \frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(q;\delta,\Delta,\sigma)\right| \leq \eta \implies \frac{\mathrm{d}^2\hat{\mathcal{F}}}{\mathrm{d}q^2}(q;\delta,\Delta,\sigma) > \eta,\ \forall\ q < \epsilon_1.$$

Then by Taylor's theorem, we have, $\forall\ q \in [0, \epsilon_1)$,

$$\begin{aligned}\hat{\mathcal{F}}(q;\delta,\Delta,\sigma) &\geq \hat{\mathcal{F}}(0;\delta,\Delta,\sigma) + \frac{\mathrm{d}\hat{\mathcal{F}}}{\mathrm{d}q}(0;\delta,\Delta,\sigma)\cdot q + \left(\inf_{x\in[0,\epsilon_1)}\frac{\mathrm{d}^2\hat{\mathcal{F}}}{\mathrm{d}q^2}(x;\delta,\Delta,\sigma)\right)\cdot\frac{q^2}{2}\\ &\geq \frac{\eta q^2}{2}.\end{aligned}$$

Hence we obtain,

$$(3) \leq \frac{C}{(n-1)\cdot m^2}\cdot\int_0^{\epsilon_1} e^{-\frac{\eta q^2}{2}\cdot m} \leq \frac{C}{(n-1)\cdot m^2} = o(1).$$

Finally we note that set η as,

$$\eta = \min\left(1, \eta_1(R), \eta_2(R), \frac{1}{6}\inf_{q\in[\epsilon_1,1]}\mathcal{F}(q;\delta,\Delta,\sigma), \frac{1}{4}\frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(0;\delta,\Delta,\sigma)\right)$$

satisfies requirements in (B.34),(B.35),(B.37), (B.39) and (B.40) and also ensures η is a fixed positive constant.

This concludes the proof of the proposition. □

B.5 Proofs from Section 4.8

This section is devoted to proving Proposition 11. Recall that the function $\mathcal{F}(q;\delta,\Delta,\sigma)$ was defined as:

$$\mathcal{F}(q;\delta,\Delta,\sigma) \stackrel{\text{def}}{=} \Xi_2(q;\sigma) - 2\Xi_1(\sigma) + \left(1-\frac{1}{\delta}\right)\ln(1-q^2) + \Delta\ln\left(1-\frac{q^2}{2}\right),$$

where the functions, Ξ_1, Ξ_2 are defined as follows:

$$\Xi_2(q;\sigma) \stackrel{\text{def}}{=} \max_{(\lambda,\phi)\in\mathbb{R}} \left(2\lambda + \phi q - \mathbb{E}_Y \ln Z_{\mathsf{TWis}}(\lambda,\phi,Y)\right),$$
$$\Xi_1(\sigma) \stackrel{\text{def}}{=} \max_{\lambda\in\mathbb{R}} \left(\lambda - \mathbb{E}_Y \ln Z_{\mathsf{TExp}}(\lambda,Y)\right).$$

In the above display the random variable $Y = |G|^2 + \sigma\epsilon$, where $G \sim \mathcal{CN}(0,1)$ and $\epsilon \sim \mathcal{N}(0,1)$. Our goal is to identify conditions on (δ,Δ,σ) such that,

$$\mathcal{F}(q;\delta,\Delta,\sigma) > \mathcal{F}(0;\delta,\Delta,\sigma)\ \forall\ q\in(0,1),\ \frac{\mathrm{d}^2}{\mathrm{d}q^2}\mathcal{F}(q;\delta,\Delta,\sigma) > 0. \tag{B.41}$$

We will not be able to solve this for a general $\sigma > 0$, but only for small enough σ since in the limit $\sigma \to 0$, the variational problems in the definition of Ξ_2, Ξ_1 simplify considerably.

We first begin with a heuristic derivation of the zero noise limit of the functions $\Xi_2(q;\sigma)$ and

$\Xi_1(\sigma)$. Recalling the definition of $Z_{\mathsf{TExp}}(\lambda, y)$ (Definition 5):

$$\begin{aligned}\ln Z_{\mathsf{TExp}}(\lambda, Y) &= \mathbb{E}_{E\sim\mathsf{Exp}(1)} e^{\lambda E}\psi_\sigma(E - |G|^2 - \sigma\epsilon) \\ &= e^{(\lambda-1)(|G|^2+\sigma\epsilon)}\mathbb{E}_{\omega\sim\mathcal{N}(0,1)} e^{\sigma(\lambda-1)\omega}\mathbf{1}_{|G|^2+\sigma\epsilon+\sigma\omega\geq 0} \\ &\overset{\sigma\to 0}{\to} e^{(\lambda-1)|G|^2}.\end{aligned}$$

This gives us,

$$\lambda - \mathbb{E}_Y \ln Z_{\mathsf{TExp}}(\lambda, Y) \overset{\sigma\to 0}{\to} 1.$$

In the zero noise limit, the variational problem in the definition of Ξ_1 is trivial. Hence, it makes sense to extend the definition of $\Xi_1(\sigma)$ to include $\sigma = 0$ as $\Xi_1(0) \overset{\text{def}}{=} 1$. Likewise, recalling Definition 6, we have,

$$\begin{aligned}&Z_{\mathsf{TWis}}(\lambda, \phi, Y) = Z_{\mathsf{TWis}}\left(\lambda, \phi, |G|^2 + \sigma\epsilon\right) \\ &= \mathbb{E}\exp\left((\lambda-1)(2Y + \sigma(\omega_1+\omega_2)) + \phi\sqrt{(Y+\sigma\omega_1)(Y+\sigma\omega_2)}\cos(\theta)\right)\mathbf{1}_{Y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0} \\ &\overset{\sigma\to 0}{\to} e^{2(\lambda-1)|G|^2}\mathbb{E}_\theta e^{\phi|G|^2\cos(\theta)} \\ &= e^{2(\lambda-1)|G|^2} I_0(|G|^2\phi).\end{aligned}$$

In the last step we used the definition of Modified Bessel function $I_0(x) \overset{\text{def}}{=} \mathbb{E}e^{x\cos\theta}$. Hence we extend the definition of $\Xi_2(q;\sigma)$ to $\sigma = 0$ as:

$$\Xi_2(q;0) \overset{\text{def}}{=} 2 + \max_{\phi\in\mathbb{R}} q\phi - \mathbb{E}_{Z\sim\mathcal{CN}(0,1)} \ln I_0(\phi|Z|^2).$$

This allows to guess the correct zero noise limit of $\mathcal{F}(q;\delta,\Delta,\sigma)$ as:

$$\mathcal{F}(q;\delta,\Delta,0) \overset{\text{def}}{=} \Xi_2(q;0) - 2\Xi_1(0) + \left(1 - \frac{1}{\delta}\right)\ln(1-q^2) + \Delta\ln\left(1 - \frac{q^2}{2}\right).$$

The remainder of this section is organized as follows:

1. In Section B.5.1 we analyze the zero noise limit function $\mathcal{F}(q;\delta,\Delta,0)$ and find a condition on (δ,Δ) such that (B.41) holds for $\mathcal{F}(q;\delta,\Delta,0)$.

2. In Section B.5.2, we show that $\Xi_1(\sigma)$ converges to $\Xi_1(0)$ and $\Xi_2(q;\sigma)$ converges to $\Xi_2(q;0)$ in an appropriate sense.

3. Finally Section B.5.3 contains the proof of Proposition 11.

Throughout this section, C denotes a universal constant that does not depend on σ. As before this constant may change from line to line.

B.5.1 Analysis in the Low Noise Limit

The following lemma shows that if $\delta < 2$, and Δ is small enough (but positive), the function $\mathcal{F}(q;\delta,\Delta,0)$ is strictly increasing.

Lemma 38 (Limiting Variational Problems)**.** *Consider the following functions for $q \in [0,1)$:*

$$\Xi_2(q;0) \stackrel{def}{=} 2 + \max_{\phi\in\mathbb{R}} q\phi - \mathbb{E}_{Z\sim\mathcal{CN}(0,1)} \ln I_0(\phi|Z|^2),$$
$$\phi_2(q;0) \stackrel{def}{=} \arg\max_{\phi\in\mathbb{R}} q\phi - \mathbb{E}_{Z\sim\mathcal{CN}(0,1)} \ln I_0(\phi|Z|^2).$$

Then we have,

1. *The function $\phi \mapsto q\phi - \mathbb{E}_{Z\sim\mathcal{CN}(0,1)} \ln I_0(\phi|Z|^2)$ has a unique maximizer $\phi_2(q;0)$ which satisfies: $0 \le \phi_2(q;0) < \infty$ for any $q \in [0,1)$. Furthermore, $\max_{q\in[0,1-\eta]} \phi_2(q;0) < \infty$ for any $\eta \in (0,1)$.*

2. *The function:*

$$f(q) \stackrel{def}{=} \Xi_2(q;0) + \left(1 - \frac{1}{\delta}\right)\ln(1-q^2) + \Delta \ln\left(1 - \frac{q^2}{2}\right),$$

is a strictly increasing function of q *with* $f(0) < f(q)\ \forall q \in [0,1)$*, provided,*

$$0 < \delta < 2,\ 0 < \Delta < \frac{2-\delta}{\delta}.$$

Proof. 1. The function $\phi \mapsto q\phi - \mathbb{E}_{Z\sim\mathcal{CN}(0,1)} \ln I_0(\phi|Z|^2)$ is strictly concave (see Fact 3, item (5), Appendix B.9). Hence, $q\phi - \mathbb{E}_{Z\sim\mathcal{CN}(0,1)} \ln I_0(\phi|Z|^2)$ has at most one maximizer. Next observe that any maximizer must lie in $[0,\infty]$. This is because $\mathbb{E}\ln I_0(|\phi||Z|^2) = \mathbb{E}\ln I_0(-|\phi||Z|^2)$ since I_0 is even (see Fact 3, Appendix B.9), but $q|\phi| \geq -q|\phi|$. This shows that if $\phi_2(q;0)$ exists, we must have, $\phi_2(q;0) \geq 0$. In order to show existence of $\phi_2(0,q)$ it is sufficient to find a solution to the first order optimality conditions:

$$\mathbb{E}|Z|^2 \cdot \frac{I_0'(\phi|Z|^2)}{I_0(\phi|Z|^2)} = q.$$

Note that:

$$\mathbb{E}|Z|^2 \cdot \frac{I_0'(\phi|Z|^2)}{I_0(\phi|Z|^2)}\bigg|_{\phi=0} = 0.$$

In order to check that $\phi_2(q;0) < \infty$, it is sufficient to show that,

$$\lim_{\phi\to\infty} q - \mathbb{E}|Z|^2 \cdot \frac{I_0'(\phi|Z|^2)}{I_0(\phi|Z|^2)} < 0.$$

By Monotone convergence theorem and the fact that $\frac{I_0'(x)}{I_0(x)} \uparrow 1$ as $x \uparrow \infty$ (see Fact 3, Appendix B.9), we have,

$$\lim_{\phi\to\infty} q - \mathbb{E}|Z|^2 \cdot \frac{I_0'(\phi|Z|^2)}{I_0(\phi|Z|^2)} = q - 1 < 0,\ \forall\ q < 1.$$

This confirms $\phi_2(q;0) < \infty$ for any $q < 1$. Further inspection of the stationarity condition:

$$\mathbb{E}|Z|^2 \cdot \frac{I_0'(\phi|Z|^2)}{I_0(\phi|Z|^2)}\bigg|_{\phi=\phi_2(q;0)} = q$$

reveals that $\phi_2(q;0)$ is an increasing function of q since the function on the left is an increasing function of ϕ (see Fact 3, Appendix B.9). Hence,

$$\max_{q\in[0,1-\eta]} \phi_2(q;0) = \phi_2(q, 1-\eta) < \infty.$$

This concludes the proof of item (1).

2. It is sufficient to show that the function

$$f(q) \stackrel{\text{def}}{=} \Xi_2(q;0) + \left(1 - \frac{1}{\delta}\right)\ln(1-q^2) + \Delta \ln\left(1 - \frac{q^2}{2}\right),$$

is strictly increasing or:

$$\frac{\mathrm{d}f(q)}{\mathrm{d}q} > 0.$$

We can compute the first derivative:

$$\frac{\mathrm{d}f(q)}{\mathrm{d}q} = \phi_2(q;0) - 2\left(1 - \frac{1}{\delta}\right)\frac{q}{1-q^2} - \frac{\Delta q}{1-\frac{q^2}{2}}.$$

Hence,

$$\frac{\mathrm{d}f(q)}{\mathrm{d}q} > 0 \Leftrightarrow \phi_2(q;0) > 2\left(1 - \frac{1}{\delta}\right)\frac{q}{1-q^2} + \frac{\Delta q}{1-\frac{q^2}{2}} \stackrel{\text{def}}{=} \phi_3(q)$$

Note that since $\phi_2(q;0)$ is the maximizing argument of the strictly concave function $q\phi -$

$\mathbb{E}\ln I_0(\phi|Z|^2)$, we have,

$$\frac{\mathrm{d}f(q)}{\mathrm{d}q} > 0 \Leftrightarrow \frac{\mathrm{d}}{\mathrm{d}\phi}\left(q\phi - \mathbb{E}\ln I_0(\phi|Z|^2)\right)\Big|_{\phi=\phi_3(q)} > 0 \Leftrightarrow q > \mathbb{E}|Z|^2 \cdot \frac{I_0'(\phi_3(q)\cdot|Z|^2)}{I_0(\phi_3(q)\cdot|Z|^2)}$$

Next we make the following sequence of observations:

(a) $\phi_3(0) = 0$, hence,

$$\mathbb{E}\frac{I_0'(\phi_3(q)\cdot|Z|^2)}{I_0(\phi_3(q)\cdot|Z|^2)}\Big|_{q=0} = 0.$$

(b) We can compute the first derivative:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}q}\mathbb{E}|Z|^2\frac{I_0'(\phi_3(q)\cdot|Z|^2)}{I_0(\phi_3(q)\cdot|Z|^2)}\Big|_{q=0} &= \frac{\mathrm{d}}{\mathrm{d}\phi}\left(\mathbb{E}|Z|^2\frac{I_0'(\phi\cdot|Z|^2)}{I_0(\phi\cdot|Z|^2)}\right)\Big|_{\phi=0}\cdot\frac{\mathrm{d}\phi_3(q)}{\mathrm{d}q}\Big|_{q=0} \\ &\overset{\text{(a)}}{=} \frac{\mathbb{E}|Z|^4}{2}\cdot\left(2\left(1-\frac{1}{\delta}\right)+\Delta\right) \\ &= 1-\left(\frac{2-\delta}{\delta}-\Delta\right) < 1,\end{aligned}$$

where the step marked (a) used Fact 3 and the definition of $\phi_3(q)$ to compute the relevant derivatives.

(c) Finally we note that, the function,

$$q \mapsto \mathbb{E}|Z|^2\cdot\frac{I_0'(\phi_3(q)\cdot|Z|^2)}{I_0(\phi_3(q)\cdot|Z|^2)},$$

is concave and increasing since $\frac{I_0'(x)}{I_0(x)}$ is concave and increasing (Fact 3, Appendix B.9) and $\phi_3(q)$ is convex and increasing.

The above three observations immediately imply:

$$\mathbb{E}|Z|^2\cdot\frac{I_0'(\phi_3(q)\cdot|Z|^2)}{I_0(\phi_3(q)\cdot|Z|^2)} < q,\ \forall\ q>0 \implies \frac{\mathrm{d}f}{\mathrm{d}q}(q) > 0,\ \forall\ q > 0.$$

Furthermore,

$$\frac{\mathrm{d}f}{\mathrm{d}q}(0) = 0.$$

Hence $f(q)$ is a stricly increasing function of q and hence so is $\mathcal{F}(q; \delta, \Delta, 0)$. This concludes the proof of item (2).

□

B.5.2 Convergence to the Low Noise Limit

The following lemma shows that $\lim_{\sigma \to 0} \Xi_1(\sigma) = \Xi_1(0) = 1$.

Lemma 39. *Recall that $\Xi_1(\sigma)$ and $\lambda_1(\sigma)$ denote the optimal value and solution of the variational problem:*

$$\Xi_1(\sigma) \stackrel{def}{=} \max_{\lambda \in \mathbb{R}} \left(\lambda - \mathbb{E}_Y \ln \mathbb{E}_{E \sim \mathsf{Exp}(1)} e^{\lambda E} \psi_\sigma(E - Y) \right),$$
$$\lambda_1(\sigma) \stackrel{def}{=} \arg\max_{\lambda \in \mathbb{R}} \left(\lambda - \mathbb{E}_Y \ln \mathbb{E}_{E \sim \mathsf{Exp}(1)} e^{\lambda E} \psi_\sigma(E - Y) \right).$$

Then we have,

1. $\lambda_1(\sigma) \leq 1$ *for all* $\sigma > 0$.
2. $\Xi_1(\sigma)$ *is a decreasing function of* σ.
3. $\lim_{\sigma \to 0} \Xi_1(\sigma) = 1$.

Proof. First we can write $\mathbb{E}_{E \sim \mathsf{Exp}(1)} e^{\lambda E} \psi_\sigma(E - Y)$ as follows:

$$\mathbb{E}_{E \sim \mathsf{Exp}(1)} e^{\lambda E} \psi_\sigma(E - Y) = e^{(\lambda - 1)Y} \mathbb{E}_{\omega \sim \mathcal{N}(0,1)} e^{\sigma(\lambda - 1)\omega} \mathbf{1}_{Y + \sigma\omega \geq 0}$$

Note that $Y \stackrel{\text{d}}{=} |Z|^2 + \epsilon,\ Z \sim \mathcal{CN}(0,1)\,,\ \epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$. Hence, we have,

$$\begin{aligned}\Xi_1(\sigma) &= \max_{\lambda \in \mathbb{R}} 1 - \mathbb{E}_Y \ln \mathbb{E}_{\omega \sim \mathcal{N}(0,1)} e^{\sigma(\lambda-1)\omega} \mathbf{1}_{Y+\sigma\omega \geq 0} \\ &= 1 - \min_{\gamma \in \mathbb{R}} \mathbb{E}_Y \ln \mathbb{E}_{\omega \sim \mathcal{N}(0,1)} e^{\sigma\gamma\omega} \mathbf{1}_{Y+\sigma\omega \geq 0}. \end{aligned} \tag{B.42}$$

Likewise,

$$\lambda_1(\sigma^2) = 1 + \arg\min_{\gamma \in \mathbb{R}} \mathbb{E}_Y \ln \mathbb{E}_{\omega \sim \mathcal{N}(0,1)} e^{\sigma\gamma\omega} \mathbf{1}_{Y+\sigma\omega \geq 0}$$

Now we consider the three claims one by one:

1. Observe that, by the Chebychev Association Inequality (Fact 2 , Appendix B.9),

$$\begin{aligned}\mathbb{E}_\omega e^{-\sigma|\gamma|\omega} \mathbf{1}_{Y+\sigma\omega \geq 0} &\leq \mathbb{E}_\omega e^{-\sigma|\gamma|\omega} \cdot \mathbb{P}(Y + \sigma\omega \geq 0) \\ &= \mathbb{E}_\omega e^{\sigma|\gamma|\omega} \cdot \mathbb{P}(Y + \sigma\omega \geq 0) \\ &\leq \mathbb{E}_\omega e^{\sigma|\gamma|\omega} \mathbf{1}_{Y+\sigma\omega \geq 0}.\end{aligned}$$

   This shows that $\lambda_1(\sigma^2) \leq 1$.

2. A gaussian integral shows that:

$$\mathbb{E}_\omega e^{\sigma\gamma\omega} \mathbf{1}_{Y+\sigma\omega \geq 0} = \frac{1}{\sqrt{2\pi}} \int_{-\frac{Y}{\sigma}}^{\infty} e^{\sigma\gamma\omega - \frac{\omega^2}{2}} = \frac{e^{\frac{\gamma^2\sigma^2}{2}}}{\sqrt{2\pi}} \int_{-\frac{Y}{\sigma}}^{\infty} e^{-\frac{(\omega-\gamma\sigma)^2}{2}} = e^{\frac{\gamma^2\sigma^2}{2}} \cdot \Phi\left(\frac{Y}{\sigma} + \gamma\sigma\right)$$

   Hence,

$$\begin{aligned}\Xi_1(\sigma) &= 1 - \min_t \left(\mathbb{E}_Y \ln \Phi\left(\frac{Y}{\sigma} + t\right) + \frac{t^2}{2} \right) \\ &= 1 - \min_t \left(\mathbb{E}_{Z,\epsilon} \ln \Phi\left(\frac{|Z|^2}{\sigma} + \epsilon + t\right) + \frac{t^2}{2} \right)\end{aligned}$$

Note that $\Phi\left(\frac{|Z|^2}{\sigma}+\epsilon+t\right)$ increases as $\sigma \downarrow 0$. Consequently, we have, $\Xi_1(\sigma)$ is a decreasing function of σ.

3. Recall that in the previous step, we showed that,

$$\Xi_1(\sigma) = 1 - \min_t \left(\mathbb{E}_{Z,\epsilon} \ln \Phi \left(\frac{|Z|^2}{\sigma} + \epsilon + t \right) + \frac{t^2}{2} \right).$$

Proposition 22 shows that for any $\sigma > 0$, the objective in the definition of Ξ_1 is cooercive. Consequently we can identify $-\infty < t_1 < t_2 < \infty$ such that,

$$\left(\mathbb{E}_{Z,\epsilon} \ln \Phi \left(|Z|^2 + \epsilon + t \right) + \frac{t^2}{2} \right) > 0, \ \forall\, t \in (-\infty, t_1) \cup (t_2, \infty).$$

Since Φ is an increasing function,

$$\left(\mathbb{E}_{Z,\epsilon} \ln \Phi \left(\frac{|Z|^2}{\sigma} + \epsilon + t \right) + \frac{t^2}{2} \right) > 0, \ \forall\, t \in (-\infty, t_1) \cup (t_2, \infty), \ \forall \sigma \leq 1.$$

On the other hand,

$$\left(\mathbb{E}_{Z,\epsilon} \ln \Phi \left(\frac{|Z|^2}{\sigma} + \epsilon + t \right) + \frac{t^2}{2} \right) \Bigg|_{t=0} \leq 0.$$

Hence we have,

$$\Xi_1(\sigma) = 1 - \min_{t \in [t_1, t_2]} \left(\mathbb{E}_{Z,\epsilon} \ln \Phi \left(\frac{|Z|^2}{\sigma} + \epsilon + t \right) + \frac{t^2}{2} \right).$$

Observing that $\left(\mathbb{E}_{Z,\epsilon} \ln \Phi \left(\frac{|Z|^2}{\sigma} + \epsilon + t \right) + \frac{t^2}{2} \right)$ is a convex function such that for every fixed

t we have,

$$\lim_{\sigma \to 0} \left(\mathbb{E}_{Z,\epsilon} \ln \Phi \left(\frac{|Z|^2}{\sigma} + \epsilon + t \right) + \frac{t^2}{2} \right) = \frac{t^2}{2}.$$

Due to convexity, this convergence can be made uniform on compact sets:

$$\lim_{\sigma \to 0} \max_{t \in [t_1, t_2]} \left| \mathbb{E}_{Z,\epsilon} \ln \Phi \left(\frac{|Z|^2}{\sigma} + \epsilon + t \right) \right| = 0.$$

This uniform convergence immediately yields $\lim_{\sigma \to 0} \Xi_1(\sigma) = 1$.

□

The following lemma analyzes the convergence of $\Xi_2(q; \sigma)$ to $\Xi_2(q; 0)$. For our purposes, it turns out, that we don't need to show that $\Xi_2(q; \sigma) \to \Xi_2(q; 0)$ as $\sigma \to 0$. It is sufficient to show the weaker result that $\Xi_2(q; \sigma)$ is asymptotically lower bounded by $\Xi_2(q; 0)$ as $\sigma \to 0$. This is the content of the following lemma.

Lemma 40. *For any $0 < \eta < 1$, we have,*

$$\liminf_{\sigma \to 0} \min_{q \in [0, 1-\eta]} \Xi_2(q, \sigma) - \Xi_2(q; 0) \geq 0.$$

Furthermore we have,

$$\lim_{\sigma \to 0} \Xi_2(0, \sigma) = \Xi_2(0, 0).$$

Proof. We lower bound $\Xi_2(q, \sigma^2)$ as follows:

$$\Xi_2(q,\sigma) \stackrel{\text{def}}{=} \max_{(\lambda,\phi)\in\mathbb{R}} \left(2\lambda + \phi q - \mathbb{E}_Y \ln \mathbb{E} e^{((\lambda-1)(2Y+\sigma(\omega_1+\omega_2))+\phi\sqrt{(Y+\sigma\omega_1)(Y+\sigma\omega_2)}\cos(\theta)} \mathbf{1}_{Y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}\right)$$
$$\geq 2 + q\phi_2(q;0) - \mathbb{E}_Y \ln \mathbb{E}_{\omega_1,\omega_2} I_0\left(\phi_2(q;0)\cdot\sqrt{(Y+\sigma\omega_1)(Y+\sigma\omega_2)}\right)\mathbf{1}_{Y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}$$
$$= \Xi_2(q;0) + \mathbb{E}\ln I_0(\phi_2(q;0)|Z|^2)$$
$$- \mathbb{E}_Y \ln \mathbb{E}_{\omega_1,\omega_2} I_0\left(\phi_2(q;0)\sqrt{(Y+\sigma\omega_1)(Y+\sigma\omega_2)}\right)\mathbf{1}_{Y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}.$$

In the above display, we recall that $\phi_2(q;0)$ was defined as,

$$\phi_2(q;0) \stackrel{\text{def}}{=} \arg\max_{\phi\in\mathbb{R}} q\phi - \mathbb{E}_{Z\sim\mathcal{CN}(0,1)} \ln I_0(\phi|Z|^2).$$

Note that for any fixed $\phi \in \mathbb{R}$, we have, by Dominated convergence, as $\sigma \to 0$, we have,

$$\mathbb{E}_Y \ln \mathbb{E}_{\omega_1,\omega_2} I_0\left(\phi\sqrt{(Y+\sigma\omega_1)(Y+\sigma\omega_2)}\right)\mathbf{1}_{Y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0} \to \mathbb{E}_{Z\sim\mathcal{CN}(0,1)} \ln I_0(\phi|Z|^2).$$

Observing that the function on the left hand side is convex in ϕ we have the above convergence holds uniformly on all compact sets. Lemma 38 guarantees that $\sup_{q\in[0,1-\eta]} |\phi_2(q;0)| < \infty$. Consequently, we have,

$$\lim_{\sigma\to 0} \mathbb{E}_Y \ln \mathbb{E}_{\omega_1,\omega_2} I_0\left(\phi_2(q;0)\cdot\sqrt{(Y+\sigma\omega_1)(Y+\sigma\omega_2)}\right)\mathbf{1}_{Y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0} = \mathbb{E}\ln I_0(\phi_2(q;0)|Z|^2),$$

where the convergence is uniform on $q \in [0, 1-\eta]$. Combining this with the lower bound on $\Xi_2(q,\sigma)$ immediately gives:

$$\liminf_{\sigma\to 0} \min_{q\in[0,1-\eta]} \Xi_2(q,\sigma) - \Xi_2(q;0) \geq 0.$$

Finally when $q = 0$ we note that, $\Xi_2(0,\sigma) = 2\Xi_1(0,\sigma)$. Lemma 39 guarantees that $\Xi_2(0,\sigma) \to 1$ as $\sigma \to 0$. Note that since $I_0(x)$ is minimized at $x = 0$ (see Fact 3, Appendix B.9), we have

$\Xi_2(0,0) = 2$. Hence we indeed have $\Xi_2(0,\sigma) \to \Xi_2(0,0)$ as $\sigma \to 0$. $\square$

B.5.3 Proof of Proposition 11

Recall that our goal is to find conditions on (δ, Δ, σ) such that $\mathcal{F}(q;\delta,\Delta,\sigma) > \mathcal{F}(0;\delta,\Delta,\sigma)$, where,

$$\mathcal{F}(q;\delta,\Delta,\sigma) \stackrel{\text{def}}{=} \Xi_2(q;\sigma) - 2\Xi_1(\sigma) + \left(1 - \frac{1}{\delta}\right)\ln(1-q^2) + \Delta \ln\left(1 - \frac{q^2}{2}\right).$$

The following lemma provides a lower bound on the curvature of $\Xi_2(q;\sigma) - 2\Xi_1(\sigma)$ in the neighborhood of $q \approx 0$.

