

Fundamental Limits of Low-Rank Matrix Estimation with Diverging Aspect Ratios

Andrea Montanari^{†*}

Yuchen Wu[†]

Abstract

We consider the problem of estimating the factors of a low-rank $n \times d$ matrix, when this is corrupted by additive Gaussian noise. A special example of our setting corresponds to clustering mixtures of Gaussians with equal (known) covariances. Simple spectral methods do not take into account the distribution of the entries of these factors and are therefore often suboptimal. Here, we characterize the asymptotics of the minimum estimation error under the assumption that the distribution of the entries is known to the statistician.

Our results apply to the high-dimensional regime $n, d \rightarrow \infty$ and $d/n \rightarrow \infty$ (or $d/n \rightarrow 0$) and generalize earlier work that focused on the proportional asymptotics $n, d \rightarrow \infty, d/n \rightarrow \delta \in (0, \infty)$. We outline an interesting signal strength regime in which $d/n \rightarrow \infty$ and partial recovery is possible for the left singular vectors while impossible for the right singular vectors.

We illustrate the general theory by deriving consequences for Gaussian mixture clustering and carrying out a numerical study on genomics data.

Contents

1	Introduction	3
1.1	Notations and conventions	5
2	Further related work	6
3	Strong signal regime	7
3.1	Estimation of $\mathbf{\Lambda}$	7
3.2	Estimation of $\mathbf{\Theta}$	8
4	Weak signal regime	8
4.1	Background: the symmetric spiked model	9
4.2	Estimation of $\mathbf{\Theta}$	9
4.3	Estimation of $\mathbf{\Lambda}$	10
5	Clustering under the Gaussian mixture model	13
5.1	Two clusters with symmetric centers	13
5.2	Two or more clusters with orthogonal centers	14
5.3	Numerical experiments	15

*Department of Electrical Engineering, Stanford University; School of Mathematics, Institute for Advanced Studies, Princeton

[†]Department of Statistics, Stanford University

6	Asymmetry in factors estimation: real world datasets	16
6.1	1000 Genomes Project	16
6.2	RNA-Seq gene expression	20
A	Preliminaries	29
A.1	Further notations and conventions	29
A.2	Wasserstein distance	29
B	Technical lemmas	29
C	Proofs for the strong signal regime	31
C.1	Proof of Theorem 3.1	31
C.2	Proof of Theorem 3.2	32
C.3	Proof of Theorem 3.3	34
D	Proof outlines for the weak signal regime	35
D.1	Proof of Theorem 4.2	35
D.2	Proof outline of Theorem 4.3	35
	D.2.1 Free energy density	36
	D.2.2 Asymptotic equivalence of free energy densities	37
D.3	Proof of Theorem 4.4	39
D.4	Reduction to bounded prior	41
D.5	Proof outline of Theorem 4.5	42
	D.5.1 Proof outline of Theorem 4.5 under condition (a)	42
	D.5.2 Proof outline of Theorem 4.5 under condition (b)	44
	D.5.3 Proof of Theorem 4.5 under condition (c)	51
E	Convergence of free energy density	52
E.1	Proof of Lemma D.2	52
	E.1.1 Upper bounding $\mathbb{E}[\ \mu_n^{(k,+)}[\sum_{l=5}^{\infty} \frac{1}{l!} h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})^l]\]$	55
	E.1.2 Upper bounding $\mathbb{E}[\ \mu_n^{(k,+)}[R^{(k)}]\]$	56
	E.1.3 Combining the upper bounds	56
E.2	Proof of Lemma D.3	58
E.3	Proof of Lemma D.4	59
E.4	Proof of Lemma D.5	60
E.5	Proof of Lemma D.6	61
F	Achieving the Bayesian MMSE	64
F.1	Proof of Lemma D.9	64
F.2	Proof of Lemma D.10	65
F.3	Proof of Lemma D.11	66
F.4	Proof of Lemma D.12	66
F.5	Proof of Lemma D.13	67
F.6	Proof of Lemma D.14	68
F.7	Proof of Lemma D.15	68
F.8	Proof of Lemma D.16	69
F.9	Proof of the first claim of Lemma D.17	70
F.10	Proof of the second claim of Lemma D.17	71

G Proofs for the Gaussian mixture clustering example	72
G.1 Proof of Proposition 5.1 claim (a)	72
G.2 Proof of Theorem 5.1 part (a)	73
G.3 Proof of Theorem 5.1 part (b)	74

1 Introduction

The problem of low-rank matrix estimation has received enormous attention within high-dimensional statistics, probability theory, random matrix theory, and computer science. The *spiked model* introduced by Johnstone [Joh01] (also known as ‘signal-plus-noise’ model or ‘deformed ensemble’) plays a central role in theoretical analysis, and has inspired a number of statistical insights [HR04, BAP05, BS06, Joh06, AW08, JL09, Ma13, DM14b, DM14a, BGN11, BGN12, JO20, MR⁺20].

In the spiked model, we observe a matrix $\mathbf{A} \in \mathbb{R}^{n \times d}$ which is given by the sum of a low-rank signal and random noise

$$\mathbf{A} = s_n \mathbf{\Lambda} \mathbf{\Theta}^\top + \mathbf{Z}. \quad (1)$$

Here, $\mathbf{\Lambda} \in \mathbb{R}^{n \times r}$, $\mathbf{\Theta} \in \mathbb{R}^{d \times r}$ are factors which we would like to estimate and \mathbf{Z} is a noise matrix with i.i.d. entries $Z_{ij} \sim \mathcal{N}(0, 1)$. Finally, $s_n \in \mathbb{R}_{>0}$ is a signal-to-noise ratio which we also assume to be known. We will consider high-dimensional asymptotics whereby $d, n \rightarrow \infty$ with r fixed. In what follows, we will denote by $\mathbf{a}_1, \dots, \mathbf{a}_n$ the rows of \mathbf{A} .

Low-rank matrix estimation is ubiquitous in high-dimensional statistics, with special cases including sparse PCA [JL09, JNRS10, BR13b, DM14b, DM14a], community detection [Abb17, DAM17], submatrix localization [KBR11, HWX17], and Gaussian mixture clustering [MV10, RV17, Nda22, CMZ⁺19]. We illustrate the broad applicability of model (1) using two examples:

Example 1.1 (Sparse PCA). In a simple model for sparse PCA [JL09], we observe vectors

$$\mathbf{a}_1, \dots, \mathbf{a}_n \sim_{iid} \mathcal{N}(0, \mathbf{\Sigma}),$$

where $\mathbf{\Sigma} = s_n^2 \mathbf{\Theta} \mathbf{\Theta}^\top + \mathbf{I}_d$ with $\mathbf{\Theta} \in \mathbb{R}^d$ a sparse vector that we would like to estimate.

This is the special case of model (1), if we let $r = 1$ and $\mathbf{\Lambda} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$.

Example 1.2 (Mixture of Gaussians with known covariance). In a mixture of Gaussian model, we observe vectors $\bar{\mathbf{a}}_1, \dots, \bar{\mathbf{a}}_n \sim_{iid} p \mathcal{N}(\mathbf{\Theta}_1, \mathbf{\Sigma}_1) + (1 - p) \mathcal{N}(\mathbf{\Theta}_2, \mathbf{\Sigma}_2)$. If the covariances coincide and are known: $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{\Sigma}$, and the population mean $\bar{\mathbf{\Theta}} := p \mathbf{\Theta}_1 + (1 - p) \mathbf{\Theta}_2$ can be estimated accurately, then we can define $\mathbf{a}_i = \mathbf{\Sigma}^{-1/2}(\bar{\mathbf{a}}_i - \bar{\mathbf{\Theta}})$. Hence the model is equivalent to observing $\mathbf{a}_1, \dots, \mathbf{a}_n \sim_{iid} p \mathcal{N}((1 - p) \mathbf{\Theta}, \mathbf{I}_d) + (1 - p) \mathcal{N}(-p \mathbf{\Theta}, \mathbf{I}_d)$, with $\mathbf{\Theta} := \mathbf{\Sigma}^{-1/2}(\mathbf{\Theta}_1 - \mathbf{\Theta}_2)$.

This is another special case of model (1), with $r = 1$ and $(\Lambda_i)_{i \leq n} \sim_{iid} p \delta_{(1-p)} + (1 - p) \delta_{-p}$. Estimating $\mathbf{\Lambda}$ amounts to estimating the cluster labels.

As demonstrated in these examples, it is often the case that the latent factors $\mathbf{\Lambda}$, $\mathbf{\Theta}$ have additional structure. In the first example $\mathbf{\Theta} \in \mathbb{R}^d$ is a sparse vector, while in the second one $\mathbf{\Lambda}$ is a vector with i.i.d. entries distributed according to a two-point mixture. In this paper, we assume a stylized model whereby the rows $\mathbf{\Lambda}$, $\mathbf{\Theta}$ are mutually independent (and independent of \mathbf{Z}) with $(\Lambda_i)_{i \leq n} \stackrel{iid}{\sim} \mu_\Lambda$ and $(\Theta_j)_{j \leq d} \stackrel{iid}{\sim} \mu_\Theta$. Here, μ_Λ and μ_Θ are fixed probability distributions on \mathbb{R}^r . We will be concerned with the problem of determining the Bayes optimal estimation error under the idealized setting in which the distributions μ_Λ, μ_Θ and the signal-to-noise ratio s_n are known to the statistician.

This setting was considered several times in recent past, see e.g., [Mio17, LM19, MV21b, BMV⁺18] and Section 2 for further references. Closely related to our work are the results of [Mio17, LM19], who determined the precise asymptotics of mutual information and (certain) estimation error metrics when $n, d \rightarrow \infty$, in the proportional regime $d/n \rightarrow \delta \in (0, \infty)$.

Our goal is to move beyond the proportional asymptotics and consider the cases $d/n \rightarrow \infty$ and $d/n \rightarrow 0$. We believe an analysis of these regimes is essential for at least two reasons.

- *First*, in modern applications, it is often the case that the dimension d is much larger than the sample size n . For instance, in genomics studies, n is the number of sequences (subjects) and is often between one hundred and a few thousands, while d is the number of gene variations under study and can be of the order of hundreds of thousands. At the other end of the spectrum, n can be significantly larger than d yet classical low-dimensional theory is not accurate.
- *Second*, in practice, the statistician does not observe a sequence of data matrices $\{\mathbf{A}_n\}$ with increasing $\{(n, d)\}$, but a single pair (n, d) . It is a natural reflex to apply proportional asymptotics results with the nominal aspect ratio $\hat{\delta} := d/n$. However, this choice does not have a rigorous justification. This is not just a mathematicians' quibble. Imagine, for a moment, what would happen if the asymptotics for $d/n^\alpha \rightarrow \delta_\alpha \in (0, \infty)$ (with $\alpha > 1$, say $\alpha = 3/2$) were different from the one for $d/n \rightarrow \delta$ followed by $\delta \rightarrow \infty$. Given data with $n = 100$ and $d = 1000$, we would obtain different theoretical prediction depending on whether we regarded it as an element of the sequence $d(n) = 10n$ (i.e., proportional asymptotics with $\delta = 10$), or $d(n) = n^{3/2}$ (i.e., $\alpha = 3/2$, $\delta_\alpha = 1$). We will see that this is not the case.

Before proceeding, we point out that recent work addresses the assumption that $s_n, \mu_\Lambda, \mu_\Theta$ are known. Namely, [ZSF22] uses empirical Bayes techniques to show that the Bayes error can be achieved even if $s_n, \mu_\Lambda, \mu_\Theta$ are unknown, under the proportional asymptotics (and in certain settings). We expect that our results should translate into similar guarantees for empirical Bayes methods in the case of diverging aspect ratios.

We next summarize our results, focusing to be definite on the case $d/n \rightarrow \infty$: analogous statements for $d/n \rightarrow 0$ follow by interchanging n and d , as well as Λ and Θ .

We show that depending on the scaling of the signal-to-noise ratio s_n , there are two interesting regimes that control the behavior of the estimation problem.

Strong signal regime. This is obtained for $s_n \asymp n^{-1/2}$, and is relatively easy to characterize analytically. Under this scaling, Λ can be estimated consistently (possibly up to a rotation), while the minimum normalized estimation error of Θ remains bounded away from 0. We characterize the limiting error of estimating Θ .

Weak signal regime: Estimation of Θ . This regime corresponds to $s_n \asymp (nd)^{-1/4}$, and most of our technical work is devoted to its analysis. We prove that, in this regime, non-trivial estimation of Θ is impossible: any estimator has asymptotically the same risk as the the null estimator $\hat{\Theta}_0 = \mathbb{E}[\Theta]$.

Weak signal regime: Mutual information. On the other hand, still in taking $s_n \asymp (nd)^{-1/4}$, estimation of Λ is non-trivial. As a first result in this direction, we characterize the asymptotic mutual information

$$\lim_{n, d \rightarrow \infty} \frac{1}{n} I(\mathbf{A}; \Lambda),$$

and show that this is non-vanishing. Further, this mutual information is asymptotically the same as for a symmetric observation model in which instead of $\mathbf{A} \in \mathbb{R}^{n \times d}$, we observe $\mathbf{Y} \in$

$\mathbb{R}^{n \times n}$ given by

$$\mathbf{Y} = \frac{q_\Theta}{n} \mathbf{\Lambda} \mathbf{\Lambda}^\top + \mathbf{W}, \quad \mathbf{W} \sim \text{GOE}(n). \quad (2)$$

(Here, $q_\Theta := r^{-1} \int \|\Theta\|^2 \mu_\Theta(d\Theta)$, we take without loss of generality $s_n = (nd)^{-1/4}$, and $\text{GOE}(n)$ denotes the distribution of a symmetric matrix with independent entries on or above the diagonal $(W_{ij})_{i \leq j}$ such that $W_{ii} \sim \mathcal{N}(0, 2/n)$ and $W_{ij} \sim \mathcal{N}(0, 1/n)$ for $i < j$.)

Weak signal regime: Estimation error. We then proceed to study the asymptotics of the Bayes optimal matrix mean square error:

$$\text{MMSE}(\mu_\Lambda, \mu_\Theta) := \lim_{n, d \rightarrow \infty} \frac{1}{n^2} \mathbb{E} \{ \|\mathbf{\Lambda} \mathbf{\Lambda}^\top - \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top | \mathbf{A}] \|_F^2 \}. \quad (3)$$

We characterize this limit in two regimes: $n \ll d \ll n^{6/5}$ or $n^3 \ll d$, and in certain cases for all $n \ll d$ (here \ll hides logarithmic factors.) In these cases we prove equivalence with model (2).

We believe that the conditions $n \ll d \ll n^{6/5}$ or $n^3 \ll d$ are artifacts of the proof. Indeed, the conclusion holds for all $n \ll d$ under a natural (unproven) continuity assumption. We leave it to future work to cover the intermediate range $n^{6/5} \lesssim d \lesssim n^3$.

Both the limiting mutual information and the asymptotic estimation error $\text{MMSE}(\mu_\Lambda, \mu_\Theta)$ are given by explicit expressions known as ‘replica symmetric formulas,’ because they are correctly predicted by the replica method in spin glass theory [MPV87, MS22]. However, the asymptotic equivalence with the symmetric model (2) is insightful in itself (i.e., independently of the fact that we can give explicit formulas for the asymptotic error and mutual information):

1. The asymptotic equivalence between model (1) and model (2) implies that the optimal estimation depends on μ_Θ only through its second moment. In other words, no substantial improvement is achieved in the regime covered by this equivalence exploiting the knowledge of the distribution of Θ .
2. The symmetric matrix \mathbf{Y} is closely related to the Gram matrix $\mathbf{Y}' = (\mathbf{A} \mathbf{A}^\top - d \mathbf{I}_n) / \sqrt{nd}$, an observation that is confirmed by inspecting the proof. This implies that there is no substantial loss of accuracy in estimating $\mathbf{\Lambda}$ uniquely on the basis of \mathbf{Y}' . This yields a substantial reduction in complexity for $n \ll d$.

We warn the reader that these conclusions do not apply in settings that are not captured here. For instance, in sparse PCA, cf. Example 1.1, one might be interested in cases in which the number of non-zeros of the principal component Θ is sub-linear in the dimension d . This case cannot be modeled as above, and requires instead to consider μ_Θ dependent on n, d .

The rest of the paper is organized as follows. We briefly review related work in Section 2. We then present our results for the strong signal regime in Section 3 and the weak signal regime in Section 4. We finally apply the general theory to the case of Gaussian mixture models in Section 5 and compare it with analysis on real data in Section 6.

1.1 Notations and conventions

For $k \in \mathbb{N}$, we define the set $[k] := \{1, 2, \dots, k\}$. We typically use lower case non-bold letters for scalars (m, n, j), and bold for vectors and matrices ($\mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{A}, \mathbf{B}, \mathbf{C}$). We use $\|\mathbf{v}\|$ to denote the Euclidean norm of a vector \mathbf{v} , and $\|\mathbf{M}\|_F$ to denote the Frobenius norm of a matrix \mathbf{M} . For

$\{c_n\}_{n \in \mathbb{N}_+}, \{d_n\}_{n \in \mathbb{N}_+} \subseteq \mathbb{R}_+$, we say $c_n \gg d_n$ if and only if $c_n/d_n \rightarrow \infty$, and for $\{e_n\}_{n \in \mathbb{N}_+} \subseteq \mathbb{R}$, we say $e_n = o_n(1)$ if and only if $e_n \rightarrow 0$ as $n \rightarrow \infty$. We denote by p-lim convergence in probability.

For $k \in \mathbb{N}_+$, we denote by S_k^+ the set of positive semi-definite matrices in $\mathbb{R}^{k \times k}$, and denote by $\mathcal{O}(k)$ the set of orthogonal matrices in $\mathbb{R}^{k \times k}$. For $\mathbf{M} \in S_k^+$, we let $\mathbf{M}^{1/2} \in S_k^+$ be any positive semi-definite matrix such that $\mathbf{M} = \mathbf{M}^{1/2} \mathbf{M}^{1/2}$.

We denote the i -th row of the factors $\mathbf{\Lambda}, \mathbf{\Theta}$ by $\mathbf{\Lambda}_i$ and $\mathbf{\Theta}_i$, respectively. We use $\mathbf{\Lambda}_0$ and $\mathbf{\Theta}_0$ to represent length- r random vectors drawn from probability distributions μ_{Λ} and μ_{Θ} . We sometimes need to write the posterior distribution of $(\mathbf{\Lambda}, \mathbf{\Theta})$ given particular observations. In this case, we use the lower case letters $\boldsymbol{\theta}, \boldsymbol{\lambda}, \boldsymbol{\lambda}_i, \boldsymbol{\theta}_i$ to represent variables corresponding to $(\mathbf{\Theta}, \mathbf{\Lambda}, \mathbf{\Lambda}_i, \mathbf{\Theta}_i)$ in the posterior distribution.

Throughout the paper, we use capital letter C to represent various numerical constants.

2 Further related work

As mentioned in the introduction, most earlier work deriving sharp asymptotics results focuses on the proportional regime $n \asymp d$, $s_n = n^{-1/2}$. In particular, [LKZ15] first obtained the limiting expression for Bayesian mean square error using non-rigorous tools from statistical mechanics. The conjectured expression was rigorously justified for special distributions $\mu_{\Lambda}, \mu_{\Theta}$ in [DM14a, DAM17]. However, the proof technique of [DM14a, DAM17] relies on the fact that approximate message passing (AMP) algorithm achieves Bayes optimality and does not apply to the general case.

Several groups developed rigorous approaches to prove the asymptotic formulas in increasing degrees of generality: spatial coupling [DMK⁺16]; the cavity method [LM19, Mio17, EAK18]; adaptive interpolation [BM19]; partial differential equation techniques [DM22].

A different line of research uses the second moment method to derive upper and lower bounds on the information-theoretic thresholds [BMV⁺18, PWB16, PWB20] for partial or exact recovery. This approach typically yields non-asymptotic bounds, under a broader class of settings but the results only determine such thresholds up to undetermined multiplicative constants. In contrast, here we attempt to obtain a characterization that is accurate up to $(1 + o_n(1))$ factors.

From a computational viewpoint, AMP-based algorithms can be shown to achieve the Bayesian error for a large region of parameters [BM11, MV21b]. One appealing fact about the AMP is that its high-dimensional behavior can be sharply characterized by *state evolution*.

Minimax guarantees were obtained by a number of groups for special cases of the low-rank model (1). Sparse PCA and Gaussian mixtures are arguably the most studied models in the literature, see e.g., [Roy17, GV19, Nda22, BR13a, CLR17]. These works often yield characterizations that hold up to usually a constant or logarithmic multiplicative gap.

Gaussian mixture models (GMM) provide a useful context for evaluating and comparing various clustering algorithms. We will use it here to illustrate the applicability of our general results. The goal can be either estimating the centers [Das99, DS07, KB16, MVW17, RV17], or recovering the underlying cluster assignments [VW04, AM05, BV08, KK10, AS12, FC18]. As we will see, in the high-dimensional weak signal regime, the cluster centers cannot be estimated, but the cluster assignments can be estimated with non-trivial accuracy.

Several algorithms were studied in detail for clustering under GMM, including semi-definite programming (SDP) [PW07, ABC⁺15, FC18, IMPV15, LLL⁺20], iterative algorithms with spectral initialization [AM05, VW04, KK10, AS12, LZ16], the method of moments [Pea94, FSO06, KMV10, MV10, HK13, BS15, HP15], and EM-based algorithms [DS07, BWY⁺17, JZB⁺16, CMZ⁺19].

Finally, in concurrent work, Donoho and Feldman recently characterized the accuracy of eigenvalue shrinkage methods in the spiked model with diverging aspect ratio [Fel21, DF22].

3 Strong signal regime

We first consider the strong signal regime in which we set $s_n = 1/\sqrt{n}$, and therefore we have

$$\mathbf{A} = \frac{1}{\sqrt{n}} \mathbf{\Lambda} \mathbf{\Theta}^\top + \mathbf{Z} \in \mathbb{R}^{n \times d}. \quad (4)$$

We define $\mathbf{Q}_\Lambda := \mathbb{E}_{\mathbf{\Lambda}_0 \sim \mu_\Lambda} [\mathbf{\Lambda}_0 \mathbf{\Lambda}_0^\top] \in S_r^+$ and $\mathbf{Q}_\Theta := \mathbb{E}_{\mathbf{\Theta}_0 \sim \mu_\Theta} [\mathbf{\Theta}_0 \mathbf{\Theta}_0^\top] \in S_r^+$. Before we proceed, we establish the following conventions for the distributions μ_Λ, μ_Θ .

Remark 3.1. Without loss of generality we can and will assume that both \mathbf{Q}_Λ and \mathbf{Q}_Θ are invertible. Furthermore, we can assume that $\mathbf{Q}_\Theta = q_\Theta \mathbf{I}_r$ for some $q_\Theta \in \mathbb{R}_{>0}$.

More precisely, we next show that —given arbitrary probability distributions $(\mu_\Theta, \mu_\Lambda)$ — the conditions of Remark 3.1 can always be satisfied by a reparameterization.

For $\mathbf{\Lambda}_0 \sim \mu_\Lambda$ and $\mathbf{\Theta}_0 \sim \mu_\Theta$, if either $\mathbf{\Lambda}_0 \stackrel{a.s.}{=} 0$ or $\mathbf{\Theta}_0 \stackrel{a.s.}{=} 0$ then estimation becomes trivial. We can therefore assume that this is not the case. By eigendecomposition of \mathbf{Q}_Λ and \mathbf{Q}_Θ , there exist $0 < k_1, k_2 \leq r$ and non-random matrices $\mathbf{M}_1 \in \mathbb{R}^{r \times k_1}$, $\mathbf{M}_2 \in \mathbb{R}^{r \times k_2}$ with full column ranks such that $\mathbf{\Lambda}_0 = \mathbf{M}_1 \mathbf{\Lambda}'_0$, $\mathbf{\Theta}_0 = \mathbf{M}_2 \mathbf{\Theta}'_0$, and $\mathbb{E}[\mathbf{\Lambda}'_0 (\mathbf{\Lambda}'_0)^\top] = \mathbf{I}_{k_1}$, $\mathbb{E}[\mathbf{\Theta}'_0 (\mathbf{\Theta}'_0)^\top] = \mathbf{I}_{k_2}$.

Assume $\mathbf{M}_1^\top \mathbf{M}_2$ has rank $k_3 \leq \min(k_1, k_2)$, and let $\mathbf{M}_1^\top \mathbf{M}_2 = \mathbf{U} \mathbf{S} \mathbf{V}^\top$ be its singular value decomposition (SVD) with $\mathbf{U} \in \mathbb{R}^{k_1 \times k_3}$, $\mathbf{V} \in \mathbb{R}^{k_2 \times k_3}$ having orthonormal columns. We then set $\bar{\mathbf{\Lambda}} \in \mathbb{R}^{n \times k_3}$ a matrix with i.i.d. rows that are copies of $\mathbf{S} \mathbf{U}^\top \mathbf{\Lambda}'_0$ and $\bar{\mathbf{\Theta}} \in \mathbb{R}^{d \times k_3}$ a matrix with i.i.d. rows that are copies of $\mathbf{V}^\top \mathbf{\Theta}'_0$. We can then write $\mathbf{\Lambda} \mathbf{\Theta}^\top = \bar{\mathbf{\Lambda}} \bar{\mathbf{\Theta}}^\top$ and the latter satisfies the conditions of Remark 3.1.

Note that this argument shows that we could assume $q_\Theta = 1$ as well, but it is convenient to keep this as a free parameter.

3.1 Estimation of $\mathbf{\Lambda}$

We first consider estimation of $\mathbf{\Lambda}$. We will show that a simple spectral estimator provides a consistent estimate up to a rotation in the r -dimensional Euclidean space. Consistency in terms of vector mean square error is not guaranteed due to potential non-identifiability issues. We propose sufficient conditions on $(\mu_\Lambda, \mu_\Theta)$, that imply consistency in terms of vector mean square error as well.

Denote by $\hat{\mathbf{\Lambda}}_s \in \mathbb{R}^{n \times r}$ the matrix whose columns are the top r eigenvectors of $\mathbf{A} \mathbf{A}^\top$, normalized so that $\hat{\mathbf{\Lambda}}_s^\top \hat{\mathbf{\Lambda}}_s / n = \mathbf{I}_r$. Denote by $\mathbf{P}, \hat{\mathbf{P}} \in \mathcal{O}(n)$ the projection matrices onto the column spaces of $\mathbf{\Lambda}$ and $\hat{\mathbf{\Lambda}}_s$, respectively. We use the following distance between the two subspaces as estimation loss

$$L^{\sin}(\hat{\mathbf{\Lambda}}_s, \mathbf{\Lambda}) := \|\mathbf{P}(\mathbf{I} - \hat{\mathbf{P}})\|_{\text{op}} = \|\hat{\mathbf{P}}(\mathbf{I} - \mathbf{P})\|_{\text{op}} = \sin \alpha(\hat{\mathbf{\Lambda}}_s, \mathbf{\Lambda}), \quad (5)$$

where $\alpha(\hat{\mathbf{\Lambda}}_s, \mathbf{\Lambda})$ is the principal angle between the two column spaces.

Theorem 3.1. Assume μ_Λ, μ_Θ have finite non-singular second moments $\mathbf{Q}_\Lambda, \mathbf{Q}_\Theta$ with $\mathbf{Q}_\Theta = q_\Theta \mathbf{I}_r$ (with no loss of generality per Remark 3.1). If $n, d \rightarrow \infty$ with $d/n \rightarrow \infty$, then under the model of Eq. (4):

1. $L^{\sin}(\hat{\mathbf{\Lambda}}_s, \mathbf{\Lambda}) \xrightarrow{P} 0$.
2. If we further assume that for some $\varepsilon > 0$ we have $\mathbb{E}_{\mathbf{\Lambda}_0 \sim \mu_\Lambda} [\|\mathbf{\Lambda}_0\|^{4+\varepsilon}] < \infty$, then there exists an estimator $\hat{\mathbf{L}} : \mathbf{A} \mapsto \hat{\mathbf{L}}(\mathbf{A}) \in \mathbb{R}^{n \times n}$, such that $\mathbb{E}[\|\hat{\mathbf{L}}(\mathbf{A}) - \mathbf{\Lambda} \mathbf{\Lambda}^\top\|_F^2] / n^2 \rightarrow 0$ as $n, d \rightarrow \infty$.
3. Let $\mathbf{\Lambda}_0 \sim \mu_\Lambda$. If we further assume that there does not exist $\mathbf{\Omega} \in \mathcal{O}(r)$, such that $\mathbf{\Omega} \neq \mathbf{I}_r$ and $\mathbf{\Omega} \mathbf{\Lambda}_0 \stackrel{d}{=} \mathbf{\Lambda}_0$, then there exists $\hat{\mathbf{\Lambda}} : \mathbf{A} \mapsto \hat{\mathbf{\Lambda}}(\mathbf{A}) \in \mathbb{R}^{n \times r}$, such that $\mathbb{E}[\|\hat{\mathbf{\Lambda}}(\mathbf{A}) - \mathbf{\Lambda}\|_F^2] / n \rightarrow 0$ as $n, d \rightarrow \infty$.

We delay the proof of Theorem 3.1 to Appendix C.1.

3.2 Estimation of Θ

Next, we turn to the estimation of Θ . According to Theorem 3.1, Λ can be estimated consistently under identifiability conditions. Therefore, a reasonable first step is to study the case in which Λ is given. This yields a lower bound on the Bayesian error of the original problem. We will see that this lower bound can be achieved asymptotically even if Λ must be estimated.

We can explicitly write the conditional distribution of Θ given (Λ, \mathbf{A}) . Using the Gaussian density formula, we see that for all $j \in [d]$, the posterior distribution of Θ_j is

$$p(d\theta_j | \Lambda, \mathbf{A}) \propto \exp \left(-\frac{1}{2n} \sum_{i=1}^n \langle \Lambda_i, \theta_j \rangle^2 + \frac{1}{\sqrt{n}} \sum_{i=1}^n A_{ij} \langle \Lambda_i, \theta_j \rangle \right) \mu_\Theta(d\theta_j). \quad (6)$$

Eq. (6) leads to the following asymptotic lower bound:

Theorem 3.2. *Consider the strong signal model of Eq. (4), assuming, without loss of generality, the setting of Remark 3.1. We let $n, d \rightarrow \infty$ simultaneously with $d/n \rightarrow \infty$, then for any estimator $\hat{\Theta} : \mathbf{A} \mapsto \hat{\Theta}(\mathbf{A}) \in \mathbb{R}^{d \times r}$, we have*

$$\liminf_{n, d \rightarrow \infty} \frac{1}{d} \mathbb{E} \left[\|\hat{\Theta}(\mathbf{A}) - \Theta\|_F^2 \right] \geq r q_\Theta - \mathbb{E} \left[\|\mathbb{E}[\Theta_0 | Q_\Lambda^{1/2} \Theta_0 + \mathbf{G}]\|^2 \right], \quad (7)$$

where $\mathbf{G} \sim \mathcal{N}(0, \mathbf{I}_r)$, $\Theta_0 \sim \mu_\Theta$ are mutually independent. Notice that the right hand side of Eq. (7) is independent of (n, d) .

If we further assume $\mathbb{E}[\|\Theta_0\|^4] < \infty$, then for any $\hat{M} : \mathbf{A} \mapsto \hat{M}(\mathbf{A}) \in \mathbb{R}^{d \times d}$, we have

$$\liminf_{n, d \rightarrow \infty} \frac{1}{d^2} \mathbb{E} \left[\|\hat{M}(\mathbf{A}) - \Theta \Theta^\top\|_F^2 \right] \geq r q_\Theta^2 - \left\| \mathbb{E} \left[\mathbb{E}[\Theta_0 | Q_\Lambda^{1/2} \Theta_0 + \mathbf{G}] \mathbb{E}[\Theta_0 | Q_\Lambda^{1/2} \Theta_0 + \mathbf{G}]^\top \right] \right\|_F^2. \quad (8)$$

We postpone the proof of Theorem 3.2 to Appendix C.2. Next, we show that the lower bound proposed in Theorem 3.2 can be achieved under identifiability conditions.

Theorem 3.3. *Under the conditions of Theorem 3.1, claim 3, there exist estimators $\hat{\Theta} : \mathbf{A} \mapsto \hat{\Theta}(\mathbf{A})$ and $\hat{M} : \mathbf{A} \mapsto \hat{M}(\mathbf{A})$, such that*

$$\begin{aligned} \lim_{n, d \rightarrow \infty} \frac{1}{d} \mathbb{E} \left[\|\hat{\Theta}(\mathbf{A}) - \Theta\|_F^2 \right] &= r q_\Theta - \mathbb{E} \left[\|\mathbb{E}[\Theta_0 | Q_\Lambda^{1/2} \Theta_0 + \mathbf{G}]\|^2 \right], \\ \lim_{n, d \rightarrow \infty} \frac{1}{d^2} \mathbb{E} \left[\|\hat{M}(\mathbf{A}) - \Theta \Theta^\top\|_F^2 \right] &= r q_\Theta^2 - \left\| \mathbb{E} \left[\mathbb{E}[\Theta_0 | Q_\Lambda^{1/2} \Theta_0 + \mathbf{G}] \mathbb{E}[\Theta_0 | Q_\Lambda^{1/2} \Theta_0 + \mathbf{G}]^\top \right] \right\|_F^2. \end{aligned}$$

We defer the proof of Theorem 3.3 to Appendix C.3. Theorems 3.2 and 3.3 together complete the analysis for the estimation of Θ in the strong signal regime.

4 Weak signal regime

In this section, we consider the weak signal regime where $s_n = 1/\sqrt[4]{nd}$. Thus the model of interest is

$$\mathbf{A} = \frac{1}{\sqrt[4]{nd}} \Lambda \Theta^\top + \mathbf{Z} \in \mathbb{R}^{n \times d}. \quad (9)$$

For convenience, we define $r_n := \sqrt[4]{d/n}$. By assumption, we see that $r_n \rightarrow \infty$ as $n, d \rightarrow \infty$.

4.1 Background: the symmetric spiked model

As mentioned in the introduction, our main technical result is that, in the weak signal regime, estimation under model (9) is equivalent to estimation under a symmetric spiked model. Under this model we observe $\mathbf{Y} \in \mathbb{R}^{n \times n}$ given by

$$\mathbf{Y} = \frac{q_\Theta}{n} \mathbf{\Lambda} \mathbf{\Lambda}^\top + \mathbf{W}, \quad (10)$$

where $\mathbf{W} \stackrel{d}{=} \text{GOE}(n)$, and $\mathbf{\Lambda}_i \stackrel{iid}{\sim} \mu_\Lambda$, independent of each other. We view $q_\Theta > 0$ as a signal-to-noise ratio parameter.

We denote the Bayesian MMSE of model (10) by

$$\text{MMSE}_n^{\text{symm}}(\mu_\Lambda; q_\Theta) := \min_{\hat{\mathbf{M}}(\cdot)} \frac{1}{n^2} \mathbb{E} \left[\left\| \hat{\mathbf{M}}(\mathbf{Y}) - \mathbf{\Lambda} \mathbf{\Lambda}^\top \right\|_F^2 \right]. \quad (11)$$

Note that the Bayesian MMSE is achieved by the posterior expectation $\hat{\mathbf{M}}(\mathbf{Y}) = \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top | \mathbf{Y}]$. We also define the normalized mutual information

$$\mathcal{I}_n^{\text{symm}}(\mu_\Lambda; q_\Theta) := \frac{1}{n} \mathbb{E} \log \frac{d\mathbb{P}_{\mathbf{\Lambda}, \mathbf{Y}}}{d(\mathbb{P}_\Lambda \times \mathbb{P}_\mathbf{Y})}(\mathbf{\Lambda}, \mathbf{Y}). \quad (12)$$

A significant amount of rigorous information is available about this model. For $s > 0$ and $\mathbf{Q} \in S_r^+$, we define the free energy functional $\mathcal{F}(s, \mathbf{Q})$ and its maximizer $\mathbf{Q}^*(s) \in S_r^+$ via

$$\mathcal{F}(s, \mathbf{Q}) := -\frac{s}{4} \|\mathbf{Q}\|_F^2 + \mathbb{E} \left\{ \log \left(\int \exp(\sqrt{s} \mathbf{z}^\top \mathbf{Q}^{1/2} \boldsymbol{\lambda} + s \boldsymbol{\lambda}^\top \mathbf{Q} \boldsymbol{\Lambda}_0 - \frac{s}{2} \boldsymbol{\lambda}^\top \mathbf{Q} \boldsymbol{\lambda}) \mu_\Lambda(d\boldsymbol{\lambda}) \right) \right\}. \quad (13)$$

$$\mathbf{Q}^*(s) \in \arg\max_{\mathbf{Q} \in S_r^+} \mathcal{F}(s, \mathbf{Q}). \quad (14)$$

In the above expression, expectation is taken over $\boldsymbol{\Lambda}_0 \sim \mu_\Lambda$ and $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I}_r)$ independent of each other. These functionals are directly related to the mutual information and the Bayes MMSE of model (10), as stated below.

Theorem 4.1 ([LM19], Corollary 42, Proposition 43). *There exists a deterministic countable set $\mathcal{D} \subseteq \mathbb{R}_{\geq 0}$ such that*

$$\begin{aligned} q_\Theta \in \mathbb{R}_{\geq 0} &\Rightarrow \lim_{n \rightarrow \infty} \mathcal{I}_n^{\text{symm}}(\mu_\Lambda; q_\Theta) = \frac{1}{4} q_\Theta^2 \|\mathbb{E}_{\boldsymbol{\Lambda}_0 \sim \mu_\Lambda} [\boldsymbol{\Lambda}_0 \boldsymbol{\Lambda}_0^\top]\|_F^2 - \sup_{\mathbf{Q} \in S_r^+} \mathcal{F}(q_\Theta^2, \mathbf{Q}), \\ q_\Theta \in \mathbb{R}_{\geq 0} \setminus \mathcal{D} &\Rightarrow \lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_\Lambda; q_\Theta) = \|\mathbb{E}_{\boldsymbol{\Lambda}_0 \sim \mu_\Lambda} [\boldsymbol{\Lambda}_0 \boldsymbol{\Lambda}_0^\top]\|_F^2 - \|\mathbf{Q}^*(s)\|_F^2. \end{aligned}$$

4.2 Estimation of Θ

We first consider estimation of Θ . We claim that in this case no estimator outperforms a naive one.

Theorem 4.2. *Consider the weak signal model of Eq. (9), assuming, without loss of generality, the setting of Remark 3.1. Let $n, d \rightarrow \infty$ simultaneously with $d/n \rightarrow \infty$. Then for any estimator $\hat{\Theta} : \mathbf{A} \mapsto \hat{\Theta}(\mathbf{A}) \in \mathbb{R}^{d \times r}$, we have*

$$\liminf_{n, d \rightarrow \infty} \frac{1}{d} \mathbb{E} [\|\hat{\Theta}(\mathbf{A}) - \Theta\|_F^2] \geq r q_\Theta - \|\mathbb{E}_{\Theta_0 \sim \mu_\Theta} [\Theta_0]\|^2.$$

If we further assume μ_Θ has bounded fourth moment, then for any $\hat{\mathbf{M}} : \mathbf{A} \mapsto \hat{\mathbf{M}}(\mathbf{A}) \in \mathbb{R}^{d \times d}$, we have

$$\liminf_{n,d \rightarrow \infty} \frac{1}{d^2} \mathbb{E}[\|\hat{\mathbf{M}}(\mathbf{A}) - \boldsymbol{\Theta}\boldsymbol{\Theta}^\top\|_F^2] \geq r q_\Theta^2 - \|\mathbb{E}_{\boldsymbol{\Theta}_0 \sim \mu_\Theta}[\boldsymbol{\Theta}_0]\|^4.$$

Notice that the above lower bounds are achieved by the null estimators $\hat{\boldsymbol{\Theta}}(\mathbf{A}) = \mathbb{E}[\boldsymbol{\Theta}] \in \mathbb{R}^{d \times r}$ and $\hat{\mathbf{M}}(\mathbf{A}) = \mathbb{E}[\boldsymbol{\Theta}\boldsymbol{\Theta}^\top] \in \mathbb{R}^{d \times d}$.

The proof of this statement is similar to the one of Theorem 3.2. Namely, we will prove that the mean square error achieved by simply taking the prior mean asymptotically agrees with the Bayesian MMSE for an estimator that has access to \mathbf{A} as additional information. The argument is summarized in Appendix D.1.

4.3 Estimation of \mathbf{A}

We finally consider the technically most interesting case, namely the estimation of \mathbf{A} in the weak signal regime. For simplicity, we will restrict ourselves to studying the matrix mean square error:

$$\text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta) := \inf_{\hat{\mathbf{M}}(\cdot)} \frac{1}{n^2} \mathbb{E} \left[\left\| \mathbf{A}\mathbf{A}^\top - \hat{\mathbf{M}}(\mathbf{A}) \right\|_F^2 \right], \quad (15)$$

where the infimum is taken over all estimators (measurable functions) $\hat{\mathbf{M}} : \mathbf{A} \mapsto \hat{\mathbf{M}}(\mathbf{A}) \in \mathbb{R}^{n \times n}$. Of course $\text{MMSE}_n^{\text{asym}}$ depends on the distributions μ_Λ, μ_Θ .

