EPFL

Thèse n° 10 543

Spectral Estimators for High-Dimensional Matrix Inference

Présentée le 31 mai 2024

Faculté informatique et communications
Laboratoire de mécanique statistique de l'inférence pour grands systèmes
Programme doctoral en informatique et communications

pour l'obtention du grade de Docteur ès Sciences

par

Farzad POURKAMALI

Acceptée sur proposition du jury

Prof. P. Thiran, président du jury
Prof. N. Macris, directeur de thèse
Prof. M. Mezard, rapporteur
Prof. M. Mondelli, rapporteur
Prof. E. Telatar, rapporteur

École polytechnique fédérale de Lausanne

2024

Thesis No. 10543

Thesis presented to the faculty of computer and communication sciences for obtaining the degree of Docteur ès Sciences

Accepted by the jury:

N. Macris
Thesis directors

M. Mezard
Expert

E. Telatar
Expert

M. Mondelli
Expert

P. Thiran
President of the jury

Ecole Polytechnique Fédérale de Lausanne, 2024

1 Abstract

A key challenge across many disciplines is to extract meaningful information from data which is often obscured by noise. These datasets are typically represented as large matrices. Given the current trend of ever-increasing data volumes, with datasets growing larger and more complex, it is necessary to develop matrix inference methodologies which provide us with the tools to deal with high-dimensional matrices.

This thesis presents a theoretical exploration of high-dimensional matrix inference problems. The high-dimensional nature of the matrices makes them amenable to the application of statistical methods in the high-dimensional limit. We primarily investigate spectral estimators, which are based on the spectral properties of matrices and constructed using their singular vectors or eigenvectors. The methodologies employed are rooted in random matrix theory and statistical physics, alongside results from the high-dimensional limits of spherical integrals. This approach provides a comprehensive theoretical framework for understanding matrix inference in the context of large-scale data.

We begin by studying low-rank estimation problems in the mismatched setting, where perfect knowledge of the priors for both signal and noise is not available. In this scenario, we derive the exact analytic expression for the asymptotic mean squared error (MSE) in the large system size limit for the particular case of Gaussian priors and additive noise for both symmetric and non-symmetric signals. Our formulas demonstrate that in the mismatched case, effective estimation is achievable, and the minimum MSE (MMSE) can be attained by selecting a non-trivial set of parameters beyond the matched parameters. Furthermore, we compare the performance of the spectral algorithms and Approximate Message Passing (AMP) in the mismatched setting.

In the latter part of the thesis, we explore the extensive-rank matrix inference

problems using the framework of rotationally invariant estimators (RIEs). In the symmetric case, we study the asymptotic mutual information and MMSE of denoising problem under Gaussian noise. Moreover, we extend RIEs to accommodate rectangular matrices for general rotational invariant noise matrices. Consequently, we derive the asymptotic MMSE in this setting. Finally, we investigate a statistical model for matrix factorization, and derive analytical formulas for the optimal RIE to reconstruct the two matrix factors, given the noisy observation of their product.

Keywords: Matrix denoising, matrix factorization, Bayesian inference, mismatched estimation, rotational invariant estimators, random matrix theory, spherical integrals, replica method

2 Résumé

Dans de nombreuses disciplines, l'un des défis majeurs consiste à extraire de l'information souvent obscurcies par le bruit. Ces données sont généralement représentées sous forme de matrices de grande taille. Compte tenu des tendances actuelles d'augmentation constante du volume, dimension et complexité de ces données, il est nécessaire de développer des méthodologies d'inférence matricielle pour se munir des outils nécessaires afin de traiter ces matrices de grande dimension.

Cette thèse présente une exploration théorique des problèmes d'inférence matricielles en haute dimension. La structure de ces matrices permettent l'application de méthodes statistiques dans la limite de haute dimension. Nous examinons principalement les estimateurs spectraux, qui reposent sur les propriétés spectrales des matrices et construits à l'aide de leurs vecteurs singuliers ou de leurs vecteurs propres. Les méthodologies employées s'inscrivent dans la théorie des matrices aléatoires et la physique statistique, ainsi que dans les résultats provenant des limites en haute dimension des intégrales sphériques. Cette approche offre un cadre théorique complet pour comprendre l'inférence matricielle dans le contexte de données à grande échelle.

Nous commençons par étudier mes problèmes d'estimation de rang faible dans un cadre non-apparié (ou "mismatch"), où la connaissance parfaite de la distribution a priori ("prior") du signal et du bruit n'est pas disponible. Dans ce scénario, nous dérivons l'expression analytique exacte de l'erreur quadratique moyenne (MSE) asymptotique dans la limite d'un systèmé de grande taille, dans le cas particulier d'un signal gaussien ainsi que d'un bruit additif pour les signaux symétriques et non symétriques. Nos formules démontrent que dans le cas non-apparié, une estimation efficace est réalisable, et l'erreur quadratique moyenne minimale (MMSE) peut être atteinte en sélectionnant un ensemble non trivial de paramètres au-delà des paramètres appariés. En outre, nous

comparons la performance des algorithmes spectraux et de "l'Approximate Message Passing" (AMP) dans le cadre non-apparié.

Dans la dernière partie de la thèse, nous explorons les problèmes d'inférence matricielle de rang extensif en utilisant le cadre des estimateurs rotationnellement invariants (RIEs). Dans le cas symétrique, nous étudions l'information mutuelle asymptotique et le "MMSE" du problème de débruitage sous bruit gaussien. De plus, nous étendons les "RIEs" pour prendre en compte les matrices rectangulaire pour les matrices de bruit invariantes rotationnelles générales. Par conséquent, nous dérivons le MMSE asymptotique dans ce cadre. Enfin, nous étudions un modèle statistique pour la factorisation de matrices et dérivons des formules analytiques pour le RIE optimal pour reconstruire les deux facteurs de matrice, étant donné l'observation bruitée de leur produit.

Mots-clés: Débruitage de matrices, factorisation de matrices, inférence bayésienne, estimation discordante, estimateurs rotationnellement invariants, théorie des matrices aléatoires, intégrales sphériques, méthode des répliques

3 Acknowledgements

I am deeply thankful to Nicolas Macris, my PhD supervisor, for his invaluable guidance and unwavering support throughout my doctoral journey. He facilitated my growth as a researcher and enriched my experience with engaging discussions on a wide range of topics, including physics and quantum mechanics. His profound knowledge and the depth of our discussions have been instrumental in my academic and personal development. I am profoundly grateful for his mentorship and the extensive learning experience he provided.

I want to express my gratitude to Jean Barbier for the enthusiasm he brings to every discussion. His eagerness to explore new problems and ideas has significantly inspired me and shaped the direction of my research. It was a privilege to have Marc Mezard, Emre Telatar, Marco Mondelli, and Patrick Thiran on my thesis committee. I appreciate the time they invested in reviewing my work and their insightful feedback.

Throughout my PhD journey, I have been fortunate to be part of a supportive and inspiring community within the Information Processing Group (IPG) and the SMILS lab. Rüdiger Urbanke, Olivier Lévêque, and Michael Gastpar have profoundly enriched my research experience with their guidance and insights. I am deeply grateful to Muriel Bardet, our administrative assistant, for her steadfast support, and to Damir Laurenzi, our IT expert, for ensuring our technical needs were always met. My experience was also enhanced by the support and companionship of my SMILS colleagues, particularly Antoine, Rodrigo, Joon, Jean, Diego, Anastasia, Anand, and Perrine, and my IPG peers, Ashok, Andreas, Aditya, Dina, Yunus, Reka, and Millen, whose friendships have been invaluable.

Throughout the ups and downs of my PhD journey, the support from my friends played a pivotal role. I am deeply grateful to Babak, Salar, Mohammadreza, Ali, Alireza, Sadra, Shayan, Yasmin, Mohammadreza, Mo-

jan, Amirhossein, Notash, Sarina, and Masood for their constant encouragement and belief in me. Additionally, special thanks to my long-distance old friends—Soheil, Hossein, Mohsen, and Amirhossein—whose distant support was equally meaningful. The joy, resilience, and inspiration they provided, transforming every challenge into enriching experiences, were invaluable. The laughter we shared, the supportive conversations, and their faith in me have been fundamental to my growth.

This thesis is affectionately dedicated to my parents, who have been my unwavering pillars throughout this journey. Their faith in me has been a source of strength, inspiring me to face challenges and pursue my dreams. To Fatemeh, my dearest, I extend heartfelt gratitude for her enduring presence and encouragement during my PhD. Her companionship has significantly enriched this journey, enhancing every step with warmth and understanding.

Lausanne, March 2024 Farzad Pourkamali

Contents

4 Introduction

Across diverse fields, from finance and biology to image processing and machine learning, data is frequently represented as large matrices. In genomics, for instance, gene expression data forms a matrix with genes as rows and different experimental conditions or samples as columns. Similarly, in financial analytics, matrices encapsulate stock price movements over time, with rows denoting time intervals and columns representing individual stock prices. The impact of noise, inaccuracies, or missing values within these matrices is substantial, affecting analyses and model performance. The inevitability of noise is a fundamental reality in data, stemming from various sources like measurement errors, environmental fluctuations, or inherent randomness. Therefore, the ability to remove or reduce noise while preserving important features or structures within these matrices is crucial for extracting valuable insights, making informed decisions, and advancing research and applications across diverse domains.

This thesis is about estimating a signal matrix of interest from its noisy observation. Instead of considering real datasets, our focus will be on utilizing random data, within the context of a probabilistic model. While this approach may diverge from real-world application, it establishes a consistent mathematical framework and enables the derivation of precise expressions that can yield practical insights. In its simplest and general form, the problem can be formulated as follows: Given the data matrix

$$\boldsymbol{Y} = \boldsymbol{S} + \boldsymbol{Z}$$

the goal is to recover the hidden signal $\boldsymbol{S}$, where $\boldsymbol{Z}$ is the noise matrix. We study this problem in large dimensions. This setting is particularly relevant for modern applications (as "real" datasets always get larger) and presents a lot of interesting theoretical challenges.

We analyze this problem under various assumptions/constraints on the signal and the noise matrices, which can be categorized into two main regimes.

One, is the low-rank regime where the rank of the signal is small compared to the dimension, and the extensive-rank regime in which the rank is comparable to the dimension (or more precisely grows with the dimension). In the following two sections, we provide a more comprehensive introduction to the problem in these two regimes, covering existing literature and outlining the main contributions of this thesis.

4.1 Low-Rank Matrix Inference

There exist two main models studied in this regime. One is often called as *spiked Wigner* model, in which one observes

$$\boldsymbol{Y} = \sqrt{\frac{\kappa}{N}}\boldsymbol{s}\boldsymbol{s}^\intercal + \boldsymbol{Z} \tag{4.1}$$

where the "spike" $\boldsymbol{s} \in \mathbb{R}^N, \boldsymbol{s} \sim P_{\boldsymbol{s}}$ is the signal and $\boldsymbol{Z} = \boldsymbol{Z}^\intercal \in \mathbb{R}^{N\times N}$ is the noise matrix with i.i.d. Gaussian entries $\mathcal{N}(0,1)$. $\kappa \in \mathbb{R}_+$ is the signal-to-noise ratio (SNR).

The second model is the non-symmetric counterpart of the model above, called *spiked Wishart*:

$$\boldsymbol{Y} = \sqrt{\frac{\kappa}{N}}\boldsymbol{s}\boldsymbol{t}^\intercal + \boldsymbol{Z} \tag{4.2}$$

where $\boldsymbol{s} \in \mathbb{R}^N, \boldsymbol{s} \sim P_{\boldsymbol{s}}$ and $\boldsymbol{t} \in \mathbb{R}^M, \boldsymbol{t} \sim P_{\boldsymbol{t}}$ are independent, and $\boldsymbol{Z} \in \mathbb{R}^{N\times M}$ has i.i.d. Gaussian entries.

In both models (4.1), (4.2), the goal is to recover the rank-one signal matrix ($\boldsymbol{s}\boldsymbol{s}^\intercal$ or $\boldsymbol{s}\boldsymbol{t}^\intercal$) from the observation. We are interested in the regime where $N \to \infty$ (in model (4.1) it is assumed that $N/M \to \alpha > 0$). In the rest of this section, we focus on the spiked Wigner model and review the basic notions of estimation (both information-theoretically and algorithmically), and briefly mention the extensions to the spiked Wishart model. Moreover, note that for simplicity we develop the discussion for the rank-one hidden signal, but the results can be extended to any finite-rank signal (by finite we mean the rank is fixed while the dimension N tends to infinity).

4.1.1 Bayesian estimation

In Bayesian inference, it is assumed that the prior distribution of the signal and the channel characteristics (distribution of the noise and SNR) are known and one estimates the signal by the posterior mean. The posterior distribution reads:

$$P(\boldsymbol{x}|\boldsymbol{Y}) = \frac{P_{\boldsymbol{s}}(\boldsymbol{x})e^{-\frac{1}{2}\|\boldsymbol{Y}-\sqrt{\frac{\kappa}{N}}\boldsymbol{x}\boldsymbol{x}^\intercal\|_{\mathrm{F}}^2}}{\int d\boldsymbol{x}\, P_{\boldsymbol{s}}(\boldsymbol{x})e^{-\frac{1}{2}\|\boldsymbol{Y}-\sqrt{\frac{\kappa}{N}}\boldsymbol{x}\boldsymbol{x}^\intercal\|_{\mathrm{F}}^2}} \tag{4.3}$$

and the Bayesian estimator is

$$\mathbb{E}[\boldsymbol{s}\boldsymbol{s}^\intercal|\boldsymbol{Y}] = \int d\boldsymbol{x}\, \boldsymbol{x}\boldsymbol{x}^\intercal P(\boldsymbol{x}|\boldsymbol{Y}) \tag{4.4}$$

The performance of the estimator is evaluated using the average quadratic loss,

$$\frac{1}{N^2}\mathbb{E}\Big[\big\|\boldsymbol{s}\boldsymbol{s}^\intercal - \mathbb{E}[\boldsymbol{s}\boldsymbol{s}^\intercal|\boldsymbol{Y}]\big\|_{\mathrm{F}}^2\Big]$$

where the outer expectation is over the signal $\boldsymbol{s}$ and the noise matrix $\boldsymbol{Z}$. For this choice of loss function, the Bayesian estimator $\mathbb{E}[\boldsymbol{s}\boldsymbol{s}^\intercal|\boldsymbol{Y}]$ is the best estimator in the sense that it has the minimum mean-squared error (MMSE). Let $\widehat{\boldsymbol{s}\boldsymbol{s}^\intercal}(.)$ denote an estimator of $\boldsymbol{s}\boldsymbol{s}^\intercal(\boldsymbol{Y})$ given the observation $\boldsymbol{Y}$. The MMSE is defined as a function of the SNR parameter as:

$$\mathrm{MMSE}_N(\kappa) := \min_{\widehat{\boldsymbol{s}\boldsymbol{s}^\intercal}(.)} \frac{1}{N^2}\mathbb{E}\big[\|\boldsymbol{s}\boldsymbol{s}^\intercal - \widehat{\boldsymbol{s}\boldsymbol{s}^\intercal}(\boldsymbol{Y})\|_{\mathrm{F}}^2\big] = \frac{1}{N^2}\mathbb{E}\Big[\big\|\boldsymbol{s}\boldsymbol{s}^\intercal - \mathbb{E}[\boldsymbol{s}\boldsymbol{s}^\intercal|\boldsymbol{Y}]\big\|_{\mathrm{F}}^2\Big] \tag{4.5}$$

The Bayes optimal estimator not only establishes a fundamental benchmark but also acts as a lower bound for the mean-square error of any estimator, including those produced by algorithms. In practical scenarios, computing the MMSE presents challenges due to the typical inaccessibility of true prior and likelihood distributions. However, when dealing with fully-specified models, it becomes viable to determine this universal lower bound and assess the performance of existing algorithms in comparison. The evaluation of MMSE involves two crucial steps: firstly, computing the normalization factor of the posterior distribution in (4.3), and secondly, determining the posterior mean as expressed in (4.4). Both of these steps entail the computation of N-dimensional integrals, which becomes intractable for large values of N.

Yet, as presented in the next subsection, there exists a connection between MMSE and the mutual information between the signal and the observation. Given the close connection between mutual information and the *free energy* in statistical mechanics, tools developed through extensive research on large disordered systems, can be leveraged to compute the mutual information (and consequently the MMSE) in high-dimensional regimes.

4.1.2 Mutual information

The mutual information is one of the central quantities in information theory that can be interpreted as a quantification of the "shared information" between two random variables. Let X and Y be two random variables with joint distribution $P_{X,Y}$ and marginal distributions P_X, P_Y. The mutual information between X, Y is defined as:

$$I(X;Y) := \mathbb{E}\big[\ln \frac{P_{X,Y}(X,Y)}{P_X(X)P_Y(Y)}\big] \tag{4.6}$$

I-MMSE relation

As mentioned before, from the estimation perspective, the importance of the mutual information comes from its connection to the MMSE. The *I-MMSE*

relation [1] is the formula that connects the derivative of the mutual information to the MMSE, where the derivative is taken with respect to the SNR. In the simplest case a scalar random variable S observed through Gaussian channel $Y = \sqrt{\kappa}S + Z$ where $Z \sim \mathcal{N}(0,1)$, then the I-MMSE relation states:

$$\frac{\partial}{\partial \kappa} I(S;Y) = \frac{1}{2}\mathrm{MMSE}(\kappa)$$

where here the MMSE is defined for the scalar inference as $\mathrm{MMSE}(\kappa) = \mathbb{E}\Big[\big(S - \mathbb{E}[S|Y]\big)^2\Big]$.

The I-MMSE relation for the spiked Wigner model reads:

$$\frac{\partial}{\partial \kappa} \frac{I(\boldsymbol{s};\boldsymbol{Y})}{N} = \frac{1}{2}\mathrm{MMSE}(\kappa) \tag{4.7}$$

The I-MMSE relation (4.7) allows us to explore the fundamental limits of an estimation problem by examining the mutual information between observations and the signal. In particular, the behavior of mutual information can be analyzed asymptotically to determine the MMSE in high-dimensional regime. This involves computing the limit of the normalized mutual information, followed by finding the derivative of this limit using (4.7), which ultimately yields the asymptotic MMSE.

Using the law of total probability $P(\boldsymbol{Y}) = \int P(\boldsymbol{Y}|\boldsymbol{x})P_{\boldsymbol{s}}(\boldsymbol{x})\,d\boldsymbol{x}$, and $P(\boldsymbol{Y}|\boldsymbol{s}) = {}^{P(\boldsymbol{s},\boldsymbol{Y})}/_{P_{\boldsymbol{s}}(\boldsymbol{s})}$, from the definition (4.6), the normalized mutual information for the spiked Wigner model reads:

$$\frac{1}{N} I(\boldsymbol{s};\boldsymbol{Y}) = -\frac{1}{N}\mathbb{E}\Big[\ln \int P(\boldsymbol{Y}|\boldsymbol{x})P_{\boldsymbol{s}}(\boldsymbol{x})\,dx\Big] + \frac{1}{N}\mathbb{E}\big[\ln P(\boldsymbol{Y}|\boldsymbol{s})\big] \tag{4.8}$$

The second term can be written as:

$$\begin{aligned}
\frac{1}{N}\mathbb{E}\big[\ln P(\boldsymbol{Y}|\boldsymbol{s})\big] &= \frac{1}{N}\mathbb{E}\Big[\ln \frac{1}{(2\pi)^{\frac{N(N+1)}{4}}} \prod_{1\le i\le j\le N} e^{-\frac{1}{2}\left(Y_{ij} - \sqrt{\kappa}s_i s_j\right)^2}\Big] \\
&= -\frac{N+1}{4}\ln 2\pi - \frac{1}{2N}\sum_{1\le i\le j\le N}\mathbb{E}[Z_{ij}^2] \\
&= -\frac{N+1}{4}\ln 2\pi - \frac{N+1}{4}
\end{aligned} \tag{4.9}$$

Therefore, computing the normalized mutual information reduces to computing the first term in (4.8), which is the normalized expected logarithm of the normalization factor of the posterior (4.3).

It turns out that the first term in (4.8) has the form of a central quantity in statistical physics, called *Free Energy*. For physical systems analyzed in statistical mechanics, computing this quantity leads to studying the properties of that system. Therefore, computing the free energy has been a central focus in this field and various approaches have been proposed. For a more thorough

introduction of statistical physics and connections to inference problems, we refer the reader to [2].

Due to this close connection between estimation and statistical mechanics, several methods developed originally in statistical physics have been used to analyze estimation problems. These methods not only lead to optimal algorithms, but they also provide heuristic formulas for mutual information and MMSE that can be proved rigorously.

The *Replica method* stands out as a highly influential technique rooted in statistical physics [3, 4], particularly in its application to spiked models. In the case of spiked models with a factorized prior on the signal $P_{\boldsymbol{s}}(\boldsymbol{s}) = \prod_{i=1}^{N} P_s(s_i)$ (and similarly for $\boldsymbol{t}$ in the Wishart case), the Replica method was employed to derive the asymptotic mutual information and MMSE in a study by [5]. These expressions had been rigorously obtained earlier for binary signals using the Guerra-Toninelli interpolation method [6] in the work of [7]. Subsequently, [8] delved into the problem extensively for general signals, utilizing Approximate Message Passing (AMP) and spatial coupling. Additional contributions include [9, 10], which applied Guerra-Toninelli interpolation and Aizenman-Sims-Starr methods, and [11, 12], which introduced the adaptive interpolation method to rigorously establish the limiting expressions of mutual information and MMSE.

4.1.3 Algorithms

There are various algorithms proposed to reconstruct the hidden low-rank signal matrix. In this section, we discuss two main algorithms spectral algorithms and Approximate Message Passing (AMP).

Spectral estimators

Spectral estimators reconstruct the signal using eigenvectors (or singular vectors) of the observation matrix. These estimators are based on the celebrated results from random matrix theory. Specifically, when the SNR κ is sufficiently high, the top eigenvalue and its corresponding eigenvector (or the top singular value and singular vector) will be correlated with the hidden rank-one signal.

For the spiked Wigner model, using heuristic replica method it was noticed in [13] that there exists a critical value of the SNR κ, beyond which the largest eigenvalue of $\boldsymbol{Y}/\sqrt{N}$ escapes from the Wigner semi-circle bulk. This was later proved rigorously in [14, 15]. The same phenomenon for the Wishart model was proved in [16]. In [17, 18], it was shown that the correlation between the top eigenvector (or singular vector) and the signal, also undergoes a phase transition at the same threshold.

Consider the spiked Wigner model with prior such that $\mathbb{E}_{P_{\boldsymbol{s}}}\left[\|\boldsymbol{s}\|^2\right] = N$, and let λ_1 be the top eigenvalue of $\boldsymbol{Y}/\sqrt{N}$ in (4.1) and $\boldsymbol{y}_1$ be the eigenvector

associated to it. Then, almost surely (a.s.) we have

$$\lambda_1 \xrightarrow{N\to\infty} \begin{cases} 2 & \text{if } \kappa \le 1 \\ \sqrt{\kappa} + \frac{1}{\sqrt{\kappa}} & \text{if } \kappa \ge 1 \end{cases}, \frac{1}{N}(\boldsymbol{s}^\intercal \boldsymbol{y}_1)^2 \xrightarrow{N\to\infty} \begin{cases} 0 & \text{if } \kappa \le 1 \\ 1 - \frac{1}{\kappa} & \text{if } \kappa \ge 1 \end{cases} \tag{4.10}$$

An analogous statement for the spiked Wishart model is proven in [18]. Using these results one can estimate the hidden signal with $\boldsymbol{y}_1\boldsymbol{y}_1^\intercal$ (with possible rescaling). A similar phase transition of the eigenvalues/eigenvectors (or singular values/vectors) is proved for the general case of rotational invariant noise matrices in [17, 18]. Moreover, the non-symmetric low-rank matrix estimation under Gaussian noise in the regime with diverging aspect-ratio of matrices ($N/M \to \infty$ or $N/M \to 0$) was studied in [19].

While these findings offer a precise understanding of the performance of the top eigenvectors (or top singular vectors) in recovering low-rank signals, these naive spectral estimators do not take into account the prior distribution or constraints on the signal. Therefore, depending on the prior, these estimators are not necessarily optimal.

Approximate Message Passing (AMP)

AMP algorithms are iterative algorithms that try to maximize the posterior distribution or to achieve the Bayes optimal estimator $\mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}]$. These algorithms are "approximation" of *belief propagation* [20] for estimation problems characterized by densely connected graphical representations. AMP originally proposed for compressed sensing [21], has been applied to low-rank matrix estimation in [22–24]. It has also been extended to general rotational invariant noise matrix in [25, 26].

A key property of AMP is that, in the high-dimensional regime, its performance can be tracked by a set of equations called *State Evolution* (SE) [27]. Interestingly, for the majority of cases involving a factorized prior on the signal P_s, AMP achieves the MMSE (in the high-dimensional setting). However, AMP requires the full knowledge of the prior and channel characteristics.

4.1.4 Mismatched inference

The Bayesian estimation framework typically relies on the assumption that the statistician has complete knowledge of the priors for both the signal and the noise. However, this assumption often does not hold true in practice. In the mismatched inference, a common approach is to assume a conventional prior (such as a Gaussian distribution) and attempt to recover the signal through Bayesian methods by calculating the posterior mean. The first part of this thesis focuses on this issue for both symmetric and non-symmetric cases.

Mismatched estimation in the symmetric case has been explored with Gaussian noise assumptions in works like [28, 29], as well as in [30], which considers scenarios where the noise is incorrectly presumed to be Gaussian.

Subsequently, the problem has been rigorously analyzed for separable priors on the signal and noise by [31]. The non-symmetric case has been examined in further studies by [32, 33], expanding our understanding of the problem.

4.2 Extensive-Rank Matrix Inference

In this regime, the rank of the signal matrix $\boldsymbol{S}$ increases with the dimension, resulting in the following observation model:

$$\boldsymbol{Y} = \sqrt{\kappa}\boldsymbol{S} + \boldsymbol{Z}, \tag{4.11}$$

where $\boldsymbol{Y}$ is the observed data matrix, $\boldsymbol{S}$ is the true signal matrix, and $\boldsymbol{Z}$ represents the noise. We refer to this model as *denoising problem*.

For the symmetric case, $\boldsymbol{S}, \boldsymbol{Z}$ are symmetric $N \times N$ matrices, and in the non-symmetric case, the signal and noise matrices are rectangular with dimensions $N \times M$. The aspect ratio of these matrices converges to a constant $\alpha > 0$ as $N \to \infty$. Note that we assume that the entries of $\boldsymbol{S}, \boldsymbol{Z}$ are of the order $O(1/\sqrt{N})$, so that the eigenvalues (singular values) are of the order $O(1)$.

This model can be seen as a generalization of the models (4.1), (4.2) as follows: For the symmetric case, the spectral decomposition of $\boldsymbol{S}$ is given by:

$$\boldsymbol{S} = \sum_{i=1}^{R} \lambda_i \boldsymbol{s}_i \boldsymbol{s}_i^\intercal,$$

where λ_i are the eigenvalues and $\boldsymbol{s}_i \in \mathbb{R}^N$ are the corresponding eigenvectors of $\boldsymbol{S}$. For the non-symmetric case, the decomposition is:

$$\boldsymbol{S} = \sum_{i=1}^{R} \gamma_i \boldsymbol{s}_i^{(\mathrm{l})} \boldsymbol{s}_i^{(\mathrm{r})\intercal},$$

where γ_i are the singular values, $\boldsymbol{s}_i^{(\mathrm{l})} \in \mathbb{R}^N$ are the left singular vectors, and $\boldsymbol{s}_i^{(\mathrm{r})} \in \mathbb{R}^M$ are the right singular vectors.

In both scenarios, the rank R of the signal matrix $\boldsymbol{S}$ grows with the dimension N, capturing the extensive-rank nature of the signal.

Similar to the low-rank case, the posterior mean estimator $\mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}]$ is Bayes-optimal and attains the MMSE:

$$\mathrm{MMSE}_N(\kappa) := \frac{1}{N}\mathbb{E}\Big[\|\boldsymbol{S} - \mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}]\|_{\mathrm{F}}^2\Big]. \tag{4.12}$$

Nevertheless, it is important to note that evaluating the N^2-dimensional integral inherent in this MMSE calculation is considerably more complex than in the low-rank scenario and generally infeasible in practice.

Furthermore, due to the intricate nature of the problem, computing the asymptotic mutual information is also formidable since conventional techniques

such as the replica method may yield incorrect results. Consequently, researchers have focused on addressing extensive-rank problems within specific constrained classes of priors, or by imposing some internal structural on the signal matrix.

Matrix denoising problem (4.11) is intimately connected to the more involved problem of matrix factorization, where one must estimate two matrices $\boldsymbol{S} \in \mathbb{R}^{N\times K}$ and $\boldsymbol{T} \in \mathbb{R}^{M\times K}$ given the observation:

$$\boldsymbol{Y} = \sqrt{\kappa}\boldsymbol{S}\boldsymbol{T} + \boldsymbol{Z}, \tag{4.13}$$

Many problems in signal processing and learning can be formulated as matrix factorization, for example sparse coding [34, 35], blind source separation [36], robust principal component analysis [37], interpretations of patterns in images [38] or also in genomics data [39].

The extensive-rank matrix factorization problem, considering matrices $\boldsymbol{S}$ and $\boldsymbol{T}$ with i.id. entries, has been previously studied via the application of the replica method coupled with generic priors and output channels, as explored in [40]. Furthermore, algorithms leveraging AMP have been proposed to tackle the factorization challenge in a series of papers [40–42]. However, recent critical evaluation by [43] highlights that earlier derivations led to only approximate solutions, which fall short of exactness in the limit due to oversimplification in the calculations. Consequently, the proposed AMP algorithms are (theoretically) sub-optimal, despite their practical efficiency as evidenced in certain applications [44].

In the work by Maillard et al. [43], the factorization problem is revisited with an improved methodology, employing high-temperature expansions with fixed order parameters to refine prior analytical efforts [40]. This approach aims to address and correct the inaccuracies contained within earlier work. Separately, [45] have introduced an innovative method that integrates the replica approach with elements of random matrix theory—termed the spectral replica method—to address the same issue. Their work leads to the derivation of variational formulas for the mutual information. However, it is important to note that while these developments offer significant theoretical insights into the factorization problem, neither provides a direct explicit algorithm for the problem.

A rather simpler model is the symmetric factorization model, in which one observes a corrupted version of the matrix $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^\intercal$, with $\boldsymbol{X}$ having i.i.d. entries drawn from a prior P_x. The observed matrix $\boldsymbol{Y} = \sqrt{\kappa}\boldsymbol{X}\boldsymbol{X}^\top + \boldsymbol{Z}$ tasks researchers with two main challenges. The first is the denoising task, which seeks to reconstruct the product $\boldsymbol{X}\boldsymbol{X}^\intercal$. This challenge has been explored in [43, 45], where variational formulas for the mutual information are derived. However, these advancements do not yield a practical algorithm for denoising, with the exception of the Gaussian prior case discussed in [43]. The second and more complex challenge is the factorization task, aiming to recover the matrix $\boldsymbol{X}$. Notably, recovery in this context is achievable only up to a permutation of the

columns of $\boldsymbol{X}$. Despite various attempts, there is currently no analytical study providing insights into the mutual information or the MMSE for factorization. On the algorithmic front, a decimation scheme based on the AMP algorithm has been proposed in [46, 47], but this method lacks computational efficiency and its claim to optimality remains conjectural.

Recently, a new version of the (non-rigorous) replica method is used to derive the asymptotic mutual information for the symmetric factorization problem in the sub-linear growth regime in [48]. More precisely, the signal has the form $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^\intercal$ with $\boldsymbol{X} \in \mathbb{R}^{n\times M}$ for $M = N^{1-\epsilon}$, $\epsilon > 0$. Interestingly, the final result coincides with the asymptotic mutual information of the rank-one case. Moreover, a decimation algorithm inspired by [46] is proposed which is conjectured to be optimal.

Given the challenging nature of matrix denoising and factorization tasks when dealing with large matrices drawn from factorized priors, researchers have shifted their focus toward a class of priors termed as *rotationally invariant*. These priors often facilitate the resolution of high-rank matrix inference problems, as they lead to models that are more amenable to analytical and computational solutions. The second part of this thesis delves into matrix inference problems with a particular emphasis on rotationally invariant priors, exploring their theoretical framework and practical implications in the context of matrix denoising and factorization.

To proceed, let us introduce *rotationally invariant* priors. A symmetric matrix $\boldsymbol{S} \in \mathbb{R}^{N\times N}$ is said to be distributed according to a rotational invariant prior if for any orthogonal matrix $\boldsymbol{O} \in \mathbb{R}^{N\times N}$, the following holds:

$$P_A(\boldsymbol{A}) = P_A(\boldsymbol{O}\boldsymbol{A}\boldsymbol{O}^\intercal). \tag{4.14}$$

Likewise, a non-symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{N\times M}$ is said to be distributed according to a bi-rotational invariant prior if for any pair of orthogonal matrices $\boldsymbol{U} \in \mathbb{R}^{N\times N}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$, we have:

$$P_A(\boldsymbol{A}) = P_A(\boldsymbol{U}\boldsymbol{A}\boldsymbol{V}^\intercal). \tag{4.15}$$

Note that matrices with i.i.d. Gaussian entries are encompassed within this category for both symmetric and non-symmetric cases.

4.2.1 Rotationally Invariant Estimators (RIEs)

Rotationally Invariant Estimators (RIEs) refer to a class of estimators that are constructed from the observation matrix by modifying its eigenvalues (or singular values) while keeping its eigenvectors (or singular vectors) unchanged.

These estimators have been studied for covariance matrix estimation in [49–52]. They have been employed in matrix denoising, see [53] for the symmetric case, and [54, 55] for the non-symmetric case. Furthermore, their application in extensive-rank matrix factorization is explored [56], which constitues the content of the chapter 11 of this thesis.

To illustrate the mechanism of RIEs, let us explain their application in symmetric matrix denoising, as investigated by [53]. In the symmetric scenario where $\boldsymbol{S} = \boldsymbol{S}^\intercal$ and $\boldsymbol{Z} = \boldsymbol{Z}^\intercal$, consider the eigenvalue decomposition of $\boldsymbol{Y}$ to be $\boldsymbol{Y} = \sum_{k=1}^{N} \lambda_k^Y \boldsymbol{y}_k \boldsymbol{y}_k^\intercal$. Accordingly, a RIE for the signal matrix $\boldsymbol{S}$ is constructed as:

$$\boldsymbol{\Xi}_S(\boldsymbol{Y}) = \sum_{k=1}^{N} \xi_k \boldsymbol{y}_k \boldsymbol{y}_k^\intercal \tag{4.16}$$

Given this structure, for any orthogonal matrix $\boldsymbol{O} \in \mathbb{R}^{N\times N}$ we have:

$$\boldsymbol{\Xi}_S(\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal) = \boldsymbol{O}\boldsymbol{\Xi}_S(\boldsymbol{Y})\boldsymbol{O}^\intercal \tag{4.17}$$

justifying the term *rotational invariant*. Indeed, one can show that any estimator with property (4.17) has the same construction given in (4.16). Therefore, both equations (4.16), (4.17) can serve as the definition for the RIE.

The goal is to minimize the squared error between the true signal matrix $\boldsymbol{S}$ and the estimator, which leads to the optimal eigenvalues ξ_k^* being the solutions to the optimization problem:

$$\min_{\xi_1,\ldots,\xi_N} \left\| \boldsymbol{S} - \boldsymbol{\Xi}_S(\boldsymbol{Y}) \right\|_{\mathrm{F}}^2,$$

Expanding the error term using (4.16), we find:

$$\begin{aligned} \|\boldsymbol{S} - \boldsymbol{\Xi}_S(\boldsymbol{Y})\|_{\mathrm{F}}^2 &= \left\|\boldsymbol{S}\right\|_{\mathrm{F}}^2 + \left\|\boldsymbol{\Xi}_S(\boldsymbol{Y})\right\|_{\mathrm{F}}^2 - 2\,\mathrm{Tr}\,\boldsymbol{S}\boldsymbol{\Xi}_S(\boldsymbol{Y}) \\ &= \left\|\boldsymbol{S}\right\|_{\mathrm{F}}^2 + \sum_{k=1}^{N} \xi_k^2 - 2\sum_{k=1}^{N} \xi_k \boldsymbol{y}_k^\intercal \boldsymbol{S} \boldsymbol{y}_k \end{aligned}$$

which yields the following equation for the optimal eigenvalues:

$$\xi_k^* = \boldsymbol{y}_k^\intercal \boldsymbol{S} \boldsymbol{y}_k, \quad \text{for } 1 \le k \le N \tag{4.18}$$

Determining these optimal eigenvalues is challenging since they depend on knowledge of the signal matrix.

In the linear-growth regime, for a rotationally invariant distributed noise matrix $P_Z(\boldsymbol{Z}) = P_Z(\boldsymbol{O}\boldsymbol{Z}\boldsymbol{O}^\intercal)$, Bun et al. [53] leveraged the replica method to derive an explicit formula for calculating the optimal eigenvalues in the large dimension limit. Although the derivation is based on the non-rigorous replica method, it leads to an algorithm that performs well in practice. Remarkably, this explicit formula does not require any prior knowledge on the signal matrix $\boldsymbol{S}$, and involves transforms of spectral measures of the noise matrix and the observation.

Assuming that $\boldsymbol{S}$ is distributed according to a rotational invariant prior, we can see that the Bayes-optimal estimator $\mathbb{E}\big[\boldsymbol{S}|\boldsymbol{Y}\big]$ satisfies the property (4.17). The posterior mean estimator can be written as:

$$\mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}] = \frac{\int d\boldsymbol{X}\, P_S(\boldsymbol{X}) \boldsymbol{X} e^{-\frac{N}{4}\left\|\boldsymbol{Y}-\sqrt{\kappa}\boldsymbol{X}\right\|_{\mathrm{F}}^2}}{\int d\boldsymbol{X}\, P_S(\boldsymbol{X}) e^{-\frac{N}{4}\left\|\boldsymbol{Y}-\sqrt{\kappa}\boldsymbol{X}\right\|_{\mathrm{F}}^2}}.$$

By rotation invariance of $P_S(\boldsymbol{X})$ under any orthogonal transformation $\boldsymbol{X} \to \boldsymbol{O}\boldsymbol{X}\boldsymbol{O}^\intercal$ with Jacobian $|\det \boldsymbol{O}| = 1$ we have

$$\begin{aligned}
\mathbb{E}[\boldsymbol{S}|\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal] &= \frac{\int d\boldsymbol{X}\, P_S(\boldsymbol{X})\boldsymbol{X} e^{-\frac{N}{4}\left\|\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal - \sqrt{\kappa}\boldsymbol{X}\right\|_{\mathrm{F}}^2}}{\int d\boldsymbol{X}\, P_S(\boldsymbol{X}) e^{-\frac{N}{4}\left\|\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal - \sqrt{\kappa}\boldsymbol{X}\right\|_{\mathrm{F}}^2}} \\
&= \frac{\int d\boldsymbol{X}\, P_S(\boldsymbol{X})\boldsymbol{O}\boldsymbol{X}\boldsymbol{O}^\intercal e^{-\frac{N}{4}\left\|\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal - \sqrt{\kappa}\boldsymbol{O}\boldsymbol{X}\boldsymbol{O}^\intercal\right\|_{\mathrm{F}}^2}}{\int d\boldsymbol{X}\, P_S(\boldsymbol{X}) e^{-\frac{N}{4}\left\|\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal - \sqrt{\kappa}\boldsymbol{O}\boldsymbol{X}\boldsymbol{O}^\intercal\right\|_{\mathrm{F}}^2}} \\
&= \boldsymbol{O}\Big\{\frac{\int d\boldsymbol{X}\, P_S(\boldsymbol{X})\boldsymbol{X} e^{-\frac{N}{4}\left\|\boldsymbol{Y} - \sqrt{\kappa}\boldsymbol{X}\right\|_{\mathrm{F}}^2}}{\int d\boldsymbol{X}\, P_S(\boldsymbol{X}) e^{-\frac{N}{4}\left\|\boldsymbol{Y} - \sqrt{\kappa}\boldsymbol{X}\right\|_{\mathrm{F}}^2}}\Big\}\boldsymbol{O}^\intercal \\
&= \boldsymbol{O}\mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}]\boldsymbol{O}^\intercal.
\end{aligned}$$

Therefore, the posterior mean estimator is a RIE.

The RIE constructed with optimal eigenvalues (4.18) has the minimum MSE among the RIE class. Moreover, the posterior mean estimator achieves the MMSE and is inside the RIE class. Therefore, we conclude that for rotational invariant prior on the signal, the optimal RIE is Bayes-optimal. Using this fact, we can access the MMSE by computing the MSE of the optimal RIE.

4.3 Main Contributions and Organization

The thesis is divided into two main parts. The first part concerns mismatched low-rank estimation problem, and the second part studies extensive-rank matrix inference problems. The next two chapters introduce tools from random matrix theory and spherical integrals which constitute an integral part of the methods and results derived in the thesis.

Chapter 5 provides an in-depth yet not exhaustive exploration of random matrix theory and the various analytical methods used to study the asymptotic behavior of large random matrices. The analyses carried out in this chapter, under a broadly applicable model of random matrices, will be consistently referenced in subsequent sections. Moreover, we give a brief introduction to Voiculescu's free probability theory which was originally proposed to understand a special class of von Neumann algebras through the concept of *freeness* [57]. Two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are considered to be mutually free if their sets of eigenbases are related through a random rotation, meaning that the eigenvectors of $\boldsymbol{A}$ and $\boldsymbol{B}$ are effectively orthogonal with high probability. The discovery by Voiculescu [58] that random matrices could asymptotically exhibit freeness has had a profound impact on random matrix theory. Furthermore, we introduce the concepts of *Free entropy* and *Free Fisher information*, which are the analogues of Shannon entropy and Fisher information, respectively. The chapter concludes with a discussion of the replica method, which is used to compute the resolvent of a large class of random matrices, and leads to optimal RIEs.

In chapter 6, we introduce the spherical integrals and discuss their asymptotic limit. For two symmetric matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{N\times N}$, the *spherical integral* is defined as:

$$\mathcal{I}_N(\boldsymbol{A}, \boldsymbol{B}) = \Big\langle \exp\big\{\frac{N}{2} \operatorname{Tr} \boldsymbol{A}\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^\intercal\big\}\Big\rangle_{\boldsymbol{U}}$$

where the average is w.r.t. the *Haar* measure over the group of (real) orthogonal $N \times N$ matrices. The spherical integrals can also be defined w.r.t. the unitary or symplectic group. These integrals are often referred to as *Harish Chandra-Itzykson-Zuber* (HCIZ) integrals in mathematical physics literature. The study of these objects dates back to the work of mathematician Harish Chandra [59] and they have since been extensively studied and developed in both physics and mathematics [60–64]. We review the asymptotics of these integrals, both symmetric and rectangular, in low-rank and high-rank regimes. These theoretical results are applied to derive the limiting mutual information in inference models and to find explicit formulas for RIEs.

Part I - Low-rank mismatched inference

In **chapter 7**, we look at the mismatched low-rank matrix estimation in the symmetric case. We consider the spiked-Wigner model (4.1) where $\boldsymbol{s} \in \mathbb{R}^N$ has i.i.d. elements distributed according to a true prior P_s and the noise matrix has i.i.d. Gaussian entries. The statistician, unaware of the prior and the channel properties, assumes a Gaussian prior for the signal and an incorrect SNR, κ'. Using a Bayesian estimation approach, he/she attempts to construct the posterior mean estimator to recover the signal. We are interested in the asymptotic MSE of the estimation with mismatched parameters. To this end, we first prove a formula relating the free energy of the system to the mismatched MSE, as stated in lemma 7.1. This formula generalizes the I-MMSE relation (4.7) to the mismatched setting, and we refer to it as the *f-MSE relation*. We derive the asymptotic free energy of the mismatched estimation for two cases of true priors, Gaussian prior (with mismatched variance) in Theorem 7.4 and Bernoulli prior in Theorem 7.6. These derivations utilize results on high-dimensional limits of spherical integrals and the integrals are evaluated carefully using Laplace method. Subsequently, we compute the asymptotic mismatched MSE using the f-MSE relation, see Theorems 7.1, 7.3. Additionally, we explore the performance of the AMP and spectral estimators in this context.

In **chapter 8**, we extend the analysis from chapter 7 to the non-symmetric case. We adapt the spiked-Wishart model (4.2) with $\boldsymbol{s} \in \mathbb{R}^N$, $\boldsymbol{t} \in \mathbb{R}^M$ having i.i.d. elements from P_s, P_t. The statistician assumes Gaussian priors for both $\boldsymbol{t}$ and $\boldsymbol{s}$ and an incorrect SNR. With this framework, we show a similar f-MSE relation in the non-symmetric scenario, lemma 8.1. For the case where the true priors are Gaussian (but statistician assumes a mismatched variance), we derive the asymptotic free energy, see Statement 8.2, from which we can compute the asymptotic mismatched MSE.

Part II - Extensive-rank inference

In **chapter 9**, we investigate symmetric extensive-rank matrix denoising problem, focusing on the denoising model (4.11) with $\boldsymbol{S} = \boldsymbol{S}^\intercal \in \mathbb{R}^{N\times N}$ distributed according to a rotationally invariant prior. For the linear-rank growth with Gaussian noise matrix, we prove in Theorem 9.1 that the asymptotic mutual information is linked to the asymptotic log-spherical integrals. Moreover, using the Bayes-optimality of the RIE, we find the asymptotic MMSE in terms of the limiting spectral measure of the observation matrix in Theorem 9.2. Then, using I-MMSE relation and basic results from free probability, we integrate the MMSE and find an explicit expression for the asymptotic mutual information (Theorem 9.3). Given the relation between the asymptotic mutual information and the spherical integrals, the explicit formula for the asymptotic mutual information can serve as an explicit formula for the asymptotic log-spherical integral in a special case. An interesting issue is whether the known information theoretic phase transitions for rank-one, and also sub-linear-rank [48], still persist in linear-rank. Our analysis suggests that only a smoothed-out trace of the transitions persists. In the sub-linear growth regime, we propose an optimal RIE for the general rotation invariant noise matrices. We compute the asymptotic MSE of this sub-linear RIE and show that it matches the rigorously derived MMSE for the particular case of Gaussian noise in [65].

In **chapter 10**, we study the extensive-rank denoising problem with non-symmetric matrices. We consider the model (4.11) with $\boldsymbol{S}, \boldsymbol{Z} \in \mathbb{R}^{N\times M}$ and $\boldsymbol{Z}$ is a general bi-rotational invariant noise matrix. We propose *rectangular RIE* which extends RIEs to accommodate rectangular matrices. Utilizing the replica method, we derive an explicit RIE formula (see eq. (10.5)), which we conjecture to be optimal for general bi-rotationally invariant noise matrices. For the particular case of Gaussian noise, we prove a trace relation in Theorem 10.3, that strongly supports the proposed RIE's optimality. Analogous to the symmetric case, if the signal $\boldsymbol{S}$ is also bi-rotationally invariant distributed, the optimal RIE is Bayes-optimal and its asymptotic MSE equals the asymptotic MMSE. We derive the asymptotic MMSE under Gaussian noise in terms of the limiting singular value distribution of the observation in Statement 10.4. Furthermore, by independent methods we show that the asymptotic mutual information between signal and the observation is linked to the asymptotic log-spherical integral, see Theorem 10.5.

In **chapter 11**, we tackle the extensive-rank matrix factorization problem (4.13). We study this problem under the assumption that $\boldsymbol{S} \in \mathbb{R}^{N\times N}$ is a symmetric matrix from a rotational invariant ensemble and both $\boldsymbol{T}, \boldsymbol{Z} \in \mathbb{R}^{N\times M}$ are distributed according to bi-rotational invariant priors. The goal is to reconstruct both factors $\boldsymbol{S}, \boldsymbol{T}$ separately, given the observation $\boldsymbol{Y} = \sqrt{\kappa}\boldsymbol{S}\boldsymbol{T} + \boldsymbol{Z}$. Under the rotational invariant assumption of the priors, we show that the RIE is Bayes-optimal to recover the factors. Using the replica method, we derive analytical formulas for two RIEs which reconstruct $\boldsymbol{S}$ and $\boldsymbol{T}$ separately given the knowledge of the priors and the observation matrix.

This thesis is concluded by **chapter 12** where we summarize its findings and present possible research directions.

Random Matrix Theory: Overview and Analytical Tools

5

This chapter serves as a review of key principles and ideas from random matrix theory, which are crucial for the rest of the thesis. While it does not aim to be an exhaustive overview of the field, it presents a brief introduction to the core concepts and findings necessary for understanding the following chapters. For a comprehensive study of random matrix theory, we direct the reader to [66] for a mathematical perspective, and to [67] by Potters and Bouchaud for an approach from the physics standpoint.

The study of random matrices originated with Wishart's work [68] in 1928, focusing on the distribution of empirical covariance matrices. This early research eventually led to the development of the Marčenko-Pastur distribution [69]. In the 1950s, Wigner introduced random matrix theory as a statistical model for the energy levels in heavy nuclei [70], which contributed to the emergence of the Wigner semi-circle distribution. random matrix theory, initially rooted in physical and statistical contexts, has evolved into a dynamic field of study, yielding numerous significant findings in recent decades. This chapter will focus on the aspects of random matrices relevant to statistical inference, excluding some topics. While, the discussion will be focused on square, symmetric matrices, we also present the tools required to study non-symmetric matrices.

We begin this chapter with an introduction to probability transforms that will be used later on.

5.1 Large Random Matrices and their Spectral Distribution

In random matrix theory, it is common to consider matrices of infinite dimension for theoretical analysis. However, in practice, we deal with large matrices of

finite dimension N. This leads to an intriguing aspect, that as we approach the limit $N \to \infty$, we find that this theoretical model provides remarkably accurate approximations for the properties of large, but finite-dimensional matrices. Specifically, it is a well-established fact that the probability distributions reflecting the fluctuations of large-scale observables often approach certain stable laws as the size increases. Therefore, we anticipate that the statistical characteristics of a random matrix $\boldsymbol{A}$ with dimension N (like its eigenvalue distribution) will exhibit a predictable or self-averaging behavior as N increases. This tendency towards deterministic behavior in the infinite dimension limit becomes a crucial tool for characterizing the matrix, assuming it is sufficiently large. Consequently, our focus will be on the $N \to \infty$ limit in the following.

In this thesis the focus is on rotational invariant ensemble. As defined in previous chapter a symmetric real matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ is said to be rotationally invariant if the probability is invariant under an orthogonal transformation, i.e. $P_A(\boldsymbol{A}) = P_A(\boldsymbol{O}\boldsymbol{A}\boldsymbol{O}^\intercal)$ for any orthogonal matrix $\boldsymbol{O} \in \mathbb{R}^{N\times N}$. A common example of an invariant measure is of the form of a Boltzmann distribution:

$$P_A(\boldsymbol{A})\, d\boldsymbol{A} \propto e^{-\frac{N}{2}\operatorname{Tr} V(\boldsymbol{A})}\, d\boldsymbol{A} \tag{5.1}$$

where $V(.)$ is called *potential* function, and $d\boldsymbol{A} = \prod_{1\le i\le j\le N} dA_{ij}$ denotes Lebesgue measure. Changing variables from the entries of $\boldsymbol{A}$ to its eigenvalues $\lambda_1, \cdots, \lambda_N$ and eigenvectors we find:

$$P_A(\boldsymbol{A})\, d\boldsymbol{A} \propto e^{-\frac{N}{2}\sum_{i=1}^{N} V(\lambda_i)} \prod_{i<j}^{N} |\lambda_i - \lambda_j| \left(\prod_{i=1}^{N} d\lambda_i\right) D\boldsymbol{O} \tag{5.2}$$

where the Vandermonde determinant $\prod_{i<j}^{N} |\lambda_i - \lambda_j|$ appears from the Jacobian of the change of variables, and $D\boldsymbol{O}$ is the *Haar* (flat) measure over orthogonal group (orthogonal matrices of dimension N).

A critical aspect of studying large matrices involves analyzing their eigenvalues (or singular values) and corresponding eigenvectors, which have been shown to have significant practical importance, as discussed in the previous section. Random matrix theory has been instrumental in providing insights into the eigenvalues and eigenvectors of matrices. The distribution of the eigenvalues $\lambda_1, \cdots, \lambda_N$ can be characterized through the Empirical Spectral Distribution (ESD):

$$\rho_A^{(N)}(x) = \frac{1}{N}\sum_{i=1}^{N} \delta(x - \lambda_i) \tag{5.3}$$

where $\delta(.)$ is the Dirac delta function.

One of the most important property of large random matrices is that under sufficient conditions, the ESD converges almost surely towards a unique deterministic probability measure $\rho_A^{(N)}(x) \to \rho_A$ as $N \to \infty$. For example, for priors like in (5.1) if $\liminf_{|x|\to\infty} V(x)/(\beta \ln |x|) > 1$ for some $\beta > 1$, then the ESD converges weakly almost surely to a well-defined measure with compact support [66].

In the next part, we give an overview of different transforms that are employed to study the spectral properties of random matrices, and in particular deal with the spectral distribution of sums of random matrices.

5.1.1 Resolvent matrix and Stieltjes transform

Stieltjes transform

We first define the Stieltjes (or Cauchy) transform of a probability density function ρ on $\mathbb{R}$ as:

$$\mathcal{G}_\rho(z) = \int_{\mathbb{R}} \frac{1}{z-t}\rho(t)\,dt \quad \text{for } z \in \mathbb{C}\backslash\text{supp}(\rho) \tag{5.4}$$

The Stieltjes transform has many interesting properties. For example, $\mathcal{G}_\rho(.)$ is an analytic function on $\mathbb{C}^+$ when range is contained in $\mathbb{C}^-$. Moreover, if the pdf ρ does not contain Dirac masses, we have:

$$\lim_{\epsilon\to 0^+} \mathcal{G}_\rho(x - \mathsf{i}\epsilon) = \pi\mathsf{H}[\rho](x) + \pi\mathsf{i}\rho(x) \tag{5.5}$$

with $\mathsf{H}[\rho](x) = \text{p.v.}\frac{1}{\pi}\int_{\mathbb{R}} \frac{\rho(t)}{x-t}dt$ the *Hilbert* transform of ρ (here p.v. stands for "principal value"). Hence, the distribution ρ can be retrieved from its Stieltjes transform.

The Stieltjes transform can also be seen as the generating function of moments of ρ. If ρ has compact support contained in $[-R, R]$ for some $R > 0$, then for $|z| > R$ $\mathcal{G}_\rho$ has the following power series:

$$\mathcal{G}_\rho(z) = \frac{1}{z} + \sum_{i=1}^{\infty} \frac{m_i^\rho}{z^{i+1}} \tag{5.6}$$

where m_i^ρ is the i^{th} moment of ρ.

Furthermore, suppose that $\{\rho_N\}_N$ is a sequence of probability measures on $\mathbb{R}$ with $\mathcal{G}_N$ the Stieltjes transform of ρ_N. If $\{\mathcal{G}_N\}_N$ converges pointwise to $\mathcal{G}$ on $\mathbb{C}^+$ with $\lim_{y\to\infty} \mathsf{i}y\mathcal{G}(\mathsf{i}y) = 1$, then there is a unique probability measure ρ on $\mathbb{R}$ such that $\rho_N \to \rho$ weakly, and $\mathcal{G}$ is the Stietjes transform of ρ. Therefore, studying the limiting Stieltjes transform of ESD of random matrices can provide us with the limiting spectrum.

Resolvent matrix

For a symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$, the *resolvent* matrix is defined as:

$$\boldsymbol{G}_A(z) := \big(z\boldsymbol{I}_N - \boldsymbol{A}\big)^{-1} \tag{5.7}$$

with $z \in \mathbb{C}\backslash\{\lambda_i\}_{i=1}^N$. Consider the eigenvalue decomposition of $\boldsymbol{A}$ to be $\boldsymbol{A} = \sum_{i=1}^N \lambda_i \boldsymbol{a}_i\boldsymbol{a}_i^\intercal$ with $\boldsymbol{a}_i \in \mathbb{R}^N$ eigenvectors of $\boldsymbol{A}$, then the resolvent of $\boldsymbol{A}$ can

alternatively be written as:

$$\boldsymbol{G}_A(z) = \sum_{i=1}^{N} \frac{1}{z - \lambda_i} \boldsymbol{a}_i \boldsymbol{a}_i^\intercal \tag{5.8}$$

Therefore, the resolvent matrix contains the complete information about the eigenvalues and eigenvectors of $\boldsymbol{A}$. It is apparent that the number of singularities in the resolvent corresponds directly to the number of eigenvalues of $\boldsymbol{A}$. Moreover, for $z \to \lambda_i$ for $1 \leq i \leq N$, the residue at the pole defines a projection onto the eigenspace associated to the eigenvalue(s) λ_i. In later chapters, we will show how this property can be used to study the eigenvectors and derive optimal RIEs.

Now, focusing on the statistics of eigenvalues, we examine the normalized trace of the resolvent

$$\frac{1}{N} \operatorname{Tr} \boldsymbol{G}_A(z) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{z - \lambda_i} \tag{5.9}$$

which can be seen as the Stieltjes transform of the ESD defined in (5.3). Now consider that the following limit exists for all $z \in \mathbb{C}^+$:

$$\frac{1}{N} \operatorname{Tr} \boldsymbol{G}_A(z) \xrightarrow{N \to \infty} \mathcal{G}(z)$$

then, the limiting spectral distribution of $\boldsymbol{A}$ can be retrieved using the relation (5.5). This approach is one of the techniques in random matrix theory to study the limiting spectral measure of an ensemble, and we will follow this approach to recover the well-known semi-circle law and Marchenko-Pastur distribution.

For $z \to \infty$, the normalized trace can be expanded as:

$$\frac{1}{N} \operatorname{Tr} \boldsymbol{G}_A(z) \overset{z \to \infty}{=} \frac{1}{z} + \frac{1}{N} \sum_{i=1}^{N} \frac{\operatorname{Tr} \boldsymbol{A}^k}{z^{k+1}}$$

Therefore, if one can compute the moments of eigenvalue distribution in the limit $N \to \infty$, it is possible to reconstruct the density of eigenvalue distribution.

R-transform

To define the *R-transform*, we introduce the functional inverse of the Stieltjes transform, also known as the *Blue transform*:

$$\mathcal{G}_\rho^{-1}\big(\mathcal{G}_\rho(z)\big) = z$$

and the R-transform is defined as:

$$\mathcal{R}_\rho(z) = \mathcal{G}_\rho^{-1}(z) - \frac{1}{z} \tag{5.10}$$

Considering the power series of $\mathcal{R}_\rho$ one finds:

$$\mathcal{R}_\rho(z) = \sum_{i=1}^{\infty} \kappa_i^\rho z^{i-1} \tag{5.11}$$

where κ_i^ρ is the *free cumulant* of order i of ρ, which for the first four terms read:

$$\begin{aligned}
\kappa_1^\rho &= m_1^\rho \\
\kappa_2^\rho &= m_2^\rho - \left(m_1^\rho\right)^2 \\
\kappa_3^\rho &= m_3^\rho - 3m_2^\rho m_1^\rho + 2\left(m_1^\rho\right)^3 \\
\kappa_4^\rho &= m_4^\rho - 4m_3^\rho m_1^\rho - 2\left(m_2^\rho\right)^2 + 10m_2^\rho\left(m_1^\rho\right)^2 - 5\left(m_1^\rho\right)^4
\end{aligned}$$

where m_i^ρ is the i^{th} moment of ρ. Note that, the free cumulants differ from 'standard' cumulants for $i \geq 4$.

It turns out that, free cumulants of limiting spectral measures of sum of random matrices are given by the sum of free cumulants of limiting spectral measure of each matrix, i.e. $\kappa_i^{\rho_{A+B}} = \kappa_i^{\rho_A} + \kappa_i^{\rho_B}$, see section 5.3.2. Therefore, the R-transform is an important tool to study the spectral measure of sums of random matrices.

5.1.2 Rectangular random matrices

In the exploration of random matrix theory, the study of rectangular random matrices presents unique challenges. Unlike their symmetric counterparts, there is no direct or unified theoretical framework for understanding the behavior of rectangular matrices. This lack of a straightforward approach stems from the inherent complexity and variability in the structure of non-square matrices. However, by leveraging the existing analytical tools and methods developed for symmetric random matrices, we can effectively investigate the statistical properties of the singular values and singular vectors of rectangular random matrices. This approach allows us to extend our understanding and apply well-established principles from the study of symmetric matrices to the more complex realm of rectangular matrices.

Consider a non-symmetric random matrix $\boldsymbol{A} \in \mathbb{R}^{N\times M}$ with $N \leq M$ [1]. As in the symmetric case, we are interested in studying the statistical properties of this matrix in the limit $N \to \infty$. A common assumption is that N, M grows proportionally with ratio converging to a constant, ${}^{N}/_{M} \to \alpha \in (0, 1]$.

The ESD of $\boldsymbol{A}$ is defined in the same as in (5.3):

$$\mu_A^{(N)}(x) = \frac{1}{N}\sum_{i=1}^{N} \delta(x - \gamma_i)$$

[1] For the purpose of this chapter, it suffices to consider the case $N \leq M$. If $N > M$, we can consider $\boldsymbol{A}^\intercal$ can exchange the role of N with M.

where $\gamma_1, \cdots, \gamma_N \geq 0$ are singular values of $\boldsymbol{A}$. We can study the singular value distribution of $\boldsymbol{A}$ by analyzing the eigenvalue distribution of the symmetric matrix $\boldsymbol{A}\boldsymbol{A}^\intercal$. Let the SVD of $\boldsymbol{A}$ be:

$$\boldsymbol{A} = \boldsymbol{U}_A \boldsymbol{\Gamma} \boldsymbol{V}_A^\intercal, \qquad \boldsymbol{\Gamma} = \left[\begin{array}{c|c} \text{diag}(\gamma_1, \cdots, \gamma_{\text{N}}) & \boldsymbol{0}_{N\times(M-N)} \end{array} \right] \in \mathbb{R}^{N\times M}$$

with orthogonal matrices $\boldsymbol{U}_A \in \mathbb{R}^{N\times N}, \boldsymbol{V}_A \in \mathbb{R}^{M\times M}$. Then, the eigenvalue decomposition of $\boldsymbol{A}\boldsymbol{A}^\intercal$ reads:

$$\boldsymbol{A}\boldsymbol{A}^\intercal = \boldsymbol{U}_A \left[\text{diag}(\gamma_1^2, \cdots, \gamma_{\text{N}}^2) \right] \boldsymbol{U}_A^\intercal$$

So, there is a one-to-one map between the singular values of $\boldsymbol{A}$ and eigenvalues of $\boldsymbol{A}\boldsymbol{A}^\intercal$. The limiting eigenvalue distribution of $\boldsymbol{A}\boldsymbol{A}^\intercal$ can be studied using the tools developed for symmetric matrices. Suppose this distribution exists, and with abuse of notation we denote it as ρ_A. Then, the limiting spectral distribution of $\boldsymbol{A}$, denoted as μ_A can be retrieved by applying a (square-root) transformation to ρ_A. Using this approach, we derive the limiting singular values distribution of a non-symmetric Gaussian matrix by applying the square-root transformation to the Marchenko-Pastur distribution in section 5.2.2.

A major limitation of this approach is that we cannot study the statistics of singular vectors of $\boldsymbol{A}$. For example, by considering $\boldsymbol{A}\boldsymbol{A}^\intercal$ we do not have any information about the right singular vectors of $\boldsymbol{A}$, $\boldsymbol{V}_A$.

Another approach is to embed a non-symmetric matrix into a symmetric matrix. Construct the symmetric matrix $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{(N+M)\times(N+M)}$ from the matrix $\boldsymbol{A}$,

$$\boldsymbol{\mathcal{A}} = \begin{bmatrix} \boldsymbol{0}_{N\times N} & \boldsymbol{A} \\ \boldsymbol{A}^\intercal & \boldsymbol{0}_{M\times M} \end{bmatrix} \tag{5.12}$$

By Theorem 7.3.3 in [71], the eigenvalue decomposition of $\boldsymbol{\mathcal{A}}$ reads:

$$\boldsymbol{\mathcal{A}} = \boldsymbol{W} \begin{bmatrix} \text{diag}(\gamma_1, \cdots, \gamma_{\text{N}}) & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -\text{diag}(\gamma_1, \cdots, \gamma_{\text{N}}) & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0}_{(M-N)\times(M-N)} \end{bmatrix} \boldsymbol{W}^\intercal \tag{5.13}$$

with

$$\boldsymbol{W} = \begin{bmatrix} \hat{\boldsymbol{U}}_A & \hat{\boldsymbol{U}}_A & \boldsymbol{0}_{N\times(M-N)} \\ \hat{\boldsymbol{V}}_A^{(1)} & -\hat{\boldsymbol{V}}_A^{(1)} & \boldsymbol{V}_A^{(2)} \end{bmatrix} \in \mathbb{R}^{(N+M)\times(N+M)}$$

where $\boldsymbol{V}_A = \begin{bmatrix} \boldsymbol{V}_A^{(1)} & \boldsymbol{V}_A^{(2)} \end{bmatrix}$ in which $\boldsymbol{V}_A^{(1)} \in \mathbb{R}^{M\times N}$, and $\hat{\boldsymbol{V}}_A^{(1)} = \frac{1}{\sqrt{2}}\boldsymbol{V}_A^{(1)}$, $\hat{\boldsymbol{U}}_A = \frac{1}{\sqrt{2}}\boldsymbol{U}_A$.

Given the decomposition in (5.13), eigenvalues of $\boldsymbol{\mathcal{A}}$ are signed singular values of $\boldsymbol{A}$ plus $M-N$ trivial zero eigenvalues. The ESD of $\boldsymbol{\mathcal{A}}$ can be written as:

$$\rho_{\mathcal{A}}^{(N)}(x) = \frac{1}{N+M}\sum_{i=1}^{2N}\delta(x-\tilde{\gamma}_i) + \frac{M-N}{N+M}\delta(x)$$

where $\tilde{\gamma}_k$ are the non-trivial zero eigenvalues of $\boldsymbol{\mathcal{A}}$, $\tilde{\gamma}_1 = \gamma_1, \ldots, \tilde{\gamma}_N = \gamma_N, \tilde{\gamma}_{N+1} = -\gamma_1, \ldots, \tilde{\gamma}_{2N} = -\gamma_N$. We can go further to get:

$$\rho_{\mathcal{A}}^{(N)}(x) = \frac{N}{N+M}\mu_A^{(N)}(x) + \frac{N}{N+M}\mu_A^{(N)}(-x) + \frac{M-N}{N+M}\delta(x)$$

In the limit $N \to \infty$, we have:

$$\rho_{\mathcal{A}}(x) = \frac{2\alpha}{1+\alpha}\bar{\mu}_A(x) + \frac{1-\alpha}{1+\alpha}\delta(x)$$

where $\bar{\mu}_A(x) = \frac{1}{2}\mu_A(x) + \frac{1}{2}\mu_A(-x)$ is the symmetrization of μ_A. Therefore, the spectral distribution of $\boldsymbol{\mathcal{A}}$ excluding trivially zero eigenvalues is the symmetrization of spectral distribution of $\boldsymbol{A}$.

Moreover, we can analyze the singular vector by constructing the resolvent of $\boldsymbol{\mathcal{A}}$. Denote the eigenvectors of $\boldsymbol{\mathcal{A}}$ by $\boldsymbol{w}_i \in \mathbb{R}^{M+N}$, $i = 1, \ldots, M+N$.

$$\begin{aligned}\boldsymbol{G}_{\mathcal{A}}(z) &= \big(z\boldsymbol{I} - \boldsymbol{\mathcal{A}}\big)^{-1} \\ &= \sum_{i=1}^{2N} \frac{1}{z - \tilde{\gamma}_i}\boldsymbol{w}_i\boldsymbol{w}_i^\intercal + \frac{1}{z}\sum_{i=2N+1}^{M+N} \boldsymbol{w}_i\boldsymbol{w}_i^\intercal\end{aligned}$$

Note that by (5.13), $\boldsymbol{w}_i$'s contain the full information about the singular vectors of $\boldsymbol{A}$, hence the resolvent $\boldsymbol{G}_{\mathcal{A}}(z)$ provides us with the full information about both singular values and singular vectors of $\boldsymbol{A}$.

The Stieltjes transform is defined in the same way as (5.4) for μ_A and can be used to retrieve μ_A by the inversion formula (5.5). An important transform for non-symmetric random matrices is the *rectangular R-transform* which is used to study the spectrum of sums of non-symmetric matrices.

Rectangular R-transform

To define the *rectangular R-transform*, we first need to define some intermediate quantities. For a probability measure μ with support contained in $[-K, K]$ with $K > 0$, we define a generating function of (even) moments $\mathcal{M}_\mu : [0, K^{-2}] \to \mathbb{R}_+$ as

$$\mathcal{M}_\mu(z) := \int \frac{1}{1 - t^2 z}\mu(t)\, dt - 1 \tag{5.14}$$

For $\alpha \in [0, 1]$, define $T^{(\alpha)}(z) = (\alpha z + 1)(z + 1)$ and

$$\mathcal{H}_\mu^{(\alpha)}(z) := zT^{(\alpha)}\big(\mathcal{M}_\mu(z)\big) \tag{5.15}$$

The *rectangular R-transform with ratio* α is then defined as:

$$\mathcal{C}_\mu^{(\alpha)}(z) := {T^{(\alpha)}}^{-1}\Big(\frac{z}{{\mathcal{H}_\mu^{(\alpha)}}^{-1}(z)}\Big) \tag{5.16}$$

Similar to the symmetric case, $\mathcal{C}_\mu^{(\alpha)}$ can be written as power series with *rectangular free cumulants*. Therefore, it can be used to characterize the limiting spectrum of sum of two non-symmetric matrices.

5.1.3 Random matrix theory in practice

In this thesis, and often in practice, one needs to estimate the limiting spectral distribution of a given large (data) matrix. As explained in 5.1.1, one can approximate the Stieltjes transform of the spectral measure, and retrieve the distribution using the inversion formula (5.5). Note that we cannot simply use the normalized trace (5.9) to evaluate $\rho(.)$ at a given a eigenvalue (or signular value) because using $\frac{1}{N}\operatorname{Tr}\boldsymbol{G}_A(z)$ in the inversion formula (5.5) gives infinity for eigenvalues of $\boldsymbol{A}$. One idea is to apply kernel methods to estimate continuous density from discrete data.

Given a set of eigenvalues (or singular values) $\lambda_1, \cdots, \lambda_N$, a smooth estimator of the density is constructed by replacing the Dirac delta function in (5.3) by a proper normalized kernel of width η, K_η.

$$\rho^{(N)} = \frac{1}{N}\sum_{i=1}^{N} K_\eta(x - \lambda_i)$$

where

$$\int K_\eta(t)\, dt = 1$$

Here, we present the Cauchy kernels which we use through the thesis, for other choices of kernel we refer the reader to section 19.5 of [67]. Cauchy's kernel is defined as:

$$K_\eta^C(t) := \frac{1}{\pi}\frac{\eta}{t^2 + \eta^2}$$

and has the following Stieltjes transform:

$$\mathcal{G}_{K_\eta^C}(z) = \frac{1}{z \pm \mathrm{i}\eta}, \quad \pm = \operatorname{sign}\operatorname{Im} z$$

Then, the smoothed Stieltjes transform of $\rho^{(N)}$ is computed as:

$$\mathcal{G}_{\rho^{(N)}}(z) \approx \frac{1}{N}\sum_{i=1}^{N}\frac{1}{z - \lambda_i - \mathrm{i}\eta}$$

where η can be chosen properly at each point, however in the numerical simulations in this thesis we use this approximation with fixed $\eta = \sqrt{1/N}$.

5.2 Semi-circlular Law and Marchenko-Pastur Law

In this section, we investigate two well-known classes of random matrices: Wigner and Wishart matrices. Our focus will be on deriving their respective limiting spectral distributions, specifically the semi-circle law for Wigner matrices and the Marchenko-Pastur distribution for Wishart matrices.

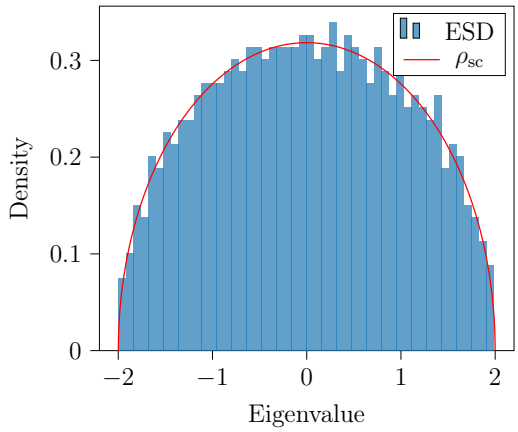


Figure 5.1: Comparison of the ESD of a Gaussian Wigner Matrix with the semi-circle Law. The histogram represents the normalized distribution of eigenvalues for a 1000×1000 Gaussian Wigner matrix. The red curve illustrates the semi-circle law, highlighting the convergence of the empirical distribution towards this theoretical prediction as the matrix size increases.

5.2.1 Wigner matrix

A *Gaussian Wigner matrix* $\boldsymbol{X} = \boldsymbol{X}^\intercal \in \mathbb{R}^{N \times N}$ is constructed with i.i.d. Gaussian entries as follows:

$$X_{ij} \sim \mathcal{N}\big(0, \frac{1}{N}(1 + \delta_{ij})\big) \quad \text{for } 1 \leq i \leq j \leq N$$

One can easily see that $\boldsymbol{X}$ is rotationally invariant distributed, and its probability measure can be described as in (5.1):

$$dP_X(\boldsymbol{X}) = C_N e^{-\frac{N}{4} \operatorname{Tr} \boldsymbol{X}^2} \prod_{i \leq j} dX_{ij} \tag{5.17}$$

where C_N is the normalizing constant. The ensemble described by (5.17) is often referred to as *Gaussian Orthogonal Ensemble (GOE)*.

We will prove that the limiting spectral distribution of GOE is the renowned *semi-circle* distribution which we define below. In figure 5.1, the ESD of a Gaussian Wigner is plotted against the theoretical semi-circle law.

Semi-circle distribution

The (standard) semi-circle distribution is the probability measure on $[-2, 2]$ with density:

$$d\rho_{sc}(t) = \frac{1}{2\pi}\sqrt{4 - t^2}\, dt \tag{5.18}$$

Since the semi-circle distribution is symmetric, the odd moments are zero. The even moments can be computed to be *Catalan numbers* $C_k = \frac{1}{k+1}\binom{2k}{k}$.

$$\frac{1}{2\pi}\int_{-2}^{2} t^n \sqrt{4 - t^2}\, dt = \begin{cases} 0, & n \text{ odd} \\ C_k, & n = 2k \text{ even} \end{cases} \tag{5.19}$$

Using the power series expansion of Stieltjes transform (5.6) and the recurrence relation of Catalan numbers, we find that $\mathcal{G}_{\rho_{\text{sc}}}(z)$ satisfies the following quadratic equation:

$$\mathcal{G}_{\rho_{\text{sc}}}(z)^2 - z\mathcal{G}_{\rho_{\text{sc}}}(z) + 1 = 0 \tag{5.20}$$

with two solutions:

$$\mathcal{G}_{\rho_{\text{sc}}}(z) = \frac{z \pm \sqrt{z^2 - 4}}{2}$$

To choose the correct sign, from (5.19) we see that for $z \to \infty$ the Stieltjes transform behaves as $1/z$. Choosing the minus sign in the solution above gives the correct asymptotic behavior. Therefore, we have:

$$\mathcal{G}_{\rho_{\text{sc}}}(z) = \frac{z - \sqrt{z^2 - 4}}{2} \tag{5.21}$$

Finally, we can compute the R-transform of the semi-circle distribution to be:

$$\mathcal{R}_{\rho_{\text{sc}}}(z) = z \tag{5.22}$$

Proof of convergence of ESD of GOE to semi-circle law

There are various methods to prove the convergence of ESD of GOE towards semi-circle distribution. One technique is to compute the moments of the matrix and show that they match the Catalan numbers. This technique involves combinatorial computations and is able to show the convergence holds for general prior on entries.

Here, we directly compute the limiting normalized trace of the resolvent of GOE and show that it satisfies (5.20). Then, using the inversion formula (5.5), we find that the limiting spectrum of GOE is the semi-circle distribution.

We will use the following two lemmas in our derivation.

Lemma 5.1. *Let $\boldsymbol{G}_X(z)$ be the resolvent matrix of a symmetric matrix $\boldsymbol{X} \in \mathbb{R}^{N\times N}$. We have:*

$$\boldsymbol{G}_X(z) = \frac{1}{z}\big(\boldsymbol{X}\boldsymbol{G}_X(z) + \boldsymbol{I}\big) \quad \text{for } z \in \mathbb{C}\backslash\mathbb{R} \tag{5.23}$$

Proof. By definition of the resolvent (5.7), we have:

$$\big(z\boldsymbol{I}_N - \boldsymbol{X}\big)\boldsymbol{G}_X(z) = \boldsymbol{I}_N$$

After expanding and rearranging the terms, we find the result. □

Lemma 5.2 (Stein's identity)**.** *Let $X_1, \cdots, X_k$ be independent Gaussian variables with $\mathbb{E}[X_i] = 0, \mathbb{V}\text{ar}(X_i) = \sigma_i^2$, and $f : \mathbb{R}^k \to \mathbb{R}$ be a continuously differentiable function. If f and its partial derivatives are of polynomial growth, then for all i we have:*

$$\mathbb{E}\big[f(X_1, \cdots, X_k)X_i\big] = \mathbb{E}\big[\frac{\partial f}{\partial X_i}(X_1, \cdots, X_k)\big] \tag{5.24}$$

Proof. In the univariate case ($k = 1$), we have:

$$\begin{aligned}
\mathbb{E}[f(X)X] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int x f(x) e^{-\frac{x^2}{2\sigma^2}}\, dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int f(x) \big[-\sigma^2 e^{-\frac{x^2}{2}} \big]'\, dx \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \int f'(x)\sigma^2 e^{-\frac{x^2}{2}}\, dx - \frac{1}{\sqrt{2\pi\sigma^2}} f(x)\sigma^2 e^{-\frac{x^2}{2}} \Big|_{-\infty}^{+\infty} \\
&= \sigma^2 \mathbb{E}[f'(X)]
\end{aligned}$$

For $k \geq 1$, we do the same steps for the i-th random variable. □

In what follows, we show that the expected normalized trace of the resolvent of Gaussian Wigner matrix $\boldsymbol{X}$ converges to the Stieltjes transform of the semi-circle distribution. By (5.23), we have:

$$\begin{aligned}
\frac{1}{N}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{G}_X(z) &= \frac{1}{N}\mathbb{E}\Big[\,\mathrm{Tr}\,\frac{1}{z}\big(\boldsymbol{X}\boldsymbol{G}_X(z) + \boldsymbol{I}\big)\Big] \\
&= \frac{1}{z} + \frac{1}{z}\frac{1}{N}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{X}\boldsymbol{G}_X(z)
\end{aligned} \tag{5.25}$$

Using Stein's identity, the second term (dropping the factor $1/z$) can be written as:

$$\begin{aligned}
\frac{1}{N}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{X}\boldsymbol{G}_X(z) &= \frac{1}{N}\mathbb{E}\sum_{i,j=1}^{N} X_{ij}\big[\boldsymbol{G}_X(z)\big]_{ji} \\
&= \frac{1}{N}\sum_{i,j=1}^{N} \mathbb{V}\mathrm{ar}(X_{ij})\mathbb{E}\Big[\frac{\partial}{\partial X_{ij}}\big[\boldsymbol{G}_X(z)\big]_{ji}\Big]
\end{aligned} \tag{5.26}$$

The partial derivatives are given by:

$$\big[\frac{\partial}{\partial X_{ii}}\boldsymbol{G}_X(z)\big]_{ii} = \big[\boldsymbol{G}_X(z)\big]_{ii}^2$$

$$\big[\frac{\partial}{\partial X_{ij}}\boldsymbol{G}_X(z)\big]_{ji} = \big[\boldsymbol{G}_X(z)\big]_{ij}^2 + \big[\boldsymbol{G}_X(z)\big]_{jj}\big[\boldsymbol{G}_X(z)\big]_{ii}$$

Plugging in (5.26), we find:

$$\begin{aligned}
&\frac{1}{N}\sum_{i,j=1}^{N}\mathbb{V}\mathrm{ar}(X_{ij})\mathbb{E}\Big[\frac{\partial}{\partial X_{ij}}\big[\boldsymbol{G}_X(z)\big]_{ji}\Big] \\
&= \frac{1}{N}\sum_{i=1}^{N}\mathbb{V}\mathrm{ar}(X_{ii})\mathbb{E}\Big[\frac{\partial}{\partial X_{ii}}\big[\boldsymbol{G}_X(z)\big]_{ii}\Big] + \frac{1}{N}\sum_{i\neq j}\mathbb{V}\mathrm{ar}(X_{ij})\mathbb{E}\Big[\frac{\partial}{\partial X_{ij}}\big[\boldsymbol{G}_X(z)\big]_{ji}\Big] \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{2}{N}\mathbb{E}\Big[\big[\boldsymbol{G}_X(z)\big]_{ii}^2\Big] + \frac{1}{N}\sum_{i\neq j}\frac{1}{N}\mathbb{E}\Big[\big[\boldsymbol{G}_X(z)\big]_{ij}^2 + \big[\boldsymbol{G}_X(z)\big]_{jj}\big[\boldsymbol{G}_X(z)\big]_{ii}\Big] \\
&= \frac{1}{N^2}\sum_{i,j=1}^{N}\mathbb{E}\Big[\big[\boldsymbol{G}_X(z)\big]_{ij}^2\Big] + \frac{1}{N^2}\sum_{i,j=1}^{N}\mathbb{E}\Big[\big[\boldsymbol{G}_X(z)\big]_{jj}\big[\boldsymbol{G}_X(z)\big]_{ii}\Big] \\
&= \frac{1}{N^2}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{G}_X(z)^2 + \frac{1}{N^2}\mathbb{E}\Big(\mathrm{Tr}\,\boldsymbol{G}_X(z)\Big)^2
\end{aligned} \tag{5.27}$$

We show that the first term in (5.27) goes to 0 as $N \to \infty$. Let $\lambda_1, \cdots, \lambda_N$ be the eigenvalues of $\boldsymbol{X}$. For $z \in \mathbb{C}\backslash\mathbb{R}$, we have:

$$(z-\lambda_i)^2 = \big(\mathrm{Re}\,z - \lambda_i\big)^2 + \big(\mathrm{Im}\,z\big)^2 \geq \big(\mathrm{Im}\,z\big)^2$$

Therefore, by Jensen inequality, we find:

$$\begin{aligned}
\Big|\frac{1}{N^2}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{G}_X(z)^2\Big| &\leq \frac{1}{N^2}\mathbb{E}\big|\,\mathrm{Tr}\,\boldsymbol{G}_X(z)^2\big| \\
&= \frac{1}{N^2}\mathbb{E}\Big[\sum_{i=1}^{N}\frac{1}{(z-\lambda_i)^2}\Big] \\
&\leq \frac{1}{N}\frac{1}{\big(\mathrm{Im}\,z\big)^2} \to 0 \,\mathrm{as}\, N\to\infty
\end{aligned} \tag{5.28}$$

The second term in (5.27), can be written as:

$$\begin{aligned}
\frac{1}{N^2}\mathbb{E}\Big(\mathrm{Tr}\,\boldsymbol{G}_X(z)\Big)^2 &= \mathbb{E}\Big[\Big(\frac{1}{N}\mathrm{Tr}\,\boldsymbol{G}_X(z)\Big)^2\Big] \\
&= \Big(\frac{1}{N}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{G}_X(z)\Big)^2 + \mathbb{V}\mathrm{ar}\Big(\frac{1}{N}\mathrm{Tr}\,\boldsymbol{G}_X(z)\Big)
\end{aligned} \tag{5.29}$$

Using the Gaussian Poincaré inequality , the variance of the normalized trace of the resolvent can be shown to converge to 0 as $N \to \infty$,

$$\mathbb{V}\mathrm{ar}\Big(\frac{1}{N}\mathrm{Tr}\,\boldsymbol{G}_X(z)\Big) \leq \frac{C_z}{N} \tag{5.30}$$

where C_z is a constant indendent of N. Putting (5.25),(5.26),(5.27), (5.28), (5.29),(5.30) together we find:

$$\frac{1}{N}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{G}_X(z) = \frac{1}{z} + \frac{1}{z}\Big(\frac{1}{N}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{G}_X(z)\Big)^2 + O\big(\frac{1}{N}\big) \tag{5.31}$$

Therefore, we conclude that $\frac{1}{N}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{G}_X(z)$ in the limit $N \to \infty$ satisfies the quadratic equation (5.20), and choosing the correct sign for the solution we obtain:

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{G}_X(z) = \mathcal{G}_{\rho_{\mathrm{sc}}}(z)$$

which implies that $\frac{1}{N}\,\mathrm{Tr}\,\boldsymbol{G}_X(z)$ converges in probability to $\mathcal{G}_{\rho_{\mathrm{sc}}}(z)$. Consequently, since $\frac{1}{N}\,\mathrm{Tr}\,\boldsymbol{G}_X(z)$ is the Stieltjes transform of ESD of $\boldsymbol{X}$, we deduce that the ESD converges weakly in probability to ρ_{sc}.

We conclude this section with a few remarks about Wigner matrices. Firstly, it is noteworthy that the convergence in probability discussed earlier can be extended to almost sure convergence, offering a stronger form of convergence. Additionally, the convergence of ESD to ρ_{sc} holds under a broader range of distributions for the entries of matrix $\boldsymbol{X}$, not just the Gaussian distribution. The primary condition for this applicability is the existence of all moments of the distribution.

5.2.2 Wishart matrix

Wishart matrices with Gaussian entries play a critical role in both statistical theory and random matrix theory. Such a matrix, typically denoted as $\boldsymbol{W} = \boldsymbol{X}\boldsymbol{X}^\intercal$, is constructed from the matrix $\boldsymbol{X} \in \mathbb{R}^{N\times M}$, whose entries are i.i.d. Gaussian random variables of variance $1/N$. In statistical contexts, the Wishart matrix is particularly significant when the columns of $\boldsymbol{X}$ represent independent samples from a multivariate normal distribution, making $\boldsymbol{W}$ a sample covariance matrix. The asymptotic behavior of these eigenvalues is characterized by the Marchenko-Pastur law, when $N \to \infty$ with fixed aspect ratio $N/M \to q \in \mathbb{R}_+$. This particular aspect makes Gaussian Wishart matrices indispensable in the analysis of high-dimensional data.

Marchenko-Pastur distribution

The Marchenko-Pastur distribution with aspect ratio $q > 0$ is the probability measure with density:

$$\rho_{\mathrm{MP}}(t) = \big(1 - \frac{1}{q}\big)^+\delta(t) + \frac{\sqrt{(t-\lambda_-)(\lambda_+ - t)}}{2\pi t} \tag{5.32}$$

where

$$\lambda_\pm = \big(\frac{1}{\sqrt{q}} \pm 1\big)^2$$

and

$$x^+ = \begin{cases} x & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

From the formula (5.32), one can see that when $q > 1$, $1 - 1/q$ proportion of eigenvalues are zero, which is expected since when $N > M$, $\boldsymbol{W}$ has $N - M$

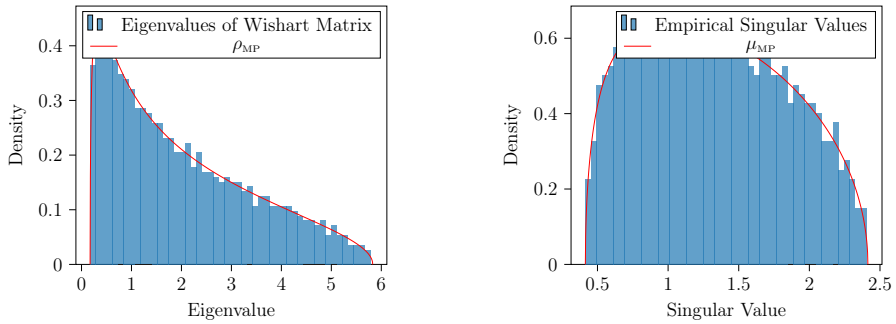


Figure 5.2: Comparison of empirical distributions for a Wishart Matrix. This figure illustrates two key distributions: the empirical singular value distribution and the empirical eigenvalue distribution of a Wishart matrix for $N = 1000, M = 2000$. Both distributions are compared with the theoretical Marchenko-Pastur law. The histogram on the left represents the distribution of singular values, showcasing their convergence towards the square root transformation of the Marchenko-Pastur law. On the right, the histogram depicts the eigenvalue distribution, aligning with the standard Marchenko-Pastur law.

trivial zero eigenvalues. In figure 5.2, the empirical eigenvalue distribution of a Gaussian Wishart matrix is plotted against the theoretical Marchenko-Pastur distribution.

The Stieltjes transform of ρ_{MP} can be shown to satsify the following quadratic equation:

$$z\mathcal{G}_{\rho_{\text{MP}}}(z)^2 - \left(z - \frac{1}{q} + 1\right)\mathcal{G}_{\rho_{\text{MP}}}(z) + 1 = 0 \tag{5.33}$$

from which we can find the (correct) solution to be:

$$\mathcal{G}_{\rho_{\text{MP}}}(z) = \frac{z - \frac{1}{q} + 1 - \sqrt{z - \lambda_+}\sqrt{z - \lambda_-}}{2z} \tag{5.34}$$

and, its R-transform is:

$$\mathcal{R}_{\rho_{\text{MP}}}(z) = \frac{1}{q}\frac{1}{1 - z} + \frac{1}{z} \tag{5.35}$$

The derivation of the Marchenko-Pastur law closely parallels that of the semi-circle law, utilizing similar methodologies such as moment calculation or the examination of the limiting normalized trace of the resolvent. Both approaches involve complex mathematical procedures, and we omit them here.

Singular value distribution

Given the eigenvalues distribution of $\boldsymbol{X}\boldsymbol{X}^\intercal$, we can find the distribution of singular values of $\boldsymbol{X}$ applying square-root transformation to ρ_{MP}. We assume that $N \leq M$ so there are N non-trivial zero singular values. With abuse of language and notation, we denote the resulting density function by μ_{MP} and refer to it as *Marchenko-Pastur* distribution.

$$\mu_{\text{MP}}(t) = \frac{\sqrt{\left(t^2 - \lambda_-\right)\left(\lambda_+ - t^2\right)}}{\pi t} \tag{5.36}$$

This distribution will be used in the analysis of denoising rectangular matrices which are corrupted by additive Gaussian noise, therefore, one important transform of this distribution is the rectangular R-transform which is given as :

$$\mathcal{C}^{(q)}_{\mu_{\mathrm{MP}}}(z) = \frac{1}{\alpha} z \quad \text{for } q \leq 1 \tag{5.37}$$

5.3 Free Probability

In inference problems, we frequently encounter matrices that are corrupted by noise, necessitating an analysis of their spectral properties. Common approaches, such as examining the Stieltjes transform, often pose significant challenges and may not always be applicable to these scenarios. This is particularly true when dealing with matrices that are sums or products of random matrices, where the complexity of interactions between matrix components complicates conventional methods. Free Probability theory emerges as a powerful framework in this context, offering robust tools for investigating the limiting spectral distribution of such composite matrices.

This section offers a concise introduction to free probability theory, a distinct approach for analyzing the asymptotic behavior of large-dimensional random matrices. Specifically, free probability presents a technique for examining the limiting spectral distribution of sums or products of symmetric random matrices. In this discussion, we focus on the basic notions of free probability as it applies to symmetric real random matrices. For a more comprehensive exploration of the topic, readers are directed to [72].

5.3.1 Freeness

Free probability, originated by Dan Voiculescu, has created a significant interplay between random matrix theory and operator algebra. Voiculescu's introduction of this theory, initially aimed to understand specific classes of von Neumann algebras [57], laid the foundation for a new calculus for non-commutative operators based on the concept of *freeness*. For simplicity, we do not go into details of algebra and work with matrices as non-commutative objects.

Consider a sequence of random symmetric matrices $\boldsymbol{A}_N, \boldsymbol{B}_N \in \mathbb{R}^{N\times N}$ with limiting spectral densities ρ_A, ρ_B, and define the linear functional

$$\varphi(\boldsymbol{A}) := \lim_{N\to\infty} \frac{1}{N} \operatorname{Tr} \boldsymbol{A}_N \tag{5.38}$$

which is the first moment of ρ_A.

Assume that $\varphi(\boldsymbol{A}) = \varphi(\boldsymbol{B}) = 0$, then $\boldsymbol{A}_N$ and $\boldsymbol{B}_N$ are called *(asymptotically) free* if:

$$\varphi(\boldsymbol{A}^{n_1}\boldsymbol{B}^{m_1}\cdots\boldsymbol{A}^{n_k}\boldsymbol{B}^{m_k}) = \varphi(\boldsymbol{A}^{n_1})\varphi(\boldsymbol{B}^{m_1})\cdots\varphi(\boldsymbol{A}^{n_k})\varphi(\boldsymbol{B}^{m_k}) \tag{5.39}$$

for any integers $n_1, \cdots, n_k$ and $m_1, \cdots, m_k$ with $k \in \mathbb{N}$. Note that if $\varphi(\boldsymbol{A}) \neq 0$ we can consider $\boldsymbol{A} - \varphi(\boldsymbol{A})\boldsymbol{I}$ (and similarly for $\boldsymbol{B}$).

For any pair of free matrices $\boldsymbol{A}_N, \boldsymbol{B}_N$, we have:

$$\varphi\big((\boldsymbol{A} - \varphi(\boldsymbol{A})\boldsymbol{I})(\boldsymbol{B} - \varphi(\boldsymbol{B})\boldsymbol{I})\big) = 0 \Rightarrow \varphi(\boldsymbol{AB}) = \varphi(\boldsymbol{A})\varphi(\boldsymbol{B})$$

Therefore, if we consider the trace operator (5.38) as the non-commutative equivalent of the expectation value for random variables, then the property of freeness (5.39) can be seen as the analogue to the moment factorization property. In general, freeness facilitates the calculation of mixed moments of matrix products using the known moments of $\boldsymbol{A}, \boldsymbol{B}$ (or ρ_A, ρ_B), analogous to how classical independence functions in classical probability theory. For example:

$$\varphi\big((\boldsymbol{A} - \varphi(\boldsymbol{A})\boldsymbol{I})(\boldsymbol{B} - \varphi(\boldsymbol{B})\boldsymbol{I})(\boldsymbol{A} - \varphi(\boldsymbol{A})\boldsymbol{I})\big) = 0 \Rightarrow \varphi(\boldsymbol{ABA}) = \varphi(\boldsymbol{A}^2)\varphi(\boldsymbol{B})$$

and by similar calculation one can find:

$$\varphi(\boldsymbol{ABAB}) = \varphi(\boldsymbol{A}^2)\varphi(\boldsymbol{B})^2 + \varphi(\boldsymbol{A})^2\varphi(\boldsymbol{B}^2) - \varphi(\boldsymbol{A})^2\varphi(\boldsymbol{B})^2$$

This theoretical framework has profound implications for the asymptotic behavior of large-dimensional random matrices. Voiculescu's seminal work [73] revealed that rotationally invariant random matrices asymptotically satisfy the criteria of freeness. This discovery has considerably impacted random matrix theory, especially in understanding the relationship between eigenbases of matrices. Under this framework, if $\boldsymbol{A}$ and $\boldsymbol{B}$ are independent self-adjoint matrices with spectral densities converging almost surely as N grows, and if $\boldsymbol{B}$ is rotationally invariant, then $\boldsymbol{A}$ and $\boldsymbol{B}$ are asymptotically free.

5.3.2 Free additive convolution

Besides computation of mixed moments of random matrices, free probability allows us to compute the limiting spectral distribution of sums and products of random matrices. Here, we only present the result for the additive case as it will be used throughout the thesis.

Consider a sequence of random matrices $\boldsymbol{M}_N = \boldsymbol{A}_N + \boldsymbol{B}_N$ that $\boldsymbol{B}_N$ is assumed to be rotational invariant, so we have:

$$\boldsymbol{M}_N = \boldsymbol{A}_N + \boldsymbol{O}\boldsymbol{B}_N\boldsymbol{O}^\intercal$$

for any orthogonal matrix $\boldsymbol{O} \in \mathbb{R}^{N\times N}$. Suppose that ESDs of $\boldsymbol{A}_N, \boldsymbol{B}_N$ converge to well-defined probability distributions ρ_A, ρ_B. As mentioned above $\boldsymbol{A}_N$ and $\boldsymbol{B}_N$ are asymptotically free, therefore we can use the law of addition for non-commutative operators to compute the limiting spectral distribution of $\boldsymbol{M}$. This law states:

$$\mathcal{R}_{\rho_M}(z) = \mathcal{R}_{\rho_A}(z) + \mathcal{R}_{\rho_B}(z) \tag{5.40}$$

The free additive convolution of two measures is denoted as $\rho_M = \rho_A \boxplus \rho_B$. In this way, the R-transform (5.10) can be viewed as the random matrix theory counterpart to the logarithm of the Fourier transform used in classic additive convolution. In the next section, we present a derivation of this relation using replica method.

The formula (5.40) provides us with a systematic tool to compute the limiting spectral distribution of a sum of two matrices. Using (5.10), we have:

$$\mathcal{G}^{-1}_{\rho_M}(z) = \mathcal{G}^{-1}_{\rho_A}(z) + \mathcal{R}_{\rho_B}(z)$$

Plugging $\mathcal{G}_{\rho_M}(z)$ as z, after rearranging we find:

$$\mathcal{G}^{-1}_{\rho_A}\big(\mathcal{G}_{\rho_M}(z)\big) = z - \mathcal{R}_{\rho_B}\big(\mathcal{G}_{\rho_M}(z)\big)$$

Applying $\mathcal{G}_{\rho_A}(z)$ on both sides, we find:

$$\mathcal{G}_{\rho_M}(z) = \mathcal{G}_{\rho_A}\Big[z - \mathcal{R}_{\rho_B}\big(\mathcal{G}_{\rho_M}(z)\big)\Big] \tag{5.41}$$

Therefore, we can compute the Stiletjes transform of ρ_M knowing ρ_A, ρ_B, from which ρ_M can be obtained using inversion formula (5.5).

Rectangular free additive convolution

In [74], the free convolution is generalized to non-symmetric matrices. Consider a sequence of $N \times M$ matrices built as:

$$\boldsymbol{M}_N = \boldsymbol{A}_N + \boldsymbol{U}\boldsymbol{B}_N\boldsymbol{V}^\intercal$$

for any orthogonal matrices $\boldsymbol{O} \in \mathbb{R}^{N\times N}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$. Suppose that ESDs (distribution of singular values) of $\boldsymbol{A}_N, \boldsymbol{B}_N$ converge to well-defined probability distributions μ_A, μ_B, and ${}^N\!/_M \to \alpha \in (0, 1]$. The law of rectangular-free convolution states that the limiting singular values distribution of $\boldsymbol{M}$ is characterized as:

$$\mathcal{C}^{(\alpha)}_{\mu_M}(z) = \mathcal{C}^{(\alpha)}_{\mu_A}(z) + \mathcal{C}^{(\alpha)}_{\mu_B}(z) \tag{5.42}$$

and this additive rectangular-free convolution is denoted as $\mu_M = \mu_A \boxplus_\alpha \mu_B$. In chapter 10, we derive this formula as a part of developing rectangular RIE, utilizing the replica method.

5.3.3 Free entropy and free Fisher information

In a series of papers [75–80], Voiculescu developed information theory for non-commutative random variables. He introduced two types of *free entropy*, see [76, 79, 81]. The first type, denoted as $\chi(X)$, quantifies the asymptotic volume of matricial microstate spaces, which is closely related to the classical entropy of random matrix models. The second type, $\chi^*(X)$, is defined through free Fisher information, which has to do with how the distribution of X interacts

with differentiation, akin to classical Fisher information. Heuristically, one expects that both types of free entropy are equal and coincide with large N limit of classical entropy of random matrix models. However, showing this equality involves several technical challenges in its proof, as discussed in [81].

Nonetheless, in the one variable case these two types of free entropy are proved to be equal. For X a self-adjoint *non-commutative* random variable associated to a probability measure ρ_X with compact support on the real line, the *free entropy* $\chi(X)$ and the *free Fisher information* $\Phi(X)$ are given as

$$\chi(X) = \iint \ln|s-t|\rho_X(s)\rho_X(t)\,ds\,dt + \frac{3}{4} + \frac{1}{2}\ln 2\pi, \tag{5.43}$$

$$\Phi(X) = \frac{4\pi^2}{3}\int \rho_X^3(s)\,ds. \tag{5.44}$$

Moreover, these two quantities are linked through the relation:

$$\chi(X) = \frac{1}{2}\int_0^\infty \left(\frac{1}{1+t} - \Phi(X+\sqrt{t}Z)\right)dt + \frac{1}{2}\ln 2\pi + \frac{1}{2} \tag{5.45}$$

where Z is non-commutative random variable associated to the semi-circle distribution ρ_{sc}, and X and Z are free (5.39).

In chapter 9, we leverage these relations to derive an explicit formula for the mutual information in symmetric matrix denoising. This underscores that, in the large N limit, matrix inference problems can effectively be modeled as scalar inference problems within the framework of free probability.

5.4 Replica Method

As mentioned before (see (5.8)), the analysis of eigenvectors can be approached through the study of the resolvent. However, methodologies like free probability primarily provide insights into the normalized trace of the resolvent, offering limited information about the structure of eigenvectors. To thoroughly examine the resolvent matrix, it is necessary to employ additional techniques, such as the Replica method, a technique adopted from statistical physics. In essence, the Replica method enables the transformation of the expectation value of a logarithm into a series of moments, computed as expectation values over multiple copies, or *replicas*, of the original system. This approach has proven highly effective in various areas, including random matrix theory and disordered systems, as evidenced in references [4], and more comprehensive reviews like [82]. It is important to note, though, that despite its effectiveness as a heuristic tool, the Replica method lacks formal mathematical rigor. Consequently, validating results derived from the Replica method with other methods, such as numerical simulations, is crucial.

As a preliminary case, we will outline the methodology for the Stieltjes transform and subsequently illustrate how this approach can be expanded to

analyze the full resolvent. We observe that the normalized trace of the resolvent of a random matrix $\boldsymbol{A}$ can be represented as follows:

$$\frac{1}{N}\operatorname{Tr}\boldsymbol{G}_A(z)=\frac{1}{N}\sum_{i=1}^{N}\frac{1}{z-\lambda_i}=\frac{1}{N}\frac{d}{dz}\ln\prod_{i=1}^{N}(z-\lambda_i)=\frac{1}{N}\frac{d}{dz}\ln\det\big(z\boldsymbol{I}-\boldsymbol{A}\big) \tag{5.46}$$

Then, using Gaussian representation of $\Big(\det\big(z\boldsymbol{I}-\boldsymbol{A}\big)\Big)^{-1/2}$, we find:

$$Z(z)\equiv\Big(\det\big(z\boldsymbol{I}-\boldsymbol{A}\big)\Big)^{-\frac{1}{2}}=\frac{1}{(2\pi)^{\frac{N}{2}}}\int\prod_{i=1}^{N}d\eta_i\,\exp\Big\{-\frac{1}{2}\boldsymbol{\eta}^{\intercal}\big(z\boldsymbol{I}-\boldsymbol{A}\big)\boldsymbol{\eta}\Big\} \tag{5.47}$$

Now, assuming that $\frac{1}{N}\operatorname{Tr}\boldsymbol{G}_A(z)$ is self-averaging (concentrates on its expectation), we can write:

$$\frac{1}{N}\operatorname{Tr}\boldsymbol{G}_A(z)=-2\frac{d}{dz}\mathbb{E}\ln Z(z) \tag{5.48}$$

where the expectation is over the prior on $\boldsymbol{A}$, P_A. However, computing the moments $\mathbb{E}Z(z)^n$ rather than $\mathbb{E}\ln Z(z)$ is typically more straightforward, and this is exactly what the Replica trick aims to facilitate. This method was originally established based on the following identity:

$$\ln Z=\lim_{n\to 0}\frac{Z^n-1}{n} \tag{5.49}$$

Hence, from (5.48) we formally obtain:

$$\frac{1}{N}\operatorname{Tr}\boldsymbol{G}_A(z)=\lim_{n\to 0}\frac{d}{dz}\frac{\mathbb{E}Z(z)^n-1}{n} \tag{5.50}$$

Thus, instead of (5.48) we need to compute n replicas of the system in (5.50). This computation involves working with integer values of n and then applying analytical continuation to extend the results to real n values, followed by taking the limit as $n\to\infty$. The limit $n\to 0$ is often taken after the limit $N\to\infty$, as in the large N limit the integrals can be evaluated using the saddle-point method. The critical assumption here is the feasibility of analytical continuation of n, which is not guaranteed, and might lead to uncontrolled approximations. Despite these concerns, the Replica method offers a straightforward heuristic for computing the resolvent matrix, which as shown below, is precise for the quantities considered in this thesis.

We end this section by deriving a resolvent relation for the additive model:

$$\boldsymbol{M}=\boldsymbol{A}+\boldsymbol{O}\boldsymbol{B}\boldsymbol{O}^{\intercal} \tag{5.51}$$

where $\boldsymbol{A},\boldsymbol{B}$ are two sequences of symmetric $N\times N$ matrices with limiting ESDs ρ_A,ρ_B, and $\boldsymbol{O}$ is a Haar-distributed matrix. For simplicity of notation we use $\boldsymbol{G}(z)\equiv\boldsymbol{G}_M(z)$ for the resolvent of a random matrix $\boldsymbol{M}$.

First, we express the entries of the resolvent $\boldsymbol{G}(z)$ using the Gaussian integral representation of an inverse matrix [83]:

$$\begin{aligned} G_{ij}(z) &= \sqrt{\frac{1}{(2\pi)^N \det\left(z\boldsymbol{I}-\boldsymbol{M}\right)}} \int \Big(\prod_{k=1}^{N} d\eta_k\Big)\, \eta_i\eta_j\, \exp\Big\{-\frac{1}{2}\boldsymbol{\eta}^{\intercal}\big(z\boldsymbol{I}-\boldsymbol{M}\big)\boldsymbol{\eta}\Big\} \\ &= \frac{\int \Big(\prod_{k=1}^{N} d\eta_k\Big)\, \eta_i\eta_j\, \exp\Big\{-\frac{1}{2}\boldsymbol{\eta}^{\intercal}\big(z\boldsymbol{I}-\boldsymbol{M}\big)\boldsymbol{\eta}\Big\}}{\int \Big(\prod_{k=1}^{N} d\eta_k\Big)\, \exp\Big\{-\frac{1}{2}\boldsymbol{\eta}^{\intercal}\big(z\boldsymbol{I}-\boldsymbol{M}\big)\boldsymbol{\eta}\Big\}} \end{aligned} \tag{5.52}$$

For z not close to the real axis, the resolvent is expected to exhibit self-averaging behavior in the limit of large N, meaning that it will not depend on the particular matrix realization. Thus, we can examine the resolvent $\boldsymbol{G}_M(z)$ by analyzing its ensemble average, denoted by $\langle . \rangle$ in the following.

$$\big\langle G_{ij}(z)\big\rangle = \Big\langle \frac{1}{\mathcal{Z}} \int \Big(\prod_{k=1}^{N} d\eta_k\Big)\, \eta_i\eta_j\, \exp\Big\{-\frac{1}{2}\boldsymbol{\eta}^{\intercal}\big(z\boldsymbol{I}-\boldsymbol{M}\big)\boldsymbol{\eta}\Big\}\Big\rangle_{\boldsymbol{M}} \tag{5.53}$$

where $\mathcal{Z}$ is the denominator in (5.52). Computing the average is, in general, non-trivial. However, the replica method provides us with a technique to overcome this issue by employing the following identity:

$$\begin{aligned} \big\langle G_{ij}(z)\big\rangle &= \lim_{n\to 0} \Big\langle \mathcal{Z}^{n-1} \int \Big(\prod_{k=1}^{N} d\eta_k\Big)\, \eta_i\eta_j\, \exp\Big\{-\frac{1}{2}\boldsymbol{\eta}^{\intercal}\big(z\boldsymbol{I}-\boldsymbol{M}\big)\boldsymbol{\eta}\Big\}\Big\rangle_{\boldsymbol{M}} \\ &= \lim_{n\to 0} \Big\langle \int \Big(\prod_{k=1}^{N}\prod_{\tau=1}^{n} d\eta_k^{(\tau)}\Big)\, \eta_i^{(1)}\eta_j^{(1)}\, \exp\Big\{-\frac{1}{2}\sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)\intercal}\big(z\boldsymbol{I}-\boldsymbol{M}\big)\boldsymbol{\eta}^{(\tau)}\Big\}\Big\rangle_{\boldsymbol{M}} \end{aligned} \tag{5.54}$$

So, the problem now is reduced to the computation of an average over n copies (or replicas) of the initial system (5.52). Note that, the identity (5.54) is valid for any random matrix $\boldsymbol{M}$ and is particularly useful when the average over the probability density P_M can be effectively computed. This equation enables us to explore the asymptotic behavior of the entries of the resolvent, thereby providing a more comprehensive insight into the spectral decomposition of $\boldsymbol{M}$ than what is offered by considering the normalized trace.

Using (5.51), we can see that P_M is the Haar measure over orthogonal

group, so (5.54) can be written as:

$$\begin{aligned}\big\langle G_{ij}(z)\big\rangle &= \lim_{n\to\infty} \\ &\Big\langle \int \Big(\prod_{k=1}^{N}\prod_{\tau=1}^{n} d\eta_k^{(\tau)}\Big)\, \eta_i^{(1)}\eta_j^{(1)}\, \exp\Big\{-\frac{1}{2}\sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)\intercal}\big(z\boldsymbol{I}-\boldsymbol{A}-\boldsymbol{O}\boldsymbol{B}\boldsymbol{O}^\intercal\big)\boldsymbol{\eta}^{(\tau)}\Big\}\Big\rangle_{\boldsymbol{O}} \\ &= \lim_{n\to\infty} \int \Big(\prod_{k=1}^{N}\prod_{\tau=1}^{n} d\eta_k^{(\tau)}\Big)\, \eta_i^{(1)}\eta_j^{(1)}\, \exp\Big\{-\frac{1}{2}\sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)\intercal}\big(z\boldsymbol{I}-\boldsymbol{A}\big)\boldsymbol{\eta}^{(\tau)}\Big\} \\ &\qquad\qquad \times \Big\langle \exp\Big\{\frac{1}{2}\sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)}\boldsymbol{\eta}^{(\tau)\intercal}\boldsymbol{O}\boldsymbol{B}\boldsymbol{O}^\intercal\big)\Big\}\Big\rangle_{\boldsymbol{O}}\end{aligned} \tag{5.55}$$

where the last term $\langle\cdots\rangle_{\boldsymbol{O}}$ is the low-rank (symmetric) spherical integral (explained in the next chapter) which in the large N limit is approximated by:

$$\Big\langle \exp\Big\{\frac{1}{2}\sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)}\boldsymbol{\eta}^{(\tau)\intercal}\boldsymbol{O}\boldsymbol{B}\boldsymbol{O}^\intercal\big)\Big\}\Big\rangle_{\boldsymbol{O}} \approx \exp\Big\{\frac{N}{2}\sum_{\tau=1}^{n} \mathcal{P}_{\rho_B}\big(\frac{1}{N}\|\boldsymbol{\eta}^{(\tau)}\|^2\big)\Big\} \tag{5.56}$$

with

$$\mathcal{P}_{\rho_B}(x) = \int_0^x \mathcal{R}_{\rho_B}(t)\, dt$$

the primitive of R-transform of ρ_B. Plugging (5.56) in (5.55), we find:

$$\begin{aligned}\big\langle G_{ij}(z)\big\rangle = \lim_{n\to\infty} &\int \Big(\prod_{k=1}^{N}\prod_{\tau=1}^{n} d\eta_k^{(\tau)}\Big)\, \eta_i^{(1)}\eta_j^{(1)} \\ &\times \exp\Big\{\sum_{\tau=1}^{n}\Big[-\frac{1}{2}\boldsymbol{\eta}^{(\tau)\intercal}\big(z\boldsymbol{I}-\boldsymbol{A}\big)\boldsymbol{\eta}^{(\tau)} + \frac{N}{2}\mathcal{P}_{\rho_B}\big(\frac{1}{N}\|\boldsymbol{\eta}^{(\tau)}\|^2\big)\Big]\Big\}\end{aligned} \tag{5.57}$$

In the next step, introducing delta functions $\delta\big(p^{(\tau)} - \frac{1}{N}\|\boldsymbol{\eta}^{(\tau)}\|^2\big)$, (5.57) can written as:

$$\begin{aligned}\big\langle G_{ij}(z)\big\rangle = \lim_{n\to\infty} &\int \Big(\prod_{k=1}^{N}\prod_{\tau=1}^{n} d\eta_k^{(\tau)}\Big)\Big(\prod_{\tau=1}^{n} dp^{(\tau)}\Big)\eta_i^{(1)}\eta_j^{(1)} \\ &\times \prod_{\tau=1}^{n}\delta\big(p^{(\tau)} - \frac{1}{N}\|\boldsymbol{\eta}^{(\tau)}\|^2\big) \\ &\times \exp\Big\{\sum_{\tau=1}^{n}\Big[-\frac{1}{2}\boldsymbol{\eta}^{(\tau)\intercal}\big(z\boldsymbol{I}-\boldsymbol{A}\big)\boldsymbol{\eta}^{(\tau)} + \frac{N}{2}\mathcal{P}_{\rho_B}\big(p^{(\tau)}\big)\Big]\Big\}\end{aligned} \tag{5.58}$$

Then, we replace each delta function with its Fourier transform $\delta\big(p^{(\tau)} - \frac{1}{N}\|\boldsymbol{\eta}^{(\tau)}\|^2\big) \propto \int d\zeta^{(\tau)} \exp\Big\{-\frac{N}{2}\zeta^{(\tau)}\big(p^{(\tau)} - \frac{1}{N}\|\boldsymbol{\eta}^{(\tau)}\|^2\big)\Big\}$. After rearranging, we

find:

$$\begin{aligned}\big\langle G_{ij}(z)\big\rangle \propto \int \Big(\prod_{\tau=1}^{n} dp^{(\tau)}\, d\zeta^{(\tau)}\Big) \exp\Big\{\frac{N}{2}\sum_{\tau=1}^{n}\Big[\mathcal{P}_{\rho_B}(p^{(\tau)}) - \zeta^{(\tau)}p^{(\tau)}\Big]\Big\} \\ \times \int \Big(\prod_{k=1}^{N}\prod_{\tau=1}^{n} d\eta_k^{(\tau)}\Big)\eta_i^{(1)}\eta_j^{(1)} \exp\Big\{-\frac{1}{2}\sum_{\tau=1}^{n}\Big[\boldsymbol{\eta}^{(\tau)\intercal}(z\boldsymbol{I}-\boldsymbol{A})\boldsymbol{\eta}^{(\tau)} \\ - \zeta^{(\tau)}\|\boldsymbol{\eta}^{(\tau)}\|^2\Big]\end{aligned} \tag{5.59}$$

The second integral in (5.59) is a Gaussian integral with matrix:

$$\boldsymbol{C}^{(\tau)} = (z-\zeta^{(\tau)})\boldsymbol{I} - \boldsymbol{A} \tag{5.60}$$

Let $\lambda_1, \cdots, \lambda_N$ be the eigenvalues of $\boldsymbol{A}$, then we have:

$$\det \boldsymbol{C}^{(\tau)} = \prod_{k=1}^{N}(z-\zeta^{(\tau)}-\lambda_k)$$

So, except for the first replica, the Gaussian integral is (up to constants):

$$\exp\Big\{-\frac{1}{2}\sum_{k=1}^{N}\ln(z-\zeta^{(\tau)}-\lambda_k)\Big\} \tag{5.61}$$

Noticing that ${\boldsymbol{C}^{(1)}}^{-1} = \boldsymbol{G}_A(z-\zeta^{(1)})$, the integral for the first replica is the above expression multiplied by $\big[\boldsymbol{G}_A(z-\zeta^{(1)})\big]_{ij}$. Putting these remarks together, the integral in (5.59) can be written as:

$$\big\langle G_{ij}(z)\big\rangle \propto \int \Big(\prod_{\tau=1}^{n} dp^{(\tau)}\, d\zeta^{(\tau)}\Big)\big[\boldsymbol{G}_A(z-\zeta^{(1)})\big]_{ij} \exp\Big\{-\frac{Nn}{2}F_0(\mathbf{p},\boldsymbol{\zeta})\Big\} \tag{5.62}$$

with

$$F_0(\mathbf{p},\boldsymbol{\zeta}) = \frac{1}{n}\sum_{\tau=1}^{n}\Big[\frac{1}{N}\sum_{k=1}^{N}\ln(z-\zeta^{(\tau)}-\lambda_k) - \mathcal{P}_{\rho_B}(p^{(\tau)}) + \zeta^{(\tau)}p^{(\tau)}\Big]$$

In the large N limit, the integral in (5.62) can be computed using the saddle-points of the function F_0. In the evaluation of this integral, we use the *replica symmetric* ansatz that assumes a saddle-point of the form:

$$\forall \tau \in \{1,\cdots,n\}: \quad p^{(\tau)} = p, \qquad \zeta^{(\tau)} = \zeta$$

One finds that the extremum of the function F_0 in the limit $N \to \infty$ is attained at:

$$p^* = \mathcal{G}_{\rho_A}(z-\zeta^*), \qquad \zeta^* = \mathcal{R}_{\rho_B}(p^*)$$

To simplify the solution, consider the normalized trace of both sides in (5.62) which gives

$$\mathcal{G}_{\rho_M}(z) = \mathcal{G}_{\rho_A}(z - \zeta^*) = p^*$$

consequently we find:

$$\zeta^* = \mathcal{R}_{\rho_B}\big(\mathcal{G}_{\rho_M}(z)\big) \tag{5.63}$$

Therefore, we have:

$$\mathcal{G}_{\rho_M}(z) = \mathcal{G}_{\rho_A}\Big[z - \mathcal{R}_{\rho_B}\big(\mathcal{G}_{\rho_M}(z)\big)\Big]$$

which coincides with (5.41).

Finally, plugging (5.63) in (5.62), we find:

$$\big\langle \boldsymbol{G}_M(z) \big\rangle = \boldsymbol{G}_A\Big(z - \mathcal{R}_{\rho_B}\big(\mathcal{G}_{\rho_M}(z)\big)\Big) \tag{5.64}$$

which generalizes (5.41) to matrix entries.

6 Spherical Integrals

In this section, we discuss *spherical integrals*, which can be seen as the analogue of Laplace/Fourier transforms in the context of random matrices. We will present results on asymptotic limits of various classes of spherical integrals .

6.1 Symmetric Spherical Integrals

For two symmetric matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{N\times N}$, the spherical integral is defined as:

$$\mathcal{I}_N(\boldsymbol{A}, \boldsymbol{B}) := \int D\boldsymbol{U} \, \exp\left\{\frac{N}{2} \operatorname{Tr} \boldsymbol{A}\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^\intercal\right\} \tag{6.1}$$

where $D\boldsymbol{U}$ is the Haar measure on orthogonal matrices of size N. These integrals can also be defined for Hermitian matrices, with integration over the Haar measure on the unitary group.

High-dimensional spherical integrals find extensive applications in statistical physics and random matrix theory. They have been explored in various scenarios, including their role in spin glass models [84, 85]. These integrals play a crucial role in tasks such as determining the density of eigenvalue distributions of random matrices [86, 87] and analyzing large deviations of eigenvalues [88–92].

For the case that the integral is over the unitary group, Harish-Chandra [59] initially derived explicit formulas for these integrals for any dimension, and this was further developed by Itzykson and Zuber [60]. For this reason, these integrals are often referred to as *Harish- Chandra-Itzykson-Zuber (HCIZ)* integrals. However, it is important to note that these formulas, which we omit here for brevity, are complex and involve determinants, making them less suitable for straightforward calculations of high-dimensional asymptotics.

We will investigate the high-dimensional limit of the integral (6.1) in two cases: one where one of the matrices has low rank, and another where the ranks of both matrices increase with the dimension N.

6.1.1 Low-rank symmetric spherical integrals

Rank-one case

The simplest case is the case in which one of the matrices is rank-one. This class of spherical integral will be used in chapter 7. The asymptotic limit of this class will be used in deriving the resolvent relation (as shown in previous chapter) which is important in deriving the optimal RIE.

Let $\boldsymbol{B}$ be a rank-one matrix with non-zero eigenvalue θ. From the definition (6.1), one may notice that the integral only depends on the eigenvalues of $\boldsymbol{A}$,$\boldsymbol{B}$, and its high-dimensional limit involves the limiting ESD of $\boldsymbol{A}$. So, with abuse of notation, we will denote this integral as:

$$\mathcal{I}_N(\theta, \boldsymbol{A}) := \int D\boldsymbol{U} \, \exp\left\{\frac{N}{2} \operatorname{Tr} \boldsymbol{A}\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^\intercal\right\} \tag{6.2}$$

Theorem 6.1 (Guionnet and Maida [93])**.** *Suppose, as $N \to \infty$, the ESD of $\boldsymbol{A}$ converges weakly a.s. to ρ_A and the minimum and maximum eigenvalues of $\boldsymbol{A}$ converge a.s. to the finite values $\lambda_{\min}$, $\lambda_{\max}$, respectively. Let $\mathcal{G}_{\min} = \lim_{z\to\lambda_{\min}} \mathcal{G}_{\rho_A}(z)$, $\mathcal{G}_{\max} = \lim_{z\to\lambda_{\max}} \mathcal{G}_{\rho_A}(z)$. Then,*

$$\mathcal{J}(\theta, \rho_A) := \lim_{N\to\infty} \frac{1}{N} \ln \mathcal{I}_N(\theta, \boldsymbol{A}) = \frac{1}{2}\theta\nu(\theta) - \frac{1}{2}\int \ln(1 + \theta\nu(\theta) - \theta t)\rho_A(t)\, dt \tag{6.3}$$

where

$$\nu(\theta) = \begin{cases} R_{\rho_A}(\theta) & \text{if } \mathcal{G}_{\min} \le \theta \le \mathcal{G}_{\max} \\ \lambda_{\min} - \frac{1}{\theta} & \text{if } \theta > \mathcal{G}_{\max} \\ \lambda_{\max} - \frac{1}{\theta} & \text{if } \theta < \mathcal{G}_{\min} \end{cases}$$

Note that, for small θ namely $\mathcal{G}_{\min} \le \theta \le \mathcal{G}_{\max}$, one can check, by comparing the derivatives, that:

$$\mathcal{J}(\theta, \rho_A) = \frac{1}{2}\int_0^\theta \mathcal{R}_{\rho_A}(t)\, dt \equiv \frac{1}{2}\mathcal{P}_{\rho_A}(\theta) \tag{6.4}$$

Finite-rank

The asymptotic limit given in Theorem 6.1, can be extended to the case where $\boldsymbol{B}$ has rank k which is finite as $N \to \infty$. Let $\theta_1, \cdots, \theta_k$ be the eigenvalues of $\boldsymbol{B}$, then it is shown in [65, 93, 94] that:

$$\lim_{N\to\infty} \frac{1}{N} \ln \mathcal{I}_N(\boldsymbol{A}, \boldsymbol{B}) = \sum_{i=1}^{k} \mathcal{J}(\theta_i, \rho_A) \tag{6.5}$$

In the replica computations, we often use this asymptotic formula where the rank is the number of replicas which we assume are finite (because eventually we take the limit $n \to \infty$). However, we only consider the case where all θ_i's are small enough to use the formula (6.4) (see for example (5.56). Although, this might lead to uncontrolled approximations, but numerical simulations show that results of a replica calculation are exact.

6.1.2 Extensive-rank symmetric spherical integrals

Sub-linear rank

In this scenario, the matrix $\mathbf{B}$ has a rank of $k(N)$ such that ${}^{k(N)}/_{N} \to 0$ as $N \to \infty$. Interestingly, it has been established that the limit of the integral in this case converges to a sum of $k(N)$ decoupled rank-one integrals, as in the finite-rank case (6.5). This result was initially demonstrated in [93] for situations where $k(N) = o(N^{-1/2-\epsilon})$ (for any positive ϵ) and the values of θ_i's below the transition threshold. Subsequently, this finding was extended in [94] to any $k(N) = o(N)$. Notably, Husson and Ko [65] recently demonstrated that the assumption regarding the convergence of maximum (and minimum) eigenvalues of $\mathbf{A}$ in Theorem 6.1 can be relaxed, and the asymptotic limit holds for any $k(N) = o(N)$.

Under conditions of Theorem 6.1, and assuming that the empirical distribution of non-zero eigenvalues of $\boldsymbol{B}$ converges to ρ_B, $\frac{1}{k}\sum_{i=1}^{k}\delta(x-\theta_i) \to \rho_B$ as $N \to \infty$, we have:

$$\lim_{N\to\infty} \frac{1}{Nk(N)} \ln \mathcal{I}_N(\boldsymbol{A}, \boldsymbol{B}) = \int \mathcal{J}(t, \rho_A)\rho_B(t)\, dt \tag{6.6}$$

Authors in [65], used this result to derive the asymptotic MMSE and mutual information in denoising symmetric matrices of sub-linear rank, which we will also use it in chapter 9.

Linear rank

The case where both matrices have rank which grows linearly with N is more involved and cannot be reduced to the rank-one case. It was first demonstrated by Matytsin [95], that using Dyson's Brownian motion, one can find:

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{N^2} \ln \mathcal{I}_N(\boldsymbol{A}, \boldsymbol{B}) = \frac{1}{2}\Bigg[-\frac{3}{4} - S(\rho_A, \rho_B) + \frac{1}{2}\int dt\, t^2\big(\rho_A(t) + \rho_B(t)\big) \\ - \frac{1}{2}\iint dt\, ds\, \ln|s-t|\big(\rho_A(t)\rho_A(s) + \rho_B(t)\rho_B(s)\big)\Bigg]\end{aligned} \tag{6.7}$$

where

$$S(\rho_A, \rho_B) = \frac{1}{2}\int dt \int d\lambda\, \rho(\lambda, t)\left[\nu^2(\lambda, t) + \frac{\pi^2}{3}\rho^2(\lambda, t)\right]$$

with $\rho(\lambda,t), \nu(\lambda,t)$ satisfying the Euler equations:

$$\begin{cases} \partial_t \rho(\lambda,t) + \partial_\lambda \big[\rho(\lambda,t)\nu(\lambda,t)\big] = 0, \\ \partial_t \nu(\lambda,t) + \nu(\lambda,t)\,\partial_\lambda \nu(\lambda,t) = \frac{\pi^2}{3}\partial_\lambda \rho^2(\lambda,t) \\ \rho(\lambda,0) = \rho_A(\lambda), \quad \nu(\lambda,1) = \rho_B(\lambda) \end{cases} \tag{6.8}$$

Guionnet and Zeitouni [61] subsequently provided a rigorous proof for the asymptotic limit of these integrals. They formulated the limit in terms of a variational problem over probability measures, which we do not present here. Instead, we only outline the conditions necessary for the existence of this limit. It is worth noting that, as discussed in [61], the resulting limit aligns with the findings of Matytsin [95] when the integral is over the unitary group.

Theorem 6.2 (Guionnet and Zeitouni [61])**.** *Assume that the support of the ESD of $\boldsymbol{A}$ is contained in a compact subset of $\mathbb{R}$, and the second moment of ESD of $\boldsymbol{B}$ is uniformly bounded. Moreover, suppose that the ESDs of $\boldsymbol{A}, \boldsymbol{B}$ converge weakly towards ρ_A, ρ_B. Then the following limit exists*

$$\mathcal{J}(\rho_A, \rho_B) := \lim_{N\to\infty} \frac{1}{N^2} \ln \mathcal{I}_N(\boldsymbol{A}, \boldsymbol{B}), \tag{6.9}$$

and $\mathcal{J}(\rho_A, \rho_B)$ is the solution of a variational problem over a space of probability measures.

In chapter 9, we use this result to prove the existence of the asymptotic mutual information for symmetric denoising problem. Moreover, we derive an explicit expression for the particular case where $\rho_B = \rho_A \boxplus \rho_{\text{sc}}$.

$$\mathcal{J}(\rho_A, \rho_B) = \frac{1}{2}\int t^2 \rho_A(t)\,dt - \frac{1}{2}\iint \ln|s-t|\,\rho_B(s)\rho_B(t)\,ds\,dt - \frac{1}{8}$$

This formula was previously derived in [43] by solving the Euler equations (6.8) for this particular case.

6.2 Rectangular Spherical Integrals

For the matrices $\boldsymbol{A} \in \mathbb{R}^{M\times N}, \boldsymbol{B} \in \mathbb{R}^{N\times M}$, the *rectangular* spherical integral is defined as:

$$\mathcal{I}_{N,M}(\boldsymbol{A}, \boldsymbol{B}) := \int D\boldsymbol{U}\, D\boldsymbol{V}\, \exp\left\{\sqrt{NM}\,\operatorname{Tr} \boldsymbol{A}\boldsymbol{U}\boldsymbol{B}\boldsymbol{V}\right\} \tag{6.10}$$

where $D\boldsymbol{U}, D\boldsymbol{V}$ are Haar measures on orthogonal matrices of size N and M respectively. These integrals can also be defined for Hermitian matrices, with integration over the Haar measure on the unitary group.

Rectangular spherical integrals are the counterparts to symmetric ones (6.1), emerging from the study of rectangular matrices. They are expected to

have similar applications, including being useful for understanding the large deviations of singular values of rectangular matrices. However, as the study of these integrals is quite recent, the full extent of their applications remains to be fully explored.

Like the symmetric case, we will investigate the high-dimensional limit of the integral (6.10) in two cases: one where one of the matrices has low rank, and another where the rank of both matrices increases with the dimension N.

6.2.1 Low-rank rectangular spherical integrals

Rank-one case

This class of spherical integrala will be used in chapter 8, and to derive the resolvent relation (as shown in previous chapter) which is important in deriving the optimal RIE.