Lemma 41 (Analysis for $q \approx 0$)**.** *There exists a universal constant C (independent of σ) such that, for any $0 \le q < 1/2,\ \sigma < 1$ we have,*

$$\Xi_2(q,\sigma) - 2\Xi_1(\sigma) \ge (1-\sigma^2)\cdot\frac{q^2}{2} - Cq^3.$$

Proof. We can write $Z_{\mathsf{TWis}}(\lambda,\phi,Y)$ (c.f. Definition 6) as:

$$\begin{aligned}
&Z_{\mathsf{TWis}}(\lambda,\phi,y)\\
&\stackrel{\text{def}}{=} \frac{1}{2\pi}\int_0^\infty\int_0^\infty\int_{-\pi}^{\pi} \exp(-(1-\lambda)(s+s') + \phi\sqrt{ss'}\cos(\theta))\cdot\psi_\sigma(s-y)\cdot\psi_\sigma(s'-y)\,\mathrm{d}\theta\,\mathrm{d}s\,\mathrm{d}s'\\
&= \mathbb{E}\exp\Big((\lambda-1)(2y+\sigma(\omega_1+\omega_2)) + \phi\sqrt{(y+\sigma\omega_1)(y+\sigma\omega_2)}\cos(\theta)\Big)\mathbf{1}_{y+\sigma\omega_1\ge 0, y+\sigma\omega_2\ge 0}.
\end{aligned}$$

In the above display ω_1,ω_2,θ are independent r.v.s with distributions $\omega_1 \sim \mathcal{N}(0,1)$, $\omega_2 \sim \mathcal{N}(0,1)$, $\theta \sim \text{Uniform}[-\pi,\pi]$. We lower bound Ξ_2 as follows:

$$\begin{aligned}
\Xi_2(q,\sigma) &= \max_{(\lambda,\phi)\in\mathbb{R}} \big(2\lambda + \phi q - \mathbb{E}_Y \ln Z_{\mathsf{TWis}}(\lambda,\phi,Y)\big)\\
&\ge \Big(2\lambda_1(\sigma^2) + q^2 - \mathbb{E}_Y \ln Z_{\mathsf{TWis}}\big(\lambda_1(\sigma), q, Y\big)\Big).
\end{aligned}$$

Next we will approximate $\ln Z_{\mathsf{TWis}}\left(\lambda_1(\sigma), q, Y\right)$ by its taylor series around $q \approx 0$. We can compute the first three derivatives:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}q} \ln Z_{\mathsf{TWis}}\left(\lambda_1(\sigma), q, y\right)\Big|_{q=0} = 0, \\
&\frac{\mathrm{d}^2}{\mathrm{d}q^2} \ln Z_{\mathsf{TWis}}\left(\lambda_1(\sigma), q, Y\right)\Big|_{q=0} \\
&\qquad = \frac{\mathbb{E}(y+\sigma\omega_1)(y+\sigma\omega_2)\cos^2(\theta) e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}}{\mathbb{E}e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}} \\
&\qquad = \frac{1}{2}\cdot\frac{\mathbb{E}(y+\sigma\omega_1)(y+\sigma\omega_2) e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}}{\mathbb{E}e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}} \\
&\qquad = \frac{1}{2}\left(\frac{\mathbb{E}(y+\sigma\omega_1) e^{(\lambda_1(\sigma)-1)(y+\sigma\omega_1)}\mathbf{1}_{y+\sigma\omega_1\geq 0}}{\mathbb{E}e^{(\lambda_1(\sigma)-1)(y+\sigma\omega_1)}\mathbf{1}_{y+\sigma\omega_1\geq 0}}\right)^2 \\
&\qquad \overset{\text{(a)}}{\leq} \frac{1}{2}\cdot\left(\mathbb{E}(y+\sigma\omega_1)\mathbf{1}_{y+\sigma\omega_1\geq 0}\right)^2 \\
&\qquad \leq \frac{1}{2}\cdot\mathbb{E}(y+\sigma\omega_1)^2 \\
&\qquad = \frac{y^2}{2}+\frac{\sigma^2}{2}.
\end{aligned}$$

In the step marked (a), we used the fact that $\lambda_1(\sigma^2) \leq 1$ (see Lemma 39) and Chebychev's Association Inequality (Fact 2, Appendix B.9). Similarly we control the third derivative:

$$\frac{\mathrm{d}^3}{\mathrm{d}q^3} \ln Z_{\mathsf{TWis}}\left(\lambda_1(\sigma), q, Y\right) = T_3 - 3T_2T_1 + 2T_1^3,$$

where, for $i = 1, 2, 3$:

$$T_i \overset{\text{def}}{=} \frac{\mathbb{E}(y+\sigma\omega_1)^{\frac{i}{2}}(y+\sigma\omega_2)^{\frac{i}{2}}\cos^i(\theta) e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))+q\sqrt{(y+\sigma\omega_1)(y+\sigma\omega_2)}\cos(\theta)}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}}{\mathbb{E}e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))+q\sqrt{(y+\sigma\omega_1)(y+\sigma\omega_2)}\cos(\theta)}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}}.$$

We can control T_i as follows, for any $q \in [0,1]$ and any $\sigma \leq 1$, we have,

$$
\begin{aligned}
|T_i| &\leq \frac{\mathbb{E}(y+\sigma\omega_1)^{\frac{i}{2}}(y+\sigma\omega_2)^{\frac{i}{2}} e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))+q\sqrt{(y+\sigma\omega_1)(y+\sigma\omega_2)}\cos(\theta)}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}}{\mathbb{E}e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))+q\sqrt{(y+\sigma\omega_1)(y+\sigma\omega_2)}\cos(\theta)}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}} \\
&\overset{\text{(a)}}{=} \frac{\mathbb{E}(y+\sigma\omega_1)^{\frac{i}{2}}(y+\sigma\omega_2)^{\frac{i}{2}} e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))} I_0(q\sqrt{(y+\sigma\omega_1)(y+\sigma\omega_2)})\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}}{\mathbb{E}e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))} I_0(q\sqrt{(y+\sigma\omega_1)(y+\sigma\omega_2)})\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}} \\
&\overset{\text{(b)}}{\leq} \frac{\mathbb{E}(y+\sigma\omega_1)^{\frac{i}{2}}(y+\sigma\omega_2)^{\frac{i}{2}} e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))} e^{q\sqrt{(y+\sigma\omega_1)(y+\sigma\omega_2)}}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}}{\mathbb{E}e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}} \\
&\leq \frac{\mathbb{E}(y+\sigma\omega_1)^{\frac{i}{2}}(y+\sigma\omega_2)^{\frac{i}{2}} e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))} e^{\frac{q}{2}\cdot(y+\sigma\omega_1+y+\sigma\omega_2)}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}}{\mathbb{E}e^{(\lambda_1(\sigma)-1)(2y+\sigma(\omega_1+\omega_2))}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}} \\
&= \left(\frac{\mathbb{E}(y+\sigma\omega_1)^{\frac{i}{2}} e^{(\lambda_1(\sigma)-1+\frac{q}{2})(y+\sigma\omega_1)}\mathbf{1}_{y+\sigma\omega_1\geq 0}}{\mathbb{E}e^{(\lambda_1(\sigma)-1+\frac{q}{2})(y+\sigma\omega_1)}\mathbf{1}_{y+\sigma\omega_1\geq 0}}\right)^2 \\
&\overset{\text{(c)}}{\leq} \left(\mathbb{E}(y+\sigma\omega_1)^{\frac{i}{2}} e^{\frac{q}{2}(y+\sigma\omega_1)}\mathbf{1}_{y+\sigma\omega_1\geq 0}\right)^2 \\
&\leq \mathbb{E}|y+\sigma\omega_1|^i e^{q(y+\sigma\omega_1)} \\
&\leq C(|y|^3+1)e^{qy}.
\end{aligned}
$$

where, C is a universal constant independent of σ. In the step marked (a), we used the definition of Modified Bessel Function (see Fact 3, Appendix B.9). In the step marked (b), we used $1 \leq I_0(x) \leq e^x$ for any $x \in \mathbb{R}$. In the step marked (c) we recalled $\lambda_1(\sigma) \leq 1$ and applied Chebychev's Association Inequality. In conclusion, we obtained the following:

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}q}\ln Z_{\mathsf{TWis}}\left(\lambda_1(\sigma),q,y\right)\Big|_{q=0} &= 0, \\
\frac{\mathrm{d}^2}{\mathrm{d}q^2}\ln Z_{\mathsf{TWis}}\left(\lambda_1(\sigma),q,y\right)\Big|_{q=0} &\leq \frac{y^2+\sigma^2}{2}, \\
\frac{\mathrm{d}^3}{\mathrm{d}q^3}\ln Z_{\mathsf{TWis}}\left(\lambda_1(\sigma),q,y\right) &\leq C(|y|^3+1)e^{qy}.
\end{aligned}
$$

This allows us to upper bound $\ln Z_{\mathsf{TWis}}\left(\lambda_1(\sigma),q,Y\right)$ for any $q < 1/2, \sigma < 1$ as follows:

$$
\mathbb{E}\ln Z_{\mathsf{TWis}}\left(\lambda_1(\sigma),q,Y\right) \leq \mathbb{E}_Y\left[\ln Z_{\mathsf{TWis}}\left(\lambda_1(\sigma),0,Y\right) + \frac{q^2}{4}\cdot(\sigma^2+Y^2) + Cq^3(|Y|^3+1)e^{\frac{Y}{2}}\right].
$$

Observing that $\mathbb{E}Y^2 = \mathbb{E}|Z|^4 + \sigma^2 = 2 + \sigma^2$. We obtain,

$$\Xi_2(q,\sigma) - 2\Xi_1(\sigma) \geq \frac{q^2}{2}\left(1-\sigma^2\right) - Cq^3.$$

□

Next we show that at $q \to 1$, $\Xi_2(q;\sigma) - 2\Xi_1(\sigma) \to \infty$ in the following lemma.

Lemma 42 (Analysis at $q \approx 1$)**.** *There exists a universal finite constant $C > 0$ (independent of σ, q) such that, for all $\sigma \leq 1$,*

$$\Xi_2(q,\sigma) - 2\Xi_1(\sigma) \geq -C - \frac{\ln(1-q)}{2},\ \forall \sigma > 0,\ \forall\, q\ \in\ [0,1).$$

Proof. We lower bound Ξ_2 as follows:

$$\begin{aligned}
\Xi_2(q,\sigma) &\stackrel{\text{def}}{=} \max_{(\lambda,\phi)\in\mathbb{R}} \left(2\lambda + \phi q - \mathbb{E}_Y \ln Z_{\mathsf{TWis}}\left(\lambda,\phi,Y\right)\right) \\
&\geq -\frac{1}{2(1-q)} + \frac{q}{2(1-q)} - \mathbb{E}_Y\left[\ln Z_{\mathsf{TWis}}\left(-\frac{1}{4(1-q)}, \frac{1}{2(1-q)}, Y\right)\right] \\
&= -2 - \mathbb{E}_Y\left[\ln Z_{\mathsf{TWis}}\left(-\frac{1}{4(1-q)}, \frac{1}{2(1-q)}, Y\right)\right].
\end{aligned}$$

Recall that,

$$\begin{aligned}
&Z_{\mathsf{TWis}}\left(-\frac{1}{4(1-q)}, \frac{1}{2(1-q)}, y\right) = \mathbb{E}e^{-(2y+\sigma(\omega_1+\omega_2))\left(\frac{1}{4(1-q)}+1\right)+\frac{\sqrt{y+\sigma\omega_1}\sqrt{y+\sigma\omega_2}\cos(\theta)}{2(1-q)}}\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0} \\
&\stackrel{\text{(a)}}{\leq} e^{-\frac{y}{2(1-q)}}\cdot \mathbb{E}e^{-\frac{\sigma(\omega_1+\omega_2)}{4(1-q)}}\cdot I_0\left(\frac{\sqrt{y+\sigma\omega_1}\sqrt{y+\sigma\omega_2}}{2(1-q)}\cdot\right)\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0} \\
&\stackrel{\text{(b)}}{\leq} e^{-\frac{y}{2(1-q)}}\cdot \mathbb{E}e^{-\frac{\sigma(\omega_1+\omega_2)}{4(1-q)}}\cdot I_0\left(\frac{2y+\sigma\omega_1+\sigma\omega_2}{4(1-q)}\right)\mathbf{1}_{y+\sigma\omega_1\geq 0, y+\sigma\omega_2\geq 0}.
\end{aligned}$$

In the above display $\omega_2, \omega_2, \theta$ are independent with $\omega \sim \mathcal{N}(0,1)$, $\omega_2 \sim \mathcal{N}(0,1)$, $\theta \sim \text{Uniform}[-\pi,\pi]$. In the step marked (a), we used the definition of Bessel Function I_0 (see Fact 3). In the step marked (b), we used AM-GM Inequality and the fact that $I_0(x)$ is increasing on $x \geq 0$ (Fact 3). Further

applying the upper bound $I_0(x) \leq Cx^{-\frac{1}{2}} \cdot e^x, x \geq 0$ (see Fact 3), gives,

$$Z_{\mathsf{TWis}}\left(-\frac{1}{4(1-q)}, \frac{1}{2(1-q)}, y\right) \leq C \cdot \sqrt{1-q} \cdot \mathbb{E}\frac{1}{\sqrt{|2y+\sigma\omega_1+\sigma\omega_2|}}.$$

Hence we have,

$$\begin{aligned}
&\mathbb{E}_Y \ln Z_{\mathsf{TWis}}\left(-\frac{1}{4(1-q)}, \frac{1}{2(1-q)}, Y\right) = \mathbb{E}_{Z,\epsilon} \ln Z_{\mathsf{TWis}}\left(-\frac{1}{4(1-q)}, \frac{1}{2(1-q)}, |Z|^2+\epsilon\right) \\
&\leq \mathbb{E}_Z \ln \mathbb{E}_\epsilon Z_{\mathsf{TWis}}\left(-\frac{1}{4(1-q)}, \frac{1}{2(1-q)}, |Z|^2+\epsilon\right) \\
&\leq \ln C + \frac{\ln(1-q)}{2} + \mathbb{E}_Z \ln \mathbb{E}_{\epsilon,\omega_1,\omega_2} \frac{1}{\sqrt{|2y+\sigma\omega_1+\sigma\omega_2|}} \\
&\overset{\text{(c)}}{\leq} \ln C + \frac{\ln(1-q)}{2} + \ln(4) - \frac{1}{2}\mathbb{E}_Z \ln |Z|^2 \\
&\overset{\text{(d)}}{\leq} C + \frac{\ln(1-q)}{2}
\end{aligned}$$

In the step marked (c), we appealed to Lemma 48. In the step marked (d), we used the fact that $\mathbb{E} \ln |Z|^2 = \int_0^\infty \ln(r) e^{-r}\, \mathrm{d}r \approx -0.58$ is finite. Hence we have the lower bound on Ξ_2:

$$\Xi_2(q,\sigma) \geq -C - \frac{\ln(1-q)}{2}.$$

Lemma 39 shows that $\Xi_1(\sigma) \leq \Xi_1(1)$ which is an absolute constant, consequently,

$$\Xi_2(q,\sigma) - 2\Xi_1(\sigma) \geq -C - \frac{\ln(1-q)}{2}$$

□

We finally put together all the different auxiliary results we have established so far and prove Proposition 11 which is restated below for convenience.

Proposition 11. *Recall that $\mathcal{F}(q;\delta,\Delta,\sigma)$ was defined as:*

$$\mathcal{F}(q;\delta,\Delta,\sigma) = \Xi_2(q;\sigma) - 2\Xi_1(\sigma) + \left(1-\frac{1}{\delta}\right)\ln(1-q^2) + \Delta\ln\left(1-\frac{q^2}{2}\right).$$

For any δ and Δ that satisfy

$$1 \le \delta < 2,\ 0 < \Delta < \frac{2-\delta}{\delta},$$

there exists a critical value of the noise level $\sigma_c(\delta,\Delta) > 0$ such that, for any $0 < \sigma < \sigma_c(\delta,\Delta)$, we have

1. *The function $\mathcal{F}(q;\delta,\Delta,\sigma)$ has a unique minimum at $q = 0$ and $\mathcal{F}(q;\delta,\Delta,\sigma) > \mathcal{F}(0;\delta,\Delta,\sigma)$ for any $q \in (0,1)$.*
2. $\frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(q;\delta,\Delta,\sigma)\Big|_{q=0} > 0$.

Proof. We will prove the above claims in 3 steps: 1) Step 1: Analysis around $q \approx 0$, 2) Step 2: Analysis around $q \approx 1$ and 3) Step 3: Analysis for all other values of q.

Step 1: $q \approx 0$. Lemma 41 guarantees the existence of a universal constant $C_1 > 0$ independent of σ, δ, Δ such that, for any $q \in [0, 0.25],\ \sigma < 1$, we have,

$$\begin{aligned}\mathcal{F}(q;\delta,\Delta,\sigma) - \mathcal{F}(0;\delta,\Delta,\sigma) &\ge \frac{q^2}{2}\cdot\left(\frac{2-\delta}{\delta} - \Delta - \sigma^2\right) - C_1 q^3 - \left(1-\frac{1}{\delta}\right)q^4 - \frac{\Delta}{2}q^4 \\ &\ge \frac{q^2}{2}\cdot\left(\frac{2-\delta}{\delta} - \Delta - \sigma^2\right) - (C_1+2)\cdot q^3\end{aligned}$$

In particular ensuring that,

$$\sigma \le \frac{1}{2}\left(\frac{2-\delta}{\delta} - \Delta\right),\ q \le \frac{1}{8(C_1+2)}\cdot\left(\frac{2-\delta}{\delta} - \Delta\right),$$

gives us,

$$\mathcal{F}(q;\delta,\Delta,\sigma) \geq \frac{q^2}{8}\cdot\left(\frac{2-\delta}{\delta}-\Delta\right). \tag{B.43}$$

Note that $\mathcal{F}(0;\delta,\Delta,\sigma)=0$. Hence, (B.43) verifies claim (1) of the proposition for small q:

$$\mathcal{F}(q;\delta,\Delta,\sigma) \geq \mathcal{F}(0;\delta,\Delta,\sigma) + \frac{q^2}{8}\cdot\left(\frac{2-\delta}{\delta}-\Delta\right),\ \forall\, q\in[0,\eta_1(\delta,\Delta)],\ \sigma\leq\sigma_1(\delta,\Delta),$$

where,

$$\eta_1(\delta,\Delta) \stackrel{\text{def}}{=} \frac{1}{8(C_1+2)}\cdot\left(\frac{2-\delta}{\delta}-\Delta\right),\ \sigma_1(\delta,\Delta)\stackrel{\text{def}}{=}\frac{1}{2}\left(\frac{2-\delta}{\delta}-\Delta\right).$$

Furthermore since,

$$\frac{\mathrm{d}\mathcal{F}}{\mathrm{d}q}(q;\delta,\Delta,\sigma)\bigg|_{q=0}=0,$$

by Taylor's Theorem,

$$\mathcal{F}(q;\delta,\Delta,\sigma)=\mathcal{F}(0;\delta,\Delta,\sigma)+\frac{q^2}{2}\cdot\frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(q;\delta,\Delta,\sigma)\bigg|_{q=0}+o(q).$$

Comparing the above display with (B.43), gives us claim (2) of the proposition:

$$\frac{\mathrm{d}^2\mathcal{F}}{\mathrm{d}q^2}(q;\delta,\Delta,\sigma)\bigg|_{q=0}\geq\frac{1}{4}\cdot\left(\frac{2-\delta}{\delta}-\Delta\right)>0.$$

Step 2: $q\approx 1$**.** Lemma 42 guarantees the existence of a universal constant C_2 such that,

$$\Xi_2(q,\sigma)-2\Xi_1(\sigma)\geq -C_2-\frac{\ln(1-q)}{2},\ \forall\sigma>0,\ \forall\, q\ \in\ [0,1),\ \forall\sigma\leq 1.$$

Consequently,

$$\begin{aligned}\mathcal{F}(q;\delta,\Delta,\sigma) &\geq -\left(\frac{2-\delta}{2\delta}\right)\ln(1-q)+\left(1-\frac{1}{\delta}\right)\ln(1+q)+\Delta\ln\left(1-\frac{q^2}{2}\right)-C_2\\ &\geq -\left(\frac{2-\delta}{2\delta}\right)\ln(1-q)-(C_2+1).\end{aligned}$$

Hence we have,

$$\mathcal{F}(q;\delta,\Delta,\sigma)\geq 1\geq 0=\mathcal{F}(0;\delta,\Delta,\sigma)\ \forall\ q\in[1-\eta_2(\delta),1),\ \sigma\leq 1,$$

where,

$$\eta_2(\delta)=\exp\left(-\frac{\delta(C_2+2)}{2-\delta}\right)>0.$$

This verifies claim (1) of the proposition for large q.

Case 3: Other values of q. In Steps (1) and (2), we have verified Claim (1) for $q\in[0,\eta_1(\delta,\Delta)]\cup[\eta_2(\delta),1]$. Now we focus our attention to:

$$q\in[\eta_1(\delta,\Delta),1-\eta_2(\delta)].$$

Note that it is sufficient to show that,

$$f(q;\delta,\Delta,\sigma)\stackrel{\text{def}}{=}\Xi_2(q,\sigma)+\left(1-\frac{1}{\delta}\right)\ln(1-q^2)+\Delta\ln\left(1-\frac{q^2}{2}\right),$$

satisfies $f(q;\delta,\Delta,\sigma)>f(0;\delta,\Delta,\sigma)\ \forall q\in[\eta_1(\delta,\Delta),1-\eta_2(\delta)]$. In Lemma 38, we had shown that the function:

$$f(q;\delta,\Delta,0)\stackrel{\text{def}}{=}\Xi_2(q;0)+\left(1-\frac{1}{\delta}\right)\ln(1-q^2)+\Delta\ln\left(1-\frac{q^2}{2}\right),$$

is strictly increasing and has the property that $f(0;\delta,\Delta,0) < f(q;\delta,\Delta,0),\ \forall\ q \in (0,1)$. Consequently, $\eta_3(\delta,\Delta)$ defined below is strictly positive:

$$\eta_3(\delta,\Delta) \stackrel{\text{def}}{=} \min_{q\in[\eta_1(\delta,\Delta),1)} f(q;\delta,\Delta,0) - f(0;\delta,\Delta,0) \tag{B.44}$$

$$= f(\eta_1(\delta,\Delta);\delta,\Delta,0) - f(0;\delta,\Delta,0) > 0. \tag{B.45}$$

Furthermore, Lemma 40 shows that $f(q;\delta,\Delta,\sigma)$ is asymptotically lower bounded by $f(q;\delta,\Delta,0)$ in the following sense:

$$\liminf_{\sigma\to 0} \min_{q\in[0,1-\eta_2(\delta)]} f(q;\delta,\Delta,\sigma) - f(q;\delta,\Delta,0) \geq 0.$$

Furthermore, it also guarantees $f(0;\delta,\Delta,\sigma) \to f(0;\delta,\Delta,0)$ as $\sigma \to 0$. Consequently there exists $\sigma_3(\delta,\Delta) > 0$ such that,

$$f(q;\delta,\Delta,\sigma) - f(q;\delta,\Delta,0) \geq -\frac{\eta_3(\delta,\Delta)}{3},\ \forall q \in [0, 1-\eta_2(\delta)],\ \forall \sigma \leq \sigma_3(\delta,\Delta), \tag{B.46}$$

$$|f(0;\delta,\Delta,\sigma) - f(0;\delta,\Delta,0)| \leq \frac{\eta_3(\delta,\Delta)}{3}\ \forall \sigma \leq \sigma_3(\delta,\Delta). \tag{B.47}$$

Hence, $\forall q \in [\eta_1(\delta,\Delta), 1-\eta_2(\delta)]$,

$$\begin{aligned}
f(q;\delta,\Delta,\sigma) &\stackrel{\text{(B.46)}}{\geq} f(q;\delta,\Delta,0) - \frac{\eta_3(\delta,\Delta)}{3} \\
&\geq f(0;\delta,\Delta,0) + \big(f(q;\delta,\Delta,0) - f(0;\delta,\Delta,0)\big) - \frac{\eta_3(\delta,\Delta)}{3} \\
&\stackrel{\text{(B.44)}}{\geq} f(0;\delta,\Delta,0) + \frac{2\eta_3(\delta,\Delta)}{3} \\
&\stackrel{\text{(B.47)}}{\geq} f(0;\delta,\Delta,\sigma) + \frac{\eta_3(\delta,\Delta)}{3} \\
&> f(0;\delta,\Delta,\sigma).
\end{aligned}$$

This concludes the proof of the proposition. □

B.6 Properties of the Tilted Exponential and Wishart Distributions

B.6.1 Properties of the Tilted Exponential Distribution

The following lemma collects some properties of $\mathsf{TExp}(\lambda, y)$ random variables which were used to prove the local CLT given in Proposition 7.

Lemma 43 (Properties of $\mathsf{TExp}(\lambda, y)$ Distribution)**.** *Let $T \sim \mathsf{TExp}(\lambda, y)$. We have,*

1. *Moment Bounds: For any $k \in \mathbb{N}$ we have:*

$$\mathbb{E}|T|^k \leq C_k \left(|y|^k + |\lambda|^k + 1 \right).$$

In the above display, C_k is a universal constant independent of y, λ but depends on k.

2. *Decay of characteristic function: For any $t \in \mathbb{R}$*

$$|\mathbb{E}e^{itT}| \leq \frac{C(1 + |y| + |\lambda|)}{|t|},$$

where C is a constant independent of t, y, λ.

Proof. 1. Since $T \geq 0$, $|T| = T$. We first observe that,

$$\begin{aligned}\mathbb{E}T^k &= \mathbb{E}T^k \mathbf{1}_{T \leq |y| + \sigma^2|\lambda|} + \mathbb{E}T^k \mathbf{1}_{T \geq |y| + \sigma^2|\lambda|} \\ &\leq (|y| + \sigma^2|\lambda|)^k + \mathbb{E}T^k \mathbf{1}_{T \geq |y| + \sigma^2|\lambda|}.\end{aligned}$$

Let $E \sim \mathsf{Exp}(1)$. Using the formula for the density of $\mathsf{TExp}(\lambda, y)$ distribution in Definition 14, it is easy to see that,

$$\mathbb{E}T^k \mathbf{1}_{T \geq |y| + \sigma^2|\lambda|} = \frac{\mathbb{E}E^k e^{\lambda E} \psi_\sigma(E - y) \mathbf{1}_{E \geq |y| + \sigma^2|\lambda|}}{\mathbb{E}e^{\lambda E} \psi_\sigma(E - y)}.$$

We observe that $f(e) = e^k$ is increasing and $g(e) = e^{\lambda e}\psi_\sigma(e - y)$ is decreasing when $e \geq |y| + \sigma^2|\lambda|$. Consequently by Chebychev's Association Inequality (Lemma 2, Appendix

B.9) we obtain,

$$\mathbb{E}T^3\mathbf{1}_{T\geq|y|+\sigma^2|\lambda|} \leq \mathbb{E}E^k = k!.$$

Hence for a suitable constant C_k, independent of λ, y we have,

$$\mathbb{E}T^k \leq C_k(|y|^k + |\lambda|^k + 1).$$

2. Let $f(u)$ denote the pdf of $\mathsf{TExp}\,(\lambda, y)$. We bound the characteristic function as follows:

$$\begin{aligned}
|\mathbb{E}e^{itT}| &= \left|\int_0^\infty e^{itu} f(u)\,\mathrm{d}u\right| \\
&= \frac{1}{|t|}\left|\int_0^\infty \frac{\mathrm{d}}{\mathrm{d}u} e^{itu} f(u)\,\mathrm{d}u\right| \\
&= \frac{1}{|t|}\cdot\left|-f(0) + \int_0^\infty f'(u)e^{itu}\,\mathrm{d}u\right| \\
&\leq \frac{f(0) + \|f'\|_1}{|t|}.
\end{aligned}$$

We further upper bound $\|f'\|_1$. Note that:

$$f'(u) = f(u)\cdot\left(\lambda - 1 - \frac{u}{\sigma^2}\right).$$

Consequently, for a suitable constant C (independent of λ, y) we obtain the estimate,

$$\begin{aligned}
\|f'\|_1 &\leq |\lambda| + 1 + \frac{\mathbb{E}T}{\sigma^2} \\
&\leq C(1 + |y| + |\lambda|).
\end{aligned}$$

In the last step, we used the estimate on $\mathbb{E}T$ from part (1) of this lemma. Next we upper

bound $f(0)$. Note that,

$$\begin{aligned}
f(0) &= \left(\int_0^\infty \exp\left(u\left(\lambda - 1 + \frac{y}{\sigma^2}\right) - \frac{u^2}{2\sigma^2} \right) \mathrm{d}u \right)^{-1} \\
&\leq \left(\int_0^1 \exp\left(u\left(\lambda - 1 + \frac{y}{\sigma^2}\right) - \frac{u^2}{2\sigma^2} \right) \mathrm{d}u \right)^{-1} \\
&\overset{\text{(a)}}{\leq} \left(\int_0^1 \exp\left(u\left(\lambda - 1 + \frac{y}{\sigma^2} - \frac{1}{2\sigma^2}\right) \right) \mathrm{d}u \right)^{-1} \\
&\leq \left(\int_0^{(|\lambda|+|y|/\sigma^2+1+1/2\sigma^2)^{-1}/2} \exp\left(u\left(\lambda - 1 + \frac{y}{\sigma^2} - \frac{1}{2\sigma^2}\right) \right) \mathrm{d}u \right)^{-1} \\
&\overset{\text{(b)}}{\leq} \left(\int_0^{(|\lambda|+|y|/\sigma^2+1+1/2\sigma^2)^{-1}/2} \left(1 + u\left(\lambda - 1 + \frac{y}{\sigma^2} - \frac{1}{2\sigma^2}\right) \right) \mathrm{d}u \right)^{-1} \\
&\overset{\text{(c)}}{\leq} C(|\lambda| + |y| + 1).
\end{aligned}$$

In the step marked (a) we used $u^2 \leq u,\ u \in (0,1)$. In the step marked (b) we used the lower bound $e^x \geq 1 + x$. Finally in the step marked (c), we observed that the integrand is larger than $1/2$ in the domain of integration. Combining the bounds on $f(0)$ and $\|f'\|_1$ gives us the required result:

$$|\mathbb{E}e^{itT}| \leq \frac{C(1 + |y| + |\lambda|)}{|t|}.$$

□

B.6.2 Properties of the Tilted Wishart Distribution

The following lemma collects some properties of the tilted Wishart distribution which were used to prove the Local CLT given in Proposition 8.

Lemma 44. *Suppose that:*

$$\boldsymbol{S} = \begin{bmatrix} r & \sqrt{rr'}e^{\mathbf{i}\theta} \\ \sqrt{rr'}e^{-\mathbf{i}\theta} & r_2 \end{bmatrix} \sim \mathsf{TWis}\left(\lambda, \phi, y\right)$$

Then, there exists a universal constant $0 < C < \infty$ depending only on σ such that:

1. *Equivalent Characterization: For any bounded measurable function f we have,*

$$\mathbb{E}f(\boldsymbol{S}) = \frac{\mathbb{E}_{\boldsymbol{g}\sim\mathcal{CN}(\mathbf{0},\boldsymbol{I}_2)} f(\boldsymbol{g}\boldsymbol{g}^{\mathsf{H}}) e^{\langle \boldsymbol{\Lambda}, \boldsymbol{g}\boldsymbol{g}^{\mathsf{H}}\rangle} \psi_\sigma(y - |g_1|^2)\psi_\sigma(y-|g_2|^2)}{\mathbb{E}_{\boldsymbol{g}\sim\mathcal{CN}(\mathbf{0},\boldsymbol{I}_2)} e^{\langle \boldsymbol{\Lambda}, \boldsymbol{g}\boldsymbol{g}^{\mathsf{H}}\rangle} \psi_\sigma(y - |g_1|^2)\psi_\sigma(y-|g_2|^2)}.$$

   In the above display, $\boldsymbol{\Lambda}$ is given by:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda & \phi/2 \\ \phi/2 & \lambda \end{bmatrix}.$$

2. *The density:*

$$\tilde{h}_{\lambda,\phi,y}(g, g') = \frac{e^{-(1-\lambda)(|g|+|g'|^2)+\phi\mathrm{Re}(g\overline{g}')}\psi_\sigma(|g|^2 - y)\psi_\sigma(|g'|^2-y)}{Z_{\mathsf{TWis}}\left(\lambda,\phi,y\right)},$$

   on $\mathbb{C}^2$ is locally bounded, that is:

$$\tilde{h}_{\lambda,\phi,y}(g_0, g_0') \le C(1 + |\lambda|^4 + |\phi|^4 + |y|^4)(1 + |g_0|^{12} + |g_0'|^{12}).$$

3. *Tail Bound: With probability $1-\epsilon$,*

$$r \le C\sqrt{1 + |y|^2 + |\phi|^2 + |\lambda|^2} + C\sqrt{\ln\frac{1}{\epsilon}}.$$

   The analogous result holds for r'.

4. *Moment Bounds: For any $k \in \mathbb{N}$, There exists a universal constant C_k depending only on k*

such that

$$\mathbb{E}r^k \le C(1 + |\lambda|^k + |\phi|^k + |y|^k).$$

5. *Decay of Characteristic Function:*

$$\left|\mathbb{E}e^{\mathbf{i}\langle \boldsymbol{T}, \boldsymbol{S}\rangle}\right| \le \frac{C \cdot (1 + |\lambda|^{20} + |\phi|^{20} + |y|^{20})}{\|\boldsymbol{T}\|^{\frac{1}{3}}}.$$

6. *For any $y, \lambda, \phi \in \mathbb{R}$, we have,*

$$0 < \lambda_{\min}\left(\boldsymbol{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, y)\right) < \lambda_{\max}\left(\boldsymbol{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, y)\right) < \infty.$$

Proof. Throughout this proof, we use C to denote constants that depend only on the noise level σ and in particular are independent of the parameters λ, ϕ, y.

1. We write $g_1 = \sqrt{r}e^{\mathbf{i}\omega}, g_2 = \sqrt{r'}e^{\mathbf{i}\omega'}$. Using standard properties of the complex gaussian distribution, we know that $r, r' \sim \mathsf{Exp}(1)$ and $\omega, \omega' \sim \text{Unif}(-\pi, \pi]$. Consequently,

$$\boldsymbol{g}\boldsymbol{g}^{\mathsf{H}} = \begin{bmatrix} r & \sqrt{rr'}e^{\mathbf{i}(\omega-\omega')} \\ \sqrt{rr'}e^{\mathbf{i}(\omega'-\omega)} & r' \end{bmatrix}$$

   Let $\theta \sim \text{Unif}(-\pi, \pi]$. Then we have $e^{\mathbf{i}(\omega-\omega')} \overset{\mathrm{d}}{=} e^{\mathbf{i}\theta}$. Consequently,

$$\begin{aligned}&\mathbb{E}f(\boldsymbol{g}\boldsymbol{g}^{\mathsf{H}})e^{\langle \boldsymbol{\Lambda}, \boldsymbol{g}\boldsymbol{g}^{\mathsf{H}}\rangle}\psi_\sigma(y - |g_1|^2)\psi_\sigma(y - |g_2|^2) \\ &\quad = \frac{1}{2\pi}\int_0^\infty \int_0^\infty \int_{-\pi}^{\pi} f(r, r', \theta)e^{-(1-\lambda)(r+r') + \phi\sqrt{rr'}\cos(\theta)}\psi_\sigma(y - r)\psi_\sigma(y - r')\,\mathrm{d}r\,\mathrm{d}r'\,\mathrm{d}\theta.\end{aligned}$$

   Comparing this with the density of $\boldsymbol{S}$ from Definition 6 gives us the claim of item (1). Note that the Tilted Wishart distribution is supported on rank-1 Hermitian matrices. In particular, it does not have a density with respect to the Lebesgue measure on Hermitian matrices. The

advantage of the alternate way of computing expectations of functions of $\boldsymbol{S}$ is that they can be computed by integrating with respect to the proper PDF $\tilde{h}_{\lambda,\phi,y}(g,g')$ on $\mathbb{C}^2$:

$$\mathbb{E}f(\boldsymbol{S}) = \int_{\mathbb{C}^2} f\left(\begin{bmatrix} |g|^2 & g\overline{g}' \\ \overline{g}g' & |g'|^2 \end{bmatrix}\right) \tilde{h}_{\lambda,\phi,y}(g,g')\,\mathrm{d}g\,\mathrm{d}g',$$

where the pdf $\tilde{h}_{\lambda,\phi,y}(g,g')$ is given by:

$$\tilde{h}_{\lambda,\phi,y}(g,g') = \frac{e^{-(1-\lambda)(|g|+|g'|^2)+\phi\mathrm{Re}(g\overline{g}')}\psi_\sigma(|g|^2-y)\psi_\sigma(|g'|^2-y)}{Z_{\mathsf{TWis}}(\lambda,\phi,y)}.$$

This density function is much nicer, in particular it is locally bounded.