In the rank-one case $r = 1$, if $\mathbb{E}_{\boldsymbol{\Theta}_0 \sim \mu_\Theta}[\boldsymbol{\Theta}_0] \neq 0$, then the naive estimator $r_n^{-1} \mathbf{A}\mathbf{y}$ with $\mathbf{y} = \mathbb{E}_{\mu_\Theta}[\boldsymbol{\Theta}_0]^{-1} \mathbf{1}_d / \sqrt{d}$ is consistent:

$$\frac{1}{n} \left\| r_n^{-1} \mathbf{A}\mathbf{y} - \mathbf{A} \right\|^2 \xrightarrow{P} 0.$$

In this case, a consistent estimate of $\mathbf{A}\mathbf{A}^\top$ naturally follows. Therefore, if $\boldsymbol{\Theta}_0$ has non-vanishing expectation, the estimation problem is significantly easier.

When $r \geq 2$, if μ_Θ has non-zero mean, the same construction leads to consistent estimation of the projection of \mathbf{A} onto the direction determined by $\mathbb{E}[\boldsymbol{\Theta}_0]$ (for $\boldsymbol{\Theta}_0 \sim \mu_\Theta$). Once this component is subtracted, the problem is effectively reduced to one in which μ_Θ has zero mean.

For the remainder of this section, we focus on the more challenging case $\mathbb{E}[\boldsymbol{\Theta}_0] = \mathbf{0}_r$. In addition, for technical reasons we will require μ_Θ to have vanishing third moment.

Assumption 4.1. We assume $\mathbb{E}[\boldsymbol{\Theta}_0] = \mathbf{0}_r$, $\mathbb{E}[\boldsymbol{\Theta}_0 \otimes \boldsymbol{\Theta}_0 \otimes \boldsymbol{\Theta}_0] = \mathbf{0}_{r \times r \times r}$, where \otimes denotes the tensor product. Furthermore, we assume that μ_Θ, μ_Λ are sub-Gaussian.

Our main results establish that, according to several criteria, estimation in the asymmetric model (9) with $n, d \rightarrow \infty, d/n \rightarrow \infty$ is equivalent to estimation in the symmetric spiked model (10).

Our first result on the relation between these models is in terms of mutual information.

Theorem 4.3. Define the mutual information per coordinate in asymmetric model of Eqs. (9), via

$$\mathbf{I}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta) := \frac{1}{n} \mathbb{E} \log \frac{d\mathbb{P}_{\mathbf{A}, \mathbf{A}}}{d(\mathbb{P}_\Lambda \times \mathbb{P}_\mathbf{A})}(\mathbf{A}, \mathbf{A}). \quad (16)$$

Further recall the definition of mutual information in the symmetric model (10) given by Eq. (12). Within the setting of Assumption 4.1, we let $n, d \rightarrow \infty$ simultaneously with $d/n \rightarrow \infty$. In addition, we require without loss of generality $\mathbf{Q}_\Theta = q_\Theta \mathbf{I}_r$, cf. Remark 3.1. Then the following limits exist and are equal

$$\lim_{n,d \rightarrow \infty} \mathbf{I}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta) = \lim_{n \rightarrow \infty} \mathbf{I}_n^{\text{symm}}(\mu_\Lambda; q_\Theta).$$

The proof of Theorem 4.3 is presented in Appendix D.2 for μ_Λ with bounded support. The generalization to μ_Λ with unbounded support is discussed in Appendix D.4.

As mentioned above, earlier work determined the asymptotics of the mutual information for the symmetric model $I_n^{\text{symm}}(\mu_\Lambda; q_\Theta)$. In particular, the next corollary follows directly from Theorem 4.3 and Theorem 4.1.

Corollary 4.1. *Recall that S_r^+ denotes the set of $r \times r$ positive semidefinite matrices, and $\mathcal{F} : \mathbb{R}_{\geq 0} \times S_r^+ \rightarrow \mathbb{R}$ is defined in Eq. (13). Under the conditions of Theorem 4.3, we have*

$$\begin{aligned} \lim_{n,d \rightarrow \infty} I_n^{\text{symm}}(\mu_\Lambda, \mu_\Theta) &= \frac{1}{4} q_\Theta^2 \|\mathbb{E}_{\mathbf{A}_0 \sim \mu_\Lambda} [\mathbf{A}_0 \mathbf{A}_0^\top]\|_F^2 - \mathcal{F}_*(q_\Theta) \\ &:= \frac{1}{4} q_\Theta^2 \|\mathbb{E}_{\mathbf{A}_0 \sim \mu_\Lambda} [\mathbf{A}_0 \mathbf{A}_0^\top]\|_F^2 - \sup_{\mathbf{Q} \in S_r^+} \mathcal{F}(q_\Theta^2, \mathbf{Q}). \end{aligned}$$

Recall the de Bruijn identity relating mutual information and minimum mean square error, see [Sta59, GSV05, DAM17]:

$$\frac{1}{4} \text{MMSE}_n^{\text{symm}}(\mu_\Lambda; \sqrt{s}) = \frac{d}{ds} I_n^{\text{symm}}(\mu_\Lambda; \sqrt{s}). \quad (17)$$

Since $\text{MMSE}_n^{\text{symm}}(\mu_\Lambda; \sqrt{s})$ is non-increasing in s , the asymptotics of $I_n^{\text{symm}}(\mu_\Lambda; \sqrt{s})$ essentially determines the asymptotics of $\text{MMSE}_n^{\text{symm}}(\mu_\Lambda; \sqrt{s})$. Namely, we have $\lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_\Lambda; \sqrt{s}) = \|\mathbb{E}_{\mathbf{A}_0 \sim \mu_\Lambda} [\mathbf{A}_0 \mathbf{A}_0^\top]\|_F^2 - 4 \frac{\partial}{\partial s} \mathcal{F}_*(\sqrt{s})$ for almost all values of s .

It would be tempting to conclude that Theorem 4.3 and Corollary 4.1 lead directly to analogous theorems relating $\text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta)$ and $\text{MMSE}_n^{\text{symm}}(\mu_\Lambda; q_\Theta)$. Establishing such a consequence is more challenging than one would naively expect because we do not have an identity analogous¹ to Eq. (17) for the asymmetric model. We can nevertheless establish the following, via a perturbation argument.

Theorem 4.4. *Under the conditions of Theorem 4.3, for all but countably many values of $q_\Theta > 0$, we have*

$$\liminf_{n,d \rightarrow \infty} \text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta) \geq \lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_\Lambda; q_\Theta). \quad (18)$$

Further, consider a modified model in which the statistician observes $(\mathbf{A}, \mathbf{Y}'(\varepsilon))$, where \mathbf{A} is given by Eq. (9), and

$$\mathbf{Y}'(\varepsilon) := \frac{\sqrt{\varepsilon}}{n} \mathbf{A} \mathbf{A}^\top + \mathbf{W}', \quad \mathbf{W}' \sim \text{GOE}(n).$$

Here, we assume \mathbf{W}' is independent of everything else. Denote by $\text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta; \varepsilon)$ the corresponding matrix mean square error. Then, for all but countably many values of $q_\Theta > 0$, we have

$$\lim_{\varepsilon \rightarrow 0^+} \limsup_{n,d \rightarrow \infty} \text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta; \varepsilon) \leq \lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_\Lambda; q_\Theta). \quad (19)$$

The proof of Theorem 4.4 is outlined in Appendix D.3 (for distributions with bounded support) and D.4 (for the general case).

¹One could differentiate the mutual information $I_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta)$ with respect to the signal-to-noise ratio parameter q_Θ , but the result is related to error in estimating $\mathbf{A} \mathbf{A}^\top$ instead of $\mathbf{A} \mathbf{A}^\top$.

Remark 4.1. In particular, Theorem 4.4 establishes that the estimation errors under the symmetric and asymmetric models coincide asymptotically, provided that the error in the perturbed model $\text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta; \varepsilon)$ is uniformly continuous (in n) as $\varepsilon \downarrow 0$. We expect this to be generically the case, but proving this remains an open problem.

The next theorem establishes a sequence of sufficient conditions under which we can prove asymptotic equivalence of estimation errors in the asymmetric and symmetric models.

Theorem 4.5. *Under the conditions of Theorem 4.3, we further assume at least one of the following conditions holds:*

- (a) $dn^{-3}(\log n)^{-6} \rightarrow \infty$.
- (b) $d(\log d)^{8/5}/n^{6/5} \rightarrow 0$ and μ_Λ has bounded support.
- (c) For the case $r = 1$, define $Y = \sqrt{\gamma}\Lambda_0 + G$ with $G \sim \mathcal{N}(0, 1)$ independent of $\Lambda_0 \sim \mu_\Lambda$, and define $l(\gamma) = \mathbb{E} \log \frac{d_{p_Y|\Lambda_0}}{d_{p_Y}}(Y, \Lambda_0)$. Let

$$\Psi(\gamma, s) = \frac{s^2}{4} + \frac{\gamma^2}{4s} - \frac{\gamma}{2} + l(\gamma).$$

Assume that the global maximum of $\gamma \mapsto \Psi(\gamma, q_\Theta)$ over $(0, \infty)$ is also the first stationary point of the same function.

Then, we have

$$\lim_{n, d \rightarrow \infty} \text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta) = \lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_\Lambda; q_\Theta). \quad (20)$$

(For condition (b), the conclusion is guaranteed to hold for all but countably many values of $q_\Theta > 0$.)

We defer the proof of Theorem 4.5 to Appendix D.5.

As anticipated in the introduction, the results presented in Section 4.2 and Section 4.3 support two key statistical insights, which we next summarize:

1. In the weak signal regime, it is possible to partially recover $\mathbf{\Lambda}$ while impossible to recover $\mathbf{\Theta}$ in any non-trivial sense. For instance, in the high-dimensional Gaussian mixture model, we might be able to estimate the labels, even if it is impossible to estimate the cluster centers.

In the next section, we will further explore the application of these results to Gaussian mixture models, while in Section 6 we will investigate such asymmetry in real world datasets.

2. In this regime, ideal estimation accuracy is asymptotically independent of the distribution of the high-dimensional factor $\mathbf{\Theta}$. As demonstrated, for instance, by Eq. (20), the only dependence on μ_Θ is through its second moment.

The three sufficient conditions given in Theorem 4.5 correspond to three different arguments. The most straightforward case is the one of condition (a). We use the fact that

$$\frac{1}{\sqrt{nd}}(\mathbf{A}\mathbf{A}^\top - d\mathbf{I}_n) = \frac{1}{n}\mathbf{\Lambda}\hat{\mathbf{Q}}_\Theta\mathbf{\Lambda}^\top + \frac{1}{\sqrt{nd}}(\mathbf{Z}\mathbf{Z}^\top - d\mathbf{I}_n) + \text{cross terms}, \quad (21)$$

where $\hat{\mathbf{Q}}_\Theta := \mathbf{\Theta}^\top \mathbf{\Theta} / d \approx q_\Theta \mathbf{I}_r$. For $d \gg n^3$, [BDER16] proved that the total variation distance between the distribution of the Wishart matrix $(\mathbf{Z}\mathbf{Z}^\top - d\mathbf{I}_n)/\sqrt{nd}$ and the one of $\mathbf{W} \sim \text{GOE}(n)$

converges to 0. While we still have to deal with the cross terms, under this condition the two models are close to each other.

For $d \ll n^3$ the Wishart and GOE distributions are asymptotically mutually singular [BDER16], and therefore proving asymptotic equality of the mean square error has to rely on a more carefully analysis. In fact, the proof of part (b) follows a different path and relies heavily on Theorem 4.4.

Finally, part (c) combines the bound of Theorem 4.4 with a matching bound that is based on the analysis of a Bayesian approximate message passing (AMP) algorithm [MV21b]. Indeed, the sufficient condition of part (c) coincides with the condition that Bayes AMP achieves Bayes optimal estimation error.

5 Clustering under the Gaussian mixture model

As an application of our theory, we consider clustering under Gaussian mixture model (GMM). Throughout, we will assume that all Gaussian components have equal covariance Σ , and that Σ is known. Without loss of generality, we can therefore assume that data are preprocessed so that $\Sigma = \mathbf{I}_d$. We will focus on the weak signal regime, because it is mathematically the most interesting regime.

The Gaussian mixture model fits our general framework, with the Λ_i 's encoding the data point labels: Λ_i takes k possible values, with k being the number of clusters. We will measure estimation accuracy using the overlap

$$\text{Overlap}_n := \max_{\pi \in \mathfrak{S}_k} \frac{1}{n} \sum_{i=1}^n \mathbb{1} \left\{ \hat{\Lambda}_i = \Lambda_i^\pi \right\}.$$

Here, \mathfrak{S}_k denotes the group of permutations over k elements, and Λ_i^π denotes the action of this group on the cluster label encodings of the i -th sample. (We will work with slightly different encodings for the cases $k = 2$ and $k \geq 3$ below.)

5.1 Two clusters with symmetric centers

As a warm-up example, we consider the case of $k = 2$ clusters with equal weights. For $i \in \{1, \dots, n\}$, we observe an independent sample

$$\mathbf{a}_i \sim \frac{1}{2} \mathbf{N}(\Theta / \sqrt[4]{nd}, \mathbf{I}_d) + \frac{1}{2} \mathbf{N}(-\Theta / \sqrt[4]{nd}, \mathbf{I}_d). \quad (22)$$

Here, $\pm \Theta / \sqrt[4]{nd} \subseteq \mathbb{R}^d$ are the cluster centers. Denoting by $(\Lambda_i)_{i \leq n} \stackrel{iid}{\sim} \text{Unif}(\{-1, +1\})$ the cluster labels, and by $\mathbf{A} \in \mathbb{R}^{n \times d}$ the matrix whose i -th row corresponds to the i -th sample, we have that \mathbf{A} follows model (9) with $r = 1$.

We further assume Θ has independent coordinates: $(\Theta_j)_{j \leq d} \stackrel{iid}{\sim} \mu_\Theta$, where μ_Θ is a centered sub-Gaussian distribution and has zero third moment.

Remark 5.1. We note that, for $r = 1$, there is no real loss of generality in assuming $(\Theta_j)_{j \leq d}$ to be i.i.d. sub-Gaussian. Indeed, model (9) is equivariant under rotations $\Theta \mapsto \Omega \Theta$, $\mathbf{A} \mapsto \mathbf{A} \Omega^\top$, where $\Omega \in \mathbb{R}^{d \times d}$ is an orthogonal matrix. Further, any loss function that depends uniquely on Λ is also invariant under the same group. Consider minimax estimation when Θ belongs to the sphere: $\|\Theta\|_2^2 = dq_\Theta$. As a consequence of the Hunt-Stein theorem, the least favorable prior is the uniform distribution over the same sphere. We expect the asymptotic Bayes risk under this prior (and therefore the minimax risk) to be the same as the risk under the prior $(\Theta_j)_{j \leq d} \stackrel{iid}{\sim} \mathbf{N}(0, q_\Theta)$.

Proposition 5.1. *Consider the Gaussian mixture model as in Eq. (22). Assume $n, d \rightarrow \infty$ simultaneously and $d/n \rightarrow \infty$, then the following results hold:*

(a) *If $q_\Theta \leq 1$, then, for any clustering estimator $\hat{\Lambda}$, as $n, d \rightarrow \infty$ we have*

$$\text{Overlap}_n \xrightarrow{P} \frac{1}{2}. \quad (23)$$

(b) *If $q_\Theta > 1$, let s_* be the largest non-negative solution of*

$$s = q_\Theta^2 \mathbb{E} \left\{ \tanh(s + \sqrt{s}G)^2 \right\}, \quad (24)$$

where $G \sim \mathcal{N}(0, 1)$. Then $s_ > 0$ and there exists an estimator achieving*

$$\text{Overlap}_n \xrightarrow{P} \Phi(\sqrt{s_*}), \quad (25)$$

where Φ denotes the cumulative distribution function for standard Gaussian distribution.

The proof of this result uses the characterization of optimal estimation in the corresponding symmetric model proven in [DAM17], and we present the proof of point (a) in Appendix G.1. The overlap in point (b) can be achieved using orthogonal invariant Bayes AMP with spectral initialization on $(\mathbf{A}\mathbf{A}^\top - d\mathbf{I}_n)/\sqrt{nd}$. This algorithm is described and analyzed in [MV21a].

5.2 Two or more clusters with orthogonal centers

We next consider the case of $k \geq 2$ clusters with approximately orthogonal centers. We denote by $\{\Theta_{\cdot i}/\sqrt[4]{nd} : i \in [k]\} \subseteq \mathbb{R}^d$ the cluster centers. Let $\Theta \in \mathbb{R}^{d \times k}$ with the i -th column given by $\Theta_{\cdot i}$. For $j \in [d]$, we let $\Theta_j \in \mathbb{R}^k$ be the j -th row of Θ . We assume $\Theta_j \stackrel{iid}{\sim} \mu_\Theta$, where μ_Θ is sub-Gaussian with vanishing first and third moments and diagonal covariance: $\text{Cov}(\Theta_1) = q_\Theta \mathbf{I}_k$.

Let \mathbf{e}_j be the j -th standard basis vector in \mathbb{R}^k . We encode the data point labels by setting $\Lambda_i = \mathbf{e}_j$ if and only if the i -th sample belongs to the j -th cluster, and consider the case of equal proportions, so that $(\Lambda_i)_{i \leq n} \stackrel{iid}{\sim} \text{Unif}(\{\mathbf{e}_1, \dots, \mathbf{e}_k\})$.

As before, we let $\mathbf{A} \in \mathbb{R}^{n \times d}$ be the matrix whose rows are i.i.d. samples \mathbf{a}_i from the Gaussian mixture model with centers $\{\Theta_{\cdot i}/\sqrt[4]{nd} : i \in [k]\}$. With these definitions, the matrix \mathbf{A} is distributed according to model (9).

Remark 5.2. While we state our results for random Θ , we can generalize Remark 5.1 to the present setting. This argument implies that the results of this section also characterize the minimax estimation error over the class of problems with orthogonal centers $\Theta^\top \Theta = d q_\Theta \mathbf{I}_k$.

Our next theorem establishes the threshold for weak recovery of the cluster labels in the high-dimensional regime $d/n \rightarrow \infty$. Recall the function $\mathcal{F}(s, \mathbf{Q})$ is defined in Eq. (13), where we take $\mu_\Lambda = \sum_{i=1}^k \delta_{\mathbf{e}_i}/k$. We let $\mathbf{Q}_0 := \mathbf{1}_k \mathbf{1}_k^\top / k^2$ and define the threshold

$$q_\Theta^{\text{info}}(k) := \inf \left\{ q_\Theta \geq 0 : \sup_{\mathbf{Q} \in \mathcal{S}_k^+} \mathcal{F}(q_\Theta^2, \mathbf{Q}) > \mathcal{F}(q_\Theta^2, \mathbf{Q}_0) \right\}. \quad (26)$$

Theorem 5.1. *Consider the Gaussian mixture model with k components of equal weights, in the high-dimensional asymptotics $d, n \rightarrow \infty$, $d/n \rightarrow \infty$. Under the above assumptions on the centers Θ , the following results hold:*

k	$q_{\Theta}^{\text{info}}(k)$
2	2
3	3
4	4
5	4.95
6	5.81
7	6.61
8	7.36

Table 1: Information-theoretic thresholds $q_{\Theta}^{\text{info}}(k)$ for $2 \leq k \leq 8$.

- (a) If $q_{\Theta} < q_{\Theta}^{\text{info}}(k)$, then for any estimator $\hat{\mathbf{A}} : \mathbb{R}^{n \times d} \rightarrow \{\mathbf{e}_j : j \in [k]\}^n$ that is a measurable function of the input \mathbf{A} , we have

$$\text{p-lim}_{n,d \rightarrow \infty} \text{Overlap}_n = \frac{1}{k}.$$

- (b) Assume either $dn^{-3}(\log n)^{-6} \rightarrow \infty$ or $dn^{-6/5}(\log d)^{8/5} \rightarrow 0$. If $q_{\Theta} > q_{\Theta}^{\text{info}}(k)$, then there exists an estimator $\hat{\mathbf{A}} : \mathbb{R}^{n \times d} \rightarrow \{\mathbf{e}_j : j \in [k]\}^n$ that is a measurable function of the input \mathbf{A} , such that

$$\liminf_{n,d \rightarrow \infty} \mathbb{E}[\text{Overlap}_n] > \frac{1}{k}.$$

We defer the proofs of parts (a) and (b) of Theorem 5.1 to Appendices G.2 and G.3, respectively. Note that [BMV⁺18, Theorem 2] implies $q_{\Theta}^{\text{info}}(k) = 2\sqrt{k \log k} \cdot (1 + o_k(1))$ as $k \rightarrow \infty$ (however, [BMV⁺18] does not establish a sharp threshold). In contrast, Theorem 5.1 derives the exact threshold for every k .

Table 1 collects values for the thresholds $q_{\Theta}^{\text{info}}(k)$ for a few values of k , as obtained by numerically evaluating Eq. (26). For $k \leq 4$, this is expected to coincide with the spectral threshold, namely $q_{\Theta}^{\text{info}}(k) = k$ [LDBB⁺16]. (Notice that the apparent discrepancy with the threshold for $k = 2$ in the previous section is due to the different normalization adopted here.)

5.3 Numerical experiments

We present in this section numerical experiments suggesting that the theory of the last sections is already relevant at moderate values of d, n . We consider several clustering methods, and compare their performances with the threshold $q_{\Theta}^{\text{info}}(k)$.

For our experiment we use built-in functions in `Python3`, for the following clustering methods:

- (1) Lloyd’s algorithm, as implemented by the function `KMeans()` in the `scikit-learn` module with option `algorithm = “lloyd”`.
- (2) Agglomerative clustering, implemented by the function `AgglomerativeClustering()` in the `scikit-learn` module with default parameters.
- (3) EM algorithm, implemented by the function `GaussianMixture()` in the `scikit-learn` module with default parameters.
- (4) A semidefinite programming (SDP) relaxation described in [PW07]. We use the `cvxpy` module for the optimization steps.

In Figure 1 we present results for these algorithms for $n = 100$, $d = 2000$, and $\mu_{\Theta} = \mathcal{N}(0, q_{\Theta})$. For each value of the pair (k, q_{Θ}) , we run 100 independent trials, and plot the average overlap versus q_{Θ} . For the case $k = 2$, we consider two slightly different settings: “Symmetric=True” corresponds to the case of two centers symmetric around the origin, as in Section 5.1, and “Symmetric=False” corresponds to the case of two approximately orthogonal centers as per Section 5.2. We also report the threshold $q_{\Theta}^{\text{info}}(k)$, its large k approximation $2\sqrt{k \log k}$, and the algorithmic threshold $q_{\Theta}^{\text{algo}}(k) = k$ (this is the conjectured threshold for efficient recovery, which coincides with the spectral threshold [LDBB⁺16, MV21b]).

Despite the small sample size, we observe that $q_{\Theta}^{\text{info}}(k)$ appears to capture the onset of non-trivial clustering accuracy across multiple algorithms.

6 Asymmetry in factors estimation: real world datasets

Previous sections imply the existence of gaps in the estimation of Λ and Θ , in the high-dimensional asymptotics $d/n \rightarrow \infty$. In summary, in the strong signal regime, Λ can be estimated consistently up to a potential rotation, while Θ can only be partially recovered. On the other hand, in the weak signal regime, Λ can be partially recovered, while no estimator achieves better asymptotic performance than a naive one in terms of the estimation of Θ .

In this section, we investigate this asymmetry in real world datasets. We focus on a problem that can be modeled as clustering with $k = 2$ clusters (this can be modeled as a GMM model, leading to Eq. (2) with $r = 1$ as described in the previous section).

6.1 1000 Genomes Project

Our first experiment involves genotype data from the 1000 Genomes Project [C⁺15]. This provides genotypes for $n = 2,504$ individuals grouped in five population groups (corresponding to their geographic origins). For our experiments, we extract $d = 100,000$ common single-nucleotide polymorphisms (SNPs). Our preprocessing steps follow from [?]. After preprocessing, we add independent Gaussian noise with variance 5 to the data matrix, to make the problem more challenging.

Principal component analysis (PCA) is often used in genome-wide association studies, in particular to explore the genetic structure of human populations [NJB⁺08, NS08]. As a first step of our experiment, for each pair of population groups, we randomly extract 30 subjects from each group without replacement. The subsampled observations form a $60 \times 100,000$ genotype matrix, the columns of which are then centered and rescaled. We next run PCA on this subset, and plot the projections onto the top 2 principal components. We display one typical outcome of PCA in Figure 2. From the figures, we see that despite the high-dimensionality, PCA still reflects the underlying population structure. We interpret this as indicating that non-trivial clustering can be achieved on these data.

To further support our conclusion, we run K-means clustering on the subsampled datasets (using `sklearn.cluster.KMeans` in Python 3 with default parameters). We then compute the overlap between the true and estimated labels (in this example labels correspond to population groups). We repeat this procedure independently 1000 times on randomly selected subsets of the data. The outcomes are recorded and displayed in the lower triangle of Figure 3. From the figures, we see that K-means clustering estimates the labels significantly better than random guessing (i.e., better than 50% accuracy) and achieves near-perfect recovery for certain pairs of population groups.

We next estimate the cluster centers Θ , for each pair of population groups. We take two non-overlapping subsets of the data (each with size 60) and run K-means on each subset: this leads

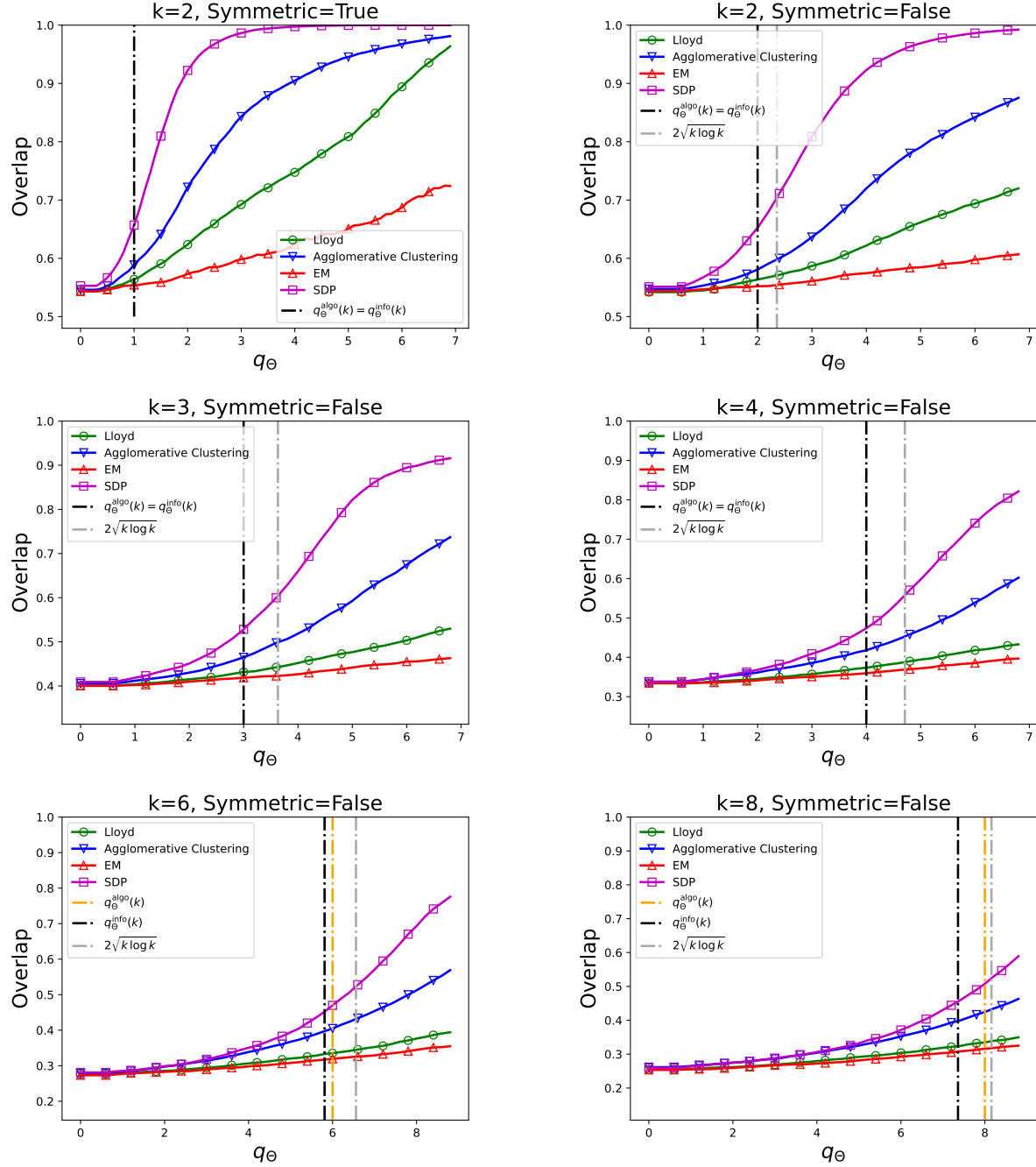


Figure 1: Average overlap achieved by several clustering algorithms on the Gaussian mixture model with $n = 100$ datapoints, $d = 2000$ dimensions, averaged over 100 instances. The black vertical line corresponds to the information-theoretic threshold for identifying clusters significantly better than random guessing; the orange vertical line corresponds to the spectral or algorithmic threshold; the grey vertical line corresponds to the approximated information-theoretic threshold $2\sqrt{k \log k}$.

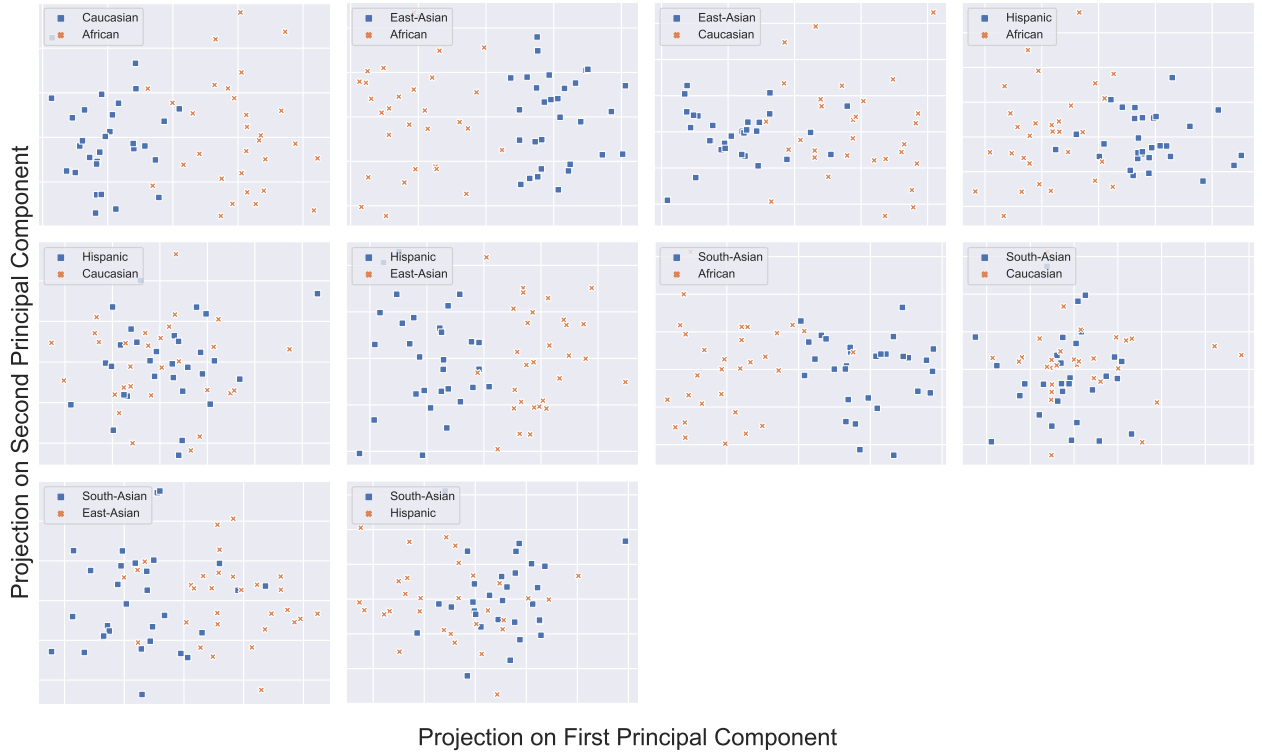


Figure 2: Illustration of PCA on a subset of 1000 Genomes Project data. In these plots, the x axis represents the projection onto the first principal component, and the y axis represents projection onto the second principal component. Point colors and shapes correspond to population groups. Each experiment involves 60 individuals in total, with 30 individuals from each of the two population groups.

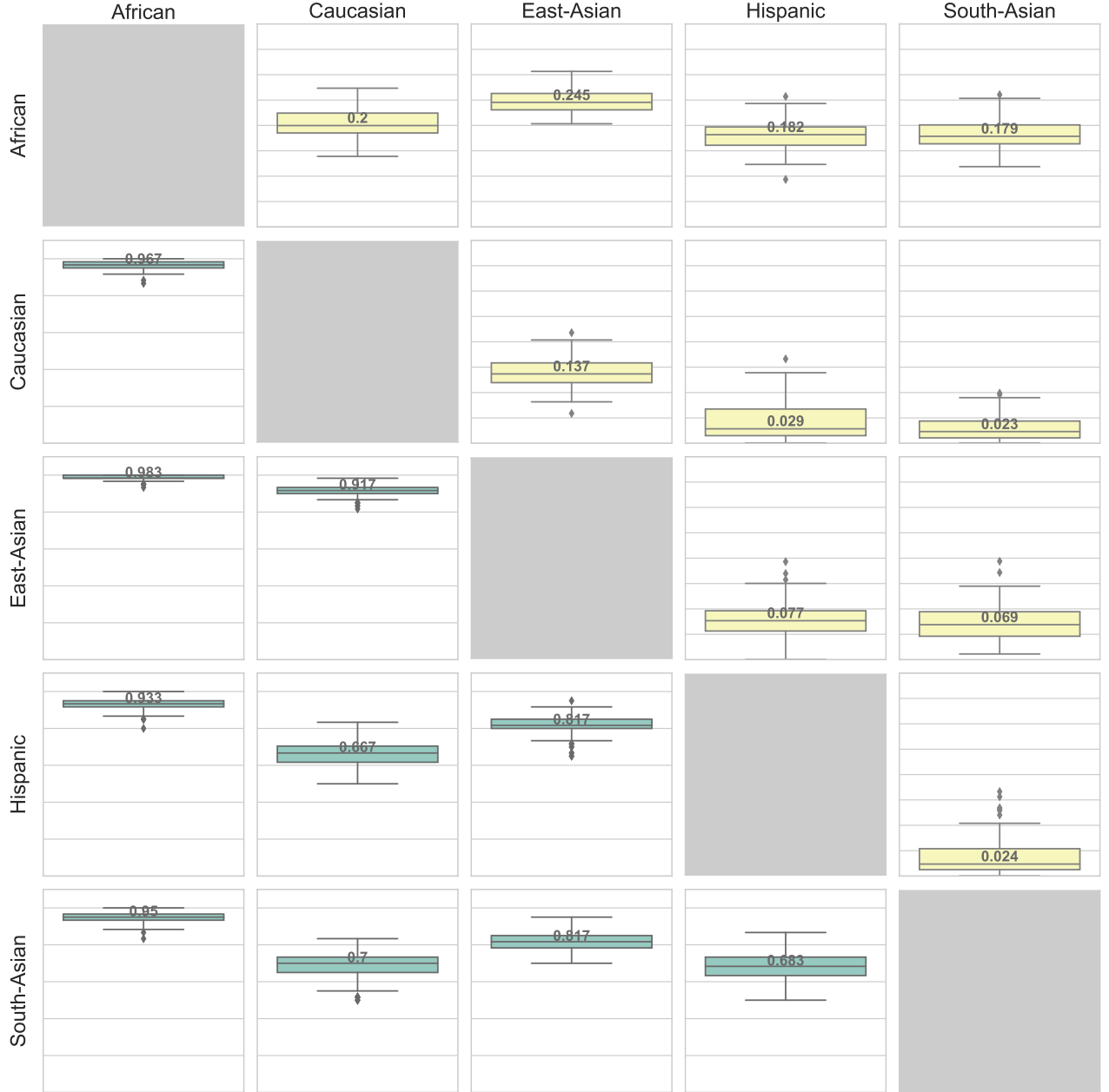


Figure 3: Simulation results derived from 1000 independent experiments for the 1000 Genomes Project dataset. The boxplots in the upper triangle display the quantiles of the normalized inner products between the estimated cluster centers. The boxplots in the lower triangle display the quantiles of overlaps between true labels and labels obtained via K-means clustering. We annotate the medians in the corresponding figures for readers' convenience.

to two distinct estimates of the cluster centers $(\hat{\Theta}_i^{(1)})_{i \in \{1,2\}}$ for data subset 1, and $(\hat{\Theta}_i^{(2)})_{i \in \{1,2\}}$ for data subset 2. We then compute the maximum normalized inner product

$$\max_{i,j \leq 2} |\langle \hat{\Theta}_i^{(1)}, \hat{\Theta}_j^{(2)} \rangle| / (\|\hat{\Theta}_i^{(1)}\|_2 \|\hat{\Theta}_j^{(2)}\|_2)$$

between the estimated cluster centers obtained via K-means from these two subsets of data.

This procedure is again repeated for 1000 times independently, and the distributions of the maximum normalized inner products are displayed in the upper triangle of Figure 3. We observe that, for several population pairs, the estimates $\hat{\Theta}^{(1)}$, $\hat{\Theta}^{(2)}$ are not significantly correlated (using initials, this is the case for the pairs C-H, C-SA, EA-H, EA-SA, H-SA). Since these estimates are obtained based on independent samples from the same population, we conclude that they are also not significantly correlated with the true centers. When this happens, the behavior of this clustering problem seems to be captured by the weak signal regime analyzed in the previous sections: clusters can be estimated in a non-trivial way, but cluster centers cannot be estimated.

For the other population pairs, the cluster centers estimates are correlated, and clustering accuracy is very high (this is the case for pairs A-C, A-EA, A-H, A-SA, with C-EA not as clear a case). This is analogous to what we observe in our model in the strong signal regime.

6.2 RNA-Seq gene expression

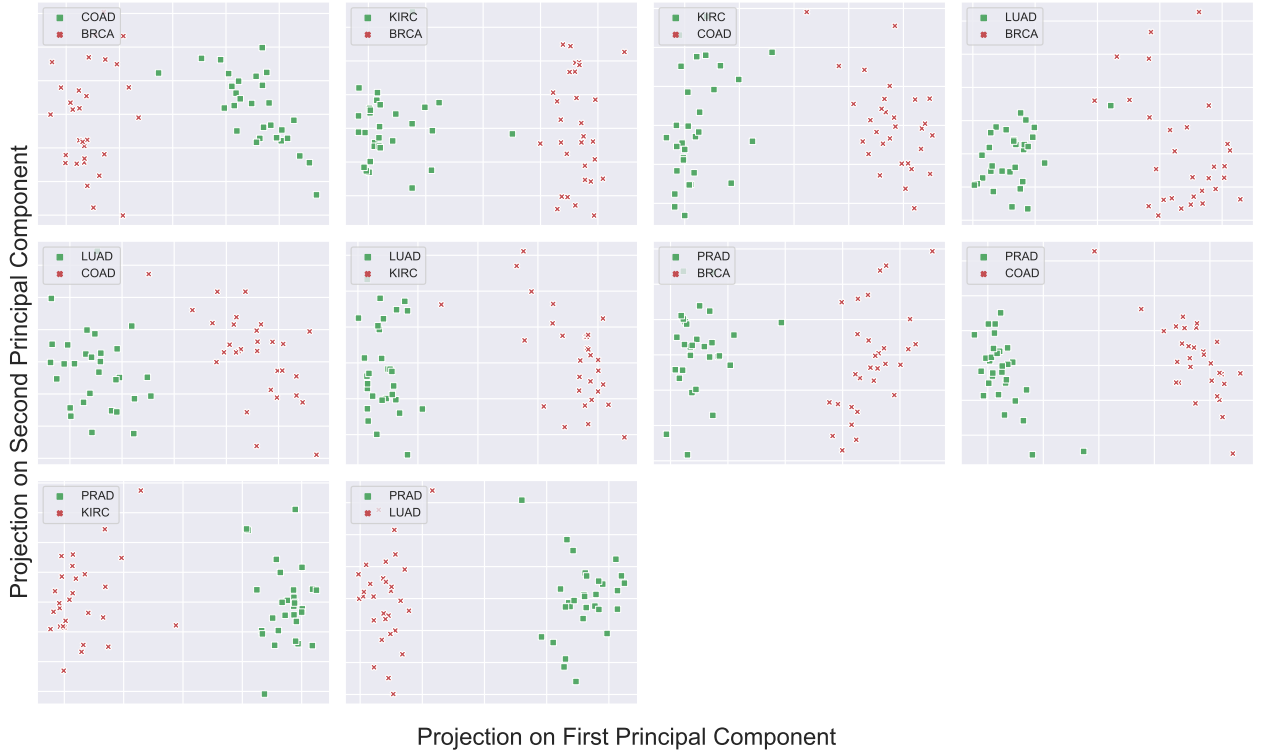


Figure 4: PCA on subsets of RNA-Seq gene expression data: each time we select 30 datapoints from each of the two cancer groups at random. Point colors and shapes stand for different cancer groups. We plot the projections of these datapoints onto the subspace defined by their first two principal components.