Let $\boldsymbol{B}$ be a rank-one matrix with non-zero singular value θ. From the definition (6.10), one may notice that the integral only depends on the singular values of $\boldsymbol{A}$,$\boldsymbol{B}$, and the its high-dimensional limit involves the limiting ESD of $\boldsymbol{A}$. With abuse of notation, we will denote this integral as:

$$\mathcal{I}_{N,M}(\theta, \boldsymbol{A}) := \int D\boldsymbol{U}\, D\boldsymbol{V}\, \exp\left\{\sqrt{NM}\, \mathrm{Tr}\, \boldsymbol{A}\boldsymbol{U}\boldsymbol{B}\boldsymbol{V}\right\} \tag{6.11}$$

Recall the transforms defined in (5.14), (5.15), (5.16). We state the asymptotic limit of $\mathcal{I}_{N,M}(\theta, \boldsymbol{A})$ below:

Statement 6.3. *Suppose that ESD of $\boldsymbol{A}$ converges weakly towards μ_A and the top singular value of $\boldsymbol{A}$ converges to $\gamma_{\max}$ as $N \to \infty$. Moreover, assume $N/M \to \alpha \in (0, 1]$. Let $\mathcal{H}_{\max} = \lim_{z \to \frac{1}{\gamma_{\max}^2}} \mathcal{H}_{\mu_A}^{(\alpha)}(z)$. Then,*

$$\begin{aligned}\mathcal{J}^{(\alpha)}(\theta, \mu_A) &:= \lim_{N\to\infty} \frac{1}{N} \ln \mathcal{I}_{N,M}(\theta, \boldsymbol{A}) \\ &= \nu - \frac{1}{2\alpha}\ln(1+\alpha\nu) - \frac{1}{2}\ln(1+\nu) - \frac{1}{2}\int \ln\big(1 - \frac{\theta^2}{T^{(\alpha)}(\nu)}t^2\big)\mu_A(t)\, dt\end{aligned}$$

with

$$\nu = \begin{cases} \mathcal{C}_{\mu_A}^{(\alpha)}(\theta^2) & \text{if } \theta^2 \le \mathcal{H}_{\max} \\ {T^{(\alpha)}}^{-1}(\theta^2\gamma_{\max}^2) & \text{if } \theta^2 > \mathcal{H}_{\max} \end{cases}$$

The expression in Statement 6.3 is derived in appendix 6.A by solving the non-rigorous (but conjectured to be exact) variational problem proposed in [96] for the asymptotic rectangular spherical integral. In [64], Statement 6.3 is rigorously proved for a smaller interval of θ, namely $\theta^2 < \frac{1}{\gamma_{\max}^2} \le H_{\max}$.

Similar to the symmetric case, for small $\theta \le \mathcal{G}_{\max}$, one can check, by comparing the derivatives, that:

$$\mathcal{J}^{(\alpha)}(\theta, \mu_A) = \int_0^\theta \frac{\mathcal{C}_{\mu_B}^{(\alpha)}(t^2)}{t}\, dt = \frac{1}{2}\int_0^{\theta^2} \frac{\mathcal{C}_{\mu_B}^{(\alpha)}(t)}{t}\, dt \equiv \frac{1}{2}\mathcal{Q}_{\mu_B}^{(\alpha)}(\theta^2) \tag{6.12}$$

In the replica computations for deriving optimal RIEs, we encounter these types of integral in which the matrix $\boldsymbol{B}$ has higher but fixed rank (≥ 1). Similar to the symmetric case, see section 6.1.1, we use the formula which is the sum over singular values of the expression on the rhs of (6.3). For $\boldsymbol{B}$ with singular values $\theta_1, \cdots, \theta_k$:

$$\lim_{N\to\infty} \frac{1}{N} \ln \mathcal{I}_{N,M}(\boldsymbol{A}, \boldsymbol{B}) = \sum_{i=1}^{k} \mathcal{J}^{(\alpha)}(\theta_i, \rho_A)$$

Although we are not aware if this generalization has been proved, we believe that the ideas found in [63] can be applied to show it holds. Furthermore, we use the formula for the case where all θ_i's are small enough to use the formula (6.12).

6.2.2 Extensive-rank rectangular spherical integrals

As mentioned before, compared to their symmetric counterparts, high-dimensional limits of rectangular spherical integrals have been less explored. However, in cases where the rank of both matrices grows with the dimension, the asymptotics of these integrals have recently been investigated, as detailed in [97].

They formulated the limit in terms of a variational problem involving probability measures, which, we do not present here. Instead, we only outline the conditions necessary for the existence of this limit which will be used derivation of asmptotic mutual information in chapter 10.

Theorem 6.4 (Guionnet and Jiaoyang [97])**.** *Suppose that the ESDs of the matrices $\boldsymbol{A}, \boldsymbol{B}$ converge weakly towards μ_A, μ_B, respectively. Moreover, assume the second moment of the ESD of $\boldsymbol{A}$ is uniformly bounded and the non-commutative entropy of the symmetrization of μ_A is finite, $\iint \ln|x-y|\, d\bar{\mu}_A(x)\, d\bar{\mu}_A(y) > -\infty$, and $\int \ln|x|\, d\bar{\mu}_A(x) > -\infty$. Then the following limit exists*

$$\mathcal{J}^{(\alpha)}(\mu_A, \mu_B) := \lim_{N\to\infty} \frac{1}{NM} \ln \mathcal{I}_{N,M}(\boldsymbol{A}, \boldsymbol{B}), \tag{6.13}$$

and $\mathcal{J}^{(\alpha)}(\mu_A, \mu_B)$ is given in terms of variational problem.

In [98], an explicit expression of this limit is derived for the case where $\boldsymbol{B} = \boldsymbol{A} + \boldsymbol{Z}$ where $\boldsymbol{Z}$ has i.i.d. Gaussian entries, $\mu_B = \mu_A \boxplus_\alpha \mu_{\text{MP}}$.

Appendix

6.A Derivation of Asymptotic Rank-One Rectangular Spherical Integral

Lemma 6.1. *(Extension of lemma 4.2 in [64]) Fix $\alpha \in [0,1]$, $\theta \in [0, \sqrt{H_{\max}})$, and define $\gamma = \mathcal{C}_\mu^{(\alpha)}(\theta^2)$. Then*

$$\mathcal{M}_\mu\big(\frac{\theta^2}{T^{(\alpha)}(\gamma)}\big) = \gamma \tag{6.14}$$

Proof. By definition, $T^{(\alpha)}(\gamma) = \frac{\theta^2}{\mathcal{H}_\mu^{(\alpha)^{-1}}(\theta^2)}$, so $\mathcal{H}_\mu^{(\alpha)^{-1}}(\theta^2) = \frac{\theta^2}{T^{(\alpha)}(\gamma)}$. Applying $\mathcal{H}_\mu^{(\alpha)}$ on the both sides, we get $\theta^2 = \mathcal{H}_\mu^{(\alpha)}\big(\frac{\theta^2}{T^{(\alpha)}(\gamma)}\big)$,

$$\frac{\theta^2}{T^{(\alpha)}(\gamma)} T^{(\alpha)}\Big(\mathcal{M}_\mu\big(\frac{\theta^2}{T^{(\alpha)}(\gamma)}\big)\Big) = \theta^2$$

It follows that $T^{(\alpha)}\Big(\mathcal{M}_\mu\big(\frac{\theta^2}{T^{(\alpha)}(\gamma)}\big)\Big) = T^{(\alpha)}(\gamma)$. Since, both $\mathcal{M}_\mu\big(\frac{\theta^2}{T^{(\alpha)}(\gamma)}\big)$ and γ are non-negative real numbers, one gets (6.14) □

Remark 6.1. *If $\mu(t) \neq \delta(t)$, then $\mathcal{M}_\mu$ is an increasing function, and $\gamma = \mathcal{C}_\mu^{(\alpha)}(\theta^2)$ is the unique solution of* (6.14).

6.A.1 Proof of Statement 6.3

We start with the following theorem:

Theorem 6.5. *(Kabashima [99]) Assume the ESD of $\boldsymbol{A}^\intercal \boldsymbol{A} \in \mathbb{R}^{M \times M}$ converges weakly towards ρ_A. Also, assume that $N, M \to \infty$ with $N/M \to \alpha$. Assume moreover that the top eigenvector of $\boldsymbol{A}^\intercal \boldsymbol{A}$ converges to $\gamma_{\max}$. Let $\theta \geq 0$ be the*

only non-zero singular value of $\boldsymbol{B}$*. Then,*

$$\lim_{M\to\infty}\frac{1}{M}\ln\mathcal{I}_{N,M}(\boldsymbol{A},\boldsymbol{B}) = \frac{1}{2}\inf_{\substack{x,y>0\\ xy\geq\theta^2\gamma_{\max}}}\left[\alpha x+y-(\alpha-1)\ln x-\int d\rho_A(\lambda)\ln(xy-\theta^2\lambda)\right]-\frac{1+\alpha}{2} \tag{6.15}$$

Our goal is to solve the variational problem in (6.15). We are interested in the limit of $\frac{1}{N}\ln\mathcal{I}_{N,M}$, which is

$$\frac{1}{2}\inf_{\substack{x,y>0\\ xy\geq\theta^2\gamma_{\max}}}\left[x+\frac{1}{\alpha}y-(1-\frac{1}{\alpha})\ln x-\frac{1}{\alpha}\int d\rho_A(\lambda)\ln(xy-\theta^2\lambda)\right]-\frac{1+\frac{1}{\alpha}}{2} \tag{6.16}$$

We also assume that, $\alpha\in[0,1]$, which implies $N\leq M$. By the assumption in Theorem 6.5, the empirical spectral law of $\boldsymbol{A}^\intercal\boldsymbol{A}$ converges to ρ_A.

$$\frac{1}{M}\sum_{i=1}^{m}\delta_{\lambda_i}\to\rho_A,\qquad \lambda_i \text{ are eigenvalues of } \boldsymbol{A}^\intercal\boldsymbol{A}$$

Since $M\geq N$, $\boldsymbol{A}^\intercal\boldsymbol{A}$ has $M-N$ zero eigenvalues. So,

$$\frac{1}{M}\big((M-N)\delta_0+\sum_{i=1}^{N}\delta_{\lambda_i'}\big)\to\rho_A,\qquad \lambda_i' \text{ are eigenvalues of } \boldsymbol{A}\boldsymbol{A}^\intercal$$

In the limit $N\to\infty$, we can write$(1-\alpha)\delta_0+\alpha\rho_A^*=\rho_A$ where ρ_A^* is the limiting spectral law of $\boldsymbol{A}\boldsymbol{A}^\intercal$.

Replacing ρ_A with $(1-\alpha)\delta_0+\alpha\rho_A^*$ in (6.16), we find

$$\begin{aligned}
&\frac{1}{2}\inf_{\substack{x,y>0\\ xy\geq\theta^2\gamma_{\max}}}\Big\{x+\frac{1}{\alpha}y-(1-\frac{1}{\alpha})\ln x-\frac{1}{\alpha}\Big[(1-\alpha)\ln x+(1-\alpha)\ln y\\
&\qquad\qquad+\alpha\int d\rho_A^*(\lambda)\ln(xy-\theta^2\lambda)\Big]\Big\}-\frac{1+\frac{1}{\alpha}}{2}\\
&=\frac{1}{2}\inf_{\substack{x,y>0\\ xy\geq\theta^2\gamma_{\max}}}\Big\{x+\frac{1}{\alpha}y-(1-\frac{1}{\alpha})\ln x-\frac{1}{\alpha}\Big[\ln x+\ln y\\
&\qquad\qquad+\alpha\int d\rho_A^*(\lambda)\ln\big(1-\frac{\theta^2}{xy}\lambda\big)\Big]\Big\}-\frac{1+\frac{1}{\alpha}}{2}\\
&=\frac{1}{2}\inf_{\substack{x,y>0\\ xy\geq\theta^2\gamma_{\max}}}\left[x+\frac{1}{\alpha}y-\ln x-\frac{1}{\alpha}\ln y-\int d\rho_A^*(\lambda)\ln\big(1-\frac{\theta^2}{xy}\lambda\big)\right]-\frac{1+\frac{1}{\alpha}}{2}
\end{aligned} \tag{6.17}$$

Since ρ_A^* is the limiting spectral law of $\boldsymbol{A}\boldsymbol{A}^\intercal$, we find that $\frac{1}{N}\sum_{i=1}^N \sigma_i^2 \to \rho_A^*$, where σ_i's are the singular values of $\boldsymbol{A}$. So, ρ_A^* is the pushforward measure (with function $f(x) = x^2$) of μ_A, which is the limiting ESD of $\boldsymbol{A}$. Assuming the the function is integrable, we can write (6.17) as

$$\frac{1}{2}\inf_{\substack{x,y>0\\ xy\geq\theta^2\sigma_{\max}^2}}\left[x+\frac{1}{\alpha}y-\ln x-\frac{1}{\alpha}\ln y-\int d\mu(t)\ln\left(1-\frac{\theta^2}{xy}t^2\right)\right]-\frac{1+\frac{1}{\alpha}}{2} \quad (6.18)$$

Define the function $f(x,y)$ as the objective function in (6.18). First, we show that $f(x,y)$ is convex.

Let $g(x,y) = x - \ln x + \frac{1}{\alpha}(y - \ln y)$. It can easily be checked that this function is convex in the whole plane, in particular in the convex set $x > 0, y > 0, xy \geq \theta^2\sigma_{\max}^2$.

Define $h(x,y) = -\int d\mu_A(t)\ln(1-\frac{\theta^2}{xy}t^2)$ on the convex set $x > 0, y > 0, xy \geq \theta^2\sigma_{\max}^2$. The Hessian of $h(x,y)$ reads

$$H_h = \begin{bmatrix} \frac{1}{x^2}\mathcal{M}_{\mu_A}\left(\frac{\theta^2}{xy}\right)+\frac{\theta^2}{x^3y}\mathcal{M}'_{\mu_A}\left(\frac{\theta^2}{xy}\right) & \frac{\theta^2}{x^2y^2}\mathcal{M}'_{\mu_A}\left(\frac{\theta^2}{xy}\right) \\ \frac{\theta^2}{x^2y^2}\mathcal{M}'_{\mu_A}\left(\frac{\theta^2}{xy}\right) & \frac{1}{y^2}\mathcal{M}_{\mu_A}\left(\frac{\theta^2}{xy}\right)+\frac{\theta^2}{xy^3}\mathcal{M}'_{\mu_A}\left(\frac{\theta^2}{xy}\right) \end{bmatrix}$$

$$\det H_h = \frac{1}{x^2y^2}\mathcal{M}_{\mu_A}^2\left(\frac{\theta^2}{xy}\right)+2\frac{\theta^2}{x^3y^3}\mathcal{M}_{\mu_A}\left(\frac{\theta^2}{xy}\right)\mathcal{M}'_{\mu_A}\left(\frac{\theta^2}{xy}\right)$$

First note that, $\mathcal{M}_{\mu_A}$ is non-negative (in particular positive if $\theta > 0$), and non-decreasing. Thus, $\det H_h$ is non-negative. Moreover, all entries of H are non-negative. Therefore, H_h is positive semi-definite, which implies $h(x,y)$ is convex. Finally, we conclude that $f(x,y) = g(x,y) + h(x,y)$ is a convex function.

To find the global minimum of $f(x,y)$, we put the derivative of $f(x,y)$ to zero.

$$\nabla f(x,y) = \begin{bmatrix} 1-\frac{1}{x}-\frac{1}{x}\mathcal{M}_{\mu_A}\left(\frac{\theta^2}{xy}\right) \\ \frac{1}{\alpha}-\frac{1}{\alpha}\frac{1}{y}-\frac{1}{y}\mathcal{M}_{\mu_A}\left(\frac{\theta^2}{xy}\right) \end{bmatrix} \equiv 0 \Rightarrow \begin{cases} \mathcal{M}_{\mu_A}\left(\frac{\theta^2}{xy}\right) = x-1 \\ \mathcal{M}_{\mu_A}\left(\frac{\theta^2}{xy}\right) = \frac{1}{\alpha}(y-1) \end{cases}$$

Denoting $x - 1 = \frac{1}{\alpha}(y-1) = \gamma$, the above equations can be written as

$$\mathcal{M}_{\mu_A}\left(\frac{\theta^2}{T^{(\alpha)}(\gamma)}\right) = \gamma$$

By lemma 6.1, if $\theta^2 < H_{\max}$, then $\gamma = \mathcal{C}_\mu^{(\alpha)}(\theta^2)$ is the unique solution of equation above. On the other hand, if $\theta^2 \geq H_{\max}$ then there is no solution to the equation above (in the proof of lemma 6.1, we use the assumption that $\theta^2 < H_{\max}$ to show that $\mathcal{H}_\mu^{(\alpha)^{-1}}(\theta^2)$ is defined. However, if $\theta^2 \geq H_{\max}$ this fails which implies that there is no solution to the equation). So, in this case by convexity of $f(x,y)$, the minimum is attained on the boundaries of domain of the function. Therefore, the minimum is attained on the curve $xy = \theta^2\sigma_{\max}^2$.

If $\theta^2 < H_{\max}$, then

$$x^* = 1 + \mathcal{C}_{\mu_A}^{(\alpha)}(\theta^2), \quad y^* = 1 + \alpha\mathcal{C}_{\mu}^{(\alpha)}(\theta^2)$$

Denoting $\mathcal{C}_{\mu_A}^{(\alpha)}(\theta^2)$ by γ, we can write (6.18) as

$$\begin{aligned}\frac{1}{2}\Bigg[1+\gamma+&\frac{1}{\alpha}(1+\alpha\gamma) - \ln(1+\gamma) - \frac{1}{\alpha}\ln(1+\alpha\gamma) \\ &- \int d\mu_A(t)\ln\big(1-\frac{\theta^2}{T^{(\alpha)}(\gamma)}t^2\big)\Bigg] - \frac{1+\frac{1}{\alpha}}{2} \\ &= \gamma - \frac{1}{2\alpha}\ln(1+\alpha\gamma) - \frac{1}{2}\ln(1+\gamma) - \frac{1}{2}\int d\mu_A(t)\ln\big(1-\frac{\theta^2}{T^{(\alpha)}(\gamma)}t^2\big)\end{aligned} \tag{6.19}$$

which coincides with the eq. (29) in [64], and is equal to $\int_0^\theta \frac{\mathcal{C}_{\mu}^{(\alpha)}(t^2)}{t}dt$.

If $\theta^2 \geq H_{\max}$, the minimizer lies on the the curve $xy = \theta^2\sigma_{\max}^2 \equiv A$. Plugging $y = \frac{A}{x}$ in (6.18) we find

$$\frac{1}{2}\inf_{x>0}\left[x + \frac{1}{\alpha}\frac{A}{x} - \ln x - \frac{1}{\alpha}\ln\frac{A}{x} - \int \mu(dt)\ln\big(1-\frac{\theta^2}{A}t^2\big)\right] - \frac{1+\frac{1}{\alpha}}{2}$$

Computing the derivative w.r.t. x, we get

$$\begin{aligned}1 - \frac{A}{\alpha}\frac{1}{x^2} - \frac{1-\frac{1}{\alpha}}{x} \equiv 0 &\Rightarrow x^2 - (1-\frac{1}{\alpha})x - \frac{A}{\alpha} = 0 \\ &\xrightarrow{\text{positive root}} x = \frac{\sqrt{(\alpha-1)^2+4\alpha A} - 1 + \alpha}{2\alpha} = T^{(\alpha)^{-1}}(A) + 1\end{aligned}$$

Therefore,

$$x^* = T^{(\alpha)^{-1}}(\theta^2\sigma_{\max}^2) + 1, \quad y^* = \alpha T^{(\alpha)^{-1}}(\theta^2\sigma_{\max}^2) + 1$$

It can be seen that $x^*y^* = T^{(\alpha)}\big(T^{(\alpha)^{-1}}(\theta^2\sigma_{\max}^2)\big) = \theta^2\sigma_{\max}^2$. Denoting $T^{(\alpha)^{-1}}(\theta^2\sigma_{\max}^2)$ by γ, $f(x^*, y^*)$ can be rewritten as (6.19), and we deduce Statement 6.3.

Part I

Low-Rank Mismatched Inference

Mismatched Estimation of Rank-One Symmetric Matrices

7

In this chapter, we investigate the mismatched estimation in the simplest case, where the signal matrix is a rank-one symmetric matrix and noise is Gaussian. Our primary objective is to compute the full asymptotic mismatched MSE in the large N limit, when a Gaussian prior is employed by statisticians for estimation. We introduce an additional temperature parameter β, in the posterior distribution, which may improve the estimation performance. For this model,

- We prove a relation which links the free energy of the system to the mismatched MSE, see lemma 7.1.
- We derive the asymptotic free energy for two cases of true priors: Gaussian prior in Theorem 7.4, and Bernoulli prior in Theorem 7.6.
- Using the f-MSE relation, we find the asymptotic mismatched MSE for Gaussian prior (Theorem 7.1) and Bernoulli prior (Theorem 7.3).
- We explore the performance of the AMP algorithm and spectral algorithms in the mismatched setting.

Part of this work was presented in [28] F. Pourkamali and N. Macris, "Mismatched estimation of symmetric rank-one matrices under gaussian noise," in International Zurich Seminar on Information and Communication (IZS 2022). Proceedings. ETH Zurich, 2022, pp. 84–88

7.1 Setting and Preliminaries

Suppose $\boldsymbol{s} \in \mathbb{R}^N$ is generated according to a prior denoted $\mathbf{P}^*$. The goal is to estimate the vector $\boldsymbol{s}$ upon observing the matrix

$$\boldsymbol{Y} = \sqrt{\frac{\kappa}{N}}\boldsymbol{s}\boldsymbol{s}^\intercal + \boldsymbol{Z} \tag{7.1}$$

where κ is the signal-to-noise-ratio (SNR), and the noise matrix $\boldsymbol{Z}$ is a symmetric matrix with i.i.d. $\mathcal{N}(0,1)$ off-diagonal and $\mathcal{N}(0,2)$ diagonal entries. This model is called the *Spiked-Wigner model*. The purpose of the scaling factor $\frac{1}{\sqrt{N}}$ is to make the inference problem neither trivially easy nor completely impossible in the large system limit.

To measure the quality of our estimate, we pick the *matrix* mean squared error (MSE). For any estimator (function of $\boldsymbol{Y}$) $\hat{\theta}$ which outputs an estimate of $\boldsymbol{s}\boldsymbol{s}^\intercal$, MSE is defined as

$$\mathrm{MSE}_N = \frac{1}{N^2}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\left[\left\|\boldsymbol{s}\boldsymbol{s}^\intercal - \hat{\theta}(\boldsymbol{Y})\right\|_{\mathrm{F}}^2\right] \tag{7.2}$$

where $\|.\|_{\mathrm{F}}$ is the Frobenius norm of a matrix. Note that it is natural to define the error in matrix form, since if P^* is a centered distribution, there is an ambiguity of sign inherited in the observation matrix; $\boldsymbol{s}$ can be replaced by $-\boldsymbol{s}$ without changing the probabilistic aspects of the observation matrix.

It can be proved that the minimum MSE (MMSE) is achieved for $\hat{\theta} = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\intercal|\boldsymbol{Y}]$, where the expectation is taken over the posterior distribution with true known parameters and prior distribution.

$$\mathrm{MMSE}_N := \frac{1}{N^2}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\left[\left\|\boldsymbol{s}\boldsymbol{s}^\intercal - \langle\boldsymbol{x}\boldsymbol{x}^\intercal\rangle_*]\right\|_{\mathrm{F}}^2\right]$$

where $\langle.\rangle_*$ denotes the expectation with respect to the posterior distribution. Here we adopt the traditional statistical mechanics notation for the internal (annealed) expectations

$$\langle f(\boldsymbol{x})\rangle_* = \frac{\int d\boldsymbol{x}\,\mathbf{P}^*(\boldsymbol{x})f(\boldsymbol{x})e^{-\frac{1}{4}\|\boldsymbol{Y}-\sqrt{\frac{\kappa}{N}}\boldsymbol{x}\boldsymbol{x}^\intercal\|_{\mathrm{F}}^2}}{\int d\boldsymbol{x}\,\mathbf{P}(\boldsymbol{x})e^{-\frac{1}{4}\|\boldsymbol{Y}-\sqrt{\frac{\kappa}{N}}\boldsymbol{x}\boldsymbol{x}^\intercal\|_{\mathrm{F}}^2}}$$

for any reasonable function $f(\boldsymbol{x})$ such that the integrals are finite. We necessarily have $\mathrm{MSE}_N \geq \mathrm{MMSE}_N$.

7.1.1 Mismatched inference

We consider the mismatched scenario in which the statistician is not aware of the prior distribution $\mathbf{P}^*$ and the SNR parameter κ. However, he knows that the channel is additive Gaussian with variance one (and two on diagonal [1]).

[1] In fact, as $N \to \infty$ the variance on the diagonal does not affect the probabilistic behavior of the problem

He assumes that the prior distribution is $\mathbf{P}$, and the SNR is κ'. Following the Bayesian approach, he chooses his estimator to be the mean of the posterior distribution. Our goal is to compute the asymptotic MSE for his estimator.

The statistician considers the posterior distribution which reads up to a normalizing factor

$$\begin{aligned}\mathbf{P}\{\boldsymbol{x}|\boldsymbol{Y}\} &\propto e^{-\frac{\beta}{4}\|\boldsymbol{Y}-\sqrt{\frac{\kappa'}{N}}\boldsymbol{x}\boldsymbol{x}^\intercal\|_{\mathrm{F}}^2}\mathbf{P}(\boldsymbol{x}) \\ &\propto e^{-\frac{\beta\kappa'}{4N}\|\boldsymbol{x}\|^4+\frac{\beta}{2}\sqrt{\frac{\kappa'}{N}}\operatorname{Tr}\boldsymbol{Y}\boldsymbol{x}\boldsymbol{x}^\intercal}\mathbf{P}(\boldsymbol{x})\end{aligned} \tag{7.3}$$

where $\mathbf{P}$ is the assumed mismatched distribution. In deriving the second line, we use the fact that $\|\boldsymbol{Y}\|_{\mathrm{F}}$ is a constant (because it is being conditioned on). Note that, by introducing the parameter β we consider a more general class of estimators. Although, in the fully matched case $\beta = 1$ is optimal, as we will see, in the mismatched case this auxiliary parameter may improve the performance of the estimator for a set of mismatched parameters.

The *partition function* is defined as the normalization factor of the posterior distribution

$$Z(\boldsymbol{Y}) = \int d\boldsymbol{x}\, e^{-\frac{\beta\kappa'}{4N}\|\boldsymbol{x}\|^4+\frac{\beta}{2}\sqrt{\frac{\kappa'}{N}}\operatorname{Tr}\boldsymbol{Y}\boldsymbol{x}\boldsymbol{x}^\intercal}\mathbf{P}(\boldsymbol{x}) \tag{7.4}$$

and the *mismatched free energy* is defined as

$$f_N(\mathbf{P}^*, \mathbf{P}, \kappa, \kappa', \beta) = -\frac{1}{N}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}[\ln Z(\boldsymbol{Y})] \tag{7.5}$$

Now we state a lemma relating mismatched free energy to MSE. Keep in mind that both mismatched free energy and MSE are functions of $\mathbf{P}^*, \mathbf{P}, \kappa, \kappa', \beta$, but for simplicity of notation we drop the arguments.

Lemma 7.1.

$$\begin{aligned}\left(2-\frac{1}{\beta}\sqrt{\frac{\kappa}{\kappa'}}\right)\frac{1}{\beta}\sqrt{\frac{\kappa}{\kappa'}}\frac{d}{d\kappa}f_N + \frac{2\beta-1}{\beta^2}\frac{d}{d\kappa'}f_N + \frac{1-\beta}{\beta}\frac{1}{\kappa'}\frac{d}{d\beta}f_N + \frac{1}{4N^2}\mathbb{E}_{\mathbf{P}^*}[\|\boldsymbol{s}\|^4] \\ = \frac{1}{4}\mathrm{MSE}_N\end{aligned} \tag{7.6}$$

Proof. Appendix 7.A. □

Remark 7.1. *Lemma* (7.1) *generalizes the classical I-MMSE relation. Here the mismatched free energy cannot be related to a mutual information. Nevertheless, note that, in the special case where* $\kappa' = \kappa, \beta = 1$ *Eq.* (7.1) *simplifies slightly and combining with the I-MMSE relation, we obtain that the difference of MSE and MMSE is directly related to a derivative of a relative entropy, equivalent to relations discussed in detail in [100] for vector channels.*

Thus, to find the asymptotic MSE, we need to compute the asymptotic (mismatched) free energy.

The central assumption in our analysis is that the statistician's prior is rotationally invariant. This assumption enables us to apply the known results from the literature on spherical integrals.

Exploiting the rotational invariance of $\mathbf{P}$ and changing variables $\boldsymbol{x} \to \boldsymbol{U}\boldsymbol{x}$, for an orthogonal matrix $\boldsymbol{U} \in \mathbb{R}^{N\times N}$, the integral in eq. (7.4) becomes ($|\det \boldsymbol{U}| = 1$):

$$\begin{aligned} Z(\boldsymbol{Y}) &= \int d\boldsymbol{x}\, e^{-\frac{\beta\kappa'}{4N}\|\boldsymbol{U}\boldsymbol{x}\|^4 + \frac{\beta}{2}\sqrt{\frac{\kappa'}{N}}\operatorname{Tr}\boldsymbol{Y}\boldsymbol{U}\boldsymbol{x}\boldsymbol{x}^\intercal\boldsymbol{U}^\intercal}\mathbf{P}(\boldsymbol{U}\boldsymbol{x}) \\ &= \int d\boldsymbol{x}\mathbf{P}(\boldsymbol{x})\, e^{-\frac{\beta\kappa'}{4N}\|\boldsymbol{x}\|^4 + \frac{\beta}{2}\sqrt{\frac{\kappa'}{N}}\operatorname{Tr}\boldsymbol{Y}\boldsymbol{U}\boldsymbol{x}\boldsymbol{x}^\intercal\boldsymbol{U}^\intercal} \end{aligned}$$

Since this holds for any orthogonal matrix $\boldsymbol{U}$, we can take the expectation over the *Haar* measure on the group of $N \times N$ orthogonal matrices.

$$Z(\boldsymbol{Y}) = \int d\boldsymbol{x}\mathbf{P}(\boldsymbol{x})\, e^{\frac{-\beta\kappa'}{4N}\|\boldsymbol{x}\|^4} \int D\boldsymbol{U} e^{\frac{\beta}{2}\sqrt{\frac{\kappa'}{N}}\operatorname{Tr}\boldsymbol{Y}\boldsymbol{U}\boldsymbol{x}\boldsymbol{x}^\intercal\boldsymbol{U}^\intercal} \tag{7.7}$$

where $D\boldsymbol{U}$ denotes the *Haar* measure.

7.1.2 Computing free energy

One clearly sees that the inner integral in eq. (7.7), is the rank-one spherical integral (6.1) in the rank-one setting. To apply the asymptotic result from theorem 6.1, we can rewrite the spherical integral in eq. (7.7) as

$$\int D\boldsymbol{U} e^{N\operatorname{Tr}\frac{\boldsymbol{Y}}{\sqrt{N}}\boldsymbol{U}\frac{\beta\sqrt{\kappa'}}{2N}\boldsymbol{x}\boldsymbol{x}^\intercal\boldsymbol{U}^\intercal} \tag{7.8}$$

The first matrix is $\frac{\boldsymbol{Y}}{\sqrt{N}} = \frac{\sqrt{\kappa}}{N}\boldsymbol{s}\boldsymbol{s}^\intercal + \frac{1}{\sqrt{N}}\boldsymbol{Z}$, where $\frac{1}{\sqrt{N}}\boldsymbol{Z}$ is the suitably normalized Wigner matrix whose limiting spectral measure is the renowned *semi-circle law* (5.18). Moreover, the spectral measure of $\frac{\boldsymbol{Y}}{\sqrt{N}}$ converges almost surely (a.s) as $N \to \infty$ to the *semi-circle law* (see e.g. proposition 1 in [101]). We have $\mathcal{G}_{\rho_{sc}}(z) = \frac{1}{2}(z - \sqrt{z^2-4})$ and $R_{\rho_{sc}}(z) = z$.

Let $\lambda_{\min}$ and $\lambda_{\max}$ be the bottom and top eigenvalue of $\frac{\boldsymbol{Y}}{\sqrt{N}}$, from the results in [102], we have (a.s.)

$$\lambda_{\min} = -2, \gamma_{\max} = \begin{cases} 2 & \text{if } \frac{\sqrt{\kappa}\|\boldsymbol{s}\|^2}{N} \le 1 \\ \frac{\sqrt{\kappa}\|\boldsymbol{s}\|^2}{N} + \frac{N}{\sqrt{\kappa}\|\boldsymbol{s}\|^2} & \text{if } \frac{\sqrt{\kappa}\|\boldsymbol{s}\|^2}{N} \ge 1 \end{cases} \tag{7.9}$$

So,

$$\mathcal{G}_{\min} = -1, \mathcal{G}_{\max} = \begin{cases} 1 & \text{if } \frac{\sqrt{\kappa}\|\boldsymbol{s}\|^2}{N} \le 1 \\ \frac{N}{\sqrt{\kappa}\|\boldsymbol{s}\|^2} & \text{if } \frac{\sqrt{\kappa}\|\boldsymbol{s}\|^2}{N} \ge 1 \end{cases} \tag{7.10}$$

On the other hand the non-zero eigenvalue of the rank-one matrix $\frac{\beta\sqrt{\kappa'}}{2N}\boldsymbol{x}\boldsymbol{x}^\intercal$ is $\theta = \frac{\beta\sqrt{\kappa'}}{2N}\|\boldsymbol{x}\|^2$. Thus, the asymptotic of the integral in eq. (7.8) is only a function of $\|\boldsymbol{x}\|$ and $\|\boldsymbol{s}\|$, and can be computed by theorem 6.1 for the different cases of the parameters.

Now assume that $\mathbf{P}(\boldsymbol{x}) \propto e^{-Ng(\|\boldsymbol{x}\|)}$. Substituting the asymptotic of the integral in eq. (7.7) in eq. (7.7), followed by an application of the Laplace method, we can find the asymptotics of the partition function, from which we are able to compute the free energy. Once we obtain the asymptotic of the free energy, we can find the MSE using lemma 7.1.

7.1.3 Approximate Message Passing (AMP) Algorithm

The approximate message passing (AMP) algorithm is an iterative algorithm rooted in statistical physics. A powerful feature of AMP is that its performance can be tracked by running a simple recursion called *State Evolution*. AMP and its state evolution were studied in [27] for the compressed sensing problem, and [103] generalized the results for the rank-one matrix estimation problem.

Following the work of [104], the AMP algorithm for the mismatched estimator in (7.3) reads

$$\begin{cases} \hat{x}_i^{t+1} = \eta_{\mathbf{P}}(a^t, b_i^t) \\ \nu_i^{t+1} = \eta'_{\mathbf{P}}(a^t, b_i^t) \\ b_i^t = \beta\sqrt{\frac{\kappa'}{N}}(\boldsymbol{Y}\hat{\boldsymbol{x}}^t)_i - \beta^2\kappa'\big(\frac{1}{N}\sum_{k=1}^N \nu_k^t\big)\hat{x}_i^{t-1} \\ a^t = \beta\kappa'\frac{\|\hat{\boldsymbol{x}}^t\|^2}{N} - \beta(\beta-1)\kappa'\frac{1}{N}\sum_{k=1}^N \nu_k^t \end{cases} \tag{7.11}$$

for $i = 1, \dots, N$, where $\eta_{\mathbf{P}}$ is called the *denoising function*, and $\eta'_{\mathbf{P}}$ is the derivative with respect to the second argument, b. The denoising function is defined as

$$\eta_{\mathbf{P}}(a, b) = \frac{\int dx\, \mathbf{P}(x)\, x\, e^{bx - \frac{1}{2}ax^2}}{\int dx\, \mathbf{P}(x)\, e^{bx - \frac{1}{2}ax^2}}$$

Note that the AMP equations are derived assuming that the (mismatched) prior $\mathbf{P}$ is i.i.d.

To derive the state evolution equations for (7.11), we introduce the following order parameters.

$$m^t = \frac{1}{N}(\boldsymbol{s}^\intercal\hat{\boldsymbol{x}}^t)^2, \quad q^t = \frac{1}{N}\|\hat{\boldsymbol{x}}^t\|^4, \quad \Sigma^t = \frac{1}{N}\sum_{i=1}^N \nu_i^t$$

m^t measures how much the AMP estimate is correlated with the true signal at iteration t, q^t is the norm of the estimate, and Σ^t is the mean of the variance of the estimate $\hat{x}_i$.

Assuming the independence of messages for large enough n, the variables b_i^t and a^t can be approximated by a normal random variables by central

limit theorem. This enables us to derive the recursion formula for the order parameters as follows:

$$\begin{cases} m^{t+1} = \mathbb{E}_{\mathbf{P}^*,\mathbf{P}_W}\Big[\eta_{\mathbf{P}}\big(\beta\kappa'\sqrt{q^t} - \beta(\beta-1)\kappa'\Sigma^t, \beta\sqrt{\kappa\kappa' m^t}s + \beta\sqrt{\kappa'\sqrt{q^t}}w\big)s\Big]^2 \\ q^{t+1} = \mathbb{E}_{\mathbf{P}^*,\mathbf{P}_W}\Big[\eta_{\mathbf{P}}\big(\beta\kappa'\sqrt{q^t} - \beta(\beta-1)\kappa'\Sigma^t, \beta\sqrt{\kappa\kappa' m^t}s + \beta\sqrt{\kappa'\sqrt{q^t}}w\big)^2\Big]^2 \\ \Sigma^{t+1} = \mathbb{E}_{\mathbf{P}^*,\mathbf{P}_W}\Big[\eta'_{\mathbf{P}}\big(\beta\kappa'\sqrt{q^t} - \beta(\beta-1)\kappa'\Sigma^t, \beta\sqrt{\kappa\kappa' m^t}s + \beta\sqrt{\kappa'\sqrt{q^t}}w\big)\Big] \end{cases} \tag{7.12}$$

where s and w are two independent random variables, w is a Gaussian variables with mean 0 and variance 1, and s is distributed according to $\mathbf{P}^*$. Iterating over the equations (7.12) allows us to to assess the MSE of the AMP estimate at each iteration.

$$\mathrm{MSE}^t = \mathbb{E}_{\mathbf{P}^*}\big[s^4\big] + q^t - 2m^t \tag{7.13}$$

Remark 7.2. *It is also possible to determine the Vector MSE of the AMP estimate at each iteration. For that purpose, the order parameters and state evolution equations are slightly different from the ones in* (7.12)*; however, since our primary object of interest is the matrix MSE, we refer the reader to the appendix for the evaluation of vector MSE.*

7.1.4 Nishimori identity

In the Bayes optimal setting in which the model is fully known, the Bayes' rule yields a set of identities which are called *Nishimori* identities. In its simplest form, it states that

$$\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\Big[\big\langle g(\boldsymbol{x}_1,\boldsymbol{x}_2)\big\rangle_*\Big] = \mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\Big[\big\langle g(\boldsymbol{s},\boldsymbol{x})\big\rangle_*\Big] \tag{7.14}$$

where g is a generic function, and $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}$ are i.i.d. according to the (true) posterior distribution, and $\boldsymbol{s}$ is generated from $\mathbf{P}^*$. Note that, the Nishimori identity holds for functions with more than two arguments, however for the matrix estimation problem the simple version, (7.14), is used.

Define the following two quantities

$$\begin{aligned} m^*_{\mathrm{stat}} &= \frac{1}{N^2}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\big[\operatorname{Tr}\boldsymbol{s}\boldsymbol{s}^\intercal\langle\boldsymbol{x}\boldsymbol{x}^\intercal\rangle_*\big] \\ q^*_{\mathrm{stat}} &= \frac{1}{N^2}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\Big[\big\|\langle\boldsymbol{x}\boldsymbol{x}^\intercal\rangle_*\big\|_{\mathrm{F}}^2\Big] \end{aligned}$$

Applying (7.14), we get

$$q^*_{\mathrm{stat}} = \mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\big[\operatorname{Tr}\langle\boldsymbol{x}_1\boldsymbol{x}_1^\intercal\rangle_*\langle\boldsymbol{x}_2\boldsymbol{x}_2^\intercal\rangle_*\big] = \mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\big[\operatorname{Tr}\boldsymbol{s}\boldsymbol{s}^\intercal\langle\boldsymbol{x}\boldsymbol{x}^\intercal\rangle_*\big] = m^*_{\mathrm{stat}}$$

Therefore, one of the primary identities we have in the Bayes optimal scenario is $q^*_{\mathrm{stat}} = m^*_{\mathrm{stat}}$.

For the mismatched case, we can similarly define

$$m_{\text{stat}} = \frac{1}{N^2}\mathbb{E}_{\mathbf{P},\mathbf{P}_{\boldsymbol{Z}}}\left[\operatorname{Tr} \boldsymbol{s}\boldsymbol{s}^\intercal \langle \boldsymbol{x}\boldsymbol{x}^\intercal\rangle_{\mathbf{P},\kappa'}\right]$$
$$q_{\text{stat}} = \frac{1}{N^2}\mathbb{E}_{\mathbf{P},\mathbf{P}_{\boldsymbol{Z}}}\left[\left\|\langle \boldsymbol{x}\boldsymbol{x}^\intercal\rangle_{\mathbf{P},\kappa'}\right\|_{\text{F}}^2\right]$$

First note that we have access to m^*_{stat}, q^*_{stat} directly from the free energy via the following relations

$$\begin{aligned} m_{\text{stat}} &= -\frac{4}{\beta}\sqrt{\frac{\kappa}{\kappa'}}\frac{d}{d\kappa}f \\ q_{\text{stat}} &= -\frac{4}{\beta}\frac{\kappa}{\kappa'}\frac{d}{d\kappa}f + \frac{4(2\beta-1)}{\beta^2}\frac{d}{d\kappa'}f + \frac{4}{\kappa'}\frac{1-\beta}{\beta}\frac{d}{d\beta}f \end{aligned} \tag{7.15}$$

As we will see, enforcing the Nishimori condition, $m_{\text{stat}} = q_{\text{stat}}$ in the mismatched case leads to the minimum mismatched MSE.

7.2 Gaussian Prior

Let s_i be i.i.d. random variables generated from $\mathcal{N}(0,\sigma^2)$ for $i = 1,\dots,N$. The statistician observes the matrix $\boldsymbol{Y}$ of the form (7.1). Considering the posterior distribution (7.3) with $\mathbf{P}$ be the centered Gaussian with variance σ'^2, we denote his estimate for $\boldsymbol{s}\boldsymbol{s}^\intercal$ by $\langle \boldsymbol{x}\boldsymbol{x}^\intercal\rangle_{\sigma',\kappa',\beta}$ to stress that the estimation is done with the mismatched parameters σ', κ', β. The asymptotic MSE of this estimator is stated in the following theorem.

Theorem 7.1. *Assume that the sequence* $(\text{MSE})_{N\geq 1}$ *converges uniformly in* $(\kappa,\kappa',\beta) \in K \subset \mathbb{R}^3_+$. *Then for all* σ,σ' *(strictly positive) and* $(\kappa,\kappa',\beta) \in K$, *the asymptotic mismatched MSE reads:*

$$\begin{aligned} &\lim_{N\to\infty} \text{MSE}_N(\sigma,\sigma',\kappa,\kappa',\beta) \\ &\quad = \begin{cases} \sigma^4 + \frac{\left[(1-2\beta)\sqrt{\kappa'}\sigma'^2+1\right]^2}{\beta^2\kappa'^2\sigma'^4} & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}, \kappa \leq \frac{1}{\sigma^4}, \\ A & \text{if } \kappa \geq \frac{1}{\sigma^4}, \sqrt{\kappa\kappa'} \geq \frac{1}{\beta\sigma^2\rho^*}, \\ \sigma^4 & \text{o.w.} \end{cases} \end{aligned} \tag{7.16}$$

where

$$\begin{aligned} A = \sigma^4\left(1-\sqrt{\frac{\kappa}{\kappa'}}\right)^2 + \frac{1}{\beta}\left[\frac{2}{\sqrt{\kappa\kappa'}} + \frac{1}{\beta}\frac{1}{\kappa'^2\sigma'^4} + \frac{2}{\kappa'}\frac{\sigma^2}{\sigma'^2}\left(1-\sqrt{\frac{\kappa}{\kappa'}}\right) - \frac{2}{\kappa\kappa'\sigma^2\sigma'^2}\right] \\ + \left(1-\frac{1}{\beta}\right)\left[\frac{2}{\kappa'} - \frac{1}{\beta}\frac{2}{\sqrt{\kappa\kappa'^3}\sigma^2\sigma'^2} + \left(1-\frac{1}{\beta}\right)\frac{1}{\kappa\kappa'\sigma^4} + \frac{2}{\sqrt{\kappa^3\kappa'\sigma^4}}\right] \end{aligned}$$

and

$$\rho^* = \begin{cases} \sigma'^2 & \text{if } \beta = 1, \\ \frac{1}{\kappa'\beta(\beta-1)}\left(\frac{1}{2\sigma'^2} - \sqrt{\frac{1}{4\sigma'^4} - \kappa'\beta(\beta-1)}\right) & \text{if } \beta \neq 1 \end{cases}$$

Proof. First, we derive the asymptotic free energy (7.5) in Theorem 7.4. This derivation utilizes the careful application of the Laplace method for computation of the integrals, incorporating results on asymptotic log-spherical integrals, as explained in section 7.1.2. Subsequently, we compute the asymptotic MSE using lemma 7.1. For a details, refer to Appendix 7.B. □

Remark 7.3. *In the fully matched case, uniform convergence of the sequence* $(\mathrm{MMSE})_{N\geq 1}$ *- except possibly at phase transition points which form a set of measure zero - follows using the concavity of mutual information with respect to* κ*. Then, using the I-MMSE relation [105], this allows to interchange limit and derivative to go from asymptotic mutual information (a.k.a. free energy) to asymptotic MMSE. For the present mismatched MSE, we use a relation similar to I-MMSE but in terms of mismatched free energies, which lack concavity w.r.t.* κ *and* κ'*. Therefore, uniform convergence is difficult to establish from general principles. However, we conjecture that it holds on a set* K *of measure one and that* (7.16) *holds almost everywhere (i.e., except possibly at phase transition lines).*

In the following, we will investigate the asymptotic MSE for different cases of parameters. Note that the MSE$= \sigma^4$ can be achieved by simply $\hat{\boldsymbol{x}} = 0$. Therefore, we are interested in studying the problem where the inference is possible, in the sense that MSE less than σ^4 is achievable. By theorem 7.1 it is statistically impossible to achieve a MSE lower than σ^4 when $\kappa < \frac{1}{\sigma^4}$, even in the matched case. Thus, the main quantity of interest is the expression A, and we try to study the behavior of this expression for different cases of parameters.

Also, we investigate the performance of AMP algorithm and compare it with the spectral methods. For mismatched prior $\mathbf{P} = \mathcal{N}(0, \sigma'^2)$, the AMP equations (7.11) reduces to

$$\begin{cases} \hat{x}_i^{t+1} = \frac{b_i^t}{a^t + \frac{1}{\sigma'^2}} \\ b_i^{\intercal} = \beta\sqrt{\frac{\kappa'}{N}}(\boldsymbol{Y}\hat{\boldsymbol{x}}^t)_i - \beta^2 \frac{\kappa'}{a^{t-1}+\frac{1}{\sigma'^2}}\hat{x}_i^{t-1} \\ a^{\intercal} = \beta\kappa' \frac{\|\hat{\boldsymbol{x}}^t\|^2}{N} - \beta(\beta-1)\frac{\kappa'}{a^{t-1}+\frac{1}{\sigma'^2}} \end{cases} \tag{7.17}$$

and the state evolution equations read

$$\begin{cases} m^{t+1} = \frac{\beta^2\kappa\kappa'\sigma^4 m^t}{\left[\beta\kappa'\sqrt{q^t} - \beta(\beta-1)\kappa'\Sigma^t + \frac{1}{\sigma'^2}\right]^2} \\ q^{t+1} = \frac{\beta^4\kappa'^2\left(\kappa\sigma^2 m^t + \sqrt{q^t}\right)^2}{\left[\beta\kappa'\sqrt{q^t} - \beta(\beta-1)\kappa'\Sigma^t + \frac{1}{\sigma'^2}\right]^4} \\ \Sigma^{t+1} = \frac{1}{\beta\kappa'\sqrt{q^t} - \beta(\beta-1)\kappa'\Sigma^t + \frac{1}{\sigma'^2}} \end{cases} \tag{7.18}$$

To evaluate the performance of AMP, we can find the fixed points of SE equations above, which we denote by $m^*_{\mathrm{AMP}}, q^*_{\mathrm{AMP}}$.

7.2.1 Bayes optimal estimation

Since the true and the mismatched prior are Gaussian, we can have the Bayes optimal scenario by substituting $\kappa' = \kappa, \sigma' = \sigma$. This leads to

$$A = \frac{2}{\kappa} + \frac{1}{\kappa^2\sigma^4}\big(3 - \frac{8}{\beta} + \frac{4}{\beta^2}\big)$$

which attains its minimum for $\beta = 1$. Putting $\beta = 1$ in (7.16) implies the MMSE for the Gaussian prior.

$$\lim_{N\to\infty} \mathrm{MMSE}_N(\sigma, \kappa) = \begin{cases} \sigma^4 & \text{if } \kappa \le \frac{1}{\sigma^4} \\ \frac{2}{\kappa} - \frac{1}{\kappa^2\sigma^4} & \text{if } \kappa \ge \frac{1}{\sigma^4} \end{cases} \tag{7.19}$$

This expression is well known and derived previously by a set of different approaches (see [8, 9, 106–108]).

The free energy for the Bayes optimal case can be obtained from theorem 7.4, from which we can find $m^*_{\mathrm{stat}}, q^*_{\mathrm{stat}}$ using (7.15). We can see that as predicted by the Nishimori identity they are equal.

$$m^*_{\mathrm{stat}} = q^*_{\mathrm{stat}} = \big(\sigma^2 - \frac{1}{\kappa\sigma^2}\big)^2$$

Performance of AMP

In this scenario, AMP is optimal, in the sense that it achieves the same MMSE. This can be checked from the fixed points of the SE equations and the relation (7.13).

$$m^*_{\mathrm{AMP}} = q^*_{\mathrm{AMP}} = \big(\sigma^2 - \frac{1}{\kappa\sigma^2}\big)^2$$

Spectral algorithm

A natural and simple algorithm for this problem is to compute the top eigenvector of $\boldsymbol{Y}$. Spectral analysis of the matrix $\boldsymbol{Y}$ yields that if $\kappa\sigma^4 > 1$ then the top eigenvector of $\boldsymbol{Y}$ is correlated with the planted signal [102]. Denote the (unit norm) top eigenvector of $\boldsymbol{Y}$ by $\boldsymbol{y}$, and the normalized $\boldsymbol{s}$ by $\bar{\boldsymbol{s}}$. In the limit $N \to \infty$, we have

$$\big(\bar{\boldsymbol{s}}^\intercal \boldsymbol{y}\big)^2 = \begin{cases} 0 & \text{if } \kappa\sigma^4 \le 1 \\ 1 - \frac{1}{\kappa\sigma^4} & \text{if } \kappa\sigma^4 \ge 1 \end{cases} \tag{7.20}$$

To minimize the matrix MSE, we can rescale $\boldsymbol{y}$ by a factor δ. With this rescaling, we have

$$\mathrm{MSE} = \lim_{N\to\infty} \frac{1}{N^2}\mathbb{E}_{\mathbf{P}^*}\Big[\|\boldsymbol{s}\boldsymbol{s}^\intercal - \delta^2\boldsymbol{y}\boldsymbol{y}^\intercal\|_{\mathrm{F}}^2\Big] = \sigma^4 + \delta^4 - 2\delta^2\sigma^2\big(1 - \frac{1}{\kappa\sigma^4}\big)$$

Minimizing over δ, we find that the optimum δ is $\sqrt{\sigma^2 - \frac{1}{\kappa\sigma^2}}$. Note that there should be also a factor of $\sqrt{N}$ in the rescaling. With this rescaling, we can see that the spectral estimate can also achieve the MMSE.

Remark 7.4. *First, note that the AMP equations form a kind of power iteration applied to the matrix* $\boldsymbol{Y}$ *(In fact this comes from the linearity of the denoising function for the Gaussian prior), so the fixed point of AMP iteration is the top eigenvector of* $\boldsymbol{Y}$*. On the other hand,* q^*_{AMP} *coincides with the optimum rescaling factor (Note that* q^*_{AMP} *is* $\|\hat{\boldsymbol{x}}\|^4$*). Therefore, the AMP can be seen as a spectral method with inherited rescaling.*

7.2.2 Mismatched estimation with $\beta = 1$

Corollary 7.2. *For* $\beta = 1$*, the asymptotic MSE in* (7.16) *reduces to*

$$\lim_{N\to\infty} \mathrm{MSE}_{\mathrm{N}} = \begin{cases} \sigma^4 + \left(\frac{1-\sqrt{\kappa'}\sigma'^2}{\kappa'\sigma'^2}\right)^2 & \text{if} \kappa \le \frac{1}{\sigma^4}, \text{and} \kappa' \ge \frac{1}{\sigma'^4} \\ \sigma^4\left(1-\sqrt{\frac{\kappa}{\kappa'}}\right)^2 + \frac{2}{\sqrt{\kappa\kappa'}} + \frac{1}{\kappa'^2\sigma'^4} & \text{if} \kappa \ge \frac{1}{\sigma^4}, \text{and} \sqrt{\kappa\kappa'} \ge \frac{1}{\sigma^2\sigma'^2} \\ \quad + \frac{2}{\kappa'}\frac{\sigma^2}{\sigma'^2}\left(1-\sqrt{\frac{\kappa}{\kappa'}}\right) - \frac{2}{\kappa\kappa'\sigma^2\sigma'^2} & \\ \sigma^4 & \text{o.w.} \end{cases} \tag{7.21}$$

Fix $\sigma = 1, \kappa = 4$, the mismatched MSE is illustrated in Fig. 7.2.1. Throughout the paper, we stick to this example for simplicity; however, the observations are generic for $\kappa\sigma^4 > 1$ (note that σ^4 could be absorbed in κ in the model definition, so it is natural to set $\sigma = 1$), and can be checked analytically from the expressions of the MSE and MMSE. We observe one phase transition line and an intermediate region where estimation better than chance is possible, in the sense that the MSE is smaller than σ^4. We refer to the caption of Fig. 1 for details. In the case $\sigma = 1$ and $\kappa < 1$, or more generally $\kappa\sigma^4 < 1$, it is easy to see from Eq. (7.21) that the intermediate region disappears and the MSE is always greater or equal to σ^4 (the phase transition line is still present technically speaking).

Figures 7.2.2 and 7.2.3 depict the behavior of the MSE along vertical and horizontal sections of Fig. 1. We clearly observe that the MSE is not monotonous and that for $\kappa' < 8$, the minimal value given by the MMSE may be achieved.

Putting $\kappa' = \kappa$, the mismatched MSE then reduces to:

$$\text{if } \sigma' \le \sigma,\ \lim_{N\to\infty} \mathrm{MSE}_N(\sigma,\sigma',\kappa) = \begin{cases} \sigma^4 & \text{if } \kappa \le \frac{1}{\sigma^2\sigma'^2} \\ \frac{2}{\kappa} - \frac{1}{\kappa^2\sigma'^2}\left(\frac{2}{\sigma^2} - \frac{1}{\sigma'^2}\right) & \text{if } \kappa \ge \frac{1}{\sigma^2\sigma'^2} \end{cases}$$

$$\text{if } \sigma' \ge \sigma,\ \lim_{N\to\infty} \mathrm{MSE}_N(\sigma,\sigma',\kappa) = \begin{cases} \sigma^4 & \text{if } \kappa \le \frac{1}{\sigma'^4} \\ \sigma^4 + \frac{1}{\kappa} - \frac{2}{\kappa^{\frac{3}{2}}\sigma'^2} + \frac{1}{\kappa^2\sigma'^4} & \text{if } \frac{1}{\sigma'^4} \le \kappa \le \frac{1}{\sigma^4} \\ \frac{2}{\kappa} - \frac{1}{\kappa^2\sigma'^2}\left(\frac{2}{\sigma^2} - \frac{1}{\sigma'^2}\right) & \text{if } \kappa \ge \frac{1}{\sigma^4} \end{cases}$$

For $\sigma = 1$ the MSE is plotted as a function of SNR for various values of σ' in Fig. 7.2.4. When $\sigma' > \sigma$, we observe that the MSE increases as the SNR increases. Although this happens when we are still in the regime of small SNR

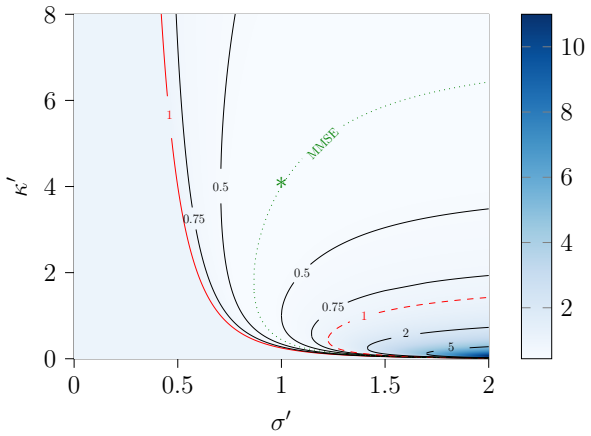


Figure 7.2.1: Plot of MSE according to Eq. (7.21) for $\sigma = 1, \kappa = 4$. The solid leftmost (red) curve is a *phase transition* line. On the left of this curve $\mathrm{MSE} = \sigma^4 = 1$. In the *intermediate region* between the solid leftmost (red) curve and the dashed (red) curve the MSE takes values less than $\sigma^4 = 1$. In this intermediate region estimation better than chance is possible. On the dotted (green) curve the MSE attains equal to $\mathrm{MMSE}(\sigma, \kappa) = \frac{2}{\kappa} - \frac{1}{\kappa^2\sigma^4} = 7/16$ (even though we do not have $\kappa' = \kappa$, $\sigma' = \sigma$ except for one point which corresponds to the true parameters, shown by $*$). This curve has vertical tangent at $\kappa' = 16/9, \sigma' = \sqrt{3}/2$. The MSE equals $\sigma^4 = 1$ on the dashed (red) line and takes higher values in the region on the right-hand side of this line. Note that this is *not* a phase transition line. Finally, we point out that the MSE is continuous throughout, and the phase transition is, therefore, a *continuous phase transition*. The analytical expressions of the phase transition line, as well as dotted and dashed lines can easily be written down from eqs. (7.21) and (7.19). For $\sigma = 1, \kappa = 4$ the dotted (resp. dashed) curves have horizontal asymptotes $\kappa' = 64/9$ (resp. $\kappa' = 16/9$).

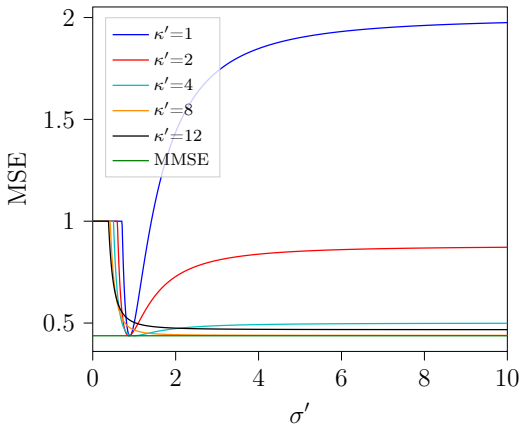


Figure 7.2.2: Behavior of the MSE as a function of σ'. Here $\sigma = 1, \kappa = 2$. The horizontal (green) level gives the value of the $\mathrm{MMSE} = \frac{2}{\kappa} - \frac{1}{\kappa^2\sigma^4} = 7/16$ in the matched case. We have $\lim_{\sigma' \to +\infty} \mathrm{MSE} = 1 - \frac{3}{\sqrt{\kappa'}} + \frac{4}{\kappa'}$ and this limiting value is decreasing (resp. increasing) for $\kappa' < 64/9$ (resp. $\kappa' > 64/9$). For $\kappa' > 16/9$ estimation better then chance is possible for large enough σ'.

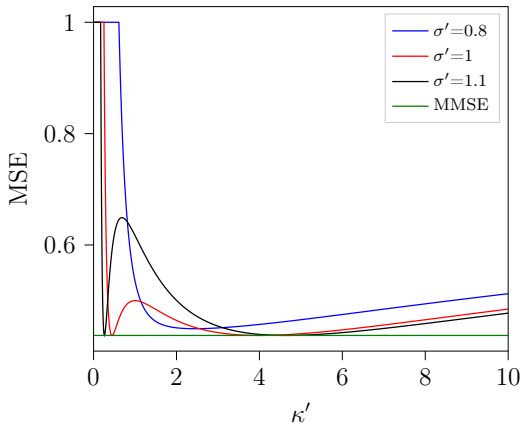


Figure 7.2.3: Behavior of the MSE as a function of κ'. Here $\sigma = 1, \kappa = 2$. The horizontal (green) level gives the value of the $\mathrm{MMSE} = \frac{2}{\kappa} - \frac{1}{\kappa^2\sigma^4} = 7/16$ in the matched case. For $\sigma' < \sqrt{3}/2$, the curve does not attain the MMSE, but for $\sigma' > \sqrt{3}/2$ the curve equals the MMSE at two points (this can also be seen in Fig. 7.2.1.). All curves have horizontal asymptote $\sigma^4 = 1$ for $\kappa \to +\infty$.

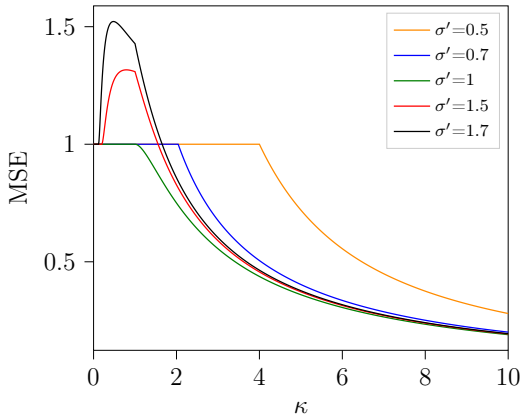


Figure 7.2.4: Behavior of the MSE for matched SNR $\kappa' = \kappa$.

and estimation is impossible, we find this behavior rather counterintuitive. A similar behavior has been observed in Fig. 1 of [100] for the vector case.

As a sanity check of our result for the matched SNR case, with a bit of work we can check explicitly that

$$\int_0^\infty [\mathrm{MSE}(\sigma, \sigma', \kappa) - \mathrm{MMSE}(\sigma, \kappa)]\, d\kappa = 4 D_{KL}(\mathcal{N}(0, \sigma^2), \mathcal{N}(0, \sigma'^2)) \tag{7.22}$$

where D_{KL} denotes the Kullback-Leibler divergence. This sum-rule was already derived in [100] for vector channels (with a factor of 2 instead of 4 in the vector case).

Remark 7.5. *The MMSE achieving curve in Fig. 7.2.1 can also be obtained by enforcing the Nishimori condition, $m_{\mathrm{stat}} = q_{\mathrm{stat}}$. By* (7.15)*, and* (7.30) *we can compute the two quantities $m_{\mathrm{stat}}, q_{\mathrm{stat}}$. Equating the two expressions, we*

find the same curve as in Fig. 1. This was observed in [109] by enforcing the Nishimori identity on the order parameters of estimation of AMP. For more details, we refer the reader to the appendix 7.C.

Remark 7.6. *Suppose instead of the MSE, we choose the angle between the estimate* $\langle \boldsymbol{x}\boldsymbol{x}^\intercal \rangle_{\sigma',\kappa'}$ *and* $\boldsymbol{s}\boldsymbol{s}^\intercal$ *as the measure of our accuracy. To find the angle, we need to compute the normalized inner product of the two matrices.*

$$\frac{\operatorname{Tr} \boldsymbol{s}\boldsymbol{s}^\intercal \langle \boldsymbol{x}\boldsymbol{x}^\intercal \rangle_{\sigma',\kappa'}}{\|\langle \boldsymbol{x}\boldsymbol{x}^\intercal \rangle_{\sigma',\kappa'}\|_{\mathrm{F}} \|\boldsymbol{s}\boldsymbol{s}^\intercal\|_{\mathrm{F}}} = \frac{m_{\mathrm{stat}}}{\sqrt{q_{\mathrm{stat}}}\sigma^2} = 1 - \frac{1}{\kappa\sigma^4}$$

This value is the value that also can be get by spectral method, see eq. (7.20)*. This suggests that, in the case that* $\kappa \geq \frac{1}{\sigma^4}, \sqrt{\kappa\kappa'} \geq \frac{1}{\sigma^2\sigma'^2}$ *the posterior mean always has the possible optimum angle with the ground, however due to the resacling factor* q_{stat} *we might get different MSEs. Therefore, the bracket estimator acts like the spectral algorithm with different rescaling* q_{stat}*. Equating* q_{stat} *with the optimum rescaling in the fully matched case* $\left(\sigma^2 - \frac{1}{\kappa\sigma^2}\right)^2$ *gives the MMSE curve.*

Performance of AMP

For $\beta = 1$, the AMP (7.17) reduces to

$$\begin{cases} \hat{x}_i^{t+1} = \frac{b_i^t}{a^t + \frac{1}{\sigma'^2}} \\ b_i^t = \sqrt{\frac{\kappa'}{N}}(\boldsymbol{Y}\hat{\boldsymbol{x}}^t)_i - \frac{\kappa'}{a^{t-1}+\frac{1}{\sigma'^2}}\hat{x}_i^{t-1} \\ a^t = \kappa' \frac{\|\hat{\boldsymbol{x}}^t\|^2}{N} \end{cases} \tag{7.23}$$

and SE (7.18) equations read

$$\begin{cases} m^{t+1} = \frac{\kappa\kappa'\sigma^4 m^t}{\left[\kappa'\sqrt{q^t}+\frac{1}{\sigma'^2}\right]^2} \\ \\ q^{t+1} = \frac{\kappa'^2\left(\kappa\sigma^2 m^t + \sqrt{q^t}\right)^2}{\left[\kappa'\sqrt{q^t}+\frac{1}{\sigma'^2}\right]^4} \end{cases} \tag{7.24}$$

By eq. (7.13), the MSE of AMP estimate can be tracked by iterating the SE equations, see Fig. 5. The MSE of the estimate of the AMP can be obtained from studying the fixed points of the SE equations (7.24) and the relation (7.13), which results in (7.21).

In the mismatched case, one may think that the number of iterations needed for the AMP to converge may help the statistician to find the optimum parameters, or at least to find the MMSE achieving curve, however as it is shown in Fig. 7.2.6, the statistician cannot infer any information from the number of iterations.

By remark 7.5, enforcing the Nishimori identity is equivalent to choose σ', κ' on the MMSE achieving curve. Running AMP with the chosen parameters

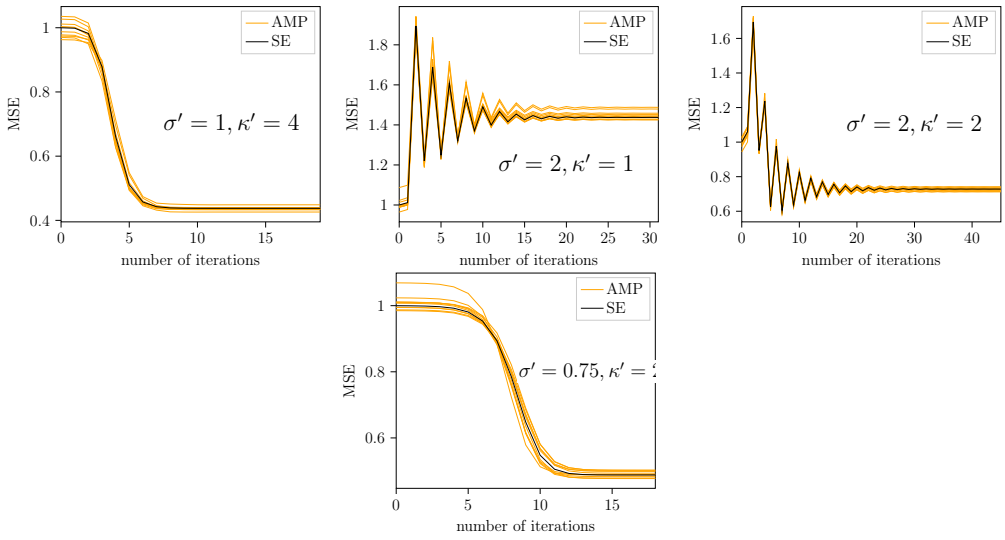


Figure 7.2.5: Value of the MSE under iterations of AMP equations (7.23) for 10 instances for $\sigma = 1, \kappa = 4, N = 10000$, compared with the SE predictions (7.24). The AMP is randomly initialized with $m = q = 1/N$. In the left most plot (matched case), MSE monotonically decreases with iteration, this can be checked by iterating the SE equations. However, this is not necessarily the case in the mismatched case. In this scenario, MSE is monotonically decreasing when $\sigma' \leq \sigma$. Moreover, we see that the AMP converges even in the regions where the estimation is worse than the random guess, and the SE tracks its performance.

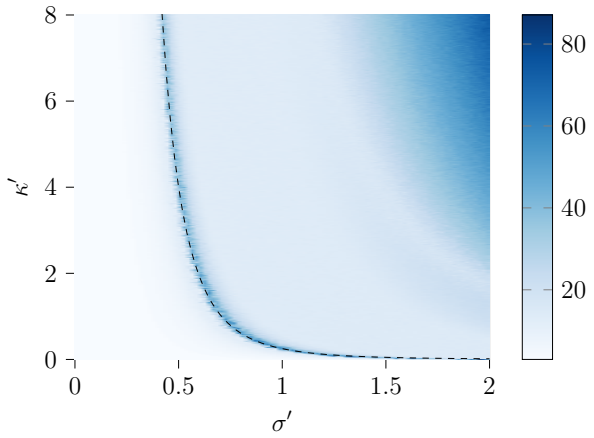


Figure 7.2.6: Averaged (over 100 instances) number iterations that AMP takes to converge for $\sigma = 1, \kappa = 4, N = 1000$. The dashed black line is where there is a phase transition in MSE. As it can be seen, the level sets of the number of iterations are not correlated with the ones of the MSE, and the number of iterations does not reveal any information.

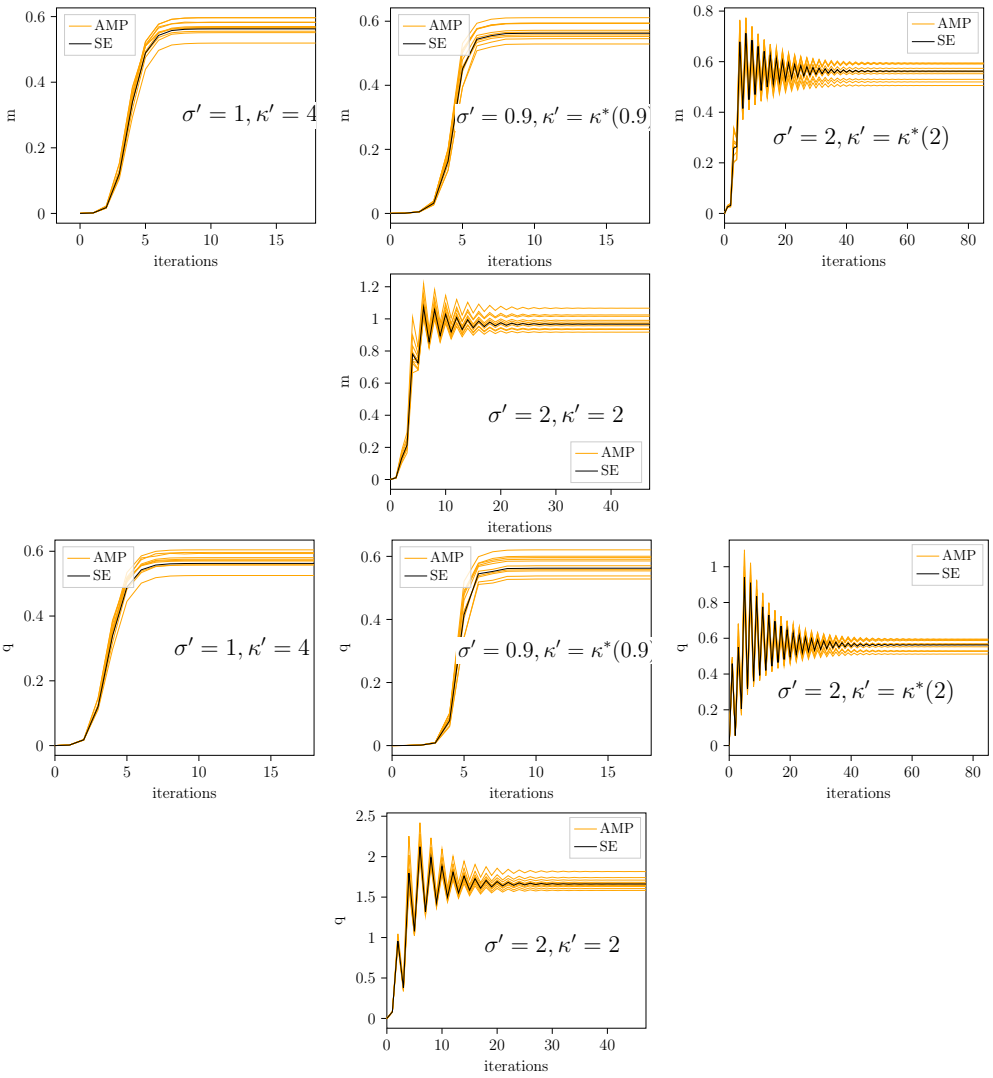


Figure 7.2.7: Value of m and q under iterations of AMP equations (7.23) for 10 instances for $\sigma = 1, \kappa = 4, N = 10000$, compared with the SE predictions (7.24). The AMP is randomly initialized with $m = q = 1/N$. In the left most plot (matched case), m and q are equal through iterations and monotonically increase until the algorithm stops. In the the two middle plots in which κ', σ' are chosen on the MMSE achieving curve, m and q are not equal through iterations, and their evolution is not necessarily monotone (depends on σ', see caption of Fig. 7.2.5), but $m = q$ holds for the estimate when the algorithm converges. And, the right most plot, illustrates an example that the Nishimori identity does not hold through iterations and the final estimate.

results in the Nishimori identity $m = q$ holding for the estimation when AMP converges. However, unlike the matched case, $m = q$ does not hold through the iterations, even if we initialize AMP with $m = q$.

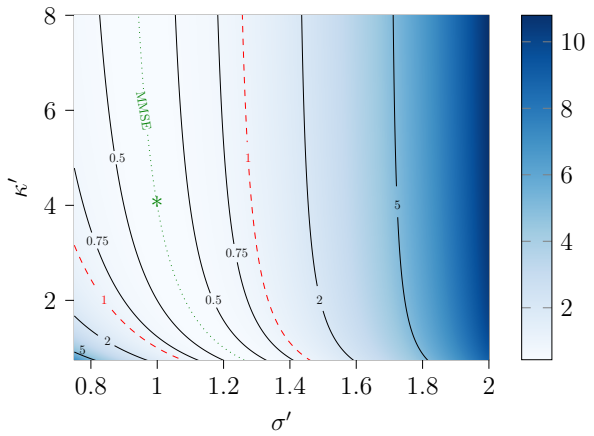


Figure 7.2.8: Plot of MSE according to Eq. (7.25) for $\sigma = 1, \kappa = 4$. The solid leftmost (red) curve is a *phase transition* line. In the *intermediate region* between the two dashed (red) curves the MSE takes values less than $\sigma^4 = 1$. In this intermediate region estimation better than chance is possible. On the dotted (green) curve the MSE attains the $\mathrm{MMSE}(\sigma, \kappa) = \frac{2}{\kappa} - \frac{1}{\kappa^2\sigma^4} = 7/16$ (even though we do not have $\kappa' = \kappa$, $\sigma' = \sigma$ except for one point which corresponds to the true parameters, shown by $*$). This curve has vertical asymptote at $\sigma' = \sqrt{3}/2$. This value is the minimum σ' that there exists an optimum κ', and is the same as the bracket estimator. The analytical expressions of the phase transition line and dotted and dashed lines can easily be written down from eqs. (7.25) and (7.19).

Performance of spectral method

As we saw in the Bayes optimal case, for the normal prior, we need to rescale the top (unit-norm) eigenvector of the the observation matrix $\boldsymbol{Y}$ by $\sqrt{\sigma^2 - \frac{1}{\kappa\sigma^2}}$ to achieve the MMSE. In the mismatched scenario, we assume that the statistician will rescale with the factor $\sqrt{\sigma'^2 - \frac{1}{\kappa'\sigma'^2}}$. This rescaling results in a MSE which we denote it by $\mathrm{MSE}_{\mathrm{spectral}}$. We have

$$\begin{aligned}
\mathrm{MSE}_{\mathrm{spectral}} &= \mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\left[\left\|\boldsymbol{s}\boldsymbol{s}^\intercal - \left(\sqrt{\sigma'^2 - \frac{1}{\kappa'\sigma'^2}}\right)^2 \boldsymbol{y}\boldsymbol{y}^\intercal\right\|_{\mathrm{F}}^2\right] \\
&= \sigma^4 + \left(\sigma'^2 - \frac{1}{\kappa'\sigma'^2}\right)^2 - 2\sigma^2\left(\sigma'^2 - \frac{1}{\kappa'\sigma'^2}\right)\left(1 - \frac{1}{\kappa\sigma^4}\right) \\
&= \left(\sigma^2 - \sigma'^2\right)^2 + \frac{1}{\kappa'^2\sigma'^4} - \frac{2}{\kappa'} + 2\frac{\sigma^2}{\sigma'^2}\frac{1}{\kappa'} + 2\frac{\sigma'^2}{\sigma^2}\frac{1}{\kappa} - \frac{2}{\kappa\kappa'\sigma^2\sigma'^2}
\end{aligned} \tag{7.25}$$

The MSE of spectral method with rescaling is depicted in Fig. 7.2.8 for $\kappa = 4, \sigma = 1$. The MSE is continuous in the plain, and there is no phase transition. Similar to the bracket estimator, the MMSE achieving curve can be found either by equating MSE and MMSE, or by equating the mismatched rescaling factor with the matched one, $\sqrt{\sigma'^2 - \frac{1}{\kappa'\sigma'^2}} = \sqrt{\sigma^2 - \frac{1}{\kappa\sigma^2}}$.

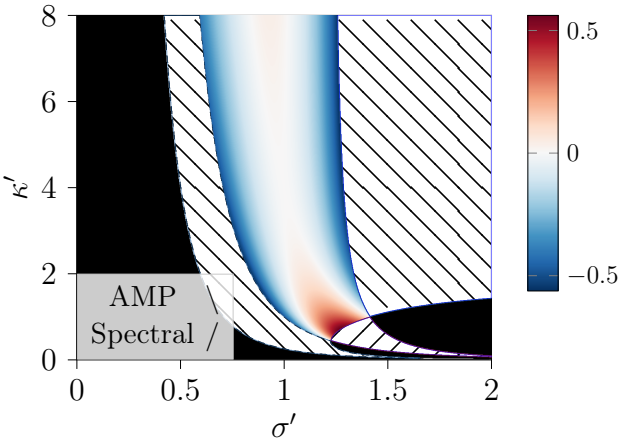


Figure 7.2.9: Comparison of AMP and spectral method. In black regions, none of the two algorithms can get an estimate better than a random guess. In regions hatched with \, AMP converges to an estimate with MSE less than 1, and the spectral method does not find an appropriate estimate. In the regions hatched with / spectral method, unlike AMP, finds a good estimate. In the intermediate region both algorithms attain MSE less than 1, and $\mathrm{MSE}_{\mathrm{AMP}} - \mathrm{MSE}_{\mathrm{Spectral}}$ is illustrated.

Comparison of AMP and spectral method

First, we compare the two algorithms in the fully mismatched case. We compare the methods in situations where at least one of them finds a reasonable estimate. In fact, in the regions where AMP obtains estimates with MSE less than 1, AMP behaves like the spectral method with different rescaling. Therefore, comparing the two algorithms is to compare the two different rescaling functions. Fig. 7.2.9 illustrates the difference between the two MSEs' in the fully mismatched case.

A remarkable point in the AMP rescaling factor is that it is data dependent, unlike the spectral method. The AMP rescaling factor is $\sqrt{\sqrt{\frac{\kappa}{\kappa'}}\sigma^2 - \frac{1}{\kappa'\sigma'^2}}$ which depends on the observation matrix via the term $\sqrt{\kappa}\sigma^2$. This value can be estimated in two ways. First, the top eigenvector of the matrix $\boldsymbol{Y}$ is approximately $\sqrt{\kappa}\sigma^2 + \frac{1}{\sqrt{\kappa}\sigma^2}$. Also, running AMP with any set of κ', σ' that $\kappa' \leq \frac{1}{\sigma'^4}$, and $\sqrt{\kappa\kappa'} \geq \frac{1}{\sigma^2\sigma'^2}$ or $\kappa' \geq \frac{1}{\sigma'^4}$ holds will converge to a vector with norm $\sqrt{\sqrt{\frac{\kappa}{\kappa'}}\sigma^2 - \frac{1}{\kappa'\sigma'^2}}$ from which $\sqrt{\kappa}\sigma^2$ can be extracted.

Now suppose one of the parameters κ, σ is known to the statistician. He can estimate the other parameter by finding the value $\sqrt{\kappa}\sigma^2$ from the matrix $\boldsymbol{Y}$, then he can apply his inference method (whether AMP or spectral) with one matched parameter and one with a small mismatched. In Fig. 7.2.10, we compare the MSE of the two methods for this scenario. In the case where κ is known, the MSE of AMP (or equivalently bracket estimator) is close to the MMSE around the true value; however the spectral method behaves poorly in the neighborhood of σ. In the other case, where σ is known, both approaches

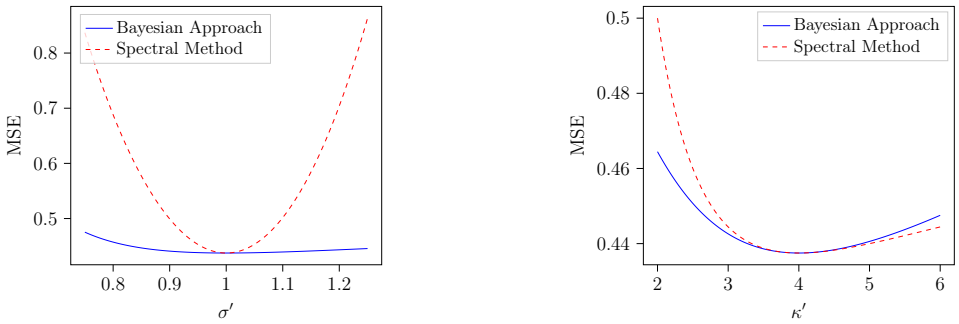


Figure 7.2.10: Comparison of the MSE of the Bayesian approach and spectral method in the partially mismatched scenario. In the left plot, the SNR parameter κ is known, and σ' is estimated from the data. In the right plot, σ is known and κ is estimated.

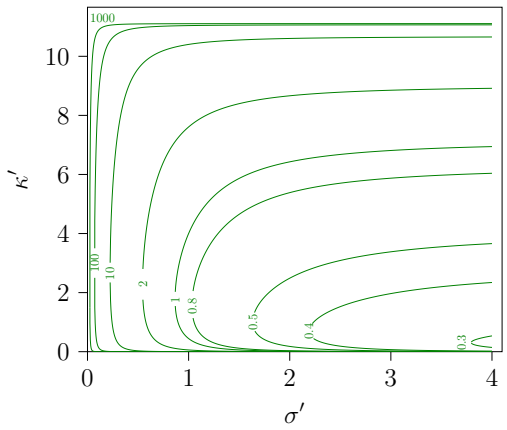


Figure 7.2.11: MMSE achieving curves for various values of β for $\sigma = 4, \kappa = 1$. For any set of (σ', κ') such that $0 < \kappa' < 100/9 = \kappa\frac{(1+\kappa\sigma^4)^2}{(1-\kappa\sigma^4)^2}$ there exists a $\beta > 0.2 = \frac{1}{1+\kappa\sigma^4}$ such that the mismatched MSE of the bracket estimator achieves the MMSE. For $\beta \to \infty$, $\kappa' = 100/9$ gives the MMSE.

are not sensitive to minor errors around the true value κ.

7.2.3 Mismatched estimation with $\beta \neq 1$

Introducing the auxiliary (temperature) parameter $\beta > 0$ to the (assumed) posterior distribution will add a degree of freedom to the problem. This leads to that for a region in the σ', κ' plane, for any pair of (σ', κ') there exists a β such that MMSE is achieved.The MMSE achieving curves for some values of β are plotted in Fig. 7.2.11. Note that for $\beta < \frac{1}{1+\kappa\sigma^4}$, the MSE is σ^4 for any pair of (σ', κ') and estimation better than random guess is not possible.

Similar to the case of $\beta = 1$, the MMSE curves can be obtained by either equating the mismatched MSE and MMSE, or enforcing the condition $m_{\text{stat}} = q_{\text{stat}}$. Moreover, computing the normalized matrix inner product of the estimate

with the true signal, we get

$$\frac{\operatorname{Tr} \boldsymbol{s}\boldsymbol{s}^\intercal \langle \boldsymbol{x}\boldsymbol{x}^\intercal \rangle_{\sigma',\kappa'}}{\|\langle \boldsymbol{x}\boldsymbol{x}^\intercal \rangle_{\sigma',\kappa'}\|_{\mathrm{F}} \|\boldsymbol{s}\boldsymbol{s}^\intercal\|_{\mathrm{F}}} = \frac{m_{\mathrm{stat}}}{\sqrt{q_{\mathrm{stat}}}\sigma^2} = 1 - \frac{1}{\kappa\sigma^4}$$

Therefore, in the case that $\kappa \geq \frac{1}{\sigma^4}$, $\sqrt{\kappa\kappa'} \geq \frac{1}{\beta\sigma^2\rho^*}$ the posterior mean always has the possible optimum angle with the ground truth, however due to the resacling factor q_{stat} we might get different MSEs. Therefore, the bracket estimator acts like the spectral algorithm with different rescaling q_{stat}. Equating q_{stat} with the optimum rescaling in the fully matched case $\left(\sigma^2 - \frac{1}{\kappa\sigma^2}\right)^2$ gives the MMSE curve.

Studying the fixed points of the SE equation (7.18) enables us to evaluate the performance of the AMP in the general case. It turns out that, similar to the case of $\beta = 1$, considering different conditions inserting the fixed points of SE equations $m_{\mathrm{AMP}}, q_{\mathrm{AMP}}$ into the eq. (7.13) leads to the same MSE as in theorem 7.1.

Moreover, the estimates of the AMP performed with the optimum β for a pair of (σ', κ') satisfies the Nishimori identity. Fig. 7.2.12 illustrates the performance of AMP and SE under various situations.

7.3 Bernoulli-Rademacher Prior

Assume s_i be i.i.d. random variables distributed according to the Bernoulli-Rademacher (BR) distribution with probabaility mass function defined as

$$p(s) = \begin{cases} \frac{\alpha}{2} & \text{if } s = \pm 1 \\ 1 - \alpha & \text{if } s = 0 \end{cases}$$

where $0 \leq \alpha \leq 1$ is the sparsity parameter. This prior is often adopted to model the sparse signals. The observed matrix is defined as in (7.1). The statistician assumes the Gaussian prior $\mathcal{N}(0, \sigma'^2)$ as the signal prior, and κ' as the SNR parameter. Since the assumed prior distribution is not the same as the true prior, we cannot be in the Bayes optimal case for any set of parameters. Therefore, we start this section with a short summary on the Bayes optimal scenario of inference with the BR prior distribution.

7.3.1 Bayes optimal estimation

From theorem 1 in [9] the matrix MMSE with the BR prior reads:

$$\mathrm{MMSE}_n = \alpha^2 - m^{*2}$$

where m^* is computed numerically through an involved recursion formula which we omit here. It turns out that, there is a phase transition at $\kappa = \frac{1}{\alpha^2}$ that for κ less than this value the MMSE is equal to α^2 and estimation better than

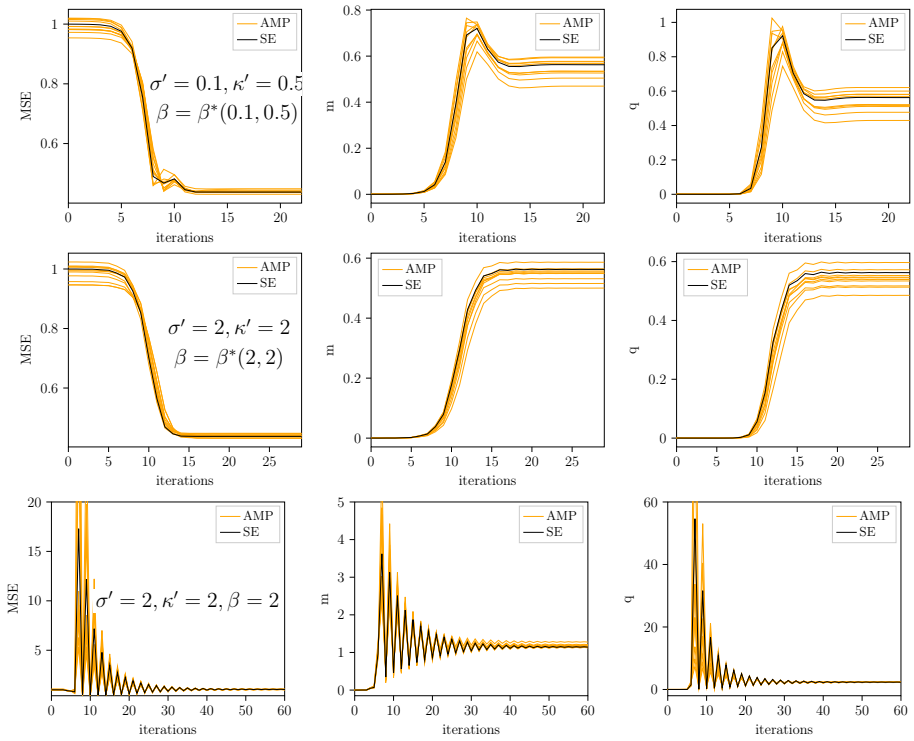


Figure 7.2.12: Performance of AMP in general case of β for 10 instances for $\sigma = 1, \kappa = 4, N = 10000$, compared with the SE predictions (7.18). In the first two rows, we applied AMP with the optimum β for the two pairs $(\sigma', \kappa') = (0.1, 0.5), (2, 2)$. The MSE of the final estimates is equal to the MMSE, and the Nishimori identity holds for the final estimates of the algorithm, although it does not hold through iterations. In the third row, AMP is applied with an arbitrary β for $\sigma' = 2, \kappa' = 2$. The MSE of the AMP and SE coincides with the theoretical one as in theorem 7.1, and MSE is greater than the MMSE. Furthermore, the Nishimori identity does not hold for the final estimates.

chance is not possible. However, for κ greater than this value the MMSE is less than α^2.

From an algorithmic point of view, AMP achieves the theoretical MMSE in the whole plane of (α, κ) except for a small region [110] which corresponds to the high sparsity regimes. In Fig. 7.3.1, we compare the performance of AMP (predicted by SE) and the spectral method (with optimum rescaling) for two values α. This figure illustrates how the knowledge of the prior distribution helps to improve on the spectral methods. Computing the best rescaling factor of the spectral is similar to the case of Gaussian prior by simply substituting σ^2 by α, see section 7.2.1. The MSE of the spectral method with the optimum rescaling factor is equal to the MMSE when the prior is Gaussian with $\sigma^2 = \alpha$.

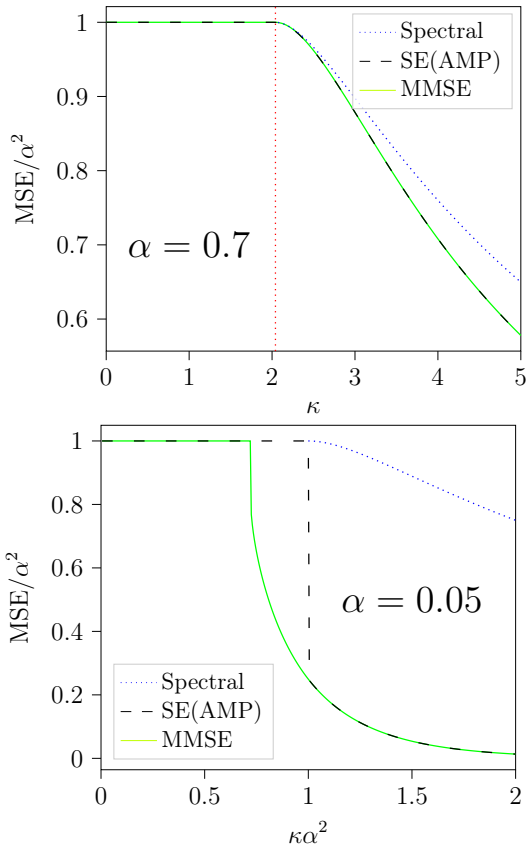


Figure 7.3.1: Comparison of MMSE, MSE of AMP (predicted by SE), and MSE of spectral method for the Bernoulli-Rademacher prior. In low sparsity regimes, the phase transition of MMSE is the same as predicted by spectral methods, $\kappa = \frac{1}{\alpha^2}$, the vertical dotted (red) line in the left plot, and AMP achieves the MMSE for all values of κ. In high sparsity regimes, $\alpha \ll 1$, the phase transition of MMSE is not continuous and AMP fails to achieve the MMSE in a small region around the phase transition point. In both cases, the MSE of spectral method is greater than the MMSE, which demonstrates how spectral method performs poorly comparing to the Bayesian approach with the knowledge of the prior distribution.

7.3.2 Mismatched estimation

Considering the posterior distribution as in (7.3) with $\mathbf{P}$ be the Gaussian distribution with variance σ'^2 leads to the following MSE in the N large limit.

Theorem 7.3. *Assume that the sequence* $(\mathrm{MSE})_{N\geq 1}$ *converges uniformly in* $(\kappa, \kappa', \beta) \in K \subset \mathbb{R}^3_+$, *then for For all* σ, σ' *(strictly positive) and* $(\kappa, \kappa', \beta) \in K$, *the asymptotic mismatched MSE is given :*

$$\lim_{N\to\infty} \mathrm{MSE}_N(\alpha,\sigma',\kappa,\kappa',\beta) = \begin{cases} \alpha^2 + \frac{\left[(1-2\beta)\sqrt{\kappa'}\sigma'^2+1\right]^2}{\beta^2\kappa'^2\sigma'^4} & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}, \kappa \le \frac{1}{\alpha^2}, \\ A & \text{if } \kappa \ge \frac{1}{\alpha^2}, \sqrt{\kappa\kappa'} \ge \frac{1}{\beta\alpha\rho^*}, \\ \sigma^4 & \text{o.w.} \end{cases} \tag{7.26}$$

where

$$A = \alpha^2\left(1-\sqrt{\frac{\kappa}{\kappa'}}\right)^2 + \frac{1}{\beta}\left[\frac{2}{\sqrt{\kappa\kappa'}} + \frac{1}{\beta}\frac{1}{\kappa'^2\sigma'^4} + \frac{2}{\kappa'}\frac{\alpha}{\sigma'^2}\left(1-\sqrt{\frac{\kappa}{\kappa'}}\right) - \frac{2}{\kappa\kappa'\alpha\sigma'^2}\right] + \left(1-\frac{1}{\beta}\right)\left[\frac{2}{\kappa'} - \frac{1}{\beta}\frac{2}{\sqrt{\kappa\kappa'^3}\alpha\sigma'^2} + \left(1-\frac{1}{\beta}\right)\frac{1}{\kappa\kappa'\alpha^2} + \frac{2}{\sqrt{\kappa^3\kappa'}\alpha^2}\right]$$

and

$$\rho^* = \begin{cases} \sigma'^2 & \text{if } \beta = 1, \\ \frac{1}{\kappa'\beta(\beta-1)}\left(\frac{1}{2\sigma'^2} - \sqrt{\frac{1}{4\sigma'^4} - \kappa'\beta(\beta-1)}\right) & \text{if } \beta \ne 1 \end{cases}$$

Proof. The proof follows similar steps as in Theorem 7.4. For details, see appendix 7.D. □

The expression in (7.26) is the same as the one in (7.16) with σ^2 replaced with α. Therefore, the discussion (and plots) on the mismatched estimation with Gaussian prior holds for the problem with Bernoulli-Rademacher prior with a major difference. Unlike the Gaussian case, the mismatched MSE (with BR prior) does not achieve the MMSE. In fact, the minimum of mismatched MSE is equal to the MSE of spectral method in the Bayes optimal scenario (equal to the MMSE of Gaussian prior) which is greater than the MMSE.

The AMP equations with the fake prior are the same as in (7.17) (note that in writing the AMP equations only the assumed prior matters). The SE equations are also similar to (7.18) with σ^2 replaced with α. Similar to the Gaussian prior, studying the fixed points of SE leads to the same MSE as in (7.3). Thus, there is no gap between the theoretical MSE and the MSE of AMP in the whole region, although they do not attain the MMSE in any cases. In Fig. 7.3.2, the performance of AMP and SE is depicted.

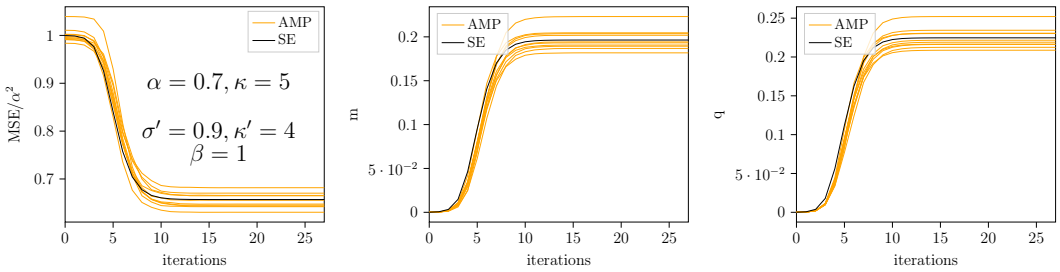


Figure 7.3.2: Performance of AMP in general case of β for 10 instances for $\sigma = 1, \kappa = 4, N = 10000$, compared with the SE predictions (7.18). In the first two rows, we applied AMP with the optimum β for the two pairs $(\sigma', \kappa') = (0.1, 0.5), (2, 2)$. The MSE of the final estimates is equal to the MMSE, and the Nishimori identity holds for the final estimates of the algorithm, although it does not hold through iterations. In the third row, AMP is applied with an arbitrary β for $\sigma' = 2, \kappa' = 2$. The MSE of the AMP and SE coincides with the theoretical one as in theorem 7.1, and MSE is greater than the MMSE. Furthermore, the Nishimori identity does not hold for the final estimates.

Appendix

7.A Proof of Lemma 7.1

From (7.5), (7.4) we have

$$f_N(\mathbf{P}^*, \mathbf{P}, \kappa, \kappa', \beta) = \\ -\frac{1}{N}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\Big[\ln \int d\boldsymbol{x}\, \mathbf{P}(\boldsymbol{x}) e^{-\frac{\beta\kappa'}{4N}\|\boldsymbol{x}\|^4 + \frac{\beta}{2}\frac{\sqrt{\kappa\kappa'}}{N}(\boldsymbol{s}^\intercal\boldsymbol{x})^2 + \frac{\beta}{2}\sqrt{\frac{\kappa'}{n}}\operatorname{Tr}\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^\intercal}\Big]$$

So,

$$\begin{aligned}
&\frac{d}{d\kappa} f_N \\
&= -\frac{1}{N}\mathbb{E}\left[\frac{\int d\boldsymbol{x}\, \mathbf{P}(\boldsymbol{x})\big[\frac{\beta}{4N}\sqrt{\frac{\kappa'}{\kappa}}(\boldsymbol{s}^\intercal\boldsymbol{x})^2\big] e^{-\frac{\beta\kappa'}{4N}\|\boldsymbol{x}\|^4 + \frac{\beta}{2}\frac{\sqrt{\kappa\kappa'}}{N}(\boldsymbol{s}^\intercal\boldsymbol{x})^2 + \frac{\beta}{2}\sqrt{\frac{\kappa'}{N}}\operatorname{Tr}\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^\intercal}}{\int d\boldsymbol{x}\, \mathbf{P}(\boldsymbol{x}) e^{-\frac{\beta\kappa'}{4N}\|\boldsymbol{x}\|^4 + \frac{\beta}{2}\frac{\sqrt{\kappa\kappa'}}{N}(\boldsymbol{s}^\intercal\boldsymbol{x})^2 + \frac{\beta}{2}\sqrt{\frac{\kappa'}{N}}\operatorname{Tr}\boldsymbol{Z}\boldsymbol{x}\boldsymbol{x}^\intercal}}\right] \\
&= -\frac{\beta}{4}\frac{1}{N^2}\sqrt{\frac{\kappa'}{\kappa}}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\Big[\big\langle(\boldsymbol{s}^\intercal\boldsymbol{x})^2\big\rangle_{\mathbf{P},\kappa',\beta}\Big]
\end{aligned} \tag{7.27}$$

and

$$\frac{d}{d\kappa'} f_N = -\frac{1}{N}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\Big[-\frac{\beta}{4n}\big\langle\|\boldsymbol{x}\|^4\big\rangle_{\mathbf{P},\kappa',\beta} + \frac{\beta}{4n}\sqrt{\frac{\kappa}{\kappa'}}\big\langle(\boldsymbol{s}^\intercal\boldsymbol{x})^2\big\rangle_{\mathbf{P},\kappa',\beta} \\ + \frac{\beta}{4}\frac{1}{\sqrt{\kappa' n}}\operatorname{Tr}\boldsymbol{Z}\langle\boldsymbol{x}\boldsymbol{x}^\intercal\rangle_{\mathbf{P},\kappa',\beta}\Big]$$

Using a standard Gaussian integration by parts trick (see lemma 5.2), we have

$$\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\big[\operatorname{Tr}\boldsymbol{Z}\langle\boldsymbol{x}\boldsymbol{x}^\intercal\rangle_{\mathbf{P},\kappa',\beta}\big] = \beta\sqrt{\frac{\kappa'}{n}}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\Big[\big\langle\|\boldsymbol{x}\|^4\big\rangle_{\mathbf{P},\kappa',\beta} - \big\|\langle\boldsymbol{x}\boldsymbol{x}^\intercal\rangle_{\mathbf{P},\kappa',\beta}\big\|_{\mathrm{F}}^2\Big]$$

Therefore,

$$\frac{d}{d\kappa'} f_N = \frac{\beta}{4}\frac{1}{N^2}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\Big[(1-\beta)\big\langle\|\boldsymbol{x}\|^4\big\rangle_{\mathbf{P},\kappa',\beta} + \beta\big\|\langle\boldsymbol{x}\boldsymbol{x}^\intercal\rangle_{\mathbf{P},\kappa',\beta}\big\|_{\mathrm{F}}^2 \\ - \sqrt{\frac{\kappa}{\kappa'}}\big\langle(\boldsymbol{s}^\intercal\boldsymbol{x})^2\big\rangle_{\mathbf{P},\kappa',\beta}\Big] \tag{7.28}$$

Again by Gaussian integration by parts trick, we get

$$\begin{aligned}\frac{d}{d\beta}f_N = \frac{1}{4}\frac{1}{N^2}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\Big[&\kappa'(1-2\beta)\big\langle\|\boldsymbol{x}\|^4\big\rangle_{\mathbf{P},\kappa',\beta} + 2\beta\kappa'\big\|\langle\boldsymbol{x}\boldsymbol{x}^\intercal\rangle_{\mathbf{P},\kappa',\beta}\big\|_{\mathrm{F}}^2 \\ &- 2\sqrt{\kappa\kappa'}\big\langle(\boldsymbol{s}^\intercal\boldsymbol{x})^2\big\rangle_{\mathbf{P},\kappa',\beta}\Big]\end{aligned} \tag{7.29}$$

Putting (7.27), (7.28), (7.29), with a bit of algebra we find that the left-hand side of (7.6) is equal to

$$\begin{aligned}&\frac{1}{4}\frac{1}{N^2}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\Big[\big\|\langle\boldsymbol{x}\boldsymbol{x}^\intercal\rangle_{\mathbf{P},\kappa',\beta}\big\|_{\mathrm{F}}^2 - 2\big\langle(\boldsymbol{s}^\intercal\boldsymbol{x})^2\big\rangle_{\mathbf{P},\kappa',\beta} + \|\boldsymbol{s}\|^4\Big] \\ &\qquad= \frac{1}{4}\frac{1}{N^2}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\Big[\big\|\langle\boldsymbol{x}\boldsymbol{x}^\intercal\rangle_{\mathbf{P},\kappa',\beta}\big\|_{\mathrm{F}}^2 - 2\operatorname{Tr}\boldsymbol{s}\boldsymbol{s}^\intercal\langle\boldsymbol{x}\boldsymbol{x}^\intercal\rangle_{\mathbf{P},\kappa',\beta} + \|\boldsymbol{s}\boldsymbol{s}^\intercal\|_{\mathrm{F}}^2\Big] \\ &\qquad= \frac{1}{4}\mathrm{MSE}_N \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square\end{aligned}$$

7.B Proof of Theorem 7.1

First, we compute the asymptotic of the free energy of the system.

Theorem 7.4. *For all $\sigma, \sigma', \kappa, \kappa', \beta$ positive, the asymptotic of free energy for the mismatched inference is given as:*

$$\begin{cases} \frac{1}{4}\frac{\rho^*}{\sigma'^2} - \frac{1}{2}\ln\rho^* - \frac{1}{4} + \ln\sigma' & \text{if } (2\beta-1)\sqrt{\kappa'} \le \frac{1}{\sigma'^2}, \kappa \le \frac{1}{\sigma^4}, \\ \frac{1}{2}\ln\beta + \frac{1}{4}\ln\kappa' + \ln\sigma' + \frac{1}{4} & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}, \kappa \le \frac{1}{\sigma^4}, \\ \quad -\frac{1}{4\beta\kappa'\sigma'^4} + \frac{1}{\sqrt{\kappa'}\sigma'^2} - \beta & \\ \frac{1}{4}\frac{\rho^*}{\sigma'^2} - \frac{1}{2}\ln\rho^* - \frac{1}{4} + \ln\sigma' & \text{if } (2\beta-1)\sqrt{\kappa'} \le \frac{1}{\sigma'^2}, \frac{1}{\sqrt{\kappa}} \le \sigma^2 \le \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}, \\ \frac{1-\beta}{4\kappa\sigma^4} - \frac{1}{4\beta\kappa'\sigma'^4} - \frac{1}{4}\beta\kappa\sigma^4 & \text{if } \kappa \ge \frac{1}{\sigma^4}, \sqrt{\kappa\kappa'} \ge \frac{1}{\beta\sigma^2\rho^*}, \\ \quad +\frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{\sigma^2}{\sigma'^2} + \frac{1}{2\sqrt{\kappa\kappa'}\sigma^2\sigma'^2} & \\ \quad +\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\sigma^2\sigma'^2 - \frac{\beta}{2} & \\ 0 & \text{o.w.} \end{cases} \tag{7.30}$$

where

$$\rho^* = \begin{cases} \sigma'^2 & \text{if } \beta = 1, \\ \frac{1}{\kappa'\beta(\beta-1)}\left(\frac{1}{2\sigma'^2} - \sqrt{\frac{1}{4\sigma'^4} - \kappa'\beta(\beta-1)}\right) & \text{if } \beta \ne 1 \end{cases}$$

Proof Sketch. We have

$$\begin{aligned} f_N &= -\frac{1}{N}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\big[\ln Z(\boldsymbol{Y})\big] \\ &= -\frac{1}{N}\int d\boldsymbol{s}\mathbf{P}^*(\boldsymbol{s})\mathbb{E}_{\mathbf{P}_{\boldsymbol{Z}}}\Big[\ln\int d\boldsymbol{x}\mathbf{P}(\boldsymbol{x})\, e^{\frac{-\beta\kappa'}{4N}\|\boldsymbol{x}\|^4 + \ln\mathcal{I}_N}\Big] \end{aligned} \tag{7.31}$$

where

$$\mathcal{I}_N = \int D\boldsymbol{U}\, e^{N\operatorname{Tr}\frac{\boldsymbol{Y}}{\sqrt{N}}\boldsymbol{U}\frac{\beta\sqrt{\kappa'}}{2N}\boldsymbol{x}\boldsymbol{x}^\intercal\boldsymbol{U}^\intercal} \tag{7.32}$$

It is not difficult to see that $\mathcal{I}_N$ is invariant under the transformation $\boldsymbol{x} \to R\boldsymbol{x}$ where R is a rotation matrix. Therefore the integrand in the $\boldsymbol{x}$-integral in (7.31) is a function of $\|\boldsymbol{x}\|$. Furthermore recalling $\boldsymbol{Y} = \sqrt{\frac{\kappa}{N}}\boldsymbol{s}\boldsymbol{s}^\intercal + \boldsymbol{Z}$ and using rotation invariance of $\mathbf{P}_{\boldsymbol{Z}}$ we see that the integrand of the $\boldsymbol{s}$-integral is a function of $\|\boldsymbol{s}\|$. Therefore we can use spherical coordinates (see appendix 7.E) to reduce the integrals in (7.31) to two one-dimensional integrals which yields

$$f_N = -\frac{2^{-\frac{N}{2}+1}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_0^{+\infty} dr\, r^{N-1} e^{-\frac{r^2}{2\sigma^2}} \times \mathbb{E}_{\mathbf{P}_{\boldsymbol{Z}}}\left[\frac{1}{N}\ln\left\{\frac{2^{-\frac{N}{2}+1}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int_0^{+\infty} d\rho\, \rho^{N-1} e^{-\frac{\rho^2}{2\sigma'^2}-\frac{\beta\kappa'}{4N}\rho^4+\ln\mathcal{I}_N}\right\}\right] \tag{7.33}$$

where $r := \|\boldsymbol{s}\|$, $\rho := \|\boldsymbol{x}\|$, and $\Gamma(.)$ is the *Gamma* function.

Changing variable $\frac{r^2}{N} \to r$, $\frac{\rho^2}{N} \to \rho$, we obtain

$$f_N = -\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_0^{+\infty} dr\, r^{\frac{N}{2}-1} e^{-N\frac{r}{2\sigma^2}} \times \mathbb{E}_{\mathbf{P}_{\boldsymbol{Z}}}\left[\frac{1}{N}\ln\left\{\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int_0^{+\infty} d\rho\, \rho^{\frac{N}{2}-1} e^{-N(\frac{\rho}{2\sigma'^2}+\frac{\beta\kappa'}{4}\rho^2-\frac{1}{N}\ln\mathcal{I}_N}\right\}\right] \tag{7.34}$$

To compute the asymptotic of the above integral, first we need to find the asymptotic of $\frac{1}{N}\ln\mathcal{I}_N$, then we can find the asymptotic of the expression in the $\mathbb{E}_{\mathbf{P}_{\boldsymbol{Z}}}$ by applying Laplace method. This asymptotic is independent of $\boldsymbol{Z}$, finally the asymptotics of f_N in (7.34) can be computed using the Laplace method again. □

Once we have the expression for the free energy, we can compute the MSE using Lemma 7.1. As explained in remark 7.3 this step uses the assumption that for $(\kappa, \kappa') \in K \subset \mathbb{R}_+^2$ the sequence $(\text{MSE})_{N\geq 1}$ converges uniformly.

In the following we explain the steps of the proof of theorem 7.4.

7.B.1 Computation of $\frac{1}{N}\ln\mathcal{I}_N$

With the change of variables $\frac{\|\boldsymbol{s}\|^2}{n} = \frac{r^2}{n} \to r$, $\frac{\|\boldsymbol{x}\|^2}{n} = \frac{\rho^2}{n} \to \rho$, the maximum and minimum eigenvalues of $\boldsymbol{Y}/\sqrt{N}$ are:

$$\lambda_{\min} = -2, \lambda_{\max} = \begin{cases} 2 & \text{if } \sqrt{\kappa}r \leq 1 \\ \sqrt{\kappa}r + \frac{1}{\sqrt{\kappa}r} & \text{if } \sqrt{\kappa}r \geq 1 \end{cases}$$

So,

$$\mathcal{G}_{\min} = -1, \mathcal{G}_{\max} = \begin{cases} 1 & \text{if } \sqrt{\kappa}r \leq 1 \\ \frac{1}{\sqrt{\kappa}r} & \text{if } \sqrt{\kappa}r \geq 1 \end{cases}$$

On the other hand the non-zero eigenvalue of the rank-one matrix $\frac{\beta\sqrt{\kappa'}}{2n}\boldsymbol{x}\boldsymbol{x}^\intercal$ is $\theta = \frac{\beta\sqrt{\kappa'}}{2N}\|\boldsymbol{x}\|^2 = \frac{\beta\sqrt{\kappa'}}{2}\rho$.

From Theorem 6.1, we have for N large enough

$$\begin{aligned}\frac{1}{N}\ln\mathcal{I}_N &= \theta\nu(\theta) - \frac{1}{2}\int \ln(1+2\theta\nu(\theta)-2\theta t)\, d\mu_{\mathrm{SC}}(t) + o(N)\\ &= \theta\nu(\theta) - \frac{1}{2}\int \ln(1+2\theta\nu(\theta)-2\theta t)\, \frac{1}{2\pi}\sqrt{4-t^2}dt + o(N)\end{aligned}$$

where

$$\nu(\theta) = \begin{cases} 2\theta & \text{if } \mathcal{G}_{\min} \le 2\theta \le \mathcal{G}_{\max} \\ \lambda_{\max} - \frac{1}{2\theta} & \text{if } 2\theta > \mathcal{G}_{\max} \\ \lambda_{\min} - \frac{1}{2\theta} & \text{if } 2\theta < \mathcal{G}_{\min} \end{cases}$$

Lemma 7.2. *For $A, B \in \mathbb{R}$, such that $A - Bx > 0$ for $-2 \le x \le 2$,*

$$\frac{1}{2\pi}\int_{-2}^{2} \ln(A-Bx)\sqrt{4-x^2}dx = \frac{A}{A+\sqrt{A^2-4B^2}} + \ln\left(A+\sqrt{A^2-4B^2}\right) - \frac{1}{2} - \ln 2$$

Using this lemma, we find the asymptotic of $\frac{1}{N}\ln\mathcal{I}_N$ under different conditions:

$$\begin{cases} \frac{\beta^2\kappa'}{4}\rho^2 & \text{for } \sqrt{\kappa}r \le 1, \beta\sqrt{\kappa'}\rho \le 1 \\ \beta\sqrt{\kappa'}\rho - \frac{1}{2}\ln\rho - \frac{1}{2}\ln\beta - \frac{1}{4}\ln\kappa' - \frac{3}{4} & \text{for } \sqrt{\kappa}r \le 1, \beta\sqrt{\kappa'}\rho > 1 \\ \frac{\beta^2\kappa'}{4}\rho^2 & \text{for } \sqrt{\kappa}r > 1, \beta\sqrt{\kappa'}\rho \le \frac{1}{\sqrt{\kappa}r} \\ \frac{\beta\sqrt{\kappa\kappa'}}{2}r\rho + \frac{1}{2}\beta\sqrt{\frac{\kappa'}{\kappa}}\frac{\rho}{r} - \frac{1}{2}\ln\rho - \frac{1}{4\kappa r^2} & \text{for } \sqrt{\kappa}r > 1, \beta\sqrt{\kappa'}\rho > \frac{1}{\sqrt{\kappa}r} \\ \quad -\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r - \frac{1}{2} & \end{cases} \tag{7.35}$$

7.B.2 Laplace method

We want to compute the limit in eq. (7.34) using Laplace method, however since we are dealing with nested integrals, we need a generalization of Laplace method. We apply the following theorem from [111] to find the asymptotics of the integrals we are interested in.

Theorem 7.5 (Olver [111])**.** *Let k, N be fixed positive numbers, and*

$$J_n = \int_0^k e^{-np(x)+t(n,x)}q(n,x)dx$$

Assume that : (i) $p'(x)$ is continuous and positive in $(0,k]$, and as $x \to 0^+$

$$p(x) = p(0) + Px^\mu + O(x^{\mu_1}), \quad p'(x) = \mu P x^{\mu-1} + O(x^{\mu_1-1})$$

(ii) For $n \ge N$, the real (or complex) functions $t(n,x), q(n,x)$ are continuous in $(0,k]$. Moreover,

$$|t(n,x)| \le Tn^\alpha x^\nu, \quad |q(n,x) - Qx^{\lambda-1}| \le Q_1 n^\beta x^{\lambda_1-1}$$

where $T, \alpha, \nu, Q, \lambda, Q_1, \beta$ *and* λ_1 *are independent of* n, x, *and*

$$\nu \geq 0, \;\; \lambda > 0, \;\; \lambda_1 > 0, \;\; \alpha < \min(1, \nu/\mu), \;\; \beta < (\lambda_1 - \lambda)/\mu$$

Then,

$$J_n = \frac{Q}{\mu}\Gamma(\frac{\lambda}{\mu})\frac{e^{-np(0)}}{(Pn)^{\lambda/\mu}}\big[1 + O(\frac{1}{n^{\omega/\mu}})\big]$$

where

$$\omega = \min(\mu_1 - \mu, \nu - \mu\alpha, \lambda_1 - \lambda - \beta)$$

Remark 7.7. *If* $x = 0$ *is the unique minimizer of* $p(x)$ *in the interval* $(0, +\infty)$, *then the same result holds for* $\int_0^{+\infty}$, *since the contribution of* $\int_k^{+\infty}$ *is exponentially negligible, for* k *a positive number.*

Remark 7.8. *If the minimizer* ($x = 0$) *in the interior of the integral domain, then the asymptotic can be evaluated by splitting the integral to two integrals* $\int_{-l}^{0}, \int_0^k$. *And, the first can be tuned to the conditions of the Theorem 7.5 by change of variables, and the final result is multiplied by two.*

7.B.3 Details of applying the Laplace method

To compute the limit in eq.(7.34), we need to apply the Laplace method twice, first to the interior integral (the integral over ρ), then to the integral over r.

For the first integral $q(N, \rho) = 1$, so we have that $Q = 1, \beta = 0, \kappa = 1, \kappa_1 = 2$. The term $t(N, \rho)$ in this integral comes from the error in the asymptotic of the $\frac{1}{N}\ln \mathcal{I}_N$, which is of the order of $o(N)$, so $\alpha = 0$. By Theorem 4 in [93], and Theorem 1 in [112], we can set $\nu = 1$. As it turns out from the functions we are dealing with in the exponent, we have $\mu_1 - \mu = 1$, thus the error of the asymptotic we find for the interior integral is of the order of $\frac{1}{N}$. Note that, since we have the (natural) logarithm of the interior integral, we will only consider the exponential term in the asymptotic, and drop other factors for simplicity.

For the integral over r, $t(N, r)$ is zero, so $\alpha = 0$, $\nu = 1$. The term $q(N, r)$ comes from the interior integral which is of the form $-\frac{1}{N}\ln C_N e^{-Np(r)}(1 + O(1/N)) = p(r) - \frac{1}{N}\ln C_N(1 + O(1/N))$, where C_N is a polynomial factor in N. For the first term, we can apply the regular Laplace method. For the second term, we can set $Q = 0$ in theorem 7.5 which results that the integral with the second term is asymptotically zero.

Based on the conditions on in eq. (7.35), we split the integral over different regions, and substitute the asymptotic of $\frac{1}{N}\ln \mathcal{I}_N$.

1) $r \leq \frac{1}{\sqrt{\kappa}}$

- $\rho \leq \frac{1}{\beta\sqrt{\kappa'}}$

$$\int_0^{\frac{1}{\beta\sqrt{\kappa'}}} \frac{d\rho}{\rho} e^{-N(\frac{\rho}{2\sigma'^2} + \beta\frac{\kappa'}{4}\rho^2 - \frac{1}{2}\ln\rho - \frac{\beta^2\kappa'}{4}\rho^2)} = \int_0^{\frac{1}{\beta\sqrt{\kappa'}}} \frac{d\rho}{\rho} e^{-N\left[\rho^2\frac{\kappa'}{4}\beta(1-\beta) + \frac{\rho}{2\sigma'^2} - \frac{1}{2}\ln\rho\right]}$$

The function in the exponent is $p_1(\rho) = \rho^2 \frac{\kappa'}{4}\beta(1-\beta) + \frac{\rho}{2\sigma'^2} - \frac{1}{2}\ln\rho$. First note that this function is undefined at $\rho = 0$, so by splitting the integral into two integral $\int_0^{\epsilon}, \int_{\epsilon}^{\frac{1}{\beta\sqrt{\kappa'}}}$, the contribution of the first part is exponentially negligible, thus we need to compute the the second one. Nevertheless, we will keep the integral $\int_0^{\frac{1}{\beta\sqrt{\kappa'}}}$ in the following.

If $\beta < 1$, it is straightforward to check that the minimum of $p_1(\rho)$ is attained at

$$\rho^* = \frac{1}{\kappa'\beta(\beta-1)}\left(\frac{1}{2\sigma'^2} - \sqrt{\frac{1}{4\sigma'^4} - \kappa'\beta(\beta-1)}\right)$$

And we have that $(2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2} \Leftrightarrow \rho^* \leq \frac{1}{\beta\sqrt{\kappa'}}$.

For $\beta > 1$, it can be verified that if $(2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}$, then $p_1'(\rho) \equiv 0$ has two solutions, one corresponding to a local minima denoted by ρ^* and one to a local maxima denoted by $\hat{\rho}$. We have

$$(2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2} \Rightarrow \rho^* \leq \frac{1}{\beta\sqrt{\kappa'}} < \hat{\rho}$$

On the other hand, if $(2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}$ there are two cases. First, $p_1'(\rho) \equiv 0$ has two solutions, but in this case $\rho^* > \frac{1}{\beta\sqrt{\kappa'}}$. Second, $p_1'(\rho) \equiv 0$ has no solutions, in this case $p_1(.)$ is monotonically decreasing. Therefore, in either case the minimum of $p_1(\rho)$ in the interval $(0, \frac{1}{\beta\sqrt{\kappa'}}]$ is attained at $\frac{1}{\beta\sqrt{\kappa'}}$. Putting all together, we find

$$\begin{aligned}\int_0^{\frac{1}{\beta\sqrt{\kappa'}}} \frac{d\rho}{\rho} e^{-N\left[\rho^2\frac{\kappa'}{4}\beta(1-\beta)+\frac{\rho}{2\sigma'^2}-\frac{1}{2}\ln\rho\right]} &\approx \begin{cases} e^{-Np_1(\rho^*)} & \text{if } (2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2} \\ e^{-Np_1(\frac{1}{\beta\sqrt{\kappa'}})} & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2} \end{cases} \\ &= \begin{cases} e^{-N(\frac{1}{4}\frac{\rho^*}{\sigma'^2}-\frac{1}{2}\ln\rho^*+\frac{1}{4})} & \text{if } (2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2} \\ e^{-N\left(-\frac{1}{4}(1-\frac{1}{\beta})+\frac{1}{2\beta}\frac{1}{\sqrt{\kappa'}\sigma'^2}+\frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'\right)} & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2} \end{cases}\end{aligned} \tag{7.36}$$

- $\rho > \frac{1}{\beta\sqrt{\kappa'}}$

$$\begin{aligned}&\int_{\frac{1}{\beta\sqrt{\kappa'}}}^{\infty} \frac{d\rho}{\rho} e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\beta\sqrt{\kappa'}\rho+\frac{1}{2}\ln\rho+\frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'+\frac{3}{4})} \\ &\qquad = e^{-N(\frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'+\frac{3}{4})} \int_{\frac{1}{\beta\sqrt{\kappa'}}}^{\infty} \frac{d\rho}{\rho} e^{-N\left[\beta\frac{\kappa'}{4}\rho^2+(\frac{1}{2\sigma'^2}-\beta\sqrt{\kappa'})\rho\right]}\end{aligned}$$

The minima of the function in the exponent is at $\rho = \frac{2}{\kappa'} - \frac{1}{\beta\kappa'\sigma'^2}$. It can be verified that

$$(2\beta-1)\sqrt{\kappa'} < \frac{1}{\sigma'^2} \Leftrightarrow \frac{2}{\kappa'} - \frac{1}{\beta\kappa'\sigma'^2} < \frac{1}{\beta\sqrt{\kappa'}}$$

Therefore, we get

$$
\begin{aligned}
&e^{-N(\frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'+\frac{3}{4})}\int_{\frac{1}{\beta\sqrt{\kappa'}}}^{\infty}\frac{d\rho}{\rho}e^{-N\left[\beta\frac{\kappa'}{4}\rho^2+(\frac{1}{2\sigma'^2}-\beta\sqrt{\kappa'})\rho\right]}\\
&\quad\approx e^{-N(\frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'+\frac{3}{4})}\begin{cases} e^{-Np_2(\frac{1}{\beta\sqrt{\kappa'}})} & \text{if } (2\beta-1)\sqrt{\kappa'}<\frac{1}{\sigma'^2}\\ e^{-Np_2(\rho^{\mathrm{opt}})} & \text{if } (2\beta-1)\sqrt{\kappa'}\geq\frac{1}{\sigma'^2}\end{cases}\\
&\quad=\begin{cases} e^{-N(\frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'-\frac{1}{4}+\frac{1}{4\beta}+\frac{1}{2\beta\sqrt{\kappa'}\sigma'^2})} & \text{if } (2\beta-1)\sqrt{\kappa'}\leq\frac{1}{\sigma'^2}\\ e^{-N(\frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'+\frac{3}{4}-\frac{1}{4\beta\kappa'\sigma'^4}+\frac{1}{\sqrt{\kappa'}\sigma'^2}-\beta)} & \text{if } (2\beta-1)\sqrt{\kappa'}>\frac{1}{\sigma'^2}\end{cases}
\end{aligned}
\tag{7.37}
$$

We need to sum the asympotics in eq. (7.36),(7.37) to find the asymptotic of $\int_0^\infty d\rho$. Since, we are considering the n large limit, we pick the dominant term under each condition. For the case $(2\beta-1)\sqrt{\kappa'}\leq\frac{1}{\sigma'^2}$ the term form eq. (7.36) is dominant, and for the case $(2\beta-1)\sqrt{\kappa'}>\frac{1}{\sigma'^2}$ the term in eq.(7.37) is dominant. Note that in both eq. (7.36),(7.37), when $(2\beta-1)\sqrt{\kappa'}=\frac{1}{\sigma'^2}$ the exponent in both cases is equal, so we can properly move the equality case of the condition from one part to the other.

We have

$$
\begin{aligned}
&\int_0^\infty\frac{d\rho}{\rho}e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{n}\ln\mathcal{I}_N)}\\
&\quad\approx\begin{cases} e^{-N(\frac{1}{4}\frac{\rho^*}{\sigma'^2}-\frac{1}{2}\ln\rho^*+\frac{1}{4})} & \text{if } (2\beta-1)\sqrt{\kappa'}\leq\frac{1}{\sigma'^2}\\ e^{-N(\frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'+\frac{3}{4}-\frac{1}{4\beta\kappa'\sigma'^4}+\frac{1}{\sqrt{\kappa'}\sigma'^2}-\beta)} & \text{if } (2\beta-1)\sqrt{\kappa'}>\frac{1}{\sigma'^2}\end{cases}
\end{aligned}
\tag{7.38}
$$

Now, we need to compute the asymptotic of

$$
\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int_0^\infty\frac{d\rho}{\rho}e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}
$$

To do so, we use the following lemma from [113] which states the asymptotic of $\Gamma(\frac{n}{2})$.