2. We first note that $\ln\tilde{h}_{\lambda,\phi,y}$ is a degree 4 polynomial in g,g'. Consequently it is local Lipchitz, that is,

$$\begin{aligned}&|\ln\tilde{h}_{\lambda,\phi,y}(g,g') - \ln\tilde{h}_{\lambda,\phi,y}(g_0,g_0')| \\ &\qquad\leq C(1+|g|+|g'|+|g_0|+|g_0'|)^3(1+|\lambda|+|\phi|+|y|)(|g-g_0|+|g'-g_0'|).\end{aligned}$$

In particular this means that there exists a large enough constant C depending only on σ such that,

$$\forall g,g':\ \max(|g-g_0|,|g'-g_0'|)\leq R,\ |\ln\tilde{h}_{\lambda,\phi,y}(g,g') - \ln\tilde{h}_{\lambda,\phi,y}(g_0,g_0')|\leq\ln(2),$$

where,

$$R \stackrel{\text{def}}{=} \frac{1}{C(1+|\lambda|+|\phi|+|y|)(1+|g_0|^3+|g_0'|^3)}.$$

We can use this to show that $\tilde{h}_{\lambda,\phi,y}$ is locally bounded. We have:

$$\begin{aligned}
\tilde{h}_{\lambda,\phi,y}(g_0, g_0') &= \frac{\tilde{h}_{\lambda,\phi,y}(g_0, g_0')}{\int_{\mathbb{C}^2} \tilde{h}_{\lambda,\phi,y}(g, g')\,\mathrm{d}g\,\mathrm{d}g'} \\
&= \left(\int_{\mathbb{C}^2} \exp\left(\ln \tilde{h}_{\lambda,\phi,y}(g, g') - \ln \tilde{h}_{\lambda,\phi,y}(g_0, g_0')\right) \mathrm{d}g\,\mathrm{d}g'\right)^{-1} \\
&\leq \left(\int_{|g-g_0|\leq R, |g'-g_0'|\leq R} \exp\left(\ln \tilde{h}_{\lambda,\phi,y}(g, g') - \ln \tilde{h}_{\lambda,\phi,y}(g_0, g_0')\right) \mathrm{d}g\,\mathrm{d}g'\right)^{-1} \\
&\leq \left(\frac{1}{2}\int_{|g-g_0|\leq R, |g'-g_0'|\leq R} \mathrm{d}g\,\mathrm{d}g'\right)^{-1} \\
&= \frac{2}{\pi^2 R^4} \\
&\leq C(1 + |\lambda|^4 + |\phi|^4 + |y|^4)(1 + |g_0|^{12} + |g_0'|^{12}).
\end{aligned}$$

3. We begin by computing the log-mgf of r:

$$\ln \mathbb{E}e^{tr} = A - B,$$

where,

$$\begin{aligned}
A &\stackrel{\text{def}}{=} \left(\frac{1}{2\pi}\int_0^\infty \int_0^\infty \int_{-\pi}^{\pi} e^{(\lambda-1+t)r+(\lambda-1)r'+\phi\sqrt{rr'}\cos(\theta)} \psi_\sigma(r_1 - y)\psi_\sigma(r_2 - y)\,\mathrm{d}\theta\,\mathrm{d}r\,\mathrm{d}r'\right) \\
B &\stackrel{\text{def}}{=} \ln Z_{\mathsf{TWis}}(\lambda, \phi, y).
\end{aligned}$$

We upper bound (A):

$$\begin{aligned}
A &= \frac{1}{2\pi}\int_0^\infty \int_0^\infty \int_{-\pi}^{\pi} e^{(\lambda-1+t)r+(\lambda-1)r'+\phi\sqrt{rr'}\cos(\theta)} \psi_\sigma(r_1 - y)\psi_\sigma(r_2 - y)\,\mathrm{d}\theta\,\mathrm{d}r\,\mathrm{d}r' \\
&\stackrel{(a)}{\leq} \int_0^\infty \int_0^\infty e^{(\lambda-1+t+|\phi|)r+(\lambda-1+|\phi|)r'} \psi_\sigma(r_1 - y)\psi_\sigma(r_2 - y)\,\mathrm{d}r\,\mathrm{d}r' \\
&\leq \int_{-\infty}^\infty \int_{-\infty}^\infty e^{(|\lambda|+1+|t|+|\phi|)(r+r')} \psi_\sigma(r_1 - y)\psi_\sigma(r_2 - y)\,\mathrm{d}r\,\mathrm{d}r' \\
&\stackrel{(b)}{=} \exp\left(2y(|\lambda| + 1 + |t| + |\phi|) + \sigma^2(|\lambda| + 1 + |t| + |\phi|)^2\right).
\end{aligned}$$

In the step marked (a), we used the fact the fact that $\sqrt{rr'}\cos(\theta) \leq r + r'$. In the step marked (b) we used the formula for the MGF of a gaussian distribution. Hence, there exists a universal constant C depending only on σ such that:

$$\ln A \leq C\left(1 + |\lambda|^2 + |\phi|^2 + |y|^2 + |t|^2\right).$$

Next we upper bound (B):

$$\begin{aligned}
B &= \ln\left(\frac{1}{2\pi}\int_0^\infty\int_0^\infty\int_{-\pi}^{\pi} e^{(\lambda-1+t)r+(\lambda-1)r'+\phi\sqrt{rr'}\cos(\theta)}\psi_\sigma(r-y)\psi_\sigma(r'-y)\,\mathrm{d}\theta\,\mathrm{d}r\,\mathrm{d}r'\right) \\
&\overset{(c)}{\geq} \frac{1}{2\pi}\int_0^\infty\int_0^\infty\int_{-\pi}^{\pi}\left((\lambda+t)r + \lambda r' + \phi\sqrt{rr'}\cos(\theta) + \ln\psi_\sigma(r-y) + \ln\psi_\sigma(r'-y)\right)e^{-r-r'} \\
&\overset{(d)}{\geq} -C\left(1 + |y|^2 + |\lambda| + |t|\right)
\end{aligned}$$

In the step marked (c) we applied Jensen's inequality and in the step marked (d) we used performed the integration involving the moments of the $\mathsf{Exp}\,(1)$ distribution and used straightforward algebraic bounds. This gives us:

$$\ln \mathbb{E}e^{tr} \leq C\left(1 + |\lambda|^2 + |\phi|^2 + |y|^2 + |t|^2\right).$$

For notational convenience we define:

$$\kappa \overset{\text{def}}{=} 1 + |\lambda|^2 + |\phi|^2 + |y|^2.$$

By Markov's Inequality we have,

$$\mathbb{P}(r > x) \leq \exp\left(C\kappa - \frac{x^2}{4C}\right).$$

Setting the tail probability to ϵ gives us the result:

$$\mathbb{P}\left(r > C\sqrt{1+|y|^2+|\phi|^2+|\lambda|^2} + C\sqrt{\ln\frac{1}{\epsilon}}\right) \leq \epsilon,$$

for a suitable constant C.

4. Integrating the tail bound obtained above gives us the moment bound:

$$\begin{aligned}
\mathbb{E}r^k &= k\int_0^\infty x^{k-1}\mathbb{P}(r>x)\,\mathrm{d}x \\
&= k\left(\int_0^{2C\sqrt{\kappa}} x^{k-1}\cdot 1\,\mathrm{d}x + \int_{2C\sqrt{\kappa}}^\infty x^{k-1}\cdot e^{C\kappa-\frac{x^2}{4C}}\right) \\
&\leq C_k\cdot\kappa^{\frac{k-1}{2}} + \int_{2C\sqrt{\kappa}}^\infty \exp\left(C\kappa-\frac{x^2}{4C}\right)\cdot k\cdot x^{k-1}\,\mathrm{d}x \\
&\overset{(a)}{\leq} C_k\cdot\kappa^{\frac{k-1}{2}} \\
&\leq C_k(1+|\lambda|^k+|\phi|^k+|y|^k).
\end{aligned}$$

In the step marked (a), we used the a bound on the truncated gaussian integral given in Lemma 47 in Appendix B.9.

5. Let $\boldsymbol{T}$ be a 2×2 Hermitian matrix. The characteristic function of the Tilted Wishart distribution evaluated at $\boldsymbol{T}$ is given by $\mathbb{E}e^{\mathbf{i}\langle \boldsymbol{T},\boldsymbol{S}\rangle}$. Let $\boldsymbol{w}\in\mathbb{C}^2$ be a random vector sampled from the pdf $h_{\lambda,\phi,y}$:

$$\boldsymbol{w} = \begin{bmatrix} w \\ w' \end{bmatrix} \sim h_{\lambda,\phi,y}.$$

Using the alternate characterization derived in item (1) of this lemma we have,

$$\mathbb{E}e^{\mathbf{i}\langle \boldsymbol{T},\boldsymbol{S}\rangle} = \mathbb{E}e^{\mathbf{i}\langle \boldsymbol{T},\boldsymbol{w}\boldsymbol{w}^{\mathsf{H}}\rangle}$$

Consider the spectral decomposition of $\boldsymbol{T}$:

$$\boldsymbol{T} = \boldsymbol{B} \begin{bmatrix} t & 0 \\ 0 & t' \end{bmatrix} \boldsymbol{B}^{\mathsf{H}}, \ \boldsymbol{B} = \begin{bmatrix} \boldsymbol{b}_1^{\mathsf{H}} \\ \boldsymbol{b}_2^{\mathsf{H}} \end{bmatrix}.$$

Since we have,

$$\|\boldsymbol{T}\|^2 = t^2 + t'^2 \implies \max(|t|, |t'|) \geq \frac{\|\boldsymbol{T}\|}{\sqrt{2}}.$$

We will assume that infact,

$$|t| \geq \frac{\|\boldsymbol{T}\|}{\sqrt{2}}.$$

Define the random vector:

$$\boldsymbol{z} = \begin{bmatrix} z \\ z' \end{bmatrix} = \boldsymbol{B}^{\mathsf{H}} \boldsymbol{g}.$$

We will often use the polar representation of $\boldsymbol{z}$:

$$\boldsymbol{z} = \begin{bmatrix} s e^{\mathbf{i}\nu} \\ s' e^{\mathbf{i}\nu'} \end{bmatrix}$$

Let $d(z, z')$ denote the density of $\boldsymbol{z}$. This density can be obtained by a simple unitary transformation of the density $h_{\lambda,\phi,y}$. While the exact formula is complicated it is easy to see that it is of the form:

$$d(s e^{\mathbf{i}\nu}, s' e^{i\nu'}) = \frac{\exp\left(\sum_{k,l:k+l\leq 4} a_{k,l}(\nu, \nu') s^k s'^l\right)}{Z_{\mathsf{TWis}}(\lambda, \phi, y)}$$

The exact formula for the coefficients $a_{k,l}(\nu)$ is not important. It is sufficient to see they

satisfy the bound:

$$|a_{k,l}(\nu,\nu')| \leq C(1+|y|+|\lambda|+|\phi|).$$

We can now analyze the decay of the characteristic function:

$$\begin{aligned}
&|\mathbb{E}e^{\mathbf{i}\langle \boldsymbol{T},\boldsymbol{S}\rangle}| = |\mathbb{E}\exp\left(\mathbf{i}t|z|^2+\mathbf{i}t'|z'|^2\right)| \\
&\leq |\mathbb{E}e^{\mathbf{i}t|z|^2+\mathbf{i}t'|z'|^2}\mathbf{1}_{|z|\leq\epsilon,|z'|<R}| + |\mathbb{E}e^{\mathbf{i}t|z|^2+\mathbf{i}t'|z'|^2}\mathbf{1}_{|z|>\epsilon,|z'|<R}| + |\mathbb{E}e^{\mathbf{i}t|z|^2+\mathbf{i}t'|z'|^2}\mathbf{1}_{|z'|>R}| \\
&\leq \underbrace{\mathbb{P}(|z|\leq\epsilon,|z'|\leq R)}_{(\mathsf{I})} + \underbrace{|\mathbb{E}\exp\left(\mathbf{i}t|z|^2+\mathbf{i}t'|z'|^2\right)\mathbf{1}_{|z|>\epsilon,|z'|<R}|}_{(\mathsf{II})} + \underbrace{\mathbb{P}(|z'|>R)}_{(\mathsf{III})}
\end{aligned}$$

In the above display $0<\epsilon<1<R$ are parameters which will be chosen later. We analyze the terms (I), (II) and (III) separately.

Analysis of (I)**:** Recall that in part (2) of this Lemma we had shown the density of $\boldsymbol{w}$ is locally bounded:

$$h_{\lambda,\phi,y}(w,w') \leq C(1+|\lambda|^4+|\phi|^4+|y|^4)(1+|w|^{12}+|w'|^{12}).$$

The density of z,z', denoted by $d(z,z')$ is a unitary transformation of the density $h_{\lambda,\phi,\phi}$. Consequently, we have the estimate:

$$d(z,z') \leq C(1+|\lambda|^4+|\phi|^4+|y|^4)\cdot R^{12},\ \forall |z|\leq\epsilon,\ |z'|\leq R. \tag{B.48}$$

Using this we can easily bound A:

$$\begin{aligned}
(1) = \mathbb{P}(|z|\leq\epsilon,|z'|\leq R) &\leq C(1+|\lambda|^4+|\phi|^4+|y|^4)\cdot R^{12}\cdot\epsilon^2\cdot R^2 \\
&\leq C\cdot(1+|\lambda|^4+|\phi|^4+|y|^4)R^{14}\cdot\epsilon^2.
\end{aligned}$$

Analysis of (II)**:** Recall that term B was given by:

$$\begin{aligned}
(\mathsf{II}) &= \left|\mathbb{E}\exp\left(\mathbf{i}t|z|^2 + \mathbf{i}t'|z'|^2\right)\mathbf{1}_{|z|>\epsilon,|z'|<R}\right| \\
&= \left|\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{0}^{R}\int_{\epsilon}^{\infty}\exp(\mathbf{i}ts^2 + \mathbf{i}t's'^2)d(se^{\mathbf{i}\nu}, s'e^{\mathbf{i}\nu'})ss'\,\mathrm{d}s\,\mathrm{d}s'\,\mathrm{d}\nu\,\mathrm{d}\nu'\right| \\
&= \frac{1}{|t|}\left|\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{0}^{R}\int_{\epsilon}^{\infty}\frac{\partial\exp(\mathbf{i}ts^2 + \mathbf{i}t's'^2)}{\partial s}d(se^{\mathbf{i}\nu}, s'e^{\mathbf{i}\nu'})s'\,\mathrm{d}s\,\mathrm{d}s'\,\mathrm{d}\nu\,\mathrm{d}\nu'\right| \\
&\overset{(a)}{\leq} (\mathsf{IIa}) + (\mathsf{IIb}).
\end{aligned}$$

In the step marked (a), we applied integration by parts and defined the terms $(\mathsf{IIa}), (\mathsf{IIb})$ as follows:

$$\begin{aligned}
(\mathsf{IIa}) &= \frac{1}{|t|}\left|\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{0}^{R}e^{\mathbf{i}t\epsilon^2 + \mathbf{i}t's'^2}d(\epsilon e^{\mathbf{i}\nu}, s'e^{\mathbf{i}\nu'})s'\,\mathrm{d}s'\,\mathrm{d}\nu\,\mathrm{d}\nu'\right|. \\
(\mathsf{IIb}) &= \frac{1}{|t|}\left|\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{0}^{R}\int_{\epsilon}^{\infty}e^{\mathbf{i}ts^2 + \mathbf{i}t's'^2}\frac{\partial}{\partial s}d(se^{\mathbf{i}\nu}, s'e^{\mathbf{i}\nu'})s'\,\mathrm{d}s\,\mathrm{d}s'\,\mathrm{d}\nu\,\mathrm{d}\nu'\right|.
\end{aligned}$$

The previously obtained bound on $d(z, z')$ immediately gives the following bound:

$$(\mathsf{IIa}) \leq \frac{C}{|t|}\cdot(1 + |\lambda|^4 + |\phi|^4 + |y|^4)\cdot R^{14}.$$

We can control (IIb) as follows:

$$
\begin{aligned}
(\mathsf{IIb}) &\overset{\text{(b)}}{\leq} \frac{1}{|t|}\left(\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{0}^{R}\int_{\epsilon}^{\infty}\left|\frac{\partial}{\partial s}d(se^{\mathbf{i}\nu}, s'e^{\mathbf{i}\nu'})\right| s'\,\mathrm{d}s\,\mathrm{d}s'\,\mathrm{d}\nu\,\mathrm{d}\nu'\right)\\
&\overset{\text{(c)}}{=} \frac{1}{|t|}\left(\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{0}^{R}\int_{\epsilon}^{\infty}\left|\sum_{k+l\leq 4} a_{ij}(\nu,\nu')s^{k-1}s'^{l}\right| d(se^{\mathbf{i}\nu}, s'e^{\mathbf{i}\nu'})s'\,\mathrm{d}s\,\mathrm{d}s'\,\mathrm{d}\nu\,\mathrm{d}\nu'\right)\\
&\leq \frac{1}{|t|}\left(\frac{1}{\epsilon}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\int_{0}^{R}\int_{\epsilon}^{\infty}\left|\sum_{k+l\leq 4} a_{ij}(\nu,\nu')s^{k-1}s'^{l}\right| d(se^{\mathbf{i}\nu}, s'e^{\mathbf{i}\nu'})ss'\,\mathrm{d}s\,\mathrm{d}s'\,\mathrm{d}\nu\,\mathrm{d}\nu'\right)\\
&\overset{\text{(d)}}{\leq} \frac{C(1+|\lambda|^4+|\phi|^4+|y|^4)}{|t|}\cdot\left(\frac{1}{\epsilon}\sum_{k+l\leq 3}\mathbb{E}|z|^k|z'|^l\right)\\
&\overset{\text{(e)}}{\leq} \frac{C(1+|\lambda|^4+|\phi|^4+|y|^4)}{|t|}\cdot\left(\frac{1+|\lambda|^2+|\phi|^2+|y|^2}{\epsilon}\right)
\end{aligned}
$$

In the step marked (b) we used the local Lipchitz bound on d. In the step marked (c) we recalled the formula for the density d (see (B.48)). In the step marked (d) we used the bound on the coefficients $a_{k,l}(\nu)$. In the step marked (e) we used the fact that the random vector $\boldsymbol{z}$ is a unitary transformation of $\boldsymbol{w}$ and the third moment of $\boldsymbol{w}$ was bounded in item (4) of this lemma. Combining the bounds on $\mathsf{IIa}, \mathsf{IIb}$ we obtain,

$$
(\mathsf{II}) \leq \frac{C(1+|\lambda|^4+|\phi|^4+|y|^4)}{\|\boldsymbol{T}\|}\cdot\left(R^{14}+\frac{1+|\lambda|^2+|\phi|^2+|y|^2}{\epsilon}\right).
$$

Analysis of (III)**:** We have:

$$\begin{aligned}(\mathsf{III}) &= \mathbb{P}\left[|z'| \geq R\right] \\ &\leq \mathbb{P}\left[\|\boldsymbol{z}\| \geq R\right] \\ &\overset{\text{(f)}}{\leq} \mathbb{P}\left[\|\boldsymbol{w}\| \geq R\right] \\ &\leq \mathbb{P}\left[|w| \geq R\right] + \mathbb{P}\left[|w'| \geq R\right] \\ &\overset{\text{(g)}}{\leq} 2\mathbb{P}\left[|w| \geq R\right].\end{aligned}$$

In the step marked (f) we used the fact that since $\boldsymbol{z}$ is a unitary transformation of $\boldsymbol{w}$, we have $\|\boldsymbol{z}\| = \|\boldsymbol{w}\|$. In the step marked (g) we used the fact that $|w|, |w|'$ are identically distributed. Finally we set R as:

$$R = C\sqrt{1 + |y|^2 + |\phi|^2 + |\lambda|^2} + C\sqrt{\ln \frac{1}{\epsilon}},$$

and apply the concentration inequality from item (3) of this lemma to obtain:

$$(3) \leq 2\epsilon.$$

Combining the bounds on $(1), (2)$ and (3) and setting $\epsilon = O(1/\|\boldsymbol{T}\|^{1/2})$ gives us the final bound on the characteristic function of $\boldsymbol{S}$:

$$\begin{aligned}\left|\mathbb{E}e^{\mathbf{i}\langle \boldsymbol{\Lambda}, \boldsymbol{S}\rangle}\right| &\leq C \cdot (1 + |\lambda|^{20} + |\phi|^{20} + |y|^{20}) \cdot \frac{\ln^{10}(\|\boldsymbol{T}\|)}{\sqrt{\|\boldsymbol{T}\|}} \\ &\leq \frac{C \cdot (1 + |\lambda|^{20} + |\phi|^{20} + |y|^{20})}{\|\boldsymbol{T}\|^{\frac{1}{3}}}.\end{aligned}$$

6. The claim $\lambda_{\max}\left(\boldsymbol{\Sigma}_{\mathsf{TWis}}(\lambda, \phi, y)\right) < \infty$ follows from the moment estimates derived in claim

(4) of the lemma. To show that $\lambda_{\min}\left(\boldsymbol{\Sigma}_{\mathsf{TWis}}(\lambda,\phi,y)\right) > 0$ we note that if

$$\lambda_{\min}\left(\boldsymbol{\Sigma}_{\mathsf{TWis}}(\lambda,\phi,y)\right) = 0,$$

we can find a matrix $\boldsymbol{T}$ with $\|\boldsymbol{T}\| = 1$ such that $\langle \boldsymbol{T}, \boldsymbol{S}\rangle$ is deterministic. If this happens then the characteristic function $\mathbb{E}e^{\mathbf{i}t\langle \boldsymbol{S},\boldsymbol{T}\rangle} = 1$ which contradicts the $O(t^{-\frac{1}{3}})$ decay proved in Claim (5) of this lemma.

□

B.7 Analysis of the Variational Problems

In this section, we study the potential functions involved in the definition of the key functions $\Xi_1(\sigma), \hat{\Xi}_1(\sigma)$ and $\Xi_2(q;\sigma), \hat{\Xi}_2(q;\sigma)$. Define the two concave potential functions:

$$V_1(\lambda; r) = \lambda r - \mathbb{E}_Y \ln Z_{\mathsf{TExp}}(\lambda, Y),\ \lambda \in \mathbb{R}$$
$$V_2(\lambda,\phi; q) = 2\alpha\lambda + \beta\phi - \mathbb{E}_Y \ln Z_{\mathsf{TWis}}(\lambda,\phi,Y),\ \lambda,\phi \in \mathbb{R}.$$

In this section, we study the two variational problems:

$$\text{P1: } \max_{\lambda\in\mathbb{R}} V_1(\lambda),$$
$$\text{P2: } \max_{\lambda,\phi\in\mathbb{R}} V_2(\lambda,\phi;q).$$

The analysis in this section will consider an arbitrary distribution on the random variable Y. The reason for doing so is to handle the following two cases in a unified way:

1. Y is sampled from the empirical distribution of the phase retrieval observations:

$$Y \sim \frac{1}{m}\sum_{i=1}^{m} \delta_{y_i}.$$

This case covers the analysis of $\hat{\Xi}_1(\sigma), \hat{\Xi}_2(q;\sigma)$.

2. $Y = |Z|^2 + \sigma\epsilon$ where $Z \sim \mathcal{CN}(0,1)$ and $\epsilon \sim \mathcal{N}(0,1)$. This case covers the analysis of $\Xi_1(\sigma), \Xi_2(q;\sigma)$.

We also note that the potential functions V_1, V_2 depend on the noise level σ even though the dependence is not explicit in our notation. In this section, we consider a fixed $\sigma > 0$ and the universal constants C of this section may depend on σ. However, they do not depend on the distribution of Y. Finally we note that the variation problem P1 is more general than we require in the sense that for the analysis of $\Xi_1, \hat{\Xi}_1$, we can set $r = 1$. The reason for studying this more general variational problem is that we can reduce the analysis of P2 to this more general variational problem.

B.7.1 Analysis of Variational Problem P1

The following proposition analyzes the variational problem P1 and shows that it has a unique minimizer which is guaranteed to lie in a ball of a certain radius.

Proposition 22 (Analysis of P1)**.** *There exists a universal constant $0 < C < \infty$ depending only on the noise level σ such that:*

1. *The following coercivity estimate holds:*

$$V_1(\lambda; r) \leq -\frac{r|\lambda|}{2C}, \ \forall |\lambda| \geq C\left(r + \frac{1}{r}\right)(\mathbb{E}|Y|^2 + 1).$$

2. *All minimizers of the variational problem lie in the compact set:*

$$\left\{\lambda : |\lambda| \leq C\left(r + \frac{1}{r}\right)(\mathbb{E}|Y|^2 + 1)\right\}.$$

3. *The function $V_1(\lambda; r)$ is strongly concave on every compact set. Consequently, the variational problem has a unique minimizer.*

Proof. Throughout this proof, C refers to a universal constant depending only on σ which may change from line to line.

1. We need to show that V_1 is coercive, that is:

$$V_1(\lambda; r) \to -\infty \text{ as } |\lambda| \to \infty.$$

In order to do so we need to obtain lower bounds on $\ln Z_{\mathsf{TExp}}\left(\pm|\lambda|, Y\right)$. First we consider:

$$\begin{aligned}
Z_{\mathsf{TExp}}\left(|\lambda|, y\right) &\overset{\text{(a)}}{=} \int_0^\infty e^{-(1-|\lambda|)u}\psi_\sigma(u-y)\,\mathrm{d}u \\
&= \frac{e^{-\frac{y^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\int_0^\infty \exp\left(|\lambda|u - \frac{u^2}{2\sigma^2} + \frac{uy}{2\sigma^2} - u\right)\mathrm{d}u \\
&\overset{\text{(b)}}{=} \frac{|\lambda|e^{-\frac{y^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\int_0^\infty \exp\left(|\lambda|^2 u\left(1-\frac{u}{2\sigma^2}\right) + \frac{|\lambda|uy}{2\sigma^2} - |\lambda|u\right)\mathrm{d}u \\
&\geq \frac{|\lambda|e^{-\frac{y^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\int_{\frac{\sigma^2}{2}}^{\sigma^2} \exp\left(|\lambda|^2 u\left(1-\frac{u}{2\sigma^2}\right) + \frac{|\lambda|uy}{2\sigma^2} - |\lambda|u\right)\mathrm{d}u \\
&\overset{\text{(c)}}{\geq} \frac{|\lambda|e^{-\frac{y^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\int_{\frac{\sigma^2}{2}}^{\sigma^2} \exp\left(\frac{|\lambda|^2\sigma^2}{4} - |\lambda|\left(\frac{|y|}{2} + \sigma^2\right)\right)\mathrm{d}u.
\end{aligned}$$

In the step marked (a), we used Definition 5. In the step marked (b), we performed a change of variable $u = |\lambda|u$. In the step marked (c), we used the fact that:

$$u\left(1-\frac{u}{2\sigma^2}\right) \leq \frac{|\lambda|^2\sigma^2}{4},\ \frac{\sigma^2}{2} \leq u \leq \sigma^2.$$

Hence we obtain, for a universal constant $0 < C < \infty$ depending only on σ^2, we have,

$$\mathbb{E}_Y \ln Z_{\mathsf{TExp}}\left(|\lambda|, y\right) \geq \ln|\lambda| + \frac{|\lambda|^2}{C} - C|\lambda|(\mathbb{E}|Y|^2 + 1).$$

Hence,

$$\begin{aligned} V_1(|\lambda|; r) &\leq |\lambda|(r + C(y^2+1)) - \ln|\lambda| - \frac{|\lambda|^2}{C} \\ &\leq -\frac{|\lambda|^2}{2C},\ \forall |\lambda| \geq 2C(r + C(\mathbb{E}|Y|^2 + 1)). \end{aligned} \tag{B.49}$$

Next we consider:

$$\begin{aligned} Z_{\mathsf{TExp}}\left(-|\lambda|, y\right) &= \int_0^\infty e^{-(1+|\lambda|)u} \psi_\sigma(u-y)\,\mathrm{d}u \\ &= |\lambda| \int_0^\infty e^{-u} \cdot \psi_\sigma\left(\frac{u}{|\lambda|} - y\right) \cdot e^{-\frac{u}{|\lambda|}}\,\mathrm{d}u \end{aligned}$$

By Jensen's Inequality,

$$\begin{aligned} \ln Z_{\mathsf{TExp}}\left(-|\lambda|, y\right) &= \ln|\lambda| + \ln\left(\mathbb{E}_{E\sim\mathsf{Exp}(1)} e^{-\frac{E}{|\lambda|}} \psi_\sigma\left(\frac{E}{|\lambda|} - y\right)\right) \\ &\geq \ln|\lambda| - \frac{1}{|\lambda|} - \frac{1}{2\sigma^2}\left(\frac{2}{|\lambda|^2} + y^2 - \frac{2y}{|\lambda|}\right) \\ &\geq \ln|\lambda| - C(y^2+1),\ \forall\ |\lambda| \geq 1. \end{aligned}$$

Consequently we have,

$$\begin{aligned} V_1(-|\lambda|; r) &\leq -r|\lambda| - \ln|\lambda| + C(\mathbb{E}|Y|^2 + 1),\ \forall\ |\lambda| \geq 1 \\ &\leq -\frac{r|\lambda|}{2},\ \forall |\lambda| \geq \frac{2C(\mathbb{E}|Y|^2+1))}{r} + 1. \end{aligned} \tag{B.50}$$

Combining the estimates in (B.49) and (B.50), we obtain that for a large enough constant C,

$$V_1(\lambda; r) \leq -\frac{r|\lambda|}{2C},\ \forall |\lambda| \geq C\left(r + \frac{1}{r}\right)(\mathbb{E}|Y|^2 + 1).$$

2. We observe that,

$$\begin{aligned} V_1(0;r) &= -\mathbb{E}_Y \ln \mathbb{E}_{E\sim \mathsf{Exp}(1)} \psi_\sigma(E-Y) \\ &\geq -\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) \\ &\geq -C. \end{aligned}$$

Hence,

$$|\lambda| \geq C\left(r+\frac{1}{r}\right)(\mathbb{E}|Y|^2+1) \implies V_1(\lambda;r) \leq V_1(0;r).$$

Hence,

$$\arg\min_{\lambda\in\mathbb{R}} V_1(\lambda;r) \subset \left\{\lambda : |\lambda| \leq C\left(r+\frac{1}{r}\right)(\mathbb{E}|Y|^2+1)\right\}.$$

3. In the light of item (2) of the lemma, it is sufficient to study the variational problem:

$$\max_{|\lambda|\leq R} V_1(\lambda;r),\ R \stackrel{\text{def}}{=} C\left(r+\frac{1}{r}\right)(\mathbb{E}|Y|^2+1).$$

In order to show uniqueness of the solution it is sufficient to show that $V_1(\lambda;r)$ is strictly concave on $|\lambda| \leq R$, for which it is sufficient to check that:

$$\min_{|\lambda|\leq R} \frac{\mathrm{d}^2 V_1}{\mathrm{d}\lambda^2}(\lambda) < 0 \Leftrightarrow \frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\mathbb{E}_Y \ln Z_{\mathsf{TExp}}(\lambda, Y) > 0.$$

Note that by convexity we have,

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\mathbb{E}_Y \ln Z_{\mathsf{TExp}}(\lambda, Y) \geq 0.$$

In order to obtain a strict inequality, suppose there is a λ_0 such that:

$$\frac{\mathrm{d}^2}{\mathrm{d}\lambda^2}\mathbb{E}_Y \ln Z_{\mathsf{TExp}}(\lambda, Y)\Big|_{\lambda=\lambda_0} = 0.$$

This means that,

$$E_Y\left[\frac{\mathbb{E}E^2 e^{\lambda_0 E}\psi_\sigma(E-Y)}{\mathbb{E}e^{\lambda_0 E}\psi_\sigma(E-Y)} - \left(\frac{\mathbb{E}Ee^{\lambda_0 E}\psi_\sigma(E-Y)}{\mathbb{E}e^{\lambda_0 E}\psi_\sigma(E-Y)}\right)^2\right] = 0.$$

Recalling Definition 5,

$$\sigma^2_{\mathsf{TExp}}(\lambda_0, Y) \overset{\text{a.s.}}{=} 0.$$

However this contradicts the decay rate property of the characteristic function of Tilted Exponential distribution proved in Lemma 43 in Appendix B.6.1 since the amplitude of the characteristic function of deterministic random variables is constant.