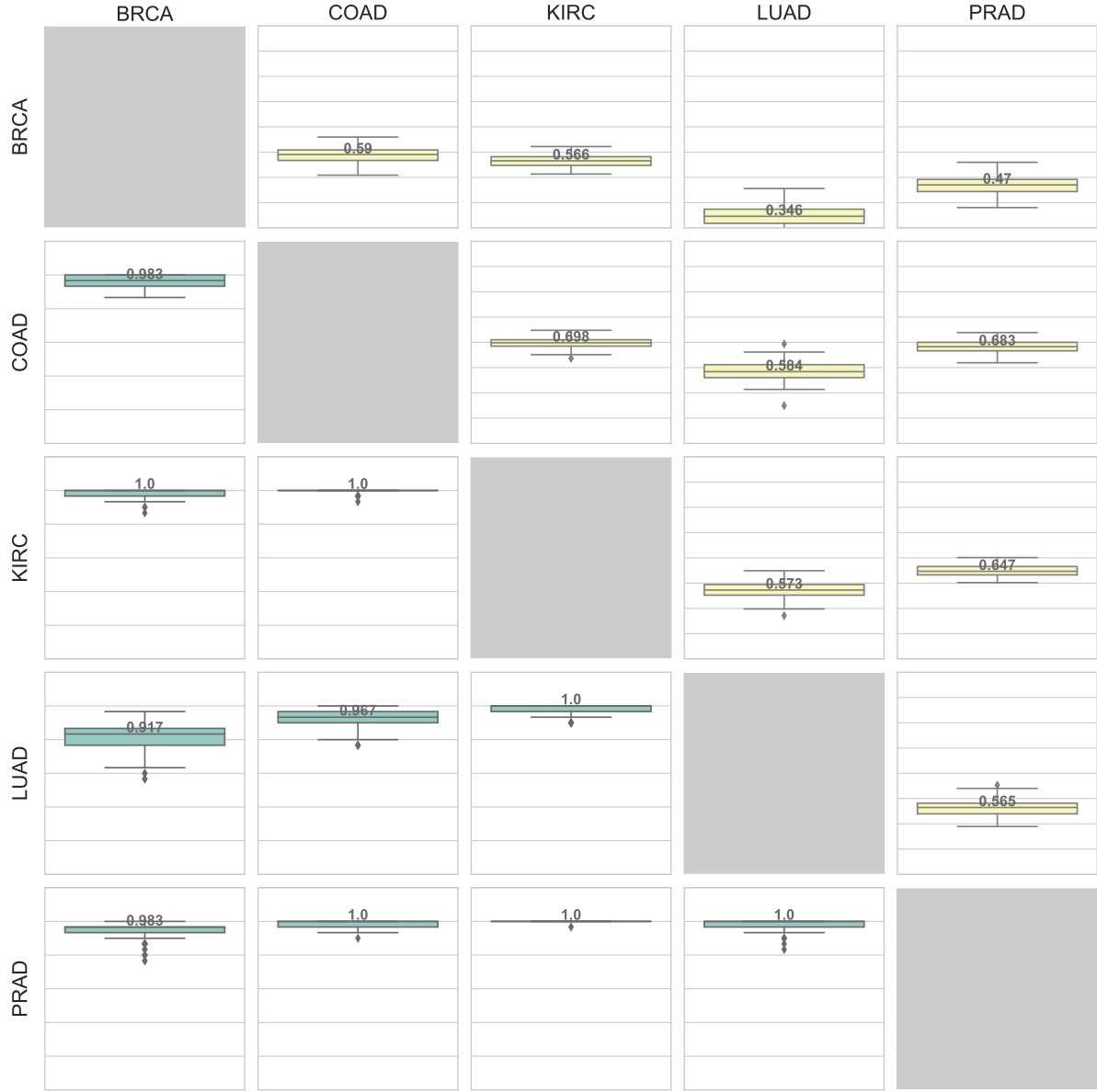


Figure 5: Simulation results derived from 1000 independent experiments for the UCI gene expression dataset. The boxplots in the upper triangle display the quantiles of the normalized inner products between estimated cluster centers. The boxplots in the lower triangle display the quantiles of the accuracy (overlap) in reconstructing the true clusters. Medians are annotated in the figures.

We carry out a similar experiment on gene expression data for different types of cancers from the UCI Machine Learning Repository² [DG17]. The dataset contains 801 samples and 20531 attributes, with the predictors being RNA-Seq gene expression levels measured by the Illumina HiSeq platform. Before proceeding, again we apply additive Gaussian noise to the data matrix, with mean zero and variance 5. We consider five different cancer types, denoted by “COAD”, “BRCA”, “KIRC”, “LUAD” and “PRAD”.

For each pair of cancer groups we subsample 30 subjects from each group, to construct a 60×20531 data matrix. We then center and rescale the columns of this matrix to unit norms. A typical outcome of PCA is presented in Figure 4. We observe that clusters corresponding to different cancer groups are well separated for each of the pairs. In Figure 5, we report the overlaps between the labels obtained from K-means clustering and the ground truth labels. The overlaps are very high for all pairs. These plots summarize the results of 1000 independent repetitions of this experiment.

In the upper half of the same figure, we present the maximum normalized inner products between the estimated cluster centers on two independent subsamples. The correlation is significantly different from zero, but far from being close to one. Once more, this is analogous to the strong signal regime in our analysis.

Acknowledgements

This work was supported by the NSF through award DMS-2031883, the Simons Foundation through Award 814639 for the Collaboration on the Theoretical Foundations of Deep Learning, the NSF grant CCF-2006489 and the ONR grant N00014-18-1-2729, and a grant from Eric and Wendy Schmidt at the Institute for Advanced Studies. Part of this work was carried out while Andrea Montanari was on partial leave from Stanford and a Chief Scientist at Ndata Inc dba Project N. The present research is unrelated to AM’s activity while on leave.

References

- [Abb17] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.
- [ABC⁺15] Pranjali Awasthi, Afonso S Bandeira, Moses Charikar, Ravishankar Krishnaswamy, Soledad Villar, and Rachel Ward. Relax, no need to round: Integrality of clustering formulations. In *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science*, pages 191–200, 2015.
- [AM05] Dimitris Achlioptas and Frank McSherry. On spectral learning of mixtures of distributions. In *International Conference on Computational Learning Theory*, pages 458–469. Springer, 2005.
- [AS12] Pranjali Awasthi and Or Sheffet. Improved spectral-norm bounds for clustering. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 37–49. Springer, 2012.
- [AW08] Arash A Amini and Martin J Wainwright. High-dimensional analysis of semidefinite relaxations for sparse principal components. In *2008 IEEE international symposium on information theory*, pages 2454–2458. IEEE, 2008.

²<https://archive.ics.uci.edu/ml/datasets/gene+expression+cancer+RNA-Seq>

- [BAP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.
- [BDER16] Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Z Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016.
- [BGN11] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low-rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.
- [BGN12] Florent Benaych-Georges and Raj Rao Nadakuditi. The singular values and vectors of low-rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012.
- [BM11] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.
- [BM19] Jean Barbier and Nicolas Macris. The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference. *Probability theory and related fields*, 174(3):1133–1185, 2019.
- [BMV⁺18] Jess Banks, Cristopher Moore, Roman Vershynin, Nicolas Verzelen, and Jiaming Xu. Information-theoretic bounds and phase transitions in clustering, sparse pca, and sub-matrix localization. *IEEE Transactions on Information Theory*, 64(7):4872–4894, 2018.
- [BR13a] Quentin Berthet and Philippe Rigollet. Computational lower bounds for sparse pca. *arXiv preprint arXiv:1304.0828*, 2013.
- [BR13b] Quentin Berthet and Philippe Rigollet. Optimal detection of sparse principal components in high dimension. *The Annals of Statistics*, 41(4):1780–1815, 2013.
- [BS06] Jinho Baik and Jack W Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of multivariate analysis*, 97(6):1382–1408, 2006.
- [BS15] Mikhail Belkin and Kaushik Sinha. Polynomial learning of distribution families. *SIAM Journal on Computing*, 44(4):889–911, 2015.
- [BV08] S Charles Brubaker and Santosh S Vempala. Isotropic pca and affine-invariant clustering. In *Building Bridges*, pages 241–281. Springer, 2008.
- [BVH⁺16] Afonso S Bandeira, Ramon Van Handel, et al. Sharp nonasymptotic bounds on the norm of random matrices with independent entries. *Annals of Probability*, 44(4):2479–2506, 2016.
- [BWY⁺17] Sivaraman Balakrishnan, Martin J Wainwright, Bin Yu, et al. Statistical guarantees for the em algorithm: From population to sample-based analysis. *Annals of Statistics*, 45(1):77–120, 2017.
- [BY88a] Zhi-Dong Bai and Yong-Qua Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a wigner matrix. *The Annals of Probability*, pages 1729–1741, 1988.

- [BY88b] Zhidong D Bai and Yong Q Yin. Convergence to the semicircle law. *The Annals of Probability*, pages 863–875, 1988.
- [C⁺15] 1000 Genomes Project Consortium et al. A global reference for human genetic variation. *Nature*, 526(7571):68, 2015.
- [CLR17] T Tony Cai, Tengyuan Liang, and Alexander Rakhlin. Computational and statistical boundaries for submatrix localization in a large noisy matrix. *The Annals of Statistics*, 45(4):1403–1430, 2017.
- [CMZ⁺19] T Tony Cai, Jing Ma, Linjun Zhang, et al. Chime: Clustering of high-dimensional gaussian mixtures with em algorithm and its optimality. *The Annals of Statistics*, 47(3):1234–1267, 2019.
- [DAM17] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the balanced binary stochastic block model. *Information and Inference: A Journal of the IMA*, 6(2):125–170, 2017.
- [Das99] Sanjoy Dasgupta. Learning mixtures of gaussians. In *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)*, pages 634–644. IEEE, 1999.
- [DF22] David L Donoho and Michael J Feldman. Optimal eigenvalue shrinkage in the semicircle limit. *arXiv:2210.04488*, 2022.
- [DG17] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.
- [DM14a] Yash Deshpande and Andrea Montanari. Information-theoretically optimal sparse pca. In *2014 IEEE International Symposium on Information Theory*, pages 2197–2201. IEEE, 2014.
- [DM14b] Yash Deshpande and Andrea Montanari. Sparse pca via covariance thresholding. *Advances in Neural Information Processing Systems*, 27, 2014.
- [DM22] Tomas Dominguez and Jean-Christophe Mourrat. Mutual information for the sparse stochastic block model. *arXiv preprint arXiv:2209.04513*, 2022.
- [DMK⁺16] Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, Lenka Zdeborová, et al. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. *Advances in Neural Information Processing Systems*, 29, 2016.
- [DS07] Sanjoy Dasgupta and Leonard J Schulman. A probabilistic analysis of em for mixtures of separated, spherical gaussians. *Journal of Machine Learning Research*, 8:203–226, 2007.
- [EAK18] Ahmed El Alaoui and Florent Krzakala. Estimation in the spiked wigner model: a short proof of the replica formula. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1874–1878. IEEE, 2018.
- [FC18] Yingjie Fei and Yudong Chen. Hidden integrality of sdp relaxations for sub-gaussian mixture models. In *Conference On Learning Theory*, pages 1931–1965. PMLR, 2018.
- [Fel21] Michael J Feldman. Spiked singular values and vectors under extreme aspect ratios. *arXiv:2104.15127*, 2021.

- [FSO06] Jon Feldman, Rocco A Servedio, and Ryan O'Donnell. Pac learning axis-aligned mixtures of gaussians with no separation assumption. In *International Conference on Computational Learning Theory*, pages 20–34. Springer, 2006.
- [GSV05] Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in gaussian channels. *IEEE transactions on information theory*, 51(4):1261–1282, 2005.
- [GV19] Christophe Giraud and Nicolas Verzelen. Partial recovery bounds for clustering with the relaxed k -means. *Mathematical Statistics and Learning*, 1(3):317–374, 2019.
- [HK13] Daniel Hsu and Sham M Kakade. Learning mixtures of spherical gaussians: moment methods and spectral decompositions. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science*, pages 11–20, 2013.
- [HP15] Moritz Hardt and Eric Price. Tight bounds for learning a mixture of two gaussians. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 753–760, 2015.
- [HR04] David C Hoyle and Magnus Rattay. Principal-component-analysis eigenvalue spectra from data with symmetry-breaking structure. *Physical Review E*, 69(2):026124, 2004.
- [HWX17] Bruce Hajek, Yihong Wu, and Jiaming Xu. Submatrix localization via message passing. *The Journal of Machine Learning Research*, 18(1):6817–6868, 2017.
- [IMPV15] Takayuki Iguchi, Dustin G Mixon, Jesse Peterson, and Soledad Villar. On the tightness of an sdp relaxation of k -means. *arXiv preprint arXiv:1505.04778*, 2015.
- [JL09] Iain M Johnstone and Arthur Yu Lu. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104(486):682–693, 2009.
- [JNRS10] Michel Journée, Yurii Nesterov, Peter Richtárik, and Rodolphe Sepulchre. Generalized power method for sparse principal component analysis. *Journal of Machine Learning Research*, 11(2), 2010.
- [JO20] Iain M Johnstone and Alexei Onatski. Testing in high-dimensional spiked models. *The Annals of Statistics*, 48(3):1231–1254, 2020.
- [Joh01] Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, pages 295–327, 2001.
- [Joh06] Iain M Johnstone. High dimensional statistical inference and random matrices. *arXiv preprint math/0611589*, 2006.
- [JZB⁺16] Chi Jin, Yuchen Zhang, Sivaraman Balakrishnan, Martin J Wainwright, and Michael I Jordan. Local maxima in the likelihood of gaussian mixture models: Structural results and algorithmic consequences. *Advances in neural information processing systems*, 29, 2016.
- [Kar03] Nouredine El Karoui. On the largest eigenvalue of wishart matrices with identity covariance when n , p and p/n tend to infinity. *arXiv preprint math/0309355*, 2003.

- [KB16] Jason M Klusowski and WD Brinda. Statistical guarantees for estimating the centers of a two-component gaussian mixture by em. *arXiv preprint arXiv:1608.02280*, 2016.
- [KBRS11] Mladen Kolar, Sivaraman Balakrishnan, Alessandro Rinaldo, and Aarti Singh. Minimax localization of structural information in large noisy matrices. *Advances in Neural Information Processing Systems*, 24, 2011.
- [KK10] Amit Kumar and Ravindran Kannan. Clustering with spectral norm and the k-means algorithm. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 299–308. IEEE, 2010.
- [KMV10] Adam Tauman Kalai, Ankur Moitra, and Gregory Valiant. Efficiently learning mixtures of two gaussians. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pages 553–562, 2010.
- [LDBB⁺16] Thibault Lesieur, Caterina De Bacco, Jess Banks, Florent Krzakala, Cris Moore, and Lenka Zdeborová. Phase transitions and optimal algorithms in high-dimensional gaussian mixture clustering. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 601–608. IEEE, 2016.
- [LKZ15] Thibault Lesieur, Florent Krzakala, and Lenka 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)*, pages 680–687. IEEE, 2015.
- [LLL⁺20] Xiaodong Li, Yang Li, Shuyang Ling, Thomas Strohmer, and Ke Wei. When do birds of a feather flock together? k-means, proximity, and conic programming. *Mathematical Programming*, 179(1):295–341, 2020.
- [LM19] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, 173(3):859–929, 2019.
- [LZ16] Yu Lu and Harrison H Zhou. Statistical and computational guarantees of lloyd’s algorithm and its variants. *arXiv preprint arXiv:1612.02099*, 2016.
- [Ma13] Zongming Ma. Sparse principal component analysis and iterative thresholding. *The Annals of Statistics*, 41(2):772–801, 2013.
- [Mio17] Léo Miolane. Fundamental limits of low-rank matrix estimation: the non-symmetric case. *arXiv preprint arXiv:1702.00473*, 2017.
- [MPV87] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond*, volume 9. World Scientific Publishing Company, 1987.
- [MR⁺20] Nicolas Macris, Cynthia Rush, et al. All-or-nothing statistical and computational phase transitions in sparse spiked matrix estimation. *Advances in Neural Information Processing Systems*, 33:14915–14926, 2020.
- [MS22] Andrea Montanari and Subhabrata Sen. A short tutorial on mean-field spin glass techniques for non-physicists. *arXiv:2204.02909*, 2022.
- [MV10] Ankur Moitra and Gregory Valiant. Settling the polynomial learnability of mixtures of gaussians. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 93–102. IEEE, 2010.

- [MV21a] Marco Mondelli and Ramji Venkataramanan. Pca initialization for approximate message passing in rotationally invariant models. *Advances in Neural Information Processing Systems*, 34:29616–29629, 2021.
- [MV21b] Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. *The Annals of Statistics*, 49(1):321–345, 2021.
- [MVW17] Dustin G Mixon, Soledad Villar, and Rachel Ward. Clustering subgaussian mixtures by semidefinite programming. *Information and Inference: A Journal of the IMA*, 6(4):389–415, 2017.
- [Nda22] Mohamed Ndaoud. Sharp optimal recovery in the two component gaussian mixture model. *The Annals of Statistics*, 50(4):2096–2126, 2022.
- [NJB⁺08] John Novembre, Toby Johnson, Katarzyna Bryc, Zoltán Kutalik, Adam R Boyko, Adam Auton, Amit Indap, Karen S King, Sven Bergmann, Matthew R Nelson, et al. Genes mirror geography within europe. *Nature*, 456(7218):98–101, 2008.
- [NS08] John Novembre and Matthew Stephens. Interpreting principal component analyses of spatial population genetic variation. *Nature genetics*, 40(5):646–649, 2008.
- [Pan13] Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer Science & Business Media, 2013.
- [Pau12] Debashis Paul. Asymptotic distribution of the smallest eigenvalue of wishart (N, n) when $N, n \rightarrow \infty$ such that $N/n \rightarrow 0$. In *Nonparametric Statistical Methods and Related Topics: A Festschrift in Honor of Professor PK Bhattacharya on the Occasion of His 80th Birthday*, pages 423–458. World Scientific, 2012.
- [Pea94] Karl Pearson. Contributions to the mathematical theory of evolution. *Philosophical Transactions of the Royal Society of London. A*, 185:71–110, 1894.
- [PW07] Jiming Peng and Yu Wei. Approximating k-means-type clustering via semidefinite programming. *SIAM journal on optimization*, 18(1):186–205, 2007.
- [PWB20] Amelia Perry, Alexander S Wein, and Afonso S Bandeira. Statistical limits of spiked tensor models. In *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, volume 56, pages 230–264. Institut Henri Poincaré, 2020.
- [PWBM16] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Optimality and sub-optimality of pca for spiked random matrices and synchronization. *arXiv preprint arXiv:1609.05573*, 2016.
- [Roy17] Martin Royer. Adaptive clustering through semidefinite programming. *Advances in Neural Information Processing Systems*, 30, 2017.
- [RV17] Oded Regev and Aravindan Vijayaraghavan. On learning mixtures of well-separated gaussians. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 85–96. IEEE, 2017.
- [Sta59] Aart J Stam. Some inequalities satisfied by the quantities of information of fisher and shannon. *Information and Control*, 2(2):101–112, 1959.

- [Ver18] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.
- [VH14] Ramon Van Handel. Probability in high dimension. Technical report, PRINCETON UNIV NJ, 2014.
- [VW04] Santosh Vempala and Grant Wang. A spectral algorithm for learning mixture models. *Journal of Computer and System Sciences*, 68(4):841–860, 2004.
- [Wed72] Per-Åke Wedin. Perturbation bounds in connection with singular value decomposition. *BIT Numerical Mathematics*, 12(1):99–111, 1972.
- [ZSF22] Xinyi Zhong, Chang Su, and Zhou Fan. Empirical bayes pca in high dimensions. *Journal of the Royal Statistical Society Series B*, pages 853–878, 2022.

Appendix A Preliminaries

A.1 Further notations and conventions

In this section, we present an incomplete summary of the notations and conventions that will be applied throughout the appendix.

For two sequences of random vectors $\{\mathbf{X}_n\}_{n \in \mathbb{N}_+} \subseteq \mathbb{R}^k$ and $\{\mathbf{Y}_n\}_{n \in \mathbb{N}_+} \subseteq \mathbb{R}^k$, we say $\mathbf{X}_n \stackrel{P}{\simeq} \mathbf{Y}_n$ if and only if $\|\mathbf{X}_n - \mathbf{Y}_n\| = o_p(1)$. For $n, k \in \mathbb{N}_+$ and matrix $\mathbf{X} \in \mathbb{R}^{n \times k}$ with the i -th row denoted by $\mathbf{x}_i \in \mathbb{R}^k$, we let $\hat{p}_{\mathbf{X}}$ be the empirical distribution of the \mathbf{x}_i 's:

$$\hat{p}_{\mathbf{X}} := \frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{x}_i},$$

where $\delta_{\mathbf{x}_i}$ is the point mass at \mathbf{x}_i .

A.2 Wasserstein distance

For two probability distributions μ_1, μ_2 over \mathbb{R}^r , recall that the Wasserstein distance between μ_1 and μ_2 is defined as

$$W_2(\mu_1, \mu_2) := \left(\inf_{\gamma \in \Gamma(\mu_1, \mu_2)} \int_{\mathbb{R}^r \times \mathbb{R}^r} \|x - y\|^2 d\gamma(x, y) \right)^{1/2}, \quad (27)$$

where $\Gamma(\mu_1, \mu_2)$ denotes the collection of all probability distributions over $\mathbb{R}^r \times \mathbb{R}^r$ with marginals μ_1 and μ_2 on the first and last r coordinates, respectively. One observation is that for matrices $\mathbf{L}_1, \mathbf{L}_2 \in \mathbb{R}^{n \times r}$, the W_2 distance between the empirical distributions of their rows is upper bounded by the Frobenius norm of their difference: $W_2(\hat{p}_{\mathbf{L}_1}, \hat{p}_{\mathbf{L}_2}) \leq \frac{1}{\sqrt{n}} \|\mathbf{L}_1 - \mathbf{L}_2\|_F$.

Appendix B Technical lemmas

Lemma B.1. *Let $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^p$ be independent τ^2 sub-Gaussian random vectors, with mean 0 and covariance $\mathbb{E}[\mathbf{x}_i \mathbf{x}_i^T] = \Sigma$. We define the sample covariance matrix $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T$. Then for any $s \geq 100$, with probability at least $1 - 2e^{-ps^2/3200}$ we have*

$$\|\hat{\Sigma} - \Sigma\|_{\text{op}} \leq s\tau^2 \sqrt{\frac{p}{n}},$$

provided that $s\sqrt{p/n} \leq 25$.

Lemma B.2 (Wedin's theorem [Wed72]). *Let $\mathbf{A}_0, \mathbf{A}_1 \in \mathbb{R}^{m \times n}$ have singular value decomposition (for $a \in \{0, 1\}$)*

$$\mathbf{A}_a = \mathbf{U}_a \Sigma_a \mathbf{V}_a^T,$$

with Σ_a containing the singular values of \mathbf{A}_a in decreasing order. Furthermore, we let $\mathbf{U}_{a,+} \in \mathbb{R}^{m \times k(a)}$, $\mathbf{V}_{a,+} \in \mathbb{R}^{n \times k(a)}$, be formed by the first $k(a)$ columns of \mathbf{U}_a , \mathbf{V}_a , respectively, such that

$$\mathbf{U}_a = [\mathbf{U}_{a,+} | \mathbf{U}_{a,-}], \quad \mathbf{V}_a = [\mathbf{V}_{a,+} | \mathbf{V}_{a,-}].$$

Let $\sigma_k(\cdot)$ denote the k -th largest singular value of a matrix. Finally assume $\Delta \equiv \sigma_{k(1)}(\mathbf{A}_1) - \sigma_{k(0)+1}(\mathbf{A}_0) > 0$. Let $\mathbf{P}_a = \mathbf{V}_{a,+} \mathbf{V}_{a,+}^T$ (respectively, $\mathbf{Q}_a = \mathbf{U}_{a,+} \mathbf{U}_{a,+}^T$) denote the projector onto the

right singular space (left singular space) corresponding to the top $k(a)$ singular values of \mathbf{A}_a . Then we have

$$\|(\mathbf{I} - \mathbf{P}_0)\mathbf{P}_1\|_{\text{op}} \leq \frac{1}{\Delta} \{ \|(\mathbf{I} - \mathbf{Q}_0)(\mathbf{A}_0 - \mathbf{A}_1)\mathbf{P}_1\|_{\text{op}} \vee \|\mathbf{Q}_1(\mathbf{A}_0 - \mathbf{A}_1)(\mathbf{I} - \mathbf{P}_0)\|_{\text{op}} \}.$$

If instead we have $\Delta \equiv \sigma_{k(0)}(\mathbf{A}_0) - \sigma_{k(1)+1}(\mathbf{A}_1) > 0$, then

$$\|\mathbf{P}_0(\mathbf{I} - \mathbf{P}_1)\|_{\text{op}} \leq \frac{1}{\Delta} \{ \|(\mathbf{I} - \mathbf{Q}_1)(\mathbf{A}_0 - \mathbf{A}_1)\mathbf{P}_0\|_{\text{op}} \vee \|\mathbf{Q}_0(\mathbf{A}_0 - \mathbf{A}_1)(\mathbf{I} - \mathbf{P}_1)\|_{\text{op}} \}.$$

Lemma B.3 (Nishimori identity, Proposition 16 in [LM19]). Let (\mathbf{X}, \mathbf{Y}) be a couple of random variables on a polish space. Let $k \geq 1$ and let $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}$ be k i.i.d. samples (given \mathbf{Y}) from the distribution $\mathbb{P}(\mathbf{X} = \cdot \mid \mathbf{Y})$, independently of every other random variables. Let us denote $\langle \cdot \rangle$ the expectation with respect to $\mathbb{P}(\mathbf{X} = \cdot \mid \mathbf{Y})$ and \mathbb{E} the expectation with respect to (\mathbf{X}, \mathbf{Y}) . Then for all continuous bounded function f ,

$$\mathbb{E}\langle f(\mathbf{Y}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}) \rangle = \mathbb{E}\langle f(\mathbf{Y}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k-1)}, \mathbf{X}) \rangle.$$

Lemma B.4. Let $\{f_n\}_{n \in \mathbb{N}_+}$ be a sequence of convex differentiable functions on \mathbb{R} , and $f_n(x) \rightarrow f(x)$ for all $x \in \mathbb{R}$. Let $D_f = \{x \in \mathbb{R} : f \text{ is differentiable at } x\}$, then $f'_n(x) \rightarrow f'(x)$ for all $x \in D_f$.

Lemma B.5. If f and g are two differentiable convex functions, then for any $b > 0$,

$$|f'(a) - g'(a)| \leq g'(a+b) - g'(a-b) + \frac{d}{b},$$

where $d = |f(a+b) - g(a+b)| + |f(a-b) - g(a-b)| + |f(a) - g(a)|$.

Proof. See [Pan13], Lemma 3.2. □

Lemma B.6. For $\mathbf{X} \in \mathbb{R}^{n \times d}$, we define $\|\mathbf{X}\|_1 = \sum_{i \in [n], j \in [d]} |X_{ij}|$. Then for $\mathbf{X}_1, \mathbf{X}_2 \in \mathbb{R}^{n \times d}$, we have $\|\mathbf{X}_1 \mathbf{X}_1^\top - \mathbf{X}_2 \mathbf{X}_2^\top\|_1 \leq n \|\mathbf{X}_1 - \mathbf{X}_2\|_F (\|\mathbf{X}_1\|_F + \|\mathbf{X}_2\|_F)$.

Proof. we let $\mathbf{x}_i^1 \in \mathbb{R}^d$ be the i -th row of \mathbf{X}_1 and we let $\mathbf{x}_i^2 \in \mathbb{R}^d$ be the i -th row of \mathbf{X}_2 . Then by triangle inequality,

$$\begin{aligned} \|\mathbf{X}_1 \mathbf{X}_1^\top - \mathbf{X}_2 \mathbf{X}_2^\top\|_1 &\leq \|\mathbf{X}_1(\mathbf{X}_1 - \mathbf{X}_2)^\top\|_1 + \|\mathbf{X}_2(\mathbf{X}_1 - \mathbf{X}_2)^\top\|_1 \\ &\leq \sum_{i,j \in [n]} (\|\mathbf{x}_i^1\|_2 + \|\mathbf{x}_i^2\|_2) \times \|\mathbf{x}_j^1 - \mathbf{x}_j^2\|_2 \\ &\leq \sum_{i \in [n]} \|\mathbf{x}_i^1\|_2 \times \sqrt{n \sum_{j \in [n]} \|\mathbf{x}_j^1 - \mathbf{x}_j^2\|_2^2} + \sum_{i \in [n]} \|\mathbf{x}_i^2\|_2 \times \sqrt{n \sum_{j \in [n]} \|\mathbf{x}_j^1 - \mathbf{x}_j^2\|_2^2} \\ &\leq \sqrt{n \sum_{i \in [n]} \|\mathbf{x}_i^1\|_2^2} \times \sqrt{n \sum_{j \in [n]} \|\mathbf{x}_j^1 - \mathbf{x}_j^2\|_2^2} + \sqrt{n \sum_{i \in [n]} \|\mathbf{x}_i^2\|_2^2} \times \sqrt{n \sum_{j \in [n]} \|\mathbf{x}_j^1 - \mathbf{x}_j^2\|_2^2} \\ &= n \|\mathbf{X}_1 - \mathbf{X}_2\|_F (\|\mathbf{X}_1\|_F + \|\mathbf{X}_2\|_F). \end{aligned}$$

□

Appendix C Proofs for the strong signal regime

C.1 Proof of Theorem 3.1

Proof of claim 1

The top r eigenvectors of $\mathbf{A}\mathbf{A}^\top$ are also the top r eigenvectors of the following matrix

$$\frac{1}{d}\mathbf{A}\mathbf{A}^\top - \mathbf{I}_n = \frac{\mathbf{\Lambda}\mathbf{\Theta}^\top\mathbf{\Theta}\mathbf{\Lambda}^\top}{nd} + \frac{1}{d\sqrt{n}}\left(\mathbf{\Lambda}\mathbf{\Theta}^\top\mathbf{Z}^\top + \mathbf{Z}\mathbf{\Theta}\mathbf{\Lambda}^\top\right) + \frac{1}{d}\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}_n.$$

We let $\mathbf{X} = (\mathbf{\Lambda}\mathbf{\Theta}^\top\mathbf{Z}^\top + \mathbf{Z}\mathbf{\Theta}\mathbf{\Lambda}^\top)/d\sqrt{n} + \mathbf{Z}\mathbf{Z}^\top/d - \mathbf{I}_n$. Applying in sequence triangle inequality, Lemma B.1 and the law of large numbers, we conclude that there exists a constant $C > 0$, such that with probability $1 - o_n(1)$

$$\begin{aligned}\|\mathbf{X}\|_{\text{op}} &\leq \frac{2}{d\sqrt{n}}\|\mathbf{\Lambda}\mathbf{\Theta}^\top\mathbf{Z}^\top\|_{\text{op}} + \left\|\frac{1}{d}\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}_n\right\|_{\text{op}} \\ &\leq \frac{2}{d\sqrt{n}}\|\mathbf{\Lambda}\|_F\|\mathbf{Z}\mathbf{\Theta}\|_F + \left\|\frac{1}{d}\mathbf{Z}\mathbf{Z}^\top - \mathbf{I}_n\right\|_{\text{op}} \\ &\leq C\sqrt{\frac{n}{d}}.\end{aligned}\tag{28}$$

Using Lemma B.2, we see that there exists another constant $\tilde{C} > 0$, such that with probability $1 - o_n(1)$ we have

$$L^{\text{sin}}(\hat{\mathbf{\Lambda}}_s, \mathbf{\Lambda}) \leq \tilde{C}\sqrt{\frac{n}{d}}.$$

This completes the proof of the first claim of the theorem since by assumption $d/n \rightarrow \infty$.

Proof of claim 2

We write $\mathbf{A}\mathbf{A}^\top/d - \mathbf{I}_n = q_\Theta\mathbf{\Lambda}\mathbf{\Lambda}^\top/n + \mathbf{W}$, where \mathbf{W} is an $n \times n$ symmetric matrix. Using Eq. (28) and the law of large numbers, we find out that $\|\mathbf{W}\|_{\text{op}} = o_P(1)$.

We denote the unique eigenvalues of $q_\Theta\mathbf{Q}_\Lambda$ by $u_1 > u_2 > \dots > u_k > 0$, where $k \leq r$. The corresponding geometric multiplicities are denoted by $s_1, \dots, s_k \in \mathbb{N}_+$. We let $\delta_l, \hat{\delta}_l$ be the l -th largest eigenvalues of $q_\Theta\mathbf{\Lambda}\mathbf{\Lambda}^\top/n$ and $q_\Theta\mathbf{\Lambda}\mathbf{\Lambda}^\top/n + \mathbf{W}$, respectively. We then see immediately that $\delta_l, \hat{\delta}_l \xrightarrow{P} \sum_{i=1}^k u_i \mathbb{1}\{\sum_{j=1}^{i-1} s_j + 1 \leq l \leq \sum_{j=1}^i s_j\}$ as $n, d \rightarrow \infty$. Let

$$\begin{aligned}\mathbf{D}_i &= \text{diag}(\delta_{\sum_{j=1}^{i-1} s_j + 1}, \dots, \delta_{\sum_{j=1}^i s_j}) \in \mathbb{R}^{s_i \times s_i}, \\ \hat{\mathbf{D}}_i &= \text{diag}(\hat{\delta}_{\sum_{j=1}^{i-1} s_j + 1}, \dots, \hat{\delta}_{\sum_{j=1}^i s_j}) \in \mathbb{R}^{s_i \times s_i}.\end{aligned}$$

The above arguments imply that $\mathbf{D}_i, \hat{\mathbf{D}}_i \xrightarrow{P} u_i \mathbf{I}_{s_i}$.

For $i \in [k]$, we define the matrices $\mathbf{V}_i, \hat{\mathbf{V}}_i \in \mathbb{R}^{n \times s_i}$, such that the columns of $\mathbf{V}_i/\sqrt{n}, \hat{\mathbf{V}}_i/\sqrt{n}$ are the eigenvectors of $q_\Theta\mathbf{\Lambda}\mathbf{\Lambda}^\top/n, q_\Theta\mathbf{\Lambda}\mathbf{\Lambda}^\top/n + \mathbf{W}$ that correspond to the top $\sum_{j=1}^{i-1} s_j + 1$ to $\sum_{j=1}^i s_j$ eigenvalues, respectively. By Wedin's theorem (Lemma B.2), we see that $L^{\text{sin}}(\mathbf{V}_i, \hat{\mathbf{V}}_i) \xrightarrow{P} 0$ for all $i \in [k]$. Combining all arguments derived, we conclude that

$$\left\|\frac{1}{n}\sum_{i=1}^r \mathbf{V}_i \mathbf{D}_i \mathbf{V}_i^\top - \frac{1}{n}\sum_{i=1}^r \hat{\mathbf{V}}_i \hat{\mathbf{D}}_i \hat{\mathbf{V}}_i^\top\right\|_F^2 \xrightarrow{P} 0.$$

Note that $\sum_{i=1}^r \mathbf{V}_i \mathbf{D}_i \mathbf{V}_i^\top / n = q_\Theta \mathbf{\Lambda} \mathbf{\Lambda}^\top / n$ and $\hat{\mathbf{D}}_i, \hat{\mathbf{V}}_i$ are functions of \mathbf{A} , thus we have found an estimator $\hat{\mathbf{L}} \in \mathbb{R}^{n \times n}$ such that $\|\hat{\mathbf{L}} - \mathbf{\Lambda} \mathbf{\Lambda}^\top\|_F^2 / n^2 = o_P(1)$. Based on this convergence, we only need to apply a standard truncation argument to show the expected mean square error vanishes. We skip the details here for the sake of simplicity.

Proof of claim 3

By claim 2 of the theorem, we see that there exists an estimate $\hat{\mathbf{L}}$ of $\mathbf{\Lambda} \mathbf{\Lambda}^\top$ which achieves consistency: $\|\hat{\mathbf{L}} - \mathbf{\Lambda} \mathbf{\Lambda}^\top\|_F / n = o_P(1)$. Let

$$\mathbf{R} := \operatorname{argmin}_{\mathbf{X} \in \mathbb{R}^{n \times r}} \|\hat{\mathbf{L}} - \mathbf{X} \mathbf{X}^\top\|_F.$$

By definition, $\|\hat{\mathbf{L}} - \mathbf{R} \mathbf{R}^\top\|_F \leq \|\hat{\mathbf{L}} - \mathbf{\Lambda} \mathbf{\Lambda}^\top\|_F$, thus $\|\hat{\mathbf{L}} - \mathbf{R} \mathbf{R}^\top\|_F / n = o_P(1)$. By triangle inequality we see that $\|\mathbf{\Lambda} \mathbf{\Lambda}^\top - \mathbf{R} \mathbf{R}^\top\|_F / n = o_P(1)$, from which we conclude that there exists $\mathbf{\Omega}_0 \in \mathcal{O}(r)$ such that

$$\frac{1}{\sqrt{n}} \|\mathbf{R} \mathbf{\Omega}_0 - \mathbf{\Lambda}\|_F = o_P(1).$$

We define

$$\mathbf{\Omega}_* := \operatorname{argmin}_{\mathbf{\Omega} \in \mathcal{O}(r)} W_2(\hat{p}_{\mathbf{R} \mathbf{\Omega}}, \mu_\Lambda), \quad \mathbf{R}_* := \mathbf{R} \mathbf{\Omega}_*.$$

Denote by $\mu_{\mathbf{\Omega} \mathbf{\Lambda}}$ the distribution of $\mathbf{\Omega} \mathbf{\Lambda}_0$ for $\mathbf{\Omega} \in \mathcal{O}(r)$ and $\mathbf{\Lambda}_0 \sim \mu_\Lambda$. We then have $\|\mathbf{\Lambda} \mathbf{\Omega}_0^\top \mathbf{\Omega}_* - \mathbf{R}_*\|_F / \sqrt{n} = o_P(1)$, which implies $W_2(\hat{p}_{\mathbf{\Lambda} \mathbf{\Omega}_0^\top \mathbf{\Omega}_*}, \hat{p}_{\mathbf{R}_*}) = o_P(1)$. Furthermore,

$$\begin{aligned} W_2(\hat{p}_{\mathbf{R}_*}, \mu_\Lambda) &\leq W_2(\hat{p}_{\mathbf{R} \mathbf{\Omega}_0}, \mu_\Lambda) \\ &\leq W_2(\hat{p}_{\mathbf{R} \mathbf{\Omega}_0}, \hat{p}_\Lambda) + W_2(\hat{p}_\Lambda, \mu_\Lambda) \\ &\leq \frac{1}{\sqrt{n}} \|\mathbf{R} \mathbf{\Omega}_0 - \mathbf{\Lambda}\|_F + W_2(\hat{p}_\Lambda, \mu_\Lambda) = o_P(1). \end{aligned}$$

Invoking triangle inequality, we have

$$\begin{aligned} W_2(\mu_{\mathbf{\Omega}_*^\top \mathbf{\Omega}_0 \mathbf{\Lambda}}, \mu_\Lambda) &\leq W_2(\mu_\Lambda, \hat{p}_{\mathbf{R}_*}) + W_2(\hat{p}_{\mathbf{R}_*}, \hat{p}_{\mathbf{\Lambda} \mathbf{\Omega}_0^\top \mathbf{\Omega}_*}) + W_2(\hat{p}_{\mathbf{\Lambda} \mathbf{\Omega}_0^\top \mathbf{\Omega}_*}, \mu_{\mathbf{\Omega}_*^\top \mathbf{\Omega}_0 \mathbf{\Lambda}}) \\ &= W_2(\mu_\Lambda, \hat{p}_{\mathbf{R}_*}) + W_2(\hat{p}_{\mathbf{R}_*}, \hat{p}_{\mathbf{\Lambda} \mathbf{\Omega}_0^\top \mathbf{\Omega}_*}) + W_2(\hat{p}_\Lambda, \mu_\Lambda). \end{aligned}$$

Combining the above results, we see that $W_2(\mu_{\mathbf{\Omega}_*^\top \mathbf{\Omega}_0 \mathbf{\Lambda}}, \mu_\Lambda) = o_P(1)$. Notice that the mapping $\mathbf{\Omega} \mapsto W_2(\mu_\Lambda, \mu_{\mathbf{\Omega} \mathbf{\Lambda}})$ is continuous on $\mathcal{O}(r)$, and $W_2(\mu_1, \mu_2) = 0$ if and only if $\mu_1 = \mu_2$. Therefore, by assumption we obtain that $\|\mathbf{\Omega}_*^\top \mathbf{\Omega}_0 - \mathbf{I}_r\|_F = o_P(1)$, thus $\|\mathbf{R}_* - \mathbf{\Lambda}\|_F / \sqrt{n} = o_P(1)$. Notice that \mathbf{R} is a function of the observation \mathbf{A} , thus \mathbf{R}_* is a function of \mathbf{A} as well. Therefore, we have constructed a consistent estimator for $\mathbf{\Lambda}$ under the metric of vector mean square error. The rest parts of the proof again follow from a standard truncation argument.