Lemma 7.3. *For $x\to\infty$,*

$$
\Gamma(x+1)=\sqrt{2\pi x}(\frac{x}{e})^x(1+O(\frac{1}{x}))
$$

Using this lemma, we have

$$
\Gamma(\frac{N}{2})\approx\sqrt{2\pi(\frac{N}{2}-1)}(\frac{N}{2}-1)^{\frac{N}{2}-1}e^{-\frac{N}{2}+1}\approx\sqrt{\pi N}(\frac{N}{2})^{\frac{N}{2}-1}e^{-\frac{N}{2}}
$$

Substituting $\Gamma(\frac{N}{2})$ with its asymptotic results in

$$\begin{aligned}
&\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int_0^\infty \frac{d\rho}{\rho}e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)} \\
&\qquad\approx \frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\sqrt{\pi N}(\frac{N}{2})^{\frac{N}{2}-1}e^{-\frac{N}{2}}}\frac{1}{\sigma'^N}\int_0^\infty \frac{d\rho}{\rho}e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)} \\
&\qquad\approx C_N e^{-N(-\frac{1}{2}+\ln\sigma')}\int_0^\infty \frac{d\rho}{\rho}e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)} \\
&\approx \begin{cases} C_N e^{-N(\frac{1}{4}\frac{\rho^*}{\sigma'^2}-\frac{1}{2}\ln\rho^*-\frac{1}{4}+\ln\sigma')} & \text{if } (2\beta-1)\sqrt{\kappa'}\le\frac{1}{\sigma'^2} \\ C_N e^{-N(\frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'+\ln\sigma'+\frac{1}{4}-\frac{1}{4\beta\kappa'\sigma'^4}+\frac{1}{\sqrt{\kappa'}\sigma'^2}-\beta)} & \text{if } (2\beta-1)\sqrt{\kappa'}>\frac{1}{\sigma'^2}\end{cases}
\end{aligned}$$

where C_N is a polynomial factor in N. This leads us to

$$\begin{aligned}
&-\frac{1}{N}\ln\left[\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int_0^\infty \frac{d\rho}{\rho}e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}\right] \\
&\approx \begin{cases} \frac{1}{4}\frac{\rho^*}{\sigma'^2}-\frac{1}{2}\ln\rho^*-\frac{1}{4}+\ln\sigma' & \text{if } (2\beta-1)\sqrt{\kappa'}\le\frac{1}{\sigma'^2} \\ \frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'+\ln\sigma'+\frac{1}{4}-\frac{1}{4\beta\kappa'\sigma'^4}+\frac{1}{\sqrt{\kappa'}\sigma'^2}-\beta & \text{if } (2\beta-1)\sqrt{\kappa'}>\frac{1}{\sigma'^2}\end{cases}
\end{aligned}$$

Define

$$q=\begin{cases} \frac{1}{4}\frac{\rho^*}{\sigma'^2}-\frac{1}{2}\ln\rho^*-\frac{1}{4}+\ln\sigma' & \text{if } (2\beta-1)\sqrt{\kappa'}\le\frac{1}{\sigma'^2} \\ \frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'+\ln\sigma'+\frac{1}{4}-\frac{1}{4\beta\kappa'\sigma'^4}+\frac{1}{\sqrt{\kappa'}\sigma'^2}-\beta & \text{if } (2\beta-1)\sqrt{\kappa'}>\frac{1}{\sigma'^2}\end{cases} \tag{7.39}$$

Thus, in this case the function $q(N,r)$ is independent of r. It remains to compute the the following limit:

$$\lim_{N\to\infty}\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_0^{\frac{1}{\sqrt{\kappa}}} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}}q$$

Applying the regular Laplace method to the integral over r, we get

$$\int_0^{\frac{1}{\sqrt{\kappa}}} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}}=\begin{cases} 2\sqrt{\frac{\pi}{N}}e^{-N(\frac{1}{2}-\ln\sigma)}\big(1+O(\frac{1}{N})\big) & \text{if } \kappa\le\frac{1}{\sigma^4} \\ \frac{2\sqrt{\kappa}}{\sqrt{\kappa}-\frac{1}{\sigma^2}}\frac{1}{N}e^{-N(\frac{1}{2\sqrt{\kappa}\sigma^2}+\frac{1}{4}\ln\kappa)}\big(1+O(\frac{1}{N})\big) & \text{if } \kappa>\frac{1}{\sigma^4}\end{cases}$$

Further, we have

$$\begin{aligned}
\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_0^{\frac{1}{\sqrt{\kappa}}} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}} &\approx \frac{1}{2}\sqrt{\frac{N}{\pi}}e^{-N(-\frac{1}{2}+\ln\sigma)}\int_0^{\frac{1}{\sqrt{\kappa}}} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}} \\
&\approx \begin{cases} 1 & \text{if } \kappa\le\frac{1}{\sigma^4} \\ C\frac{1}{\sqrt{N}}e^{-N(\frac{1}{2\sqrt{\kappa}\sigma^2}+\ln\kappa^{\frac{1}{4}}\sigma-\frac{1}{2})} & \text{if } \kappa>\frac{1}{\sigma^4}\end{cases}
\end{aligned}$$

where C is some constant.

Since $\frac{1}{x} + \ln x - 1$ is positive for $x \geq 1$, we find

$$\lim_{N\to\infty} \frac{2^{-\frac{N}{2}} N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})} \frac{1}{\sigma^N} \int_0^{\frac{1}{\sqrt{\kappa}}} dr\, r^{\frac{N}{2}-1} e^{-N\frac{r}{2\sigma^2}} q = \begin{cases} q & \text{if } \kappa \leq \frac{1}{\sigma^4} \\ 0 & \text{if } \kappa > \frac{1}{\sigma^4} \end{cases}$$

Substituting q from eq.(7.39), we obtain

$$\begin{aligned} &\lim_{N\to\infty} \frac{2^{-\frac{N}{2}} N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})} \frac{1}{\sigma^N} \int_0^{\frac{1}{\sqrt{\kappa}}} dr\, r^{\frac{N}{2}-1} e^{-N\frac{r}{2\sigma^2}} \\ &\quad \times \left[-\frac{1}{N} \ln \left(\frac{2^{-\frac{N}{2}} N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})} \frac{1}{\sigma'^N} \int_0^\infty \frac{d\rho}{\rho} e^{-N(\frac{\rho}{2\sigma'^2} + \beta\frac{\kappa'}{4}\rho^2 - \frac{1}{2}\ln\rho - \frac{1}{N}\ln \mathcal{I}_N)} \right) \right] \\ &= \begin{cases} \frac{1}{4}\frac{\rho^*}{\sigma'^2} - \frac{1}{2}\ln\rho^* - \frac{1}{4} + \ln\sigma' & \text{if } (2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}, \kappa \leq \frac{1}{\sigma^4} \\ \frac{1}{2}\ln\beta + \frac{1}{4}\ln\kappa' + \ln\sigma' + \frac{1}{4} & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}, \kappa \leq \frac{1}{\sigma^4}, \\ \quad -\frac{1}{4\beta\kappa'\sigma'^4} + \frac{1}{\sqrt{\kappa'}\sigma'^2} - \beta & \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{7.40}$$

To compute the remaining part of integral, we need to consider two cases.

Throughout the following, suppose $(2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}$. This implies that $p_1'(\rho) \equiv 0$ has two solutions $\rho^*, \hat{\rho}$.

2) $\frac{1}{\sqrt{\kappa}} < r \leq \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}$

- $\rho \leq \frac{1}{\beta\sqrt{\kappa\kappa'}r}$

$$\begin{aligned} &\int_0^{\frac{1}{\beta\sqrt{\kappa\kappa'}r}} \frac{d\rho}{\rho} e^{-N(\frac{\rho}{2\sigma'^2} + \beta\frac{\kappa'}{4}\rho^2 - \frac{1}{2}\ln\rho - \frac{\beta^2\kappa'}{4}\rho^2)} \\ &\qquad = \int_0^{\frac{1}{\beta\sqrt{\kappa\kappa'}r}} \frac{d\rho}{\rho} e^{-N\left[\rho^2\frac{\kappa'}{4}\beta(1-\beta) + \frac{\rho}{2\sigma'^2} - \frac{1}{2}\ln\rho\right]} \end{aligned}$$

In both cases $\beta < 1$ and $\beta > 1$, $p_1'(\rho) = 0$ has solution and p is minimized at ρ^*. We are in the region where $r \leq \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}$, so ρ^* is in the integral domain, and we get

$$\int_0^{\frac{1}{\beta\sqrt{\kappa\kappa'}r}} \frac{d\rho}{\rho} e^{-N\left[\rho^2\frac{\kappa'}{4}\beta(1-\beta) + \frac{\rho}{2\sigma'^2} - \frac{1}{2}\ln\rho\right]} \approx e^{-N(\frac{1}{4}\frac{\rho^*}{\sigma'^2} - \frac{1}{2}\ln\rho^* + \frac{1}{4})} \tag{7.41}$$

- $\rho > \frac{1}{\beta\sqrt{\kappa\kappa'}r}$

$$\begin{aligned} &\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}^\infty \frac{d\rho}{\rho} e^{-N(\frac{\rho}{2\sigma'^2} + \beta\frac{\kappa'}{4}\rho^2 - \frac{1}{2}\ln\rho - \beta\frac{\sqrt{\kappa\kappa'}}{2}r\rho - \frac{1}{2}\beta\sqrt{\frac{\kappa'}{\kappa}}\frac{\rho}{r} + \frac{1}{2}\ln\rho + \frac{1}{4\kappa r^2} + \frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r + \frac{1}{2})} \\ &= e^{-N(\frac{1}{4\kappa r^2} + \frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r + \frac{1}{2})} \int_{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}^\infty \frac{d\rho}{\rho} e^{-N\left[\beta\frac{\kappa'}{4}\rho^2 + (\frac{1}{2\sigma'^2} - \beta\frac{\sqrt{\kappa\kappa'}r}{2} - \beta\frac{1}{2}\sqrt{\frac{\kappa'}{\kappa}}\frac{1}{r})\rho\right]} \end{aligned}$$

The function in the exponent is $p_3(\rho) = \beta\frac{\kappa'}{4}\rho^2 + (\frac{1}{2\sigma'^2} - \beta\frac{\sqrt{\kappa\kappa'}r}{2} - \beta\frac{1}{2}\sqrt{\frac{\kappa'}{\kappa}}\frac{1}{r})\rho$, whose minimum is at $\rho^{\text{opt}} = -\frac{1}{\beta\kappa'\sigma'^2} + \sqrt{\frac{\kappa}{\kappa'}}r + \frac{1}{\sqrt{\kappa\kappa'}r}$. It can be checked that, in this region $\rho^{\text{opt}} < \frac{1}{\beta\sqrt{\kappa\kappa'}r}$. Thus, $p_3(.)$ is minimized at $\frac{1}{\beta\sqrt{\kappa\kappa'}r}$ (in the integral domain). So, we get

$$\begin{aligned} e^{-N(\frac{1}{4\kappa r^2}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r+\frac{1}{2})} \int_{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}^{\infty} \frac{d\rho}{\rho} e^{-N\left[\beta\frac{\kappa'}{4}\rho^2+(\frac{1}{2\sigma'^2}-\beta\frac{\sqrt{\kappa\kappa'}r}{2}-\beta\frac{1}{2}\sqrt{\frac{\kappa'}{\kappa}}\frac{1}{r})\rho\right]} \\ \approx e^{-N(\frac{1}{4\kappa r^2}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r+\frac{1}{2})} e^{-np_3(\frac{1}{\beta\sqrt{\kappa\kappa'}r})} \\ = e^{-N\left[-\frac{1}{4\kappa r^2}(1-\frac{1}{\beta})+\frac{1}{2\beta\sqrt{\kappa\kappa'}\sigma'^2 r}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r\right]} \end{aligned} \tag{7.42}$$

Comparing eq. (7.41),(7.42), we obtain

$$\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}^{\infty} \frac{d\rho}{\rho}\, e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)} \approx e^{-N(\frac{1}{4}\frac{\rho^*}{\sigma'^2}-\frac{1}{2}\ln\rho^*+\frac{1}{4})}$$

which leads to

$$\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int \frac{d\rho}{\rho}\, e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)} \approx e^{-N(\frac{1}{4}\frac{\rho^*}{\sigma'^2}-\frac{1}{2}\ln\rho^*-\frac{1}{4}+\ln\sigma')}$$

$$\begin{aligned} q &= -\frac{1}{N}\ln\left[\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int\frac{d\rho}{\rho}\, e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}\right] \\ &\approx \frac{1}{4}\frac{\rho^*}{\sigma'^2} - \frac{1}{2}\ln\rho^* - \frac{1}{4} + \ln\sigma' \end{aligned} \tag{7.43}$$

It remains to compute the limit

$$\lim_{N\to\infty} \frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_{\frac{1}{\sqrt{\kappa}}}^{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}}q$$

Applying the regular Laplace method to the integral over r, we get

$$\begin{aligned} &\int_{\frac{1}{\sqrt{\kappa}}}^{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}} \\ &\qquad\approx \begin{cases} 2\sqrt{\frac{\pi}{N}}e^{-N(\frac{1}{2}-\ln\sigma)} & \text{if } \frac{1}{\sqrt{\kappa}} \le \alpha \le \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ \frac{1}{N}e^{-N(\frac{1}{2\sqrt{\kappa}\sigma^2}+\frac{1}{4}\ln\kappa)} & \text{if } \kappa < \frac{1}{\sigma^4} \\ \frac{1}{N}e^{-N(\frac{1}{2\beta\sqrt{\kappa\kappa'}\sigma^2\rho^*}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\rho^*)} & \text{if } \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} < \sigma^2 \end{cases} \end{aligned}$$

Further, we have

$$\begin{aligned}
&\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_{\frac{1}{\sqrt{\kappa}}}^{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}} \\
&\qquad\approx \frac{1}{2}\sqrt{\frac{N}{\pi}}e^{-N(-\frac{1}{2}+\ln\sigma)}\int_{\frac{1}{\sqrt{\kappa}}}^{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}} \\
&\qquad\approx \begin{cases} 1 & \text{if } \frac{1}{\sqrt{\kappa}} \le \sigma^2 \le \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ Ce^{-N(\frac{1}{2\sqrt{\kappa}\sigma^2}+\frac{1}{2}\ln\sqrt{\kappa}\sigma^2-\frac{1}{2})} & \text{if } \kappa < \frac{1}{\sigma^4} \\ Ce^{-N(\frac{1}{2\beta\sqrt{\kappa\kappa'}\sigma^2\rho^*}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\sigma^2\rho^*-\frac{1}{2})} & \text{if } \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} < \sigma^2 \end{cases}
\end{aligned}$$

where C is some constant.

Since $\frac{1}{x}+\ln x-1$ is positive for $x \ge 1$, we find

$$\lim_{N\to\infty}\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_{\frac{1}{\sqrt{\kappa}}}^{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}}q = \begin{cases} q & \text{if } \frac{1}{\sqrt{\kappa}} \le \sigma^2 \le \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ 0 & \text{if } \kappa < \frac{1}{\sigma^4} \\ 0 & \text{if } \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} < \sigma^2 \end{cases}$$

Therefore, for $(2\beta-1)\sqrt{\kappa'} \le \frac{1}{\sigma'^2}$,

$$\begin{aligned}
&\lim_{N\to\infty}\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_{\frac{1}{\sqrt{\kappa}}}^{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}} \\
&\qquad\times\left[-\frac{1}{N}\ln\left(\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int\frac{d\rho}{\rho}\,e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}\right)\right] \qquad (7.44) \\
&\qquad= \begin{cases} \frac{1}{4}\frac{\rho^*}{\sigma'^2}-\frac{1}{2}\ln\rho^*-\frac{1}{4}+\ln\sigma' & \text{if } \frac{1}{\sqrt{\kappa}} \le \sigma^2 \le \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ 0 & \text{if o.w.} \end{cases}
\end{aligned}$$

3) $r > \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}$

- $\rho \le \frac{1}{\beta\sqrt{\kappa\kappa'}r}$

$$\begin{aligned}
&\int_0^{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}\frac{d\rho}{\rho}e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{\beta^2\kappa'}{4}\rho^2)} \\
&\qquad= \int_0^{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}\frac{d\rho}{\rho}e^{-N\left[\rho^2\frac{\kappa'}{4}\beta(1-\beta)+\frac{\rho}{2\sigma'^2}-\frac{1}{2}\ln\rho\right]}
\end{aligned}$$

We are in the region where $r > \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}$, so $p_1(.)$ is minimized at $\frac{1}{\beta\sqrt{\kappa\kappa'}r}$, and we get

$$\begin{aligned}
\int_0^{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}\frac{d\rho}{\rho}e^{-N\left[\rho^2\frac{\kappa'}{4}\beta(1-\beta)+\frac{\rho}{2\sigma'^2}-\frac{1}{2}\ln\rho\right]} &\approx e^{-Np_1(\frac{1}{\beta\sqrt{\kappa\kappa'}r})} \\
&= e^{-N\left[-\frac{1}{4\kappa r^2}(1-\frac{1}{\beta})+\frac{1}{2\beta\sqrt{\kappa\kappa'}\sigma'^2 r}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r\right]}
\end{aligned} \qquad (7.45)$$

• $\rho > \frac{1}{\beta\sqrt{\kappa\kappa'}r}$

$$\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}^{\infty} \frac{d\rho}{\rho}\, e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\beta\frac{\sqrt{\kappa\kappa'}}{2}r\rho-\frac{1}{2}\beta\sqrt{\frac{\kappa'}{\kappa}}\frac{\rho}{r}+\frac{1}{2}\ln\rho+\frac{1}{4\kappa r^2}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r+\frac{1}{2})}$$
$$= e^{-N(\frac{1}{4\kappa r^2}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r+\frac{1}{2})}\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}^{\infty}\frac{d\rho}{\rho}e^{-N\left[\beta\frac{\kappa'}{4}\rho^2+(\frac{1}{2\sigma'^2}-\beta\frac{\sqrt{\kappa\kappa'}r}{2}-\beta\frac{1}{2}\sqrt{\frac{\kappa'}{\kappa}}\frac{1}{r})\rho\right]}$$

The function in the exponent is minimized is at $\rho^{\text{opt}} = -\frac{1}{\beta\kappa'\sigma'^2}+\sqrt{\frac{\kappa}{\kappa'}}r+\frac{1}{\sqrt{\kappa\kappa'}r}$. It can be checked that, in this region $\rho^{\text{opt}} > \frac{1}{\beta\sqrt{\kappa\kappa'}r}$. So, we get

$$\begin{aligned} e^{-N(\frac{1}{4\kappa r^2}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r+\frac{1}{2})}&\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}^{\infty}\frac{d\rho}{\rho}e^{-N\left[\beta\frac{\kappa'}{4}\rho^2+(\frac{1}{2\sigma'^2}-\beta\frac{\sqrt{\kappa\kappa'}r}{2}-\beta\frac{1}{2}\sqrt{\frac{\kappa'}{\kappa}}\frac{1}{r})\rho\right]} \\ &\approx e^{-N(\frac{1}{4\kappa r^2}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r+\frac{1}{2})}e^{-np_3(\rho^{\text{opt}})} \\ &= e^{-N\left[-\frac{1}{4\kappa r^2}(\beta-1)-\frac{1}{4\beta\kappa'\sigma'^4}-\frac{1}{4}\beta\kappa r^2+\frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{1}{\sigma'^2}r+\frac{1}{2\sqrt{\kappa\kappa'}\sigma'^2}\frac{1}{r}-\frac{1}{2}(\beta-1)+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r\right]} \end{aligned} \tag{7.46}$$

Comparing the exponent in eq.(7.45),(7.46)

$$\begin{aligned} &\int d\rho\,\rho^{\frac{N}{2}-1}e^{-N\frac{\rho}{2\sigma'^2}-N\beta\frac{\kappa'}{4}\rho^2+\ln\mathcal{I}_N} \\ &\quad\approx e^{-N\left[-\frac{1}{4\kappa r^2}(\beta-1)-\frac{1}{4\beta\kappa'\sigma'^4}-\frac{1}{4}\beta\kappa r^2+\frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{1}{\sigma'^2}r+\frac{1}{2\sqrt{\kappa\kappa'}\sigma'^2}\frac{1}{r}-\frac{1}{2}(\beta-1)+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r\right]} \end{aligned}$$

which leads to

$$\begin{aligned} q(r) &= -\frac{1}{N}\ln\left[\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int\frac{d\rho}{\rho}\,e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}\right] \\ &= -\frac{1}{4\kappa r^2}(\beta-1)-\frac{1}{4\beta\kappa'\sigma'^4}-\frac{1}{4}\beta\kappa r^2+\frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{1}{\sigma'^2}r+\frac{1}{2\sqrt{\kappa\kappa'}\sigma'^2}\frac{1}{r} \\ &\qquad -\frac{1}{2}(\beta-1)+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r \end{aligned} \tag{7.47}$$

Now, we need to compute the integral

$$\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}}^{\infty} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}}q(r)$$

This integral asymptotically is

$$\begin{cases} 2\sqrt{\frac{\pi}{N}}e^{-N(\frac{1}{2}-\ln\sigma)}q(\sigma^2) & \text{if } \sigma^2 \geq \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ \frac{1}{N}e^{-N(\frac{1}{2\beta\sqrt{\kappa\kappa'}\sigma^2\rho^*}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\rho^*)}q(\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}) & \text{if } \sigma^2 < \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \end{cases}$$

This leads to

$$\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}}^{\infty} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}}q(r)$$
$$\approx \begin{cases} q(\sigma^2) & \text{if } \sigma^2 \geq \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ e^{-n(\frac{1}{2\beta\sqrt{\kappa\kappa'}\sigma^2\rho^*}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\sigma^2\rho^*-\frac{1}{2})}q(\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}) & \text{if } \sigma^2 < \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \end{cases}$$

So,

$$\lim_{N\to\infty}\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}}^{\infty} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}}q(r) = \begin{cases} q(\sigma^2) & \text{if } \sigma^2 \geq \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ 0 & \text{if } \sigma^2 < \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \end{cases}$$

Since $\frac{1}{x}+\ln x - 1$ is positive for $x \geq 1$.

Inserting $q(\sigma^2)$ from eq.(7.47), we find that for $(2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}$,

$$\lim_{N\to\infty}\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}}^{\infty} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}}$$
$$\times\left[-\frac{1}{N}\ln\left(\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int\frac{d\rho}{\rho}\,e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}\right)\right] \quad (7.48)$$
$$= \begin{cases} \frac{1-\beta}{4\kappa\sigma^4}-\frac{1}{4\beta\kappa'\sigma'^4}-\frac{1}{4}\beta\kappa\sigma^4+\frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{\sigma^2}{\sigma'^2} & \text{if } \sigma^2 \geq \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ \quad +\frac{1}{2\sqrt{\kappa\kappa'}\sigma^2\sigma'^2}-\frac{\beta}{2}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\sigma^2\sigma'^2 & \\ 0 & \text{if } \sigma^2 < \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \end{cases}$$

Now assume $(2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}$.

4) $r > \frac{1}{\sqrt{\kappa}}$

• $\rho < \frac{1}{\beta\sqrt{\kappa\kappa'}r}$

$$\int_0^{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}\frac{d\rho}{\rho}e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{\beta^2\kappa'}{4}\rho^2)}$$
$$=\int_0^{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}\frac{d\rho}{\rho}e^{-N\left[\rho^2\frac{\kappa'}{4}\beta(1-\beta)+\frac{\rho}{2\sigma'^2}-\frac{1}{2}\ln\rho\right]}$$

If $\beta < 1$, then $p_1(.)$ is minimized at ρ^*, but since $(2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}$ and $r > \frac{1}{\sqrt{\kappa}}$, it can be checked that $\rho^* > \frac{1}{\beta\sqrt{\kappa\kappa'}r}$.

If $\beta > 1$, then $p_1(.)$ is either a monotonically decreasing function, or attains its minimum at ρ^* (but $\rho^* > \frac{1}{\beta\sqrt{\kappa\kappa'}r}$). Therefore, in any case p_1 is minimized (over the integral domain) at $\frac{1}{\beta\sqrt{\kappa\kappa'}r}$.

$$\int_0^{\frac{1}{\beta\sqrt{\kappa\kappa'}r}} \frac{d\rho}{\rho} e^{-N\left[\rho^2\frac{\kappa'}{4}\beta(1-\beta)+\frac{\rho}{2\sigma'^2}-\frac{1}{2}\ln\rho\right]} \approx e^{-N\left[-\frac{1}{4\kappa r^2}(1-\frac{1}{\beta})+\frac{1}{2\beta\sqrt{\kappa\kappa'}\sigma'^2 r}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r\right]} \tag{7.49}$$

- $\rho \geq \frac{1}{\beta\sqrt{\kappa\kappa'}r}$

$$\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}^{\infty} \frac{d\rho}{\rho} e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\beta\frac{\sqrt{\kappa\kappa'}}{2}r\rho-\frac{1}{2}\beta\sqrt{\frac{\kappa'}{\kappa}}\frac{\rho}{r}+\frac{1}{2}\ln\rho+\frac{1}{4\kappa r^2}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r+\frac{1}{2})}$$
$$= e^{-N(\frac{1}{4\kappa r^2}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r+\frac{1}{2})} \int_{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}^{\infty} \frac{d\rho}{\rho} e^{-N\left[\beta\frac{\kappa'}{4}\rho^2+(\frac{1}{2\sigma'^2}-\beta\frac{\sqrt{\kappa\kappa'}r}{2}-\beta\frac{1}{2}\sqrt{\frac{\kappa'}{\kappa}}\frac{1}{r})\rho\right]}$$

The function in the exponent is minimized is at $\rho^{\text{opt}} = -\frac{1}{\beta\kappa'\sigma'^2} + \sqrt{\frac{\kappa}{\kappa'}}r + \frac{1}{\sqrt{\kappa\kappa'}r}$. It can be checked that, in this region $\rho^{\text{opt}} > \frac{1}{\beta\sqrt{\kappa\kappa'}r}$. So, we get

$$\begin{aligned} & e^{-N(\frac{1}{4\kappa r^2}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r+\frac{1}{2})} \int_{\frac{1}{\beta\sqrt{\kappa\kappa'}r}}^{\infty} \frac{d\rho}{\rho} e^{-N\left[\beta\frac{\kappa'}{4}\rho^2+(\frac{1}{2\sigma'^2}-\beta\frac{\sqrt{\kappa\kappa'}r}{2}-\beta\frac{1}{2}\sqrt{\frac{\kappa'}{\kappa}}\frac{1}{r})\rho\right]} \\ & \approx e^{-N(\frac{1}{4\kappa r^2}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r+\frac{1}{2})} e^{-np_3(\rho^{\text{opt}})} \\ & = e^{-N\left[-\frac{1}{4\kappa r^2}(\beta-1)-\frac{1}{4\beta\kappa'\sigma'^4}-\frac{1}{4}\beta\kappa r^2+\frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{1}{\sigma'^2}r+\frac{1}{2\sqrt{\kappa\kappa'}\sigma'^2}\frac{1}{r}-\frac{1}{2}(\beta-1)+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r\right]} \end{aligned} \tag{7.50}$$

Picking the dominant term from eq. (7.49), (7.50), we have

$$\int d\rho\, \rho^{\frac{N}{2}-1} e^{-N\frac{\rho}{2\sigma'^2}-N\beta\frac{\kappa'}{4}\rho^2+\ln\mathcal{I}_N} \approx e^{-N\left[-\frac{1}{4\kappa r^2}(\beta-1)-\frac{1}{4\beta\kappa'\sigma'^4}-\frac{1}{4}\beta\kappa r^2+\frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{1}{\sigma'^2}r+\frac{1}{2\sqrt{\kappa\kappa'}\sigma'^2}\frac{1}{r}-\frac{1}{2}(\beta-1)+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r\right]}$$

which leads to

$$\begin{aligned} q(r) &= -\frac{1}{N}\ln\left[\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int\frac{d\rho}{\rho}\, e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}\right] \\ &= -\frac{1}{4\kappa r^2}(\beta-1) - \frac{1}{4\beta\kappa'\sigma'^4} - \frac{1}{4}\beta\kappa r^2 + \frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{1}{\sigma'^2}r + \frac{1}{2\sqrt{\kappa\kappa'}\sigma'^2}\frac{1}{r} \\ &\quad - \frac{1}{2}(\beta-1) + \frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r \end{aligned} \tag{7.51}$$

Now, we need to compute the integral

$$\int_{\frac{1}{\sqrt{\kappa}}}^{\infty} dr\, r^{\frac{N}{2}-1} e^{-N\frac{r}{2\sigma^2}} q(r)$$

This integral asymptotically is

$$\begin{cases} 2\sqrt{\frac{\pi}{N}}e^{-N(\frac{1}{2}-\ln\sigma)}q(\sigma^2) & \text{if } \sigma^2 \geq \frac{1}{\sqrt{\kappa}} \\ \frac{1}{N}e^{-N(\frac{1}{2\beta\sqrt{\kappa\kappa'}\sigma^2\rho^*}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\rho^*)}q(\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}) & \text{if } \sigma^2 < \frac{1}{\sqrt{\kappa}} \end{cases}$$

This leads to

$$\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_{\frac{1}{\sqrt{\kappa}}}^{\infty} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}}q(r)$$
$$\approx \begin{cases} q(\sigma^2) & \text{if } \sigma^2 \geq \frac{1}{\sqrt{\kappa}} \\ e^{-N(\frac{1}{2\beta\sqrt{\kappa\kappa'}\sigma^2\rho^*}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\sigma^2\rho^*-\frac{1}{2})}q(\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}) & \text{if } \sigma^2 < \frac{1}{\sqrt{\kappa}} \end{cases}$$

So,

$$\lim_{N\to\infty}\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_{\frac{1}{\sqrt{\kappa}}}^{\infty} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}}q(r) = \begin{cases} q(\sigma^2) & \text{if } \sigma^2 \geq \frac{1}{\sqrt{\kappa}} \\ 0 & \text{if } \sigma^2 < \frac{1}{\sqrt{\kappa}} \end{cases}$$

Since $\frac{1}{x}+\ln x - 1$ is positive for $x \neq 1$.

Inserting $q(\sigma^2)$ from eq.(7.51), we find that for $(2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}$,

$$\lim_{N\to\infty}\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma^N}\int_{\frac{1}{\sqrt{\kappa}}}^{\infty} dr\, r^{\frac{N}{2}-1}e^{-N\frac{r}{2\sigma^2}}$$
$$\times\left[-\frac{1}{N}\ln\left(\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int\frac{d\rho}{\rho}\,e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}\right)\right] \quad (7.52)$$
$$= \begin{cases} \frac{1-\beta}{4\kappa\sigma^4}-\frac{1}{4\beta\kappa'\sigma'^4}-\frac{1}{4}\beta\kappa\sigma^4+\frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{\sigma^2}{\sigma'^2} & \text{if } \sigma^2 \geq \frac{1}{\sqrt{\kappa}} \\ \quad +\frac{1}{2\sqrt{\kappa\kappa'}\sigma^2\sigma'^2}-\frac{\beta}{2}+\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\sigma^2\sigma'^2 & \\ 0 & \text{if } \sigma^2 < \frac{1}{\sqrt{\kappa}} \end{cases}$$

Finally, from eq.(7.40),(7.44),(7.48),(7.52), we have $\lim_{N\to\infty} f_N =:$

$$\begin{cases} \frac{1}{4}\frac{\rho^*}{\sigma'^2}-\frac{1}{2}\ln\rho^*-\frac{1}{4}+\ln\sigma' & \text{if } (2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}, \kappa \leq \frac{1}{\sigma^4}, \\ \frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'+\ln\sigma'+\frac{1}{4} & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}, \kappa \leq \frac{1}{\sigma^4}, \\ \quad -\frac{1}{4\beta\kappa'\sigma'^4}+\frac{1}{\sqrt{\kappa'}\sigma'^2}-\beta & \\ \frac{1}{4}\frac{\rho^*}{\sigma'^2}-\frac{1}{2}\ln\rho^*-\frac{1}{4}+\ln\sigma' & \text{if } (2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}, \frac{1}{\sqrt{\kappa}} \leq \sigma^2 \leq \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}, \\ \frac{1-\beta}{4\kappa\sigma^4}-\frac{1}{4\beta\kappa'\sigma'^4}-\frac{1}{4}\beta\kappa\sigma^4 & \text{if } (2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}, \sqrt{\kappa\kappa'} \geq \frac{1}{\beta\sigma^2\rho^*}, \\ \quad +\frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{\sigma^2}{\sigma'^2}+\frac{1}{2\sqrt{\kappa\kappa'}\sigma^2\sigma'^2} & \\ \quad +\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\sigma^2\sigma'^2-\frac{\beta}{2} & \\ \frac{1-\beta}{4\kappa\sigma^4}-\frac{1}{4\beta\kappa'\sigma'^4}-\frac{1}{4}\beta\kappa\sigma^4+\frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{\sigma^2}{\sigma'^2} & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}, \kappa \geq \frac{1}{\sigma^4}, \\ \quad +\frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{\sigma^2}{\sigma'^2}+\frac{1}{2\sqrt{\kappa\kappa'}\sigma^2\sigma'^2} & \\ \quad +\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\sigma^2\sigma'^2-\frac{\beta}{2} & \\ 0 & \text{if o.w.} \end{cases} \quad (7.53)$$

with

$$\rho^* = \frac{1}{\kappa'\beta(\beta-1)}\left(\frac{1}{2\sigma'^2}-\sqrt{\frac{1}{4\sigma'^4}-\kappa'\beta(\beta-1)}\right)$$

Remark 7.9. *ρ^* was defined as the minimizer of $p_1(.)$. For $\beta = 1$ we have $p_1(\rho) = \frac{\rho}{2\sigma'^2} - \frac{1}{2}\ln\rho$, so $\rho^* = \sigma'^2$ and all the steps we went through are valid by simply substituting $\rho^* = \sigma'^2$.*

The two conditions $(2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}, \sqrt{\kappa\kappa'} \geq \frac{1}{\beta\sigma^2\rho^*}$ or $(2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}, \kappa \geq \frac{1}{\sigma^4}$ can be combined into $\kappa \geq \frac{1}{\sigma^4}, \sqrt{\kappa\kappa'} \geq \frac{1}{\beta\sigma^2\rho^*}$.

7.C MMSE Achieving Curve for $\beta = 1$

7.C.1 Equating MSE and MMSE

For $\beta = 1$, the asymptotic mismatched MSE reads

$$\begin{cases} \sigma^4 + \left(\frac{1-\sqrt{\kappa'}\sigma'^2}{\kappa'\sigma'^2}\right)^2 & \text{if} \kappa \leq \frac{1}{\sigma^4}, \text{and} \kappa' \geq \frac{1}{\sigma'^4} \\ \sigma^4\left(1-\sqrt{\frac{\kappa}{\kappa'}}\right)^2 + \frac{2}{\sqrt{\kappa\kappa'}} + \frac{1}{\kappa'^2\sigma'^4} & \text{if} \kappa' \leq \frac{1}{\sigma'^4}, \text{ and } \sqrt{\kappa\kappa'} \geq \frac{1}{\sigma^2\sigma'^2} \\ \quad + \frac{2}{\kappa'}\frac{\sigma^2}{\sigma'^2}\left(1-\sqrt{\frac{\kappa}{\kappa'}}\right) - \frac{2}{\kappa\kappa'\sigma^2\sigma'^2} & \\ \sigma^4\left(1-\sqrt{\frac{\kappa}{\kappa'}}\right)^2 + \frac{2}{\sqrt{\kappa\kappa'}} + \frac{1}{\kappa'^2\sigma'^4} & \text{if} \kappa' \geq \frac{1}{\sigma'^4}, \text{ and } \kappa \geq \frac{1}{\sigma^4} \\ \quad + \frac{2}{\kappa'}\frac{\sigma^2}{\sigma'^2}\left(1-\sqrt{\frac{\kappa}{\kappa'}}\right) - \frac{2}{\kappa\kappa'\sigma^2\sigma'^2} & \\ \sigma^4 & \text{o.w.} \end{cases}$$

We consider the case where $\kappa\sigma^4 > 1$ and estimation better than guess is possible. Equating MSE and MMSE (7.19) for parameters κ', σ', we find:

$$\begin{aligned} \kappa' &= \frac{\kappa\sigma^2}{1-\kappa\sigma^4}\frac{1}{\sigma'^2} \pm \frac{1}{2}\frac{\kappa^2\sigma^5}{(\kappa\sigma^4-1)^2}\frac{1}{\sigma'}\sqrt{4(1-\kappa\sigma^4)+\kappa^2\sigma^6\sigma'^2} + \frac{1}{2}\frac{\kappa^3\sigma^8}{(\kappa\sigma^4-1)^2} \\ \sigma' &\geq 2\frac{\sqrt{\kappa\sigma^4-1}}{\kappa\sigma^3} \end{aligned} \tag{7.54}$$

From (7.54), we see that the curve has vertical tangent at $\kappa' = \frac{\kappa^3\sigma^8}{4(\kappa\sigma^4-1)^2}, \sigma' = 2\frac{\sqrt{\kappa\sigma^4-1}}{\kappa\sigma^3}$. Moreover, the asymptotic of the curve at $\sigma' \to \infty$ is $\kappa' = \frac{\kappa^3\sigma^8}{(\kappa\sigma^4-1)^2}$.

7.C.2 Enforcing Nishimori condition

For $\beta = 1$, the asymptotic free energy is

$$\begin{cases} -\frac{1}{4\kappa'\sigma'^4} + \frac{1}{\sqrt{\kappa'}\sigma'^2} - \frac{3}{4} + \ln\kappa'^{\frac{1}{4}}\sigma' & \text{if } \kappa \leq \frac{1}{\sigma^4}, \text{ and } \kappa' \geq \frac{1}{\sigma'^4} \\ \frac{1}{2}\ln\sqrt{\kappa\kappa'}\sigma^2\sigma'^2 - \frac{1}{4\kappa'\sigma'^4} - \frac{\kappa\sigma^4}{4} & \text{if } \kappa' \leq \frac{1}{\sigma'^4}, \text{ and } \sqrt{\kappa\kappa'} \geq \frac{1}{\sigma^2\sigma'^2} \\ \quad + \sqrt{\frac{\kappa}{\kappa'}}\frac{\sigma^2}{2\sigma'^2} + \frac{1}{2\sqrt{\kappa\kappa'}\sigma^2\sigma'^2} - \frac{1}{2} & \\ \frac{1}{2}\ln\sqrt{\kappa\kappa'}\sigma^2\sigma'^2 - \frac{1}{4\kappa'\sigma'^4} - \frac{\kappa\sigma^4}{4} & \text{if } \kappa' \geq \frac{1}{\sigma'^4}, \text{ and } \kappa \geq \frac{1}{\sigma^4} \\ \quad + \sqrt{\frac{\kappa}{\kappa'}}\frac{\sigma^2}{2\sigma'^2} + \frac{1}{2\sqrt{\kappa\kappa'}\sigma^2\sigma'^2} - \frac{1}{2} & \\ 0 & \text{if o.w.} \end{cases}$$

Using (7.15), we have

$$m_{\text{stat}} = -4\sqrt{\frac{\kappa}{\kappa'}}\frac{d}{d\kappa}f = -\frac{1}{\sqrt{\kappa\kappa'}} + \sqrt{\frac{\kappa}{\kappa'}}\sigma^4 - \frac{1}{\kappa'}\frac{\sigma^2}{\sigma'^2} + \frac{1}{\kappa\kappa'\sigma^2\sigma'^2}$$

$$q_{\text{stat}} = -4\frac{\kappa}{\kappa'}\frac{d}{d\kappa}f + 4\frac{d}{d\kappa'}f = \left(\sqrt{\frac{\kappa}{\kappa'}}\sigma^2 - \frac{1}{\kappa\sigma'^2}\right)^2$$

Imposing Nishimori identity, $m_{\text{stat}} = q_{\text{stat}}$ we find the same relation as (7.54) for κ', σ'.

Remark 7.10. *As the fixed points of the state evolution equations* (7.24)*, m^*_{AMP} and q^*_{AMP} coincides with their corresponding statistical parameters, the same discussion is true for $m^*_{\text{AMP}}, q^*_{\text{AMP}}$.*

7.C.3 Spectral rescaling

By remark 7.5, the bracket estimator performs as the spectral algorithm with the rescaling factor $q_{\text{stat}} = \left(\sqrt{\frac{\kappa}{\kappa'}}\sigma^2 - \frac{1}{\kappa\sigma'^2}\right)^2$. The rescaling factor in the fully known model is $\delta^2 = \left(\sigma^2 - \frac{1}{\kappa\sigma^2}\right)^2$. Equating q_{stat} with true factor gives the same curve as eq. (7.54).

7.D Proof of Theorem 7.3

First, we compute the asymptotic of the free energy of the system with the true Bernoulli-Rademacher prior.

Theorem 7.6. *For all $\alpha, \sigma', \kappa, \kappa', \beta$ positive, the asymptotic of free energy for the mismatched inference reads:*

$$\begin{cases} \frac{1}{4}\frac{\rho^*}{\sigma'^2} - \frac{1}{2}\ln\rho^* - \frac{1}{4} + \frac{1}{2}\ln\alpha & \text{if } (2\beta-1)\sqrt{\kappa'} \le \frac{1}{\sigma'^2}, \kappa \le \frac{1}{\alpha^2}, \\ \frac{1}{2}\ln\beta + \frac{1}{4}\ln\kappa' + \ln\sigma' + \frac{1}{4} & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}, \kappa \le \frac{1}{\alpha^2}, \\ \quad -\frac{1}{4\beta\kappa'\sigma'^4} + \frac{1}{\sqrt{\kappa'}\sigma'^2} - \beta & \\ \frac{1}{4}\frac{\rho^*}{\sigma'^2} - \frac{1}{2}\ln\rho^* - \frac{1}{4} + \frac{1}{2}\ln\alpha & \text{if } (2\beta-1)\sqrt{\kappa'} \le \frac{1}{\sigma'^2}, \frac{1}{\sqrt{\kappa}} \le \alpha \le \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}, \\ \frac{1-\beta}{4\kappa\alpha^2} - \frac{1}{4\beta\kappa'\sigma'^4} - \frac{1}{4}\beta\kappa\alpha^2 - \frac{\beta}{2} & \text{if } \kappa \ge \frac{1}{\alpha^2}, \sqrt{\kappa\kappa'} \ge \frac{1}{\beta\sigma^2\rho^*}, \\ \quad +\frac{1}{2\sqrt{\kappa\kappa'}\alpha\sigma'^2} + \frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{\alpha}{\sigma'^2} & \\ \quad\quad +\frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\alpha\sigma'^2 & \\ 0 & \text{o.w.} \end{cases} \tag{7.55}$$

where

$$\rho^* = \begin{cases} \sigma'^2 & \text{if } \beta = 1, \\ \frac{1}{\kappa'\beta(\beta-1)}\left(\frac{1}{2\sigma'^2} - \sqrt{\frac{1}{4\sigma'^4} - \kappa'\beta(\beta-1)}\right) & \text{if } \beta \ne 1 \end{cases}$$

Proof. We follow the same steps as the proof of the theorem 7.4 for the Gaussian prior. We have

$$\begin{aligned} f_N &= -\frac{1}{N}\mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\big[\ln Z(\boldsymbol{Y})\big] \\ &= \mathbb{E}_{\mathbf{P}^*,\mathbf{P}_{\boldsymbol{Z}}}\Big[-\frac{1}{N}\ln\int d\boldsymbol{x}\mathbf{P}(\boldsymbol{x})\, e^{\frac{-\beta\kappa'}{4N}\|\boldsymbol{x}\|^4+\ln\mathcal{I}_N}\Big] \end{aligned} \tag{7.56}$$

where

$$\mathcal{I}_N = \int D\boldsymbol{U} e^{n\,\mathrm{Tr}\,\frac{\boldsymbol{Y}}{\sqrt{n}}\boldsymbol{U}\frac{\beta\sqrt{\kappa'}}{2n}\boldsymbol{x}\boldsymbol{x}^T\boldsymbol{U}^T} \tag{7.57}$$

The integrand in the $\boldsymbol{x}$-integral in (7.56) is a function of $\|\boldsymbol{x}\|$. Furthermore, the function in the expectation $\mathbb{E}_{\mathbf{P}^*}$ is a function of $\|\boldsymbol{s}\|$. Therefore we can use spherical coordinates (see appendix 7.E) to reduce the interior integral in (7.56) to a one-dimensional integral, and the expectation over $\mathbf{P}^*$ to the expectation over the norm $\|\boldsymbol{s}\| = r$.

$$f_N = \mathbb{E}_{\mathbf{P}_r,\mathbf{P}_{\boldsymbol{Z}}}\Big[-\frac{1}{N}\ln\Big\{\frac{2^{-\frac{N}{2}+1}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int_0^{+\infty} d\rho\,\rho^{N-1}e^{-\frac{\rho^2}{2\sigma'^2}-\frac{\beta\kappa'}{4N}\rho^4+\ln\mathcal{I}_N}\Big\}\Big] \tag{7.58}$$

where $r := \|\boldsymbol{s}\|$, $\rho := \|\boldsymbol{x}\|$, and $\Gamma(.)$ is the *Gamma* function. r^2 is distributed according to the Binomial distribution, $B(n,\alpha)$. Changing variable $\frac{r^2}{n} \to r$, $\frac{\rho^2}{n} \to \rho$, we obtain

$$f_N = \mathbb{E}_{\mathbf{P}_r,\mathbf{P}_{\boldsymbol{Z}}}\Big[\frac{1}{N}\ln\Big\{\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int_0^{+\infty} d\rho\,\rho^{\frac{N}{2}-1}e^{-N(\frac{\rho}{2\sigma'^2}+\frac{\beta\kappa'}{4}\rho^2-\frac{1}{N}\ln\mathcal{I}_N}\Big\}\Big] \tag{7.59}$$

where r has the following distribution:

$$\mathbf{P}\{r=\frac{k}{N}\} = \binom{N}{k}\alpha^k(1-\alpha)^{N-k}, \quad k=0,1,\dots,N$$

□

The term $\frac{1}{N}\ln\mathcal{I}_N$ is the same as the case of Gaussian prior, see (7.35).

7.D.1 Details of applying the Laplace method

To compute the limit in eq.(7.59), we split the interior integral and the expectation (over r) based on the different cases of the involved parameters.

1) $r \le \frac{1}{\sqrt{\kappa}}$

With similar calculations in 7.B.3 we can find the asymptotic of the function in the expectation which leads us to the following expression that we need to compute.

$$\mathbb{E}_{\mathbf{P}_r,\mathbf{P}_{\boldsymbol{Z}}}\big[q\,\boldsymbol{I}\{r\le\frac{1}{\sqrt{\kappa}}\}\big]$$

where

$$q = \begin{cases} \frac{1}{4}\frac{\rho^*}{\sigma'^2} - \frac{1}{2}\ln\rho^* - \frac{1}{4} + \ln\sigma' & \text{if } (2\beta-1)\sqrt{\kappa'} \le \frac{1}{\sigma'^2} \\ \frac{1}{2}\ln\beta + \frac{1}{4}\ln\kappa' + \ln\sigma' + \frac{1}{4} - \frac{1}{4\beta\kappa'\sigma'^4} + \frac{1}{\sqrt{\kappa'}\sigma'^2} - \beta & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2} \end{cases} \tag{7.60}$$

q is independent of the $\boldsymbol{Z}$, therefore, we have

$$\mathbb{E}_{\mathbf{P}_r}\big[q\,\boldsymbol{I}\{r \le \frac{1}{\sqrt{\kappa}}\}\big] = \sum_{i=0}^{\lfloor \frac{N}{\sqrt{\kappa}} \rfloor} \binom{N}{i} \alpha^i (1-\alpha)^{N-i} q = q \sum_{i=0}^{\lfloor \frac{N}{\sqrt{\kappa}} \rfloor} \binom{N}{i} \alpha^i (1-\alpha)^{N-i} \tag{7.61}$$

By central limit theorem, we can approximate the Binomial distribution with a normal one, and for $N \to \infty$ we have

$$\binom{N}{i} \alpha^i (1-\alpha)^{N-i} \sim \frac{1}{\sqrt{2\pi N\alpha(1-\alpha)}} e^{-\frac{(i-N\alpha)^2}{2N\alpha(1-\alpha)}}$$

Substituting in (7.61), we get

$$\begin{aligned} \mathbb{E}_{\mathbf{P}_r}\big[q\,\boldsymbol{I}\{r \le \frac{1}{\sqrt{\kappa}}\}\big] &= q \sum_{i=0}^{\lfloor \frac{N}{\sqrt{\kappa}} \rfloor} \frac{1}{\sqrt{2\pi N\alpha(1-\alpha)}} e^{-\frac{(i-N\alpha)^2}{2N\alpha(1-\alpha)}} \\ &= q \sum_{j=0}^{\frac{1}{\sqrt{\kappa}}} \frac{1}{\sqrt{2\pi N\alpha(1-\alpha)}} e^{-N\frac{(j-\alpha)^2}{2\alpha(1-\alpha)}} \end{aligned} \tag{7.62}$$

where in the last equality we changed the index $\frac{i}{N} \to j$, and approximate $\frac{1}{N}\lfloor \frac{N}{\sqrt{\kappa}} \rfloor$ with $\frac{1}{\sqrt{\kappa}}$ for N large enough. Now, for $N \to \infty$ the sum can be interpreted as a Riemann sum.

$$q \sum_{j=0}^{\frac{1}{\sqrt{\kappa}}} \frac{1}{\sqrt{2\pi N\alpha(1-\alpha)}} e^{-N\frac{(j-\alpha)^2}{2\alpha(1-\alpha)}} = q\,N \int_0^{\frac{1}{\sqrt{\kappa}}} \frac{d\,r}{\sqrt{2\pi N\alpha(1-\alpha)}} e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}} \tag{7.63}$$

Thus, it remains to compute the the following limit:

$$\lim_{N\to\infty} q \sqrt{\frac{N}{2\pi\alpha(1-\alpha)}} \int_0^{\frac{1}{\sqrt{\kappa}}} d\,r\, e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}}$$

Applying the regular Laplace method to the integral over r, we get

$$\int_0^{\frac{1}{\sqrt{\kappa}}} d\,r\, e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}} = \begin{cases} \sqrt{\frac{2\pi}{N\frac{1}{\alpha(1-\alpha)}}}\big(1 + O(\frac{1}{N})\big) & \text{if } \kappa \le \frac{1}{\alpha^2} \\ \frac{\alpha(1-\alpha)}{\alpha - \frac{1}{\sqrt{\kappa}}} \frac{1}{N} e^{-N\frac{(\frac{1}{\sqrt{\kappa}}-\alpha)^2}{2\alpha(1-\alpha)}} \big(1 + O(\frac{1}{N})\big) & \text{if } \kappa > \frac{1}{\alpha^2} \end{cases}$$

Therefore, we have

$$\lim_{N\to\infty} q\sqrt{\frac{N}{2\pi\alpha(1-\alpha)}}\int_0^{\frac{1}{\sqrt{\kappa}}} dr\, e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}} = \begin{cases} q & \text{if } \kappa \le \frac{1}{\alpha^2} \\ 0 & \text{if } \kappa > \frac{1}{\alpha^2} \end{cases}$$

Substituting q from eq.(7.60), we obtain

$$\begin{aligned}\lim_{N\to\infty} -\frac{1}{N}\mathbb{E}_{\mathbf{P}_r,\mathbf{P}_{\boldsymbol{Z}}}\Bigg[\boldsymbol{I}\{r \le \frac{1}{\sqrt{\kappa}}\} \\ \times \ln\left(\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int_0^\infty \frac{d\rho}{\rho}e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}\right)\Bigg] \\ = \begin{cases} \frac{1}{4}\frac{\rho^*}{\sigma'^2}-\frac{1}{2}\ln\rho^*-\frac{1}{4}+\ln\sigma' & \text{if } (2\beta-1)\sqrt{\kappa'} \le \frac{1}{\sigma'^2}, \kappa \le \frac{1}{\alpha^2} \\ \frac{1}{2}\ln\beta+\frac{1}{4}\ln\kappa'+\ln\sigma'+\frac{1}{4} & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}, \kappa \le \frac{1}{\alpha^2}, \\ \quad -\frac{1}{4\beta\kappa'\sigma'^4}+\frac{1}{\sqrt{\kappa'}\sigma'^2}-\beta & \\ 0 & \text{otherwise} \end{cases}\end{aligned} \tag{7.64}$$

To compute the remaining part of integral, we need to consider two cases. Throughout the following, suppose $(2\beta-1)\sqrt{\kappa'} \le \frac{1}{\sigma'^2}$.

2) $\frac{1}{\sqrt{\kappa}} < r \le \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}$

In this case, the asymptotic of the interior integral is same as in (7.43), and by the discussion in previous section we can approximate $\mathbf{P}_r$ by a Gaussian distribution with mean α, and variance $\alpha(1-\alpha)$. Therefore, in the end, we need to compute the following limit.

$$\lim_{N\to\infty}\sqrt{\frac{N}{2\pi\alpha(1-\alpha)}}\int_{\frac{1}{\sqrt{\kappa}}}^{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}} dr\, e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}}q$$

where

$$\begin{aligned} q &= -\frac{1}{N}\ln\left[\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int\frac{d\rho}{\rho}\,e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}\right] \\ &\approx \frac{1}{4}\frac{\rho^*}{\sigma'^2}-\frac{1}{2}\ln\rho^*-\frac{1}{4}+\ln\sigma' \end{aligned}$$

Applying the regular Laplace method to the integral over r, we get

$$\int_{\frac{1}{\sqrt{\kappa}}}^{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}} dr\, e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}} \approx \begin{cases} \sqrt{\frac{2\pi}{N\frac{1}{\alpha(1-\alpha)}}} & \text{if } \frac{1}{\sqrt{\kappa}} \le \alpha \le \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ \frac{1}{N}e^{-N\frac{(\frac{1}{\sqrt{\kappa}}-\alpha)^2}{2\alpha(1-\alpha)}} & \text{if } \kappa < \frac{1}{\alpha^2} \\ \frac{1}{N}e^{-N\frac{(\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}-\alpha)^2}{2\alpha(1-\alpha)}} & \text{if } \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} < \alpha \end{cases}$$

which leads to

$$\lim_{N\to\infty}\sqrt{\frac{N}{2\pi\alpha(1-\alpha)}}\int_{\frac{1}{\sqrt{\kappa}}}^{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}} dr\, e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}}q = \begin{cases} q & \text{if } \frac{1}{\sqrt{\kappa}} \le \alpha \le \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ 0 & \text{if } \kappa < \frac{1}{\alpha^2} \\ 0 & \text{if } \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} < \alpha \end{cases}$$

Therefore for $(2\beta-1)\sqrt{\kappa'} \le \frac{1}{\sigma'^2}$,

$$\begin{aligned} -\frac{1}{N}\mathbb{E}_{\mathbf{P}_r,\mathbf{P}_{\boldsymbol{Z}}}\Bigg[\boldsymbol{I}\{\frac{1}{\sqrt{\kappa}} < r \le \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}\} \\ \times \ln\left(\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int\frac{d\rho}{\rho}\, e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}\right)\Bigg] \qquad (7.65) \\ = \begin{cases} \frac{1}{4}\frac{\rho^*}{\sigma'^2} - \frac{1}{2}\ln\rho^* - \frac{1}{4} + \ln\sigma' & \text{if } \frac{1}{\sqrt{\kappa}} \le \alpha \le \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ 0 & \text{if o.w.} \end{cases} \end{aligned}$$

3) $r > \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}$

In this case, the asymptotic of the interior integral is same as in (7.47), and by the discussion in previous section we can approximate $\mathbf{P}_r$ by a Gaussian distribution with mean α, and variance $\alpha(1-\alpha)$. Therefore, in the end, we need to compute the following limit. which leads to

$$\lim_{N\to\infty}\sqrt{\frac{N}{2\pi\alpha(1-\alpha)}}\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}}^{\infty} dr\, e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}}q(r)$$

where

$$\begin{aligned} q(r) = -\frac{1}{4\kappa r^2}(\beta-1) - \frac{1}{4\beta\kappa'\sigma'^4} - \frac{1}{4}\beta\kappa r^2 + \frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{1}{\sigma'^2}r \\ + \frac{1}{2\sqrt{\kappa\kappa'}\sigma'^2}\frac{1}{r} - \frac{1}{2}(\beta-1) + \frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r \end{aligned}$$

We have

$$\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}}^{\infty} dr\, e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}}q(r) \approx \begin{cases} \sqrt{\frac{2\pi}{N\frac{1}{\alpha(1-\alpha)}}}q(\alpha) & \text{if } \alpha \ge \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ \frac{1}{N}e^{-N\frac{(\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}-\alpha)^2}{2\alpha(1-\alpha)}}q(\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}) & \text{if } \alpha < \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \end{cases}$$

This leads to

$$\lim_{N\to\infty}\sqrt{\frac{N}{2\pi\alpha(1-\alpha)}}\int_{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}}^{\infty} dr\, e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}}q(r) \approx \begin{cases} q(\alpha) & \text{if } \alpha \ge \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ 0 & \text{if } \alpha < \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \end{cases}$$

Inserting $q(\sigma^2)$, we find that for $(2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}$,

$$\begin{aligned}\lim_{N\to\infty} -\frac{1}{N}\mathbb{E}_{\mathbf{P}_r,\mathbf{P}_{\boldsymbol{Z}}}\Bigg[\boldsymbol{I}\{\frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} < r\} \\ \times \ln\left(\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int\frac{d\rho}{\rho}\,e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}\right)\Bigg] \\ = \begin{cases} \frac{1-\beta}{4\kappa\alpha^2} - \frac{1}{4\beta\kappa'\sigma'^4} - \frac{1}{4}\beta\kappa\alpha^2 + \frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{\alpha}{\sigma'^2} & \text{if } \alpha \geq \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \\ \quad +\frac{1}{2\sqrt{\kappa\kappa'}\alpha\sigma'^2} - \frac{\beta}{2} + \frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\alpha\sigma'^2 & \\ 0 & \text{if } \alpha < \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*} \end{cases}\end{aligned} \tag{7.66}$$

Now assume $(2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}$.

4) $r > \frac{1}{\sqrt{\kappa}}$

We need to compute the following limit

$$\lim_{N\to\infty}\sqrt{\frac{N}{2\pi\alpha(1-\alpha)}}\int_{\frac{1}{\sqrt{\kappa}}}^{\infty} dr\, e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}}q(r)$$

where (from (7.51))

$$\begin{aligned}q(r) = -\frac{1}{4\kappa r^2}(\beta-1) - \frac{1}{4\beta\kappa'\sigma'^4} - \frac{1}{4}\beta\kappa r^2 + \frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{1}{\sigma'^2}r \\ + \frac{1}{2\sqrt{\kappa\kappa'}\sigma'^2}\frac{1}{r} - \frac{1}{2}(\beta-1) + \frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}r\end{aligned}$$

We have

$$\int_{\frac{1}{\sqrt{\kappa}}}^{\infty} dr\, e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}}q(r) \approx \begin{cases} \sqrt{\frac{2\pi}{N\frac{1}{\alpha(1-\alpha)}}}q(\alpha) & \text{if } \alpha \geq \frac{1}{\sqrt{\kappa}} \\ \frac{1}{N}e^{-N\frac{(\frac{1}{\sqrt{\kappa}}-\alpha)^2}{2\alpha(1-\alpha)}}q(\frac{1}{\sqrt{\kappa}}) & \text{if } \alpha < \frac{1}{\sqrt{\kappa}} \end{cases}$$

which implies

$$\lim_{N\to\infty}\sqrt{\frac{N}{2\pi\alpha(1-\alpha)}}\int_{\frac{1}{\sqrt{\kappa}}}^{\infty} dr\, e^{-N\frac{(r-\alpha)^2}{2\alpha(1-\alpha)}}q(r) = \begin{cases} q(\alpha) & \text{if } \alpha \geq \frac{1}{\sqrt{\kappa}} \\ 0 & \text{if } \alpha < \frac{1}{\sqrt{\kappa}} \end{cases}$$

Substituting $q(\alpha)$, we find that for $(2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}$,

$$\begin{aligned}
&\lim_{N\to\infty} -\frac{1}{N}\mathbb{E}_{\mathbf{P}_r,\mathbf{P}_{\boldsymbol{Z}}}\Bigg[\boldsymbol{I}\{\frac{1}{\sqrt{\kappa}} < r\} \\
&\quad \times \ln\left(\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{\sigma'^N}\int\frac{d\rho}{\rho}\, e^{-N(\frac{\rho}{2\sigma'^2}+\beta\frac{\kappa'}{4}\rho^2-\frac{1}{2}\ln\rho-\frac{1}{N}\ln\mathcal{I}_N)}\right)\Bigg] \qquad (7.67)\\
&= \begin{cases} \frac{1-\beta}{4\kappa\alpha^2} - \frac{1}{4\beta\kappa'\sigma'^4} - \frac{1}{4}\beta\kappa\alpha^2 + \frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{\alpha}{\sigma'^2} & \text{if } \alpha \geq \frac{1}{\sqrt{\kappa}} \\ \quad + \frac{1}{2\sqrt{\kappa\kappa'}\alpha\sigma'^2} - \frac{\beta}{2} + \frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\alpha\sigma'^2 & \\ 0 & \text{if } \alpha < \frac{1}{\sqrt{\kappa}} \end{cases}
\end{aligned}$$

Finally, from eq.(7.64),(7.65),(7.66),(7.67), we have $\lim_{N\to\infty} f_N =:$

$$\begin{cases}
\frac{1}{4}\frac{\rho^*}{\sigma'^2} - \frac{1}{2}\ln\rho^* - \frac{1}{4} + \ln\sigma' & \text{if } (2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}, \kappa \leq \frac{1}{\alpha^2}, \\
\frac{1}{2}\ln\beta + \frac{1}{4}\ln\kappa' + \ln\sigma' + \frac{1}{4} & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}, \kappa \leq \frac{1}{\alpha^2}, \\
\quad - \frac{1}{4\beta\kappa'\sigma'^4} + \frac{1}{\sqrt{\kappa'}\sigma'^2} - \beta & \\
\frac{1}{4}\frac{\rho^*}{\sigma'^2} - \frac{1}{2}\ln\rho^* - \frac{1}{4} + \ln\sigma' & \text{if } (2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}, \frac{1}{\sqrt{\kappa}} \leq \alpha \leq \frac{1}{\beta\sqrt{\kappa\kappa'}\rho^*}, \\
\frac{1-\beta}{4\kappa\alpha^2} - \frac{1}{4\beta\kappa'\sigma'^4} - \frac{1}{4}\beta\kappa\alpha^2 & \text{if } (2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}, \sqrt{\kappa\kappa'} \geq \frac{1}{\beta\alpha\rho^*} \\
\quad + \frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{\alpha}{\sigma'^2} + \frac{1}{2\sqrt{\kappa\kappa'}\alpha\sigma'^2} & \\
\quad + \frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\alpha\sigma'^2 - \frac{\beta}{2} & \\
\frac{1-\beta}{4\kappa\alpha^2} - \frac{1}{4\beta\kappa'\sigma'^4} - \frac{1}{4}\beta\kappa\alpha^2 & \text{if } (2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}, \kappa \geq \frac{1}{\alpha^2}, \\
\quad + \frac{1}{2}\sqrt{\frac{\kappa}{\kappa'}}\frac{\alpha}{\sigma'^2} + \frac{1}{2\sqrt{\kappa\kappa'}\alpha\sigma'^2} & \\
\quad + \frac{1}{2}\ln\beta\sqrt{\kappa\kappa'}\alpha\sigma'^2 - \frac{\beta}{2} & \\
0 & \text{if o.w.}
\end{cases} \qquad (7.68)$$

with

$$\rho^* = \frac{1}{\kappa'\beta(\beta-1)}\left(\frac{1}{2\sigma'^2} - \sqrt{\frac{1}{4\sigma'^4} - \kappa'\beta(\beta-1)}\right)$$

Remark 7.11. *ρ^* was defined as the minimizer of $p_1(.)$. For $\beta = 1$ we have $p_1(\rho) = \frac{\rho}{2\sigma'^2} - \frac{1}{2}\ln\rho$, so $\rho^* = \sigma'^2$ and all the steps we went through are valid by simply substituting $\rho^* = \sigma'^2$.*

The two conditions $(2\beta-1)\sqrt{\kappa'} \leq \frac{1}{\sigma'^2}, \sqrt{\kappa\kappa'} \geq \frac{1}{\beta\alpha\rho^*}$ or $(2\beta-1)\sqrt{\kappa'} > \frac{1}{\sigma'^2}, \kappa \geq \frac{1}{\alpha^2}$ can be combined into $\kappa \geq \frac{1}{\alpha^2}, \sqrt{\kappa\kappa'} \geq \frac{1}{\beta\alpha\rho^*}$.

7.E N-Dimensional Spherical Coordinates

$$d\boldsymbol{s} = ds_1 ds_2 \dots ds_N$$

Consider the variables $0 \leq r, 0 \leq \theta \leq 2\pi, \phi_1, \phi_2, \ldots, \phi_{N-2}$, where $0 \leq \phi_i \leq \pi$ with the following transformation:

$$\begin{aligned}
s_1 &= r\cos\phi_1 \\
s_2 &= r\sin\phi_1\ \cos\phi_2 \\
s_3 &= r\sin\phi_1\ \sin\phi_2\, cos\phi_3 \\
&\vdots \\
s_{N-2} &= r\sin\phi_1\ \ldots\sin\phi_{N-1}\ \cos\phi_{N-2} \\
s_{N-1} &= r\sin\phi_1\ \ldots\sin\phi_{N-1}\ \sin\phi_{N-2}\ \cos\theta \\
s_{N-1} &= r\sin\phi_1\ \ldots\sin\phi_{N-1}\ \sin\phi_{N-2}\ \sin\theta
\end{aligned}$$

The Jacobian of transformation is $r^{N-1}\prod_{j=1}^{N-1}\sin^{N-1-j}\phi_j$.Thus we have

$$ds_1 ds_2 \ldots ds_N = dr\, d\theta\, d\phi_1 \ldots d\phi_{N-2}\, r^{N-1}\prod_{j=1}^{N-1}\sin^{N-1-j}\phi_j$$

If the integrand is only function of $\|\boldsymbol{s}\|^2$, we can get rid of the integral over the angles.

$$\int_0^{2\pi} d\theta \prod_{j=1}^{N-2}\int_0^{\pi}\sin^{N-1-j}\phi_j\, d\phi_j = \frac{2\pi^{\frac{N}{2}}}{\Gamma\left(\frac{N}{2}\right)}$$

Mismatched Estimation of Rank-One Non-Symmetric Matrices

8

In this chapter, we delve into the analysis of mismatched estimation focusing on a scenario where the signal matrix is a rank-one non-symmetric matrix and the noise follows a Gaussian distribution. Our main goal is to calculate the full asymptotic of the mismatched MSE in the large N limit, specifically when statisticians use a Gaussian prior for estimation.

- We prove a relation which links the free energy of the system to the mismatched MSE in the non-symmetric case, see Theorem 8.1.
- We derive the asymptotic free energy for the case where true prior is Gaussian in Statement 8.2.
- Using the f-MSE relation, we derive the asymptotic mismatched MSE for Gaussian prior.

8.1 Problem Setting

Suppose the ground-truth vectors $\boldsymbol{s}^* \in \mathbb{R}^N$,$\boldsymbol{t}^* \in \mathbb{R}^M$ are distributed according to $\mathbf{P}_s$, $\mathbf{P}_t$, respectively. The matrix estimation problem is the task to infer the matrix $\boldsymbol{s}^*\boldsymbol{t}^{*\intercal}$ from the noisy observation:

$$\boldsymbol{Y} = \sqrt{\frac{\kappa}{N}}\boldsymbol{s}^*\boldsymbol{t}^{*\intercal} + \boldsymbol{Z} \tag{8.1}$$

This work was presented in [32] F. Pourkamali and N. Macris, "Mismatched estimation of non-symmetric rank-one matrices under gaussian noise," in 2022 IEEE International Symposium on Information Theory (ISIT). IEEE, 2022, pp. 1288–1293.

where κ is the signal-to-noise-ratio (SNR), and the noise matrix $\boldsymbol{Z} \in \mathbb{R}^{N\times M}$ has i.i.d. entries (independent of $\boldsymbol{s}^*, \boldsymbol{t}^*$) distributed according to $\mathcal{N}(0,1)$. We assume that M scales like N, and $N/M \to \alpha \in (0,1]$ as $N \to \infty$. The normalization $1/\sqrt{N}$ makes the inference problem nontrivial in the limit $N \to \infty$.

The statistician is aware that the channel is additive Gaussian with variance 1. But, he does not know the SNR κ and the prior distributions, and he assumes (not necessarily) untrue priors $\mathbf{Q}_u$, $\mathbf{Q}_v$, κ' instead. Following the Bayesian inference framework, he chooses the mean of the posterior distribution as his estimate for $\boldsymbol{s}^*\boldsymbol{t}^{*\intercal}$. The goal is to compute the MSE of this estimation in the $N \to \infty$ limit. The MSE is defined as

$$\mathrm{MSE}_N = \frac{1}{NM}\mathbb{E}\Big[\big\|\boldsymbol{s}^*\boldsymbol{t}^{*\intercal} - \langle \boldsymbol{s}\boldsymbol{t}^{\intercal}\rangle\big\|_{\mathrm{F}}^2\Big] \tag{8.2}$$

where $\mathbb{E}$ is the expectation over the true prior distributions $\mathbf{P}_s, \mathbf{P}_t$, and the noise distribution $\mathbf{P}_Z$, and $\|.\|_{\mathrm{F}}$ is the *Frobenius* norm of a matrix. Adopting the traditional statistical mechanics notation, we use $\langle . \rangle$ to denote the expectation with respect to the posterior distribution from the statistician's point of view with the mismatched priors $\mathbf{Q}_s, \mathbf{Q}_t$ and the SNR κ'.

$$\langle g(\boldsymbol{s},\boldsymbol{t})\rangle = \frac{\iint d\boldsymbol{s}\, d\boldsymbol{t}\, g(\boldsymbol{s},\boldsymbol{t})\, \mathbf{Q}_u(\boldsymbol{s})\, \mathbf{Q}_v(\boldsymbol{t})\, e^{-\frac{1}{2}\left\|\boldsymbol{Y}-\sqrt{\frac{\kappa'}{N}}\boldsymbol{s}\boldsymbol{t}^{\intercal}\right\|_{\mathrm{F}}^2}}{\iint d\boldsymbol{s}\, d\boldsymbol{t}\, \mathbf{Q}_u(\boldsymbol{s})\, \mathbf{Q}_v(\boldsymbol{t})\, e^{-\frac{1}{2}\left\|\boldsymbol{Y}-\sqrt{\frac{\kappa'}{N}}\boldsymbol{s}\boldsymbol{t}^{\intercal}\right\|_{\mathrm{F}}^2}}$$

for a reasonable function $g(\boldsymbol{s},\boldsymbol{t})$ such that the integrals exist.

It is understood that the MSE in (8.2) is a function of $\alpha, \kappa, \mathbf{P}_s, \mathbf{P}_t, \kappa', \mathbf{Q}_s, \mathbf{Q}_t$. Also, note that since we are interested in the large size limit, we only use N in the subscript for the MSE.

In the matched case, i.e. $\kappa' = \kappa, \mathbf{Q}_s = \mathbf{P}_s, \mathbf{Q}_t = \mathbf{P}_t$, the MSE reduces to the MMSE which is the best achievable error:

$$\mathrm{MMSE}_N = \frac{1}{NM}\mathbb{E}\Big[\big\|\boldsymbol{s}^*\boldsymbol{t}^{*\intercal} - \langle \boldsymbol{s}\boldsymbol{t}^{\intercal}\rangle_{\mathrm{matched}}\big\|_{\mathrm{F}}^2\Big] \tag{8.3}$$

Notation: We often drop the N subscript to denote the asymptotic large N limit. We may also drop the parameter dependency for notational simplicity. We use $\mathbb{E}$ to denote the expectation over the true priors and the noise distribution unless the distribution is specified as a subscript. $\langle . \rangle$ denotes the expectation over the posterior distribution with mismatched priors and parameters, unless the parameters are indicated as a subscript.

Remark 8.1. *Studying the problem for the case $\alpha \in (0,1]$ suffices to capture the problem in the general case. Suppose the observation matrix be $\boldsymbol{Y} \in \mathbb{R}^{N\times M}$ as in* (8.1) *such that $N > M$ (so $\alpha > 1$). Exchanging the role of M, N and rescaling κ to κ/α, we can apply our method to the matrix $\boldsymbol{Y}^{\intercal}$ with the aspect ratio $1/\alpha \in (0,1)$.*

8.2 Main Result

In the matched setting, one method for obtaining the MMSE is to compute the mutual information between the signal and the observation, and proceed with the well-known I-MMSE relationship [1]. However, mutual information cannot be defined in a mismatched setting, and such relationships do not exist.

In the mismatched setting, the average free energy (a similar quantity to mutual information) is the key to accessing the mismatched MSE. Free energy is defined as the logarithm of the normalizing factor of the posterior distribution. Theorem 8.1 states the relation of mismatched average free energy and MSE (in the following, we often refer to the mismatched average free energy simply as free energy).

8.2.1 Asymptotic mismatched free energy and MSE

Given an observation matrix $\boldsymbol{Y}$ and the model (8.1), the posterior distribution from the standpoint of the statistician, reads up to a normalizing factor as

$$\begin{aligned}\mathbf{P}(\boldsymbol{s},\boldsymbol{t}|\boldsymbol{Y}) &\propto e^{-\frac{1}{2}\left\|\boldsymbol{Y}-\sqrt{\frac{\kappa'}{N}}\boldsymbol{s}\boldsymbol{t}^{\intercal}\right\|_{\mathrm{F}}^{2}}\mathbf{Q}_s(\boldsymbol{s})\mathbf{Q}_t(\boldsymbol{t})\\ &\propto e^{-\frac{\kappa'}{2N}\|\boldsymbol{s}\|^2\|\boldsymbol{t}\|^2+\sqrt{\frac{\kappa'}{N}}\operatorname{Tr}\boldsymbol{Y}\boldsymbol{t}\boldsymbol{s}^{\intercal}}\mathbf{Q}_s(\boldsymbol{s})\mathbf{Q}_t(\boldsymbol{t})\end{aligned} \tag{8.4}$$

where $\mathbf{Q}_s$,$\mathbf{Q}_t$ are the priors the statistician has assumed. To derive the second line, we drop the term $\left\|\boldsymbol{Y}\right\|_{\mathrm{F}}^{2}$ in the exponent since it is a constant term (because the probability is conditioned on $\boldsymbol{Y}$).

The *partition function* is defined as the normalization factor of the posterior distribution

$$Z(\boldsymbol{Y}) := \iint d\boldsymbol{s}d\boldsymbol{t}\,\mathbf{Q}_s(\boldsymbol{s})\mathbf{Q}_t(\boldsymbol{t})e^{-\frac{\kappa'}{2N}\|\boldsymbol{s}\|^2\|\boldsymbol{t}\|^2+\sqrt{\frac{\kappa'}{N}}\operatorname{Tr}\boldsymbol{Y}\boldsymbol{t}\boldsymbol{s}^{\intercal}} \tag{8.5}$$

and the *mismatched free energy* is defined as

$$f_N(\alpha,\kappa,\mathbf{P}_s,\mathbf{P}_t,\kappa',\mathbf{Q}_s,\mathbf{Q}_t) := -\frac{1}{N}\mathbb{E}\big[\ln Z(\boldsymbol{Y})\big] \tag{8.6}$$

Now, we state the relation between the free energy and MSE.

Theorem 8.1. *For the model* (8.1)*, with any true priors* $\mathbf{P}_s,\mathbf{P}_t$ *and assumed priors* $\mathbf{Q}_s,\mathbf{Q}_t$ *with bounded second moment, we have*

$$\frac{d}{d\kappa'}f_N+\left(2-\sqrt{\frac{\kappa}{\kappa'}}\right)\sqrt{\frac{\kappa}{\kappa'}}\frac{d}{d\kappa}f_N+\frac{1}{2}\frac{1}{N^2}\mathbb{E}_{\mathbf{P}_s}\big[\|\boldsymbol{s}\|^2\big]\mathbb{E}_{\mathbf{P}_t}\big[\|\boldsymbol{t}\|^2\big]=\frac{1}{2}\frac{M}{N}\mathrm{MSE}_N \tag{8.7}$$

The proof of Theorem 8.1 is given in section 8.3. Eq. (8.7), generalizes the classical I-MMSE relation [1], and holds (with a change of constant factors) for other mismatched inference problems such as symmetric matrix estimation [28], and vector channels.

Remark 8.2. *Denote the mismatched distribution of* $\boldsymbol{Y}$ *by* $\mathbf{Q}_{\boldsymbol{Y}}$*. We have*

$$f_N + \frac{\kappa}{2}\frac{1}{N^2}\mathbb{E}_{\mathbf{P}_s}\big[\|\boldsymbol{s}\|^2\big]\mathbb{E}_{\mathbf{P}_t}\big[\|\boldsymbol{t}\|^2\big] - \frac{1}{N}I(\boldsymbol{s}^*, \boldsymbol{t}^*; \boldsymbol{Y}) = \frac{1}{N}D(\mathbf{P}_{\boldsymbol{Y}}\|\mathbf{Q}_{\boldsymbol{Y}}) \tag{8.8}$$

Then, (8.7) *links the difference of MSE and MMSE to a derivative of relative entropy, equivalent to relations discussed in detail in [114] for vector channels. The relations of free energy and mutual information and other information theory notions are discussed in detail in Appendix 8.A.*

8.2.2 Gaussian priors

In this section, we consider the case in which the priors (both true and assumed ones) are Gaussian and derive the mismatched free energy using the non-rigorous results in spherical integrals. Suppose $\boldsymbol{s}$ and $\boldsymbol{t}$ are generated with i.i.d. elements from $\mathcal{N}(0, \sigma_s^2)$, $\mathcal{N}(0, \sigma_t^2)$, and statistician assumes incorrect priors $\mathcal{N}(0, {\sigma'_s}^2)$, $\mathcal{N}(0, {\sigma'_t}^2)$.

Statement 8.2. *For all* $\alpha \in (0, 1], \kappa, \sigma_s, \sigma_t, \kappa', \sigma'_s, \sigma'_t > 0$*, the asymptotic free energy of mismatched inference model (with Gaussian priors) is given by:*

$$\lim_{N\to\infty} f_N = \begin{cases} \frac{1}{2\alpha}\Big(F + (\alpha-1)\ln\big(1+\sqrt{\alpha}\big) & \text{if } \kappa\sigma_s^2\sigma_t^2 < \sqrt{\alpha}, \kappa'{\sigma'_s}^2{\sigma'_t}^2 > \sqrt{\alpha}, \\ \qquad -\alpha\ln\sqrt{\alpha} + \sqrt{\alpha}\Big) & \\ \frac{1}{2\alpha}\Big(F + (\alpha-1)\ln(\alpha+\kappa\sigma^2) & \text{if } \kappa\sigma_s^2\sigma_t^2 > \sqrt{\alpha}, \kappa\kappa'\sigma_s^2\sigma_t^2{\sigma'_s}^2{\sigma'_t}^2 > \alpha, \\ \qquad -\alpha\ln\alpha + \ln\kappa\sigma^2 + \frac{\alpha}{\kappa\sigma^2}\Big) & \\ 0 & \text{otherwise} \end{cases} \tag{8.9}$$

where

$$F = -\sqrt{A} + (1-\alpha)\ln\big(1-\alpha+\sqrt{A}\big) + \ln\kappa'{\sigma'_s}^2{\sigma'_t}^2 - \frac{\alpha}{\kappa'{\sigma'_s}^2{\sigma'_t}^2} + \gamma_{\max}^2 + (\alpha-1)\ln\Big(-2\alpha + \kappa'{\sigma'_s}^2{\sigma'_t}^2\big(1-\alpha+\sqrt{B}\big)\Big)$$

with

$$A = \frac{4\gamma_{\max}^2}{\kappa'{\sigma'_s}^2{\sigma'_t}^2}\Big(\alpha + \kappa'{\sigma'_s}^2{\sigma'_t}^2\big(\gamma_{\max}^2 - \sqrt{B}\big)\Big) + (\alpha-1)^2$$

$$B = (\alpha-1)^2 + 4\alpha\frac{\gamma_{\max}^2}{\kappa'{\sigma'_s}^2{\sigma'_t}^2}$$

$$\gamma_{\max} = \begin{cases} 1+\sqrt{\alpha} & \text{if } \kappa\sigma_s^2\sigma_t^2 < \sqrt{\alpha}, \\ \sqrt{\frac{(1+\kappa\sigma_s^2\sigma_t^2)(\alpha+\kappa\sigma_s^2\sigma_t^2)}{\kappa\sigma_s^2\sigma_t^2}} & \text{if } \kappa\sigma_s^2\sigma_t^2 > \sqrt{\alpha} \end{cases}$$

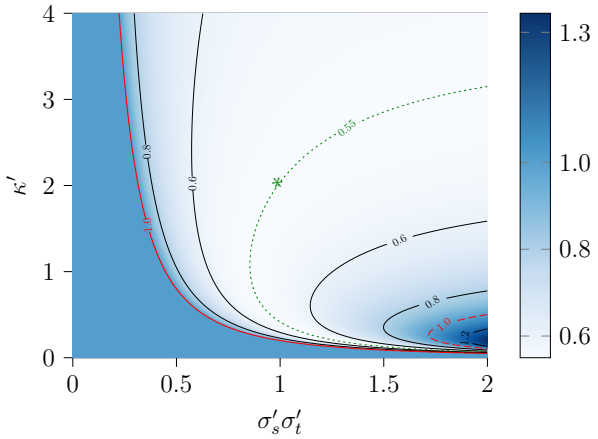


Figure 8.2.1: Plot of mismatched MSE for Gaussian priors with $\alpha = 0.4, \sigma_s = \sigma_t = 1, \kappa = 2$. The solid leftmost (red) curve is a *phase transition* (the derivative of MSE is discontinuous) line. On the left of this curve, $\text{MSE} = \sigma_s^2\sigma_t^2 = 1$. In the *intermediate* region, between the solid leftmost (red) and the dashed (red) curves the MSE is less than $\sigma_s^2\sigma_t^2 = 1$ and the estimation better than chance is possible. On the dotted (green) curve the MSE attains its minimum which is equal to the MMSE, although we do not have $\kappa' = \kappa, \sigma_s\sigma_t = \sigma'_s\sigma'_t$ except for one point (which is the point where we are in fully matched case) depicted by *. The MSE is $\sigma_s\sigma_t = 1$ on the dashed (red) line, and take higher values on the right of this line. MSE is analytic on this curve, and this line is not a phase transition. Finally, we point out that the MSE is continuous throughout and the phase transition is therefore a continuous phase transition.

Proof. The derivation is sketched in section 8.3.2, with details provided in appendix 8.B. □

Now, we are able to compute the asymptotic MSE using Theorem 8.1. We assume that the sequence $(\text{MSE})_{N\geq 1}$ converges uniformly for $(\kappa, \kappa') \in K \subset \mathbb{R}_+^2$. This allows us to interchange limit and derivative to go from asymptotic mismatched free energy to asymptotic mismatched MSE. Due to the complexity of the expression, we omit the detailed presentation of the MSE.

Remark 8.3. *In the matched setting, concavity of mutual information with respect to κ implies uniform convergence of the sequence $(\text{MMSE})_{N\geq 1}$, which enables us to interchange limit and derivative when using I-MMSE relation.*

The MSE is illustrated for the case of $\alpha = 0.4, \sigma_s = \sigma_t = 1, \kappa = 2$ in Fig. 1. The observed behavior is generic when we are above the information theory threshold, $\kappa\sigma_s^2\sigma_t^2 > \sqrt{\alpha}$. In this case, as we see in Fig. 8.2.1, there is an intermediate region where the MSE is less than $\sigma_s^2\sigma_t^2$, and estimation better than chance is possible. We refer to the caption of Fig. 8.2.1 for details. In the case $\kappa\sigma_s^2\sigma_t^2 < \sqrt{\alpha}$, the mutual information between the signal and the observed matrix is zero and the MSE is always greater or equal to $\sigma_s^2\sigma_t^2$.

Fully matched setting

Setting $\kappa' = \kappa$, $\sigma_s'^2 \sigma_t'^2 = \sigma_s^2 \sigma_t^2$ in (8.9), and using (8.8), we find the asymptotic average mutual information:

$$\begin{aligned} I(\boldsymbol{s}^*, \boldsymbol{t}^*; \boldsymbol{Y}) &= f^{\text{fully matched}} + \frac{\kappa}{2\alpha}\sigma_s^2\sigma_t^2 \\ &= \begin{cases} \frac{\kappa}{2\alpha}\sigma_s^2\sigma_t^2 & \text{if } \kappa\sigma_s^2\sigma_t^2 < \sqrt{\alpha} \\ \frac{1}{2\alpha}\Big[\frac{\alpha}{\kappa\sigma_s^2\sigma_t^2} + (1-\alpha)\ln\frac{1+\kappa\sigma_s^2\sigma_t^2}{\alpha+\kappa\sigma_s^2\sigma_t^2} & \text{if } \kappa\sigma_s^2\sigma_t^2 > \sqrt{\alpha} \\ \qquad +(1+\alpha)\ln\kappa\sigma_s^2\sigma_t^2 - \alpha\ln\alpha\Big] \end{cases} \end{aligned} \tag{8.10}$$

From the mutual information, we can find the asymptotic MMSE using the I-MMSE relation. The expressions for mutual information and MMSE are well-known and obtained previously using a variety of methods [10, 11, 115].

Limit $\alpha \to 0$

Setting $n = 1$, the model (8.1) reduces to a modified vector channel of the form

$$\boldsymbol{y} = \sqrt{\kappa} u \boldsymbol{t} + \boldsymbol{z} \tag{8.11}$$

where u is a random variable distributed according to $\mathcal{N}(0, \sigma_s^2)$, and $\boldsymbol{t} \in \mathbb{R}^M$ has i.i.d. elements generated from $\mathcal{N}(0, \sigma_t^2)$. Similar to the matrix case, we can define the partition function for this model, and the free energy is defined as $f_M^{\text{vector}} = -\frac{1}{M}\mathbb{E}\big[\ln Z(\boldsymbol{y})\big]$ (because N is fixed and M tends to ∞). With a slight abuse of notation that we can index the matrix free energy with M, we can see the asymptotic (mismatched) free energy of vector channel as $\lim_{M\to\infty}\lim_{N/M\to 0}\frac{N}{M}f_M$. Assuming that we are permitted to interchange the limits, we find $\lim_{M\to\infty} f_M^{\text{vector}} = \lim_{\alpha\to 0}\alpha f$, where f is the asymptotic mismatched free energy of the matrix model.

$$\lim_{M\to\infty} f_M^{\text{vector}} = \frac{1}{2}\big(\ln(1+\kappa\sigma_s^2\sigma_t^2) - \kappa\sigma_s^2\sigma_t^2\big) \tag{8.12}$$

This result seems rather counter-intuitive, since the mismatched free energy does not depend on the mismatched parameters and it is equal to the matched free energy. But, independent calculations confirms this behavior.

Case of $\alpha = 1$

Consider the symmetric rank-one matrix estimation problem defined as

$$\boldsymbol{Y}^{\text{sym.}} = \sqrt{\frac{\kappa}{N}}\boldsymbol{s}^*\boldsymbol{s}^{*\intercal} + \boldsymbol{Z} \tag{8.13}$$

where $\boldsymbol{s}^* \in \mathbb{R}^N$ has i.i.d. elements distributed according to $\mathcal{N}(0, \sigma^2)$, and $\boldsymbol{Z}$ is a symmetric Gaussian noise matrix with variance 1. The asymptotic average

mutual information for this model reads

$$I(\boldsymbol{s}^*; \boldsymbol{Y}^{\text{sym.}}) = \begin{cases} \frac{1}{4}\kappa\sigma^4 & \text{if } \kappa\sigma^4 < 1 \\ \frac{1}{4}\Big(\frac{1}{\kappa\sigma^4} + 2\ln\kappa\sigma^4\Big) & \text{if } \kappa\sigma^4 \geq 1 \end{cases} \tag{8.14}$$

In the non-symmetric case, let $\alpha = 1$ and $\sigma_s = \sigma_t = \sigma$. Comparing (8.10) and (8.14), we find $2I(\boldsymbol{s}^*; \boldsymbol{Y}^{\text{sym.}}) = I(\boldsymbol{s}^*, \boldsymbol{t}^*; \boldsymbol{Y})$.

The same relation also holds for the mismatched free energy. The mismatched free energy for the symmetric model is computed in previous chapter 7.4 with mismatched parameters κ', σ'. In the non-symmetric case, with $\sigma'_s = \sigma'_t = \sigma'$ we have

$$2f^{\text{sym.}}(\kappa, \sigma, \kappa', \sigma') = f(\alpha = 1, \kappa, \sigma, \sigma, \kappa', \sigma', \sigma')$$

8.3 Analysis

8.3.1 Proof of Theorem 8.1

From (8.1),(8.5), (8.6), we can write

$$f_N = -\frac{1}{N}\mathbb{E}\Big[\ln \iint d\boldsymbol{s}\, d\boldsymbol{t}\, \mathbf{Q}_s(\boldsymbol{s})\mathbf{Q}_t(\boldsymbol{t}) \times e^{-\frac{\kappa'}{2N}\|\boldsymbol{s}\|^2\|\boldsymbol{t}\|^2 + \frac{\sqrt{\kappa\kappa'}}{N}(\boldsymbol{s}^{*\intercal}\boldsymbol{s})(\boldsymbol{t}^{*\intercal}\boldsymbol{t}) + \sqrt{\frac{\kappa'}{n}}\operatorname{Tr}\boldsymbol{Z}\boldsymbol{t}\boldsymbol{s}^\intercal}\Big]$$

From which, we have

$$\frac{d}{d\kappa}f_N = -\frac{1}{2}\frac{1}{N^2}\sqrt{\frac{\kappa'}{\kappa}}\mathbb{E}\Big[\big\langle(\boldsymbol{s}^{*\intercal}\boldsymbol{s})(\boldsymbol{t}^{*\intercal}\boldsymbol{t})\big\rangle\Big]$$

Using a standard Gaussian integration by parts trick, we get

$$\frac{d}{d\kappa'}f_N = \frac{1}{2}\frac{1}{N^2}\mathbb{E}\Big[\big\|\langle\boldsymbol{s}\boldsymbol{t}^\intercal\rangle\big\|_{\text{F}}^2 - \sqrt{\frac{\kappa}{\kappa'}}\big\langle(\boldsymbol{s}^{*\intercal}\boldsymbol{s})(\boldsymbol{t}^{*\intercal}\boldsymbol{t})\big\rangle\Big]$$

Plugging these two equations into (8.7), we find

$$\begin{aligned}\frac{1}{2}\frac{1}{N^2}\mathbb{E}\Big[\big\|\langle\boldsymbol{s}\boldsymbol{t}^\intercal\rangle\big\|_{\text{F}}^2 - 2\big\langle(\boldsymbol{s}^{*\intercal}\boldsymbol{s})(\boldsymbol{t}^{*\intercal}\boldsymbol{t})\big\rangle + \|\boldsymbol{s}^*\|^2\|\boldsymbol{t}^*\|^2\Big] \\ = \frac{1}{2}\frac{1}{N^2}\mathbb{E}\Big[\big\|\langle\boldsymbol{s}\boldsymbol{t}^\intercal\rangle\big\|_{\text{F}}^2 - 2\operatorname{Tr}\boldsymbol{s}^*\boldsymbol{t}^{*\intercal}\langle\boldsymbol{t}\boldsymbol{s}^\intercal\rangle + \|\boldsymbol{s}^*\boldsymbol{t}^{*\intercal}\|_{\text{F}}^2\Big] \\ = \frac{1}{2}\frac{1}{N^2}\mathbb{E}\Big[\big\|\boldsymbol{s}^*\boldsymbol{t}^{*\intercal} - \langle\boldsymbol{s}\boldsymbol{t}^\intercal\rangle\big\|_{\text{F}}^2\Big] = \frac{1}{2}\frac{M}{N}\text{MSE}_N\end{aligned}$$

8.3.2 Asymptotic mismatched free energy for Gaussian priors

The main idea to compute the free energy, is to exploit the rotational invariance of the mismatched priors $\mathbf{Q}_s$,$\mathbf{Q}_t$. Changing variables $\boldsymbol{s} \to \boldsymbol{U}^\intercal\boldsymbol{s}$, $\boldsymbol{t} \to \boldsymbol{V}\boldsymbol{t}$ for

orthogonal matrices $\boldsymbol{U} \in \mathbb{R}^{N\times N}$, $\boldsymbol{V} \in \mathbb{R}^{M\times M}$, the integral in (8.5) becomes ($|\det \boldsymbol{U}| = |\det \boldsymbol{V}| = 1$):

$$\begin{aligned} Z(\boldsymbol{Y}) &= \iint d\boldsymbol{s}\, d\boldsymbol{t}\, \mathbf{Q}_s(\boldsymbol{U}^\intercal \boldsymbol{s})\mathbf{Q}_t(\boldsymbol{V}\boldsymbol{t}) e^{-\frac{\kappa'}{2N}\|\boldsymbol{U}^\intercal \boldsymbol{s}\|^2\|\boldsymbol{V}\boldsymbol{t}\|^2 + \sqrt{\frac{\kappa'}{N}}\operatorname{Tr}\boldsymbol{Y}\boldsymbol{V}\boldsymbol{t}\boldsymbol{s}^\intercal\boldsymbol{U}} \\ &= \iint d\boldsymbol{s}\, d\boldsymbol{t}\, \mathbf{Q}_s(\boldsymbol{s})\mathbf{Q}_t(\boldsymbol{t}) e^{-\frac{\kappa'}{2N}\|\boldsymbol{s}\|^2\|\boldsymbol{t}\|^2 + \sqrt{\frac{\kappa'}{N}}\operatorname{Tr}\boldsymbol{Y}\boldsymbol{V}\boldsymbol{t}\boldsymbol{s}^\intercal\boldsymbol{U}} \end{aligned}$$

We have this relation for any orthogonal matrices $\boldsymbol{U}, \boldsymbol{V}$, so it also holds for the expectation over *Haar* measure on group of orthogonal matrices.

$$Z(\boldsymbol{Y}) = \iint d\boldsymbol{s}\, d\boldsymbol{t}\, \mathbf{Q}_s(\boldsymbol{s})\mathbf{Q}_t(\boldsymbol{t}) e^{-\frac{\kappa'}{2N}\|\boldsymbol{s}\|^2\|\boldsymbol{t}\|^2} \int D\boldsymbol{V} \int D\boldsymbol{U} e^{\sqrt{\frac{\kappa'}{N}}\operatorname{Tr}\boldsymbol{Y}\boldsymbol{V}\boldsymbol{t}\boldsymbol{s}^\intercal\boldsymbol{U}} \tag{8.15}$$

where $D\boldsymbol{U}$, $D\boldsymbol{V}$ denote the *Haar* measure on group of $n \times n$ and $m \times m$ orthogonal matrices, respectively.

We can write the interior integral in (8.15) as in the definition of the rank-one rectangular spherical integral (6.10) with $\boldsymbol{A} = \frac{\boldsymbol{Y}}{\sqrt{M}}$, $\boldsymbol{B} = \frac{\sqrt{\kappa'}}{N}\boldsymbol{t}\boldsymbol{s}^\intercal$ (which is rank-one with non-zero singular-value $\frac{\sqrt{\kappa'}}{N}\|\boldsymbol{s}\|\|\boldsymbol{t}\|$), so (8.15) can be rewritten as

$$Z(\boldsymbol{Y}) = \iint d\boldsymbol{s}\, d\boldsymbol{t}\, \mathbf{Q}_s(\boldsymbol{s})\mathbf{Q}_t(\boldsymbol{t}) e^{-\frac{\kappa'}{2N}\|\boldsymbol{s}\|^2\|\boldsymbol{t}\|^2 + \ln \mathcal{I}_{N,M}(\frac{\sqrt{\kappa'}}{N}\|\boldsymbol{s}\|\|\boldsymbol{t}\|, \frac{\boldsymbol{Y}}{\sqrt{M}})} \tag{8.16}$$

Since the priors $\mathbf{Q}_s$,$\mathbf{Q}_t$ are Gaussian, and the function in the exponent is only a function of $\|\boldsymbol{s}\|$, $\|\boldsymbol{t}\|$, we can use spherical coordinates to reduce the integrals into a two-dimensional integral over $\|\boldsymbol{s}\| \equiv \rho_s$, $\|\boldsymbol{t}\| \equiv \rho_t$. Following by a change of variables $\frac{\rho_s^2}{N} \to \rho_s$, $\frac{\rho_t^2}{M} \to \rho_t$, we can write:

$$\begin{aligned} Z(\boldsymbol{Y}) = \frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{{\sigma'_s}^N}\frac{2^{-\frac{M}{2}}M^{\frac{M}{2}}}{\Gamma(\frac{M}{2})}\frac{1}{{\sigma'_t}^M}\iint_0^{+\infty}\frac{d\rho_s d\rho_t}{\rho_s\rho_t} \\ \times e^{-N\left[\frac{1}{2}(\frac{\rho_s}{{\sigma'_s}^2} - \ln\rho_s) + \frac{1}{2}\frac{M}{N}\left(\frac{\rho_t}{{\sigma'_t}^2} - \ln\rho_t\right) + \frac{\kappa'}{2}\frac{M}{N}\rho_s\rho_t - \frac{1}{N}\ln\mathcal{I}_{N,M}\left(\sqrt{\kappa'\frac{M}{N}\rho_s\rho_t}, \frac{\boldsymbol{Y}}{\sqrt{M}}\right)\right]} \end{aligned} \tag{8.17}$$

By statement 6.3 converges to a deterministic function $\mathcal{J}^{(\alpha)}(\sqrt{\kappa'/\alpha\rho_s\rho_t}, \mu_{\mathrm{MP}})$ where μ_{MP} is the limiting singular law of $\frac{\boldsymbol{Y}}{\sqrt{M}}$. Note that since we are interested in the large N limit, we replaced $M = \frac{N}{\alpha}$.

The limiting singular law of the matrix $\frac{\boldsymbol{Z}}{\sqrt{M}}$ is the *Marchenko-Pastur law* with density $\mu_{\mathrm{MP}} = \frac{\sqrt{4\alpha - (t^2-1-\alpha)^2}}{\pi\alpha t}$ which support is $[1-\sqrt{\alpha}, 1+\sqrt{\alpha}]$ [18]. Moreover, by [74] since $\frac{\boldsymbol{Y}}{\sqrt{M}}$ is the rank-one deformation of $\frac{\boldsymbol{Z}}{\sqrt{M}}$ has the same limiting distribution, but the top singular-value of $\boldsymbol{Y}$ can be different. By [18], the top singular-value of $\frac{\boldsymbol{Y}}{\sqrt{M}}$ converges (a.s.) as $N \to \infty$ to $\gamma_{\max}$ given in (8.9).

The prefactors in (8.17) are independent of the $\boldsymbol{Y}$ and we find:

$$\lim_{N\to\infty} -\frac{1}{N}\ln\frac{2^{-\frac{N}{2}}N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})}\frac{1}{{\sigma'_s}^N}\frac{2^{-\frac{M}{2}}M^{\frac{M}{2}}}{\Gamma(\frac{M}{2})}\frac{1}{{\sigma'_t}^M} = -\frac{1}{2}(1+\frac{1}{\alpha}) + \ln\sigma'_s + \frac{1}{\alpha}\ln\sigma'_t$$

It remains to find the $\lim_{N\to\infty} \mathbb{E}\big[-\frac{1}{N}\ln K_B(\boldsymbol{Y})\big]$, where $K_B(\boldsymbol{Y})$ is the integral in (8.17). Let us define the function

$$\begin{aligned}\psi(\rho_s,\rho_t) = \frac{1}{2}\Big(\frac{\rho_s}{{\sigma'_s}^2} - \ln\rho_s\Big) + \frac{1}{2\alpha}\Big(\frac{\rho_t}{{\sigma'_t}^2} - \ln\rho_t\Big) + \frac{\kappa'}{2\alpha}\rho_s\rho_t \\ - \mathcal{J}^{(\alpha)}\Big(\sqrt{\frac{\kappa'}{\alpha}\rho_s\rho_t}, \mu_{\mathrm{MP}}\Big)\end{aligned} \tag{8.18}$$

We show in appendix 8.B that $\lim_{N\to\infty} \mathbb{E}\big[-\frac{1}{N}\ln K_N(\boldsymbol{Y})\big]$ is bounded above and below by $\min_{\rho_s,\rho_t\geq 0}\psi(\rho_s,\rho_t)$. Therefore, we have:

$$\lim_{N\to\infty} f_N = \min_{\rho_s,\rho_t\geq 0}\psi(\rho_s,\rho_t) - \frac{1}{2}(1+\frac{1}{\alpha}) + \ln\sigma'_s + \frac{1}{\alpha}\ln\sigma'_t \tag{8.19}$$

from which we can derive (8.9).

Appendix

8.A Relations between Free Energy and Information Theory Notions

8.A.1 Matched setting

Denote the true distribution of $\boldsymbol{Y}$ by $\mathbf{P}_{\boldsymbol{Y}}$.

$$\mathbf{P}_{\boldsymbol{Y}}(\boldsymbol{Y}) = \frac{1}{C_{\boldsymbol{Z}}} \iint d\boldsymbol{s}\, d\boldsymbol{t} \mathbf{P}_s(\boldsymbol{s}) \mathbf{P}_t(\boldsymbol{t})\, e^{-\frac{1}{2}\left\|\boldsymbol{Y} - \sqrt{\frac{\kappa}{N}}\boldsymbol{s}\boldsymbol{t}^\intercal\right\|_{\mathrm{F}}^2}$$

where $C_{\boldsymbol{Z}}$ is the normalizing constant of the distribution of the noise matrix $\boldsymbol{Z}$. From the definition of matched partition function (setting $\mathbf{Q}_s = \mathbf{P}_s, \mathbf{Q}_t = \mathbf{P}_t, \kappa' = \kappa$ in (8.5)), we can write $\mathbf{P}_{\boldsymbol{Y}}(\boldsymbol{Y}) = \frac{1}{C_{\boldsymbol{Z}}} e^{-\frac{1}{2}\left\|\boldsymbol{Y}\right\|_{\mathrm{F}}^2} Z^{\text{fully matched}}(\boldsymbol{Y})$. So, the matched free energy can written as

$$\begin{aligned} f_N^{\text{fully matched}} &= -\frac{1}{N}\mathbb{E}_{\mathbf{P}_{\boldsymbol{Y}}}\left[\ln Z^{\text{fully matched}}(\boldsymbol{Y})\right] \\ &= -\frac{1}{N}\mathbb{E}_{\mathbf{P}_{\boldsymbol{Y}}}\left[\ln \mathbf{P}_{\boldsymbol{Y}}(\boldsymbol{Y})\right] - \frac{1}{2}\frac{1}{N}\mathbb{E}_{\mathbf{P}_{\boldsymbol{Y}}}\left[\left\|\boldsymbol{Y}\right\|_{\mathrm{F}}^2\right] - \frac{1}{N}\ln C_{\boldsymbol{Z}} \\ &= \frac{1}{N}h(\boldsymbol{Y}) - \frac{1}{2}\frac{1}{N}\left(\frac{\kappa}{N}\mathbb{E}\left[\|\boldsymbol{s}^*\|^2\right]\mathbb{E}\left[\|\boldsymbol{t}^*\|^2\right] + \mathbb{E}\left[\left\|\boldsymbol{Z}\right\|_{\mathrm{F}}^2\right]\right) - \frac{1}{N}\ln C_{\boldsymbol{Z}} \end{aligned} \tag{8.20}$$

where $h(\boldsymbol{Y})$ is the differential entropy of $\boldsymbol{Y}$. In deriving the last line, we use the independence of $\boldsymbol{s}^*, \boldsymbol{t}^*, \boldsymbol{Z}$.

The average mutual information is

$$\begin{aligned} \frac{1}{N}I(\boldsymbol{Y}; \boldsymbol{s}^*, \boldsymbol{t}^*) &= \frac{1}{N}h(\boldsymbol{Y}) - \frac{1}{N}h(\boldsymbol{Y}|\boldsymbol{s}^*, \boldsymbol{t}^*) \\ &= \frac{1}{N}h(\boldsymbol{Y}) - \frac{1}{N}h(\boldsymbol{Z}) \end{aligned} \tag{8.21}$$

$\boldsymbol{Z}$ is Gaussian matrix with i.i.d. entries, so $h(\boldsymbol{Z}) = \frac{1}{2}\mathbb{E}\left[\left\|\boldsymbol{Z}\right\|_{\mathrm{F}}^2\right] + \ln C_{\boldsymbol{Z}}$.

Putting (8.20), (8.21) together, we find

$$f_N^{\text{fully matched}} + \frac{\kappa}{2}\frac{1}{N^2}\mathbb{E}\left[\|\boldsymbol{s}^*\|^2\right]\mathbb{E}\left[\|\boldsymbol{t}^*\|^2\right] = \frac{1}{N}I(\boldsymbol{Y}; \boldsymbol{s}^*, \boldsymbol{t}^*) \tag{8.22}$$

8.A.2 Mismatched setting

Let $\mathbf{Q}_{\boldsymbol{Y}}$ denote the mismatched distribution of $\boldsymbol{Y}$.

$$\mathbf{Q}_{\boldsymbol{Y}}\big(\boldsymbol{Y}\big) = \frac{1}{C_{\boldsymbol{Z}}} \iint d\boldsymbol{s}\, d\boldsymbol{t} \mathbf{Q}_s(\boldsymbol{s})\mathbf{Q}_t(\boldsymbol{t})\, e^{-\frac{1}{2}\left\|\boldsymbol{Y}-\sqrt{\frac{\kappa'}{N}}\boldsymbol{s}\boldsymbol{t}^\intercal\right\|_{\mathrm{F}}^2}$$

From the definition of mismatched partition function (8.5), we can write $\mathbf{Q}_{\boldsymbol{Y}}\big(\boldsymbol{Y}\big) = \frac{1}{C_{\boldsymbol{Z}}} e^{-\frac{1}{2}\left\|\boldsymbol{Y}\right\|_{\mathrm{F}}^2} Z(\boldsymbol{Y})$. So, the mismatched free energy can written as

$$\begin{aligned} f_N &:= -\frac{1}{N}\mathbb{E}_{\mathbf{P}_{\boldsymbol{Y}}}\big[\ln Z(\boldsymbol{Y})\big] \\ &= -\frac{1}{N}\mathbb{E}_{\mathbf{P}_{\boldsymbol{Y}}}\Big[\ln \mathbf{Q}_{\boldsymbol{Y}}\big(\boldsymbol{Y}\big)\Big] - \frac{1}{N}\frac{1}{2}\mathbb{E}_{\mathbf{P}_{\boldsymbol{Y}}}\Big[\left\|\boldsymbol{Y}\right\|_{\mathrm{F}}^2\Big] - \frac{1}{N}\ln C_{\boldsymbol{Z}} \end{aligned}$$

The relative entropy between the two distributions $\mathbf{P}_{\boldsymbol{Y}}, \mathbf{Q}_{\boldsymbol{Y}}$ reads

$$\begin{aligned} \frac{1}{N}D\big(\mathbf{P}_{\boldsymbol{Y}}\|\mathbf{Q}_{\boldsymbol{Y}}\big) &= \frac{1}{N}\mathbb{E}_{\mathbf{P}_{\boldsymbol{Y}}}\Big[\frac{\mathbf{P}_{\boldsymbol{Y}}\big(\boldsymbol{Y}\big)}{\mathbf{Q}_{\boldsymbol{Y}}\big(\boldsymbol{Y}\big)}\Big] \\ &= -\frac{1}{N}h(\boldsymbol{Y}) + f_N + \frac{1}{N}\frac{1}{2}\mathbb{E}_{\mathbf{P}_{\boldsymbol{Y}}}\Big[\left\|\boldsymbol{Y}\right\|_{\mathrm{F}}^2\Big] + \frac{1}{N}\ln C_{\boldsymbol{Z}} \\ &= f_N - f_N^{\text{fully matched}} \qquad \text{(From second equality in (8.20))} \end{aligned}$$

From (8.22), we can also write

$$f_N + \frac{\kappa}{2}\frac{1}{N^2}\mathbb{E}\big[\|\boldsymbol{s}\|^2\big]\mathbb{E}\big[\|\boldsymbol{t}\|^2\big] = \frac{1}{N}D(\mathbf{P}_{\boldsymbol{Y}}\|\mathbf{Q}_{\boldsymbol{Y}}) + \frac{1}{N}I\big(\boldsymbol{Y};\boldsymbol{s}^*,\boldsymbol{t}^*\big) \tag{8.23}$$

8.B Derivation of Eq. (8.19)

In this section, we present the detailed derivation of asymptotic mismatched free energy. We start with (8.16):

$$\begin{aligned} Z(\boldsymbol{Y}) &= \iint d\boldsymbol{s}\, d\boldsymbol{t}\, \mathbf{Q}_s(\boldsymbol{s})\mathbf{Q}_t(\boldsymbol{t}) e^{-\frac{\kappa'}{2N}\|\boldsymbol{s}\|^2\|\boldsymbol{t}\|^2 + \ln \mathcal{I}_{N,M}\left(\frac{\sqrt{\kappa'}}{N}\|\boldsymbol{s}\|\|\boldsymbol{t}\|, \frac{\boldsymbol{Y}}{\sqrt{M}}\right)} \\ &\overset{(a)}{=} \frac{1}{(2\pi)^{\frac{N}{2}}{\sigma'_s}^N}\frac{1}{(2\pi)^{\frac{M}{2}}{\sigma'_t}^M}\iint d\boldsymbol{s}\, d\boldsymbol{t} \\ &\qquad \times e^{-\frac{\|\boldsymbol{s}\|^2}{2{\sigma'_s}^2} - \frac{\|\boldsymbol{t}\|^2}{2{\sigma'_t}^2} - \frac{\kappa'}{2N}\|\boldsymbol{s}\|^2\|\boldsymbol{t}\|^2 + \ln \mathcal{I}_{N,M}\left(\frac{\sqrt{\kappa'}}{N}\|\boldsymbol{s}\|\|\boldsymbol{t}\|, \frac{\boldsymbol{Y}}{\sqrt{M}}\right)} \\ &\overset{(b)}{=} \frac{2^{-\frac{N}{2}+1}}{\Gamma(\frac{N}{2})}\frac{1}{{\sigma'_s}^N}\frac{2^{-\frac{M}{2}+1}}{\Gamma(\frac{M}{2})}\frac{1}{{\sigma'_t}^M}\iint_0^{+\infty} d\rho_s d\rho_t\, \rho_s^{N-1}\rho_t^{M-1} \\ &\qquad \times e^{-\frac{\rho_s^2}{2{\sigma'_s}^2} - \frac{\rho_t^2}{2{\sigma'_t}^2} - \frac{\beta\kappa'}{2N}\rho_s^2\rho_t^2 + \ln \mathcal{I}_{N,M}\left(\frac{\sqrt{\kappa'}}{N}\rho_s\rho_t, \frac{\boldsymbol{Y}}{\sqrt{M}}\right)} \end{aligned} \tag{8.24}$$

In (a), we write the explicit distribution functions of $\mathbf{Q}_s, \mathbf{Q}_t$, and in (b) we used the spherical coordinates 7.E to change the integrals into a two-dimensional

integral. With change of variables $\frac{\rho_s^2}{N} \to \rho_s$, $\frac{\rho_t^2}{M} \to \rho_t$, we have

$$\begin{aligned} Z(\boldsymbol{Y}) &= \frac{2^{-\frac{N}{2}} N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})} \frac{1}{{\sigma'_s}^N} \frac{2^{-\frac{M}{2}} M^{\frac{M}{2}}}{\Gamma(\frac{M}{2})} \frac{1}{{\sigma'_t}^M} \iint_0^{+\infty} d\rho_s d\rho_t \, \rho_s^{\frac{N}{2}-1} \rho_t^{\frac{M}{2}-1} \\ &\quad \times e^{-N\frac{\rho_s}{2{\sigma'_s}^2} - M\frac{\rho_t}{2{\sigma'_t}^2} - \frac{\beta\kappa'}{2N} N\rho_s M\rho_t + \ln \mathcal{I}_{N,M}(\frac{\sqrt{\kappa'}}{N}\sqrt{MN\rho_s\rho_t}, \frac{\boldsymbol{Y}}{\sqrt{M}})} \\ &= \frac{2^{-\frac{N}{2}} N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})} \frac{1}{{\sigma'_s}^N} \frac{2^{-\frac{M}{2}} M^{\frac{M}{2}}}{\Gamma(\frac{M}{2})} \frac{1}{{\sigma'_t}^M} \iint_0^{+\infty} \frac{d\rho_s d\rho_t}{\rho_s \rho_t} \\ &\quad \times e^{-N\left[\frac{1}{2}(\frac{\rho_s}{{\sigma'_s}^2} - \ln\rho_s) + \frac{1}{2}\frac{M}{N}\left(\frac{\rho_t}{{\sigma'_t}^2} - \ln\rho_t\right) + \frac{\kappa'}{2}\frac{M}{N}\rho_s\rho_t - \frac{1}{N}\mathcal{I}_{N,M}\left(\sqrt{\kappa'\frac{M}{N}\rho_s\rho_t}, \frac{\boldsymbol{Y}}{\sqrt{M}}\right)\right]} \end{aligned} \tag{8.25}$$

where $\frac{1}{N}\mathcal{I}_{N,M}\big(\sqrt{\kappa'\frac{M}{N}\rho_s\rho_t}, \frac{\boldsymbol{Y}}{\sqrt{M}}\big)$, by statement 6.3 converges to a deterministic function $\mathcal{J}^{(\alpha)}(.;\mu_{\mathrm{MP}})$ where μ_{MP} is the limiting singular law of $\frac{\boldsymbol{Y}}{\sqrt{M}}$.

We are interested in $\lim_{N\to\infty} f_N = \lim_{N\to\infty} \mathbb{E}\big[-\frac{1}{N}\ln Z(\boldsymbol{Y})\big]$. First, we compute the asymptotic of the integral in (8.25), denoted from now on by $K(\boldsymbol{Y})$. Let us define the function

$$\begin{aligned} \psi(\rho_s,\rho_t) = \frac{1}{2}\big(\frac{\rho_s}{{\sigma'_s}^2} - \ln\rho_s\big) + \frac{1}{2\alpha}\big(\frac{\rho_t}{{\sigma'_t}^2} - \ln\rho_t\big) + \frac{\kappa'}{2\alpha}\rho_s\rho_t \\ - \mathcal{J}^{(\alpha)}\big(\sqrt{\frac{\kappa'}{\alpha}\rho_s\rho_t}, \mu_{\mathrm{MP}}\big) \end{aligned} \tag{8.26}$$

We show that $\lim_{N\to\infty}\mathbb{E}\big[-\frac{1}{N}\ln K(\boldsymbol{Y})\big] = \min_{\rho_s,\rho_t\geq 0}\psi(\rho_s,\rho_t)$. Assume that the minimum of $\psi(\rho_s,\rho_t)$ is attained at (ρ_s^*,ρ_t^*).

To show this limit, we need the following lemma:

Lemma 8.1. *For $\theta \geq 0$,*

$$\lim_{n\to\infty}\mathbb{E}\Big[\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\theta, \frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big] = \mathcal{J}^{(\alpha)}(\theta,\mu_{\mathrm{MP}})$$

and this convergence is uniform over compacts.

Proof. By Statement 6.3, $\lim_{N\to\infty}\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\theta,\frac{\boldsymbol{Y}}{\sqrt{M}}\big) = \mathcal{J}^{(\alpha)}(\theta,\mu_{\mathrm{MP}})$, and this convergence is an almost sure convergence.

Let $\frac{\boldsymbol{Y}}{\sqrt{M}} = \boldsymbol{U}_{\boldsymbol{Y}}\boldsymbol{\Gamma}_{\boldsymbol{Y}}\boldsymbol{V}_{\boldsymbol{Y}}$ be singular value decomposition of $\frac{\boldsymbol{Y}}{\sqrt{M}}$, and $\boldsymbol{\Sigma}_\theta$ be an $M\times N$ with all zeros entries except the first entry which is θ. We can write:

$$\begin{aligned} \mathcal{I}_{N,M}\big(\theta,\frac{\boldsymbol{Y}}{\sqrt{M}}\big) &= \int D\boldsymbol{U}\int D\boldsymbol{V} e^{\sqrt{MN}\,\mathrm{Tr}\,\boldsymbol{\Gamma}_{\boldsymbol{Y}}\boldsymbol{U}\boldsymbol{\Sigma}_\theta\boldsymbol{V}} \\ &= \int D\boldsymbol{U}\int D\boldsymbol{t} e^{\sqrt{MN}\theta\sum_{i=1}^n \gamma_i U_{i1}V_{1i}} \\ &\leq \int D\boldsymbol{U}\int D\boldsymbol{V} e^{\sqrt{mn}\theta\gamma_{\max}^{(\boldsymbol{Y})}} \\ &= e^{\sqrt{MN}\theta\gamma_{\max}^{(\boldsymbol{Y})}} \end{aligned}$$

where $\gamma_{\max}^{(\boldsymbol{Y})}$ is the top singular-value of $\frac{\boldsymbol{Y}}{\sqrt{M}}$. So, we have that $\frac{1}{N}\ln \mathcal{I}_{N,M}\big(\theta, \frac{\boldsymbol{Y}}{\sqrt{M}}\big) \leq \sqrt{\frac{M}{N}}\theta\gamma_{\max}^{(\boldsymbol{Y})}$. By [116] (Theorem II.13), for any N, M we have

$$\mathbf{P}\Big\{\gamma_{\max}^{(\boldsymbol{Y})} \geq 1 + \sqrt{\frac{N}{M}} + t\Big\} < e^{-M\frac{t^2}{2}}$$

From this and since $\gamma_{\max}^{(\boldsymbol{Y})}$ is a positive random variable, we can deduce that the sequence $\frac{1}{N}\ln \mathcal{I}_{N,M}\big(\gamma_{\max}^{(\boldsymbol{Y})}\big)_{N,M}$ is uniformly integrable. Because $\frac{1}{N}\ln \mathcal{I}_{N,M}\big(\theta, \frac{\boldsymbol{Y}}{\sqrt{M}}\big) \leq \sqrt{\frac{M}{N}}\theta\gamma_{\max}^{(\boldsymbol{Y})}$, so $\Big(\frac{1}{N}\ln \mathcal{I}_{N,M}\big(\theta, \frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big)_{N,M}$ is also uniformly integrable. Therefore, the almost sure convergence of $\frac{1}{N}\ln \mathcal{I}_{N,M}\big(\theta, \frac{\boldsymbol{Y}}{\sqrt{M}}\big)$ to $\mathcal{J}^{(\alpha)}(\theta, \mu_{\mathrm{MP}})$ implies $\lim_{N\to\infty}\mathbb{E}\Big[\frac{1}{N}\ln \mathcal{I}_{N,M}\big(\theta, \frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big] = \mathcal{J}^{(\alpha)}(\theta, \mu_{\mathrm{MP}})$.

Moreover, similar to Lemma 14 in [62], one can show that for $\theta, \theta' \geq 0$:

$$\Big|\frac{1}{N}\ln \mathcal{I}_{N,M}\big(\theta, \frac{\boldsymbol{Y}}{\sqrt{M}}\big) - \frac{1}{N}\ln \mathcal{I}_{N,M}\big(\theta', \frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big| \leq \sqrt{\frac{M}{N}}\gamma_{\max}^{(\boldsymbol{Y})}|\theta - \theta'|$$

Therefore, using Jensen's inequality, we can write

$$\begin{aligned}
&\Big|\mathbb{E}\Big[\frac{1}{N}\ln \mathcal{I}_{N,M}\big(\theta, \frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big] - \mathbb{E}\Big[\frac{1}{N}\ln \mathcal{I}_{N,M}\big(\theta', \frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big]\Big| \\
&\qquad \leq \mathbb{E}\Big[\Big|\frac{1}{N}\ln \mathcal{I}_{N,M}\big(\theta, \frac{\boldsymbol{Y}}{\sqrt{M}}\big) - \frac{1}{N}\ln \mathcal{I}_{N,M}\big(\theta', \frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big|\Big] \\
&\qquad \leq \mathbb{E}\big[\sqrt{\frac{M}{N}}\gamma_{\max}^{(\boldsymbol{Y})}|\theta - \theta'|\big] \\
&\qquad = \sqrt{\frac{M}{N}}\mathbb{E}\big[\gamma_{\max}^{(\boldsymbol{Y})}\big]|\theta - \theta'| \\
&\qquad \leq \sqrt{\frac{M}{N}}\mathbb{E}\big[\sqrt{\frac{\kappa}{NM}}\|\boldsymbol{s}^*\|\|\boldsymbol{t}^*\| + \gamma_{\max}^{(\boldsymbol{Z})}\big]|\theta - \theta'| \\
&\qquad \leq \sqrt{\frac{M}{N}}\Big(\sqrt{\kappa}\sigma_s\sigma_t + 1 + \sqrt{\frac{N}{M}}\Big)|\theta - \theta'| \\
&\qquad \leq C|\theta - \theta'|
\end{aligned} \tag{8.27}$$

for a constant C and for N large enough. $\gamma_{\max}^{(\boldsymbol{Z})}$ denotes the top singular-value of the matrix $\boldsymbol{Z} \in \mathbb{R}^{N\times M}$, and by [116] $\mathbb{E}\big[\gamma_{\max}^{(\boldsymbol{Z})}\big] \leq 1 + \sqrt{\frac{N}{M}}$. In deriving the last inequality, we also used the fact that for a Gaussian vector $\boldsymbol{x} \in \mathbb{R}^N$ with i.i.d. elements of variance σ^2, $\mathbb{E}\big[\|\boldsymbol{x}\|\big] \leq \sqrt{N}\sigma$.

Therefore, the family of functions $\Big\{\mathbb{E}\Big[\frac{1}{N}\ln \mathcal{I}_{N,M}\big(\theta, \frac{\boldsymbol{Y}}{\sqrt{m}}\big)\Big]\Big\}$ is uniformly equicontinuous. So, the point-wise convergence to $\mathcal{J}^{(\alpha)}(\theta, \mu_{\mathrm{MP}})$ is uniform over compacts. □

8.B.1 Upper bound

For any N, M, we have:

$$\begin{aligned} K(\boldsymbol{Y}) &= \iint_0^{+\infty} \frac{d\rho_s d\rho_t}{\rho_s \rho_t} \\ &\quad \times e^{-N\left[\frac{1}{2}(\frac{\rho_s}{\sigma_s'^2}-\ln\rho_s)+\frac{1}{2}\frac{M}{N}\left(\frac{\rho_t}{\sigma_t'^2}-\ln\rho_t\right)+\frac{\kappa'}{2}\frac{M}{N}\rho_s\rho_t-\frac{1}{N}\ln\mathcal{I}_{N,M}\left(\sqrt{\kappa'\frac{M}{N}\rho_s\rho_t},\frac{\boldsymbol{Y}}{\sqrt{M}}\right)\right]} \\ &\geq \frac{1}{\rho_s^*\rho_t^*} e^{-N\left[\frac{1}{2}(\frac{\rho_s^*}{\sigma_s'^2}-\ln\rho_s^*)+\frac{1}{2}\frac{M}{N}\left(\frac{\rho_t^*}{\sigma_t'^2}-\ln\rho_t^*\right)+\frac{\kappa'}{2}\frac{M}{N}\rho_s^*\rho_t^*-\frac{1}{N}\ln\mathcal{I}_{N,M}\left(\sqrt{\kappa'\frac{M}{N}\rho_s^*\rho_t^*},\frac{\boldsymbol{Y}}{\sqrt{M}}\right)\right]} \end{aligned} \tag{8.28}$$

So,

$$\begin{aligned} \mathbb{E}\big[-\frac{1}{N}\ln K(\boldsymbol{Y})\big] \leq \frac{1}{2}(\frac{\rho_s^*}{\sigma_s'^2}-\ln\rho_s^*)+\frac{1}{2}\frac{M}{N}(\frac{\rho_t^*}{\sigma_t'^2}-\ln\rho_t^*)+\frac{\kappa'}{2}\frac{M}{N}\rho_s^*\rho_t^* \\ -\mathbb{E}\Big[\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\sqrt{\kappa'\frac{M}{N}\rho_s^*\rho_t^*},\frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big]+\frac{1}{N}\ln\rho_s^*\rho_t^* \end{aligned} \tag{8.29}$$

Since the convergence of $\mathbb{E}\Big[\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\theta,\frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big] \xrightarrow{N\to\infty} \mathcal{J}^{(\alpha)}(\theta,\mu_{\mathrm{MP}})$ is uniform over compacts, and $\sqrt{\kappa'\frac{M}{N}\rho_s^*\rho_t^*} \xrightarrow{N\to\infty} \sqrt{\frac{\kappa'}{\alpha}\rho_s^*\rho_t^*}$, so

$$\mathbb{E}\Big[\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\sqrt{\kappa'\frac{M}{N}\rho_s^*\rho_t^*},\frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big] \xrightarrow{N\to\infty} \mathcal{J}^{(\alpha)}(\sqrt{\frac{\kappa'}{\alpha}\rho_s^*\rho_t^*},\mu_{\mathrm{MP}})$$

Therefore, we obtain

$$\lim_{N\to\infty}\mathbb{E}\big[-\frac{1}{N}\ln K(\boldsymbol{Y})\big] \leq \psi(\rho_s^*,\rho_t^*) \tag{8.30}$$

8.B.2 Lower bound

Denote $a_N(\rho_s,\rho_t)=\frac{1}{2}(\frac{\rho_s}{\sigma_s'^2}-\ln\rho_s)+\frac{1}{2}\frac{M}{N}\big(\frac{\rho_t}{\sigma_t'^2}-\ln\rho_t\big)$, and $b_N(\rho_s,\rho_t)=\frac{\kappa'}{2}\frac{M}{N}\rho_s\rho_t-\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\sqrt{\kappa'\frac{M}{N}\rho_s\rho_t},\frac{\boldsymbol{Y}}{\sqrt{M}}\big)$. Let (ρ_s',ρ_t') be the minimizer of $a_N(\rho_s,\rho_t)+b_N(\rho_s,\rho_t)$. Since $\frac{N}{M}\to\alpha$, for sufficiently large N, we have $\rho_s'<N$, $\rho_t'<N$.

We can split the integral in $K(\boldsymbol{Y})$ over different regions in $\mathbb{R}_+^2$, $D_1=(\rho_s'-\epsilon,\rho_s'+\epsilon)\times(\rho_t'-\epsilon,\rho_t'+\epsilon)$ for small constant $\epsilon>0$, $D_2=[0,N]\times[0,N]\backslash D_1$, $D_3=\mathbb{R}_+^2\backslash(D_1\cup D_2)$. For the first part, choosing ϵ small enough, we have:

$$\begin{aligned} &\iint_{D_1}\frac{d\rho_s d\rho_t}{\rho_s\rho_t}e^{-N\left[a_N(\rho_s,\rho_t)+b_N(\rho_s,\rho_t)\right]} \\ &\qquad\leq e^{-N\min_{\rho_s,\rho_t\in D_1}\left[a_N(\rho_s,\rho_t)+b_N(\rho_s,\rho_t)\right]}\iint_{D_1}\frac{d\rho_s d\rho_t}{\rho_s\rho_t} \\ &\qquad= e^{-N\min_{\rho_s,\rho_t\in D_1}\left[a_N(\rho_s,\rho_t)+b_N(\rho_s,\rho_t)\right]}\ln\frac{\rho_s'+\epsilon}{\rho_s'-\epsilon}\ln\frac{\rho_t'+\epsilon}{\rho_t'-\epsilon} \\ &\qquad\leq e^{-N\min_{\rho_s,\rho_t\in D_1}\left[a_N(\rho_s,\rho_t)+b_N(\rho_s,\rho_t)\right]} \end{aligned}$$

D_2 is a closed, bounded, and connected subset of $\mathbb{R}_+^2$, so by integral mean value theorem, there exists a point $(\hat{\rho}_s, \hat{\rho}_t)$ s.t.

$$\begin{aligned}\iint_{D_2} \frac{d\rho_s d\rho_t}{\rho_s \rho_t} e^{-N\left[a_N(\rho_s,\rho_t)+b_N(\rho_s,\rho_t)\right]} &\leq N^2 \frac{e^{-N\left[a_N(\hat{\rho}_s,\hat{\rho}_t)+b_N(\hat{\rho}_s,\hat{\rho}_t)\right]}}{\hat{\rho}_s \hat{\rho}_t} \\ &= C_{N,2} e^{-N\left[a_N(\hat{\rho}_s,\hat{\rho}_t)+b_N(\hat{\rho}_s,\hat{\rho}_t)\right]}\end{aligned}$$

For the last part, we have

$$\begin{aligned}&\iint_{D_3} \frac{d\rho_s d\rho_t}{\rho_s \rho_t} e^{-N\left[a_N(\rho_s,\rho_t)+b_N(\rho_s,\rho_t)\right]} \\ &\qquad \leq e^{-N \min_{\rho_s,\rho_t \in D_3} b_N(\rho_s,\rho_t)} \iint_{D_3} \frac{d\rho_s d\rho_t}{\rho_s \rho_t} e^{-N a_N(\rho_s,\rho_t)}\end{aligned}$$

For N large enough, since $M \geq N$, the integral in the RHS can be upper bounded as below:

$$\begin{aligned}&\iint_{D_3} \frac{d\rho_s d\rho_t}{\rho_s \rho_t} e^{-N\frac{1}{2}\left(\frac{\rho_s}{{\sigma'_s}^2}-\ln \rho_s\right)-M\frac{1}{2}\left(\frac{\rho_t}{{\sigma'_t}^2}-\ln \rho_t\right)} \\ &\qquad \leq \iint_{D_3} \frac{d\rho_s d\rho_t}{\rho_s \rho_t} e^{-N\left[\frac{1}{2}\left(\frac{\rho_s}{{\sigma'_s}^2}-\ln \rho_s\right)+\frac{1}{2}\left(\frac{\rho_t}{{\sigma'_t}^2}-\ln \rho_t\right)\right]}\end{aligned}$$

This integral can be well approximated by Laplace method (for N large). The minimum of the function in the exponent (in D_3 is attained on the boundaries at either $({\sigma'_s}^2, N)$ or $({\sigma'_t}^2, N)$. Without loss of generality, assume that the minimum is at $({\sigma'_s}^2, N)$.

$$\begin{aligned}&\iint_{D_3} \frac{d\rho_s d\rho_t}{\rho_s \rho_t} e^{-N\left[\frac{1}{2}\left(\frac{\rho_s}{{\sigma'_s}^2}-\ln \rho_s\right)+\frac{1}{2}\left(\frac{\rho_t}{{\sigma'_t}^2}-\ln \rho_t\right)\right]} \\ &\qquad = C_{N,3} e^{-N\left[\frac{1}{2}(1-\ln {\sigma'_s}^2)+\frac{1}{2}\left(\frac{N}{{\sigma'_t}^2}-\ln N\right)\right]} \left(1+O\left(\frac{1}{N}\right)\right)\end{aligned}$$

where $C_{N,3}$ is a polynomial in N corresponding to the terms in Laplace approximation. So, we have

$$\begin{aligned}&\iint_{D_3} \frac{d\rho_s d\rho_t}{\rho_s \rho_t} e^{-N\left[a_N(\rho_s,\rho_t)+b_N(\rho_s,\rho_t)\right]} \\ &\qquad \leq C_{N,3} e^{-N\left[\frac{1}{2}(1-\ln {\sigma'_s}^2)+\frac{1}{2}\left(\frac{N}{{\sigma'_t}^2}-\ln N\right)+\min_{\rho_s,\rho_t \in D_3} b_N(\rho_s,\rho_t)\right]}\end{aligned}$$

Putting all together, for N large we have

$$K(\boldsymbol{Y}) \leq e^{-N \min\left[a_N(\rho_s,\rho_t)+b_N(\rho_s,\rho_t)\right]} \left(1 + C_{N,2} e^{-N r_{N,2}} + C_{N,3} e^{-N r_{N,3}}\right)$$

where $r_{N,2}$,$r_{N,3}$ are positive constants (depending on N).