□

B.7.2 Analysis of Variational Problem P2

The following proposition analyzes the variational problem P2 and shows that it has a unique minimizer which is guaranteed to lie in a ball of a certain radius.

Proposition 23 (Analysis of P2)**.** *Suppose that $q \in (0,1)$. There exists a universal constant $0 < C < \infty$ depending only on the noise level σ such that:*

1. *The following coercivity estimate holds:*

$$V_2(\lambda, \phi; q) \leq -\frac{(1-q)}{2C}\cdot(|\lambda|+|\phi|),\ |\lambda|+|\phi| \geq C\left(1+q+\frac{1}{1-q}\right)(\mathbb{E}Y^2+1).$$

2. *All minimizers of the variational problem lie in the compact set:*

$$\left\{(\lambda,\phi)\in\mathbb{R}^2 : |\lambda|+|\phi| \le C\left(1+q+\frac{1}{1-q}\right)\left(\mathbb{E}|Y|^2+1\right)\right\}.$$

3. *The function $V_2(\lambda,\phi;q)$ is strongly concave on any compact set. Consequently, the variational problem has a unique minimizer.*

Proof. Throughout this proof, C refers to a universal constant depending only on σ which may change from line to line. It will be helpful to write the variational problem in the following matrix notation. Define,

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda & \frac{\phi}{2} \\ \frac{\phi}{2} & \lambda \end{bmatrix}$$

Then the problem P2 can be rewritten as:

$$\max_{\boldsymbol{\Lambda}} V_2(\boldsymbol{\Lambda}),\ V_2(\boldsymbol{\Lambda}) = \langle \boldsymbol{\Lambda}, \boldsymbol{Q}\rangle - \mathbb{E}_Y \ln \mathbb{E}_{\boldsymbol{g}\sim\mathcal{CN}(\mathbf{0},\boldsymbol{I}_2)} \exp(\langle \boldsymbol{\Lambda}, \boldsymbol{g}\boldsymbol{g}^{\mathsf{H}}\rangle)\psi_\sigma(Y-|g_1|^2)\psi_\sigma(Y-|g_2|^2),$$

where,

$$\boldsymbol{Q} = \begin{bmatrix} 1 & q \\ q & 1 \end{bmatrix}.$$

To obtain the above display, we recalled the definition of the normalizing constant of the Tilted Wishart Distribution (Definition 6).

1. In order to obtain a coercivity estimate we need to lower bound $\ln Z_{\mathsf{TWis}}(\lambda,\phi,y)$. Our lower bound will depend only on the spectrum of $\boldsymbol{\Lambda}$. We consider the eigendecomposition of $\boldsymbol{\Lambda}$:

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{b}_1^{\mathsf{H}} \\ \boldsymbol{b}_2^{\mathsf{H}} \end{bmatrix} \cdot \begin{bmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 \end{bmatrix}$$

In the above display $\gamma_1 \geq \gamma_2$ are the ordered eigenvalues of $\mathbf{\Lambda}$. We have the following lower bound on $\ln Z_{\mathsf{TWis}}(\lambda, \phi, y)$:

$$
\begin{aligned}
&\ln Z_{\mathsf{TWis}}(\lambda, \phi, y) = \ln \mathbb{E} \exp(\langle \mathbf{\Lambda}, \boldsymbol{g}\boldsymbol{g}^{\mathsf{H}} \rangle) \psi_\sigma(y - |g_1|^2) \psi_\sigma(y - |g_2|^2) \\
&= \ln \mathbb{E} \exp(\gamma_1 |g_1|^2 + \gamma_2 |g_2|^2) \psi_\sigma(y - |\boldsymbol{b}_1^{\mathsf{H}} \boldsymbol{g}|^2) \psi_\sigma(y - |\boldsymbol{b}_2^{\mathsf{H}} \boldsymbol{g}|^2) \\
&= \ln \mathbb{E} \exp\left(\gamma_1 |g_1|^2 + \gamma_2 |g_2|^2 - \frac{1}{2\sigma^2} (2y^2 - 2y(|\boldsymbol{b}_1^{\mathsf{H}} \boldsymbol{g}|^2 + |\boldsymbol{b}_2^{\mathsf{H}} \boldsymbol{g}|^2) + |\boldsymbol{b}_1^{\mathsf{H}} \boldsymbol{g}|^4 + |\boldsymbol{b}_2^{\mathsf{H}} \boldsymbol{g}|^4) \right) \\
&\overset{\text{(a)}}{\geq} \ln \mathbb{E} \exp\left(\gamma_1 |g_1|^2 + \gamma_2 |g_2|^2 - \frac{1}{2\sigma^2} (2y^2 - 2y(|g_1|^2 + |g_2|^2) + 2|g_1|^4 + 2|g_2|^4) \right) \\
&\geq \ln \mathbb{E} \exp(\gamma_1 |g_1|^2) \psi_\sigma\left(\frac{y}{\sqrt{2}} - \sqrt{2}|g_1|^2 \right) + \ln \mathbb{E} \exp(\gamma_2 |g_2|^2) \psi_\sigma\left(\frac{y}{\sqrt{2}} - \sqrt{2}|g_2|^2 \right) - \frac{y^2}{2\sigma^2} \\
&= \ln \mathbb{E}_{E \sim \mathsf{Exp}(1)} e^{\gamma_1 E} \psi_\sigma\left(\frac{y}{\sqrt{2}} - \sqrt{2}E \right) + \ln \mathbb{E}_{E \sim \mathsf{Exp}(1)} e^{\gamma_2 E} \psi_\sigma\left(\frac{y}{\sqrt{2}} - \sqrt{2}E \right) - \frac{y^2}{2\sigma^2}
\end{aligned}
$$

In the step marked (a), we used the fact that,

$$\|\boldsymbol{B}\boldsymbol{g}\|_2^2 = \|\boldsymbol{g}\|^2, \ \|\boldsymbol{B}\boldsymbol{g}\|_4^4 \leq \|\boldsymbol{g}\|_2^4 \leq 2(|g_1|^4 + |g_2|^4).$$

Next note that,

$$\langle \mathbf{\Lambda}, \boldsymbol{Q} \rangle \leq \gamma_1 \lambda_1(\boldsymbol{Q}) + \gamma_2 \lambda_2(\boldsymbol{Q}),$$

where $\lambda_1(\boldsymbol{Q}) \geq \lambda_2(\boldsymbol{Q})$ are the ordered eigenvalues of $\boldsymbol{Q}$. It is easy to check that $\lambda_1(\boldsymbol{Q}) = 1 + q$ and $\lambda_2(\boldsymbol{Q}) = 1 - q$ which means,

$$\langle \mathbf{\Lambda}, \boldsymbol{Q} \rangle \leq \gamma_1(1 + q) + \gamma_2(1 - q).$$

This gives us,

$$V_2(\mathbf{\Lambda}; q) \leq \gamma_1(1+q) - \mathbb{E}\ln \mathbb{E}e^{\gamma_1 E}\psi_\sigma\left(\frac{Y}{\sqrt{2}} - \sqrt{2}E\right) + \gamma_2(1-q) \\ - \mathbb{E}\ln \mathbb{E}e^{\gamma_2 E}\psi_\sigma\left(\frac{y}{\sqrt{2}} - \sqrt{2}E\right) + \frac{\mathbb{E}Y^2}{\sigma^2}$$

Utilizing the coercivity estimates from Proposition 22, we obtain,

$$\left(\gamma_1 \cdot (1+q) - \mathbb{E}_Y \ln \mathbb{E}e^{\gamma_1 E}\psi_\sigma\left(\frac{Y}{\sqrt{2}} - \sqrt{2}E\right)\right) \leq -\frac{(1+q)\cdot|\gamma_1|}{2C}, \\ \left(\gamma_2(1-q) - \mathbb{E}_Y \ln \mathbb{E}e^{\gamma_2 E}\psi_\sigma\left(\frac{Y}{\sqrt{2}} - \sqrt{2}E\right)\right) \leq -\frac{(1-q)|\gamma_2|}{2C},$$

for all:

$$|\gamma_1| \geq C\left(1+q+\frac{1}{1+q}\right)\left(\mathbb{E}Y^2+1\right), \\ |\gamma_2| \geq C\left((1-q)+\frac{1}{1-q}\right)\left(\mathbb{E}Y^2+1\right).$$

Since,

$$\|\mathbf{\Lambda}\|^2 \geq t \implies \max(\gamma_1^2, \gamma_2^2) \geq \frac{t}{2},\ \forall t,$$

we obtain,

$$V_2(\mathbf{\Lambda}; q) \leq -\frac{1-q}{2C}\cdot\|\mathbf{\Lambda}\|,\ \|\mathbf{\Lambda}\| \geq C\left(1+q+\frac{1}{1-q}\right)(\mathbb{E}Y^2+1).$$

This is equivalent to the estimate:

$$V_2(\lambda, \phi; q) \leq -\frac{1-q}{2C}\cdot(|\lambda|+|\phi|),\ |\lambda|+|\phi| \geq C\cdot\left(1+q+\frac{1}{1-q}\right)(\mathbb{E}Y^2+1).$$

This concludes the proof of item (1) in the statement of the lemma.

2. The proof is analogous to the proof of item (2) in Proposition 22.

3. The proof is analogous to the proof of item (3) in Proposition 22.

□

B.8 Background on Characteristic Functions

In this section we collect some basic facts about characteristic functions (CF). Most of these results are taken from Chapter XV of Feller [76]. The characteristic function is simply the Fourier transform of the probability density function.

Definition 11 (Characteristic Function)**.** *Let f be a probability density function on $\mathbb{R}$. Then the characteristic function of f is defined as:*

$$\psi(t) = \int_{\mathbb{R}} e^{\mathbf{i}tx} f(x)\,\mathrm{d}x.$$

If the characteristic function is absolutely integrable, the probability density function can be recovered from it using the Fourier inversion formula.

Theorem 7 (Fourier Inversion of CFs)**.** *Let ψ be the CF of density f. Then,*

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \psi(t) e^{-\mathbf{i}tx}\,\mathrm{d}t.$$

The moments of the PDF can be recovered from the Taylors expansion of the CF.

Theorem 8 (Taylors Series of CF)**.** *Let X be a random variable with probability density function f. Let ψ be the CF of f. We have, for any $t \in \mathbb{R}$.*

$$\left| \psi(t) - \left(1 + \sum_{k=1}^{n-1} \frac{\mathbb{E}X^k}{k!} \cdot (\mathbf{i}t)^k \right) \right| \leq \frac{\mathbb{E}|X|^n t^n}{n!}$$

The following bound on CFs will be useful in the proofs of the local central limit theorems.

Lemma 45 (Bounds on CF)**.** *Let ψ be a multivariate CF and suppose that, there exists $0 < c < 1$ and $b > 0$ such that,*

$$|\psi(\boldsymbol{t})| \leq c \; \forall \; \|\boldsymbol{t}\| > b. \tag{B.51}$$

Then, for any $\|\boldsymbol{t}\| \leq b$ we have,

$$|\psi(\boldsymbol{t})| \leq 1 - \frac{1-c^2}{8b^2}\|\boldsymbol{t}\|^2.$$

Proof. A univariate version is given as Theorem 1 in Chapter 1 of Petrov [88]. A multivariate version is given as Theorem 1.8.13 in Ushakov [89]. □

Finally we state a Multivariate Berry-Eseen bound due to Bhattacharya [83].

Theorem 9 (A Multivariate Berry-Eseen Bound, [83])**.** *Let $X_1, X_2 \ldots X_n$ be independent random vectors in $\mathbb{R}^k$. Suppose that:*

$$\mathbb{E}X_i = \mathbf{0}, \; \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}X_i X_i^{\mathsf{T}} = \boldsymbol{I}_k.$$

Define:

$$\rho_3 \stackrel{def}{=} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\|X_i\|^3.$$

Then, there exists a universal constant C_k depending only on the dimension k, such that for any bounded, Lipchitz function f we have,

$$\left| \mathbb{E}f\left(\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \right) - \mathbb{E}_{Z \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_k)} f(Z) \right| \leq \frac{C_k \cdot \rho_3 \cdot (\|f\|_\infty + \|f\|_{Lip})}{\sqrt{n}}.$$

B.9 Some Miscellaneous Results

This appendix collects some miscellaneous facts and results that are useful in our analysis. The first is a classical correlation inequality.

Fact 2 (Chebychev Association Inequality, [70]). *Let A, B be r.v.s and $B \geq 0$. Suppose f, g are two non-decreasing functions. Then, $\mathbb{E}[B]\mathbb{E}[Bf(A)g(A)] \geq \mathbb{E}[f(A)B]\mathbb{E}[g(A)B]$.*

The following collects some useful properties of Modified Bessel Function of the first kind. These results can be found in the standard references [90, 91]. Item (5) of the following is relatively less known and is due to Watson [92, Appendix A].

Fact 3 (Properties of Modified Bessel Function of the First Kind). *For $x \in \mathbb{R}$, the Modified Bessel Function of the First Kind, denoted by, $I_0(x)$ is defined as:*

$$I_0(x) \stackrel{def}{=} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{x\cos(\theta)}.$$

It satisfies the following properties:

1. *$I_0(x)$ is an increasing function on $x \geq 0$ and $I_0(0) = 1$.*

2. *$I_0(x)$ is an even function.*

3. *There exists a universal constant C such that,*

$$I_0(x) \leq \frac{Ce^x}{\sqrt{x}}, \ \forall x \geq 0.$$

4. *I_0 is infinitely differentiable.*

5. *The function $\frac{I_0'}{I_0}$ is an increasing concave function with,*

$$\frac{I_0'(0)}{I_0(0)} = 0, \ \lim_{x\to\infty} \frac{I_0'(x)}{I_0(x)} = 1,$$

and,

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(\frac{I_0'(z)}{I_0(z)}\right)\Big|_{z=0} = \frac{1}{2}.$$

The following lemma is about a bivariate Gaussian integral.

Lemma 46. *Let Z_1, Z_2 be distributed as:*

$$\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right).$$

Then the integral:

$$J(a,b) \stackrel{def}{=} \mathbb{E}_{Z_1,Z_2}\psi_1(a-Z_1)\psi_1(b-Z_2),$$

is given by:

$$J(a,b) = \frac{1}{4\pi\sqrt{1-\rho^2/4}} \cdot \exp\left(-\frac{a^2+b^2-\rho ab}{4(1-\rho^2/4)}\right).$$

Proof. Note that $J(a,b)$ is the Joint PDF of the random variables (A,B) with distribution:

$$A = Z_1 + \epsilon_1,\ B = Z_2 + \epsilon_2,\ \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}\right),\ \epsilon_1 \sim \mathcal{N}(0,1),\ \epsilon_2 \sim \mathcal{N}(0,1).$$

We can directly find the distribution of A, B from this description:

$$\begin{bmatrix} A \\ B \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & \rho \\ \rho & 2 \end{bmatrix}\right).$$

Hence by the formula for the bivariate Gaussian pdf,

$$J(a,b) = \frac{1}{4\pi\sqrt{1-\rho^2/4}} \cdot \exp\left(-\frac{a^2+b^2-\rho ab}{4(1-\rho^2/4)}\right).$$

□

We will also find the following bound on truncated Gaussian integrals useful.

Lemma 47 (Truncated Gaussian Integrals)**.** *Suppose that $a, A > 0$ and $k \in \mathbb{N}$. Then, we have,*

$$\int_a^\infty x^k e^{-\frac{x^2}{2A^2}}\,\mathrm{d}x \le C_k \cdot A \cdot (A^k + a^k) \cdot e^{-\frac{a^2}{2}}.$$

In the above display C_k is a universal constant depending only on k.

Proof. Let us first consider the case when $A = 1$. Then we have,

$$\begin{aligned}
\int_a^\infty x^k e^{-\frac{x^2}{2}}\,\mathrm{d}x &\overset{\text{(a)}}{=} 2^{\frac{k-1}{2}} \int_{a^2/2}^\infty u^{\frac{k-1}{2}} e^{-u}\,\mathrm{d}u \\
&\overset{\text{(b)}}{=} 2^{\frac{k-1}{2}} \cdot e^{-\frac{a^2}{2}} \cdot \int_0^\infty \left(x + \frac{a^2}{2}\right)^{\frac{k-1}{2}} e^{-x}\,\mathrm{d}x \\
&\overset{\text{(c)}}{\le} 2^{k-1} \cdot e^{-\frac{a^2}{2}} \cdot \int_0^\infty \left(x^{\frac{k-1}{2}} + \frac{a^{k-1}}{2^{\frac{k-1}{2}}}\right) e^{-x}\,\mathrm{d}x \\
&\le 2^{k-1} \cdot e^{-\frac{a^2}{2}} \cdot \left(\sqrt{\int_0^\infty e^{-x} x^{k-1}\,\mathrm{d}x} + \frac{a^{k-1}}{2^{\frac{k-1}{2}}}\right) \\
&\le 2^{k-1} \cdot e^{-\frac{a^2}{2}} \cdot \left(\sqrt{(k-1)!} + \frac{a^{k-1}}{2^{\frac{k-1}{2}}}\right) \\
&\le C_k(1 + a^k) e^{-\frac{a^2}{2}}
\end{aligned}$$

In the step marked (a), we substituted $u = x^2/2$ in the step marked (b) we substituted $u = x + a$. In the step marked (c) we used the inequality $(a+b)^k \le 2^k(a^k + b^k),\ a, b \ge 0\ k \ge 0$. Making the

substitution $x = Ax$ in the above bound gives us:

$$\int_a^\infty x^k e^{-\frac{x^2}{2A^2}} \, \mathrm{d}x \leq C_k \cdot A \cdot (A^k + a^k) \cdot e^{-\frac{a^2}{2}}.$$

This concludes the proof. □

The following lemma contains a useful upper bound on $\mathbb{E}|G|^{-\frac{1}{2}}$ where $G \sim \mathcal{N}(0, 1)$.

Lemma 48 (Fractional Moments of Gaussian Distribution)**.** *Let $G \sim \mathcal{N}\left(\mu, \sigma^2\right)$. Then we have,*

$$\mathbb{E}\frac{1}{\sqrt{|G|}} \leq \frac{4}{\sqrt{|\mu|}}.$$

Proof. We have,

$$\begin{aligned}
\mathbb{E}\frac{1}{\sqrt{|G|}} &\leq \mathbb{E}\frac{1}{\sqrt{|G|}}\mathbf{1}_{|G|\leq 0.5|\mu|} + \mathbb{E}\frac{1}{\sqrt{|G|}}\mathbf{1}_{|G|>0.5|\mu|} \\
&\leq \int_{-0.5|\mu|}^{0.5|\mu|} \frac{1}{\sqrt{|x|}} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x + \frac{\sqrt{2}}{\sqrt{|\mu|}} \\
&\leq \frac{e^{-\frac{\mu^2}{8\sigma^2}}}{\sqrt{2\pi\sigma^2}} \cdot \int_{-0.5|\mu|}^{0.5|\mu|} \frac{1}{\sqrt{|x|}} \, \mathrm{d}x + \frac{\sqrt{2}}{\sqrt{|\mu|}} \\
&= \frac{2\sqrt{|\mu|}e^{-\frac{\mu^2}{8\sigma^2}}}{\sqrt{\pi\sigma^2}} + \frac{\sqrt{2}}{\sqrt{|\mu|}} \\
&\leq \frac{4}{\sqrt{|\mu|}}
\end{aligned}$$

In the last step we used the fact that $\max_{x\geq 0} \sqrt{x}e^{-x} \leq \frac{1}{\sqrt{2e}}$. □

Appendix C: Omitted Proofs from Chapter 5

C.1 Proof of Lemmas 21 and 22

C.1.1 Proof of Lemma 21

Proof of Lemma 21. Recall that, $\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} = \boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^{\mathsf{T}},\ \boldsymbol{\Psi} = \boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} - \mathbb{E}[\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}|\boldsymbol{U}] = \boldsymbol{U}(\boldsymbol{B} - \kappa \boldsymbol{I}_m)\boldsymbol{U}^{\mathsf{T}}$ where $\boldsymbol{B}$ is a uniformly random $m \times m$ diagonal matrix with exactly n entries set to 1 and the remaining entries set to 0. Using the concentration inequality of Lemma 17:

$$\mathbb{P}\left(|(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})_{ij} - \mathbb{E}(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})_{ij}| > \epsilon \;\middle|\; \boldsymbol{U}\right) \leq 4\exp\left(-\frac{\epsilon^2}{8m\|\boldsymbol{U}\|_\infty^4}\right). \tag{C.1}$$

Setting $\epsilon = \sqrt{32 \cdot m \cdot \|\boldsymbol{U}\|_\infty^4 \cdot \log(m)}$ in (C.1) we obtain,

$$\mathbb{P}\left(|(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})_{ij} - \mathbb{E}(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})_{ij}| > \sqrt{32 \cdot m \cdot \|\boldsymbol{U}\|_\infty^4 \cdot \log(m)} \;\middle|\; \boldsymbol{U}\right) \leq \frac{4}{m^4}.$$

By a union bound, $\mathbb{P}(\mathcal{E}^c|\boldsymbol{U}) \leq 4/m^2 \to 0$. In order to prove the claim of the lemma for the subsampled Haar model, we first note that by Fact 7 we have,

$$\mathbb{P}\left(|O_{ij}| > \sqrt{\frac{8\log(m)}{m}}\right) \leq \frac{2}{m^4}.$$

By a union bound $\mathbb{P}(\|\boldsymbol{O}\|_\infty > \sqrt{8\log(m)/m}) \leq 2m^{-2}$. This gives us:

$$\begin{aligned}\mathbb{P}\left(\left\{\|\boldsymbol{O}\|_\infty \leq \sqrt{\frac{8\log(m)}{m}}\right\} \cap \mathcal{E}\right) &\geq 1 - \mathbb{P}\left(\|\boldsymbol{O}\|_\infty > \sqrt{\frac{8\log(m)}{m}}\right) - \mathbb{P}(\mathcal{E}^c) \\ &\geq 1 - \frac{2}{m^2} - \mathbb{E}\mathbb{P}(\mathcal{E}^c|\boldsymbol{U}) \\ &\geq 1 - \frac{6}{m^2}.\end{aligned}$$

This concludes the proof of the lemma. □

C.1.2 Proof of Lemma 22

Proof of Lemma 22. Consider any alternating product $\mathcal{A}$ (see Definition 7):

$$\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}) = (\boldsymbol{\Psi})q_1(\boldsymbol{Z})(\boldsymbol{\Psi})\cdots q_k(\boldsymbol{Z}).$$

Note that in the above expression, we have assumed the alternating product is of Type 2 but the following argument applies to all the other types too. We define:

$$\mathcal{A}_i = (\boldsymbol{\Psi})q_1(\boldsymbol{Z})(\boldsymbol{\Psi})q_2(\boldsymbol{Z})\cdots(\boldsymbol{\Psi})q_i(\boldsymbol{Z})(\boldsymbol{\Psi})q_{i+1}(\boldsymbol{Z}')(\boldsymbol{\Psi})q_{i+2}(\boldsymbol{Z}')\cdots(\boldsymbol{\Psi})q_k(\boldsymbol{Z}').$$

Then we can express $\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}') - \mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})$ as a telescoping sum:

$$\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}) - \mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}') = \sum_{i=1}^{k}(\mathcal{A}_i - \mathcal{A}_{i-1}).$$

Hence,

$$\left|\frac{\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})}{m} - \frac{\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}')}{m}\right| \leq \frac{1}{m}\sum_{i=1}^{k}|\mathsf{Tr}(\mathcal{A}_i - \mathcal{A}_{i-1})|.$$

Next we observe that:

$$
\begin{aligned}
&|\mathsf{Tr}(\mathcal{A}_i - \mathcal{A}_{i-1})| \\
&= |\mathsf{Tr}((\boldsymbol{\Psi})q_1(\boldsymbol{Z}) \cdots (\boldsymbol{\Psi})q_{i-1}(\boldsymbol{Z}) \cdot (q_i(\boldsymbol{Z}) - q_i(\boldsymbol{Z}')) \cdot (\boldsymbol{\Psi})q_{i+1}(\boldsymbol{Z}') \cdots (\boldsymbol{\Psi})q_k(\boldsymbol{Z}'))| \\
&\leq \left\|(\boldsymbol{\Psi})q_1(\boldsymbol{Z}) \cdots (\boldsymbol{\Psi})q_{i-1}(\boldsymbol{Z}) \cdot (\boldsymbol{\Psi})q_{i+1}(\boldsymbol{Z}') \cdots (\boldsymbol{\Psi})q_k(\boldsymbol{Z}')\right\|_{\mathsf{op}} \cdot \left(\sum_{j=1}^{m} |q_i(z_j) - q_i(z'_j)|\right) \\
&\leq \|(\boldsymbol{\Psi})\|_{\mathsf{op}}\|q_1(\boldsymbol{Z})\|_{\mathsf{op}} \cdots \|(\boldsymbol{\Psi})\|_{\mathsf{op}}\|q_k(\boldsymbol{Z}')\|_{\mathsf{op}} \cdot \left(\sum_{j=1}^{m} |q_i(z_j) - q_i(z'_j)|\right) \\
&\overset{\text{(a)}}{\leq} \left(\prod_{j=1}^{k} \|q_j\|_{\infty}\right) \cdot \|q_i\|_{\mathsf{Lip}} \cdot \left(\sum_{j=1}^{m} |z_j - z'_j|\right) \\
&\leq \sqrt{m} \cdot C(\mathcal{A}) \cdot \|\boldsymbol{Z} - \boldsymbol{Z}'\|_{\mathsf{Fr}}.
\end{aligned}
$$

In the step marked (a), we observed that: $\|(\boldsymbol{\Psi})\|_{\mathsf{op}} = \|\boldsymbol{U}(\overline{\boldsymbol{B}})\boldsymbol{U}^{\mathsf{T}}\|_{\mathsf{op}} \leq \max(|\kappa)|, |1-\kappa|) \leq 1$. Similarly, $\|q_j(\boldsymbol{Z})\|_{\mathsf{op}} \leq \|q_j\|_{\infty} \overset{\text{def}}{=} \sup_{\xi \in \mathbb{R}} |q_j(\xi)|$. We also recalled the functions q_i are assumed to be Lipchitz and denoted the Lipchitz constant of q_i by $\|q_i\|_{\mathsf{Lip}}$. Hence we obtain:

$$
\left|\frac{\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})}{m} - \frac{\mathsf{Tr}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}')}{m}\right| \leq \frac{k \cdot C(\mathcal{A})}{\sqrt{m}} \cdot \|\boldsymbol{Z} - \boldsymbol{Z}'\|_{\mathsf{Fr}}.
$$

This concludes the proof of the lemma. □

C.2 Proof of Proposition 19

The proof of Proposition 19 is very similar to the proof of Proposition 18 and hence we will be brief in our arguments.

As discussed in the proof of Proposition 18, we will assume that alternating form is of Type 1. The other types are handled as outlined in Remark 15. Furthermore, in light of Lemma 16 we can further assume that all polynomials $p_i(\psi) = \psi$. Hence we assume that $\mathcal{A}$ is of the form:

$$
\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}) = \boldsymbol{\Psi}q_1(\boldsymbol{Z})\boldsymbol{\Psi} \cdots q_{k-1}(\boldsymbol{Z})\boldsymbol{\Psi}.
$$

The proof of Proposition 19 consists of various steps which will be organized as separate lemmas. We begin by recall that

$$\boldsymbol{z} \sim \mathcal{N}\left(0, \frac{\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}}{\kappa}\right).$$

Define the event:

$$\mathcal{E} = \left\{ \max_{i \neq j} |(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}|)_{ij} \leq \sqrt{\frac{2048 \cdot \log^3(m)}{m}},\ \max_{i \in [m]} |(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})_{ii} - \kappa| \leq \sqrt{\frac{2048 \cdot \log^3(m)}{m}} \right\} \quad \text{(C.2)}$$

By Lemma 21 we know that $\mathbb{P}(\mathcal{E}^c) \to 0$ for both the subsampled Haar sensing and the subsampled Hadamard model. We define the normalized random vector $\widetilde{\boldsymbol{z}}$ as:

$$\widetilde{z}_i = \frac{z_i}{\sigma_i},\ \sigma_i^2 = \frac{(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})_{ii}}{\kappa}$$

Note that conditional on $\boldsymbol{A}$, $\widetilde{\boldsymbol{z}}$ is a zero mean Gaussian vector with:

$$\mathbb{E}[\widetilde{z}_i^2 | \boldsymbol{A}] = 1,\ \mathbb{E}[\widetilde{z}_i \widetilde{z}_j | \boldsymbol{A}] = \frac{(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})_{ij}/\kappa}{\sigma_i \sigma_j}.$$

We define the diagonal matrix $\widetilde{\boldsymbol{Z}} = \text{Diag}\,(\widetilde{\boldsymbol{z}})$.

Lemma 49. *We have,*

$$\lim_{m \to \infty} \frac{\mathbb{E}(\boldsymbol{z}^{\mathsf{T}} \mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}) \boldsymbol{z})^2}{m^2} = \lim_{m \to \infty} \frac{\mathbb{E}(\widetilde{\boldsymbol{z}}^{\mathsf{T}} \mathcal{A}(\boldsymbol{\Psi}, \widetilde{\boldsymbol{Z}}) \widetilde{\boldsymbol{z}})^2}{m^2} \mathbb{I}\left(\mathcal{E}\right),$$

provided the latter limit exists.

The proof of this lemma is analogous the proof of Lemma 26 and is omitted. The advantage of Lemma 49 is that $\widetilde{z}_i \sim \mathcal{N}\left(0, 1\right)$ and on the event $\mathcal{E}$ the coordinates of $\widetilde{\boldsymbol{z}}$ have weak correlations. Consequently, Mehler's Formula (Proposition 15) can be used to analyze the leading order term in $\mathbb{E}[\widetilde{\boldsymbol{z}}^{\mathsf{T}} \mathcal{A}(\boldsymbol{\Psi}, \widetilde{\boldsymbol{Z}}) \widetilde{\boldsymbol{z}}\, \mathbb{I}\left(\mathcal{E}\right)]$. Before we do so, we do one additional preprocessing step.

Lemma 50. *We have,*

$$\lim_{m\to\infty} \frac{\mathbb{E}(\widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}})\widetilde{\boldsymbol{z}})^2}{m^2}\mathbb{I}\left(\mathcal{E}\right) = \lim_{m\to\infty} \frac{\mathbb{E}\,\mathsf{Tr}(\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}-\widetilde{\boldsymbol{Z}}^2)\cdot\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}-\widetilde{\boldsymbol{Z}}^2))\mathbb{I}\left(\mathcal{E}\right)}{m^2},$$

provided the latter limit exists.