C.2 Proof of Theorem 3.2

Proof of claim 1

We first prove Eq. (7). Define $\mathbf{A}_0 := (\sum_{i=1}^n \mathbf{\Lambda}_i \mathbf{\Lambda}_i^\top / n)^{-1/2} \in S_r^+$. We note that under the assumptions of remark 3.1, with high probability \mathbf{A}_0 is well-defined. For $j \in [d]$, we let $\mathbf{B}_0^j := \frac{1}{\sqrt{n}} \mathbf{A}_0^2 \sum_{i=1}^n A_{ij} \mathbf{\Lambda}_i \in \mathbb{R}^r$. We denote by \mathbf{M}_r the set of symmetric invertible matrices in $\mathbb{R}^{r \times r}$. Let $\mathbf{G} \sim \mathcal{N}(0, \mathbf{I}_r)$, independent of $\mathbf{\Theta}_0 \sim \mu_\Theta$. We define the mapping $f_\Theta : \mathbf{M}_r \times \mathbb{R}^r \rightarrow \mathbb{R}^r$ such that

$$f_\Theta(\mathbf{A}, \mathbf{B}) := \mathbb{E}[\mathbf{\Theta}_0 \mid \mathbf{\Theta}_0 + \mathbf{A} \mathbf{G} = \mathbf{B}] = \frac{\int \boldsymbol{\theta} \exp(-\frac{1}{2} \boldsymbol{\theta}^\top \mathbf{A}^{-2} \boldsymbol{\theta} + \mathbf{B}^\top \mathbf{A}^{-2} \boldsymbol{\theta}) \mu_\Theta(d\boldsymbol{\theta})}{\int \exp(-\frac{1}{2} \boldsymbol{\theta}^\top \mathbf{A}^{-2} \boldsymbol{\theta} + \mathbf{B}^\top \mathbf{A}^{-2} \boldsymbol{\theta}) \mu_\Theta(d\boldsymbol{\theta})}.$$

Let $\hat{\Theta}_j^B := \mathbb{E}[\Theta_j \mid \mathbf{A}, \Lambda] = f_\Theta(\mathbf{A}_0, \mathbf{B}_0^j)$, then $\hat{\Theta}_j^B$ achieves Bayesian mean square error.

Dominated convergence theorem reveals that $f_\Theta(\cdot, \cdot)$ is continuous. By the law of large numbers and central limit theorem, we see that $(\mathbf{A}_0, \mathbf{B}_0^j) \xrightarrow{d} (\mathbf{Q}_\Lambda^{-1/2}, \Theta_0 + \mathbf{Q}_\Lambda^{-1/2} \mathbf{G})$ as $n, d \rightarrow \infty$. Using Skorokhod's representation theorem, there exist $(\mathbf{A}_n, \mathbf{B}_n^j)$ and $(\mathbf{A}_\infty, \mathbf{B}_\infty^j)$ being random vectors defined on the same probability space, such that $(\mathbf{A}_n, \mathbf{B}_n^j) \xrightarrow{a.s.} (\mathbf{A}_\infty, \mathbf{B}_\infty^j)$, $(\mathbf{A}_n, \mathbf{B}_n^j) \stackrel{d}{=} (\mathbf{A}_0, \mathbf{B}_0^j)$, and $\mathbf{A}_\infty = \mathbf{Q}_\Lambda^{-1/2}$, $\mathbf{B}_\infty^j \stackrel{d}{=} \Theta_0 + \mathbf{Q}_\Lambda^{-1/2} \mathbf{G}$. Therefore, $f_\Theta(\mathbf{A}_0, \mathbf{B}_0^j) \stackrel{d}{=} f_\Theta(\mathbf{A}_n, \mathbf{B}_n^j) \xrightarrow{a.s.} f_\Theta(\mathbf{A}_\infty, \mathbf{B}_\infty^j)$. Since $\|f_\Theta(\mathbf{A}_n, \mathbf{B}_n^j)\|^2 \stackrel{d}{=} \|\mathbb{E}[\Theta_j \mid \mathbf{A}, \Lambda]\|^2$, we conclude that the set of random variables $\{\|f_\Theta(\mathbf{A}_n, \mathbf{B}_n^j)\|^2 : n \in \mathbb{N}_+\}$ is uniformly integrable. Therefore, we have $\|f_\Theta(\mathbf{A}_n, \mathbf{B}_n^j)\|^2 \xrightarrow{L_1} \|f_\Theta(\mathbf{A}_\infty, \mathbf{B}_\infty^j)\|^2$. This further implies that as $n, d \rightarrow \infty$

$$\begin{aligned} \mathbb{E}[\|\Theta_j - \hat{\Theta}_j^B\|_F^2] &= \mathbb{E}_{\Theta_0 \sim \mu_\Theta}[\|\Theta_0\|^2] - \mathbb{E}[\|f_\Theta(\mathbf{A}_n, \mathbf{B}_n^j)\|^2] \rightarrow \mathbb{E}[\|\Theta_0\|^2] - \mathbb{E}[\|f_\Theta(\mathbf{A}_\infty, \mathbf{B}_\infty)\|^2] \\ &= \mathbb{E}[\|\Theta_0\|^2] - \mathbb{E}[\|\mathbb{E}[\Theta_0 \mid \mathbf{Q}_\Lambda^{1/2} \Theta_0 + \mathbf{G}]\|^2], \end{aligned}$$

thus completing the proof of the first claim.

Proof of claim 2

Next, we prove Eq. (8). For $k, j \in [d]$, $k \neq j$, notice that Θ_k and Θ_j are conditionally independent conditioning on (\mathbf{A}, Λ) . Then we have $\mathbb{E}[\Theta_k^\top \Theta_j \mid \mathbf{A}, \Lambda] = \mathbb{E}[\Theta_k \mid \mathbf{A}, \Lambda]^\top \mathbb{E}[\Theta_j \mid \mathbf{A}, \Lambda] = f_\Theta(\mathbf{A}_0, \mathbf{B}_0^k)^\top f_\Theta(\mathbf{A}_0, \mathbf{B}_0^j)$, thus

$$\mathbb{E}[(\Theta_k^\top \Theta_j - \mathbb{E}[\Theta_k^\top \Theta_j \mid \mathbf{A}, \Lambda])^2] = r q_\Theta^2 - \mathbb{E}[(f_\Theta(\mathbf{A}_0, \mathbf{B}_0^k)^\top f_\Theta(\mathbf{A}_0, \mathbf{B}_0^j))^2].$$

By the law of large numbers and the central limit theorem, we have $(\mathbf{A}_0, \mathbf{B}_0^j, \mathbf{B}_0^k) \xrightarrow{d} (\mathbf{Q}_\Lambda^{-1/2}, \Theta_1 + \mathbf{Q}_\Lambda^{-1/2} \mathbf{G}_1, \Theta_2 + \mathbf{Q}_\Lambda^{-1/2} \mathbf{G}_2)$, where $\Theta_1, \Theta_2 \sim \mu_\Theta$, $\mathbf{G}_1, \mathbf{G}_2 \sim \mathcal{N}(0, \mathbf{I}_r)$ are mutually independent. By Skorokhod's representation theorem, there exist $(\mathbf{A}_n, \mathbf{B}_n^j, \mathbf{B}_n^k)$ and $(\mathbf{A}_\infty, \mathbf{B}_\infty^j, \mathbf{B}_\infty^k)$ being random vectors on the same probability space, such that $(\mathbf{A}_n, \mathbf{B}_n^j, \mathbf{B}_n^k) \xrightarrow{a.s.} (\mathbf{A}_\infty, \mathbf{B}_\infty^j, \mathbf{B}_\infty^k)$, $(\mathbf{A}_n, \mathbf{B}_n^j, \mathbf{B}_n^k) \stackrel{d}{=} (\mathbf{A}_0, \mathbf{B}_0^j, \mathbf{B}_0^k)$, and $\mathbf{A}_\infty = \mathbf{Q}_\Lambda^{-1/2}$, $(\mathbf{B}_\infty^j, \mathbf{B}_\infty^k) \stackrel{d}{=} (\Theta_1 + \mathbf{Q}_\Lambda^{-1/2} \mathbf{G}_1, \Theta_2 + \mathbf{Q}_\Lambda^{-1/2} \mathbf{G}_2)$. Therefore, as $n, d \rightarrow \infty$

$$\begin{aligned} & (f_\Theta(\mathbf{A}_0, \mathbf{B}_0^k)^\top f_\Theta(\mathbf{A}_0, \mathbf{B}_0^j))^2 \\ & \stackrel{d}{=} (f_\Theta(\mathbf{A}_n, \mathbf{B}_n^k)^\top f_\Theta(\mathbf{A}_n, \mathbf{B}_n^j))^2 \xrightarrow{a.s.} (f_\Theta(\mathbf{A}_\infty, \mathbf{B}_\infty^k)^\top f_\Theta(\mathbf{A}_\infty, \mathbf{B}_\infty^j))^2. \end{aligned}$$

Notice that

$$\begin{aligned} (f_\Theta(\mathbf{A}_0, \mathbf{B}_0^k)^\top f_\Theta(\mathbf{A}_0, \mathbf{B}_0^j))^2 & \leq \|f_\Theta(\mathbf{A}_0, \mathbf{B}_0^k)\|^2 \|f_\Theta(\mathbf{A}_0, \mathbf{B}_0^j)\|^2 \\ & \leq \mathbb{E}[\|\Theta_k\|^2 \mid \mathbf{A}, \Lambda] \mathbb{E}[\|\Theta_j\|^2 \mid \mathbf{A}, \Lambda] \\ & = \mathbb{E}[\|\Theta_k\|^2 \|\Theta_j\|^2 \mid \mathbf{A}, \Lambda]. \end{aligned}$$

Therefore, the set of random variables $\{(f_\Theta(\mathbf{A}_0, \mathbf{B}_0^k)^\top f_\Theta(\mathbf{A}_0, \mathbf{B}_0^j))^2 : n \in \mathbb{N}_+\}$ is uniformly integrable. This further implies that $(f_\Theta(\mathbf{A}_0, \mathbf{B}_0^k)^\top f_\Theta(\mathbf{A}_0, \mathbf{B}_0^j))^2 \xrightarrow{L_1} (f_\Theta(\mathbf{A}_\infty, \mathbf{B}_\infty^k)^\top f_\Theta(\mathbf{A}_\infty, \mathbf{B}_\infty^j))^2$, thus as $n, d \rightarrow \infty$

$$\mathbb{E}[(\Theta_k^\top \Theta_j - \mathbb{E}[\Theta_k^\top \Theta_j \mid \mathbf{A}, \Lambda])^2] \rightarrow r q_\Theta^2 - \left\| \mathbb{E} \left[\mathbb{E}[\Theta_0 \mid \mathbf{Q}_\Lambda^{1/2} \Theta_0 + \mathbf{G}] \mathbb{E}[\Theta_0 \mid \mathbf{Q}_\Lambda^{1/2} \Theta_0 + \mathbf{G}]^\top \right] \right\|_F^2,$$

which concludes the proof of the second claim of the theorem.

C.3 Proof of Theorem 3.3

By Theorem 3.1 claim 3, we see that there exists estimate $\hat{\mathbf{\Lambda}}$ of $\mathbf{\Lambda}$, such that $\|\mathbf{\Lambda} - \hat{\mathbf{\Lambda}}\|_F / \sqrt{n} \xrightarrow{P} 0$. Notice with high probability $\|\mathbf{A}\|_{\text{op}} \leq C\sqrt{d}$ for some constant $C > 0$ that depends uniquely on $(\mu_{\Lambda}, \mu_{\Theta})$, we then conclude that

$$\frac{1}{\sqrt{nd}} \|\mathbf{A}^\top \hat{\mathbf{\Lambda}} - \mathbf{A}^\top \mathbf{\Lambda}\|_F \leq \frac{\|\mathbf{A}\|_{\text{op}}}{\sqrt{d}} \cdot \frac{\|\mathbf{\Lambda} - \hat{\mathbf{\Lambda}}\|_F}{\sqrt{n}} = o_P(1).$$

Since \mathbf{Z} is independent of $\mathbf{\Lambda}$, we immediately see that there exists $\mathbf{g} \in \mathbb{R}^{d \times r}$ that has i.i.d. standard Gaussian entries and is independent of $(\mathbf{\Lambda}, \mathbf{\Theta})$, such that

$$\frac{1}{\sqrt{d}} \left\| \frac{1}{\sqrt{n}} \mathbf{A}^\top \hat{\mathbf{\Lambda}} \mathbf{Q}_{\Lambda}^{-1/2} - \mathbf{\Theta} \mathbf{Q}_{\Lambda}^{1/2} - \mathbf{g} \right\|_F = o_P(1). \quad (29)$$

Proof of the first result

We let $\mathbf{G} \sim \mathcal{N}(0, \mathbf{I}_r)$, $\mathbf{\Theta}_0 \sim \mu_{\Theta}$, independent of each other. Define the mapping $F : \mathbb{R}^r \rightarrow \mathbb{R}^r$, such that

$$F(\mathbf{y}) := \mathbb{E}[\mathbf{\Theta}_0 \mid \mathbf{Q}_{\Lambda}^{1/2} \mathbf{\Theta}_0 + \mathbf{G} = \mathbf{y}].$$

Dominated convergence theorem straightforwardly implies that F is continuous on \mathbb{R}^r . Therefore, for any $w \in (0, 1)$, we see that there exists a mapping $F^w : \mathbb{R}^r \rightarrow \mathbb{R}^r$, such that F^w is Lipschitz continuous. In addition,

$$\mathbb{E} \left[\|F(\mathbf{Q}_{\Lambda}^{1/2} \mathbf{\Theta}_0 + \mathbf{G}) - F^w(\mathbf{Q}_{\Lambda}^{1/2} \mathbf{\Theta}_0 + \mathbf{G})\|^2 \right] \leq w^2$$

We denote the Lipschitz constant of F^w by $L_w > 0$. Let $\mathbf{g}_i \in \mathbb{R}^r$ be the i -th row of \mathbf{g} . The law of large numbers gives the following convergence:

$$\frac{1}{d} \sum_{i=1}^d \|\mathbf{\Theta}_i - F(\mathbf{Q}_{\Lambda}^{1/2} \mathbf{\Theta}_i + \mathbf{g}_i)\|^2 \xrightarrow{P} \mathbb{E}[\|\mathbf{\Theta}_0\|^2] - \mathbb{E} \left[\left\| \mathbb{E}[\mathbf{\Theta}_0 \mid \mathbf{Q}_{\Lambda}^{1/2} \mathbf{\Theta}_0 + \mathbf{G}] \right\|^2 \right].$$

Therefore, as $n, d \rightarrow \infty$

$$\left| \frac{1}{d} \sum_{i=1}^d \|\mathbf{\Theta}_i - F^w(\mathbf{Q}_{\Lambda}^{1/2} \mathbf{\Theta}_i + \mathbf{g}_i)\|^2 - \mathbb{E}[\|\mathbf{\Theta}_0\|^2] + \mathbb{E} \left[\left\| \mathbb{E}[\mathbf{\Theta}_0 \mid \mathbf{Q}_{\Lambda}^{1/2} \mathbf{\Theta}_0 + \mathbf{G}] \right\|^2 \right] \right| \leq Cw + o_P(1),$$

where $C > 0$ is a constant depending only on μ_{Θ} . We denote by \mathbf{v}_i the i -th row of $\mathbf{A}^\top \hat{\mathbf{\Lambda}} \mathbf{Q}_{\Lambda}^{-1/2} / \sqrt{n}$. By assumption we have

$$\frac{1}{d} \sum_{i=1}^d \|F^w(\mathbf{Q}_{\Lambda}^{1/2} \mathbf{\Theta}_i + \mathbf{g}_i) - F^w(\mathbf{v}_i)\|^2 \leq \frac{L_w^2}{d} \sum_{i=1}^d \|\mathbf{Q}_{\Lambda}^{1/2} \mathbf{\Theta}_i + \mathbf{g}_i - \mathbf{v}_i\|^2,$$

which is $o_P(1)$ according to Eq. (29). Combining the above analysis, we conclude that for any $w \in (0, 1)$, there exists $n_w \in \mathbb{N}_+$, such that for $n \geq n_w$, there exists estimator $\hat{\mathbf{\Theta}}_w \in \mathbb{R}^{d \times r}$, such that with probability at least $1 - w$

$$\frac{1}{d} \sum_{i=1}^d \|\mathbf{\Theta}_i - F^w(\mathbf{v}_i)\|^2 \leq \mathbb{E}[\|\mathbf{\Theta}_0\|^2] - \mathbb{E} \left[\left\| \mathbb{E}[\mathbf{\Theta}_0 \mid \mathbf{Q}_{\Lambda}^{1/2} \mathbf{\Theta}_0 + \mathbf{G}] \right\|^2 \right] + 2w + o_P(1),$$

Since w is arbitrary, the rest parts of the proof follow from a simple truncation argument.

Proof of the second result

By analyzing the second moment we obtain that

$$\begin{aligned} & \frac{1}{d^2} \sum_{i,j \in [d]} |\boldsymbol{\Theta}_i^\top \boldsymbol{\Theta}_j - F(\mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_i + \mathbf{g}_i)^\top F(\mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_j + \mathbf{g}_j)|^2 \\ &= r q_\Theta^2 - \left\| \mathbb{E} \left[\mathbb{E}[\boldsymbol{\Theta}_0 \mid \mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_0 + \mathbf{G}] \mathbb{E}[\boldsymbol{\Theta}_0 \mid \mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_0 + \mathbf{G}]^\top \right] \right\|_F^2 + o_P(1). \end{aligned}$$

Since F is continuous, then for any $w \in (0, 1)$, there exists $\tilde{F}^w : \mathbb{R}^{2r} \rightarrow \mathbb{R}$ such that \tilde{F}^w is \tilde{L}_w -Lipschitz continuous. Furthermore,

$$\mathbb{E} \left[|F(\mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_i + \mathbf{g}_i)^\top F(\mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_j + \mathbf{g}_j) - \tilde{F}^w(\mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_i + \mathbf{g}_i, \mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_j + \mathbf{g}_j)|^2 \right] \leq w^2.$$

Again through analysis of the second moment we have

$$\begin{aligned} & \frac{1}{d^2} \sum_{i,j \in [d]} |\boldsymbol{\Theta}_i^\top \boldsymbol{\Theta}_j - \tilde{F}^w(\mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_i + \mathbf{g}_i, \mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_j + \mathbf{g}_j)|^2 \\ & \leq r q_\Theta^2 - \left\| \mathbb{E} \left[\mathbb{E}[\boldsymbol{\Theta}_0 \mid \mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_0 + \mathbf{G}] \mathbb{E}[\boldsymbol{\Theta}_0 \mid \mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_0 + \mathbf{G}]^\top \right] \right\|_F^2 + \tilde{C}w + o_P(1), \end{aligned}$$

where $\tilde{C} > 0$ is a constant depending uniquely on μ_Θ . By Lipschitzness we have

$$\begin{aligned} & \frac{1}{d^2} \sum_{i,j \in [d]} |\tilde{F}^w(\mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_i + \mathbf{g}_i, \mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_j + \mathbf{g}_j) - \tilde{F}^w(\mathbf{v}_i, \mathbf{v}_j)|^2 \\ & \leq \frac{\tilde{L}_w^2}{d^2} \sum_{i,j \in [d]} \left\{ \|\mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_i + \mathbf{g}_i - \mathbf{v}_i\|^2 + \|\mathbf{Q}_\Lambda^{1/2} \boldsymbol{\Theta}_j + \mathbf{g}_j - \mathbf{v}_j\|^2 \right\}, \end{aligned}$$

which by Eq. (29) is $o_P(1)$. Since w is arbitrary, again the claim follows by applying standard truncation argument.

Appendix D Proof outlines for the weak signal regime

D.1 Proof of Theorem 4.2

Assume $\boldsymbol{\Lambda}$ is given, then for any $j \in [d]$, the posterior distribution of $\boldsymbol{\Theta}_j$ given $(\mathbf{A}, \boldsymbol{\Lambda})$ can be expressed as

$$p(d\boldsymbol{\theta}_j | \boldsymbol{\Lambda}, \mathbf{A}) \propto \exp \left(-\frac{1}{2\sqrt{nd}} \sum_{i=1}^n \langle \boldsymbol{\Lambda}_i, \boldsymbol{\theta}_j \rangle^2 + \frac{1}{\sqrt[4]{nd}} \sum_{i=1}^n A_{ij} \langle \boldsymbol{\Lambda}_i, \boldsymbol{\theta}_j \rangle^2 \right) \mu_\Theta(d\boldsymbol{\theta}_j).$$

From the above equation we see that the posterior of $\boldsymbol{\Theta}$ given $(\mathbf{A}, \boldsymbol{\Lambda})$ is a product distribution over \mathbb{R}^d , thus greatly simplifies the analysis. The rest of the proof is similar to that of Theorem 3.2, and we skip it for simplicity.

D.2 Proof outline of Theorem 4.3

In this section we outline the proof of Theorem 4.3. We leave the proofs of technical lemmas to Appendix E. For the sake of simplicity, here we consider only $r = 1$. We comment that cases with $r \geq 2$ can be proven similarly.

D.2.1 Free energy density

Note that the posterior distributions that correspond to the symmetric and asymmetric models can be expressed as follows:

$$\begin{aligned} d\mathbb{P}(\mathbf{\Lambda} = \boldsymbol{\lambda} \mid \mathbf{Y}) &= \frac{e^{H_{s,n}(\boldsymbol{\lambda})} \mu_{\mathbf{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda})}{\int e^{H_{s,n}(\boldsymbol{\lambda})} \mu_{\mathbf{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda})}, \\ d\mathbb{P}(\mathbf{\Lambda} = \boldsymbol{\lambda}, \boldsymbol{\Theta} = \boldsymbol{\theta} \mid \mathbf{A}) &= \frac{e^{H_n(\boldsymbol{\lambda}, \boldsymbol{\theta})} \mu_{\boldsymbol{\Theta}}^{\otimes d}(d\boldsymbol{\theta}) \mu_{\mathbf{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda})}{\int e^{H_n(\boldsymbol{\lambda}, \boldsymbol{\theta})} \mu_{\boldsymbol{\Theta}}^{\otimes d}(d\boldsymbol{\theta}) \mu_{\mathbf{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda})}, \end{aligned}$$

where $\mu_{\mathbf{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda})$ ($\mu_{\boldsymbol{\Theta}}^{\otimes d}(d\boldsymbol{\theta})$) is the product distribution over \mathbb{R}^n (\mathbb{R}^d) with each coordinate having marginal distribution $\mu_{\mathbf{\Lambda}}$ ($\mu_{\boldsymbol{\Theta}}$), and $H_{s,n}, H_n$ are the Hamiltonians that correspond to models (10) and (9), respectively:

$$\begin{aligned} H_{s,n}(\boldsymbol{\lambda}) &:= \frac{q_{\boldsymbol{\Theta}}^2}{2n} \langle \mathbf{\Lambda}, \boldsymbol{\lambda} \rangle^2 + \frac{q_{\boldsymbol{\Theta}}}{2} \boldsymbol{\lambda}^\top \mathbf{W} \boldsymbol{\lambda} - \frac{q_{\boldsymbol{\Theta}}^2}{4n} \|\boldsymbol{\lambda}\|^4, \\ H_n(\boldsymbol{\lambda}, \boldsymbol{\theta}) &:= \frac{1}{\sqrt{nd}} \langle \mathbf{\Lambda}, \boldsymbol{\lambda} \rangle \langle \boldsymbol{\Theta}, \boldsymbol{\theta} \rangle + \frac{1}{\sqrt{4nd}} \boldsymbol{\lambda}^\top \mathbf{Z} \boldsymbol{\theta} - \frac{1}{2\sqrt{nd}} \|\boldsymbol{\lambda}\|^2 \|\boldsymbol{\theta}\|^2. \end{aligned} \tag{30}$$

Following the terminology of statistical mechanics, the *free energy density* is defined as the expected log-partition function (also known as log normalizing constant):

$$\begin{aligned} \Psi_n^s &:= \frac{1}{n} \mathbb{E} \log \int e^{H_{s,n}(\boldsymbol{\lambda})} \mu_{\mathbf{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda}), \\ \Psi_n &:= \frac{1}{n} \mathbb{E} \log \int e^{H_n(\boldsymbol{\lambda}, \boldsymbol{\theta})} \mu_{\mathbf{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda}) \mu_{\boldsymbol{\Theta}}^{\otimes d}(d\boldsymbol{\theta}). \end{aligned}$$

The lemma below connects free energy densities with the corresponding mutual informations.

Lemma D.1. *The following equations hold:*

$$\begin{aligned} \Psi_n^s &= \frac{q_{\boldsymbol{\Theta}}^2 \mathbb{E}[\mathbf{\Lambda}_0^2]^2}{4} - \mathbb{I}_n^{\text{symm}}(\mu_{\mathbf{\Lambda}}; q_{\boldsymbol{\Theta}}) + o_n(1), \\ \Psi_n &= \frac{q_{\boldsymbol{\Theta}}^2 \mathbb{E}[\mathbf{\Lambda}_0^2]^2}{4} - \mathbb{I}_n^{\text{asym}}(\mu_{\mathbf{\Lambda}}, \mu_{\boldsymbol{\Theta}}) + o_n(1). \end{aligned}$$

Proof. By definition, the mutual information that corresponds to the symmetric model can be reformulated as

$$\begin{aligned} \mathbb{I}_n^{\text{symm}}(\mu_{\mathbf{\Lambda}}; q_{\boldsymbol{\Theta}}) &= \frac{1}{n} \mathbb{E} \left\{ \log \frac{d\mu_{\mathbf{\Lambda}}^{\otimes n}(\mathbf{\Lambda}) \cdot \exp(H_{s,n}(\mathbf{\Lambda}))}{d\mu_{\mathbf{\Lambda}}^{\otimes n}(\mathbf{\Lambda}) \cdot \int \exp(H_{s,n}(\boldsymbol{\lambda})) d\mu_{\mathbf{\Lambda}}^{\otimes n}(\boldsymbol{\lambda})} \right\} \\ &= \frac{q_{\boldsymbol{\Theta}}^2 \mathbb{E}[\mathbf{\Lambda}_0^2]^2}{4} - \Psi_n^s. \end{aligned}$$

The asymmetric mutual information is a slightly more complicated, which we write below

$$\begin{aligned} &\mathbb{I}_n^{\text{asym}}(\mu_{\mathbf{\Lambda}}, \mu_{\boldsymbol{\Theta}}) \\ &= \frac{1}{n} \mathbb{E} \left\{ \log \frac{d\mu_{\mathbf{\Lambda}}^{\otimes n}(\mathbf{\Lambda}) \cdot \int \exp(H_n(\mathbf{\Lambda}, \boldsymbol{\theta})) d\mu_{\boldsymbol{\Theta}}^{\otimes n}(\boldsymbol{\theta})}{d\mu_{\mathbf{\Lambda}}^{\otimes n}(\mathbf{\Lambda}) \cdot \int \exp(H_n(\boldsymbol{\lambda}, \boldsymbol{\theta})) d\mu_{\mathbf{\Lambda}}^{\otimes n}(\boldsymbol{\lambda}) d\mu_{\boldsymbol{\Theta}}^{\otimes n}(\boldsymbol{\theta})} \right\} \\ &= \frac{1}{n} \sum_{i=1}^d \mathbb{E} \left\{ \log \int \exp \left(\frac{1}{\sqrt{nd}} \|\mathbf{\Lambda}\|^2 \boldsymbol{\Theta}_i \boldsymbol{\theta}_i + \frac{1}{\sqrt{4nd}} \langle \mathbf{Z}_{\cdot i}, \mathbf{\Lambda} \rangle \boldsymbol{\theta}_i - \frac{1}{2\sqrt{nd}} \|\mathbf{\Lambda}\|^2 \boldsymbol{\theta}_i^2 \right) \mu_{\boldsymbol{\Theta}}(d\boldsymbol{\theta}_i) \right\} - \Psi_n \end{aligned}$$

$$= \frac{d}{n} \mathbb{E} \left\{ \log \int \exp \left(\frac{1}{\sqrt{nd}} \|\Lambda\|^2 \Theta_1 \theta_1 + \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot 1}, \Lambda \rangle \theta_1 - \frac{1}{2\sqrt{nd}} \|\Lambda\|^2 \theta_1^2 \right) \mu_{\Theta}(\mathrm{d}\theta_1) \right\} - \Psi_n.$$

Define

$$F(q) = \mathbb{E} \left\{ \log \int \exp \left(q \Theta_0 \theta + \sqrt{q} \mathbf{G} \theta - \frac{1}{2} q \theta^2 \right) \mu_{\Theta}(\mathrm{d}\theta) \right\},$$

where the expectation is taken over $\Theta_0 \sim \mu_{\Theta}$, $\mathbf{G} \sim \mathcal{N}(0, 1)$ that are independent of each other. Applying Stein's lemma, we obtain that for $q > 0$, $F(q)$ has second order continuous derivatives satisfying

$$F'(q) = \frac{1}{2} \mathbb{E} \left\{ \mathbb{E}[\Theta_0 \mid \sqrt{q} \Theta_0 + \mathbf{G}]^2 \right\},$$

$$F''(q) = \mathbb{E} \left\{ \frac{1}{2} \mathbb{E}[\Theta_0^2 \mid \sqrt{q} \Theta_0 + \mathbf{G}]^2 + \frac{1}{2} \mathbb{E}[\Theta_0 \mid \sqrt{q} \Theta_0 + \mathbf{G}]^4 - \mathbb{E}[\Theta_0 \mid \sqrt{q} \Theta_0 + \mathbf{G}]^2 \mathbb{E}[\Theta_0^2 \mid \sqrt{q} \Theta_0 + \mathbf{G}] \right\}.$$

Since μ_{Θ} has mean zero, we conclude that F also has second order continuous derivatives at zero, and $F'(0) = 0$, $F''(0) = q_{\Theta}^2/2$. These arguments imply that

$$\mathbb{I}_n^{\text{asym}}(\mu_{\Lambda}, \mu_{\Theta}) = \frac{d}{n} \mathbb{E} \left\{ F \left(\frac{1}{\sqrt{nd}} \|\Lambda\|^2 \right) \right\} - \Psi_n = \frac{q_{\Theta}^2 \mathbb{E}[\Lambda_0^2]^2}{4} - \Psi_n + o_n(1),$$

which concludes the proof of the lemma. □

From Lemma D.1 we see that in order to prove the theorem, it suffices to show that the free energy densities agree asymptotically:

$$\lim_{n \rightarrow \infty} \Psi_n^s = \lim_{n, d \rightarrow \infty} \Psi_n. \quad (31)$$

D.2.2 Asymptotic equivalence of free energy densities

We then proceed to prove Eq. (31). We will start with the additional constraint that μ_{Λ} has bounded support. Later in Appendix D.4, we show that proofs for general μ_{Λ} can be reduced to the bounded case.

Assumption D.1. *We assume that $\text{support}(\mu_{\Lambda}) \subseteq [-K, K]$, with $K > 0$ being some fixed constant that is independent of n, d .*

For $h, s \geq 0$, we define the perturbations

$$\mathbf{Y}'(h) = \frac{\sqrt{h}}{n} \Lambda \Lambda^{\top} + \mathbf{W}',$$

$$\mathbf{x}'(s) = \sqrt{s} \Lambda + \mathbf{g}',$$

where $\mathbf{W}' \stackrel{d}{=} \text{GOE}(n)$ and $\mathbf{g}' \stackrel{d}{=} \mathcal{N}(0, \mathbf{I}_n)$. Furthermore, we require that $(\mathbf{W}', \mathbf{g}', \Lambda, \Theta, \mathbf{Z}, \mathbf{W})$ are mutually independent. We define the Hamiltonians associated with the perturbations $\mathbf{Y}'(h)$ and $\mathbf{x}'(s)$ respectively as follows:

$$H_n(\lambda; \mathbf{Y}'(h)) := \frac{h}{2n} \langle \Lambda, \lambda \rangle^2 + \frac{\sqrt{h}}{2} \lambda^{\top} \mathbf{W}' \lambda - \frac{h}{4n} \|\lambda\|^4, \quad (32)$$

$$H_n(\boldsymbol{\lambda}; \mathbf{x}'(s)) := \sqrt{s} \langle \boldsymbol{\lambda}, \mathbf{g}' \rangle + s \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle - \frac{s}{2} \|\boldsymbol{\lambda}\|^2. \quad (33)$$

The posterior distribution of $\boldsymbol{\Lambda}$ given $(\mathbf{A}, \mathbf{Y}'(h), \mathbf{x}'(s))$ can be expressed as

$$\mu(d\boldsymbol{\lambda}) = \frac{1}{Z_n(h, s)} \mu_{\boldsymbol{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda}) \int \exp(H_n(\boldsymbol{\lambda}, \boldsymbol{\theta}) + H_n(\boldsymbol{\lambda}; \mathbf{Y}'(h)) + H_n(\boldsymbol{\lambda}; \mathbf{x}'(s))) \mu_{\boldsymbol{\Theta}}^{\otimes d}(d\boldsymbol{\theta}),$$

where $Z_n(h, s)$ is the normalizing constant:

$$Z_n(h, s) = \int \exp(H_n(\boldsymbol{\lambda}, \boldsymbol{\theta}) + H_n(\boldsymbol{\lambda}; \mathbf{Y}'(h)) + H_n(\boldsymbol{\lambda}; \mathbf{x}'(s))) \mu_{\boldsymbol{\Theta}}^{\otimes d}(d\boldsymbol{\theta}) \mu_{\boldsymbol{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda}).$$

Note that $Z_n(h, s)$ is random and depends on $(\mathbf{A}, \mathbf{Y}'(h), \mathbf{x}'(s))$. We define the free energy density that corresponds to observations $(\mathbf{A}, \mathbf{Y}'(h), \mathbf{x}'(s))$ as

$$\Phi_n(h, s) := \frac{1}{n} \mathbb{E} [\log Z_n(h, s)]. \quad (34)$$

The next equation follows from Gaussian integration by parts and Nishimori identity (Lemma B.3):

$$\frac{\partial}{\partial h} \Phi_n(h, s) = \frac{1}{4n^2} \mathbb{E} \left[\langle \boldsymbol{\Lambda} \boldsymbol{\Lambda}^\top, \mathbb{E}[\boldsymbol{\Lambda} \boldsymbol{\Lambda}^\top \mid \mathbf{A}, \mathbf{Y}'(h), \mathbf{x}'(s)] \rangle \right]. \quad (35)$$

Eq. (35) holds for all $h > 0, s \geq 0$, and is directly related to the MMSE in the perturbed model. The rest parts of the proof will be devoted to proving convergence of $\Phi_n(h, s)$ as $n, d \rightarrow \infty$.

To this end, we first show that asymptotically speaking, the free energy density depends on $\mu_{\boldsymbol{\Theta}}$ only through its second moment. More precisely, we can replace $\mu_{\boldsymbol{\Theta}}$ with a Gaussian distribution which has mean zero and variance $q_{\boldsymbol{\Theta}}$. This vastly simplifies further computation.

Lemma D.2. *For $k \in [d]$, we let $P_{\boldsymbol{\Theta}, k}$ be a distribution over \mathbb{R}^d with independent coordinates, such that $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_d) \sim P_{\boldsymbol{\Theta}, k}$ if and only if $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_k \stackrel{iid}{\sim} \mu_{\boldsymbol{\Theta}}$ and $\boldsymbol{\theta}_{k+1}, \dots, \boldsymbol{\theta}_d \stackrel{iid}{\sim} \mathcal{N}(0, q_{\boldsymbol{\Theta}})$. We define*

$$\Phi_n^{(k)}(h, s) := \frac{1}{n} \mathbb{E} \left[\log \left(\int \exp(H_n(\boldsymbol{\lambda}, \boldsymbol{\theta}) + H_n(\boldsymbol{\lambda}; \mathbf{Y}'(h)) + H_n(\boldsymbol{\lambda}; \mathbf{x}'(s))) \mu_{\boldsymbol{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda}) dP_{\boldsymbol{\Theta}, k}(\boldsymbol{\theta}) \right) \right].$$

In the above expression, the expectation is taken over $(\boldsymbol{\Lambda}, \boldsymbol{\Theta}, \mathbf{Z}, \mathbf{W}', \mathbf{g}')$. Notice that by definition $\Phi_n(h, s) = \Phi_n^{(d)}(h, s)$. Then under the conditions of Theorem 4.3 and in addition Assumption D.1, as $n, d \rightarrow \infty$ we have $\Phi_n(h, s) - \Phi_n^{(0)}(h, s) = o_n(1)$ for all fixed $h, s \geq 0$.

Lemma D.2 can be proved via a Lindeberg type argument, and we postpone the details to Appendix E.1. According to Lemma D.2, in order to derive the limiting expression of $\Phi_n(h, s)$, it suffices to compute the limit of $\Phi_n^{(0)}(h, s)$ instead, which can be done via Gaussian integration techniques.

Lemma D.3. *For fixed $h, s \geq 0$, we define*

$$\begin{aligned} \tilde{H}_n(\boldsymbol{\lambda}; \mathbf{Y}'(h), \mathbf{x}'(s)) &:= \frac{q_{\boldsymbol{\Theta}}^2}{2n} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^2 + \frac{q_{\boldsymbol{\Theta}}}{2\sqrt{nd}} \|\mathbf{Z}^\top \boldsymbol{\lambda}\|^2 - \frac{dq_{\boldsymbol{\Theta}}}{2\sqrt{nd}} \|\boldsymbol{\lambda}\|^2 - \frac{q_{\boldsymbol{\Theta}}^2}{4n} \|\boldsymbol{\lambda}\|^4 \\ &\quad + H_n(\boldsymbol{\lambda}; \mathbf{Y}'(h)) + H_n(\boldsymbol{\lambda}; \mathbf{x}'(s)), \\ \tilde{\Phi}_n(h, s) &:= \frac{1}{n} \mathbb{E} \left[\log \left(\int \exp(\tilde{H}_n(\boldsymbol{\lambda}; \mathbf{Y}'(h), \mathbf{x}'(s))) \mu_{\boldsymbol{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda}) \right) \right]. \end{aligned}$$

Then under the conditions of Theorem 4.3 and Assumption D.1, as $n, d \rightarrow \infty$, we have $\tilde{\Phi}_n(h, s) - \Phi_n^{(0)}(h, s) = o_n(1)$.

We defer the proof of Lemma D.3 to Appendix E.2. Under the asymptotics $n, d \rightarrow \infty$, $d/n \rightarrow \infty$, according to [BY88b], the matrix $(\mathbf{Z}\mathbf{Z}^\top - d\mathbf{I}_n)/\sqrt{nd}$ behaves like a $\text{GOE}(n)$ matrix. Replacing $(\mathbf{Z}\mathbf{Z}^\top - d\mathbf{I}_n)/\sqrt{nd}$ with a $\text{GOE}(n)$ matrix in the definition of $\tilde{\Phi}_n(h, s)$, we see that this allows us to approximate $\tilde{\Phi}_n(h, s)$ via the free energy density of the symmetric model (10). Such heuristics can be made rigorous via the following lemma:

Lemma D.4. *Recall that \mathbf{Y} is defined in Eq. (10). For $h, s \geq 0$, we define the free energy density $\Phi_n^Y(h, s)$ that corresponds to the observations $(\mathbf{Y}, \mathbf{Y}'(h), \mathbf{x}'(s))$ as*

$$\begin{aligned} \Phi_n^Y(h, s) &:= \frac{1}{n} \mathbb{E} \left[\log \left(\int \exp(H_n^Y(\boldsymbol{\lambda}) + H_n(\boldsymbol{\lambda}; \mathbf{Y}'(h)) + H_n(\boldsymbol{\lambda}; \mathbf{x}'(s))) \mu_{\Lambda}^{\otimes n}(d\boldsymbol{\lambda}) \right) \right], \\ H_n^Y(\boldsymbol{\lambda}) &:= \frac{q_{\Theta}^2}{2n} \langle \mathbf{A}, \boldsymbol{\lambda} \rangle^2 + \frac{q_{\Theta}}{2} \boldsymbol{\lambda}^\top \mathbf{W} \boldsymbol{\lambda} - \frac{q_{\Theta}^2}{4n} \|\boldsymbol{\lambda}\|^4. \end{aligned} \quad (36)$$

Then under the conditions of Theorem 4.3 and Assumption D.1, as $n, d \rightarrow \infty$, we have $\Phi_n^Y(h, s) - \tilde{\Phi}_n(h, s) = o_n(1)$.

We defer the proof of Lemma D.4 to Appendix E.3. Combining Lemmas D.2 to D.4, we conclude that as $n, d \rightarrow \infty$, for all fixed $h, s \geq 0$, we have $\Phi_n(h, s) - \Phi_n^Y(h, s) = o_n(1)$. This relates the asymmetric model to the symmetric model through their free energy densities. The following lemma summarizes this result and lists several additional useful properties for future reference.

Lemma D.5. *Under the conditions of Theorem 4.3 and Assumption D.1, for all fixed $h, s \geq 0$, the following claims hold:*

1. As $n, d \rightarrow \infty$, we have $\Phi_n(h, s) = \Phi_n^Y(h, s) + o_n(1)$.
2. The following mappings $x \mapsto \Phi_n(h, x)$, $x \mapsto \Phi_n(x, s)$, $x \mapsto \Phi_n^Y(h, x)$, $x \mapsto \Phi_n^Y(x, s)$ are all convex on $[0, \infty)$ and differentiable on $(0, \infty)$.
3. $\lim_{n \rightarrow \infty} \Phi_n^Y(0, s)$ exists for all

$$(q_{\Theta}^2, s) \in \{(tx, (1-t)xq^*(x)) : x \geq 0, q^*(x) \text{ exists and is unique}, t \in [0, 1]\},$$

where $\mathcal{F}(\cdot, \cdot)$ is defined in Eq. (13) and

$$q^*(x) := \operatorname{argmax}_{q \geq 0} \mathcal{F}(x, q). \quad (37)$$

Remark D.1. By [LM19, Proposition 17], $q^*(x)$ exists and is unique for all but countably many $x > 0$. Claims 2 and 3 do not rely on Assumption D.1.