Therefore,

$$\begin{aligned}
&\mathbb{E}\big[-\frac{1}{N}\ln K(\boldsymbol{Y})\big] \\
&\geq \mathbb{E}\Big[\min\big[a_N(\rho_s,\rho_t)+b_N(\rho_s,\rho_t)\big]-\frac{1}{N}\ln\big(1+C_{N,2}e^{-Nr_{N,2}}+C_{N,3}e^{-Nr_{N,3}}\big)\Big]
\end{aligned}$$

Second term is bounded by $\frac{1}{N}\ln C_N$, where C_N is polynomial in N. By an application dominated convergence theorem and continuity of log function, this term vanishes in $N\to\infty$. For the first term, we have

$$\begin{aligned}
&\mathbb{E}\big[\min a_N(\rho_s,\rho_t)+b_N(\rho_s,\rho_t)\big] \\
&=\mathbb{E}\Big[\min_{\rho_s,\rho_t}\frac{1}{2}\big(\frac{\rho_s}{\sigma_s'^2}-\ln\rho_s\big)+\frac{1}{2}\frac{M}{N}\big(\frac{\rho_t}{\sigma_t'^2}-\ln\rho_t\big)+\frac{\kappa'}{2}\frac{M}{N}\rho_s\rho_t \\
&\qquad\qquad -\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\sqrt{\kappa'\frac{M}{N}\rho_s\rho_t},\frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big] \\
&=\mathbb{E}\Big[\min_{\rho_s,\rho_t}\frac{1}{2}\big(\frac{\rho_s}{\sigma_s'^2}-\ln\rho_s\big)+\frac{1}{2}\frac{M}{N}\big(\frac{\rho_t}{\sigma_t'^2}-\ln\rho_t\big)+\frac{\kappa'}{2}\frac{M}{N}\rho_s\rho_t \\
&\qquad\qquad -\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\sqrt{\kappa'\frac{M}{N}\rho_s\rho_t},\frac{\boldsymbol{Y}}{\sqrt{M}}\big)+\psi(\rho_s,\rho_t)-\psi(\rho_s,\rho_t)\Big] \\
&\geq \min_{\rho_s,\rho_t}\psi(\rho_s,\rho_t) \\
&\qquad+\min_{\rho_s,\rho_t}\frac{1}{2}\frac{M}{N}\big(\frac{\rho_t}{\sigma_t'^2}-\ln\rho_t\big)+\frac{\kappa'}{2}\frac{M}{N}\rho_s\rho_t-\Big[\frac{1}{2\alpha}\big(\frac{\rho_t}{\sigma_t'^2}-\ln\rho_t\big)+\frac{\kappa'}{2\alpha}\rho_s\rho_t\Big] \\
&\qquad+\mathbb{E}\Big[\min_{\rho_s,\rho_t}\mathcal{J}^{(\alpha)}\big(\sqrt{\frac{\kappa'}{\alpha}\rho_s\rho_t},\mu_{\mathrm{MP}}\big)-\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\sqrt{\kappa'\frac{M}{N}\rho_s\rho_t},\frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big]
\end{aligned}$$

In the limit $N\to\infty$, it can easily be shown that the second term vanishes. We show that the third term also vanishes. Similar to the proof of compact convergence of $\mathbb{E}\Big[\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\theta,\frac{\boldsymbol{Y}}{\sqrt{m}}\big)\Big]$, we have:

$$\begin{aligned}
&\mathbb{E}\Big[\min_{\rho_s,\rho_t}\mathcal{J}^{(\alpha)}\big(\sqrt{\frac{\kappa'}{\alpha}\rho_s\rho_t},\mu_{\mathrm{MP}}\big)-\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\sqrt{\kappa'\frac{M}{N}\rho_s\rho_t},\frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big] \\
&\quad\geq\mathbb{E}\Big[\min_{\rho_s,\rho_t}\mathcal{J}^{(\alpha)}\big(\sqrt{\frac{\kappa'}{\alpha}\rho_s\rho_t},\mu_{\mathrm{MP}}\big)-\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\sqrt{\frac{\kappa'}{\alpha}\rho_s\rho_t},\frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big] \\
&\qquad+\mathbb{E}\Big[\min_{\rho_s,\rho_t}\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\sqrt{\frac{\kappa'}{\alpha}\rho_s\rho_t},\frac{\boldsymbol{Y}}{\sqrt{M}}\big)-\frac{1}{N}\ln\mathcal{I}_{N,M}\big(\sqrt{\kappa'\frac{M}{N}\rho_s\rho_t},\frac{\boldsymbol{Y}}{\sqrt{M}}\big)\Big]
\end{aligned}$$

where both terms converges to 0 as $N\to\infty$ (see (8.27)).

Thus, we find

$$\lim_{N\to\infty}\mathbb{E}\big[-\frac{1}{N}\ln K(\boldsymbol{Y})\big]\geq\psi(\rho_s^*,\rho_t^*) \tag{8.31}$$

From (8.30),(8.31), we have

$$\lim_{N\to\infty}\mathbb{E}\big[-\frac{1}{N}\ln K(\boldsymbol{Y})\big]=\psi(\rho_s^*,\rho_t^*) \tag{8.32}$$

8.B.3 Constant terms

Lemma 8.2. *For $N \to \infty$,*

$$\Gamma\big(\frac{N}{2}\big) = \sqrt{\pi N}\big(\frac{N}{2}\big)^{\frac{N}{2}-1} e^{-\frac{N}{2}} \Big(1 + O\big(\frac{1}{N}\big)\Big)$$

Proof. From [113], $\Gamma(\frac{N}{2})$ can be approximated by

$$\sqrt{2\pi(\frac{N}{2}-1)}\big(\frac{N}{2}-1\big)^{\frac{N}{2}-1} e^{-\frac{N}{2}+1}\Big(1 + O\big(\frac{1}{N}\big)\Big),$$

from which the result follows. □

Using this lemma, we have:

$$\lim_{N\to\infty} -\frac{1}{N}\ln \frac{2^{-\frac{N}{2}} N^{\frac{N}{2}}}{\Gamma(\frac{N}{2})} \frac{1}{{\sigma'_s}^N} \frac{2^{-\frac{M}{2}} M^{\frac{M}{2}}}{\Gamma(\frac{M}{2})} \frac{1}{{\sigma'_t}^M} = -\frac{1}{2}(1+\frac{1}{\alpha}) + \ln \sigma'_s + \frac{1}{\alpha}\ln \sigma'_t \quad (8.33)$$

From (8.32),(8.33), we get

$$\lim_{N\to\infty} f_N = \min_{\rho_s,\rho_t} \psi(\rho_s,\rho_t) - \frac{1}{2}(1+\frac{1}{\alpha}) + \ln \sigma'_s + \frac{1}{\alpha}\ln \sigma'_t \quad (8.34)$$

8.C Computation of free energy

In this section, we present the computation of free energy from the expression $\min_{\rho_s,\rho_t} \psi(\rho_s,\rho_t)$, where $\psi(.,.)$ is defined in (8.18). First, we find the expression of the function $\mathcal{J}^{(\alpha)}(.;\mu_{\mathrm{MP}})$.

8.C.1 $\mathcal{J}^{(\alpha)}(.,\mu_{\mathrm{MP}})$

From Statement 6.3, we have that

$$\mathcal{J}^{(\alpha)}(.;\mu_{\mathrm{MP}}) = \nu - \frac{1}{2\alpha}\ln(1+\alpha\nu) - \frac{1}{2}\ln(1+\nu) - \frac{1}{2}\int d\mu_{\mathrm{MP}}(t) \ln\big(1 - \frac{\theta^2}{T^{(\alpha)}(\nu)} t^2\big)$$

with

$$\nu = \begin{cases} C^{(\alpha)}_{\mu_{\mathrm{MP}}}(\theta^2) & \text{if } \theta^2 < \mathcal{H}_{\max} \\ {T^{(\alpha)}}^{-1}(\theta^2 \gamma^2_{\max}) & \text{if } \theta^2 \geq \mathcal{H}_{\max} \end{cases}$$

First, we compute rectangular R-transform for the Marchenko-Pastur distribution $\mu_{\mathrm{MP}} = \frac{\sqrt{4\alpha - (t^2-1-\alpha)^2}}{\pi\alpha t}$. We have

$$\mathcal{H}^{(\alpha)}_{\mu_{\mathrm{MP}}}(z) = \frac{1-(1+\alpha)z - \sqrt{\big[1-(1+\alpha)z\big]^2 - 4\alpha z^2}}{2\alpha z}, \quad {\mathcal{H}^{(\alpha)}_{\mu_{\mathrm{MP}}}}^{-1}(z) = \frac{z}{T^{(\alpha)}(z)}$$

So, we have that $C^{(\alpha)}_{\mu_{\mathrm{MP}}}(z) = z$.

By [18], the top singular-value of the matrix $\boldsymbol{Y}$ as $n \to \infty$ converges to $\gamma_{\max} = \begin{cases} 1+\sqrt{\alpha} & \text{if } \kappa\sigma_s^2\sigma_t^2 \le \sqrt{\alpha} \\ \sqrt{\frac{(1+\kappa\sigma_s^2\sigma_t^2)(\alpha+\kappa\sigma_s^2\sigma_t^2)}{\kappa\sigma_s^2\sigma_t^2}} & \text{if } \kappa\sigma_s^2\sigma_t^2 \ge \sqrt{\alpha} \end{cases}$. Thus, we can compute

$$\mathcal{H}_{\max} = \begin{cases} \frac{1}{\sqrt{\alpha}} & \text{if } \kappa\sigma_s^2\sigma_t^2 \le \sqrt{\alpha} \\ \frac{1}{\kappa\sigma_s^2\sigma_t^2} & \text{if } \kappa\sigma_s^2\sigma_t^2 \ge \sqrt{\alpha} \end{cases}$$

Moreover, we have

$$\begin{aligned} \int d\mu_{\mathrm{MP}}(t) \ln\left(1-zt^2\right) = &-\frac{1-(1+\alpha)z-\sqrt{\left[1-(1+\alpha)z\right]^2-4\alpha z^2}}{2\alpha z} + \ln 2 \\ &- \frac{\alpha+1}{2\alpha}\ln\left(1-(1+\alpha)z+\sqrt{\left[1-(1+\alpha)z\right]^2-4\alpha z^2}\right) \\ &+ \frac{1-\alpha}{2\alpha}\ln\left(\alpha+1-(-1+\alpha)^2 z-(-1+\alpha)\sqrt{\left[1-(1+\alpha)z\right]^2-4\alpha z^2}\right) \end{aligned} \tag{8.35}$$

Putting all together, $\mathcal{J}^{(\alpha)}(\theta, \mu_{\mathrm{MP}})$ can be expressed as the following:
If $\kappa\sigma_s^2\sigma_t^2 \le \sqrt{\alpha}$:

$$\mathcal{J}^{(\alpha)}(\theta,\mu_{\mathrm{MP}}) = \begin{cases} \frac{1}{2}\theta^2 & \text{if } \theta^2 \le \frac{1}{\sqrt{\alpha}} \\ \frac{\sqrt{\alpha}+1}{2\alpha}\left(\sqrt{4\alpha\theta^2+\left(1-\sqrt{\alpha}\right)^2}-1\right) - \ln\theta - \frac{1}{2} & \text{if } \theta^2 \ge \frac{1}{\sqrt{\alpha}} \\ \quad + \frac{1-\alpha}{2\alpha}\ln\left(\frac{2}{\sqrt{4\alpha\theta^2+\left(1-\sqrt{\alpha}\right)^2}-\sqrt{\alpha}+1}\right) - \frac{\ln\alpha}{4} & \end{cases}$$

If $\kappa\sigma_s^2\sigma_t^2 \ge \sqrt{\alpha}$, $\gamma_{\max}^2 = \frac{(1+\kappa\sigma_s^2\sigma_t^2)(\alpha+\kappa\sigma_s^2\sigma_t^2)}{\kappa\sigma_s^2\sigma_t^2}$:

$$\mathcal{J}^{(\alpha)}(\theta,\mu_{\mathrm{MP}}) = \begin{cases} \frac{1}{2}\theta^2 & \text{if } \theta^2 \le \frac{1}{\kappa\sigma_s^2\sigma_t^2} \\ -\frac{1-\alpha}{2\alpha}\ln\left(1-\alpha+\sqrt{(-1+\alpha)^2+4\alpha\theta^2\gamma_{\max}^2}\right) & \text{if } \theta^2 \ge \frac{1}{\kappa\sigma_s^2\sigma_t^2} \\ \quad -\frac{1}{2\kappa\sigma_s^2\sigma_t^2} + \frac{\sqrt{(-1+\alpha)^2+4\alpha\theta^2\gamma_{\max}^2}}{2\alpha} - \frac{1+\alpha}{2\alpha} & \\ \quad -\frac{\ln\kappa\sigma_s^2\sigma_t^2}{2\alpha} + \frac{1-\alpha}{2\alpha}\ln 2 - \ln\theta & \\ \quad +\frac{1-\alpha}{2\alpha}\ln(\alpha+\kappa\sigma_s^2\sigma_t^2) & \end{cases}$$

8.C.2 Minimizing of $\psi(\rho_s, \rho_t)$

The calculations in this part are carried out using Mathematica [117]. Considering different conditions, based on these calculations, the minimizer of $\psi(\rho_s, \rho_t)$ is unique.

1) $\kappa\sigma_s^2\sigma_t^2 \le \sqrt{\alpha}$

If $\kappa'\sigma_s'^2\sigma_s'^2 \le \sqrt{\alpha}$:

$$\begin{cases} \rho_s^* &= \sigma_s'^2 \\ \rho_t^* &= \sigma_t'^2 \end{cases}$$

If $\kappa' {\sigma'_s}^2 {\sigma'_s}^2 \geq \sqrt{\alpha}$:

$$\begin{cases} \rho_s^* &= -\frac{1}{\kappa' {\sigma'_t}^2} + \frac{{\sigma'_s}^2}{2\alpha}\left[\alpha - 1 + \sqrt{(\alpha-1)^2 + 4\alpha \frac{(1+\sqrt{\alpha})^2}{\kappa' {\sigma'_s}^2 {\sigma'_t}^2}}\right] \\ \rho_t^* &= -\frac{\alpha}{\kappa' {\sigma'_s}^2} + \frac{{\sigma'_t}^2}{2}\left[1 - \alpha + \sqrt{(\alpha-1)^2 + 4\alpha \frac{(1+\sqrt{\alpha})^2}{\kappa' {\sigma'_s}^2 {\sigma'_t}^2}}\right] \end{cases}$$

2) $\kappa \sigma_s^2 \sigma_t^2 \geq \sqrt{\alpha}$

If $\kappa' {\sigma'_s}^2 {\sigma'_s}^2 \leq \frac{\alpha}{\kappa \sigma_s^2 \sigma_t^2}$:

$$\begin{cases} \rho_s^* &= {\sigma'_s}^2 \\ \rho_t^* &= {\sigma'_t}^2 \end{cases}$$

If $\kappa' {\sigma'_s}^2 {\sigma'_s}^2 \geq \frac{\alpha}{\kappa \sigma_s^2 \sigma_t^2}$:

$$\begin{cases} \rho_s^* &= -\frac{1}{\kappa' {\sigma'_t}^2} + \frac{{\sigma'_s}^2}{2\alpha}\left[\alpha - 1 + \sqrt{(\alpha-1)^2 + 4\alpha \frac{(1+\kappa\sigma_s^2\sigma_t^2)(\alpha+\kappa\sigma_s^2\sigma_t^2)}{\kappa\kappa' {\sigma'_s}^2 {\sigma'_t}^2 \sigma_s^2 \sigma_t^2}}\right] \\ \rho_t^* &= -\frac{\alpha}{\kappa' {\sigma'_s}^2} + \frac{{\sigma'_t}^2}{2}\left[1 - \alpha + \sqrt{(\alpha-1)^2 + 4\alpha \frac{(1+\kappa\sigma_s^2\sigma_t^2)(\alpha+\kappa\sigma_s^2\sigma_t^2)}{\kappa\kappa' {\sigma'_s}^2 {\sigma'_t}^2 \sigma_s^2 \sigma_t^2}}\right] \end{cases}$$

Part II

Extensive-Rank Inference

Extensive-Rank Symmetric Matrix Denoising

9

This chapter focuses on estimating a large, real symmetric signal matrix $\boldsymbol{S} \in \mathbb{R}^{N\times N}$ from noisy observations. The observations are modeled as:

$$\boldsymbol{Y} = \sqrt{\kappa}\boldsymbol{S} + \boldsymbol{Z} \tag{9.1}$$

Here, $\mathbf{Z}$ is a symmetric Gaussian matrix, and $\kappa > 0$ is proportional to the SNR.

We study this problem for the case $\boldsymbol{S}$ follows a rotationally invariant prior and its rank diverges as $N \to \infty$. Two growth regimes are considered: linear (rank grows linearly with N) and sub-linear.

In the linear growth regime:

- We show that the asymptotic mutual information is linked to the asymptotic of the log-spherical integral, Theorem 9.1.
- Leveraging the optimality of RIEs, in Theorem 9.2 we derive the asymptotic MMSE is in terms of the limiting ESD of $\boldsymbol{Y}$.
- Using the I-MMSE relation, we derive an explicit expression for the asymptotic mutual information in Theorem 9.3.

In the sub-linear growth regime:

- We derive the optimal RIE, see (9.52).
- We calculate the asymptotic MMSE and mutual information for this regime, aligning with findings in [65], which supports the optimality of the proposed RIE in this regime, see section 9.6.2.

The content of this chapter is based on a joint work with Prof. Barbier in [48] F. Pourkamali, J. Barbier, and N. Macris, "Matrix inference in growing rank regimes," arXiv preprint arXiv:2306.01412, 2023.

9.1 Linear Rank Matrix Denoising

We consider $\boldsymbol{S} = \boldsymbol{S}^\intercal \in \mathbb{R}^{N\times N}$ distributed according to a rotation invariant prior, i.e., $dP_{S,N}(\boldsymbol{O}\boldsymbol{S}\boldsymbol{O}^\intercal) = dP_{S,N}(\boldsymbol{S})$ for any $N \times N$ orthogonal matrix $\boldsymbol{O}$. The matrix $\boldsymbol{S}$ is corrupted by Gaussian Wigner noise $\boldsymbol{Y} = \sqrt{\kappa}\boldsymbol{S} + \boldsymbol{Z}$ and we are interested in the asymptotic mutual information and MMSE of this estimation problem.

The empirical spectral distribution of $\boldsymbol{S}$ is denoted as $\rho_S^{(N)}(x)dx = \frac{1}{N}\sum_{i\le N}\delta(x-\lambda_i^S)dx$ where (λ_i^S) are the eigenvalues of $\boldsymbol{S}$. For the rigorous analysis throughout the linear rank case we shall assume the following:

Assumption 9.1. *The empirical spectral distribution $\rho_S^{(N)}$ converges almost surely weakly to a well-defined probability measure ρ_S with support in $[-C, C]$ for some finite $C > 0$ independent of N. Moreover, the second moment of $\rho_S^{(N)}$ is almost surely bounded.*

Signal instances satisfying this assumption can be constructed as $\boldsymbol{S} = \boldsymbol{O}\boldsymbol{\Lambda}\boldsymbol{O}^\intercal$ with $\boldsymbol{O}$ uniformly sampled over the manifold of orthogonal matrices (or Haar distributed) and $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1^S, \dots, \lambda_N^S)$ i.i.d. eigenvalues distributed according to ρ_S with compact support. Note that for measures ρ_S containing a weight $\epsilon \in [0, 1]$ at 0 the random matrices $\boldsymbol{S}$ have rank $M = (1-\epsilon)N$. When $M = N$ there is another popular way to construct rotation invariant matrix ensembles, namely by setting $dP_{S,N}(\boldsymbol{S}) \propto \exp(-\frac{N}{2}\operatorname{Tr} V(\boldsymbol{S}))d\boldsymbol{S}$ where $V(\boldsymbol{S})$ is a rotation invariant "matrix potential". For such priors almost sure weak convergence of the empirical spectral distribution is proved in [66] whenever $\liminf_{|x|\to\infty} V(x)/(\beta \ln|x|) > 1$ for some $\beta > 1$. Moreover, under some additional conditions, the largest eigenvalue of such a random matrix satisfies a large deviation principle, which implies the almost sure boundedness of the top eigenvalue [66]. Therefore, our assumption holds for a large class of ensembles described by rotation invariant potentials.

Recall the spherical integral defined in (6.1), and denote the log-spherical integral as $\mathcal{J}_N(\boldsymbol{A}, \boldsymbol{B}) := \frac{1}{N}\ln \mathcal{I}_N(\boldsymbol{A}, \boldsymbol{B})$. Let

$$\mathcal{J}[\rho_{\sqrt{\kappa}S}, \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}] = \lim_{N\to+\infty} \mathcal{J}_N(\sqrt{\kappa}\boldsymbol{S}, \boldsymbol{Y})$$

where $\rho_{\sqrt{\kappa}S}$ is the limiting spectral distribution of $\sqrt{\kappa}\boldsymbol{S}$, ρ_{sc} is the Wigner semi-circle distribution, and $\rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}$ is the free convolution of $\rho_{\kappa S}$ and ρ_{sc} which is the limiting ESD of $\boldsymbol{Y}$.

Our first result for matrix denoising in the linear rank regime is a rigorous formula for the mutual information.

Theorem 9.1 (**Mutual Information for linear rank matrix denoising**). *Under assumption 9.1,*

$$\frac{I_N(\boldsymbol{S}; \boldsymbol{Y})}{N^2} \xrightarrow{N\to\infty} \frac{\kappa}{2}\int x^2 \rho_S(x)\, dx - \mathcal{J}[\rho_{\sqrt{\kappa}S}, \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}]. \tag{9.2}$$

Proof. Proof steps are outlined in section 9.3. □

Although we know that the limiting log-spherical integral is given by a variational problem, see [61], its computation requires going through Matytsin's formalism [95] and is, in general, highly non-trivial. Here, we will provide another formula (Theorem 9.3) for the mutual information, which is much simpler and explicit and which, in turn, also provides an expression for the log-spherical integral.

The route to this program goes first through the class of RIE. An estimator $\hat{\Xi}(\boldsymbol{Y})$ is called *rotation invariant* if for any orthogonal matrix $\boldsymbol{O}$, $\boldsymbol{O}\hat{\Xi}(\boldsymbol{Y})\boldsymbol{O}^\intercal = \hat{\Xi}(\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal)$. We may define the best possible reconstruction error *within the RIE class* as

$$\mathrm{MMSE}_{\mathrm{RIE,N}}(\kappa) = \min_{\hat{\Xi}\in\mathrm{RIE}} \frac{1}{N}\mathbb{E}\|\boldsymbol{S} - \hat{\Xi}(\boldsymbol{Y})\|_{\mathrm{F}}^2. \tag{9.3}$$

Obviously $\mathrm{MMSE}_N(\kappa) \le \mathrm{MMSE}_{\mathrm{RIE,N}}(\kappa)$. However, it is easy to check explicitly that the MMSE estimator $\mathbb{E}[\boldsymbol{S} \mid \boldsymbol{Y}]$ belongs to the RIE class. Thus we also have $\mathrm{MMSE}_{\mathrm{RIE,N}}(\kappa) \le \mathrm{MMSE}_N(\kappa)$ and hence $\mathrm{MMSE}_N(\kappa) = \mathrm{MMSE}_{\mathrm{RIE,N}}(\kappa)$.

Because of rotation invariance, $\boldsymbol{Y}$ and $\hat{\Xi} \in \mathrm{RIE}$ can be diagonalized in the same basis $(\boldsymbol{y}_i)$. Thus any RIE is expressed as $\hat{\Xi}(\boldsymbol{Y}) = \sum \hat{\xi}_i \boldsymbol{y}_i\boldsymbol{y}_i^\intercal$ where $(\hat{\xi}_i)$ are the eigenvalues of the estimator. Therefore (9.3) requires minimizing over $(\hat{\xi}_i)$ only. In [53] the heuristic replica method is used to show that *in the large N limit*, the optimal $(\hat{\xi}_i)$ can be expressed only in terms of the limiting spectral measure ρ_Y of the data $\boldsymbol{Y}$ and its eigenvalues (λ_i^Y):

$$\hat{\Xi}^*(\boldsymbol{Y}) = \sum_{i\le N} \xi_i^* \boldsymbol{y}_i\boldsymbol{y}_i^\intercal, \quad \xi_i^* = \frac{1}{\sqrt{\kappa}}\big(\lambda_i^Y - 2\pi \mathsf{H}[\rho_Y](\lambda_i^Y)\big), \tag{9.4}$$

where $\mathsf{H}[\rho_Y](z) := \mathrm{P.V}\frac{1}{\pi}\int \rho_Y(x)/(z-x)dx$ is the *Hilbert* transform of ρ_Y. Note that ρ_Y is given by the free convolution $\rho_Y = \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}$ which is a continuous density due to the smoothing effect of the semi-circle law [118]. Based on this result we make the following assumption here:

Assumption 9.2. *The estimator* (9.4) *is asymptotically optimal in the RIE class, i.e.,*

$$\mathrm{MMSE}_{\mathrm{RIE,N}}(\kappa) = \frac{1}{N}\mathbb{E}\|\boldsymbol{S} - \hat{\Xi}^*(\boldsymbol{Y})\|_{\mathrm{F}}^2 + o_N(1). \tag{9.5}$$

Using (9.4) and (9.5) we prove the following explicit formula for the MMSE in linear rank matrix denoising:

Theorem 9.2 (**MMSE for linear rank matrix denoising**). *Under assumptions 9.1, 9.2 we have*

$$\mathrm{MMSE}_N(\kappa) \xrightarrow{N\to\infty} \frac{1}{\kappa}\Big(1 - \frac{4\pi^2}{3}\int \rho_Y^3(x)\,dx\Big) \tag{9.6}$$

where the data spectral density $\rho_Y = \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}$. Moreover, $\mathrm{MMSE}(\kappa) := \lim_{N\to\infty}\mathrm{MMSE}_N(\kappa)$ *is continuous in $\kappa > 0$.*

Proof. Section 9.4. □

This is an explicit formula that can be used to concretely compute the MMSE(κ) curves for various models of rotation invariant signal ensembles, and in particular, allows to investigate the existence and nature of phase transitions[1]. The continuity of MMSE(κ) guarantees that there is no first order phase transition. In low-rank matrix denoising (as well as other inference problems) when there is no first order phase transition (but possibly higher order continuous transitions) the model does not display an algorithmically "hard phase" for low complexity algorithms (e.g., message passing algorithms are optimal). The present linear-rank rotation invariant case is no exception to this picture. Indeed equation (9.4) suggests an optimal spectral algorithm to estimate the signal: given an observation $\boldsymbol{Y}$ one computes its eigenvalues and an estimate of the Hilbert transform replacing the integral by an empirical sum to use in (9.4). This algorithm is optimal since as remarked above $\mathrm{MMSE}_N(\kappa) = \mathrm{MMSE}_{\mathrm{RIE,N}}(\kappa)$. Finally, we also mention that the optimality of RIE was also discussed in [43] in a different manner where the authors show heuristically that the posterior mean $\mathbb{E}[\boldsymbol{S} \mid \boldsymbol{Y}]$ equals $\hat{\Xi}^*(\boldsymbol{Y})$ as $N \to +\infty$, and also before in [51].

We now proceed to deduce a simpler formula for the mutual information (than in Thm. 9.1) using the I-MMSE relation [1]

$$\mathrm{MMSE}_N(\kappa) = 4\frac{d}{d\kappa}\frac{I_N(\boldsymbol{S};\boldsymbol{Y})}{N^2} \tag{9.7}$$

and free probability. Using the concavity of the mutual information w.r.t. the SNR, (9.7) also holds as $N \to +\infty$. One can thus permute the limit $N \to +\infty$ with the derivative w.r.t. κ. Therefore, it suffices to compute the integral over the *asymptotic* MMSE to find the *asymptotic* mutual information; this is done using basic results from free probability leading to:

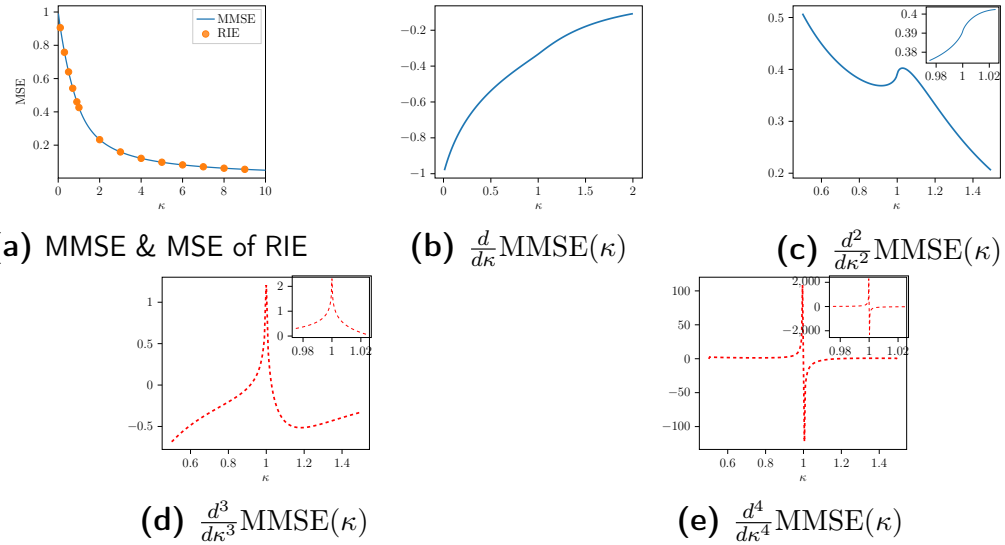
Theorem 9.3 (**Explicit Mutual Information for linear rank matrix denoising**). *Let $\rho_Y = \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}$. Under assumptions 9.1, 9.2 we have*

$$\frac{I_N(\boldsymbol{S};\boldsymbol{Y})}{N^2} \xrightarrow{N\to\infty} \frac{1}{2}\iint \ln|s-t|\rho_Y(s)\rho_Y(t)\,ds\,dt + \frac{1}{8}. \tag{9.8}$$

Proof. Section 9.5. □

In appendix 9.A, we extend Theorem 9.3 to the case where the noise matrix the realization of a non-Gaussian rotation invariant ensemble. While we are not quite able to treat this case, we can generalize this theorem to the setting $\boldsymbol{Y}_\epsilon = \sqrt{\kappa}\boldsymbol{S} + \boldsymbol{Z}_\epsilon$ where $\boldsymbol{Z}_\epsilon = \boldsymbol{Z} + \sqrt{\epsilon}\boldsymbol{\zeta}$ with $\boldsymbol{\zeta}$ from the Gaussian Wigner ensemble, and $\epsilon > 0$ (so the noise is non-Gaussian rotation invariant).

[1] Phase transitions are non-analyticity points in the asymptotic mutual information as a function of SNR. This is a concave and continuous function, and k-th order phase transitions correspond to discontinuities in the k-th derivative. In particular for a first order transition the MMSE is discontinuous because of the I-MMSE relation (9.7).



(a) MMSE & MSE of RIE (b) $\frac{d}{d\kappa}\mathrm{MMSE}(\kappa)$ (c) $\frac{d^2}{d\kappa^2}\mathrm{MMSE}(\kappa)$

(d) $\frac{d^3}{d\kappa^3}\mathrm{MMSE}(\kappa)$ (e) $\frac{d^4}{d\kappa^4}\mathrm{MMSE}(\kappa)$

Figure 9.2.1: MMSE for the Rademacher spectral distribution. From left to right: Plot (a) $\mathrm{MMSE}(\kappa)$ computed from (9.6) and $\mathrm{MSE}_{\mathrm{N,RIE}}(\kappa)$ points computed from (9.4) for $N = 1000$ averaged over 20 runs (error bars are invisible). Plots (b) and (c): first and second derivatives of $\mathrm{MMSE}(\kappa)$ computed using their integral representation (integral computed numerically). Plots (d) and (e): first and second numerical differentiation of (c). These suggest that the $\mathrm{MMSE}''(\kappa)$ has a vertical tangent at $\kappa_c = 1$, and a possible phase transition (if present) would be 4-th order. A numerical analysis in appendix 9.B.1 is compatible with a weak singularity at $\kappa_c = 1$ of the form $(\kappa - 1)^3 \ln|\kappa - 1|$.

9.2 Numerical Examples

As first example in the linear rank regime, we consider the case where $\rho_S = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}$. Using the technique introduced in [118], we obtain an explicit analytical expression for $\rho_Y = \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{SC}}$. For $\kappa \geq 1$ the support of ρ_Y consists of two disjoint intervals, and for $\kappa < 1$ we get a single interval. Therefore, we expect that, if a phase transition in the mutual information and MMSE exists, it should happen at $\kappa_c = 1$. As noted before, because of Theorem 9.2 we know that $\mathrm{MMSE}(\kappa)$ is continuous, so a phase transition can only be second or higher order. Furthermore, from the explicit formula for ρ_Y we get integral representations of the first few derivatives of $\mathrm{MMSE}(\kappa)$. Fig. 9.2.1 displays the results of precise numerical integrations for the MMSE and its first two derivatives, while the third derivative is computed by numerically differentiating the second derivative. These plots suggest that there is 4-th order phase transition for this example model at $\kappa_c = 1$. All details and additional figures can be found appendix 9.B.1.

In a second example we consider $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^\intercal/N$ where $\boldsymbol{X} \in \mathbb{R}^{N\times M}$ has i.i.d. standard Gaussian entries. The limiting spectral distribution of $\boldsymbol{S}$ when the aspect ratio $N/M \to q$ (fixed) is the *Marchenko-Pastur* law. For this model it is not difficult to directly compute ρ_Y (see apendix 9.B.3). Its support is a single interval for $q \leq 1$, and two disjoint intervals for $q > 1$ and $\kappa > q(q^{1/3} - 1)^{-3}$.

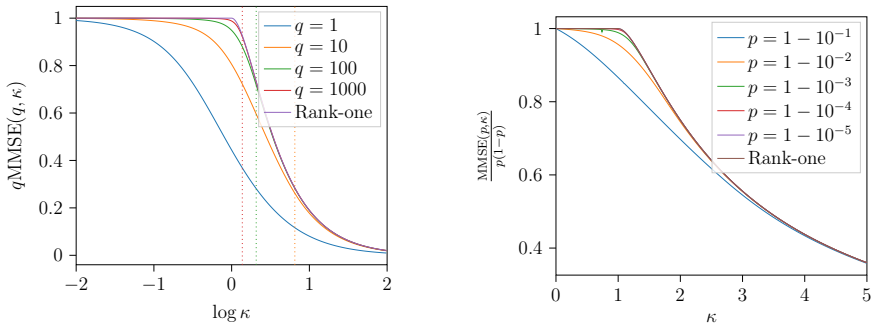


Figure 9.2.2: MMSE in the linear-rank regime with sparse spectral priors. The MMSE of the rank-one problem is also plotted for comparison. (left) Signal with Marchenko-Pastur spectral distribution for large q's.The vertical dashed lines corresponds to the critical value where the support of ρ_Y splits. (right) Signal with rank $(1-p)N$ and Bernoulli spectral distribution, $\rho_S = p\delta_0 + (1-p)\delta_{+1}$, for p's close to 1.

However, when the intervals merge in this case there does not seem to exist a phase transition, at least on low order derivatives (investigated numerically). Fig. 9.2.2 (left) shows the MMSE as a function of $\log \kappa$.

We observe in the left part of Fig. 9.2.2 that as $q \to +\infty$ the MMSE tends to the one of the rank-one version of matrix denoising with Gaussian prior. In particular, we recover the second-order phase transition of the rank-one problem at $\kappa = 1$, which matches the famous BBP transition [16]. This convergence towards the rank-one prediction can also be observed in a model with Bernoulli spectral distribution, see right part of Fig.9.2.2.

9.3 Proof Steps of Theorem 9.1

We present the steps needed to prove theorem 9.1. It is convenient to decompose Assumption 1 in the main text in two parts.

Assumption 1.A The empirical spectral distribution of $\boldsymbol{S}$ converges almost surely weakly to a well-defined probability measure ρ_S with support included in $[-C, C]$ for some finite positive constant $0 < C < +\infty$ independent of N.
Assumption 1.B The second moment of the empirical spectral distribution of $\boldsymbol{S}$ is almost surely bounded.

Remark 9.1. *These assumptions taken together imply that the second moment of the empirical measure $\rho_S^{(N)}$ converges almost surely to the second moment of ρ_S.*

Remark 9.2. *By Theorem 7.12 in ref. [119], these assumptions are equivalent to the convergence of the empirical distribution in the Wasserstein-2 metric to ρ_S.*

We start from the posterior distribution of the model, which is proportional to

$$\begin{aligned} P_N(\boldsymbol{X}|\boldsymbol{Y}) &\propto e^{-\frac{N}{4}\|\boldsymbol{Y}-\sqrt{\kappa}\boldsymbol{X}\|_{\mathrm{F}}^2} P_{S,N}(\boldsymbol{X}) \\ &\propto e^{\frac{N}{2}\operatorname{Tr}\left[\sqrt{\kappa}\boldsymbol{X}\boldsymbol{Y}-\frac{\kappa}{2}\boldsymbol{X}^2\right]} P_{S,N}(\boldsymbol{X}). \end{aligned} \tag{9.9}$$

The partition function is defined as the normalizing factor of the posterior distribution (9.9)

$$Z(\boldsymbol{Y}) = \int d\boldsymbol{X} e^{\frac{N}{2}\operatorname{Tr}\left[\sqrt{\kappa}\boldsymbol{X}\boldsymbol{Y}-\frac{\kappa}{2}\boldsymbol{X}^2\right]} P_{S,N}(\boldsymbol{X}) \tag{9.10}$$

and the free energy is defined as

$$F_N(\kappa) = -\frac{1}{N^2}\mathbb{E}_Y\big[\ln Z(\boldsymbol{Y})\big]. \tag{9.11}$$

One can easily see that the free energy is linked to the average mutual information via the identity

$$\frac{1}{N^2} I_N(\boldsymbol{S};\boldsymbol{Y}) = F_N(\kappa) + \frac{\kappa}{4N}\mathbb{E}\big[\operatorname{Tr}\boldsymbol{S}^2\big] \tag{9.12}$$

in which $\frac{1}{N}\mathbb{E}\big[\operatorname{Tr}\boldsymbol{S}^2\big]$ converges to the second moment of ρ_S by assumption. Therefore, to prove theorem 9.1 it is enough to show

$$\lim_{N\to\infty} F_N(\kappa) = \frac{\kappa}{4}\int x^2\rho_S(x)\,dx - \mathcal{J}[\rho_{\sqrt{\kappa}S},\, \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}]. \tag{9.13}$$

The proof of (9.13) is done in two main steps. First we we show that such a limit holds for the free energy of an *independent eigenvalue model*. Second, using the *pseudo-Lipschitz* continuity of the free energy w.r.t. to the prior distribution, we deduce that the same limit holds for the free energy of the original model.

We make the convention that in the eigen-decomposition of a $N \times N$ matrix $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\intercal$ with $\boldsymbol{\Lambda} = \operatorname{diag}(\boldsymbol{\lambda})$, the eignevalues $\lambda_1, \cdots, \lambda_N$ are in non-decreasing order.

9.3.1 An independent eigenvalue model

Suppose $\boldsymbol{\lambda}^0 \in \mathbb{R}^N$ is generated with i.i.d. elements from ρ_S and is ordered in non-decreasing way. Fix $\boldsymbol{\lambda}^0$ once for all. Let $\tilde{\boldsymbol{S}} \in \mathbb{R}^{N\times N}$ the matrix constructed as $\boldsymbol{U}\tilde{\boldsymbol{\Lambda}}\boldsymbol{U}^\intercal$ where $\boldsymbol{U}$ is distributed according to the Haar measure, and $\tilde{\boldsymbol{\Lambda}} = \operatorname{diag}(\tilde{\boldsymbol{\lambda}})$ is a diagonal matrix. The distribution of the matrix $\tilde{\boldsymbol{S}}$ is

$$dP_{\tilde{S},N}(\tilde{\boldsymbol{S}}) = d\mu_N(\boldsymbol{U}) dp_{\tilde{S},N}(\tilde{\boldsymbol{\lambda}}) = d\mu_N(\boldsymbol{U}) \prod_{i=1}^{N} \delta(\tilde{\lambda}_i - \lambda_i^0)\, d\tilde{\boldsymbol{\lambda}}. \tag{9.14}$$

The independent eigenvalue model is defined as an inference model where the matrix $\tilde{\boldsymbol{S}}$ is observed through an AWGN channel $\tilde{\boldsymbol{Y}} = \sqrt{\kappa}\tilde{\boldsymbol{S}} + \tilde{\boldsymbol{Z}}$ with SNR proportional to κ and $\tilde{\boldsymbol{Z}}$ a symmetric Gaussian Wigner matrix. The associated partition function and the free energy are defined in the same way as in (9.10),(9.11)

$$\tilde{Z}(\tilde{\boldsymbol{Y}}) = \int d\boldsymbol{X} e^{\frac{N}{2}\operatorname{Tr}\left[\sqrt{\kappa}\boldsymbol{X}\tilde{\boldsymbol{Y}} - \frac{\kappa}{2}\boldsymbol{X}^2\right]} P_{\tilde{S},N}(\boldsymbol{X}). \tag{9.15}$$

$$\tilde{F}_N(\kappa) = -\frac{1}{N^2}\mathbb{E}_{\tilde{Y}}\big[\ln \tilde{Z}(\tilde{\boldsymbol{Y}})\big]. \tag{9.16}$$

For this independent eigenvalue model, we have

Proposition 9.4. *For ρ_S with compact support and any $\kappa > 0$, ρ_S-almost surely*

$$\lim_{N\to\infty} \tilde{F}_N(\kappa) = \frac{\kappa}{4}\int x^2 \rho_S(x)\,dx - \mathcal{J}[\rho_{\sqrt{\kappa}S},\, \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}]. \tag{9.17}$$

Proof. We start from the partition function (9.15),

$$\begin{aligned}
\tilde{Z}(\tilde{\boldsymbol{Y}}) &= \int d\boldsymbol{X} e^{\frac{N}{2}\operatorname{Tr}\left[\sqrt{\kappa}\boldsymbol{X}\tilde{\boldsymbol{Y}} - \frac{\kappa}{2}\boldsymbol{X}^2\right]} P_{\tilde{S},N}(\boldsymbol{X}) \\
&= \int d\boldsymbol{\lambda} d\mu_N(\boldsymbol{U}) \prod_{i=1}^{N} \delta(\lambda_i - \lambda_i^0)\, e^{\frac{N}{2}\operatorname{Tr}[\sqrt{\kappa}\boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\intercal\tilde{\boldsymbol{Y}} - \frac{\kappa}{2}\boldsymbol{\Lambda}^2]} \\
&= e^{-\frac{N}{4}\kappa\operatorname{Tr}\boldsymbol{\Lambda}^{0\,2}} \int d\mu_N(\boldsymbol{U})\, e^{\frac{N}{2}\operatorname{Tr}[\sqrt{\kappa}\boldsymbol{U}\boldsymbol{\Lambda}^0\boldsymbol{U}^\intercal\tilde{\boldsymbol{Y}}]} \\
&= e^{-\frac{N}{4}\kappa\operatorname{Tr}\boldsymbol{\Lambda}^{0\,2}} \mathcal{I}_N\big(\sqrt{\kappa}\boldsymbol{\Lambda}^0, \tilde{\boldsymbol{Y}}\big)
\end{aligned} \tag{9.18}$$

Recall that $\tilde{\boldsymbol{Y}} = \sqrt{\kappa}\boldsymbol{U}\boldsymbol{\Lambda}^0\boldsymbol{U}^\intercal + \tilde{\boldsymbol{Z}}$, so the free energy can be written as

$$\begin{aligned}
\tilde{F}_N(\kappa) &= \mathbb{E}_{\tilde{\boldsymbol{Y}}}\Big[\frac{\kappa}{4N}\operatorname{Tr}\boldsymbol{\Lambda}^{0\,2} - \mathcal{J}_N\big(\sqrt{\kappa}\boldsymbol{\Lambda}^0, \tilde{\boldsymbol{Y}}\big)\Big] \\
&= \frac{\kappa}{4N}\operatorname{Tr}\boldsymbol{\Lambda}^{0\,2} - \mathbb{E}_{\boldsymbol{U}}\mathbb{E}_{\tilde{\boldsymbol{Z}}}\Big[\mathcal{J}_N\big(\sqrt{\kappa}\boldsymbol{\Lambda}^0, \tilde{\boldsymbol{Y}}\big)\Big] \\
&\overset{(a)}{=} \frac{\kappa}{4N}\operatorname{Tr}\boldsymbol{\Lambda}^{0\,2} - \mathbb{E}_{\boldsymbol{U}}\mathbb{E}_{\tilde{\boldsymbol{Z}}}\Big[\mathcal{J}_N\big(\sqrt{\kappa}\boldsymbol{\Lambda}^0, \sqrt{\kappa}\boldsymbol{U}\boldsymbol{\Lambda}^0\boldsymbol{U}^\intercal + \boldsymbol{U}\tilde{\boldsymbol{Z}}\boldsymbol{U}^\intercal\big)\Big] \\
&\overset{(b)}{=} \frac{\kappa}{4N}\operatorname{Tr}\boldsymbol{\Lambda}^{0\,2} - \mathbb{E}_{\boldsymbol{U}}\mathbb{E}_{\tilde{\boldsymbol{Z}}}\Big[\mathcal{J}_N\big(\sqrt{\kappa}\boldsymbol{\Lambda}^0, \sqrt{\kappa}\boldsymbol{\Lambda}^0 + \tilde{\boldsymbol{Z}}\big)\Big] \\
&= \frac{\kappa}{4N}\operatorname{Tr}\boldsymbol{\Lambda}^{0\,2} - \mathbb{E}_{\tilde{\boldsymbol{Z}}}\Big[\mathcal{J}_N\big(\sqrt{\kappa}\boldsymbol{\Lambda}^0, \sqrt{\kappa}\boldsymbol{\Lambda}^0 + \tilde{\boldsymbol{Z}}\big)\Big]
\end{aligned} \tag{9.19}$$

where in (a), we use rotational invariance of the noise matrix $\boldsymbol{Z}$, and in (b), we use the fact that $\mathcal{J}_N$ is invariant under rotation by $\boldsymbol{U}$.

By the strong law of large numbers the first term in (9.19) converges to $\frac{\kappa}{4}\int x^2\rho_S(x)\,dx$ almost surely. Finally proposition 9.4 follows from the subsequent lemma.

Lemma 9.1. *For any $\kappa \in \mathbb{R}_+$, the sequence $\mathbb{E}_{\tilde{\boldsymbol{Z}}}\Big[\mathcal{J}_N\big(\sqrt{\kappa}\boldsymbol{\Lambda}^0, \sqrt{\kappa}\boldsymbol{\Lambda}^0 + \tilde{\boldsymbol{Z}}\big)\Big]$ converges to $\mathcal{J}[\rho_{\sqrt{\kappa}S}, \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}]$ as $N \to \infty$, ρ_S-almost surely.*

This lemma is based on an important result on the convergence of log-spherical integrals [61]. We refer to appendix 9.C.1 for the details. □

9.3.2 Pseudo-Lipschitz continuity of the free energy

Consider any two rotationally invariant matrix ensembles $P_N^{(1)}$, $P_N^{(2)}$. Let $\boldsymbol{S} \sim P_N^{(1)}(\boldsymbol{S})$, $\tilde{\boldsymbol{S}} \sim P_N^{(2)}(\tilde{\boldsymbol{S}})$ with eigendecompositions $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\intercal$, $\tilde{\boldsymbol{S}} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{U}}^\intercal$ and

$$\begin{aligned} dP_N^{(1)}(\boldsymbol{S}) &= d\mu_N(\boldsymbol{U})\ P_N^{(1)}(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}, \\ dP_N^{(2)}(\tilde{\boldsymbol{S}}) &= d\mu_N(\tilde{\boldsymbol{U}})\ P_N^{(2)}(\tilde{\boldsymbol{\lambda}})\, d\tilde{\boldsymbol{\lambda}} \end{aligned} \tag{9.20}$$

where $P_N^{(1)}(\boldsymbol{\lambda})$, $P_N^{(2)}(\tilde{\boldsymbol{\lambda}})$ are the joint probability density functions for the eigenvalues, induced by the priors. Now consider the two inference problems corresponding to reconstructing the signals from outputs of an AWGN channel and define as before the corresponding free energies $F_N^{(1)}(\kappa), F_N^{(2)}(\kappa)$. Then we have

Proposition 9.5. *For all $\kappa > 0$ and N*

$$\big|F_N^{(1)}(\kappa) - F_N^{(2)}(\kappa)\big| \le \frac{\kappa}{4N}\Big(\sqrt{\mathbb{E}_{\boldsymbol{\lambda}}\big[\|\boldsymbol{\lambda}\|_2^2\big]} + \sqrt{\mathbb{E}_{\tilde{\boldsymbol{\lambda}}}\big[\|\tilde{\boldsymbol{\lambda}}\|_2^2\big]}\Big)\sqrt{\mathbb{E}_{\boldsymbol{\lambda},\tilde{\boldsymbol{\lambda}}}\big[\|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}\|_2^2\big]}. \tag{9.21}$$

Proof. The proof is based on an interpolation between the two matrix ensembles. We refer to appendix 9.C.2. □

9.3.3 The distance between the original and independent eigenvalue models

Consider the original and independent eigenvalue models, in other words, the models with prior distributions

$$\begin{aligned} dP_{S,N}(\boldsymbol{S}) &= d\mu_N(\boldsymbol{U})\, P_{S,N}(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}^S, \\ dP_{\tilde{S},N}(\tilde{\boldsymbol{S}}) &= d\mu_N(\boldsymbol{U}) \prod_{i=1}^N \delta(\tilde{\lambda}_i - \lambda_i^0)\, d\tilde{\boldsymbol{\lambda}} \end{aligned} \tag{9.22}$$

where $P_{S,N}(\boldsymbol{\lambda})$ is the joint p.d.f. of eigenvalues of $\boldsymbol{S}$, and $\boldsymbol{\lambda}^0$ is generated with i.i.d. elements from ρ_S. Denote $\prod_{i=1}^N \delta(\tilde{\lambda}_i - \lambda_i^0)$ by $P_{\tilde{S},N}(\tilde{\boldsymbol{\lambda}})$.

Lemma 9.2. *Under assumption 1 for* $\boldsymbol{\lambda} \sim P_{S,N}(\boldsymbol{\lambda})$ *and* $\tilde{\boldsymbol{\lambda}} \sim P_{\tilde{S},N}(\tilde{\boldsymbol{\lambda}})$ *we have*

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{\boldsymbol{\lambda},\tilde{\boldsymbol{\lambda}}}\big[\|\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}}\|^2\big] = 0. \tag{9.23}$$

Proof. Since $P_{\tilde{\boldsymbol{S}},N}(\tilde{\boldsymbol{\lambda}})$ is a delta distribution, we can write

$$\mathbb{E}_{\boldsymbol{\lambda},\tilde{\boldsymbol{\lambda}}}\big[\|\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}}\|^2\big] = \mathbb{E}_{\boldsymbol{\lambda}}\big[\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^0\|^2\big]. \tag{9.24}$$

For a vector $\boldsymbol{\lambda}$, denote the empirical distribution of its components by $\hat{\mu}_{\boldsymbol{\lambda}}$. The Wasserstein-2 distance between two empirical distributions, $\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0}$ is defined as

$$W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0}) = \sqrt{\inf_{\kappa\in\kappa(\hat{\mu}_{\boldsymbol{\lambda}},\hat{\mu}_{\boldsymbol{\lambda}^0})} \mathbb{E}_{\kappa(x,y)}\big[(x-y)^2\big]}$$

with $\kappa(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0})$ the set of couplings of $(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0})$. By lemma 9.8 in appendix 9.C.3 , we have

$$W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0}) = \sqrt{\min_{\pi\in\mathcal{S}_N} \frac{1}{N}\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^0_\pi\|^2}$$

where $\boldsymbol{\lambda}^0_\pi$ is the permuted version of $\boldsymbol{\lambda}^0$, and $\mathcal{S}_N$ is the group of all permutations of N elements. So, for given $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}^0$ (which have a non-decreasing order), we have (considering the identity permutation)

$$\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^0\|^2 \geq N W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0})^2. \tag{9.25}$$

Now we recall recall the *rearrangement inequality*: for real numbers $x_1 \leq x_2 \leq \cdots \leq x_n$, $y_1 \leq y_2 \leq \cdots \leq y_n$, for every permutation $\pi \in \mathcal{S}_N$ we have [120]

$$x_n y_1 + \ldots + x_1 y_n \leq x_{\pi(1)} y_1 + \ldots + x_{\pi(n)} y_n \leq x_1 y_1 + \ldots + x_n y_n$$

For any permutation of $\boldsymbol{\lambda}^0$ (in particular the one which achieves the minimum in (9.25)), using the rearrangement inequality, we get (recall that $\boldsymbol{\lambda}^0$ is ordered in non-decreasing order)

$$\begin{aligned}\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^0_\pi\|^2 &= \|\boldsymbol{\lambda}\|^2 + \|\boldsymbol{\lambda}^0_\pi\|^2 - 2\boldsymbol{\lambda}^\intercal\boldsymbol{\lambda}^0_\pi \\ &\geq \|\boldsymbol{\lambda}\|^2 + \|\boldsymbol{\lambda}^0_\pi\|^2 - 2\boldsymbol{\lambda}^\intercal\boldsymbol{\lambda}^0 \\ &= \|\boldsymbol{\lambda}-\boldsymbol{\lambda}^0\|^2\end{aligned}$$

and consequently

$$\|\boldsymbol{\lambda}-\boldsymbol{\lambda}^0\|^2 \leq N W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0})^2. \tag{9.26}$$

Finally from (9.24), (9.25),(9.26), we obtain

$$\mathbb{E}_{\boldsymbol{\lambda},\tilde{\boldsymbol{\lambda}}}\big[\|\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}}\|^2\big] = \mathbb{E}_{\boldsymbol{\lambda}}\big[N W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0})^2\big]. \tag{9.27}$$

Lemma 9.9 in appendix 9.C.3 allows to conclude the proof. □

9.3.4 Concluding the proof

By proposition 9.5, the free energies $F_N(\kappa)$ (defined in (9.11)) and $\tilde{F}_N(\kappa)$ (defined in (9.16)), satisfy

$$\left|F_N^{(1)}(\kappa) - F_N^{(2)}(\kappa)\right| \leq \frac{\kappa}{4N}\left(\sqrt{\mathbb{E}_{\boldsymbol{\lambda}}\left[\|\boldsymbol{\lambda}\|^2\right]} + \sqrt{\mathbb{E}_{\tilde{\boldsymbol{\lambda}}}\left[\|\tilde{\boldsymbol{\lambda}}\|^2\right]}\right)\sqrt{\mathbb{E}_{\boldsymbol{\lambda},\tilde{\boldsymbol{\lambda}}}\left[\|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}\|^2\right]}. \tag{9.28}$$

The term $\frac{1}{N}\|\boldsymbol{\lambda}\|^2 = \frac{1}{N}\sum \lambda_i^2$ is the second moment of the empirical spectral distribution of $\boldsymbol{S}$, which is almost surely bounded by assumption 1.B. So, $\frac{1}{N}\mathbb{E}_{\boldsymbol{\lambda}}\left[\|\boldsymbol{\lambda}\|^2\right]$ is bounded uniformly in N. Moreover, $\frac{1}{N}\mathbb{E}\left[\|\tilde{\boldsymbol{\lambda}}\|^2\right] = \frac{1}{N}\sum \lambda^{0}{}_i^2$ is also bounded uniformly in N. By lemma 9.2, $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{\boldsymbol{\lambda},\tilde{\boldsymbol{\lambda}}}\left[\|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}\|^2\right] = 0$. Therefore

$$\lim_{N\to\infty} |F_N(\kappa) - \tilde{F}_N(\kappa)| = 0. \tag{9.29}$$

Proposition 9.4 together with (9.29) conclude the proof. □

9.4 Proof of Theorem 9.2 and Rotation Invariance of the Bayes Estimator

In this section we show how to use the RIE class in order to prove Theorem 9.2 under the extra assumption 1.B. Recall that an estimator $\hat{\Xi}(\boldsymbol{Y})$ is called *rotation invariant* if for any orthogonal matrix $\boldsymbol{O}\hat{\Xi}(\boldsymbol{Y})\boldsymbol{O}^\intercal = \hat{\Xi}(\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal)$. Such an estimator has the same eigenvectors as the matrix $\boldsymbol{Y}$ and denoting the eigenvectors of $\boldsymbol{Y}$ by $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N$, it can be expressed as $\hat{\Xi}(\boldsymbol{Y}) = \sum_{i=1}^N \hat{\xi}_i \boldsymbol{y}_i \boldsymbol{y}_i^\intercal$ where $\hat{\xi}_1, \ldots, \hat{\xi}_N$ are the eigenvalues of the estimator. The best RIE estimator corresponds to eigenvalues chosen to minimize the mean-square-error in the RIE class and a heuristic calculation using the replica method leads to

$$\begin{cases} \hat{\Xi}^*(\boldsymbol{Y}) = \sum_{i=1}^N \xi_i^* \boldsymbol{y}_i \boldsymbol{y}_i^\intercal \\ \xi_i^* = \sum_{j=1}^N \lambda_j^S \left(\boldsymbol{s}_j^\intercal \boldsymbol{y}_i\right)^2 = \frac{1}{\sqrt{\kappa}}\left(\lambda_i^Y - 2\pi \mathsf{H}[\rho_Y](\lambda_i^Y)\right) \end{cases} \tag{9.30}$$

where $(\lambda_i^Y)_{1\leq i\leq N}$ are the eigenvalues of $\boldsymbol{Y}$, and $\mathsf{H}[\rho_Y]$ is the *Hilbert* transform of the limiting spectral distribution of $\boldsymbol{Y}$ defined as:

$$\mathsf{H}[\rho_Y](z) := \mathrm{PV}\frac{1}{\pi}\int \frac{\rho_Y(x)}{z - x}\, dx. \tag{9.31}$$

We will need the following properties of the Hilbert transform. A proof can be found in lemma 3.1 of [121].

Lemma 9.3. *If $f : \mathbb{R} \to \mathbb{R}$ is compactly supported and sufficiently regular, then one has the identities*

$$\int_{\mathbb{R}} f(x)\left(\mathsf{H}[f](x)\right)^2 dx = \frac{1}{3}\int_{\mathbb{R}} f^3(x)\, dx, \tag{9.32}$$

$$\int_{\mathbb{R}} \mathsf{H}[f](x) x f(x)\, dx = \frac{1}{2\pi}\Big(\int_{\mathbb{R}} f(x)\, dx\Big)^2. \tag{9.33}$$

Proof of theorem 9.2. As explained in the main text the best RIE estimator is optimal in the sense $\mathrm{MMSE}_N(\kappa) = \mathrm{MMSE}_{\mathrm{RIE,N}}(\kappa)$ so our proof proceeds by a computation of the limit of the r.h.s (the optimality follows from rotation invariance of the MMSE estimator $\mathbb{E}(\boldsymbol{S}|\boldsymbol{Y})$ and this rotation invariance is checked later for completeness). From assumption 2, it suffices to compute the limit of $\frac{1}{N}\mathbb{E}\|\boldsymbol{S} - \hat{\Xi}^*(\boldsymbol{Y})\|^2$. Denoting the eigenvectors of $\boldsymbol{S}$ by $\boldsymbol{s}_1, \dots, \boldsymbol{s}_N$. Expanding the MSE, we find

$$\begin{aligned}\big\|\boldsymbol{S} - \hat{\Xi}^*(\boldsymbol{Y})\big\|_{\mathrm{F}}^2 &= \sum_{i=1}^N \Big[\lambda_i^{S\,2} + \xi_i^{*2} - 2\xi_i^* \sum_{j=1}^N \lambda_j^S \big(\boldsymbol{s}_j^\intercal \boldsymbol{y}_i\big)^2\Big] \\ &= \sum_{i=1}^N (\lambda_i^{S\,2} - \xi_i^{*2})\end{aligned} \tag{9.34}$$

where we used (9.30) to get the last equality. Using again (9.30) we have

$$\begin{aligned}\big\|\boldsymbol{S} - \hat{\Xi}^*(\boldsymbol{Y})\big\|_{\mathrm{F}}^2 &= \sum_{i=1}^N (\lambda_i^{S\,2} - \xi_i^{*2}) \\ &= \sum_{i=1}^N \lambda_i^{S\,2} - \frac{1}{\kappa}\big(\lambda_i^Y - 2\pi \mathsf{H}[\rho_Y](\lambda_i^Y)\big)^2 \\ &= \sum_{i=1}^N \lambda_i^{S\,2} - \frac{1}{\kappa}\sum_{i=1}^N \lambda_i^{Y\,2} - \frac{4\pi^2}{\kappa}\sum_{i=1}^N \big(\mathsf{H}[\rho_Y](\lambda_i^Y)\big)^2 + \frac{4\pi}{\kappa}\sum_{i=1}^N \mathsf{H}[\rho_Y](\lambda_i^Y)\lambda_i^Y\end{aligned} \tag{9.35}$$

From linearity of expectation and, as the Hilbert transform $\mathsf{H}[\rho_Y]$ is continuous on the support of ρ_Y [118], we find

$$\begin{aligned}&\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\big\|\boldsymbol{S} - \hat{\Xi}^*(\boldsymbol{Y})\big\|_{\mathrm{F}}^2 \\ &\quad= \int x^2 \rho_S(x)\, dx - \frac{1}{\kappa}\int x^2 \rho_Y(x)\, dx - \frac{4\pi^2}{\kappa}\int \rho_Y(x)\big(\mathsf{H}[\rho_Y](x)\big)^2\, dx \\ &\qquad + \frac{4\pi}{\kappa}\int \mathsf{H}[\rho_Y](x) x \rho_Y(x)\, dx\end{aligned} \tag{9.36}$$

By the independence of $\boldsymbol{S}$ and $\boldsymbol{Z}$, we have:

$$\begin{aligned}\int x^2 \rho_Y(x)\, dx &= \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{Y}^2 \\ &= \lim_{N\to\infty} \frac{1}{N}\kappa \mathbb{E}\,\mathrm{Tr}\,\boldsymbol{S}^2 + \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{Z}^2 + \lim_{N\to\infty} \frac{2}{N}\mathbb{E}\,\mathrm{Tr}\,\boldsymbol{S}\boldsymbol{Z}\end{aligned}$$

$$= \kappa \int x^2 \rho_S(x) + 1 \tag{9.37}$$

Finally, using the identities in lemma 9.3, we get

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\big\|\boldsymbol{S} - \hat{\boldsymbol{S}}\big\|_{\mathrm{F}}^2 = -\frac{1}{\kappa} - \frac{4\pi^2}{3\kappa}\int \rho_Y^3(x)\,dx + \frac{2}{\kappa} = \frac{1}{\kappa}\Big(1 - \frac{4\pi^2}{3}\int \rho_Y^3(x)\,dx\Big). \tag{9.38}$$

□

For completeness we provide a check that the MMSE estimator belongs to the RIE class. The posterior mean given $\boldsymbol{Y}$ is

$$\mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}] = \frac{\int d\boldsymbol{X}\, P_{S,N}(\boldsymbol{X})\boldsymbol{X} e^{-\frac{N}{4}\left\|\boldsymbol{Y}-\sqrt{\kappa}\boldsymbol{X}\right\|_{\mathrm{F}}^2}}{\int d\boldsymbol{X}\, P_{S,N}(\boldsymbol{X}) e^{-\frac{N}{4}\left\|\boldsymbol{Y}-\sqrt{\kappa}\boldsymbol{X}\right\|_{\mathrm{F}}^2}}. \tag{9.39}$$

By rotation invariance of $P_{S,N}(\boldsymbol{X})$ under any orthogonal transformation $\boldsymbol{X} \to \boldsymbol{O}\boldsymbol{X}\boldsymbol{O}^\intercal$ with Jacobian $|\det \boldsymbol{O}| = 1$ we have

$$\begin{aligned}
\mathbb{E}[\boldsymbol{S}|\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal] &= \frac{\int d\boldsymbol{X}\, P_{S,N}(\boldsymbol{X})\boldsymbol{X} e^{-\frac{N}{4}\left\|\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal-\sqrt{\kappa}\boldsymbol{X}\right\|_{\mathrm{F}}^2}}{\int d\boldsymbol{X}\, P_{S,N}(\boldsymbol{X}) e^{-\frac{N}{4}\left\|\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal-\sqrt{\kappa}\boldsymbol{X}\right\|_{\mathrm{F}}^2}} \\
&= \frac{\int d\boldsymbol{X}\, P_{S,N}(\boldsymbol{X})\boldsymbol{O}\boldsymbol{X}\boldsymbol{O}^\intercal e^{-\frac{N}{4}\left\|\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal-\sqrt{\kappa}\boldsymbol{O}\boldsymbol{X}\boldsymbol{O}^\intercal\right\|_{\mathrm{F}}^2}}{\int d\boldsymbol{X}\, P_{S,N}(\boldsymbol{X}) e^{-\frac{N}{4}\left\|\boldsymbol{O}\boldsymbol{Y}\boldsymbol{O}^\intercal-\sqrt{\kappa}\boldsymbol{O}\boldsymbol{X}\boldsymbol{O}^\intercal\right\|_{\mathrm{F}}^2}} \\
&= \boldsymbol{O}\Big\{\frac{\int d\boldsymbol{X}\, P_{S,N}(\boldsymbol{X})\boldsymbol{X} e^{-\frac{N}{4}\left\|\boldsymbol{Y}-\sqrt{\kappa}\boldsymbol{X}\right\|_{\mathrm{F}}^2}}{\int d\boldsymbol{X}\, P_{S,N}(\boldsymbol{X}) e^{-\frac{N}{4}\left\|\boldsymbol{Y}-\sqrt{\kappa}\boldsymbol{X}\right\|_{\mathrm{F}}^2}}\Big\}\boldsymbol{O}^\intercal \\
&= \boldsymbol{O}\mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}]\boldsymbol{O}^\intercal.
\end{aligned} \tag{9.40}$$

Therefore the posterior mean estimator is an RIE.

9.5 Proof of Theorem 9.3: Explicit Expression of the Mutual Information

We derive an explicit expression for the asymptotic mutual information using the I-MMSE relation [1] and basic results in free probability. This derivation is completely independent of Theorem 9.1 in which the asymptotic mutual information is expressed using the asymptotic spherical integral, $\mathcal{J}[\rho_{\sqrt{\kappa}S}, \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}]$.

Free probability (see [72, 122]) was initially introduced to study operator algebras, but has gained considerable importance in other realms due to its connection with the asymptotic behavior of random matrices. While free

probability has been exploited in linear estimation and wireless communication problems [114, 123–125], the connection with matrix inference setting studied here is to the best of our knowledge new.

Let X be a self-adjoint *non-commutative* random variable X associated to a probability measure μ_X with compact support on the real line. According to [75, 76] the *free entropy* $\chi(X)$ and the *free Fisher information* $\Phi(X)$ are given as

$$\chi(X) = \iint \ln|s-t|\mu_X(s)\mu_X(t)\,ds\,dt + \frac{3}{4} + \frac{1}{2}\ln 2\pi, \tag{9.41}$$

$$\Phi(X) = \frac{4\pi^2}{3}\int \mu_X^3(s)\,ds. \tag{9.42}$$

Moreover, these two quantities are related through the relation:

$$\chi(X) = \frac{1}{2}\int_0^\infty \Big(\frac{1}{1+t} - \Phi(X+\sqrt{t}Z)\Big)\,dt + \frac{1}{2}\ln 2\pi + \frac{1}{2} \tag{9.43}$$

where Z is a semicircular non-commutative random variable, and X and Z are *free*. We apply these relations to $X = S$ and Z, two free non-commutative random variables (not to confused with the $N\times N$ matrices $\boldsymbol{S}$ and $\boldsymbol{Z}$) associated to the probability measures ρ_S and $\rho_{\rm sc}$. Since S and Z are free, the sum $\sqrt{\kappa}S + Z$ is a non-commutative random variable associated to the measure $\rho_{\sqrt{\kappa}S} \boxplus \rho_{\rm sc} = \rho_Y$. Clearly then (9.6) can be written in the free probability language as

$$\mathrm{MMSE}(\kappa) = \frac{1}{\kappa}\Big(1 - \Phi(\sqrt{\kappa}S + Z)\Big). \tag{9.44}$$

Proof of Theorem 9.3. An important property of the Gaussian channel is the I-MMSE relation relating the MMSE to the derivative of the mutual information w.r.t the SNR. The concavity of the mutual information w.r.t. SNR, implies that this relation also holds in the limit $N\to\infty$. Integrating this relation we have

$$I(\boldsymbol{S};\boldsymbol{Y}) = \frac{1}{4}\int_0^\kappa \mathrm{MMSE}(t)\,dt + \text{constant} \tag{9.45}$$

where $I(\boldsymbol{S};\boldsymbol{Y}) := \lim_{N\to\infty}\frac{1}{N^2}I_N(\boldsymbol{S};\boldsymbol{Y})$ and $\mathrm{MMSE}(t) = \lim_{N\to+\infty}\mathrm{MMSE}_N(t)$.

Since for $\kappa = 0$, $I(\boldsymbol{S};\boldsymbol{Y}) = 0$ the integration constant vanishes. Therefore, we just need to compute the integral over the asymptotic MMSE given by

Theorem 9.2. Using (9.44), we have

$$\begin{aligned}
I(\boldsymbol{S};\boldsymbol{Y}) &= \frac{1}{4}\int_0^{\kappa}\Big(\frac{1}{t}-\frac{1}{t}\Phi(\sqrt{t}S+Z)\Big)\,dt \\
&= \frac{1}{4}\int_0^{\kappa}\Big(\frac{1}{t}-\frac{1}{t^2}\Phi(S+\sqrt{\frac{1}{t}}Z)\Big)\,dt \\
&= \frac{1}{4}\int_{\frac{1}{\kappa}}^{\infty}\Big(\frac{1}{x}-\Phi(S+\sqrt{x}Z)\Big)\,dx \quad (\frac{1}{t}\to x) \\
&\overset{(a)}{=} \frac{1}{4}\int_{\frac{1}{\kappa}}^{\infty}\Big(\frac{1}{x}-\kappa\Phi(\sqrt{\kappa}S+\sqrt{\kappa x}Z)\Big)\,dx \\
&= \frac{1}{4}\int_{1}^{\infty}\Big(\frac{1}{y}-\Phi(\sqrt{\kappa}S+\sqrt{y}Z)\Big)\,dy \quad (\kappa x\to y) \\
&= \frac{1}{4}\int_{0}^{\infty}\Big(\frac{1}{t+1}-\Phi(\sqrt{\kappa}S+\sqrt{t+1}Z)\Big)\,dt \quad (y\to t+1) \\
&\overset{(b)}{=} \frac{1}{4}\int_{0}^{\infty}\Big(\frac{1}{t+1}-\Phi(\sqrt{\kappa}S+Z_0+\sqrt{t}Z)\Big)\,dt \\
&= \frac{1}{2}\chi(\sqrt{\kappa}S+Z_0)-\frac{1}{4}\ln 2\pi-\frac{1}{4} \quad (\text{from (9.43)})
\end{aligned} \tag{9.46}$$

where in (a), we use the relation $\Phi(cX)=\frac{1}{c^2}\Phi(X)$ for $c>0$, and in (b) Z and Z_0 two (mutually) free semi-circular non-commutative random variables. Therefore, we obtain

$$\lim_{N\to\infty}\frac{1}{N^2}I_N(\boldsymbol{S};\boldsymbol{Y})=\frac{1}{2}\iint \ln|s-t|\rho_Y(s)\rho_Y(t)\,ds\,dt+\frac{1}{8} \tag{9.47}$$

where $\rho_Y=\rho_{\sqrt{\kappa}S}\boxplus\rho_{\rm sc}$. □

9.6 Sub-Linear Rank RIE and MSE

The signal matrix $\boldsymbol{S}\in\mathbb{R}^{N\times N}$ is taken from a rotationally invariant prior and observed through an additive channel,

$$\boldsymbol{Y}=\sqrt{\kappa}\boldsymbol{S}+\boldsymbol{Z}$$

where the noise $\boldsymbol{Z}\in\mathbb{R}^{N\times N}$ is also distributed according to a rotationally invariant ensemble. Here we allow non-Gaussian noise. It is assumed that $\boldsymbol{S}$ has $M=\lfloor N^{\alpha}\rfloor$ non-zero i.i.d eigenvalues sampled from a distribution $\rho_S(x)$ with finite second moment and bounded support. By construction the empirical distribution of non-zeros eigenvalues $\frac{1}{M}\sum_{i=1}^{M}\delta(x-\lambda_i^S)$ tends weakly to $\rho_S(x)$. Moreover we assume that $\boldsymbol{Z}$ has a limiting spectral distribution ρ_Z. Since we are in the sub-linear regime, the limiting spectral measure of $\boldsymbol{Y}$ (normalized by $1/N$) is the same as the one of $\boldsymbol{Z}$. But note that $\boldsymbol{Y}$ may have sub-linear number of eigenvalues outside the support.

We propose a sub-linear rank RIE and an associated algorithm to estimate $\boldsymbol{S}$ from $\boldsymbol{Y}$, which we conjecture to be optimal. Arguing as in the linear case it is not difficult to see that the *optimal* RIE (i.e., the one minimizing the MSE in the general RIE class)must have a MSE which equals the MMSE. We conjecture that the *proposed* RIE, and associated algorithm, are indeed the optimal. Evidence for this conjecture comes from numerics discussed below (at least for some range of α), but also from the particular case of Gaussian noise. Indeed for Gaussian noise we have the I-MMSE relation so by integrating the MMSE we find the mutual information. Thus our proposed sub-linear rank RIE predicts an expression for the mutual information. On the other hand the mutual information has recently been rigorously computed from the asymptotics of sub-linear rank spherical integrals [65]. By comparing the expressions obtained by these two independent approaches we validate the conjecture analytically, at least for Gaussian noise.

9.6.1 Sub-linear rank RIE

Let the eigen-decomposition of $\boldsymbol{S}$ and $\boldsymbol{Y}$ be $\boldsymbol{S} = \sum_{i=1}^{M} \lambda_i^S \boldsymbol{s}_i \boldsymbol{s}_i^\intercal$, $\boldsymbol{Y} = \sum_{i=1}^{N} \lambda_i^Y \boldsymbol{y}_i \boldsymbol{y}_i^\intercal$. For a RIE $\Xi(\boldsymbol{Y}) = \sum_{i=1}^{N} \xi_i \boldsymbol{y}_i \boldsymbol{y}_i^\intercal$, the MSE can be written as:

$$\begin{aligned} \frac{1}{M}\|\boldsymbol{S} - \Xi(\boldsymbol{Y})\|_F^2 &= \frac{1}{M}\Big\| \sum_{i=1}^{M} \lambda_i^S \boldsymbol{s}_i \boldsymbol{s}_i^\intercal - \sum_{i=1}^{N} \xi_i \boldsymbol{y}_i \boldsymbol{y}_i^\intercal \Big\|_F^2 \\ &= \frac{1}{M}\sum_{i=1}^{M} \lambda_i^{S2} + \frac{1}{M}\sum_{i=1}^{N} \xi_i^2 - \frac{2}{M}\sum_{i=1}^{N} \xi_i \sum_{j=1}^{M} \lambda_j^S \big(\boldsymbol{s}_j^\intercal \boldsymbol{y}_i\big)^2. \end{aligned} \tag{9.48}$$

Minimizing (9.48) over ξ_i's, the optimum is achieved at

$$\xi_i^* = \sum_{j=1}^{M} \lambda_j^S \big(\boldsymbol{s}_j^\intercal \boldsymbol{y}_i\big)^2 = \boldsymbol{y}_i^\intercal \boldsymbol{S} \boldsymbol{y}_i. \tag{9.49}$$

This estimator is called *oracle estimator* since it requires the knowledge of the signal.

To compute the optimal eigenvalues we need to know the overlap between the eigenvectors of the signal and the observation. For finite-rank additive perturbation, $\boldsymbol{X}$ of a rotationally invariant matrix $\boldsymbol{Z}$, the eigenvalues and the overlap between the eigenvectors of the perturbation and the perturbed matrix $\boldsymbol{X} + \boldsymbol{Z}$ has been computed rigorously in [17]. In [126] this is extended to sub-linear rank perturbations. Essentially the same formulas (as for finite-rank perturbations) give the eigenvalues of the perturbed matrix and overlap in the large size limit.

These results could presumably be used to get a rigorous computation of the asymptotic MSE predicted by the oracle estimator but this is left for future work. Here we use these results in an heuristic way to propose a specific RIE

and associated algorithm. We use the notation $a \to b$ to mean that $|a-b|$ tends to zero with high probability as $N \to +\infty$. Theorem 2.7 in [126] suggests

$$\left(\boldsymbol{s}_i^\intercal \boldsymbol{y}_i\right)^2 \to \begin{cases} \frac{-1}{\kappa {\lambda_i^S}^2 \mathcal{G}'_{\rho_Z}\left(\mathcal{G}_{\rho_Z}^{-1}\left(\frac{1}{\sqrt{\kappa}\lambda_i^S}\right)\right)} & \text{if } \frac{1}{\sqrt{\kappa}\lambda_i^S} \in \left(\mathcal{G}_{\rho_Z}(a^-), \mathcal{G}_{\rho_Z}(b^+)\right), \\ 0 & \text{else.} \end{cases} \tag{9.50}$$

where $\mathcal{G}_{\rho_Z}(z)$ is the *Cauchy transform* of ρ_Z, constrained on $\mathbb{R}\backslash\operatorname{supp}\rho_Z$, and a, b are the infimum and supremum of the support of ρ_Z, and $\mathcal{G}_{\rho_Z}(a^-) \equiv \lim_{z\to a^-} \mathcal{G}_{\rho_Z}(z)$ and similarly for b^+. The overlap in (9.50) is expressed in terms of eigenvalues of $\boldsymbol{S}$, but since the corresponding eigenvalue of $\boldsymbol{Y}$ is affected by λ_i^S, we can express the overlap in terms of eigenvalues of $\boldsymbol{Y}$. Theorem 2.1 in [126] suggests

$$\lambda_i^Y \to \begin{cases} \mathcal{G}_{\rho_Z}^{-1}\left(\frac{1}{\sqrt{\kappa}\lambda_i^S}\right) & \text{if } \frac{1}{\sqrt{\kappa}\lambda_i^S} \in \left(\mathcal{G}_{\rho_Z}(a^-), \mathcal{G}_{\rho_Z}(b^+)\right), \\ b & \text{if } \frac{1}{\sqrt{\kappa}\lambda_i^S} > \mathcal{G}_{\rho_Z}(b^+), \\ a & \text{if } \frac{1}{\sqrt{\kappa}\lambda_i^S} > \mathcal{G}_{\rho_Z}(a^-). \end{cases} \tag{9.51}$$

From (9.51), we can see that if an eigenvalue of $\boldsymbol{Y}$ is outside the support of ρ_Z, then the corresponding eigenvalue of the signal can computed as $\lambda_i^S \approx \frac{1}{\sqrt{\kappa}\mathcal{G}_{\rho_Z}(\lambda_i^Y)}$. From (9.50), (9.51) we deduce that for an eigenvalue $\lambda_i^Y \in (a, b)$ there is no spike aligned with $\boldsymbol{y}_i$, because otherwise λ_i^Y would be outside of the support of ρ_Z. So, the corresponding ξ_i^*'s are zero for the eigenvalues in (a, b). On the other hand, since there are M spikes, at most M ξ_i^*'s are non-zero which makes the whole expression in (9.48) for the optimum ξ_i^*'s $O(1)$.

Finally the proposed RIE estimator is naturally constructed as follows

$$\begin{cases} \hat{\Xi}^*(\boldsymbol{Y}) = \sum_{i=1}^N \xi_i^* \boldsymbol{y}_i \boldsymbol{y}_i^\intercal, \\ \xi_i^* = -\frac{1}{\sqrt{\kappa}} \mathbb{I}\left(\lambda_i^Y \notin [a, b]\right) \frac{\mathcal{G}_{\rho_Z}\left(\lambda_i^Y\right)}{\mathcal{G}'_{\rho_Z}\left(\lambda_i^Y\right)}. \end{cases} \tag{9.52}$$

This provides an algorithm with the following steps to reconstruct the signal

(i) Compute spectral data $(\lambda_i^Y, \boldsymbol{y}_i)$ from the matrix $\boldsymbol{Y}$.

(ii) Apply the function f_Z to the eigenvalues

$$f_Z(x) = \begin{cases} -\frac{1}{\sqrt{\kappa}} \frac{\mathcal{G}_{\rho_Z}(x)}{\mathcal{G}'_{\rho_Z}(x)} & \text{if } x \notin [a, b] \\ 0 & \text{else.} \end{cases}$$

(iii) Construct the estimate as $\hat{\boldsymbol{S}} = \sum_{i=1}^N f_Z(\lambda_i^Y) \boldsymbol{y}_i \boldsymbol{y}_i^\intercal$.

The second algorithmic step requires knowledge of the limiting distribution of the noise which can in principle be computed from its distribution.

9.6.2 Gaussian noise

For $\rho_Z = \frac{1}{2\pi}\sqrt{4-x^2}$, we have that $\mathcal{G}_{\rho_{\rm sc}}(z) = \frac{z-\text{sign}(z)\sqrt{z^2-4}}{2}$. Thus, given matrix $\boldsymbol{Y}$ the estimator reads:

$$\begin{cases} \hat{\Xi}^*(\boldsymbol{Y}) = \sum_{i=1}^N \xi_i^* \boldsymbol{y}_i \boldsymbol{y}_i^\intercal, \\ \xi_i^* = \frac{1}{\sqrt{\kappa}} \mathbb{I}\big(|\lambda_i^Y| > 2\big)\, \text{sign}(\lambda_i^Y)\, \sqrt{{\lambda_i^Y}^2 - 4}. \end{cases}$$

MSE

To compute the MSE, we use (9.50) which is in terms of the eigenvalues of $\boldsymbol{S}$. We have

$$\big(\boldsymbol{s}_i^\intercal \boldsymbol{y}_i\big)^2 \to \begin{cases} 1 - \frac{1}{\kappa {\lambda_i^S}^2} & \text{if } \kappa {\lambda_i^S}^2 \geq 1, \\ 0 & \text{else}, \end{cases} \tag{9.53}$$

so the optimal eigenvalues of the estimator are

$$\xi_i^* \to \begin{cases} \lambda_i^S - \frac{1}{\kappa \lambda_i^S} & \text{if } \kappa {\lambda_i^S}^2 \geq 1, \\ 0 & \text{else}. \end{cases} \tag{9.54}$$

The MSE can be written as,

$$\frac{1}{M}\|\boldsymbol{S} - \Xi^*(\boldsymbol{Y})\|_F^2 = \frac{1}{M}\sum_{i=1}^M {\lambda_i^S}^2 - \frac{1}{M}\sum_{i=1}^N {\xi_i^*}^2 \tag{9.55}$$

where, we used that, in the limit $N \to \infty$, at most M number of ξ_i^*'s are non-zero. Taking the limit $N \to \infty$ (or $M \to \infty$), we find

$$\lim_{N\to\infty} \frac{1}{M}\|\boldsymbol{S} - \Xi^*(\boldsymbol{Y})\|_F^2 = \int x^2 \rho_S(x)\, dx - \int_{|x|\geq \frac{1}{\sqrt{\kappa}}} \big(x - \frac{1}{\kappa x}\big)^2 \rho_S(x)\, dx. \tag{9.56}$$

Mutual information in Gaussian noise

As already explained the MSE computed above should be equal to the MMSE, thus by integrating over κ we should recover the mutual information. Moreover, the mutual information can be computed using the limit of the spherical integrals of sub-linear rank [65]. In this section we explicitly check for a few priors that the two expressions indeed coincide.

We first present the main formula of [65]. The asymptotic mutual information between $\boldsymbol{S}$ and the observation $\boldsymbol{Y} = \sqrt{\kappa}\boldsymbol{S} + \boldsymbol{Z}$ is

$$\lim_{N\to\infty} \frac{1}{MN} I_N(\boldsymbol{S}; \boldsymbol{Y}) = \frac{\kappa}{2}\int x^2 \rho_S(x)\, dx - \lim_{N\to\infty} \frac{1}{MN} \ln \int D\boldsymbol{U} e^{\frac{N}{2}\,\text{Tr}\,\sqrt{\kappa}\boldsymbol{S}\boldsymbol{U}\boldsymbol{Y}\boldsymbol{U}^\intercal}. \tag{9.57}$$

The integral $\int D\boldsymbol{U} e^{\frac{N}{2}\operatorname{Tr}\sqrt{\kappa}\boldsymbol{S}\boldsymbol{U}\boldsymbol{Y}\boldsymbol{U}^\intercal}$ is the spherical integral of sub-linear rank, and its asymptotic limit has been studied in [65]. For matrix $\boldsymbol{A}$ with M positive eigenvalues, by Theorem 2.5 in [65], we have that

$$\lim_{N\to\infty}\frac{1}{MN}\ln\int D\boldsymbol{U}e^{\frac{N}{2}\operatorname{Tr}\boldsymbol{A}\boldsymbol{U}\boldsymbol{B}\boldsymbol{U}^\intercal}=\frac{1}{2}\lim_{N\to\infty}\frac{1}{M}\sum_{i=1}^{M}K(\theta_i,\lambda_i,\mu_B) \tag{9.58}$$

where θ_i's are non-zero eigenvalues of $\boldsymbol{A}$, λ_i's are the top eigenvalues of $\boldsymbol{B}$, and μ_B is the limiting spectral distribution of $\boldsymbol{B}$. Note that the result of [65] also covers the case of negative eigenvalues, however for simplicity we only consider matrices with positive eigenvalues. The function K is defined as

$$K(\theta,\lambda,\mu)=\theta\lambda'+(v-\lambda')\mathcal{G}_\mu(v)-\ln|\theta|-\int\ln|v-x|\,d\mu(x)-1$$

where $\lambda'=\max(\lambda,r(\mu))$ ($r(\mu)$ is the rightmost point of the support of μ),

$$v:=v(\lambda,\theta)=\begin{cases}\lambda' & \text{when } 0\le\mathcal{G}_\mu(\lambda')\le\theta \text{ or } \theta\le\mathcal{G}_\mu(\lambda')\le 0\\ \mathcal{G}_\mu^{-1}(\theta) & \text{else}\end{cases}$$

and $\mathcal{G}_\mu$ is the Stieltjes transform of μ.

Example 1: Sub-linear Wishart signal

Consider the signal matrix $\boldsymbol{S}$ to be $\frac{1}{N}\boldsymbol{X}\boldsymbol{X}^\intercal$, with $\boldsymbol{X}\in\mathbb{R}^{N\times M}$ has i.i.d. standard Gaussian entries, and $M=\lfloor N^\alpha\rfloor$. In the limit $N\to\infty$, one can show that the limiting distribution of non-zero eigenvalues of $\boldsymbol{S}$ is $\delta(x-1)$. For this example, the MSE given by (9.56) reads

$$\mathrm{MSE}(\kappa)=\begin{cases}1 & \text{if } \kappa\le 1,\\ \frac{1}{\kappa}\big(2-\frac{1}{\kappa}\big) & \text{if } \kappa\ge 1.\end{cases} \tag{9.59}$$

Integrating over κ, we find the mutual information to be:

$$\lim_{N\to\infty}\frac{1}{MN}I_N(\boldsymbol{S};\boldsymbol{Y})=\begin{cases}\frac{\kappa}{4} & \text{if } \kappa\le 1,\\ \frac{1}{4}\frac{1}{\kappa}+\frac{1}{2}\ln\kappa & \text{if } \kappa\ge 1.\end{cases} \tag{9.60}$$

which is the mutual information in the rank-one case when the prior for the spike is a Gaussian vector.

Now, we compute the mutual information using spherical integrals. All the non-zeros eigenvalues of sub-linear rank matrix $\sqrt{\kappa}\boldsymbol{S}$, θ_i's, converge to a single number, $\sqrt{\kappa}$. By [17], the limiting top eigenvalues of $\boldsymbol{Y}$, λ_i's, can also be computed. So, all the summands in r.h.s. of (9.58) are equal in the limit, and we have:

$$\lim_{N\to\infty}\frac{1}{MN}\ln\int D\boldsymbol{U}e^{\frac{N}{2}\operatorname{Tr}\sqrt{\kappa}\boldsymbol{S}\boldsymbol{U}\boldsymbol{Y}\boldsymbol{U}^\intercal}=\begin{cases}\frac{1}{2}K(\sqrt{\kappa},2,\rho_{\mathrm{sc}}) & \text{if } \kappa<1,\\ \frac{1}{2}K(\sqrt{\kappa},\sqrt{\kappa}+\frac{1}{\sqrt{\kappa}},\rho_{\mathrm{sc}}) & \text{if } \kappa\ge 1.\end{cases} \tag{9.61}$$

Case $\kappa < 1$: we have $\lambda' = 2$ and $v = \mathcal{G}_{\rho_{\rm sc}}^{-1}(\sqrt{\kappa}) = \sqrt{\kappa} + \frac{1}{\sqrt{\kappa}}$. Thus

$$\begin{aligned} K(\sqrt{\kappa}, 2, \rho_{\rm sc}) &= 2\sqrt{\kappa} + \big(\sqrt{\kappa} + \frac{1}{\sqrt{\kappa}} - 2\big)\mathcal{G}_{\rho_{\rm sc}}\big(\sqrt{\kappa} + \frac{1}{\sqrt{\kappa}}\big) - \ln\sqrt{\kappa} \\ &\qquad - \int_{-2}^{2} \frac{\sqrt{4-x^2}}{2\pi} \ln\big|\sqrt{\kappa} + \frac{1}{\sqrt{\kappa}} - x\big|\, dx - 1 \\ &= 2\sqrt{\kappa} + \big(\sqrt{\kappa} + \frac{1}{\sqrt{\kappa}} - 2\big)\sqrt{\kappa} - \frac{1}{2}\ln\kappa - \big(\frac{\kappa}{2} - \frac{1}{2}\ln\kappa\big) - 1 \\ &= \frac{\kappa}{2} \end{aligned} \tag{9.62}$$

where we used the integral formula

$$\begin{aligned} \frac{1}{2\pi}\int_{-2}^{2} \ln(A - Bx)\sqrt{4-x^2}dx = \frac{A}{A + \sqrt{A^2 - 4B^2}} + \ln\big(A + \sqrt{A^2 - 4B^2}\big) \\ - \frac{1}{2} - \ln 2 \end{aligned}$$

Case $\kappa \geq 1$: we have $\lambda' = \sqrt{\kappa} + \frac{1}{\sqrt{\kappa}}$ and $v = \lambda' = \sqrt{\kappa} + \frac{1}{\sqrt{\kappa}}$. Thus

$$\begin{aligned} K(\sqrt{\kappa}, \sqrt{\kappa} + \frac{1}{\sqrt{\kappa}}, \rho_{\rm sc}) &= \sqrt{\kappa}\big(\sqrt{\kappa} + \frac{1}{\sqrt{\kappa}}\big) - \ln\sqrt{\kappa} \\ &\qquad - \int_{-2}^{2} \frac{\sqrt{4-x^2}}{2\pi} \ln\big|\sqrt{\kappa} + \frac{1}{\sqrt{\kappa}} - x\big|\, dx - 1 \\ &= \kappa + 1 - \frac{1}{2}\ln\kappa - \big(\frac{1}{2}\frac{1}{\kappa} + \frac{1}{2}\ln\kappa\big) - 1 \\ &= \kappa - \frac{1}{2}\frac{1}{\kappa} - \ln\kappa. \end{aligned} \tag{9.63}$$

From (9.61), (9.62), and (9.63), we obtain

$$\lim_{N\to\infty} \frac{1}{MN} \ln \int D\boldsymbol{U} e^{\frac{N}{2}\operatorname{Tr}\sqrt{\kappa}\boldsymbol{S}\boldsymbol{U}\boldsymbol{Y}\boldsymbol{U}^\intercal} = \begin{cases} \frac{\kappa}{4} & \text{if } \kappa < 1, \\ \frac{\kappa}{2} - \frac{1}{4}\frac{1}{\kappa} - \frac{1}{2}\ln\kappa & \text{if } \kappa \geq 1. \end{cases} \tag{9.64}$$

Replacing this result in (9.57) we get

$$\lim_{N\to\infty} \frac{1}{MN} I_N(\boldsymbol{S}; \boldsymbol{Y}) = \begin{cases} \frac{\kappa}{4} & \text{if } \kappa < 1, \\ \frac{1}{4}\frac{1}{\kappa} + \frac{1}{2}\ln\kappa & \text{if } \kappa \geq 1. \end{cases}$$

which is the same as the mutual information computed from the MSE, (9.60).

In Fig. 9.6.1, we compare the performance of the sub-linear RIE and the oracle estimator (9.49) with $M = \lfloor\sqrt{N}\rfloor$ for various values of N. We observe that the performance of the RIE is very close to the one of oracle estimator (which requires the knowledge of the signal). Moreover, the MSE is close to the theoretical predictions.

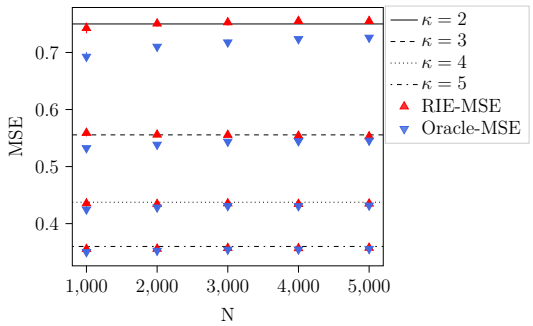


Figure 9.6.1: Comparison of the sub-linear RIE and the oracle estimator (9.49) for Gaussian noise with sub-linear Wishart signal with $M = \lfloor\sqrt{N}\rfloor$. Horizontal lines are MMSE computed from (9.59). Points are averaged over 10 experiments (error bars might be invisible).

Example 2: Uniform distribution

Let ρ_S be the uniform distribution on $[1, 2]$. From (9.56), the MSE can be computed to be

$$\mathrm{MSE}(\kappa) = \begin{cases} \frac{7}{3} & \text{for } 0 \le \kappa \le \frac{1}{4}, \\ -\frac{1}{3} + \frac{4}{\kappa} - \frac{8}{3\kappa^{3/2}} + \frac{1}{2\kappa^2} & \text{for } \frac{1}{4} \le \kappa \le 1, \\ \frac{2}{\kappa} - \frac{1}{2\kappa^2} & \text{for } 1 \le \kappa. \end{cases} \tag{9.65}$$

Integrating over κ, we find the mutual information to be:

$$\lim_{N\to\infty} \frac{1}{MN} I_N(\boldsymbol{S}; \boldsymbol{Y}) = \begin{cases} \frac{7}{12}\kappa & \text{for } 0 \le \kappa \le \frac{1}{4}, \\ -\frac{\kappa}{12} + \ln\kappa + 2(\ln 2 - 1) + \frac{4}{3\sqrt{\kappa}} - \frac{1}{8\kappa} & \text{for } \frac{1}{4} \le \kappa \le 1, \\ \frac{1}{2}\ln\kappa + 2\ln 2 - 1 + \frac{1}{8\kappa} & \text{for } 1 \le \kappa. \end{cases} \tag{9.66}$$

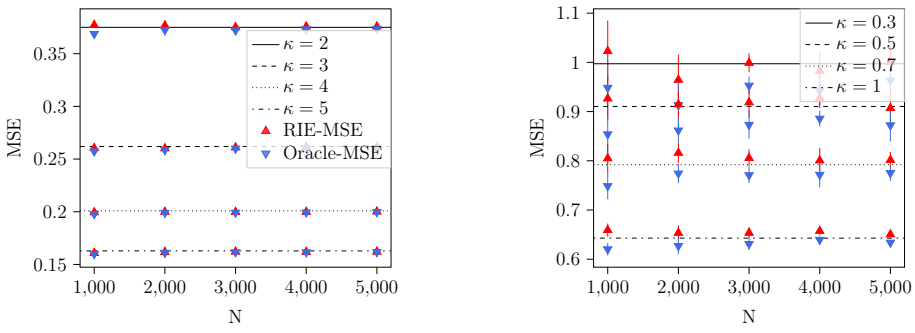
Now, we proceed with the computation of the mutual information using the asymptotic of spherical integrals. In the limit $N \to \infty$, r.h.s. of (9.58) becomes an integral (expectation w.r.t. ρ_S). For $\rho_S = \mathcal{U}\big([1, 2]\big)$ we have

$$\lim_{N\to\infty} \frac{1}{MN} \ln \int D\boldsymbol{U} e^{\frac{N}{2}\operatorname{Tr}\sqrt{\kappa}\boldsymbol{S}\boldsymbol{U}\boldsymbol{Y}\boldsymbol{U}^\intercal} = \frac{1}{2}\int_1^2 K\Big(\sqrt{\kappa}x, h\big(\sqrt{\kappa}x\big), \rho_{\mathrm{sc}}\Big)\, dx \tag{9.67}$$

with

$$h\big(\sqrt{\kappa}x\big) = \begin{cases} 2 & \text{if } \sqrt{\kappa}x \le 1, \\ \sqrt{\kappa}x + \frac{1}{\sqrt{\kappa}x} & \text{if } \sqrt{\kappa}x \ge 1. \end{cases}$$

$$K\Big(\sqrt{\kappa}x, h\big(\sqrt{\kappa}x\big), \rho_{\mathrm{sc}}\Big) = \begin{cases} \frac{1}{2}\kappa x^2 & \text{for } 0 < \kappa \le \frac{1}{4}, \\ \frac{1}{2}\kappa x^2 & \text{for } \frac{1}{4} \le \kappa \le 1 \text{ and } x \le \frac{1}{\sqrt{\kappa}}, \\ \kappa x^2 - \frac{1}{2}\frac{1}{\kappa x^2} - \ln\kappa x^2 & \text{for } \frac{1}{4} \le \kappa \le 1 \text{ and } x \ge \frac{1}{\sqrt{\kappa}}, \\ \kappa x^2 - \frac{1}{2}\frac{1}{\kappa x^2} - \ln\kappa x^2 & \text{for } 1 \le \kappa. \end{cases}$$



(a) high-SNR, $M = \lfloor \sqrt{N} \rfloor$ **(b)** low-SNR, $M = \lfloor \sqrt{N} \rfloor$

Figure 9.6.2: Comparison of the sub-linear RIE and the oracle estimator (9.49) for Gaussian noise with sub-linear signal with $\rho_S = \mathcal{U}([1,2])$. Horizontal lines are MSE computed from (9.65). Points are averaged over 10 experiments (error bars might be invisible).

Thus we find:

$$\lim_{N\to\infty} \frac{1}{MN} \ln \int D\boldsymbol{U} e^{\frac{N}{2} \operatorname{Tr} \sqrt{\kappa} \boldsymbol{S}\boldsymbol{U}\boldsymbol{Y}\boldsymbol{U}^\intercal} = \begin{cases} \frac{7}{12}\kappa & \text{for } 0 \le \kappa \le \frac{1}{4}, \\ \frac{5}{4}\kappa - \ln\kappa + 2(1-\ln 2) - \frac{4}{3\sqrt{\kappa}} + \frac{1}{8\kappa} & \text{for } \frac{1}{4} \le \kappa \le 1, \\ \frac{7}{6}\kappa - \frac{1}{2}\kappa + 1 - 2\ln 2 - \frac{1}{8\kappa} & \text{for } 1 \le \kappa. \end{cases} \tag{9.68}$$

Replacing in (9.57), we find the same mutual information as in (9.66).

In Fig. 9.6.2, we compare the performance of the sub-linear RIE and the oracle estimator (9.49) with $M = \lfloor \sqrt{N} \rfloor$ for various values of N as well as with theoretical predictions for a uniformly distributed signal.

9.6.3 Uniform noise

We now consider non-Gaussian noise, a noise matrix with eigenvalues which are uniformly distributed $\rho_Z = \mathcal{U}\big([1,2]\big)$. We have $\mathcal{G}_{\mathcal{U}([1,2])}(z) = \ln \frac{z-1}{z-2}$. Thus the eigenvalues of the proposed estimator are:

$$\xi_i^* = \frac{1}{\sqrt{\kappa}} \mathbb{I}\big(\lambda_i^Y \notin [1,2]\big) \ln \frac{\lambda_i^Y - 1}{\lambda_i^Y - 2} \big(\lambda_i^Y - 1\big)\big(\lambda_i^Y - 2\big).$$

MSE

Writing the overlap in terms of the eigenvalues of the signal, for $1 \le i \le M$ we have

$$\big(\boldsymbol{s}_i^\intercal \boldsymbol{y}_i\big)^2 \to \frac{1}{4} \frac{1}{\kappa \lambda_i^{S^2}} \big(\operatorname{csch} \frac{1}{2\sqrt{\kappa}\lambda_i^S}\big)^2. \tag{9.69}$$

Note that, $\mathcal{G}_{\mathcal{U}([1,2])}(1^-) = -\infty$ and $\mathcal{G}_{\mathcal{U}([1,2])}(2^+) = +\infty$, so for any $\kappa > 0$ we have an outlier eigenvalue for each spike in the signal. From (9.69), the eigenvalues

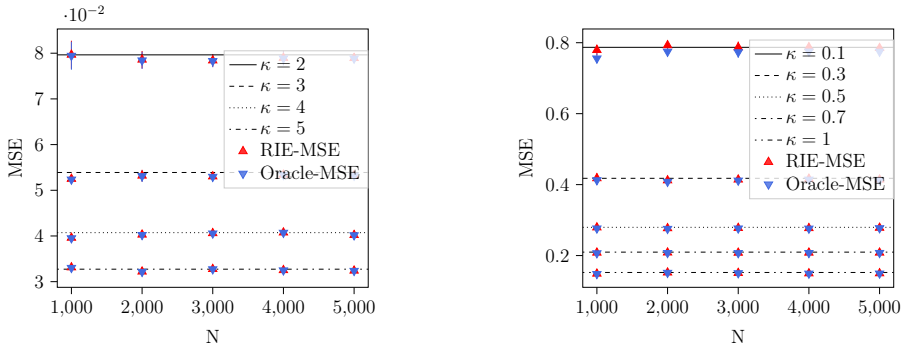


Figure 9.6.3: Comparison of the sub-linear RIE and the oracle estimator (9.49) for noise with uniform spectral distribution and sub-linear Wishart signal, for $M = \lfloor \sqrt{N} \rfloor$. Horizontal lines are MSE computed from $1 - \frac{1}{16}\frac{1}{\kappa^2}\big(\operatorname{csch}\frac{1}{2\sqrt{\kappa}}\big)^4$. Points are averaged over 10 experiments (error bars might be invisible). Note that unlike the Gaussian noise, for any SNR the estimation is possible and MSE is less than 1.

of the estimator are

$$\xi_i^* \to \frac{1}{4}\frac{1}{\kappa\lambda_i^S}\Big(\operatorname{csch}\frac{1}{2\sqrt{\kappa}\lambda_i^S}\Big)^2. \tag{9.70}$$

Using (9.55) we find in the asymptotic limit

$$\lim_{N\to\infty}\frac{1}{M}\|\boldsymbol{S} - \Xi^*(\boldsymbol{Y})\|_F^2 = \int\Big[x^2 - \frac{1}{16}\frac{1}{\kappa^2 x^2}\Big(\operatorname{csch}\frac{1}{2\sqrt{\kappa}x}\Big)^4\Big]\rho_S(x)\,dx. \tag{9.71}$$

Example: Sub-linear Wishart Signal

As mentioned in section 9.6.2, the limiting measure of the signal in this case is $\rho_S = \delta_{+1}$, so the MSE is $1 - \frac{1}{16}\frac{1}{\kappa^2}\big(\operatorname{csch}\frac{1}{2\sqrt{\kappa}}\big)^4$. In Fig. 9.6.3, we compare the performance of the sub-linear RIE and the oracle estimator (9.49) for $M = \lfloor \sqrt{N} \rfloor$ for various values of N. We observe that the performance of the RIE is very close to the one of oracle estimator (which requires the knowledge of the signal). Moreover, the MSE is close to the theoretical predictions.

Appendix

9.A Discussion of Models with Rotation Invariant Noise

Suppose that the noise matrix $\boldsymbol{Z}$ in our basic model is the realization of a rotation invariant ensemble. While we are not quite able to treat this case, we can generalize Theorem 9.3 to the setting $\boldsymbol{Y}_\epsilon = \sqrt{\kappa}\boldsymbol{S} + \boldsymbol{Z}_\epsilon$ where $\boldsymbol{Z}_\epsilon = \boldsymbol{Z} + \sqrt{\epsilon}\boldsymbol{\zeta}$ with $\boldsymbol{\zeta}$ from the Gaussian Wigner ensemble, and $\epsilon > 0$ (so the noise is non-Gaussian rotation invariant). We call this model the *Additive Rotation Invariant Noise* (ARIN) model.

Proposition 9.6 (**Explicit Mutual Information for the ARIN model**)**.** *Assume that the conditions in assumption 1 hold for both* $\boldsymbol{S}$ *and* $\boldsymbol{Z}$*. Then, we have for the ARIN model:*

$$\begin{aligned}\frac{I_N(\boldsymbol{S};\boldsymbol{Y}_\epsilon)}{N^2} \xrightarrow{N\to\infty} \frac{1}{2}\iint \ln|s-t|\rho_{Y_\epsilon}(s)\rho_{Y_\epsilon}(t)\,ds\,dt \\ - \frac{1}{2}\iint \ln|s-t|\rho_{Z_\epsilon}(s)\rho_{Z_\epsilon}(t)\,ds\,dt.\end{aligned} \tag{9.72}$$

The proof leverages only on the simple formula for the mutual information for Gaussian noise in theorem 9.3, and does not really hinge on assumption 9.2. Finally we would like to take the limit $\epsilon \to 0$. This however is a subtle problem which would require more specific hypothesis on the rotation invariant ensemble of $\boldsymbol{Z}$. For example it is quite apparent that one would need the existence of a density for the limiting empirical measures ρ_Z and ρ_{Y_0}. Moreover, this also requires an argument to permute the limits $N \to +\infty$ and $\epsilon \to 0$.

Proof. Set $\boldsymbol{X} = \sqrt{\kappa}\boldsymbol{S} + \boldsymbol{Z}$. We have the information theoretic equalities ($\mathcal{H}_N$ are Shannon entropies)

$$\begin{aligned} I_N(\boldsymbol{Y}_\epsilon;\boldsymbol{S}) &= \mathcal{H}_N(\boldsymbol{Y}_\epsilon) - \mathcal{H}_N(\boldsymbol{Y}_\epsilon|\boldsymbol{S}) \\ &= \mathcal{H}_N(\boldsymbol{Y}_\epsilon) - \mathcal{H}_N(\boldsymbol{Z}_\epsilon) \\ &= [\mathcal{H}_N(\boldsymbol{Y}_\epsilon) - \mathcal{H}_N(\boldsymbol{\zeta})] - [\mathcal{H}_N(\boldsymbol{Z}_\epsilon) - \mathcal{H}_N(\boldsymbol{\zeta})] \\ &= [\mathcal{H}_N(\boldsymbol{Y}_\epsilon) - \mathcal{H}_N(\boldsymbol{Y}_\epsilon|\boldsymbol{X})] - [\mathcal{H}_N(\boldsymbol{Z}_\epsilon) - \mathcal{H}_N(\boldsymbol{Z}_\epsilon|\boldsymbol{Z})] \\ &= I_N(\boldsymbol{Y}_\epsilon;\boldsymbol{X}) - I_N(\boldsymbol{Z}_\epsilon;\boldsymbol{Z}). \end{aligned} \tag{9.73}$$

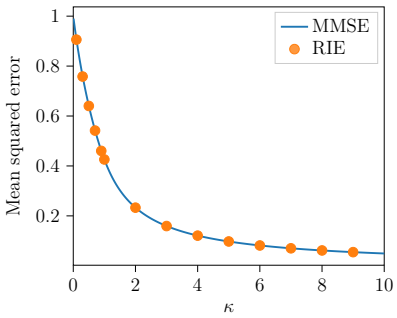


Figure 9.B.1: MMSE and the MSE of RIE (9.4) for the signal with spectrum $\rho_S = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}$. The MMSE is continuous w.r.t. κ. The RIE is computed for $N = 1000$, and the results are averaged over 20 runs (error bars are invisible).

The two mutual informations in the last line correspond to the inference models for two AWGN models with strength ϵ, namely $\boldsymbol{Y}_\epsilon = \boldsymbol{X} + \sqrt{\epsilon}\boldsymbol{\zeta}$ and $\boldsymbol{Z}_\epsilon = \boldsymbol{Z} + \sqrt{\epsilon}\boldsymbol{\zeta}$, and inputs from rotation invariant ensembles. By the formula in Theorem 9.3 which holds for these channels we obtain

$$\begin{aligned}\lim_{N\to+\infty} \frac{1}{N^2} I_N(\boldsymbol{Y}_\epsilon; \boldsymbol{S}) = \frac{1}{2}\iint \ln|s-t|\rho_{Y_\epsilon}(s)\rho_{Y_\epsilon}(t)dsdt \\ - \frac{1}{2}\iint \ln|s-t|\rho_{Z_\epsilon}(s)\rho_{Z_\epsilon}(t)dsdt.\end{aligned} \tag{9.74}$$

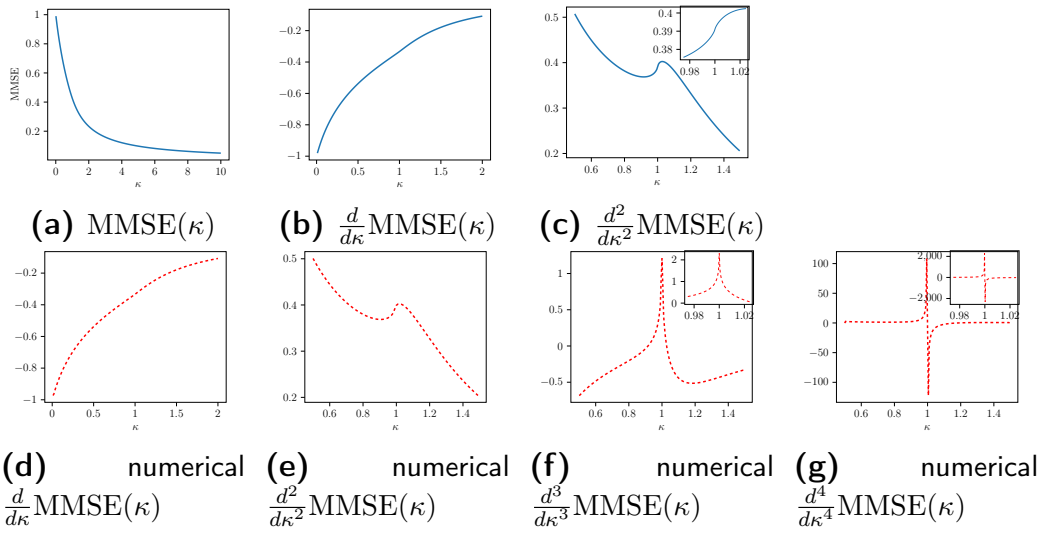
□

9.B Examples and Numerical Calculations for Linear Ranks

9.B.1 Signal with Rademacher spectrum

In this example, we consider the case where $\rho_S = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}$. Using the technique introduced in [118], we compute $\rho_Y = \rho_{\sqrt{\kappa}S} \boxplus \rho_{\text{sc}}$ in Appendix 9.D. Fig. 9.D.1 shows the support of ρ_Y which consists of two disjoint intervals when $\kappa \geq 1$, and one single open set when $\kappa < 1$. Therefore, we expect that a phase transition, if it exists, should happen at a value $\kappa_c = 1$. By Theorem 9.2 the MMSE is a continuous function of κ, and the phase transition (if it exists) is of the second or higher order. The MMSE and the performance of the RIE for this example are plotted in figure 9.B.1.

From the expression of ρ_Y in Appendix 9.D and from Theorem 9.2 we provide integral representations of the derivatives of the MMSE w.r.t. κ. These integrals are computed numerically and the result illustrated in figures 9.B.2 (a-b-c) for the MMSE and its first and second derivatives. For the third and the fourth derivatives the integral representation become unwieldy, so we only computed it numerically from the second derivative, as shown in figures 9.B.2 (f),(g). Based on these plots, we see that the third derivative (of the MMSE)



(a) $\mathrm{MMSE}(\kappa)$ **(b)** $\frac{d}{d\kappa}\mathrm{MMSE}(\kappa)$ **(c)** $\frac{d^2}{d\kappa^2}\mathrm{MMSE}(\kappa)$

(d) numerical $\frac{d}{d\kappa}\mathrm{MMSE}(\kappa)$ **(e)** numerical $\frac{d^2}{d\kappa^2}\mathrm{MMSE}(\kappa)$ **(f)** numerical $\frac{d^3}{d\kappa^3}\mathrm{MMSE}(\kappa)$ **(g)** numerical $\frac{d^4}{d\kappa^4}\mathrm{MMSE}(\kappa)$

Figure 9.B.2: Analysis of the MMSE in example 9.B.1. In plots (a),(b),(c) MMSE and its first and second derivatives are computed using the the expression (9.38). Plots (d), (e), (f) are the numerical differentiation of plots (a),(b), (c) respectively. The fourth derivative of the MMSE is computed from the curve in plot (c) by numerical second order differentiation. The first three plots shows that the MMSE, its first and second derivatives are continuous. But, plots (f), (g) suggests that the $\frac{d^2}{d\kappa^2}\mathrm{MMSE}(\kappa)$ has a vertical tangent at $\kappa_c = 1$, and MMSE has a phase transition of third order at this point.

at $\kappa_c = 1$ does not seem to exist, and therefore the free energy (9.11) (or mutual information) might have a fourth order phase transition at this point. Further numerical analysis in appendix 9.D is compatible with a behavior of the function $\mathrm{MMSE}(\kappa)$ close to the point $\kappa_c = 1$ of the form

$$\begin{aligned}\mathrm{MMSE}(\kappa) \approx \mathrm{MMSE}(1) + \mathrm{MMSE}'(1)(\kappa-1) + \frac{1}{2}\mathrm{MMSE}''(1)(\kappa-1)^2 \\ + \alpha(\kappa-1)^3\big(\ln|\kappa-1| + \beta\big) + o\big((\kappa-1)^3\big)\end{aligned} \tag{9.75}$$

with $\alpha \approx -0.06125$, $\beta \approx 1.411$.

9.B.2 Signal with Bernoulli spectrum

Let $\rho_S = p\delta_0 + (1-p)\delta_{+1}$. The corresponding signal matrix is not full-rank but it has a rank linear in N. For this prior, the spectrum of $\rho_Y = \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}$ is computed in Appendix 9.E using a similar technique as in the previous example. Depending on the SNR parameter κ the support of ρ_Y can be a single interval or is composed of two disjoint intervals as shown on figure 9.E.1. The MMSE and the MSE of RIE are illustrated in figure 9.B.3 for the two values $p = 0.9$ and $p = 0.3$.

In figure 9.B.4, the suitably normalized MMSE is plotted for the highly sparse case where p tends to 1. The MMSE is normalized by dividing by $p(1-p)$. We observe that as $p \to 1$ the MMSE approaches the MMSE of the

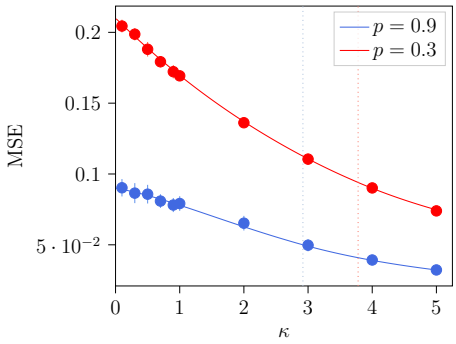


Figure 9.B.3: MMSE computed for the signal with Bernoulli spectrum $\rho_S = p\delta_0 + (1-p)\delta_{+1}$ for $p = 0.3$ and 0.9. The MMSE is continuous w.r.t. κ. The vertical dashed lines corresponds to the values of κ, i.e., 2.92 for $p = 0.9$ and 3.78 for $p = 0.3$, where the disjoint intervals of the support of ρ_Y merge. We do not observe any phase transition for any low order derivative at these values. The MSE of RIE is computed for $N = 1000$, and the results are averaged over 20 runs.

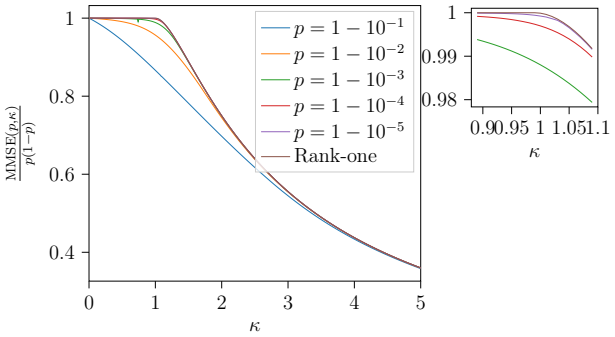


Figure 9.B.4: Normalized MMSE for the signal with spectrum $\rho_S = p\delta_0 + (1-p)\delta_{+1}$ for $p \to 1$. The MMSE of the rank-one problem is also plotted for comparison.

rank-one symmetric matrix estimation problem, which has a phase transition at $\kappa_c = 1$.

9.B.3 Wishart matrix

In this example, we consider the signal matrix $\boldsymbol{S}$ to be $\frac{1}{N}\boldsymbol{X}\boldsymbol{X}^\intercal$, where $\boldsymbol{X} \in \mathbb{R}^{N\times M}$ has i.i.d. standard Gaussian entries. We look at the limit of aspect ratio $\frac{N}{M} \to q$. Then, the limiting spectral distribution of $\boldsymbol{S}$ is a rescaling of the usual *Marchenko-Pastur* distribution by the factor α:

$$\rho_S(x) = \left(1 - \frac{1}{q}\right)^+ \delta(x) + \frac{\sqrt{\left(x - \left(\frac{1}{\sqrt{q}} - 1\right)^2\right)\left(\left(\frac{1}{\sqrt{q}} + 1\right)^2 - x\right)}}{2\pi x}. \tag{9.76}$$

The limiting spectral distribution of $\boldsymbol{Y}$ is computed in Appendix 9.F. For $q > 1$ the support of ρ_Y is the union of two disjoint intervals if $\kappa > q\left(\sqrt[3]{q} - 1\right)^{-3}$, and is a single interval otherwise. As in the previous example, we expect that in

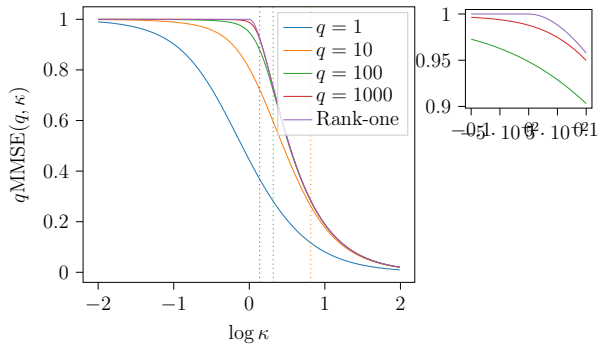


Figure 9.B.5: Normalized MMSE for the signal with the Marchenko-Pastur spectral distribution for large q's . The MMSE of the rank-one problem is also plotted for comparison. The vertical dashed lines corresponds to the critical value where the support of ρ_Y splits into two intervals. We do not find any phase transition on low order derivatives

the high sparse regime, the MMSE behaves like the low-rank case. In figure 9.B.5 the MMSE is illustrated large q's.

9.B.4 Finite-rank deformation of a Wigner matrix as signal

First, consider the case where $\boldsymbol{S}$ is a standard Wigner matrix, then by independence of $\boldsymbol{S}$ and $\boldsymbol{Z}$, $\boldsymbol{Y}$ is also a Wigner matrix with variance $\kappa + 1$, and ρ_Y is a semi-circle law of variance $\kappa + 1$. From (9.47), we find that $\frac{1}{N^2} I_N(\boldsymbol{S}; \boldsymbol{Y})$ converges to $\frac{1}{4}\ln(\kappa+1)$. This limit could also be obtained using the Gaussianity of entries of the matrices.

Now, let $\boldsymbol{S}$ be a finite rank deformation of a Wigner matrix, $\boldsymbol{S} = \boldsymbol{A}+\boldsymbol{\zeta}$, where $\boldsymbol{A}$ is a *finite-rank* symmetric matrix, and $\boldsymbol{\zeta} \in \mathbb{R}^{N\times N}$ is a symmetric Gaussian matrix with variance $\frac{1}{N}$ for non-diagonal, and $\frac{2}{N}$ for diagonal entries. We have the observation $\boldsymbol{Y} = \sqrt{\kappa}\boldsymbol{S} + \boldsymbol{Z}$. Since $\boldsymbol{A}$ has finite rank, the limiting spectral measure of $\boldsymbol{S}$ is the semicircle law, and average mutual information $\frac{1}{N^2} I_N(\boldsymbol{S}; \boldsymbol{Y})$ converges to $\frac{1}{4}\ln(\kappa+1)$, (9.47). Since $\boldsymbol{Z}$ and $\boldsymbol{\zeta}$ are independent, the observation matrix $\boldsymbol{Y}$ has the same distribution as the matrix $\tilde{\boldsymbol{Y}} = \sqrt{\kappa}\boldsymbol{A} + \sqrt{\kappa+1}\tilde{\boldsymbol{Z}}$, where $\tilde{\boldsymbol{Z}}$ is a symmetric Gaussian matrix, so $\mathcal{H}(\boldsymbol{Y}) = \mathcal{H}(\tilde{\boldsymbol{Y}})$. Define the matrix $\boldsymbol{Y}' = \sqrt{\frac{\kappa}{\kappa+1}}\boldsymbol{A} + \boldsymbol{Z}'$. By symmetry of the matrices, we have:

$$\mathcal{H}_N(\boldsymbol{Y}') = \mathcal{H}_N(\boldsymbol{Y}) - \frac{N(N+1)}{2}\ln\sqrt{\kappa+1}. \tag{9.77}$$

Here $\boldsymbol{Y}'$ has the form of the observation matrix in the low-rank matrix estimation, which has been extensively studied in [8, 9], and in particular, it has been shown that under suitable assumptions, the average mutual information $\frac{1}{N} I_N(\boldsymbol{A}; \boldsymbol{Y}')$ converges to a finite value. By definition of mutual information,

and (9.77), we have:

$$\begin{aligned}\frac{1}{N}I_N(\boldsymbol{A};\boldsymbol{Y}') &= \frac{1}{N}\mathcal{H}_N(\boldsymbol{Y}') - \frac{1}{N}\mathcal{H}_N(\boldsymbol{Z}') \\ &= \frac{1}{N}\mathcal{H}_N(\boldsymbol{Y}) - \frac{1}{N}\mathcal{H}_N(\boldsymbol{Z}') - \frac{N+1}{4}\ln(\kappa+1) \\ &= \frac{1}{N}I_N(\boldsymbol{S};\boldsymbol{Y}) - \frac{N+1}{4}\ln(\kappa+1).\end{aligned} \tag{9.78}$$

Dividing by N and taking the limit, we find the asymptotic mutual information

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N^2}I_N(\boldsymbol{S};\boldsymbol{Y}) &= \lim_{N\to\infty}\frac{1}{N^2}I_N(\boldsymbol{A};\boldsymbol{Y}') + \frac{N+1}{4N}\ln(\kappa+1) \\ &= \frac{1}{4}\ln(\kappa+1).\end{aligned} \tag{9.79}$$

where we used the fact that $\frac{1}{N}I_N(\boldsymbol{A};\boldsymbol{Y}')$ has a finite limit. Moreover, from (9.78) the finite-size correction term for the asymptotic average mutual information can be derived as

$$\lim_{N\to\infty} N\big(\frac{1}{N^2}I_N(\boldsymbol{S};\boldsymbol{Y}) - \frac{1}{4}\ln(\kappa+1)\big) = \lim_{N\to\infty}\frac{1}{N}I_N(\boldsymbol{A};\boldsymbol{Y}') + \frac{1}{4}\ln(\kappa+1). \tag{9.80}$$

Therefore when $\boldsymbol{S}$ is a finite rank deformation of the Wigner matrix, the finite-size correction term is directly related to the mutual information in the low-rank matrix estimation problem, which may exhibit a phase transition.

Now, let $\boldsymbol{A}$ be a *rank-one* matrix. $\boldsymbol{S} = \frac{\sqrt{\eta}}{N}\boldsymbol{x}\boldsymbol{x}^\intercal + \boldsymbol{\zeta}$ where $\boldsymbol{x} \in \mathbb{R}^N$ has i.i.d. components distributed according to the standard normal distribution. The matrix $\boldsymbol{Y}' = \frac{\sqrt{\frac{\eta\kappa}{\kappa+1}}}{N}\boldsymbol{x}\boldsymbol{x}^\intercal + \boldsymbol{Z}'$ is a rescaling of the rank-one model studied in [9], and since mutual information is invariant under rescaling, using Theorem 1 in [9], we have:

$$\lim_{N\to\infty}\frac{1}{N}I_N(\boldsymbol{x};\boldsymbol{Y}') = \begin{cases}\frac{1}{4}\frac{\eta\kappa}{\kappa+1} & \text{if } \frac{\eta\kappa}{\kappa+1} \le 1, \\ \frac{1}{4}\frac{\kappa+1}{\eta\kappa} + \frac{1}{2}\ln\frac{\eta\kappa}{\kappa+1} & \text{else.}\end{cases} \tag{9.81}$$

Fix $\eta \le 1$. For all $\kappa > 0$, we have $\frac{\eta\kappa}{\kappa+1} \le 1$ and there is no phase transition in the mutual information. On the other hand, for $\eta > 1$ mutual information has a phase transition at $\kappa = \frac{1}{\eta-1}$. From random matrix theory [17], we know that For $\eta \le 1$, in the limit $N \to \infty$, all the eigenvalues of $\boldsymbol{S}$ are inside the bulk, whereas for $\eta > 1$ one eigenvalue (which is the largest one) is outside the bulk. Therefore, for this particular signal, we can relate the phase transition of the correction term to the existence of an eigenvalue outside the bulk of ρ_S in the asymptotic limit.

9.C Further details for proof of Theorem 9.1

9.C.1 Proof of lemma 9.1

To prove lemma 9.1 we first need four preliminary lemmas.