Proof Sketch. Observe that we can write:

$$\begin{aligned}
&(\widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}\widetilde{\boldsymbol{z}})^2 = \mathsf{Tr}(\mathcal{A}\cdot\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}\cdot\mathcal{A}\cdot\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}})\\
&= \mathsf{Tr}(\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}-\widetilde{\boldsymbol{Z}}^2+\widetilde{\boldsymbol{Z}}^2)\cdot\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}-\widetilde{\boldsymbol{Z}}^2+\widetilde{\boldsymbol{Z}}^2))\\
&= \mathsf{Tr}(\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}-\widetilde{\boldsymbol{Z}}^2)\cdot\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}-\widetilde{\boldsymbol{Z}}^2)) + \mathsf{Tr}(\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2\cdot\mathcal{A}\cdot\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}) + \mathsf{Tr}(\mathcal{A}\cdot\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}\cdot\widetilde{\boldsymbol{Z}}^2\cdot\mathcal{A})\\
&\qquad\qquad - \mathsf{Tr}(\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2\cdot\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2)\\
&= \mathsf{Tr}(\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}-\widetilde{\boldsymbol{Z}}^2)\cdot\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}-\widetilde{\boldsymbol{Z}}^2)) + 2\widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2\cdot\mathcal{A}\cdot\widetilde{\boldsymbol{z}} - \mathsf{Tr}(\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2\cdot\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2).
\end{aligned}$$

Next we note that:

$$|\widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2\cdot\mathcal{A}\cdot\widetilde{\boldsymbol{z}}| \le \|\widetilde{\boldsymbol{z}}\|^2\cdot\|\mathcal{A}\|_{\mathsf{op}}^2\cdot\left(\max_{i\in[m]}|\widetilde{z}_i|^2\right) \le O_P(m)\cdot O(1)\cdot O_P(\mathrm{polylog}(m)),$$

Hence it can be shown that,

$$\frac{\mathbb{E}|\widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2\cdot\mathcal{A}\cdot\widetilde{\boldsymbol{z}}|}{m^2}\to 0.$$

Similarly,

$$\begin{aligned}
|\mathsf{Tr}(\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2\cdot\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2)| &\le m\|\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2\cdot\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2\|_{\mathsf{op}} \le m\|\mathcal{A}\|_{\mathsf{op}}^2\cdot\left(\max_{i\in[m]}|\widetilde{z}_i|^4\right)\\
&\le O(m)\cdot O(1)\cdot O_P(\mathrm{polylog}(m)),
\end{aligned}$$

and hence one expects that,

$$\frac{\mathbb{E}|\mathsf{Tr}(\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2\cdot\mathcal{A}\cdot\widetilde{\boldsymbol{Z}}^2)|}{m^2}\to 0.$$

We omit the detailed arguments. This concludes the proof of the lemma. □

Note that, so far, we have shown that:

$$\lim_{m\to\infty}\frac{\mathbb{E}(\boldsymbol{z}^\mathsf{T}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\boldsymbol{z})^2}{m^2}=\lim_{m\to\infty}\frac{\mathbb{E}\ \mathsf{Tr}(\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^\mathsf{T}-\widetilde{\boldsymbol{Z}}^2)\cdot\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^\mathsf{T}-\widetilde{\boldsymbol{Z}}^2))\mathbb{I}\left(\mathcal{E}\right)}{m^2},$$

provided the latter limit exists. We now focus on analyzing the RHS. We expand

$$\mathsf{Tr}(\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^\mathsf{T}-\widetilde{\boldsymbol{Z}}^2)\cdot\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^\mathsf{T}-\widetilde{\boldsymbol{Z}}^2))=$$

$$\sum_{\substack{a_{1:2k+2}\in[m]\\ a_1\neq a_{2k+2}\\ a_{k+1}\neq a_{k+2}}}(\boldsymbol{\Psi})_{a_1,a_2}q_1(\widetilde{z}_{a_2})\cdots(\boldsymbol{\Psi})_{a_k,a_{k+1}}\widetilde{z}_{a_{k+1}}\widetilde{z}_{a_{k+2}}(\boldsymbol{\Psi})_{a_{k+2},a_{k+3}}q_1(\widetilde{z}_{a_{k+3}})\cdots(\boldsymbol{\Psi})_{a_{2k+1},a_{2k+2}}\widetilde{z}_{a_{2k+2}}\widetilde{z}_{a_1}.$$

This can be written compactly in terms of matrix moments (Definition 8) as follows: Let $\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}(2k+2)$ denote the graph formed by combining two disconnected copies of the simple line graph on vertices $[1:k+1]$ and $[k+2:2k+2]$:

$$(\boldsymbol{\ell}_{k+1}^{\otimes 2})_{ij}=\begin{cases}1: & |i-j|=1,\ \{i,j\}\neq\{k+1,k+2\},\\ 0: & \text{otherwise}\end{cases}.$$

Recall the notation for partitions introduced in Section 5.6.1. Observe that:

$$\{(a_1\dots a_{2k+2})\in[m]^{2k+2}:\ a_1\neq a_{2k+2},\ a_{k+1}\neq a_{k+2}\}=\bigsqcup_{\pi\in\mathcal{P}_0([2k+2])}\mathcal{C}(\pi),$$

where,

$$\mathcal{P}_0([2k+2]) \stackrel{\text{def}}{=} \{\pi \in \mathcal{P}(2k+2) : \pi(1) \neq \pi(2k+2),\ \pi(k+1) \neq \pi(k+2)\}.$$

Recalling Definition 8, we have,

$$(\boldsymbol{\Psi})_{a_1,a_2} \cdots (\boldsymbol{\Psi})_{a_k,a_{k+1}} (\boldsymbol{\Psi})_{a_{k+2},a_{k+3}} \cdots (\boldsymbol{\Psi})_{a_{2k+1},a_{2k+2}} = \mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})$$

Hence,

$$\frac{\mathbb{E}\ \mathsf{Tr}(\mathcal{A} \cdot (\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2) \cdot \mathcal{A} \cdot (\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2))\mathbb{I}(\mathcal{E})}{m^2} =$$
$$\frac{1}{m^2} \sum_{\substack{\pi \in \mathcal{P}_0(2k+2) \\ \boldsymbol{a} \in \mathcal{C}(\pi)}} \mathbb{E}\ \mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a}) \cdot (\widetilde{z}_{a_1} q_1(\widetilde{z}_{a_2}) \cdots \widetilde{z}_{a_{k+1}} \widetilde{z}_{a_{k+2}} q_1(\widetilde{z}_{a_{k+3}}) \cdots \widetilde{z}_{a_{2k+2}}) \cdot \mathbb{I}(\mathcal{E}).$$

By the tower property,

$$\mathbb{E}\ \mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a}) \cdot (\widetilde{z}_{a_1} q_1(\widetilde{z}_{a_2}) \cdots \widetilde{z}_{a_{k+1}} \widetilde{z}_{a_{k+2}} q_1(\widetilde{z}_{a_{k+3}}) \cdots \widetilde{z}_{a_{2k+2}}) \cdot \mathbb{I}(\mathcal{E}) =$$
$$\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a}) \cdot \mathbb{E}[\widetilde{z}_{a_1} q_1(\widetilde{z}_{a_2}) \cdots \widetilde{z}_{a_{k+1}} \widetilde{z}_{a_{k+2}} q_1(\widetilde{z}_{a_{k+3}}) \cdots \widetilde{z}_{a_{2k+2}} | \boldsymbol{A}] \mathbb{I}(\mathcal{E})\right].$$

We will now use Mehler's formula (Proposition 15) to evaluate $\mathbb{E}[\cdots|\boldsymbol{A}]$ upto leading order. Note that some of the random variables $\widetilde{z}_{a_{1:2k+2}}$ are equal (as given by the partition π). Hence we group them together and recenter the resulting functions. The blocks corresponding to $a_1, a_{k+1}, a_{k+2}, a_{2k+2}$ need to be treated specially due to the presence of $\widetilde{z}_{a_1}, \widetilde{z}_{a_{k+1}}, \widetilde{z}_{a_{k+2}}, \widetilde{z}_{a_{2k+2}}$ in the above expectations. Hence, we introduce the following notations: We introduce the following notations:

$$\mathscr{F}_1(\pi) = \pi(1),\ \mathscr{L}_1(\pi) = \pi(k+1),\ \mathscr{F}_2(\pi) = \pi(k+2),\ \mathscr{L}_2(\pi) = \pi(2k+2)$$
$$\mathscr{S}(\pi) = \{i \in [1:2k+2] \backslash \{1, k+1, k+2, 2k+2\} : |\pi(i)| = 1\}.$$

We label all the remaining blocks of π as $\mathcal{V}_1, \mathcal{V}_2 \dots \mathcal{V}_{|\pi|-|\mathscr{S}(\pi)|-4}$. Hence the partition π is given by:

$$\pi = \mathscr{F}_1(\pi) \sqcup \mathscr{L}_1(\pi) \sqcup \mathscr{F}_2(\pi) \sqcup \mathscr{L}_2(\pi) \sqcup \left(\bigsqcup_{i \in \mathscr{S}(\pi)} \{i\} \right) \sqcup \left(\bigsqcup_{t=1}^{|\pi|-|\mathscr{S}(\pi)|-4} \mathcal{V}_i \right).$$

To simplify notation, we additionally define:

$$q_{k+1+i}(\xi) \stackrel{\text{def}}{=} q_i(\xi),\ i = 1, 2 \dots k-1.$$

Note that:

$$\widetilde{z}_{a_1} \widetilde{z}_{a_{k+1}} \widetilde{z}_{a_{k+2}} \widetilde{z}_{a_{2k+2}} \prod_{\substack{i=1 \\ i \neq k, k+1}}^{2k} q_i(\widetilde{z}_{a_{i+1}}) =$$

$$Q_{\mathscr{F}_1}(\widetilde{z}_{a_1}) Q_{\mathscr{L}_1}(\widetilde{z}_{a_{k+1}}) Q_{\mathscr{F}_2}(\widetilde{z}_{a_{k+2}}) Q_{\mathscr{L}_2}(\widetilde{z}_{a_{2k+2}}) \left(\prod_{i \in \mathscr{S}(\pi)} q_{i-1}(\widetilde{z}_{a_i}) \right) \prod_{i=1}^{|\pi|-|\mathscr{S}(\pi)|-4} (Q_{\mathcal{V}_i}(z_{a_{\mathcal{V}_i}}) + \mu_{\mathcal{V}_i}),$$

where,

$$Q_{\mathscr{F}_1}(\xi) = \xi \cdot \prod_{i \in \mathscr{F}_1(\pi), i \neq 1} q_{i-1}(\xi),$$

$$Q_{\mathscr{L}_1}(\xi) = \xi \cdot \prod_{i \in \mathscr{L}_1(\pi), i \neq k+1} q_{i-1}(\xi),$$

$$Q_{\mathscr{F}_2}(\xi) = \xi \cdot \prod_{i \in \mathscr{F}_2(\pi), i \neq k+2} q_{i-1}(\xi),$$

$$Q_{\mathscr{L}_2}(\xi) = \xi \cdot \prod_{i \in \mathscr{L}_2(\pi), i \neq 2k+2} q_{i-1}(\xi),$$

$$\mu_{\mathcal{V}_i} = \mathbb{E}_{\xi \sim \mathcal{N}(0,1)} \left[\prod_{j \in \mathcal{V}_i} q_{j-1}(\xi) \right],$$

$$Q_{\mathcal{V}_i}(\xi) = \prod_{j \in \mathcal{V}_i} q_{j-1}(\xi) - \mu_{\mathcal{V}_i},$$

With this notation in place we can apply Mehler's formula. The result is summarized in the following lemma.

Lemma 51. *For any $\pi \in \mathcal{P}_0([2k+2])$ and any $\boldsymbol{a} \in \mathcal{C}(\pi)$ we have,*

$$\mathbb{I}\left(\mathcal{E}\right)\left|\mathbb{E}[\widetilde{z}_{a_1}q_1(\widetilde{z}_{a_2})\cdots\widetilde{z}_{a_{k+1}}\widetilde{z}_{a_{k+2}}q_1(\widetilde{z}_{a_{k+3}})\cdots\widetilde{z}_{a_{2k+2}}|\boldsymbol{A}] - \sum_{\boldsymbol{w}\in\mathcal{G}_2(\pi)} G(\boldsymbol{w},\pi)\cdot\mathcal{M}(\boldsymbol{\Psi},\boldsymbol{w},\pi,\boldsymbol{a})\right| \\ \leq C(\mathcal{A})\cdot\left(\frac{\log^3(m)}{m\kappa^2}\right)^{\frac{3+|\mathscr{S}(\pi)|}{2}},$$

where, $\mathcal{M}(\boldsymbol{\Psi},\boldsymbol{w},\pi,\boldsymbol{a})$ is the matrix moment as defined in Definition 8,

$$G(\boldsymbol{w},\pi) = \frac{1}{\kappa^{\|\boldsymbol{w}\|}\boldsymbol{w}!}\left(\hat{Q}_{\mathscr{F}_1}(1)\hat{Q}_{\mathscr{L}_1}(1)\hat{Q}_{\mathscr{F}_2}(1)\hat{Q}_{\mathscr{L}_2}(1)\prod_{i\in\mathscr{S}(\pi)}\hat{q}_{i-1}(2)\right)\left(\prod_{i\in[|\pi|-|\mathscr{S}(\pi)|-4]}\mu_{\mathcal{V}_i}\right)$$

$$\mathcal{G}_2(\pi) \stackrel{def}{=} \big\{\boldsymbol{w}\in\mathcal{G}(2k+2) : \mathsf{d}_i(\boldsymbol{w}) = 1\ \forall\ i\ \in\ \{1,k+1,k+2,2k+2\}, \\ \mathsf{d}_i(\boldsymbol{w}) = 2\ \forall\ i\ \in\ \mathscr{S}(\pi),\ \mathsf{d}_i(\boldsymbol{w}) = 0\ \forall\ i\ \notin\ \{1,k+1,k+2,2k+2\}\cup\mathscr{S}(\pi)\big\},$$

The proof of the lemma involves instantiating Mehler's formula for this situation and identifying the leading order term. Since the proof is analogous to the proof of Lemma 28 provided in Appendix C.4.3, we omit it.

We return to our analysis of:

$$\frac{\mathbb{E}\ \mathsf{Tr}(\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2)\cdot\mathcal{A}\cdot(\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2))\mathbb{I}\left(\mathcal{E}\right)}{m^2} = \\ \frac{1}{m^2}\sum_{\substack{\pi\in\mathcal{P}_0(2k+2)\\ \boldsymbol{a}\in\mathcal{C}(\pi)}}\mathbb{E}\ \mathcal{M}(\boldsymbol{\Psi},\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi,\boldsymbol{a})\cdot(\widetilde{z}_{a_1}q_1(\widetilde{z}_{a_2})\cdots\widetilde{z}_{a_{k+1}}\widetilde{z}_{a_{k+2}}q_1(\widetilde{z}_{a_{k+3}})\cdots\widetilde{z}_{a_{2k+2}})\cdot\mathbb{I}\left(\mathcal{E}\right).$$

We define the following subsets of $\mathcal{P}_0(2k+2)$ as:

$$\mathcal{P}_1([2k+2]) \stackrel{\text{def}}{=} \big\{\pi \in \mathcal{P}_0(2k+2) : \ |\pi(i)| = 1, \ \forall\, i \ \in \ \{1, k+1, k+2, 2k+2\}, \qquad |\pi(j)| \leq 2 \ \forall\, j \ \in \ [k+1]\big\}, \tag{C.4a}$$

$$\mathcal{P}_2([2k+2]) \stackrel{\text{def}}{=} \mathcal{P}_0([2k+2]) \backslash \mathcal{P}_1([2k+2]), \tag{C.4b}$$

and the error term which was controlled in Lemma 28:

$$\epsilon(\boldsymbol{\Psi}, \boldsymbol{a}) \stackrel{\text{def}}{=}$$
$$\mathbb{I}\left(\mathcal{E}\right) \left(\mathbb{E}[\widetilde{z}_{a_1} q_1(\widetilde{z}_{a_2}) \cdots \widetilde{z}_{a_{k+1}} \widetilde{z}_{a_{k+2}} q_1(\widetilde{z}_{a_{k+3}}) \cdots \widetilde{z}_{a_{2k+2}} | \boldsymbol{A}] - \sum_{\boldsymbol{w} \in \mathcal{G}_2(\pi)} G(\boldsymbol{w}, \pi) \cdot \mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) \right)$$

.

With these definitions we consider the decomposition:

$$\frac{\mathbb{E}\ \mathsf{Tr}(\mathcal{A} \cdot (\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2) \cdot \mathcal{A} \cdot (\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}} - \widetilde{\boldsymbol{Z}}^2)) \mathbb{I}\left(\mathcal{E}\right)}{m^2} =$$
$$\frac{1}{m^2} \sum_{\pi \in \mathcal{P}_1([2k+2])} \sum_{a \in \mathcal{C}(\pi)} \sum_{\boldsymbol{w} \in \mathcal{G}_2(\pi)} G(\boldsymbol{w}, \pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right] - \mathsf{I} + \mathsf{II} + \mathsf{III},$$

where,

$$\mathsf{I} = \frac{1}{m^2} \sum_{\pi \in \mathcal{P}_0([2k+2])} \sum_{a \in \mathcal{C}(\pi)} \sum_{\boldsymbol{w} \in \mathcal{G}_2(\pi)} G(\boldsymbol{w}, \pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a}) \mathbb{I}\left(\mathcal{E}^c\right)\right],$$
$$\mathsf{II} = \frac{1}{m^2} \sum_{\pi \in \mathcal{P}_0(2k+2])} \sum_{a \in \mathcal{C}(\pi)} \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a}) \epsilon(\boldsymbol{\Psi}, \boldsymbol{a}) \mathbb{I}\left(\mathcal{E}\right)\right],$$
$$\mathsf{III} = \frac{1}{m^2} \sum_{\pi \in \mathcal{P}_2([2k+2])} \sum_{a \in \mathcal{C}(\pi)} \sum_{\boldsymbol{w} \in \mathcal{G}_2(\pi)} G(\boldsymbol{w}, \pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right].$$

We will show that $\mathsf{I}, \mathsf{II}, \mathsf{III} \to 0$. Showing this involves the following components:

1. Bounds on matrix moments $\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right]$ which have been developed in Lemma 18.

2. Controlling the size of the set $|\mathcal{C}(\pi)|$ (since we sum over $\boldsymbol{a} \in \mathcal{C}(\pi)$ in the above terms). Since,

$$|\mathcal{C}(\pi)| = m(m-1)\cdots(m-|\pi|+1) \asymp m^{|\pi|},$$

we need to develop bounds on $|\pi|$. This is done in the following lemma. In contrast, the sums over $\pi \in \mathcal{P}_0([2k+2])$ and $\boldsymbol{w} \in \mathcal{G}_1(\pi)$ are not a cause of concern since $|\mathcal{P}_0([2k+2])|, |\mathcal{G}_1(\pi)|$ depend only on k (which is held fixed) and not on m.

Lemma 52. *For any $\pi \in \mathcal{P}_1([2k+2])$ we have,*

$$|\pi| = \frac{2k+6+|\mathscr{S}(\pi)|}{2} \implies |\mathcal{C}(\pi)| \leq m^{\frac{2k+6+|\mathscr{S}(\pi)|}{2}}.$$

For any $\pi \in \mathcal{P}_2([2k+2])$, we have,

$$|\pi| \leq \frac{2k+5+|\mathscr{S}(\pi)|}{2} \implies |\mathcal{C}(\pi)| \leq m^{\frac{2k+5+|\mathscr{S}(\pi)|}{2}}.$$

Proof. Consider any $\pi \in \mathcal{P}_0([2k+2])$. Recall that the disjoint blocks of $|\pi|$ were given by:

$$\pi = \mathscr{F}_1(\pi) \sqcup \mathscr{L}_1(\pi) \sqcup \mathscr{F}_2(\pi) \sqcup \mathscr{L}_2(\pi) \sqcup \left(\bigsqcup_{i \in \mathscr{S}(\pi)} \{i\} \right) \sqcup \left(\bigsqcup_{t=1}^{|\pi|-|\mathscr{S}(\pi)|-4} \mathcal{V}_i \right).$$

Hence,

$$2k+2 = |\mathscr{F}_1(\pi)| + |\mathscr{F}_2(\pi)| + |\mathscr{L}_1(\pi)| + |\mathscr{L}_2(\pi)| + |\mathscr{S}(\pi)| + \sum_{t=1}^{|\pi|-|\mathscr{S}(\pi)|-4} |\mathcal{V}_i|.$$

Note that:

$$|\mathscr{F}_1(\pi)| \geq 1 \qquad \text{(Since } 1 \in \mathscr{F}_1(\pi)\text{)} \tag{C.5a}$$

$$|\mathscr{F}_2(\pi)| \geq 1 \qquad \text{(Since } k+2 \in \mathscr{F}_2(\pi)\text{)} \tag{C.5b}$$

$$|\mathscr{L}_1(\pi)| \geq 1 \qquad \text{(Since } k+1 \in \mathscr{L}_1(\pi)\text{)} \tag{C.5c}$$

$$|\mathscr{L}_2(\pi)| \geq 1 \qquad \text{(Since } 2k+2 \in \mathscr{L}_1(\pi)\text{)} \tag{C.5d}$$

$$|\mathcal{V}_i| \geq 2 \qquad \text{(Since } \mathcal{V}_i \text{ are not singletons).} \tag{C.5e}$$

Hence,

$$2k+2 \geq 4 + 2|\pi| - |\mathscr{S}(\pi)| - 8,$$

which implies,

$$|\pi| \leq \frac{2k+6+|\mathscr{S}(\pi)|}{2}, \tag{C.6}$$

and hence,

$$|\mathcal{C}(\pi)| \leq m^{|\pi|} \leq m^{\frac{2k+6+|\mathscr{S}(\pi)|}{2}}.$$

Finally observe that:

1. For any $\pi \in \mathcal{P}_2([2k+2])$ each of the inequalities in (C.5) are exactly tight by the definition of $\mathcal{P}_1([k+1])$ in (C.4), and hence,

$$|\pi| = \frac{2k+6+|\mathscr{S}(\pi)|}{2}.$$

2. For any $\pi \in \mathcal{P}_2([2k+2])$, one of the inequalities in (C.5) must be strict (see (C.4)). Hence,

when $\pi \in \mathcal{P}_2([k+1])$ we have the improved bound:

$$|\pi| \leq \frac{2k+5+|\mathscr{S}(\pi)|}{2}.$$

This proves the claims of the lemma. □

We will now show that $\mathsf{I}, \mathsf{II}, \mathsf{III} \to 0$.

Lemma 53. *We have,*

$$\mathsf{I} \to 0,\ \mathsf{II} \to 0,\ \mathsf{III} \to 0 \ \text{ as } m \to \infty,$$

and hence,

$$\lim_{m\to\infty} \frac{\mathbb{E}(\boldsymbol{z}^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z})\boldsymbol{z})^2}{m^2} = \lim_{m\to\infty} \frac{1}{m^2} \sum_{\pi\in\mathcal{P}_1([2k+2])} \sum_{a\in\mathcal{C}(\pi)} \sum_{\boldsymbol{w}\in\mathcal{G}_2(\pi)} G(\boldsymbol{w},\pi)\cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right],$$

provided the latter limit exists.

Proof. First note that for any $\boldsymbol{w} \in \mathcal{G}_1(\pi)$, we have,

$$\|\boldsymbol{w}\| = \frac{1}{2}\sum_{i=1}^{2k+2} \mathsf{d}_i(\boldsymbol{w}) = \frac{1+1+1+1+2|\mathscr{S}(\pi)|}{2} = 2+|\mathscr{S}(\pi)| \ \text{ (See Lemma 51).}$$

Furthermore recalling the definition of $\boldsymbol{\ell}_{k+1}^{\otimes 2}$, $\|\boldsymbol{\ell}_{k+1}^{\otimes 2}\| = 2k$. Now we apply Lemma 18 to obtain:

$$
\begin{aligned}
|\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\mathbb{I}\left(\mathcal{E}^c\right)\right]| &\leq \sqrt{\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, 2\boldsymbol{w} + 2\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right]}\sqrt{\mathbb{P}(\mathcal{E}^c)} \\
&\leq \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+2+2k}{2}} \cdot \sqrt{\mathbb{P}(\mathcal{E}^c)}, \\
&\overset{\text{(a)}}{\leq} \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+2+2k}{2}} \cdot \frac{C_k}{m}. \\
\mathbb{E}|\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})| &\leq \left(\frac{C_k \log^2(m)}{m}\right)^{k}, \\
\mathbb{E}\left[|\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w} + \boldsymbol{\ell}_{k+1}, \pi, \boldsymbol{a})|\right] &\leq \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+2+2k}{2}}
\end{aligned}
$$

In the step marked (a) we used Lemma 21. Further recall that by Lemma 28 we have,

$$
|\epsilon(\boldsymbol{\Psi}, \boldsymbol{a})| \leq C(\mathcal{A}) \cdot \left(\frac{\log^3(m)}{m\kappa^2}\right)^{\frac{3+|\mathscr{S}(\pi)|}{2}}.
$$

Using these estimates, we obtain,

$$
\begin{aligned}
|\mathsf{I}| &\leq \frac{C(\mathcal{A})\cdot}{m^2} \cdot \sum_{\pi:\mathcal{P}_0([2k+2])} |\mathcal{C}(\pi)| \cdot \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+2+2k}{2}} \cdot \frac{C_k}{m} \\
&\leq \frac{C(\mathcal{A})\cdot}{m^2} \cdot \sum_{\pi:\mathcal{P}_0([2k+2])} m^{\frac{2k+6+|\mathscr{S}(\pi)|}{2}} \cdot \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+2+2k}{2}} \cdot \frac{C_k}{m} \\
&= O\left(\frac{\mathrm{polylog}(m)}{m}\right) \\
|\mathsf{II}| &\leq \frac{C(\mathcal{A})}{m^2} \cdot \left(\frac{C_k \log^2(m)}{m}\right)^{k} \cdot \sum_{\pi:\mathcal{P}_0([2k+2])} |\mathcal{C}(\pi)| \cdot \left(\frac{\log^3(m)}{m\kappa^2}\right)^{\frac{3+|\mathscr{S}(\pi)|}{2}} \\
&\leq \frac{C(\mathcal{A})}{m^2} \cdot \left(\frac{C_k \log^2(m)}{m}\right)^{k} \cdot \sum_{\pi:\mathcal{P}_0([2k+2])} m^{\frac{2k+6+|\mathscr{S}(\pi)|}{2}} \cdot \left(\frac{\log^3(m)}{m\kappa^2}\right)^{\frac{3+|\mathscr{S}(\pi)|}{2}} \\
&= O\left(\frac{\mathrm{polylog}(m)}{\sqrt{m}}\right) \\
|\mathsf{III}| &\leq \frac{C(\mathcal{A})\cdot}{m^2} \cdot \sum_{\pi:\mathcal{P}_2([2k+2])} |\mathcal{C}(\pi)| \cdot \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+1+k}{2}} \\
&\leq \frac{C(\mathcal{A})\cdot}{m^2} \cdot \sum_{\pi:\mathcal{P}_2([2k+2])} m^{\frac{2k+5+|\mathscr{S}(\pi)|}{2}} \cdot \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+2+2k}{2}} \\
&= O\left(\frac{\mathrm{polylog}(m)}{\sqrt{m}}\right).
\end{aligned}
$$

This concludes the proof of this lemma. □

Next, we consider the decomposition:

$$
\begin{aligned}
&\frac{1}{m^2} \sum_{\pi\in\mathcal{P}_1([2k+2])} \sum_{a\in\mathcal{C}(\pi)} \sum_{\boldsymbol{w}\in\mathcal{G}_2(\pi)} G(\boldsymbol{w},\pi)\cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right] = \\
&\frac{1}{m^2} \sum_{\pi\in\mathcal{P}_1([2k+2])} \sum_{\substack{\boldsymbol{w}\in\mathcal{G}_2(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}_{\mathsf{DA}}(\pi)}} \sum_{a\in\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)} G(\boldsymbol{w},\pi)\cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right] + \mathsf{IV} + \mathsf{V},
\end{aligned}
$$

where,

$$\mathsf{IV} \stackrel{\text{def}}{=} \frac{1}{m^2} \sum_{\pi \in \mathcal{P}_1([2k+2])} \sum_{a \in \mathcal{C}(\pi)} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_2(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2} \notin \mathcal{G}_{\mathsf{DA}}(\pi)}} G(\boldsymbol{w}, \pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right],$$

$$\mathsf{V} \stackrel{\text{def}}{=} \frac{1}{m^2} \sum_{\pi \in \mathcal{P}_1([2k+2])} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_2(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2} \in \mathcal{G}_{\mathsf{DA}}(\pi)}} \sum_{a \in \mathcal{C}(\pi) \setminus \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi)} G(\boldsymbol{w}, \pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right].$$

Lemma 54. *We have,* $\mathsf{IV} \to 0, \mathsf{V} \to 0$ *as* $m \to \infty$*, and hence,*

$$\lim_{m\to\infty} \frac{\mathbb{E}(\boldsymbol{z}^{\mathsf{T}}\mathcal{A}\boldsymbol{z})^2}{m^2} = \lim_{m\to\infty} \frac{1}{m^2} \sum_{\pi \in \mathcal{P}_1([2k+2])} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_2(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2} \in \mathcal{G}_{\mathsf{DA}}(\pi)}} \sum_{a \in \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi)} G(\boldsymbol{w}, \pi) \cdot \mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right],$$

provided the latter limit exists.

Proof. We will prove this in two steps.

Step 1: $\mathsf{IV} \to 0$**.** We consider the two sensing models separately:

1. Subsampled Hadamard Sensing: In this case, Proposition 17 tells us that if $\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2} \notin \mathcal{G}_{\mathsf{DA}}(\pi)$, then,

$$\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right] = 0$$

   and hence $\mathsf{IV} = 0$.

2. Subsampled Haar Sensing: Observe that, since $\|\boldsymbol{w}\| + \|\boldsymbol{\ell}_{k+1}^{\otimes 2}\| = 2 + |\mathscr{S}(\pi)| + 2k$, we have,

$$\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right] = \frac{\mathbb{E}\left[\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right]}{m^{\frac{2+|\mathscr{S}(\pi)|+2k}{2}}}.$$

By Proposition 16 we know that,

$$\left|\mathbb{E}\left[\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right] - \prod_{\substack{s,t\in[|\pi|]\\ s\le t}} \mathbb{E}\left[Z_{st}^{W_{st}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)}\right]\right| \le \frac{K_1 \log^{K_2}(m)}{m^{\frac{1}{4}}},$$

$\forall\, m \ge K_3$, where K_1, K_2, K_3 are universal constants depending only on k. Note that since $\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2} \notin \mathcal{G}_{\mathsf{DA}}(\pi)$, must have some $s \in [|\pi|]$ such that:

$$W_{ss}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi) \ge 1.$$

Recall that, $\mathsf{d}_i(\boldsymbol{w}) = 0$ for any $i \notin \{1, k+1, k+2, 2k+2\} \cup \mathscr{S}(\pi)$ (since $\boldsymbol{w} \in \mathcal{G}_2(\pi)$) and furthermore, $|\pi(i)| = 1 \forall\ i \in \{1, k+1, k+2, 2k+2\} \cup \mathscr{S}(\pi)$ (since $\pi \in \mathcal{P}_1(2k+2)$). Hence, we have $\boldsymbol{w} \in \mathcal{G}_{\mathsf{DA}}(\pi)$ and in particular, $W_{ss}(\boldsymbol{w}, \pi) = 0$. Consequently, we must have $W_{ss}(\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi) \ge 1$. Recall the definition of $\boldsymbol{\ell}_{k+1}^{\otimes 2}$, since $W_{ss}(\boldsymbol{\ell}_{k+1}, \pi) \ge 1$ we must have that for some $i \in [2k+2]$, we have, $\pi(i) = \pi(i+1) = \mathcal{V}_s$. However, since $\pi \in \mathcal{P}_1(2k+2)$, $|\mathcal{V}_s| \le 2$, and hence $\mathcal{V}_s = \{i, i+1\}$. This means that $W_{ss}(\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi) = 1 = W_{ss}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi)$. Consequently since $\mathbb{E}Z_{ss} = 0$, we have,

$$\prod_{\substack{s,t\in[|\pi|]\\ s\le t}} \mathbb{E}\left[Z_{st}^{W_{st}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)}\right] = 0,$$

or,

$$\left|\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right]\right| = \frac{\mathrm{polylog}(m)}{m^{\frac{2+|\mathscr{S}(\pi)|+2k}{2}+\frac{1}{4}}}.$$

Recalling Lemma 52,

$$|\mathcal{C}(\pi)| \le m^{|\pi|} \le m^{\frac{2k+6+|\mathscr{S}(\pi)|}{2}},$$

we obtain,

$$|\mathsf{IV}| \leq \frac{C(\mathcal{A})}{m^2} \sum_{\pi \in \mathcal{P}_1([2k+2])} |\mathcal{C}(\pi)| \cdot \frac{\text{polylog}(m)}{m^{\frac{2+|\mathscr{S}(\pi)|+2k}{2}+\frac{1}{4}}} = O\left(\frac{\text{polylog}(m)}{m^{\frac{1}{4}}}\right) \to 0.$$

Step 2: $\mathsf{V} \to 0$. Using Lemma 20, we know that

$$|\mathcal{C}(\pi) \backslash \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi)| \leq O(m^{|\pi|-1})$$

In Lemma 52, we showed that for any $\pi \in \mathcal{P}_1([k+1])$,

$$|\pi| = \frac{2k+6+|\mathscr{S}(\pi)|}{2}.$$

Hence,

$$|\mathcal{C}(\pi) \backslash \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi)| \leq O(m^{\frac{2k+4+|\mathscr{S}(\pi)|}{2}}).$$

We already know from Lemma 18 that,

$$|\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi, \boldsymbol{a})\right]| \leq \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{\|\boldsymbol{w}\|+\|\boldsymbol{\ell}_{k+1}^{\otimes 2}\|}{2}} \leq \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+2+2k}{2}},$$

This gives us:

$$\begin{aligned}|\mathsf{V}| &\leq \frac{C}{m^2} \sum_{\pi \in \mathcal{P}_1([2k+2])} \sum_{\substack{\boldsymbol{w} \in \mathcal{G}_2(\pi) \\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2} \in \mathcal{G}_{\mathsf{DA}}(\pi)}} |\mathcal{C}(\pi) \backslash \mathcal{L}_{\mathsf{CF}}(\boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi)| \left(\frac{C_k \log^2(m)}{m}\right)^{\frac{|\mathscr{S}(\pi)|+2+2k}{2}} \\ &= O\left(\frac{\text{polylog}(m)}{m}\right)\end{aligned}$$

which goes to zero as claimed.