We delay the proof of Lemma D.5 to Appendix E.4. We note that Theorem 4.3 is an immediate consequence of Lemma D.5.

D.3 Proof of Theorem 4.4

In this section, we will apply Lemma D.5 to prove Theorem 4.4. Using Lemma B.3 and Gaussian integration by parts, for all $h > 0$ we have

$$\begin{aligned} \frac{\partial}{\partial h} \Phi_n(h, 0) &= \frac{1}{4n^2} \mathbb{E} \left[\langle \mathbf{A} \mathbf{A}^\top, \mathbb{E}[\mathbf{A} \mathbf{A}^\top \mid \mathbf{A}, \mathbf{Y}'(h)] \rangle \right], \\ \frac{\partial}{\partial h} \Phi_n^Y(h, 0) &= \frac{1}{4n^2} \mathbb{E} \left[\langle \mathbf{A} \mathbf{A}^\top, \mathbb{E}[\mathbf{A} \mathbf{A}^\top \mid \mathbf{Y}, \mathbf{Y}'(h)] \rangle \right]. \end{aligned} \quad (38)$$

Recall that \mathcal{F} is defined in Eq. (13). We let

$$D := \{s > 0 \mid \mathcal{F}(s, \cdot) \text{ has a unique maximizer } q^*(s)\}. \quad (39)$$

By proposition 17 in [LM19], D is equal to $(0, +\infty)$ minus a countable set, and is precisely the set of $s > 0$ at which the function $\phi : s \mapsto \sup_{q \geq 0} \mathcal{F}(s, q)$ is differentiable. Furthermore, by [LM19, Theorem 13], for all $h \geq 0$,

$$\lim_{n \rightarrow \infty} \Phi_n^Y(h, 0) = \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q). \quad (40)$$

By the first claim of Lemma D.5, $\Phi_n(h, 0) = \Phi_n^Y(h, 0) + o_n(1)$, thus $\lim_{n \rightarrow \infty} \Phi_n(h, 0) = \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q)$. By the second claim of Lemma D.5, the mappings $h \mapsto \Phi_n(h, 0)$, $h \mapsto \Phi_n^Y(h, 0)$ are convex and differentiable on $(0, \infty)$. Next, we apply Lemma B.4 to function sequences $\{h \mapsto \Phi_n(h, 0)\}_{n \geq 1}$, $\{h \mapsto \Phi_n^Y(h, 0)\}_{n \geq 1}$, and conclude that for all but countably many values of $h > 0$,

$$\lim_{n \rightarrow \infty} \frac{\partial}{\partial h} \Phi_n(h, 0) = \lim_{n \rightarrow \infty} \frac{\partial}{\partial h} \Phi_n^Y(h, 0) = \phi'(h + q_\Theta^2). \quad (41)$$

From the above equation we see that the mapping $\lambda \mapsto \phi'(\lambda)$ is non-decreasing on D . Therefore, for all but countably many $q_\Theta > 0$, ϕ' is continuous at $q_\Theta^2 \in D$. For these q_Θ , we immediately see that for any $\varepsilon > 0$, there exists $h_\varepsilon > 0$ depending uniquely on $(q_\Theta, \varepsilon, \mu_\Lambda)$, such that $\phi'(q_\Theta^2 + h_\varepsilon) \leq \phi'(q_\Theta^2) + \varepsilon$, and ϕ is differentiable at $q_\Theta^2 + h_\varepsilon$. According to Eq. (41), there exists $n_\varepsilon \in \mathbb{N}_+$, such that for all $n \geq n_\varepsilon$,

$$\left| \frac{\partial}{\partial h} \Phi_n(h_\varepsilon, 0) - \phi'(h_\varepsilon + q_\Theta^2) \right| \leq \varepsilon.$$

According to Eq. (38),

$$\begin{aligned} \frac{\partial}{\partial h} \Phi_n(h, 0) &\geq \frac{1}{4n^2} \mathbb{E} \left[\langle \mathbf{A} \mathbf{A}^\top, \mathbb{E}[\mathbf{A} \mathbf{A}^\top \mid \mathbf{A}] \rangle \right] \\ &= \frac{1}{4n^2} (n \mathbb{E}_{\mathbf{A}_0 \sim \mu_\Lambda} [\mathbf{A}_0^4] + n(n-1) \mathbb{E}_{\mathbf{A}_0 \sim \mu_\Lambda} [\mathbf{A}_0^2]^2 - n^2 \text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta)). \end{aligned}$$

Invoking Proposition 17 and Corollary 18 from [LM19], for all $q_\Theta^2 \in D$

$$\phi'(q_\Theta^2) = \frac{1}{4} q^*(q_\Theta^2)^2 = \frac{1}{4} \mathbb{E}_{\mathbf{A}_0 \sim \mu_\Lambda} [\mathbf{A}_0^2]^2 - \frac{1}{4} \lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_\Lambda; q_\Theta) + o_n(1).$$

Combining all arguments above, we obtain that

$$\liminf_{n, d \rightarrow \infty} \text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta) \geq \lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_\Lambda; q_\Theta) - 8\varepsilon.$$

Since ε is arbitrary, we then complete the proof of the first claim of the theorem.

We then proceed to prove the second claim. For $q_\Theta^2, q_\Theta^2 + \eta \in D$ satisfying $0 < \eta < \varepsilon$, by Eqs. (38) and (41) we have

$$\begin{aligned} \lim_{n, d \rightarrow \infty} \text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta; \eta) &= \mathbb{E}_{\mathbf{A}_0 \sim \mu_\Lambda} [\mathbf{A}_0^2]^2 - 4\phi'(q_\Theta^2 + \eta) \\ &\leq \mathbb{E}_{\mathbf{A}_0 \sim \mu_\Lambda} [\mathbf{A}_0^2]^2 - 4\phi'(q_\Theta^2) \\ &= \lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_\Lambda; q_\Theta). \end{aligned}$$

Note that $\lim_{n, d \rightarrow \infty} \text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta; \eta) \geq \limsup_{n, d \rightarrow \infty} \text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta; \varepsilon)$, the proof of the second claim immediately follows.

D.4 Reduction to bounded prior

In this section, we show that in order to prove Theorem 4.3 and Theorem 4.4, it suffices to prove the theorems under Assumption D.1.

Since μ_Λ is sub-Gaussian, for any $\varepsilon > 0$, there exists $K_\varepsilon > 0$, such that if we let $\bar{\Lambda}_0 := \Lambda_0 \mathbb{1}\{|\Lambda_0| \leq K_\varepsilon\}$, then $\mathbb{E}_{\Lambda_0 \sim \mu_\Lambda}[(\Lambda_0 - \bar{\Lambda}_0)^4] < \varepsilon$ and $\mu_\Lambda([-K_\varepsilon, K_\varepsilon]) > 1 - \varepsilon^2$. For all $i \in [n]$, we define $\bar{\Lambda}_i := \Lambda_i \mathbb{1}\{|\Lambda_i| \leq K_\varepsilon\}$ and $\bar{\lambda}_i := \lambda_i \mathbb{1}\{|\lambda_i| \leq K_\varepsilon\}$. Let $\bar{\Lambda} = (\bar{\Lambda}_i)_{i \leq n} \in \mathbb{R}^n$ and $\bar{\lambda} = (\bar{\lambda}_i)_{i \leq n} \in \mathbb{R}^n$. We introduce the truncated Hamiltonians:

$$\begin{aligned}\bar{H}_n^\varepsilon(\bar{\lambda}, \theta) &:= \frac{1}{\sqrt{nd}} \langle \bar{\Lambda}, \bar{\lambda} \rangle \langle \Theta, \theta \rangle + \frac{1}{\sqrt[4]{nd}} \bar{\lambda}^\top \mathbf{Z} \theta - \frac{1}{2\sqrt{nd}} \|\bar{\lambda}\|^2 \|\theta\|^2, \\ \bar{H}_n^{Y,\varepsilon}(\bar{\lambda}) &:= \frac{q_\Theta^2}{2n} \langle \bar{\Lambda}, \bar{\lambda} \rangle^2 + \frac{q_\Theta}{2} \bar{\lambda}^\top \mathbf{W} \bar{\lambda} - \frac{q_\Theta^2}{4n} \|\bar{\lambda}\|^4.\end{aligned}$$

Recall that $\mathbf{W}' \stackrel{d}{=} \text{GOE}(n)$ and is independent of (\mathbf{W}, \mathbf{Z}) . For $s, q, h \geq 0$, we define the truncated versions of Φ_n^Y , Φ_n and \mathcal{F} as

$$\begin{aligned}\bar{\Phi}_n^{Y,\varepsilon}(h) &:= \frac{1}{n} \mathbb{E} \left[\log \left(\int \exp \left(\bar{H}_n^{Y,\varepsilon}(\bar{\lambda}) + \frac{h}{2n} \langle \bar{\Lambda}, \bar{\lambda} \rangle^2 + \frac{\sqrt{h}}{2} \bar{\lambda}^\top \mathbf{W}' \bar{\lambda} - \frac{h}{4n} \|\bar{\lambda}\|^4 \right) \mu_{\bar{\Lambda}}^{\otimes n}(\mathrm{d}\bar{\lambda}) \right) \right], \\ \bar{\Phi}_n^\varepsilon(h) &:= \frac{1}{n} \mathbb{E} \left[\log \left(\int \exp \left(\bar{H}_n^\varepsilon(\bar{\lambda}, \theta) + \frac{h}{2n} \langle \bar{\Lambda}, \bar{\lambda} \rangle^2 + \frac{\sqrt{h}}{2} \bar{\lambda}^\top \mathbf{W}' \bar{\lambda} - \frac{h}{4n} \|\bar{\lambda}\|^4 \right) \mu_{\bar{\Lambda}}^{\otimes n}(\mathrm{d}\bar{\lambda}) \mu_{\Theta}^{\otimes d}(\mathrm{d}\theta) \right) \right], \\ \bar{\mathcal{F}}^\varepsilon(s, q) &:= -\frac{s}{4} q^2 + \mathbb{E}_{Z \sim \mathcal{N}(0,1), \Lambda_0 \sim \mu_\Lambda} \left[\log \left(\int \exp \left(\sqrt{sq} Z \bar{\lambda} + sq \bar{\lambda} \bar{\Lambda}_0 - \frac{s}{2} q \bar{\lambda}^2 \right) \mu_{\bar{\Lambda}}(\mathrm{d}\bar{\lambda}) \right) \right].\end{aligned}$$

In the above display, $\mu_{\bar{\Lambda}}$ stands for the law of $\bar{\Lambda}_0$. The following lemma states that $\bar{\Phi}_n^\varepsilon(h)$ is close to $\Phi_n(h, 0)$ for small ε .

Lemma D.6. *Under the conditions of Theorem 4.3, there exists a constant $C_0 > 0$, which is a function of $(\mu_\Lambda, \mu_\Theta)$ only, such that for n, d large enough, the following inequality holds for all $h \in [0, 1]$:*

$$|\Phi_n(h, 0) - \bar{\Phi}_n^\varepsilon(h)| \leq C_0 \sqrt[4]{\varepsilon}.$$

The proof of Lemma D.6 is deferred to Section E.5. Furthermore, according to Lemma 46 from [LM19], \mathcal{F} is also close to $\bar{\mathcal{F}}^\varepsilon$ for ε small.

Lemma D.7 (Lemma 46 from [LM19]). *Under the conditions of Theorem 4.3, there exists a constant $K' > 0$ that depends only on μ_Λ , such that*

$$\left| \sup_{q \geq 0} \mathcal{F}(s, q) - \sup_{q \geq 0} \bar{\mathcal{F}}^\varepsilon(s, q) \right| \leq s K' \varepsilon.$$

Invoking the convergence results of free energy density for the symmetric spiked model [LM19], we have

$$\left| \bar{\Phi}_n^{Y,\varepsilon}(h) - \sup_{q \geq 0} \bar{\mathcal{F}}^\varepsilon(q_\Theta^2 + h, q) \right| = o_n(1).$$

Applying Lemma D.5 to the truncated distribution $\mu_{\bar{\Lambda}}$, we obtain that $|\bar{\Phi}_n^{Y,\varepsilon}(h) - \bar{\Phi}_n^\varepsilon(h)| = o_n(1)$ for all $h \geq 0$. Using this result and Lemma D.6, D.7, we derive that for all $h \in [0, 1]$,

$$|\Phi_n(h, 0) - \Phi_n^Y(h, 0)|$$

$$\begin{aligned}
&\leq |\Phi_n(h, 0) - \bar{\Phi}_n^\varepsilon(h)| + |\bar{\Phi}_n^\varepsilon(h) - \bar{\Phi}_n^{Y, \varepsilon}(h)| + |\bar{\Phi}_n^{Y, \varepsilon}(h) - \sup_{q \geq 0} \bar{\mathcal{F}}^\varepsilon(q_\Theta^2 + h, q)| \\
&\quad + |\sup_{q \geq 0} \bar{\mathcal{F}}^\varepsilon(q_\Theta^2 + h, q) - \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q)| + |\sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q) - \Phi_n^Y(h, 0)| \\
&\leq C_0 \sqrt[4]{\varepsilon} + (q_\Theta^2 + 1)K'\varepsilon + o_n(1).
\end{aligned}$$

Since ε is arbitrary, we then have the following lemma:

Lemma D.8. *Under the conditions of Theorem 4.3, for all $h \in [0, 1]$, as $n, d \rightarrow \infty$ we have*

$$\lim_{n, d \rightarrow \infty} |\Phi_n(h, 0) - \Phi_n^Y(h, 0)| = 0.$$

Theorem 4.3 is a direct consequence of Lemma D.8. The remainder proof of Theorem 4.4 follows exactly the same procedure as stated in Section D.3, and here we skip it for the sake of simplicity.

D.5 Proof outline of Theorem 4.5

We state the proof outline of Theorem 4.5 in this section. The proofs of supporting lemmas are delayed to Appendix F. For the sake of simplicity, here we only consider the rank-one case $r = 1$. We comment that proof for $r \geq 2$ can be conducted analogously.

D.5.1 Proof outline of Theorem 4.5 under condition (a)

Since μ_Λ is sub-Gaussian, there exists a constant $K_0 > 0$ depending only on μ_Λ , such that for all $x > 0$

$$\mathbb{P}(|\Lambda_0| \geq x) \leq 2 \exp(-x^2/K_0^2). \quad (42)$$

For all $i \in [n]$, we define $\bar{\Lambda}_i := \Lambda_i \mathbb{1}\{|\Lambda_i| \leq 2K_0\sqrt{\log n}\}$, $\bar{\Lambda} := (\bar{\Lambda}_1, \dots, \bar{\Lambda}_n)^\top \in \mathbb{R}^n$ and $\bar{\mathbf{A}} := \bar{\Lambda}\Theta^\top/\sqrt{nd} + \mathbf{Z} \in \mathbb{R}^{n \times d}$. The next lemma says that truncation does not decrease the MMSE too much.

Lemma D.9. *Under the conditions of Theorem 4.5 (a), as $n, d \rightarrow \infty$ we have*

$$\frac{1}{n^2} \mathbb{E} \left[\left\| \Lambda \Lambda^\top - \mathbb{E}[\Lambda \Lambda^\top | \mathbf{A}] \right\|_F^2 \right] \leq \frac{1}{n^2} \mathbb{E} \left[\left\| \bar{\Lambda} \bar{\Lambda}^\top - \mathbb{E}[\bar{\Lambda} \bar{\Lambda}^\top | \bar{\mathbf{A}}] \right\|_F^2 \right] + o_n(1).$$

We leave the proof of Lemma D.9 to Appendix F.1. By Lemma D.9, in order to prove the theorem, it suffices to show that under the current conditions, for all but countably many values of $q_\Theta > 0$,

$$\limsup_{n, d \rightarrow \infty} \frac{1}{n^2} \mathbb{E} \left[\left\| \bar{\Lambda} \bar{\Lambda}^\top - \mathbb{E}[\bar{\Lambda} \bar{\Lambda}^\top | \bar{\mathbf{A}}] \right\|_F^2 \right] \leq \lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_\Lambda; q_\Theta). \quad (43)$$

We let $\mathbf{Z}\Theta = \|\Theta\| \cdot \mathbf{g}$, where $\mathbf{g} \sim \mathcal{N}(0, \mathbf{I}_n)$ is independent of (Θ, Λ) . We denote by \mathbf{Z}' an independent copy of \mathbf{Z} , such that \mathbf{Z}' is further independent of $(\Theta, \Lambda, \mathbf{g})$. Furthermore, we can choose \mathbf{Z}' such that $(\Theta, \Lambda, \mathbf{g}, \mathbf{Z}\mathbf{P}_\Theta^\perp \mathbf{Z}^\top) = (\Theta, \Lambda, \mathbf{g}, \mathbf{Z}'\mathbf{P}_\Theta^\perp \mathbf{Z}'^\top)$, where \mathbf{P}_Θ^\perp denotes the projection onto the null space of Θ .

We define $\mathbf{Y}_1 := (\bar{\mathbf{A}}\bar{\mathbf{A}}^\top - d\mathbf{I}_n)/\sqrt{d}$. Then \mathbf{Y}_1 admits the following decomposition:

$$\mathbf{Y}_1 = \frac{\|\Theta\|^2}{\sqrt{nd}} \bar{\Lambda} \bar{\Lambda}^\top + \frac{\|\Theta\|}{n^{1/4}d^{3/4}} \bar{\Lambda} \mathbf{g}^\top + \frac{\|\Theta\|}{n^{1/4}d^{3/4}} \mathbf{g} \bar{\Lambda}^\top + \frac{1}{\sqrt{d}} (\mathbf{Z}\mathbf{P}_\Theta^\perp \mathbf{Z}^\top - d\mathbf{I}_n) + \frac{1}{\sqrt{d}} \mathbf{g} \mathbf{g}^\top$$

$$= \frac{1}{\sqrt{n}} \left(q_{\Theta}^{1/2} \bar{\Lambda} + r_n^{-1} \mathbf{g} \right) \left(q_{\Theta}^{1/2} \bar{\Lambda} + r_n^{-1} \mathbf{g} \right)^{\top} + \frac{1}{\sqrt{d}} (\mathbf{Z}' \mathbf{Z}'^{\top} - d \mathbf{I}_n) + \mathbf{E}.$$

In the above display, the $n \times n$ symmetric matrix \mathbf{E} is defined as follows:

$$\begin{aligned} \mathbf{E} := & -\frac{1}{\sqrt{d} \|\boldsymbol{\Theta}\|^2} \mathbf{Z}' \boldsymbol{\Theta} \boldsymbol{\Theta}^{\top} \mathbf{Z}'^{\top} + \frac{1}{\sqrt{n}} \left(\frac{\|\boldsymbol{\Theta}\|^2}{d} - q_{\Theta} \right) \bar{\Lambda} \bar{\Lambda}^{\top} \\ & + \frac{1}{n^{1/4} d^{1/4}} \left(\frac{\|\boldsymbol{\Theta}\|}{\sqrt{d}} - q_{\Theta}^{1/2} \right) \left(\bar{\Lambda} \mathbf{g}^{\top} + \mathbf{g} \bar{\Lambda}^{\top} \right) + \frac{1}{\sqrt{d}} \mathbf{g} \mathbf{g}^{\top}. \end{aligned}$$

We define the set

$$\Omega_1 := \left\{ \left| \frac{1}{d} \|\boldsymbol{\Theta}\|^2 - q_{\Theta} \right| \leq \frac{C_1 \sqrt{\log n}}{\sqrt{d}}, \frac{1}{\sqrt{d}} |(\mathbf{Z}' \boldsymbol{\Theta})_i| \leq C_1 \sqrt{\log n}, \frac{1}{\sqrt{d}} |g_i| \leq C_1 \sqrt{\log n} \text{ for all } i \in [n] \right\},$$

where $C_1 > 0$ is a constant depending only on μ_{Θ} . Since $\mu_{\Theta}, \mu_{\Lambda}$ are sub-Gaussian distributions, if we choose C_1 large enough, then we have $\mathbb{P}(\Omega_1) = 1 - o_n(1)$. Using the definition of Ω_1 , we conclude that there exists a constant $C_2 > 0$ that depends only on (C_1, K_0, μ_{Θ}) , such that on Ω_1 we have $|E_{ij}| \leq C_2 \log n / \sqrt{d}$ for all $i, j \in [n]$. For some absolute constant $C_3 > 0$, we let $\bar{\mathbf{g}} \in \mathbb{R}^n$ such that $\bar{g}_i := g_i \mathbb{1}\{|g_i| \leq C_3 \sqrt{\log n}\}$ for all $i \in [n]$. Direct computation reveals that for C_3 large enough, we have $\mathbb{P}(\bar{\mathbf{g}} \neq \mathbf{g}) \rightarrow 0$ as $n, d \rightarrow \infty$.

Define

$$\mathbf{Y}_2 := \frac{1}{\sqrt{n}} \left(q_{\Theta}^{1/2} \bar{\Lambda} + r_n^{-1} \bar{\mathbf{g}} \right) \left(q_{\Theta}^{1/2} \bar{\Lambda} + r_n^{-1} \bar{\mathbf{g}} \right)^{\top} + \mathbf{G},$$

where $\mathbf{G} \sim \sqrt{n} \text{GOE}(n)$ and is independent of $(\Lambda, \bar{\Lambda}, \boldsymbol{\Theta}, \mathbf{g}, \bar{\mathbf{g}})$. By [BDER16, Theorem 4], under condition (a), there exists a coupling such that as $n, d \rightarrow \infty$, with probability $1 - o_n(1)$ we have $(\mathbf{Z}' \mathbf{Z}'^{\top} - d \mathbf{I}_n) / \sqrt{d} = \mathbf{G}$. We define $\Omega_2 := \Omega_1 \cap \{\mathbf{Y}_2 = \mathbf{Y}_1 - \mathbf{E}\}$, then we see that $\mathbb{P}(\Omega_2) \rightarrow 1$ as $n, d \rightarrow \infty$.

For $\mathbf{X} \in \mathbb{R}^{n \times n}$, we define

$$\mathbf{M}_n(\mathbf{X}) := \frac{1}{n} \mathbb{E} \left[\left(q_{\Theta}^{1/2} \bar{\Lambda} + r_n^{-1} \bar{\mathbf{g}} \right) \left(q_{\Theta}^{1/2} \bar{\Lambda} + r_n^{-1} \bar{\mathbf{g}} \right)^{\top} \mid \mathbf{Y}_2 = \mathbf{X} \right].$$

Then for any $i, j, k, s \in [n]$, we have

$$\begin{aligned} \frac{\partial \mathbf{M}_n(\mathbf{X})_{ks}}{\partial X_{ij}} &= \frac{1}{n^{3/2}} \mathbb{E} \left[\left(q_{\Theta}^{1/2} \bar{\Lambda}_i + r_n^{-1} \bar{g}_i \right) \left(q_{\Theta}^{1/2} \bar{\Lambda}_j + r_n^{-1} \bar{g}_j \right) \left(q_{\Theta}^{1/2} \bar{\Lambda}_k + r_n^{-1} \bar{g}_k \right) \left(q_{\Theta}^{1/2} \bar{\Lambda}_s + r_n^{-1} \bar{g}_s \right) \mid \mathbf{Y}_2 = \mathbf{X} \right] \\ &- \frac{1}{n^{3/2}} \mathbb{E} \left[\left(q_{\Theta}^{1/2} \bar{\Lambda}_i + r_n^{-1} \bar{g}_i \right) \left(q_{\Theta}^{1/2} \bar{\Lambda}_j + r_n^{-1} \bar{g}_j \right) \mid \mathbf{Y}_2 = \mathbf{X} \right] \mathbb{E} \left[\left(q_{\Theta}^{1/2} \bar{\Lambda}_k + r_n^{-1} \bar{g}_k \right) \left(q_{\Theta}^{1/2} \bar{\Lambda}_s + r_n^{-1} \bar{g}_s \right) \mid \mathbf{Y}_2 = \mathbf{X} \right]. \end{aligned}$$

Note that on Ω_2 we have $|\bar{\Lambda}_i| \leq 2K_0 \sqrt{\log n}$, $|\bar{g}_i| \leq C_3 \sqrt{\log n}$, and $|E_{ij}| \leq C_2 \log n / \sqrt{d}$ for all $i, j \in [n]$. Therefore, we conclude that there exists a constant $C_4 > 0$ depending only on $(K_0, \mu_{\Theta}, C_1, C_2, C_3)$, such that on Ω_2 we have

$$|\mathbf{M}_n(\mathbf{Y}_1)_{ks} - \mathbf{M}_n(\mathbf{Y}_1 - \mathbf{E})_{ks}| \leq \frac{C_4 \sqrt{n} (\log n)^3}{\sqrt{d}}.$$

Therefore, we obtain that

$$\mathbb{E} \left[\|\mathbf{M}_n(\mathbf{Y}_1 - \mathbf{E}) - \mathbf{M}_n(\mathbf{Y}_1)\|_F^2 \mathbb{1}_{\Omega_2} \right] \leq \frac{C_4^2 n^3 (\log n)^6}{d}.$$

The right hand side of the above equation vanishes as $n, d \rightarrow \infty$ under condition (a). Therefore, we derive that $\|\mathbf{M}_n(\mathbf{Y}_2) - \mathbf{M}_n(\mathbf{Y}_1)\|_F = o_P(1)$. Note that \mathbf{Y}_1 is a function of $\bar{\mathbf{A}}$. Using standard truncation argument, we conclude that in order to prove Eq. (43), it suffices to show

$$\limsup_{n,d \rightarrow \infty} \mathbb{E} \left[\left\| q_{\Theta}^{-1} \mathbf{M}_n(\mathbf{Y}_2) - \frac{1}{n} \bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^{\top} \right\|_F^2 \right] \leq \lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_{\Lambda}; q_{\Theta}),$$

which we prove in Lemma D.10 below. The proof of this lemma is deferred to Appendix F.2.

Lemma D.10. *Under the conditions of Theorem 4.5 (a), for all but countably many values of $q_{\Theta} > 0$, we have*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left\| q_{\Theta}^{-1} \mathbf{M}_n(\mathbf{Y}_2) - \frac{1}{n} \bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^{\top} \right\|_F^2 \right] = \lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_{\Lambda}; q_{\Theta}).$$

D.5.2 Proof outline of Theorem 4.5 under condition (b)

Truncation

By assumption, there exists $0 < K_1 < \infty$ such that $\text{support}(\mu_{\Lambda}) \subseteq [-K_1, K_1]$. Since μ_{Θ} is sub-Gaussian, there exists $K_2 > 0$ which depends only on μ_{Θ} , such that for all $x > 0$

$$\mathbb{P}_{\Theta_0 \sim \mu_{\Theta}}(|\Theta_0| \geq x) \leq 2 \exp(-x^2/K_2^2). \quad (44)$$

For $j \in \{0\} \cup [d]$, we define $\bar{\Theta}_j := \Theta_j \mathbb{1}\{|\Theta_j| \leq 2K_2\sqrt{\log d}\}$, $\bar{\Theta} := (\bar{\Theta}_1, \dots, \bar{\Theta}_d)^{\top} \in \mathbb{R}^d$ and $\bar{\mathbf{A}} := \mathbf{\Lambda} \bar{\Theta}^{\top} / \sqrt[4]{nd} + \mathbf{Z}$. Note that $\bar{\mathbf{A}}$ defined here is not to be confused with $\bar{\mathbf{A}}$ defined in Appendix D.5.1. The following lemma states that truncation does not decrease the asymptotic matrix MMSE.

Lemma D.11. *Under the conditions of Theorem 4.5 (b), as $n, d \rightarrow \infty$ we have*

$$\frac{1}{n^2} \mathbb{E} \left[\left\| \mathbf{\Lambda} \mathbf{\Lambda}^{\top} - \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^{\top} \mid \mathbf{A}] \right\|_F^2 \right] \leq \frac{1}{n^2} \mathbb{E} \left[\left\| \mathbf{\Lambda} \mathbf{\Lambda}^{\top} - \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^{\top} \mid \bar{\mathbf{A}}] \right\|_F^2 \right] + o_n(1).$$

We postpone the proof of the lemma to Appendix F.3. By Lemma D.11, in order to prove Theorem 4.5 under condition (b), we only need to show for all but countably many values of $q_{\Theta} > 0$,

$$\limsup_{n,d \rightarrow \infty} \frac{1}{n^2} \mathbb{E} \left[\left\| \mathbf{\Lambda} \mathbf{\Lambda}^{\top} - \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^{\top} \mid \bar{\mathbf{A}}] \right\|_F^2 \right] \leq \lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_{\Lambda}; q_{\Theta}).$$

Model with extra perturbation

For $s, h, a, a' \geq 0$ and $\{\varepsilon_n\}_{n \geq 1}, \{\varepsilon'_n\}_{n \geq 1} \subseteq \mathbb{R}_+$, we introduce a perturbed model sequence, such that for each n , we observe $(\bar{\mathbf{A}}(s), \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h))$ defined as follows:

$$\bar{\mathbf{A}}(s) := \frac{\sqrt{s}}{\sqrt[4]{nd}} \mathbf{\Lambda} \bar{\Theta}^{\top} + \mathbf{Z}, \quad (45)$$

$$\mathbf{x}'(a') := a' \sqrt{\varepsilon'_n} \mathbf{\Lambda} + \mathbf{g}', \quad (46)$$

$$\bar{\mathbf{x}}(a) := a \sqrt{\frac{n\varepsilon_n}{d}} \bar{\Theta} + \mathbf{g}, \quad (47)$$

$$\mathbf{Y}'(h) := \frac{\sqrt{h}}{n} \mathbf{\Lambda} \mathbf{\Lambda}^{\top} + \mathbf{W}', \quad (48)$$

where $\mathbf{g} \sim \mathcal{N}(0, \mathbf{I}_d)$, $\mathbf{g}' \sim \mathcal{N}(0, \mathbf{I}_n)$, $\mathbf{W}' \sim \text{GOE}(n)$, mutually independent and are independent of everything else. Note that $\bar{\mathbf{A}}(1) = \bar{\mathbf{A}}$. Furthermore, we assume that $\varepsilon_n, \varepsilon'_n \rightarrow 0^+$ as $n, d \rightarrow \infty$. We can associate to the observation (45) the Hamiltonian

$$\bar{H}_n^{[s]}(\boldsymbol{\lambda}, \bar{\boldsymbol{\theta}}) = \sum_{i \in [n], j \in [d]} \left\{ \frac{s}{\sqrt{nd}} \boldsymbol{\Lambda}_i \boldsymbol{\lambda}_i \bar{\boldsymbol{\Theta}}_j \bar{\boldsymbol{\theta}}_j + \frac{\sqrt{s}}{\sqrt[4]{nd}} Z_{ij} \boldsymbol{\lambda}_i \bar{\boldsymbol{\theta}}_j - \frac{s}{2\sqrt{nd}} \boldsymbol{\lambda}_i^2 \bar{\boldsymbol{\theta}}_j^2 \right\}, \quad (49)$$

where $\bar{\boldsymbol{\theta}}_j = \boldsymbol{\theta}_j \mathbb{1}\{|\boldsymbol{\theta}_j| \leq 2K_2\sqrt{\log d}\}$. For the sake of simplicity, we let $\bar{H}_n(\boldsymbol{\lambda}, \bar{\boldsymbol{\theta}}) = \bar{H}_n^{[1]}(\boldsymbol{\lambda}, \bar{\boldsymbol{\theta}})$. Similarly, we can associate to the observations (46) and (47) the following Hamiltonians, respectively:

$$\begin{aligned} \bar{H}_{n,\boldsymbol{\lambda}}^{(pert)}(\boldsymbol{\lambda}) &= \sum_{i=1}^n \left\{ \sqrt{\varepsilon'_n} a' \boldsymbol{\lambda}_i g'_i + a'^2 \varepsilon'_n \boldsymbol{\Lambda}_i \boldsymbol{\lambda}_i - \frac{a'^2 \varepsilon'_n}{2} \boldsymbol{\lambda}_i^2 \right\}, \\ \bar{H}_{n,\boldsymbol{\theta}}^{(pert)}(\bar{\boldsymbol{\theta}}) &= \sum_{j=1}^d \left\{ \sqrt{\frac{n\varepsilon_n}{d}} a \bar{\boldsymbol{\theta}}_j g_j + \frac{a^2 n \varepsilon_n}{d} \bar{\boldsymbol{\Theta}}_j \bar{\boldsymbol{\theta}}_j - \frac{a^2 n \varepsilon_n}{2d} \bar{\boldsymbol{\theta}}_j^2 \right\}. \end{aligned}$$

We then define the “total” Hamiltonian, which corresponds to all observations in the perturbed model as

$$\bar{H}_n^{(tot)}(\boldsymbol{\lambda}, \bar{\boldsymbol{\theta}}) := \bar{H}_n^{[s]}(\boldsymbol{\lambda}, \bar{\boldsymbol{\theta}}) + \bar{H}_{n,\boldsymbol{\lambda}}^{(pert)}(\boldsymbol{\lambda}) + \bar{H}_{n,\boldsymbol{\theta}}^{(pert)}(\bar{\boldsymbol{\theta}}) + H_n(\boldsymbol{\lambda}; \mathbf{Y}'(h)),$$

where we recall that $H_n(\boldsymbol{\lambda}; \mathbf{Y}'(h))$ is defined in Eq. (32). The posterior distribution of $(\boldsymbol{\Lambda}, \bar{\boldsymbol{\Theta}})$ given observations $(\bar{\mathbf{A}}(s), \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h))$ can be expressed as

$$\mu(d\boldsymbol{\lambda}, d\bar{\boldsymbol{\theta}} \mid \bar{\mathbf{A}}(s), \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h)) \propto \exp(\bar{H}_n^{(tot)}(\boldsymbol{\lambda}, \bar{\boldsymbol{\theta}})) \mu_{\boldsymbol{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda}) \mu_{\bar{\boldsymbol{\Theta}}}^{\otimes d}(d\bar{\boldsymbol{\theta}}).$$

We define the free energy functionals corresponding to the total Hamiltonian as

$$\begin{aligned} \bar{\phi}_n(s, a, a', h) &:= \frac{1}{n} \log \int \exp(\bar{H}_n^{(tot)}(\boldsymbol{\lambda}, \bar{\boldsymbol{\theta}})) \mu_{\boldsymbol{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda}) \mu_{\bar{\boldsymbol{\Theta}}}^{\otimes d}(d\bar{\boldsymbol{\theta}}), \\ \bar{\Phi}_n(s, a, a', h) &:= \mathbb{E} [\bar{\phi}_n(s, a, a', h)], \end{aligned} \quad (50)$$

where $\mu_{\bar{\boldsymbol{\Theta}}}^{\otimes d}$ is the product distribution over \mathbb{R}^d with each coordinate distributed as $\bar{\boldsymbol{\Theta}}_0$.

Truncation does not change the asymptotic free energy density

Next, we show that truncation does not change the asymptotic free energy density.

Lemma D.12. *For $s, h \geq 0$, we define the following unperturbed Hamiltonian and free energy density*

$$\begin{aligned} H_n^{[s]}(\boldsymbol{\lambda}, \boldsymbol{\theta}) &:= \sum_{i \in [n], j \in [d]} \left\{ \frac{s}{\sqrt{nd}} \boldsymbol{\Lambda}_i \boldsymbol{\lambda}_i \boldsymbol{\Theta}_j \boldsymbol{\theta}_j + \frac{\sqrt{s}}{\sqrt[4]{nd}} Z_{ij} \boldsymbol{\lambda}_i \boldsymbol{\theta}_j - \frac{s}{2\sqrt{nd}} \boldsymbol{\lambda}_i^2 \boldsymbol{\theta}_j^2 \right\}, \\ \Phi_n(s, 0, 0, h) &:= \frac{1}{n} \mathbb{E} \left[\log \int \exp(H_n^{[s]}(\boldsymbol{\lambda}, \boldsymbol{\theta}) + H_n(\boldsymbol{\lambda}; \mathbf{Y}'(h))) \mu_{\boldsymbol{\Lambda}}^{\otimes n}(d\boldsymbol{\lambda}) \mu_{\bar{\boldsymbol{\Theta}}}^{\otimes d}(d\bar{\boldsymbol{\theta}}) \right]. \end{aligned}$$

Then for all fixed $S_0 > 0$, under the conditions of Theorem 4.5 (b), as $n, d \rightarrow \infty$ we have

$$\sup_{h \geq 0, S_0 \geq s \geq 0} |\Phi_n(s, 0, 0, h) - \bar{\Phi}_n(s, 0, 0, h)| = o_n(1).$$

The proof of Lemma D.12 is given in Appendix F.4. We further characterize the convergence of free energy density in Lemma D.13, again postponing the proof to Appendix F.5.

Lemma D.13. *Recall that $\mathcal{F}(\cdot, \cdot)$ is defined in Eq. (13). Under the conditions of Theorem 4.5 (b), if we further assume that $\varepsilon_n, \varepsilon'_n \rightarrow 0^+$, then for all fixed $s, h \geq 0$, as $n, d \rightarrow \infty$*

$$\lim_{n, d \rightarrow \infty} \sup_{a, a' \in [0, 10]} \left| \bar{\Phi}_n(s, a, a', h) - \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 s^2 + h, q) \right| = 0.$$

Recall that D is defined in Eq. (39). For all $h + q_\Theta^2 \in D$, the mapping $s \mapsto \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 s^2 + h, q)$ is differentiable at $s = 1$. Furthermore, for all fixed $a, a', h \geq 0$, the mappings $s \mapsto \bar{\Phi}_n(s, a, a', h)$, $s \mapsto \bar{\Phi}_n(s, 0, 0, h)$ are convex differentiable on $(0, \infty)$. By Lemmas D.12 and D.13, as $n, d \rightarrow \infty$, $\bar{\Phi}_n(s, a, a', h) \rightarrow \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 s^2 + h, q)$ and $\bar{\Phi}_n(s, 0, 0, h) \rightarrow \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 s^2 + h, q)$. Then we apply Lemma B.4 and conclude that for $h + q_\Theta^2 \in D$, we have

$$\left| \frac{\partial}{\partial s} \bar{\Phi}_n(s, a, a', h) \Big|_{s=1} - \frac{\partial}{\partial s} \bar{\Phi}_n(s, 0, 0, h) \Big|_{s=1} \right| = o_n(1).$$

Using Gaussian integration by parts and Nishimori identity, we further derive that for $h + q_\Theta^2 \in D$

$$\lim_{n, d \rightarrow \infty} \frac{1}{2n\sqrt{nd}} \mathbb{E} \left[\|\mathbb{E}[\mathbf{\Lambda} \bar{\Theta}^\top \mid \bar{\mathbf{A}}(1), \bar{\mathbf{x}}(a), \mathbf{x}'(a'), \mathbf{Y}'(h)] - \mathbb{E}[\mathbf{\Lambda} \bar{\Theta}^\top \mid \bar{\mathbf{A}}(1), \mathbf{Y}'(h)]\|_F^2 \right] = 0. \quad (51)$$

By Jensen's inequality, for all $a, a' \in [1, 2]$

$$\begin{aligned} & \frac{1}{2n\sqrt{nd}} \mathbb{E} \left[\|\mathbb{E}[\mathbf{\Lambda} \bar{\Theta}^\top \mid \bar{\mathbf{A}}(1), \bar{\mathbf{x}}(a), \mathbf{x}'(a'), \mathbf{Y}'(h)] - \mathbb{E}[\mathbf{\Lambda} \bar{\Theta}^\top \mid \bar{\mathbf{A}}(1), \mathbf{Y}'(h)]\|_F^2 \right] \\ & \leq \frac{1}{2n\sqrt{nd}} \mathbb{E} \left[\|\mathbb{E}[\mathbf{\Lambda} \bar{\Theta}^\top \mid \bar{\mathbf{A}}(1), \bar{\mathbf{x}}(2), \mathbf{x}'(2), \mathbf{Y}'(h)] - \mathbb{E}[\mathbf{\Lambda} \bar{\Theta}^\top \mid \bar{\mathbf{A}}(1), \mathbf{Y}'(h)]\|_F^2 \right]. \end{aligned}$$

By Eq. (51), the last line above converges to zero as $n, d \rightarrow \infty$ for all $q_\Theta^2 + h \in D$. In this case, we have

$$\lim_{n, d \rightarrow \infty} \int_1^2 \int_1^2 \frac{1}{2n\sqrt{nd}} \mathbb{E} \left[\|\mathbb{E}[\mathbf{\Lambda} \bar{\Theta}^\top \mid \bar{\mathbf{A}}(1), \bar{\mathbf{x}}(a), \mathbf{x}'(a'), \mathbf{Y}'(h)] - \mathbb{E}[\mathbf{\Lambda} \bar{\Theta}^\top \mid \bar{\mathbf{A}}(1), \mathbf{Y}'(h)]\|_F^2 \right] da da' = 0. \quad (52)$$

We define

$$\begin{aligned} \mathbf{\Lambda}_{s, a, a', h} &:= \mathbb{E}[\mathbf{\Lambda} \mid \bar{\mathbf{A}}(1), \bar{\mathbf{x}}(a), \mathbf{x}'(a'), \mathbf{Y}'(h)] \in \mathbb{R}^n, \\ \bar{\Theta}_{s, a, a', h} &:= \mathbb{E}[\bar{\Theta} \mid \bar{\mathbf{A}}(1), \bar{\mathbf{x}}(a), \mathbf{x}'(a'), \mathbf{Y}'(h)] \in \mathbb{R}^d, \\ \mathbf{M}_{s, a, a', h} &:= \mathbb{E}[\mathbf{\Lambda} \bar{\Theta}^\top \mid \bar{\mathbf{A}}(1), \bar{\mathbf{x}}(a), \mathbf{x}'(a'), \mathbf{Y}'(h)] \in \mathbb{R}^{n \times d}. \end{aligned}$$

Invoking Stein's lemma, we see that the following equation holds:

$$\frac{\partial}{\partial a} \mathbb{E} [\|\mathbf{\Lambda}_{1, a, a', h}\|^2] = \frac{2an\varepsilon_n}{d} \mathbb{E} [\|\mathbf{M}_{1, a, a', h} - \mathbf{\Lambda}_{1, a, a', h} \bar{\Theta}_{1, a, a', h}^\top\|_F^2].$$

Using the above equation, we obtain that

$$\begin{aligned} & \int_1^2 \int_1^2 \frac{1}{2n\sqrt{nd}} \mathbb{E} [\|\mathbf{M}_{1, a, a', h} - \mathbf{\Lambda}_{1, a, a', h} \bar{\Theta}_{1, a, a', h}^\top\|_F^2] da da' \\ & \leq \frac{d^{1/2}}{2\varepsilon_n n^{5/2}} \int_1^2 \{ \mathbb{E} [\|\mathbf{\Lambda}_{1, 2, a', h}\|^2] - \mathbb{E} [\|\mathbf{\Lambda}_{1, 1, a', h}\|^2] \} da' \\ & \leq \frac{d^{1/2} \mathbb{E}_{\mathbf{\Lambda}_0 \sim \mu_\Lambda} [\mathbf{\Lambda}_0^2]}{\varepsilon_n n^{3/2}}. \end{aligned} \quad (53)$$

Overlap concentration

The next lemmas show that if we draw two independent samples from the posterior distribution of $\bar{\Theta}$ (Λ) given $(\bar{\mathbf{A}}(1), \bar{\mathbf{x}}(a), \mathbf{x}'(a'), \mathbf{Y}'(h))$, then their normalized inner product concentrates. This phenomenon is referred to as *overlap concentration* in the literature of statistical mechanics. In what follows, we prove overlap concentration for $\bar{\Theta}$. Since the proof is similar, in order to avoid redundancy, we skip the counterpart proof for Λ . For the sake of simplicity, we denote by $\langle \cdot \rangle_{s,a,a',h}$ the expectation with respect to the posterior distribution $\mathbb{P}(\cdot \mid \bar{\mathbf{A}}(s), \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h))$.