Lemma 9.4. *For any N, and $N \times N$ symmetric matrices $\boldsymbol{A}, \boldsymbol{B}$ with spectral radius $r^{(\boldsymbol{A})}, r^{(\boldsymbol{B})}$*

$$-\frac{1}{2} r^{(\boldsymbol{A})} r^{(\boldsymbol{B})} \leq \mathcal{J}_N(\boldsymbol{A}, \boldsymbol{B}) \leq \frac{1}{2} r^{(\boldsymbol{A})} r^{(\boldsymbol{B})}.$$

Proof. Let $\boldsymbol{A} = \boldsymbol{U_A} \boldsymbol{\Lambda_A} \boldsymbol{U_A^\intercal}$, $\boldsymbol{B} = \boldsymbol{U_B} \boldsymbol{\Lambda_B} \boldsymbol{U_B^\intercal}$ be the eigendecomposition of $\boldsymbol{A}$, $\boldsymbol{B}$. We can write

$$\mathcal{I}_N(\boldsymbol{A}, \boldsymbol{B}) = \int D\boldsymbol{U} e^{\frac{N}{2} \operatorname{Tr} \boldsymbol{U} \boldsymbol{\Lambda_A} \boldsymbol{U}^\intercal \boldsymbol{\Lambda_B}} = \int D\boldsymbol{U} e^{\frac{N}{2} \sum_{i,j}^N \lambda_i^{(\boldsymbol{A})} \lambda_j^{(\boldsymbol{B})} U_{ij}^2}.$$

For each for all $i, j \in \{1, \ldots, N\}$, $-r^{(\boldsymbol{A})} r^{(\boldsymbol{B})} \leq \lambda_i^{(\boldsymbol{A})} \lambda_j^{(\boldsymbol{B})} \leq r^{(\boldsymbol{A})} r^{(\boldsymbol{B})}$. Therefore, we get

$$\begin{aligned}
\mathcal{I}_N(\boldsymbol{A}, \boldsymbol{B}) &= \int D\boldsymbol{U} e^{\frac{N}{2} \sum_{i,j}^N \lambda_i^{(\boldsymbol{A})} \lambda_j^{(\boldsymbol{B})} U_{ij}^2} \\
&\leq \int D\boldsymbol{U} e^{\frac{N}{2} r^{(\boldsymbol{A})} r^{(\boldsymbol{B})} \sum_{i,j}^N U_{ij}^2} \\
&= \int D\boldsymbol{U} e^{\frac{N^2}{2} r^{(\boldsymbol{A})} r^{(\boldsymbol{B})}} \\
&= e^{\frac{N^2}{2} r^{(\boldsymbol{A})} r^{(\boldsymbol{B})}}.
\end{aligned}$$

Similarly we can obtain $\mathcal{I}_N(\boldsymbol{A}, \boldsymbol{B}) \geq e^{-\frac{N^2}{2} r^{(\boldsymbol{A})} r^{(\boldsymbol{B})}}$. Therefore

$$-\frac{1}{2} r^{(\boldsymbol{A})} r^{(\boldsymbol{B})} \leq \mathcal{J}_N(\boldsymbol{A}, \boldsymbol{B}) \leq \frac{1}{2} r^{(\boldsymbol{A})} r^{(\boldsymbol{B})}.$$

□

Lemma 9.5. *Let $r^{(\tilde{\tilde{\boldsymbol{Y}}})}$ denote the spectral radius of the matrix $\tilde{\tilde{\boldsymbol{Y}}} = \sqrt{\kappa} \boldsymbol{\Lambda}^0 + \tilde{\boldsymbol{Z}}$. For $k > 2 + \sqrt{\kappa} C$, we have*

$$\mathbf{P}\{r^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq k\} \leq 4 e^{-\frac{N}{4}(k - \sqrt{\kappa} C - 2)^2}.$$

Proof. Denote the top and bottom eigenvalues of $\tilde{\tilde{\boldsymbol{Y}}}$ by $\lambda_{\max}^{(\tilde{\tilde{\boldsymbol{Y}}})}$, $\lambda_{\min}^{(\tilde{\tilde{\boldsymbol{Y}}})}$. By Weyls' inequality,

$$\begin{aligned}
\lambda_{\max}^{(\tilde{\tilde{\boldsymbol{Y}}})} &\leq \sqrt{\kappa} \max_i \lambda_i^0 + \lambda_{\max}^{(\tilde{\boldsymbol{Z}})} \leq \sqrt{\kappa} C + \lambda_{\max}^{(\tilde{\boldsymbol{Z}})} \\
\lambda_{\min}^{(\tilde{\tilde{\boldsymbol{Y}}})} &\geq \sqrt{\kappa} \min_i \lambda_i^0 + \lambda_{\min}^{(\tilde{\boldsymbol{Z}})} \geq -\sqrt{\kappa} C + \lambda_{\min}^{(\tilde{\boldsymbol{Z}})}
\end{aligned}$$

where $\lambda_{\max}^{(\tilde{\boldsymbol{Z}})}$,$\lambda_{\min}^{(\tilde{\boldsymbol{Z}})}$ are the top and bottom eigenvalues of $\tilde{\boldsymbol{Z}}$. Thus, we can write

$$\begin{aligned}
\mathbf{P}\{\lambda_{\max}^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq k\} &\leq \mathbf{P}\big\{\lambda_{\max}^{(\tilde{\boldsymbol{Z}})} + \sqrt{\kappa} C \geq k\big\} \\
&= \mathbf{P}\{\lambda_{\max}^{(\tilde{\boldsymbol{Z}})} \geq k - \sqrt{\kappa} C\}.
\end{aligned}$$

By [116] (Theorem II.11), for $k - \sqrt{\kappa}C > 2$, we have

$$\mathbf{P}\{\lambda_{\max}^{(\tilde{\boldsymbol{Z}})} \geq k - \sqrt{\kappa}C\} \leq e^{-\frac{N}{4}(k-\sqrt{\kappa}C-2)^2}$$

and therefore we get

$$\mathbf{P}\{\lambda_{\max}^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq k\} \leq e^{-\frac{N}{4}(k+\sqrt{\kappa}C-2)^2}. \tag{9.82}$$

For the bottom eigenvalue we have

$$\begin{aligned}\mathbf{P}\{\lambda_{\min}^{(\tilde{\tilde{\boldsymbol{Y}}})} \leq -k\} &\leq \mathbf{P}\big\{ -\sqrt{\kappa}C + \lambda_{\min}^{(\tilde{\boldsymbol{Z}})} \leq -k\big\} \\ &= \mathbf{P}\big\{\lambda_{\min}^{(\tilde{\boldsymbol{Z}})} \leq \sqrt{\kappa}C - k\big\} \\ &\leq e^{-\frac{N}{4}(k-\sqrt{\kappa}C-2)^2}.\end{aligned} \tag{9.83}$$

From (9.82), (9.83), we get

$$\begin{aligned}\mathbf{P}\{r^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq k\} &\leq \mathbf{P}\{|\lambda_{\max}^{(\tilde{\tilde{\boldsymbol{Y}}})}| \geq k\} + \mathbf{P}\{|\lambda_{\min}^{(\tilde{\tilde{\boldsymbol{Y}}})}| \geq k\} \\ &\leq \mathbf{P}\{\lambda_{\max}^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq k\} + \mathbf{P}\{\lambda_{\max}^{(\tilde{\tilde{\boldsymbol{Y}}})} \leq -k\} + \mathbf{P}\{\lambda_{\min}^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq k\} + \mathbf{P}\{\lambda_{\min}^{(\tilde{\tilde{\boldsymbol{Y}}})} \leq -k\} \\ &\leq 2\mathbf{P}\{\lambda_{\max}^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq k\} + 2\mathbf{P}\{\lambda_{\min}^{(\tilde{\tilde{\boldsymbol{Y}}})} \leq -k\} \\ &\leq 4e^{-\frac{N}{4}(k-\sqrt{\kappa}C-2)^2}\end{aligned}$$

for $k > 2 + \sqrt{\kappa}C$. □

Lemma 9.6. *For any polynomial function g, and k a sufficiently large constant, we have that*

$$\lim_{n\to\infty} \mathbb{E}\big[g(r^{(\tilde{\tilde{\boldsymbol{Y}}})})\mathbb{I}\{r^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq k\}\big] = 0.$$

Proof. By linearity of expectation, it is enough to consider the case $g(x) = x^i$. Let $X = {r^{(\tilde{\tilde{\boldsymbol{Y}}})}}^i \mathbb{I}\{r^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq k\}$ a non-negative random variable. We have

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty \mathbf{P}(X \geq x)dx \\ &= i\int_0^\infty \mathbf{P}(X \geq x^i)x^{i-1}dx \\ &= i\int_0^\infty \mathbf{P}\big({r^{(\tilde{\tilde{\boldsymbol{Y}}})}}^i \mathbb{I}\{r^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq k\} \geq x^i\big)x^{i-1}dx \\ &= i\int_0^\infty \mathbf{P}(r^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq k, r^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq x)x^{i-1}dx \\ &= i\int_0^k \mathbf{P}(r^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq k)x^{i-1}dx + i\int_k^\infty \mathbf{P}(r^{(\tilde{\tilde{\boldsymbol{Y}}})} \geq x)x^{i-1}dx \\ &\leq 4e^{-\frac{N}{4}(k-\sqrt{\kappa}C-2)^2}k^i + 4i\int_k^\infty e^{-\frac{N}{4}(x-\sqrt{\kappa}C-2)^2}dx \\ &\leq 4e^{-\frac{N}{4}(k-\sqrt{\kappa}C-2)^2}k^i + 4i\int_0^\infty e^{-\frac{N}{4}(x-\sqrt{\kappa}C-2)^2}x^{i-1}dx.\end{aligned}$$

The first term converges to 0 as $N \to \infty$. The second term involves moments of a Gaussian with variance $\frac{1}{2N}$, one can see that the second term also converges to 0. Thus, $\lim_{N\to\infty} \mathbb{E}[X] = 0$. □

Lemma 9.7. *For the sequence of random matrices $\tilde{\tilde{\boldsymbol{Y}}}$, the log spherical integral $\mathcal{J}_N\big(\sqrt{\kappa}\boldsymbol{\Lambda}^0, \tilde{\tilde{\boldsymbol{Y}}}\big)$ converges almost surely to a well-defined limit, denoted by $\mathcal{J}[\rho_{\sqrt{\kappa}S}, \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}]$.*

Proof. By assumption 1.A the support of $\hat{\mu}^{(N)}_{\sqrt{\kappa}\boldsymbol{\Lambda}^0}$ is included in a compact subset of $\mathbb{R}$,$[-\sqrt{\kappa}C, \sqrt{\kappa}C]$, for all $N \in \mathbb{N}$. Moreover, by the law of large numbers, $\hat{\mu}^{(N)}_{\sqrt{\kappa}\boldsymbol{\Lambda}^0}$ converges weakly towards $\rho_{\sqrt{\kappa}S}$.

Consider the sequence $\sqrt{\kappa}\boldsymbol{\Lambda}^0 + \tilde{\boldsymbol{Z}}$. For each matrix in the sequence, the second moment of the empirical spectral distribution is:

$$\frac{1}{N}\operatorname{Tr}(\sqrt{\kappa}\boldsymbol{\Lambda}^0 + \tilde{\boldsymbol{Z}})^2 = \frac{1}{N}\kappa \operatorname{Tr} \boldsymbol{\Lambda}^{0^2} + \frac{2}{N}\sqrt{\kappa} \operatorname{Tr} \boldsymbol{\Lambda}^0 \tilde{\boldsymbol{Z}} + \frac{1}{N} \operatorname{Tr} \tilde{\boldsymbol{Z}}^2.$$

The first term is bounded (for all N) by the construction of $\boldsymbol{\lambda}^0$. The last term is also bounded since the second moment of the sequence of Wigner matrices converges to 1 almost surely. For the second term, we have

$$\begin{aligned}\frac{1}{N} \operatorname{Tr} \boldsymbol{\Lambda}^0 \tilde{\boldsymbol{Z}} &\leq \frac{1}{N}\sqrt{\sum \kappa_i^{0^2}}\sqrt{\operatorname{Tr} \tilde{\boldsymbol{Z}}^2} \\ &\leq C\sqrt{\frac{1}{N} \operatorname{Tr} \tilde{\boldsymbol{Z}}^2}\end{aligned}$$

which is bounded for all N, since $\frac{1}{N} \operatorname{Tr} \tilde{\boldsymbol{Z}}^2$ is a convergent sequence (a.s.). Therefore, the sequence of matrices $\sqrt{\kappa}\boldsymbol{\Lambda}^0 + \tilde{\boldsymbol{Z}}$ has bounded second moment for all N almost surely. Moreover, according to [66], by the independence of $\boldsymbol{\lambda}^0$ and $\tilde{\boldsymbol{Z}}$, the empirical spectral distribution of this sequence converges weakly, almost surely to the free additive convolution of $\rho_{\sqrt{\kappa}S}$ with the semi-circle law ρ_{sc}.

Therefore, the conditions of theorem 1 in [61] hold a.s. for the sequence $\boldsymbol{\Lambda}^0, \sqrt{\kappa}\boldsymbol{\Lambda}^0 + \tilde{\boldsymbol{Z}}$. Hence, $\mathcal{J}_N\big(\sqrt{\kappa}\boldsymbol{\Lambda}^0, \tilde{\tilde{\boldsymbol{Y}}}\big)$ has a well-defined limit which is a function of $\rho_{\sqrt{\kappa}S}$ and $\rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}$, and is dented by$\mathcal{J}[\rho_{\sqrt{\kappa}S}, \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}]$. □

Now, we are ready to prove lemma 9.1.

Proof. of lemma 9.1. For simplicity of notation, we denote $\mathcal{J}_N(\sqrt{\kappa}\boldsymbol{\Lambda}^0, \tilde{\tilde{\boldsymbol{Y}}})$ by $\mathcal{J}_N$, and $\mathcal{J}[\rho_{\sqrt{\kappa}S}, \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}]$ by $\mathcal{J}$. By Jensen's inequality (note that the expectation is over the matrix $\tilde{\boldsymbol{Z}}$), we have

$$\big|\mathbb{E}[\mathcal{J}_N] - \mathcal{J}\big| \leq \mathbb{E}\big[|\mathcal{J}_N - \mathcal{J}|\big]. \tag{9.84}$$

Let $X_N \equiv \mathcal{J}_N - \mathcal{J}$. For $\epsilon > 0$ We can write

$$\begin{aligned}\mathbb{E}\big[|X_N|\big] &= \mathbb{E}\big[|X_N| \mathbb{I}\{|X_N| \leq \epsilon\}\big] + \mathbb{E}\big[|X_N| \mathbb{I}\{|X_N| > \epsilon\}\big] \\ &\leq \epsilon + \mathbb{E}\big[|X_N| \mathbb{I}\{|X_N| > \epsilon\}\big].\end{aligned} \tag{9.85}$$

By lemma 9.4, $|\mathcal{J}_N| \leq \frac{1}{2}\sqrt{\kappa}Cr^{(\tilde{\tilde{\boldsymbol{Y}}})}$, so the second term in (9.85) can be bounded as,

$$\mathbb{E}\big[|X_N|\,\mathbb{I}\{|X_N| > \epsilon\}\big] \leq \mathbb{E}\big[|W_N|\,\mathbb{I}\{|X_N| > \epsilon\}\big] \tag{9.86}$$

where

$$W_N = \max\Big\{\big|\mathcal{J} - \frac{1}{2}\sqrt{\kappa}Cr^{(\tilde{\tilde{\boldsymbol{Y}}})}\big|, \big|\mathcal{J} + \frac{1}{2}\sqrt{\kappa}Cr^{(\tilde{\tilde{\boldsymbol{Y}}})}\big|\Big\} = \frac{1}{2}\sqrt{\kappa}Cr^{(\tilde{\tilde{\boldsymbol{Y}}})} + \operatorname{sign}(\mathcal{J})\mathcal{J}.$$

For any positive constant t, we have

$$\begin{aligned}
&\mathbb{E}\big[|W_N|\,\mathbb{I}\{|X_N| > \epsilon\}\big] \\
&= \mathbb{E}\big[|W_N|\,\mathbb{I}\{|X_N| > \epsilon\}\,\mathbb{I}\{|W_N| \leq t\}\big] + \mathbb{E}\big[|W_N|\,\mathbb{I}\{|X_N| > \epsilon\}\,\mathbb{I}\{|W_N| > t\}\big] \\
&\leq \mathbb{E}\big[|W_N|\,\mathbb{I}\{|X_N| > \epsilon\}\mathbb{I}\{|W_N| \leq t\}\big] + \mathbb{E}\big[|W_N|\,\mathbb{I}\{|W_N| > t\}\big]
\end{aligned} \tag{9.87}$$

For the first term in (9.87) we can write

$$\begin{aligned}
\mathbb{E}\big[|W_N|\,\mathbb{I}\{|X_N| > \epsilon\}\mathbb{I}\{|W_N| \leq t\}\big] &\leq t\mathbb{E}\big[\mathbb{I}\{|X_N| > \epsilon\}\big] \\
&\leq t\,\mathbf{P}\big(|X_N| > \epsilon\big)
\end{aligned} \tag{9.88}$$

and the second term in (9.87) can be rewritten as

$$\mathbb{E}\big[|W_N|\,\mathbb{I}\{|W_N| > t\}\big] = \mathbb{E}\bigg[|W_N|\,\mathbb{I}\Big\{r^{(\tilde{\tilde{\boldsymbol{Y}}})} > \frac{2}{\sqrt{\kappa}C}\big(t - \operatorname{sign}(\mathcal{J})\mathcal{J}\big)\Big\}\bigg]. \tag{9.89}$$

From (9.86), (9.87), (9.88), we obtain

$$\begin{aligned}
\mathbb{E}\big[|X_N|\,\mathbb{I}\{|X_N| > \epsilon\}\big] \leq\; & t\,\mathbf{P}\big(|X_N| > \epsilon\big) \\
& + \mathbb{E}\bigg[|W_N|\,\mathbb{I}\Big\{r^{(\tilde{\tilde{\boldsymbol{Y}}})} > \frac{2}{\sqrt{\kappa}C}\big(t - \operatorname{sign}(\mathcal{J})\mathcal{J}\big)\Big\}\bigg].
\end{aligned} \tag{9.90}$$

Notice that W_N is a polynomial function of $r^{(\tilde{\tilde{\boldsymbol{Y}}})}$, so by lemma 9.6, vanishes as $N \to \infty$ for sufficiently large constant t. By lemma 9.7, $\mathbf{P}\big(|X_N| > \epsilon\big) \xrightarrow{N\to\infty} 0$. For a fixed $t > 0$, the first term in (9.90) goes to 0 in the limit $N \to \infty$. Therefore, taking the limit of both sides in (9.85), for any $\epsilon > 0$, we find:

$$\lim_{N\to\infty} \mathbb{E}\big[|X_N|\big] \leq \epsilon. \tag{9.91}$$

From which, by (9.84), we deduce that $\lim_{N\to\infty} \mathbb{E}[\mathcal{J}_N] = \mathcal{J}$. □

9.C.2 Proof of proposition 9.5

Consider two matrices with the same eigenvectors, $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\intercal$ and $\tilde{\boldsymbol{S}} = \boldsymbol{U}\tilde{\boldsymbol{\Lambda}}\boldsymbol{U}^\intercal$, where $\boldsymbol{U}$ is a Haar orthogonal matrix, and $\boldsymbol{\lambda}$, $\tilde{\boldsymbol{\lambda}}$ are distributed

according to $P_N^{(1)}(\boldsymbol{\lambda}), P_N^{(2)}(\tilde{\boldsymbol{\lambda}})$, respectively. For two such matrices, we write $(\boldsymbol{S}, \tilde{\boldsymbol{S}}) \sim Q_N(\boldsymbol{U}, \boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}})$, where $Q_N(\boldsymbol{U}, \boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}})$ is the joint p.d.f. of $\boldsymbol{U}, \boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}$,

$$dQ_N(\boldsymbol{U}, \boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}) = d\mu_N(\boldsymbol{U})\ P_N^{(1)}(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}\ P_N^{(2)}(\tilde{\boldsymbol{\lambda}})\, d\tilde{\boldsymbol{\lambda}}.$$

For $t \in [0,1]$ an interpolating parameter, consider the following observation model:

$$\begin{cases} \boldsymbol{Y}_1^{(t)} = \sqrt{\kappa t}\boldsymbol{S} + \boldsymbol{Z}_1 \\ \boldsymbol{Y}_2^{(t)} = \sqrt{\kappa(1-t)}\tilde{\boldsymbol{S}} + \boldsymbol{Z}_2 \end{cases} \tag{9.92}$$

where $\boldsymbol{Z}_1, \boldsymbol{Z}_2$ are Wigner matrices independent of each other, and $(\boldsymbol{S}, \tilde{\boldsymbol{S}}) \sim Q_N(\boldsymbol{U}, \boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}})$. The free energy for this model can be written as :

$$\begin{aligned} F_N(t) &= -\frac{1}{N^2}\mathbb{E}_{\boldsymbol{Y}_1^{(t)}, \boldsymbol{Y}_2^{(t)}}\Bigg[\ln \int dQ_N(\boldsymbol{U}, \boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}) \\ &\qquad \times e^{\frac{N}{2}\operatorname{Tr}[\sqrt{\kappa t}\boldsymbol{X}\boldsymbol{Y}_1^{(t)} - \frac{\kappa t}{2}\boldsymbol{X}^2 + \sqrt{\kappa(1-t)}\tilde{\boldsymbol{X}}\boldsymbol{Y}_2^{(t)} - \frac{\kappa(1-t)}{2}\tilde{\boldsymbol{X}}^2]}\Bigg] \\ &= -\frac{1}{N^2}\mathbb{E}_{\boldsymbol{Y}_1^{(t)}, \boldsymbol{Y}_2^{(t)}}\Bigg[\ln \int dQ_N(\boldsymbol{U}, \boldsymbol{\lambda}, \tilde{\boldsymbol{\lambda}}) \\ &\qquad \times e^{\frac{N}{2}\operatorname{Tr}[\kappa t\boldsymbol{X}\boldsymbol{S} + \sqrt{\kappa t}\boldsymbol{X}\boldsymbol{Z}_1 - \frac{\kappa t}{2}\boldsymbol{X}^2 + \kappa(1-t)\tilde{\boldsymbol{X}}\tilde{\boldsymbol{S}} + \sqrt{\kappa(1-t)}\tilde{\boldsymbol{X}}\boldsymbol{Z}_2 - \frac{\kappa(1-t)}{2}\tilde{\boldsymbol{X}}^2]}\Bigg] \end{aligned}$$

where $\boldsymbol{X}, \tilde{\boldsymbol{X}}$ has the same eigenspace, $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\intercal$, $\tilde{\boldsymbol{X}} = \boldsymbol{U}\tilde{\boldsymbol{\Lambda}}\boldsymbol{U}^\intercal$. Note that, for $t = 0$ the only term depending on $\boldsymbol{\lambda}$ (in both the inner and outer expectation) is the pdf $P_N^{(1)}(\boldsymbol{\lambda})$ and we can integrate over $\boldsymbol{\lambda}$ in both of the expectations, to get $F_N(0) = F_N^{(2)}(\kappa)$. Similarly, we have $F_N(1) = F_N^{(1)}(\kappa)$.

Taking the derivative w.r.t. t, we get:

$$\begin{aligned} \frac{d}{dt}F_N(t) = -\frac{1}{N}\mathbb{E}\Big[&\frac{\kappa}{2}\operatorname{Tr}\langle \boldsymbol{X}\boldsymbol{S}\rangle_t + \frac{1}{4}\sqrt{\frac{\kappa}{t}}\operatorname{Tr}\boldsymbol{Z}_1\langle \boldsymbol{X}\rangle_t - \frac{\kappa}{4}\operatorname{Tr}\langle \boldsymbol{X}^2\rangle_t \\ &- \frac{\kappa}{2}\operatorname{Tr}\langle \tilde{\boldsymbol{X}}\tilde{\boldsymbol{S}}\rangle_t - \frac{1}{4}\sqrt{\frac{\kappa}{1-t}}\operatorname{Tr}\boldsymbol{Z}_2\langle \tilde{\boldsymbol{X}}\rangle_t + \frac{\kappa}{4}\operatorname{Tr}\langle \tilde{\boldsymbol{X}}^2\rangle_t\Big] \end{aligned}$$

where $\langle . \rangle_t$ denotes the expectation with respect to the posterior distribution of the model (9.92). By integration by parts, we have

$$\begin{aligned} \mathbb{E}\big[\operatorname{Tr}\boldsymbol{Z}_1\langle \boldsymbol{X}\rangle_t\big] &= \sqrt{\kappa t}\mathbb{E}\Big[\operatorname{Tr}\langle \boldsymbol{X}^2\rangle_t - \operatorname{Tr}\langle \boldsymbol{X}\rangle_t^2\Big], \\ \mathbb{E}\big[\operatorname{Tr}\boldsymbol{Z}_2\langle \tilde{\boldsymbol{X}}\rangle_t\big] &= \sqrt{\kappa(1-t)}\mathbb{E}\Big[\operatorname{Tr}\langle \tilde{\boldsymbol{X}}^2\rangle_t - \operatorname{Tr}\langle \tilde{\boldsymbol{X}}\rangle_t^2\Big] \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{d}{dt}F_N(t) &= -\frac{1}{N}\frac{\kappa}{4}\mathbb{E}\Big[2\operatorname{Tr}\langle \boldsymbol{X}\boldsymbol{S}\rangle_t - \operatorname{Tr}\langle \boldsymbol{X}\rangle_t^2 - 2\operatorname{Tr}\langle \tilde{\boldsymbol{X}}\tilde{\boldsymbol{S}}\rangle_t + \operatorname{Tr}\langle \tilde{\boldsymbol{X}}\rangle_t^2\Big] \\ &= \frac{1}{N}\frac{\kappa}{4}\mathbb{E}\big[\operatorname{Tr}[\langle \boldsymbol{X}\boldsymbol{S}\rangle_t - \langle \tilde{\boldsymbol{X}}\tilde{\boldsymbol{S}}\rangle_t]\big] \quad \text{(By a Nishimori identity).} \end{aligned}$$

We have

$$\begin{aligned}
\frac{4N}{\kappa}\Big|\frac{d}{dt}F_N(t)\Big| &= \left|\mathbb{E}\left[\Big\langle \operatorname{Tr}\left[\boldsymbol{S}(\boldsymbol{X}-\tilde{\boldsymbol{X}})-(\tilde{\boldsymbol{S}}-\boldsymbol{S})\tilde{\boldsymbol{X}}\right]\Big\rangle_t\right]\right| \\
&\le \mathbb{E}\left[\left\langle\left|\operatorname{Tr}\left[\boldsymbol{S}(\boldsymbol{X}-\tilde{\boldsymbol{X}})-(\tilde{\boldsymbol{S}}-\boldsymbol{S})\tilde{\boldsymbol{X}}\right]\right|\right\rangle_t\right] \quad \text{(By Jensen)} \\
&\le \mathbb{E}\left[\left\langle\left|\operatorname{Tr}\boldsymbol{S}(\boldsymbol{X}-\tilde{\boldsymbol{X}})\right|\right\rangle_t\right]+\mathbb{E}\left[\left\langle\left|\operatorname{Tr}(\tilde{\boldsymbol{S}}-\boldsymbol{S})\tilde{\boldsymbol{X}}\right|\right\rangle_t\right] \\
&\le \mathbb{E}\Big[\|\boldsymbol{S}\|_F\langle\|\boldsymbol{X}-\tilde{\boldsymbol{X}}\|_F\rangle_t\Big]+\mathbb{E}\Big[\|\boldsymbol{S}-\tilde{\boldsymbol{S}}\|_F\langle\|\tilde{\boldsymbol{X}}\|_F\rangle_t\Big] \\
&\le \sqrt{\mathbb{E}\Big[\|\boldsymbol{S}\|_F^2\Big]\mathbb{E}\Big[\big\langle\|\boldsymbol{X}-\tilde{\boldsymbol{X}}\|_F\big\rangle_t^2\Big]} \quad \text{(By Cauchy–Schwarz)} \\
&\qquad + \sqrt{\mathbb{E}\Big[\|\boldsymbol{S}-\tilde{\boldsymbol{S}}\|_F^2\Big]\mathbb{E}\Big[\big\langle\|\tilde{\boldsymbol{X}}\|_F\big\rangle_t^2\Big]} \\
&\le \sqrt{\mathbb{E}\big[\|\boldsymbol{S}\|_F^2\big]\mathbb{E}\Big[\big\langle\|\boldsymbol{X}-\tilde{\boldsymbol{X}}\|_F^2\big\rangle_t\Big]} \quad \text{(By Cauchy–Schwarz)} \\
&\qquad + \sqrt{\mathbb{E}\big[\|\boldsymbol{S}-\tilde{\boldsymbol{S}}\|_F^2\big]\mathbb{E}\Big[\big\langle\|\tilde{\boldsymbol{X}}\|_F^2\big\rangle_t\Big]} \\
&= \sqrt{\mathbb{E}\big[\|\boldsymbol{S}\|_F^2\big]\mathbb{E}\big[\|\boldsymbol{S}-\tilde{\boldsymbol{S}}\|_F^2\big]} \quad \text{(By Nishimori)} \\
&\qquad + \sqrt{\mathbb{E}\big[\|\boldsymbol{S}-\tilde{\boldsymbol{S}}\|_F^2\big]\mathbb{E}\big[\|\tilde{\boldsymbol{S}}\|_F^2\big]} \\
&= \Big(\sqrt{\mathbb{E}\big[\|\boldsymbol{S}\|_F^2\big]}+\sqrt{\mathbb{E}\big[\|\tilde{\boldsymbol{S}}\|_F^2\big]}\Big)\sqrt{\mathbb{E}\big[\|\boldsymbol{S}-\tilde{\boldsymbol{S}}\|_F^2\big]} \\
&= \Big(\sqrt{\mathbb{E}_{\boldsymbol{\lambda}}\big[\|\boldsymbol{\lambda}\|^2\big]}+\sqrt{\mathbb{E}_{\tilde{\boldsymbol{\lambda}}}\big[\|\tilde{\boldsymbol{\lambda}}\|^2\big]}\Big)\sqrt{\mathbb{E}_{\boldsymbol{\lambda},\tilde{\boldsymbol{\lambda}}}\big[\|\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}}\|^2\big]}.
\end{aligned}$$

We obtain the result by integrating over t from 0 to 1. □

9.C.3 Proof of technical lemmas

Lemma 9.8. *Given two vectors $\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^N$, denote their empirical distributions by μ,ν respectively. We have*

$$W_2(\mu,\nu)=\sqrt{\min_{\pi\in\mathcal{S}_N}\frac{1}{N}\|\boldsymbol{u}-\boldsymbol{v}_\pi\|^2}$$

where $\boldsymbol{v}_\pi$ is a permutation of $\boldsymbol{v}$.

Proof. By definition we have

$$W_2(\mu,\nu)^2=\inf_{\kappa\in\kappa(\mu,\nu)}\mathbb{E}_{\kappa(x,y)}\big[(x-y)^2\big].$$

Any measure in $\kappa(\mu,\nu)$ can be represented by a doubly stochastic $N\times N$ matrix. Thus, we have

$$W_2(\mu,\nu)^2 = \inf_{\boldsymbol{P}\in\mathcal{B}_N} \frac{1}{N}\sum_{i,j} P_{ij}(u_i - v_j)^2$$

where $\mathcal{B}_N$ denotes the set of doubly stochastic matrices. This minimization problem is a linear optimization problem on the bounded convex set $\mathcal{B}_N$. By Choquet's theorem, the solutions to this problem exist and are the extremal points of $\mathcal{B}_N$, which are permutation matrices (by Birkhoff's theorem). Therefore, the minimization can be written on the set of permutation matrices to get:

$$W_2(\mu,\nu)^2 = \min_{\pi\in\mathcal{S}_N} \frac{1}{N}\sum_{i,j}(u_i - v_{\pi(i)})^2.$$

□

Lemma 9.9. *Suppose $\boldsymbol{\lambda}\in\mathbb{R}^N$ is distributed according to $P_{S,N}(\boldsymbol{\lambda})$, and $\boldsymbol{\lambda}^0$ is generated with i.i.d. elements from ρ_S. Let $\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0}$ be their empirical distribution. We have:*

$$\lim_{N\to\infty} \mathbb{E}_{\boldsymbol{\lambda}}\big[W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0})^2\big] = 0.$$

Proof. By the triangle inequality

$$W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0}) \le W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \rho_S) + W_2(\hat{\mu}_{\boldsymbol{\lambda}^0}, \rho_S). \tag{9.93}$$

The first term approaches 0 as $N\to\infty$ almost surely, by remark 9.2. By lemma 9.10, the second term also converges 0 as $N\to\infty$. Therefore, we have $W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0})\to 0$ almost surely. Consequently, we have that $W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0})^2\to 0$ almost surely. Denote $W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0})^2$ by X_N which is a non-negative random variable. We have:

$$\begin{aligned}\mathbb{E}[X_N] &= \mathbb{E}\big[X_N\,\mathbb{I}\{X_N\le\epsilon\}\big] + \mathbb{E}\big[X_N\,\mathbb{I}\{X_N>\epsilon\}\big]\\ &\le \epsilon + \mathbb{E}\big[X_N\,\mathbb{I}\{X_N>\epsilon\}\big].\end{aligned} \tag{9.94}$$

By definition, one can see that $W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0})^2 \le 2(m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}} + m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}^0}})$, where $m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}} = \frac{1}{N}\sum\lambda_i^2$ and $m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}^0}} = \frac{1}{N}\sum\lambda_i^{0\,2}$. For the second term in (9.94) we have

$$\begin{aligned}\mathbb{E}\big[X_N\,\mathbb{I}\{X_N>\epsilon\}\big] &\le 2\mathbb{E}\big[(m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}} + m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}^0}})\,\mathbb{I}\{X_N>\epsilon\}\big]\\ &= 2m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}^0}}\mathbb{E}\big[\mathbb{I}\{X_N>\epsilon\}\big] + 2\mathbb{E}\big[m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}}\,\mathbb{I}\{X_N>\epsilon\}\big]\\ &= 2m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}^0}}\mathbf{P}[X_N>\epsilon] + 2\mathbb{E}\big[m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}}\,\mathbb{I}\{X_N>\epsilon\}\big].\end{aligned} \tag{9.95}$$

For the last term in (9.95) is decomposed as

$$\begin{aligned}&\mathbb{E}\big[m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}}\,\mathbb{I}\{X_N>\epsilon\}\big]\\ &\quad= \mathbb{E}\big[m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}}\,\mathbb{I}\{X_N>\epsilon\}\,\mathbb{I}\{m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}}\le t\}\big] + \mathbb{E}\big[m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}}\,\mathbb{I}\{X_N>\epsilon\}\,\mathbb{I}\{m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}}>t\}\big]\\ &\quad\le t\mathbf{P}[X_N>\epsilon] + \mathbb{E}\big[m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}}\,\mathbb{I}\{m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}}>t\}\big]\end{aligned} \tag{9.96}$$

where t is a fixed sufficiently large constant such that $t > C$. From (9.94), (9.95), (9.96), we get:

$$\mathbb{E}[X_N] \leq \epsilon + 2m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}^0}} \mathbf{P}[X_N > \epsilon] + 2t\mathbf{P}[X_N > \epsilon] + 2\mathbb{E}\big[m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}} \mathbb{I}\{m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}} > t\}\big]. \quad (9.97)$$

Since $W_2(\hat{\mu}_{\boldsymbol{\lambda}}, \hat{\mu}_{\boldsymbol{\lambda}^0})^2 \to 0$ almost surely, $\mathbf{P}[X_N > \epsilon]$ approaches 0 as $N \to \infty$. By construction $m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}^0}}$ is bounded (by the constant C), so the second and the third terms approaches 0 as $N \to \infty$. The last term also converges 0, by lemma 9.11. Therefore, for arbitrary $\epsilon > 0$ we find

$$\lim_{N\to\infty} \mathbb{E}[X_N] \leq \epsilon.$$

□

Lemma 9.10. *Let $X_1, \ldots, X_N$ be i.i.d. random variables distributed according to distribution μ, which has finite support. Let μ_N denote their empirical distribution. Then*

$$\lim_{N\to\infty} W_2(\mu_N, \mu) = 0 \quad \textit{almost surely.}$$

Proof. By the law of large numbers, $\mu_N \to \mu$ almost surely. Moreover, since μ has bounded support, the second moment of μ_N converges to the one of μ. By Theorem 7.12 in [119], we have the convergence in the Wasserstein-2 metric almost surely. □

Lemma 9.11. *Under assumption 1.B, for t large enough, we have:*

$$\lim_{N\to\infty} \mathbb{E}\big[m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}} \mathbb{I}\{m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}} > t\}\big] = 0.$$

Proof. Boundedness and a.s. convergence of $m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}}$ imply that $\lim_{N\to\infty} \mathbb{E}[m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}}] = m^{(2)}_{\rho_S}$, which is bounded since ρ_S has compact support. Denote

$$X_N = m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}} \mathbb{I}\{m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}} \leq t\}$$

Then, $X_N \to m^{(2)}_{\rho_S}$ a.s. and by bounded convergence we also have that $\lim_{N\to\infty} \mathbb{E}[X_N] = m^{(2)}_{\rho_S}$. Therefore,

$$\lim_{N\to\infty} \mathbb{E}\big[m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}} \mathbb{I}\{m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}} > t\}\big] = \lim_{N\to\infty} \mathbb{E}\big[m^{(2)}_{\hat{\mu}_{\boldsymbol{\lambda}}} - X_N\big] = 0.$$

□

9.D Derivation of the limiting spectral distribution for the model 9.B.1

Suppose we want to find the density $\mu(x)$, which is the free convolution of the density $\rho(x)$ with the semi-circle density $\rho_{\text{sc}}(x)$. This density is given in [118]

by

$$\begin{aligned}
&\mu\big(\psi(u)\big) = \frac{v(u)}{\pi} \\
&\psi(u) = u + \int_{\mathbb{R}} \frac{u-x}{(u-x)^2+v(u)^2}\rho(x)\,dx, \\
&v(u) = \inf\Big\{w \geq 0 \Big| \int_{\mathbb{R}} \frac{\rho(x)}{(u-x)^2+w^2}\,dx \leq 1\Big\}.
\end{aligned} \tag{9.98}$$

This result will be used repeatedly.

To compute the density $\rho_Y = \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}$ where $\rho_{\sqrt{\kappa}S}(x) = \frac{1}{2}\delta(x+\sqrt{\kappa}) + \frac{1}{2}\delta(x-\sqrt{\kappa})$, we compute functions $v(u)$ and $\psi(u)$

$$\begin{aligned}
&\text{if } \kappa < 1: \\
&v(u) = \begin{cases} \frac{1}{\sqrt{2}}\sqrt{1-2(u^2+\kappa)+\sqrt{1+16\kappa u^2}} & \text{if } |u| \leq \frac{1}{\sqrt{2}}\sqrt{1+2\kappa+\sqrt{1+8\kappa}} \\ 0 & \text{else.} \end{cases} \\
&\psi(u) = \begin{cases} \frac{1+8u^2-\sqrt{1+16\kappa u^2}}{4u} & \text{if } |u| \leq \frac{1}{\sqrt{2}}\sqrt{1+2\kappa+\sqrt{1+8\kappa}} \\ \frac{u(u^2-\kappa+1)}{u^2-\kappa} & \text{else.} \end{cases}
\end{aligned} \tag{9.99}$$

$$\begin{aligned}
&\text{if } \kappa \geq 1: \\
&v(u) = \begin{cases} \frac{1}{\sqrt{2}}\sqrt{1-2(u^2+\kappa)+\sqrt{1+16\kappa u^2}} & \text{if } u \text{ satisfies (9.101)} \\ 0 & \text{else.} \end{cases} \\
&\psi(u) = \begin{cases} \frac{1+8u^2-\sqrt{1+16\kappa u^2}}{4u} & \text{if } u \text{ satisfies (9.101)} \\ \frac{u(u^2-\kappa+1)}{u^2-\kappa} & \text{else.} \end{cases}
\end{aligned} \tag{9.100}$$

with the condition:

$$\frac{1}{\sqrt{2}}\sqrt{1+2\kappa-\sqrt{1+8\kappa}} \leq |u| \leq \frac{1}{\sqrt{2}}\sqrt{1+2\kappa+\sqrt{1+8\kappa}} \tag{9.101}$$

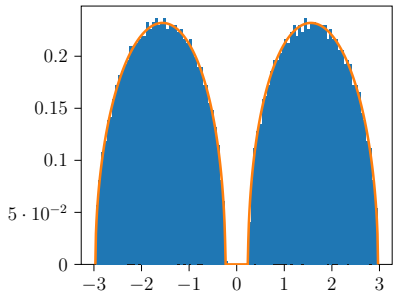
Solving the equation $\rho_Y\big(\psi(u)\big) = \frac{v(u)}{\pi}$, we find:

$$\text{if } \kappa < 1: \quad \rho_Y(x) = \begin{cases} \frac{1}{\sqrt{2\pi}}\sqrt{1-2\big(\kappa+\frac{1}{2304}A^2\big)+\sqrt{1+\frac{\kappa}{144}A^2}} & \text{if } |x| \leq U(\kappa) \\ 0 & \text{else.} \end{cases}$$

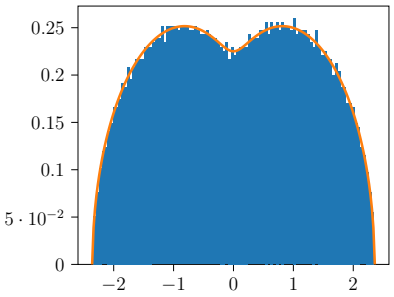
$$\begin{aligned}
&\text{if } \kappa \geq 1: \\
&\rho_Y(x) = \begin{cases} \frac{1}{\sqrt{2\pi}}\sqrt{1-2\big(\kappa+\frac{1}{2304}A^2\big)+\sqrt{1+\frac{\kappa}{144}A^2}} & \text{if } L(\kappa) \leq |x| \leq U(\kappa) \\ 0 & \text{else.} \end{cases}
\end{aligned} \tag{9.102}$$

where

$$A = 16x + \frac{32 \times 2^{1/3}(-3+3\kappa+x^2)}{B} + 2^{2/3}B$$



(a) $\kappa = 2$



(b) $\kappa = 0.5$

Figure 9.D.1: The continuous line is the asymptotic spectral density of the observation matrix $\boldsymbol{Y}$ in example 9.B.1. $\rho_Y(x) = \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}$ where $\rho_{\sqrt{\kappa}S}(x) = \frac{1}{2}\delta(x + \sqrt{\kappa}) + \frac{1}{2}\delta(x - \sqrt{\kappa})$ for $\kappa = 2$ and $\kappa = 0.5$. Compared to the histogram of a realization of size $N = 5000$.

$$B = \sqrt[3]{576x + 1152\kappa x - 128x^3 + 64\sqrt{x^2(9 + 18\kappa - 2x^2)^2 - 4(-3 + 3\kappa + x^2)^3}}$$

$$L(\kappa) = \frac{\big(-3 + \sqrt{1 + 8\kappa}\big)\sqrt{1 + 2\kappa - \sqrt{1 + 8\kappa}}}{\sqrt{2}\big(-1 + \sqrt{1 + 8\kappa}\big)}$$

$$U(\kappa) = \frac{\big(3 + \sqrt{1 + 8\kappa}\big)\sqrt{1 + 2\kappa + \sqrt{1 + 8\kappa}}}{\sqrt{2}\big(1 + \sqrt{1 + 8\kappa}\big)}$$

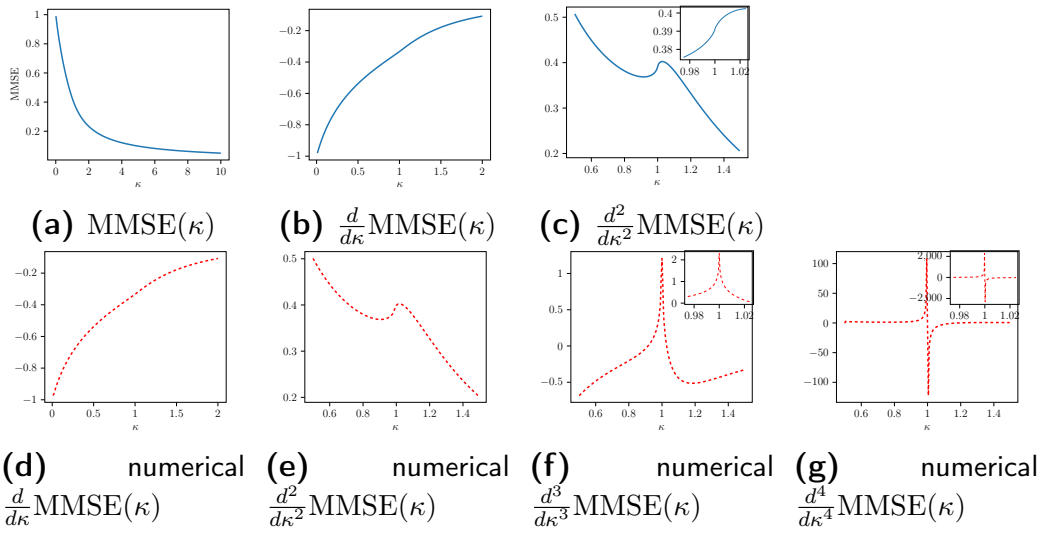
From (9.102), one can see that the support of ρ_Y is constituted of two disjoint intervals for $\kappa \geq 1$, and of one single interval for $\kappa < 1$.

Once we have the expression (9.102), we get from Theorem 9.2 an explicit integral representation for MMSE(κ), as well as for the derivatives. We show the details here for $\kappa \geq 1$. Since in this example $\rho_Y(x)$ is symmetric we have for $\kappa > 1$

$$\mathrm{MMSE}(\kappa) = \frac{1}{\kappa}\Big(1 - \frac{8\pi^2}{3}\int_{L(\kappa)}^{U(\kappa)} \rho_Y^3(x)\,dx\Big). \tag{9.103}$$

By the Leibniz integral rule (all derivatives are w.r.t κ) we have using that ρ_Y vanishes at the end-points of the interval $[L(\kappa), U(\kappa)]$

$$\begin{aligned}\mathrm{MMSE}'(\kappa) &= -\frac{1}{\kappa^2} + \frac{1}{\kappa^2}\frac{8\pi^2}{3}\int_{L(\kappa)}^{U(\kappa)} \rho_Y^3(x)\,dx \\ &\quad - \frac{1}{\kappa}\frac{8\pi^2}{3}\Big(\rho_Y^3\big(U(\kappa)\big)U'(\kappa) - \rho_Y^3\big(L(\kappa)\big)L'(\kappa) \\ &\quad + \int_{L(\kappa)}^{U(\kappa)} 3\rho_Y^2(x)\rho_Y'(x)\,dx\Big) \\ &= -\frac{1}{\kappa^2} + \frac{1}{\kappa^2}\frac{8\pi^2}{3}\int_{L(\kappa)}^{U(\kappa)} \rho_Y^3(x)\,dx - \frac{8\pi^2}{\kappa}\int_{L(\kappa)}^{U(\kappa)} \rho_Y^2(x)\rho_Y'(x)\,dx.\end{aligned} \tag{9.104}$$



(a) MMSE(κ) **(b)** $\frac{d}{d\kappa}$MMSE(κ) **(c)** $\frac{d^2}{d\kappa^2}$MMSE(κ)

(d) numerical $\frac{d}{d\kappa}$MMSE(κ) **(e)** numerical $\frac{d^2}{d\kappa^2}$MMSE(κ) **(f)** numerical $\frac{d^3}{d\kappa^3}$MMSE(κ) **(g)** numerical $\frac{d^4}{d\kappa^4}$MMSE(κ)

Figure 9.D.2: Analysis of the MMSE in example 9.B.1. In (a), MMSE is plotted from (9.103) for $0 < \kappa \leq 10$ with step-size $h = 0.01$. The numerical first derivative of the curve in (a) is illustrated in (d), which is computed using five-point stencil [127] $f'(x) \approx \frac{-f(x+2h)+8f(x+h)-8f(x-h)+f(x-2h)}{12h}$. In (b), the first derivative is plotted from (9.104) with step-size $h = 0.01$, and its numerical first derivative is plotted in (e). The second derivative of MMSE, computed from (9.105), is depicted in (c) with step-size $h = 0.005$. The inset plot is with step-size $h = 0.00025$. The third derivative of MMSE in (f) is obtained from the numerical differentiation of the curve in (c). The fourth derivative is computed using the five-point stencil $f''(x) \approx \frac{-f(x+2h)+16f(x+h)-30f(x)+16f(x-h)-f(x-2h)}{12h^2}$ from the curve in (c).

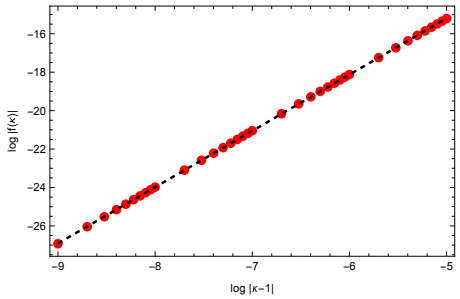
Moreover, $\rho_Y^2(x)\rho_Y'(x)$ can also be checked to vanishes at the end-points of the interval $[L(\kappa), U(\kappa)]$ so

$$\begin{aligned}\mathrm{MMSE}''(\kappa) = \frac{2}{\kappa^3} - \frac{1}{\kappa^3}\frac{16\pi^2}{3}\int_{L(\kappa)}^{U(\kappa)} \rho_Y^3(x)\,dx + \frac{16\pi^2}{\kappa^2}\int_{L(\kappa)}^{U(\kappa)} \rho_Y^2(x)\rho_Y'(x)\,dx \\ - \frac{8\pi^2}{\kappa}\int_{L(\kappa)}^{U(\kappa)} \left[2\rho_Y(x)\big(\rho_Y'(x)\big)^2 + \rho_Y^2(x)\rho_Y''(x)\right] dx.\end{aligned} \tag{9.105}$$

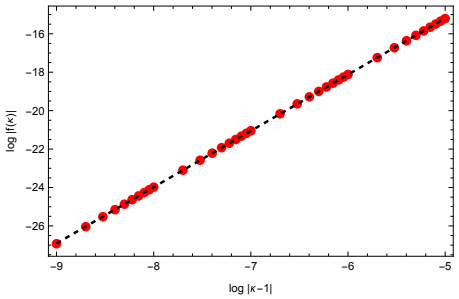
A similar and somewhat simpler calculation also provides integral representations for $\kappa < 1$.

It is not clear how to compute these integrals analytically but precise results can be obtained from numerical integration. Integral represenations of higher derivatives can also be obtained in principle but become unwieldy. In fact the numerical integration of the formula for the second derivative is precise enough to get a good numerical calculation of the third and fourth derivatives. All numerical results are summearized in figure 9.D.2.

Plots (f),(g) in figure 9.D.2 suggest the existence of a third-order phase transition at $\kappa_c = 1$. Note that this is the point where the support of ρ_Y



(a) $\log|f(\kappa)|$ for $\kappa \to 1^-$. The fitted line has slope $\alpha \approx 2.929$, and $\log|c| \approx -0.5511$.



(b) $\log|f(\kappa)|$ for $\kappa \to 1^+$. The fitted line has slope $\alpha \approx 2.929$, and $\log|c| \approx -0.5511$.

Figure 9.D.3: $\log|f(\kappa)|$ as a function of $\log|\kappa - 1|$

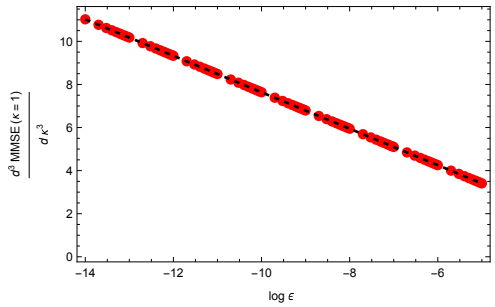


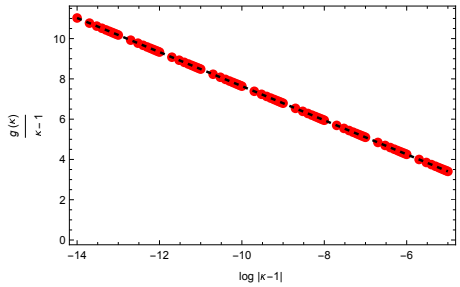
Figure 9.D.4: Numerical third derivative of MMSE at $\kappa = 1$ computed from $\frac{\mathrm{MMSE}''(1+\epsilon)-\mathrm{MMSE}''(1-\epsilon)}{2\epsilon}$ as a function of $\log\epsilon$. A linear function $-0.8463\log\epsilon - 0.8249$ is fitted to the points.

transitions from a single to two intervals. Since the singularity seems to appear in teh third derivative we define the function

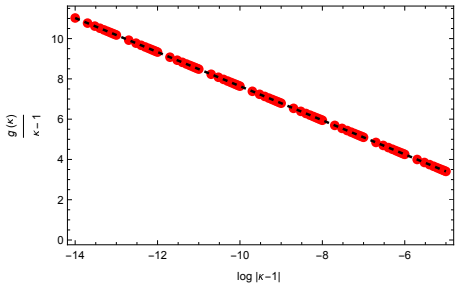
$$f(\kappa) = \mathrm{MMSE}(\kappa) - \mathrm{MMSE}(1) + \mathrm{MMSE}'(1)(\kappa-1) + \frac{1}{2}\mathrm{MMSE}''(1)(\kappa-1)^2.$$

If one tries the ansatz $f(\kappa) = c|\kappa-1|^\alpha$ with $2 < \alpha \leq 3$, or in other words $\log|f(\kappa)| \approx \log|c| + \alpha\log|\kappa-1|$ we find $\alpha \approx 2.929$. This is shown in figure 9.D.3 where $f(\kappa)$ is plotted on a log-log scale on both sides of $\kappa_c = 1^\pm$. However, the appearance of this exponent is not consistent with the fact that the expression for ρ_Y (9.102) is fully algebraic and an excat integration could only give an integer exponent or a logarithmic singularity. To further investigate the behavior of the MMSE, we study the third numerical derivative obtained from the curve of $\mathrm{MMSE}''(\kappa)$ using the relation $\frac{d^3}{d\kappa^3}\mathrm{MMSE}(1) \approx \frac{\mathrm{MMSE}''(1+\epsilon)-\mathrm{MMSE}''(1-\epsilon)}{2\epsilon}$. As plotted in figure 9.D.4, $\frac{d^3}{d\kappa^3}\mathrm{MMSE}(1)$ diverges linearly as ϵ decays exponentially. This suggests that the behavior of the correction term for the second derivative is of the form $a(\kappa-1)\big(\log|\kappa-1| + b\big)$. Define the function

$$g(\kappa) = \mathrm{MMSE}''(\kappa) - \mathrm{MMSE}''(1) \tag{9.106}$$



(a) $\frac{g(\kappa)}{\kappa-1}$ for $\kappa \to 1^-$. The fitted line has slope $a \approx -0.8463$, and $ab \approx -0.8249$.



(b) $\frac{g(\kappa)}{\kappa-1}$ for $\kappa \to 1^+$. The fitted line has slope $a \approx -0.8463$, and $ab \approx -0.8249$.

Figure 9.D.5: $\frac{g(\kappa)}{\kappa-1}$ as a function of $\log|\kappa-1|$

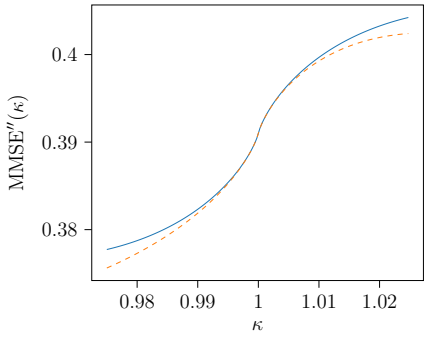


Figure 9.D.6: Comparison of the second derivative of the $\mathrm{MMSE}(\kappa)$ and the expansion (9.108) (plotted with dashed line).

for κ close to 1, we have:

$$\frac{g(\kappa)}{\kappa-1} \approx a \log|\kappa-1| + ab \tag{9.107}$$

From the plots in figure 9.D.5, we deduce that $a \approx -0.8463$ and $b \approx 0.9746$. Therefore, the $\mathrm{MMSE}''(\kappa)$ can be described by the following expansion close to the point $\kappa = 1$

$$\mathrm{MMSE}''(\kappa) = \mathrm{MMSE}''(1) + a(\kappa-1)\big(\log|\kappa-1| + b\big) + o\big((\kappa-1)\big). \tag{9.108}$$

From this expansion, we conjecture that the MMSE has a third-order phase transition at $\kappa = 1$.

9.E Derivation of the limiting spectral distribution for the model 9.B.2

We indicate the main steps to compute the density $\rho_Y = \rho_{\sqrt{\kappa}S} \boxplus \rho_{SC}$ where $\rho_{\sqrt{\kappa}S}(x) = p\delta(x) + (1-p)\delta(x-\sqrt{\kappa})$. Following the same procedure as in the

previous example, functions $v(u)$ and $\psi(u)$ can be derived as :

$$v(u) = \begin{cases} \frac{1}{2}\Big[-2u^2 + 2\sqrt{\kappa}u - \kappa + 1 \\ \quad + \sqrt{\sqrt{\kappa}\big(\sqrt{\kappa} - 2u\big)\big(-2\sqrt{\kappa}u + \kappa + 4p - 2\big) + 1}\Big]^{\frac{1}{2}} & \text{if } u \in \mathrm{Supp}(v) \\ 0 & \text{else.} \end{cases}$$

$$\psi(u) = \begin{cases} \frac{-8u^2 + 6\sqrt{\kappa}u + \sqrt{\sqrt{\kappa}\big(\sqrt{\kappa}-2u\big)\big(-2\sqrt{\kappa}u+\kappa+4p-2\big)+1} - \kappa - 1}{2\big(\sqrt{\kappa}-2u\big)} & \text{if } u \in \mathrm{Supp}(v) \\ u + \frac{p}{u} + \frac{1-p}{u-\sqrt{\kappa}} & \text{else.} \end{cases}$$

where

$$\mathrm{Supp}(v) = \Big\{ u \big| g(u) < 0 \Big\}$$

and

$$g(u) = u^4 - 2\sqrt{\kappa}u^3 + (\kappa - 1)u^2 + 2p\sqrt{\kappa}u - p\kappa. \tag{9.109}$$

Solving the equation $\rho_Y\big(\psi(u)\big) = \frac{v(u)}{\pi}$, we find the analytical expression for $\rho_Y(x)$ which we omit here.

The set $\mathrm{Supp}(v)$ determines the support of ρ_Y. For a given $0 < p < 1$ the degree four polynomial $g(u)$ has either two or four real roots, depending on κ. The former case corresponds to the situation where the support of ρ_Y is a single interval, and the latter corresponds to the case where the support of ρ_Y is a union of two intervals. Using Theorem 3.7 in [128], a critical value κ_c is found such that for $\kappa < \kappa_c$ the polynomial $g(u)$ has two real roots and for $\kappa > \kappa_c$ it has four real roots. We have

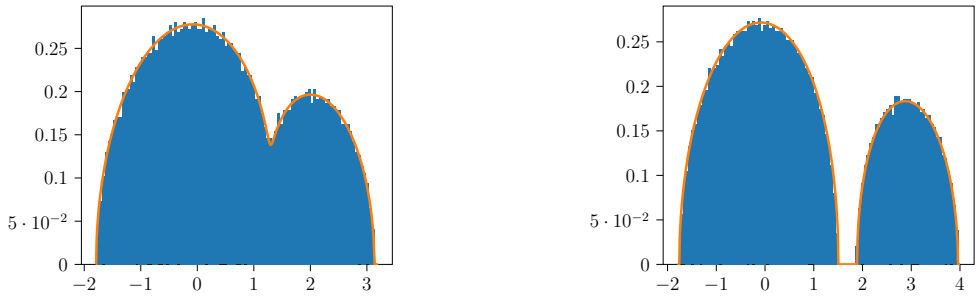
$$\kappa_c = 1 + 3\sqrt[3]{p^2(1-p)} + 3\sqrt[3]{p(1-p)^2}. \tag{9.110}$$

An example is illutrated on figure 9.E.1. But contrary to the previous example we have not identified any singularity in $\mathrm{MMSE}(\kappa)$ due to the merging of the two intervals.

9.F Derivation of the limiting spectral distribution for the model 9.B.3

In this example, we find the limiting spectral measure $\rho_Y = \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{sc}}$ directly using the free additive convolution formula. The limiting spectral measure of $\boldsymbol{S}$ is the Marchenko-Pastur law rescaled by factor q, which we denote by ρ_{MP}. The R-transform is $\mathcal{R}_{\rho_{\mathrm{MP}}} = \frac{1}{q}\frac{1}{1-z}$. Using the relation $\mathcal{R}_{a*\mu}(z) = a\mathcal{R}_\mu(az)$, we have that $R_{\rho_{\sqrt{\kappa}S}}(z) = \frac{\sqrt{\kappa}}{q}\frac{1}{1-\sqrt{\kappa}z}$. The free additive convolution formula is $R_{\rho_{\sqrt{\kappa}S}}(z) + R_{\rho_{\mathrm{sc}}}(z) = R_{\rho_Y}(z)$. Substituting z by the inverse of the Cauchy transform of ρ_Y, $G^{-1}_{\rho_Y}$, and using that $R_{\rho_{\mathrm{sc}}}(z) = z$ we get:

$$G_{\rho_Y}(z) + \frac{\sqrt{\kappa}}{q}\frac{1}{1 - \sqrt{\kappa}G_{\rho_Y}(z)} + \frac{1}{G_{\rho_Y}(z)} = z. \tag{9.111}$$



(a) $\kappa = 3$ **(b)** $\kappa = 7$

Figure 9.E.1: The continuous line is plotted from the analytical expression for the asymptotic spectral density of the observation matrix $\boldsymbol{Y}$ in example 9.B.2 for $p = 0.7$. $\rho_Y(x) = \rho_{\sqrt{\kappa}S} \boxplus \rho_{\mathrm{SC}}$ where $\rho_{\sqrt{\kappa}S}(x) = 0.7\delta(x) + 0.3\delta(x - \sqrt{\kappa})$ for $\kappa = 3$ and $\kappa = 7$. The critical value for $p = 0.7$ is $\kappa^c \approx 3.78$. This is compared to the histogram of a realization of size $N = 5000$.

Solving this equation for $G_{\rho_Y}(z)$, and using the Stieltjes inversion formula, $\mu(x) = \frac{1}{\pi} \lim_{\epsilon \to 0} \Im \boldsymbol{\mathcal{S}}_\mu(x + i\epsilon)$, we find the density of ρ_Y to be

$$\rho_Y(x) = \begin{cases} \frac{\sqrt[3]{2A^2} - 2\left(q\kappa(x^2-3) - q\sqrt{\kappa}x + q + 3\kappa\right)}{\pi 2^{\frac{5}{3}} \sqrt{3q\kappa}\, \sqrt[3]{A}} & \text{if } x \in \mathrm{Supp}(\rho_Y) \\ 0 & \text{else.} \end{cases} \tag{9.112}$$

where

$$A = q^{\frac{3}{2}}\left(\sqrt{\kappa}x - 2\right)\left(\kappa(2x^2 - 9) + \sqrt{\kappa}x - 1\right) + 9\sqrt{q}\kappa\left(\sqrt{\kappa}x + 1\right) + \sqrt{f(x)} \tag{9.113}$$

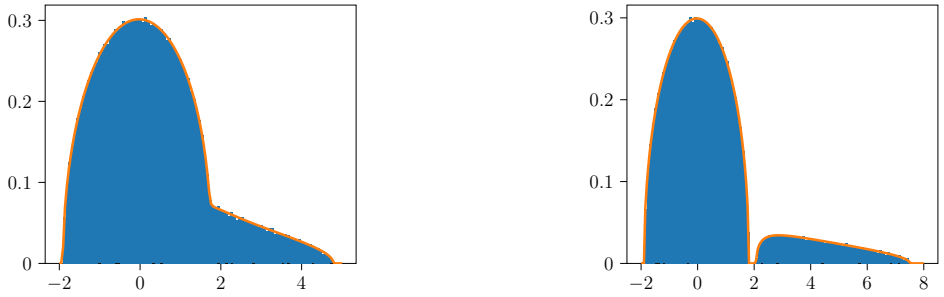
where

$$\begin{aligned} f(x) = q\Big(q\left(\sqrt{\kappa}x - 2\right)\left(2\kappa x^2 + \sqrt{\kappa}x - 9\kappa - 1\right) + 9(\kappa^{\frac{3}{2}}x + \kappa)\Big)^2 \\ - 4\Big(q\kappa(x^2 - 3) - q\sqrt{\kappa}x + q + 3\kappa\Big)^3 \end{aligned} \tag{9.114}$$

and

$$\mathrm{Supp}(\rho_Y) = \Big\{x \Big| f(x) \geq 0\Big\}. \tag{9.115}$$

If $q \leq 1$, then the support of $\rho_{\sqrt{\kappa}S}$ is only a single interval, and the support of ρ_Y is also an interval. However, for $q > 1$, $\rho_{\sqrt{\kappa}S}$ has a delta at zero, and the support of ρ_Y can be a single interval or union of two intervals, depending on κ. For fixed $q > 1$, the intervals merge at the critical value $\kappa_c = \frac{q}{\left(\sqrt[3]{q}-1\right)^3}$: if $\kappa \leq \kappa_c$, $\mathrm{Supp}(\rho_Y)$ is a single interval, while if $\kappa > \kappa_c$, $\mathrm{Supp}(\rho_Y)$ is the union of two intervals. In Fig. 9.F.1, the density is plotted for $q = 8$, for which $\kappa_c = 8$.



(a) $\kappa = 6$ **(b)** $\kappa = 16$

Figure 9.F.1: Asymptotic spectral density of the observation matrix $\boldsymbol{Y}$ in example 9.B.3 for $q = 8$. $\kappa^c = 8$. It is Compared to the empirical density of a realization of size $N = 10000$.

Extensive-Rank Non-Symmetric Matrix Denoising

10

In this chapter, we consider denoising a *non-symmetric rectangular* matrix under additive bi-rotational invariant noise. We consider the model:

$$\boldsymbol{Y} = \sqrt{\kappa}\boldsymbol{S} + \boldsymbol{Z}$$

where $\boldsymbol{S}, \boldsymbol{Z} \in \mathbb{R}^{N\times M}$ are non-symmetric matrices,$\boldsymbol{Z}$ is distributed according to a bi-rotationally invariant distribution, and $\boldsymbol{S}$ has rank which grows linearly with N.

For this model,

- We extend the rotational invariant estimators to rectangular matrices. We conjecture that the proposed estimator is optimal among the RIE class under general bi-rotational invariant noise, see (10.5).
- For the particular case of Gaussian noise:
  - We prove a trace relation which gives a solid justification for the optimality of the proposed RIE (Theorem 10.3).
  - Using the optimality of the RIE, we derive the asymptotic Bayes-optimal error in terms of the limiting singular value distribution of the observation matrix, presented in Statement 10.4.
  - We prove by independent methods that the mutual information between $\boldsymbol{S}$ and $\boldsymbol{Y}$ is linked to the asymptotic log-spherical integral (Theorem 10.5).
- We provide numerical simulations under various settings, which

Part of this work was presented in [54] F. Pourkamali and N. Macris, "Rectangular rotational invariant estimator for general additive noise matrices," in 2023 IEEE International Symposium on Information Theory (ISIT), 2023, pp. 2081–2086.

– support the optimality of the proposed (general) RIE, as suggested by the derivation based on (non-rigorous) methods from statistical physics.
– suggest that RIE is not limited to the rotational invariant signals, and can be applied regardless of of the prior to get non-trivial (although non-optimal) estimates.

10.1 Denoising Model and Rotational Invariant Estimators

Let $\boldsymbol{S} \in \mathbb{R}^{N\times M}$ be the signal matrix that we aim to estimate from the observation matrix $\boldsymbol{Y}$:

$$\boldsymbol{Y} = \sqrt{\kappa}\boldsymbol{S} + \boldsymbol{Z} \tag{10.1}$$

where $\boldsymbol{Z} \in \mathbb{R}^{N\times M}$ is a bi-rotationally invariant matrix, i.e. $P_Z(\boldsymbol{Z}) = P_Z(\boldsymbol{U}\boldsymbol{Z}\boldsymbol{V}^\intercal)$ for any orthogonal matrices $\boldsymbol{U}, \boldsymbol{V}$, and $\kappa \in \mathbb{R}_+$ is proportional to the signal-to-noise-ratio (SNR). We assume that M scales like N, and $N/M \to \alpha$. Moreover, we assume that the empirical singular value distributions (ESD) of $\boldsymbol{Z}$ and $\boldsymbol{Y}$ have well-defined limiting measures as $N \to \infty$. We denote them μ_Z, μ_Y respectively and refer to them as *limiting ESD*. Studying the problem for the case $\alpha \in (0, 1]$ suffices. Indeed, suppose the observation matrix $\boldsymbol{Y} \in \mathbb{R}^{N\times M}$ has dimensions $N > M$ (so $\alpha > 1$), then exchanging the role of M, N, we can apply our results to the matrix $\boldsymbol{Y}^\intercal$ with aspect ratio $1/\alpha \in (0, 1)$.

10.1.1 Rectangular RIE Class

Given the observation $\boldsymbol{Y}$, the class of *Rotational Invariant Estimators (RIE)* $\boldsymbol{\Xi}_S(\boldsymbol{Y})$ of $\boldsymbol{S}$ have the same singular vectors than $\boldsymbol{Y}$. More precisely, consider the SVD of $\boldsymbol{Y}$ to be:

$$\boldsymbol{Y} = \boldsymbol{U}_Y \boldsymbol{\Gamma} \boldsymbol{V}_Y^\intercal, \qquad \boldsymbol{\Gamma} = \left[\ \mathrm{diag}(\gamma_1, \cdots, \gamma_{\mathrm{N}}) \ \middle|\ \boldsymbol{0}_{N\times(M-N)}\ \right] \in \mathbb{R}^{N\times M}$$

with $\gamma_1, \cdots, \gamma_N \geq 0$ singular values of $\boldsymbol{Y}$, and orthogonal matrices $\boldsymbol{U}_Y \in \mathbb{R}^{N\times N}, \boldsymbol{V}_Y \in \mathbb{R}^{M\times M}$. RIEs $\boldsymbol{\Xi}_S(\boldsymbol{Y})$ are constructed by definition as :

$$\boldsymbol{\Xi}_S(\boldsymbol{Y}) = \sum_{j=1}^{N} \xi_j \boldsymbol{u}_j \boldsymbol{v}_j^\intercal \tag{10.2}$$

where $\boldsymbol{u}_j, \boldsymbol{v}_j$ are columns of $\boldsymbol{U}_Y, \boldsymbol{V}_Y$. The goal is to have the minimum squared error, therefore the optimal singular values are the solution to the following optimization problem:

$$\min_{\xi_1,\cdots,\xi_N} \|\boldsymbol{S} - \boldsymbol{\Xi}_S(\boldsymbol{Y})\|_{\mathrm{F}}^2 \tag{10.3}$$

One can easily see that the solutions to optimization problem (10.3) are:

$$\xi_j^* = \boldsymbol{u}_j^\intercal \boldsymbol{S} \boldsymbol{v}_j \qquad \text{for} \quad 1 \leq j \leq N \tag{10.4}$$

The particular estimator constructed with the singular values (10.4) is denoted by $\Xi_S^*(\boldsymbol{Y})$, and is called *oracle estimator* as it involves the signal matrix $\boldsymbol{S}$.

10.1.2 Algorithmic RIE

Our main contribution is the derivation of an explicit formula for the optimal singular values (10.4) which only involves the observation matrix and the knowledge of spectral measure of the noise. This formula leads to an algorithm for the estimation, which we conjecture, has in the asymptotic limit a performance matching the one of the oracle estimator (in the sense of the mean-square-error (10.3).

The optimal singular values can be approximated for sufficiently large N, as:

$$\widehat{\xi_j^*} = \frac{1}{\sqrt{\kappa}}\Bigg[\gamma_j - \frac{1}{\pi\bar{\mu}_Y(\gamma_j)}\operatorname{Im}\mathcal{C}_{\mu_Z}^{(\alpha)}\Bigg(\frac{1-\alpha}{\gamma_j}\pi\mathsf{H}[\bar{\mu}_Y](\gamma_j) + \alpha\big(\pi\mathsf{H}[\bar{\mu}_Y](\gamma_j)\big)^2 - \alpha\big(\pi\bar{\mu}_Y(\gamma_j)\big)^2 + \mathrm{i}\pi\bar{\mu}_Y(\gamma_j)\Big(\frac{1-\alpha}{\gamma_j} + 2\alpha\pi\mathsf{H}[\bar{\mu}_Y](\gamma_j)\Big)\Bigg)\Bigg] \tag{10.5}$$

where $\bar{\mu}_Y(\gamma) = \frac{1}{2}(\mu_Y(\gamma)+\mu_Y(-\gamma))$ is the symmetrization of the limiting ESD of $\boldsymbol{Y}$, $\mathcal{C}_{\mu_Z}^{(\alpha)}$ is the rectangular R-transform of μ_Z, and $\mathsf{H}[\bar{\mu}_Y]$ is the Hilbert transform of $\bar{\mu}_Y$. Derivation of the estimator (10.5) is sketched in section 10.4.1.

The algorithm to estimate $\boldsymbol{S}$ proceeds as follows:

1. Compute the SVD of $\boldsymbol{Y}$, $\boldsymbol{Y} = \boldsymbol{U}_Y\boldsymbol{\Gamma}\boldsymbol{V}_Y^\intercal$.
2. Approximate $\mathcal{G}_{\bar{\mu}_Y}(z)$ from the singular values of $\boldsymbol{Y}$, from which $\bar{\mu}_Y(\gamma)$ and $\mathsf{H}[\bar{\mu}_Y](\gamma)$ can be evaluated using (5.5).
3. Compute $\widehat{\xi_j^*}$ as in (10.5), and construct the estimator $\widehat{\Xi_S^*}(\boldsymbol{Y}) = \sum_{j=1}^N \widehat{\xi_j^*}\boldsymbol{u}_j\boldsymbol{v}_j^\intercal$.

10.1.3 Bayes optimality and MMSE

From the Bayesian estimation point of view, considering a prior distribution for the signal $P_S(\boldsymbol{S})$, one wishes to minimize the average mean-squared-error (MSE), which is defined for an estimator $\Theta_S : \mathbb{R}^{N\times M} \to \mathbb{R}^{N\times M}$ as

$$\mathrm{MSE}_{\Theta_S} = \frac{1}{N}\mathbb{E}\big\|\boldsymbol{S} - \Theta_S(\boldsymbol{Y})\big\|_{\mathrm{F}}^2 \quad ,$$

where the expectation is over $\boldsymbol{S}, \boldsymbol{Z}$. It is well known that the estimator which has the minimum MSE is the posterior mean estimator $\Theta_S^*(\boldsymbol{Y}) = \mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}]$.

Note that for model (10.1) the oracle estimator (10.4) is the best estimator among the *RIE class* (in the sense that it minimizes the MSE in this class). Furthermore the derivation of the explicit estimator (10.5) does not involve

Bayesian methodology and does not require any knowledge of the prior of the signal.

However, if the prior on the signal is bi-rotationally invariant, i.e. $P_S(\boldsymbol{S}) = P_S(\boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\intercal)$ for any orthogonal matrices $\boldsymbol{U}, \boldsymbol{V}$, these estimators are intimately related to the Bayesian one. As shown in section 10.4.2 for bi-rotationally invariant signal distributions the posterior mean estimator $\mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}]$ belongs to the RIE class. Since the oracle estimator has minimum MSE among the RIE class, we have that $\mathrm{MSE}_{\boldsymbol{\Xi}_S^*} \leq \mathrm{MSE}_{\Theta_S^*}$. On the other hand, by definition, we have $\mathrm{MSE}_{\boldsymbol{\Xi}_S^*} \leq \mathrm{MSE}_{\Theta_S^*}$. Therefore, the oracle estimator is Bayes-optimal under bi-rotational invariant prior and achieves the MMSE.

Moreover, the "exact" analytical derivation of the explicit estimator (10.5) suggests that it has the same performance as the oracle estimator as $N \to \infty$. Therefore, the above algorithm should be asymptotically Bayes-optimal with an asymptotic MSE equal to the MMSE. Denoting the rhs in (10.5) as a *function* of singular values of $\boldsymbol{Y}$, $\widehat{\xi^*} : \mathrm{supp}(\mu_Y) \to \mathbb{R}$, we are led to the following result.

Statement 10.1 (**MMSE**)**.** *Suppose that $\boldsymbol{S}, \boldsymbol{Z}$ have bi-rotational invariant priors, and Assume their ESDs converge to well-defined measures μ_S, μ_Z with bounded second moments. We have:*

$$\lim_{N\to\infty} \mathrm{MMSE}_N(\kappa) = \int x^2 \mu_S(x)\, dx - \int \widehat{\xi^*}(x)^2 \mu_Y(x)\, dx \tag{10.6}$$

where μ_Y is the limiting ESD of $\boldsymbol{Y}$, $\mu_Y = \mu_S \boxplus_\alpha \mu_Z$.

Remark 10.1. *Note that, for non-rotation invariant priors the estimator* (10.5) *still can be applied, however it may results in a sub-optimal estimate of the signal. However, this estimate can be used as a "warmed-up" spectral initialization for more efficient algorithms (see for example [23, 24]).*

10.2 Gaussian Noise

10.2.1 RIE

In this section, we consider the case of Gaussian noise matrix, more precisely we suppose that $\boldsymbol{Z} \in \mathbb{R}^{N\times M}$ has i.i.d. Gaussian entries of variance $1/N$, and $\kappa \in \mathbb{R}_+$ is proportional to the signal-to-noise-ratio (SNR). We make the following assumptions:

Assumption 10.1. *The operator norm of $\boldsymbol{S}$, and the ratio M/N are bounded by some numerical constant $K > 0$ independent of N, M.*

Recall that the resolvent of the matrix $\boldsymbol{Y}\boldsymbol{Y}^\intercal$, evaluated at z^2 is defined as:

$$\boldsymbol{G}_{\boldsymbol{Y}\boldsymbol{Y}^\intercal}(z^2) = \left(z^2\boldsymbol{I} - \boldsymbol{Y}\boldsymbol{Y}^\intercal\right)^{-1}$$

Now, define two random functions of $z \in \mathbb{C}\backslash\mathbb{R}$ as:

$$G(z) = \frac{1}{N}\,\mathrm{Tr}\,\boldsymbol{G}_{\boldsymbol{Y}\boldsymbol{Y}^\intercal}(z^2), \quad L(z) = \frac{1}{N}\,\mathrm{Tr}\,\boldsymbol{G}_{\boldsymbol{Y}\boldsymbol{Y}^\intercal}(z^2)\boldsymbol{Y}\boldsymbol{S}^\intercal \tag{10.7}$$

Proposition 10.2. *For $1 \leq j \leq N$, and for any $\epsilon > 0$ such that $[\gamma_j - \epsilon, \gamma_j + \epsilon] \cap \{\gamma_1, \cdots, \gamma_N\} = \{\gamma_j\}$, the optimal singular values* (10.4) *satisfy*

$$\xi_j^* = \lim_{\eta \to 0} \frac{\int_{\gamma_j - \epsilon}^{\gamma_j + \epsilon} \operatorname{Im} L(x + \mathrm{i}\eta)\, dx}{\int_{\gamma_j - \epsilon}^{\gamma_j + \epsilon} \operatorname{Im}\big\{(x + \mathrm{i}\eta) G(x + \mathrm{i}\eta)\big\}\, dx} \tag{10.8}$$

The proof of the above Proposition is presented in section 10.4.3. Note that (10.8) is an exact formula for the optimal singular values ξ_j^*, but in practice given an explicit expression for $L(z)$ we use the following approximation to evaluate ξ_j^*:

$$\xi_j^* \approx \frac{\operatorname{Im} L(z)}{\operatorname{Im}\big\{z G(z)\big\}} \equiv \widehat{\xi}_j^* \qquad \text{for } z = \gamma_j + \mathrm{i}\eta, \quad \text{with } \eta \ll 1 \tag{10.9}$$

The definition of the function $L(z)$ in the numerator of the estimator (10.9), involves the signal matrix. Therefore, to use the estimator, we need to find a way to estimate this function only from the data. In the following theorem, we give an asymptotic approximation of $L(z)$, which we prove in section 10.4.4.

Theorem 10.3 (Estimation of $L(z)$)**.** *Let $\alpha_0 = N/M$. For any $z \in \mathbb{C}\backslash\mathbb{R}$ with $\big|\operatorname{Im} z\big| < 1$, the function $L(z)$ defined in* (10.7) *satisfies*

$$L(z) = \frac{1}{\sqrt{\kappa}} \Big[G(z)\big(z^2 + 1 - \frac{1}{\alpha_0}\big) - z^2 G^2(z) - 1 \Big] + \epsilon_N \tag{10.10}$$

where the error term ϵ_N is bounded as:

$$\epsilon_N \leq \frac{C_K + X}{N \big|\operatorname{Im} z\big|^3}$$

with C_K a constant depending on K, and X is a complex sub-Gaussian random variables with finite sub-Gaussian norm depending on K.

Remark 10.2. *We believe that adopting the methodology developed in [129, 130] the approximation* (10.10) *can be improved, with the error term controlled by $\big(N|\operatorname{Im} z|\big)^{-1}$. We numerically verify this conjecture in section 10.3.2.*

Algorithm

Using the explicit expression (10.10) for $L(z)$, we are led to the following approximation to evaluate ξ_j^*:

$$\widehat{\xi}_j^* = \frac{1}{\sqrt{\kappa}} \frac{\operatorname{Im}\Big\{ G(z)\big(z^2 + 1 - \frac{1}{\alpha_0}\big) - z^2 G^2(z) \Big\}}{\operatorname{Im}\big\{z G(z)\big\}} \qquad \text{for } z = \gamma_j + \mathrm{i}\frac{1}{\sqrt{N}} \tag{10.11}$$

Remark 10.3 (On imaginary part of z in (10.9))**.** *Approximating the exact formula* (10.8) *with the expression* (10.9) *is more accurate when z is closer to the real line (small η). On the other hand, for z close to the real line the error of the approximation* (10.10) *gets worse. Considering that the error in* (10.10) *is controlled by $\big(N|\mathrm{Im}\, z|\big)^{-1}$, we can see that $\eta = 1/N^{\epsilon}$ with any $0 < \epsilon < 1$ should work properly as N increases. We study the effect of choice of ϵ in the numerical section, see Fig. 10.3.3b.*

Remark 10.4 (Relation to the formula (10.5))**.** *First note that $zG(z)$ in the denominator of* (10.11) *is the Stieltjes transform of empirical symmetric spectral measure of $\boldsymbol{Y}$, and in the limit $\eta \to 0$, $\mathrm{Re}\big\{zG(z)\big\} = \pi \mathsf{H}[\bar{\mu}_Y](x)$. Therefore, from* (10.10)*, for $z = \gamma_j + \mathsf{i}\eta$ with $\eta \ll 1$, we have*

$$\begin{aligned}\frac{\mathrm{Im}\, L(z)}{\mathrm{Im}\big\{zG(z)\big\}} &= \frac{1}{\sqrt{\kappa}} \frac{\mathrm{Re}\big\{\mathrm{G(z)}\big\} 2\gamma_j \eta + \mathrm{Im}\big\{\mathrm{G(z)}\big\}\big(\gamma_j^2 - \eta^2 + 1 - \frac{1}{\alpha_0}\big)}{\gamma_j \mathrm{Im}\big\{G(z)\big\} + \mathrm{Re}\big\{G(z)\big\}\eta} \\ &\qquad - \frac{1}{\sqrt{\kappa}} 2\mathrm{Re}\big\{zG(z)\big\} \\ &\approx \frac{1}{\sqrt{\kappa}}\Big[\gamma_j + \big(1 - \frac{1}{\alpha_0}\big)\frac{1}{\gamma_j} - 2\pi \mathsf{H}[\bar{\mu}_Y](\gamma_j)\Big]\end{aligned} \tag{10.12}$$

For $\boldsymbol{Z}$ with i.i.d. Gaussian entries of variance $1/N$ and with the assumption that $\alpha_0 \to \alpha$ we have $\mathcal{C}^{(\alpha)}_{\mu_Z}(z) = \frac{1}{\alpha} z$, and thus the expression in (10.5) *reduces to* (10.12)*.*

10.2.2 MMSE

Given the rather simple expression for the optimal singular values, we can compute the asymptotic MMSE for the particular case of Gaussian noise, see section 10.4.5 for the derivation.

Statement 10.4 (**Gaussian MMSE**)**.** *Assume that the prior on $\boldsymbol{S}$ is bi-rotational invariant, and the ESD of $\boldsymbol{S}$ converges to a well-defined limiting measure μ_S with compact support and bounded second moment. We have:*

$$\lim_{N\to\infty} \mathrm{MMSE}_N(\kappa) = \frac{1}{\kappa}\Big[\frac{1}{\alpha} - \big(\frac{1}{\alpha} - 1\big)^2 \int \frac{\mu_Y(x)}{x^2}\, dx - \frac{\pi^2}{3}\int \mu_Y(x)^3\, dx\Big] \tag{10.13}$$

where μ_Y is the limiting ESD of $\boldsymbol{Y}$ and $\mu_Y = \mu_S \boxplus_\alpha \mu_{\mathrm{MP}}$.

Remark 10.5. *In the symmetric case, the asymptotic MMSE of a Gaussian channel is linked to the free Fisher information of non-commutative random variables [48]. Using this link, we can deduce the continuity of the MMSE as a function of SNR, which rules out the existence of the first-order phase transitions. Moreover, using the I-MMSE relation [1], this link also implies a rather explicit expression for the asymptotic mutual information. We believe that similar relations hold for the rectangular case and the MMSE should be a*

continuous function of κ. However in the rectangular case free probability [131] is much less developed than its symmetric counterpart, and these considerations are beyond the scope of the present paper.

10.2.3 Mutual information

In this subsection we prove that the asymptotic mutual information is linked to the asymptotic rectangular spherical integral. The rectangular spherical integral is defined for two matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{N\times M}$ as:

$$\mathcal{I}_{N,M}(\boldsymbol{A},\boldsymbol{B}) := \iint D\boldsymbol{U}\, D\boldsymbol{V}\, e^{N \operatorname{Tr} \boldsymbol{A}^\intercal \boldsymbol{U}\boldsymbol{B}\boldsymbol{V}^\intercal}$$

where $D\boldsymbol{U}, D\boldsymbol{V}$ denote the Haar measures over the groups of $N \times N$ and $M \times M$ orthogonal matrices. The asymptotic behavior of these integrals has been studied in [132] which proves that the limit $\lim_{N\to\infty} \frac{1}{N^2} \ln \mathcal{I}_{N,M}(\boldsymbol{A},\boldsymbol{B})$ exists and equals a variational formula given in terms of limiting ESD of $\boldsymbol{A}, \boldsymbol{B}$.

We make the following assumptions on the prior of $\boldsymbol{S}$:

Assumption 10.2. *The empirical singular value distribution of $\boldsymbol{S}$ converges almost surely weakly to a well-defined probability density function $\mu_S(x)$ with compact support in $[C_1, C_2]$ with $C_1, C_2 \in \mathbb{R}_{\geq 0}$. Moreover, the symmetrization of μ_S has bounded second moment $\int x^2 \, d\bar{\mu}_S < \infty$, finite non-commutative entropy $\iint \ln|x-y| \, d\bar{\mu}_S(x)\, d\bar{\mu}_S(y) > -\infty$, and $\int \ln|x| \, d\bar{\mu}_S(x) > -\infty$.*

Assumption 10.3. *The second moment of $\mu_S^{(N)}$ is almost surely bounded.*

Let

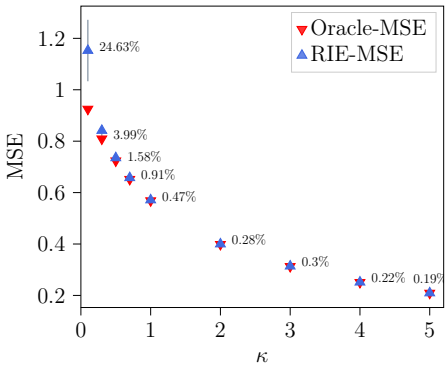
$$\mathcal{J}^{(\alpha)}[\mu_{\sqrt{\kappa}S}, \mu_{\sqrt{\kappa}S} \boxplus_\alpha \mu_{\mathrm{MP}}] = \lim_{N\to+\infty} \frac{1}{NM} \ln \mathcal{I}_{N,M}(\sqrt{\kappa}\boldsymbol{S}, \boldsymbol{Y})$$

where $\mu_{\sqrt{\kappa}S}$ is the limiting spectral distribution of $\sqrt{\kappa}\boldsymbol{S}$ and μ_{MP} is the Marchenko-Pastur distribution.

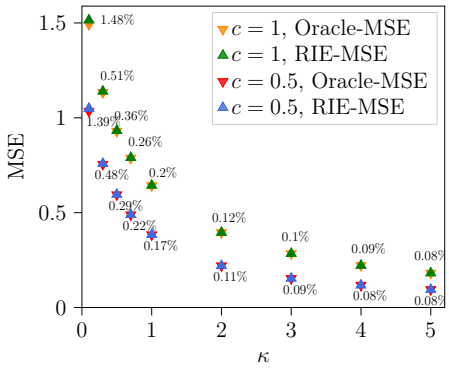
Theorem 10.5 (Mutual Information). *Under assumptions 10.2,10.3 and $M/N \leq K$, we have:*

$$\lim_{N\to\infty} \frac{1}{MN} I_N(\boldsymbol{S};\boldsymbol{Y}) = \kappa\alpha \int x^2 \mu_S(x)\, dx - \mathcal{J}^{(\alpha)}[\mu_{\sqrt{\kappa}S}, \mu_{\sqrt{\kappa}S} \boxplus_\alpha \mu_{\mathrm{MP}}] \quad (10.14)$$

Remark 10.6. *In [98], the asymptotic log-spherical integral $\mathcal{J}^{(\alpha)}[\mu_{\sqrt{\kappa}S}, \mu_{\sqrt{\kappa}S} \boxplus_\alpha \mu_{\mathrm{MP}}]$ is computed explicitly, which together with Theorem 10.5 gives an an explicit expression for the asymptotic mutual information under Gaussian noise.*



(a) Uniform spectral noise matrix, $N = M = 1000$



(b) Sum of rank-one factors noise matrix, $N = 1000, M = 2000, L = cN$

Figure 10.3.1: Performance of the algorithmic RIE based on (10.5) as compared to the oracle one. Signal matrix $\boldsymbol{S} \in \mathbb{R}^{N\times M}$ has i.i.d. Gaussian entries of variance $1/N$. Results are averaged over 10 runs (error bars are invisible). Average relative error is also reported. In both examples, the Hilbert transform of the observation is computed numerically using Cauchy kernel method in [67].

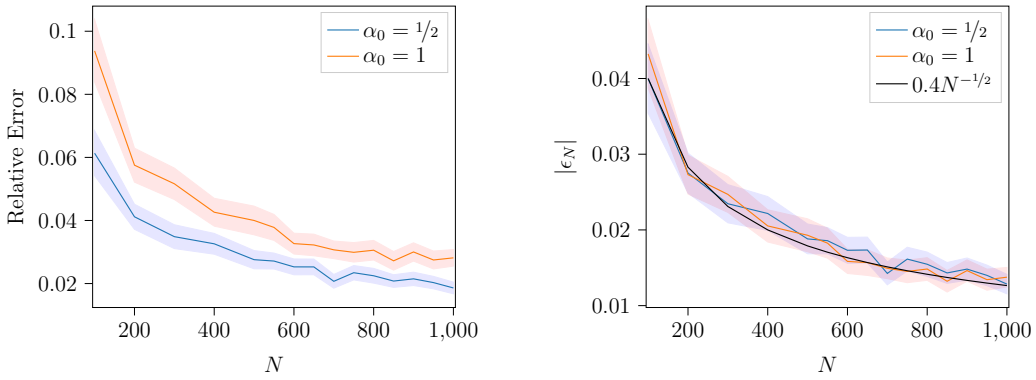
10.3 Numerical Simulations

10.3.1 General bi-rotational invariant noise

In Fig. 10.3.1, the performance of the algorithmic RIE based on (10.5) is compared against the oracle estimator (10.4) for two cases of noise distribution and Gaussian signal matrix, i.e. $\boldsymbol{S}$ is a matrix with i.i.d. Gaussian entries of variance $1/N$.

Uniform spectral noise. For this prior, the noise matrix $\boldsymbol{Z} \in \mathbb{R}^{N\times N}$ is constructed as $\boldsymbol{Z} = \boldsymbol{U}\mathrm{diag}(r_1, \cdots, r_N)\boldsymbol{V}^\intercal$, where $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{R}^{N\times N}$ are independent Haar distributed matrices, and the singular values $r_1, \cdots, r_N$ are chosen independently uniformly from $[0, 2]$. The limiting spectral measure of the uniform noise distribution $\mathcal{U}_{[0,2]}$ has rectangular R-transform $\mathcal{C}^{(1)}_{\mathcal{U}_{[0,2]}}(z) = 2\sqrt{z}\coth\left(2\sqrt{z}\right) - 1$.

Sum of rank-one factors. For the noise matrix we take a sum of rank-one matrices, $\boldsymbol{Z} = \sum_{k=1}^{L} \boldsymbol{u}_k\boldsymbol{v}_k^\intercal$, where $\boldsymbol{u}_k$'s and $\boldsymbol{v}_k$'s are independent uniform random vectors of the unit norm in $\mathbb{R}^N, \mathbb{R}^M$. Denoting the limiting ratio $L/N \to c$, the liming symmetrized ESD of $\boldsymbol{Z}$ is the rectangular analogue of the symmetrized Poisson distribution with parameter c, with rectangular R-transform $\mathcal{C}^{(\alpha)}(z) = cz/1-z$ (see section 4.3, Proposition 6.1 in [133]).



(a) Relative error **(b)** Error term

Figure 10.3.2: Validity of the estimation (10.10). Plots are average of 100 experiments and 95% confidence interval is also depicted. The signal matrix $\boldsymbol{S} \in \mathbb{R}^{N\times M}$ has i.i.d. Gaussian entries of variance $1/N$, and $M = N/\alpha_0$. the expressions are evaluated for $z = 1 + \mathrm{i}/\sqrt{N}$. In the left panel, the relative error (10.15) is plotted for various values of N. On the same simulations, the error term is plotted in the right panel which behaves as $N^{-1/2}$ which matches with the conjecture of remark 10.2.

10.3.2 Gaussian noise

Validity of Theorem 10.3

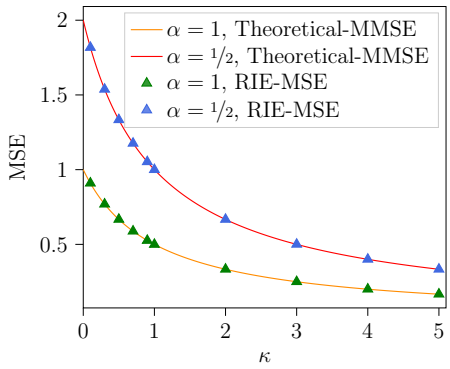
In Fig. 10.3.2, we numerically verify Theorem 10.3 and check the behavior of the error term ϵ_N. For simplicity, we set the SNR parameter to one, $\kappa = 1$. In Fig. 10.3.2a, the relative error is plotted,

$$\frac{|\epsilon_N|}{|L(z)|} = \frac{\left|L(z) - \left[G(z)\left(z^2 + 1 - \frac{1}{\alpha_0}\right) - z^2G^2(z) - 1\right]\right|}{|L(z)|} \tag{10.15}$$

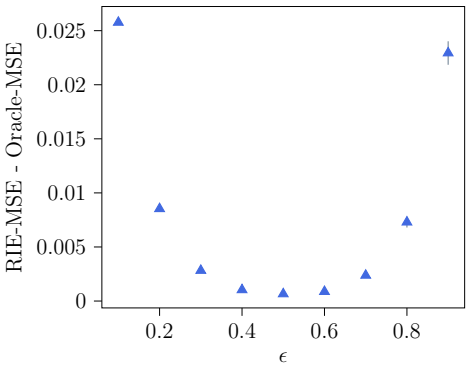
for the case of a signal matrix with i.i.d. Gaussian entries of variance $1/N$. In Fig. 10.3.2b, the behavior of the error term is depicted, which verifies the conjecture stated in remark 10.2, namely that the error is controlled by $\left(N|\mathrm{Im}\, z|\right)^{-1}$.

Gaussian Signal

If we consider the signal matrix to have i.i.d. Gaussian entries of variance $1/N$, then each entry of $\boldsymbol{Y}$ can be viewed as an independent scalar AWGN channel. For this scalar channel, the MMSE equals $\frac{1}{N}\frac{1}{1+\kappa}$ [1]. Therefore, the (normalized) MMSE of the matrix problem is $\frac{M}{N}\frac{1}{1+\kappa} \to \frac{1}{\alpha}\frac{1}{1+\kappa}$ for $N \to \infty$. As a sanity check of Statement 10.4, using the fact that μ_Y is the Marchenko-Pastur (MP) law rescaled with $\sqrt{\kappa+1}$, we compute the MMSE analytically (with the help of *Mathematica* [117]) and find it equal to $\frac{1}{\alpha}\frac{1}{1+\kappa}$. In Fig. 10.3.3a, MSE of RIE is compared to the theoretical MMSE for $\alpha = 1, \alpha = 1/2$. Note that, for



(a) Performance of the RIE (10.5) for the Gaussian signal and noise. Signal and noise matrices $\boldsymbol{S}, \boldsymbol{Z} \in \mathbb{R}^{N\times M}$ have i.i.d. Gaussian entries of variance $1/N$. The MMSE is plotted for two aspect ratios $\alpha = 1, 1/2$, and the RIE (10.5) is applied to $N = 1000$. Results are averaged over 10 runs.



(b) Performance of the RIE using (10.11) for the Gaussian signal. The formula (10.11) is used to estimate the optimal singular values with $z = \gamma_j + \mathrm{i}N^{-\epsilon}$. RIE is applied to $N = 1000, M = 2000, \kappa = 2$, and results are averaged over 10 runs.

this example, we use the RIE (10.5), and the Hilbert transform used in RIE is the exact Hilbert transform of the symmetrization of MP law rescaled with $\sqrt{\kappa+1}$, which is $\mathsf{H}_{\bar{\mu}_Y}(x) = \frac{x}{2+2\kappa} - \frac{1-\alpha}{2\alpha x}$.

In Fig. 10.3.3b, we investigate the performance of the RIE using the relation (10.11) for various values of the imaginary part of z. In this plot, the difference of the MSEs of the RIE and the oracle estimator for the Gaussian signal and noise matrices is depicted. The RIE is applied with $z = \gamma_j + \mathrm{i}N^{-\epsilon}$. A few remarks about this plot are in order. First, it supports the conjecture stated in remark 10.2 that the error term in (10.3) is controlled by $\left(N|\mathrm{Im}\, z|\right)^{-1}$. Moreover, we can see that as the imaginary part of z increases (ϵ decreases) the difference increases. For this regime, the error term in (10.3) becomes small, however the approximation (10.11) of the exact formula (10.8) is inaccurate. On the other hand, for z with small imaginary part, this approximation is more accurate, but the error of the estimation in (10.3) becomes large.

Signal with sparse spectrum

The signal matrix $\boldsymbol{S} \in \mathbb{R}^{N\times M}$ is constructed as

$$\boldsymbol{S} = \boldsymbol{U}[\mathrm{diag}(\sigma_1, \cdots, \sigma_N), \boldsymbol{0}_{M-N}]\boldsymbol{V}^\intercal$$

where $\boldsymbol{U} \in \mathbb{R}^{N\times N}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$ are independent Haar distributed matrices, and the singular values $\sigma_1, \cdots, \sigma_N$ are independent Bernoulli random variables, $\mu_S = p\delta_0 + (1-p)\delta_{+1}$ for $0 \le p \le 1$. In Fig. 10.3.4 MSE of RIE is compared to the MSE of oracle estimator with $\alpha = 1/2$ for $p = 0.2, 0.9$. We observe that, in the high-sparsity regime $p = 0.9$, the model behaves like finite-rank signal and the MSE is close to the rank-one MMSE computed in [10].

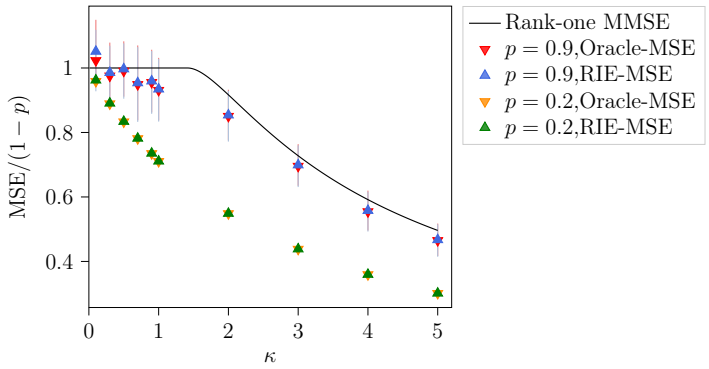


Figure 10.3.4: Signal with Bernoulli spectrum. MSE is normalized by the norm of the signal, $1-p$. The RIE is applied to $N = 1000, M = 2000$, and the results are averaged over 10 runs (error bars might be invisible).

10.3.3 Non-rotational invariant signal distribution

We consider $\boldsymbol{S}$ to have i.i.d. entries from the Bernoulli-Rademacher distribution,

$$S_{i,j} = \begin{cases} +\frac{1}{\sqrt{N}} & \text{with probability } \frac{1-p}{2} \\ 0 & \text{with probability } p \\ -\frac{1}{\sqrt{N}} & \text{with probability } \frac{1-p}{2} \end{cases}, \qquad \forall \quad 1 \le i \le N, \quad 1 \le j \le M$$

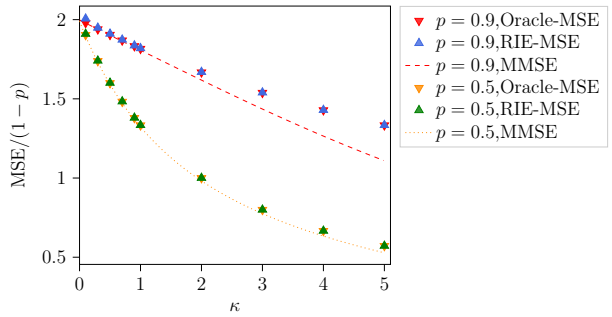
With normalization $1/\sqrt{N}$, the spectrum of $\boldsymbol{S}$ does not grow with the dimension and has a finite support, thus we can apply our estimator to reconstruct $\boldsymbol{S}$. *Note that the prior of $\boldsymbol{S}$ is not rotationally invariant, and neither the oracle estimator nor the RIE are optimal.* In Fig. 10.3.5, the performance of the RIE is compared with the oracle estimator for two cases of noise priors. Note that under Gaussian noise, the MMSE can be computed simply by considering the MMSE of scalar channel, and the MMSE is also plotted. We can see that RIE, although it is sub-optimal, can give a non-trivial estimate of the signal for non-rotationally invariant priors.

10.4 Analytical Derivations and Proofs

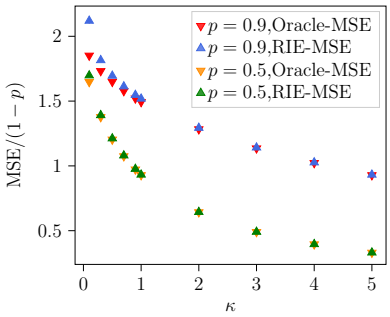
10.4.1 Derivation sketch of the explicit rectangular RIE

Let the SVD of the signal be $\boldsymbol{S} = \sum_{k=1}^{N} \sigma_k \boldsymbol{s}_k^{(l)} \boldsymbol{s}_k^{(r)\intercal}$ where $\boldsymbol{s}_k^{(r)}/\boldsymbol{s}_k^{(l)}$ is the right/left singular vector of $\boldsymbol{S}$ corresponding to the k-th singular value σ_k. From (10.4), the optimal singular values of the oracle RIE can be written as:

$$\xi_j^* = \boldsymbol{u}_j^\intercal \boldsymbol{S} \boldsymbol{v}_j = \sum_{k=1}^{N} \sigma_k \left(\boldsymbol{u}_j^\intercal \boldsymbol{s}_k^{(l)}\right)\left(\boldsymbol{v}_j^\intercal \boldsymbol{s}_k^{(r)}\right)$$



(a) Gaussian noise, $N = 1000, M = 2000$



(b) Sum of rank-one factors noise matrix, $N = 1000, M = 2000, c = 1$

Figure 10.3.5: Performance of the RIE ((10.5) and (10.11)) and oracle estimator for non-rotational invariant signal. Signal matrix $\boldsymbol{S} \in \mathbb{R}^{N\times M}$ has i.i.d. Bernoulli-Rademacher entries (divided by $1/\sqrt{N}$. Results are averaged over 10 runs (error bars are invisible).

The main assumption is that in the large-N limit, ξ_j^*'s can be approximated by the expectation, $\widehat{\xi_i^*} = \sum_{j=1}^N \sigma_j \left\langle \left(\boldsymbol{u}_j^\intercal \boldsymbol{s}_k^{(l)}\right)\left(\boldsymbol{v}_j^\intercal \boldsymbol{s}_k^{(r)}\right)\right\rangle$, where the expectation $\langle - \rangle$ is over the singular vectors of the observation $\boldsymbol{Y}$.

Therefore, to compute the optimal singular vales, we need to find the overlap $\left\langle \left(\boldsymbol{u}_j^\intercal \boldsymbol{s}_k^{(l)}\right)\left(\boldsymbol{v}_j^\intercal \boldsymbol{s}_k^{(r)}\right)\right\rangle$ between singular vectors of $\boldsymbol{S}$ and singular vectors of $\boldsymbol{Y}$. In what follows, we will see that (a rescaling of) this quantity can be expressed in terms of j-th singular value of $\boldsymbol{Y}$ and k-th singular value of $\boldsymbol{S}$ and the limiting measures, indeed. Thus, we will use the notation $O(\gamma_j, \sigma_k) = N\left\langle \left(\boldsymbol{u}_j^\intercal \boldsymbol{s}_k^{(l)}\right)\left(\boldsymbol{v}_j^\intercal \boldsymbol{s}_k^{(r)}\right)\right\rangle$ in the following and write

$$\widehat{\xi_j^*} = \frac{1}{N}\sum_{k=1}^{N} \sigma_k O(\gamma_j, \sigma_k) \tag{10.16}$$

In the next section, we discuss how this overlap can be computed from the resolvent of the "Hermitized" version of $\boldsymbol{Y}$.

Relation between overlap and the resolvent

Construct the symmetric matrix $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{(N+M)\times(N+M)}$ from the matrix $\boldsymbol{Y}$,

$$\boldsymbol{\mathcal{Y}} = \begin{bmatrix} \mathbf{0}_{N\times N} & \boldsymbol{Y} \\ \boldsymbol{Y}^\intercal & \mathbf{0}_{M\times M} \end{bmatrix} \tag{10.17}$$

By Theorem 7.3.3 in [71], the eigen-decomposition of $\boldsymbol{\mathcal{Y}}$ reads:

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{W} \begin{bmatrix} \mathrm{diag}(\gamma_1, \cdots, \gamma_N) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathrm{diag}(\gamma_1, \cdots, \gamma_N) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix} \boldsymbol{W}^\intercal \tag{10.18}$$

$$\boldsymbol{W} = \begin{bmatrix} \hat{\boldsymbol{U}}_Y & \hat{\boldsymbol{U}}_Y & \boldsymbol{0}_{N\times(M-N)} \\ \hat{\boldsymbol{V}}_Y^{(1)} & -\hat{\boldsymbol{V}}_Y^{(1)} & \boldsymbol{V}_Y^{(2)} \end{bmatrix}$$

with $\boldsymbol{V}_Y = \begin{bmatrix} \boldsymbol{V}_Y^{(1)} & \boldsymbol{V}_Y^{(2)} \end{bmatrix}$ in which $\boldsymbol{V}_Y^{(1)} \in \mathbb{R}^{M\times N}$, and $\hat{\boldsymbol{V}}_Y^{(1)} = \frac{1}{\sqrt{2}}\boldsymbol{V}_Y^{(1)}$, $\hat{\boldsymbol{U}}_Y = \frac{1}{\sqrt{2}}\boldsymbol{U}_Y$. Denote the eigenvectors of $\boldsymbol{\mathcal{Y}}$ by $\boldsymbol{w}_i \in \mathbb{R}^{M+N}$, $i = 1, \ldots, M+N$. Define the resolvent of $\boldsymbol{\mathcal{Y}}$

$$\boldsymbol{G}_{\mathcal{Y}}(z) = \begin{bmatrix} z\boldsymbol{I} - \boldsymbol{\mathcal{Y}} \end{bmatrix}^{-1}$$

For $z = x - \mathsf{i}\eta$, we have:

$$\boldsymbol{G}_{\mathcal{Y}}(x - \mathsf{i}\eta) = \sum_{k=1}^{2N} \frac{x + \mathsf{i}\eta}{(x - \tilde{\gamma}_k)^2 + \eta^2} \boldsymbol{w}_k \boldsymbol{w}_k^\intercal + \frac{x + \mathsf{i}\eta}{x^2 + \eta^2} \sum_{k=2N+1}^{M+N} \boldsymbol{w}_k \boldsymbol{w}_k^\intercal$$

where $\tilde{\gamma}_k$ are the non-trivially zero eigenvalues of $\boldsymbol{\mathcal{Y}}$, which are in fact the (signed) singular values of $\boldsymbol{Y}$, $\tilde{\gamma}_1 = \gamma_1, \ldots, \tilde{\gamma}_N = \gamma_N, \tilde{\gamma}_{N+1} = -\gamma_1, \ldots, \tilde{\gamma}_{2N} = -\gamma_N$. Define set of vectors $\boldsymbol{r}_i, \boldsymbol{l}_i \in \mathbb{R}^{N+M}$ for $i = 1, \ldots, N$ as:

$$\boldsymbol{r}_i = \begin{bmatrix} \boldsymbol{0}_N \\ \boldsymbol{s}_i^{(r)} \end{bmatrix} \qquad \boldsymbol{l}_i = \begin{bmatrix} \boldsymbol{s}_i^{(l)} \\ \boldsymbol{0}_M \end{bmatrix}$$

We have

$$\begin{aligned} \boldsymbol{r}_j^\intercal \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathsf{i}\eta)\big) \boldsymbol{l}_j = &\sum_{k=1}^{2N} \frac{\eta}{(x - \tilde{\gamma}_k)^2 + \eta^2} (\boldsymbol{r}_j^\intercal \boldsymbol{w}_k)(\boldsymbol{w}_k^\intercal \boldsymbol{l}_j) \\ &+ \frac{x + \mathsf{i}\eta}{x^2 + \eta^2} \sum_{k=2N+1}^{M+N} (\boldsymbol{r}_j^\intercal \boldsymbol{w}_k)(\boldsymbol{w}_k^\intercal \boldsymbol{l}_j) \end{aligned} \tag{10.19}$$

Given the structure of $\boldsymbol{w}_k$'s in (10.18):

$$(\boldsymbol{r}_j^\intercal \boldsymbol{w}_k)(\boldsymbol{w}_k^\intercal \boldsymbol{l}_j) = \begin{cases} \frac{1}{2}\big(\boldsymbol{u}_k^\intercal \boldsymbol{s}_j^{(l)}\big)\big(\boldsymbol{v}_k^\intercal \boldsymbol{s}_j^{(r)}\big) & \text{for } 1 \le k \le N \\ -\frac{1}{2}\big(\boldsymbol{u}_k^\intercal \boldsymbol{s}_j^{(l)}\big)\big(\boldsymbol{v}_k^\intercal \boldsymbol{s}_j^{(r)}\big) & \text{for } N+1 \le k \le 2N \\ 0 & \text{for } 2N+1 \le k \le M+N \end{cases}$$

Taking an average over singular vectors of $\boldsymbol{Y}$ in (10.19), we find:

$$\boldsymbol{r}_j^\intercal \big\langle \operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathsf{i}\eta) \big\rangle \boldsymbol{l}_j = \frac{1}{N} \sum_{k=1}^{2N} \frac{\eta}{(x - \tilde{\gamma}_k)^2 + \eta^2} (-1)^{\mathbb{I}(k>N)} O(\gamma_k, \sigma_j) \tag{10.20}$$

Now, taking the limit $N \to \infty$, we obtain:

$$\boldsymbol{r}_j^\intercal \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathsf{i}\eta)\big) \boldsymbol{l}_j \xrightarrow{N\to\infty} \int_{\mathbb{R}} \frac{\eta}{(x - t)^2 + \eta^2} O(t, \sigma_j) \bar{\mu}_Y(t)\, dt$$

where $O(t, \sigma_j)$ is extended (continuously) to arbitrary values inside the support of $\bar{\mu}_Y$ (the symmetrized limiting singular value distribution of $\boldsymbol{Y}$) with the

property that $O(-t, \sigma_j) = -O(t, \sigma_j)$. Sending $\eta \to 0$, we find the following formula valid in the large N limit:

$$\boldsymbol{r}_j^\intercal \big\langle \operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathrm{i}\eta) \big\rangle \boldsymbol{l}_j \approx \pi \bar{\mu}_Y(x) O(x, \sigma_j) \tag{10.21}$$

Eq. (10.21) is important because it enables us to investigate the overlap through the resolvent of $\boldsymbol{\mathcal{Y}}$. In the next section, we derive a relation between this resolvent and the signal $\boldsymbol{S}$ which will allow us to find a formula for the optimal singular values ξ_i^*'s in terms of the singular values of the observation matrix $\boldsymbol{Y}$.

Resolvent relation

To derive a resolvent relation between the observation and the signal, we consider the model

$$\boldsymbol{Y} = \boldsymbol{S} + \boldsymbol{U}\boldsymbol{Z}\boldsymbol{V}^\intercal \tag{10.22}$$

with $\boldsymbol{Z}$ a fixed matrix with limiting singular value distribution μ_Z, and $\boldsymbol{U} \in \mathbb{R}^{N\times N}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$ random orthogonal matrices. Note that for convenience the SNR parameter has been absorbed into $\boldsymbol{S}$, so to obtain the estimator for model (10.1), this estimator should be divided by $\sqrt{\kappa}$ eventually.

In Appendix 10.A, we derive the relation (10.23) for the resolvent $\boldsymbol{G}_{\mathcal{Y}}(z)$, in which $\langle . \rangle$ is the expectation w.r.t. the singular vectors of $\boldsymbol{Y}$, and $\boldsymbol{G}_{S^\intercal S}$ is the resolvent matrix of $\boldsymbol{S}^\intercal \boldsymbol{S}$.

$$\begin{aligned} \langle \boldsymbol{G}_{\mathcal{Y}}(z) \rangle &= \left\langle \begin{bmatrix} z^{-1}\boldsymbol{I}_N + z^{-1}\boldsymbol{Y}\boldsymbol{G}_{Y^\intercal Y}(z^2)\boldsymbol{Y}^\intercal & \boldsymbol{Y}\boldsymbol{G}_{Y^\intercal Y}(z^2) \\ \boldsymbol{G}_{Y^\intercal Y}(z^2)\boldsymbol{Y}^\intercal & z\boldsymbol{G}_{Y^\intercal Y}(z^2) \end{bmatrix} \right\rangle \\ &\approx \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}^\intercal & \boldsymbol{C} \end{bmatrix} \end{aligned} \tag{10.23}$$

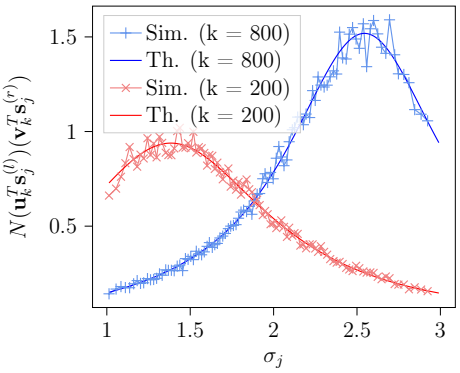
where each block is:

$$\begin{aligned} \boldsymbol{A} &= (z - \zeta_1^*)^{-1}\boldsymbol{I}_N + (z - \zeta_1^*)^{-1}\boldsymbol{S}\boldsymbol{G}_{S^\intercal S}\big((z - \zeta_2^*)(z - \zeta_1^*)\big)\boldsymbol{S}^\intercal \\ \boldsymbol{B} &= \boldsymbol{S}\boldsymbol{G}_{S^\intercal S}\big((z - \zeta_2^*)(z - \zeta_1^*)\big) \\ \boldsymbol{C} &= (z - \zeta_1^*)\boldsymbol{G}_{S^\intercal S}\big((z - \zeta_2^*)(z - \zeta_1^*)\big) \end{aligned}$$

with

$$\begin{aligned} &\zeta_a^* = z\frac{Z(z)}{\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big) + 1}, \quad \zeta_b^* = \alpha z \frac{Z(z)}{\alpha\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big) + 1}, \\ &Z(z) = \mathcal{C}_{\mu_Z}^{(\alpha)}\left(\frac{1}{z^2}T^{(\alpha)}\Big(\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big)\Big)\right) \end{aligned} \tag{10.24}$$

As a sanity check, by considering the normalized trace of the first block on both sides of (10.23), one can recover the free rectangular addition formula $\mathcal{C}_{\mu_S}^{(\alpha)}(u) + \mathcal{C}_{\mu_Z}^{(\alpha)}(u) = \mathcal{C}_{\mu_Y}^{(\alpha)}(u)$ for $u = \frac{1}{z^2}T^{(\alpha)}\Big(\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big)\Big)$ (see Appendix 10.B).



(a) Computation of the rescaled overlap. Both $\boldsymbol{S}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$, and $N/M = 0.25$. The simulation results are average of 1000 experiments with fixed $\boldsymbol{S}$, and $N = 1000, M = 4000$. Some of the simulation points are dropped for clarity.

Overlap and optimal singular values

From the lower-left block of (10.23), we get:

$$\begin{aligned}\boldsymbol{r}_j^\intercal \langle \boldsymbol{G}_{\mathcal{Y}}(z) \rangle \boldsymbol{l}_j &= \boldsymbol{s}_j^{(r)\intercal} \boldsymbol{G}_{S^\intercal S}\big((z-\zeta_2^*)(z-\zeta_1^*)\big)\boldsymbol{S}^\intercal \boldsymbol{s}_j^{(l)} \\ &= \frac{\sigma_j}{(z-\zeta_2^*)(z-\zeta_1^*) - {\sigma_j}^2}\end{aligned}$$

and using (10.21), we find:

$$O(\gamma, \sigma) \approx \frac{1}{\pi \bar{\mu}_Y(\gamma)} \lim_{z \to \gamma - \mathrm{i}0^+} \mathrm{Im}\, \frac{\sigma}{(z-\zeta_2^*)(z-\zeta_1^*) - \sigma^2} \tag{10.25}$$

where σ is in the support of the limiting singular value distribution of $\boldsymbol{S}$, μ_S. In Fig. 10.4.1a we illustrate on an example that theoretical predictions for the overlaps from (10.25) are in good agreement with numerical simulations.

The optimal estimator for singular values reads:

$$\begin{aligned}\widehat{\xi_j^*} = \frac{1}{N}\sum_{k=1}^N \sigma_k O(\gamma_j, \sigma_k) &\approx \frac{1}{N\pi\bar{\mu}_Y(\gamma_j)} \lim_{z \to \gamma_j - \mathrm{i}0^+} \mathrm{Im} \sum_{k=1}^N \frac{\sigma_k^2}{(z-\zeta_2^*)(z-\zeta_1^*) - \sigma_k^2} \\ &= \frac{1}{N\pi\bar{\mu}_Y(\gamma_j)} \lim_{z \to \gamma_j - \mathrm{i}0^+} \mathrm{Im}\, \mathrm{Tr}\, \boldsymbol{S}\boldsymbol{G}_{S^\intercal S}\big((z-\zeta_2^*)(z-\zeta_1^*)\big)\boldsymbol{S}^\intercal\end{aligned} \tag{10.26}$$

Comparing the left-upper blocks in the first and second lines of (10.23) we find

$$\boldsymbol{S}\boldsymbol{G}_{S^\intercal S}\big((z-\zeta_2^*)(z-\zeta_1^*)\big)\boldsymbol{S}^\intercal = \Big\langle -\frac{\zeta_1^*}{z}\boldsymbol{I}_N + \big(1-\frac{\zeta_1^*}{z}\big)\boldsymbol{Y}\boldsymbol{G}_{Y^\intercal Y}(z^2)\boldsymbol{Y}^\intercal \Big\rangle \tag{10.27}$$

The trace of the r.h.s of (10.27) is (with multiplication by $1/N$)

$$\begin{aligned}\frac{1}{N}\sum_{k=1}^N \Big[\frac{{\gamma_k}^2}{z^2-{\gamma_k}^2}\big(1-\frac{\zeta_1^*}{z}\big) - \frac{\zeta_1^*}{z}\Big] &= -\frac{\zeta_1^*}{z}\frac{1}{N}\sum_{k=1}^N \Big[\frac{{\gamma_k}^2}{z^2-{\gamma_k}^2}+1\Big] + \frac{1}{N}\sum_{k=1}^N \frac{{\gamma_k}^2}{z^2-{\gamma_k}^2} \\ &\approx -\zeta_1^* z \mathcal{G}_{\rho_Y}(z^2) + \mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big) \\ &= -\zeta_1^* z \mathcal{G}_{\rho_Y}(z^2) + z^2 \mathcal{G}_{\rho_Y}(z^2) - 1\end{aligned}$$

The last expression on the r.h.s can be expressed in terms of the symmetrized limiting spectral distribution of $\boldsymbol{Y}$. Indeed if we denote the Stieltjes of $\bar{\mu}_Y$ by $\mathcal{G}_{\bar{\mu}_Y}(z)$, using the relation $z\mathcal{G}_{\rho_Y}(z^2) = \mathcal{G}_{\bar{\mu}_Y}(z)$, the above trace implies with (10.27):

$$\frac{1}{N}\operatorname{Tr}\boldsymbol{S}\boldsymbol{G}_{S^\intercal S}\big((z-\zeta_2^*)(z-\zeta_1^*)\big)\boldsymbol{S} \approx -\zeta_1^*\mathcal{G}_{\bar{\mu}_Y}(z) + z\mathcal{G}_{\bar{\mu}_Y}(z) - 1$$

Moreover ζ_1^* in (10.24) can be written as,

$$\zeta_1^* = \frac{1}{\mathcal{G}_{\bar{\mu}_Y}(z)}\mathcal{C}_{\mu_Z}^{(\alpha)}\left(\frac{1}{z}\mathcal{G}_{\bar{\mu}_Y}(z)\Big(1-\alpha+\alpha z\mathcal{G}_{\bar{\mu}_Y}(z)\Big)\right) \tag{10.28}$$

Replacing these results in (10.26) we easily deduce (10.29) for the optimal singular values of the RIE.

$$\begin{aligned}\widehat{\xi_j^*} &= \frac{1}{\pi\bar{\mu}_Y(\gamma_j)}\operatorname{Im}\Bigg[\gamma_j\mathcal{G}_{\bar{\mu}_Y}(\gamma_j-\mathsf{i}0^+) \\ &\qquad - \mathcal{C}_{\mu_Z}^{(\alpha)}\left(\frac{1}{\gamma_j}\mathcal{G}_{\bar{\mu}_Y}(\gamma_j-\mathsf{i}0^+)\Big(1-\alpha+\alpha\gamma_j\mathcal{G}_{\bar{\mu}_Y}(\gamma_j-\mathsf{i}0^+)\Big)\right)\Bigg] \\ &= \gamma_j - \frac{1}{\pi\bar{\mu}_Y(\gamma_j)}\operatorname{Im}\mathcal{C}_{\mu_Z}^{(\alpha)}\left(\frac{1-\alpha}{\gamma_j}\pi\mathsf{H}[\bar{\mu}_Y](\gamma_j) + \alpha\big(\pi\mathsf{H}[\bar{\mu}_Y](\gamma_j)\big)^2\right. \\ &\qquad \left. - \alpha\big(\pi\bar{\mu}_Y(\gamma_j)\big)^2 + \mathsf{i}\pi\bar{\mu}_Y(\gamma_j)\big(\frac{1-\alpha}{\gamma_j} + 2\alpha\pi\mathsf{H}[\bar{\mu}_Y](\gamma_j)\big)\right)\end{aligned} \tag{10.29}$$

10.4.2 Optimality of oracle estimator

In this section, we show that for rotational invariant priors, the posterior mean estimator belongs to the RIE class. We proceed by presenting an equivalent definition of the RIE and then show that posterior mean estimator satisfies this definition.

Lemma 10.1. *Given the observation matrix $\boldsymbol{Y}$, let $\Theta_S(\boldsymbol{Y})$ be an estimator for $\boldsymbol{S}$. Then $\Theta_S(\boldsymbol{Y})$ is a RIE if and only if for any orthogonal matrices $\boldsymbol{U}\in\mathbb{R}^{N\times N}, \boldsymbol{V}\in\mathbb{R}^{M\times M}$:*

$$\Theta_S(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal) = \boldsymbol{U}\Theta_S(\boldsymbol{Y})\boldsymbol{V}^\intercal \tag{10.30}$$

Proof. If $\Theta_S(\boldsymbol{Y})$ is a RIE, then this property clearly follows from the definition (10.2). Let us now show the converse.

Suppose that an estimator $\Theta_S(\boldsymbol{Y})$ satisfies (10.30). First, we show that if the observation matrix is diagonal, then the estimator is also diagonal. Consider the observation matrix to be $\boldsymbol{Y}^{\text{diag}} = \big[\ \operatorname{diag}(\mathrm{y}_1,\ldots,\mathrm{y}_N)\ \big|\ \boldsymbol{0}_{N\times(M-N)}\ \big]$. Let $\boldsymbol{I}_k^-\in\mathbb{R}^{N\times N}, \boldsymbol{J}_k^-\in\mathbb{R}^{M\times M}$ be diagonal matrices with diagonal entries all one except the k-th entry which is -1. Note that for $1\le k\le N$, we have

$\boldsymbol{Y}^{\text{diag}} = \boldsymbol{I}_k^- \boldsymbol{Y}^{\text{diag}} \boldsymbol{J}_k^-$. Moreover matrices $\boldsymbol{I}_k^-, \boldsymbol{J}_k^-$ are orthogonal thus or any $1 \le k \le N$, from (10.30) we have:

$$\Theta_S(\boldsymbol{Y}^{\text{diag}}) = \Theta_S(\boldsymbol{I}_k^- \boldsymbol{Y}^{\text{diag}} \boldsymbol{J}_k^-) = \boldsymbol{I}_k^- \Theta_S(\boldsymbol{Y}^{\text{diag}}) \boldsymbol{J}_k^- \tag{10.31}$$

This implies that all entries on the k-th row and k-th column of $\Theta_S(\boldsymbol{Y}^{\text{diag}})$ is zero except the k-th entry on the diagonal. Since this holds for any k, we conclude that $\Theta_S(\boldsymbol{Y}^{\text{diag}})$ is diagonal.

Now, for a given general observation matrix $\boldsymbol{Y} = \boldsymbol{U}_Y \boldsymbol{\Gamma} \boldsymbol{V}_Y^\intercal$, put $\boldsymbol{U} = \boldsymbol{U}_Y^\intercal, \boldsymbol{V} = \boldsymbol{V}_Y^\intercal$ in the property (10.30). We have:

$$\Theta_S(\boldsymbol{\Gamma}) = \boldsymbol{U}_Y^\intercal \Theta_S(\boldsymbol{Y}) \boldsymbol{V}_Y$$

From the argument above, the matrix on the l.h.s is diagonal. Consequently, the matrix $\boldsymbol{U}_Y^\intercal \Theta_S(\boldsymbol{Y}) \boldsymbol{V}_Y$ is diagonal which implies that the columns of $\boldsymbol{U}_Y, \boldsymbol{V}_Y$ are the left and right singular vectors of $\Theta_S(\boldsymbol{Y})$. Therefore, $\Theta_S(\boldsymbol{Y})$ is a RIE. □

Now, we prove that the posterior mean estimator $\Theta_S^*(\boldsymbol{Y}) = \mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}]$ has the property (10.30), and thus belongs to the RIE class. For simplicity, we drop the SNR factor $\sqrt{\kappa}$. For any orthogonal matrices $\boldsymbol{U} \in \mathbb{R}^{N \times N}, \boldsymbol{V} \in \mathbb{R}^{M \times M}$, we have:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{S}|\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal] &= \frac{\int d\tilde{\boldsymbol{S}}\, \tilde{\boldsymbol{S}}\, P_S(\tilde{\boldsymbol{S}}) P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \tilde{\boldsymbol{S}})}{\int d\tilde{\boldsymbol{S}}\, P_S(\tilde{\boldsymbol{S}}) P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \tilde{\boldsymbol{S}})} \\
&\overset{\text{(a)}}{=} \frac{\int\, d\tilde{\boldsymbol{S}}\, \boldsymbol{U}\tilde{\boldsymbol{S}}\boldsymbol{V}^\intercal\, P_S(\tilde{\boldsymbol{S}}) P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{U}\tilde{\boldsymbol{S}}\boldsymbol{V}^\intercal)}{\int\, d\tilde{\boldsymbol{S}}\, P_S(\tilde{\boldsymbol{S}}) P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{U}\tilde{\boldsymbol{S}}\boldsymbol{V}^\intercal)} \\
&\overset{\text{(b)}}{=} \boldsymbol{U} \Big\{ \frac{\int d\tilde{\boldsymbol{S}}\, \tilde{\boldsymbol{S}}\, P_S(\tilde{\boldsymbol{S}}) P_Z(\boldsymbol{Y} - \tilde{\boldsymbol{S}})}{\int d\tilde{\boldsymbol{S}}\, P_S(\tilde{\boldsymbol{S}}) P_Z(\boldsymbol{Y} - \tilde{\boldsymbol{S}})} \Big\} \boldsymbol{V}^\intercal \\
&= \boldsymbol{U}\mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}]\boldsymbol{V}^\intercal
\end{aligned}$$

where in (a), we changed variables $\tilde{\boldsymbol{S}} \to \boldsymbol{U}\tilde{\boldsymbol{S}}\boldsymbol{V}^\intercal$, used $|\det \boldsymbol{U}| = |\det \boldsymbol{V}| = 1$, and bi-rotational invariance of P_S, $P_S(\tilde{\boldsymbol{S}}) = P_S(\boldsymbol{U}\tilde{\boldsymbol{S}}\boldsymbol{V}^\intercal)$. In (b), we used the bi-rotational invariance property of P_Z, namely $P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{U}\tilde{\boldsymbol{S}}\boldsymbol{V}^\intercal) = P_Z(\boldsymbol{Y} - \tilde{\boldsymbol{S}})$.

10.4.3 Proof of Proposition 10.2

Define the two measures:

$$\nu := \frac{1}{2N} \sum_{j=1}^{N} \boldsymbol{u}_j^\intercal \boldsymbol{S} \boldsymbol{v}_j \big(\delta_{\gamma_j} - \delta_{-\gamma_j}\big), \qquad \tau := \frac{1}{2N} \sum_{j=1}^{N} \big(\delta_{\gamma_j} + \delta_{-\gamma_j}\big)$$

Using the Stieltjes inversion formula, for any $\epsilon > 0$ such that $[\gamma_j - \epsilon, \gamma_j + \epsilon] \cap \{\gamma_1, \cdots, \gamma_N\} = \{\gamma_j\}$, the optimal singular value ξ_j^* can be expressed as:

$$\xi_j^* = \lim_{\eta \to 0} \frac{\int_{\gamma_j - \epsilon}^{\gamma_j + \epsilon} \operatorname{Im} \mathcal{G}_\nu(x + \mathrm{i}\eta)\, dx}{\int_{\gamma_j - \epsilon}^{\gamma_j + \epsilon} \operatorname{Im} \big\{\mathcal{G}_\tau(x + \mathrm{i}\eta)\big\}\, dx} \tag{10.32}$$

with $\mathcal{G}_\nu, \mathcal{G}_\tau$ the Stieltjes transforms of ν, τ. The first Stiltjes transform can be written as:

$$\begin{aligned}
\mathcal{G}_\nu(z) &= \frac{1}{2N}\sum_{j=1}^{N} \boldsymbol{u}_j^\intercal \boldsymbol{S}\boldsymbol{v}_j \big(\frac{1}{z-\gamma_j} - \frac{1}{z+\gamma_j}\big) \\
&= \frac{1}{N}\sum_{j=1}^{N} \frac{\gamma_j}{z^2-\gamma_j^2} \boldsymbol{u}_j^\intercal \boldsymbol{S}\boldsymbol{v}_j \\
&= \frac{1}{N}\sum_{j=1}^{N} \frac{\gamma_j}{z^2-\gamma_j^2} \operatorname{Tr} \boldsymbol{v}_j\boldsymbol{u}_j^\intercal \boldsymbol{S} \\
&= \frac{1}{N}\sum_{j=1}^{N} \frac{\gamma_j}{z^2-\gamma_j^2} \operatorname{Tr} \boldsymbol{S}^\intercal\boldsymbol{u}_j\boldsymbol{v}_j^\intercal \\
&= \frac{1}{N} \operatorname{Tr} \boldsymbol{S}^\intercal \sum_{j=1}^{N} \frac{\gamma_j}{z^2-\gamma_j^2} \boldsymbol{u}_j\boldsymbol{v}_j^\intercal \\
&= \frac{1}{N} \operatorname{Tr} \boldsymbol{S}^\intercal \big(z^2\boldsymbol{I} - \boldsymbol{Y}\boldsymbol{Y}^\intercal\big)^{-1}\boldsymbol{Y} \\
&= \frac{1}{N} \operatorname{Tr} \boldsymbol{G}_{\boldsymbol{Y}\boldsymbol{Y}^\intercal}(z^2)\boldsymbol{Y}\boldsymbol{S}^\intercal \\
&= L(z)
\end{aligned} \tag{10.33}$$

Similarly, we get

$$\mathcal{G}_\tau(z) = zG(z) \tag{10.34}$$

Finally (10.8) follows from (10.32), (10.33), (10.34). □

10.4.4 Proof of Theorem 10.3

For simplicity of notation, we drop the z-dependence of the random functions $G(z), L(z)$. Let

$$g := \mathbb{E}\, G, \qquad l := \mathbb{E}\, L$$

where the expectation is over the noise matrix $\boldsymbol{Z}$ in (10.1). We will need the following lemma whose proof is deferred to subsection 10.4.4.