This concludes the proof of the lemma. □

So far we have shown that:

$$\lim_{m\to\infty} \frac{\mathbb{E}(\boldsymbol{z}^\mathsf{T}\mathcal{A}\boldsymbol{z})^2}{m^2} =$$
$$\lim_{m\to\infty} \frac{1}{m^2} \sum_{\pi\in\mathcal{P}_1([2k+2])} \sum_{\substack{\boldsymbol{w}\in\mathcal{G}_2(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}_{\mathsf{DA}}(\pi)}} \sum_{a\in\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)} G(\boldsymbol{w},\pi)\cdot\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi},\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi,\boldsymbol{a})\right].$$

provided the latter limit exists. In the following lemma we explicitly calculate the limit on the RHS and hence show that it exists and is same for the subsampled Haar and subsampled Hadamard sensing models.

Lemma 55. *For both the subsampled Haar sensing and Hadamard sensing model, we have,*

$$\lim_{m\to\infty} \frac{\mathbb{E}(\boldsymbol{z}^\mathsf{T}\mathcal{A}\boldsymbol{z})^2}{m^2} = \sum_{\pi\in\mathcal{P}_1([2k+2])} \sum_{\substack{\boldsymbol{w}\in\mathcal{G}_2(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}_{\mathsf{DA}}(\pi)}} G(\boldsymbol{w},\pi)\cdot\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi),$$

where,

$$\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi) \stackrel{def}{=} \prod_{\substack{s,t\in[|\pi|]\\ s<t}} \mathbb{E}\left[Z^{W_{st}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)}\right],\ Z\sim\mathcal{N}\left(0,\kappa(1-\kappa)\right).$$

Proof. By Propositions 17 (for the subsampled Hadamard model) and 16 (for the subsampled Haar model) we know that, if $\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}_{\mathsf{DA}}(\pi),\ \boldsymbol{a}\in\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)$, we have,

$$\mathcal{M}(\sqrt{m}\boldsymbol{\Psi},\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi,\boldsymbol{a}) = \mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)+\epsilon(\boldsymbol{w},\pi,\boldsymbol{a}),$$

where

$$|\epsilon(\boldsymbol{w},\pi,\boldsymbol{a})| \le \frac{K_1\log^{K_2}(m)}{m^{\frac{1}{4}}},\ \forall\ m\ge K_3,$$

for some constants K_1, K_2, K_3 depending only on k. Hence, we can consider the decomposition:

$$\frac{1}{m^2}\sum_{\pi\in\mathcal{P}_1([2k+2])}\sum_{\substack{\boldsymbol{w}\in\mathcal{G}_2(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}_{\mathsf{DA}}(\pi)}}\sum_{a\in\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)} G(\boldsymbol{w},\pi)\cdot\mathbb{E}\left[\mathcal{M}(\boldsymbol{\Psi},\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi,\boldsymbol{a})\right]$$

$$= \mathsf{VI}+\mathsf{VII},$$

where,

$$\mathsf{VI} \stackrel{\text{def}}{=} \frac{1}{m^2}\sum_{\pi\in\mathcal{P}_1([2k+2])}\sum_{\substack{\boldsymbol{w}\in\mathcal{G}_2(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}_{\mathsf{DA}}(\pi)}}\sum_{a\in\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)} G(\boldsymbol{w},\pi)\cdot\frac{\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)}{m^{\frac{2+\mathscr{S}(\pi)+2k}{2}}},$$

$$\mathsf{VII} \stackrel{\text{def}}{=} \frac{1}{m^2}\sum_{\pi\in\mathcal{P}_1([2k+2])}\sum_{\substack{\boldsymbol{w}\in\mathcal{G}_2(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}_{\mathsf{DA}}(\pi)}}\sum_{a\in\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)} G(\boldsymbol{w},\pi)\cdot\frac{\epsilon(\boldsymbol{w},\pi,\boldsymbol{a})}{m^{\frac{2+\mathscr{S}(\pi)+2k}{2}}}$$

We can upper bound $|\mathsf{VII}|$ as follows:

$$|\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)| \le |\mathcal{C}(\pi)| \le m^{\frac{2k+6+|\mathscr{S}(\pi)|}{2}},$$

$$|\mathsf{VII}| \le \frac{C(\mathcal{A})}{m^2}\cdot C_k\cdot|\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)|\cdot\frac{1}{m^{\frac{2+|\mathscr{S}(\pi)|+2k}{2}}}\cdot\frac{K_1\log^{K_2}(m)}{m^{\frac{1}{4}}}$$

$$= O\left(\frac{\mathrm{polylog}(m)}{m^{\frac{1}{4}}}\right)\to 0.$$

We can compute:

$$\lim_{m\to\infty}(\mathsf{VI}) = \lim_{m\to\infty}\frac{1}{m^2}\sum_{\pi\in\mathcal{P}_1([2k+2])}\sum_{\substack{\boldsymbol{w}\in\mathcal{G}_2(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}_{\mathsf{DA}}(\pi)}}\sum_{a\in\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)} G(\boldsymbol{w},\pi)\cdot\frac{\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)}{m^{\frac{2+\mathscr{S}(\pi)+2k}{2}}}$$

$$= \lim_{m\to\infty}\frac{1}{m^2}\sum_{\pi\in\mathcal{P}_1([2k+2])}\sum_{\substack{\boldsymbol{w}\in\mathcal{G}_2(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}_{\mathsf{DA}}(\pi)}} G(\boldsymbol{w},\pi)\cdot\frac{\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)}{m^{\frac{2+\mathscr{S}(\pi)+2k}{2}}}\cdot|\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)|$$

$$= \sum_{\pi\in\mathcal{P}_1([2k+2])}\sum_{\substack{\boldsymbol{w}\in\mathcal{G}_2(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}_{\mathsf{DA}}(\pi)}} G(\boldsymbol{w},\pi)\cdot\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)\cdot\frac{m^{|\pi|}}{m^{\frac{6+\mathscr{S}(\pi)+2k}{2}}}\cdot\frac{|\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)|}{m^{|\pi|}}$$

$$\overset{\text{(a)}}{=} \sum_{\pi\in\mathcal{P}_1([2k+2])}\sum_{\substack{\boldsymbol{w}\in\mathcal{G}_2(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}_{\mathsf{DA}}(\pi)}} G(\boldsymbol{w},\pi)\cdot\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)\cdot\frac{|\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)|}{m^{|\pi|}}$$

$$\overset{\text{(b)}}{=} \sum_{\pi\in\mathcal{P}_1([2k+2])}\sum_{\substack{\boldsymbol{w}\in\mathcal{G}_2(\pi)\\ \boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}\in\mathcal{G}_{\mathsf{DA}}(\pi)}} G(\boldsymbol{w},\pi)\cdot\mu(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi).$$

In the step marked (a) we used the fact that $|\pi| = (6+|\mathscr{S}(\pi)|+2k)/2$ for any $\pi\in\mathcal{P}_1([2k+2])$ (Lemma 52) and in step (b) we used Lemma 20 ($|\mathcal{L}_{\mathsf{CF}}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2},\pi)|/m^{|\pi|}\to 1$). This proves the claim of the lemma and Proposition 19. □

We can actually significantly simply the combinatorial sum obtained in Lemma 55 which we do so in the following lemma.

Lemma 56. *For both the subsampled Haar sensing and Hadamard sensing models, we have,*

$$\lim_{m\to\infty}\frac{\mathbb{E}(\boldsymbol{z}^{\mathsf{T}}\mathcal{A}\boldsymbol{z})^2}{m^2} = (1-\kappa)^{2k}\cdot\prod_{i=1}^{k-1}\hat{q}_i^2(2).$$

In particular, Proposition 19 holds.

Proof. We claim that the only partition with a non-zero contribution is:

$$\pi = \bigsqcup_{i=1}^{2k+2}\{i\}.$$

In order to see this suppose π is not entirely composed of singleton blocks. Define:

$$i_\star \stackrel{\text{def}}{=} \min\{i \in [2k+2] : |\pi(i)| > 1\}.$$

Note $i_\star > 1$ since we know that $|\pi(1)| = |\mathscr{F}_1(\pi)| = 1$ for any $\pi \in \mathcal{P}_1(2k+2)$. Since $\pi \in \mathcal{P}_1([2k+2])$ we must have $|\pi(i_\star)| = 2$, hence denote:

$$\pi(i_\star) = \{i_\star, j_\star\}.$$

for some $j_\star > i_\star + 1$ ($i_\star \leq j_\star$ since it is the first index which is not in a singleton block, and $j_\star \neq i_\star + 1$ since otherwise $\boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}$ will not be disassortative. Similarly we know that $i_\star, j_\star \neq k+1, k+2, 2k+2$ because $|\pi(k+1)| = |\pi(k+2)| = |\pi(2k+2)| = 1$ since $\pi \in \mathcal{P}_1([2k+2])$. Let us label the first few blocks of π as:

$$\mathcal{V}_1 = \{1\},\ \mathcal{V}_2 = \{2\}, \ldots \mathcal{V}_{i_\star - 1} = \{i_\star - 1\},\ \mathcal{V}_{i_\star} = \{i_\star, j_\star\}.$$

Next we compute:

$$\begin{aligned}
W_{i_\star-1,i_\star}(\boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi) &= W_{i_\star-1,i_\star}(\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi) + W_{i_\star-1,i_\star}(\boldsymbol{w}, \pi) \\
&\stackrel{\text{(a)}}{=} W_{i_\star-1,i_\star}(\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi) \\
&\stackrel{\text{(b)}}{=} \mathbf{1}_{i_\star-1 \in \mathcal{V}_{i_\star-1}} + \mathbf{1}_{i_\star+1 \in \mathcal{V}_{i_\star-1}} + \mathbf{1}_{j_\star-1 \in \mathcal{V}_{i_\star-1}} + \mathbf{1}_{j_\star+1 \in \mathcal{V}_{i_\star-1}} \\
&\stackrel{\text{(c)}}{=} \mathbf{1}_{i_\star-1 = i_\star-1} + \mathbf{1}_{i_\star+1 = i_\star-1} + \mathbf{1}_{j_\star-1 = i_\star-1} + \mathbf{1}_{j_\star+1 = i_\star-1} \\
&\stackrel{\text{(d)}}{=} 1.
\end{aligned}$$

In the step marked (a), we used the fact that since $\boldsymbol{w} \in \mathcal{G}_2(\pi)$ and $|\pi(i_\star)| = |\pi(j_\star)| = 2$, we must have $d_{i_\star}(\boldsymbol{w}) = d_{j_\star}(\boldsymbol{w}) = 0$ and $W_{i_\star-1,i_\star}(\boldsymbol{w}, \pi) = 0$. In the step marked (b) we used the definition of $\boldsymbol{\ell}_{k+1}^{\otimes 2}$. In the step marked (c) we used the fact that $\mathcal{V}_{i_\star-1} = \{i_{\star-1}\}$. In the step marked (d) we used the fact that $j_\star > i_\star + 1$.

Hence we have shown that for any $\pi \neq \sqcup_{i=1}^{2k+2}\{i\}$, we have

$$\mu(\boldsymbol{w}, \pi) = 0 \ \forall \ \boldsymbol{w} \text{ such that } \boldsymbol{w} \in \mathcal{G}_2(\pi), \ \boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2} \in \mathcal{G}_{\mathsf{DA}}(\pi).$$

Next, let $\pi = \sqcup_{i=1}^{2k+2}\{i\}$. We observe for any $\boldsymbol{w}$ such that $\boldsymbol{w} \in \mathcal{G}_2(\pi), \ \boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2} \in \mathcal{G}_{\mathsf{DA}}(\pi)$, we have,

$$\begin{aligned}\mu(\boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi) &= \prod_{\substack{s,t \in [|\pi|] \\ s<t}} \mathbb{E}\left[Z^{W_{st}(\boldsymbol{w}+\boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi)}\right], \ Z \sim \mathcal{N}\left(0, \kappa(1-\kappa)\right) \\ &= \prod_{\substack{i,j \in [2k+2] \\ i<j}} \mathbb{E}\left[Z^{w_{ij} + (\ell_{k+1})_{ij}, \pi)}\right], \ Z \sim \mathcal{N}\left(0, \kappa(1-\kappa)\right)\end{aligned}$$

Note that since $\mathbb{E}Z = 0$, for $\mu(\boldsymbol{w} + \boldsymbol{\ell}_{k+1}^{\otimes 2}, \pi) \neq 0$ we must have:

$$w_{ij} \geq (\boldsymbol{\ell}_{k+1}^{\otimes 2})_{ij}, \ \forall \ i, j \ \in \ [2k+2].$$

However since $\boldsymbol{w} \in \mathcal{G}_2(\pi)$ we have,

$$\begin{aligned}&\mathsf{d}_1(\boldsymbol{w}) = \mathsf{d}_{k+1}(\boldsymbol{w}) = \mathsf{d}_{k+2}(\boldsymbol{w}) = \mathsf{d}_{2k+2}(\boldsymbol{w}) = 1, \\ &\mathsf{d}_i(\boldsymbol{w}) = 2 \ \forall \ i \ \in \ [2k+2] \backslash \{1, k+1, k+2, 2k+2\},\end{aligned}$$

hence $\boldsymbol{w} = \boldsymbol{\ell}_{k+1}^{\otimes 2}$. Hence, recalling the formula for $g(\boldsymbol{w}, \pi)$ from Lemma 28 we obtain:

$$\lim_{m \to \infty} \frac{\mathbb{E}(\boldsymbol{z}^{\mathsf{T}} \mathcal{A} \boldsymbol{z})^2}{m^2} = (1-\kappa)^{2k} \cdot \prod_{i=1}^{k-1} \hat{q}_i^2(2).$$

This proves the statement of the lemma and also Proposition 18 (see Remark 15 regarding how the analysis extends to other types). $\square$

C.3 Proofs from Section 5.6.4

C.3.1 Proof of Lemma 18

Proof of Lemma 18. Recall that,

$$\begin{aligned}
\mathbb{E}|\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})| &= \mathbb{E} \prod_{\substack{i,j \in [k] \\ i<j}} |\Psi_{a_i,a_j}^{w_{ij}}| \\
&\overset{(a)}{\leq} \sum_{\substack{i,j \in [k] \\ i<j}} \frac{w_{ij}}{\|\boldsymbol{w}\|} \mathbb{E}|\Psi_{a_i,a_j}^{\|\boldsymbol{w}\|_1}| \\
&\leq \max_{i,j \in [m]} \mathbb{E}|\Psi_{ij}|^{\|\boldsymbol{w}\|},
\end{aligned}$$

where step (a) follows from the AM-GM inequality. We now consider the subsampled Haar and Hadamard cases separately.

Hadamard Case: By Lemma 17, Ψ_{ij} is subgaussian with with variance proxy bounded by C/m for some universal constant C. Hence,

$$\mathbb{E}|\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})| \leq \left(\frac{C\|\boldsymbol{w}\|}{m}\right)^{\frac{\|\boldsymbol{w}\|}{2}}.$$

Haar Case: By Lemma 17, conditional on $\boldsymbol{O}$, Ψ_{ij} is sub-Gaussian. The variance proxy is bounded by $Cm\|\boldsymbol{o}_i\|_\infty^2\|\boldsymbol{o}_j\|_\infty^2$. Hence,

$$\begin{aligned}
\mathbb{E}|\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})| &\leq \max_{i,j \in [m]} \mathbb{E}|\Psi_{ij}|^{\|\boldsymbol{w}\|} \\
&= \max_{i,j \in [m]} \mathbb{E}[\mathbb{E}[|\Psi_{ij}|^{\|\boldsymbol{w}\|}|\boldsymbol{O}]] \\
&\leq \max_{i,j \in [m]} (C\|\boldsymbol{w}\|m)^{\frac{\|\boldsymbol{w}\|}{2}} \mathbb{E}\left[\|\boldsymbol{o}_i\|_\infty^{\|\boldsymbol{w}\|}\|\boldsymbol{o}_j\|_\infty^{\|\boldsymbol{w}\|}\right] \\
&\leq \max_{i,j \in [m]} (C\|\boldsymbol{w}\|m)^{\frac{\|\boldsymbol{w}\|}{2}} \left(\mathbb{E}\|\boldsymbol{o}_i\|_\infty^{2\|\boldsymbol{w}\|} + \mathbb{E}\|\boldsymbol{o}_j\|_\infty^{2\|\boldsymbol{w}\|}\right).
\end{aligned}$$

Note that $\boldsymbol{o}_i \overset{\text{d}}{=} \boldsymbol{o}_j \overset{\text{d}}{=} \boldsymbol{u} \sim \text{Unif}\,(\mathbb{S}_{m-1})$. Applying Fact 8 gives us,

$$\mathbb{E}|\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})| \leq \left(\sqrt{\frac{C\|\boldsymbol{w}\| \log^2(m)}{m}} \right)^{\|\boldsymbol{w}\|}.$$

□

C.3.2 Proofs of Propositions 16 and 17

This section is dedicated to the proof of Propositions 16 and 17. We consider the following general setup. Let $\boldsymbol{v}_1, \boldsymbol{v}_2 \cdots, \boldsymbol{v}_m$ be fixed vectors in $\mathbb{R}^d$ for a fixed $d \in \mathbb{N}$. Define the statistic:

$$\boldsymbol{T} = \sqrt{m} \sum_{i=1}^{m} \overline{B}_{ii} \boldsymbol{v}_i,$$

where $\overline{\boldsymbol{B}}$ denotes a diagonal matrix whose n diagonal entries are set to $1 - \kappa$ uniformly at random and the remaining $m - n$ are set to $-\kappa$.

Analogously, we define the statistic:

$$\hat{\boldsymbol{T}} = \sqrt{m} \sum_{i=1}^{m} \hat{B}_{ii} \boldsymbol{v}_i,$$

where,

$$\hat{B}_{ii} \overset{\text{i.i.d.}}{\sim} \begin{cases} 1 - \kappa : & \text{with prob. } \kappa \\ -\kappa : & \text{with prob. } 1 - \kappa \end{cases}.$$

As in the proof of Lemma 17 we $\overline{\boldsymbol{B}}$ and $\hat{\boldsymbol{B}}$ in the same probability space as follows:

1. We first sample $\overline{\boldsymbol{B}}$. Let $S = \{i \in [m] : \overline{B}_{ii} = 1 - \kappa\}$
2. Next sample $N \sim \mathsf{Binom}(m, \kappa)$.
3. Sample a subset $\hat{S} \subset [m]$ with $|\hat{S}| = N$ as follows:

- If $N \leq n$, then set $\hat{S}$ to be a uniformly random subset of S of size N.
- If $N > n$ first sample a uniformly random subset A of S^c of size $N - n$ and set $\hat{S} = S \cup A$

4. Set $\hat{\boldsymbol{B}}$ as follows:

$$\hat{B}_{ii} = \begin{cases} -\kappa & : i \notin \hat{S} \\ 1 - \kappa & : i \in \hat{S}. \end{cases}.$$

We stack the vectors $\boldsymbol{v}_{1:m}$ along the rows of a matrix $\boldsymbol{V} \in \mathbb{R}^{m \times d}$ and refer to the columns of $\boldsymbol{V}$ as $\boldsymbol{V}_1, \boldsymbol{V}_2 \cdots \boldsymbol{V}_d$:

$$\boldsymbol{V} = [\boldsymbol{V}_1, \boldsymbol{V}_2 \cdots \boldsymbol{V}_d] = \begin{bmatrix} \boldsymbol{v}_1^\mathsf{T} \\ \boldsymbol{v}_2^\mathsf{T} \\ \vdots \\ \boldsymbol{v}_m^\mathsf{T} \end{bmatrix}.$$

Lastly we introduce the matrix $\hat{\boldsymbol{\Sigma}} \in \mathbb{R}^{d \times d}$:

$$\hat{\boldsymbol{\Sigma}} \stackrel{\text{def}}{=} \mathbb{E}[\hat{\boldsymbol{T}}\hat{\boldsymbol{T}}^\mathsf{T} | \boldsymbol{V}] = m\kappa(1-\kappa)\boldsymbol{V}^\mathsf{T}\boldsymbol{V}.$$

These definitions are intended to capture the matrix moments $\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})$ as follows: Consider any $k \in \mathbb{N}, \pi \in \mathcal{P}([k]), \boldsymbol{w} \in \mathcal{G}(k)$ and any $\boldsymbol{a} \in \mathcal{C}(\pi)$. Let the disjoint blocks of π be given by $\pi = \mathcal{V}_1 \sqcup \mathcal{V}_2 \cdots \sqcup \mathcal{V}_{|\pi|}$.

In order to capture $\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})$ in the subsampled Hadamard case $\boldsymbol{\Psi} = \boldsymbol{H}\overline{\boldsymbol{B}}\boldsymbol{H}^\mathsf{T}$ and the subsampled Haar case $\boldsymbol{\Psi} = \boldsymbol{O}\overline{\boldsymbol{B}}\boldsymbol{O}^\mathsf{T}$ we will set $\boldsymbol{V}_{1:d}$ as follows:

1. In the subsampled Haar case, we set:

$$\{\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots \boldsymbol{V}_d\} = \{(\boldsymbol{o}_{a_{\mathcal{V}_s}} \odot \boldsymbol{o}_{a_{\mathcal{V}_t}}) - \delta(s,t)\hat{\boldsymbol{e}} : s,t \in [|\pi|],\ s \leq t,\ W_{st}(\boldsymbol{w},\pi) > 0\},$$

where,

$$\boldsymbol{e}^{\mathsf{T}} = \left(\frac{1}{m}, \frac{1}{m} \cdots \frac{1}{m}\right),\ \delta(s,t) = \begin{cases} 1: & s = t \\ & \\ 0: & s \neq t \end{cases}.$$

If for some $i \in [d]$ and some $s,t \in [|\pi|]$ we have $\boldsymbol{V}_i = \boldsymbol{o}_{a_{\mathcal{V}_s}} \odot \boldsymbol{o}_{a_{\mathcal{V}_t}} - \delta(s,t)\hat{\boldsymbol{e}}$, we will abuse notation and often refer to $\boldsymbol{V}_i$ as $\boldsymbol{V}_{st}$. Likewise the corresponding entries of $\boldsymbol{T}, \hat{\boldsymbol{T}}, T_i, \hat{T}_i$ will be referred to as $T_{st}, \hat{T}_{st}$.

2. In the subsampled Hadamard case, we set:

$$\{\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots \boldsymbol{V}_d\} = \{\boldsymbol{h}_{a_{\mathcal{V}_s}} \odot \boldsymbol{h}_{a_{\mathcal{V}_t}} - \delta(s,t)\hat{\boldsymbol{e}} : s,t \in [|\pi|],\ s \leq t,\ W_{st}(\boldsymbol{w},\pi) > 0\}.$$

If for some $i \in [d]$ and some $s,t \in [|\pi|]$ we have $\boldsymbol{V}_i = \boldsymbol{h}_{a_{\mathcal{V}_s}} \odot \boldsymbol{h}_{a_{\mathcal{V}_t}} - \delta(s,t)\hat{\boldsymbol{e}}$, we will abuse notation and often refer to $\boldsymbol{V}_i$ as $\boldsymbol{V}_{st}$. Likewise the corresponding entries of $\boldsymbol{T}, \hat{\boldsymbol{T}}$: $T_i, \hat{T}_i$ will be referred to as $T_{st}, \hat{T}_{st}$.

With the above conventions and the observation that $\sum_{i=1}^{m} \overline{B}_{ii} = 0$ we have:

$$\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) = \prod_{\substack{s,t \in [|\pi|] \\ s \leq t \\ W_{st}(\boldsymbol{w},\pi) > 0}} T_{st}^{W_{st}(\boldsymbol{w},\pi)}.$$

The remainder of this section is organized as follows:

1. First, in Lemma 57 we show that $\hat{\boldsymbol{\Sigma}}$ converges to a fixed deterministic matrix $\boldsymbol{\Sigma}$ and bound the rate of convergence in terms of $\mathbb{E}\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathsf{Fr}}^2$.

2. In Lemma 58 we upper bound $\mathbb{E}\|\hat{\boldsymbol{T}} - \boldsymbol{T}\|_2^2$. Consequently a Gaussian approximation result for $\hat{\boldsymbol{T}}$ implies a Gaussian approximation result for $\boldsymbol{T}$.

3. In Lemma 59, we use a standard Berry Eseen bound of Bhattacharya [83] to derive a Gaussian approximation result for $\hat{\boldsymbol{T}}$ since it is a weighted sum of i.i.d. centered random variables.

4. Finally we conclude by using the above lemmas to provide a proof for Propositions 17 and 16.

Lemma 57. *1. For the Hadamard case suppose $\boldsymbol{w}$ is disassortative with respect to π and $\boldsymbol{a}$ is a conflict free labelling of $(\boldsymbol{w}, \pi)$. Then,*

$$\hat{\boldsymbol{\Sigma}} = \kappa(1-\kappa)\boldsymbol{I}_d.$$

2. For the Haar case there exists a universal constant $C < \infty$ such that for any partition $\pi \in \mathcal{P}([k])$, any weight matrix $\boldsymbol{w} \in \mathcal{G}(k)$ and any labelling $\boldsymbol{a} \in \mathcal{C}(\pi)$ we have,

$$\mathbb{E}\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathsf{Fr}}^2 \leq \frac{C \cdot k^4 \cdot (\kappa^2(1-\kappa)^2)}{m}.$$

where the matrix $\boldsymbol{\Sigma}$ is a diagonal matrix whose diagonal entries are given by:

$$\Sigma_{st,st} = \begin{cases} \kappa(1-\kappa): & s \neq t \\ 2\kappa(1-\kappa): & s = t \end{cases}.$$

Proof. Recall that,

$$\hat{\boldsymbol{\Sigma}} = m\kappa(1-\kappa)\boldsymbol{V}^{\mathsf{T}}\boldsymbol{V}.$$

We consider the Hadamard and the Haar case separately.

Hadamard Case: Consider two pairs (s,t) and (s',t') such that:

$$s \le t,\ W_{st}(\boldsymbol{w},\pi) > 0,\ s,t\ \in\ [|\pi|].$$

and the analogous assumptions on the pair (s',t'). Then the entry $\hat{\Sigma}_{st,s't'}$ is given by:

$$\begin{aligned}
\hat{\Sigma}_{st,s't'} &= m\kappa(1-\kappa)\langle \boldsymbol{V}_{st}, \boldsymbol{V}_{s't'}\rangle \\
&= m\kappa(1-\kappa)\langle \boldsymbol{h}_{a_{\mathcal{V}_s}} \odot \boldsymbol{h}_{a_{\mathcal{V}_t}} - \delta(s,t)\hat{\boldsymbol{e}}, \boldsymbol{h}_{a_{\mathcal{V}'_s}} \odot \boldsymbol{h}_{a_{\mathcal{V}'_t}} - \delta(s',t')\hat{\boldsymbol{e}}\rangle \\
&\overset{\text{(a)}}{=} \kappa(1-\kappa)\langle \boldsymbol{h}_{a_{\mathcal{V}_s}\oplus a_{\mathcal{V}_t}} - \sqrt{m}\delta(s,t)\hat{\boldsymbol{e}}, \boldsymbol{h}_{a_{\mathcal{V}'_s}\oplus a_{\mathcal{V}'_t}} - \sqrt{m}\delta(s',t')\hat{\boldsymbol{e}}\rangle \\
&\overset{\text{(b)}}{=} \kappa(1-\kappa)\langle \boldsymbol{h}_{a_{\mathcal{V}_s}\oplus a_{\mathcal{V}_t}}, \boldsymbol{h}_{a_{\mathcal{V}'_s}\oplus a_{\mathcal{V}'_t}}\rangle \\
&\overset{\text{(c)}}{=} \kappa(1-\kappa)\delta(s,s')\delta(t,t').
\end{aligned}$$

In the step marked (a) we appealed to Lemma 19. In the step marked (b), we noted that $\hat{\boldsymbol{e}} = \boldsymbol{h}_1/\sqrt{m}$ and $\hat{\boldsymbol{e}} \perp \boldsymbol{h}_{a_{\mathcal{V}_s}\oplus a_{\mathcal{V}_t}}$ unless $s = t$ which is ruled out by the fact that $\boldsymbol{w}$ is disassortative with respect to π i.e. $W_{ss}(\boldsymbol{w},\pi) = 0$. In the step marked (c) we used the fact that $\boldsymbol{a}$ is a conflict free labelling. Consequently, we have shown that $\hat{\boldsymbol{\Sigma}} = \kappa(1-\kappa)\boldsymbol{I}_d$.

Haar case: By the bias-variance decomposition:

$$\mathbb{E}\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathsf{Fr}}^2 = \mathbb{E}\|\hat{\boldsymbol{\Sigma}} - \mathbb{E}\hat{\boldsymbol{\Sigma}}\|_{\mathsf{Fr}}^2 + \|\mathbb{E}\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathsf{Fr}}^2.$$

We will first compute $\mathbb{E}\hat{\boldsymbol{\Sigma}}$. Consider the $(st, s't')$ entry of $\hat{\boldsymbol{\Sigma}}$:

$$\begin{aligned}
\hat{\Sigma}_{st,s't'} &= m\kappa(1-\kappa)\langle \boldsymbol{V}_{st}, \boldsymbol{V}_{s't'}\rangle \\
&= m\kappa(1-\kappa)\langle \boldsymbol{o}_{a_{\mathcal{V}_s}} \odot \boldsymbol{o}_{a_{\mathcal{V}_t}} - \delta(s,t)\hat{\boldsymbol{e}}, \boldsymbol{o}_{a_{\mathcal{V}'_s}} \odot \boldsymbol{o}_{a_{\mathcal{V}'_t}} - \delta(s',t')\hat{\boldsymbol{e}}\rangle \\
&= m\kappa(1-\kappa)\left[\sum_{i=1}^{m}\left((\boldsymbol{o}_{a_{\mathcal{V}_s}})_i(\boldsymbol{o}_{a_{\mathcal{V}_t}})_i - \frac{\delta(s,t)}{m}\right)\left((\boldsymbol{o}_{a_{\mathcal{V}'_s}})_i(\boldsymbol{o}_{a_{\mathcal{V}'_t}})_i - \frac{\delta(s',t')}{m}\right)\right].
\end{aligned}$$

Note that $\boldsymbol{O}_i$ is a uniformly random unit vector. Hence we can compute $\mathbb{E}\hat{\boldsymbol{\Sigma}}$ using Fact 6.