Lemma D.14. *For $\bar{\theta} \in \mathbb{R}^d$, we define*

$$U(\bar{\theta}) := \sum_{j=1}^d \left\{ \frac{1}{\sqrt{\varepsilon_n n d}} \bar{\theta}_j g_j + \frac{2a}{d} \bar{\theta}_j \bar{\Theta}_j - \frac{a}{d} \bar{\theta}_j^2 \right\}.$$

Let $\bar{\theta}^{(1)}, \bar{\theta}^{(2)}, \bar{\theta} \in \mathbb{R}^d$ be independent samples drawn from the posterior distribution $\mathbb{P}(\bar{\Theta} = \cdot \mid \bar{\mathbf{A}}(1), \bar{\mathbf{x}}(a), \mathbf{x}'(a'), \mathbf{Y}'(h))$. Then under the conditions of Theorem 4.5 (b), for all $a, a' \in [0.1, 10]$ and $h \geq 0$, we have

$$\begin{aligned} & \mathbb{E}[\langle (\bar{\theta}^{(1)})^\top \bar{\theta}^{(2)} / d - \mathbb{E}[\langle (\bar{\theta}^{(1)})^\top \bar{\theta}^{(2)} / d \rangle_{1,a,a',h}] \rangle_{1,a,a',h}^2] \\ & \leq 40K_2^2 \log d \mathbb{E}[\langle |U(\bar{\theta}) - \mathbb{E}[U(\bar{\theta})]_{1,a,a',h}| \rangle_{1,a,a',h}]. \end{aligned} \quad (54)$$

Using Eq. (54), we see that in order to prove overlap concentration, we only need to show that the right hand side of Eq. (54) is sufficiently small, which is accomplished via the following lemmas:

Lemma D.15. *We let $\bar{\theta} \in \mathbb{R}^d$ be a sample drawn from the posterior distribution $\mathbb{P}(\bar{\Theta} = \cdot \mid \bar{\mathbf{A}}(1), \bar{\mathbf{x}}(a), \mathbf{x}'(a'), \mathbf{Y}'(h))$. For $h \geq 0$, we define*

$$v_n(h) := \sup_{1/2 \leq a, a' \leq 3} \mathbb{E}[|\bar{\phi}_n(1, a, a', h) - \mathbb{E}[\bar{\phi}_n(1, a, a', h)]|],$$

where we recall that $\bar{\phi}_n$ is defined in Eq. (50). Then under the conditions of Theorem 4.5 (b), if we further assume that $\varepsilon_n \rightarrow 0^+$ and $nd^{-1/2}\varepsilon_n \rightarrow \infty$ as $n, d \rightarrow \infty$, then there exists a numerical constant $C > 0$, such that for n, d large enough

$$\int_1^2 \int_1^2 \mathbb{E}[\langle |U(\bar{\theta}) - \mathbb{E}[U(\bar{\theta})]_{1,a,a',h}| \rangle_{1,a,a',h}] da da' \leq CK_2 \sqrt{(v_n(h) + n^{-1})\varepsilon_n^{-1} \log d}.$$

Lemma D.16. *Under the conditions of Theorem 4.5 (b), if we further assume $\varepsilon_n, \varepsilon'_n \rightarrow 0^+$ as $n, d \rightarrow \infty$, then there exists a numerical constant $C_1 > 0$ such that for all n, d large enough and $0 \leq h \leq 1$*

$$v_n(h) \leq C_1 K_1^2 K_2^2 d^{1/2} n^{-1} \log d.$$

We defer the proofs of Lemmas D.14 to D.16 to Sections F.6 to F.8, respectively. Combining Lemmas D.14 to D.16, we deduce that under the conditions of these lemmas, for $\bar{\theta}^{(1)}, \bar{\theta}^{(2)} \in \mathbb{R}^d$ that are two independent samples from the posterior distribution $\mathbb{P}(\bar{\Theta} = \cdot \mid \bar{\mathbf{A}}(1), \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h))$, there exists a numerical constant $C > 0$, such that for n, d large enough

$$\frac{d}{n} \int_1^2 \int_1^2 \mathbb{E}[\langle (\bar{\theta}_1^\top \bar{\theta}_2 / d - \mathbb{E}[\langle \bar{\theta}_1^\top \bar{\theta}_2 / d \rangle_{1,a,a',h}])^2 \rangle_{1,a,a',h}] da da' \leq CK_1 K_2^4 \varepsilon_n^{-1/2} (\log d)^2 d^{5/4} n^{-3/2}.$$

Under condition (b), we see that there exists $\varepsilon_n \rightarrow 0^+$, such that $\varepsilon_n^{-1/2} (\log d)^2 d^{5/4} n^{-3/2} \rightarrow 0$ and $nd^{-1/2}\varepsilon_n \rightarrow \infty$ as $n, d \rightarrow \infty$. We summarize the overlap concentration results in Theorem D.1 below, which also contains concentration argument for Λ (that we skip the proof).

Theorem D.1 (Overlap concentration). *Let $\bar{\boldsymbol{\theta}}^{(1)}, \bar{\boldsymbol{\theta}}^{(2)} \in \mathbb{R}^d$ be two independent samples drawn from the posterior distribution $\mathbb{P}(\bar{\boldsymbol{\Theta}} = \cdot \mid \bar{\mathbf{A}}(1), \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h))$, and $\boldsymbol{\lambda}^{(1)}, \boldsymbol{\lambda}^{(2)} \in \mathbb{R}^n$ be two independent samples drawn from the posterior distribution $\mathbb{P}(\boldsymbol{\Lambda} = \cdot \mid \bar{\mathbf{A}}(1), \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h))$. Under the conditions of Theorem 4.5 (b), there exist $\varepsilon_n, \varepsilon'_n \rightarrow 0^+$ such that for all $h \in [0, 1]$, as $n, d \rightarrow \infty$*

$$\begin{aligned} \frac{(\log d)^2 d^{5/4}}{n^{3/2} \varepsilon_n^{1/2}} &\rightarrow 0, \quad \frac{n \varepsilon_n}{d^{1/2}} \rightarrow \infty, \quad \frac{(\log d)^{1/2} d^{1/4}}{n^{1/2} \varepsilon'_n^{1/2}} \rightarrow 0, \quad \frac{n \varepsilon'_n}{d^{1/2} \log d} \rightarrow \infty, \\ \frac{d}{n} \int_1^2 \int_1^2 \mathbb{E}[\langle (\bar{\boldsymbol{\theta}}_1^\top \bar{\boldsymbol{\theta}}_2 / d - \mathbb{E}[\langle \bar{\boldsymbol{\theta}}_1^\top \bar{\boldsymbol{\theta}}_2 / d \rangle_{1,a,a',h}])^2 \rangle_{1,a,a',h}] da da' &\rightarrow 0, \\ \int_1^2 \int_1^2 \mathbb{E}[\langle (\boldsymbol{\lambda}_1^\top \boldsymbol{\lambda}_2 / n - \mathbb{E}[\langle \boldsymbol{\lambda}_1^\top \boldsymbol{\lambda}_2 / n \rangle_{1,a,a',h}])^2 \rangle_{1,a,a',h}] da da' &\rightarrow 0. \end{aligned}$$

Corollary D.1. *Under the conditions of Theorem D.1, for all $h \in [0, 1]$, as $n, d \rightarrow \infty$ we have*

$$\begin{aligned} \frac{d}{n} \int_1^2 \int_1^2 \mathbb{E}[\langle (\langle \bar{\boldsymbol{\theta}}_1^\top \bar{\boldsymbol{\theta}}_2 / d \rangle_{1,a,a',h} - \mathbb{E}[\langle \bar{\boldsymbol{\theta}}_1^\top \bar{\boldsymbol{\theta}}_2 / d \rangle_{1,a,a',h}])^2 \rangle_{1,a,a',h}] da da' &\rightarrow 0, \\ \int_1^2 \int_1^2 \mathbb{E}[\langle (\langle \boldsymbol{\lambda}_1^\top \boldsymbol{\lambda}_2 / n \rangle_{1,a,a',h} - \mathbb{E}[\langle \boldsymbol{\lambda}_1^\top \boldsymbol{\lambda}_2 / n \rangle_{1,a,a',h}])^2 \rangle_{1,a,a',h}] da da' &\rightarrow 0. \end{aligned}$$

Remark D.2. In Theorem D.1 and Corollary D.1, we can replace the interval $[1, 2]$ with $[a, b]$ for any fixed $0 < a < b < \infty$.

Proof of the theorem

In the rest parts of the proof, we always assume that $\{\varepsilon_n\}_{n \in \mathbb{N}_+}$ and $\{\varepsilon'_n\}_{n \in \mathbb{N}_+}$ are chosen as in Theorem D.1. Under this assumption we have $\varepsilon_n^{-1} (\log d)^4 d^{5/2} n^{-3} \rightarrow 0$ and $d^{-2} n^{3/2} (\log d)^{-4} \rightarrow 0$, thus $d^{1/2} n^{-3/2} \varepsilon_n^{-1} \rightarrow 0$ as $n, d \rightarrow \infty$. Plugging this result into Eqs. (52) and (53), we obtain that

$$\lim_{n, d \rightarrow \infty} \int_1^2 \int_1^2 \frac{1}{2n\sqrt{nd}} \mathbb{E} [\|\boldsymbol{\Lambda}_{1,a,a',h} \bar{\boldsymbol{\Theta}}_{1,a,a',h} - \mathbf{M}_{1,0,0,h}\|_F^2] da da' = 0. \quad (55)$$

By Lemma D.12 and Lemma D.13 we see that $|\Phi_n(s, 0, 0, h) - \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 s^2 + h, q)| = o_n(1)$ for all fixed $s, h \geq 0$. Notice that $s \mapsto \Phi_n(s, 0, 0, h)$ is convex and differentiable. Furthermore, for all but countably many values of $q_\Theta^2 + h > 0$, the mapping $s \mapsto \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 s^2 + h, q)$ is differentiable at $s = 1$. Using Gaussian integration by parts, Lemmas B.3 and B.4 we conclude that for these $q_\Theta^2 + h$ we have

$$\lim_{n, d \rightarrow \infty} \frac{1}{2n\sqrt{nd}} \mathbb{E} [\|\mathbb{E}[\boldsymbol{\Lambda} \bar{\boldsymbol{\Theta}}^\top \mid \bar{\mathbf{A}}(1), \mathbf{Y}'(h)]\|_F^2] = \frac{\partial}{\partial s} \sup_{q \geq 0} \mathcal{F}(s^2 q_\Theta^2 + h, q) \Big|_{s=1}. \quad (56)$$

Define $D_\Theta(h) := \frac{\partial}{\partial s} \sup_{q \geq 0} \mathcal{F}(s^2 q_\Theta^2 + h, q) \Big|_{s=1}$. For all but countably many $q_\Theta > 0$ the mapping $s \mapsto \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 s^2, q)$ is differentiable at $s = 1$, thus $D_\Theta(0)$ is well-defined. In this case, if $D_\Theta(0) = 0$, then by [LM19] we have $\lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_\Lambda, q_\Theta) = \mathbb{E}_{\boldsymbol{\Lambda}_0 \sim \mu_\Lambda} [\boldsymbol{\Lambda}_0^2]^2$, which is achieved by $\mathbf{0}_{n \times n}$. Using Theorem 4.4 we deduce that $\lim_{n, d \rightarrow \infty} \text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta) = \mathbb{E}_{\boldsymbol{\Lambda}_0 \sim \mu_\Lambda} [\boldsymbol{\Lambda}_0^2]^2$, which concludes the proof of the theorem.

In the following parts of the proof we will assume $D_\Theta(0) > 0$. Let $a_1 < a_2$, $a'_1 < a'_2$, and $a \sim \text{Unif}[a_1, a_2]$ and $a' \sim \text{Unif}[a'_1, a'_2]$. Similar to the derivation of Eq. (55) we have

$$\lim_{n,d \rightarrow \infty} \frac{1}{2n\sqrt{nd}} \mathbb{E} [\|\mathbf{\Lambda}_{1,a,a',h} \bar{\mathbf{\Theta}}_{1,a,a',h} - \mathbf{M}_{1,0,0,h}\|_F^2] = 0, \quad (57)$$

where the expectation is taken over $a \sim \text{Unif}[a_1, a_2]$ and $a' \sim \text{Unif}[a'_1, a'_2]$. By Eq. (56), Eq. (57) and triangle inequality, for all but countably many $q_\Theta^2, q_\Theta^2 + h$ we have

$$\begin{aligned} & \limsup_{n,d \rightarrow \infty} \left\{ \frac{1}{2\sqrt{n^3 d}} \mathbb{E} [\|\mathbf{\Lambda}_{1,a,a',h} \bar{\mathbf{\Theta}}_{1,a,a',h} - \mathbf{M}_{1,0,0,0}\|_F^2] \right\}^{1/2} \\ & \leq \frac{1}{2\sqrt{n^3 d}} \left(\limsup_{n,d \rightarrow \infty} \left\{ \mathbb{E} [\|\mathbf{\Lambda}_{1,a,a',h} \bar{\mathbf{\Theta}}_{1,a,a',h} - \mathbf{M}_{1,0,0,h}\|_F^2] \right\}^{1/2} + \limsup_{n,d \rightarrow \infty} \left\{ \mathbb{E} [\|\mathbf{M}_{1,0,0,h} - \mathbf{M}_{1,0,0,0}\|_F^2] \right\}^{1/2} \right) \\ & = (D_\Theta(h) - D_\Theta(0))^{1/2}. \end{aligned} \quad (58)$$

The following lemmas establish several useful properties of $\mathbf{\Lambda}_{1,a,a',h}$ and $\bar{\mathbf{\Theta}}_{1,a,a',h}$.

Lemma D.17. *Recall that D is defined in Eq. (39). For all fixed $a_*, a'_* \in (0, 5]$, $h \in [0, 1]$, under the assumptions of Theorem 4.5 (b), if we further assume that $h + q_\Theta^2 \in D$, then as $n, d \rightarrow \infty$ we have*

$$\begin{aligned} \frac{1}{n} \mathbb{E} [\|\mathbb{E}[\mathbf{\Lambda} \mid \bar{\mathbf{A}}(1), \mathbf{x}'(a'_*), \bar{\mathbf{x}}(a_*), \mathbf{Y}'(h)]\|^2] &= 2 \left(\frac{\partial}{\partial h} \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q) \right)^{1/2} + o_n(1), \\ \frac{1}{\sqrt{nd}} \mathbb{E} [\|\mathbb{E}[\bar{\mathbf{\Theta}} \mid \bar{\mathbf{A}}(1), \mathbf{x}'(a'_*), \bar{\mathbf{x}}(a_*), \mathbf{Y}'(h)]\|^2] &= 2q_\Theta^2 \left(\frac{\partial}{\partial h} \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q) \right)^{1/2} + o_n(1). \end{aligned}$$

The proof of Lemma D.17 is deferred to Sections F.9 and F.10, respectively. By Lemma D.17 and Corollary D.1, we conclude that for all $0 < a_1 < a_2 < 5$ and $0 < a'_1 < a'_2 < 5$, if we let $a \sim \text{Unif}[a_1, a_2]$ and $a' \sim \text{Unif}[a'_1, a'_2]$, then for all $h + q_\Theta^2 \in D$ we have

$$\frac{1}{\sqrt{nd}} \|\mathbb{E}[\bar{\mathbf{\Theta}} \mid \bar{\mathbf{A}}(1), \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h)]\|^2 = 2q_\Theta^2 \left(\frac{\partial}{\partial h} \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q) \right)^{1/2} + o_P(1). \quad (59)$$

Let $C_0(q_\Theta, h) := 2q_\Theta^2 \left(\frac{\partial}{\partial h} \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q) \right)^{1/2}$. We define the mapping $\mathbf{M} : \mathbb{R}^{n \times n} \times \mathbb{R}^+ \mapsto \mathbb{R}^{n \times n}$, such that $\mathbf{M}(\mathbf{X}; b)_{ij} = X_{ij} \mathbb{1}\{|X_{ij}| \leq b\}$. We further define $\mathbf{M}_0 : \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^{n \times n}$ such that for $\mathbf{X} \in \mathbb{R}^{n \times d}$

$$\mathbf{M}_0(\mathbf{X}) = \mathbf{M} \left(\frac{1}{\sqrt{nd}} \mathbb{E}[\mathbf{\Lambda} \bar{\mathbf{\Theta}}^\top \mid \bar{\mathbf{A}} = \mathbf{X}] \mathbb{E}[\bar{\mathbf{\Theta}} \bar{\mathbf{\Lambda}}^\top \mid \bar{\mathbf{A}} = \mathbf{X}]; 2K_1^2 C_0(q_\Theta, h) \right).$$

By triangle inequality

$$\begin{aligned} & \frac{1}{n} \mathbb{E} [\|\mathbf{M}_0(\bar{\mathbf{A}}) - C_0(q_\Theta, h) \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}, \mathbf{Y}'(h)]\|_F^2]^{1/2} \\ & \leq \frac{C_0(q_\Theta, h)}{n} \mathbb{E} [\|\mathbf{\Lambda}_{1,a,a',h} \mathbf{\Lambda}_{1,a,a',h}^\top - \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}, \mathbf{Y}'(h)]\|_F^2]^{1/2} \\ & \quad + \frac{1}{n} \mathbb{E} [\|C_0(q_\Theta, h) \mathbf{\Lambda}_{1,a,a',h} \mathbf{\Lambda}_{1,a,a',h}^\top - \mathbf{M}_0(\bar{\mathbf{A}})\|_F^2]^{1/2}, \end{aligned} \quad (60)$$

where the expectation is taken over $(a, a', \mathbf{\Lambda}, \bar{\mathbf{\Theta}}, \mathbf{Z}, \mathbf{g}, \mathbf{g}', \mathbf{W}')$. Using Gaussian integration by parts we obtain

$$\frac{\partial}{\partial a'} \mathbb{E} [\|\mathbf{\Lambda}_{1,a,a',h}\|^2] = \varepsilon'_n a' \mathbb{E} \left[\|\mathbb{E}[\mathbf{\Lambda}\mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}, \bar{\mathbf{x}}(a), \bar{\mathbf{x}}'(a'), \mathbf{Y}'(h)] - \mathbf{\Lambda}_{1,a,a',h} \mathbf{\Lambda}_{1,a,a',h}^\top\|_F^2 \right]. \quad (61)$$

By Eq. (61) we have

$$\begin{aligned} & \frac{1}{n^2} \mathbb{E} \left[\|\mathbf{\Lambda}_{1,a,a',h} \mathbf{\Lambda}_{1,a,a',h} - \mathbb{E}[\mathbf{\Lambda}\mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}, \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h)]\|_F^2 \right] \\ &= \frac{1}{n^2(a_2 - a_1)(a'_2 - a'_1)} \int_{a_1}^{a_2} \int_{a'_1}^{a'_2} (\varepsilon'_n a')^{-1} \frac{\partial}{\partial a'} \mathbb{E} [\|\mathbf{\Lambda}_{1,a,a',h}\|^2] da' da \\ &\leq \frac{1}{n^2 \varepsilon'_n a'_1 (a_2 - a_1)(a'_2 - a'_1)} \int_{a_1}^{a_2} \mathbb{E} [\|\mathbf{\Lambda}_{1,a,a'_2,h}\|^2] da \\ &\leq \frac{K_1^2}{n(a'_2 - a'_1) \varepsilon'_n a'_1}, \end{aligned} \quad (62)$$

which vanishes as $n, d \rightarrow \infty$. For $\mathbf{X} \in \mathbb{R}^{n \times n}$, we define $\|\mathbf{X}\|_1 := \sum_{i,j \in [n]} |X_{ij}|$. Then we have

$$\begin{aligned} & \frac{1}{n^2} \mathbb{E} \left[\|C_0(q_\Theta, h) \mathbf{\Lambda}_{1,a,a',h} \mathbf{\Lambda}_{1,a,a',h}^\top - \mathbf{M}_0(\bar{\mathbf{A}})\|_F^2 \right] \\ &\stackrel{(i)}{\leq} \frac{3K_1^2 C_0(q_\Theta, h)}{n^2} \mathbb{E} \left[\|C_0(q_\Theta, h) \mathbf{\Lambda}_{1,a,a',h} \mathbf{\Lambda}_{1,a,a',h}^\top - \mathbf{M}_0(\bar{\mathbf{A}})\|_1 \right] \\ &\stackrel{(ii)}{\leq} \frac{3K_1^2 C_0(q_\Theta, h)}{n^2} \mathbb{E} \left[\left\| C_0(q_\Theta, h) \mathbf{\Lambda}_{1,a,a',h} \mathbf{\Lambda}_{1,a,a',h}^\top - \frac{1}{\sqrt{nd}} \mathbf{M}_{1,0,0,0} \mathbf{M}_{1,0,0,0}^\top \right\|_1 \right] \\ &\stackrel{(iii)}{\leq} \frac{3K_1^2 C_0(q_\Theta, h)}{n^2} \mathbb{E} \left[\left\| \frac{\|\bar{\mathbf{\Theta}}_{1,a,a',h}\|^2}{\sqrt{nd}} \mathbf{\Lambda}_{1,a,a',h} \mathbf{\Lambda}_{1,a,a',h}^\top - \frac{1}{\sqrt{nd}} \mathbf{M}_{1,0,0,0} \mathbf{M}_{1,0,0,0}^\top \right\|_1 \right] \end{aligned} \quad (63)$$

$$+ 3K_1^4 C_0(q_\Theta, h) \mathbb{E} \left[\left| \frac{\|\bar{\mathbf{\Theta}}_{1,a,a',h}\|^2}{\sqrt{nd}} - C_0(q_\Theta, h) \right| \right], \quad (64)$$

where in (i) we use the assumption that μ_Λ has bounded support, and (ii) is by the fact that for all $|x| \leq C_0(q_\Theta, h) K_1^2$, $y \in \mathbb{R}$,

$$|x - y \mathbb{1}\{|y| \leq 2K_1^2 C_0(q_\Theta, h)\}| \leq |x - y|.$$

Lastly, (iii) is by triangle inequality. Applying Lemma B.6 and Hölder's inequality we see that

$$\begin{aligned} & \frac{3K_1^2 C_0(q_\Theta, h)}{n^2} \times \mathbb{E} \left[\left\| \frac{\|\bar{\mathbf{\Theta}}_{1,a,a',h}\|^2}{\sqrt{nd}} \mathbf{\Lambda}_{1,a,a',h} \mathbf{\Lambda}_{1,a,a',h}^\top - \frac{1}{\sqrt{nd}} \mathbf{M}_{1,0,0,0} \mathbf{M}_{1,0,0,0}^\top \right\|_1 \right] \\ &\leq \frac{6K_1^2 C_0(q_\Theta, h)}{n\sqrt{nd}} \times \mathbb{E} \left[\left\| \mathbf{M}_{1,0,0,0} - \mathbf{\Lambda}_{1,a,a',h} \bar{\mathbf{\Theta}}_{1,a,a',h}^\top \right\|_F^2 \right]^{1/2} \\ &\quad \times \left(\mathbb{E} \left[\left\| \mathbf{M}_{1,0,0,0} \right\|_F^2 \right]^{1/2} + \mathbb{E} \left[\left\| \mathbf{\Lambda}_{1,a,a',h} \bar{\mathbf{\Theta}}_{1,a,a',h}^\top \right\|_F^2 \right]^{1/2} \right). \end{aligned} \quad (65)$$

By Lemma D.13, for all $a, a' \in [0, 5]$, as $n, d \rightarrow \infty$ we have $\bar{\Phi}_n(1, a, a', h) \rightarrow \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q)$. Notice that the mapping $h \mapsto \bar{\Phi}_n(1, a, a', h)$ is convex and differentiable, thus for $q_\Theta^2 + h \in D$ we can apply Lemma B.4 and conclude that

$$\lim_{n,d \rightarrow \infty} \frac{1}{4n^2} \mathbb{E} [\|\mathbb{E}[\mathbf{\Lambda}\mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}, \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h)]\|_F^2] = \frac{\partial}{\partial h} \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q). \quad (66)$$

Leveraging triangle inequality, Eqs. (56) and (58), we obtain that for all $a, a' \in [0, 5]$, $q_\Theta^2 + h, q_\Theta^2 \in D$

$$\begin{aligned} \limsup_{n,d \rightarrow \infty} \frac{1}{n\sqrt{nd}} \mathbb{E}[\|\mathbf{M}_{1,0,0,0}\|_F^2] &= 2D_\Theta(0), \\ \limsup_{n,d \rightarrow \infty} \frac{1}{n\sqrt{nd}} \mathbb{E}[\|\mathbf{\Lambda}_{1,a,a',h} \bar{\mathbf{\Theta}}_{1,a,a',h}^\top\|_F^2] &= 2D_\Theta(h), \\ \lim_{n,d \rightarrow \infty} \frac{1}{n\sqrt{nd}} \mathbb{E}[\|\mathbf{M}_{1,0,0,0} - \mathbf{\Lambda}_{1,a,a',h} \bar{\mathbf{\Theta}}_{1,a,a',h}^\top\|_F^2] &= 2(D_\Theta(h) - D_\Theta(0)). \end{aligned} \quad (67)$$

By Theorem D.1 and Eq. (59) we have

$$\limsup_{n,d \rightarrow \infty} \mathbb{E} \left[\left\| \frac{1}{\sqrt{nd}} \|\bar{\mathbf{\Theta}}_{1,a,a',h}\|^2 - C_0(q_\Theta, h) \right\| \right] = 0. \quad (68)$$

We plug Eqs. (65), (67) and (68) into Eq. (63) and obtain that

$$\begin{aligned} \limsup_{n,d \rightarrow \infty} \frac{1}{n^2} \mathbb{E} \left[\|C_0(q_\Theta, h) \mathbf{\Lambda}_{1,a,a',h} \mathbf{\Lambda}_{1,a,a',h}^\top - \mathbf{M}_0(\bar{\mathbf{A}})\|_F^2 \right] \\ \leq 24K_1^2 C_0(q_\Theta, h) \times (D_\Theta(h) - D_\Theta(0))^{1/2}. \end{aligned} \quad (69)$$

Using triangle inequality

$$\begin{aligned} &\frac{1}{n} \mathbb{E} \left[\|\mathbf{\Lambda} \mathbf{\Lambda}^\top - C_0(q_\Theta, h)^{-1} \mathbf{M}_0(\bar{\mathbf{A}})\|_F^2 \right]^{1/2} \\ &\leq \frac{1}{n} \mathbb{E} \left[\|\mathbf{\Lambda} \mathbf{\Lambda}^\top - \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}, \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h)]\|_F^2 \right]^{1/2} + \frac{1}{n} \mathbb{E} \left[\|\mathbf{\Lambda}_{1,a,a',h} \mathbf{\Lambda}_{1,a,a',h}^\top - C_0(q_\Theta, h)^{-1} \mathbf{M}_0(\bar{\mathbf{A}})\|_F^2 \right]^{1/2} \\ &\quad + \frac{1}{n} \mathbb{E} \left[\|\mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}, \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h)] - \mathbf{\Lambda}_{1,a,a',h} \mathbf{\Lambda}_{1,a,a',h}^\top\|_F^2 \right]^{1/2}. \end{aligned}$$

We plug Eqs. (62), (66) and (69) into the above equation and conclude that

$$\begin{aligned} &\limsup_{n,d \rightarrow \infty} \frac{1}{n^2} \mathbb{E} \left[\|\mathbf{\Lambda} \mathbf{\Lambda}^\top - C_0(q_\Theta, h)^{-1} \hat{\mathbf{M}}(\bar{\mathbf{A}})\|_F^2 \right] \\ &\leq \mathbb{E}_{\mathbf{\Lambda}_0 \sim \mu_\Lambda} [\mathbf{\Lambda}_0^2] - 4 \frac{\partial}{\partial h} \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q) + 2K_1^2 \sqrt{24K_1^2 C_0(q_\Theta, h)^{-1} (D_\Theta(h) - D_\Theta(0))^{1/2}} \\ &\quad + 24K_1^2 C_0(q_\Theta, h)^{-1} (D_\Theta(h) - D_\Theta(0))^{1/2}, \end{aligned}$$

which is an upper bound for $\limsup_{n,d \rightarrow \infty} \mathbb{E} \left[\|\mathbf{\Lambda} \mathbf{\Lambda}^\top - \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}]\|_F^2 \right] / n^2$. Recall that $D_\Theta(0) > 0$, thus $C_0(q_\Theta, h)^{-1} < C_0(q_\Theta, 0)^{-1} < \infty$. For all but countably many $q_\Theta > 0$ the mapping $h \mapsto D_\Theta(h)$ is continuous at 0. For these q_Θ , if we take $h \rightarrow 0^+$ while maintaining $h + q_\Theta^2 \in D$ then we derive that

$$\limsup_{n,d \rightarrow \infty} \frac{1}{n^2} \mathbb{E} \left[\|\mathbf{\Lambda} \mathbf{\Lambda}^\top - \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}]\|_F^2 \right] \leq \mathbb{E}_{\mathbf{\Lambda}_0 \sim \mu_\Lambda} [\mathbf{\Lambda}_0^2] - 4 \frac{\partial}{\partial h} \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q) \Big|_{h=0},$$

thus concludes the proof of the theorem using Lemma D.11.

D.5.3 Proof of Theorem 4.5 under condition (c)

We define $\tilde{\mathbf{Y}} = (\mathbf{A} \mathbf{A}^\top - d \mathbf{I}_n) / \sqrt{nd}$ and $\tilde{\mathbf{Y}}' = q_\Theta \mathbf{\Lambda} \mathbf{\Lambda}^\top / n + (\mathbf{Z}^\top \mathbf{Z} - d \mathbf{I}_n) / \sqrt{nd}$. One can verify that as $n, d \rightarrow \infty$ we have $\|\tilde{\mathbf{Y}} - \tilde{\mathbf{Y}}'\|_{op} = o_P(1)$. We then run rotationally invariant AMP with spectral initialization based on $\tilde{\mathbf{Y}}$. According to [MV21a], for large enough number of iterations this algorithm achieves Bayesian MMSE, thus completing the proof of the theorem under condition (c).

Appendix E Convergence of free energy density

E.1 Proof of Lemma D.2

For $k \in [d]$, we define

$$H_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) := \frac{1}{\sqrt{nd}} \sum_{i=1}^n \sum_{j=1, j \neq k}^d \Lambda_i \lambda_i \Theta_j \theta_j + \frac{1}{\sqrt{nd}} \sum_{i=1}^n \sum_{j=1, j \neq k}^d Z_{ij} \lambda_i \theta_j - \frac{1}{2\sqrt{nd}} \sum_{i=1}^n \sum_{j=1, j \neq k}^d \lambda_i^2 \theta_j^2 \\ + H_n(\boldsymbol{\lambda}; \mathbf{Y}'(h)) + H_n(\boldsymbol{\lambda}; \mathbf{x}'(s)).$$

Furthermore, we introduce the following distributions

$$\mu_n^{(k,+)}(d\boldsymbol{\lambda}, d\boldsymbol{\theta}) := \frac{1}{Z_n^{(k,+)}} \exp\left(H_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})\right) \mu_\Lambda^{\otimes n}(d\boldsymbol{\lambda}) P_{\Theta,k}(d\boldsymbol{\theta}), \\ Z_n^{(k,+)} := \int \exp\left(H_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})\right) \mu_\Lambda^{\otimes n}(d\boldsymbol{\lambda}) P_{\Theta,k}(d\boldsymbol{\theta}), \\ \mu_n^{(k,-)}(d\boldsymbol{\lambda}, d\boldsymbol{\theta}) := \frac{1}{Z_n^{(k,-)}} \exp\left(H_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})\right) \mu_\Lambda^{\otimes n}(d\boldsymbol{\lambda}) P_{\Theta,k-1}(d\boldsymbol{\theta}), \\ Z_n^{(k,-)} := \int \exp\left(H_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})\right) \mu_\Lambda^{\otimes n}(d\boldsymbol{\lambda}) P_{\Theta,k-1}(d\boldsymbol{\theta}).$$

Note that $\mu_n^{(k,+)}$, $\mu_n^{(k,-)}$, $Z_n^{(k,+)}$ and $Z_n^{(k,-)}$ are all random objects. The following lemma is a straightforward consequence of the above definitions.

Lemma E.1. *The following statements are true for all $k \in [d]$:*

1. $Z_n^{(k,+)} = Z_n^{(k,-)}$.
2. We denote by $\mathbf{Z}_{\cdot,k} \in \mathbb{R}^n$ the k -th column of \mathbf{Z} , then $(\mu_n^{(k,+)}, \mu_n^{(k,-)}, Z_n^{(k,+)}, Z_n^{(k,-)})$ are independent of $(\mathbf{Z}_{\cdot,k}, \boldsymbol{\Theta}_k)$.
3. We let $\boldsymbol{\theta}_{-k} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_{k-1}, \boldsymbol{\theta}_{k+1}, \dots, \boldsymbol{\theta}_d) \in \mathbb{R}^{d-1}$. For $(\boldsymbol{\lambda}, \boldsymbol{\theta}) \sim \mu_n^{(k,+)}$, we have $\boldsymbol{\theta}_k \sim \mu_\Theta$ and is independent of $(\boldsymbol{\theta}_{-k}, \boldsymbol{\lambda})$. Similarly, for $(\boldsymbol{\lambda}, \boldsymbol{\theta}) \sim \mu_n^{(k,-)}$, we have $\boldsymbol{\theta}_k \sim \mathcal{N}(0, q_\Theta)$, and is independent of $(\boldsymbol{\theta}_{-k}, \boldsymbol{\lambda})$.

We define

$$h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) := \frac{1}{\sqrt{nd}} \sum_{i=1}^n \Lambda_i \lambda_i \Theta_k \theta_k + \frac{1}{\sqrt{nd}} \sum_{i=1}^n Z_{ik} \lambda_i \theta_k - \frac{1}{2\sqrt{nd}} \sum_{i=1}^n \lambda_i^2 \theta_k^2.$$

For some random variable X , we denote by $\mu_n^{(k,+)}[X]$, $\mu_n^{(k,-)}[X]$ the expectations of X evaluated under distributions $\mu_n^{(k,+)}$ and $\mu_n^{(k,-)}$, respectively. Using Lemma E.1, we obtain that

$$\Phi_n^{(k)} - \Phi_n^{(k-1)} \\ = \frac{1}{n} \mathbb{E} \left[\log \left(\int \exp(h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})) \mu_n^{(k,+)}(d\boldsymbol{\lambda}, d\boldsymbol{\theta}) Z_n^{(k,+)} \right) \right] - \frac{1}{n} \mathbb{E} \left[\log \left(\int \exp(h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})) \mu_n^{(k,-)}(d\boldsymbol{\lambda}, d\boldsymbol{\theta}) Z_n^{(k,-)} \right) \right] \\ \stackrel{(i)}{=} \frac{1}{n} \mathbb{E} \left[\log \left(\mu_n^{(k,+)} \left[\exp(h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})) \right] \right) \right] - \frac{1}{n} \mathbb{E} \left[\log \left(\mu_n^{(k,-)} \left[\exp(h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})) \right] \right) \right], \quad (70)$$

where (i) is by result 1 in Lemma E.1. We consider the following Taylor expansion:

$$\exp\left(h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})\right) = \sum_{l=0}^{\infty} \frac{1}{l!} h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})^l = 1 + \sum_{l=1}^4 c_l^{(k)} \boldsymbol{\theta}_k^l + R^{(k)} + \sum_{l=5}^{\infty} \frac{1}{l!} h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})^l, \quad (71)$$

where

$$\begin{aligned} c_1^{(k)} &= \frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k + \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle, \\ c_2^{(k)} &= \frac{1}{2} \left(\frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k + \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \right)^2 - \frac{1}{2\sqrt{nd}} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle, \\ c_3^{(k)} &= -\frac{1}{2nd} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k - \frac{1}{2n^{3/4}d^{3/4}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle + \frac{1}{6} \left(\frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k + \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \right)^3, \\ c_4^{(k)} &= \frac{1}{8nd} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^2 - \frac{1}{4\sqrt{nd}} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle \left(\frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k + \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \right)^2 \\ &\quad + \frac{1}{24} \left(\frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k + \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \right)^4, \\ R^{(k)} &= -\frac{\langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^3}{48(nd)^{3/2}} \boldsymbol{\theta}_k^6 + \frac{1}{8nd} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^2 \left(\frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k + \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \right) \boldsymbol{\theta}_k^5 + \frac{1}{384n^2d^2} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^4 \boldsymbol{\theta}_k^8 \\ &\quad - \frac{1}{12\sqrt{nd}} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle \left(\frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k + \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \right)^3 \boldsymbol{\theta}_k^5 \\ &\quad + \frac{1}{16nd} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^2 \left(\frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k + \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \right)^2 \boldsymbol{\theta}_k^6 \\ &\quad - \frac{1}{48(nd)^{3/2}} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^3 \left(\frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k + \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \right) \boldsymbol{\theta}_k^7. \end{aligned}$$

The next lemma characterizes convergence of power series:

Lemma E.2. *For n, d large enough, the following quantities almost surely exist and are finite:*

$$\sum_{l=5}^{\infty} \frac{1}{l!} \mu_n^{(k,+)} \left[|h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})|^l \right], \quad \sum_{l=5}^{\infty} \frac{1}{l!} \mu_n^{(k,-)} \left[|h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})|^l \right].$$

Proof. We will only prove the lemma for $\mu_n^{(k,-)}$. Proof for $\mu_n^{(k,+)}$ is analogous and we skip it for simplicity. By the power mean inequality we have

$$\begin{aligned} &\sum_{l=5}^{\infty} \frac{1}{l!} |h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})|^l \\ &\leq \sum_{l=5}^{\infty} \frac{3^l}{l!} \left\{ \left| \frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k \boldsymbol{\theta}_k \right|^l + \left| \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \boldsymbol{\theta}_k \right|^l + \left| \frac{1}{2\sqrt{nd}} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\theta}_k^2 \right|^l \right\}. \end{aligned}$$

Next, we take the expectation of the last line above with respect to $\mu_n^{(k,-)}$, which gives

$$\sum_{l=5}^{\infty} \frac{1}{l!} \mu_n^{(k,-)} \left[|h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})|^l \right]$$

$$\stackrel{(i)}{\leq} \sum_{l=5}^{\infty} \left\{ \frac{3^l n^{l/2} K^{2l} |\boldsymbol{\Theta}_k|^{l!!} q_{\Theta}^{l/2}}{l! d^{l/2}} + \frac{3^l n^{3l/4} \|\mathbf{Z}_{\cdot k}\|_{\infty}^l K^l l!! q_{\Theta}^{l/2}}{l! d^{l/4}} + \frac{3^l n^{l/2} K^{2l} (2l)!! q_{\Theta}^l}{l! d^{l/2}} \right\} \stackrel{(ii)}{<} \infty,$$

where (i) is by Assumption D.1 and the third result of Lemma E.1. In order to prove (ii), we only need to use the following fact: For n, d large enough we have

$$\frac{6q_{\Theta} K^2 n^{1/2}}{d^{1/2}} < 1.$$

□

According to Lemma E.2, we can take the expectation of Eq. (71) with respect to $\mu_n^{(k,+)}$, which gives

$$\begin{aligned} & \mu_n^{(k,+)} \left[\exp \left(h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right) \right] \\ &= 1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} + \mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4] + \mu_n^{(k,+)}[R^{(k)}] + \mu_n^{(k,+)} \left[\sum_{l=5}^{\infty} \frac{1}{l!} h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})^l \right]. \end{aligned} \quad (72)$$

In the above derivation, we use the fact that under $\mu_n^{(k,+)}$, we have $\boldsymbol{\theta}_k \stackrel{d}{=} \mu_{\Theta}$, which has zero first and third moments. Using Assumption D.1, we conclude that $c_2^{(k)} \geq -\frac{1}{2}\sqrt{\frac{n}{d}}K^2$. Furthermore, notice that

$$\frac{3}{8nd} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^2 - \frac{1}{4\sqrt{nd}} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle \left(\frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k + \frac{1}{\sqrt{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \right)^2 + \frac{1}{24} \left(\frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k + \frac{1}{\sqrt{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \right)^4$$

is non-negative, thus we have $c_4^{(k)} \geq -\frac{1}{4nd} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^2 \geq -\frac{n}{4d} K^4$. Since $\boldsymbol{\theta}_k$ has zero expectation under $\mu_n^{(k,+)}$, we then conclude that $\mu_n^{(k,+)}[h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})] = -\frac{1}{2\sqrt{nd}} \mu_n^{(k,+)}[\langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\theta}_k^2] \geq -\frac{1}{2}\sqrt{\frac{n}{d}}K^2 q_{\Theta}$. By Jensen's inequality

$$\mu_n^{(k,+)} \left[\exp \left(h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right) \right] \geq \exp \left(\mu_n^{(k,+)} \left[h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right] \right) \geq \exp \left(-\frac{1}{2}\sqrt{\frac{n}{d}}K^2 q_{\Theta} \right).$$

Note that the function $x \mapsto \log(x)$ is concave. We next plug the lower bounds derived above into Eq. (72), and obtain

$$\begin{aligned} & \left| \log \left(\mu_n^{(k,+)} \left[\exp \left(h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right) \right] \right) - \log \left(1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} + \mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4] \right) \right| \\ & \leq \left| \sum_{l=5}^{\infty} \mu_n^{(k,+)} \left[\frac{1}{l!} h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})^l \right] + \mu_n^{(k,+)}[R^{(k)}] \right| \times \\ & \quad \max \left\{ \mu_n^{(k,+)} \left[\exp \left(h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right) \right]^{-1}, \left(1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} + \mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4] \right)^{-1} \right\} \\ & \leq \underbrace{\left| \sum_{l=5}^{\infty} \mu_n^{(k,+)} \left[\frac{1}{l!} h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})^l \right] + \mu_n^{(k,+)}[R^{(k)}] \right|}_{I} \times \\ & \quad \underbrace{\max \left\{ \exp \left(\frac{1}{2}\sqrt{\frac{n}{d}}K^2 q_{\Theta} \right), \left(1 - \frac{1}{2}\sqrt{\frac{n}{d}}K^2 q_{\Theta} - \frac{n}{4d}K^4 \mathbb{E}[\boldsymbol{\Theta}_0^4] \right)^{-1} \right\}}_{II}. \end{aligned} \quad (73)$$

Since $d \gg n$ and $(K, q_\Theta, \mathbb{E}[\Theta_0^4])$ are independent of (n, d) , we obtain that term II above converges to 1 as $n, d \rightarrow \infty$. Next, we will provide an upper bound for term I. To this end, we upper bound $\mathbb{E}[\mu_n^{(k,+)}[\sum_{l=5}^\infty \frac{1}{l!} h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})^l]]$ and $\mathbb{E}[\mu_n^{(k,+)}[R^{(k)}]]$ in Section E.1.1 and Section E.1.2, respectively, and combine them to finish the proof in Section E.1.3.