Lemma 10.2. *There is a numerical constant $c > 0$ (depending on K) such that for any $z \in \mathbb{C}\backslash\mathbb{R}$ and for any $t > 0$, we have:*

$$\mathbf{P}\big(|G-g| \geq t\big) \leq 2e^{-c\left(tN\left|\operatorname{Im} z\right|^3\right)^2}$$

The same is also true for L, l.

Consider the decomposition

$$L = l + (L - l)$$

from the lemma above, $L - l$ is a sub-Gaussian random variable with sub-Gaussian norm $O\Big(\frac{1}{N\big|\operatorname{Im} z\big|^3}\Big)$. Therefore, to prove the Theorem, it suffices to show that

$$l = g\big(z^2 - \frac{1}{\alpha_0}\big) - g\big(z^2 g - 1\big) - 1 + O\Big(\frac{1}{N\big|\operatorname{Im} z\big|^3}\Big) \tag{10.35}$$

Let $\boldsymbol{G} := \boldsymbol{G}_{YY^\intercal}(z^2)$. We start by expanding the following matrix products:

$$\boldsymbol{G}\boldsymbol{Y}\boldsymbol{Y}^\intercal = \kappa \boldsymbol{G}\boldsymbol{S}\boldsymbol{S}^\intercal + \sqrt{\kappa}\boldsymbol{G}\boldsymbol{S}\boldsymbol{Z}^\intercal + \sqrt{\kappa}\boldsymbol{G}\boldsymbol{Z}\boldsymbol{S}^\intercal + \boldsymbol{G}\boldsymbol{Z}\boldsymbol{Z}^\intercal$$

$$\boldsymbol{G}\boldsymbol{Y}\boldsymbol{S}^\intercal = \sqrt{\kappa}\boldsymbol{G}\boldsymbol{S}\boldsymbol{S}^\intercal + \boldsymbol{G}\boldsymbol{Z}\boldsymbol{S}^\intercal$$

Using the identity $z^2\boldsymbol{G} - \boldsymbol{I} = \boldsymbol{G}\boldsymbol{Y}\boldsymbol{Y}^\intercal$, we have:

$$\boldsymbol{G}\boldsymbol{Y}\boldsymbol{S}^\intercal = \frac{1}{\sqrt{\kappa}}\Big(z^2\boldsymbol{G} - \boldsymbol{I} - \boldsymbol{G}\boldsymbol{Z}\boldsymbol{Z}^\intercal\Big) - \boldsymbol{G}\boldsymbol{S}\boldsymbol{Z}^\intercal \tag{10.36}$$

Taking expectation and trace of the both sides:

$$\mathbb{E}\ \operatorname{Tr}\boldsymbol{G}\boldsymbol{Y}\boldsymbol{S}^\intercal = \frac{1}{\sqrt{\kappa}}z^2\mathbb{E}\ \operatorname{Tr}\boldsymbol{G} - \frac{1}{\sqrt{\kappa}}N - \mathbb{E}\ \operatorname{Tr}\boldsymbol{G}\boldsymbol{S}\boldsymbol{Z}^\intercal - \frac{1}{\sqrt{\kappa}}\mathbb{E}\operatorname{Tr}\ \boldsymbol{G}\boldsymbol{Z}\boldsymbol{Z}^\intercal \tag{10.37}$$

The next step is to compute the last two terms in (10.37) through a use of gaussian integration by parts.

• Expansion of $\mathbb{E}\ \operatorname{Tr}\boldsymbol{G}\boldsymbol{S}\boldsymbol{Z}^\intercal$: Using cyclicity of the trace and the fact that $\boldsymbol{G}$ is symmetric, we have:

$$\mathbb{E}\ \operatorname{Tr}\boldsymbol{G}\boldsymbol{S}\boldsymbol{Z}^\intercal = \mathbb{E}\ \operatorname{Tr}\boldsymbol{G}\boldsymbol{Z}\boldsymbol{S}^\intercal = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{N}\mathbb{E}\,G_{ik}Z_{kj}\,S_{ij} \tag{10.38}$$

Gaussian integration by parts yields:

$$\mathbb{E}\,G_{ik}Z_{kj} = \frac{1}{N}\mathbb{E}\Big(\frac{\partial \boldsymbol{G}}{\partial Z_{kj}}\Big)_{ik} \tag{10.39}$$

We have:

$$\frac{\partial \boldsymbol{G}}{\partial Z_{ab}} = -\boldsymbol{G}\frac{\partial\big(z^2\boldsymbol{I} - \boldsymbol{Y}\boldsymbol{Y}^\intercal\big)}{\partial Z_{ab}}\boldsymbol{G} = \boldsymbol{G}\big(\boldsymbol{Y}\boldsymbol{J}^{ab\intercal} + \boldsymbol{J}^{ab}\boldsymbol{Y}^\intercal\big)\boldsymbol{G}$$

with $\boldsymbol{J}^{ab} \in \mathbb{R}^{N\times M}$, $J^{ab}_{ij} = \delta\{i = a, j = b\}$. Thus, we find:

$$\begin{aligned}
\Big(\frac{\partial \boldsymbol{G}}{\partial Z_{kj}}\Big)_{ik} &= \big[\boldsymbol{G}\boldsymbol{Y}\boldsymbol{J}^{kj\intercal}\boldsymbol{G}\big]_{ik} + \big[\boldsymbol{G}\boldsymbol{J}^{kj}\boldsymbol{Y}^\intercal\boldsymbol{G}\big]_{ik} \\
&= \sum_{a,b,c} G_{ia}\,Y_{ab}\,J^{kj}_{cb}\,G_{ck} + \sum_{a,b,c} G_{ia}\,J^{kj}_{ab}\,Y_{cb}\,G_{ck} \\
&= \sum_{a} G_{ia}\,Y_{aj}\,G_{kk} + \sum_{c} G_{ik}\,Y_{cj}\,G_{ck} \\
&= G_{kk}\,\big(\boldsymbol{G}\boldsymbol{Y}\big)_{ij} + G_{ik}\,(\boldsymbol{G}\boldsymbol{Y})_{kj}
\end{aligned} \tag{10.40}$$

Joining (10.40) with (10.39), (10.38) can be written further to be:

$$\begin{aligned}\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}\boldsymbol{S}\boldsymbol{Z}^\intercal &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{N}\Big[G_{kk}\left(\boldsymbol{G}\boldsymbol{Y}\right)_{ij} + G_{ik}\left(\boldsymbol{G}\boldsymbol{Y}\right)_{kj}\Big]S_{ij} \\ &= \frac{1}{N}\mathbb{E}\Big[\big(\mathrm{Tr}\,\boldsymbol{G}\big)\big(\mathrm{Tr}\,\boldsymbol{G}\boldsymbol{Y}\boldsymbol{S}^\intercal\big)\Big] + \frac{1}{N}\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}\boldsymbol{G}\boldsymbol{Y}\boldsymbol{S}^\intercal\end{aligned} \tag{10.41}$$

- Expansion of $\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}\boldsymbol{Z}\boldsymbol{Z}^\intercal$:

$$\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}\boldsymbol{Z}\boldsymbol{Z}^\intercal = \sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{N}\mathbb{E}\,G_{ik}Z_{kj}\,Z_{ij} \tag{10.42}$$

Using again gaussian integration by parts:

$$\begin{aligned}\mathbb{E}\,G_{ik}Z_{kj}Z_{ij} &= \frac{1}{N}\mathbb{E}\,\frac{\partial G_{ik}Z_{ij}}{\partial Z_{kj}} \\ &= \frac{1}{N}\mathbb{E}\,Z_{ij}\Big(\frac{\partial \boldsymbol{G}}{\partial Z_{kj}}\Big)_{ik} + \delta\{i=k\}\frac{1}{N}\mathbb{E}\,G_{ik} \\ &= \frac{1}{N}\mathbb{E}\,Z_{ij}\Big(G_{kk}\left(\boldsymbol{G}\boldsymbol{Y}\right)_{ij} + G_{ik}\left(\boldsymbol{G}\boldsymbol{Y}\right)_{kj}\Big) + \delta\{i=k\}\frac{1}{N}\mathbb{E}\,G_{ik}\end{aligned} \tag{10.43}$$

Plugging in (10.42), we find

$$\begin{aligned}&\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}\boldsymbol{Z}\boldsymbol{Z}^\intercal \\ &= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{N}\mathbb{E}\,Z_{ij}G_{kk}\left(\boldsymbol{G}\boldsymbol{Y}\right)_{ij} + \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{M}\sum_{k=1}^{N}\mathbb{E}\,Z_{ij}G_{ik}\left(\boldsymbol{G}\boldsymbol{Y}\right)_{kj} \\ &\qquad + \frac{1}{N}\sum_{i,j,k}\delta\{k=i\}\mathbb{E}\,G_{ik} \\ &= \frac{1}{N}\mathbb{E}\Big[\big(\mathrm{Tr}\,\boldsymbol{G}\big)\big(\mathrm{Tr}\,\boldsymbol{G}\boldsymbol{Y}\boldsymbol{Z}^\intercal\big)\Big] + \frac{1}{N}\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}\boldsymbol{G}\boldsymbol{Y}\boldsymbol{Z}^\intercal + \frac{M}{N}\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}\end{aligned} \tag{10.44}$$

Replacing (10.41) and (10.44) in (10.37), we find:

$$
\begin{aligned}
&\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}\boldsymbol{Y}\boldsymbol{S}^\intercal \\
&= \frac{1}{\sqrt{\kappa}} z^2 \mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G} - \frac{1}{N}\mathbb{E}\Big[\big(\mathrm{Tr}\,\boldsymbol{G}\big)\big(\mathrm{Tr}\,\boldsymbol{G}\boldsymbol{Y}\boldsymbol{S}^\intercal\big)\Big] - \frac{1}{\sqrt{\kappa}}\frac{1}{N}\mathbb{E}\Big[\big(\mathrm{Tr}\,\boldsymbol{G}\big)\big(\mathrm{Tr}\,\boldsymbol{G}\boldsymbol{Y}\boldsymbol{Z}^\intercal\big)\Big] \\
&\qquad - \frac{1}{N}\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}\boldsymbol{G}\boldsymbol{Y}\boldsymbol{S}^\intercal - \frac{1}{\sqrt{\kappa}}\frac{1}{N}\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}\boldsymbol{G}\boldsymbol{Y}\boldsymbol{Z}^\intercal - \frac{1}{\sqrt{\kappa}}\frac{M}{N}\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G} - \frac{1}{\sqrt{\kappa}}N \\
&= \frac{1}{\sqrt{\kappa}}\big(z^2 - \frac{1}{\alpha_0}\big)\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G} - \frac{1}{\sqrt{\kappa}}\frac{1}{N}\mathbb{E}\Big[\big(\mathrm{Tr}\,\boldsymbol{G}\big)\big(\mathrm{Tr}\,\boldsymbol{G}\boldsymbol{Y}\boldsymbol{Y}^\intercal\big)\Big] \\
&\qquad\qquad - \frac{1}{\sqrt{\kappa}}\frac{1}{N}\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}\boldsymbol{G}\boldsymbol{Y}\boldsymbol{Y}^\intercal - \frac{1}{\sqrt{\kappa}}N \\
&= \frac{1}{\sqrt{\kappa}}\big(z^2 - \frac{1}{\alpha_0}\big)\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G} - \frac{1}{\sqrt{\kappa}}\frac{1}{N}\mathbb{E}\Big[\big(\mathrm{Tr}\,\boldsymbol{G}\big)\big(z^2\,\mathrm{Tr}\,\boldsymbol{G} - N\big)\Big] \\
&\qquad\qquad - \frac{1}{\sqrt{\kappa}}\frac{1}{N}\mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}\big(z^2\boldsymbol{G} - \boldsymbol{I}\big) - \frac{1}{\sqrt{\kappa}}N
\end{aligned}
\tag{10.45}
$$

Dividing by N and rearranging terms we find:

$$
l = \frac{1}{\sqrt{\kappa}}\big(z^2 + 1 - \frac{1}{\alpha_0}\big)g - \frac{1}{\sqrt{\kappa}} z^2 \mathbb{E}\, G^2 - 1 - \frac{1}{N^2} z^2 \mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}^2 - \frac{1}{N} g \tag{10.46}
$$

Using lemma 10.2,

$$
\begin{aligned}
\mathbb{E}\, G^2 &= g^2 + \mathbb{E}\,\big(g - G\big)^2 + 2g\mathbb{E}\,\big(g - G\big) \\
&= g^2 + O\Big(\frac{1}{N\big|\mathrm{Im}\, z\big|^3}\Big)
\end{aligned}
\tag{10.47}
$$

and,

$$
\begin{aligned}
\big|\frac{1}{N^2} z^2 \mathbb{E}\ \mathrm{Tr}\,\boldsymbol{G}^2\big| &\le \mathbb{E}\big|\frac{1}{N^2} z^2\, \mathrm{Tr}\,\boldsymbol{G}^2\big| \\
&\le \frac{1}{N^2}\mathbb{E}\big|\sum_{k=1}^{N} \frac{z^2}{\big(z^2 - \gamma_k^2\big)^2}\big| \\
&\le \frac{1}{N^2}\mathbb{E}\sum_{k=1}^{N} \big|\frac{z^2}{\big(z^2 - \gamma_k^2\big)^2}\big| \\
&\le \frac{1}{N^2}\mathbb{E}\sum_{k=1}^{N} \frac{1}{\big(\mathrm{Im}\, z\big)^2} = O\Big(\frac{1}{N\big(\mathrm{Im}\, z\big)^2}\Big)
\end{aligned}
\tag{10.48}
$$

Similarly, we have that

$$
\frac{1}{N} g = O\Big(\frac{1}{N\big(\mathrm{Im}\, z\big)^2}\Big) \tag{10.49}
$$

Combining (10.46), (10.47),(10.48), (10.49), we obtain the result:

$$
l = \frac{1}{\sqrt{\kappa}}\Big[\big(z^2 + 1 - \frac{1}{\alpha_0}\big)g - z^2 g^2 - 1\Big] + O\Big(\frac{1}{N\big|\mathrm{Im}\, z\big|^3}\Big)
$$

This completes the proof of Theorem 10.3.

Proof of Lemma 10.2

To prove lemma 10.2, we use a Gaussian concentration inequality:

Theorem 10.6 (Gaussian concentration inequality)**.** *Let $X = (X_1, ..., X_n)$ be a vector of n independent Gaussian random variables of variance σ^2. Let $f : \mathbb{R}^n \to \mathbb{R}$ denote an L-Lipschitz function (w.r.t Euclidean norm in $\mathbb{R}^n$). Then, for any $t > 0$,*

$$\mathbb{P}\big(|f(X) - \mathbb{E}f(X)| \geq t\big) \leq 2e^{-\frac{1}{\sigma^2}\frac{t^2}{2L^2}}$$

that is, $f(X) - \mathbb{E}f(X)$ is sub-Gaussian with sub-Gaussian norm $O\big(\frac{1}{L\sigma}\big)$.

Using the above result, it suffices to show that $G(z), L(z)$ as functions of the noise matrix $\boldsymbol{Z} \in \mathbb{R}^{N\times M}$ are Lipschitz with constant $O\big(\frac{1}{\sqrt{N}\left|\operatorname{Im} z\right|^3}\big)$.

Consider the Hermitization $\boldsymbol{\mathcal{Y}}$ of $\boldsymbol{Y}$ in (10.17). Given the decomposition (10.18), we have:

$$\begin{aligned}
G(z) &= \frac{1}{N}\operatorname{Tr}\left(z^2\boldsymbol{I} - \boldsymbol{Y}\boldsymbol{Y}^\intercal\right)^{-1} \\
&= \frac{1}{N}\sum_{k=1}^{N}\frac{1}{z^2-\gamma_k^2} \\
&= \frac{1}{N}\sum_{k=1}^{N}\frac{1}{2z}\big(\frac{1}{z-\gamma_k} + \frac{1}{z+\gamma_k}\big) \\
&= \frac{1}{N}\frac{1}{2z}\Big[\sum_{k=1}^{N}\big(\frac{1}{z-\gamma_k} + \frac{1}{z+\gamma_k}\big) + (M-N)\frac{1}{z}\Big] - \frac{M-N}{2N}\frac{1}{z^2} \\
&= \frac{1}{N}\frac{1}{2z}\operatorname{Tr}\left(z\boldsymbol{I} - \boldsymbol{\mathcal{Y}}\right)^{-1} - \frac{M-N}{2N}\frac{1}{z^2}
\end{aligned} \tag{10.50}$$

and

$$\begin{aligned}
L(z) &= \frac{1}{N}\operatorname{Tr}\left(z^2\boldsymbol{I} - \boldsymbol{Y}\boldsymbol{Y}^\intercal\right)^{-1}\boldsymbol{Y}\boldsymbol{S}^\intercal \\
&= \frac{1}{N}\operatorname{Tr}\Big(\sum_{k=1}^{N}\frac{1}{z^2-\gamma_k^2}\boldsymbol{u}_k\boldsymbol{u}_k^\intercal\Big)\boldsymbol{Y}\boldsymbol{S}^\intercal \\
&= \frac{1}{N}\sum_{k=1}^{N}\frac{1}{z^2-\gamma_k^2}\operatorname{Tr}\boldsymbol{u}_k\boldsymbol{u}_k^\intercal\boldsymbol{Y}\boldsymbol{S}^\intercal \\
&= \frac{1}{N}\sum_{k=1}^{N}\frac{\gamma_k}{z^2-\gamma_k^2}\operatorname{Tr}\boldsymbol{u}_k\boldsymbol{v}_k^\intercal\boldsymbol{S}^\intercal
\end{aligned} \tag{10.51}$$

On the other hand, letting $\boldsymbol{\mathcal{S}}$ be the Hertmitization of $\boldsymbol{S}$, and $\boldsymbol{w}_k$ the k-th column of $\boldsymbol{W}$ in (10.18), we have:

$$\boldsymbol{w}_k^\intercal \boldsymbol{\mathcal{S}} \boldsymbol{w}_k = \begin{cases} \boldsymbol{u}_k^\intercal \boldsymbol{S} \boldsymbol{v}_k & \text{for } 1 \leq \mathrm{k} \leq \mathrm{N} \\ -\boldsymbol{u}_{k-N}^\intercal \boldsymbol{S} \boldsymbol{v}_{k-N} & N+1 \leq k \leq 2N \\ 0 & 2N+1 \leq k \leq M+N \end{cases}$$

Therefore, denoting eigenvalues of $\boldsymbol{\mathcal{Y}}$ by $\tilde{\gamma_k}$, we find

$$\begin{aligned} \frac{1}{N} \operatorname{Tr}(z\boldsymbol{I} - \boldsymbol{\mathcal{Y}})^{-1} \boldsymbol{\mathcal{S}} &= \frac{1}{N} \sum_{k=1}^{M+N} \frac{1}{z - \tilde{\gamma_k}} \operatorname{Tr} \boldsymbol{w}_k \boldsymbol{w}_k^\intercal \boldsymbol{\mathcal{S}} \\ &= \frac{1}{N} \Big[\sum_{k=1}^{N} \frac{1}{z - \gamma_k} \operatorname{Tr} \boldsymbol{u}_k \boldsymbol{v}_k^\intercal \boldsymbol{S} - \sum_{k=1}^{N} \frac{1}{z + \gamma_k} \operatorname{Tr} \boldsymbol{u}_k \boldsymbol{v}_k^\intercal \boldsymbol{S} \Big] \\ &= \frac{1}{N} \sum_{k=1}^{N} \big(\frac{1}{z - \gamma_k} - \frac{1}{z + \gamma_k} \big) \operatorname{Tr} \boldsymbol{u}_k \boldsymbol{v}_k^\intercal \boldsymbol{S}^\intercal \\ &= \frac{2}{N} \sum_{k=1}^{N} \frac{\gamma_k}{z^2 - \gamma_k^2} \operatorname{Tr} \boldsymbol{u}_k \boldsymbol{v}_k^\intercal \boldsymbol{S}^\intercal \end{aligned} \tag{10.52}$$

From (10.51), (10.52), we obtain:

$$L(z) = \frac{1}{2N} \operatorname{Tr}(z\boldsymbol{I} - \boldsymbol{\mathcal{Y}})^{-1} \boldsymbol{\mathcal{S}} \tag{10.53}$$

Now, to show Lipschitz continuity of the functions we consider a variation of the noise matrix $\boldsymbol{Z} \to \boldsymbol{Z} + \delta_Z$. From now on, variables evaluated at $\boldsymbol{Z} + \delta_Z$ are denoted with a "tilde" symbol, for example:

$$\tilde{G}(z) = \frac{1}{N} \operatorname{Tr} \left(z^2 \boldsymbol{I} - \tilde{\boldsymbol{Y}} \tilde{\boldsymbol{Y}}^\intercal \right)^{-1} = \frac{1}{N} \operatorname{Tr} \left(z^2 \boldsymbol{I} - (\boldsymbol{S} + \boldsymbol{Z} + \delta_Z)(\boldsymbol{S} + \boldsymbol{Z} + \delta_Z)^\intercal \right)^{-1}$$

We have:

$$\begin{aligned} |G(z) - \tilde{G}(z)| &= \frac{1}{N} \left| \operatorname{Tr} \left[\left(z^2 \boldsymbol{I} - \boldsymbol{Y} \boldsymbol{Y}^\intercal \right)^{-1} - \left(z^2 \boldsymbol{I} - \tilde{\boldsymbol{Y}} \tilde{\boldsymbol{Y}}^\intercal \right)^{-1} \right] \right| \\ &\overset{\text{(a)}}{=} \frac{1}{N} \frac{1}{2|z|} \left| \operatorname{Tr} \left[\left(z\boldsymbol{I} - \boldsymbol{\mathcal{Y}} \right)^{-1} - \left(z\boldsymbol{I} - \tilde{\boldsymbol{\mathcal{Y}}} \right)^{-1} \right] \right| \\ &\overset{\text{(b)}}{=} \frac{1}{N} \frac{1}{2|z|} \left| \operatorname{Tr} \left(z\boldsymbol{I} - \boldsymbol{\mathcal{Y}} \right)^{-1} (\tilde{\boldsymbol{\mathcal{Y}}} - \boldsymbol{\mathcal{Y}}) \left(z\boldsymbol{I} - \tilde{\boldsymbol{\mathcal{Y}}} \right)^{-1} \right| \\ &\overset{\text{(c)}}{\leq} \frac{\sqrt{M+N}}{N} \frac{1}{2|z|} \left\| \left(z\boldsymbol{I} - \boldsymbol{\mathcal{Y}} \right)^{-1} (\tilde{\boldsymbol{\mathcal{Y}}} - \boldsymbol{\mathcal{Y}}) \left(z\boldsymbol{I} - \tilde{\boldsymbol{\mathcal{Y}}} \right)^{-1} \right\|_{\mathrm{F}} \\ &\overset{\text{(d)}}{\leq} \frac{\sqrt{M+N}}{N} \frac{1}{2|z|} \left\| \left(z\boldsymbol{I} - \boldsymbol{\mathcal{Y}} \right)^{-1} \right\|_{\mathrm{op}} \left\| \left(z\boldsymbol{I} - \tilde{\boldsymbol{\mathcal{Y}}} \right)^{-1} \right\|_{\mathrm{op}} \left\| (\tilde{\boldsymbol{\mathcal{Y}}} - \boldsymbol{\mathcal{Y}}) \right\|_{\mathrm{F}} \\ &\overset{\text{(e)}}{\leq} C \frac{1}{\sqrt{N}} \frac{1}{\left| \operatorname{Im} z \right|^3} \left\| \delta_Z \right\|_{\mathrm{F}} \end{aligned} \tag{10.54}$$

where in (a) we use the identity in (10.50), in (b) we use the following resolvent formula, namely that for any square matrices $\boldsymbol{A}, \boldsymbol{B}$:

$$\big(z\boldsymbol{I}-\boldsymbol{A}\big)^{-1}-\big(z\boldsymbol{I}-\boldsymbol{B}\big)^{-1}=\big(z\boldsymbol{I}-\boldsymbol{A}\big)^{-1}\big(\boldsymbol{A}-\boldsymbol{B}\big)\big(z\boldsymbol{I}-\boldsymbol{B}\big)^{-1} \quad ,$$

in (c) we use the inequality for any matrix $\boldsymbol{A}\in\mathbb{R}^{N\times N}$:

$$\big|\operatorname{Tr}\boldsymbol{A}\big|\leq\sqrt{N}\|\boldsymbol{A}\|_{\mathrm{F}},$$

in (d) we use a non-commutative Hölder-type inequality (see e.g. [134], Thm 2.8), namely for any product $\boldsymbol{A}_1\cdots\boldsymbol{A}_k$ of matrices with any size and any $i=1,...,k$,

$$\big\|\boldsymbol{A}_1\cdots\boldsymbol{A}_k\big\|_{\mathrm{F}}\leq\prod_{j\neq i}\|\boldsymbol{A}_j\|_{\mathrm{op}}\ \|\boldsymbol{A}_i\|_{\mathrm{F}},$$

and finally in (e) the constant C depends only on K and we use that the operator norm of $\big(z\boldsymbol{I}-\tilde{\boldsymbol{\mathcal{Y}}}\big)^{-1}$ is bounded by $\operatorname{Im} z$:

$$\Big\|\big(z\boldsymbol{I}-\tilde{\boldsymbol{\mathcal{Y}}}\big)^{-1}\Big\|_{\mathrm{op}}=\max_{\{0,\pm\gamma_1,\cdots,\pm\gamma_k\}}\frac{1}{|z-x|}\leq\frac{1}{|z-\operatorname{Re} z|}=\frac{1}{\big|\operatorname{Im} z\big|}$$

Similarly for $L(z)$, we have:

$$\begin{aligned}
|L(z)-\tilde{L}(z)|&=\frac{1}{2N}\Big|\operatorname{Tr}\Big[\big(z\boldsymbol{I}-\boldsymbol{\mathcal{Y}}\big)^{-1}\boldsymbol{\mathcal{S}}-\big(z\boldsymbol{I}-\tilde{\boldsymbol{\mathcal{Y}}}\big)^{-1}\boldsymbol{\mathcal{S}}\Big]\Big|\\
&=\frac{1}{2N}\Big|\operatorname{Tr}\Big[\big(z\boldsymbol{I}-\boldsymbol{\mathcal{Y}}\big)^{-1}-\big(z\boldsymbol{I}-\tilde{\boldsymbol{\mathcal{Y}}}\big)^{-1}\Big]\boldsymbol{\mathcal{S}}\ \Big|\\
&\leq\frac{1}{2N}\Big\|\big(z\boldsymbol{I}-\boldsymbol{\mathcal{Y}}\big)^{-1}-\big(z\boldsymbol{I}-\tilde{\boldsymbol{\mathcal{Y}}}\big)^{-1}\Big\|_{\mathrm{F}}\|\boldsymbol{\mathcal{S}}\|_{\mathrm{F}} \qquad (10.55)\\
&\leq\frac{\sqrt{2N}}{2N}\Big\|\big(z\boldsymbol{I}-\boldsymbol{\mathcal{Y}}\big)^{-1}-\big(z\boldsymbol{I}-\tilde{\boldsymbol{\mathcal{Y}}}\big)^{-1}\Big\|_{\mathrm{F}}\|\boldsymbol{S}\|_{\mathrm{op}}\\
&\leq\frac{C'}{\sqrt{N}}\frac{1}{\big(\operatorname{Im} z\big)^2}\big\|\delta_Z\big\|_{\mathrm{F}}
\end{aligned}$$

with C' a positive constant depending only on K.

10.4.5 Computation of MMSE for the Gaussian noise - Statement 10.4

From (10.6) and (10.12) we see that to compute the MMSE we must compute the following expectation:

$$\int\left(x-\frac{1-\alpha}{\alpha}\frac{1}{x}-2\pi\mathsf{H}[\bar{\mu}_Y](x)\right)^2\mu_Y(x)\,dx$$

In the following, using properties of the Hilbert transform, we show this integral equals:

$$\int x^2\mu_Y(x)\,dx+\big(\frac{1}{\alpha}-1\big)^2\int\frac{\mu_Y(x)}{x^2}dx+\frac{\pi^2}{3}\int\mu_Y(x)^3\,dx-\frac{2}{\alpha} \qquad (10.56)$$

Putting these relations together, we deduce (for Gaussian noise):

$$\int \xi^*(x)^2 \mu_Y(x)\,dx = \int x^2 \mu_S(x)\,dx \\ - \frac{1}{\kappa}\Big[\frac{1}{\alpha} - \big(\frac{1}{\alpha}-1\big)^2 \int \frac{\mu_Y(x)}{x^2}\,dx - \frac{\pi^2}{3}\int \mu_Y(x)^3\,dx\Big]$$

Replacing this identity in (10.6), we get (10.13).

Derivation of (10.56)

For simplicity we denote $\mathsf{H}[\bar{\mu}_Y](x)$ by $\bar{\mathsf{H}}(x)$. Expanding the square in the integrand, we find

$$x^2 + \big(\frac{1-\alpha}{\alpha}\big)^2 \frac{1}{x^2} - 2\frac{1-\alpha}{\alpha} + 4\pi^2\big(\bar{\mathsf{H}}(x)\big)^2 - 4\pi x \bar{\mathsf{H}}(x) + 4\pi \frac{1-\alpha}{\alpha}\frac{\bar{\mathsf{H}}(x)}{x} \quad (10.57)$$

To compute the expectation of the last three terms, we need the following properties of the Hilbert transform.

Lemma 10.3. *If $f : \mathbb{R} \to \mathbb{R}$ is compactly supported and sufficiently regular, then one has the identities*

$$\int_{\mathbb{R}} f(x)\big(\mathsf{H}[f](x)\big)^2\,dx = \frac{1}{3}\int_{\mathbb{R}} f^3(x)\,dx \quad (10.58)$$

$$\int_{\mathbb{R}} \mathsf{H}[f](x) x f(x)\,dx = \frac{1}{2\pi}\Big(\int_{\mathbb{R}} f(x)\,dx\Big)^2 \quad (10.59)$$

$$\int_{\mathbb{R}} \frac{\mathsf{H}[f](x)}{x} f(x)\,dx = -\frac{1}{2\pi}\Big(\int_{\mathbb{R}} \frac{f(x)}{x}\,dx\Big)^2 \quad (10.60)$$

Proof. The proof of the first two properties can be found in Lemma 3.1 of [121]. To prove the last equality, we apply the same idea as in remark 3.2 of this paper to write:

$$\begin{aligned}\int_{\mathbb{R}} \frac{\mathsf{H}[f](x)}{x} f(x)\,dx &= \frac{1}{2\pi}\iint \big(\frac{1}{x}-\frac{1}{y}\big)\frac{1}{x-y} f(x)f(y)\,dx\,dy \\ &= -\frac{1}{2\pi}\iint \frac{1}{xy} f(x)f(y)\,dx\,dy \\ &= -\frac{1}{2\pi}\Big(\int \frac{f(x)}{x}\,dx\Big)^2\end{aligned}$$

□

We remark that the Hilbert transform of an even function is an odd function (see e.g. [135]), in other words for the symmetrized measure $\bar{\mu}_Y$ we have $\bar{\mathsf{H}}(x) = -\bar{\mathsf{H}}(x)$. From (10.58) we have:

$$\int \big(\bar{\mathsf{H}}(x)\big)^2 \bar{\mu}_Y(x)\,dx = \frac{1}{3}\int \bar{\mu}_Y(x)^3\,dx \quad (10.61)$$

The l.h.s can be written as:

$$\begin{aligned}\frac{1}{2}\int_{\mathbb{R}_+}\big(\bar{\mathsf{H}}(x)\big)^2\mu_Y(x)\,dx + \frac{1}{2}\int_{\mathbb{R}_-}\big(\bar{\mathsf{H}}(x)\big)^2\mu_Y(-x)\,dx \\ &= \frac{1}{2}\int_{\mathbb{R}_+}\big(\bar{\mathsf{H}}(x)\big)^2\mu_Y(x)\,dx + \frac{1}{2}\int_{\mathbb{R}_+}\big(\bar{\mathsf{H}}(-x)\big)^2\mu_Y(x)\,dx \\ &= \frac{1}{2}\int_{\mathbb{R}_+}\big(\bar{\mathsf{H}}(x)\big)^2\mu_Y(x)\,dx + \frac{1}{2}\int_{\mathbb{R}_+}\big(\bar{\mathsf{H}}(x)\big)^2\mu_Y(x)\,dx \\ &= \int_{\mathbb{R}_+}\big(\bar{\mathsf{H}}(x)\big)^2\mu_Y(x)\,dx\end{aligned}$$

The rhs in (10.61) equals $\frac{1}{12}\int \mu_Y(x)^3\,dx$. Therefore, the expectation of the fourth term in (10.57) is:

$$4\pi^2\int\big(\bar{\mathsf{H}}(x)\big)^2\mu_Y(x)\,dx = \frac{\pi^2}{3}\int\mu_Y(x)^3\,dx \tag{10.62}$$

Similarly, using symmetry properties of $\bar{H}$ and $\bar{\mu}_Y$ we have that:

$$\int x\bar{H}(x)\bar{\mu}_Y(x)\,dx = \int x\bar{H}(x)\mu_Y(x)\,dx$$

Thus, by (10.59), the expectation of the fifth term in (10.57) is:

$$-4\pi\int x\bar{H}(x)\mu_Y(x)\,dx = -2\Big(\int_{\mathbb{R}}\bar{\mu}_Y(x)\,dx\Big)^2 = -2. \tag{10.63}$$

Again, by symmetry we have:

$$\int\frac{\bar{H}(x)}{x}\bar{\mu}_Y(x)\,dx = \int\frac{\bar{H}(x)}{x}\mu_Y(x)\,dx$$

Thus, by (10.60), the expectation of the last term in (10.57) is:

$$\int\frac{\bar{H}(x)}{x}\mu_Y(x)\,dx = \Big(\int_{\mathbb{R}}\frac{\bar{\mu}_Y(x)}{x}\,dx\Big)^2 = 0 \tag{10.64}$$

where we used that $\frac{\bar{\mu}_Y(x)}{x}$ is an odd function.

Finally putting together (10.57), (10.62), (10.63), (10.64) we get (10.56).

10.4.6 Proof of Theorem 10.5

We start from the posterior distribution of the model (10.1) which reads (up to some constants):

$$\begin{aligned}P(\boldsymbol{X}|\boldsymbol{Y}) &\propto e^{-\frac{N}{2}\|\boldsymbol{Y}-\sqrt{\kappa}\boldsymbol{X}\|_{\mathrm{F}}^2}P_S(\boldsymbol{X}) \\ &\propto e^{N\,\mathrm{Tr}\left[\sqrt{\kappa}\boldsymbol{X}\boldsymbol{Y}^\intercal-\frac{\kappa}{2}\boldsymbol{X}\boldsymbol{X}^\intercal\right]}P_S(\boldsymbol{X})\end{aligned} \tag{10.65}$$

The partition function is defined as the normalizing factor of the posterior distribution (10.65):

$$Z(\boldsymbol{Y}) = \int d\boldsymbol{X}\, e^{N \operatorname{Tr}\left[\sqrt{\kappa}\boldsymbol{X}\boldsymbol{Y}^\intercal - \frac{\kappa}{2}\boldsymbol{X}\boldsymbol{X}^\intercal\right]} P_S(\boldsymbol{X}) \tag{10.66}$$

and the free energy is defined as:

$$F_N(\kappa) = -\frac{1}{MN}\mathbb{E}_Y\big[\ln Z(\boldsymbol{Y})\big] \tag{10.67}$$

One can easily see that the free energy is linked to the (average) mutual information via the relation:

$$\frac{1}{MN} I_N(\boldsymbol{S};\boldsymbol{Y}) = F_N(\kappa) + \frac{\kappa}{2M}\mathbb{E}\big[\operatorname{Tr}\boldsymbol{S}\boldsymbol{S}^\intercal\big]$$

in which $\frac{1}{M}\mathbb{E}\big[\operatorname{Tr}\boldsymbol{S}^\intercal\boldsymbol{S}\big]$ converges to the second moment of μ_S rescaled by the factor α. Therefore, to prove theorem 10.5, it is enough to show that

$$\lim_{N\to\infty} F_N(\kappa) = \frac{\kappa}{2}\alpha \int x^2 \mu_S(x)\, dx - \mathcal{J}^{(\alpha)}[\mu_{\sqrt{\kappa}S}, \mu_{\sqrt{\kappa}S} \boxplus_\alpha \mu_{\mathrm{MP}}]$$

To prove this limit, first, we show that this limit also holds for the free energy of a simpler model. Then, using the *pseudo-Lipschitz* continuity of the free energy w.r.t. to a distance between two models which converges to 0 as $N \to \infty$, we deduce that the same limit holds for the free energy of the original model.

An independent singular value model

Suppose $\boldsymbol{\sigma}^0 \in \mathbb{R}^N$ is generated with i.i.d. elements from μ_S, and is ordered in non-decreasing way. Fix $\boldsymbol{\sigma}^0$ once for all. Let $\tilde{\boldsymbol{S}} \in \mathbb{R}^{N\times M}$ the matrix contructed as $\boldsymbol{U}\tilde{\boldsymbol{\Sigma}}\boldsymbol{V}^\intercal$ where $\boldsymbol{U} \in \mathbb{R}^{N\times N}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$ are independent and distributed according to the Haar measure, and $\tilde{\boldsymbol{\Sigma}} \in \mathbb{R}^{N\times M}$ with $\tilde{\boldsymbol{\sigma}}$ on its main diagonal for $\tilde{\boldsymbol{\sigma}} \in \mathbb{R}^N$. The distribution of the matrix $\tilde{\boldsymbol{S}}$ is :

$$dP_{\tilde{S}}(\tilde{\boldsymbol{S}}) = d\mu_N(\boldsymbol{U})\, d\mu_M(\boldsymbol{V}) dp_{\tilde{S}}(\tilde{\boldsymbol{\sigma}}) = d\mu_N(\boldsymbol{U})\, d\mu_M(\boldsymbol{V}) \prod_{i=1}^{N} \delta(\tilde{\sigma}_i - \sigma_i^0)\, d\tilde{\boldsymbol{\sigma}} \tag{10.68}$$

Matrix $\tilde{\boldsymbol{S}}$ is observed through an AWGN channel as in (10.1), $\tilde{\boldsymbol{Y}} = \sqrt{\kappa}\tilde{\boldsymbol{S}} + \tilde{\boldsymbol{Z}}$. The partition function and the free energy can be defined in the same way as in (10.66),(10.67) denoted by $\tilde{Z}(\tilde{\boldsymbol{Y}})$, $\tilde{F}_N(\kappa)$ respectively.

Proposition 10.7. *For μ_S with compact support, and any $\kappa > 0$, we have μ_S-almost surely*

$$\lim_{N\to\infty} \tilde{F}_N(\kappa) = \frac{\kappa}{2}\alpha \int x^2 \mu_S(x)\, dx - \mathcal{J}^{(\alpha)}[\mu_{\sqrt{\kappa}S}, \mu_{\sqrt{\kappa}S} \boxplus_\alpha \mu_{\mathrm{MP}}]$$

Proof Appendix 10.C.1.

Pseudo-Lipschitz continuity of the free energy

Consider two bi-rotationally invariant matrix ensembles $P^{(1)}$, $P^{(2)}$, i.e. for $\boldsymbol{S} \sim P^{(1)}(\boldsymbol{S})$, $\tilde{\boldsymbol{S}} \sim P^{(2)}(\tilde{\boldsymbol{S}})$ with SVDs $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\intercal$, $\tilde{\boldsymbol{S}} = \tilde{\boldsymbol{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{V}}^\intercal$

$$dP_N^{(1)}(\boldsymbol{S}) \propto d\mu_N(\boldsymbol{U})\, d\mu_M(\boldsymbol{V})\, p^{(1)}(\boldsymbol{\sigma})\, d\boldsymbol{\sigma}$$
$$dP_N^{(2)}(\tilde{\boldsymbol{S}}) \propto d\mu_N(\tilde{\boldsymbol{U}})\, d\mu_M(\tilde{\boldsymbol{V}})\, p^{(2)}(\tilde{\boldsymbol{\sigma}})\, d\tilde{\boldsymbol{\sigma}}$$

where $p^{(1)}(\boldsymbol{\sigma})$, $p^{(2)}(\tilde{\boldsymbol{\sigma}})$ are the joint probability density functions for the singular values, induced by the priors. Suppose each of these distributions to be the prior of an inference problem in model (10.1). The free energy can be defined similarly for each of the priors, which are denoted by $F_N^{(1)}(\kappa), F_N^{(2)}(\kappa)$ respectively. Then, we have

Proposition 10.8. *For all $\kappa > 0$ and N :*

$$\left|F_N^{(1)}(\kappa) - F_N^{(2)}(\kappa)\right| \leq \frac{\kappa}{2N}\left(\sqrt{\mathbb{E}_{\boldsymbol{\sigma}}\left[\|\boldsymbol{\sigma}\|^2\right]} + \sqrt{\mathbb{E}_{\tilde{\boldsymbol{\sigma}}}\left[\|\tilde{\boldsymbol{\sigma}}\|^2\right]}\right)\sqrt{\mathbb{E}_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}}\left[\|\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}\|^2\right]} \tag{10.69}$$

Proof Appendix 10.C.2.

The distance between two models

Recall that

$$dP_S(\boldsymbol{S}) \propto d\mu_N(\boldsymbol{U})\, d\mu_M(\boldsymbol{V}) p_S(\boldsymbol{\sigma})\, d\boldsymbol{\sigma}$$

where $p_S(\boldsymbol{\sigma})$ is the joint p.d.f. of singular values of $\boldsymbol{S}$. Moreover, $dP_{\tilde{S}}(\tilde{\boldsymbol{S}})$ is defined in (10.68) with $p_{\tilde{S}}(\tilde{\boldsymbol{\sigma}}) \equiv \prod_{i=1}^N \delta(\tilde{\sigma}_i - \sigma_i^0)$, where $\boldsymbol{\sigma}^0$ is generated with i.i.d. elements from μ_S

Lemma 10.4. *Under assumptions 10.2, 10.3, for $\boldsymbol{\sigma} \sim p_S(\boldsymbol{\sigma})$, $\tilde{\boldsymbol{\sigma}} \sim p_{\tilde{S}}(\tilde{\boldsymbol{\sigma}})$, we have:*

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}}\left[\|\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}\|^2\right] = 0 \tag{10.70}$$

Proof Appendix 10.C.3.

Concluding the proof

By proposition 10.8, the distance between the free energies $F_N(\kappa)$ (defined in (10.67)) and $\tilde{F}_N(\kappa)$ is upper bounded by rhs in (10.69). The term $\frac{1}{N}\|\boldsymbol{\sigma}\|^2 = \frac{1}{N}\sum \sigma_i^2$ is the second moment of the empirical spectral distribution of $\boldsymbol{S}$, which is almost surely bounded by assumption 10.3. So, $\frac{1}{N}\mathbb{E}_{\boldsymbol{\sigma}}\left[\|\boldsymbol{\sigma}\|^2\right]$ is bounded. Moreover, $\frac{1}{N}\mathbb{E}\left[\|\tilde{\boldsymbol{\sigma}}\|^2\right] = \frac{1}{N}\sum {\sigma^0}_i^2$ which is bounded by C_2^2. By proposition 10.4, $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}}\left[\|\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}\|^2\right] = 0$. Therefore $\lim_{N\to\infty} |F_N(\kappa) - \tilde{F}_N(\kappa)| = 0$ and Proposition 10.7 gives the result. □

Appendix

10.A Derivation of the Resolvent Relation

From (10.22), we have

$$\begin{aligned}\boldsymbol{\mathcal{Y}} &= \begin{bmatrix} \mathbf{0} & \boldsymbol{S} \\ \boldsymbol{S}^{\intercal} & \mathbf{0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{U} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V} \end{bmatrix} \begin{bmatrix} \mathbf{0} & \boldsymbol{Z} \\ \boldsymbol{Z}^{\intercal} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}^{\intercal} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{V}^{\intercal} \end{bmatrix} \\ &= \boldsymbol{\mathcal{S}} + \boldsymbol{O}\boldsymbol{\mathcal{Z}}\boldsymbol{O}^{\intercal}\end{aligned} \tag{10.71}$$

Let $\boldsymbol{G}(z) \equiv \boldsymbol{G}_{\mathcal{Y}}(z)$. First, we express the entries of $\boldsymbol{G}(z)$ using the Gaussian integral representation of an inverse matrix [83]:

$$\begin{aligned}G_{ij}(z) &= \sqrt{\frac{1}{(2\pi)^{N+M}\det(z\boldsymbol{I}-\boldsymbol{\mathcal{Y}})}} \int \Big(\prod_{k=1}^{M+N} d\eta_k \Big)\, \eta_i \eta_j \, \exp\Big\{ -\frac{1}{2}\boldsymbol{\eta}^{\intercal}(z\boldsymbol{I}-\boldsymbol{\mathcal{Y}})\boldsymbol{\eta} \Big\} \\ &= \frac{\displaystyle\int \Big(\prod_{k=1}^{M+N} d\eta_k \Big)\, \eta_i \eta_j \, \exp\Big\{ -\frac{1}{2}\boldsymbol{\eta}^{\intercal}(z\boldsymbol{I}-\boldsymbol{\mathcal{Y}})\boldsymbol{\eta} \Big\}}{\displaystyle\int \Big(\prod_{k=1}^{M+N} d\eta_k \Big) \, \exp\Big\{ -\frac{1}{2}\boldsymbol{\eta}^{\intercal}(z\boldsymbol{I}-\boldsymbol{\mathcal{Y}})\boldsymbol{\eta} \Big\}}\end{aligned} \tag{10.72}$$

For z not close to the real axis, the resolvent is expected to exhibit self-averaging behavior in the limit of large N, meaning that it will not depend on the particular matrix realization. Thus, we can examine the resolvent $\boldsymbol{G}_{\mathcal{Y}}(z)$ by analyzing its ensemble average, denoted by $\langle . \rangle$ in the following.

$$\big\langle G_{ij}(z) \big\rangle = \Big\langle \frac{1}{\mathcal{Z}} \int \Big(\prod_{k=1}^{M+N} d\eta_k \Big)\, \eta_i \eta_j \, \exp\Big\{ -\frac{1}{2}\boldsymbol{\eta}^{\intercal}(z\boldsymbol{I}-\boldsymbol{\mathcal{Y}})\boldsymbol{\eta} \Big\} \Big\rangle \tag{10.73}$$

where $\mathcal{Z}$ is the denominator in (10.72). Computing the average is, in general, non-trivial. However, the replica method provides us with a technique to

overcome this issue by employing the following identity:

$$\begin{aligned}\langle G_{ij}(z)\rangle &= \lim_{n\to 0}\Big\langle \mathcal{Z}^{n-1}\int\Big(\prod_{k=1}^{M+N} d\eta_k\Big)\,\eta_i\eta_j\,\exp\Big\{-\frac{1}{2}\boldsymbol{\eta}^\intercal(z\boldsymbol{I}-\boldsymbol{\mathcal{Y}})\boldsymbol{\eta}\Big\}\Big\rangle \\ &= \lim_{n\to 0}\Big\langle \int\Big(\prod_{k=1}^{M+N}\prod_{\tau=1}^{n} d\eta_k^{(\tau)}\Big)\,\eta_i^{(1)}\eta_j^{(1)}\,\exp\Big\{-\frac{1}{2}\sum_{\tau=1}^{n}\boldsymbol{\eta}^{(\tau)\intercal}(z\boldsymbol{I}-\boldsymbol{\mathcal{Y}})\boldsymbol{\eta}^{(\tau)}\Big\}\Big\rangle\end{aligned} \tag{10.74}$$

For the expression in the exponent, we have:

$$\begin{aligned}-\frac{1}{2}\sum_{\tau=1}^{n}\sum_{k,l=1}^{M+N}\eta_k^{(\tau)}\big(z\delta_{kl}-\boldsymbol{\mathcal{Y}}_{kl}\big)\eta_l^{(\tau)} &= -\frac{1}{2}\sum_{\tau=1}^{n}\sum_{k,l=1}^{M+N}\eta_k^{(\tau)}\big(z\delta_{kl}-\boldsymbol{\mathcal{S}}_{kl}\big)\eta_l^{(\tau)} \\ &\quad+\frac{1}{2}\sum_{\tau=1}^{n}\sum_{k,l=1}^{M+N}\eta_k^{(\tau)}(\boldsymbol{O}\boldsymbol{\mathcal{Z}}\boldsymbol{O}^\intercal)_{kl}\eta_l^{(\tau)}\end{aligned} \tag{10.75}$$

The first term in the RHS can be written as

$$-\frac{1}{2}\sum_{\tau=1}^{n}\boldsymbol{\eta}^{(\tau)\intercal}\big(z\boldsymbol{I}_{N+M}-\boldsymbol{\mathcal{S}}\big)\boldsymbol{\eta}^{(\tau)} \tag{10.76}$$

Given the structure (10.71) for $\boldsymbol{O}\boldsymbol{\mathcal{Z}}\boldsymbol{O}^\intercal$, the second sum in (10.75) can be written as:

$$\sum_{k=1}^{N}\sum_{l=N+1}^{M+N}\eta_k^{(\tau)}\big(\boldsymbol{U}\boldsymbol{Z}\boldsymbol{V}^\intercal\big)_{k,l-N}\eta_l^{(\tau)}+\sum_{k=N+1}^{M+N}\sum_{l=1}^{N}\eta_k^{(\tau)}\big(\boldsymbol{V}\boldsymbol{Z}^\intercal\boldsymbol{U}^\intercal\big)_{k-N,l}\eta_l^{(\tau)} \tag{10.77}$$

Split each replica $\boldsymbol{\eta}^{(\tau)}$ into two vectors $\boldsymbol{a}^{(\tau)}\in\mathbb{R}^N, \boldsymbol{b}^{(\tau)}\in\mathbb{R}^M$, $\boldsymbol{\eta}^{(\tau)}=\begin{bmatrix}\boldsymbol{a}^{(\tau)}\\ \boldsymbol{b}^{(\tau)}\end{bmatrix}$. The expression in (10.77) can be rewritten as $2\operatorname{Tr}\boldsymbol{b}^{(\tau)}\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{U}\boldsymbol{Z}\boldsymbol{V}^\intercal$. So, we have (dropping the limit term for brevity):

$$\begin{aligned}\langle G_{ij}(z)\rangle = \int\Big(\prod_{k=1}^{M+N}\prod_{\tau=1}^{n} d\eta_k^{(\tau)}\Big)\,\eta_i^{(1)}\eta_j^{(1)}\exp\Big\{-\frac{1}{2}\sum_{\tau=1}^{n}\boldsymbol{\eta}^{(\tau)\intercal}\big(z\boldsymbol{I}_{N+M}-\boldsymbol{\mathcal{S}}\big)\boldsymbol{\eta}^{(\tau)}\Big\} \\ \times\Big\langle\exp\Big\{\sum_{\tau=1}^{n}\operatorname{Tr}\boldsymbol{b}^{(\tau)}\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{U}\boldsymbol{Z}\boldsymbol{V}^\intercal\Big\}\Big\rangle_{\boldsymbol{U},\boldsymbol{V}}\end{aligned} \tag{10.78}$$

Using the formula for the rectangular spherical integral [64] (see (6.4)) for the last term we find:

$$\begin{aligned}&\Big\langle\exp\Big\{\sum_{\tau=1}^{n}\operatorname{Tr}\boldsymbol{b}^{(\tau)}\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{U}\boldsymbol{Z}\boldsymbol{V}^\intercal\Big\}\Big\rangle_{\boldsymbol{U},\boldsymbol{V}} \\ &\qquad\approx\exp\Big\{\frac{N}{2}\sum_{\tau=1}^{n}\mathcal{Q}_{\mu_Z}\big(\frac{1}{NM}\|\boldsymbol{a}^{(\tau)}\|^2\|\boldsymbol{b}^{(\tau)}\|^2\big)\Big\}\end{aligned}$$

where we used that for each replica, the non-zero singular value of $\boldsymbol{b}^{(\tau)}\boldsymbol{a}^{(\tau)\mathsf{T}}$ is $\|\boldsymbol{a}^{(\tau)}\|\|\boldsymbol{b}^{(\tau)}\|$.

Therefore, we find

$$\langle G_{ij}(z)\rangle = \int \Big(\prod_{k=1}^{M+N} \prod_{\tau=1}^{n} d\eta_k^{(\tau)} \Big)\, \eta_i^{(1)}\eta_j^{(1)} \times \exp\Big\{ \sum_{\tau=1}^{n} \Big[-\frac{1}{2}\boldsymbol{\eta}^{(\tau)\mathsf{T}}\big(z\boldsymbol{I}-\boldsymbol{\mathcal{S}}\big)\boldsymbol{\eta}^{(\tau)} + \frac{N}{2}\mathcal{Q}_{\mu_Z}\big(\frac{1}{NM}\|\boldsymbol{a}^{(\tau)}\|^2\|\boldsymbol{b}^{(\tau)}\|^2\big)\Big]\Big\} \tag{10.79}$$

Introducing delta functions $\delta\big(p_1^{(\tau)} - \frac{1}{N}\|\boldsymbol{a}^{(\tau)}\|^2\big)$, $\delta\big(p_2^{(\tau)} - \frac{1}{M}\|\boldsymbol{b}^{(\tau)}\|^2\big)$, (10.79) can be written as:

$$\begin{aligned}\langle G_{ij}(z)\rangle = \int \Big(\prod_{k=1}^{M+N} \prod_{\tau=1}^{n} d\eta_k^{(\tau)} \Big) \Big(\prod_{\tau=1}^{n} dp_1^{(\tau)}\, dp_2^{(\tau)} \Big)\, \eta_i^{(1)}\eta_j^{(1)} \\ \times \prod_{\tau=1}^{n} \delta\big(p_1^{(\tau)} - \frac{1}{N}\|\boldsymbol{a}^{(\tau)}\|^2\big)\delta\big(p_2^{(\tau)} - \frac{1}{M}\|\boldsymbol{b}^{(\tau)}\|^2\big) \\ \times \exp\Big\{ \sum_{\tau=1}^{n} \Big[-\frac{1}{2}\boldsymbol{\eta}^{(\tau)\mathsf{T}}\big(z\boldsymbol{I}-\boldsymbol{\mathcal{S}}\big)\boldsymbol{\eta}^{(\tau)} + \frac{N}{2}\mathcal{Q}_{\mu_Z}\big(p_1^{(\tau)}p_2^{(\tau)}\big)\Big]\Big\}\end{aligned} \tag{10.80}$$

In the next step, we replace each delta with its Fourier transform $\delta\big(p_1^{(\tau)} - \frac{1}{N}\|\boldsymbol{a}^{(\tau)}\|^2\big) \propto \int d\zeta_1^{(\tau)} \exp\Big\{ -\frac{N}{2}\zeta_1^{(\tau)}\big(p_1^{(\tau)} - \frac{1}{N}\|\boldsymbol{a}^{(\tau)}\|^2\big)\Big\}$. After rearranging, we find:

$$\begin{aligned}\langle G_{ij}(z)\rangle \propto \int \Big(\prod dp_1^{(\tau)}\, dp_2^{(\tau)}\, d\zeta_1^{(\tau)}\, d\zeta_2^{(\tau)} \Big) \\ \times \exp\Big\{ \frac{N}{2}\sum_{\tau=1}^{n} \big[\mathcal{Q}_{\mu_Z}(p_1^{(\tau)}p_2^{(\tau)}) - \zeta_1^{(\tau)}p_1^{(\tau)} - \frac{1}{\alpha}\zeta_2^{(\tau)}p_2^{(\tau)}\big]\Big\} \\ \times \int \Big(\prod_{k=1}^{M+N} \prod_{\tau=1}^{n} d\eta_k^{(\tau)} \Big)\, \eta_i^{(1)}\eta_j^{(1)} \\ \times \exp\Big\{ \sum_{\tau=1}^{n} \Big[-\frac{1}{2}\boldsymbol{\eta}^{(\tau)\mathsf{T}}\big(z\boldsymbol{I}-\boldsymbol{\mathcal{S}}\big)\boldsymbol{\eta}^{(\tau)} + \frac{1}{2}\zeta_1^{(\tau)}\|\boldsymbol{a}^{(\tau)}\|^2 + \frac{1}{2}\zeta_2^{(\tau)}\|\boldsymbol{b}^{(\tau)}\|^2\Big]\Big\}\end{aligned} \tag{10.81}$$

The second integral in (10.81) is a Gaussian integral with matrix

$$\boldsymbol{M}^{(\tau)} = \begin{bmatrix} (z-\zeta_1^{(\tau)})\boldsymbol{I}_N & -\boldsymbol{S} \\ -\boldsymbol{S}^{\mathsf{T}} & (z-\zeta_2^{(\tau)})\boldsymbol{I}_M \end{bmatrix} \tag{10.82}$$

Using the formula for determinant of block matrices, we have

$$\begin{aligned}\det \boldsymbol{M}^{(\tau)} &= \det\left[(z-\zeta_1^{(\tau)})\boldsymbol{I}_N - (z-\zeta_2^{(\tau)})^{-1}\boldsymbol{S}\boldsymbol{S}^\intercal\right]\det\left[(z-\zeta_2^{(\tau)})\boldsymbol{I}_M\right]\\ &= (z-\zeta_2^{(\tau)})^{M-N}\prod_{k=1}^{N}\left[(z-\zeta_1^{(\tau)})(z-\zeta_2^{(\tau)}) - {\sigma_k}^2\right]\end{aligned}$$

Except for the first replica, the Gaussian integral is (up to constants):

$$\exp\left\{-\frac{1}{2}\Big[(M-N)\ln(z-\zeta_2^{(\tau)}) + \sum_{k=1}^{N}\ln\left\{(z-\zeta_1^{(\tau)})(z-\zeta_2^{(\tau)}) - {\sigma_k}^2\right\}\Big]\right\}$$

And, the integral for the first replica is the above expression multiplied by $\big(\boldsymbol{M}^{(1)^{-1}}\big)_{ij}$. By Proposition 2.8.7 [136]), $\boldsymbol{M}^{(1)^{-1}}$ can be written as

$$\boldsymbol{M}^{(1)^{-1}} = \begin{bmatrix}\boldsymbol{A} & \boldsymbol{B}\\ \boldsymbol{B}^\intercal & \boldsymbol{C}\end{bmatrix} \tag{10.83}$$

with blocks given as:

$$\begin{aligned}\boldsymbol{A} &= \frac{1}{z-\zeta_1^{(1)}}\boldsymbol{I}_N + \frac{1}{z-\zeta_1^{(1)}}\boldsymbol{S}\boldsymbol{G}_{S^\intercal S}\big((z-\zeta_2^{(1)})(z-\zeta_1^{(1)})\big)\boldsymbol{S}^\intercal\\ \boldsymbol{B} &= \boldsymbol{S}\boldsymbol{G}_{S^\intercal S}\big((z-\zeta_2^{(1)})(z-\zeta_1^{(1)})\big)\\ \boldsymbol{C} &= (z-\zeta_1^{(1)})\boldsymbol{G}_{S^\intercal S}\big((z-\zeta_2^{(1)})(z-\zeta_1^{(1)})\big)\end{aligned}$$

with $\boldsymbol{G}_{S^\intercal S}$ the resolvent of the matrix $\boldsymbol{S}^\intercal\boldsymbol{S}$.

Putting all this together, the integral in (10.81) can be written as

$$\begin{aligned}\langle G_{ij}(z)\rangle \propto \int\left(\prod_{\tau=1}^{n} dp_1^{(\tau)}\, dp_2^{(\tau)}\, d\zeta_1^{(\tau)}\, d\zeta_2^{(\tau)}\right)\big(\boldsymbol{M}^{(1)^{-1}}\big)_{ij}\\ \times\exp\left\{-\frac{Nn}{2}F_0\big(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2\big)\right\}\end{aligned} \tag{10.84}$$

with

$$\begin{aligned}&F_0\big(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{\zeta}_1,\boldsymbol{\zeta}_2\big)\\ &\quad= \frac{1}{n}\sum_{\tau=1}^{n}\Big[\frac{1}{N}\sum_{k=1}^{N}\ln\left\{(z-\zeta_1^{(\tau)})(z-\zeta_2^{(\tau)}) - {\sigma_k}^2\right\} - \big(1-\frac{1}{\alpha}\big)\ln(z-\zeta_2^{(\tau)})\\ &\qquad\qquad - \mathcal{Q}_{\mu_Z}(p_1^{(\tau)}p_2^{(\tau)}) + \zeta_1^{(\tau)}p_1^{(\tau)} + \frac{1}{\alpha}\zeta_2^{(\tau)}p_2^{(\tau)}\Big]\end{aligned}$$

In the large N limit, the integral in (10.84) can be computed using the saddle-points of the function F_0. In the evaluation of this integral, we use the *replica symmetric* ansatz that assumes a saddle-point of the form:

$$\forall\tau\in\{1,\cdots,n\}:\quad p_1^{(\tau)} = p_1,\quad p_2^{(\tau)} = p_2,\quad \zeta_1^{(\tau)} = \zeta_1,\quad \zeta_2^{(\tau)} = \zeta_2$$

One finds that the extremum of the function is then at:

$$\begin{cases} p_1^* = (z-\zeta_2^*)\mathcal{G}_{\rho_S}\big((z-\zeta_1^*)(z-\zeta_2^*)\big) \\ p_2^* = (1-\alpha)\frac{1}{z-\zeta_2^*} + \alpha(z-\zeta_1^*)\mathcal{G}_{\rho_S}\big((z-\zeta_1^*)(z-\zeta_2^*)\big) \\ \zeta_1^* = \frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(p_1^* p_2^*)}{p_1^*} \\ \zeta_2^* = \alpha\frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(p_1^* p_2^*)}{p_2^*} \end{cases} \tag{10.85}$$

where $\mathcal{G}_{\rho_S}$ is the Stieltjes transform of the matrix $\boldsymbol{S}\boldsymbol{S}^\intercal$, whose limiting eigenvalue distribution is the squared transform of the limiting singular value distribution of $\boldsymbol{S}$.

To simplify the solution, we compute the normalized trace of both sides in (10.84). First on the r.h.s we compute the trace of the matrix $\boldsymbol{M}^{-1}$ in (10.83) plugging ζ_1^*, ζ_2^*. The trace of the first block is:

$$\begin{aligned} \frac{1}{N}\frac{1}{z-\zeta_1^*}\sum_{k=1}^{N}\Big[1+\frac{{\sigma_k}^2}{(z-\zeta_2^*)(z-\zeta_1^*)-{\sigma_k}^2}\Big] & \\ = \frac{1}{N}(z-\zeta_2^*)\sum_{k=1}^{N}\frac{1}{(z-\zeta_2^*)(z-\zeta_1^*)-{\sigma_k}^2} & \\ \approx (z-\zeta_2^*)\mathcal{G}_{\rho_S}\big((z-\zeta_2^*)(z-\zeta_1^*)\big) & \\ = p_1^* & \end{aligned} \tag{10.86}$$

Similarly, the trace of the last block can be computed to be p_b^*.

The matrix in the l.h.s of (10.84) is $\boldsymbol{G}_{\mathcal{Y}}(z)$, which has the blocks

$$\begin{aligned} \boldsymbol{G}_{\mathcal{Y}}(z) &= \big(z\boldsymbol{I}-\boldsymbol{\mathcal{Y}}\big)^{-1} \\ &= \begin{bmatrix} z^{-1}\boldsymbol{I}_N + z^{-1}\boldsymbol{Y}\boldsymbol{G}_{Y^\intercal Y}(z^2)\boldsymbol{Y}^\intercal & \boldsymbol{Y}\boldsymbol{G}_{Y^\intercal Y}(z^2) \\ \boldsymbol{G}_{Y^\intercal Y}(z^2)\boldsymbol{Y}^\intercal & z\boldsymbol{G}_{Y^\intercal Y}(z^2) \end{bmatrix} \end{aligned} \tag{10.87}$$

The trace of the first block is:

$$\begin{aligned} \frac{1}{N}\frac{1}{z}\sum_{k=1}^{N}\big[1+\frac{{\gamma_k}^2}{z^2-{\gamma_k}^2}\big] = \frac{1}{N}z\sum_{k=1}^{N}\frac{1}{z^2-{\gamma_k}^2} & \\ \approx z\mathcal{G}_{\rho_Y}(z^2) & \end{aligned} \tag{10.88}$$

Therefore, from (10.86), we find $p_1^* = z\mathcal{G}_{\rho_Y}(z^2)$. The trace of the last block can be evaluated to be $\alpha z\mathcal{G}_{\rho_Y}(z^2) + (1-\alpha)\frac{1}{z}$. So, $p_2^* = \alpha z\mathcal{G}_{\rho_Y}(z^2) + (1-\alpha)\frac{1}{z}$.

Thus we find

$$\begin{cases} p_1^* = z\mathcal{G}_{\rho_Y}(z^2) = \frac{1}{z}\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big) + \frac{1}{z} \\ p_2^* = \alpha z\mathcal{G}_{\rho_Y}(z^2) + (1-\alpha)\frac{1}{z} = \alpha\frac{1}{z}\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big) + \frac{1}{z} \end{cases} \tag{10.89}$$

and

$$p_1^* p_2^* = \frac{1}{z^2}T^{(\alpha)}\Big(\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big)\Big),$$

which implies

$$\zeta_1^* = z\frac{\mathcal{C}_{\mu_Z}^{(\alpha)}\Big(\frac{1}{z^2}T^{(\alpha)}\Big(\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big)\Big)\Big)}{\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big)+1}, \quad \zeta_2^* = \alpha z\frac{\mathcal{C}_{\mu_Z}^{(\alpha)}\Big(\frac{1}{z^2}T^{(\alpha)}\Big(\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big)\Big)\Big)}{\alpha\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big)+1} \tag{10.90}$$

10.B Derivation of Rectangular Free Convolution

Consider the normalized trace of the first block on each side in (10.23). The trace of the first block of the lhs is computed in (10.88) which is $\frac{1}{z}\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big)+\frac{1}{z}$. The trace of the first block in rhs is computed in (10.86) which is $(z-\zeta_2^*)\mathcal{G}_{\rho_S}\big((z-\zeta_2^*)(z-\zeta_1^*)\big)$.

$$\begin{aligned}\frac{1}{z}\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big)+\frac{1}{z} &= (z-\zeta_2^*)\mathcal{G}_{\rho_S}\big((z-\zeta_2^*)(z-\zeta_1^*)\big)\\ &= (z-\zeta_2^*)\frac{1}{(z-\zeta_2^*)(z-\zeta_1^*)}\Big(\mathcal{M}_{\mu_S}\big(\frac{1}{(z-\zeta_2^*)(z-\zeta_1^*)}\big)+1\Big)\\ &= \frac{1}{z-\zeta_1^*}\mathcal{M}_{\mu_S}\big(\frac{1}{(z-\zeta_2^*)(z-\zeta_1^*)}\big)+\frac{1}{z-\zeta_1^*}\end{aligned}$$

From which, we get:

$$(z-\zeta_1^*)\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big)+z-\zeta_1^* = z\mathcal{M}_{\mu_S}\big(\frac{1}{(z-\zeta_2^*)(z-\zeta_1^*)}\big)+z$$

Taking the ζ_1^* to the rhs, and plugging the expression for ζ_1^* from (10.90), after a bit of algebra we find:

$$\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big) = \mathcal{M}_{\mu_S}\big(\frac{1}{(z-\zeta_2^*)(z-\zeta_1^*)}\big)+\mathcal{C}_{\mu_Z}^{(\alpha)}\Big(\frac{1}{z^2}T^{(\alpha)}\Big(\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big)\Big)\Big)$$

Let $\frac{1}{z^2}T^{(\alpha)}\Big(\mathcal{M}_{\mu_Y}\big(\frac{1}{z^2}\big)\Big)=u$. Then, $\frac{1}{z^2}=\mathcal{H}_{\mu_Y}^{(\alpha)^{-1}}(u)$. Moreover, from the definition one can see that $\mathcal{M}_{\mu_Y}\Big(\mathcal{H}_{\mu_Y}^{(\alpha)^{-1}}(u)\Big)=\mathcal{C}_{\mu_Y}^{(\alpha)}(u)$. So, (10.91) can be written as:

$$\mathcal{C}_{\mu_Y}^{(\alpha)}(u) = \mathcal{M}_{\mu_S}\big(\frac{1}{(z-\zeta_2^*)(z-\zeta_1^*)}\big)+\mathcal{C}_{\mu_Z}^{(\alpha)}(u) \tag{10.91}$$

From (10.90),

$$\begin{aligned}(z-\zeta_2^*)(z-\zeta_1^*) &= z^2\Big(1-\frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(u)}{\mathcal{C}_{\mu_Y}^{(\alpha)}(u)+1}\Big)\Big(1-\frac{\alpha\mathcal{C}_{\mu_Z}^{(\alpha)}(u)}{\alpha\mathcal{C}_{\mu_Y}^{(\alpha)}(u)+1}\Big)\\ &= z^2\Big[1-\mathcal{C}_{\mu_Z}^{(\alpha)}(u)\Big(\frac{1}{\mathcal{C}_{\mu_Y}^{(\alpha)}(u)+1}+\frac{\alpha}{\alpha\mathcal{C}_{\mu_Y}^{(\alpha)}(u)+1}\Big)+\frac{\alpha\big(\mathcal{C}_{\mu_Z}^{(\alpha)}(u)\big)^2}{T^{(\alpha)}\big(\mathcal{C}_{\mu_Z}^{(\alpha)}(u)\big)}\Big]\\ &= \frac{z^2}{T^{(\alpha)}\big(\mathcal{C}_{\mu_Y}^{(\alpha)}(u)\big)}T^{(\alpha)}\big(\mathcal{C}_{\mu_Y}^{(\alpha)}(u)-\mathcal{C}_{\mu_Z}^{(\alpha)}(u)\big)\end{aligned}$$

The first factor, using the definition of $\mathcal{C}_{\mu_Y}^{(\alpha)}(u)$, is:

$$\begin{aligned}\frac{z^2}{T^{(\alpha)}\big(\mathcal{C}_{\mu_Y}^{(\alpha)}(u)\big)} &= \frac{1}{\frac{1}{z^2}}\frac{1}{T^{(\alpha)}\big(\mathcal{C}_{\mu_Y}^{(\alpha)}(u)\big)} \\ &= \frac{1}{\mathcal{H}_{\mu_Y}^{(\alpha)^{-1}}(u)}\frac{1}{\frac{u}{\mathcal{H}_{\mu_Y}^{(\alpha)^{-1}}(u)}} = \frac{1}{u}\end{aligned}$$

So, (10.91) can be written as

$$\mathcal{C}_{\mu_Y}^{(\alpha)}(u) - \mathcal{C}_{\mu_Z}^{(\alpha)}(u) = \mathcal{M}_{\mu_S}\Big(\frac{u}{T^{(\alpha)}\big(\mathcal{C}_{\mu_Y}^{(\alpha)}(u) - \mathcal{C}_{\mu_Z}^{(\alpha)}(u)\big)}\Big)$$

One can see that, if the limiting singular value distribution of $\boldsymbol{S}$, is not $\delta(x)$, the unique solution to the equation $\mathcal{M}_{\mu_S}\big(\frac{u}{T^{(\alpha)}(x)}\big) = x$, is $x = \mathcal{C}_{\mu_S}^{(\alpha)}(u)$ (see lemma 4.2 in [64] for a particular case). Therefore, we find:

$$\mathcal{C}_{\mu_Y}^{(\alpha)}(u) - \mathcal{C}_{\mu_Z}^{(\alpha)}(u) = \mathcal{C}_{\mu_S}^{(\alpha)}(u) \tag{10.92}$$

as we expected.

10.C Detailed proof of Theorem 10.5

10.C.1 Proof of proposition 10.7

We start from the partition function,

$$\begin{aligned}\tilde{Z}(\tilde{\boldsymbol{Y}}) &= \int d\boldsymbol{X}\, e^{N\,\mathrm{Tr}\left[\sqrt{\kappa}\boldsymbol{X}^\intercal\tilde{\boldsymbol{Y}} - \frac{\kappa}{2}\boldsymbol{X}^\intercal\boldsymbol{X}\right]} P_{\tilde{S}}(\boldsymbol{X}) \\ &= \iiint d\boldsymbol{\sigma}\, d\mu_N(\boldsymbol{U})\, d\mu_M(\boldsymbol{V}) \prod_{i=1}^{N}\delta(\tilde{\sigma}_i - \sigma_i^0)\, e^{N\,\mathrm{Tr}[\sqrt{\kappa}\boldsymbol{V}\tilde{\boldsymbol{\Sigma}}^\intercal\boldsymbol{U}^\intercal\tilde{\boldsymbol{Y}} - \frac{\kappa}{2}\tilde{\boldsymbol{\Sigma}}\tilde{\boldsymbol{\Sigma}}^\intercal]} \\ &= e^{-\frac{N}{2}\kappa\,\mathrm{Tr}\,\boldsymbol{\Sigma}^{0\intercal}\boldsymbol{\Sigma}^0}\iint d\mu_N(\boldsymbol{U})\, d\mu_M(\boldsymbol{V})\, e^{N\,\mathrm{Tr}[\sqrt{\kappa}\boldsymbol{\Sigma}^{0\intercal}\boldsymbol{U}\tilde{\boldsymbol{Y}}\boldsymbol{V}^\intercal]} \\ &= e^{-\frac{N}{2}\kappa\,\mathrm{Tr}\,\boldsymbol{\Sigma}^{0\intercal}\boldsymbol{\Sigma}^0}\mathcal{I}_{N,M}\big(\sqrt{\kappa}\boldsymbol{\Sigma}^0, \tilde{\boldsymbol{Y}}\big)\end{aligned} \tag{10.93}$$

where, we change variables $\boldsymbol{U} \to \boldsymbol{U}^\intercal, \boldsymbol{V} \to \boldsymbol{V}^\intercal$ in third line to match the definition of the spherical integral.

Recall that $\tilde{\boldsymbol{Y}} = \sqrt{\kappa}\boldsymbol{U}\boldsymbol{\Sigma}^0\boldsymbol{V}^\intercal + \tilde{\boldsymbol{Z}}$, so denoting $\frac{1}{NM}\ln\mathcal{I}_{N,M}\big(\sqrt{\kappa}\boldsymbol{\Sigma}^0, \tilde{\boldsymbol{Y}}\big)$ by $\mathcal{J}_{N,M}\big(\sqrt{\kappa}\boldsymbol{\Sigma}^0, \tilde{\boldsymbol{Y}}\big)$, the free energy can be written as:

$$\begin{aligned}\tilde{F}_N(\kappa) &= \mathbb{E}_{\tilde{\boldsymbol{Y}}}\Big[\frac{\kappa}{2M}\,\mathrm{Tr}\,\boldsymbol{\Sigma}^{0\intercal}\boldsymbol{\Sigma}^0 - \mathcal{J}_{N,M}\big(\sqrt{\kappa}\boldsymbol{\Sigma}^0, \tilde{\boldsymbol{Y}}\big)\Big] \\ &= \frac{\kappa}{2M}\sum_{i=1}^{N}\sigma_i^{0\,2} - \mathbb{E}_{\boldsymbol{U},\boldsymbol{V},\tilde{\boldsymbol{Z}}}\Big[\mathcal{J}_{N,M}\big(\sqrt{\kappa}\boldsymbol{\Sigma}^0, \sqrt{\kappa}\boldsymbol{U}\boldsymbol{\Sigma}^0\boldsymbol{V}^\intercal + \tilde{\boldsymbol{Z}}\big)\Big]\end{aligned} \tag{10.94}$$

By bi-rotational invariance of $\tilde{\boldsymbol{Z}}$, the second term equals

$$\mathbb{E}_{\boldsymbol{U},\boldsymbol{V},\tilde{\boldsymbol{Z}}}\Big[\mathcal{J}_{N,M}\big(\sqrt{\kappa}\boldsymbol{\Sigma}^0, \sqrt{\kappa}\boldsymbol{U}\boldsymbol{\Sigma}^0\boldsymbol{V}^\intercal + \boldsymbol{U}\tilde{\boldsymbol{Z}}\boldsymbol{V}^\intercal\big)\Big]$$

and then, both matrices $\boldsymbol{U}, \boldsymbol{V}$ can be absorbed into the integration in $\mathcal{J}_{N,M}$. So, the free energy equals:

$$\tilde{F}_N(\kappa) = \frac{\kappa}{2M}\sum_{i=1}^{N}\sigma_i^{0\,2} - \mathbb{E}_{\tilde{\boldsymbol{Z}}}\Big[\mathcal{J}_{N,M}\big(\sqrt{\kappa}\boldsymbol{\Sigma}^0, \sqrt{\kappa}\boldsymbol{\Sigma}^0 + \tilde{\boldsymbol{Z}}\big)\Big]$$

By the strong law of large numbers, the first term in (10.94) converges to $\frac{\kappa}{2}\alpha\int x^2\mu_S(x)\,dx$ almost surely, and proposition 10.7 follows from the following lemma.

Lemma 10.5. *For any $\kappa \in \mathbb{R}_+$, the sequence $\mathbb{E}_{\tilde{\boldsymbol{Z}}}\Big[\mathcal{J}_{N,M}\big(\sqrt{\kappa}\boldsymbol{\Sigma}^0, \sqrt{\kappa}\boldsymbol{\Sigma}^0 + \tilde{\boldsymbol{Z}}\big)\Big]$ converges to $\mathcal{J}^{(\alpha)}[\mu_{\sqrt{\kappa}S}, \mu_{\sqrt{\kappa}S} \boxplus_\alpha \mu_{\mathrm{MP}}]$ as $N \to \infty$, μ_S-almost surely.*

Proof. We first show that the assumptions of Theorem 1.1 in [132] holds a.s. for the sequence $\sqrt{\kappa}\boldsymbol{\Sigma}^0, \sqrt{\kappa}\boldsymbol{\Sigma}^0 + \tilde{\boldsymbol{Z}}$, so $\mathcal{J}_{N,M}$ converges to $\mathcal{J}^{(\alpha)}$ a.s. .

By assumption 10.2, the symmetrized ESD of $\sqrt{\kappa}\boldsymbol{\Sigma}^0$ converges weakly to $\bar{\mu}_{\sqrt{\kappa}S}$ by construction. The ESD of $\tilde{\boldsymbol{Z}}$ converges a.s. to the MP law μ_{MP}, so by independence of $\boldsymbol{\Sigma}^0, \tilde{\boldsymbol{Z}}$ the limiting ESD of $\sqrt{\kappa}\boldsymbol{\Sigma}^0 + \tilde{\boldsymbol{Z}}$ is the rectangular free convolution of $\mu_{\sqrt{\kappa}S}, \mu_{\mathrm{MP}}$ denoted by $\mu_{\sqrt{\kappa}S} \boxplus_\alpha \mu_{\mathrm{MP}}$. Moreover, by assumptions 10.2, 10.3, the second moment of ESD of $\sqrt{\kappa}\boldsymbol{\Sigma}^0$ is boounded and $\bar{\mu}_S$ has finite non-commutative entropy $\iint \ln|x-y|\,d\bar{\mu}_S(x)\,d\bar{\mu}_S(y) > -\infty$, and $\int \ln|x|\,d\bar{\mu}_S(x) > -\infty$. Therefore, the sequence $\mathcal{J}_{N,M}\big(\sqrt{\kappa}\boldsymbol{\Sigma}^0, \sqrt{\kappa}\boldsymbol{\Sigma}^0 + \tilde{\boldsymbol{Z}}\big)$ converges a.s. to $\mathcal{J}^{(\alpha)}[\mu_{\sqrt{\kappa}S}, \mu_{\sqrt{\kappa}S} \boxplus_\alpha \mu_{\mathrm{MP}}]$.

Now, we prove that the limit also holds under the expectation $\mathbb{E}_{\tilde{\boldsymbol{Z}}}$. For simplicity of notation we denote $\mathcal{J}_{N,M}\big(\sqrt{\kappa}\boldsymbol{\Sigma}^0, \sqrt{\kappa}\boldsymbol{\Sigma}^0 + \tilde{\boldsymbol{Z}}\big)$ by $\mathcal{J}_N$, and $\mathcal{J}^{(\alpha)}[\mu_{\sqrt{\kappa}S}, \mu_{\sqrt{\kappa}S} \boxplus_\alpha \mu_{\mathrm{MP}}]$ by $\mathcal{J}$. By Jensen's inequality (note that the expectation is over the matrix $\tilde{\boldsymbol{Z}}$), we have

$$\big|\mathbb{E}[\mathcal{J}_N] - \mathcal{J}\big| \le \mathbb{E}\big[|\mathcal{J}_N - \mathcal{J}|\big]. \tag{10.95}$$

Let $X_N \equiv \mathcal{J}_N - \mathcal{J}$. For $\epsilon > 0$ We can write

$$\begin{aligned}\mathbb{E}\big[|X_N|\big] &= \mathbb{E}\big[|X_N|\,\mathbb{I}\{|X_N| \le \epsilon\}\big] + \mathbb{E}\big[|X_N|\,\mathbb{I}\{|X_N| > \epsilon\}\big] \\ &\le \epsilon + \mathbb{E}\big[|X_N|\,\mathbb{I}\{|X_N| > \epsilon\}\big].\end{aligned} \tag{10.96}$$

By lemma 10.6, $|\mathcal{J}_N| \le \sqrt{\kappa}C_2\tilde{\gamma}_N$, where $\tilde{\gamma}_N$ is the top singular value of $\sqrt{\kappa}\boldsymbol{\Sigma}^0 + \tilde{\boldsymbol{Z}}$. The second term in (10.96) can be bounded as,

$$\mathbb{E}\big[|X_N|\,\mathbb{I}\{|X_N| > \epsilon\}\big] \le \mathbb{E}\big[|W_N|\,\mathbb{I}\{|X_N| > \epsilon\}\big] \tag{10.97}$$

where

$$W_N = \max\Big\{\big|\mathcal{J} - \sqrt{\kappa}C_2\tilde{\gamma}_N\big|, \big|\mathcal{J} + \sqrt{\kappa}C_2\tilde{\gamma}_N\big|\Big\} = \sqrt{\kappa}C_2\tilde{\gamma}_N + \mathrm{sign}(\mathcal{J})\mathcal{J}.$$

For any positive constant t, we have

$$\begin{aligned}
&\mathbb{E}\big[|W_N|\,\mathbb{I}\{|X_N|>\epsilon\}\big] \\
&\quad= \mathbb{E}\big[|W_N|\,\mathbb{I}\{|X_N|>\epsilon\}\,\mathbb{I}\{|W_N|\le t\}\big] + \mathbb{E}\big[|W_N|\,\mathbb{I}\{|X_N|>\epsilon\}\,\mathbb{I}\{|W_N|>t\}\big] \\
&\quad\le \mathbb{E}\big[|W_N|\,\mathbb{I}\{|X_N|>\epsilon\}\mathbb{I}\{|W_N|\le t\}\big] + \mathbb{E}\big[|W_N|\,\mathbb{I}\{|W_N|>t\}\big]
\end{aligned} \tag{10.98}$$

For the first term in (10.98) we can write

$$\begin{aligned}
\mathbb{E}\big[|W_N|\,\mathbb{I}\{|X_N|>\epsilon\}\mathbb{I}\{|W_N|\le t\}\big] &\le t\mathbb{E}\big[\mathbb{I}\{|X_N|>\epsilon\}\big] \\
&\le t\,\mathbb{P}\big(|X_N|>\epsilon\big)
\end{aligned} \tag{10.99}$$

and the second term in (10.98) can be rewritten as

$$\mathbb{E}\big[|W_N|\,\mathbb{I}\{|W_N|>t\}\big] = \mathbb{E}\bigg[|W_N|\,\mathbb{I}\Big\{\tilde{\gamma}_N > \frac{2}{\sqrt{\kappa}C_2}\big(t-\operatorname{sign}(\mathcal{J})\mathcal{J}\big)\Big\}\bigg]. \tag{10.100}$$

From (10.97), (10.98), (10.99), we obtain

$$\begin{aligned}
&\mathbb{E}\big[|X_N|\,\mathbb{I}\{|X_N|>\epsilon\}\big] \\
&\qquad\le t\,\mathbb{P}\big(|X_N|>\epsilon\big) + \mathbb{E}\bigg[|W_N|\,\mathbb{I}\Big\{\tilde{\gamma}_N > \frac{2}{\sqrt{\kappa}C_2}\big(t-\operatorname{sign}(\mathcal{J})\mathcal{J}\big)\Big\}\bigg].
\end{aligned} \tag{10.101}$$

Notice that W_N is a polynomial function of $\tilde{\gamma}_N$, so by lemma 10.8, vanishes as $N\to\infty$ for sufficiently large constant t. By almost sure convergence of $\mathcal{J}_N$ to $\mathcal{J}$, $\mathbb{P}\big(|X_N|>\epsilon\big) \xrightarrow{N\to\infty} 0$. For a fixed $t>0$, the first term in (10.101) goes to 0 in the limit $N\to\infty$. Therefore, taking the limit of both sides in (10.96), for any $\epsilon>0$, we find:

$$\lim_{N\to\infty} \mathbb{E}\big[|X_N|\big] \le \epsilon. \tag{10.102}$$

From which, by (10.95), we deduce that $\lim_{N\to\infty}\mathbb{E}[\mathcal{J}_N] = \mathcal{J}$. □

Technical lemmas

Lemma 10.6. *For any N, $M\ge N$, and $N\times M$ symmetric matrices $\boldsymbol{A},\boldsymbol{B}$ with top singular values σ_N^A,σ_N^B*

$$-\sigma_N^A\sigma_N^B \le \mathcal{J}_{N,M}(\boldsymbol{A},\boldsymbol{B}) \le \sigma_N^A\sigma_N^B.$$

Proof. Let $\boldsymbol{A}=\boldsymbol{U}_A\boldsymbol{\Sigma}_A\boldsymbol{V}_A^\intercal$, $\boldsymbol{B}=\boldsymbol{U}_B\boldsymbol{\Sigma}_B\boldsymbol{V}_B^\intercal$ be the SVD of $\boldsymbol{A}$, $\boldsymbol{B}$. We can write

$$\mathcal{I}_{N,M}(\boldsymbol{A},\boldsymbol{B}) = \iint D\boldsymbol{U}\,D\boldsymbol{V}\,e^{N\operatorname{Tr}\boldsymbol{\Sigma}_A^\intercal\boldsymbol{U}\boldsymbol{\Sigma}_B\boldsymbol{V}^\intercal} = \iint D\boldsymbol{U}\,D\boldsymbol{V}\,e^{N\sum_{i,j=1}^N\sigma_i^A\sigma_j^B U_{ij}V_{ij}}$$

The term in the exponent can be bounded as:

$$\begin{aligned}\sum_{i,j=1}^{N} \sigma_i^A \sigma_j^B U_{ij} V_{ij} &\leq \sum_{i,j=1}^{N} \left|\sigma_i^A \sigma_j^B U_{ij} V_{ij}\right| \\ &\leq \sigma_N^A \sigma_N^B \sum_{i,j=1}^{N} \left|U_{ij} V_{ij}\right| \\ &\leq \|\boldsymbol{U}\|_{\mathrm{F}} \|\boldsymbol{V}\|_{\mathrm{F}} \sigma_N^A \sigma_N^B \\ &\leq \sqrt{NM}\, \sigma_N^A \sigma_N^B\end{aligned} \tag{10.103}$$

which implies $\mathcal{I}_{N,M}(\boldsymbol{A}, \boldsymbol{B}) \leq e^{N\sqrt{NM}\sigma_N^A \sigma_N^B}$. Similarly, we get $\mathcal{I}_{N,M}(\boldsymbol{A}, \boldsymbol{B}) \geq e^{-N\sqrt{NM}\sigma_N^A \sigma_N^B}$. Therefore, we obtain

$$-\sqrt{\frac{N}{M}}\sigma_N^A \sigma_N^B \leq \mathcal{J}_{N,M}(\boldsymbol{A}, \boldsymbol{B}) \leq \sqrt{\frac{N}{M}}\sigma_N^A \sigma_N^B$$

The result follows since $N \leq M$. □

Lemma 10.7. *Let $\tilde{\gamma}_N$ be the top singular value of the matrix $\sqrt{\kappa}\boldsymbol{\Sigma}^0 + \tilde{\boldsymbol{Z}}$. For $k > 1 + \sqrt{K} + \sqrt{\kappa}C_2$, we have*

$$\mathbb{P}\{\tilde{\gamma}_N \geq k\} \leq e^{-\frac{N}{2}(k - \sqrt{\kappa}C_2 - 1 - \sqrt{K})^2}.$$

Proof. By triangle inequality, we have:

$$\tilde{\gamma}_N \leq \sqrt{\kappa} \max \sigma_i + \gamma_N^{\tilde{Z}} \leq \sqrt{\kappa}C_2 + \gamma_N^{\tilde{Z}}$$

where $\gamma_N^{\tilde{Z}}$ is the top singular value of $\tilde{\boldsymbol{Z}}$. Thus, we can write

$$\begin{aligned}\mathbb{P}\{\tilde{\gamma}_N \geq k\} &\leq \mathbb{P}\big\{\gamma_N^{\tilde{Z}} + \sqrt{\kappa}C_2 \geq k\big\} \\ &= \mathbb{P}\{\gamma_N^{\tilde{Z}} \geq k - \sqrt{\kappa}C_2\}.\end{aligned}$$

By [116] (Theorem II.13), for $k - \sqrt{\kappa}C_2 > 1 + \sqrt{K} > 1 + \sqrt{M/N}$, we have

$$\mathbb{P}\{\tilde{\gamma}_N \geq k - \sqrt{\kappa}C_2\} \leq e^{-\frac{N}{2}(k - \sqrt{\kappa}C_2 - 1 - \sqrt{K})^2}$$

and therefore we get the result. □

Lemma 10.8. *For any polynomial function g, and k a sufficiently large constant, we have that*

$$\lim_{n \to \infty} \mathbb{E}\big[g(\tilde{\gamma}_N)\mathbb{I}\{\tilde{\gamma}_N \geq k\}\big] = 0.$$

Proof. See Lemma H.3 in [48]. □

10.C.2 Proof of proposition 10.8

Consider two matrices with the same singular vectors, $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\intercal$, $\tilde{\boldsymbol{S}} = \boldsymbol{U}\tilde{\boldsymbol{\Sigma}}\boldsymbol{V}^\intercal$, where $\boldsymbol{U} \in \mathbb{R}^{N\times N}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$ are Haar orthogonal matrices, and $\boldsymbol{\sigma}$, $\tilde{\boldsymbol{\sigma}}$ are distributed according to $P_N^{(1)}(\boldsymbol{\sigma}), P_N^{(2)}(\tilde{\boldsymbol{\sigma}})$, respectively. For two such matrices, we write $(\boldsymbol{S}, \tilde{\boldsymbol{S}}) \sim Q_{N,M}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})$ which is the joint p.d.f. of $\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}$,

$$dQ_{N,M}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) = d\mu_N(\boldsymbol{U})\, d\mu_M(\boldsymbol{V})\, P_N^{(1)}(\boldsymbol{\sigma})\, d\boldsymbol{\sigma}\, P_N^{(2)}(\tilde{\boldsymbol{\sigma}})\, d\tilde{\boldsymbol{\sigma}}$$

For $t \in [0,1]$, consider the following observation model:

$$\begin{cases} \boldsymbol{Y}_1^{(t)} = \sqrt{\kappa t}\boldsymbol{S} + \boldsymbol{Z}_1 \\ \boldsymbol{Y}_2^{(t)} = \sqrt{\kappa(1-t)}\tilde{\boldsymbol{S}} + \boldsymbol{Z}_2 \end{cases} \tag{10.104}$$

where $\boldsymbol{Z}_1, \boldsymbol{Z}_2 \in \mathbb{R}^{N\times M}$ are Gaussian matrices as in (10.1), independent of each other. $(\boldsymbol{S}, \tilde{\boldsymbol{S}}) \sim Q_{N,M}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}})$. The free energy for this model can be written as

$$\begin{aligned} F_N(t) &= -\frac{1}{MN}\mathbb{E}_{\boldsymbol{Y}_1^{(t)}, \boldsymbol{Y}_2^{(t)}}\Bigg[\ln \int dQ_{N,M}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) \\ &\qquad \times e^{N \operatorname{Tr}[\sqrt{\kappa t}\boldsymbol{X}^\intercal \boldsymbol{Y}_1^{(t)} - \frac{\kappa t}{2}\boldsymbol{X}^\intercal\boldsymbol{X} + \sqrt{\kappa(1-t)}\tilde{\boldsymbol{X}}^\intercal \boldsymbol{Y}_2^{(t)} - \frac{\kappa(1-t)}{2}\tilde{\boldsymbol{X}}^\intercal\tilde{\boldsymbol{X}}]}\Bigg] \\ &= -\frac{1}{MN}\mathbb{E}_{\boldsymbol{Y}_1^{(t)}, \boldsymbol{Y}_2^{(t)}}\Bigg[\ln \int dQ_{N,M}(\boldsymbol{U}, \boldsymbol{V}, \boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}) \\ &\qquad \times e^{N \operatorname{Tr}[\kappa t \boldsymbol{X}^\intercal \boldsymbol{S} + \sqrt{\kappa t}\boldsymbol{X}^\intercal \boldsymbol{Z}_1 - \frac{\kappa t}{2}\boldsymbol{X}^\intercal\boldsymbol{X} + \kappa(1-t)\tilde{\boldsymbol{X}}^\intercal\tilde{\boldsymbol{S}} + \sqrt{\kappa(1-t)}\tilde{\boldsymbol{X}}^\intercal \boldsymbol{Z}_2 - \frac{\kappa(1-t)}{2}\tilde{\boldsymbol{X}}^\intercal\tilde{\boldsymbol{X}}]}\Bigg] \end{aligned} \tag{10.105}$$

where the singular vectors of $\boldsymbol{X}, \tilde{\boldsymbol{X}}$ are the same, $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\intercal$, $\tilde{\boldsymbol{X}} = \boldsymbol{U}\tilde{\boldsymbol{\Sigma}}\boldsymbol{V}^\intercal$. Note that, for $t = 0$ the only term depending on $\boldsymbol{\sigma}$ (in both the inner and outer expectation) is the pdf $P_N^{(1)}(\boldsymbol{\sigma})$ and we can integrate over $\boldsymbol{\sigma}$ in both of the expectations, to get $F_N(0) = F_N^{(2)}(\kappa)$. Similarly, we have $F_N(1) = F_N^{(1)}(\kappa)$.

Taking the derivative w.r.t. t, we get

$$\begin{aligned} \frac{d}{dt}F_N(t) = -\frac{1}{M}\mathbb{E}\Big[&\kappa \operatorname{Tr}\langle \boldsymbol{X}^\intercal\boldsymbol{S}\rangle_t + \frac{1}{2}\sqrt{\frac{\kappa}{t}}\operatorname{Tr}\boldsymbol{Z}_1^\intercal\langle\boldsymbol{X}\rangle_t - \frac{\kappa}{2}\operatorname{Tr}\langle\boldsymbol{X}^\intercal\boldsymbol{X}\rangle_t \\ &- \kappa \operatorname{Tr}\langle\tilde{\boldsymbol{X}}^\intercal\tilde{\boldsymbol{S}}\rangle_t - \frac{1}{2}\sqrt{\frac{\kappa}{1-t}}\operatorname{Tr}\boldsymbol{Z}_2^\intercal\langle\tilde{\boldsymbol{X}}\rangle_t + \frac{\kappa}{2}\operatorname{Tr}\langle\tilde{\boldsymbol{X}}^\intercal\tilde{\boldsymbol{X}}\rangle_t\Big] \end{aligned} \tag{10.106}$$

where $\langle . \rangle_t$ denotes the expectation with respect to the posterior distribution of the model (10.104). By integration by parts, we have

$$\begin{aligned} \mathbb{E}\big[\operatorname{Tr}\boldsymbol{Z}_1^\intercal\langle\boldsymbol{X}\rangle_t\big] &= \sqrt{\kappa t}\,\mathbb{E}\Big[\operatorname{Tr}\langle\boldsymbol{X}^\intercal\boldsymbol{X}\rangle_t - \operatorname{Tr}\langle\boldsymbol{X}\rangle_t^\intercal\langle\boldsymbol{X}\rangle_t\Big] \\ \mathbb{E}\big[\operatorname{Tr}\boldsymbol{Z}_2^\intercal\langle\tilde{\boldsymbol{X}}\rangle_t\big] &= \sqrt{\kappa(1-t)}\,\mathbb{E}\Big[\operatorname{Tr}\langle\tilde{\boldsymbol{X}}^\intercal\tilde{\boldsymbol{X}}\rangle_t - \operatorname{Tr}\langle\tilde{\boldsymbol{X}}\rangle_t^\intercal\langle\tilde{\boldsymbol{X}}\rangle_t\Big] \end{aligned}$$

Therefore (10.106) can be written as:

$$\begin{aligned}\frac{d}{dt}F_N(t) &= -\frac{1}{M}\frac{\kappa}{2}\mathbb{E}\Big[2\,\mathrm{Tr}\langle \boldsymbol{X}^\intercal \boldsymbol{S}\rangle_t - \mathrm{Tr}\langle \boldsymbol{X}\rangle_t^\intercal\langle \boldsymbol{X}\rangle_t - 2\,\mathrm{Tr}\langle \tilde{\boldsymbol{X}}^\intercal \tilde{\boldsymbol{S}}\rangle_t + \mathrm{Tr}\langle \tilde{\boldsymbol{X}}\rangle_t^\intercal\langle \tilde{\boldsymbol{X}}\rangle_t\Big] \\ &= -\frac{1}{M}\frac{\kappa}{2}\mathbb{E}\big[\,\mathrm{Tr}[\langle \boldsymbol{X}^\intercal \boldsymbol{S}\rangle_t - \langle \tilde{\boldsymbol{X}}^\intercal \tilde{\boldsymbol{S}}\rangle_t]\big] \quad \text{(By Nishimori)}\end{aligned} \tag{10.107}$$

We have:

$$\begin{aligned}\frac{2M}{\kappa}\Big|\frac{d}{dt}F_N(t)\Big| &= \left|\mathbb{E}\left[\Big\langle \mathrm{Tr}\left[\boldsymbol{S}^\intercal(\boldsymbol{X}-\tilde{\boldsymbol{X}}) - (\tilde{\boldsymbol{S}}^\intercal - \boldsymbol{S})\tilde{\boldsymbol{X}}\right]\Big\rangle_t\right]\right| \\ &\le \mathbb{E}\left[\left\langle\left|\,\mathrm{Tr}\left[\boldsymbol{S}^\intercal(\boldsymbol{X}-\tilde{\boldsymbol{X}}) - (\tilde{\boldsymbol{S}}^\intercal - \boldsymbol{S})\tilde{\boldsymbol{X}}\right]\right|\right\rangle_t\right] \\ &\le \mathbb{E}\left[\left\langle\left|\,\mathrm{Tr}\,\boldsymbol{S}^\intercal(\boldsymbol{X}-\tilde{\boldsymbol{X}})\right|\right\rangle_t\right] + \mathbb{E}\left[\left\langle\left|\,\mathrm{Tr}(\tilde{\boldsymbol{S}}-\boldsymbol{S})\tilde{\boldsymbol{X}}^\intercal]\right|\right\rangle_t\right] \\ &\le \mathbb{E}\Big[\|\boldsymbol{S}\|_{\mathrm{F}}\langle\|\boldsymbol{X}-\tilde{\boldsymbol{X}}\|_{\mathrm{F}}\rangle_t\Big] + \mathbb{E}\Big[\|\boldsymbol{S}-\tilde{\boldsymbol{S}}\|_{\mathrm{F}}\langle\|\tilde{\boldsymbol{X}}\|_{\mathrm{F}}\rangle_t\Big] \\ &\le \sqrt{\mathbb{E}\Big[\|\boldsymbol{S}\|_{\mathrm{F}}^2\Big]\mathbb{E}\Big[\big\langle\|\boldsymbol{X}-\tilde{\boldsymbol{X}}\|_{\mathrm{F}}\big\rangle_t^2\Big]} + \sqrt{\mathbb{E}\Big[\|\boldsymbol{S}-\tilde{\boldsymbol{S}}\|_{\mathrm{F}}^2\Big]\mathbb{E}\Big[\big\langle\|\tilde{\boldsymbol{X}}\|_{\mathrm{F}}\big\rangle_t^2\Big]} \\ &\le \sqrt{\mathbb{E}\big[\|\boldsymbol{S}\|_{\mathrm{F}}^2\big]\mathbb{E}\Big[\big\langle\|\boldsymbol{X}-\tilde{\boldsymbol{X}}\|_{\mathrm{F}}^2\big\rangle_t\Big]} + \sqrt{\mathbb{E}\big[\|\boldsymbol{S}-\tilde{\boldsymbol{S}}\|_{\mathrm{F}}^2\big]\mathbb{E}\Big[\big\langle\|\tilde{\boldsymbol{X}}\|_{\mathrm{F}}^2\big\rangle_t\Big]} \\ &= \sqrt{\mathbb{E}\big[\|\boldsymbol{S}\|_{\mathrm{F}}^2\big]\mathbb{E}\big[\|\boldsymbol{S}-\tilde{\boldsymbol{S}}\|_{\mathrm{F}}^2\big]} + \sqrt{\mathbb{E}\big[\|\boldsymbol{S}-\tilde{\boldsymbol{S}}\|_{\mathrm{F}}^2\big]\mathbb{E}\big[\|\tilde{\boldsymbol{S}}\|_{\mathrm{F}}^2\big]} \\ &= \Big(\sqrt{\mathbb{E}\big[\|\boldsymbol{S}\|_{\mathrm{F}}^2\big]} + \sqrt{\mathbb{E}\big[\|\tilde{\boldsymbol{S}}\|_{\mathrm{F}}^2\big]}\Big)\sqrt{\mathbb{E}\big[\|\boldsymbol{S}-\tilde{\boldsymbol{S}}\|_{\mathrm{F}}^2\big]} \\ &= \Big(\sqrt{\mathbb{E}_{\boldsymbol{\sigma}}\big[\|\boldsymbol{\sigma}\|^2\big]} + \sqrt{\mathbb{E}_{\tilde{\boldsymbol{\sigma}}}\big[\|\tilde{\boldsymbol{\sigma}}\|^2\big]}\Big)\sqrt{\mathbb{E}_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}}\big[\|\boldsymbol{\sigma}-\tilde{\boldsymbol{\sigma}}\|^2\big]}\end{aligned} \tag{10.108}$$

We obtain the result by integrating (10.108), over t from 0 to 1, and using $N \le M$. □

10.C.3 Proof of lemma 10.4

First, note that by rotational invariance, $p_S(\boldsymbol{\sigma})$ is invariant under permutations, so without loss of generality, we can assume $\boldsymbol{\sigma}$ is in non-decreasing order.

Since $p_{\tilde{S}}(\tilde{\boldsymbol{\sigma}})$ is a delta distribution, we can easily write

$$\mathbb{E}_{\boldsymbol{\sigma},\tilde{\boldsymbol{\sigma}}}\big[\|\boldsymbol{\sigma}-\tilde{\boldsymbol{\sigma}}\|^2\big] = \mathbb{E}_{\boldsymbol{\sigma}}\big[\|\boldsymbol{\sigma}-\boldsymbol{\sigma}^0\|^2\big] \tag{10.109}$$

For a vector $\boldsymbol{\sigma}$, denote the empirical distribution of its components by $\hat{\mu}_{\boldsymbol{\sigma}}$. The Wasserstein-2 distance between two empirical distributions, $\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0}$ is defined as

$$W_2(\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0}) = \sqrt{\inf_{\gamma\in\Gamma(\hat{\mu}_{\boldsymbol{\sigma}},\hat{\mu}_{\boldsymbol{\sigma}^0})}\mathbb{E}_{\gamma(x,y)}\big[(x-y)^2\big]}$$

with $\Gamma(\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0})$ is the set of couplings of $\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0}$. By lemma H.5 in [48] ,the Wasserstein-2 distance can be written as

$$W_2(\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0}) = \sqrt{\min_{\pi \in \mathcal{S}_N} \frac{1}{N} \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_\pi^0\|^2} \tag{10.110}$$

where $\boldsymbol{\sigma}_\pi^0$ is the permuted version of $\boldsymbol{\sigma}^0$, and $\mathcal{S}_N$ is the set of all N-permutations. So, for a given $\boldsymbol{\sigma}$ and $\boldsymbol{\sigma}^0$ (which have a non-decreasing order), we have (considering the identity permutation)

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}^0\|^2 \geq N W_2(\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0})^2 \tag{10.111}$$

On the other hand, for any permutation of $\boldsymbol{\sigma}^0$ (in particular, the one which achieves the minimum in (10.110)), we have

$$\begin{aligned} \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_\pi^0\|^2 &= \|\boldsymbol{\sigma}\|^2 + \|\boldsymbol{\sigma}_\pi^0\|^2 - 2\boldsymbol{\sigma}^\intercal \boldsymbol{\sigma}_\pi^0 \\ &\geq \|\boldsymbol{\sigma}\|^2 + \|\boldsymbol{\sigma}_\pi^0\|^2 - 2\boldsymbol{\sigma}^\intercal \boldsymbol{\sigma}^0 = \|\boldsymbol{\sigma} - \boldsymbol{\sigma}^0\|^2 \end{aligned}$$

where we used rearrangement inequality [120] to get the inequality in the second line. So,

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}^0\|^2 \leq N W_2(\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0})^2 \tag{10.112}$$

From (10.109), (10.111),(10.112), we have

$$\mathbb{E}_{\boldsymbol{\sigma}, \tilde{\boldsymbol{\sigma}}} \left[\|\boldsymbol{\sigma} - \tilde{\boldsymbol{\sigma}}\|^2 \right] = \mathbb{E}_{\boldsymbol{\sigma}} \left[N W_2(\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0})^2 \right] \tag{10.113}$$

Lemma 10.9 concludes the proof.

Lemma 10.9. *Suppose $\boldsymbol{\sigma} \in \mathbb{R}^N$ is distributed according to $p_S(\boldsymbol{\sigma})$, and $\boldsymbol{\sigma}^0$ is generated with i.i.d. elements from μ_S. Let $\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0}$ be their empirical distribution. We have:*

$$\lim_{N \to \infty} \mathbb{E}_{\boldsymbol{\sigma}} \left[W_2(\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0})^2 \right] = 0$$

Proof. By triangle inequality, we have:

$$W_2(\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0}) \leq W_2(\hat{\mu}_{\boldsymbol{\sigma}}, \mu_S) + W_2(\hat{\mu}_{\boldsymbol{\sigma}^0}, \mu_S) \tag{10.114}$$

From the weak convergence and convergence of second moment, assumptions 10.2 and 10.3 imply that the second moment of the empirical spectral distribution converges almost surely to the one of μ_S. Thus, by [119](Theorem 7.12), the empirical singular value distribution in the Wasserstein-2 metric to μ_S. Hence, the first term approaches 0 as $N \to \infty$ almost surely.

By law of large numbers and since the support of μ_S is bounded, the second term also converges 0 as $N \to \infty$. Therefore, we have $W_2(\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0}) \to 0$ almost surely. Consequently, we have that $W_2(\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0})^2 \to 0$ almost surely.

One can see that:

$$\begin{aligned} W_2(\hat{\mu}_{\boldsymbol{\sigma}}, \hat{\mu}_{\boldsymbol{\sigma}^0})^2 &\leq \frac{2}{N} \sum \sigma_i^2 + \frac{2}{N} \sum \sigma_i^{0\,2} \\ &\leq 2 m_{\hat{\mu}_{\boldsymbol{\sigma}}}^{(2)} + 2 C_2^2 \end{aligned} \tag{10.115}$$

with $m_{\hat{\mu}_{\boldsymbol{\sigma}}}^{(2)}$ the second moment of $\hat{\mu}_{\boldsymbol{\sigma}}$ which is almost surely bounded by assumption. Therefore, the result follows by using dominated convergence theorem. □

11 Extensive-Rank Matrix Factorization

In this chapter, we consider the extensive-rank matrix factorization (MF) problem. We approach the problem from a Bayesian perspective and assume that an observation or data matrix

$$\boldsymbol{Y} = \sqrt{\kappa}\boldsymbol{S}\boldsymbol{T} + \boldsymbol{Z}$$

is given to a statistician who knows the prior distributions of $\boldsymbol{S}$ and $\boldsymbol{T}$ as well as the prior of the additive noise matrix $\boldsymbol{Z}$ and the SNR $\kappa > 0$. The task of the statistician is to construct estimators $\boldsymbol{\Xi}_S(\cdot)$, $\boldsymbol{\Xi}_T(\cdot)$ for the matrix factors $\boldsymbol{S}$, $\boldsymbol{T}$, that ideally, minimize the average MSE $\mathbb{E}\|\boldsymbol{S} - \boldsymbol{\Xi}_S(\boldsymbol{Y})\|_{\mathrm{F}}^2$ and $\mathbb{E}\|\boldsymbol{T} - \boldsymbol{\Xi}_T(\boldsymbol{Y})\|_{\mathrm{F}}^2$ (the expectation is w.r.t $\boldsymbol{S}, \boldsymbol{T}, \boldsymbol{Z}$). We consider priors which are rotation invariant for all three matrices $\boldsymbol{S}$, $\boldsymbol{T}$, $\boldsymbol{Z}$ and for $\boldsymbol{S}$ we furthermore impose that it is square and symmetric. We look at the asymptotic regime where all matrix dimensions and ranks tend to infinity at the same speed.

We propose:

- an explicit RIE to estimate $\boldsymbol{S}$, which requires the knowledge of the priors of *both* $\boldsymbol{S}, \boldsymbol{T}$ and of the noise $\boldsymbol{Z}$ (see (11.6)). Moreover, under the assumption that $\boldsymbol{S}$ is positive-semi-definite, a *sub-optimal* RIE can be derived which *does not* require any prior on $\boldsymbol{S}$ (see (11.9)).
- an explicit RIE to estimate $\boldsymbol{T}$ (see (11.11)), which requires the knowledge of the priors of the noise $\boldsymbol{Z}$ and $\boldsymbol{S}$ *only* (the prior of $\boldsymbol{Y}$ is not required).
- combined with the singular value decomposition (SVD) of the observation matrix, our explicit RIEs provide a spectral algorithm to reconstruct both factors $\boldsymbol{S}$ and $\boldsymbol{T}$.

Content of this chapter was presented in [56] F. Pourkamali and N. Macris, "Bayesian extensive-rank matrix factorization with rotational invariant priors," in Thirty-seventh Conference on Neural Information Processing Systems, 2023.

11.1 Matrix Factorization Model and Rotation Invariant Estimators

11.1.1 Matrix Factorization Model

Let $\boldsymbol{S} = \boldsymbol{S}^\intercal \in \mathbb{R}^{N\times N}$ a symmetric matrix distributed according to a rotationally invariant prior $P_S(\boldsymbol{S})$, i.e., for any orthogonal matrix $\boldsymbol{O} \in \mathbb{R}^{N\times N}$ we have $P_S(\boldsymbol{OSO}^\intercal) = P_S(\boldsymbol{S})$. Let also $\boldsymbol{T} \in \mathbb{R}^{N\times M}$ be distributed according to a bi-rotationally invariant prior $P_T(\boldsymbol{T})$, i.e. for any orthogonal matrices $\boldsymbol{U} \in \mathbb{R}^{N\times N}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$ we have $P_T(\boldsymbol{UTV}^\intercal) = P_T(\boldsymbol{T})$. We observe the data matrix $\boldsymbol{Y} \in \mathbb{R}^{N\times M}$,

$$\boldsymbol{Y} = \sqrt{\kappa}\boldsymbol{ST} + \boldsymbol{Z} \tag{11.1}$$

where $\boldsymbol{Z} \in \mathbb{R}^{N\times M}$ is also bi-rotationally invariant distributed, and $\kappa \in \mathbb{R}_+$ is proportional to the signal-to-noise-ratio (SNR). The goal is to recover *both factors* $\boldsymbol{S}$ and $\boldsymbol{T}$ from the data matrix $\boldsymbol{Y}$. For definiteness, we consider the regime $M \geq N$ with aspect ratio $N/M \to \alpha \in (0, 1]$ as $N \to \infty$. The case of $\alpha > 1$ can be analyzed in the same manner and is presented in appendix 11.E. Furthermore, we assume that the entries of $\boldsymbol{S}, \boldsymbol{T}$ and $\boldsymbol{Z}$ are of the order $O(1/\sqrt{N})$. This scaling is such that the eigenvalues of $\boldsymbol{S}$ and singular values of $\boldsymbol{T}, \boldsymbol{Z}$ and $\boldsymbol{Y}$ are of the order $O(1)$ as $N \to \infty$.

Assumption 11.1. *The empirical eigenvalue distribution of $\boldsymbol{S}$ converge weakly to measure ρ_S, and the ESD of $\boldsymbol{T}, \boldsymbol{Z}$ converge weakly to measures μ_T, μ_Z with bounded support on the real line. Moreover, these measures are known to the statistician. He can deduce (in principle) these measures from the priors on $\boldsymbol{S}, \boldsymbol{T}, \boldsymbol{Z}$.*

Remark 11.1. *In a general formulation of matrix factorization the hidden matrices have dimensions $\boldsymbol{S} \in \mathbb{R}^{N\times H}, \boldsymbol{T} \in \mathbb{R}^{H\times M}$, and in the Bayesian framework with bi-rotational invariant priors for both factors, the optimal estimators are trivially the zero matrix. Indeed, from bi-rotational invariance we have $P_S(-\boldsymbol{S}) = P_S(\boldsymbol{S})$, $P_T(-\boldsymbol{T}) = P_T(\boldsymbol{T})$, which implies that the Bayesian estimate is zero. Here, by imposing that $\boldsymbol{S} \in \mathbb{R}^{N\times N}$ is symmetric and $P_S(\boldsymbol{OSO}^\intercal) = P_S(\boldsymbol{S})$, we can break this symmetry and find non-trivial estimators. This is due to the fact that the map $\boldsymbol{S} \to -\boldsymbol{S}$ cannot be realized as a (real) orthogonal transformation, so $P_S(-\boldsymbol{S}) = P_S(\boldsymbol{S})$ does not hold in general (various examples are given in section 11.3 and appendices). Of course, if the prior is even, e.g. Wigner ensemble, again the Bayesian posterior estimate is trivially zero for both factors. As we will see our RIEs are consistent with these observations.*

Related literature and discussion

In extensive-rank regimes, when the rank grows like the matrix dimensions, despite various attempts there is no solid theory of MF. One approach is

based on Approximate Message Passing (AMP) methods developed in [40–42]. Despite acceptable performance in practical settings [44], as pointed out in [43] the AMP algorithms developed in these works are (theoretically) sub-optimal. Other approaches rooted in statistical physics have been considered in [43,45–47] but have not led to explicit reconstructions of matrix factors or algorithms. A practical probabilistic approach to MF problem is based on variational Bayesian approximations [137–139], in which one tries to approximate the posterior distribution with proper distribution. In [140] it is shown that under Gaussian priors, the solution to the MF problem is a reweighted SVD of the observation matrix. We point out here that these estimators can be seen as a RIE and therefore there seems to be a rather close relation between the RIE studied here and the variational Bayesian approach. This also suggests that adapting RIEs to real data is an interesting direction for future research. Finally, let us also mention optimization approaches where one constructs estimators by following a gradient flow (or gradient descent) trajectory of a training loss of the type $\|\boldsymbol{Y} - \boldsymbol{S}\boldsymbol{T}\|_{\mathrm{F}}^2 + \text{reg. term}$ (see [141], [142] for analysis in rotation invariant models). Benchmarking these various other algorithmic approaches against our explicit RIEs (conjectured to be optimal) is outside the scope of this work and is left for future work.

Constraints such as sparsity or non-negativity of the matrix entries which have important applications [38] are not covered by our theory. Despite this drawback, we believe that the proposed estimators are important both for theoretical and practical purposes. Even in non-rotation invariant problems our explicit RIEs may serve as sub-optimal estimators, and as we show in an example they can be used as a "warmed-up" spectral initialization for more efficient algorithms (see for example [23, 24] for related ideas in other contexts). The methodology developed here may open up the way to further analysis in inference and learning problems perhaps also in the context of neural networks where extensive rank weight matrices must be estimated.

11.1.2 Rotation Invariant Estimators

To recover matrices $\boldsymbol{S}, \boldsymbol{T}$ from $\boldsymbol{Y}$, we consider two denoising problems. One is recovering $\boldsymbol{S}$ by treating both $\boldsymbol{T}, \boldsymbol{Z}$ as "noise" matrices, and the other is estimating $\boldsymbol{T}$ by treating $\boldsymbol{S}, \boldsymbol{Z}$ as "noise". As will become clear the procedure is not iterative, and the two denoising problems are solved independently and simultaneously. In the following, for each of these two problems, we introduce two rotation invariant classes of estimators and discuss their optimum *Oracle* estimators. We then provide an explicit construction and algorithm for RIEs which we conjecture have the optimum performance of Oracle estimators in the large N limit.

RIE class for $\boldsymbol{S}$

Consider the SVD of $\boldsymbol{Y} = \boldsymbol{U}_Y \boldsymbol{\Gamma} \boldsymbol{V}_Y^\intercal$, where $\boldsymbol{U}_Y \in \mathbb{R}^{N\times N}$, $\boldsymbol{V}_Y \in \mathbb{R}^{M\times M}$ are orthogonal, and $\boldsymbol{\Gamma} \in \mathbb{R}^{N\times M}$ is a diagonal matrix with singular values of $\boldsymbol{Y}$ on its diagonal, $\big(\gamma_i\big)_{1\le i\le N}$. A rotational invariant estimator for $\boldsymbol{S}$ is denoted $\boldsymbol{\Xi}_S(\boldsymbol{Y})$, and is constructed as:

$$\boldsymbol{\Xi}_S(\boldsymbol{Y}) = \boldsymbol{U}_Y \operatorname{diag}(\xi_{s1}, \dots, \xi_{sN})\, \boldsymbol{U}_Y^\intercal \tag{11.2}$$

where $\xi_{s1}, \dots, \xi_{sN}$ are the eigenvalues of the estimator.

First, we derive an *Oracle estimator* by minimizing the squared error $\frac{1}{N}\big\|\boldsymbol{S} - \boldsymbol{\Xi}_S(\boldsymbol{Y})\big\|_{\mathrm{F}}^2$ for a given instance, over the RIE class or equivalently over the choice of the eigenvalues $\big(\xi_{si}\big)_{1\le i\le N}$. Let the eigen-decomposition of $\boldsymbol{S}$ be $\boldsymbol{S} = \sum_{i=1}^N \lambda_i\, \boldsymbol{s}_i \boldsymbol{s}_i^\intercal$ with $\boldsymbol{s}_i \in \mathbb{R}^N$ eigenvectors of $\boldsymbol{S}$. The error can be expanded as:

$$\frac{1}{N}\big\|\boldsymbol{S} - \boldsymbol{\Xi}_S(\boldsymbol{Y})\big\|_{\mathrm{F}}^2 = \frac{1}{N}\sum_{i=1}^N \lambda_i^2 + \frac{1}{N}\sum_{i=1}^N \xi_{si}^{\,2} - \frac{2}{N}\sum_{i=1}^N \xi_{si} \sum_{j=1}^N \lambda_j \left(\boldsymbol{u}_i^\intercal \boldsymbol{s}_j\right)^2$$

where $\boldsymbol{u}_i$'s are columns of $\boldsymbol{U}_Y$. Minimizing over ξ_{si}'s, we find the optimum among the RIE class:

$$\boldsymbol{\Xi}_S^*(\boldsymbol{Y}) = \sum_{i=1}^N \xi_{s\,i}^*\, \boldsymbol{u}_i \boldsymbol{u}_i^\intercal, \quad \xi_{s\,i}^* = \sum_{j=1}^N \lambda_j \left(\boldsymbol{u}_i^\intercal \boldsymbol{s}_j\right)^2 = \boldsymbol{u}_i^\intercal \boldsymbol{S} \boldsymbol{u}_i \tag{11.3}$$

Expression (11.3) defines the Oracle estimator which requires the knowledge of signal matrix $\boldsymbol{S}$. Surprisingly, in the large N limit, the optimal eigenvalues $\big(\xi_{s\,i}^*\big)_{1\le i\le N}$ can be computed from the observation matrix and knowledge of the measures ρ_S, μ_T, μ_Z. In the next section, we show that this leads to an *explicitly computable* (or algorithmic) RIE, which we conjecture to be optimal as $N \to \infty$, in the sense that its performance matches the one of the Oracle estimator.

Now we remark that the Oracle estimator is not only optimal within the rotation invariant class but is also Bayesian optimal. From the Bayesian estimation point of view, one wishes to minimize the average mean squared error (MSE) $\mathrm{MSE}_{\hat{\boldsymbol{S}}} \equiv \frac{1}{N}\mathbb{E}\big\|\boldsymbol{S} - \hat{\boldsymbol{S}}(\boldsymbol{Y})\big\|_{\mathrm{F}}^2$, where the expectation is over $\boldsymbol{S}, \boldsymbol{T}, \boldsymbol{Z}$, and $\hat{\boldsymbol{S}}(\boldsymbol{Y})$ is an estimator of $\boldsymbol{S}$. The MSE is minimized for $\hat{\boldsymbol{S}}^*(\boldsymbol{Y}) = \mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}]$ which is the posterior mean. Therefore, the posterior mean estimator has the minimum MSE (MMSE) among all possible estimators, in particular $\mathrm{MSE}_{\hat{\boldsymbol{S}}^*} \le \mathrm{MSE}_{\boldsymbol{\Xi}_S^*}$ for any N. In appendix 11.A.1, we show that, for rotational invariant priors, the posterior mean estimator is inside the RIE class. Thus, since $\boldsymbol{\Xi}_S^*(\boldsymbol{Y})$ is optimum among the RIE class $\mathrm{MSE}_{\boldsymbol{\Xi}_S^*} \le \mathrm{MSE}_{\hat{\boldsymbol{S}}^*}$. Therefore, we conclude that the Oracle estimator (11.3) is Bayesian optimal in the sense that $\mathrm{MSE}_{\boldsymbol{\Xi}_S^*} = \mathrm{MSE}_{\hat{\boldsymbol{S}}^*} = \mathrm{MMSE}$.

RIE class for $\boldsymbol{T}$

Estimators for $\boldsymbol{T}$ from the rotation invariant class are denoted $\boldsymbol{\Xi}_T(\boldsymbol{Y})$, and are constructed as:

$$\boldsymbol{\Xi}_T(\boldsymbol{Y}) = \boldsymbol{U}_Y \left[\begin{array}{c|c} \mathrm{diag}(\xi_{t1}, \ldots, \xi_{tN}) & \boldsymbol{0}_{N\times(M-N)} \end{array} \right] \boldsymbol{V}_Y^\intercal \tag{11.4}$$

where $\xi_{t1}, \ldots, \xi_{tN}$ are the singular values of the estimator.

Let the SVD of $\boldsymbol{T}$ be $\boldsymbol{T} = \sum_{i=1}^N \sigma_i \boldsymbol{t}_i^{(l)} \boldsymbol{t}_i^{(r)\intercal}$ with $\boldsymbol{t}_i^{(l)} \in \mathbb{R}^N, \boldsymbol{t}_i^{(r)} \in \mathbb{R}^M$ the left and right singular vectors of $\boldsymbol{T}$. To derive an *Oracle estimator*, we proceed as above. Expanding the error, we have:

$$\frac{1}{N}\left\|\boldsymbol{T} - \boldsymbol{\Xi}_T(\boldsymbol{Y})\right\|_{\mathrm{F}}^2 = \frac{1}{N}\sum_{i=1}^N \sigma_i^2 + \frac{1}{N}\sum_{i=1}^N \xi_{ti}^{\,2} - \frac{2}{N}\sum_{i=1}^N \xi_{ti} \sum_{j=1}^N \sigma_j \left(\boldsymbol{u}_i^\intercal \boldsymbol{t}_j^{(l)}\right)\left(\boldsymbol{v}_i^\intercal \boldsymbol{t}_j^{(r)}\right)$$

where $\boldsymbol{v}_i$'s are columns of $\boldsymbol{V}_Y$. Minimizing over ξ_{ti}'s, we find the optimum among the RIE class:

$$\boldsymbol{\Xi}_T^*(\boldsymbol{Y}) = \sum_{i=1}^N \xi_{t\,i}^*\, \boldsymbol{u}_i \boldsymbol{v}_i^\intercal, \quad \xi_{t\,i}^* = \sum_{j=1}^N \sigma_j \left(\boldsymbol{u}_i^\intercal \boldsymbol{t}_j^{(l)}\right)\left(\boldsymbol{v}_i^\intercal \boldsymbol{t}_j^{(r)}\right) = \boldsymbol{u}_i^\intercal \boldsymbol{T} \boldsymbol{v}_i \tag{11.5}$$

Expression (11.5) defines the Oracle estimator which requires the knowledge of signal matrix $\boldsymbol{T}$. Like for the case of $\boldsymbol{S}$, in the large N limit we can derive the optimal singular values $\left(\xi_{t\,i}^*\right)_{1\le i\le N}$ in terms of the singular values of observation matrix and knowledge of the measures ρ_S, μ_Z. This leads to an *explicitly computable* (or algorithmic) RIE, which is conjectured to be optimal as $N \to \infty$, in the sense that it has the same performance as the Oracle estimator. Note that unlike the estimator for $\boldsymbol{S}$, we do not need the knowledge of μ_T.

In appendix 11.A.2, we show that for bi-rotationally invariant priors the posterior mean estimator $\hat{\boldsymbol{T}}^*(\boldsymbol{Y}) = \mathbb{E}[\boldsymbol{T}|\boldsymbol{Y}]$ belongs to the RIE class, which (by similar arguments to the case of $\boldsymbol{S}$) implies that the Oracle estimator (11.5) is Bayesian optimal.

11.2 Algorithmic RIEs for the Matrix Factors

In this section, we present our explicit RIEs for $\boldsymbol{S}, \boldsymbol{T}$ and the corresponding algorithm. We conjecture that their performance matches the one of Oracles estimators in the large N limit and they are therefore Bayesian optimal in this limit.

11.2.1 Explicit RIE for S

The RIE for $\boldsymbol{S}$ is constructed as $\widehat{\boldsymbol{\Xi}^*_S}(\boldsymbol{Y}) = \sum_{i=1}^N \widehat{\xi^*_{s\,i}} \boldsymbol{u}_i \boldsymbol{u}_i^\intercal$ with eigenvalues $\big(\widehat{\xi^*_{s\,i}}\big)_{1\le i\le N}$:

$$\widehat{\xi^*_{s\,i}} = \frac{1}{2\kappa\pi\bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z\to\gamma_i - \mathrm{i}0^+} \left\{ \frac{1}{\zeta_3} \Big[\mathcal{G}_{\rho_S}\Big(\sqrt{\frac{z-\zeta_1}{\kappa\zeta_3}} \Big) + \mathcal{G}_{\rho_S}\Big(-\sqrt{\frac{z-\zeta_1}{\kappa\zeta_3}} \Big) \Big] \right\} \tag{11.6}$$

where γ_i is the i-th singular value of $\boldsymbol{Y}$, $\bar{\mu}_Y$ is the symmetrized limiting ESD of $\boldsymbol{Y}$, and

$$\zeta_1 = \frac{1}{\mathcal{G}_{\bar{\mu}_Y}(z)} \mathcal{C}^{(\alpha)}_{\mu_Z}\Big(\mathcal{G}_{\bar{\mu}_Y}(z) \big[\alpha \mathcal{G}_{\bar{\mu}_Y}(z) + \frac{1-\alpha}{z} \big] \Big) \tag{11.7}$$

and ζ_3 satisfies [1]:

$$(z-\zeta_1)\mathcal{G}_{\bar{\mu}_Y}(z) - 1 = \mathcal{C}^{(\alpha)}_{\mu_T}\Big(\frac{1}{\zeta_3} \big[\alpha \mathcal{G}_{\bar{\mu}_Y}(z) + \frac{1-\alpha}{z} \big] \big[(z-\zeta_1)\mathcal{G}_{\bar{\mu}_Y}(z) - 1 \big] \Big) \tag{11.8}$$

Remark 11.2. *If ρ_S is a symmetric measure, $\rho_S(x) = \rho_S(-x)$, then $\mathcal{G}_{\rho_S}(-z) = -\mathcal{G}_{\rho_S}(z)$. This implies that the optimal eigenvalues $\big(\widehat{\xi^*_{s\,i}}\big)_{1\le i\le N}$ in* (11.6) *are all zero, and $\widehat{\boldsymbol{\Xi}^*_S}(\boldsymbol{Y}) = \boldsymbol{0}$, see figure 11.4.1.*

An estimator for $\boldsymbol{S}^2$

It is interesting to note that we can construct a RIE for $\boldsymbol{S}^2$ as $\widehat{\boldsymbol{\Xi}^*_{S^2}}(\boldsymbol{Y}) = \sum_{i=1}^N \widehat{\xi^*_{s^2 i}} \boldsymbol{u}_i \boldsymbol{u}_i^\intercal$ with eigenvalues $\big(\widehat{\xi^*_{s^2 i}}\big)_{1\le i\le N}$:

$$\widehat{\xi^*_{s^2 i}} = \frac{1}{\kappa}\frac{1}{\pi\bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z\to\gamma_i - \mathrm{i}0^+} \frac{z-\zeta_1}{\zeta_3}\mathcal{G}_{\bar{\mu}_Y}(z) - \frac{1}{\zeta_3} \tag{11.9}$$

with ζ_1, ζ_3 as in (11.7), (11.8). Note that, ζ_1, ζ_3 can be evaluated using the observation matrix and the knowledge of μ_T, μ_Z, and therefore this time the statistician *does not need to know the prior of* $\boldsymbol{S}$. Furthermore, assuming that $\boldsymbol{S}$ is positive semi-definite (PSD), we can construct a sub-optimal RIE for $\boldsymbol{S}$ by using $\sqrt{\widehat{\xi^*_{s^2 i}}}$ for the eigenvalues of the estimator.