We obtain:

$$\frac{\mathbb{E}\hat{\Sigma}_{st,s't'}}{\kappa(1-\kappa)} = \begin{cases} 2 - \frac{6}{m+2} : & s = s' = t = t' \\ \frac{2}{(m-1)(m+2)} : & s = t, s' = t', s \neq s' \\ 1 + \frac{2}{(m-1)(m+2)} : & s = s', t = t', s \neq t \\ 0 : & \text{otherwise} \end{cases}.$$

Hence, the bias term can be bounded by:

$$\|\mathbb{E}\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathsf{Fr}}^2 \leq \frac{36 \cdot k^4 \cdot \kappa^2(1-\kappa)^2}{(m+2)^2}.$$

On the other hand, applying the Poincare Inequality (Fact 9) and a tedious calculation involving 6th moments of a random unit vector (see for example Proposition 2.5 of Meckes [93]) shows that,

$$\mathrm{Var}(\hat{\Sigma}_{st,s't'}) \leq \frac{C \cdot \kappa^2(1-\kappa)^2}{m},$$

for some universal constant C. Hence,

$$\mathbb{E}\|\hat{\boldsymbol{\Sigma}} - \mathbb{E}\hat{\boldsymbol{\Sigma}}\|_{\mathsf{Fr}}^2 \leq \frac{C \cdot k^4 \cdot \kappa^2(1-\kappa)^2}{m},$$

for some universal constant C, and consequently the claim of the lemma holds.

□

Lemma 58. *We have,*

$$\mathbb{E}\left[\|\boldsymbol{T} - \hat{\boldsymbol{T}}\|_2^2\right] \leq \frac{Ck^3}{\sqrt{m}},$$

for a universal constant C.

Proof. Let $\overline{\boldsymbol{b}}, \hat{\boldsymbol{b}} \in \mathbb{R}^m$ be the vectors formed by the diagonals of $\overline{\boldsymbol{B}}, \hat{\boldsymbol{B}}$, respectively. Define:

$$p_1 = \mathbb{P}(\overline{b}_1 \neq \hat{b}_1),\ p_2 = \mathbb{P}(\overline{b}_1 \neq \hat{b}_1,\ \overline{b}_2 \neq \hat{b}_2).$$

We have,

$$\begin{aligned}
\mathbb{E}\left[\|\boldsymbol{T} - \hat{\boldsymbol{T}}\|_2^2 \mid \boldsymbol{V}\right] &= m\mathbb{E}\left[(\overline{\boldsymbol{b}} - \hat{\boldsymbol{b}})^\mathsf{T}\boldsymbol{V}\boldsymbol{V}^\mathsf{T}(\overline{\boldsymbol{b}} - \hat{\boldsymbol{b}})\right] \\
&= m\mathsf{Tr}\left(\boldsymbol{V}\boldsymbol{V}^\mathsf{T}\mathbb{E}\left[(\overline{\boldsymbol{b}} - \hat{\boldsymbol{b}})(\overline{\boldsymbol{b}} - \hat{\boldsymbol{b}})^\mathsf{T}\right]\right) \\
&= m\mathsf{Tr}\left(\boldsymbol{V}\boldsymbol{V}^\mathsf{T}(1-2\kappa)^2\left(p_2\mathbf{1}\mathbf{1}^\mathsf{T} + (p_1 - p_2)\boldsymbol{I}_m\right)\right) \\
&= m(1-2\kappa)^2\left(p_2\left\|\boldsymbol{V}^\mathsf{T}\mathbf{1}\right\|_2^2 + (p_1 - p_2)\mathsf{Tr}\left(\boldsymbol{V}\boldsymbol{V}^\mathsf{T}\right)\right).
\end{aligned}$$

Now, since $\boldsymbol{V}^\mathsf{T}$ has centered coordinate-wise product of columns of an orthogonal matrix we have $\boldsymbol{V}^\mathsf{T}\mathbf{1} = 0$. Hence,

$$\mathbb{E}\left[\|\boldsymbol{T} - \hat{\boldsymbol{T}}\|_2^2 \mid \boldsymbol{V}\right] = (p_1 - p_2)\mathsf{Tr}\left(\boldsymbol{V}\boldsymbol{V}^\mathsf{T}\right).$$

Next we compute $p_1 = \mathbb{P}(\overline{b}_1 \neq \hat{b}_1)$. Observe that conditional on N, the symmetric difference $S \triangle \hat{S}$ is a uniformly random set of size $|N - n|$. Hence,

$$\mathbb{P}(\overline{b}_1 \neq \hat{b}_1 | N) = \mathbb{P}(1 \in S \triangle \hat{S} | N) = \frac{|n - N|}{m}.$$

Therefore

$$p_1 = \frac{\mathbb{E}\left[N - n\right]}{m} \leq \frac{\sqrt{\mathrm{Var}(N))}}{m} = \frac{\sqrt{\kappa(1-\kappa)}}{\sqrt{m}}.$$

Hence, we obtain

$$\mathbb{E}\left[\|\boldsymbol{T} - \hat{\boldsymbol{T}}\|_2^2 | \boldsymbol{V}\right] \leq \frac{(1-2\kappa)^2}{\sqrt{m \cdot \kappa(1-\kappa)}} \cdot \mathsf{Tr}(\hat{\boldsymbol{\Sigma}}). \tag{C.7}$$

By Lemma 57 we have,

$$\begin{aligned}\mathbb{E}\mathsf{Tr}(\hat{\boldsymbol{\Sigma}}) &\leq \mathbb{E}\mathsf{Tr}(\boldsymbol{\Sigma}) + \sqrt{d \cdot \mathbb{E}\|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathsf{Fr}}^2} \\ &\leq C\kappa(1-\kappa)k^3.\end{aligned}$$

where constant $C_{\kappa,d}$ depends only on κ, d. And hence,

$$\mathbb{E}\left[\|\boldsymbol{T} - \hat{\boldsymbol{T}}\|_2^2\right] \leq \frac{Ck^3}{\sqrt{m}},$$

for a universal constant C. □

Lemma 59. *Under the assumptions and notations of Lemma 57 for both the subsampled Haar sensing and the subsampled Hadamard sensing models, we have, for any bounded Lipschitz function $f : \mathbb{R}^d \to \mathbb{R}$:*

$$\mathbb{E}\left|\mathbb{E}[f(\hat{\boldsymbol{T}})|\boldsymbol{V}] - \mathbb{E}f(\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z})\right| \leq \frac{C_k \cdot (\|f\|_\infty + \|f\|_{\mathsf{Lip}})}{\sqrt{m}}. \tag{C.8}$$

where $\boldsymbol{Z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$, C_k is a constant depending only on k.

Proof. Note that $\hat{\boldsymbol{T}} = \sqrt{m}\boldsymbol{V}^\mathsf{T}\hat{\boldsymbol{b}}$ and $\sqrt{m}\hat{\boldsymbol{\Sigma}}^{\frac{-1}{2}}\boldsymbol{V}^\mathsf{T}\hat{\boldsymbol{b}}$ has the identity covariance matrix. Hence, by the Berry Eseen bound of Bhattacharya [83] for any bounded and Lipschitz function g we have

$$\left|\mathbb{E}\left[g\left(\hat{\boldsymbol{\Sigma}}^{\frac{-1}{2}}\hat{\boldsymbol{T}}\right)\right] - \mathbb{E}\left[\boldsymbol{Z}\right]\right| \leq \frac{C_d \cdot \rho_3' \cdot \left(\|g\|_\infty + \|g\|_{Lip}\right)}{\sqrt{m}}, \tag{C.9}$$

where C_d is a constant only dependent on d and

$$\begin{aligned}\rho_3' &= m^2 \sum_{i=1}^{m} \mathbb{E}\left[\hat{b}_i \|\hat{\boldsymbol{\Sigma}}^{\frac{-1}{2}} \boldsymbol{v}_i\|_2^3 | \boldsymbol{V}\right] \\ &= m^2 \left(\kappa(1-\kappa)^3 + (1-\kappa)\kappa^3\right) \sum_{i=1}^{m} \|\hat{\boldsymbol{\Sigma}}^{\frac{-1}{2}} \boldsymbol{v}_i\|_2^3 \\ &\leq m^2 \cdot \sqrt{d} \cdot \|\hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\|_{\mathsf{op}}^3 \cdot (\kappa(1-\kappa)) \cdot \sum_{i=1}^{m} \|\boldsymbol{v}_i\|_3^3\end{aligned}$$

.

Define $g(\boldsymbol{X}) \stackrel{\text{def}}{=} f(\hat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\boldsymbol{X})$,hence , $g(\hat{\boldsymbol{\Sigma}}^{\frac{-1}{2}}\boldsymbol{V}^{\mathsf{T}}\hat{\boldsymbol{b}}) = f(\hat{\boldsymbol{T}})$. Moreover, $\|g\|_\infty \leq \|f\|_\infty$ and $\|g\|_{\mathsf{Lip}} \leq \|\boldsymbol{\Sigma}\|_{\mathsf{op}}^{\frac{1}{2}} \|f\|_{\mathsf{Lip}}$. Hence we obtain:

$$\begin{aligned}&\left|\mathbb{E}[f(\hat{\boldsymbol{T}})|\boldsymbol{V}] - \mathbb{E}f(\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z})\right| \leq \\ &\qquad C_d(\kappa(1-\kappa)) \cdot m^{\frac{3}{2}} \cdot (\|f\|_\infty + \|\hat{\boldsymbol{\Sigma}}\|_{\mathsf{op}}^{\frac{1}{2}} \|f\|_{\mathsf{Lip}}) \cdot \|\hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\|_{\mathsf{op}}^3 \cdot \sum_{i=1}^{m} \|\boldsymbol{v}_i\|_3^3. \qquad \text{(C.10)}\end{aligned}$$

We define the event:

$$\mathcal{E} \stackrel{\text{def}}{=} \left\{ \boldsymbol{V} : \|\hat{\boldsymbol{\Sigma}} - \boldsymbol{\Sigma}\|_{\mathsf{Fr}}^2 \leq \frac{\kappa^2(1-\kappa)^2}{4} \right\}.$$

By Markov Inequality and Lemma 57, we know that, $\mathbb{P}(\mathcal{E}^c) \leq Ck^4/m$ for some universal constant C. Hence,

$$\mathbb{E}\left|\mathbb{E}[f(\hat{\boldsymbol{T}})|\boldsymbol{V}] - \mathbb{E}f(\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z})\right| \leq \frac{2C \cdot \|f\|_\infty \cdot k^4}{m} + \mathbb{E}\left|\mathbb{E}[f(\hat{\boldsymbol{T}})|\boldsymbol{V}] - \mathbb{E}f(\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z})\right| \mathbb{I}(\mathcal{E}).$$

On the event $\mathcal{E}$ we have,

$$\begin{aligned}
\|\hat{\boldsymbol{\Sigma}}\|_{\mathsf{op}} &\leq \|\boldsymbol{\Sigma}\|_{\mathsf{op}} + \frac{\kappa(1-\kappa)}{2} \leq \frac{5\kappa(1-\kappa)}{2}, \\
\|\hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\|_{\mathsf{op}} &\leq \|\boldsymbol{\Sigma}^{-\frac{1}{2}}\|_{\mathsf{op}} + \|\hat{\boldsymbol{\Sigma}}^{-\frac{1}{2}} - \boldsymbol{\Sigma}^{-\frac{1}{2}}\|_{\mathsf{op}} \overset{\text{(a)}}{\leq} \frac{1}{\kappa(1-\kappa)} + \frac{1}{2} \leq \frac{9}{8(\kappa(1-\kappa))}, \\
\mathbb{E}\|\boldsymbol{v}_i\|^3 &= \sum_{j=1}^{d} \mathbb{E}|v_{ij}|^3 \overset{\text{(b)}}{\leq} \frac{Cd}{m^3}.
\end{aligned}$$

In the step marked (a) we used the continuity estimate for matrix square root in Fact 10. In the step marked (b), we recalled the definition of $\boldsymbol{v}_i$ and used the moment bounds for a coordinate of a random unit vector from Fact 6. Substituting these estimates in (C.10) we obtain:

$$\mathbb{E}\left|\mathbb{E}[f(\hat{\boldsymbol{T}})|\boldsymbol{V}] - \mathbb{E}f(\hat{\boldsymbol{\Sigma}}^{1/2}\boldsymbol{Z})\right| \leq \frac{2C \cdot \|f\|_\infty \cdot k^4}{m} + \frac{C_k \cdot (\|f\|_\infty + \|f\|_{\mathsf{Lip}})}{\sqrt{m}}.$$

□

Using the above lemmas, we can now provide a proof of Propositions 17 and 16.

Proof of Propositions 17 and 16. Define the polynomial $p(\boldsymbol{z})$ as:

$$p(\boldsymbol{z}) \overset{\text{def}}{=} \prod_{\substack{s,t \in [|\pi|] \\ s \leq t \\ W_{st}(\boldsymbol{w},\pi) > 0}} z_{st}^{W_{st}(\boldsymbol{w},\pi)},$$

and the indicator function:

$$\mathbb{I}\left(\mathcal{E}\right)(\boldsymbol{z}) \overset{\text{def}}{=} \begin{cases} 1: & \boldsymbol{z} \in \mathcal{E} \\ 0: & \boldsymbol{z} \notin \mathcal{E} \end{cases},$$

where:

$$\mathcal{E} \overset{\text{def}}{=} \left\{ \max_{s,t} |z_{st}| \leq \left(2048 \log^3(m)\right)^{\frac{1}{2}} \right\}.$$

Recall that we had,

$$\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) = \prod_{\substack{s,t\in[|\pi|]\\ s\leq t\\ W_{st}(\boldsymbol{w},\pi)>0}} T_{st}^{W_{st}(\boldsymbol{w},\pi)} = p(\boldsymbol{T}),$$

and in Lemma 21 we showed that,

$$\mathbb{P}(\boldsymbol{T} \notin \mathcal{E}) \leq \frac{C}{m^2}.$$

We additionally define the function $\widetilde{p}(\boldsymbol{z}) \stackrel{\text{def}}{=} p(\boldsymbol{z})\mathbb{I}\left(\mathcal{E}\right)(\boldsymbol{z})$. observe that:

$$\|\widetilde{p}\|_{\infty} \leq \left(2048\log^3(m)\right)^{\frac{\|\boldsymbol{w}\|}{2}}, \ \|\widetilde{p}\|_{\mathsf{Lip}} \leq \|\boldsymbol{w}\|\left(2048\log^3(m)\right)^{\frac{\|\boldsymbol{w}\|}{2}}.$$

Let $\boldsymbol{Z} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}_d\right)$. Then, we can write:

$$\begin{aligned}
\left|\mathbb{E}p(\boldsymbol{T}) - \mathbb{E}p(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z})\right| &\leq \left|\mathbb{E}\widetilde{p}(\boldsymbol{T}) - \mathbb{E}\widetilde{p}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z})\right| + |\mathbb{E}p(\boldsymbol{T})\mathbb{I}\left(\mathcal{E}^c\right)(\boldsymbol{T})| + |\mathbb{E}p(\boldsymbol{T})\mathbb{I}\left(\mathcal{E}^c\right)(\boldsymbol{\Sigma}^{\frac{1}{2}}Z)| \\
&\leq \underbrace{\left|\mathbb{E}\widetilde{p}(\boldsymbol{T}) - \mathbb{E}\widetilde{p}(\hat{\boldsymbol{T}})\right|}_{\text{(I)}} + \underbrace{\left|\mathbb{E}\widetilde{p}(\boldsymbol{T}) - \mathbb{E}\widetilde{p}(\hat{\boldsymbol{\Sigma}}^{\frac{1}{2}}\boldsymbol{Z})\right|}_{\text{(II)}} + \underbrace{\left|\mathbb{E}\widetilde{p}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z}) - \mathbb{E}\widetilde{p}(\hat{\boldsymbol{\Sigma}^{\frac{1}{2}}}\boldsymbol{Z})\right|}_{\text{(III)}} \\
&\qquad + \underbrace{|\mathbb{E}p(\boldsymbol{T})\mathbb{I}\left(\mathcal{E}^c\right)(\boldsymbol{T})|}_{\text{(IV)}} + \underbrace{|\mathbb{E}p(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z})\mathbb{I}\left(\mathcal{E}^c\right)(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z})|}_{\text{(V)}}.
\end{aligned}$$

We control each of these terms separately.

Analysis of (I)**:** In order to control I observe that:

$$\begin{aligned}
\text{(I)} &\leq \|\widetilde{p}\|_{\mathsf{Lip}}\mathbb{E}\|\boldsymbol{T} - \hat{\boldsymbol{T}}\|_2 \\
&\leq \|\widetilde{p}\|_{\mathsf{Lip}} \cdot (\mathbb{E}\|\boldsymbol{T} - \hat{\boldsymbol{T}}\|_2^2)^{\frac{1}{2}} \\
&\leq C \cdot \|\boldsymbol{w}\| \cdot \left(2048\log^3(m)\right)^{\frac{\|\boldsymbol{w}\|}{2}} \cdot \frac{\sqrt{k^3}}{m^{\frac{1}{4}}}.
\end{aligned}$$

In the last step, we appealed to Lemma 58.

Analysis of (II)**:** In order to control I, recall that:

$$\|\widetilde{p}\|_\infty \leq \left(2048\log^3(m)\right)^{\frac{\|\boldsymbol{w}\|}{2}},\ \|\widetilde{p}\|_{\mathsf{Lip}} \leq \|\boldsymbol{w}\|\left(2048\log^3(m)\right)^{\frac{\|\boldsymbol{w}\|}{2}}.$$

Hence, by Lemma 59 we have,

$$(\mathsf{II}) \leq \frac{C_k\cdot(2048\log^3(m))^{\frac{\|\boldsymbol{w}\|}{2}}(1+\|\boldsymbol{w}\|)}{\sqrt{m}}.$$

Analysis of (III)**:** Again using the Lipchitz bound on $\widetilde{p}$ we have,

$$\begin{aligned}
(\mathsf{III}) &\leq \mathbb{E}|\widetilde{p}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z}) - \widetilde{p}(\hat{\boldsymbol{\Sigma}^{\frac{1}{2}}}\boldsymbol{Z})| \\
&\leq \|\boldsymbol{w}\|\left(2048\log^3(m)\right)^{\frac{\|\boldsymbol{w}\|}{2}}\cdot\mathbb{E}\|(\hat{\boldsymbol{\Sigma}}^{\frac{1}{2}}-\boldsymbol{\Sigma}^{\frac{1}{2}})\boldsymbol{Z}\|_2 \\
&\leq \|\boldsymbol{w}\|\left(2048\log^3(m)\right)^{\frac{\|\boldsymbol{w}\|}{2}}\cdot\sqrt{\mathbb{E}\|(\hat{\boldsymbol{\Sigma}}^{\frac{1}{2}}-\boldsymbol{\Sigma}^{\frac{1}{2}})\boldsymbol{Z}\|_2^2} \\
&\leq \|\boldsymbol{w}\|\left(2048\log^3(m)\right)^{\frac{\|\boldsymbol{w}\|}{2}}\cdot\sqrt{\mathbb{E}\|\hat{\boldsymbol{\Sigma}}^{\frac{1}{2}}-\boldsymbol{\Sigma}^{\frac{1}{2}}\|_{\mathsf{Fr}}^2} \\
&\overset{\text{(a)}}{\leq} \|\boldsymbol{w}\|\left(2048\log^3(m)\right)^{\frac{\|\boldsymbol{w}\|}{2}}\cdot\frac{k^2}{\lambda_{\max}(\boldsymbol{\Sigma})}\cdot\mathbb{E}\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_{\mathsf{Fr}}^2 \\
&\overset{\text{(b)}}{\leq} \frac{C\cdot k^6\cdot\|\boldsymbol{w}\|(2048\log^3(m))^{\frac{\|\boldsymbol{w}\|}{2}}}{m}.
\end{aligned}$$

In the step marked (a) we used the fact that the continuity estimate for matrix square roots given in Fact 10. In the step marked (b) we recalled the definition of $\boldsymbol{\Sigma}$ and observed that $\lambda_{\max}(\boldsymbol{\Sigma}) \geq \kappa(1-\kappa)$ for the subsampled Haar and the Hadamard sensing model. We also used the bound on $\mathbb{E}\|\hat{\boldsymbol{\Sigma}}-\boldsymbol{\Sigma}\|_{\mathsf{Fr}}^2$ obtained in Lemma 57.

Analysis of (IV)**:** We can control (III) as follows:

$$\begin{aligned}(\mathsf{IV}) &\leq \sqrt{\mathbb{E}p^2(\boldsymbol{T})} \cdot \sqrt{\mathbb{P}(\boldsymbol{T} \notin \mathcal{E})} \\ &\overset{\text{(c)}}{\leq} \frac{C\sqrt{\mathbb{E}\mathcal{M}(\sqrt{m}\boldsymbol{\Psi}, 2\boldsymbol{w}, \pi, \boldsymbol{a})}}{m} \\ &\overset{\text{(d)}}{\leq} \frac{(C\|\boldsymbol{w}\| \log^2(m))^{\frac{\|\boldsymbol{w}\|}{2}}}{m}\end{aligned}$$

In the step marked (c) we recalled that $\mathbb{P}(\boldsymbol{T} \notin \mathcal{E}) \leq C/m^2$ and expressed $p^2(\boldsymbol{T})$ as a matrix moment. In the step marked (d) we used the bounds on matrix moments obtained in Lemma 18.

Analysis of (IV)**:** We recall that $\boldsymbol{\Sigma}$ was a diagonal matrix with $|\Sigma_{ii}| \leq 2\kappa(1-\kappa) \leq 1$. Hence,

$$\begin{aligned}(\mathsf{V}) &\leq \sqrt{\mathbb{E}p^2(\boldsymbol{\Sigma}^{\frac{1}{2}})} \cdot \sqrt{\mathbb{P}(\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{Z} \notin \mathcal{E})} \\ &\overset{\text{(e)}}{\leq} \frac{k\|\boldsymbol{w}\|^{\frac{\|\boldsymbol{w}\|}{2}}}{m}.\end{aligned}$$

In the step marked (e) we used standard moment and tail bounds on Gaussian random variables.

Combining the bounds on $\mathsf{I} - \mathsf{V}$ immediately yields the claims of Proposition 17 and 16. □

C.4 Missing Proofs from Section 5.8

C.4.1 Proof of Lemma 25

Proof of Lemma 25. We will assume that $\mathcal{A}$ is of Type 1 (the proof of the other types is analogous):

$$\mathcal{A}(\boldsymbol{\Psi}, \boldsymbol{Z}) = p_1(\boldsymbol{\Psi})q_1(\boldsymbol{Z})p_2(\boldsymbol{\Psi}) \cdots q_{k-1}(\boldsymbol{Z})p_k(\boldsymbol{\Psi}).$$

Define for any $i \in [k]$:

$$\mathcal{A}_0 \stackrel{\text{def}}{=} p_1(\boldsymbol{\Psi}) q_1(\text{Diag}\,(\boldsymbol{z})) p_2(\boldsymbol{\Psi}) \cdots q_{k-1}(\text{Diag}\,(\boldsymbol{z})) p_k(\boldsymbol{\Psi}),$$
$$\mathcal{A}_i \stackrel{\text{def}}{=} p_1(\boldsymbol{\Psi}) q_1(\text{Diag}\,(\widetilde{\boldsymbol{z}})) \cdots q_i(\text{Diag}\,(\widetilde{\boldsymbol{z}})) p_{i+1}(\boldsymbol{\Psi}) q_{i+1}(\text{Diag}\,(\boldsymbol{z})) \cdots q_{k-1}(\text{Diag}\,(\boldsymbol{z})) p_k(\boldsymbol{\Psi}).$$

where $\boldsymbol{\Psi} = \boldsymbol{U}\overline{\boldsymbol{B}}\boldsymbol{U}^{\mathsf{T}}$. Observe that we can write:

$$\begin{aligned}
&\boldsymbol{z}^{\mathsf{T}}\mathcal{A}(\boldsymbol{U}\overline{\boldsymbol{B}}\boldsymbol{U}^{\mathsf{T}}, \text{Diag}\,(\boldsymbol{z}))\boldsymbol{z} - \widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}(\boldsymbol{U}\overline{\boldsymbol{B}}\boldsymbol{U}^{\mathsf{T}}, \text{Diag}\,(\widetilde{\boldsymbol{z}}))\widetilde{\boldsymbol{z}} = \boldsymbol{z}^{\mathsf{T}}\mathcal{A}_0\boldsymbol{z} - \widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}_{k-1}\widetilde{\boldsymbol{z}} \\
&= \boldsymbol{z}^{\mathsf{T}}\mathcal{A}_0\boldsymbol{z} - \boldsymbol{z}^{\mathsf{T}}\mathcal{A}_{k-1}\boldsymbol{z} + \boldsymbol{z}^{\mathsf{T}}\mathcal{A}_{k-1}\boldsymbol{z} + \widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}_{k-1}\widetilde{\boldsymbol{z}} \\
&= \left(\sum_{i=0}^{k-2} \boldsymbol{z}^{\mathsf{T}}(\mathcal{A}_i - \mathcal{A}_{i+1})\boldsymbol{z}\right) + \langle \mathcal{A}_{k-1}, \boldsymbol{z}\boldsymbol{z}^{\mathsf{T}} - \widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}\rangle.
\end{aligned}$$

We bound each of these terms separately. First observe that:

$$\begin{aligned}
|\boldsymbol{z}^{\mathsf{T}}(\mathcal{A}_i - \mathcal{A}_{i+1})\boldsymbol{z}| &\leq \|\boldsymbol{z}\|_2^2 \cdot \|\mathcal{A}_i - \mathcal{A}_{i+1}\|_{\mathsf{op}} \\
&\leq C(\mathcal{A}) \cdot \|\boldsymbol{z}\|_2^2 \cdot \|\boldsymbol{z} - \widetilde{\boldsymbol{z}}\|_\infty.
\end{aligned}$$

Next we note that,

$$\begin{aligned}
|\langle \mathcal{A}_{k-1}, \boldsymbol{z}\boldsymbol{z}^{\mathsf{T}} - \widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}\rangle| &\leq 2\|\mathcal{A}_{k-1}\|_{\mathsf{op}} \cdot \|\boldsymbol{z}\boldsymbol{z}^{\mathsf{T}} - \widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}\|_{\mathsf{op}} \\
&= C(\mathcal{A}) \cdot \|\boldsymbol{z} - \widetilde{\boldsymbol{z}}\|_2 \cdot (\|\boldsymbol{z}\|_2 + \|\widetilde{\boldsymbol{z}}\|_2).
\end{aligned}$$

This gives is the estimate:

$$\left|\frac{\boldsymbol{z}^{\mathsf{T}}\mathcal{A}(\boldsymbol{U}\overline{\boldsymbol{B}}\boldsymbol{U}^{\mathsf{T}}, \text{Diag}\,(\boldsymbol{z}))\boldsymbol{z}}{m} - \frac{\widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}(\boldsymbol{U}\overline{\boldsymbol{B}}\boldsymbol{U}^{\mathsf{T}}, \text{Diag}\,(\widetilde{\boldsymbol{z}}))\widetilde{\boldsymbol{z}}}{m}\right| \leq$$
$$\frac{C(\mathcal{A})}{m} \cdot \left(\|\boldsymbol{z}\|_2^2 \cdot \|\boldsymbol{z} - \widetilde{\boldsymbol{z}}\|_\infty + \|\boldsymbol{z} - \widetilde{\boldsymbol{z}}\|_2 \cdot (\|\boldsymbol{z}\|_2 + \|\widetilde{\boldsymbol{z}}\|_2)\right),$$

where $C(\mathcal{A})$ denotes a finite constant depending only on the $\|\|_\infty$ norms and Lipchitz constants of the functions appearing in $\mathcal{A}$. $\square$

C.4.2 Proof of Lemma 26

Proof of Lemma 26. Using the continuity estimate from Lemma 25 we know that on the event $\mathcal{E}$,

$$\begin{aligned}
&\left|\frac{\boldsymbol{z}^\mathsf{T}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\boldsymbol{z}}{m} - \frac{\widetilde{\boldsymbol{z}}^\mathsf{T}\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}})\widetilde{\boldsymbol{z}}}{m}\right| \leq \frac{C(\mathcal{A})}{m}\cdot\left(\|\boldsymbol{z}\|_2^2\cdot\|\boldsymbol{z}-\widetilde{\boldsymbol{z}}\|_\infty + \|\boldsymbol{z}-\widetilde{\boldsymbol{z}}\|_2\cdot(\|\boldsymbol{z}\|_2+\|\widetilde{\boldsymbol{z}}\|_2)\right)\\
&\leq \frac{C(\mathcal{A})}{m}\cdot\left(\|\boldsymbol{z}\|_2^2\cdot\|\boldsymbol{z}\|_\infty + \|\boldsymbol{z}\|_2\cdot(\|\boldsymbol{z}\|_2+\|\widetilde{\boldsymbol{z}}\|_2)\right)\cdot\left(\max_{i\in[m]}\left|\frac{1}{\sigma_i}-1\right|\right)\\
&\leq \frac{C(\mathcal{A})}{m\kappa}\cdot\left(\|\boldsymbol{z}\|_2^2\cdot\|\boldsymbol{z}\|_\infty + \|\boldsymbol{z}\|_2\cdot(\|\boldsymbol{z}\|_2+\|\widetilde{\boldsymbol{z}}\|_2)\right)\cdot\sqrt{\frac{\log^3(m)}{m}}
\end{aligned}$$

Hence,

$$\begin{aligned}
\left|\mathbb{E}\frac{\boldsymbol{z}^\mathsf{T}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\boldsymbol{z}}{m} - \mathbb{E}\frac{\widetilde{\boldsymbol{z}}^\mathsf{T}\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}})\widetilde{\boldsymbol{z}}}{m}\mathbb{I}\left(\mathcal{E}\right)\right| &\leq \left|\mathbb{E}\frac{\boldsymbol{z}^\mathsf{T}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\boldsymbol{z}}{m}\mathbb{I}\left(\mathcal{E}^c\right)\right|\\
&\quad + \frac{C(\mathcal{A})\log^{\frac{3}{2}}(m)}{m\sqrt{m}\kappa}\cdot\left(\mathbb{E}\|\boldsymbol{z}\|_2^2\cdot\|\boldsymbol{z}\|_\infty + \mathbb{E}\|\boldsymbol{z}\|_2\cdot(\|\boldsymbol{z}\|_2+\|\widetilde{\boldsymbol{z}}\|_2)\right).
\end{aligned}$$

Observe that $\boldsymbol{z}^\mathsf{T}\mathcal{A}\boldsymbol{z} \leq \|\mathcal{A}\|_{\mathsf{op}}\|\boldsymbol{z}\|^2 \leq C(\mathcal{A})\|\boldsymbol{z}\|_2^2 \leq C(\mathcal{A})\|\boldsymbol{x}\|_2^2$. Hence,

$$\begin{aligned}
\left|\mathbb{E}\frac{\boldsymbol{z}^\mathsf{T}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\boldsymbol{z}}{m}\mathbb{I}\left(\mathcal{E}^c\right)\right| &\leq C(\mathcal{A})\frac{\sqrt{\mathbb{E}\|\boldsymbol{x}_\star\|_2^4\cdot\mathbb{P}(\mathcal{E}^c)}}{m} \leq \frac{C(\mathcal{A})\sqrt{\mathbb{P}(\mathcal{E}^c)}}{\kappa^2} \to 0,\\
\mathbb{E}\|\boldsymbol{z}\|_2^2 + \mathbb{E}\|\boldsymbol{z}\|_2\|\widetilde{\boldsymbol{z}}\|_2 &\leq 2\mathbb{E}\|\boldsymbol{z}\|_2^2 + \mathbb{E}\|\widetilde{\boldsymbol{z}}\|_2^2 \leq 2\mathbb{E}\|\boldsymbol{x}_\star\|_2^2 + \mathbb{E}\|\widetilde{\boldsymbol{z}}\|_2^2 = \frac{2m}{\kappa} + m,\\
\mathbb{E}\|\boldsymbol{z}\|_2^2\cdot\|\boldsymbol{z}\|_\infty &\leq m\mathbb{E}\|\boldsymbol{z}\|_\infty^3 \leq m\left(\mathbb{E}\|\boldsymbol{z}\|_9^9\right)^{\frac{1}{3}} \leq Cm^{\frac{4}{3}}.
\end{aligned}$$

This gives us,

$$\left|\mathbb{E}\frac{\boldsymbol{z}^\mathsf{T}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\boldsymbol{z}}{m} - \mathbb{E}\frac{\widetilde{\boldsymbol{z}}^\mathsf{T}\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}})\widetilde{\boldsymbol{z}}}{m}\mathbb{I}\left(\mathcal{E}\right)\right| \to 0,$$

and hence we have shown,

$$\lim_{m\to\infty} \frac{\mathbb{E}\boldsymbol{z}^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi},\boldsymbol{Z})\boldsymbol{z}}{m} = \lim_{m\to\infty} \mathbb{E}\frac{\widetilde{\boldsymbol{z}}^{\mathsf{T}}\mathcal{A}(\boldsymbol{\Psi},\widetilde{\boldsymbol{Z}})\widetilde{\boldsymbol{z}}}{m}\mathbb{I}\left(\mathcal{E}\right),$$

provided the latter limit exists. □

C.4.3 Proof of Lemma 28

Proof of Lemma 28. Recall that:

$$\widetilde{z}_{a_1}\widetilde{z}_{a_{k+1}}\prod_{i=1}^{k} q_i(\widetilde{z}_{a_i}) = Q_{\mathscr{F}}(\widetilde{z}_{a_1})\cdot Q_{\mathscr{L}}(\widetilde{z}_{a_{k+1}})\left(\prod_{i\in\mathscr{S}(\pi)} q_{i-1}(\widetilde{z}_{a_i})\right)\prod_{i=1}^{|\pi|-|\mathscr{S}(\pi)|-2}(Q_{\mathcal{V}_i}(z_{a_{\mathcal{V}_i}})+\mu_{\mathcal{V}_i})$$

Hence,

$$\mathbb{E}[\widetilde{z}_{a_1}q_1(\widetilde{z}_{a_2})q_2(\widetilde{z}_{a_3})\cdots q_{k-1}(\widetilde{z}_{a_k})\widetilde{z}_{a_{k+1}}|\boldsymbol{A}] = \sum_{V\subset[|\pi|-|\mathscr{S}(\pi)|-2]} \mathbb{E}\left[Q_{\mathscr{F}}(\widetilde{z}_{a_1})Q_{\mathscr{L}}(\widetilde{z}_{a_{k+1}})\left(\prod_{i\in\mathscr{S}(\pi)} q_{i-1}(\widetilde{z}_{a_i})\right)\prod_{i\in V}(Q_{\mathcal{V}_i}(\widetilde{z}_{a_{\mathcal{V}_i}}))\Big|\boldsymbol{A}\right]\left(\prod_{i\notin V}\mu_{\mathcal{V}_i}\right) \tag{C.11}$$

We now apply Mehler's formula to estimate the above conditional expectations. We first check the conditions for Mehler's formula:

1. The random variables $\widetilde{\boldsymbol{z}}$ are marginally $\mathcal{N}(0,1)$. Define $\boldsymbol{\Sigma} = \mathbb{E}[\widetilde{\boldsymbol{z}}\widetilde{\boldsymbol{z}}^{\mathsf{T}}|\boldsymbol{A}]$. $\widetilde{\boldsymbol{z}}$ and are weakly correlated on the event $\mathcal{E}$ since:

$$\begin{aligned}\max_{i\neq j}|\Sigma_{ij}| &= \left|\frac{(\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}})_{ij}/\kappa}{\sigma_i\sigma_j}\right| \\ &= \left|\frac{(\boldsymbol{\Psi})_{ij}/\kappa}{\sigma_i\sigma_j}\right| \\ &\leq C\sqrt{\frac{\log^3(m)}{m\kappa^2}},\ \text{for } m \text{ large enough},\end{aligned}$$

where C denotes a universal constant.