E.1.1 Upper bounding $\mathbb{E}[\mu_n^{(k,+)}[\sum_{l=5}^\infty \frac{1}{l!} h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})^l]]$

Since μ_Θ is sub-Gaussian, there exists a constant $C > 0$ depending only on μ_Θ , such that for all $p \in \mathbb{N}_+$, $\mathbb{E}_{\Theta_0 \sim \mu_\Theta}[|\Theta_0|^p] \leq C^p p^{p/2}$ and $\mathbb{E}_{G \sim \mathcal{N}(0,1)}[|G|^p] \leq C^p p^{p/2}$. Then for n, d large enough, we have

$$\begin{aligned}
& \mathbb{E} \left[\left| \sum_{l=5}^\infty \frac{1}{l!} \mu_n^{(k,+)} \left[h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})^l \right] \right| \right] \\
& \stackrel{(i)}{\leq} \sum_{l=5}^\infty \frac{1}{l!} \mathbb{E} \left[\mu_n^{(k,+)} \left[\left| \frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k \boldsymbol{\theta}_k + \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \boldsymbol{\theta}_k - \frac{1}{2\sqrt{nd}} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\theta}_k^2 \right|^l \right] \right] \\
& \stackrel{(ii)}{\leq} \sum_{l=5}^\infty \frac{1}{l!} \mathbb{E} \left[\mu_n^{(k,+)} \left[\left| \frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k \boldsymbol{\theta}_k + \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \boldsymbol{\theta}_k - \frac{1}{2\sqrt{nd}} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\theta}_k^2 \right|^{2l} \right] \right]^{1/2} \\
& \stackrel{(iii)}{\leq} \sum_{l=5}^\infty \frac{3^l}{l!} \times \left\{ \mathbb{E} \left[\mu_n^{(k,+)} \left[\left| \frac{1}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\Theta}_k \boldsymbol{\theta}_k \right|^{2l} \right] \right]^{1/2} + \mathbb{E} \left[\mu_n^{(k,+)} \left[\left| \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \boldsymbol{\theta}_k \right|^{2l} \right] \right]^{1/2} + \right. \\
& \quad \left. \mathbb{E} \left[\mu_n^{(k,+)} \left[\left| \frac{1}{2\sqrt{nd}} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle \boldsymbol{\theta}_k^2 \right|^{2l} \right] \right]^{1/2} \right\} \\
& \stackrel{(iv)}{\leq} \sum_{l=5}^\infty \frac{3^l}{l!} \times \left\{ \frac{K^{2l} C^{2l} n^{l/2}}{d^{l/2}} \times (2l)^l + \frac{K^l C^{2l} n^{l/4}}{d^{l/4}} \times (2l)^l + \frac{K^{2l} C^{2l} n^{l/2}}{d^{l/2}} \times (2l)^l \right\} \\
& \stackrel{(v)}{\leq} F_{K,C} \times \frac{n^{5/4}}{d^{5/4}}, \tag{74}
\end{aligned}$$

where $F_{K,C} > 0$ is a constant that depends only on K and C . In the above inequalities, (i) is by triangle inequality, (ii) is by Hölder's inequality and (iii) is by power mean inequality. Argument (iv) is via a combination of the following facts: (1) $\text{Support}(\boldsymbol{\Lambda}) \subseteq [-K, K]$, (2) μ_Θ is sub-Gaussian, (3) the random distribution $\mu_n^{(k,+)}$ is independent of $(\mathbf{Z}_{\cdot k}, \boldsymbol{\Theta}_k)$.

For illustration, in the following parts of the proof we upper bound the second summand in the second to last line of Eq. (74). The proofs for the first and third summands follow analogously.

By Lemma E.1, we see that $\mu_n^{(k,+)}$ is independent of $\mathbf{Z}_{\cdot k}$, and $\boldsymbol{\theta}_k$ is independent of $\boldsymbol{\lambda}$ under $\mu_n^{(k,+)}$. Therefore, we have

$$\mathbb{E} \left[\mu_n^{(k,+)} \left[\left| \frac{1}{\sqrt[4]{nd}} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle \boldsymbol{\theta}_k \right|^{2l} \right] \right] = \sum_{s_1=1}^n \cdots \sum_{s_{2l}=1}^n \mathbb{E} [Z_{s_1 k} Z_{s_2 k} \cdots Z_{s_{2l} k}] C_{s_1, s_2, \dots, s_{2l}}, \tag{75}$$

where $C_{s_1, s_2, \dots, s_{2l}} = n^{-l/2} d^{-l/2} \mathbb{E}[\mu_n^{(k,+)}[\lambda_{s_1} \lambda_{s_2} \cdots \lambda_{s_{2l}}]] \mathbb{E}[\Theta_0^{2l}]$. By sub-Gaussian property we have

$$|C_{s_1, s_2, \dots, s_{2l}}| \leq K^{2l} C^{2l} (2l)^l n^{-l/2} d^{-l/2}.$$

Consider all terms that take the form of $\mathbb{E}[Z_{s_1 k} Z_{s_2 k} \cdots Z_{s_{2l} k}]$. If such term is non-zero, then it must be positive. Using property of Gaussian distribution, we have

$$\sum_{s_1=1}^n \cdots \sum_{s_{2l}=1}^n \mathbb{E}[Z_{s_1 k} Z_{s_2 k} \cdots Z_{s_{2l} k}] = \mathbb{E}[(Z_{1k} + \cdots + Z_{nk})^{2l}] = n^l (2l-1)!! \leq n^l C^{2l} (2l)^l.$$

Therefore, the right hand side of Eq. (75) has value no larger than

$$n^l C^{2l} (2l)^l \times K^{2l} C^{2l} (2l)^l n^{-l/2} d^{-l/2} = K^{2l} C^{4l} (2l)^{2l} n^{l/2} d^{-l/2},$$

which leads to the desired upper bound for the second summand. Finally, we use Stirling formula and the assumption that $d \gg n$ to prove argument (v).

E.1.2 Upper bounding $\mathbb{E}[|\mu_n^{(k,+)}[R^{(k)}]|]$

Similar to the proof in Section E.1.1, we conclude that for n, d large enough

$$\mathbb{E}[|\mu_n^{(k,+)}[R^{(k)}]|] \leq F'_{K,C} \times \frac{n^{5/4}}{d^{5/4}}, \quad (76)$$

where $F'_{K,C} > 0$ is a constant depending only on K and C . The derivation of the above upper bound is similar to the derivation of the upper bound for $\mathbb{E}[|\mu_n^{(k,+)}[\sum_{l=5}^{\infty} \frac{1}{l!} h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta})^l]|]$ given in Eq. (74), and we skip the details here for the sake of simplicity.

E.1.3 Combining the upper bounds

Combining Eqs. (73), (74) and (76), we obtain that for n, d large enough

$$\left| \sum_{k=1}^d \frac{1}{n} \mathbb{E} \left[\log \left(\mu_n^{(k,+)} \left[\exp \left(h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right) \right] \right) - \log \left(1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} + \mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4] \right) \right] \right| \leq 2(F_{K,C} + F'_{K,C}) \times \frac{n^{1/4}}{d^{1/4}}. \quad (77)$$

In what follows, we show that the following quantity is small:

$$\left| \sum_{k=1}^d \frac{1}{n} \mathbb{E} \left[\log \left(1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} + \mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4] \right) \right] - \sum_{k=1}^d \frac{1}{n} \mathbb{E} \left[\log \left(1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} \right) \right] \right|. \quad (78)$$

Again we use the concavity of the mapping $x \mapsto \log(x)$, which gives

$$\begin{aligned} & \log \left(1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} \right) + \mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4] - \frac{\left| \mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4] \left(\mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} + \mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4] \right) \right|}{1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} + \mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4]} \\ & \leq \log \left(1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} \right) + \frac{\mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4]}{1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} + \mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4]} \\ & \leq \log \left(1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} + \mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4] \right) \leq \log \left(1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta} \right) + \frac{\mu_n^{(k,+)}[c_4^{(k)}] \mathbb{E}[\boldsymbol{\Theta}_0^4]}{1 + \mu_n^{(k,+)}[c_2^{(k)}] q_{\Theta}} \end{aligned}$$

$$\leq \log \left(1 + \mu_n^{(k,+)} [c_2^{(k)}] q_\Theta \right) + \mu_n^{(k,+)} [c_4^{(k)}] \mathbb{E}[\Theta_0^4] + \frac{|\mu_n^{(k,+)} [c_2^{(k)}] \mu_n^{(k,+)} [c_4^{(k)}] q_\Theta \mathbb{E}[\Theta_0^4]|}{1 + \mu_n^{(k,+)} [c_2^{(k)}] q_\Theta}. \quad (79)$$

By Lemma E.1, $\mu_n^{(k,+)}$ and $\mathbf{Z}_{\cdot k}$ are independent of each other, thus

$$\mathbb{E} \left[\mu_n^{(k,+)} \left[\frac{1}{8nd} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^2 - \frac{1}{4nd} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle^2 + \frac{1}{24nd} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle^4 \right] \right] = 0. \quad (80)$$

Next, we plug Eq. (80) into the definition of $c_4^{(k)}$, then apply Lemma E.1 claim 3, which gives

$$\begin{aligned} & \left| \mathbb{E}[\mu_n^{(k,+)} [c_4^{(k)}]] \right| \\ &= \left| -\frac{\mathbb{E} \left[\mu_n^{(k,+)} [\langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle \langle \boldsymbol{\lambda}, \boldsymbol{\Lambda} \rangle^2 \Theta_k^2] \right]}{4n^{3/2} d^{3/2}} + \frac{\mathbb{E} \left[\mu_n^{(k,+)} [\langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^4 \Theta_k^4] \right]}{24n^2 d^2} + \frac{\mathbb{E} \left[\mu_n^{(k,+)} [\langle \boldsymbol{\lambda}, \boldsymbol{\Lambda} \rangle^2 \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle^2 \Theta_k^2] \right]}{4n^{3/2} d^{3/2}} \right| \\ &\leq \frac{n^{3/2}}{2d^{3/2}} K^6 q_\Theta + \frac{n^2}{24d^2} K^8 \mathbb{E}[\Theta_0^4]. \end{aligned} \quad (81)$$

In addition, we have the following lemma:

Lemma E.3. *There exist constants $A_1(K, \mu_\Theta), A_2(K, \mu_\Theta) > 0$, which are functions of (K, μ_Θ) only, such that*

$$\mathbb{E}[\mu_n^{(k,+)} [|c_2^{(k)}|^2]] \leq A_1(K, \mu_\Theta) \times \frac{n}{d}, \quad \mathbb{E}[\mu_n^{(k,+)} [|c_4^{(k)}|^2]] \leq A_2(K, \mu_\Theta) \times \frac{n^2}{d^2}.$$

Proof. Straightforward computation reveals that there exist $A'_1(K, \mu_\Theta), A'_2(K, \mu_\Theta) > 0$ depending only on (K, μ_Θ) , such that

$$\begin{aligned} \mathbb{E}[\mu_n^{(k,+)} [|c_2^{(k)}|^2]] &\leq A'_1(K, \mu_\Theta) \mathbb{E} \left[\mu_n^{(k,+)} \left[\frac{1}{n^2 d^2} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^4 \Theta_k^4 + \frac{1}{nd} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle^4 + \frac{1}{nd} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^2 \right] \right] \\ \mathbb{E}[\mu_n^{(k,+)} [|c_4^{(k)}|^2]] &\leq A'_2(K, \mu_\Theta) \mathbb{E} \left[\mu_n^{(k,+)} \left[\frac{1}{n^2 d^2} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^4 + \frac{1}{n^3 d^3} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^2 \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^4 \Theta_k^4 + \frac{1}{n^2 d^2} \langle \boldsymbol{\lambda}, \boldsymbol{\lambda} \rangle^2 \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle^4 + \right. \right. \\ &\quad \left. \left. \frac{1}{n^4 d^4} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^8 \Theta_k^8 + \frac{1}{n^2 d^2} \langle \mathbf{Z}_{\cdot k}, \boldsymbol{\lambda} \rangle^8 \right] \right]. \end{aligned}$$

The rest of the proof follows from Lemma E.1 and the assumption that $d \gg n$. □

Recall that $c_2^{(k)} \geq -\frac{1}{2} \sqrt{\frac{n}{d}} K^2$ and $c_4^{(k)} \geq -\frac{n}{4d} K^4$. Then for n, d large enough, using Lemma E.3, we obtain

$$\begin{aligned} \mathbb{E} \left[\frac{|\mu_n^{(k,+)} [c_2^{(k)}] \mu_n^{(k,+)} [c_4^{(k)}] q_\Theta \mathbb{E}[\Theta_0^4]|}{1 + \mu_n^{(k,+)} [c_2^{(k)}] q_\Theta} \right] &\leq \frac{q_\Theta \mathbb{E}[\Theta_0^4]}{1 - \frac{1}{2} \sqrt{\frac{n}{d}} K^2 q_\Theta} \times \mathbb{E}[\mu_n^{(k,+)} [|c_2^{(k)}|^2]]^{1/2} \mathbb{E}[\mu_n^{(k,+)} [|c_4^{(k)}|^2]]^{1/2}, \\ &\leq \frac{2q_\Theta \mathbb{E}[\Theta_0^4] A_1(K, \mu_\Theta)^{1/2} A_2(K, \mu_\Theta)^{1/2} n^{3/2}}{d^{3/2}}, \end{aligned} \quad (82)$$

$$\mathbb{E} \left[\frac{|\mu_n^{(k,+)} [c_4^{(k)}] \mathbb{E}[\Theta_0^4] \left(\mu_n^{(k,+)} [c_2^{(k)}] q_\Theta + \mu_n^{(k,+)} [c_4^{(k)}] \mathbb{E}[\Theta_0^4] \right)|}{1 + \mu_n^{(k,+)} [c_2^{(k)}] q_\Theta + \mu_n^{(k,+)} [c_4^{(k)}] \mathbb{E}[\Theta_0^4]} \right]$$

$$\begin{aligned}
&\leq \frac{\mathbb{E}[\Theta_0^4]^2 \mathbb{E}[\mu_n^{(k,+)}[|c_4^{(k)}|^2]]}{1 - \frac{1}{2}\sqrt{\frac{n}{d}}K^2q_\Theta - \frac{n}{4d}K^4\mathbb{E}[\Theta_0^4]} + \frac{\mathbb{E}[\Theta_0^4]q_\Theta \mathbb{E}[\mu_n^{(k,+)}[|c_2^{(k)}|^2]]^{1/2} \mathbb{E}[\mu_n^{(k,+)}[|c_4^{(k)}|^2]]^{1/2}}{1 - \frac{1}{2}\sqrt{\frac{n}{d}}K^2q_\Theta - \frac{n}{4d}K^4\mathbb{E}[\Theta_0^4]} \\
&\leq \frac{2\mathbb{E}[\Theta_0^4]^2 A_2(K, \mu_\Theta)n^2}{d^2} + \frac{2\mathbb{E}[\Theta_0^4]q_\Theta A_1(K, \mu_\Theta)^{1/2} A_2(K, \mu_\Theta)^{1/2} n^{3/2}}{d^{3/2}}.
\end{aligned} \tag{83}$$

Next, we plug Eqs. (81) to (83) into Eq. (79), then sum over $k \in [d]$. This implies the existence of $C(K, \mu_\Theta) > 0$, which is a constant depending only on (K, μ_Θ) , such that for n, d large enough

$$\begin{aligned}
&\sum_{k=1}^d \frac{1}{n} \mathbb{E} \left[\log \left(1 + \mu_n^{(k,+)}[c_2^{(k)}]q_\Theta \right) \right] - \frac{C(K, \mu_\Theta)n^{1/2}}{d^{1/2}} \\
&\leq \sum_{k=1}^d \frac{1}{n} \mathbb{E} \left[\log \left(1 + \mu_n^{(k,+)}[c_2^{(k)}]q_\Theta + \mu_n^{(k,+)}[c_4^{(k)}]\mathbb{E}[\Theta_0^4] \right) \right] \\
&\leq \sum_{k=1}^d \frac{1}{n} \mathbb{E} \left[\log \left(1 + \mu_n^{(k,+)}[c_2^{(k)}]q_\Theta \right) \right] + \frac{C(K, \mu_\Theta)n^{1/2}}{d^{1/2}}.
\end{aligned} \tag{84}$$

Combining Eqs. (77) and (84), we derive that

$$\sum_{k=1}^d \frac{1}{n} \mathbb{E} \left[\log \left(1 + \mu_n^{(k,+)}[c_2^{(k)}]q_\Theta \right) \right] - \sum_{k=1}^d \frac{1}{n} \mathbb{E} \left[\log \left(\mu_n^{(k,+)} \left[\exp \left(h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right) \right] \right) \right] = o_n(1). \tag{85}$$

Similarly, we can prove that

$$\sum_{k=1}^d \frac{1}{n} \mathbb{E} \left[\log \left(1 + \mu_n^{(k,-)}[c_2^{(k)}]q_\Theta \right) \right] - \sum_{k=1}^d \frac{1}{n} \mathbb{E} \left[\log \left(\mu_n^{(k,-)} \left[\exp \left(h_n^{(k)}(\boldsymbol{\lambda}, \boldsymbol{\theta}) \right) \right] \right) \right] = o_n(1). \tag{86}$$

Since $c_2^{(k)}$ is independent of $\boldsymbol{\theta}_k$, by Lemma E.1 we have $\mu_n^{(k,+)}[c_2^{(k)}] = \mu_n^{(k,-)}[c_2^{(k)}]$. Finally, we combine Eqs. (70), (85) and (86), which gives $\Phi_n^{(d)} - \Phi_n^{(0)} = o_n(1)$. Thus, we have completed the proof of Lemma D.2.

E.2 Proof of Lemma D.3

In this section we prove Lemma D.3. Applying Gaussian integration by parts, we obtain that

$$\Phi_n^{(0)}(h, s) = \frac{1}{n} \mathbb{E} \left[\log \left(\int \exp \left(\tilde{H}'_n(\boldsymbol{\lambda}; \mathbf{Y}'(h), \mathbf{x}'(s)) + H_n(\boldsymbol{\lambda}; \mathbf{Y}'(h)) + H_n(\boldsymbol{\lambda}; \mathbf{x}'(s)) \right) \mu_\Lambda^{\otimes n}(\mathrm{d}\boldsymbol{\lambda}) \right) \right],$$

where

$$\begin{aligned}
\tilde{H}'_n(\boldsymbol{\lambda}; \mathbf{Y}'(h), \mathbf{x}'(s)) &= \frac{\langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^2 \|\boldsymbol{\Theta}\|^2 / (nd) + 2\langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \langle \boldsymbol{\lambda}, \mathbf{Z}\boldsymbol{\Theta} \rangle / (nd)^{3/4} + \|\mathbf{Z}^\top \boldsymbol{\lambda}\|^2 / \sqrt{nd}}{2q_\Theta^{-1} + 2\|\boldsymbol{\lambda}\|^2 / \sqrt{nd}} \\
&\quad - \frac{d}{2} \log \left(1 + \frac{q_\Theta}{\sqrt{nd}} \|\boldsymbol{\lambda}\|^2 \right).
\end{aligned}$$

By triangle inequality,

$$\sup_{\|\boldsymbol{\lambda}\|_\infty \leq K} \left| \frac{\langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^2 \|\boldsymbol{\Theta}\|^2 / (nd)}{2q_\Theta^{-1} + 2\|\boldsymbol{\lambda}\|^2 / \sqrt{nd}} - \frac{q_\Theta^2}{2n} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^2 \right|$$

$$\begin{aligned}
&\leq \sup_{\|\boldsymbol{\lambda}\|_\infty \leq K} \left| \frac{\langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^2 (\|\boldsymbol{\Theta}\|^2 - dq_\Theta) / (nd)}{2q_\Theta^{-1} + 2\|\boldsymbol{\lambda}\|^2 / \sqrt{nd}} \right| + \sup_{\|\boldsymbol{\lambda}\|_\infty \leq K} \left| \frac{q_\Theta^2 \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^2 \|\boldsymbol{\lambda}\|^2 / (n^{3/2} d^{1/2})}{2q_\Theta^{-1} + 2\|\boldsymbol{\lambda}\|^2 / \sqrt{nd}} \right| \\
&\leq \frac{nK^4 q_\Theta}{2d} \|\boldsymbol{\Theta}\|^2 - dq_\Theta + \frac{q_\Theta^3 n^{3/2} K^6}{2d^{1/2}}.
\end{aligned} \tag{87}$$

Furthermore, notice that the following inequalities hold:

$$\sup_{\|\boldsymbol{\lambda}\|_\infty \leq K} \left| \frac{\langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \langle \boldsymbol{\lambda}, \mathbf{Z}\boldsymbol{\Theta} \rangle / (nd)^{3/4}}{q_\Theta^{-1} + \|\boldsymbol{\lambda}\|^2 / \sqrt{nd}} \right| \leq \frac{n^{3/4} q_\Theta K^3}{d^{3/4}} \|\mathbf{Z}\boldsymbol{\Theta}\|, \tag{88}$$

$$\sup_{\|\boldsymbol{\lambda}\|_\infty \leq K} \left| \frac{\|\mathbf{Z}^\top \boldsymbol{\lambda}\|^2 / \sqrt{nd}}{2q_\Theta^{-1} + 2\|\boldsymbol{\lambda}\|^2 / \sqrt{nd}} - \frac{\|\mathbf{Z}^\top \boldsymbol{\lambda}\|^2 / \sqrt{nd}}{2q_\Theta^{-1}} + \frac{\|\mathbf{Z}^\top \boldsymbol{\lambda}\|^2 \|\boldsymbol{\lambda}\|^2 / (nd)}{2q_\Theta^{-2}} \right| \leq \frac{n^2 K^6 q_\Theta^3}{2d\sqrt{nd}} \|\mathbf{Z}\mathbf{Z}^\top\|_{\text{op}}, \tag{89}$$

$$\sup_{\|\boldsymbol{\lambda}\|_\infty \leq K} \left| \frac{d}{2} \log \left(1 + \frac{q_\Theta}{\sqrt{nd}} \|\boldsymbol{\lambda}\|^2 \right) - \frac{dq_\Theta}{2\sqrt{nd}} \|\boldsymbol{\lambda}\|^2 + \frac{q_\Theta^2}{4n} \|\boldsymbol{\lambda}\|^4 \right| \leq \frac{q_\Theta^3 n^{3/2} K^6}{6d^{1/2}}, \tag{90}$$

$$\sup_{\|\boldsymbol{\lambda}\|_\infty \leq K} \left| \frac{q_\Theta^2 \|\mathbf{Z}^\top \boldsymbol{\lambda}\|^2 \|\boldsymbol{\lambda}\|^2}{2nd} - \frac{q_\Theta^2}{2n} \|\boldsymbol{\lambda}\|^4 \right| \leq \frac{nq_\Theta^2 K^4}{2d} \|\mathbf{Z}\mathbf{Z}^\top - d\mathbf{I}_n\|_{\text{op}}, \tag{91}$$

where in Eq. (90), we use the fact that for all $x \in [0, \infty)$, there exists $y \in [0, x]$ such that

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3(1+y)^3}.$$

Next, we combine Eqs. (87) to (91), and conclude that

$$\begin{aligned}
&\left| \Phi_n^{(0)}(h, s) - \tilde{\Phi}_n(h, s) \right| \\
&\leq \frac{1}{n} \mathbb{E} \left[\frac{nK^4 q_\Theta \|\boldsymbol{\Theta}\|^2 - dq_\Theta}{2d} + \frac{q_\Theta^3 n^{3/2} K^6}{2d^{1/2}} + \frac{n^{3/4} q_\Theta K^3 \|\mathbf{Z}\boldsymbol{\Theta}\|}{d^{3/4}} + \frac{n^{3/2} K^6 q_\Theta^3 \|\mathbf{Z}\mathbf{Z}^\top\|_{\text{op}}}{2d^{3/2}} \right. \\
&\quad \left. + \frac{q_\Theta^3 n^{3/2} K^6}{6d^{1/2}} + \frac{nq_\Theta^2 K^4 \|\mathbf{Z}\mathbf{Z}^\top - d\mathbf{I}_n\|_{\text{op}}}{2d} \right].
\end{aligned} \tag{92}$$

Using Lemma B.1, we see that

$$\begin{aligned}
\mathbb{E} \left[\left\| \frac{1}{d} \mathbf{Z}\mathbf{Z}^\top - \mathbf{I}_n \right\|_{\text{op}} \right] &\leq 100 \sqrt{\frac{n}{d}} + \int_{100\sqrt{n/d}}^\infty \exp \left(-\frac{dx^2}{3200} \right) dx \\
&\leq 100 \sqrt{\frac{n}{d}} + \frac{40}{\sqrt{d}} \int_0^\infty \exp \left(-\frac{y^2}{2} \right) dy.
\end{aligned} \tag{93}$$

Finally, we combine Eq. (92), (93), the assumption that $d \gg n$, and conclude that as $n, d \rightarrow \infty$, $|\Phi_n^{(0)}(h, s) - \tilde{\Phi}_n(h, s)| = o_n(1)$, thus completing the proof of Lemma D.3.

E.3 Proof of Lemma D.4

Recall that $\mathbf{W} \stackrel{d}{=} \text{GOE}(n)$. Then for all fixed orthogonal matrix $\mathbf{O} \in \mathbb{R}^{n \times n}$, $\mathbf{O}^\top \mathbf{W} \mathbf{O} \stackrel{d}{=} \mathbf{W}$ and $\mathbf{O}^\top (\mathbf{Z}\mathbf{Z}^\top - d\mathbf{I}_n) \mathbf{O} \stackrel{d}{=} (\mathbf{Z}\mathbf{Z}^\top - d\mathbf{I}_n)$. By orthogonal invariance, we can couple $(\mathbf{Z}\mathbf{Z}^\top - d\mathbf{I}_n) / \sqrt{nd}$ with \mathbf{W} such that they admit the following eigen-decomposition:

$$\frac{1}{\sqrt{nd}} (\mathbf{Z}\mathbf{Z}^\top - d\mathbf{I}_n) = \boldsymbol{\Omega}^\top \mathbf{S}_1 \boldsymbol{\Omega}, \quad \mathbf{W} = \boldsymbol{\Omega}^\top \mathbf{S}_2 \boldsymbol{\Omega}. \tag{94}$$

In the above display, $\mathbf{\Omega}$ is Haar-distributed on the orthogonal matrix group $\mathcal{O}(n)$, \mathbf{S}_1 and \mathbf{S}_2 are diagonal matrices containing ascendingly ordered eigenvalues of matrices $(\mathbf{Z}\mathbf{Z}^\top - d\mathbf{I}_n)/\sqrt{nd}$ and \mathbf{W} , respectively. Furthermore, $\mathbf{S}_1, \mathbf{S}_2$ are both independent of $\mathbf{\Omega}$. Direct computation implies the following inequality:

$$\begin{aligned} \left| \tilde{\Phi}_n(h, s) - \Phi_n^Y(h, s) \right| &\leq \frac{1}{n} \mathbb{E} \left[\sup_{\|\boldsymbol{\lambda}\|_\infty \leq K} \left| \frac{q_\Theta}{2} \boldsymbol{\lambda}^\top \left(\mathbf{W} - \frac{1}{\sqrt{nd}} (\mathbf{Z}\mathbf{Z}^\top - d\mathbf{I}_n) \right) \boldsymbol{\lambda} \right| \right] \\ &\leq \frac{q_\Theta K^2}{2} \mathbb{E} [\|\mathbf{S}_1 - \mathbf{S}_2\|_{\text{op}}]. \end{aligned}$$

Let $\sigma_i(\mathbf{S}_j)$ be the i -th largest eigenvalue of \mathbf{S}_j for $j \in [2]$, then $\|\mathbf{S}_1 - \mathbf{S}_2\|_{\text{op}} = \max_{i \in [n]} |\sigma_i(\mathbf{S}_1) - \sigma_i(\mathbf{S}_2)|$. We denote by $\text{ESD}(\mathbf{M})$ the empirical spectral distribution of matrix \mathbf{M} . Then using random matrix theory, $\text{ESD}(\mathbf{S}_1)$ and $\text{ESD}(\mathbf{S}_2)$ both converge almost surely to the semicircle law (see [BY88b]). Furthermore, according to the results in [BY88a, Pau12, Kar03], asymptotically speaking, we have $\sigma_1(\mathbf{S}_1), \sigma_1(\mathbf{S}_2) \xrightarrow{a.s.} 2$ and $\sigma_n(\mathbf{S}_1), \sigma_n(\mathbf{S}_2) \xrightarrow{a.s.} -2$. Therefore, we see that $\|\mathbf{S}_1 - \mathbf{S}_2\|_{\text{op}} \xrightarrow{a.s.} 0$ as $n, d \rightarrow \infty$.

By Theorem 1.1 in [BVH⁺16], for all $0 < \varepsilon \leq 1/2$

$$\mathbb{E}[\|\mathbf{W}\|_{\text{op}}] \leq (1 + \varepsilon) \left\{ 2\sqrt{1 + \frac{1}{n}} + \frac{6}{\sqrt{\log(1 + \varepsilon)}} \sqrt{\frac{2 \log n}{n}} \right\}.$$

In the above equation, we first let $n \rightarrow \infty$ then let $\varepsilon \rightarrow 0^+$, which gives $\limsup_{n \rightarrow \infty} \mathbb{E}[\|\mathbf{S}_2\|_{\text{op}}] \leq 2$. By Fatou's lemma, we further have $\liminf_{n \rightarrow \infty} \mathbb{E}[\|\mathbf{S}_2\|_{\text{op}}] \geq 2$, thus $\lim_{n \rightarrow \infty} \mathbb{E}[\|\mathbf{S}_2\|_{\text{op}}] = 2$. By Lemma B.1, for any $\varepsilon > 0$, there exists $M > 0$, such that for n, d large enough

$$\mathbb{E} \left[\|\mathbf{S}_1\|_{\text{op}} \mathbb{1}_{\{\|\mathbf{S}_1\|_{\text{op}} \geq M\}} \right] < \varepsilon.$$

Dominated convergence theorem gives $\limsup_{n \rightarrow \infty} \mathbb{E}[\|\mathbf{S}_1\|_{\text{op}} \mathbb{1}_{\{\|\mathbf{S}_1\|_{\text{op}} < M\}}] \leq 2$, thus we have $\limsup_{n \rightarrow \infty} \mathbb{E}[\|\mathbf{S}_1\|_{\text{op}}] \leq 2 + \varepsilon$. On the other hand, Fatou's lemma implies $\liminf_{n \rightarrow \infty} \mathbb{E}[\|\mathbf{S}_1\|_{\text{op}}] \geq 2$, thus $\lim_{n \rightarrow \infty} \mathbb{E}[\|\mathbf{S}_1\|_{\text{op}}] = 2$. Finally, notice that $\|\mathbf{S}_1\|_{\text{op}} + \|\mathbf{S}_2\|_{\text{op}} - \|\mathbf{S}_1 - \mathbf{S}_2\|_{\text{op}} \geq 0$. We then apply Scheffé's lemma to both $\|\mathbf{S}_1 - \mathbf{S}_2\|_{\text{op}}$ and $\|\mathbf{S}_1\|_{\text{op}} + \|\mathbf{S}_2\|_{\text{op}} - \|\mathbf{S}_1 - \mathbf{S}_2\|_{\text{op}}$, which gives $\mathbb{E}[\|\mathbf{S}_1 - \mathbf{S}_2\|_{\text{op}}] \rightarrow 0$. This concludes the proof of Lemma D.4.

E.4 Proof of Lemma D.5

The first claim is a direct consequence of Lemmas D.2 to D.4. As for the second claim, it is straightforward that the free energy densities $\Phi_n(h, s)$ and $\Phi_n^Y(h, s)$ are well-defined on $[0, \infty) \times [0, \infty)$ and differentiable for all $h, s \in (0, \infty)$.

By Nishimori identity (Lemma B.3) and Gaussian integration by parts, we see that for $h, s > 0$,

$$\begin{aligned} \frac{\partial}{\partial h} \Phi_n(h, s) &= \frac{1}{4n^2} \mathbb{E} \left[\langle \mathbf{\Lambda} \mathbf{\Lambda}^\top, \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \mathbf{A}, \mathbf{Y}'(h), \mathbf{x}'(s)] \rangle \right], \\ \frac{\partial}{\partial s} \Phi_n(h, s) &= \frac{1}{2n} \mathbb{E} \left[\langle \mathbf{\Lambda}, \mathbb{E}[\mathbf{\Lambda} \mid \mathbf{A}, \mathbf{Y}'(h), \mathbf{x}'(s)] \rangle \right]. \end{aligned}$$

Recall that h, s stand for the signal-to-noise ratios in the perturbed model. Therefore, we obtain that for fixed $s \geq 0$, $\frac{\partial}{\partial h} \Phi_n(h, s)$ is increasing in h and for fixed $h \geq 0$, $\frac{\partial}{\partial s} \Phi_n(h, s)$ is increasing in s .

As a result, for all fixed $h, s \geq 0$, the mappings $x \mapsto \Phi_n(h, x)$, $x \mapsto \Phi_n(x, s)$ are convex on $(0, \infty)$. Since these mappings are obviously continuous, we obtain that they are convex functions

on $[0, \infty)$. Similarly, we can show that for all fixed $h, s \geq 0$, $x \mapsto \Phi_n^Y(h, x)$, $x \mapsto \Phi_n^Y(x, s)$ are convex functions on $[0, \infty)$. This concludes the proof of the second claim.

Finally, we prove the third claim. This proof is based on Guerra's interpolation technique. For $t \in [0, 1]$, $x, q \in \mathbb{R}_+$, we define

$$\begin{aligned} H_n^G(\boldsymbol{\lambda}; t, x, q) &:= \frac{xt}{2n} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^2 + \frac{\sqrt{xt}}{2} \boldsymbol{\lambda}^\top \mathbf{W}' \boldsymbol{\lambda} - \frac{xt}{4n} \|\boldsymbol{\lambda}\|^4 + \sqrt{(1-t)xq} \langle \mathbf{g}', \boldsymbol{\lambda} \rangle \\ &\quad + (1-t)xq \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle - \frac{(1-t)xq}{2} \|\boldsymbol{\lambda}\|^2, \\ \Psi_n^G(t, x, q) &:= \frac{1}{n} \mathbb{E} \left[\log \int \exp(H_n^G(\boldsymbol{\lambda}; t, x, q)) \mu_{\boldsymbol{\Lambda}}^{\otimes n}(\mathrm{d}\boldsymbol{\lambda}) \right]. \end{aligned}$$

By [LM19, Theorem 13], we see that $\lim_{n \rightarrow \infty} \Psi_n^G(1, x, q) = \sup_{y \geq 0} \mathcal{F}(x, y)$. Using Lemma B.3 and Gaussian integration by parts, we see that

$$\begin{aligned} \frac{\partial}{\partial t} \Psi_n^G(t, x, q) &= \frac{1}{n} \mathbb{E} \left[\frac{x}{4n} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^2 - \frac{xq}{2} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \mid \frac{\sqrt{xt}}{n} \boldsymbol{\Lambda} \boldsymbol{\Lambda}^\top + \mathbf{W}', \sqrt{(1-t)xq} \boldsymbol{\Lambda} + \mathbf{g}' \right] \\ &= \frac{x}{4} \mathbb{E} \left[\left(\frac{1}{n} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle - q \right)^2 \mid \frac{\sqrt{xt}}{n} \boldsymbol{\Lambda} \boldsymbol{\Lambda}^\top + \mathbf{W}', \sqrt{(1-t)xq} \boldsymbol{\Lambda} + \mathbf{g}' \right] - \frac{xq^2}{4} \\ &\geq -\frac{xq^2}{4}. \end{aligned}$$

Direct computation reveals that

$$\Psi_n^G(0, x, q) = \mathcal{F}(x, q) + \frac{xq^2}{4}.$$

Therefore, for all $x, q \geq 0$,

$$\Psi_n^G(1, x, q) = \Psi_n^G(0, x, q) + \int_0^1 \frac{\partial}{\partial t} \Psi_n^G(t, x, q) \mathrm{d}t \geq \mathcal{F}(x, q). \quad (95)$$

According to [LM19, Proposition 17], for all but countably many $x > 0$, $\mathcal{F}(x, \cdot)$ has a unique maximizer $q^*(x)$. For these x , we plug $q = q^*(x)$ into Eq. (95), which implies for all but countably many $x > 0$ and all $t \in [0, 1]$, $\lim_{n \rightarrow \infty} \Psi_n^G(t, x, q^*(x)) = \mathcal{F}(x, q^*(x)) + xq^*(x)^2(1-t)/4$. Notice that $\Psi_n^G(t, x, q^*(x)) = \Phi_n^Y(0, s)$ if $xt = q_\Theta^2$ and $(1-t)xq^*(x) = s$. This concludes the proof of the third claim of the lemma.