Case of Gaussian $\boldsymbol{T}, \boldsymbol{Z}$

If $\boldsymbol{T}, \boldsymbol{Z}$ have i.i.d. Gaussian entries with variance $1/N$, then $\mathcal{C}^{(\alpha)}_{\mu_T}(z) = \mathcal{C}^{(\alpha)}_{\mu_Z}(z) = z/\alpha$. Consequently, ζ_1, ζ_3 can easily be computed to be $\zeta_1 = \zeta_3 = \mathcal{G}_{\bar{\mu}_Y}(z) + (1-\alpha)/(\alpha z)$, thus the estimator (11.6) can be evaluated from the observation matrix. In particular, the estimator (11.9) simplifies to:

$$\widehat{\xi^*_{s^2 i}} = \frac{1}{\kappa}\left[-1 + \frac{1}{\alpha\Big(\pi^2 \bar{\mu}_Y(\gamma_i)^2 + \big(\pi \mathsf{H}[\bar{\mu}_Y](\gamma_i) + \frac{1-\alpha}{\alpha\gamma_i} \big)^2 \Big)} \right] \tag{11.10}$$

[1] ζ_1, ζ_3 are the only parameters which appear in the final estimator. However, in derivation of the RIE, we have defined other parameters which do not appear in the final estimator and we omit them here.

11.2.2 Explicit RIE for $\boldsymbol{T}$

Our explicit RIE for $\boldsymbol{T}$ is constructed as $\widehat{\boldsymbol{\Xi}_T^*}(\boldsymbol{Y}) = \sum_{i=1}^N \widehat{\xi_{t\,i}^*}\boldsymbol{u}_i\boldsymbol{v}_i^\intercal$ with singular values $\big(\widehat{\xi_{t\,i}^*}\big)_{1\leq i\leq N}$:

$$\widehat{\xi_{t\,i}^*} = \frac{1}{\sqrt{\kappa}}\frac{1}{\pi\bar{\mu}_Y(\gamma_i)}\,\mathrm{Im}\lim_{z\to\gamma_i-\mathrm{i}0^+} q_4 \tag{11.11}$$

where γ_i is the i-th singular value of $\boldsymbol{Y}$, and q_4 is the solution to the following system of equations [2]:

$$\begin{cases} \beta_1 = \frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(q_1q_2)}{q_1} + \frac{1}{2}\sqrt{\frac{q_3}{q_1}}\Big(\mathcal{R}_{\rho_S}\big(q_4+\sqrt{q_1q_3}\big) - \mathcal{R}_{\rho_S}\big(q_4-\sqrt{q_1q_3}\big)\Big) \\ \beta_4 = \frac{1}{2}\Big(\mathcal{R}_{\rho_S}\big(q_4+\sqrt{q_1q_3}\big) + \mathcal{R}_{\rho_S}\big(q_4-\sqrt{q_1q_3}\big)\Big) \\ q_1 = \mathcal{G}_{\bar{\mu}_Y}(z), \quad q_2 = \alpha\mathcal{G}_{\bar{\mu}_Y}(z) + (1-\alpha)\frac{1}{z} \\ q_3 = \frac{(z-\beta_1)^2}{\beta_4^2}\mathcal{G}_{\bar{\mu}_Y}(z) - \frac{z-\beta_1}{\beta_4^2}, \quad q_4 = \frac{z-\beta_1}{\beta_4}\mathcal{G}_{\bar{\mu}_Y}(z) - \frac{1}{\beta_4} \end{cases} \tag{11.12}$$

Similarly to the estimator derived for $\boldsymbol{S}$, if ρ_S is a symmetric measure then the optimal singular values for the estimator of $\boldsymbol{T}$ are all zero, see remark 11.5.

If $\boldsymbol{S}$ is a shifted Wigner matrix, i.e. $\boldsymbol{S} = \boldsymbol{F} + c\boldsymbol{I}$ with $\boldsymbol{F} = \boldsymbol{F}^\intercal \in \mathbb{R}^{N\times N}$ having i.i.d. Gaussian entries with variance $1/N$ and $c \neq 0$ a real number, then $\mathcal{R}_{\rho_S}(z) = z + c$. Moreover, if $\boldsymbol{Z}$ is Gaussian matrix with variance $1/N$, then the set of equations (11.12) simplifies to a great extent, and we can compute q_4 analytically in terms of $\mathcal{G}_{\bar{\mu}_Y}(z)$, see appendix 11.C.4.

11.2.3 Algorithmic nature of the RIEs

The explicit RIEs (11.6) and (11.11) proposed in this section, provide spectral algorithms to estimate the matrix factors from the data matrix (and the priors). An essential ingredient that must be extracted from the data matrix is $\mathcal{G}_{\bar{\mu}_Y}(z)$. This quantity can be approximated from the observation matrix using Cauchy kernel method introduced in [67](see section 19.5.2), from which $\bar{\mu}_Y(.)$ can be approximated using (5.5). Therefore, given an observation matrix $\boldsymbol{Y}$, the spectral algorithm proceeds as follows:

1. Compute the SVD of $\boldsymbol{Y}$.

2. Approximate $\mathcal{G}_{\bar{\mu}_Y}(z)$ from the singular values of $\boldsymbol{Y}$.

3. Construct the RIEs for $\boldsymbol{S}, \boldsymbol{T}$ as proposed in paragraphs 11.2.1, 11.2.2.

[2] Like the case for $\boldsymbol{S}$, we omit some of the parameters which do not appear in the final estimator.

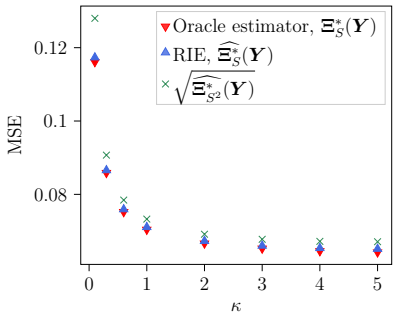


Figure 11.3.1: MSE of estimating $\boldsymbol{S}$. MSE is normalized by the norm of the signal, $\|\boldsymbol{S}\|_{\mathrm{F}}^2$. $\boldsymbol{S}$ is a Wishart matrix with aspect ratio $1/4$, $\boldsymbol{S} = \mathbf{H}\mathbf{H}^\intercal$ with $\mathbf{H} \in \mathbb{R}^{N\times 4N}$ having i.i.d. Gaussian entries of variance $1/N$. Both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$. RIE is applied to $N = 2000, M = 4000$, and the results are averaged over 10 runs (error bars are invisible).

11.3 Numerical Results

11.3.1 Performance of RIE for $\boldsymbol{S}$

We consider the case where $\boldsymbol{T}, \boldsymbol{Z}$ both have i.i.d. Gaussian entries of variance $1/N$, and $\boldsymbol{S}$ is a Wishart matrix, $\boldsymbol{S} = \mathbf{H}\mathbf{H}^\intercal$ with $\mathbf{H} \in \mathbb{R}^{N\times 4N}$ having i.i.d. Gaussian entries of variance $1/N$. For various SNRs, we examine the performance of two proposed estimators, the RIE (11.6), and the square-root of the estimator (11.9) (since $\boldsymbol{S}$ is PSD), which is sub-optimal. In figure 11.3.1, the MSEs of these algorithmic estimators are compared with the one of Oracle estimator (11.3). We see that the average performance of the algorithmic RIE $\widehat{\Xi^*_S}(\boldsymbol{Y})$ is very close to the (optimal) Oracle estimator $\Xi^*_S(\boldsymbol{Y})$ (relative errors are small and provided in the appendices) and we believe that the slight mismatch is due to the numerical approximations and finite-size effects. Note that, although the estimator $\sqrt{\widehat{\Xi^*_{X^2}}(\boldsymbol{Y})}$ is sub-optimal, it does not use any prior knowledge of $\boldsymbol{S}$. For more examples, details of the numerical experiments and the relative error of the estimators, we refer to appendix 11.B.3.

11.3.2 Performance of RIE for $\boldsymbol{T}$

We consider the case where $\boldsymbol{Z}$ has i.i.d. Gaussian entries of variance $1/N$, and $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 3$. Matrix $\boldsymbol{T}$ is constructed as $\boldsymbol{T} = \boldsymbol{U}_T\boldsymbol{\Sigma}\boldsymbol{V}_T^\intercal$ with $\boldsymbol{U}_T \in \mathbb{R}^{N\times N}, \boldsymbol{V}_T \in \mathbb{R}^{M\times M}$ are Haar distributed, and the singular values are generated independently from the uniform distribution on $[1, 3]$. MSEs of the RIE (11.11) and the Oracle estimator (11.5) are illustrated in figure 11.3.2. We see that the performance of the algorithmic RIE $\widehat{\Xi^*_T}(\boldsymbol{Y})$ is very close to the optimal estimator $\Xi^*_T(\boldsymbol{Y})$.

Non-rotational invariant prior In another example, which we omit here, with the same settings for $\boldsymbol{S}, \boldsymbol{Z}$, we consider the case where $\boldsymbol{T}$ is a sparse

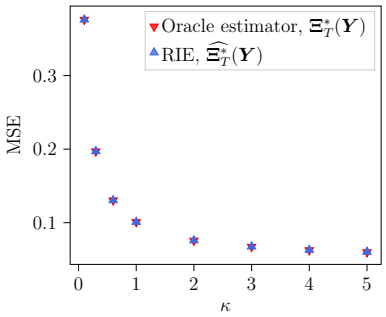


Figure 11.3.2: MSE of estimating $\boldsymbol{T}$. MSE is normalized by the norm of the signal, $\|\boldsymbol{T}\|_{\mathrm{F}}^2$. $\boldsymbol{T}$ has uniform spectral density, $\mathcal{U}\big([1,3]\big)$. $\boldsymbol{S}$ is a shifted Wigner matrix with $c=3$, and $\boldsymbol{Z}$ is a $N \times M$ matrix with i.i.d. Gaussian entries of variance $1/N$. RIE is applied to $N=2000, M=4000$, and the results are averaged over 10 runs (error bars are invisible).

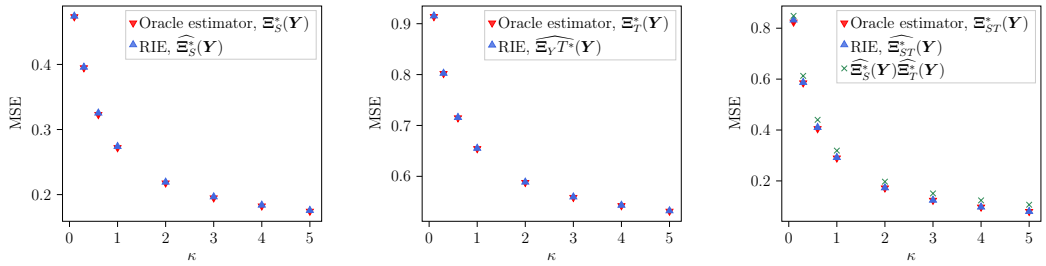
matrix with entries distributed according to Bernoulli-Rademacher prior. The RIE is not optimal in this setting (since the prior is not bi-rotational invariant), however applying a simple thresholding function on the matrix constructed by RIE yields an estimate with lower MSE. This observation suggests that for the case of general priors, the RIEs can provide a spectral initialization for more efficient estimators. For more details and examples, see appendix 11.C.4.

11.3.3 Comparing RIEs of matrix factorization and denoising

The proposed RIEs, namely (11.6) and (11.11), simplify greatly when the matrices $\boldsymbol{Z}, \boldsymbol{T}$ are Gaussian, and $\boldsymbol{S}$ is a shifted Wigner matrix. We perform experiments with these priors, where for a given observation matrix $\boldsymbol{Y}$, we look at the RIEs of $\boldsymbol{S}$, $\boldsymbol{T}$ for the *MF problem*, and simultaneously at the RIE of the product $\boldsymbol{ST}$ as a whole for the *denoising problem* with formulas introduced in [54] (which can also be obtained by taking $\boldsymbol{S}$ to be the identity matrix, see appendix 11.C.3). Figure 11.3.3 illustrates these experiments. In particular the MSE of the denoising-RIE matches well the one of the associated Oracle estimator, and as expected is lower than the MSE of the product of MF-RIEs.

11.4 Derivation of the Explicit RIEs

In this section, we sketch the derivation of our explicit RIE for $\boldsymbol{S}$. The RIE for $\boldsymbol{T}$ is derived similarly, but requires more involved analysis and is presented in appendix 11.C. For simplicity, we take the SNR parameter in (11.1) to be 1, so the model is $\boldsymbol{Y} = \boldsymbol{ST} + \boldsymbol{Z}$. The optimal eigenvalues are constructed as $\xi_{s\,i}^* = \sum_{j=1}^N \lambda_j \big(\boldsymbol{u}_i^\intercal \boldsymbol{s}_j\big)^2$. We assume that in the large N limit, $\xi_{s\,i}^*$ can be



(a) Estimating $\boldsymbol{S}$ **(b)** Estimating $\boldsymbol{T}$ **(c)** Estimating $\boldsymbol{ST}$

Figure 11.3.3: MSE of factorization problem. MSE is normalized by the norm of the signal. $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 1$, and both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$. RIE is applied to $N = 2000, M = 4000$. In each run, the observation matrix $\boldsymbol{Y}$ is generated according to (11.1), and the factors $\boldsymbol{S}$, $\boldsymbol{T}$ are estimated simultaneously from $\boldsymbol{Y}$. Results are averaged over 10 runs (error bars are invisible).

approximated by its expectation and we introduce

$$\widehat{\xi}^*_{s\,i} = \sum_{j=1}^{N} \lambda_j \, \mathbb{E}\Big[\big(\boldsymbol{u}_i^\intercal \boldsymbol{s}_j\big)^2\Big] \tag{11.13}$$

where the expectation is over the (left) singular vectors of the observation matrix $\boldsymbol{Y}$. Therefore, to compute these eigenvalues, we need to find the mean squared overlap $\mathbb{E}\Big[\big(\boldsymbol{u}_i^\intercal \boldsymbol{s}_j\big)^2\Big]$ between eigenvectors of $\boldsymbol{S}$ and singular vectors of $\boldsymbol{Y}$. In what follows, we will see that (a rescaling of) this quantity can be expressed in terms of i-th singular value of $\boldsymbol{Y}$ and j-th eigenvector of $\boldsymbol{S}$ (and the limiting measures, indeed). Thus, we will use the notation $O_S(\gamma_i, \lambda_j) := N\mathbb{E}\Big[\big(\boldsymbol{u}_i^\intercal \boldsymbol{s}_j\big)^2\Big]$ in the following. In the next section, we discuss how the overlap can be computed from the resolvent of the "Hermitized" version of $\boldsymbol{Y}$.

11.4.1 Relation between overlap and resolvent

Construct the matrix $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{(N+M)\times(N+M)}$ from the observation matrix:

$$\boldsymbol{\mathcal{Y}} = \begin{bmatrix} \boldsymbol{0}_{N\times N} & \boldsymbol{Y} \\ \boldsymbol{Y}^\intercal & \boldsymbol{0}_{M\times M} \end{bmatrix}$$

By Theorem 7.3.3 in [71], $\boldsymbol{\mathcal{Y}}$ has the following eigen-decomposition:

$$\boldsymbol{\mathcal{Y}} = \begin{bmatrix} \hat{\boldsymbol{U}}_Y & \hat{\boldsymbol{U}}_Y & \boldsymbol{0} \\ \hat{\boldsymbol{V}}_Y^{(1)} & -\hat{\boldsymbol{V}}_Y^{(1)} & \boldsymbol{V}_Y^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Gamma}_N & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{\Gamma}_N & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{U}}_Y & \hat{\boldsymbol{U}}_Y & \boldsymbol{0} \\ \hat{\boldsymbol{V}}_Y^{(1)} & -\hat{\boldsymbol{V}}_Y^{(1)} & \boldsymbol{V}_Y^{(2)} \end{bmatrix}^\intercal \tag{11.14}$$

with $\boldsymbol{V}_Y = \begin{bmatrix} \boldsymbol{V}_Y^{(1)} & \boldsymbol{V}_Y^{(2)} \end{bmatrix}$ in which $\boldsymbol{V}_Y^{(1)} \in \mathbb{R}^{M\times N}$. And, $\hat{\boldsymbol{V}}_Y^{(1)} = \frac{1}{\sqrt{2}}\boldsymbol{V}_Y^{(1)}$, $\hat{\boldsymbol{U}}_Y = \frac{1}{\sqrt{2}}\boldsymbol{U}_Y$. Eigenvalues of $\boldsymbol{\mathcal{Y}}$ are signed singular values of $\boldsymbol{Y}$, therefore the

limiting eigenvalue distribution of $\boldsymbol{\mathcal{Y}}$ (ignoring zero eigenvalues) is the same as the limiting symmetrized singular value distribution of $\boldsymbol{Y}$. Define the resolvent of $\boldsymbol{\mathcal{Y}}$,

$$\boldsymbol{G}_{\mathcal{Y}}(z) = (z\boldsymbol{I} - \boldsymbol{\mathcal{Y}})^{-1}$$

We assume that as $N \to \infty$ and z is not too close to the real axis, the matrix $\boldsymbol{G}_{\mathcal{Y}}(z)$ concentrates around its mean. Consequently, the value of $\boldsymbol{G}_{\mathcal{Y}}(z)$ becomes uncorrelated with the particular realization of $\boldsymbol{Y}$. Specifically, as $N \to \infty$, $\boldsymbol{G}_{\mathcal{Y}}(z)$ converges to a deterministic matrix for any fixed value of $z \in \mathbb{C}\backslash\mathbb{R}$ (independent of N). Denote the eigenvectors of $\boldsymbol{\mathcal{Y}}$ by $\boldsymbol{y}_i \in \mathbb{R}^{M+N}$, $i = 1, \ldots, M+N$. For $z = x - \mathsf{i}\epsilon$ with $x \in \mathbb{R}$ and small ϵ, we have:

$$\boldsymbol{G}_{\mathcal{Y}}(x - \mathsf{i}\epsilon) = \sum_{k=1}^{2N} \frac{x + \mathsf{i}\epsilon}{(x - \tilde{\gamma}_k)^2 + \epsilon^2} \boldsymbol{y}_k \boldsymbol{y}_k^{\mathsf{T}} + \frac{x + \mathsf{i}\epsilon}{x^2 + \epsilon^2} \sum_{k=2N+1}^{N+M} \boldsymbol{y}_k \boldsymbol{y}_k^{\mathsf{T}}$$

where $\tilde{\gamma}_k$ are the eigenvalues of $\boldsymbol{\mathcal{Y}}$, which are in fact the (signed) singular values of $\boldsymbol{Y}$, $\tilde{\gamma}_1 = \gamma_1, \ldots, \tilde{\gamma}_N = \gamma_N, \tilde{\gamma}_{N+1} = -\gamma_1, \ldots, \tilde{\gamma}_{2N} = -\gamma_N$.

Define the vectors $\tilde{\boldsymbol{s}}_i = [\boldsymbol{s}_i^{\mathsf{T}}, \boldsymbol{0}_M]^{\mathsf{T}}$ for $\boldsymbol{s}_i$ eigenvectors of $\boldsymbol{S}$. We have

$$\tilde{\boldsymbol{s}}_i^{\mathsf{T}} \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathsf{i}\epsilon)\big) \tilde{\boldsymbol{s}}_i = \sum_{k=1}^{2N} \frac{\epsilon}{(x - \tilde{\gamma}_k)^2 + \epsilon^2} \big(\tilde{\boldsymbol{s}}_i^{\mathsf{T}} \boldsymbol{y}_k\big)^2 + \frac{\epsilon}{x^2 + \epsilon^2} \sum_{k=2N+1}^{N+M} \big(\tilde{\boldsymbol{s}}_i^{\mathsf{T}} \boldsymbol{y}_k\big)^2 \tag{11.15}$$

Given the structure of $\boldsymbol{y}_k$'s in (11.14), $\big(\tilde{\boldsymbol{s}}_i^{\mathsf{T}} \boldsymbol{y}_j\big)^2 = \frac{1}{2}\big(\boldsymbol{s}_i^{\mathsf{T}} \boldsymbol{u}_j\big)^2 = \big(\tilde{\boldsymbol{s}}_i^{\mathsf{T}} \boldsymbol{y}_{j+N}\big)^2$ for $1 \le j \le N$, and the second sum in (11.15) is zero. We assume that in the limit of large N this quantity concentrates on $O_S(\gamma_j, \lambda_i)$ and depends only on the singular values and eigenvalue pairs (γ_j, λ_i). We thus have:

$$\tilde{\boldsymbol{s}}_i^{\mathsf{T}} \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathsf{i}\epsilon)\big) \tilde{\boldsymbol{s}}_i \xrightarrow{N\to\infty} \int_{\mathbb{R}} \frac{\epsilon}{(x-t)^2 + \epsilon^2} O_S(t, \lambda_i) \bar{\mu}_Y(t)\, dt \tag{11.16}$$

where the overlap function $O_S(t, \lambda_i)$ is extended (continuously) to arbitrary values within the support of $\bar{\mu}_Y$ (the symmetrized limiting singular value distribution of $\boldsymbol{Y}$) with the property that $O_S(t, \lambda_i) = O_S(-t, \lambda_i)$ for $t \in \operatorname{supp}(\mu_Y)$. Sending $\epsilon \to 0$, we find

$$\tilde{\boldsymbol{s}}_i^{\mathsf{T}} \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathsf{i}\epsilon)\big) \tilde{\boldsymbol{s}}_i \to \pi \bar{\mu}_Y(x) O_S(x, \lambda_i) \tag{11.17}$$

This is a crucial relation as it allows us to study the overlap by means of the resolvent of $\boldsymbol{\mathcal{Y}}$. In the next section, we establish a connection between this resolvent and the signal $\boldsymbol{S}$, which enables us to determine the optimal eigenvalues values $\widehat{\xi}_{si}^*$ in terms of the singular values of $\boldsymbol{Y}$.

11.4.2 Resolvent relation

To derive the resolvent relation between $\boldsymbol{Y}$ and $\boldsymbol{S}$, we fix the matrix $\boldsymbol{S}$ and consider the model

$$\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{U}_1 \boldsymbol{T} \boldsymbol{V}_1^{\mathsf{T}} + \boldsymbol{U}_2 \boldsymbol{Z} \boldsymbol{V}_2^{\mathsf{T}}$$

with $\boldsymbol{T}, \boldsymbol{Z} \in \mathbb{R}^{N\times M}$ fixed matrices with limiting singular value distribution μ_T, μ_Z, and $\boldsymbol{U}_1, \boldsymbol{U}_2 \in \mathbb{R}^{N\times N}, \boldsymbol{V}_1, \boldsymbol{V}_2 \in \mathbb{R}^{M\times M}$ independent random Haar matrices. Indeed, if we substitute the SVD of the matrices $\boldsymbol{T}, \boldsymbol{Z}$ in model (11.1) we find the latter model. Now, the average over the singular vectors of $\boldsymbol{Y}$ (with fixed $\boldsymbol{S}$) is equivalent to the average over the matrices $\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{V}_1, \boldsymbol{V}_2$. In appendix 11.B.1, using the Replica trick, we derive the following relation in the limit $N \to \infty$:

$$\left\langle \boldsymbol{G}_{\mathcal{Y}}(z) \right\rangle = \begin{bmatrix} \zeta_3^{-1} \boldsymbol{G}_{S^2}\left(\frac{z-\zeta_1}{\zeta_3}\right) & \boldsymbol{0} \\ \boldsymbol{0} & (z-\zeta_2)^{-1} \boldsymbol{I}_M \end{bmatrix} \tag{11.18}$$

with $\zeta_1, \zeta_2, \zeta_3$ satisfying set of equations (11.37). $\langle . \rangle$ is the expectation w.r.t. the singular vectors of $\boldsymbol{Y}$ (or equivalently over $\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{V}_1, \boldsymbol{V}_2$), and $\boldsymbol{G}_{S^2}$ is the resolvent of $\boldsymbol{S}^2$. As stated earlier, we assume that the resolvent $\boldsymbol{G}_{\mathcal{Y}}(z)$ concentrates in the limit $N \to \infty$, therefore we drop the brackets in the following computation.

11.4.3 Overlaps and optimal eigenvalues

From (11.17), (11.18), we find:

$$\begin{aligned} O_S(\gamma, \lambda_i) &\approx \frac{1}{\pi \bar{\mu}_Y(\gamma)} \operatorname{Im} \lim_{z \to \gamma - \mathrm{i}0^+} \boldsymbol{s}_i^\intercal \, \zeta_3^{-1} \boldsymbol{G}_{S^2}\left(\frac{z-\zeta_1}{\zeta_3}\right) \boldsymbol{s}_i \\ &= \frac{1}{\pi \bar{\mu}_Y(\gamma)} \operatorname{Im} \lim_{z \to \gamma - \mathrm{i}0^+} \frac{1}{z - \zeta_1 - \zeta_3 \lambda_i^2} \end{aligned} \tag{11.19}$$

In Fig. 11.4.1 we illustrate that the theoretical predictions (11.19) are in good agreement with numerical simulations for a particular case of $\boldsymbol{S}$ a Wigner matrix, and $\boldsymbol{T}, \boldsymbol{Z}$ with i.i.d. Gaussian entries.

Once we have the overlap, we can compute the optimal eigenvalues to be

$$\widehat{\xi}_{si}^{*} \approx \frac{1}{N} \sum_{j=1}^{N} \lambda_j O_S(\gamma_i, \lambda_j) \approx \frac{1}{\pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z \to \gamma_i - \mathrm{i}0^+} \frac{1}{N} \sum_{j=1}^{N} \frac{\lambda_j}{z - \zeta_1 - \zeta_3 \lambda_j^2} \tag{11.20}$$

With a bit of algebra, we find the estimator in (11.6) in the limit $N \to \infty$, see appendix 11.B.2.

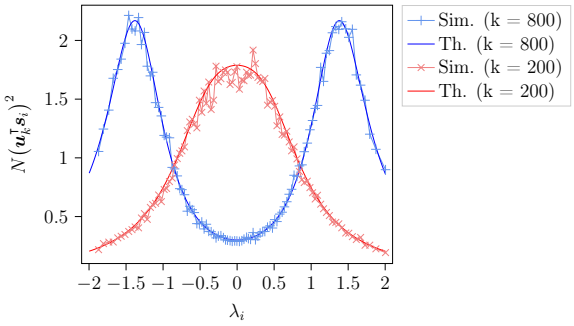


Figure 11.4.1: Comparison of the theoretical prediction (11.19) of the rescaled overlap with the numerical simulation. The rescaled overlap between 200-th and 800-th left singular vector of $\boldsymbol{Y}$ and the eigenvectors of $\boldsymbol{S}$ is illustrated. $\boldsymbol{S} = \boldsymbol{S}^\intercal \in \mathbb{R}^{N\times N}$ has i.i.d. Gaussian entries with variance $1/\sqrt{N}$ and is fixed. Both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$. The simulation results are average of 1000 experiments with fixed $\boldsymbol{S}$, and $N = 1000, M = 2000$. Some of the simulation points are dropped for clarity.
One can see that the overlap is an even function of eigenvalues λ_i, so the optimal eigenvalues $\xi^*_{s\,i} = \sum_{j=1}^N \lambda_j \left(\boldsymbol{u}_i^\intercal \boldsymbol{s}_j\right)^2$ are all zero, as discussed in remark 11.2.

Appendix

11.A Posterior Mean Estimator is in the RIE Class

In this section, we show that for rotational invariant priors, the posterior mean estimator is inside the RIE class. For each of the estimators of $\boldsymbol{S}, \boldsymbol{T}$, we present an equivalent definition of the RIE, then we show that posterior mean estimator satisfies this definition.

11.A.1 S Estimator

Lemma 11.1. *Given the observation matrix $\boldsymbol{Y}$, let $\hat{\boldsymbol{S}}(\boldsymbol{Y})$ be an estimator of $\boldsymbol{S}$. Then $\hat{\boldsymbol{S}}(\boldsymbol{Y})$ is a RIE if and only if for any orthogonal matrices $\boldsymbol{U} \in \mathbb{R}^{N\times N}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$:*

$$\hat{\boldsymbol{S}}(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal) = \boldsymbol{U}\hat{\boldsymbol{S}}(\boldsymbol{Y})\boldsymbol{U}^\intercal \tag{11.21}$$

Proof. If $\hat{\boldsymbol{S}}(\boldsymbol{Y})$ is a RIE, then the property (11.21) clearly follows from the definition (11.2). Now we turn to the converse.

Suppose that an estimator $\hat{\boldsymbol{S}}(\boldsymbol{Y})$ satisfies (11.21). First, we show that if the observation matrix is diagonal, then the estimator is also diagonal. Consider the observation matrix to be $\boldsymbol{Y}^{\text{diag}} = \left[\ \text{diag}(\text{y}_1, \ldots, \text{y}_\text{N}) \ \middle|\ \boldsymbol{0}_{N\times(M-N)}\ \right]$. Let $\boldsymbol{I}_k^- \in \mathbb{R}^{N\times N}, \boldsymbol{J}_k^- \in \mathbb{R}^{M\times M}$ be diagonal matrices with diagonal entries all one except the k-th entry which is -1. Note that for $1 \leq k \leq N$, we have $\boldsymbol{Y}^{\text{diag}} = \boldsymbol{I}_k^- \boldsymbol{Y}^{\text{diag}} \boldsymbol{J}_k^-$. Moreover, matrices $\boldsymbol{I}_k^-, \boldsymbol{J}_k^-$ are indeed orthogonal. For any $1 \leq k \leq N$, from the property we have:

$$\hat{\boldsymbol{S}}(\boldsymbol{Y}^{\text{diag}}) = \hat{\boldsymbol{S}}(\boldsymbol{I}_k^- \boldsymbol{Y}^{\text{diag}} \boldsymbol{J}_k^-) = \boldsymbol{I}_k^- \hat{\boldsymbol{S}}(\boldsymbol{Y}^{\text{diag}}) \boldsymbol{I}_k^- \tag{11.22}$$

This implies that all entries on the k-th row and k-th column of $\hat{\boldsymbol{S}}(\boldsymbol{Y}^{\text{diag}})$ are zero except the k-th entry on the diagonal. Since this holds for any k, we conclude that $\hat{\boldsymbol{S}}(\boldsymbol{Y}^{\text{diag}})$ is diagonal.

Now, for a given general observation matrix with SVD $\boldsymbol{Y} = \boldsymbol{U}_Y \boldsymbol{\Gamma} \boldsymbol{V}_Y^\intercal$, put $\boldsymbol{U} = \boldsymbol{U}_Y^\intercal, \boldsymbol{V} = \boldsymbol{V}_Y^\intercal$ in the property (11.21). We have:

$$\hat{\boldsymbol{S}}(\boldsymbol{\Gamma}) = \boldsymbol{U}_Y^\intercal \hat{\boldsymbol{S}}(\boldsymbol{Y}) \boldsymbol{U}_Y$$

From the argument above, the matrix on the lhs is diagonal. Consequently, the matrix $\boldsymbol{U}_Y^\intercal \hat{\boldsymbol{S}}(\boldsymbol{Y})\boldsymbol{U}_Y$ is diagonal which implies that the columns of $\boldsymbol{U}_Y$ are eigenvectors of $\hat{\boldsymbol{S}}(\boldsymbol{Y})$. Therefore, $\hat{\boldsymbol{S}}(\boldsymbol{Y})$ is a RIE. □

Now, we prove that the posterior mean estimator $\hat{\boldsymbol{S}}^*(\boldsymbol{Y}) = \mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}]$ has the property (11.21), and therefore belongs to the RIE class. For simplicity, we drop the SNR factor $\sqrt{\kappa}$. For any orthogonal matrices $\boldsymbol{U} \in \mathbb{R}^{N\times N}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$, we have:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{S}|\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal] &= \frac{\int d\boldsymbol{T}\, d\boldsymbol{X}\, \boldsymbol{X}\, P_S(\boldsymbol{X})P_T(\boldsymbol{T})P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{X}\boldsymbol{T})}{\int d\boldsymbol{T}\, d\boldsymbol{X}\, P_S(\boldsymbol{X})P_T(\boldsymbol{T})P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{X}\boldsymbol{T})} \\
&\overset{\text{(a)}}{=} \frac{\int d\boldsymbol{T}\, d\boldsymbol{X}\, \boldsymbol{U}\boldsymbol{X}\boldsymbol{U}^\intercal\, P_S(\boldsymbol{X})P_T(\boldsymbol{T})P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{U}\boldsymbol{X}\boldsymbol{U}^\intercal\boldsymbol{T})}{\int d\boldsymbol{T}\, d\boldsymbol{X}\, P_S(\boldsymbol{X})P_T(\boldsymbol{T})P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{U}\boldsymbol{X}\boldsymbol{U}^\intercal\boldsymbol{T})} \\
&\overset{\text{(b)}}{=} \frac{\int d\boldsymbol{T}\, d\boldsymbol{X}\, \boldsymbol{U}\boldsymbol{X}\boldsymbol{U}^\intercal\, P_S(\boldsymbol{X})P_T(\boldsymbol{T})P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{U}\boldsymbol{X}\boldsymbol{U}^\intercal\boldsymbol{U}\boldsymbol{T}\boldsymbol{V}^\intercal)}{\int d\boldsymbol{T}\, d\boldsymbol{X}\, P_S(\boldsymbol{X})P_T(\boldsymbol{T})P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{U}\boldsymbol{X}\boldsymbol{U}^\intercal\boldsymbol{U}\boldsymbol{T}\boldsymbol{V}^\intercal)} \\
&\overset{\text{(c)}}{=} \boldsymbol{U}\Big\{\frac{\int d\boldsymbol{T}\, d\boldsymbol{X}\, \boldsymbol{X}\, P_S(\boldsymbol{X})P_T(\boldsymbol{T})P_Z(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{T})}{\int d\boldsymbol{T}\, d\boldsymbol{X}\, P_S(\boldsymbol{X})P_T(\boldsymbol{T})P_Z(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{T})}\Big\}\boldsymbol{U}^\intercal \\
&= \boldsymbol{U}\mathbb{E}[\boldsymbol{S}|\boldsymbol{Y}]\boldsymbol{U}^\intercal
\end{aligned}$$

where in (a), we changed variables $\boldsymbol{X} \to \boldsymbol{U}\boldsymbol{X}\boldsymbol{U}^\intercal$, used $|\det \boldsymbol{U}| = 1$, and rotational invariance of P_S, $P_S(\boldsymbol{X}) = P_S(\boldsymbol{U}\boldsymbol{X}\boldsymbol{U}^\intercal)$. In (b), we changed variables $\boldsymbol{T} \to \boldsymbol{U}\boldsymbol{T}\boldsymbol{V}^\intercal$, used $|\det \boldsymbol{U}| = |\det \boldsymbol{V}| = 1$, and bi-rotational invariance of P_T, $P_T(\boldsymbol{T}) = P_T(\boldsymbol{U}\boldsymbol{T}\boldsymbol{V}^\intercal)$. In (c), we used the bi-rotational invariance property of P_Z, namely $P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{U}\boldsymbol{X}\boldsymbol{T}\boldsymbol{V}^\intercal) = P_Z(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{T})$.

11.A.2 T Estimator

Lemma 11.2. *Given the observation matrix $\boldsymbol{Y}$, let $\hat{\boldsymbol{T}}(\boldsymbol{Y})$ be an estimator for $\boldsymbol{T}$. Then $\hat{\boldsymbol{T}}(\boldsymbol{Y})$ is a RIE if and only if for any orthogonal matrices $\boldsymbol{U} \in \mathbb{R}^{N\times N}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$:*

$$\hat{\boldsymbol{T}}(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal) = \boldsymbol{U}\hat{\boldsymbol{T}}(\boldsymbol{Y})\boldsymbol{V}^\intercal \tag{11.23}$$

Proof. If $\hat{\boldsymbol{T}}(\boldsymbol{Y})$ is a RIE, then this property clearly follows from the definition (11.4). Let us now show the converse.

Suppose that an estimator $\hat{\boldsymbol{T}}(\boldsymbol{Y})$ satisfies (11.23). First, we show that if the observation matrix is diagonal, then the estimator is also diagonal. Consider the observation matrix to be $\boldsymbol{Y}^{\text{diag}} = \big[\ \text{diag}(\mathrm{s}_1, \ldots, \mathrm{s}_N) \ \big|\ \boldsymbol{0}_{N\times(M-N)}\ \big]$. Let $\boldsymbol{I}_k^- \in \mathbb{R}^{N\times N}, \boldsymbol{J}_k^- \in \mathbb{R}^{M\times M}$ be diagonal matrices with diagonal entries all one except the k-th entry which is -1. Note that for $1 \le k \le N$, we have $\boldsymbol{Y}^{\text{diag}} = \boldsymbol{I}_k^-\boldsymbol{Y}^{\text{diag}}\boldsymbol{J}_k^-$. Moreover, matrices $\boldsymbol{I}_k^-, \boldsymbol{J}_k^-$ are indeed orthogonal. For any $1 \le k \le N$, from the property we have:

$$\hat{\boldsymbol{T}}(\boldsymbol{Y}^{\text{diag}}) = \hat{\boldsymbol{T}}(\boldsymbol{I}_k^-\boldsymbol{Y}^{\text{diag}}\boldsymbol{J}_k^-) = \boldsymbol{I}_k^-\hat{\boldsymbol{T}}(\boldsymbol{Y}^{\text{diag}})\boldsymbol{J}_k^- \tag{11.24}$$

This implies that all entries on the k-th row and k-th column of $\hat{\boldsymbol{T}}(\boldsymbol{Y}^{\text{diag}})$ is zero except the k-th entry on the diagonal. Since this holds for any k, we conclude that $\hat{\boldsymbol{T}}(\boldsymbol{Y}^{\text{diag}})$ is diagonal.

Now, for a given general observation matrix $\boldsymbol{Y} = \boldsymbol{U}_Y \boldsymbol{\Gamma} \boldsymbol{V}_Y^\intercal$, put $\boldsymbol{U} = \boldsymbol{U}_Y^\intercal, \boldsymbol{V} = \boldsymbol{V}_Y^\intercal$ in the property (11.23). We have:

$$\hat{\boldsymbol{T}}(\boldsymbol{\Gamma}) = \boldsymbol{U}_Y^\intercal \hat{\boldsymbol{T}}(\boldsymbol{Y}) \boldsymbol{V}_Y$$

From the argument above, the matrix on the lhs is diagonal. Consequently, the matrix $\boldsymbol{U}_Y^\intercal \hat{\boldsymbol{T}}(\boldsymbol{Y}) \boldsymbol{V}_Y$ is diagonal which implies that the columns of $\boldsymbol{U}_Y, \boldsymbol{V}_Y$ are the left and right singular vectors of $\hat{\boldsymbol{T}}(\boldsymbol{Y})$. Therefore, $\hat{\boldsymbol{T}}(\boldsymbol{Y})$ is a RIE. □

Now, we prove that the posterior mean estimator $\hat{\boldsymbol{T}}^*(\boldsymbol{Y}) = \mathbb{E}[\boldsymbol{T}|\boldsymbol{Y}]$ has the property (11.23), and it is inside the RIE class. For simplicity, we drop the SNR factor $\sqrt{\kappa}$. For any orthogonal matrices $\boldsymbol{U} \in \mathbb{R}^{N\times N}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$, we have:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{T}|\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal] &= \frac{\int d\boldsymbol{S}\, d\boldsymbol{X}\, \boldsymbol{X}\, P_S(\boldsymbol{S}) P_T(\boldsymbol{X}) P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{S}\boldsymbol{X})}{\int d\boldsymbol{S}\, d\boldsymbol{X}\, P_S(\boldsymbol{S}) P_T(\boldsymbol{X}) P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{S}\boldsymbol{X})} \\
&\overset{\text{(a)}}{=} \frac{\int d\boldsymbol{S}\, d\boldsymbol{X}\, \boldsymbol{U}\boldsymbol{X}\boldsymbol{V}^\intercal\, P_S(\boldsymbol{S}) P_T(\boldsymbol{X}) P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{S}\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}^\intercal)}{\int d\boldsymbol{S}\, d\boldsymbol{X}\, P_S(\boldsymbol{S}) P_T(\boldsymbol{X}) P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{S}\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}^\intercal)} \\
&\overset{\text{(b)}}{=} \frac{\int d\boldsymbol{S}\, d\boldsymbol{X}\, \boldsymbol{U}\boldsymbol{X}\boldsymbol{V}^\intercal\, P_S(\boldsymbol{S}) P_T(\boldsymbol{X}) P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^\intercal\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}^\intercal)}{\int d\boldsymbol{S}\, d\boldsymbol{X}\, P_S(\boldsymbol{S}) P_T(\boldsymbol{X}) P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^\intercal\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}^\intercal)} \\
&\overset{\text{(c)}}{=} \boldsymbol{U} \Big\{ \frac{\int d\boldsymbol{S}\, d\boldsymbol{X}\, \boldsymbol{X}\, P_S(\boldsymbol{S}) P_T(\boldsymbol{X}) P_Z(\boldsymbol{Y} - \boldsymbol{S}\boldsymbol{X})}{\int d\boldsymbol{S}\, d\boldsymbol{X}\, P_S(\boldsymbol{S}) P_T(\boldsymbol{X}) P_Z(\boldsymbol{Y} - \boldsymbol{S}\boldsymbol{X})} \Big\} \boldsymbol{V}^\intercal \\
&= \boldsymbol{U}\mathbb{E}[\boldsymbol{T}|\boldsymbol{Y}]\boldsymbol{V}^\intercal
\end{aligned}$$

where in (a), we changed variables $\boldsymbol{X} \to \boldsymbol{U}\boldsymbol{X}\boldsymbol{V}^\intercal$, used $|\det \boldsymbol{U}| = |\det \boldsymbol{V}| = 1$, and bi-rotational invariance of P_T, $P_T(\boldsymbol{X}) = P_T(\boldsymbol{U}\boldsymbol{X}\boldsymbol{V}^\intercal)$. In (b), we changed variables $\boldsymbol{S} \to \boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^\intercal$, used $|\det \boldsymbol{U}| = 1$, and rotational invariance of P_S, $P_S(\boldsymbol{S}) = P_S(\boldsymbol{U}\boldsymbol{S}\boldsymbol{U}^\intercal)$. In (c), we used the bi-rotational invariance property of P_Z, namely $P_Z(\boldsymbol{U}\boldsymbol{Y}\boldsymbol{V}^\intercal - \boldsymbol{U}\boldsymbol{S}\boldsymbol{X}\boldsymbol{V}^\intercal) = P_Z(\boldsymbol{Y} - \boldsymbol{S}\boldsymbol{X})$.

11.B Derivation of the RIE for S

In this section, we consider estimating $\boldsymbol{S}$, and treat both $\boldsymbol{T}$ and $\boldsymbol{Z}$ as noise. We consider $\boldsymbol{S}$ to be fixed, and the observation model:

$$\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{U}_1\boldsymbol{T}\boldsymbol{V}_1^\intercal + \boldsymbol{U}_2\boldsymbol{Z}\boldsymbol{V}_2^\intercal \tag{11.25}$$

where $\boldsymbol{T}, \boldsymbol{Z} \in \mathbb{R}^{N\times M}$ are fixed matrices with limiting singular value distribution μ_T, μ_Z, and $\boldsymbol{U}_1, \boldsymbol{U}_2 \in \mathbb{R}^{N\times N}, \boldsymbol{V}_1, \boldsymbol{V}_2 \in \mathbb{R}^{M\times M}$ are independent random Haar matrices.

Construct the hermitization $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{(N+M)\times(N+M)}$ from $\boldsymbol{Y}$ as

$$\boldsymbol{\mathcal{Y}} = \begin{bmatrix} \boldsymbol{0}_{N\times N} & \boldsymbol{Y} \\ \boldsymbol{Y}^\intercal & \boldsymbol{0}_{M\times M} \end{bmatrix}$$

For simplicity of notation, we use $\boldsymbol{W} \equiv \boldsymbol{S}\boldsymbol{U}_1\boldsymbol{T}\boldsymbol{V}_1^\intercal$, $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{(N+M)\times(N+M)}$ the hermitization of $\boldsymbol{W}$, and $\widetilde{\boldsymbol{\mathcal{Z}}}$ denotes the hermitization of the matrix $\boldsymbol{U}_2\boldsymbol{Z}\boldsymbol{V}_2^\intercal$.

11.B.1 Resolvent relation

We want to find a relation between $\boldsymbol{G}(z) \equiv \boldsymbol{G}_{\boldsymbol{\mathcal{Y}}}(z)$, and the signal matrix $\boldsymbol{S}$. We have

$$\begin{aligned}\langle G_{ij}(z)\rangle &= \lim_{n\to\infty} \int \Big(\prod_{k=1}^{N+M} \prod_{\tau=1}^{n} d\eta_k^{(\tau)} \Big) \eta_i^{(1)} \eta_j^{(1)} \\ &\qquad \times \Big\langle \exp\Big\{ -\frac{1}{2} \sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)\intercal} (z\boldsymbol{I} - \boldsymbol{\mathcal{Y}}) \boldsymbol{\eta}^{(\tau)} \Big\} \Big\rangle_{\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{V}_1, \boldsymbol{V}_2} \\ &= \lim_{n\to\infty} \int \Big(\prod_{k=1}^{N+M} \prod_{\tau=1}^{n} d\eta_k^{(\tau)} \Big) \eta_i^{(1)} \eta_j^{(1)} \exp\Big\{ -\frac{z}{2} \sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)\intercal} \boldsymbol{\eta}^{(\tau)} \Big\} \\ &\qquad \times \Big\langle \exp\Big\{ \frac{1}{2} \sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)\intercal} \boldsymbol{\mathcal{W}} \boldsymbol{\eta}^{(\tau)} \Big\} \Big\rangle_{\boldsymbol{U}_1, \boldsymbol{V}_1} \Big\langle \exp\Big\{ \frac{1}{2} \sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)\intercal} \widetilde{\boldsymbol{\mathcal{Z}}} \boldsymbol{\eta}^{(\tau)} \Big\} \Big\rangle_{\boldsymbol{U}_2, \boldsymbol{V}_2}\end{aligned} \tag{11.26}$$

Split each replica $\boldsymbol{\eta}^{(\tau)}$ into two vectors $\boldsymbol{a}^{(\tau)} \in \mathbb{R}^N, \boldsymbol{b}^{(\tau)} \in \mathbb{R}^M$, $\boldsymbol{\eta}^{(\tau)} = \begin{bmatrix} \boldsymbol{a}^{(\tau)} \\ \boldsymbol{b}^{(\tau)} \end{bmatrix}$. The exponent in the first bracket in (11.26) can be written as:

$$\begin{aligned}\boldsymbol{\eta}^{(\tau)\intercal} \boldsymbol{\mathcal{W}} \boldsymbol{\eta}^{(\tau)} &= \boldsymbol{a}^{(\tau)\intercal} \boldsymbol{S} \boldsymbol{U}_1 \boldsymbol{T} \boldsymbol{V}_1^\intercal \boldsymbol{b}^{(\tau)} + \boldsymbol{b}^{(\tau)\intercal} \boldsymbol{V}_1 \boldsymbol{T}^\intercal \boldsymbol{U}_1^\intercal \boldsymbol{S} \boldsymbol{a}^{(\tau)} \\ &= 2\boldsymbol{a}^{(\tau)\intercal} \boldsymbol{S} \boldsymbol{U}_1 \boldsymbol{T} \boldsymbol{V}_1^\intercal \boldsymbol{b}^{(\tau)} \\ &= 2 \operatorname{Tr} \boldsymbol{b}^{(\tau)} \boldsymbol{a}^{(\tau)\intercal} \boldsymbol{S} \boldsymbol{U}_1 \boldsymbol{T} \boldsymbol{V}_1^\intercal\end{aligned} \tag{11.27}$$

Using the formula for the rectangular spherical integral [64] (see (6.12)), we find:

$$\begin{aligned}&\Big\langle \exp\Big\{ \sum_{\tau=1}^{n} \operatorname{Tr} \boldsymbol{b}^{(\tau)} \boldsymbol{a}^{(\tau)\intercal} \boldsymbol{S} \boldsymbol{U}_1 \boldsymbol{T} \boldsymbol{V}_1^\intercal \Big\} \Big\rangle_{\boldsymbol{U}_1, \boldsymbol{V}_1} \\ &\qquad \approx \exp\Big\{ \frac{N}{2} \sum_{\tau=1}^{n} \mathcal{Q}_{\mu_T}^{(\alpha)} \big(\frac{1}{NM} \|\boldsymbol{S}\boldsymbol{a}^{(\tau)}\|^2 \|\boldsymbol{b}^{(\tau)}\|^2 \big) \Big\}\end{aligned} \tag{11.28}$$

with $\mathcal{Q}_{\mu_T}^{(\alpha)}(x) = \int_0^x \frac{\mathcal{C}_{\mu_T}^{(\alpha)}(t)}{t}\, dt$. In (11.28), we used that $\boldsymbol{b}^{(\tau)} \boldsymbol{a}^{(\tau)\intercal} \boldsymbol{S}$ is a rank-one matrix with non-zero singular value $\|\boldsymbol{b}^{(\tau)}\| \|\boldsymbol{S}\boldsymbol{a}^{(\tau)}\|$.

Similarly, for the second bracket in (11.26) we can write:

$$\begin{aligned}\boldsymbol{\eta}^{(\tau)\intercal} \widetilde{\boldsymbol{\mathcal{Z}}} \boldsymbol{\eta}^{(\tau)} &= \boldsymbol{a}^{(\tau)\intercal} \boldsymbol{U}_2 \boldsymbol{Z} \boldsymbol{V}_2^\intercal \boldsymbol{b}^{(\tau)} + \boldsymbol{b}^{(\tau)\intercal} \boldsymbol{V}_2 \boldsymbol{Z}^\intercal \boldsymbol{U}_2^\intercal \boldsymbol{a}^{(\tau)} \\ &= 2\boldsymbol{a}^{(\tau)\intercal} \boldsymbol{U}_2 \boldsymbol{Z} \boldsymbol{V}_2^\intercal \boldsymbol{b}^{(\tau)} \\ &= 2 \operatorname{Tr} \boldsymbol{b}^{(\tau)} \boldsymbol{a}^{(\tau)\intercal} \boldsymbol{U}_2 \boldsymbol{Z} \boldsymbol{V}_2^\intercal\end{aligned} \tag{11.29}$$

which using the formula of rectangular spherical integrals, implies

$$\begin{aligned}&\Big\langle \exp\Big\{ \sum_{\tau=1}^{n} \operatorname{Tr} \boldsymbol{b}^{(\tau)} \boldsymbol{a}^{(\tau)\intercal} \boldsymbol{U}_2 \boldsymbol{Z} \boldsymbol{V}_2^\intercal \Big\} \Big\rangle_{\boldsymbol{U}_2, \boldsymbol{V}_2} \\ &\qquad \approx \exp\Big\{ \frac{N}{2} \sum_{\tau=1}^{n} \mathcal{Q}_{\mu_Z}^{(\alpha)} \big(\frac{1}{NM} \|\boldsymbol{a}^{(\tau)}\|^2 \|\boldsymbol{b}^{(\tau)}\|^2 \big) \Big\}\end{aligned} \tag{11.30}$$

From (11.26), (11.28), (11.30), we find:

$$\begin{aligned}\langle G_{ij}(z)\rangle = \lim_{n\to\infty}\int \Big(\prod_{k=1}^{N+M}\prod_{\tau=1}^{n} d\eta_k^{(\tau)}\Big)\,\eta_i^{(1)}\eta_j^{(1)} \\ \times \exp\Big\{-\frac{1}{2}\sum_{\tau=1}^{n} z\|\boldsymbol{\eta}^{(\tau)}\|^2 - N\mathcal{Q}_{\mu_T}^{(\alpha)}\Big(\frac{\|\boldsymbol{S}\boldsymbol{a}^{(\tau)}\|^2\|\boldsymbol{b}^{(\tau)}\|^2}{NM}\Big) \\ - N\mathcal{Q}_{\mu_Z}^{(\alpha)}\Big(\frac{\|\boldsymbol{a}^{(\tau)}\|^2\|\boldsymbol{b}^{(\tau)}\|^2}{NM}\Big)\Big\}\end{aligned} \tag{11.31}$$

Now, we introduce delta functions $\delta\big(p_1^{(\tau)} - \frac{\|\boldsymbol{a}^{(\tau)}\|^2}{N}\big)$, $\delta\big(p_2^{(\tau)} - \frac{\|\boldsymbol{b}^{(\tau)}\|^2}{M}\big)$, and $\delta\big(p_3^{(\tau)} - \frac{\|\boldsymbol{S}\boldsymbol{a}^{(\tau)}\|^2}{N}\big)$, and using them, the integral in (11.31) can be written as (for brevity we drop the limit term):

$$\begin{aligned}\langle G_{ij}(z)\rangle = \int \Big(\prod_{k=1}^{N+M}\prod_{\tau=1}^{n} d\eta_k^{(\tau)}\Big)\Big(\prod_{\tau=1}^{n} dp_1^{(\tau)}\,dp_2^{(\tau)}\,dp_3^{(\tau)}\Big)\,\eta_i^{(1)}\eta_j^{(1)} \\ \times \prod_{\tau=1}^{n}\delta\Big(p_1^{(\tau)} - \frac{\|\boldsymbol{a}^{(\tau)}\|^2}{N}\Big)\,\delta\Big(p_2^{(\tau)} - \frac{\|\boldsymbol{b}^{(\tau)}\|^2}{M}\Big)\,\delta\Big(p_3^{(\tau)} - \frac{\|\boldsymbol{S}\boldsymbol{a}^{(\tau)}\|^2}{N}\Big) \\ \times \exp\Big\{-\frac{1}{2}\sum_{\tau=1}^{n} z\|\boldsymbol{\eta}^{(\tau)}\|^2 - N\mathcal{Q}_{\mu_T}^{(\alpha)}(p_2^{(\tau)}p_3^{(\tau)}) - N\mathcal{Q}_{\mu_Z}^{(\alpha)}(p_1^{(\tau)}p_2^{(\tau)})\Big\}\end{aligned} \tag{11.32}$$

In the next step, we replace each delta with its Fourier transform, $\delta\big(p_1^{\tau} - \frac{1}{N}\|\boldsymbol{a}^{\tau}\|^2\big) \propto \int d\zeta_1^{\tau}\exp\Big\{-\frac{N}{2}\zeta_1^{\tau}\big(p_1^{\tau} - \frac{1}{N}\|\boldsymbol{a}^{\tau}\|^2\big)\Big\}$. After rearranging, we find:

$$\begin{aligned}\langle G_{ij}(z)\rangle \propto \int \Big(\prod_{\tau=1}^{n} dp_1^{(\tau)}\,dp_2^{(\tau)}\,dp_3^{(\tau)}\,d\zeta_1^{(\tau)}\,d\zeta_2^{(\tau)}\,d\zeta_3^{(\tau)}\Big) \\ \times \exp\Big\{\frac{N}{2}\sum_{\tau=1}^{n}\mathcal{Q}_{\mu_T}^{(\alpha)}(p_2^{(\tau)}p_3^{(\tau)}) + \mathcal{Q}_{\mu_Z}^{(\alpha)}(p_1^{(\tau)}p_2^{(\tau)}) \\ - \zeta_1^{(\tau)}p_1^{(\tau)} - \frac{1}{\alpha}\zeta_2^{(\tau)}p_2^{(\tau)} - \zeta_3^{(\tau)}p_3^{(\tau)}\Big\} \\ \times \int \Big(\prod_{k=1}^{N+M}\prod_{\tau=1}^{n} d\eta_k^{(\tau)}\Big)\,\eta_i^{(1)}\eta_j^{(1)} \\ \times \exp\Big\{-\frac{1}{2}\sum_{\tau=1}^{n} z\|\boldsymbol{\eta}^{(\tau)}\|^2 - \zeta_1^{(\tau)}\|\boldsymbol{a}^{(\tau)}\|^2 \\ - \zeta_2^{(\tau)}\|\boldsymbol{b}^{(\tau)}\|^2 - \zeta_3^{(\tau)}\|\boldsymbol{S}\boldsymbol{a}^{(\tau)}\|^2\Big\}\end{aligned} \tag{11.33}$$

The inner integral in (11.33) is a Gaussian integral, and can be written as:

$$\int \Big(\prod_{k=1}^{N+M} \prod_{\tau=1}^{n} d\eta_k^{(\tau)} \Big)\, \eta_i^{(1)} \eta_j^{(1)} \times \exp\left\{ \sum_{\tau=1}^{n} -\frac{1}{2} \boldsymbol{\eta}^{(\tau)\intercal} \begin{bmatrix} (z-\zeta_1^{(\tau)})\boldsymbol{I}_N - \zeta_3^{(\tau)} \boldsymbol{S}^2 & \boldsymbol{0} \\ \boldsymbol{0} & (z-\zeta_2^{(\tau)})\boldsymbol{I}_M \end{bmatrix} \boldsymbol{\eta}^{(\tau)} \right\} \tag{11.34}$$

Denote the matrix in the exponent by $\boldsymbol{C}_S^{(\tau)}$. Its determinant reads:

$$\det \boldsymbol{C}_S^{(\tau)} = (z-\zeta_2^{(\tau)})^M \prod_{k=1}^{N} (z - \zeta_1^{(\tau)} - \zeta_3^{(\tau)} \lambda_k^2)$$

where λ_k's are eigenvalues of $\boldsymbol{S}$. So replacing the formula for the Gaussian integrals, (11.33) can be written as:

$$\langle G_{ij}(z) \rangle \propto \int \Big(\prod_{\tau=1}^{n} dp_1^{(\tau)}\, dp_2^{(\tau)}\, dp_3^{(\tau)}\, d\zeta_1^{(\tau)}\, d\zeta_2^{(\tau)}\, d\zeta_3^{(\tau)} \Big) \Big(\boldsymbol{C}_S^{(1)^{-1}} \Big)_{ij} \times \exp\Big\{ -\frac{Nn}{2} F_0^S(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3) \Big\} \tag{11.35}$$

with

$$\begin{aligned} &F_0^S(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3, \boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \boldsymbol{\zeta}_3) \\ &\quad = \frac{1}{n} \sum_{\tau=1}^{n} \Bigg[\frac{1}{N} \sum_{k=1}^{N} \ln(z - \zeta_1^{(\tau)} - \zeta_3^{(\tau)} \lambda_k^2) + \frac{1}{\alpha} \ln(z - \zeta_2^{(\tau)}) \\ &\qquad - \mathcal{Q}_{\mu_T}^{(\alpha)}(p_2^{(\tau)} p_3^{(\tau)}) - \mathcal{Q}_{\mu_Z}^{(\alpha)}(p_1^{(\tau)} p_2^{(\tau)}) + \zeta_1^{(\tau)} p_1^{(\tau)} + \frac{1}{\alpha} \zeta_2^{(\tau)} p_2^{(\tau)} + \zeta_3^{(\tau)} p_3^{(\tau)} \Bigg] \end{aligned} \tag{11.36}$$

In the large N limit, the integral in (11.35) can be computed using the saddle-points of the function F_0^S. In the evaluation of this integral, we use the *replica symmetric* ansatz that assumes a saddle-point of the form:

$$\forall \tau \in \{1, \cdots, n\} : \quad \begin{cases} p_1^\tau = p_1, & p_2^\tau = p_2, & p_3^\tau = p_3 \\ \zeta_1^\tau = \zeta_1, & \zeta_2^\tau = \zeta_2, & \zeta_3^\tau = \zeta_3 \end{cases}$$

The saddle point is a solution of the set of equations:

$$\begin{cases} \zeta_1^* = \frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(p_1^* p_2^*)}{p_1^*}, \quad \zeta_2^* = \frac{\alpha}{p_2^*} \big(\mathcal{C}_{\mu_Z}^{(\alpha)}(p_1^* p_2^*) + \mathcal{C}_{\mu_T}^{(\alpha)}(p_2^* p_3^*) \big), \quad \zeta_3^* = \frac{\mathcal{C}_{\mu_T}^{(\alpha)}(p_2^* p_3^*)}{p_3^*} \\ p_1^* = \frac{1}{\zeta_3^*} \mathcal{G}_{\rho_{S^2}}\big(\frac{z-\zeta_1^*}{\zeta_3^*}\big), \quad p_2^* = \frac{1}{z-\zeta_2^*}, \quad p_3^* = \frac{z-\zeta_1^*}{\zeta_3^{*2}} \mathcal{G}_{\rho_{S^2}}\big(\frac{z-\zeta_1^*}{\zeta_3^*}\big) - \frac{1}{\zeta_3^*} \end{cases} \tag{11.37}$$

Now, since the relation (11.35) and the solutions (11.37) hold for arbitrary indices i, j, we can state the relation in matrix form. The inverse of $\boldsymbol{C}_S^{*-1}$, and the block structure of $\boldsymbol{G}_{\mathcal{Y}}(z)$ are computed in sections 11.F. From (11.104), (11.105) we have (for sufficiently large N):

$$\begin{aligned}\big\langle \boldsymbol{G}_{\mathcal{Y}}(z)\big\rangle_{\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{V}_1,\boldsymbol{V}_2} &= \left\langle \begin{bmatrix} \frac{1}{z}\boldsymbol{I}_N + \frac{1}{z}\boldsymbol{Y}\boldsymbol{G}_{Y^\intercal Y}(z^2)\boldsymbol{Y}^\intercal & \boldsymbol{Y}\boldsymbol{G}_{Y^\intercal Y}(z^2) \\ \boldsymbol{G}_{Y^\intercal Y}(z^2)\boldsymbol{Y}^\intercal & z\boldsymbol{G}_{Y^\intercal Y}(z^2) \end{bmatrix} \right\rangle \\ &= \begin{bmatrix} \frac{1}{\zeta_3^*}\boldsymbol{G}_{S^2}\big(\frac{z-\zeta_1^*}{\zeta_3^*}\big) & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{z-\zeta_2^*}\boldsymbol{I}_M \end{bmatrix}\end{aligned} \tag{11.38}$$

With this relation, we proceed to simplify the equations (11.37).

The normalized trace of the upper-left blocks of $\big\langle \boldsymbol{G}_{\mathcal{Y}}(z)\big\rangle_{\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{V}_1,\boldsymbol{V}_2}$ is:

$$\begin{aligned}\frac{1}{N}\sum_{k=1}^{N}\big[\frac{1}{z}+\frac{1}{z}\frac{\gamma_k^2}{z^2-\gamma_k^2}\big] &= \frac{1}{z}\frac{1}{N}\sum_{k=1}^{N}\big[1+\frac{\gamma_k^2}{z^2-\gamma_k^2}\big] \\ &= z\frac{1}{N}\sum_{k=1}^{N}\frac{1}{z^2-\gamma_k^2} \\ &= \frac{1}{2N}\sum_{k=1}^{N}\big[\frac{1}{z-\gamma_k}+\frac{1}{z+\gamma_k}\big] = \mathcal{G}_{\bar{\mu}_Y}(z)\end{aligned} \tag{11.39}$$

and the normalized trace of the upper-left block in $\boldsymbol{C}_S^{*-1}$ is $\frac{1}{\zeta_3^*}\mathcal{G}_{\rho_{S^2}}\big(\frac{z-\zeta_1^*}{\zeta_3^*}\big) = p_1^*$. Therefore, we have $p_1^* = \mathcal{G}_{\bar{\mu}_Y}(z)$.

The normalized trace of lower-right block of $\big\langle \boldsymbol{G}_{\mathcal{Y}}(z)\big\rangle_{\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{V}_1,\boldsymbol{V}_2}$ reads:

$$\begin{aligned}&\frac{1}{M}z\Big[\sum_{k=1}^{N}\frac{1}{z^2-\gamma_k^2}+(M-N)\frac{1}{z^2}\Big] \\ &\quad= \frac{N}{M}\mathcal{G}_{\bar{\mu}_Y}(z)+\frac{M-N}{M}\frac{1}{z} = \alpha\mathcal{G}_{\bar{\mu}_Y}(z)+(1-\alpha)\frac{1}{z}\end{aligned} \tag{11.40}$$

and the normalized trace of the lower-right block in $\boldsymbol{C}_S^{*-1}$ is $\frac{1}{z-\zeta_2^*} = p_2^*$. Therefore, we have $p_2^* = \alpha\mathcal{G}_{\bar{\mu}_Y}(z)+(1-\alpha)\frac{1}{z}$. Moreover, we also have that $\zeta_2^* = \alpha z\frac{z\mathcal{G}_{\bar{\mu}_Y}(z)-1}{\alpha z\mathcal{G}_{\bar{\mu}_Y}(z)+1-\alpha}$.

Therefore, the saddle point equations (11.37) can be rewritten in a simplified form, which does not involve ρ_{S^2}, as:

$$\begin{cases}\zeta_1^* = \frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(p_1^*p_2^*)}{p_1^*}, \quad \zeta_2^* = \alpha z\frac{z\mathcal{G}_{\bar{\mu}_Y}(z)-1}{\alpha z\mathcal{G}_{\bar{\mu}_Y}(z)+1-\alpha}, \quad \zeta_3^* = \frac{\mathcal{C}_{\mu_T}^{(\alpha)}(p_2^*p_3^*)}{p_3^*} \\ \\ p_1^* = \mathcal{G}_{\bar{\mu}_Y}(z), \quad p_2^* = \alpha\mathcal{G}_{\bar{\mu}_Y}(z)+(1-\alpha)\frac{1}{z}, \quad p_3^* = \frac{z-\zeta_1^*}{\zeta_3^*}\mathcal{G}_{\bar{\mu}_Y}(z)-\frac{1}{\zeta_3^*}\end{cases} \tag{11.41}$$

Note that ζ_1^*, ζ_2^* can be computed from the observation matrix, and we only need to find ζ_3^* satisfying the following equation:

$$(z-\zeta_1^*)\mathcal{G}_{\bar{\mu}_Y}(z)-1 = \mathcal{C}_{\mu_T}^{(\alpha)}\Big(\frac{1}{\zeta_3^*}\big[\alpha\mathcal{G}_{\bar{\mu}_Y}(z)+\frac{1-\alpha}{z}\big]\big[(z-\zeta_1^*)\mathcal{G}_{\bar{\mu}_Y}(z)-1\big]\Big) \tag{11.42}$$

11.B.2 Overlaps and optimal eigenvalues

We restate the relation between the resolvent and the overlaps from the main text (11.17). For $\tilde{\boldsymbol{s}}_i = [\boldsymbol{s}_i^\intercal, \boldsymbol{0}_M]^\intercal$ with $\boldsymbol{s}_i$ eigenvectors of $\boldsymbol{S}$, we have:

$$\tilde{\boldsymbol{s}}_i^\intercal \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathrm{i}\epsilon)\big) \tilde{\boldsymbol{s}}_i \approx \pi \bar{\mu}_Y(x) O_S(x, \lambda_i) \tag{11.43}$$

From (11.43), (11.38), we find:

$$\begin{aligned} O_S(\gamma, \lambda_i) &\approx \frac{1}{\pi \bar{\mu}_Y(\gamma)} \operatorname{Im} \lim_{z \to \gamma - \mathrm{i}0^+} \boldsymbol{s}_i^\intercal \, \zeta_3^{*-1} \boldsymbol{G}_{S^2}\big(\frac{z - \zeta_1^*}{\zeta_3^*}\big) \, \boldsymbol{s}_i \\ &= \frac{1}{\pi \bar{\mu}_Y(\gamma)} \operatorname{Im} \lim_{z \to \gamma - \mathrm{i}0^+} \frac{1}{z - \zeta_1^* - \zeta_3^* \lambda_i^2} \end{aligned} \tag{11.44}$$

Once we have the overlap, we can compute the optimal eigenvalues from (11.13) in section 11.4. Note that, until now we had absorbed $\sqrt{\kappa}$ into $\boldsymbol{S}$. Therefore, we should use (11.44) with $O_S(\gamma, \sqrt{\kappa}\lambda_i)$. This leads to:

$$\begin{aligned} \widehat{\xi^*_{s\,i}} &\approx \frac{1}{N} \sum_{j=1}^{N} \lambda_j O_S(\gamma_i, \sqrt{\kappa}\lambda_j) \\ &\approx \frac{1}{\pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z \to \gamma_i - \mathrm{i}0^+} \frac{1}{N} \sum_{j=1}^{N} \frac{\lambda_j}{z - \zeta_1^* - \zeta_3^* \kappa \lambda_j^2} \\ &= \frac{1}{\pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z \to \gamma_i - \mathrm{i}0^+} \frac{1}{\kappa \zeta_3^*} \frac{1}{N} \sum_{j=1}^{N} \frac{\lambda_j}{\frac{z - \zeta_1^*}{\kappa \zeta_3^*} - \lambda_j^2} \\ &= \frac{1}{\kappa \pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z \to \gamma_i - \mathrm{i}0^+} \frac{1}{\zeta_3^*} \left(\frac{1}{2} \frac{1}{N} \sum_{j=1}^{N} \frac{1}{\sqrt{\frac{z - \zeta_1^*}{\kappa \zeta_3^*}} - \lambda_j} - \frac{1}{2} \frac{1}{N} \sum_{j=1}^{N} \frac{1}{\sqrt{\frac{z - \zeta_1^*}{\kappa \zeta_3^*}} + \lambda_j} \right) \\ &\approx \frac{1}{\kappa \pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z \to \gamma_i - \mathrm{i}0^+} \left\{ \frac{1}{2} \frac{1}{\zeta_3^*} \mathcal{G}_{\rho_S}\Big(\sqrt{\frac{z - \zeta_1^*}{\kappa \zeta_3^*}}\Big) - \frac{1}{2} \frac{1}{\zeta_3^*} \mathcal{G}_{\rho_{-S}}\Big(\sqrt{\frac{z - \zeta_1^*}{\kappa \zeta_3^*}}\Big) \right\} \\ &= \frac{1}{2\kappa \pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z \to \gamma_i - \mathrm{i}0^+} \left\{ \frac{1}{\zeta_3^*} \Big[\mathcal{G}_{\rho_S}\Big(\sqrt{\frac{z - \zeta_1^*}{\kappa \zeta_3^*}}\Big) + \mathcal{G}_{\rho_S}\Big(- \sqrt{\frac{z - \zeta_1^*}{\kappa \zeta_3^*}}\Big) \Big] \right\} \end{aligned} \tag{11.45}$$

Estimating $\boldsymbol{S}^2$

The resolvent relation we have found in (11.38) is in terms of $\boldsymbol{G}_{S^2}$. Therefore, like other RIEs in other problems [53, 54], we can express the estimator for $\boldsymbol{S}^2$ without any knowledge about ρ_S or ρ_{S^2}. One can see that, the optimal RIE for $\boldsymbol{S}^2$ is constructed in the same way as for $\boldsymbol{S}$ with eigenvalues denoted by $\widehat{\xi^*_{s^2\,i}}$. To compute the optimal eigenvalues, we absorb $\sqrt{\kappa}$ into $\boldsymbol{S}$ and we use the exact expression in (11.44). In the end, we only need to divide by κ to find an

estimator for the true $\boldsymbol{S}^2$.

$$
\begin{aligned}
\widehat{\xi^*_{s^2 i}} &\approx \frac{1}{N}\sum_{j=1}^{N} \lambda_j^2 O_S(\gamma_i, \lambda_j) \\
&\approx \frac{1}{\pi\bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z\to\gamma_i - \mathrm{i}0^+} \frac{1}{N}\sum_{j=1}^{N} \frac{\lambda_j^2}{z-\zeta_1^*-\zeta_3^*\lambda_j^2} \\
&= \frac{1}{\pi\bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z\to\gamma_i - \mathrm{i}0^+} \frac{1}{\zeta_3^*}\frac{1}{N}\sum_{j=1}^{N} \frac{\lambda_j^2}{\frac{z-\zeta_1^*}{\zeta_3^*}-\lambda_j^2} \\
&= \frac{1}{\pi\bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z\to\gamma_i - \mathrm{i}0^+} -\frac{1}{\zeta_3^*}\frac{1}{N}\sum_{j=1}^{N} \frac{\frac{z-\zeta_1^*}{\zeta_3^*}-\lambda_j^2-\frac{z-\zeta_1^*}{\zeta_3^*}}{\frac{z-\zeta_1^*}{\zeta_3^*}-\lambda_j^2} \\
&= \frac{1}{\pi\bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z\to\gamma_i - \mathrm{i}0^+} -\frac{1}{\zeta_3^*}\frac{1}{N}\sum_{j=1}^{N} \Big[1-\frac{z-\zeta_1^*}{\zeta_3^*}\frac{1}{\frac{z-\zeta_1^*}{\zeta_3^*}-\lambda_j^2}\Big] \\
&\approx \frac{1}{\pi\bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z\to\gamma_i - \mathrm{i}0^+} -\frac{1}{\zeta_3^*} + \frac{z-\zeta_1^*}{{\zeta_3^*}^2}\mathcal{G}_{\rho_{S^2}}\big(\frac{z-\zeta_1^*}{\zeta_3^*}\big) \\
&\overset{(a)}{=} \frac{1}{\pi\bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z\to\gamma_i - \mathrm{i}0^+} p_3^* \\
&\overset{(b)}{=} \frac{1}{\pi\bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z\to\gamma_i - \mathrm{i}0^+} \frac{z-\zeta_1^*}{\zeta_3^*}\mathcal{G}_{\bar{\mu}_Y}(z) - \frac{1}{\zeta_3^*}
\end{aligned}
\tag{11.46}
$$

where in (a) we used (11.37), and for (b) we used (11.41). Thus, the optimal eigenvalues for $\boldsymbol{S}^2$ read:

$$
\widehat{\xi^*_{s^2 i}} = \frac{1}{\kappa}\frac{1}{\pi\bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z\to\gamma_i - \mathrm{i}0^+} \frac{z-\zeta_1^*}{\zeta_3^*}\mathcal{G}_{\bar{\mu}_Y}(z) - \frac{1}{\zeta_3^*} \tag{11.47}
$$

Note that the parameters ζ_1^*, ζ_3^* can be computed from (11.41), (11.42), without the knowledge of ρ_S or ρ_{S^2}.

Remark 11.3. *The main barrier to find an estimator for $\boldsymbol{S}$ is that the resolvent relation* (11.38) *is in terms of $\mathcal{G}_{\rho_{S^2}}$. Moreover, in the estimator for $\boldsymbol{S}$, second equality in* (11.45)*, we have the sum $\sum_{j=1}^{N} \frac{\lambda_j}{z-\zeta_1^*-\kappa\zeta_3^*\lambda_j^2}$ which cannot be written in terms of $\mathcal{G}_{\rho_{S^2}}$.*

Remark 11.4. *If we add the assumption that the matrix $\boldsymbol{S}$ is positive semi-definite, without any further knowledge on the prior, we can use $\sqrt{\widehat{\xi^*_{s^2 i}}}$ for the eigenvalues of $\boldsymbol{\Xi}_S(\boldsymbol{Y})$. However, note that, this estimator is sub-optimal for $\boldsymbol{S}$ as $\sqrt{\sum_{j=1}^{N} \lambda_j^2 \left(\boldsymbol{u}_i^\intercal \boldsymbol{s}_j\right)^2} \neq \sum_{j=1}^{N} \lambda_j \left(\boldsymbol{u}_i^\intercal \boldsymbol{s}_j\right)^2$.*

11.B.3 Numerical examples

In this section, we will illustrate the derived formulas (11.38), (11.44), and (11.45) with numerical experiments.

We consider matrices $\boldsymbol{T}, \boldsymbol{Z} \in \mathbb{R}^{N\times M}$ to have i.i.d. Gaussian entries, so $\mathcal{C}^{(\alpha)}_{\mu_T}(z) = \mathcal{C}^{(\alpha)}_{\mu_Z}(z) = \frac{1}{\alpha}z$ which leads to a simplification of saddle point equations (11.41):

$$\begin{cases} \zeta_1^* = \frac{1}{\alpha}p_2^*, \quad \zeta_2^* = \alpha z \frac{z\mathcal{G}_{\bar{\mu}_Y}(z)-1}{\alpha z\mathcal{G}_{\bar{\mu}_Y}(z)+1-\alpha}, \quad \zeta_3^* = \frac{1}{\alpha}p_2^* \\ p_1^* = \mathcal{G}_{\bar{\mu}_Y}(z), \quad p_2^* = \alpha\mathcal{G}_{\bar{\mu}_Y}(z) + (1-\alpha)\frac{1}{z}, \quad p_3^* = \frac{z-\zeta_1^*}{\zeta_3^*}\mathcal{G}_{\bar{\mu}_Y}(z) - \frac{1}{\zeta_3^*} \end{cases} \tag{11.48}$$

Resolvent relation

We take $\kappa = 1$. In model (11.25), without loss of generality we can consider $\boldsymbol{S}$ to be diagonal. In figures 11.B.1 and 11.B.2 respectively, we consider the $\boldsymbol{S}$ to be a diagonal matrix obtained by taking the eigenvalues of a Wigner matrix and a Wishart matrix respectively.

Note that μ_Y and $\mathcal{G}_{\bar{\mu}_Y}(z)$ can be computed analytically using tools from random matrix theory, but the computation is highly involved. In our experiments, we use instead a numerical estimation of $\mathcal{G}_{\bar{\mu}_Y}(z)$ obtained from the observation matrix with the help of a Cauchy kernel to compute the parameters ζ_1^*, ζ_3^* (see section 5.1.3, and [67] for details on the Cauchy kernel method).

Unlike the simpler models [51] for which the fluctuations are derived to be of the order $1/\sqrt{N}$, based on our derivation we cannot assess the order of fluctuations. However, from our numerics we observe that the fluctuations are of the order $o(N)$. Moreover, fluctuations near the edge points of density are larger (in particular for the last row in both figures 11.B.1, 11.B.2), which is due to the fact that the limiting measures have higher fluctuations on their edge-points.

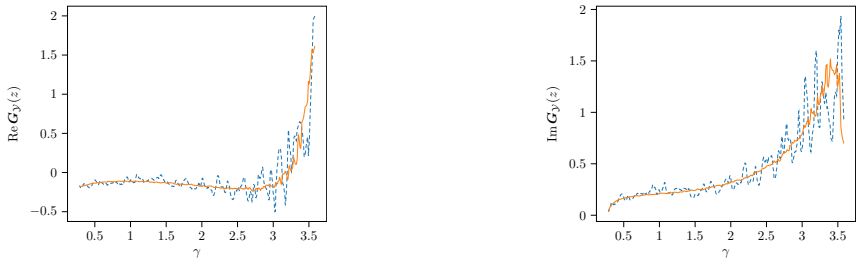
Another observation, from comparison of figures 11.B.1, 11.B.2, is that the fluctuations for the first example are relatively larger than the second one. One possible guess could be that this is due to the symmetry of ρ_S in the first example. However based on more extensive numerical observations (which we omit here) we speculate that this issue is in fact related to the existence of small eigenvalues of $\boldsymbol{S}$. In other words, if $\boldsymbol{S}$ has eigenvalue 0 or small eigenvalues, we have higher fluctuations in the relation (11.43).

Overlaps

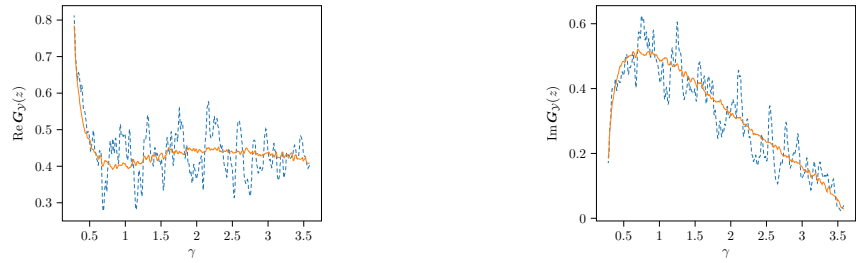
To illustrate the formula for the overlap (11.44), we fix the matrix $\boldsymbol{S}$ and run experiments over various realization of the model (11.25). For each experiment, we record the overlap of k-th left singular vector of $\boldsymbol{Y}$ and the eigenvectors of $\boldsymbol{S}$. To compute the theoretical prediction, we find $\zeta_1^* = \zeta_3^*$ for $z = \bar{\gamma}_k - \mathrm{i}0^+$ where $\bar{\gamma}_k$ is the average of k-th singular value of $\boldsymbol{Y}$ in the experiments.

To find $\zeta_1^* = \zeta_3^*$, we use the set of equations (11.37) which for $\boldsymbol{T}, \boldsymbol{Z}$ Gaussian can be written as:

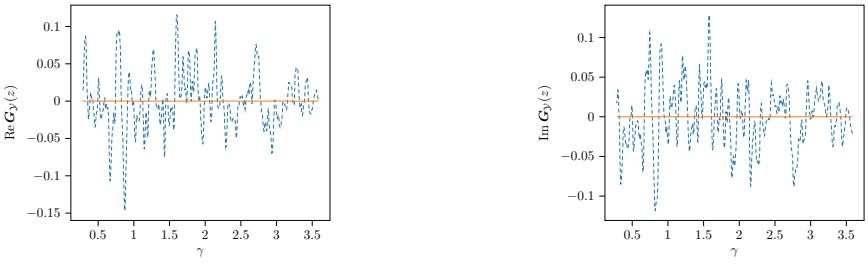
$$\begin{cases} \zeta_1^* = \frac{1}{\alpha}p_2^*, \quad \zeta_2^* = p_1^* + p_3^*, \quad \zeta_3^* = \frac{1}{\alpha}p_2^* \\ \\ p_1^* = \frac{1}{\zeta_1^*}\mathcal{G}_{\rho_{S^2}}\big(\frac{z}{\zeta_1^*} - 1\big), \quad p_2^* = \frac{1}{z-\zeta_2^*}, \quad p_3^* = \frac{z-\zeta_1^*}{\zeta_1^{*2}}\mathcal{G}_{\rho_{S^2}}\big(\frac{z}{\zeta_1^*} - 1\big) - \frac{1}{\zeta_1^*} \end{cases} \tag{11.49}$$



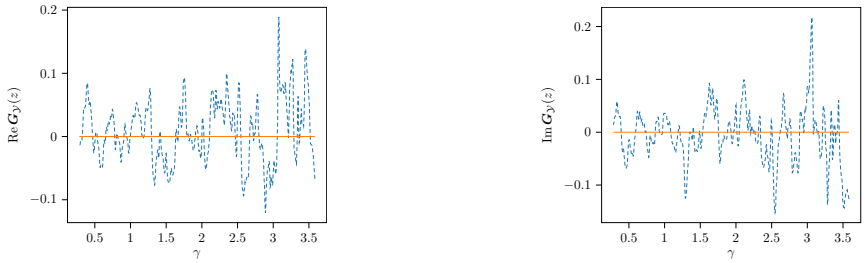
(a) Entry $i = j = 1$, first diagonal entry in the upper-left block



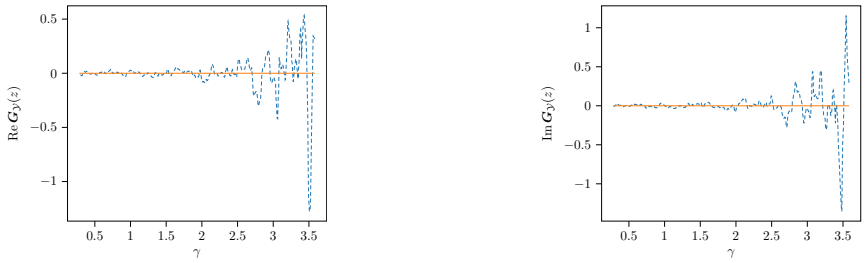
(b) Entry $i = j = 5000$, last diagonal entry in the lower-right block



(c) Entry $i = 4999, j = 5000$, a non-diagonal entry in the lower-right block

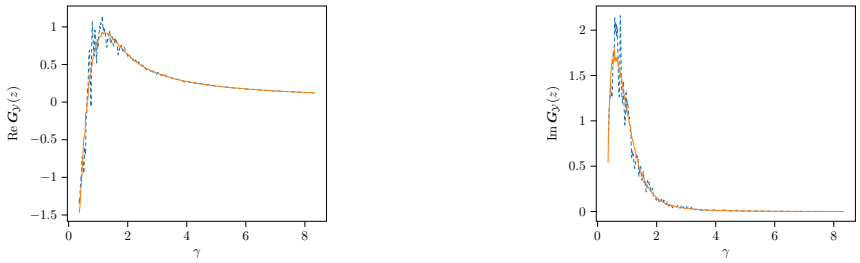


(d) Entry $i = 1, j = 5000$, an entry in the upper-right block



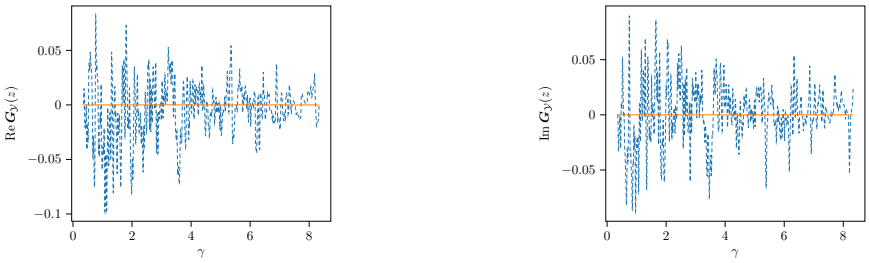
(e) Entry $i = 1, j = 2$, a non-diagonal entry in the upper-left block

Figure 11.B.1: Illustration of (11.38). $\boldsymbol{S}$ is diagonal matrix from the eigenvalues of a Wigner matrix and $\boldsymbol{T}, \boldsymbol{Z}$ are Gaussian matrices with $N = 2000, M = 3000$. The empirical estimate of $\boldsymbol{G}_Y(z)$ (dashed blue line) is computed for $z = \gamma_i - \mathrm{i}\sqrt{\frac{1}{2N}}$ for $1 \leq i \leq N$. Theoretical estimate (solid orange line) computed from the rhs of (11.38) with parameters obtained from the generated matrix. Note that, the theoretical estimate has also fluctuations because the parameters ζ_1^*, ζ_3^* are given by the numerical estimate of $\mathcal{G}_{\bar{\mu}_Y}(z)$.

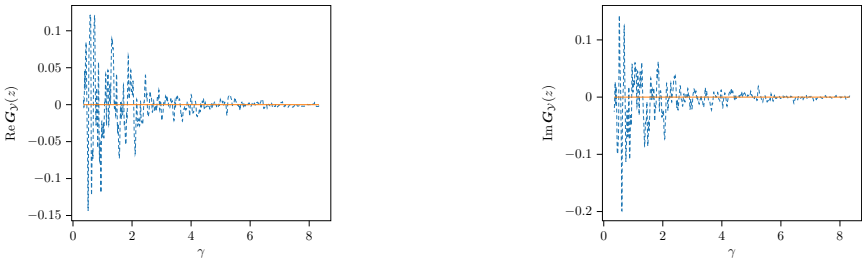


(a) Entry $i = j = 1$, first diagonal entry in the upper-left block

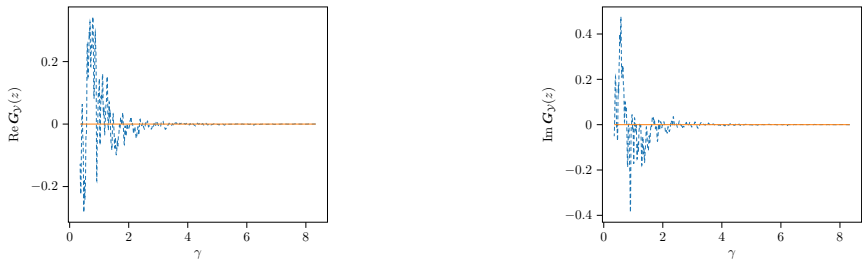
(b) Entry $i = j = 5000$, last diagonal entry in the lower-right block



(c) Entry $i = 4999, j = 5000$, a non-diagonal entry in the lower-right block



(d) Entry $i = 1, j = 5000$, an entry in the upper-right block



(e) Entry $i = 1, j = 2$, a non-diagonal entry in the upper-left block

Figure 11.B.2: Illustration of (11.38).$\boldsymbol{S}$ is diagonal matrix from the eigenvalues of a Wishart matrix with aspect ratio $1/2$ and $\boldsymbol{T}, \boldsymbol{Z}$ are Gaussian matrices with $N = 2000, M = 3000$. The empirical estimate of $\boldsymbol{G}_Y(z)$ (dashed blue line) is computed for $z = \gamma_i - \mathrm{i}\sqrt{\frac{1}{2N}}$ for $1 \leq i \leq N$. The Theoretical estimate (solid orange line) is computed from the rhs of (11.38) with parameters obtained from the generated matrix. Note that, the theoretical estimate has also fluctuations because the parameters ζ_1^*, ζ_3^* are given by the numerical estimate of $\mathcal{G}_{\bar{\mu}_Y}(z)$.

Now we proceed to simplify the solution above:

$$\zeta_2^* = p_1^* + p_3^* = \frac{z}{\zeta_1^{*2}} \mathcal{G}_{\rho_{S^2}}\big(\frac{z}{\zeta_1^*} - 1\big) - \frac{1}{\zeta_1^*}$$

$$p_2^* = \frac{1}{z - \zeta_2^*} = \frac{\zeta_1^*}{\zeta_1^* z - \frac{z}{\zeta_1^*} \mathcal{G}_{\rho_{S^2}}\big(\frac{z}{\zeta_1^*} - 1\big) + 1}$$

$$\begin{aligned}
\zeta_1^* = \frac{1}{\alpha} p_2^* &\Longrightarrow \zeta_1^* z - \frac{z}{\zeta_1^*} \mathcal{G}_{\rho_{S^2}}\big(\frac{z}{\zeta_1^*} - 1\big) + 1 = \frac{1}{\alpha} \\
&\Rightarrow \mathcal{G}_{\rho_{S^2}}\big(\frac{z}{\zeta_1^*} - 1\big) = \zeta_1^{*2} + \big(1 - \frac{1}{\alpha}\big)\frac{\zeta_1^*}{z} \\
&\Rightarrow \frac{z}{\zeta_1^*} - 1 = \mathcal{G}^{-1}_{\rho_{S^2}}\Big(\zeta_1^{*2} + \big(1 - \frac{1}{\alpha}\big)\frac{\zeta_1^*}{z}\Big) \\
&\Rightarrow \frac{z}{\zeta_1^*} - 1 - \frac{1}{\zeta_1^{*2} + \big(1 - \frac{1}{\alpha}\big)\frac{\zeta_1^*}{z}} = \mathcal{R}_{\rho_{S^2}}\Big(\zeta_1^{*2} + \big(1 - \frac{1}{\alpha}\big)\frac{\zeta_1^*}{z}\Big)
\end{aligned} \tag{11.50}$$

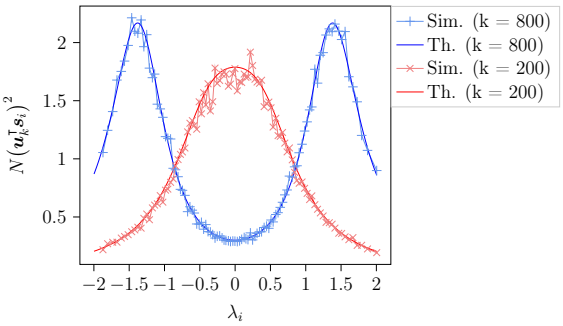
Thus, ζ_1^* is the solution to (11.50). For each example, we solve this equation and compare the obtained theoretical overlap against the average over the experiments.

Wigner $\boldsymbol{S}$. Let $\boldsymbol{S} \in \mathbb{R}^{N \times N}$ be a Wigner matrix, then $\mathcal{R}_{\rho_{S^2}}(z) = \frac{1}{1-z}$. Solving (11.50), we can compute the overlap using (11.44). In Fig. 11.B.3a, we compare the theoretical computation with simulations for $N = 1000, M = 2000$. As in previous cases $\bar{\mu}_Y(\gamma)$ is approximated using a Cauchy kernel [67].

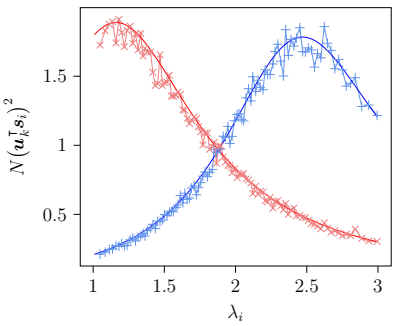
Square root Wishart $\boldsymbol{S}$. Let $\boldsymbol{S} \in \mathbb{R}^{N \times N}$ be the square root of a Wishart matrix $\boldsymbol{S} = \sqrt{\frac{1}{N}\mathbf{H}\mathbf{H}^\intercal}$ with $\mathbf{H} \in \mathbb{R}^{N \times N'}$ having i.i.d. Gaussian entries. Then $\mathcal{R}_{\rho_{S^2}}(z) = \frac{1}{\alpha'}\frac{1}{1-z}$, $\alpha' = {}^{N}/_{N'}$. Solving (11.50), we can compute the overlap using (11.44). In Fig. 11.B.3b, we compare the theoretical computation with simulations for $N = 1000, N' = 4000, M = 2000$.

RIE performance

In this section, we investigate the performance of our proposed estimators for $\boldsymbol{S}$. We compare performances of the optimal RIE (11.45) with the one of Oracle estimator (11.3). Moreover, we illustrate the performance of the estimator for $\boldsymbol{S}^2$ (11.46), and the sub-optimal estimator of $\boldsymbol{S}$ derived from it, see remark 11.4.



(a) Wigner $\boldsymbol{S}$



(b) Square root Wishart $\boldsymbol{S}$ with $\alpha' = 1/4$

Figure 11.B.3: Computation of the rescaled overlap. Both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$, and aspect ratio $N/M = 1/2$. The simulation results are averaged over 1000 experiments with fixed $\boldsymbol{S}$, and $N = 1000, M = 2000$. Some of the simulation points are dropped for clarity.

For $\boldsymbol{T}, \boldsymbol{Z}$ with Gaussian i.i.d. entries, (11.47) simplifies to:

$$
\begin{aligned}
\widehat{\xi^*_{s^2 i}} &= \frac{1}{\kappa}\frac{1}{\pi\bar{\mu}_Y(\gamma_i)}\,\mathrm{Im}\lim_{z\to\gamma_i - \mathrm{i}0^+}\frac{z-\zeta_1^*}{\zeta_3^*}\mathcal{G}_{\bar{\mu}_Y}(z) - \frac{1}{\zeta_3^*} \\
&= \frac{1}{\kappa}\frac{1}{\pi\bar{\mu}_Y(\gamma_i)}\,\mathrm{Im}\lim_{z\to\gamma_i - \mathrm{i}0^+}\frac{z}{\zeta_1^*}\mathcal{G}_{\bar{\mu}_Y}(z) - \mathcal{G}_{\bar{\mu}_Y} - \frac{1}{\zeta_1^*} \\
&= \frac{1}{\kappa}\frac{1}{\pi\bar{\mu}_Y(\gamma_i)}\,\mathrm{Im}\lim_{z\to\gamma_i - \mathrm{i}0^+}\frac{z}{\mathcal{G}_{\bar{\mu}_Y}(z) + \frac{1-\alpha}{\alpha}\frac{1}{z}}\mathcal{G}_{\bar{\mu}_Y}(z) - \mathcal{G}_{\bar{\mu}_Y}(z) - \frac{1}{\mathcal{G}_{\bar{\mu}_Y}(z) + \frac{1-\alpha}{\alpha}\frac{1}{z}} \\
&= \frac{1}{\kappa}\frac{1}{\pi\bar{\mu}_Y(\gamma_i)}\,\mathrm{Im}\left\{\frac{\gamma_i}{\pi\mathsf{H}[\bar{\mu}_Y](\gamma_i) + \pi\mathrm{i}\bar{\mu}_Y(\gamma_i) + \frac{1-\alpha}{\alpha}\frac{1}{\gamma_i}}\left(\pi\mathsf{H}[\bar{\mu}_Y](\gamma_i) + \pi\mathrm{i}\bar{\mu}_Y(\gamma_i)\right)\right. \\
&\qquad\qquad \left. - \left(\pi\mathsf{H}[\bar{\mu}_Y](\gamma_i) + \pi\mathrm{i}\bar{\mu}_Y(\gamma_i)\right) - \frac{1}{\pi\mathsf{H}[\bar{\mu}_Y](\gamma_i) + \pi\mathrm{i}\bar{\mu}_Y(\gamma_i) + \frac{1-\alpha}{\alpha}\frac{1}{\gamma_i}}\right\} \\
&= \frac{1}{\kappa}\frac{1}{\pi\bar{\mu}_Y(\gamma_i)}\,\pi\bar{\mu}_Y(\gamma_i)\left(-1 + \frac{1}{\alpha\left(\pi^2\bar{\mu}_Y(\gamma_i)^2 + \left(\pi\mathsf{H}[\bar{\mu}_Y](\gamma_i) + \frac{-1+\frac{1}{\alpha}}{\gamma_i}\right)^2\right)}\right) \\
&= \frac{1}{\kappa}\left[-1 + \frac{1}{\alpha\left(\pi^2\bar{\mu}_Y(\gamma_i)^2 + \left(\pi\mathsf{H}[\bar{\mu}_Y](\gamma_i) + \frac{-1+\frac{1}{\alpha}}{\gamma_i}\right)^2\right)}\right]
\end{aligned}
\tag{11.51}
$$

For our first example, we consider two priors for $\boldsymbol{S}$:

Shifted Wigner $\boldsymbol{S}$. We consider $\boldsymbol{S} = \boldsymbol{F} + c\boldsymbol{I}$ where $\boldsymbol{F} = \boldsymbol{F}^\intercal \in \mathbb{R}^{N\times N}$ has i.i.d. entries with variance $1/N$, and $c \neq 0$ is a real number. Then, the spectrum of $\boldsymbol{S}$ is a shifted version of the Wigner law

$$
\rho_S(\lambda) = \frac{\sqrt{4-(\lambda-c)^2}}{2\pi}, \quad \text{for } c-2 < \lambda < c+2,
$$

and the Stieltjes transform reads:

$$\mathcal{G}_{\rho_S}(z) = \frac{z - c - \sqrt{(z-2-c)(z+2-c)}}{2}$$

Wishart S. Take $\boldsymbol{S} = \frac{1}{N}\mathbf{H}\mathbf{H}^\intercal$ with $\mathbf{H} \in \mathbb{R}^{N\times N'}$ having i.i.d. Gaussian entries, with $N/N' = \alpha' \leq 1$. Then, the spectrum of $\boldsymbol{S}$ is the renowned *Marchenko-Pastur* distribution:

$$\rho_S(\lambda) = \frac{\sqrt{\left[\lambda - \left(\frac{1}{\sqrt{\alpha'}} - 1\right)^2\right]\left[\left(\frac{1}{\sqrt{\alpha'}} + 1\right)^2 - \lambda\right]}}{2\pi\lambda}$$

for $\left(\frac{1}{\sqrt{\alpha'}} - 1\right)^2 < \lambda < \left(\frac{1}{\sqrt{\alpha'}} + 1\right)^2$, and the Stieltjes transform reads:

$$\mathcal{G}_{\rho_S}(z) = \frac{z - \left(\frac{1}{\alpha'} - 1\right) - \sqrt{\left[z - \left(\frac{1}{\sqrt{\alpha'}} - 1\right)^2\right]\left[z - \left(\frac{1}{\sqrt{\alpha'}} + 1\right)^2\right]}}{2z}$$

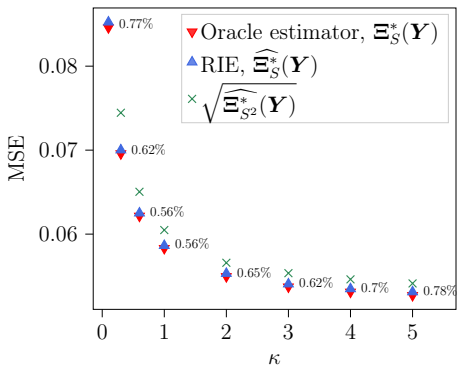
In Figure 11.B.4, the MSE of Oracle estimator, RIE (11.45), and $\sqrt{\boldsymbol{S}^2}$-RIE is illustrated for shifted Wigner $\boldsymbol{S}$ with $c = 3$, and Wishart with aspect-ratio $\alpha' = 1/4$. We see that the performance of RIE is close to the one of Oracle estimator, which implies the optimality of the proposed estimator (11.45). Moreover, we observe the sub-optimality of estimating $\boldsymbol{S}$ using $\sqrt{\widehat{\boldsymbol{\Xi}^*_{S^2}}(\boldsymbol{Y})}$. Note that, in the low-SNR regime, the estimated eigenvalues $\widehat{\xi^*_{s^2 i}}$ might be negative which makes the estimator $\sqrt{\widehat{\boldsymbol{\Xi}^*_{S^2}}(\boldsymbol{Y})}$ undefined, so the MSE is not depicted in this case.

In Figure 11.B.5, the MSE of estimating $\boldsymbol{S}^2$ is shown. We see that in the high-SNR regimes the RIE (11.51) has the same performance as the Oracle estimator.

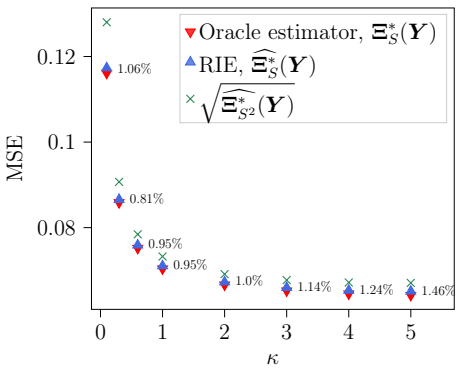
Bernoulli spectral distribution. In this case, the matrix $\boldsymbol{S}$ is constructed as $\boldsymbol{S} = \boldsymbol{U}_S\boldsymbol{\Lambda}\boldsymbol{U}_S^\intercal$ with $\boldsymbol{U}_S$ a $N \times N$ orthogonal matrix distributed according to Haar measure on orthogonal matrices, and $\boldsymbol{\Lambda} = \mathrm{diag}(\boldsymbol{\lambda})$ where $\boldsymbol{\lambda}$ has i.i.d. Bernoulli elements. Thus, $\rho_S = p\delta_0 + (1-p)\delta_{+1}$ for $p \in (0,1)$, and the Stieltjes transform is:

$$\mathcal{G}_{\rho_S}(z) = p\frac{1}{z} + (1-p)\frac{1}{z-1}$$

For this prior, we have that $\boldsymbol{S} = \boldsymbol{S}^2$, so both estimators $\widehat{\boldsymbol{\Xi}^*_S}(\boldsymbol{Y})$ and $\widehat{\boldsymbol{\Xi}^*_{S^2}}(\boldsymbol{Y})$ should have the same performance. However, note that $\widehat{\boldsymbol{\Xi}^*_{S^2}}(\boldsymbol{Y})$ does not use any knowledge of ρ_S. In Figure 11.B.6, the MSE is illustrated for these two estimators for two sparsity parameter, $p = 0.5$ and 0.9. We observe that, except for for the low-SNR regimes, both estimators have the same MSE. The poor performance of $\widehat{\boldsymbol{\Xi}^*_{S^2}}(\boldsymbol{Y})$ in the low-SNR regimes might be due to the fact that,

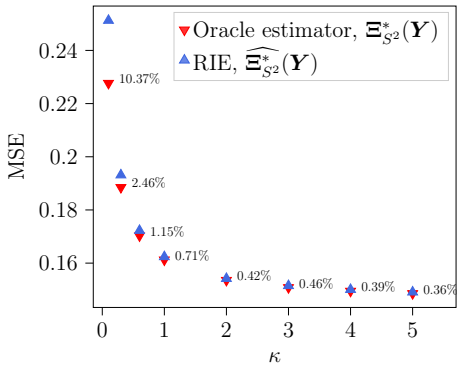


(a) Shifted Wigner, $c = 3$

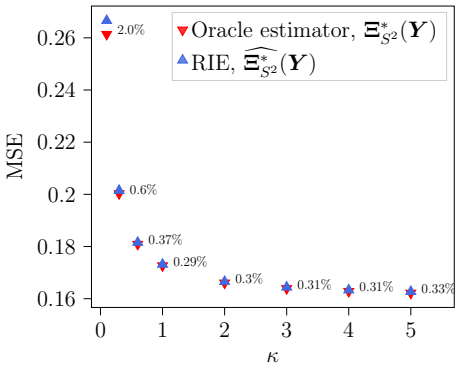


(b) Wishart, $\alpha' = 1/4$

Figure 11.B.4: Estimating $\boldsymbol{S}$. The MSE is normalized by the norm of the signal, $\|\boldsymbol{S}\|_{\mathrm{F}}^2$. Both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$, and aspect ratio $N/M = 1/2$. The RIE is applied to $N = 2000, M = 4000$, and the results are averaged over 10 runs (error bars are invisible). Average relative error between RIE $\widehat{\boldsymbol{\Xi}^*_S}(\boldsymbol{Y})$ and Oracle estimator is also reported.

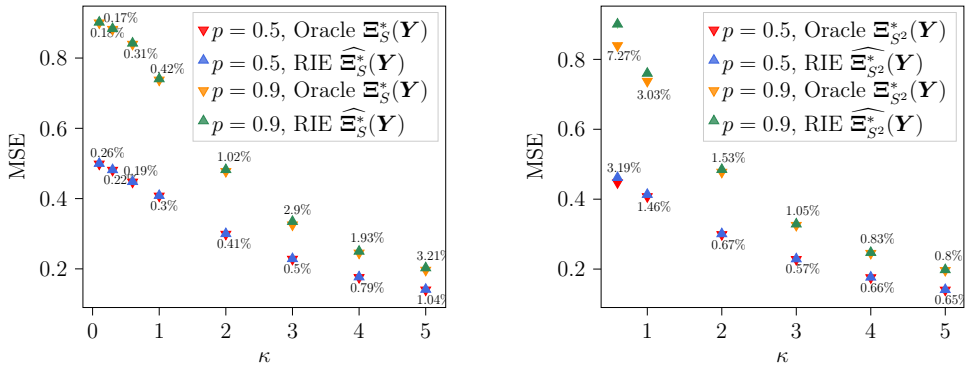


(a) Shifted Wigner, $c = 3$



(b) Wishart, $\alpha' = 1/4$

Figure 11.B.5: Estimating $\boldsymbol{S}^2$. The MSE is normalized by the norm of the signal, $\|\boldsymbol{S}^2\|_{\mathrm{F}}^2$. Both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$, and aspect ratio $N/M = 1/2$. The RIE is applied to $N = 2000, M = 4000$, and the results are averaged over 10 runs (error bars are invisible). Average relative error between RIE $\widehat{\boldsymbol{\Xi}^*_S}(\boldsymbol{Y})$ and Oracle estimator is also reported.



(a) Estimating $\boldsymbol{S}$ **(b)** Estimating $\boldsymbol{S}^2$

Figure 11.B.6: Estimating $\boldsymbol{S}$ and $\boldsymbol{S}^2$ with Bernoulli spectral prior distribution. The MSE is normalized by the norm of the signal, $\|\boldsymbol{S}\|_{\mathrm{F}}^2 = \|\boldsymbol{S}^2\|_{\mathrm{F}}^2$. Both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$, and aspect ratio $N/M = 1/2$. The RIE is applied to $N = 2000, M = 4000$, and the results are averaged over 10 runs (error bars are invisible). Average relative error between RIE $\widehat{\Xi^*_S}(\boldsymbol{Y})$ and Oracle estimator is also reported.

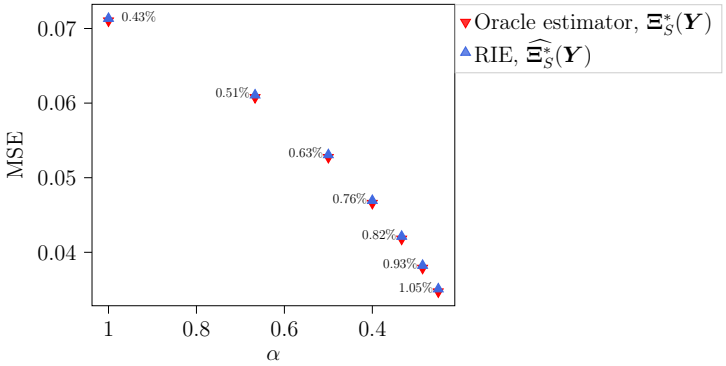


Figure 11.B.7: MSE of estimating $\boldsymbol{S}$ as a function of aspect-ratio α, prior on $\boldsymbol{S}$ is shifted Wigner with $c = 3$, and $\kappa = 5$. MSE is normalized by the norm of the signal, $\|\boldsymbol{S}\|_{\mathrm{F}}^2$. Both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$. The RIE is applied to $N = 2000, M = 1/\alpha N$, and the results are averaged over 10 runs (error bars are invisible). Average relative error between RIE $\widehat{\Xi^*_S}(\boldsymbol{Y})$ and Oracle estimator is also reported.

some of the estimated eigenvalues $\widehat{\xi^*_{s^2}}_i$ are negative although the true eigenvalue is 0. This makes the estimation more difficult for the sparser prior, see Figure 11.B.6b. However, this problem is resolved in $\widehat{\Xi^*_S}(\boldsymbol{Y})$ by taking the knowledge of $\mathcal{G}_{\rho_S}(z)$ into account.

Effect of aspect-ratio α. In Figure 11.B.7, we consider $\boldsymbol{S}$ to be shifted Wigner with $c = 3$, and the MSE is depicted for various values of the aspect-ratio α. As expected, as M increases (α decreases) and we have more observation or more data samples, the estimation error decreases.

11.C Derivation of the RIE for $\boldsymbol{T}$

In this section, we present the derivation of the optimal RIE for $\boldsymbol{T}$. For simplicity, the SNR parameter in (11.1) is absorbed into $\boldsymbol{T}$, so the model is $\boldsymbol{Y} = \boldsymbol{S}\boldsymbol{T} + \boldsymbol{Z}$. Therefore, the final estimator should be divided by $1/\sqrt{\kappa}$ to give an estimate of the original $\boldsymbol{T}$.

The optimal singular values are constructed as $\xi^*_{t\,i} = \sum_{j=1}^{N} \sigma_j \left(\boldsymbol{u}_i^\intercal \boldsymbol{t}_j^{(l)}\right)\left(\boldsymbol{v}_i^\intercal \boldsymbol{t}_j^{(r)}\right)$. We assume that, for large N, $\xi^*_{t\,i}$ can be approximated by its expectation:

$$\widehat{\xi}^*_{t\,i} \approx \sum_{j=1}^{N} \sigma_j \,\mathbb{E}\Big[\left(\boldsymbol{u}_i^\intercal \boldsymbol{t}_j^{(l)}\right)\left(\boldsymbol{v}_i^\intercal \boldsymbol{t}_j^{(r)}\right)\Big]$$

where the expectation is over the singular vectors of the observation matrix $\boldsymbol{Y}$. Therefore, to compute the optimal singular values, we need to find the mean overlap $\mathbb{E}\Big[\left(\boldsymbol{u}_i^\intercal \boldsymbol{t}_j^{(l)}\right)\left(\boldsymbol{v}_i^\intercal \boldsymbol{t}_j^{(r)}\right)\Big]$ between singular vectors of $\boldsymbol{T}$ and singular vectors of $\boldsymbol{Y}$. In the following we will see that (a rescaling of) this quantity can be expressed in terms of i-th singular value of $\boldsymbol{Y}$ and j-th singular value of $\boldsymbol{T}$ (and the limiting measures, indeed). Thus, we will use the notation $O_T(\gamma_i, \sigma_j) := N\mathbb{E}\Big[\left(\boldsymbol{u}_i^\intercal \boldsymbol{t}_j^{(l)}\right)\left(\boldsymbol{v}_i^\intercal \boldsymbol{t}_j^{(r)}\right)\Big]$ in what follows. In the nest section, we discuss how the overlap can be computed from the resolvent of the Hermitized matrix of $\boldsymbol{Y}$.

11.C.1 Relation between overlap and the resolvent

Construct the matrix $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{(N+M)\times(N+M)}$ from the observation matrix:

$$\boldsymbol{\mathcal{Y}} = \begin{bmatrix} \boldsymbol{0}_{N\times N} & \boldsymbol{Y} \\ \boldsymbol{Y}^\intercal & \boldsymbol{0}_{M\times M} \end{bmatrix}$$

By Theorem 7.3.3 in [71], $\boldsymbol{\mathcal{Y}}$ has the following eigen-decomposition:

$$\boldsymbol{\mathcal{Y}} = \begin{bmatrix} \hat{\boldsymbol{U}}_Y & \hat{\boldsymbol{U}}_Y & \boldsymbol{0} \\ \hat{\boldsymbol{V}}_Y^{(1)} & -\hat{\boldsymbol{V}}_Y^{(1)} & \boldsymbol{V}_Y^{(2)} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Gamma}_N & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{\Gamma}_N & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{U}}_Y & \hat{\boldsymbol{U}}_Y & \boldsymbol{0} \\ \hat{\boldsymbol{V}}_Y^{(1)} & -\hat{\boldsymbol{V}}_Y^{(1)} & \boldsymbol{V}_Y^{(2)} \end{bmatrix}^\intercal \tag{11.52}$$

with $\boldsymbol{V}_Y = \begin{bmatrix} \boldsymbol{V}_Y^{(1)} & \boldsymbol{V}_Y^{(2)} \end{bmatrix}$ in which $\boldsymbol{V}_Y^{(1)} \in \mathbb{R}^{M\times N}$. And, $\hat{\boldsymbol{V}}_Y^{(1)} = \frac{1}{\sqrt{2}}\boldsymbol{V}_Y^{(1)}$, $\hat{\boldsymbol{U}}_Y = \frac{1}{\sqrt{2}}\boldsymbol{U}_Y$. Eigenvalues of $\boldsymbol{\mathcal{Y}}$ are signed singular values of $\boldsymbol{Y}$, therefore the limiting eigenvalue distribution of $\boldsymbol{\mathcal{Y}}$ (ignoring zero eigenvalues) is the same as the limiting symmetrized singular value distribution of $\boldsymbol{Y}$.

Define the resolvent of $\boldsymbol{\mathcal{Y}}$

$$\boldsymbol{G}_{\mathcal{Y}}(z) = \left(\; z\boldsymbol{I} - \boldsymbol{\mathcal{Y}} \;\right)^{-1}$$

Denote the eigenvectors of $\boldsymbol{\mathcal{Y}}$ by $\boldsymbol{y}_i \in \mathbb{R}^{M+N}$, $i = 1, \ldots, M+N$. For $z = x - \mathsf{i}\epsilon$ with $x \in \mathbb{R}$ and $\epsilon \gg \frac{1}{N}$, we have:

$$\boldsymbol{G}_{\mathcal{Y}}(x - \mathsf{i}\epsilon) = \sum_{k=1}^{2N} \frac{x + \mathsf{i}\epsilon}{(x - \tilde{\gamma}_k)^2 + \epsilon^2} \boldsymbol{y}_k \boldsymbol{y}_k^\intercal + \frac{x + \mathsf{i}\epsilon}{x^2 + \epsilon^2} \sum_{k=2N+1}^{N+M} \boldsymbol{y}_k \boldsymbol{y}_k^\intercal$$

where $\tilde{\gamma}_k$ are the eigenvalues of $\boldsymbol{\mathcal{Y}}$, which are in fact the (signed) singular values of $\boldsymbol{Y}$, $\tilde{\gamma}_1 = \gamma_1, \ldots, \tilde{\gamma}_N = \gamma_N, \tilde{\gamma}_{N+1} = -\gamma_1, \ldots, \tilde{\gamma}_{2N} = -\gamma_N$.

Define the vectors $\boldsymbol{r}_i = \begin{bmatrix} \boldsymbol{0}_N \\ \boldsymbol{t}_i^{(r)} \end{bmatrix}$, $\boldsymbol{l}_i = \begin{bmatrix} \boldsymbol{t}_i^{(l)} \\ \boldsymbol{0}_M \end{bmatrix}$ for $\boldsymbol{t}_i^{(r)}, \boldsymbol{t}_i^{(l)}$ right/ left singular vectors of $\boldsymbol{T}$, we have

$$\boldsymbol{r}_i^\intercal \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathrm{i}\epsilon)\big)\boldsymbol{l}_i = \sum_{k=1}^{2N} \frac{\epsilon}{(x-\tilde{\gamma}_k)^2 + \epsilon^2} \big(\boldsymbol{r}_i^\intercal \boldsymbol{y}_k\big)\big(\boldsymbol{l}_i^\intercal \boldsymbol{y}_k\big) + \frac{x + \mathrm{i}\epsilon}{x^2 + \epsilon^2} \sum_{k=2N+1}^{N+M} \big(\boldsymbol{r}_i^\intercal \boldsymbol{y}_k\big)\big(\boldsymbol{l}_i^\intercal \boldsymbol{y}_k\big) \tag{11.53}$$

Given the structure of $\boldsymbol{y}_k$'s in (11.52), we have:

$$\big(\boldsymbol{r}_i^\intercal \boldsymbol{y}_k\big)\big(\boldsymbol{l}_i^\intercal \boldsymbol{y}_k\big) = \begin{cases} \frac{1}{2}\big(\boldsymbol{u}_k^\intercal \boldsymbol{t}_i^{(l)}\big)\big(\boldsymbol{v}_k^\intercal \boldsymbol{t}_i^{(r)}\big) & \text{for } 1 \le k \le N \\ -\frac{1}{2}\big(\boldsymbol{u}_{k-N}^\intercal \boldsymbol{t}_i^{(l)}\big)\big(\boldsymbol{v}_{k-N}^\intercal \boldsymbol{t}_i^{(r)}\big) & \text{for } N+1 \le k \le 2N \\ 0 & \text{for } 2N+1 \le k \le N+M \end{cases}$$

In the limit of large N, the latter quantity is also self-averaging, due to the fact that as $N \to \infty$, these overlaps exhibit asymptotic independence, enabling the law of large numbers to be applied here. We can thus state that:

$$\boldsymbol{r}_i^\intercal \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathrm{i}\epsilon)\big)\boldsymbol{l}_i \xrightarrow{N\to\infty} \int_{\mathbb{R}} \frac{\epsilon}{(x-t)^2 + \epsilon^2} O_T(t, \sigma_i)\bar{\mu}_Y(t)\, dt \tag{11.54}$$

where the overlap function $O_T(t, \sigma_i)$ is extended (continuously) to arbitrary values within the support of $\bar{\mu}_Y$ with the property that $O_T(-t, \sigma_i) = -O_T(t, \sigma_i)$ for $t \in \operatorname{supp}(\mu_Y)$. Sending $\epsilon \to 0$, we find

$$\boldsymbol{r}_i^\intercal \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathrm{i}\epsilon)\big)\boldsymbol{l}_i \approx \pi \bar{\mu}_Y(x) O_T(x, \sigma_i) \tag{11.55}$$

In the next section, we establish a connection between the resolvent $\boldsymbol{G}_{\mathcal{Y}}(z)$ and the signal $\boldsymbol{T}$, which enables us to determine the overlap and consequently the optimal singular values values $\widehat{\xi}^*_{t\,i}$ in terms of the singular values of the observation matrix $\boldsymbol{Y}$.

11.C.2 Resolvent relation for $\boldsymbol{T}$

In this section, we consider estimating $\boldsymbol{T}$, and treat both $\boldsymbol{S}$ and $\boldsymbol{Z}$ as noise. We consider the model to be:

$$\boldsymbol{Y} = \boldsymbol{O}\boldsymbol{S}\boldsymbol{O}^\intercal \boldsymbol{T} + \boldsymbol{U}\boldsymbol{Z}\boldsymbol{V}^\intercal \tag{11.56}$$

where $\boldsymbol{S} = \boldsymbol{S}^\intercal \in \mathbb{R}^{N\times N}, \boldsymbol{Z} \in \mathbb{R}^{N\times M}$ are fixed matrices with limiting eigenvalue/singular value distribution ρ_S, μ_Z, and $\boldsymbol{O}, \boldsymbol{U} \in \mathbb{R}^{N\times N}, \boldsymbol{V} \in \mathbb{R}^{M\times M}$

are independent random Haar matrices. For simplicity of notation, we use $\boldsymbol{W} \equiv \boldsymbol{O}\boldsymbol{S}\boldsymbol{O}^\intercal\boldsymbol{T}$, and $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{(N+M)\times(N+M)}$ the hermitization of $\boldsymbol{W}$. And $\widetilde{\boldsymbol{\mathcal{Z}}}$ denotes the hermitization of the matrix $\boldsymbol{U}\boldsymbol{Z}\boldsymbol{V}^\intercal$.

As in the case for $\boldsymbol{S}$, we express the entries of $\boldsymbol{G}(z) \equiv \boldsymbol{G}_{\mathcal{Y}}(z)$ using Gaussian integral representation, and after applying the replica trick, we find:

$$\begin{aligned}
&\langle G_{ij}(z)\rangle \\
&= \lim_{n\to\infty} \int \Big(\prod_{k=1}^{N+M} \prod_{\tau=1}^{n} d\eta_k^{(\tau)} \Big) \eta_i^{(1)}\eta_j^{(1)} \Big\langle \exp\big\{ -\frac{1}{2}\sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)\intercal}(z\boldsymbol{I} - \boldsymbol{\mathcal{Y}})\boldsymbol{\eta}^{(\tau)}\big\}\Big\rangle_{\boldsymbol{O},\boldsymbol{U},\boldsymbol{V}} \\
&= \lim_{n\to\infty} \int \big(\prod_{k=1}^{N+M} \prod_{\tau=1}^{n} d\eta_k^{(\tau)}\big)\, \eta_i^{(1)}\eta_j^{(1)} \exp\big\{ -\frac{z}{2}\sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)\intercal}\boldsymbol{\eta}^{(\tau)}\big\} \\
&\qquad\qquad \times \Big\langle \exp\big\{\frac{1}{2}\sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)\intercal}\boldsymbol{\mathcal{W}}\boldsymbol{\eta}^{(\tau)}\big\}\Big\rangle_{\boldsymbol{O}} \Big\langle \exp\big\{\frac{1}{2}\sum_{\tau=1}^{n} \boldsymbol{\eta}^{(\tau)\intercal}\widetilde{\boldsymbol{\mathcal{Z}}}\boldsymbol{\eta}^{(\tau)}\big\}\Big\rangle_{\boldsymbol{U},\boldsymbol{V}}
\end{aligned} \tag{11.57}$$

Split each replica $\boldsymbol{\eta}^{(\tau)}$ into two vectors $\boldsymbol{a}^{(\tau)} \in \mathbb{R}^N, \boldsymbol{b}^{(\tau)} \in \mathbb{R}^M$, $\boldsymbol{\eta}^{(\tau)} = \begin{bmatrix} \boldsymbol{a}^{(\tau)} \\ \boldsymbol{b}^{(\tau)} \end{bmatrix}$. The exponent in the first bracket in (11.57) can be written as :

$$\begin{aligned}
\boldsymbol{\eta}^{(\tau)\intercal}\boldsymbol{\mathcal{W}}\boldsymbol{\eta}^{(\tau)} &= \boldsymbol{a}^{(\tau)\intercal}\boldsymbol{O}\boldsymbol{S}\boldsymbol{O}^\intercal\boldsymbol{T}\boldsymbol{b}^{(\tau)} + \boldsymbol{b}^{(\tau)\intercal}\boldsymbol{T}^\intercal\boldsymbol{O}\boldsymbol{S}\boldsymbol{O}^\intercal\boldsymbol{a}^{(\tau)} \\
&= \operatorname{Tr}\boldsymbol{O}\boldsymbol{S}\boldsymbol{O}^\intercal\big(\underbrace{\boldsymbol{T}\boldsymbol{b}^{(\tau)}\boldsymbol{a}^{(\tau)\intercal} + \boldsymbol{a}^{(\tau)}\boldsymbol{b}^{(\tau)\intercal}\boldsymbol{T}^\intercal}_{\tilde{\boldsymbol{T}}^{(\tau)}}\big)
\end{aligned} \tag{11.58}$$

where $\tilde{\boldsymbol{T}}^{(\tau)}$ is a symmetric $N \times N$ matrix with two non-zero eigenvalues $\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{T}\boldsymbol{b}^{(\tau)} \pm \|\boldsymbol{a}^{(\tau)}\|\|\boldsymbol{T}\boldsymbol{b}^{(\tau)}\|$ by lemma 11.3.