2. Let $S \subset [m]$ with $|S| \leq k+2$. Let $\mathbf{\Sigma}_{S,S}$ denote the principal submatrix of $\mathbf{\Sigma}$ formed by picking rows and columns in S. Then by Gershgorin's Circle theorem, on the event $\mathcal{E}$,

$$\begin{aligned}\lambda_{\min}(\mathbf{\Sigma}) &\geq 1 - (k+1)\max_{i \neq j}|\Sigma_{ij}| \\ &\geq 1 - C(k+1)\sqrt{\frac{\log^3(m)}{m\kappa^2}} \\ &\geq \frac{1}{2}, \quad \text{for } m \text{ large enough.}\end{aligned}$$

3. Note that for $\xi \sim \mathcal{N}(0,1)$, we have,

$$\mathbb{E}Q_{\mathscr{F}}(\xi) = 0,\ \mathbb{E}Q_{\mathscr{L}}(\xi) = 0 \ \text{ (Since they are odd functions, see (5.24), (5.26)),}$$

$$\mathbb{E}q_{i-1}(\xi) = \mathbb{E}\xi q_{i-1}(\xi) = 0\ \forall\ i \in \mathscr{S}(\pi) \ \text{ (They are centered, even functions, see Def. 7),}$$

$$\mathbb{E}Q_{\mathcal{V}_i}(\xi) = \mathbb{E}\xi Q_{\mathcal{V}_i}(\xi) = 0\ \forall\ i\ \in\ [|\pi| - |\mathscr{S}(\pi)| - 2] \ \text{ (See (5.28))}$$

Hence applying the first non-zero term in Mehler's Expansion (Proposition 15) of the conditional expectation:

$$\mathbb{E}\left[Q_{\mathscr{F}}(\widetilde{z}_{a_1}) \cdot Q_{\mathscr{L}}(\widetilde{z}_{a_{k+1}}) \cdot \left(\prod_{i \in \mathscr{S}(\pi)} q_{i-1}(\widetilde{z}_{a_i})\right) \cdot \prod_{i \in V}(Q_{\mathcal{V}_i}(\widetilde{z}_{a_{\mathcal{V}_i}}))\Big| \boldsymbol{A}\right]$$

has total weight $\|\boldsymbol{w}\|$ given by:

$$\|\boldsymbol{w}\| \geq \frac{1+1+2|\mathscr{S}(\pi)| + 2|V|}{2} = 1 + |\mathscr{S}(\pi)| + |V|.$$

Hence, by Proposition 15 we have,

$$\mathbb{I}(\mathcal{E}) \cdot \left| \mathbb{E}\left[Q_{\mathscr{F}}(\widetilde{z}_{a_1}) \cdot Q_{\mathscr{L}}(\widetilde{z}_{a_{k+1}}) \cdot \left(\prod_{i \in \mathscr{S}(\pi)} q_{i-1}(\widetilde{z}_{a_i}) \right) \cdot \prod_{i \in V} (Q_{\mathcal{V}_i}(\widetilde{z}_{a_{\mathcal{V}_i}})) \middle| \boldsymbol{A} \right] \right|$$
$$\leq C(\mathcal{A}) (\max_{i \neq j} |\Sigma_{i,j}|)^{1+|\mathscr{S}(\pi)|+|V|} \leq C(\mathcal{A}) \cdot \left(\frac{\log^2(m)}{m\kappa^2} \right)^{\frac{1+|\mathscr{S}(\pi)|+|V|}{2}}, \quad \text{(C.12)}$$

where $C(\mathcal{A})$ denotes a finite constant depending only on the functions $q_{1:k}$. When $V = \emptyset$ we will also need to estimate the leading order term more accurately. Define,

$$\mathcal{G}_1(\pi) \stackrel{\text{def}}{=} \big\{ \boldsymbol{w} \in \mathcal{G}(k+1) : \mathsf{d}_1(\boldsymbol{w}) = 1,\ \mathsf{d}_{k+1}(\boldsymbol{w}) = 1,\ \mathsf{d}_i(\boldsymbol{w}) = 2\ \forall\ i\ \in\ \mathscr{S}(\pi),$$
$$\mathsf{d}_i(\boldsymbol{w}) = 0\ \forall\ i\ \notin\ \{1, k+1\} \cup \mathscr{S}(\pi) \big\}.$$

By Mehler's formula, on the event $\mathcal{E}$, we have:

$$\left| \mathbb{E}\left[Q_{\mathscr{F}}(\widetilde{z}_{a_1}) \cdot Q_{\mathscr{L}}(\widetilde{z}_{a_{k+1}}) \cdot \left(\prod_{i \in \mathscr{S}(\pi)} q_{i-1}(\widetilde{z}_{a_i}) \right) \middle| \boldsymbol{A} \right] - \sum_{\boldsymbol{w} \in \mathcal{G}_1(\pi)} \hat{g}(\boldsymbol{w}, \boldsymbol{\Psi}) \cdot \mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) \right|$$
$$\leq C(\mathcal{A}) \cdot \left(\frac{\log^3(m)}{m\kappa^2} \right)^{\frac{2+|\mathscr{S}(\pi)|}{2}},$$

where,

$$\hat{g}(\boldsymbol{w}, \boldsymbol{\Psi}) = \frac{1}{\boldsymbol{w}!} \cdot \left(\prod_{i=1}^{k+1} \frac{1}{\sigma_{a_i}^{\mathsf{d}_i(\boldsymbol{w})}} \right) \cdot \left(\hat{Q}_{\mathscr{F}}(1) \hat{Q}_{\mathscr{L}}(1) \prod_{i \in \mathscr{S}(\pi)} \hat{q}_{i-1}(2) \right) \frac{1}{\kappa^{\|\boldsymbol{w}\|}},$$

and $\mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a})$ are matrix moments as defined in Definition 8. Note that the coefficients $\hat{g}(\boldsymbol{w}, \boldsymbol{\Psi})$ depend on $\boldsymbol{\Psi}$ since,

$$\sigma_i^2 = 1 + \frac{\boldsymbol{\Psi}_{ii}}{\kappa},$$

but we can remove this dependence. On the event $\mathcal{E}$, note that,

$$\max_{i\in[m]} |\sigma_{ii}^2 - 1| \leq C\sqrt{\frac{\log^3(m)}{m\kappa^2}}.$$

Hence defining:

$$\hat{g}(\boldsymbol{w},\pi) = \frac{1}{\boldsymbol{w}!}\cdot\left(\hat{Q}_{\mathscr{F}}(1)\hat{Q}_{\mathscr{L}}(1)\prod_{i\in\mathscr{S}(\pi)}\hat{q}_{i-1}(2)\right)\frac{1}{\kappa^{\|\boldsymbol{w}\|}},$$

we have, for m large enough and on the event $\mathcal{E}$,

$$|\hat{g}(\boldsymbol{w},\pi) - \hat{g}(\boldsymbol{w},\boldsymbol{\Psi})| \leq C_k\sqrt{\frac{\log^3(m)}{m\kappa^2}}.$$

Furthermore, we have the estimate,

$$\begin{aligned}|\mathcal{M}(\boldsymbol{\Psi},\boldsymbol{w},\pi,\boldsymbol{a})| &\leq (\max_{i,j}|\Psi_{ij}|)^{\|\boldsymbol{w}\|_1}\\ &\overset{\text{(a)}}{\leq} C\left(\frac{\log^3(m)}{m\kappa^2}\right)^{\frac{1+|\mathscr{S}(\pi)|}{2}},\end{aligned}$$

where in the step (a), we used the definition of the event $\mathcal{E}$ in (5.23) and the fact that $\|\boldsymbol{w}\| = 1 + |\mathscr{S}(\pi)|$ for any $\boldsymbol{w} \in \mathcal{G}_1(\pi)$. Hence we obtain, on the event $\mathcal{E}$,

$$\begin{aligned}\left|\mathbb{E}\left[Q_{\mathscr{F}}(\widetilde{z}_{a_1})\cdot Q_{\mathscr{L}}(\widetilde{z}_{a_{k+1}})\cdot\left(\prod_{i\in\mathscr{S}(\pi)}q_{i-1}(\widetilde{z}_{a_i})\right)\Big|\boldsymbol{A}\right] - \sum_{\boldsymbol{w}\in\mathcal{G}_1(\pi)}\hat{g}(\boldsymbol{w},\pi)\cdot\mathcal{M}(\boldsymbol{\Psi},\boldsymbol{w},\pi,\boldsymbol{a})\right|&\\ \leq C(\mathcal{A})\cdot\left(\frac{\log^3(m)}{m\kappa^2}\right)^{\frac{2+|\mathscr{S}(\pi)|}{2}},&\end{aligned}$$

Combining this estimate with (C.11) and (C.12) gives us:

$$\mathbb{I}\left(\mathcal{E}\right) \cdot \left| \mathbb{E}[\widetilde{z}_{a_1} q_1(\widetilde{z}_{a_2}) q_2(\widetilde{z}_{a_3}) \cdots q_{k-1}(\widetilde{z}_{a_k}) \widetilde{z}_{a_{k+1}} | \boldsymbol{A}] - \sum_{\boldsymbol{w} \in \mathcal{G}_1(\pi)} g(\boldsymbol{w}, \pi) \cdot \mathcal{M}(\boldsymbol{\Psi}, \boldsymbol{w}, \pi, \boldsymbol{a}) \right| \\ \leq C(\mathcal{A}) \cdot \left(\frac{\log^3(m)}{m\kappa^2} \right)^{\frac{2+|\mathscr{S}(\pi)|}{2}},$$

where,

$$g(\boldsymbol{w}, \pi) = \frac{1}{\kappa^{\|\boldsymbol{w}\|} \boldsymbol{w}!} \cdot \left(\hat{Q}_{\mathscr{F}}(1) \hat{Q}_{\mathscr{L}}(1) \prod_{i \in \mathscr{S}(\pi)} \hat{q}_{i-1}(2) \right) \cdot \left(\prod_{i \in [|\pi| - |\mathscr{S}(\pi)| - 2]} \mu_{\mathcal{V}_i} \right)$$

$$\mathcal{G}_1(\pi) \stackrel{\text{def}}{=} \big\{ \boldsymbol{w} \in \mathcal{G}(k+1) : \mathsf{d}_1(\boldsymbol{w}) = 1,\ \mathsf{d}_{k+1}(\boldsymbol{w}) = 1,\ \mathsf{d}_i(\boldsymbol{w}) = 2\ \forall\ i\ \in\ \mathscr{S}(\pi), \\ \mathsf{d}_i(\boldsymbol{w}) = 0\ \forall\ i\ \notin\ \{1, k+1\} \cup \mathscr{S}(\pi) \big\},$$

and $C(\mathcal{A})$ denotes a constant depending only on the functions appearing in $\mathcal{A}$ and k. This was precisely the claim of Lemma 28. □

C.5 Proof of Proposition 15

Proof of Proposition 15. Let $\psi(\boldsymbol{z}; \boldsymbol{\Sigma})$ denote the density of a k dimensional zero mean Gaussian vector with positive definite covariance matrix $\boldsymbol{\Sigma}$ i.e. $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. Suppose that $\Sigma_{ii} = 1\ \forall\ i \in [k]$. In this situation Slepian [62] has found an explicit expression for the Taylor series expansion of $\psi(\boldsymbol{z}; \boldsymbol{\Sigma})$ around $\boldsymbol{\Sigma} = \boldsymbol{I}_k$ given by:

$$\psi(\boldsymbol{z}; \boldsymbol{\Sigma}) = \sum_{\boldsymbol{w} \in \mathcal{G}(k)} \frac{D_{\boldsymbol{\Sigma}}^{\boldsymbol{w}}\ \psi(\boldsymbol{z}; \boldsymbol{I}_k)}{\boldsymbol{w}!} \cdot \left(\prod_{i<j} \Sigma_{ij}^{w_{ij}} \right),$$

where $D^{\boldsymbol{w}}_{\boldsymbol{\Sigma}}\ \psi(\boldsymbol{z};\boldsymbol{I}_k)$ denotes the derivative:

$$D^{\boldsymbol{w}}_{\boldsymbol{\Sigma}}\ \psi(\boldsymbol{z};\boldsymbol{I}_k) \stackrel{\text{def}}{=} \frac{\partial^{\|\boldsymbol{w}\|}}{\partial\Sigma_{12}^{w_{12}}\ \partial\Sigma_{13}^{w_{13}}\cdots\partial\Sigma_{23}^{w_{23}}\ \partial\Sigma_{24}^{w_{24}}\cdots\partial\Sigma_{k-1,k}^{w_{k-1,k}}}\ \psi(\boldsymbol{z};\boldsymbol{\Sigma})\Big|_{\boldsymbol{\Sigma}=\boldsymbol{I}_k}$$
$$= \left(\prod_{i=1}^{k} H_{\mathsf{d}_i(\boldsymbol{w})}(z_i)\right)\cdot\psi(\boldsymbol{z};\boldsymbol{I}_k).$$

We intend to integrate the Taylor series for $\psi(\boldsymbol{z};\boldsymbol{\Sigma})$ to obtain the expansion for the expectation in Proposition 15. In order to do so we need to understand the truncation error in the Taylor Series. By Taylor's Theorem, we know that:

$$\psi(\boldsymbol{z};\boldsymbol{\Sigma}) - \sum_{\boldsymbol{w}\in\mathcal{G}(k):\|\boldsymbol{w}\|\le t}\frac{D^{\boldsymbol{w}}_{\boldsymbol{\Sigma}}\ \psi(\boldsymbol{z};\boldsymbol{I}_k)}{\boldsymbol{w}!}\cdot\left(\prod_{i<j}\Sigma_{ij}^{w_{ij}}\right) = \sum_{\boldsymbol{w}\in\mathcal{G}(k):\|\boldsymbol{w}\|=t+1}\frac{D^{\boldsymbol{w}}_{\boldsymbol{\Sigma}}\ \psi(\boldsymbol{z};\boldsymbol{\Sigma}_\gamma)}{\boldsymbol{w}!}\cdot\boldsymbol{\Sigma}^{\boldsymbol{w}}, \tag{C.14}$$

where $\boldsymbol{\Sigma}_\gamma = \gamma\boldsymbol{\Sigma} + (1-\gamma)\boldsymbol{I}_k$ for some $\gamma \in (0,1)$. Slepian has further showed the following remarkable identity:

$$D^{\boldsymbol{w}}_{\boldsymbol{\Sigma}}\ \psi(\boldsymbol{z};\boldsymbol{\Sigma}) = \frac{\partial^{2\|\boldsymbol{w}\|}}{\partial z_1^{\mathsf{d}_1(\boldsymbol{w})}\ \partial z_2^{\mathsf{d}_2(\boldsymbol{w})}\cdots\partial z_k^{\mathsf{d}_k(\boldsymbol{w})}}\ \psi(\boldsymbol{z};\boldsymbol{\Sigma}).$$

An inductive calculation shows that the ratio:

$$\frac{1}{\psi(\boldsymbol{z};\boldsymbol{\Sigma})}\ \frac{\partial^{2\|\boldsymbol{w}\|}}{\partial z_1^{\mathsf{d}_1(\boldsymbol{w})}\ \partial z_2^{\mathsf{d}_2(\boldsymbol{w})}\cdots\partial z_k^{\mathsf{d}_k(\boldsymbol{w})}}\ \psi(\boldsymbol{z};\boldsymbol{\Sigma}),$$

is a polynomial of degree $4\|\boldsymbol{w}\|$ in the variables $z_1, z_2 \ldots z_k, \{(\boldsymbol{\Sigma}^{-1})_{ij}\}_{i<j}$. Hence:

$$\left|\frac{1}{\psi(\boldsymbol{z};\boldsymbol{\Sigma})}\ \frac{\partial^{2\|\boldsymbol{w}\|}}{\partial z_1^{\mathsf{d}_1(\boldsymbol{w})}\ \partial z_2^{\mathsf{d}_2(\boldsymbol{w})}\cdots\partial z_k^{\mathsf{d}_k(\boldsymbol{w})}}\ \psi(\boldsymbol{z};\boldsymbol{\Sigma})\right| \le$$
$$C_{\|\boldsymbol{w}\|}\cdot\left(1+\sum_{i<j}|(\boldsymbol{\Sigma}^{-1})_{ij}|^{4\|\boldsymbol{w}\|}+\sum_{i=1}^{k}|z_i|^{4\|\boldsymbol{w}\|}\right),$$

where $C_{\|\boldsymbol{w}\|}$ denotes a constant depending only on $\|\boldsymbol{w}\|$. Observing that:

$$(\boldsymbol{\Sigma}^{-1})_{ij} \leq \|\boldsymbol{\Sigma}^{-1}\|_{\mathsf{op}} = \frac{1}{\lambda_{\min}(\boldsymbol{\Sigma})} < \infty.$$

This gives us:

$$\left| \frac{1}{\psi(\boldsymbol{z};\boldsymbol{\Sigma})} \frac{\partial^{2\|\boldsymbol{w}\|}}{\partial z_1^{\mathsf{d}_1(\boldsymbol{w})} \partial z_2^{\mathsf{d}_2(\boldsymbol{w})} \cdots \partial z_k^{\mathsf{d}_k(\boldsymbol{w})}} \psi(\boldsymbol{z};\boldsymbol{\Sigma}) \right| \leq C_{\|\boldsymbol{w}\|} \left(1 + \frac{k^2}{\lambda_{\min}^{4\|\boldsymbol{w}\|}(\boldsymbol{\Sigma})} + \sum_{i=1}^{k} |z_i|^{4\|\boldsymbol{w}\|} \right).$$

Substituting this estimate in (C.14) gives us:

$$\begin{aligned} &\left| \psi(\boldsymbol{z};\boldsymbol{\Sigma}) - \sum_{\boldsymbol{w}\in\mathcal{G}(k):\|\boldsymbol{w}\|\leq t} \frac{D_{\boldsymbol{\Sigma}}^{\boldsymbol{w}}\, \psi(\boldsymbol{z};\boldsymbol{I}_k)}{\boldsymbol{w}!} \cdot \boldsymbol{\Sigma}^{\boldsymbol{w}} \right| \\ &\qquad \leq C_{t,k} \cdot \left(1 + \frac{k^2}{\lambda_{\min}^{4t+4}(\boldsymbol{\Sigma}_\gamma)} + \sum_{i=1}^{k} |z_i|^{4t+4} \right) \cdot \left(\max_{i\neq j} |\Sigma_{ij}| \right)^{t+1} \cdot \psi(\boldsymbol{z};\boldsymbol{\Sigma}_\gamma). \end{aligned}$$

Note that $\lambda_{\min}(\boldsymbol{\Sigma}_\gamma) = \gamma + (1-\gamma)\lambda_{\min}(\boldsymbol{\Sigma}) \geq \min(1, \lambda_{\min}(\boldsymbol{\Sigma}))$. Hence,

$$\begin{aligned} &\left| \psi(\boldsymbol{z};\boldsymbol{\Sigma}) - \sum_{\boldsymbol{w}\in\mathcal{G}(k):\|\boldsymbol{w}\|\leq t} \frac{D_{\boldsymbol{\Sigma}}^{\boldsymbol{w}}\, \psi(\boldsymbol{z};\boldsymbol{I}_k)}{\boldsymbol{w}!} \cdot \boldsymbol{\Sigma}^{\boldsymbol{w}} \right| \\ &\qquad \leq C_{t,k} \cdot \left(1 + \frac{k^2}{\min(\lambda_{\min}^{4t+4}(\boldsymbol{\Sigma}), 1)} + \sum_{i=1}^{k} |z_i|^{4t+4} \right) \cdot \left(\max_{i\neq j} |\Sigma_{ij}| \right)^{t+1} \cdot \psi(\boldsymbol{z};\boldsymbol{\Sigma}_\gamma). \end{aligned}$$

Using this expansion to compute the expectation of $\prod_{i=1}^k f_i(z_i)$ we obtain:

$$\left| \mathbb{E}\left[\prod_{i=1}^{k} f_i(z_i) \right] - \sum_{\substack{\boldsymbol{w}\in\mathcal{G}(k) \\ \|\boldsymbol{w}\|\leq t}} \left(\prod_{i=1}^{k} \hat{f}_i(\mathsf{d}_i(\boldsymbol{w})) \right) \cdot \frac{\boldsymbol{\Sigma}^{\boldsymbol{w}}}{\boldsymbol{w}!} \right| \leq C \left(1 + \frac{1}{\lambda_{\min}^{4t+4}(\boldsymbol{\Sigma})} \right) \left(\max_{i\neq j} |\Sigma_{ij}| \right)^{t+1},$$

where $C = C_{t,k,f_{1:k}}$ denotes a constant depending only on t, k and the functions $f_{1:k}$. In obtaining the above estimate we use the fact that since the functions f_i have polynomial growth and

marginally $z_i \sim \mathcal{N}(0,1)$ under the measure $\mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_\gamma\right)$ (since $(\Sigma_\gamma)_{ii} = 1$) we have,

$$\mathbb{E}_{\boldsymbol{z}\sim\mathcal{N}\left(\mathbf{0},\mathbf{\Sigma}_\gamma\right)}\left[|z_i|^{4t+4}\prod_{j=1}^{k}|f_j(z_j)|\right] \leq \sum_{j=1}^{k}\mathbb{E}_{\boldsymbol{z}\sim\mathcal{N}\left(\mathbf{0},\mathbf{\Sigma}_\gamma\right)}\left[|z_i|^{4t+4}|f_j(z_j)|^k\right] = C_{t,k,f_{1:k}} < \infty.$$

□

C.6 Some Miscellaneous Facts

Fact 4 (Hanson-Wright Inequality [94])**.** *Let $\boldsymbol{x} = (x_1, x_2 \ldots, x_n) \in \mathbb{R}^n$ be a random vector with independent 1-subgaussian, zero mean components. Let $\boldsymbol{A}$ be an $n \times n$ matrix. Then, for every $t \geq 0$,*

$$\mathbb{P}\left(|\boldsymbol{x}^\mathsf{T}\boldsymbol{A}\boldsymbol{x} - \mathbb{E}\boldsymbol{x}^\mathsf{T}\boldsymbol{A}\boldsymbol{x}| > t\right) \leq 2\exp\left(-c\min\left(\frac{t^2}{\|\boldsymbol{A}\|_{\mathsf{Fr}}^2}, \frac{t}{\|\boldsymbol{A}\|_{\mathsf{op}}}\right)\right).$$

Fact 5 (Gaussian Poincare Inequality)**.** *Let $\boldsymbol{x} \sim \mathcal{N}(0, \boldsymbol{I}_n)$. Then, for any L-Lipchitz function $f : \mathbb{R}^n \to \mathbb{R}$ we have,*

$$\mathrm{Var}(f(\boldsymbol{x})) \leq L^2.$$

Fact 6 (Moments of a Random Unit vector, Lemma 2.22 & Proposition 2.5 of [93])**.** *Let $\boldsymbol{x} \sim$ Unif$(\mathbb{S}_{n-1})$. Let i, j, k, ℓ be distinct indices. Then:*

$$\mathbb{E}x_i^4 = \frac{3}{n(n+2)},\ \mathbb{E}x_i^2x_j^2 = \frac{n+1}{n(n-1)(n+2)}\ \mathbb{E}x_i^3x_j = 0\ \mathbb{E}x_ix_jx_k^2 = 0,\ \mathbb{E}x_ix_jx_kx_l = 0.$$

Furthermore, there exists a universal constant C such that, for any $t \in \mathbb{N}$:

$$\mathbb{E}|x_i|^t \leq \left(\frac{Ct}{m}\right)^{\frac{t}{2}}.$$

Fact 7 (Concentration on the Sphere, Ball [95])**.** *Let* $\boldsymbol{x} \sim$ *Unif*$(\mathbb{S}_{n-1})$*. Then*

$$\mathbb{P}\left(|x_1| \geq \epsilon\right) \leq 2e^{-n\epsilon^2/2}.$$

Fact 8 (ℓ_∞ norm of a random unit vector)**.** $\boldsymbol{x} \sim$ *Unif*$(\mathbb{S}_{n-1})$*. Then*

$$\mathbb{E}\|\boldsymbol{x}\|_\infty^t \leq \left(\frac{C\log(n)}{n}\right)^{\frac{t}{2}},$$

for a universal constant C*.*

Proof. For a random unit vector we can control $\mathbb{E}\|\boldsymbol{x}\|_\infty^t$ as follows. Let $q \in \mathbb{N}$ be a parameter to be set suitably. Then,

$$\begin{aligned}
\mathbb{E}\|\boldsymbol{x}\|_\infty^t &\leq \left(\mathbb{E}\|\boldsymbol{x}\|_\infty^{qt}\right)^{\frac{1}{q}} \\
&\leq \left(\sum_{i=1}^n \mathbb{E}|x_i|^{qt}\right)^{\frac{1}{q}} \\
&\overset{(a)}{=} \left(n\mathbb{E}|x_1|^{qt}\right)^{\frac{1}{q}} \\
&\overset{(b)}{=} n^{\frac{1}{q}} \cdot q^{\frac{t}{2}} \cdot \left(\frac{Ct}{n}\right)^{\frac{t}{2}} \\
&\overset{(c)}{\leq} e^t \cdot (2\log(n))^{\frac{t}{2}} \cdot \left(\frac{C}{n}\right)^{\frac{t}{2}}.
\end{aligned}$$

In the step marked (a) we used the fact that the coordinates of a random unit vector are exchangeable, in (b) we used the fact that u_1 is C/m-subgaussian (see Fact 7) and in (c) we set $q = \lfloor \frac{2\log(n)}{t} \rfloor$. □

Fact 9 (Poincare Inequality for Haar Measure, Gromov and Milman [96])**.** *Consider the following setups:*

1. *Let* $\boldsymbol{O} \sim$ *Unif*$\big(\mathbb{O}(m)\big)$ *and* $f : \mathbb{R}^{m\times m} \to \mathbb{R}$ *be a function such that:*

$$f(\boldsymbol{O}) = f(\boldsymbol{O}\boldsymbol{D}),\ \boldsymbol{D} = \textit{Diag}\left(1, 1, 1, \ldots, 1, \mathsf{sign}(\det(\boldsymbol{O}))\right), \tag{C.15}$$

then,

$$\mathrm{Var}(f(\boldsymbol{O})) \leq \frac{8}{m} \cdot \mathbb{E}\|\nabla f(\boldsymbol{O})\|_{\mathsf{Fr}}^2.$$

for any $m \geq 4$*.*

2. *Let* $\boldsymbol{O} \sim$ *Unif*$\big(\mathbb{U}(m)\big)$ *and* $f : \mathbb{C}^{m\times m} \to \mathbb{R}$*. Then,*

$$\mathrm{Var}(f(\boldsymbol{O})) \leq \frac{8}{m} \cdot \mathbb{E}\|\nabla f(\boldsymbol{O})\|_{\mathsf{Fr}}^2.$$

Proof. This result is due to Gromov and Milman [96]. Our reference for these inequalities was the book of Meckes [93]. Theorem 5.16 of Meckes shows that Haar measures on $\mathbb{SO}(m), \mathbb{U}(m)$ satisfy Log-sobolev inequality with constant $8/m$. It is well known that Log-Sobolev Inequality implies the Poincare Inequality (see for e.g. Lemma 8.12 in Handel [97]). Note that, in the real case we only obtain the Poincare inequality for the Haar measure on $\mathbb{SO}(m)$, condition (C.15) ensures the result still holds for $\boldsymbol{O} \sim$ Unif $\big(\mathbb{O}(m)\big)$. □

Fact 10 (Continuity of Matrix Square Root [98, Lemma 2.2])**.** *For any two symmetric positive semi-definite matrices* $\boldsymbol{M}_1, \boldsymbol{M}_2$ *we have,*

$$\|\boldsymbol{M}_1^{\frac{1}{2}} - \boldsymbol{M}_2^{\frac{1}{2}}\|_{\mathsf{op}} \leq \frac{\|\boldsymbol{M}_1 - \boldsymbol{M}_2\|_{\mathsf{op}}}{\sqrt{\lambda_{\min}(\boldsymbol{M}_1)}}.$$