Using the formula for the spherical integral [62] (see (6.4)), we find:

$$\begin{aligned}
&\Big\langle \exp\big\{\frac{1}{2}\sum_{\tau=1}^{n} \operatorname{Tr}\boldsymbol{O}\boldsymbol{S}\boldsymbol{O}^\intercal\tilde{\boldsymbol{T}}^{(\tau)}\big\}\Big\rangle_{\boldsymbol{O}} \\
&\qquad \approx \exp\Big\{\frac{N}{2}\sum_{\tau=1}^{n} \mathcal{P}_{\rho_S}\Big(\frac{1}{N}\big(\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{T}\boldsymbol{b}^{(\tau)} + \|\boldsymbol{a}^{(\tau)}\|\|\boldsymbol{T}\boldsymbol{b}^{(\tau)}\|\big)\Big) \\
&\qquad\qquad + \mathcal{P}_{\rho_S}\Big(\frac{1}{N}\big(\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{T}\boldsymbol{b}^{(\tau)} - \|\boldsymbol{a}^{(\tau)}\|\|\boldsymbol{T}\boldsymbol{b}^{(\tau)}\|\big)\Big)\Big\}
\end{aligned} \tag{11.59}$$

By the same computation as previous section, for the second bracket we have:

$$\Big\langle \exp\big\{\sum_{\tau=1}^{n} \operatorname{Tr}\boldsymbol{b}^{(\tau)}\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{U}\boldsymbol{Z}\boldsymbol{V}^\intercal\big\}\Big\rangle_{\boldsymbol{U},\boldsymbol{V}} \approx \exp\Big\{\frac{N}{2}\sum_{\tau=1}^{n} \mathcal{Q}_{\mu_Z}^{(\alpha)}\big(\frac{1}{NM}\|\boldsymbol{a}^{(\tau)}\|^2\|\boldsymbol{b}^{(\tau)}\|^2\big)\Big\} \tag{11.60}$$

From (11.57), (11.59), (11.60), we find:

$$\begin{aligned}\langle G_{ij}(z)\rangle = \lim_{n\to\infty} \int \big(\prod_{k=1}^{N+M} \prod_{\tau=1}^{n} d\eta_k^{(\tau)} \big)\, \eta_i^{(1)} \eta_j^{(1)} \\ \times \exp\Big\{ -\frac{1}{2} \sum_{\tau=1}^{n} \Big[z\|\boldsymbol{\eta}^{(\tau)}\|^2 - N\mathcal{Q}_{\mu_Z}^{(\alpha)}\big(\frac{1}{NM}\|\boldsymbol{a}^{(\tau)}\|^2\|\boldsymbol{b}^{(\tau)}\|^2\big) \\ - N\mathcal{P}_{\rho_S}\Big(\frac{1}{N}\big(\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{T}\boldsymbol{b}^{(\tau)} + \|\boldsymbol{a}^{(\tau)}\|\|\boldsymbol{T}\boldsymbol{b}^{(\tau)}\|\big)\Big) \\ - N\mathcal{P}_{\rho_S}\Big(\frac{1}{N}\big(\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{T}\boldsymbol{b}^{(\tau)} - \|\boldsymbol{a}^{(\tau)}\|\|\boldsymbol{T}\boldsymbol{b}^{(\tau)}\|\big)\Big)\Big]\Big\}\end{aligned} \tag{11.61}$$

Now, we introduce delta functions (for brevity we drop the limit term):

$$\begin{aligned}\langle G_{ij}(z)\rangle = \int \big(\prod_{k=1}^{N+M} \prod_{\tau=1}^{n} d\eta_k^{(\tau)} \big) \big(\prod_{\tau=1}^{n} dp_1^{(\tau)}\, dq_2^{(\tau)}\, dq_3^{(\tau)}\, dq_4^{(\tau)} \big)\, \eta_i^{(1)} \eta_j^{(1)} \\ \times \prod_{\tau=1}^{n} \delta\big(q_1^{(\tau)} - \frac{1}{N}\|\boldsymbol{a}^{(\tau)}\|^2\big)\, \delta\big(q_2^{(\tau)} - \frac{1}{M}\|\boldsymbol{b}^{(\tau)}\|^2\big) \\ \times \delta\big(q_3^{(\tau)} - \frac{1}{N}\|\boldsymbol{T}\boldsymbol{b}^{(\tau)}\|^2\big)\, \delta\big(q_4^{(\tau)} - \frac{1}{N}\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{T}\boldsymbol{b}^{(\tau)}\big) \\ \times \exp\Big\{ -\frac{1}{2} \sum_{\tau=1}^{n} z\|\boldsymbol{\eta}^{(\tau)}\|^2 - N\mathcal{Q}_{\mu_Z}^{(\alpha)}(q_1^{(\tau)} q_2^{(\tau)}) \\ - N\mathcal{P}_{\rho_S}\big(q_4^{(\tau)} + \sqrt{q_1^{(\tau)} q_3^{(\tau)}}\big) - N\mathcal{P}_{\rho_S}\big(q_4^{(\tau)} - \sqrt{q_1^{(\tau)} q_3^{(\tau)}}\big)\Big\}\end{aligned} \tag{11.62}$$

In the next step, we replace each delta with its Fourier transform. Note that for the parameters q_1, q_2, q_3 we use $\delta\big(q_1^\tau - \frac{1}{N}\|\boldsymbol{a}^\tau\|^2\big) \propto \int d\beta_1^\tau \exp\Big\{ -\frac{N}{2}\beta_1^\tau\big(q_1^\tau - \frac{1}{N}\|\boldsymbol{a}^\tau\|^2\big)\Big\}$, and for q_4 we use $\delta\big(q_4^{(\tau)} - \frac{1}{N}\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{T}\boldsymbol{b}^{(\tau)}\big) \propto \int d\beta_1^\tau \exp\Big\{ -N\beta_1^\tau\big(q_4^{(\tau)} -$

$\frac{1}{N}\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{T}\boldsymbol{b}^{(\tau)}\big)\Big\}$. After rearranging, we find:

$$\begin{aligned}\langle G_{ij}(z)\rangle \propto \int \big(\prod_{\tau=1}^{n} dq_1^{(\tau)}\, dq_2^{(\tau)}\, dq_3^{(\tau)}\, dq_4^{(\tau)}\, d\beta_1^{(\tau)}\, d\beta_2^{(\tau)}\, d\beta_3^{(\tau)}\, d\beta_4^{(\tau)}\big)\\ \times \exp\Big\{\frac{N}{2}\sum_{\tau=1}^{n}\mathcal{Q}_{\mu_Z}^{(\alpha)}(q_1^{(\tau)}q_2^{(\tau)}) + \mathcal{P}_{\rho_S}\big(q_4^{(\tau)}+\sqrt{q_1^{(\tau)}q_3^{(\tau)}}\big)+\mathcal{P}_{\rho_S}\big(q_4^{(\tau)}-\sqrt{q_1^{(\tau)}q_3^{(\tau)}}\big)\\ -\beta_1^{(\tau)}q_1^{(\tau)}-\frac{1}{\alpha}\beta_2^{(\tau)}q_2^{(\tau)}-\beta_3^{(\tau)}q_3^{(\tau)}-2\beta_4^{(\tau)}q_4^{(\tau)}\Big\}\\ \times\int\big(\prod_{k=1}^{N+M}\prod_{\tau=1}^{n}d\eta_k^{(\tau)}\big)\,\eta_i^{(1)}\eta_j^{(1)}\exp\Big\{-\frac{1}{2}\sum_{\tau=1}^{n}z\|\boldsymbol{\eta}^{(\tau)}\|-\beta_1^{(\tau)}\|\boldsymbol{a}^{(\tau)}\|^2-\beta_2^{(\tau)}\|\boldsymbol{b}^{(\tau)}\|^2\\ -\beta_3^{(\tau)}\|\boldsymbol{T}\boldsymbol{b}^{(\tau)}\|^2-2\beta_4^{(\tau)}\boldsymbol{a}^{(\tau)\intercal}\boldsymbol{T}\boldsymbol{b}^{(\tau)}\Big\}\end{aligned} \tag{11.63}$$

The inner integral is a Gaussian integral, and can be written as:

$$\begin{aligned}\int\big(\prod_{k=1}^{N+M}\prod_{\tau=1}^{n}d\eta_k^{(\tau)}\big)\,\eta_i^{(1)}\eta_j^{(1)}\\ \times\exp\Big\{\sum_{\tau=1}^{n}-\frac{1}{2}\boldsymbol{\eta}^{(\tau)\intercal}\begin{bmatrix}(z-\beta_1^{(\tau)})\boldsymbol{I}_N & -\beta_4^{(\tau)}\boldsymbol{T}\\ -\beta_4^{(\tau)}\boldsymbol{T}^\intercal & (z-\beta_2^{(\tau)})\boldsymbol{I}_M-\beta_3^{(\tau)}\boldsymbol{T}^\intercal\boldsymbol{T}\end{bmatrix}\boldsymbol{\eta}^{(\tau)}\Big\}\end{aligned} \tag{11.64}$$

Denote the matrix in the exponent by $\boldsymbol{C}_T^{(\tau)}$. Using the formula for determinant of block matrices (see proposition 2.8.4 in [136]), we have::

$$\begin{aligned}\det\boldsymbol{C}_T^{(\tau)} &= \det\Big[(z-\beta_1^{(\tau)})\boldsymbol{I}_N-{\beta_4^{(\tau)}}^2\boldsymbol{T}\big((z-\beta_2^{(\tau)})\boldsymbol{I}_M-\beta_3^{(\tau)}\boldsymbol{T}^\intercal\boldsymbol{T}\big)^{-1}\boldsymbol{T}^\intercal\Big]\\ &\qquad\times\det\big[(z-\beta_2^{(\tau)})\boldsymbol{I}_M-\beta_3^{(\tau)}\boldsymbol{T}^\intercal\boldsymbol{T}\big]\\ &=\prod_{k=1}^{N}\big[z-\beta_1^{(\tau)}-{\beta_4^{(\tau)}}^2\frac{\sigma_k^2}{z-\beta_2^{(\tau)}-\beta_3^{(\tau)}\sigma_k^2}\big]\prod_{k=1}^{N}\big(z-\beta_2^{(\tau)}-\beta_3^{(\tau)}\sigma_k^2\big)\big(z-\beta_2^{(\tau)}\big)^{M-N}\\ &=\big(z-\beta_2^{(\tau)}\big)^{M-N}\prod_{k=1}^{N}\big[(z-\beta_1^{(\tau)})(z-\beta_2^{(\tau)}-\beta_3^{(\tau)}\sigma_k^2)-{\beta_4^{(\tau)}}^2\sigma_k^2\big]\\ &=\big(z-\beta_2^{(\tau)}\big)^{M-N}\prod_{k=1}^{N}\Big[(z-\beta_1^{(\tau)})(z-\beta_2^{(\tau)})-\big({\beta_4^{(\tau)}}^2+\beta_3^{(\tau)}(z-\beta_1^{(\tau)})\big)\sigma_k^2\Big]\end{aligned}$$

where σ_k's are the singular values of $\boldsymbol{T}$. So computing the Gaussian integrals, (11.63) can be written as:

$$\begin{aligned}\langle G_{ij}(z)\rangle \propto \int\big(\prod_{\tau=1}^{n}dq_1^{(\tau)}\,dq_2^{(\tau)}\,dq_3^{(\tau)}\,dq_4^{(\tau)}\,d\beta_1^{(\tau)}\,d\beta_2^{(\tau)}\,d\beta_3^{(\tau)}\,d\beta_4^{(\tau)}\big)\big(\boldsymbol{C}_T^{(1)^{-1}}\big)_{ij}\\ \times\exp\big\{-\frac{Nn}{2}F_0^T(\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3,\boldsymbol{q}_4,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4)\big\}\end{aligned} \tag{11.65}$$

with

$$
\begin{aligned}
F_0^T(\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3,\boldsymbol{q}_4,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\boldsymbol{\beta}_3,\boldsymbol{\beta}_4) = \frac{1}{n}\sum_{\tau=1}^{n}\Bigg[&\big(\frac{1}{\alpha}-1\big)\ln(z-\beta_2^{(\tau)}) \\
&+\frac{1}{N}\sum_{k=1}^{N}\ln\Big((z-\beta_1^{(\tau)})(z-\beta_2^{(\tau)})-\big({\beta_4^{(\tau)}}^2+\beta_3^{(\tau)}(z-\beta_1^{(\tau)})\big)\sigma_k^2\Big) \\
&-\mathcal{Q}_{\mu_Z}^{(\alpha)}(q_1^{(\tau)}q_2^{(\tau)})-\mathcal{P}_{\rho_S}\big(q_4^{(\tau)}+\sqrt{q_1^{(\tau)}q_3^{(\tau)}}\big)-\mathcal{P}_{\rho_S}\big(q_4^{(\tau)}-\sqrt{q_1^{(\tau)}q_3^{(\tau)}}\big) \\
&+\beta_1^{(\tau)}q_1^{(\tau)}+\frac{1}{\alpha}\beta_2^{(\tau)}q_2^{(\tau)}+\beta_3^{(\tau)}q_3^{(\tau)}+2\beta_4^{(\tau)}q_4^{(\tau)}\Bigg]
\end{aligned}
\tag{11.66}
$$

We will evaluate the integral (11.63) using saddle-points of the function F_0^T. From the *replica symmetric ansatz* at the saddle-point we have:

$$
\forall\tau\in\{1,\cdots,n\}:\quad
\begin{cases}
q_1^\tau=q_1, \quad q_2^\tau=q_2, \quad q_3^\tau=q_3, \quad q_4^\tau=q_4\\
\beta_1^\tau=\beta_1, \quad \beta_2^\tau=\beta_2, \quad \beta_3^\tau=\beta_3, \quad \beta_4^\tau=\beta_4
\end{cases}
$$

Finally, we find the solution to be:

$$
\begin{cases}
\beta_1^* = \frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(q_1^*q_2^*)}{q_1^*}+\frac{1}{2}\sqrt{\frac{q_3^*}{q_1^*}}\Big(\mathcal{R}_{\rho_S}\big(q_4^*+\sqrt{q_1^*q_3^*}\big)-\mathcal{R}_{\rho_S}\big(q_4^*-\sqrt{q_1^*q_3^*}\big)\Big)\\
\beta_2^* = \alpha\frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(q_1^*q_2^*)}{q_2^*}\\
\beta_3^* = \frac{1}{2}\sqrt{\frac{q_1^*}{q_3^*}}\Big(\mathcal{R}_{\rho_S}\big(q_4^*+\sqrt{q_1^*q_3^*}\big)-\mathcal{R}_{\rho_S}\big(q_4^*-\sqrt{q_1^*q_3^*}\big)\Big)\\
\beta_4^* = \frac{1}{2}\Big(\mathcal{R}_{\rho_S}\big(q_4^*+\sqrt{q_1^*q_3^*}\big)+\mathcal{R}_{\rho_S}\big(q_4^*-\sqrt{q_1^*q_3^*}\big)\Big)\\
q_1^* = \frac{(z-\beta_2^*){\beta_4^*}^2}{Z_2(z)^2}\mathcal{G}_{\rho_T}\big(\frac{Z_1(z)}{Z_2(z)}\big)+\frac{\beta_3^*}{Z_2(z)}\\
q_2^* = \alpha\frac{z-\beta_1^*}{Z_2(z)}\mathcal{G}_{\rho_T}\big(\frac{Z_1(z)}{Z_2(z)}\big)+\frac{1-\alpha}{z-\beta_2^*}\\
q_3^* = \frac{(z-\beta_1^*)Z_1(z)}{Z_2(z)^2}\mathcal{G}_{\rho_T}\big(\frac{Z_1(z)}{Z_2(z)}\big)-\frac{z-\beta_1^*}{Z_2(z)}\\
q_4^* = \frac{\beta_4^*Z_1(z)}{Z_2(z)^2}\mathcal{G}_{\rho_T}\big(\frac{Z_1(z)}{Z_2(z)}\big)-\frac{\beta_4^*}{Z_2(z)}
\end{cases}
\tag{11.67}
$$

with

$$
\begin{cases}
Z_1(z)=(z-\beta_1^*)(z-\beta_2^*)\\
Z_2(z)={\beta_4^*}^2+\beta_3^*(z-\beta_1^*)
\end{cases}
$$

where ρ_T is the limiting eigenvalue distribution of $\boldsymbol{T}\boldsymbol{T}^\intercal$.

The relation (11.65) and the solutions (11.67) hold for arbitrary indices i,j, so we can state the relation in the matrix form. Computing the inverse of ${\boldsymbol{C}_T^*}^{-1}$ (see section 11.F), we have:

$$
\begin{aligned}
\big\langle \boldsymbol{G}_{\mathcal{Y}}(z)\big\rangle_{\boldsymbol{O},\boldsymbol{U},\boldsymbol{V}} &= \left\langle \begin{bmatrix} \frac{1}{z}\boldsymbol{I}_N+\frac{1}{z}\boldsymbol{Y}\boldsymbol{G}_{Y^\intercal Y}(z^2)\boldsymbol{Y}^\intercal & \boldsymbol{Y}\boldsymbol{G}_{Y^\intercal Y}(z^2)\\ \boldsymbol{G}_{Y^\intercal Y}(z^2)\boldsymbol{Y}^\intercal & z\boldsymbol{G}_{Y^\intercal Y}(z^2)\end{bmatrix}\right\rangle \\
&= \begin{bmatrix} \frac{1}{z-\beta_1^*}\boldsymbol{I}_N+\frac{{\beta_4^*}^2}{(z-\beta_1^*)Z_2(z)}\boldsymbol{T}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big)\boldsymbol{T}^\intercal & \frac{\beta_4^*}{Z_2(z)}\boldsymbol{T}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big)\\ \frac{\beta_4^*}{Z_2(z)}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big)\boldsymbol{T}^\intercal & \frac{z-\beta_1^*}{Z_2(z)}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big)\end{bmatrix}
\end{aligned}
\tag{11.68}
$$

With this relation, we can further simplify the solution (11.67).

We start with comparing the trace of upper-left block in (11.68). The normalized trace of the first block in $\langle \boldsymbol{G}_{\mathcal{Y}}(z) \rangle_{\boldsymbol{O},\boldsymbol{U},\boldsymbol{V}}$ is computed in (11.39) to be $\mathcal{G}_{\bar{\mu}_Y}(z)$. The normalized trace of the upper-left block in $\boldsymbol{C}_T^{*\,-1}$ is:

$$\begin{aligned}
&\frac{1}{N}\operatorname{Tr}\Big[(z-\beta_1^*)^{-1}\boldsymbol{I}_N + \frac{\beta_4^{*2}}{(z-\beta_1^*)Z_2(z)}\boldsymbol{T}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big)\boldsymbol{T}^\intercal\Big] \\
&\quad= \frac{1}{N}\frac{1}{z-\beta_1^*}\sum_{k=1}^{N}\Big[1+\frac{\beta_4^{*2}}{Z_2(z)}\frac{\sigma_k^2}{\frac{Z_1(z)}{Z_2(z)}-\sigma_k^2}\Big] \\
&\quad= \frac{1}{N}\frac{1}{z-\beta_1^*}\sum_{k=1}^{N}\Big[\frac{\beta_4^{*2}Z_1(z)}{Z_2^2(z)}\frac{1}{\frac{Z_1(z)}{Z_2(z)}-\sigma_k^2}+1-\frac{\beta_4^{*2}}{Z_2(z)}\Big] \\
&\quad= \frac{1}{N}\frac{1}{z-\beta_1^*}\frac{\beta_4^{*2}Z_1(z)}{Z_2^2(z)}\sum_{k=1}^{N}\frac{1}{\frac{Z_1(z)}{Z_2(z)}-\sigma_k^2}+\frac{1}{z-\beta_1^*}\frac{\beta_3^*(z-\beta_1^*)}{Z_2(z)} \\
&\quad= \frac{(z-\beta_2^*)\beta_4^{*2}}{Z_2(z)^2}\mathcal{G}_{\rho_T}\big(\frac{Z_1(z)}{Z_2(z)}\big)+\frac{\beta_3^*}{Z_2(z)} \\
&\quad= q_1^*
\end{aligned} \tag{11.69}$$

Thus, $q_1^* = \mathcal{G}_{\bar{\mu}_Y}(z)$.

The normalized trace of the lower-right block of $\langle \boldsymbol{G}_{\mathcal{Y}}(z) \rangle_{\boldsymbol{O},\boldsymbol{U},\boldsymbol{V}}$ is $\alpha\mathcal{G}_{\bar{\mu}_Y}(z) + (1-\alpha)\frac{1}{z}$ (see (11.40)). The normalized trace of the lower-right block in $\boldsymbol{C}_T^{*\,-1}$ is:

$$\begin{aligned}
&\frac{1}{M}\operatorname{Tr}\Big[\frac{z-\beta_1^*}{Z_2(z)}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big)\Big] \\
&\quad= \frac{1}{M}\frac{z-\beta_1^*}{Z_2(z)}\sum_{k=1}^{N}\frac{1}{\frac{Z_1(z)}{Z_2(z)}-\sigma_k^2}+\frac{M-N}{M}\frac{z-\beta_1^*}{Z_2(z)}\frac{Z_2(z)}{Z_1(z)} \\
&\quad= \frac{N}{M}\frac{1}{N}\frac{z-\beta_1^*}{Z_2(z)}\sum_{k=1}^{N}\frac{1}{\frac{Z_1(z)}{Z_2(z)}-\sigma_k^2}+\frac{M-N}{M}\frac{z-\beta_1^*}{Z_1(z)} \\
&\quad= \alpha\frac{z-\beta_1^*}{Z_2(z)}\mathcal{G}_{\rho_T}\big(\frac{Z_1(z)}{Z_2(z)}\big)+\frac{1-\alpha}{z-\beta_2^*} \\
&\quad= q_2^*
\end{aligned} \tag{11.70}$$

So, $q_2^* = \alpha\mathcal{G}_{\bar{\mu}_Y}(z) + (1-\alpha)\frac{1}{z}$.

With a bit of algebra, we can express the parameters q_3^*, q_4^* in terms of $q_1^*, \beta_1^*, \beta_4^*$:

$$q_3^* = \frac{(z-\beta_1^*)^2}{\beta_4^{*2}}q_1^* - \frac{z-\beta_1^*}{\beta_4^{*2}}, \quad q_4^* = \frac{z-\beta_1^*}{\beta_4^*}q_1^* - \frac{1}{\beta_4^*} \tag{11.71}$$

Therefore, the solution can be written without involving $\mathcal{G}_{\rho_T}$, as:

$$\begin{cases} \beta_1^* = \frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(q_1^* q_2^*)}{q_1^*} + \frac{1}{2}\sqrt{\frac{q_3^*}{q_1^*}}\Big(\mathcal{R}_{\rho_S}\big(q_4^* + \sqrt{q_1^* q_3^*}\big) - \mathcal{R}_{\rho_S}\big(q_4^* - \sqrt{q_1^* q_3^*}\big)\Big) \\ \beta_2^* = \alpha\frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(q_1^* q_2^*)}{q_2^*} \\ \beta_3^* = \frac{1}{2}\sqrt{\frac{q_1^*}{q_3^*}}\Big(\mathcal{R}_{\rho_S}\big(q_4^* + \sqrt{q_1^* q_3^*}\big) - \mathcal{R}_{\rho_S}\big(q_4^* - \sqrt{q_1^* q_3^*}\big)\Big) \\ \beta_4^* = \frac{1}{2}\Big(\mathcal{R}_{\rho_S}\big(q_4^* + \sqrt{q_1^* q_3^*}\big) + \mathcal{R}_{\rho_S}\big(q_4^* - \sqrt{q_1^* q_3^*}\big)\Big) \\ q_1^* = \mathcal{G}_{\bar{\mu}_Y}(z) \\ q_2^* = \alpha\mathcal{G}_{\bar{\mu}_Y}(z) + (1-\alpha)\frac{1}{z} \\ q_3^* = \frac{(z-\beta_1^*)^2}{\beta_4^{*2}}\mathcal{G}_{\bar{\mu}_Y}(z) - \frac{z-\beta_1^*}{\beta_4^{*2}} \\ q_4^* = \frac{z-\beta_1^*}{\beta_4^*}\mathcal{G}_{\bar{\mu}_Y}(z) - \frac{1}{\beta_4^*} \end{cases} \tag{11.72}$$

Remark 11.5. *The simplifications in* (11.71) *are derived with the assumption that* $\beta_4^* \neq 0$. *However, in the initial set of equations* (11.67)*, if* ρ_S *is symmetric measure then* $\beta_4^* = q_4^* = 0$ *is a solution. If* ρ_S *is symmetric, then* $\mathcal{R}_{\rho_S}(-z) = -\mathcal{R}_{\rho_S}(z)$*, and plugging* $q_4^* = 0$ *in the expression for* β_4^* *in* (11.67)*, we find that* $\beta_4^* = 0$.

11.C.3 Overlaps and the optimal singular values

From (11.55), (11.68), we find:

$$\begin{aligned} O_T(\gamma, \sigma_i) &\approx \frac{1}{\pi\bar{\mu}_Y(\gamma)}\,\mathrm{Im}\lim_{z\to\gamma-\mathrm{i}0^+}\frac{\beta_4^*}{Z_2(z)}\boldsymbol{t}_i^{(r)\intercal}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big)\boldsymbol{T}^\intercal\boldsymbol{t}_i^{(l)} \\ &= \frac{1}{\pi\bar{\mu}_Y(\gamma)}\,\mathrm{Im}\lim_{z\to\gamma-\mathrm{i}0^+}\beta_4^*\frac{\sigma_i}{Z_1(z) - Z_2(z)\sigma_i^2} \end{aligned} \tag{11.73}$$

From the overlap, we can compute the optimal singular values:

$$\begin{aligned} \widehat{\xi}_{t\,i}^* &\approx \frac{1}{N}\sum_{j=1}^{N}\sigma_j O_T(\gamma_i, \sigma_j) \\ &\approx \frac{1}{\pi\bar{\mu}_Y(\gamma_i)}\,\mathrm{Im}\lim_{z\to\gamma_i-\mathrm{i}0^+}\frac{1}{N}\sum_{j=1}^{N}\beta_4^*\frac{\sigma_j^2}{Z_1(z) - Z_2(z)\sigma_j^2} \\ &= \frac{1}{\pi\bar{\mu}_Y(\gamma_i)}\,\mathrm{Im}\lim_{z\to\gamma_i-\mathrm{i}0^+}\frac{1}{N}\frac{\beta_4^*}{Z_2(z)}\sum_{j=1}^{N}\frac{\sigma_j^2}{\frac{Z_1(z)}{Z_2(z)} - \sigma_j^2} \\ &= \frac{1}{\pi\bar{\mu}_Y(\gamma_i)}\,\mathrm{Im}\lim_{z\to\gamma_i-\mathrm{i}0^+}\frac{1}{N}\frac{\beta_4^*}{Z_2(z)}\sum_{j=1}^{N}\left[\frac{\frac{Z_1(z)}{Z_2(z)}}{\frac{Z_1(z)}{Z_2(z)} - \sigma_j^2} - 1\right] \\ &\approx \frac{1}{\pi\bar{\mu}_Y(\gamma_i)}\,\mathrm{Im}\lim_{z\to\gamma_i-\mathrm{i}0^+}\frac{\beta_4^* Z_1(z)}{Z_2(z)^2}\mathcal{G}_{\rho_T}\big(\frac{Z_1(z)}{Z_2(z)}\big) - \frac{\beta_4^*}{Z_2(z)} \\ &= \frac{1}{\pi\bar{\mu}_Y(\gamma_i)}\,\mathrm{Im}\lim_{z\to\gamma_i-\mathrm{i}0^+}q_4^* \end{aligned} \tag{11.74}$$

where in the last equality we used the solution we have found in (11.67). Note that, based on (11.72), we do not need to have any knowledge about ρ_T to compute q_4^*. In the end, we need to divide the estimator by $\sqrt{\kappa}$ as we have absorbed it into $\boldsymbol{T}$.

Recovering the rectangular RIE for a denoising problem

Note that if in the model (11.56), we put $\boldsymbol{S} = \boldsymbol{I}$ the model reduces to the additive denoising of $\boldsymbol{T}$, and we recover the estimator derived in previous chapter (10.5) ([54]) for the rectangular case.

For $\boldsymbol{S} = \boldsymbol{I}$, $\mathcal{R}_{\rho_S}(z) = 1$, so (11.72) reduces to:

$$\begin{cases} \beta_1^* = \frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(q_1^* q_2^*)}{q_1^*}, \quad \beta_2^* = \alpha \frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(q_1^* q_2^*)}{q_2^*}, \quad \beta_3^* = 0, \quad \beta_4^* = 1 \\ q_1^* = \mathcal{G}_{\bar{\mu}_Y}(z), \quad q_2^* = \alpha \mathcal{G}_{\bar{\mu}_Y}(z) + (1-\alpha)\frac{1}{z} \\ q_3^* = (z-\beta_1^*)^2 \mathcal{G}_{\bar{\mu}_Y}(z) - (z-\beta_1^*), \quad q_4^* = (z-\beta_1^*)\mathcal{G}_{\bar{\mu}_Y}(z) - 1 \end{cases} \tag{11.75}$$

From (11.74), we have:

$$\begin{aligned} \widehat{\xi}_{t\,i}^* &= \frac{1}{\pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z \to \gamma_i - \mathsf{i}0^+} q_4^* \\ &= \frac{1}{\pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z \to \gamma_i - \mathsf{i}0^+} z\mathcal{G}_{\bar{\mu}_Y}(z) - \beta_1^* \mathcal{G}_{\bar{\mu}_Y}(z) - 1 \\ &= \frac{1}{\pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z \to \gamma_i - \mathsf{i}0^+} z\mathcal{G}_{\bar{\mu}_Y}(z) - \frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(q_1^* q_2^*)}{q_1^*} \mathcal{G}_{\bar{\mu}_Y}(z) - 1 \\ &= \frac{1}{\pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z \to \gamma_i - \mathsf{i}0^+} z\mathcal{G}_{\bar{\mu}_Y}(z) - \mathcal{C}_{\mu_Z}^{(\alpha)}(q_1^* q_2^*) - 1 \\ &= \frac{1}{\pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z \to \gamma_i - \mathsf{i}0^+} z\mathcal{G}_{\bar{\mu}_Y}(z) - \mathcal{C}_{\mu_Z}^{(\alpha)}\Big(\mathcal{G}_{\bar{\mu}_Y}(z)\big(\alpha \mathcal{G}_{\bar{\mu}_Y}(z) + (1-\alpha)\frac{1}{z}\big)\Big) - 1 \\ &\overset{(a)}{=} \frac{1}{\pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \Bigg[\gamma_i \mathcal{G}_{\bar{\mu}_Y}(\gamma_i - \mathsf{i}0^+) \\ &\qquad - \mathcal{C}_{\mu_Z}^{(\alpha)}\bigg(\frac{1}{\gamma_i}\mathcal{G}_{\bar{\mu}_Y}(\gamma_i - \mathsf{i}0^+)\Big(1 - \alpha + \alpha\gamma_i \mathcal{G}_{\bar{\mu}_Y}(\gamma_i - \mathsf{i}0^+)\Big)\bigg)\Bigg] \\ &\overset{(b)}{=} \gamma_i - \frac{1}{\pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \mathcal{C}_{\mu_Z}^{(\alpha)}\Bigg(\frac{1-\alpha}{\gamma_i}\pi \mathsf{H}[\bar{\mu}_Y](\gamma_i) + \alpha\big(\pi \mathsf{H}[\bar{\mu}_Y](\gamma_i)\big)^2 \\ &\qquad - \alpha\big(\pi \bar{\mu}_Y(\gamma_i)\big)^2 + \mathsf{i}\pi \bar{\mu}_Y(\gamma_i)\big(\frac{1-\alpha}{\gamma_i} + 2\alpha\pi \mathsf{H}[\bar{\mu}_Y](\gamma_i)\big)\Bigg) \end{aligned} \tag{11.76}$$

where in (a) we used the analyticity of rectangular R-transform [74], and in (b), we used Plemelj formula (5.5). Note that, the final estimator should be divided by the $\sqrt{\kappa}$.

11.C.4 Examples

Throughout the numerical experiments, we consider the matrix $\boldsymbol{Z}$ to have i.i.d. Gaussian entries with variance $1/N$, so $\mathcal{C}_{\mu_Z}^{(\alpha)}(z) = \frac{1}{\alpha}z$. And, $\boldsymbol{S} = \boldsymbol{F} + c\boldsymbol{I}$ where $\boldsymbol{F} = \boldsymbol{F}^\intercal \in \mathbb{R}^{N\times N}$ has i.i.d. entries with variance $1/N$, and $c \neq 0$ is a real number, so $\mathcal{R}_{\rho_S}(z) = z + c$. With these choices, the solution (11.72) simplifies to:

$$\begin{cases} \beta_1^* = \frac{1}{\alpha}q_2^* + q_3^*, \quad \beta_2^* = q_1^*, \quad \beta_3^* = q_1^*, \quad \beta_4^* = q_4^* + c \\ q_1^* = \mathcal{G}_{\bar{\mu}_Y}(z), \quad q_2^* = \alpha\mathcal{G}_{\bar{\mu}_Y}(z) + (1-\alpha)\frac{1}{z} \\ q_3^* = \frac{(z-\beta_1^*)^2}{\beta_4^{*2}}\mathcal{G}_{\bar{\mu}_Y}(z) - \frac{z-\beta_1^*}{\beta_4^{*2}}, \quad q_4^* = \frac{z-\beta_1^*}{\beta_4^*}\mathcal{G}_{\bar{\mu}_Y}(z) - \frac{1}{\beta_4^*} \end{cases} \tag{11.77}$$

Note that in (11.77), q_1^*, q_2^* are given in terms of the observation, so to find the solution we only need to find the parameters q_3^*, q_4^*. In (11.77), one can see that we have the relation $q_3^* = \frac{z-\beta_1^*}{\beta_4^*}q_4^*$. Writing the parameters β_1^*, β_4^* in terms of q_2^*, q_3^*, q_4^*, after a bit of algebra we have the following relation:

$$q_3^* = \frac{z - \frac{1}{\alpha}q_2^*}{2q_4^* + c}q_4^* \tag{11.78}$$

In the expression for q_4^* in (11.77), using (11.78) we can rewrite β_1^*, β_4^* in terms of q_2^*, q_4^*. After some manipulations we find that q_4^* is the solution to the following cubic equation:

$$\begin{aligned} 2x^3 + 3c\,x^2 + \Big[c^2 + 2 - \big(z - \mathcal{G}_{\bar{\mu}_Y}(z) - \frac{1-\alpha}{\alpha}\frac{1}{z}\big)\mathcal{G}_{\bar{\mu}_Y}(z)\Big]\,x \\ - c\Big[\big(z - \mathcal{G}_{\bar{\mu}_Y}(z) - \frac{1-\alpha}{\alpha}\frac{1}{z}\big)\mathcal{G}_{\bar{\mu}_Y}(z) - 1\Big] = 0 \end{aligned} \tag{11.79}$$

Based on our numerical simulations, we pick the following root for q_4^*:

$$q_4^* = -\frac{c}{2} - \frac{12 - 3c^2 + 6A}{3\sqrt[3]{B}} + \frac{\sqrt[3]{B}}{12} \tag{11.80}$$

with

$$\begin{aligned} A &= \mathcal{G}_{\bar{\mu}_Y}(z)^2 - \frac{\mathcal{G}_{\bar{\mu}_Y}(z)}{z}\big(1 - \frac{1}{\alpha}\big) - \mathcal{G}_{\bar{\mu}_Y}(z)z \\ B &= -216cA + 4\sqrt{4\big(12 - 3c^2 + 6A\big)^3 + 54^2c^2A^2} \end{aligned}$$

Once we have q_4^*, we can find q_3^* using (11.78). In the end, $\beta_1^*, \cdots, \beta_4^*$ can be evaluated. Note that, for the RIE, only q_4^* is required. Other parameters are used to evaluate the resolvent relation (11.68) and the overlap (11.73).

Resolvent relation

We take $\kappa = 1$. In model (11.56), without loss of generality we can consider $\boldsymbol{T}$ to be diagonal.

In figure 11.C.1, $\boldsymbol{T}$ is the diagonal matrix obtained from the singular values of a Gaussian matrix with i.i.d. entries of variance $1/N$. In figure 11.C.2, the non-zero entries (on main diagonal) of $\boldsymbol{T}$ are uniformly distributed in $[1, 3]$. As in previous cases, $\mu_Y, \mathcal{G}_{\bar{\mu}_Y}(z)$ are estimated numerically using Cauchy kernel, from which the parameters $\beta_1^*, \cdots, \beta_4^*$ are computed.

Overlap

To illustrate the formula for the overlap (11.73), we fix the matrix $\boldsymbol{T}$ and run experiments over various realization of the model (11.56). For each experiment, we record the overlap of k-th singular vectors left and right) of $\boldsymbol{Y}$ and singular vectors of $\boldsymbol{T}$. To compute the theoretical prediction, we evaluate the parameters $\beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*$, for $z = \bar{\gamma}_k - \mathrm{i}0^+$ where $\bar{\gamma}_k$ is the average of k-th singular value of $\boldsymbol{Y}$ in the experiments.

In figure 11.C.3a, the overlap is shown for $\boldsymbol{T}$ with i.i.d. Gaussian entries of variance $\frac{1}{N}$, so μ_T is the Marchenko-Pastur law with aspect-ratio α. In figure 11.C.3b, matrix $\boldsymbol{T}$ is constructed as $\boldsymbol{T} = \boldsymbol{U}_T \boldsymbol{\Sigma} \boldsymbol{V}_T^\intercal$, where $\boldsymbol{U}_T \in \mathbb{R}^{N\times N}, \boldsymbol{V}_T \in \mathbb{R}^{M\times M}$ are Haar distributed orthogonal matrices, and singular values $\sigma_1, \cdots, \sigma_N$ are chosen independently uniformly from $[1, 3]$, so $\mu_T = \mathcal{U}\big([1, 3]\big)$.

RIE performance

In this section, we investigate the performance of our proposed estimators for $\boldsymbol{T}$. To construct the RIE for $\boldsymbol{T}$, we only need q_4^* which we use (11.80). We compare performances of the optimal RIE (11.74) with the one of oracle estimator (11.5).

In figures 11.C.4,11.C.5, the MSE of RIE and the oracle estimator is plotted for three cases of priors: $\boldsymbol{T}$ with Gaussian entries, $\boldsymbol{T}$ with uniform spectral density, and $\boldsymbol{T}$ with Bernoulli spectral density. In all cases, observe that the RIE has the same performance as the oracle estimator.

Effect of aspect-ratio α. In Figure 11.C.6, we take $\boldsymbol{T}$ to have Gaussian entries (with variance $\frac{1}{N}$), and the MSE is depicted for various values of the aspect-ratio α. We see that as M increases (α decreases) the estimation error (of $\boldsymbol{T}$) decreases.

Sparse $\boldsymbol{T}$: a non-rotation invariant example. We consider $\boldsymbol{T}$ to have i.i.d. entries from the Bernoulli-Rademacher distribution,

$$T_{i,j} = \begin{cases} +\frac{1}{\sqrt{N}} & \text{with probability } \frac{1-p}{2} \\ 0 & \text{with probability } p \\ -\frac{1}{\sqrt{N}} & \text{with probability } \frac{1-p}{2} \end{cases}, \qquad \forall \quad 1 \le i \le N, \quad 1 \le j \le M$$

With the normalization $1/\sqrt{N}$, the spectrum of $\boldsymbol{T}$ does not grow with the dimension and has a finite support, thus we can apply our estimator to reconstruct $\boldsymbol{T}$.

(a) Entry $i = j = 1$, first diagonal entry in the upper-left block

(b) Entry $i = 1, j = 2$, non-diagonal entry in the upper-left block

(c) Entry $i = j = 2001$, first diagonal entry in the lower-right block

(d) Entry $i = 2001, j = 2002$, non-diagonal entry in the lower-right block

(e) Entry $i = 1, j = 2001$, first entry in the upper-right block

(f) Entry $i = 1, j = 2002$, second entry in the upper-right block

Figure 11.C.1: Illustration of (11.68). $\boldsymbol{T} \in \mathbb{R}^{N \times M}$ is a diagonal matrix obtained from the singular values of a $N \times M$ matrix with i.i.d. entries of variance $1/N$, $\boldsymbol{S} = \boldsymbol{S}^\intercal$ is shifted Wigner matrix with $c = 3$, and $\boldsymbol{Z}$ is a Gaussian matrices with. The empirical estimate of $\boldsymbol{G}_{\mathcal{Y}}(z)$ (dashed blue line) is computed for $z = \gamma_i - \mathrm{i}\sqrt{\frac{1}{2N}}$ for $1 \leq i \leq N$, for $N = 2000, M = 4000$. Theoretical one (solid orange line) is computed from the rhs of (11.68) with parameters computed from the generated matrix. Note that, the theoretical one has also fluctuations because the parameters $\beta_1^*, \cdots \beta_4^*$ are computed from the numerical estimate of $\mathcal{G}_{\tilde{\mu}_Y}(z)$.

(a) Entry $i = j = 1$, first diagonal entry in the upper-left block

(b) Entry $i = 1, j = 2$, non-diagonal entry in the upper-left block

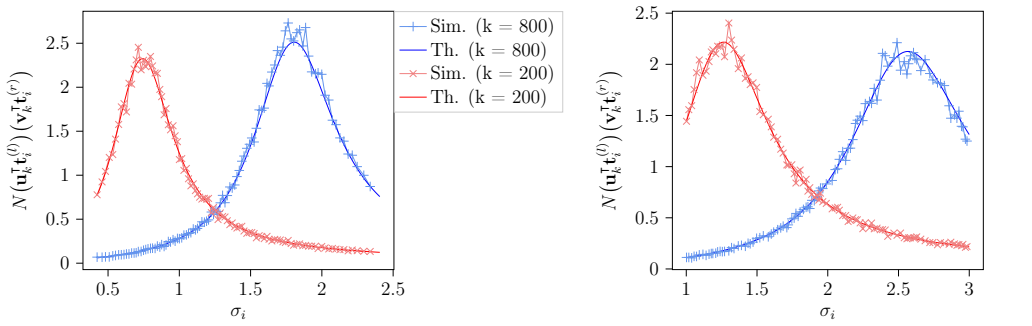
(c) Entry $i = j = 2001$, first diagonal entry in the lower-right block

(d) Entry $i = 2001, j = 2002$, non-diagonal entry in the lower-right block

(e) Entry $i = 1, j = 2001$, first entry in the upper-right block

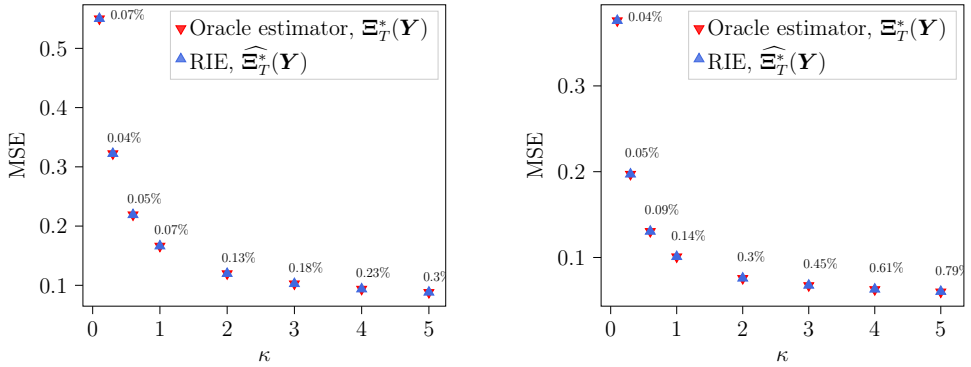
(f) Entry $i = 1, j = 2002$, second entry in the upper-right block

Figure 11.C.2: Illustration of (11.68). $\boldsymbol{T} \in \mathbb{R}^{N\times M}$ is a diagonal matrix with (main) diagonal entries uniformly distributed in $[1, 3]$, $\boldsymbol{S} = \boldsymbol{S}^\intercal$ is shifted Wigner matrix with $c = 3$, and $\boldsymbol{Z}$ is a Gaussian matrices with. The empirical estimate of $\boldsymbol{G}_{\mathcal{Y}}(z)$ (dashed blue line) is computed for $z = \gamma_i - \mathrm{i}\sqrt{\frac{1}{2N}}$ for $1 \le i \le N$, for $N = 2000, M = 4000$. Theoretical one (solid orange line) is computed from the rhs of (11.68) with parameters computed from the generated matrix. Note that, the theoretical one has also fluctuations because the parameters $\beta_1^*, \cdots \beta_4^*$ are computed from the numerical estimate of $\mathcal{G}_{\bar{\mu}_Y}(z)$.



(a) Gaussian $\boldsymbol{T}$ (b) $\boldsymbol{T}$ with Uniform spectrum, $\mathcal{U}([1,3])$

Figure 11.C.3: Computation of the rescaled overlap. $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 3$, and $\boldsymbol{Z}$ has i.i.d. Gaussian entries of variance $1/N$, and $N/M = 1/2$. The simulation results are average of 1000 experiments with fixed $\boldsymbol{T}$, and $N = 1000, M = 2000$. Some of the simulation points are dropped for clarity.



(a) Gaussian (b) Uniform spectral, $\mathcal{U}([1,3])$

Figure 11.C.4: Estimating $\boldsymbol{T}$. MSE is normalized by the norm of the signal, $\|\boldsymbol{T}\|_{\mathrm{F}}^2$. $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 3$, and $\boldsymbol{Z}$ has i.i.d. Gaussian entries of variance $1/N$, and $N/M = 1/2$. The RIE is applied to $N = 2000, M = 4000$, and the results are averaged over 10 runs (error bars are invisible). Average relative error between RIE $\widehat{\boldsymbol{\Xi}_T^*}(\boldsymbol{Y})$ and Oracle estimator is also reported.

Note that the prior of $\boldsymbol{T}$ is not rotationally invariant, and neither the oracle estimator nor the RIE are optimal. Therefore, taking the prior into account, we apply a thresholding function on the entries of the matrix obtained from the RIE, $\widehat{\boldsymbol{\Xi}_T^*}(\boldsymbol{Y})$. We apply the following function on each entry of the estimator:

$$f_h(x) = \begin{cases} +\frac{1}{\sqrt{N}} & \text{if } x > \frac{h}{\sqrt{N}} \\ 0 & \text{if } |x| \leq \frac{h}{\sqrt{N}} \\ -\frac{1}{\sqrt{N}} & \text{if } x < -\frac{h}{\sqrt{N}} \end{cases}, \qquad \text{for } h \in [0,1]$$

In figure 11.C.7, the MSE of the oracle estimator, RIE, and RIE+$f_p(x)$ (with $h = p$) is plotted. A few remarks on this figure are in order. First, RIEs are not limited to rotationally invariant priors and can give non-trivial estimates

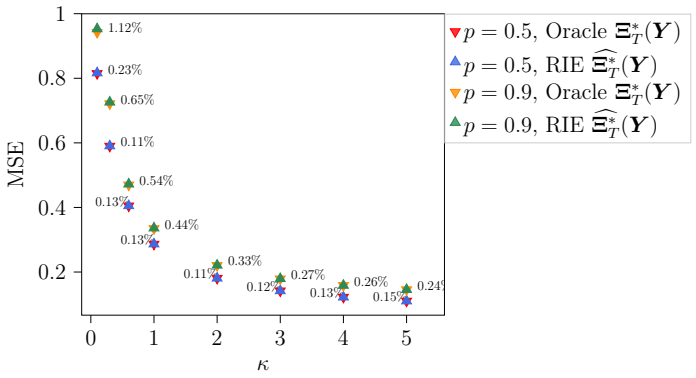


Figure 11.C.5: Estimating $\boldsymbol{T}$ with Bernoulli spectral prior. MSE is normalized by the norm of the signal, $\|\boldsymbol{T}\|_{\mathrm{F}}^2$. $\boldsymbol{T}$ has Bernoulli spectral distribution with parameter p. $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 3$, and $\boldsymbol{Z}$ has i.i.d. Gaussian entries of variance $1/N$, and $N/M = 1/2$. The RIE is applied to $N = 2000, M = 4000$, and the results are averaged over 10 runs (error bars are invisible).Average relative error between RIE $\widehat{\Xi_T^*}(\boldsymbol{Y})$ and Oracle estimator is also reported.

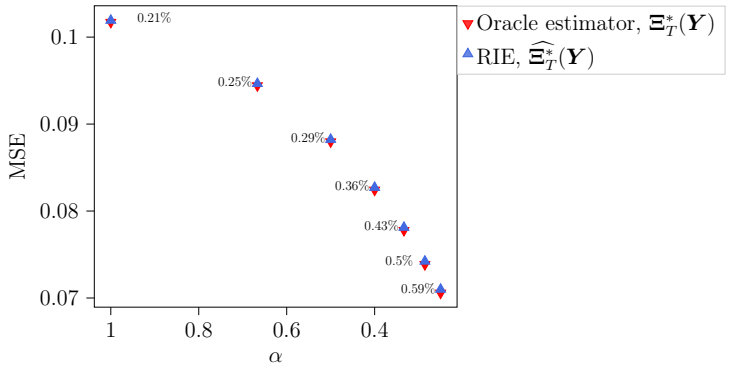
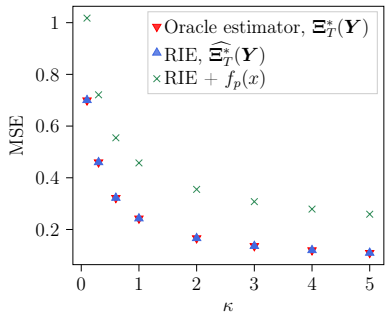
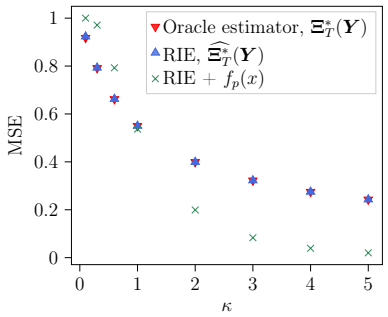


Figure 11.C.6: MSE of estimating $\boldsymbol{T}$ as a function of aspect-ratio α, $\boldsymbol{T}$ has Gaussain entries of variance $1/N$, and $\kappa = 5$. MSE is normalized by the norm of the signal, $\|\boldsymbol{T}\|_{\mathrm{F}}^2$. $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 3$, and $\boldsymbol{Z}$ has i.i.d. Gaussian entries of variance $1/N$. The RIE is applied to $N = 2000, M = 1/\alpha N$, and the results are averaged over 10 runs (error bars are invisible). Average relative error between RIE $\widehat{\Xi_T^*}(\boldsymbol{Y})$ and Oracle estimator is also reported.

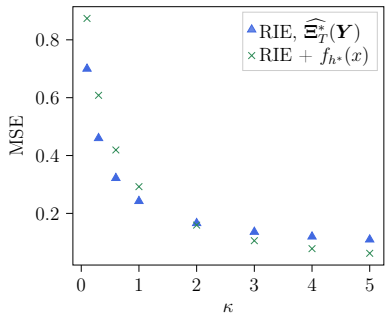


(a) $p = 0.5$

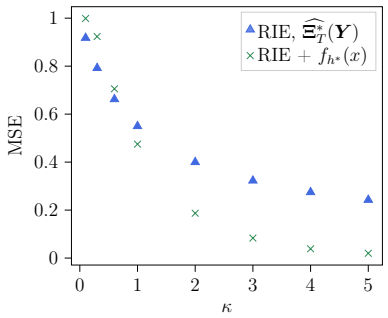


(b) $p = 0.9$

Figure 11.C.7: Estimating $\boldsymbol{T}$ with Bernoulli-Rademacher *entries*. MSE is normalized by the norm of the signal, $\|\boldsymbol{T}\|_{\mathrm{F}}^2$. $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 3$, and $\boldsymbol{Z}$ has i.i.d. Gaussian entries of variance $1/N$, and $N/M = 1/2$. The RIE is applied to $N = 2000, M = 4000$, and the results are averaged over 10 runs (error bars are invisible).



(a) $p = 0.5$



(b) $p = 0.9$

Figure 11.C.8: Estimating $\boldsymbol{T}$ with Bernoulli-Rademacher *entries*. MSE is normalized by the norm of the signal, $\|\boldsymbol{T}\|_{\mathrm{F}}^2$. $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 3$, and $\boldsymbol{Z}$ has i.i.d. Gaussian entries of variance $1/N$, and $N/M = 1/2$. The RIE is applied to $N = 2000, M = 4000$, and thresholding function is applied with the best h among $\{0, 0.1, \cdots, 1\}$. Results are averaged over 10 runs (error bars are invisible).

for non-rotationally invariant priors, although they are sub-optimal. The RIE's output can be refined, or used as a warmed-up initialization for other algorithms to get a better estimate.

In figure 11.C.8, for one experiment, the MSE is plotted for RIE and RIE+$f(x)$ with the best h among $\{0, 0.1, \cdots, 1\}$. We observe that for the particular case of Bernoulli-Rademacher prior, the thresholding stage can improve the MSE when SNR is greater than 1, however the parameter h should be chosen properly.

11.D Comparison of RIEs for Matrix Factorization and Denoising

For estimating $\boldsymbol{S}$, we have derived the estimator (11.45) for general priors ρ_S, μ_T, μ_Z. This estimator simplifies greatly, with parameters in (11.48), when both μ_T, μ_Z are Marchenko-Pastur distribution, i.e. both $\boldsymbol{T}, \boldsymbol{Z}$ having i.i.d. Gaussian entries of variance $1/N$. Similarly, although the RIE for $\boldsymbol{T}$ in (11.74) is derived for the general priors, it reduces to a rather simple estimator if ρ_S, μ_Z are taken to be shifted Wigner, and Marchenko-Pastur distribution, respectively. Therefore, in our numerical examples on factorization problem, we consider $\boldsymbol{S}$ to be a shifted Wigner matrix, and $\boldsymbol{T}, \boldsymbol{Z}$ to be Gaussian matrices.

In each experiment, the factors $\boldsymbol{S}$, $\boldsymbol{T}$ are estimated simultaneously using RIE from the observation matrix $\boldsymbol{Y}$. In addition to the MSE of estimating each factor, we also compute the MSE of estimating the product $\boldsymbol{ST}$. We compare the MSE of the product with the MSE of the oracle estimator and the RIE introduced in [54] for the denoising problem. The oracle estimator for the denoising is constructed as:

$$\boldsymbol{\Xi}^*_{ST}(\boldsymbol{Y}) = \sum_{i=1}^{N} \xi^*_{sti}\, \boldsymbol{u}_i \boldsymbol{v}_i^\intercal, \quad \xi^*_{sti} = \boldsymbol{u}_i^\intercal \boldsymbol{S}\boldsymbol{T} \boldsymbol{v}_i \tag{11.81}$$

where $\boldsymbol{u}_i, \boldsymbol{v}_i$'s are left/right singular vectors of $\boldsymbol{Y}$. In the RIE proposed in [54], the singular values are estimated by (see section 11.C.3)

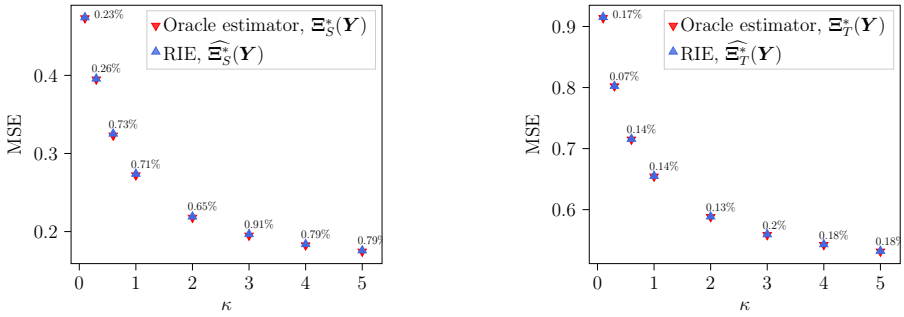
$$\widehat{\xi^*_{sti}} = \frac{1}{\sqrt{\kappa}} \Bigg[\gamma_i - \frac{1}{\pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \mathcal{C}^{(\alpha)}_{\mu_Z} \Bigg(\frac{1-\alpha}{\gamma_i} \pi \mathsf{H}[\bar{\mu}_Y](\gamma_i) + \alpha \big(\pi \mathsf{H}[\bar{\mu}_Y](\gamma_i)\big)^2 - \alpha \big(\pi \bar{\mu}_Y(\gamma_i)\big)^2 + \mathsf{i} \pi \bar{\mu}_Y(\gamma_i) \big(\frac{1-\alpha}{\gamma_i} + 2\alpha \pi \mathsf{H}[\bar{\mu}_Y](\gamma_i) \big) \Bigg) \Bigg] \tag{11.82}$$

Note that, in general the MSE of the denoising RIE $\widehat{\boldsymbol{\Xi}^*_{ST}}(\boldsymbol{Y})$, is less than the MSE of the prdouct of the estimated factors $\widehat{\boldsymbol{\Xi}^*_{S}}(\boldsymbol{Y})\widehat{\boldsymbol{\Xi}^*_{T}}(\boldsymbol{Y})$.

In figures 11.D.1,11.D.2, the MSE of estimating the factors is illustrated for $c = 1$ and $c = 3$ respectively. The MSE of estimating the product is shown in figure 11.D.3.

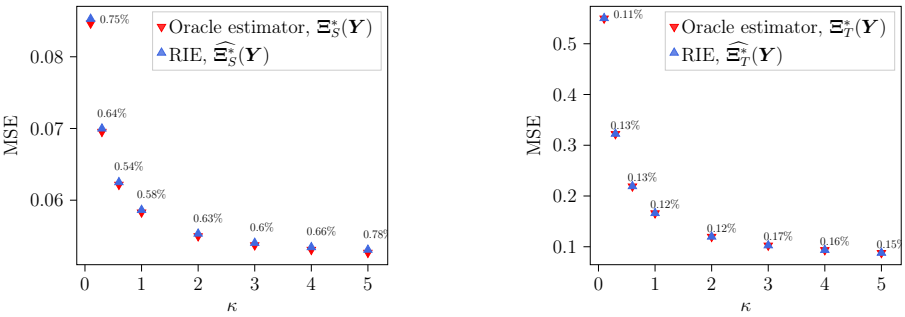
11.E Case of $\alpha \geq 1$

In this section we consider the case where $M \leq N$ and $N/M \to \alpha \geq 1$ as $N \to \infty$. Throughout this section $\boldsymbol{\Gamma} \in \mathbb{R}^{N \times M}$ is a (tall) matrix with $\boldsymbol{\Gamma}_M$ in its upper $M \times M$ block, and the rest zero entries. $\boldsymbol{\Gamma}_M$ is diagonal matrix constructed from $\boldsymbol{\gamma} \in \mathbb{R}^M$ which are the singular values of $\boldsymbol{Y}$.



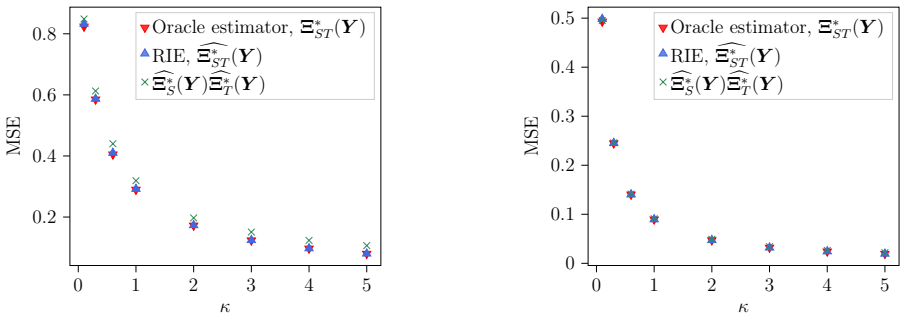
(a) Estimating $\boldsymbol{S}$ **(b)** Estimating $\boldsymbol{T}$

Figure 11.D.1: MSE of factorization problem. MSE is normalized by the norm of the signal. $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 1$, and both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$, and $N/M = 1/2$. The RIE is applied to $N = 2000, M = 4000$. In each run, the observation matrix $\boldsymbol{Y}$ is generated according to (11.1), and the factors $\boldsymbol{S}$, $\boldsymbol{T}$ are estimated simultaneously from $\boldsymbol{Y}$. Results are averaged over 10 runs (error bars are invisible). Average relative error between RIEs and Oracle estimators is also reported.



(a) Estimating $\boldsymbol{S}$ **(b)** Estimating $\boldsymbol{T}$

Figure 11.D.2: MSE of factorization problem. MSE is normalized by the norm of the signal. $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 3$, and both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$, and $N/M = 1/2$. The RIE is applied to $N = 2000, M = 4000$. In each run, the observation matrix $\boldsymbol{Y}$ is generated according to (11.1), and the factors $\boldsymbol{S}$, $\boldsymbol{T}$ are estimated simultaneously from $\boldsymbol{Y}$. Results are averaged over 10 runs (error bars are invisible). Average relative error between RIEs and Oracle estimators is also reported.



(a) $c = 1$ **(b)** $c = 3$

Figure 11.D.3: MSE of the product of the factors. MSE is normalized by the norm of the signal $\|\boldsymbol{ST}\|_{\mathrm{F}}^2$. $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 1, c = 3$, and both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$, and $N/M = 1/2$. The RIE is applied to $N = 2000, M = 4000$. Results are averaged over 10 runs (error bars are invisible).

Similar to the case of $\alpha \leq 1$, resolvent of the matrix $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{(N+M)\times(N+M)}$ plays a central role in deriving the RIEs. For the case of $M \geq N$, with $\boldsymbol{Y} = \boldsymbol{U}_Y \boldsymbol{\Gamma} \boldsymbol{V}_Y^\intercal$, the matrix $\boldsymbol{\mathcal{Y}}$ has the following eigen-decomposition:

$$\boldsymbol{\mathcal{Y}} = \begin{bmatrix} \hat{\boldsymbol{U}}_Y^{(1)} & -\hat{\boldsymbol{U}}_Y^{(1)} & \boldsymbol{U}_Y^{(2)} \\ \hat{\boldsymbol{V}}_Y & -\hat{\boldsymbol{V}}_Y & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Gamma}_M & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{\Gamma}_M & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{U}}_Y^{(1)} & -\hat{\boldsymbol{U}}_Y^{(1)} & \boldsymbol{U}_Y^{(2)} \\ \hat{\boldsymbol{V}}_Y & -\hat{\boldsymbol{V}}_Y & \boldsymbol{0} \end{bmatrix}^\intercal \tag{11.83}$$

with $\boldsymbol{U}_Y = \begin{bmatrix} \boldsymbol{U}_Y^{(1)} & \boldsymbol{U}_Y^{(2)} \end{bmatrix}$ in which $\boldsymbol{U}_Y^{(1)} \in \mathbb{R}^{N\times M}$. And, $\hat{\boldsymbol{U}}_Y^{(1)} = \frac{1}{\sqrt{2}}\boldsymbol{U}_Y^{(1)}$, $\hat{\boldsymbol{V}}_Y = \frac{1}{\sqrt{2}}\boldsymbol{V}_Y$. The resolvent of $\boldsymbol{\mathcal{Y}}$ can be written as:

$$\boldsymbol{G}_{\mathcal{Y}}(x - \mathrm{i}\epsilon) = \sum_{k=1}^{2M} \frac{x + \mathrm{i}\epsilon}{(x - \tilde{\gamma}_k)^2 + \epsilon^2} \boldsymbol{y}_k \boldsymbol{y}_k^\intercal + \frac{x + \mathrm{i}\epsilon}{x^2 + \epsilon^2} \sum_{k=2M+1}^{M+N} \boldsymbol{y}_k \boldsymbol{y}_k^\intercal$$

where $\tilde{\gamma}_k$ are the eigenvalues of $\boldsymbol{\mathcal{Y}}$, which are in fact the (signed) singular values of $\boldsymbol{Y}$, $\tilde{\gamma}_1 = \gamma_1, \ldots, \tilde{\gamma}_M = \gamma_M, \tilde{\gamma}_{M+1} = -\gamma_1, \ldots, \tilde{\gamma}_{2M} = -\gamma_M$.

11.E.1 Estimating S

The RIE for $\boldsymbol{S}$ is constructed in the same way as in the case of $\alpha \leq 1$, (11.2). However, in the present case the observation matrix $\boldsymbol{Y}$ has M (non-trivially zero) singular values and we need to estimate N eigenvalues for the RIE. As it will be clear, the $N - M$ eigenvalues are chosen to be equal.

Relation between overlap and the resolvent

Define the vectors $\tilde{\boldsymbol{s}}_i = [\boldsymbol{s}_i^\intercal, \mathbf{0}_M]^\intercal$ for $\boldsymbol{s}_i$ eigenvectors of $\boldsymbol{S}$. We have

$$\tilde{\boldsymbol{s}}_i^\intercal \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathrm{i}\epsilon)\big) \tilde{\boldsymbol{s}}_i = \sum_{k=1}^{2M} \frac{\epsilon}{(x-\tilde{\gamma}_k)^2 + \epsilon^2} \big(\tilde{\boldsymbol{s}}_i^\intercal \boldsymbol{y}_k\big)^2 + \frac{\epsilon}{x^2+\epsilon^2} \sum_{k=2M+1}^{M+N} \big(\tilde{\boldsymbol{s}}_i^\intercal \boldsymbol{y}_k\big)^2 \tag{11.84}$$

Given the structure of $\boldsymbol{y}_k$'s in (11.83), we have:

$$\big(\tilde{\boldsymbol{s}}_i^\intercal \boldsymbol{y}_k\big)^2 = \begin{cases} \frac{1}{2}\big(\boldsymbol{s}_i^\intercal \boldsymbol{u}_k\big)^2 & \text{for } 1 \leq k \leq M \\ \frac{1}{2}\big(\boldsymbol{s}_i^\intercal \boldsymbol{u}_{k-M}\big)^2 & \text{for } M+1 \leq k \leq 2M \\ \big(\boldsymbol{s}_i^\intercal \boldsymbol{u}_{k-M}\big)^2 & \text{for } 2M+1 \leq k \leq M+N \end{cases}$$

We assume that in the limit of large N this quantity concentrates on $O_S(\gamma_j, \lambda_i)$ and depends only on the singular values and eigenvalue pairs (γ_j, λ_i). This assumption implies that the singular vectors associated with 0 singular values ($\boldsymbol{u}_j$ for $M+1 \leq j \leq N$) all have the same overlap with the eigenvectors of $\boldsymbol{S}$, $O_S(0, \lambda_i)$. We thus have:

$$\begin{aligned} \tilde{\boldsymbol{s}}_i^\intercal \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathrm{i}\epsilon)\big) \tilde{\boldsymbol{s}}_i \xrightarrow{N\to\infty} \frac{1}{\alpha} \int_{\mathbb{R}} \frac{\epsilon}{(x-t)^2+\epsilon^2} O_S(t, \lambda_i) \bar{\mu}_Y(t)\, dt \\ + \big(1 - \frac{1}{\alpha}\big) \frac{\epsilon}{x^2+\epsilon^2} O_S(0, \lambda_i) \end{aligned} \tag{11.85}$$

where the overlap function $O_S(t, \lambda_i)$ is extended (continuously) to arbitrary values within the support of $\bar{\mu}_Y$ (the symmetrized limiting singular value distribution of $\boldsymbol{Y}$) with the property that $O_S(t, \lambda_i) = O_S(-t, \lambda_i)$ for $t \in \operatorname{supp}(\mu_Y)$. Sending $\epsilon \to 0$, we find

$$\tilde{\boldsymbol{s}}_i^\intercal \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathrm{i}\epsilon)\big) \tilde{\boldsymbol{s}}_i \to \frac{1}{\alpha} \pi \bar{\mu}_Y(x) O_S(x, \lambda_i) + \big(1 - \frac{1}{\alpha}\big) \pi \delta(x) O_S(x, \lambda_i) \tag{11.86}$$

Resolvent relation

We derive the resolvent relation for the same model as in (11.25). The derivation is similar to the procedure explained in section 11.B.1, and we omit here. The final resolvent relation is the same as (11.38), with parameters satisfying:

$$\begin{cases} \zeta_1^* = \frac{1}{\alpha} \frac{\mathcal{C}_{\mu_Z}^{(1/\alpha)}(p_1^* p_2^*)}{p_1^*}, \quad \zeta_2^* = \frac{1}{p_2^*}\big(\mathcal{C}_{\mu_Z}^{(1/\alpha)}(p_1^* p_2^*) + \mathcal{C}_{\mu_T}^{(1/\alpha)}(p_2^* p_3^*)\big), \quad \zeta_3^* = \frac{1}{\alpha} \frac{\mathcal{C}_{\mu_T}^{(1/\alpha)}(p_2^* p_3^*)}{p_3^*} \\ p_1^* = \frac{1}{\zeta_3^*} \mathcal{G}_{\rho_{S^2}}\big(\frac{z-\zeta_1^*}{\zeta_3^*}\big), \quad p_2^* = \frac{1}{z-\zeta_2^*}, \quad p_3^* = \frac{z-\zeta_1^*}{\zeta_3^{*2}} \mathcal{G}_{\rho_{S^2}}\big(\frac{z-\zeta_1^*}{\zeta_3^*}\big) - \frac{1}{\zeta_3^*} \end{cases} \tag{11.87}$$

Again, with the same procedure as (11.39),(11.40), the saddle point equations (11.87) can be rewritten in a simplified form, which does not involve ρ_{S^2}, as:

$$\begin{cases} \zeta_1^* = \frac{1}{\alpha} \frac{\mathcal{C}_{\mu_Z}^{(1/\alpha)}(p_1^* p_2^*)}{p_1^*}, \quad \zeta_2^* = z - \frac{1}{\mathcal{G}_{\bar{\mu}_Y}(z)}, \quad \zeta_3^* = \frac{1}{\alpha} \frac{\mathcal{C}_{\mu_T}^{(1/\alpha)}(p_2^* p_3^*)}{p_3^*} \\ p_1^* = \frac{1}{\alpha} \mathcal{G}_{\bar{\mu}_Y}(z) + \big(1 - \frac{1}{\alpha}\big) \frac{1}{z}, \quad p_2^* = \mathcal{G}_{\bar{\mu}_Y}(z) \\ p_3^* = \frac{z-\zeta_1^*}{\alpha \zeta_3^*} \mathcal{G}_{\bar{\mu}_Y}(z) + \frac{z-\zeta_1^*}{\zeta_3^*} \big(1 - \frac{1}{\alpha}\big) \frac{1}{z} - \frac{1}{\zeta_3^*} \end{cases} \tag{11.88}$$

with $\bar{\mu}_Y$ the limiting ESD of non-trivial singular values of $\boldsymbol{Y}$. Note that ζ_1^*, ζ_2^* can be computed from the observation matrix, and we only need to find ζ_3^* satisfying the following equation:

$$\begin{aligned}(z-\zeta_1^*)\big[\frac{1}{\alpha}\mathcal{G}_{\bar{\mu}_Y}(z)+\big(1-\frac{1}{\alpha}\big)\frac{1}{z}\big]-1 \\ = \frac{1}{\alpha}\mathcal{C}_{\mu_T}^{(1/\alpha)}\Big(\frac{1}{\zeta_3^*}\mathcal{G}_{\bar{\mu}_Y}(z)(z-\zeta_1^*)\big[\frac{1}{\alpha}\mathcal{G}_{\bar{\mu}_Y}(z)+\big(1-\frac{1}{\alpha}\big)\frac{1}{z}\big]\Big)\end{aligned} \tag{11.89}$$

Note that both sets of equations (11.86), (11.88) and (11.43), (11.41) match for $\alpha = 1$.

Overlaps and optimal eigenvalues

From (11.86), (11.38), for γ a non-trivially zero singular value of $\boldsymbol{Y}$ we find:

$$\begin{aligned}O_S(\gamma,\lambda_i) &\approx \frac{\alpha}{\pi\bar{\mu}_Y(\gamma)}\operatorname{Im}\lim_{z\to\gamma-\mathrm{i}0^+}\boldsymbol{s}_i^\intercal\,\zeta_3^{*-1}\boldsymbol{G}_{S^2}\big(\frac{z-\zeta_1^*}{\zeta_3^*}\big)\,\boldsymbol{s}_i \\ &= \frac{\alpha}{\pi\bar{\mu}_Y(\gamma)}\operatorname{Im}\lim_{z\to\gamma-\mathrm{i}0^+}\frac{1}{z-\zeta_1^*-\zeta_3^*\lambda_i^2}\end{aligned} \tag{11.90}$$

And, in the case of $M > N$, for zero singular values we have:

$$\begin{aligned}O_S(0,\lambda_i) &\approx \frac{\alpha}{(\alpha-1)\pi}\operatorname{Im}\lim_{z\to-\mathrm{i}0^+}\boldsymbol{s}_i^\intercal\,\zeta_3^{*-1}\boldsymbol{G}_{S^2}\big(\frac{z-\zeta_1^*}{\zeta_3^*}\big)\,\boldsymbol{s}_i \\ &= \frac{\alpha}{(\alpha-1)\pi}\operatorname{Im}\lim_{z\to-\mathrm{i}0^+}\frac{1}{z-\zeta_1^*-\zeta_3^*\lambda_i^2}\end{aligned} \tag{11.91}$$

Finally, the optimal eigenvalues can be derived in the same way as in (11.45). For $1 \le i \le M$, we have:

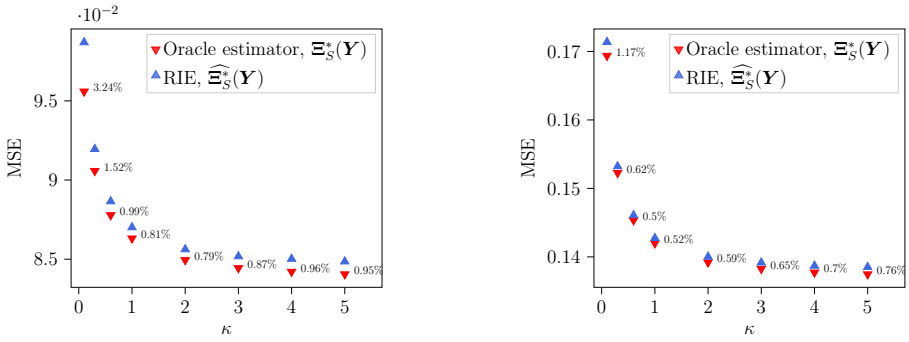
$$\widehat{\xi}_{s\,i}^* = \frac{\alpha}{2\kappa\pi\bar{\mu}_Y(\gamma_i)}\operatorname{Im}\lim_{z\to\gamma_i-\mathrm{i}0^+}\Big\{\frac{1}{\zeta_3^*}\Big[\mathcal{G}_{\rho_S}\Big(\sqrt{\frac{z-\zeta_1^*}{\kappa\zeta_3^*}}\Big)+\mathcal{G}_{\rho_S}\Big(-\sqrt{\frac{z-\zeta_1^*}{\kappa\zeta_3^*}}\Big)\Big]\Big\} \tag{11.92}$$

And, for all $M+1 \le i \le N$:

$$\widehat{\xi}_{s\,i}^* = \frac{\alpha}{2\kappa(\alpha-1)\pi}\operatorname{Im}\lim_{z\to-\mathrm{i}0^+}\Big\{\frac{1}{\zeta_3^*}\Big[\mathcal{G}_{\rho_S}\Big(\sqrt{\frac{z-\zeta_1^*}{\kappa\zeta_3^*}}\Big)+\mathcal{G}_{\rho_S}\Big(-\sqrt{\frac{z-\zeta_1^*}{\kappa\zeta_3^*}}\Big)\Big]\Big\} \tag{11.93}$$

Numerical examples

For matrices $\boldsymbol{T}, \boldsymbol{Z} \in \mathbb{R}^{N\times M}$ with i.i.d. Gaussian entries of variance $1/N$ and $M > N$, we have that $\mathcal{C}_{\mu_T}^{(1/\alpha)}(z) = \mathcal{C}_{\mu_Z}^{(1/\alpha)}(z) = z$ which leads to a simplification



(a) Shifted Wigner, $c = 3$ **(b)** Wishart, $\alpha' = 1/4$

Figure 11.E.1: Estimating $\boldsymbol{S}$. The MSE is normalized by the norm of the signal, $\|\boldsymbol{S}\|_{\mathrm{F}}^2$. Both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$, and aspect ratio $N/M = 2$. The RIE is applied to $N = 2000, M = 1000$, and the results are averaged over 10 runs (error bars are invisible). Average relative error between RIE $\widehat{\boldsymbol{\Xi}_S^*}(\boldsymbol{Y})$ and Oracle estimator is also reported.

of equations (11.88):

$$\begin{cases} \zeta_1^* = \frac{1}{\alpha} p_2^*, \quad \zeta_2^* = z - \frac{1}{\mathcal{G}_{\bar{\mu}_Y}(z)}, \quad \zeta_3^* = \frac{1}{\alpha} p_2^* \\ p_1^* = \frac{1}{\alpha} \mathcal{G}_{\bar{\mu}_Y}(z) + \left(1 - \frac{1}{\alpha}\right) \frac{1}{z}, \quad p_2^* = \mathcal{G}_{\bar{\mu}_Y}(z) \\ p_3^* = \frac{z - \zeta_1^*}{\alpha \zeta_3^*} \mathcal{G}_{\bar{\mu}_Y}(z) + \frac{z - \zeta_1^*}{\zeta_3^*} \left(1 - \frac{1}{\alpha}\right) \frac{1}{z} - \frac{1}{\zeta_3^*} \end{cases} \tag{11.94}$$

Therefore, $\zeta_1^* = \zeta_3^* = \frac{1}{\alpha} \mathcal{G}_{\bar{\mu}_Y}(z)$.

In Figure 11.E.1, the MSE of the Oracle estimator and the RIE (11.92), (11.93) is illustrated for shifted Wigner $\boldsymbol{S}$ with $c = 3$, and Wishart with aspect-ratio $\alpha' = 1/4$.

Effect of aspect-ratio α. In Figure 11.E.2, we take $\boldsymbol{S}$ to be a shifted Wigner matrix with $c = 3$, and the MSE is depicted for various values of the aspect-ratio $\alpha > 1$. We see that as M decreases (α increases) the estimation error (of $\boldsymbol{T}$) increases.

11.E.2 Estimating $\boldsymbol{T}$

Relation between overlap and the resolvent

For the vectors $\boldsymbol{r}_i = \begin{bmatrix} \boldsymbol{0}_N \\ \boldsymbol{t}_i^{(r)} \end{bmatrix}$, $\boldsymbol{l}_i = \begin{bmatrix} \boldsymbol{t}_i^{(l)} \\ \boldsymbol{0}_M \end{bmatrix}$ with $\boldsymbol{t}_i^{(r)}, \boldsymbol{t}_i^{(l)}$ right/ left singular vectors of $\boldsymbol{T}$, we have

$$\begin{aligned} \boldsymbol{r}_i^\intercal \left(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathrm{i}\epsilon)\right) \boldsymbol{l}_i = & \sum_{k=1}^{2M} \frac{\epsilon}{(x - \tilde{\gamma}_k)^2 + \epsilon^2} \left(\boldsymbol{r}_i^\intercal \boldsymbol{y}_k\right) \left(\boldsymbol{l}_i^\intercal \boldsymbol{y}_k\right) \\ & + \frac{\epsilon}{x^2 + \epsilon^2} \sum_{k=2M+1}^{M+N} \left(\boldsymbol{r}_i^\intercal \boldsymbol{y}_k\right) \left(\boldsymbol{l}_i^\intercal \boldsymbol{y}_k\right) \end{aligned} \tag{11.95}$$

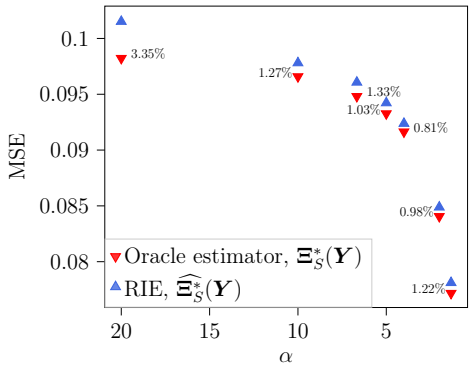


Figure 11.E.2: MSE of estimating $\boldsymbol{S}$ as a function of aspect-ratio $\alpha > 1$, prior on $\boldsymbol{S}$ is shifted Wigner with $c = 3$, and $\kappa = 5$. MSE is normalized by the norm of the signal, $\|\boldsymbol{S}\|_{\mathrm{F}}^2$. Both $\boldsymbol{T}$ and $\boldsymbol{Z}$ are $N \times M$ matrices with i.i.d. Gaussian entries of variance $1/N$. The RIE is applied to $N = 2000, M = 1/\alpha N$, and the results are averaged over 10 runs (error bars are invisible). Average relative error between RIE $\widehat{\Xi}_S^*(\boldsymbol{Y})$ and Oracle estimator is also reported.

Given the structure of $\boldsymbol{y}_k$'s in (11.83), we have:

$$\big(\boldsymbol{r}_i^\intercal \boldsymbol{y}_k\big)\big(\boldsymbol{l}_i^\intercal \boldsymbol{y}_k\big) = \begin{cases} \frac{1}{2}\big(\boldsymbol{u}_k^\intercal \boldsymbol{t}_i^{(l)}\big)\big(\boldsymbol{v}_k^\intercal \boldsymbol{t}_i^{(r)}\big) & \text{for } 1 \le k \le M \\ -\frac{1}{2}\big(\boldsymbol{u}_{k-M}^\intercal \boldsymbol{t}_i^{(l)}\big)\big(\boldsymbol{v}_{k-M}^\intercal \boldsymbol{t}_i^{(r)}\big) & \text{for } M+1 \le k \le 2M \\ 0 & \text{for } 2M+1 \le k \le N+M \end{cases}$$

Therefore, in the limit $N \to \infty$, we have:

$$\boldsymbol{r}_i^\intercal \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathrm{i}\epsilon)\big)\boldsymbol{l}_i \xrightarrow{N\to\infty} \frac{1}{\alpha}\int_{\mathbb{R}} \frac{\epsilon}{(x-t)^2+\epsilon^2} O_T(t, \sigma_i)\bar{\mu}_Y(t)\, dt \tag{11.96}$$

where the overlap function $O_T(t, \lambda_i)$ is extended (continuously) to arbitrary values within the support of $\bar{\mu}_Y$ with the property that $O_T(-t, \lambda_i) = -O_T(t, \lambda_i)$ for $t \in \operatorname{supp}(\mu_Y)$. Sending $\epsilon \to 0$, we find

$$\boldsymbol{r}_i^\intercal \big(\operatorname{Im} \boldsymbol{G}_{\mathcal{Y}}(x - \mathrm{i}\epsilon)\big)\boldsymbol{l}_i \approx \frac{1}{\alpha}\pi\bar{\mu}_Y(x) O_T(x, \sigma_i) \tag{11.97}$$

Resolvent relation

The resolvent relation for the model (11.56) with $M < N$ is the same as in (11.68) with parameters satisfying:

$$\begin{cases} \beta_1^* = \frac{1}{\alpha}\frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(q_1^* q_2^*)}{q_1^*} + \frac{1}{2}\sqrt{\frac{q_3^*}{q_1^*}}\Big(\mathcal{R}_{\rho_S}\big(q_4^* + \sqrt{q_1^* q_3^*}\big) - \mathcal{R}_{\rho_S}\big(q_4^* - \sqrt{q_1^* q_3^*}\big)\Big) \\ \beta_2^* = \frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(q_1^* q_2^*)}{q_2^*} \\ \beta_3^* = \frac{1}{2}\sqrt{\frac{q_1^*}{q_3^*}}\Big(\mathcal{R}_{\rho_S}\big(q_4^* + \sqrt{q_1^* q_3^*}\big) - \mathcal{R}_{\rho_S}\big(q_4^* - \sqrt{q_1^* q_3^*}\big)\Big) \\ \beta_4^* = \frac{1}{2}\Big(\mathcal{R}_{\rho_S}\big(q_4^* + \sqrt{q_1^* q_3^*}\big) + \mathcal{R}_{\rho_S}\big(q_4^* - \sqrt{q_1^* q_3^*}\big)\Big) \\ q_1^* = \frac{1}{\alpha}\frac{(z-\beta_2^*)\beta_4^{*2}}{Z_2(z)^2}\mathcal{G}_{\rho_T}\big(\frac{Z_1(z)}{Z_2(z)}\big) + \frac{1}{\alpha}\frac{\beta_3^*}{Z_2(z)} + \frac{\alpha-1}{\alpha}\frac{1}{z-\beta_1^*} \\ q_2^* = \frac{z-\beta_1^*}{Z_2(z)}\mathcal{G}_{\rho_T}\big(\frac{Z_1(z)}{Z_2(z)}\big) \\ q_3^* = \frac{1}{\alpha}\frac{(z-\beta_1^*)Z_1(z)}{Z_2(z)^2}\mathcal{G}_{\rho_T}\big(\frac{Z_1(z)}{Z_2(z)}\big) - \frac{1}{\alpha}\frac{z-\beta_1^*}{Z_2(z)} \\ q_4^* = \frac{1}{\alpha}\frac{\beta_4^* Z_1(z)}{Z_2(z)^2}\mathcal{G}_{\rho_T}\big(\frac{Z_1(z)}{Z_2(z)}\big) - \frac{1}{\alpha}\frac{\beta_4^*}{Z_2(z)} \end{cases} \tag{11.98}$$

with

$$\begin{cases} Z_1(z) = (z-\beta_1^*)(z-\beta_2^*) \\ Z_2(z) = \beta_4^{*2} + \beta_3^*(z-\beta_1^*) \end{cases}$$

With the same procedure as (11.69),(11.70), the saddle point equations (11.98) can be rewritten in a simplified form:

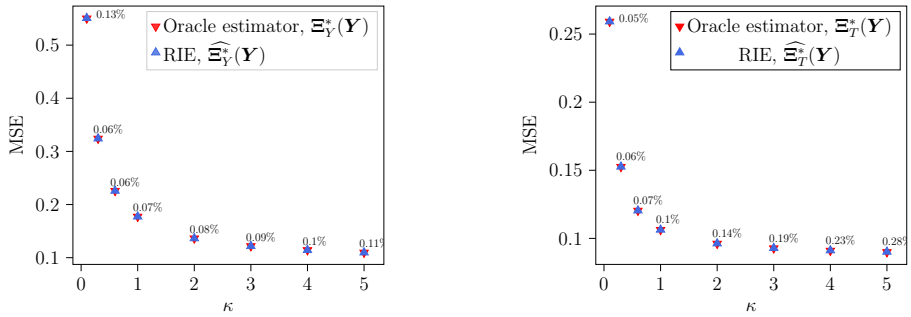
$$\begin{cases} \beta_1^* = \frac{1}{\alpha}\frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(q_1^* q_2^*)}{q_1^*} + \frac{1}{2}\sqrt{\frac{q_3^*}{q_1^*}}\Big(\mathcal{R}_{\rho_S}\big(q_4^* + \sqrt{q_1^* q_3^*}\big) - \mathcal{R}_{\rho_S}\big(q_4^* - \sqrt{q_1^* q_3^*}\big)\Big) \\ \beta_2^* = \frac{\mathcal{C}_{\mu_Z}^{(\alpha)}(q_1^* q_2^*)}{q_2^*} \\ \beta_3^* = \frac{1}{2}\sqrt{\frac{q_1^*}{q_3^*}}\Big(\mathcal{R}_{\rho_S}\big(q_4^* + \sqrt{q_1^* q_3^*}\big) - \mathcal{R}_{\rho_S}\big(q_4^* - \sqrt{q_1^* q_3^*}\big)\Big) \\ \beta_4^* = \frac{1}{2}\Big(\mathcal{R}_{\rho_S}\big(q_4^* + \sqrt{q_1^* q_3^*}\big) + \mathcal{R}_{\rho_S}\big(q_4^* - \sqrt{q_1^* q_3^*}\big)\Big) \\ q_1^* = \frac{1}{\alpha}\mathcal{G}_{\bar{\mu}_Y}(z) + \big(1-\frac{1}{\alpha}\big)\frac{1}{z} \\ q_2^* = \mathcal{G}_{\bar{\mu}_Y}(z) \\ q_3^* = \frac{(z-\beta_1^*)^2}{\beta_4^{*2}}q_1^* - \frac{z-\beta_1^*}{\beta_4^{*2}} \\ q_4^* = \frac{z-\beta_1^*}{\beta_4^*}q_1^* - \frac{1}{\beta_4^*} \end{cases} \tag{11.99}$$

Note that both sets of equations (11.97), (11.99) and (11.55), (11.72) match for $\alpha = 1$.

Overlaps and optimal singular values

From (11.68), (11.97), we have:

$$\begin{aligned} O_T(\gamma, \sigma_i) &\approx \frac{\alpha}{\pi\bar{\mu}_Y(\gamma)}\operatorname{Im}\lim_{z\to\gamma-\mathrm{i}0^+}\frac{\beta_4^*}{Z_2(z)}\boldsymbol{t}_i^{(r)\intercal}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big)\boldsymbol{T}^\intercal\boldsymbol{t}_i^{(l)} \\ &= \frac{\alpha}{\pi\bar{\mu}_Y(\gamma)}\operatorname{Im}\lim_{z\to\gamma-\mathrm{i}0^+}\beta_4^*\frac{\sigma_i}{Z_1(z) - Z_2(z)\sigma_i^2} \end{aligned} \tag{11.100}$$



(a) Gaussian **(b)** Uniform spectral, $\mathcal{U}\big([1,3]\big)$

Figure 11.E.3: Estimating $\boldsymbol{T}$. MSE is normalized by the norm of the signal, $\|\boldsymbol{T}\|_{\mathrm{F}}^2$. $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 3$, and $\boldsymbol{Z}$ has i.i.d. Gaussian entries of variance $1/N$, and $N/M = 2$. The RIE is applied to $N = 2000, M = 1000$, and the results are averaged over 10 runs (error bars are invisible).

Similar to (11.74), we can compute the optimal singular values to be:

$$\widehat{\xi}_{t\,i}^* = \frac{\alpha}{\pi \bar{\mu}_Y(\gamma_i)} \operatorname{Im} \lim_{z \to \gamma_i - \mathrm{i}0^+} q_4^* \tag{11.101}$$

Numerical examples

We consider the matrix $\boldsymbol{Z}$ to have i.i.d. Gaussian entries with variance $1/N$, so $\mathcal{C}_{\mu_Z}^{(1/\alpha)}(z) = z$. And, $\boldsymbol{S} = \boldsymbol{F} + c\boldsymbol{I}$ where $\boldsymbol{F} = \boldsymbol{F}^\intercal \in \mathbb{R}^{N\times N}$ has i.i.d. entries with variance $1/N$, and $c \neq 0$ is a real number, so $\mathcal{R}_{\rho_S}(z) = z + c$. With these choices, the solution (11.99) simplifies to:

$$\begin{cases} \beta_1^* = \frac{1}{\alpha} q_2^* + q_3^*, \quad \beta_2^* = q_1^*, \quad \beta_3^* = q_1^*, \quad \beta_4^* = q_4^* + c \\ q_1^* = \frac{1}{\alpha}\mathcal{G}_{\bar{\mu}_Y}(z) + \big(1 - \frac{1}{\alpha}\big)\frac{1}{z}, \quad q_2^* = \mathcal{G}_{\bar{\mu}_Y}(z) \\ q_3^* = \frac{(z-\beta_1^*)^2}{\beta_4^{*2}} q_1^* - \frac{z-\beta_1^*}{\beta_4^{*2}}, \quad q_4^* = \frac{z-\beta_1^*}{\beta_4^*} q_1^* - \frac{1}{\beta_4^*} \end{cases} \tag{11.102}$$

After a bit of algebra, we find that q_4^* is the solution to the following qubic equation:

$$\begin{aligned} 2x^3 + 3c\,x^2 + \Big[c^2 + 2 - \big(z - \frac{1}{\alpha}\mathcal{G}_{\bar{\mu}_Y}(z)\big)\big(\frac{1}{\alpha}\mathcal{G}_{\bar{\mu}_Y}(z) + \frac{\alpha-1}{\alpha z}\big)\Big]\, x \\ - c\Big[\big(z - \frac{1}{\alpha}\mathcal{G}_{\bar{\mu}_Y}(z)\big)\big(\frac{1}{\alpha}\mathcal{G}_{\bar{\mu}_Y}(z) + \frac{\alpha-1}{\alpha z}\big) - 1\Big] = 0 \end{aligned} \tag{11.103}$$

In figure 11.E.3 the MSE of RIE and the oracle estimator is plotted for two cases of priors: $\boldsymbol{T}$ with Gaussian entries and $\boldsymbol{T}$ with uniform spectral density.

Effect of aspect-ratio α. In Figure 11.E.4, we take $\boldsymbol{T}$ to have Gaussian entries (with variance $\frac{1}{N}$), and the MSE is depicted for various values of the aspect-ratio $\alpha > 1$. We see that as M decreases (α increases) the estimation error (of $\boldsymbol{T}$) increases.

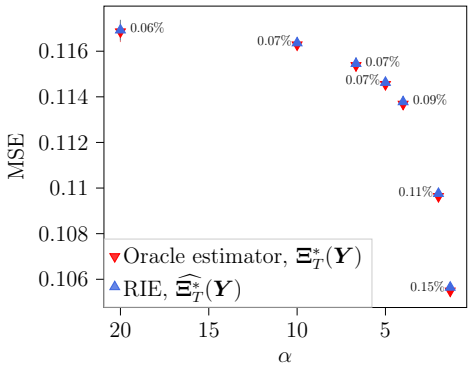


Figure 11.E.4: MSE of estimating $\boldsymbol{T}$ as a function of aspect-ratio $\alpha > 1$, $\boldsymbol{T}$ has Gaussain entries of variance $1/N$, and $\kappa = 5$. MSE is normalized by the norm of the signal, $\|\boldsymbol{T}\|_{\mathrm{F}}^2$. $\boldsymbol{S}$ is a shifted Wigner matrix with $c = 3$, and $\boldsymbol{Z}$ has i.i.d. Gaussian entries of variance $1/N$. The RIE is applied to $N = 2000, M = 1/\alpha N$, and the results are averaged over 10 runs (error bars are invisible). Average relative error between RIE $\widehat{\boldsymbol{\Xi}_T^*}(\boldsymbol{Y})$ and Oracle estimator is also reported.

11.F Auxiliary Lemmas and Calculations

Proposition 11.1 (**Inverse of a block matrix**, Bernstein [136])**.** *For a block matrix* $\boldsymbol{F} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{bmatrix}$ *with* $\boldsymbol{A} \in \mathbb{R}^{N\times N}, \boldsymbol{B} \in \mathbb{R}^{N\times M}, \boldsymbol{C} \in \mathbb{R}^{M\times N}, \boldsymbol{D} \in \mathbb{R}^{M\times M}$, *if* $\boldsymbol{A}$ *and* $\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}$, *are non-singular, then,*

$$\boldsymbol{F}^{-1} = \begin{bmatrix} \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & -\boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \\ -(\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & (\boldsymbol{D} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} \end{bmatrix}$$

Block structure of $\boldsymbol{G}_{\mathcal{Y}}(z)$ The matrix $\boldsymbol{G}_{\mathcal{Y}}(z)$ is:

$$\boldsymbol{G}_{\mathcal{Y}}(z) = \big(z\boldsymbol{I} - \boldsymbol{\mathcal{Y}}\big)^{-1} = \begin{bmatrix} z\boldsymbol{I}_N & -\boldsymbol{Y} \\ -\boldsymbol{Y}^\intercal & z\boldsymbol{I}_M \end{bmatrix}^{-1}$$

Using Proposition 11.1, first we need to compute the inverse matrix $\big(z\boldsymbol{I}_M - (-\boldsymbol{S}^\intercal)(z\boldsymbol{I}_N)^{-1}(-\boldsymbol{S})\big)^{-1}$ which simply reads:

$$\big(z\boldsymbol{I}_M - \frac{1}{z}\boldsymbol{Y}^\intercal\boldsymbol{Y}\big)^{-1} = z\big(z^2\boldsymbol{I}_M - \boldsymbol{Y}^\intercal\boldsymbol{Y}\big)^{-1} = z\boldsymbol{G}_{\boldsymbol{Y}^\intercal\boldsymbol{Y}}(z^2)$$

Consequently, we find:

$$\boldsymbol{G}_{\mathcal{Y}}(z) = \begin{bmatrix} \frac{1}{z}\boldsymbol{I}_N + \frac{1}{z}\boldsymbol{Y}\boldsymbol{G}_{\boldsymbol{Y}^\intercal\boldsymbol{Y}}(z^2)\boldsymbol{Y}^\intercal & \boldsymbol{Y}\boldsymbol{G}_{\boldsymbol{Y}^\intercal\boldsymbol{Y}}(z^2) \\ \boldsymbol{G}_{\boldsymbol{Y}^\intercal\boldsymbol{Y}}(z^2)\boldsymbol{Y}^\intercal & z\boldsymbol{G}_{\boldsymbol{Y}^\intercal\boldsymbol{Y}}(z^2) \end{bmatrix} \tag{11.104}$$

Inverse of $\boldsymbol{C}_S^*$ For $\boldsymbol{C}_S^*$ since the blocks $\boldsymbol{B}, \boldsymbol{C}$ are zero, the inverse is simply:

$$\begin{aligned}
\boldsymbol{C}_X^{*\,-1} &= \begin{bmatrix} \left[(z-\zeta_1^*)\boldsymbol{I}_N - \zeta_3^*\boldsymbol{S}^2\right]^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \left[(z-\zeta_2^*)\boldsymbol{I}_M\right]^{-1} \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\zeta_3^*}\left[\frac{z-\zeta_1^*}{\zeta_3^*}\boldsymbol{I}_N - \boldsymbol{S}^2\right]^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{z-\zeta_2^*}\boldsymbol{I}_M \end{bmatrix} \\
&= \begin{bmatrix} \frac{1}{\zeta_3^*}\boldsymbol{G}_{S^2}\left(\frac{z-\zeta_1^*}{\zeta_3^*}\right) & \boldsymbol{0} \\ \boldsymbol{0} & \frac{1}{z-\zeta_2^*}\boldsymbol{I}_M \end{bmatrix}
\end{aligned} \tag{11.105}$$

Inverse of $\boldsymbol{C}_T^*$ Let the block structure of $\boldsymbol{C}_T^*$ be as in Proposition 11.1, then

$$\begin{aligned}
(\boldsymbol{D}-\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B})^{-1} &= \Big((z-\beta_2^*)\boldsymbol{I}_M - \beta_3^*\boldsymbol{T}^\intercal\boldsymbol{T} - \frac{\beta_4^{*2}}{z-\beta_1^*}\boldsymbol{T}^\intercal\boldsymbol{T}\Big)^{-1} \\
&= \Big((z-\beta_2^*)\boldsymbol{I}_M - \big(\beta_3^* + \frac{\beta_4^{*2}}{z-\beta_1^*}\big)\boldsymbol{T}^\intercal\boldsymbol{T}\Big)^{-1} \\
&= (z-\beta_1^*)\Big(Z_1(z)\boldsymbol{I}_M - Z_2(z)\boldsymbol{T}^\intercal\boldsymbol{T}\Big)^{-1} \\
&= \frac{z-\beta_1^*}{Z_2(z)}\Big(\frac{Z_1(z)}{Z_2(z)}\boldsymbol{I}_M - \boldsymbol{T}^\intercal\boldsymbol{T}\Big)^{-1} \\
&= \frac{z-\beta_1^*}{Z_2(z)}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big)
\end{aligned}$$

where $\boldsymbol{G}_{T^\intercal T}$ is the resolvent of the matrix $\boldsymbol{T}^\intercal\boldsymbol{T}$. So, we have

$$\boldsymbol{C}_T^{*\,-1} = \begin{bmatrix} (z-\beta_1^*)^{-1}\boldsymbol{I}_N + \frac{\beta_4^{*2}}{(z-\beta_1^*)Z_2(z)}\boldsymbol{T}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big)\boldsymbol{T}^\intercal & \frac{\beta_4^*}{Z_2(z)}\boldsymbol{T}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big) \\ \frac{\beta_4^*}{Z_2(z)}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big)\boldsymbol{T}^\intercal & \frac{z-\beta_1^*}{Z_2(z)}\boldsymbol{G}_{T^\intercal T}\big(\frac{Z_1(z)}{Z_2(z)}\big) \end{bmatrix}$$

Lemma 11.3. *Consider two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^N$. The symmetric matrix $\boldsymbol{x}\boldsymbol{y}^\intercal + \boldsymbol{y}\boldsymbol{x}^\intercal$ has rank at most two with non-zero eigenvalues $\boldsymbol{x}^\intercal\boldsymbol{y} \pm \|\boldsymbol{x}\|\|\boldsymbol{y}\|$.*

Proof. Construct the matrices $\boldsymbol{A} \in \mathbb{R}^{2\times N}, \boldsymbol{B} \in \mathbb{R}^{N\times 2}$ as follows:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{x}^\intercal \\ \boldsymbol{y}^\intercal \end{bmatrix}, \quad \boldsymbol{B} = \begin{bmatrix} \boldsymbol{y} & \boldsymbol{x} \end{bmatrix}$$

Then, we have that $\boldsymbol{x}\boldsymbol{y}^\intercal + \boldsymbol{y}\boldsymbol{x}^\intercal = \boldsymbol{B}\boldsymbol{A}$. Using the lemma 11.4, we have that:

$$z^2 \det\left(z\boldsymbol{I}_N - \boldsymbol{B}\boldsymbol{A}\right) = z^N \det\left(z\boldsymbol{I}_2 - \boldsymbol{A}\boldsymbol{B}\right)$$

So, the characteristic polynomial of $\boldsymbol{x}\boldsymbol{y}^\intercal + \boldsymbol{y}\boldsymbol{x}^\intercal$ is $z^{N-2}\det\left(z\boldsymbol{I}_2 - \boldsymbol{A}\boldsymbol{B}\right)$, which implies that the $\boldsymbol{x}\boldsymbol{y}^\intercal + \boldsymbol{y}\boldsymbol{x}^\intercal$ has eigenvalue 0 with multiplicity $N-2$, plus the eigenvalues of the 2×2 matrix $\boldsymbol{A}\boldsymbol{B}$. The matrix $\boldsymbol{A}\boldsymbol{B}$ is:

$$\boldsymbol{A}\boldsymbol{B} = \begin{bmatrix} \boldsymbol{x}^\intercal\boldsymbol{y} & \|\boldsymbol{x}\|^2 \\ \|\boldsymbol{y}\|^2 & \boldsymbol{x}^\intercal\boldsymbol{y} \end{bmatrix}$$

which has two eigenvalues $\boldsymbol{x}^\intercal\boldsymbol{y} \pm \|\boldsymbol{x}\|\|\boldsymbol{y}\|$. □

Lemma 11.4. *For matrices* $\boldsymbol{A} \in \mathbb{R}^{M\times N}, \boldsymbol{B} \in \mathbb{R}^{N\times M}$*, we have:*

$$z^M \det\big(z\boldsymbol{I}_N - \boldsymbol{B}\boldsymbol{A}\big) = z^N \det\big(z\boldsymbol{I}_M - \boldsymbol{A}\boldsymbol{B}\big)$$

Proof. Construct the matrices $\boldsymbol{C}, \boldsymbol{D} \in \mathbb{R}^{(M+N)\times(M+N)}$ as follows:

$$\boldsymbol{C} = \begin{bmatrix} z\boldsymbol{I}_M & \boldsymbol{A} \\ \boldsymbol{B} & \boldsymbol{I}_N \end{bmatrix}, \quad \boldsymbol{D} = \begin{bmatrix} \boldsymbol{I}_M & \boldsymbol{0}_{M\times N} \\ -\boldsymbol{B} & z\boldsymbol{I}_N \end{bmatrix}$$

We have:

$$\det \boldsymbol{C}\boldsymbol{D} = z^N \det\big(z\boldsymbol{I}_M - \boldsymbol{A}\boldsymbol{B}\big), \quad \det \boldsymbol{D}\boldsymbol{C} = z^M \det\big(z\boldsymbol{I}_N - \boldsymbol{B}\boldsymbol{A}\big)$$

The result follows from the fact that $\det \boldsymbol{C}\boldsymbol{D} = \det \boldsymbol{D}\boldsymbol{C}$. □

12 Conclusion and Future Directions

In this thesis, we leveraged diverse tools from random matrix theory, statistical physics and information theory alongside results on high-dimensional limits of spherical integrals to study high-dimensional matrix inference problems. In part I, we investigated mismatched estimation of finite-rank signal matrices corrupted by Gaussian noise both in symmetric and non-symmetric scenarios. We derive the asymptotic MSE of estimation assuming incorrect Gaussian prior, and we compare performance of the spectral algorithms and AMP. In part II, we explored the matrix inference problems in growing-rank regimes through the lens of rotation invariant estimators (RIE). In the symmetric case, using the optimality of RIE we deduce the asymptotic MMSE of denoising problem under Gaussian noise with rotational invariant priors. In the non-symmetric case, we derive explicit optimal RIE for the denoising problem, and consequently we computed the asymptotic MMSE under Gaussian noise. Finally, we explored a solvable model of the matrix factorization problem in the extensive-rank regime and derived analytical formulas for the optimal RIEs to reconstruct the two matrix factors.

In inference problems, the assumption of perfectly known priors (on both signal and the noise) is often unrealistic, highlighting the significance of mismatched estimation in practical scenarios. In chapters 7 and 8, we considered a simple model in this scenario where the statistician assumes Gaussian priors. However, considering models (in both low and extensive-rank regimes) where the estimation is conducted with more general priors (which are possibly incorrect) is an interesting research problem with practical importance.

The problems we studied in this thesis involved scenarios with proportional aspect ratios, which align with conventional settings in high-dimensional matrix inference. However, in many practical cases the number of samples is much larger than the dimension or vice versa, for example in genomics the number of

sequences (subjects) is a few thousands, while the number of genes can be of the order of hundreds of thousands. This necessitates a shift in focus towards diverging aspect ratios, which requires developing tools in random matrix theory and high-dimensional statistics suitable for such regimes. Developing these tools will enable us to effectively tackle the unique challenges posed by large-scale datasets, paving the way for more robust and applicable statistical models in the field of high-dimensional data analysis.

In the extensive-rank regime, our exploration included the application of RIEs to denoising problems and matrix factorization, as detailed in chapters 9, 10, and 11. A major limitation in the current methodology of RIEs is their lack of consideration for the internal structure of the signal. The framework developed so far, while effective in various problems, does not account for the internal constraints and priors on the signal itself. This represents a significant area for future research and development. Incorporating an understanding of the signal's internal structure into RIEs could enhance their effectiveness and applicability, which may also broaden their utility in more complex inference scenarios.

As we conclude our discussion, it is important to acknowledge that despite various progress made, several crucial problems in the extensive-rank regime remain open. One such example is the analysis of denoising models like $\boldsymbol{S} = \boldsymbol{X}\boldsymbol{X}^\intercal$ with separable priors (non-rotational invariant) on entries of $\boldsymbol{X}$. This particular problem presents unique challenges and opportunities for advancing our understanding of high-dimensional inference. The exploration of such models is not only critical for deepening our theoretical knowledge but also for practical applications in fields like machine learning and data science. Addressing these open problems will require innovative approaches and possibly the development of new theoretical tools, continuing the evolution of the interdisciplinary collaboration between statistical physics, random matrix theory and high-dimensional statistics. Tackling these challenges will not only push the boundaries of existing theories but also pave the way for advancements in the analysis and application of extensive-rank problems.

Bibliography

[1] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE transactions on information theory*, vol. 51, no. 4, pp. 1261–1282, 2005.

[2] M. Mezard and A. Montanari, *Information, physics, and computation*. Oxford University Press, 2009.

[3] S. F. Edwards and P. W. Anderson, "Theory of spin glasses," *Journal of Physics F: Metal Physics*, vol. 5, no. 5, p. 965, 1975.

[4] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*. World Scientific Publishing Company, 1987, vol. 9.

[5] T. Lesieur, F. Krzakala, and L. Zdeborová, "Mmse of probabilistic low-rank matrix estimation: Universality with respect to the output channel," in *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2015, pp. 680–687.

[6] F. Guerra and F. L. Toninelli, "Quadratic replica coupling in the sherrington-kirkpatrick mean field spin glass model," *Journal of Mathematical Physics*, vol. 43, no. 7, p. 3704Ð3716, Jul 2002.

[7] S. Korada and N. Macris, "Exact solution of the gauge symmetric p-spin glass model on a complete graph," *Journal of Statistical Physics*, vol. 136, 07 2009.

[8] M. Dia, N. Macris, F. Krzakala, T. Lesieur, L. Zdeborová *et al.*, "Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[9] M. Lelarge and L. Miolane, "Fundamental limits of symmetric low-rank matrix estimation," *Probability Theory and Related Fields*, vol. 173, no. 3, pp. 859–929, 2019.

[10] L. Miolane, "Fundamental limits of low-rank matrix estimation: the non-symmetric case," *arXiv preprint arXiv:1702.00473*, 2017.

[11] J. Barbier and N. Macris, "The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference," *Probability theory and related fields*, vol. 174, no. 3, pp. 1133–1185, 2019.

[12] ——, "The adaptive interpolation method for proving replica formulas. applications to the curie–weiss and wigner spike models," *Journal of Physics A: Mathematical and Theoretical*, vol. 52, no. 29, p. 294002, 2019.

[13] S. F. Edwards and R. C. Jones, "The eigenvalue spectrum of a large symmetric random matrix," *Journal of Physics A: Mathematical and General*, vol. 9, no. 10, p. 1595, 1976.

[14] D. Féral and S. Péché, "The largest eigenvalue of rank one deformation of large wigner matrices," *Communications in mathematical physics*, vol. 272, pp. 185–228, 2007.

[15] M. Capitaine, C. Donati-Martin, and D. Féral, "The largest eigenvalues of finite rank deformation of large wigner matrices: Convergence and nonuniversality of the fluctuations." *Annals of Probability*, vol. 37, pp. 1–47, 2007.

[16] J. Baik, G. B. Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Annals of Probability*, vol. 33, pp. 1643–1697, 2004.

[17] F. Benaych-Georges and R. R. Nadakuditi, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *Advances in Mathematics*, vol. 227, no. 1, pp. 494–521, 2011.

[18] ——, "The singular values and vectors of low rank perturbations of large rectangular random matrices," *Journal of Multivariate Analysis*, vol. 111, pp. 120–135, 2012.

[19] A. Montanari and Y. Wu, "Fundamental limits of low-rank matrix estimation with diverging aspect ratios," *arXiv preprint arXiv:2211.00488*, 2022.

[20] J. Pearl, *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan kaufmann, 1988.

[21] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[22] A. K. Fletcher and S. Rangan, "Iterative reconstruction of rank-one matrices in noise," *Information and Inference: A Journal of the IMA*, vol. 7, no. 3, pp. 531–562, 2018.

[23] A. Montanari and R. Venkataramanan, "Estimation of low-rank matrices via approximate message passing," *The Annals of Statistics*, vol. 49, no. 1, 2021.

[24] 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.

[25] Z. Fan, "Approximate message passing algorithms for rotationally invariant matrices," *The Annals of Statistics*, vol. 50, no. 1, pp. 197–224, 2022.

[26] J. Barbier, F. Camilli, M. Mondelli, and M. Sáenz, "Fundamental limits in structured principal component analysis and how to reach them," *Proceedings of the National Academy of Sciences*, vol. 120, no. 30, p. e2302028120, 2023.

[27] 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.

[28] F. Pourkamali and N. Macris, "Mismatched estimation of symmetric rank-one matrices under gaussian noise," in *International Zurich Seminar on Information and Communication (IZS 2022). Proceedings*. ETH Zurich, 2022, pp. 84–88.

[29] F. Camilli, P. Contucci, and E. Mingione, "An inference problem in a mismatched setting as a spin-glass model with mattis interaction," *arXiv preprint arXiv:2107.11689*, 2021.

[30] J. Barbier, T. Hou, M. Mondelli, and M. Sáenz, "The price of ignorance: how much does it cost to forget noise structure in low-rank matrix estimation?" *Advances in Neural Information Processing Systems*, vol. 35, pp. 36 733–36 747, 2022.

[31] A. Guionnet, J. Ko, F. Krzakala, and L. Zdeborová, "Estimating rank-one matrices with mismatched prior and noise: universality and large deviations," *arXiv preprint arXiv:2306.09283*, 2023.

[32] F. Pourkamali and N. Macris, "Mismatched estimation of non-symmetric rank-one matrices under gaussian noise," in *2022 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2022, pp. 1288–1293.

[33] T. Fu, Y. Liu, J. Barbier, M. Mondelli, S. Liang, and T. Hou, "Mismatched estimation of non-symmetric rank-one matrices corrupted by structured noise," *arXiv preprint arXiv:2302.03306*, 2023.

[34] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[35] ——, "Sparse coding with an overcomplete basis set: A strategy employed by v1?" *Vision research*, vol. 37, no. 23, pp. 3311–3325, 1997.

[36] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on signal processing*, vol. 45, no. 2, pp. 434–444, 1997.

[37] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM (JACM)*, vol. 58, no. 3, pp. 1–37, 2011.

[38] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[39] A. V. Kossenkov and M. F. Ochs, "Matrix factorisation methods applied in microarray data analysis," *International journal of data mining and bioinformatics*, vol. 4, no. 1, pp. 72–90, 2010.

[40] Y. Kabashima, F. Krzakala, M. Mézard, A. Sakata, and L. Zdeborová, "Phase transitions and sample complexity in bayes-optimal matrix factorization," *IEEE Transactions on information theory*, vol. 62, no. 7, pp. 4228–4265, 2016.

[41] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—part i: Derivation," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 5839–5853, 2014.

[42] Q. Zou, H. Zhang, and H. Yang, "Multi-layer bilinear generalized approximate message passing," *IEEE Transactions on Signal Processing*, vol. 69, pp. 4529–4543, 2021.

[43] A. Maillard, F. Krzakala, M. Mézard, and L. Zdeborová, "Perturbative construction of mean-field equations in extensive-rank matrix factorization and denoising," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2022, no. 8, p. 083301, 2022.

[44] J. T. Parker, P. Schniter, and V. Cevher, "Bilinear generalized approximate message passing—part ii: Applications," *IEEE Transactions on Signal Processing*, vol. 62, no. 22, pp. 5854–5867, 2014.

[45] J. Barbier and N. Macris, "Statistical limits of dictionary learning: random matrix theory and the spectral replica method," *Physical Review E*, vol. 106, no. 2, p. 024136, 2022.

[46] F. Camilli and M. Mézard, "The decimation scheme for symmetric matrix factorization," *arXiv preprint arXiv:2307.16564*, 2023.

[47] ——, "Matrix factorization with neural networks," *Physical Review E*, vol. 107, no. 6, p. 064308, 2023.

[48] F. Pourkamali, J. Barbier, and N. Macris, "Matrix inference in growing rank regimes," *arXiv preprint arXiv:2306.01412*, 2023.

[49] C. Stein, "Estimation of a covariance matrix," in *39th Annual Meeting IMS, Atlanta, GA, 1975*, 1975.

[50] A. Takemura, "An orthogonally invariant minimax estimator of the covariance matrix of a multivariate normal population," *Tsukuba journal of mathematics*, vol. 8, no. 2, pp. 367–376, 1984.

[51] J. Bun, J.-P. Bouchaud, and M. Potters, "Cleaning large correlation matrices: tools from random matrix theory," *Physics Reports*, vol. 666, pp. 1–109, 2017.

[52] F. Benaych-Georges, J.-P. Bouchaud, and M. Potters, "Optimal cleaning for singular values of cross-covariance matrices," *The Annals of Applied Probability*, vol. 33, no. 2, pp. 1295–1326, 2023.

[53] J. Bun, R. Allez, J.-P. Bouchaud, and M. Potters, "Rotational invariant estimator for general noisy matrices," *IEEE Transactions on Information Theory*, vol. 62, no. 12, pp. 7475–7490, 2016.

[54] F. Pourkamali and N. Macris, "Rectangular rotational invariant estimator for general additive noise matrices," in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 2081–2086.

[55] I. D. Landau, G. C. Mel, and S. Ganguli, "Singular vectors of sums of rectangular random matrices and optimal estimators of high-rank signals: The extensive spike model," *arXiv preprint arXiv:2306.00340*, 2023.

[56] F. Pourkamali and N. Macris, "Bayesian extensive-rank matrix factorization with rotational invariant priors," in *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[57] D. Voiculescu, "Symmetries of some reduced free product c*-algebras," in *Operator Algebras and their Connections with Topology and Ergodic Theory: Proceedings of the OATE Conference held in Buşteni, Romania, Aug. 29–Sept. 9, 1983*. Springer, 2006, pp. 556–588.

[58] ——, "Limit laws for random matrices and free products," *Inventiones mathematicae*, vol. 104, no. 1, pp. 201–220, 1991.

[59] Harish-Chandra, "Differential operators on a semisimple lie algebra," *American Journal of Mathematics*, pp. 87–120, 1957.

[60] C. Itzykson and J.-B. Zuber, "The planar approximation. ii," *Journal of Mathematical Physics*, vol. 21, no. 3, pp. 411–421, 1980.

[61] A. Guionnet and O. Zeitouni, "Large deviations asymptotics for spherical integrals," *Journal of functional analysis*, vol. 188, no. 2, pp. 461–515, 2002.

[62] A. Guionnet and M. Maïda, "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.

[63] A. Guionnet and J. Husson, "Asymptotics of k dimensional spherical integrals and applications," *ALEA*, vol. 19, pp. 769–797, 2022.

[64] F. Benaych-Georges, "Rectangular R-transform as the limit of rectangular spherical integrals," *Journal of Theoretical Probability*, vol. 24, no. 4, pp. 969–987, 2011.

[65] J. Husson and J. Ko, "Spherical integrals of sublinear rank," *arXiv preprint arXiv:2208.03642*, 2022.

[66] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*. Cambridge university press, 2010.

[67] M. Potters and J.-P. Bouchaud, *A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists*. Cambridge University Press, 2020.

[68] J. Wishart, "The generalised product moment distribution in samples from a normal multivariate population," *Biometrika*, pp. 32–52, 1928.

[69] V. A. Marchenko and L. A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Matematicheskii Sbornik*, vol. 114, no. 4, pp. 507–536, 1967.

[70] E. P. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions i," *The Collected Works of Eugene Paul Wigner: Part A: The Scientific Papers*, pp. 524–540, 1993.

[71] R. A. Horn and C. R. Johnson, *Matrix analysis*. Cambridge university press, 2012.

[72] J. A. Mingo and R. Speicher, *Free probability and random matrices*. Springer, 2017, vol. 35.

[73] D. Voiculescu, K. J. Dykema, and A. Nica, *Free random variables*. American Mathematical Soc., 1992, no. 1.

[74] F. Benaych-Georges, "Rectangular random matrices, related convolution," *Probability Theory and Related Fields*, vol. 144, no. 3, pp. 471–515, 2009.

[75] D. Voiculescu, "The analogues of entropy and of fisher's information measure in free probability theory, i," *Communications in mathematical physics*, vol. 155, no. 1, pp. 71–92, 1993.

[76] ——, "The analogues of entropy and of fisher's information measure in free probability theory, ii," *Inventiones mathematicae*, vol. 118, no. 1, pp. 411–440, 1994.

[77] ——, "The analogues of entropy and of fisher's information measure in free probability theory iii: The absence of cartan subalgebras," *Geometric & Functional Analysis GAFA*, vol. 6, no. 1, pp. 172–199, 1996.

[78] ——, *Free probability theory*. American Mathematical Soc., 1997, vol. 12.

[79] ——, "The analogues of entropy and fisher's information in free probability theory, v: Noncommutative hilbert transforms," *Invent. Math*, vol. 132, pp. 182–227, 1998.

[80] ——, "The analogues of entropy and of fisher's information measure in free probability theory: Vi. liberation and mutual free information," *Advances in Mathematics*, vol. 146, no. 2, pp. 101–166, 1999.

[81] ——, "Free entropy," *Bulletin of the London Mathematical Society*, vol. 34, no. 3, pp. 257–278, 2002.

[82] F. Morone, F. Caltagirone, E. Harrison, and G. Parisi, "Replica theory and spin glasses," *arXiv preprint arXiv:1409.2722*, 2014.

[83] J. Zinn-Justin, *Quantum field theory and critical phenomena*. Oxford university press, 2021, vol. 171.

[84] J. M. Kosterlitz, D. J. Thouless, and R. C. Jones, "Spherical model of a spin-glass," *Physical Review Letters*, vol. 36, no. 20, p. 1217, 1976.

[85] G. Parisi and M. Potters, "Mean-field equations for spin models with orthogonal interaction matrices," *Journal of Physics A: Mathematical and General*, vol. 28, no. 18, p. 5267, 1995.

[86] J.-B. Zuber, "Horn's problem and harish-chandra's integrals. probability density functions," *Annales de l'Institut Henri Poincaré D*, vol. 5, no. 3, pp. 309–338, 2018.

[87] R. Coquereaux, C. McSwiggen, and J.-B. Zuber, "On horn's problem and its volume function," *Communications in Mathematical Physics*, vol. 376, no. 3, pp. 2409–2439, 2020.

[88] G. Biroli and A. Guionnet, "Large deviations for the largest eigenvalues and eigenvectors of spiked gaussian random matrices," *Electronic Communications in Probability*, vol. 25, 2020.

[89] S. T. Belinschi, A. Guionnet, and J. Huang, "Large deviation principles via spherical integrals," *Probability and Mathematical Physics*, 2020.

[90] A. Guionnet and J. Husson, "Large deviations for the largest eigenvalue of rademacher matrices," *The Annals of Probability*, 2018.

[91] F. Augeri, A. Guionnet, and J. Husson, "Large deviations for the largest eigenvalue of sub-gaussian matrices," *Communications in Mathematical Physics*, vol. 383, pp. 997 – 1050, 2019.

[92] N. A. Cook, R. Ducatez, and A. Guionnet, "Full large deviation principles for the largest eigenvalue of sub-gaussian wigner matrices," *arXiv preprint arXiv:2302.14823*, 2023.

[93] A. Guionnet and M. Maïda, "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. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022123604003623

[94] B. Collins and P. Śniady, "New scaling of itzykson–zuber integrals," in *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, vol. 43, no. 2. Elsevier, 2007, pp. 139–146.

[95] A. Matytsin, "On the large-N limit of the Itzykson-Zuber integral," *Nuclear Physics B*, vol. 411, no. 2-3, pp. 805–820, 1994.

[96] Y. Kabashima, "Inference from correlated patterns: a unified theory for perceptron learning and linear vector channels," in *Journal of Physics: Conference Series*, vol. 95, no. 1. IOP Publishing, 2008, p. 012001.

[97] A. Guionnet and J. Huang, "Asymptotics of rectangular spherical integrals," *Journal of Functional Analysis*, vol. 285, no. 11, p. 110144, 2023.

[98] E. Troiani, V. Erba, F. Krzakala, A. Maillard, and L. Zdeborová, "Optimal denoising of rotationally invariant rectangular matrices," in *Mathematical and Scientific Machine Learning*. PMLR, 2022, pp. 97–112.

[99] Y. Kabashima, "An integral formula for large random rectangular matrices and its application to analysis of linear vector channels," in *2008 6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks and Workshops*. IEEE, 2008, pp. 620–624.

[100] S. Verdu, "Mismatched estimation and relative entropy," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3712–3720, 2010.

[101] M. Capitaine and C. Donati-Martin, "Spectrum of deformed random matrices and free probability," Jul. 2016, working paper or preprint. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01346371

[102] F. Benaych-Georges and R. R. Nadakuditi, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *Advances in Mathematics*, vol. 227, no. 1, pp. 494–521, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0001870811000570

[103] S. Rangan and A. K. Fletcher, "Iterative estimation of constrained rank-one matrices in noise," in *2012 IEEE International Symposium on Information Theory Proceedings*, 2012, pp. 1246–1250.

[104] T. Lesieur, F. Krzakala, and L. Zdeborová, "Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2017, no. 7, p. 073403, 2017.

[105] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in gaussian channels," *Information Theory, IEEE Transactions on*, vol. 51, pp. 1261 – 1282, 05 2005.

[106] Y. Deshpande and A. Montanari, "Information-theoretically optimal sparse pca," in *2014 IEEE International Symposium on Information Theory*. IEEE, 2014, pp. 2197–2201.

[107] Y. Deshpande, E. Abbe, and A. Montanari, "Asymptotic mutual information for the two-groups stochastic block model," 2015.

[108] J. Barbier and N. Macris, "The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference," *Probability Theory and Related Fields*, vol. 174, pp. 1133–1185, 2019.

[109] F. Antenucci, F. Krzakala, P. Urbani, and L. Zdeborova, "Approximate survey propagation for statistical inference," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2019, no. 2, p. 023401, 2019.

[110] J. Barbier, M. Dia, N. Macris, F. Krzakala, and L. Zdeborová, "Rank-one matrix estimation: analysis of algorithmic and information theoretic limits by the spatial coupling method," *arXiv preprint arXiv:1812.02537*, 2018.

[111] F. Olver, "9 - integrals: Further methods," in *Asymptotics and Special Functions*, F. Olver, Ed. Academic Press, 1974, pp. 322–361. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780125258500500146

[112] J. Huang, "Asymptotic expansion of spherical integral," *Journal of Theoretical Probability*, vol. 32, no. 2, pp. 1051–1075, 2019.

[113] M. Abramowitz, I. A. Stegun, and R. H. Romer, "Handbook of mathematical functions with formulas, graphs, and mathematical tables," 1988.

[114] S. Verdú, "Mismatched estimation and relative entropy," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3712–3720, 2010.

[115] C. Luneau, N. Macris, and J. Barbier, "High-dimensional rank-one non-symmetric matrix decomposition: the spherical case," in *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2020, pp. 2646–2651.

[116] K. R. Davidson and S. J. Szarek, "Local operator theory, random matrices and banach spaces," *Handbook of the geometry of Banach spaces*, vol. 1, no. 317-366, p. 131, 2001.

[117] W. R. Inc., "Mathematica, Version 13.2," champaign, IL, 2022. [Online]. Available: https://www.wolfram.com/mathematica

[118] P. Biane, "On the free convolution with a semi-circular distribution," *Indiana University Mathematics Journal*, pp. 705–718, 1997.

[119] C. Villani, *Topics in optimal transportation*. American Mathematical Soc., 2021, vol. 58.

[120] G. H. Hardy, J. E. Littlewood, G. Pólya, G. Pólya *et al.*, *Inequalities*. Cambridge university press, 1952.

[121] D. Shlyakhtenko and T. T. W. an appendix by David Jekel, "Fractional free convolution powers," *Indiana University Mathematics Journal*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:221507737

[122] A. Nica and R. Speicher, *Lectures on the combinatorics of free probability*. Cambridge University Press, 2006, vol. 13.

[123] R. R. Müller, "Random matrices, free probability and the replica method," in *2004 12th European Signal Processing Conference*. IEEE, 2004, pp. 189–196.

[124] A. Nica, R. Speicher, A. Tulino, and D. Voiculescu, "Free probability extensions and applications," in *Proc. BIRS*. Citeseer, 2008, pp. 3–10.

[125] G. Reeves, H. D. Pfister, and A. Dytso, "Mutual information as a function of matrix snr for linear gaussian channels," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 1754–1758.

[126] J. Huang, "Mesoscopic perturbations of large random matrices," *Random Matrices: Theory and Applications*, vol. 7, no. 02, p. 1850004, 2018.

[127] T. Sauer, *Numerical analysis*. Addison-Wesley Publishing Company, 2011.

[128] S. Janson, "Resultant and discriminant of polynomials," *Notes, September*, vol. 22, 2007.

[129] L. ErdHos, B. Schlein, and H.-T. Yau, "Semicircle law on short scales and delocalization of eigenvectors for wigner random matrices," *Annals of Probability*, vol. 37, pp. 815–852, 2007.

[130] L. Erdős and H.-T. Yau, *A dynamical approach to random matrix theory*. American Mathematical Soc., 2017, vol. 28.

[131] F. Benaych-Georges, "Rectangular random matrices, entropy, and fisher's information," *Journal of Operator Theory*, pp. 371–419, 2009.

[132] A. Guionnet and J. Huang, "Asymptotics of rectangular spherical integrals," *Journal of Functional Analysis*, p. 110144, 2023.

[133] F. Benaych-Georges, "Infinitely divisible distributions for rectangular free convolution: classification and matricial interpretation," *Probability Theory and Related Fields*, vol. 139, pp. 143–189, 2007.

[134] B. Simon, *Trace ideals and their applications*. American Mathematical Soc., 2005, no. 120.

[135] F. R. Kschischang, "The Hilbert transform," *University of Toronto*, vol. 83, p. 277, 2006.

[136] D. S. Bernstein, "Matrix mathematics," in *Matrix Mathematics*. Princeton university press, 2009.

[137] C. Bishop, "Variational principal components," *IET Conference Proceedings*, pp. 509–514(5), January 1999. [Online]. Available: https://digital-library.theiet.org/content/conferences/10.1049/cp_19991160

[138] Y. J. Lim and Y. W. Teh, "Variational Bayesian approach to movie rating prediction," in *Proceedings of KDD cup and workshop*, vol. 7, 2007, pp. 15–21.

[139] A. Ilin and T. Raiko, "Practical approaches to principal component analysis in the presence of missing values," *The Journal of Machine Learning Research*, vol. 11, pp. 1957–2000, 2010.

[140] S. Nakajima, M. Sugiyama, S. D. Babacan, and R. Tomioka, "Global analytic solution of fully-observed variational bayesian matrix factorization," *The Journal of Machine Learning Research*, vol. 14, no. 1, pp. 1–37, 2013.

[141] S. Tarmoun, G. Franca, B. D. Haeffele, and R. Vidal, "Understanding the dynamics of gradient flow in overparameterized linear models," in *International Conference on Machine Learning*. PMLR, 2021, pp. 10 153–10 161.

[142] A. Bodin and N. Macris, "Gradient flow on extensive-rank positive semidefinite matrix denoising," in *2023 IEEE Information Theory Workshop (ITW)*, 2023, pp. 365–370.

EPFL IC IINFCOM LTHC INR 036 Station 14
CH-1015 Lausanne
+41787117083
farzad.pourkamali@epfl.ch
Github Linkedin Webpage

Farzad Pourkamali

Education

- **École Polytechnique Fédérale de Lausanne** — Lausanne, Switzerland
  - Ph.D., School of Computer and Communication Sciences — 2019, 2024
  - High-dimensional statistics, Machine learning
- **Sharif University of Technology** — Tehran, Iran
  - B.Sc., Double Majors In Electrical Engineering And Theoretical Mathematics — 2014- 2019

Publications

- **F. Pourkamali**, N. Macris, "Mismatched Estimation of Rank-One Symmetric Matrices Under Gaussian Noise", In International Zurich Seminar on Information and Communication (IZS 2022). Proceedings, pp. 84-88. ETH Zurich, 2022.
- **F. Pourkamali**, N. Macris, "Mismatched Estimation of Rank-One Non-Symmetric Matrices Under Gaussian Noise", In 2022 IEEE International Symposium on Information Theory (ISIT). IEEE, 2022.
- **F. Pourkamali**, N. Macris, "Rectangular Rotational Invariant Estimator for General Additive Noise Matrices", In 2023 IEEE International Symposium on Information Theory (ISIT). IEEE, 2023.
- **F. Pourkamali**, J. Barbier, N. Macris, "Matrix Inference in Growing Rank Regimes", *Submitted to the IEEE Transactions on Information Theory .*
- **F. Pourkamali**, N. Macris, "Bayesian Extensive-Rank Matrix Factorization with Rotational Invariant Priors", *Accepted as* SPOTLIGHT *in Conference on Neural Information Processing Systems (NeurIPS) 2023.*
- **F. Pourkamali**, N. Macris, "Rectangular Rotational Invariant Estimator for High-Rank Matrix Estimation", *Submitted to Information and Inference: A Journal of the IMA.*
- **F. Pourkamali**, N. Macris, "Solution to the Three-Matrix Model", *To be submitted.*

Research Experiences

- **Doctoral Assistant at EPFL** — Lausanne, Switzerland
  - Thesis Advisor : Professor Nicolas Macris — Sep 2019- present
  - Working on high-dimensional statistics and machine learning problems such as matrix/tensor inference.
- **Research Assistant at Sharif University of Technology** — Tehran, Iran
  - Under the supervision of Professor Arash Amini & Dr. Mehdi Molkaraie — Sep 2018- June 2019
  - Investigating the run-time and accuracy of sequential Monte Carlo methods applied to various models, such as two-dimensional channels and Ising mode, for computing information theoretic quantities.
- **Research Intern at Summer@EPFL** — Lausanne, Switzerland
  - Working in Information and Network Dynamics Laboratory 2 (INDY2) — July-Sep 2018
  - Under the supervision of Professor Patrick Thiran
  - Investigating the source localization in real-world networks with dynamic and static settings.
- **Research Assistant at Sharif University of Technology** — Tehran, Iran
  - Working in Learning and Intelligent Systems (LIS) — Oct 2017- Feb 2019
  - Under the supervision of Professor Saber Salehkaleybar
  - Designing an algorithm for learning Bayesian networks based on the PC algorithm, requiring less

number of samples compared to the original PC.

- **Research Assistant at Chinese University of Hong Kong (CUHK)** Hong Kong
  - Working under the supervision of Professor Pascal O. Vontobel July-Sep 2017
  - Studying the Bethe approximation for the permanent of a non-negative matrix.

Selected Courses

Deep Learning	5.75/6	Prof. Fleuret
Deep Learning For Natural Language Processing	5.25/6	Dr. Henderson
Statistical Physics For Optimization & learning	5.5/6	Prof. Krzakala, Zdeborová
Information Theory and Coding	6/6	Prof. Telatar
Combinatorial Statistics	5.75/6	Prof. Abbe
Distributed Systems	20/20	Prof. Salehkaleybar
Convex Optimization	19.5/20	Prof. Alishahi

Honors and Awards

- **Research Fellowship** at *École Polytechnique Fédérale de Lausanne* Switzerland, 2018
- **Research Fellowship** at *Chinese University of Hong Kong* Hong Kong, 2017
- Recipient of the grant for undergraduate studies from the **Iranian National Elites Foundation** for outstanding academic success 2014 - 2019
- **Ranked 28 out of nearly 200,000 participants** Iran
  in University National Entrance Exam for B.Sc. degree 2014

Teaching & Mentoring Expriences

- Teaching Assistant
  - Learning Theory [Prof. Urbanke, Prof. Macris] 2023S,2022S
  - Markov Chains and Algorithmic Applications [Prof. Macris, Dr. Leveque] 2022F, 2021F
  - Quantum computation [Prof. Macris] 2021S
  - Quantum information processing [Prof. Macris] 2020F
  - Probabilities and statistics [Prof. Abbe] 2020S
- Students mentored
  - Boukil Farouk, semester project, Two temperature matrix estimation 2021F
  - Mike Sinsoillier, Bachelor thesis, Quantum methods for sampling and understanding quantum advantage 2021S

Professional Activities

- Reviewer for IEEE Transactions on Information Theory
- Reviewer for IEEE International Symposium on Information Theory (ISIT) 2021,2022, 2023, 2024
- Student Committee of MWTS 2017, Sharif University of Technology Tehran, Iran
  2nd Conference on Modern Wireless Telecommunication Systems 2017

Programming Languages

- Python, c++, MATLAB, R
- Machine Learning Packages: Pytorch, scikit-learn

Languages

English: Fluent **French**: Intermediate **Farsi**: Native