E.5 Proof of Lemma D.6

For $t \in [0, 1]$, we define the interpolated Hamiltonian as

$$\begin{aligned} H_{n,t}^\varepsilon(\boldsymbol{\lambda}, \boldsymbol{\theta}; h) &:= \frac{t}{\sqrt{nd}} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle \langle \boldsymbol{\Theta}, \boldsymbol{\theta} \rangle + \frac{\sqrt{t}}{\sqrt[4]{nd}} \boldsymbol{\lambda}^\top \mathbf{Z} \boldsymbol{\theta} - \frac{t}{2\sqrt{nd}} \|\boldsymbol{\lambda}\|^2 \|\boldsymbol{\theta}\|^2 + \\ &\quad \frac{1-t}{\sqrt{nd}} \langle \bar{\boldsymbol{\Lambda}}, \bar{\boldsymbol{\lambda}} \rangle \langle \boldsymbol{\Theta}, \boldsymbol{\theta} \rangle + \frac{\sqrt{1-t}}{\sqrt[4]{nd}} \bar{\boldsymbol{\lambda}}^\top \mathbf{Z}' \boldsymbol{\theta} - \frac{1-t}{2\sqrt{nd}} \|\bar{\boldsymbol{\lambda}}\|^2 \|\boldsymbol{\theta}\|^2 + \\ &\quad \frac{ht}{2n} \langle \boldsymbol{\Lambda}, \boldsymbol{\lambda} \rangle^2 + \frac{\sqrt{ht}}{2} \boldsymbol{\lambda}^\top \mathbf{W} \boldsymbol{\lambda} - \frac{ht}{4n} \|\boldsymbol{\lambda}\|^4 + \\ &\quad \frac{h(1-t)}{2n} \langle \bar{\boldsymbol{\Lambda}}, \bar{\boldsymbol{\lambda}} \rangle^2 + \frac{\sqrt{h(1-t)}}{2} \bar{\boldsymbol{\lambda}}^\top \mathbf{W}' \bar{\boldsymbol{\lambda}} - \frac{h(1-t)}{4n} \|\bar{\boldsymbol{\lambda}}\|^4. \end{aligned}$$

where $\mathbf{Z}' = (Z'_{ij})_{i \in [n], j \in [d]}$ is an independent copy of \mathbf{Z} and is independent of everything else. We emphasize that $\mathbf{Z}, \mathbf{Z}', \mathbf{W}, \mathbf{W}', \mathbf{\Lambda}, \mathbf{\Theta}$ are mutually independent. Notice that $H_{n,t}^\varepsilon(\boldsymbol{\lambda}, \boldsymbol{\theta}; h)$ is the Hamiltonian that corresponds to observations $(\mathbf{A}_1, \mathbf{A}_2, \mathbf{Y}_1, \mathbf{Y}_2)$

$$\begin{aligned} \mathbf{A}_1 &= \frac{\sqrt{t}}{\sqrt[4]{nd}} \mathbf{\Lambda} \mathbf{\Theta}^\top + \mathbf{Z}, & \mathbf{A}_2 &= \frac{\sqrt{1-t}}{\sqrt[4]{nd}} \bar{\mathbf{\Lambda}} \mathbf{\Theta}^\top + \mathbf{Z}', \\ \mathbf{Y}_1 &= \frac{\sqrt{ht}}{n} \mathbf{\Lambda} \mathbf{\Lambda}^\top + \mathbf{W}, & \mathbf{Y}_2 &= \frac{\sqrt{h(1-t)}}{n} \bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^\top + \mathbf{W}'. \end{aligned}$$

We define the corresponding free energy density

$$\Phi_{n,t}^\varepsilon(h) := \frac{1}{n} \mathbb{E} \left[\log \left(\int \exp(H_{n,t}^\varepsilon(\boldsymbol{\lambda}, \boldsymbol{\theta}; h)) \mu_{\mathbf{\Lambda}}^{\otimes n}(\mathrm{d}\boldsymbol{\lambda}) \mu_{\mathbf{\Theta}}^{\otimes d}(\mathrm{d}\boldsymbol{\theta}) \right) \right].$$

At the endpoints, we have $\Phi_{n,0}^\varepsilon(h) = \bar{\Phi}_n^\varepsilon(h)$ and $\Phi_{n,1}^\varepsilon(h) = \Phi_n(h, 0)$. For simplicity, we denote by $\langle \cdot \rangle_{h,\varepsilon,t}$ the expectation with respect to the posterior distribution $\mathbb{P}(\mathbf{\Lambda} = \cdot, \mathbf{\Theta} = \cdot \mid \mathbf{A}_1, \mathbf{A}_2, \mathbf{Y}_1, \mathbf{Y}_2)$. Using Gaussian integration by parts and Nishimori identity (Lemma B.3), we have

$$\begin{aligned} \frac{\partial}{\partial t} \Phi_{n,t}^\varepsilon(h) &= \frac{1}{2n\sqrt{nd}} \sum_{i \in [n], j \in [d]} \mathbb{E} [\langle (\mathbf{\Lambda}_i \boldsymbol{\lambda}_i - \bar{\mathbf{\Lambda}}_i \bar{\boldsymbol{\lambda}}_i) \mathbf{\Theta}_j \boldsymbol{\theta}_j \rangle_{h,\varepsilon,t}] + \frac{h}{4n^2} \mathbb{E} [\langle (\mathbf{\Lambda}^\top \boldsymbol{\lambda})^2 \rangle_{h,\varepsilon,t}] - \frac{h}{4n^2} \mathbb{E} [\langle (\bar{\mathbf{\Lambda}}^\top \bar{\boldsymbol{\lambda}})^2 \rangle_{h,\varepsilon,t}] \\ &= \frac{1}{2\sqrt{nd}} \mathbb{E} [\langle (\mathbf{\Lambda}_1 \boldsymbol{\lambda}_1 - \bar{\mathbf{\Lambda}}_1 \bar{\boldsymbol{\lambda}}_1) \langle \mathbf{\Theta}, \boldsymbol{\theta} \rangle \rangle_{h,\varepsilon,t}] + \frac{h}{4n^2} \sum_{i \in [n], j \in [n]} \mathbb{E} [\langle \mathbf{\Lambda}_i \boldsymbol{\lambda}_j - \bar{\mathbf{\Lambda}}_i \bar{\boldsymbol{\lambda}}_j \rangle_{h,\varepsilon,t}]. \end{aligned}$$

Next, we provide upper bound for the above partial derivative. Invoking Holder's inequality, we see that

$$\begin{aligned} & \left| \frac{\partial}{\partial t} \Phi_{n,t}^\varepsilon(h) \right| \\ & \leq \frac{1}{2\sqrt{nd}} \mathbb{E} [\langle (\mathbf{\Lambda}_1 \boldsymbol{\lambda}_1 - \bar{\mathbf{\Lambda}}_1 \bar{\boldsymbol{\lambda}}_1)^2 \rangle_{h,\varepsilon,t}^{1/2} \langle \langle \mathbf{\Theta}, \boldsymbol{\theta} \rangle^2 \rangle_{h,\varepsilon,t}^{1/2}] + \frac{h}{2n^2} \sum_{i \in [n], j \in [n]} \mathbb{E} [(\mathbf{\Lambda}_i \boldsymbol{\lambda}_j - \bar{\mathbf{\Lambda}}_i \bar{\boldsymbol{\lambda}}_j)^2]^{1/2} \mathbb{E} [\mathbf{\Lambda}_i^2 \mathbf{\Lambda}_j^2]^{1/2} \\ & \leq \frac{1}{2\sqrt{nd}} \mathbb{E} [\langle (\mathbf{\Lambda}_1 \boldsymbol{\lambda}_1 - \bar{\mathbf{\Lambda}}_1 \bar{\boldsymbol{\lambda}}_1)^2 \rangle_{h,\varepsilon,t}^{1/2}] \mathbb{E} [\langle \langle \mathbf{\Theta}, \boldsymbol{\theta} \rangle^2 \rangle_{h,\varepsilon,t}^{1/2}] + h \mathbb{E} [\mathbf{\Lambda}_1^4]^{3/4} \mathbb{E} [(\mathbf{\Lambda}_1 - \bar{\mathbf{\Lambda}}_1)^4]^{1/4}. \end{aligned}$$

We denote by $\langle \cdot \rangle_{h,\varepsilon,t,*}$ the expectation with respect to the posterior distribution

$$\mathbb{P}(\mathbf{\Theta} = \cdot \mid \mathbf{A}_1, \mathbf{A}_2, \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{\Lambda}, \bar{\mathbf{\Lambda}}).$$

Direct computation gives the following inequality:

$$\begin{aligned} \mathbb{E} [\langle \langle \mathbf{\Theta}, \boldsymbol{\theta} \rangle^2 \rangle_{h,\varepsilon,t,*}] &= d \mathbb{E} [\langle \mathbf{\Theta}_1^2 \boldsymbol{\theta}_1^2 \rangle_{h,\varepsilon,t,*}] + d(d-1) \mathbb{E} [\langle \mathbf{\Theta}_1 \mathbf{\Theta}_2 \boldsymbol{\theta}_1 \boldsymbol{\theta}_2 \rangle_{h,\varepsilon,t,*}] \\ &\leq d \mathbb{E} [\mathbf{\Theta}_1^4] + d(d-1) \mathbb{E} [\langle \mathbf{\Theta}_1 \mathbf{\Theta}_2 \boldsymbol{\theta}_1 \boldsymbol{\theta}_2 \rangle_{h,\varepsilon,t,*}]. \end{aligned} \tag{96}$$

Recall that $r_n = d^{1/4} n^{-1/4}$. We define the mapping

$$F_{\mathbf{\Theta}}(\delta) := r_n^2 \mathbb{E} [\mathbb{E} [\mathbf{\Theta}_0 \mid r_n^{-1} \delta \mathbf{\Theta}_0 + \mathbf{G}]^2],$$

where $\mathbf{\Theta}_0 \sim \mu_{\mathbf{\Theta}}$, $\mathbf{G} \sim \mathbf{N}(0, 1)$ and $\mathbf{\Theta}_0 \perp \mathbf{G}$. Notice that

$$\mathbb{E} [\langle \mathbf{\Theta}_1 \mathbf{\Theta}_2 \boldsymbol{\theta}_1 \boldsymbol{\theta}_2 \rangle_{h,\varepsilon,t,*}] \leq \frac{n}{d} \mathbb{E} \left[F_{\mathbf{\Theta}} \left(\sqrt{(t \|\mathbf{\Lambda}\|_2^2 + (1-t) \|\bar{\mathbf{\Lambda}}\|_2^2) / n} \right)^2 \right]. \tag{97}$$

Direct computation gives

$$\begin{aligned} & \frac{d}{d\delta} \mathbb{E} [\Theta_0 \mid \delta \Theta_0 + \mathbf{G}] \\ &= (2\delta \Theta_0 + \mathbf{G}) \text{Var}[\Theta_0 \mid \delta \Theta_0 + \mathbf{G}] - \delta \mathbb{E}[\Theta_0^3 \mid \delta \Theta_0 + \mathbf{G}] + \delta \mathbb{E}[\Theta_0^2 \mid \delta \Theta_0 + \mathbf{G}] \mathbb{E}[\Theta_0 \mid \delta \Theta_0 + \mathbf{G}], \quad (98) \end{aligned}$$

Next, we apply triangle inequality to upper bound the right hand side of Eq. (98), which gives

$$\begin{aligned} & \left| \frac{d}{d\delta} \mathbb{E} [\Theta_0 \mid r_n^{-1} \delta \Theta_0 + \mathbf{G}] \right| \\ & \leq r_n^{-1} \times \{ (2r_n^{-1} \delta |\Theta_0| + |\mathbf{G}|) \text{Var} [\Theta_0 \mid r_n^{-1} \delta \Theta_0 + \mathbf{G}] \\ & \quad + r_n^{-1} \delta \mathbb{E}[|\Theta_0|^3 \mid r_n^{-1} \delta \Theta_0 + \mathbf{G}] + r_n^{-1} \delta \mathbb{E}[\Theta_0^2 \mid r_n^{-1} \delta \Theta_0 + \mathbf{G}] \mathbb{E}[|\Theta_0| \mid r_n^{-1} \delta \Theta_0 + \mathbf{G}] \}. \end{aligned}$$

Leveraging the above formulas and Hölder's inequality, we obtain that for n, d large enough

$$\begin{aligned} F_\Theta(\delta) &= \sqrt{\frac{d}{n}} \mathbb{E} [\mathbb{E}[\Theta_0 \mid r_n^{-1} \delta \Theta_0 + \mathbf{G}]^2] \\ &\leq \sqrt{\frac{d}{n}} \mathbb{E} \left[\left(\int_0^\delta \left| \frac{d}{dx} \mathbb{E} [\Theta_0 \mid r_n^{-1} x \Theta_0 + \mathbf{G}] \right| dx \right)^2 \right] \\ &\leq \sqrt{\frac{d}{n}} \mathbb{E} \left[\delta \int_0^\delta \left| \frac{d}{dx} \mathbb{E} [\Theta_0 \mid r_n^{-1} x \Theta_0 + \mathbf{G}] \right|^2 dx \right] \\ &\leq 4\delta \mathbb{E} \left[\int_0^\delta 4r_n^{-2} x^2 \Theta_0^2 \text{Var} [\Theta_0 \mid r_n^{-1} x \Theta_0 + \mathbf{G}]^2 + \mathbf{G}^2 \text{Var} [\Theta_0 \mid r_n^{-1} x \Theta_0 + \mathbf{G}]^2 \right. \\ &\quad \left. + r_n^{-2} x^2 \mathbb{E}[|\Theta_0|^3 \mid r_n^{-1} x \Theta_0 + \mathbf{G}]^2 + r_n^{-2} x^2 \mathbb{E}[\Theta_0^2 \mid r_n^{-1} x \Theta_0 + \mathbf{G}]^2 \mathbb{E}[|\Theta_0| \mid r_n^{-1} x \Theta_0 + \mathbf{G}]^2 dx \right] \\ &\leq C_{\mu_\Theta} (\delta^4 + 1). \quad (99) \end{aligned}$$

In the above display, $C_{\mu_\Theta} > 0$ is a constant that depends only on μ_Θ .

We define the set $S = \{\|\mathbf{\Lambda}\|_2^2 \leq n\mathbb{E}[\mathbf{\Lambda}_0^2] + n\|\mathbf{\Lambda}_0^2\|_{\Psi_1}\}$, where $\|\cdot\|_{\Psi_1}$ is the sub-exponential norm of $\mathbf{\Lambda}_0^2$. Then by Bernstein's inequality [Ver18, Theorem 2.8.1], we can conclude that there exists a constant $C_{\mu_\Lambda} > 0$ depending only on μ_Λ , such that for all $s \geq 1$,

$$\mathbb{P} (\|\mathbf{\Lambda}\|_2^2 \geq n\mathbb{E}[\mathbf{\Lambda}_0^2] + sn\|\mathbf{\Lambda}_0^2\|_{\Psi_1}) \leq 2 \exp(-C_{\mu_\Lambda} ns).$$

Therefore, for n, d large enough we have

$$\begin{aligned} & \frac{n}{d} \mathbb{E} \left[F_\Theta \left(\sqrt{(t\|\mathbf{\Lambda}\|_2^2 + (1-t)\|\bar{\mathbf{\Lambda}}\|_2^2)/n} \right)^2 \right] \\ &= \frac{n}{d} \mathbb{E} \left[F_\Theta \left(\sqrt{(t\|\mathbf{\Lambda}\|_2^2 + (1-t)\|\bar{\mathbf{\Lambda}}\|_2^2)/n} \right)^2 \mathbb{1}_S \right] + \frac{n}{d} \mathbb{E} \left[F_\Theta \left(\sqrt{(t\|\mathbf{\Lambda}\|_2^2 + (1-t)\|\bar{\mathbf{\Lambda}}\|_2^2)/n} \right)^2 \mathbb{1}_{S^c} \right] \\ &\stackrel{(i)}{\leq} \frac{4C_{\mu_\Theta}^2 n}{d} + \frac{2C_{\mu_\Theta}^2 n}{d} (\mathbb{E}[\mathbf{\Lambda}_0^2] + \|\mathbf{\Lambda}_0^2\|_{\Psi_1})^4 + \frac{2nC_{\mu_\Theta}^2}{d} \mathbb{E} \left[(\|\mathbf{\Lambda}\|_2^2/n)^4 \mathbb{1}_{S^c} \right] \\ &\leq \frac{4C_{\mu_\Theta}^2 n}{d} + \frac{2C_{\mu_\Theta}^2 n}{d} (\mathbb{E}[\mathbf{\Lambda}_0^2] + \|\mathbf{\Lambda}_0^2\|_{\Psi_1})^4 + \\ &\quad \frac{2nC_{\mu_\Theta}^2}{d} \int_1^\infty 4\mathbb{P} (\|\mathbf{\Lambda}\|_2^2/n \geq \mathbb{E}[\mathbf{\Lambda}_0^2] + s\|\mathbf{\Lambda}_0^2\|_{\Psi_1}) (\mathbb{E}[\mathbf{\Lambda}_0^2] + s\|\mathbf{\Lambda}_0^2\|_{\Psi_1})^3 \|\mathbf{\Lambda}_0^2\|_{\Psi_1} ds \quad (100) \end{aligned}$$

$$\begin{aligned}
&\leq \frac{4C_{\mu_\Theta}^2 n}{d} + \frac{2C_{\mu_\Theta}^2 n}{d} (\mathbb{E}[\mathbf{\Lambda}_0^2] + \|\mathbf{\Lambda}_0^2\|_{\Psi_1})^4 + \frac{2nC_{\mu_\Theta}^2}{d} \int_1^\infty 8 \exp(-C_{\mu_\Lambda} ns) (\mathbb{E}[\mathbf{\Lambda}_0^2] + s\|\mathbf{\Lambda}_0^2\|_{\Psi_1})^3 \|\mathbf{\Lambda}_0^2\|_{\Psi_1} ds \\
&\leq \frac{C_1 n}{d},
\end{aligned} \tag{101}$$

where $C_1 > 0$ is a constant depending only on $(\mu_\Theta, \mu_\Lambda)$, and in (i) we use Eq. (99). Furthermore,

$$\begin{aligned}
\mathbb{E} \left[\langle (\mathbf{\Lambda}_1 \mathbf{\lambda}_1 - \bar{\mathbf{\Lambda}}_1 \bar{\mathbf{\lambda}}_1)^2 \rangle_{h,\varepsilon,t} \right] &\leq 2\mathbb{E} [\mathbf{\Lambda}_1^2 \langle (\mathbf{\lambda}_1 - \bar{\mathbf{\lambda}}_1)^2 \rangle_{h,\varepsilon,t}] + 2\mathbb{E} [(\mathbf{\Lambda}_1 - \bar{\mathbf{\Lambda}}_1)^2 \langle \bar{\mathbf{\lambda}}_1^2 \rangle_{h,\varepsilon,t}] \\
&\leq 2\mathbb{E} [\mathbf{\Lambda}_1^4]^{1/2} \mathbb{E} [\langle (\mathbf{\lambda}_1 - \bar{\mathbf{\lambda}}_1)^2 \rangle_{h,\varepsilon,t}^2]^{1/2} + 2\mathbb{E} [(\mathbf{\Lambda}_1 - \bar{\mathbf{\Lambda}}_1)^4]^{1/2} \mathbb{E} [\langle \bar{\mathbf{\lambda}}_1^2 \rangle_{h,\varepsilon,t}^2]^{1/2} \\
&\leq 2\mathbb{E} [\mathbf{\Lambda}_1^4]^{1/2} \mathbb{E} [(\mathbf{\Lambda}_1 - \bar{\mathbf{\Lambda}}_1)^4]^{1/2} + 2\mathbb{E} [(\mathbf{\Lambda}_1 - \bar{\mathbf{\Lambda}}_1)^4]^{1/2} \mathbb{E} [\bar{\mathbf{\Lambda}}_1^4] \\
&\leq C_2 \sqrt{\varepsilon},
\end{aligned} \tag{102}$$

where $C_2 > 0$ is a constant depending only on μ_Λ . Finally, we combine Eqs. (96), (97), (101) and (102), and conclude that

$$\left| \frac{\partial}{\partial t} \Phi_{n,t}^\varepsilon(h) \right| \leq C_0 \sqrt[4]{\varepsilon}$$

for all $t, h \in [0, 1]$, where $C_0 > 0$ is a constant depending only on $(\mu_\Theta, \mu_\Lambda)$. This concludes the proof of the lemma.

Appendix F Achieving the Bayesian MMSE

In this section we prove the technical lemmas required to prove Theorem 4.5.

F.1 Proof of Lemma D.9

We define the set

$$\Omega := \left\{ |\mathbf{\Lambda}_i| \leq 2K_0 \sqrt{\log n} : i \in [n] \right\}.$$

By Eq. (42) we see that $\mathbb{P}(\Omega^c) \leq 2n^{-3}$. Furthermore, on Ω we have $\mathbf{A} = \bar{\mathbf{A}}$. For matrix $\mathbf{X} \in \mathbb{R}^{n \times d}$, we define the mapping $\bar{\mathbf{M}} : \mathbb{R}^{n \times d} \rightarrow \mathbb{R}^{n \times n}$ such that $\bar{\mathbf{M}}(\mathbf{X}) = \mathbb{E}[\bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^\top \mid \bar{\mathbf{A}} = \mathbf{X}]$. Then we have

$$\begin{aligned}
\frac{1}{n} \mathbb{E} \left[\left\| \mathbf{\Lambda} \mathbf{\Lambda}^\top - \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \mathbf{A}] \right\|_F^2 \right]^{1/2} &\stackrel{(i)}{\leq} \frac{1}{n} \mathbb{E} \left[\left\| \mathbf{\Lambda} \mathbf{\Lambda}^\top - \bar{\mathbf{M}}(\mathbf{A}) \right\|_F^2 \right]^{1/2} \\
&\stackrel{(ii)}{\leq} \frac{1}{n} \mathbb{E} \left[\left\| \bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^\top - \bar{\mathbf{M}}(\mathbf{A}) \right\|_F^2 \right]^{1/2} + \frac{1}{n} \mathbb{E} \left[\left\| \bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^\top - \mathbf{\Lambda} \mathbf{\Lambda}^\top \right\|_F^2 \right]^{1/2},
\end{aligned} \tag{103}$$

where (i) is by the fact that the posterior expectation achieves Bayesian MMSE, and (ii) is by triangle inequality. Applying triangle inequality and Hölder's inequality, we have

$$\begin{aligned}
\frac{1}{n} \mathbb{E} \left[\left\| \bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^\top - \bar{\mathbf{M}}(\mathbf{A}) \right\|_F^2 \right]^{1/2} &\leq \frac{1}{n} \mathbb{E} \left[\left\| \bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^\top - \bar{\mathbf{M}}(\mathbf{A}) \right\|_F^2 \mathbb{1}_\Omega \right]^{1/2} + \frac{1}{n} \mathbb{E} \left[\left\| \bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^\top - \bar{\mathbf{M}}(\mathbf{A}) \right\|_F^2 \mathbb{1}_{\Omega^c} \right]^{1/2} \\
&\leq \frac{1}{n} \mathbb{E} \left[\left\| \bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^\top - \bar{\mathbf{M}}(\bar{\mathbf{A}}) \right\|_F^2 \right]^{1/2} + \frac{1}{n} \mathbb{E} \left[\left\| \bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^\top - \bar{\mathbf{M}}(\mathbf{A}) \right\|_F^4 \right]^{1/4} \mathbb{P}(\Omega^c)^{1/4}.
\end{aligned} \tag{104}$$

Direct computation reveals that $\mathbb{E}[\|\bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^\top - \mathbf{\Lambda} \mathbf{\Lambda}^\top\|_F^2]^{1/2}/n = o_n(1)$ as $n, d \rightarrow \infty$, and $\mathbb{E}[\|\bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^\top - \bar{\mathbf{M}}(\mathbf{A})\|_F^4]^{1/4}/n \leq 8K_0^2 \log n$. As a result, we conclude that $\mathbb{E}[\|\bar{\mathbf{\Lambda}} \bar{\mathbf{\Lambda}}^\top - \bar{\mathbf{M}}(\mathbf{A})\|_F^4]^{1/4} \mathbb{P}(\Omega^c)^{1/4}/n = o_n(1)$. Combining these analysis with Eqs. (103) and (104) concludes the proof of the lemma.

F.2 Proof of Lemma D.10

Let $\mathbf{W}, \mathbf{W}' \stackrel{iid}{\sim} \text{GOE}(n)$ that are independent of $\mathbf{\Lambda}$. For $t \in [0, 1]$, $s \geq 0$, we define

$$\begin{aligned}\mathbf{Y}_{a,t}^{(s)} &:= \frac{q_\Theta \sqrt{(1-t)s} \mathbf{\Lambda} \mathbf{\Lambda}^\top}{n} + \mathbf{W}, \\ \mathbf{Y}_{b,t}^{(s)} &:= \frac{\sqrt{ts}}{n} (q_\Theta^{1/2} \bar{\mathbf{\Lambda}} + r_n^{-1} \bar{\mathbf{g}}) (q_\Theta^{1/2} \bar{\mathbf{\Lambda}} + r_n^{-1} \bar{\mathbf{g}})^\top + \mathbf{W}'.\end{aligned}$$

For $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we define the corresponding truncated vectors $\bar{\mathbf{x}}, \bar{\mathbf{y}} \in \mathbb{R}^n$ such that $\bar{x}_i = x_i \mathbb{1}\{|x_i| \leq 2K_0 \sqrt{\log n}\}$ and $\bar{y}_i = y_i \mathbb{1}\{|y_i| \leq C_3 \sqrt{\log n}\}$ for all $i \in [n]$. The Hamiltonian that corresponds to $(\mathbf{Y}_{a,t}^{(s)}, \mathbf{Y}_{b,t}^{(s)})$ can be expressed as

$$\begin{aligned}H_{n,t}^{(s)}(\mathbf{x}, \mathbf{y}) &:= \frac{ts}{2n} ((q_\Theta^{1/2} \bar{\mathbf{\Lambda}} + r_n^{-1} \bar{\mathbf{g}})^\top (q_\Theta^{1/2} \bar{\mathbf{x}} + r_n^{-1} \bar{\mathbf{y}}))^2 + \frac{\sqrt{ts}}{2} (q_\Theta^{1/2} \bar{\mathbf{x}} + r_n^{-1} \bar{\mathbf{y}})^\top \mathbf{W}' (q_\Theta^{1/2} \bar{\mathbf{x}} + r_n^{-1} \bar{\mathbf{y}}) - \frac{ts}{4n} \|q_\Theta^{1/2} \bar{\mathbf{x}} + r_n^{-1} \bar{\mathbf{y}}\|^4 \\ &\quad + \frac{(1-t)sq_\Theta^2}{2n} (\mathbf{\Lambda}^\top \mathbf{x})^2 + \frac{\sqrt{(1-t)s} q_\Theta}{2} \mathbf{x}^\top \mathbf{W} \mathbf{x} - \frac{(1-t)sq_\Theta^2}{4n} \|\mathbf{x}\|^4.\end{aligned}$$

The corresponding free energy density can be written as

$$G_n(t, s) := \frac{1}{n} \mathbb{E} \left[\log \left(\int \exp \left(H_{n,t}^{(s)}(\mathbf{x}, \mathbf{y}) \right) P_{\mathbf{\Lambda}}^{\otimes n}(\mathrm{d}\mathbf{x}) P_{\mathbf{N}(0,1)}^{\otimes n}(\mathrm{d}\mathbf{y}) \right) \right],$$

where $P_{\mathbf{N}(0,1)}^{\otimes n}$ is the distribution of $\mathbf{N}(\mathbf{0}, \mathbf{I}_n)$. Invoking Lemma B.3 and Gaussian integration by parts, we obtain that

$$\begin{aligned}\frac{\partial}{\partial t} G_n(t, s) &= \frac{s}{4n^2} \mathbb{E} \left[\left\| \mathbb{E}[(q_\Theta^{1/2} \bar{\mathbf{\Lambda}} + r_n^{-1} \bar{\mathbf{g}})(q_\Theta^{1/2} \bar{\mathbf{\Lambda}} + r_n^{-1} \bar{\mathbf{g}})^\top \mid \mathbf{Y}_{a,t}^{(s)}, \mathbf{Y}_{b,t}^{(s)}] \right\|_F^2 \right] \\ &\quad - \frac{sq_\Theta^2}{4n^2} \mathbb{E} \left[\left\| \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \mathbf{Y}_{a,t}^{(s)}, \mathbf{Y}_{b,t}^{(s)}] \right\|_F^2 \right].\end{aligned}$$

Leveraging Hölder's inequality, we see that

$$\begin{aligned}& \left| \frac{\partial}{\partial t} G_n(t, s) \right| \\ & \leq \left| \mathbb{E} \left[\frac{s}{4} \mathbb{E} \left[(q_\Theta^{1/2} \bar{\mathbf{\Lambda}}_1 + r_n^{-1} \bar{g}_1)(q_\Theta^{1/2} \bar{\mathbf{\Lambda}}_2 + r_n^{-1} \bar{g}_2) \mid \mathbf{Y}_{a,t}^{(s)}, \mathbf{Y}_{b,t}^{(s)} \right]^2 - \frac{sq_\Theta^2}{4} \mathbb{E} \left[\bar{\mathbf{\Lambda}}_1 \bar{\mathbf{\Lambda}}_2 \mid \mathbf{Y}_{a,t}^{(s)}, \mathbf{Y}_{b,t}^{(s)} \right]^2 \right] \right| + o_n(1) \\ & \leq \frac{s}{4} \mathbb{E} \left[\left(q_\Theta \bar{\mathbf{\Lambda}}_1 \bar{\mathbf{\Lambda}}_2 + r_n^{-1} q_\Theta^{1/2} \bar{\mathbf{\Lambda}}_1 \bar{g}_2 + r_n^{-1} q_\Theta^{1/2} \bar{\mathbf{\Lambda}}_2 \bar{g}_1 + r_n^{-2} \bar{g}_1 \bar{g}_2 - q_\Theta \bar{\mathbf{\Lambda}}_1 \bar{\mathbf{\Lambda}}_2 \right)^2 \right]^{1/2} \times \\ & \quad \mathbb{E} \left[\left(q_\Theta \bar{\mathbf{\Lambda}}_1 \bar{\mathbf{\Lambda}}_2 + r_n^{-1} q_\Theta^{1/2} \bar{\mathbf{\Lambda}}_1 \bar{g}_2 + r_n^{-1} q_\Theta^{1/2} \bar{\mathbf{\Lambda}}_2 \bar{g}_1 + r_n^{-2} \bar{g}_1 \bar{g}_2 + q_\Theta \bar{\mathbf{\Lambda}}_1 \bar{\mathbf{\Lambda}}_2 \right)^2 \right]^{1/2} + o_n(1).\end{aligned}\tag{105}$$

The upper bound given in the last line of Eq. (105) is independent of t and converges to 0 as $n, d \rightarrow \infty$. Therefore, we conclude that as $n, d \rightarrow \infty$

$$\sup_{t \in (0,1), s \in [0,2]} \left| \frac{\partial}{\partial t} G_n(t, s) \right| \rightarrow 0,$$

which further implies that $|G_n(1, s) - G_n(0, s)| = o_n(1)$ for all $s \in [0, 2]$. Recall that $\mathcal{F}(\cdot, \cdot)$ is defined in Eq. (13). Using [LM19, Theorem 13], we have $G_n(0, s) = \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 s, q) + o_n(1)$. Observe that

$s \mapsto G_n(1, s)$ is convex differentiable on $(0, \infty)$, and converges point-wisely to $\sup_{q \geq 0} \mathcal{F}(q_\Theta^2 s, q)$ as $n, d \rightarrow \infty$, the later is differentiable at $s = 1$ for all but countably many values of $q_\Theta > 0$ according to [LM19, Proposition 17]. Invoking Lemma B.4, Lemma B.3 and Gaussian integration by parts, we conclude that for all but countably many $q_\Theta > 0$

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \mathbb{E} \left[\left\| q_\Theta^{-1} \mathbf{M}_n(\mathbf{Y}_2) - q_\Theta^{-1} (q_\Theta^{1/2} \bar{\mathbf{A}} + r_n^{-1} \bar{\mathbf{g}}) (q_\Theta^{1/2} \bar{\mathbf{A}} + r_n^{-1} \bar{\mathbf{g}})^\top \right\|_F^2 \right] = \lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_\Lambda; q_\Theta).$$

Notice that $\lim_{n \rightarrow \infty} \mathbb{E}[\|(q_\Theta^{1/2} \bar{\mathbf{A}} + r_n^{-1} \bar{\mathbf{g}})(q_\Theta^{1/2} \bar{\mathbf{A}} + r_n^{-1} \bar{\mathbf{g}})^\top / n - q_\Theta \bar{\mathbf{A}} \bar{\mathbf{A}}^\top / n\|_F^2] = 0$, then the proof of the lemma follows immediately from triangle inequality.

F.3 Proof of Lemma D.11

We define the set

$$\Omega := \left\{ |\Theta_j| \leq 2K_2 \sqrt{\log d} : j \in [d] \right\}.$$

By Eq. (44) we have $\mathbb{P}(\Omega^c) \leq 2d^{-3}$. Furthermore, on the set Ω we have $\mathbf{A} = \bar{\mathbf{A}}$. For $\mathbf{X} \in \mathbb{R}^{n \times d}$, we define the mapping $\mathbf{M}(\mathbf{X}) := \mathbb{E}[\mathbf{A} \mathbf{A}^\top \mid \bar{\mathbf{A}} = \mathbf{X}]$. Leveraging triangle inequality and Hölder's inequality, we obtain that

$$\begin{aligned} \frac{1}{n} \mathbb{E} \left[\left\| \mathbf{A} \mathbf{A}^\top - M(\mathbf{A}) \right\|_F^2 \right]^{1/2} &\leq \frac{1}{n} \mathbb{E} \left[\left\| \mathbf{A} \mathbf{A}^\top - M(\mathbf{A}) \right\|_F^2 \mathbb{1}_\Omega \right]^{1/2} + \frac{1}{n} \mathbb{E} \left[\left\| \mathbf{A} \mathbf{A}^\top - M(\mathbf{A}) \right\|_F^2 \mathbb{1}_{\Omega^c} \right]^{1/2} \\ &\leq \frac{1}{n} \mathbb{E} \left[\left\| \mathbf{A} \mathbf{A}^\top - M(\bar{\mathbf{A}}) \right\|_F^2 \right]^{1/2} + \frac{1}{n} \mathbb{E} \left[\left\| \mathbf{A} \mathbf{A}^\top - M(\mathbf{A}) \right\|_F^4 \right]^{1/4} \mathbb{P}(\Omega^c)^{1/4}. \end{aligned}$$

Since the posterior expectation minimizes the expected l^2 risk, we then have

$$\frac{1}{n^2} \mathbb{E} \left[\left\| \mathbf{A} \mathbf{A}^\top - M(\mathbf{A}) \right\|_F^2 \right] \geq \frac{1}{n^2} \mathbb{E} \left[\left\| \mathbf{A} \mathbf{A}^\top - \mathbb{E}[\mathbf{A} \mathbf{A}^\top \mid \mathbf{A}] \right\|_F^2 \right].$$

By the bounded-support assumption, we see that $\mathbb{E}[\|\mathbf{A} \mathbf{A}^\top - M(\mathbf{A})\|_F^4]^{1/4}/n \leq 2K_1^2$. Thus, as $n, d \rightarrow \infty$,

$$\frac{1}{n} \mathbb{E} \left[\left\| \mathbf{A} \mathbf{A}^\top - M(\mathbf{A}) \right\|_F^4 \right]^{1/4} \mathbb{P}(\Omega^c)^{1/4} \rightarrow 0,$$

which completes the proof of the lemma.

F.4 Proof of Lemma D.12

For $t \in [0, 1]$, we define the interpolated Hamiltonian as

$$\begin{aligned} H_{n,t}^{[s]}(\boldsymbol{\lambda}, \boldsymbol{\theta}; h) &:= \sum_{i \in [n], j \in [d]} \left\{ \frac{ts}{\sqrt{nd}} \boldsymbol{\Lambda}_i \boldsymbol{\lambda}_i \boldsymbol{\Theta}_j \boldsymbol{\theta}_j + \frac{\sqrt{ts}}{\sqrt{nd}} Z_{ij} \boldsymbol{\lambda}_i \boldsymbol{\theta}_j - \frac{ts}{2\sqrt{nd}} \boldsymbol{\lambda}_i^2 \boldsymbol{\theta}_j^2 \right\} + \\ &\quad \sum_{i \in [n], j \in [d]} \left\{ \frac{s(1-t)}{\sqrt{nd}} \boldsymbol{\Lambda}_i \boldsymbol{\lambda}_i \bar{\boldsymbol{\Theta}}_j \bar{\boldsymbol{\theta}}_j + \frac{\sqrt{s(1-t)}}{\sqrt{nd}} Z'_{ij} \boldsymbol{\lambda}_i \bar{\boldsymbol{\theta}}_j - \frac{s(1-t)}{2\sqrt{nd}} \boldsymbol{\lambda}_i^2 \bar{\boldsymbol{\theta}}_j^2 \right\} + H_n(\boldsymbol{\lambda}; \mathbf{Y}'(h)), \end{aligned}$$

where $\mathbf{Z}' = (Z'_{ij})_{i \in [n], j \in [d]}$ is an independent copy of \mathbf{Z} and is independent of everything else. Note that $H_{n,t}^{[s]}(\boldsymbol{\lambda}, \boldsymbol{\theta}; h)$ is the Hamiltonian corresponding to the observations $(\mathbf{A}_1^{(s,t)}, \mathbf{A}_2^{(s,t)}, \mathbf{Y}'(h))$, where

$\mathbf{A}_1^{(s,t)} = \sqrt{ts}\mathbf{\Lambda}\mathbf{\Theta}^\top/\sqrt[4]{nd} + \mathbf{Z}$, $\mathbf{A}_2^{(s,t)} = \sqrt{(1-t)s}\mathbf{\Lambda}\bar{\mathbf{\Theta}}^\top/\sqrt[4]{nd} + \mathbf{Z}'$ and $\mathbf{Y}'(h) = \sqrt{h}\mathbf{\Lambda}\mathbf{\Lambda}^\top/n + \mathbf{W}'$. Here, we recall that $\mathbf{W}' \sim \text{GOE}(n)$, and \mathbf{W}' , \mathbf{Z} , \mathbf{Z}' are mutually independent. We define the free energy density corresponding to the Hamiltonian $H_{n,t}^{[s]}(\boldsymbol{\lambda}, \boldsymbol{\theta}; h)$ as

$$\Phi_{n,t}^{[s]}(h) := \frac{1}{n} \mathbb{E} \left[\log \left(\int \exp \left(H_{n,t}^{[s]}(\boldsymbol{\lambda}, \boldsymbol{\theta}; h) \right) \mu_{\mathbf{\Lambda}}^{\otimes n}(\mathrm{d}\boldsymbol{\lambda}) \mu_{\bar{\mathbf{\Theta}}}^{\otimes d}(\mathrm{d}\boldsymbol{\theta}) \right) \right].$$

At the endpoints, we have $\Phi_{n,0}^{[s]}(h) = \bar{\Phi}_n(s, 0, 0, h)$ and $\Phi_{n,1}^{[s]}(h) = \Phi_n(s, 0, 0, h)$. We denote by $\langle \cdot \rangle_{t,h}^{[s]}$ the expectation with respect to the posterior distribution $\mathbb{P}(\cdot \mid \mathbf{A}_1^{(s,t)}, \mathbf{A}_2^{(s,t)}, \mathbf{Y}'(h))$. Then we have

$$\begin{aligned} \left| \frac{\partial}{\partial t} \Phi_{n,t}^{[s]}(h) \right| &\stackrel{(i)}{=} \frac{s}{2n\sqrt{nd}} \sum_{i \in [n], j \in [d]} \mathbb{E} \left[\langle \boldsymbol{\Lambda}_i \boldsymbol{\lambda}_i (\boldsymbol{\Theta}_j \boldsymbol{\theta}_j - \bar{\boldsymbol{\Theta}}_j \bar{\boldsymbol{\theta}}_j) \rangle_{t,h}^{[s]} \right] \\ &\stackrel{(ii)}{\leq} \frac{s}{2n\sqrt{nd}} \sum_{i \in [n], j \in [d]} \mathbb{E} \left[\langle \boldsymbol{\Lambda}_i^2 \boldsymbol{\lambda}_i^2 \rangle_{t,h}^{[s]} \right]^{1/2} \mathbb{E} \left[\langle (\boldsymbol{\Theta}_j \boldsymbol{\theta}_j - \bar{\boldsymbol{\Theta}}_j \bar{\boldsymbol{\theta}}_j)^2 \rangle_{t,h}^{[s]} \right]^{1/2}, \end{aligned} \quad (106)$$

where (i) is by Gaussian integration by parts and Nishimori identity (Lemma B.3), and (ii) is by Hölder's inequality. For all $j \in [d]$, using power mean inequality and Hölder's inequality, we have

$$\begin{aligned} \mathbb{E} \left[\langle (\boldsymbol{\Theta}_j \boldsymbol{\theta}_j - \bar{\boldsymbol{\Theta}}_j \bar{\boldsymbol{\theta}}_j)^2 \rangle_{t,h}^{[s]} \right] &\leq 2\mathbb{E}[\boldsymbol{\Theta}_j^2 \langle (\boldsymbol{\theta}_j - \bar{\boldsymbol{\theta}}_j)^2 \rangle_{t,h}^{[s]}] + 2\mathbb{E}[(\boldsymbol{\Theta}_j - \bar{\boldsymbol{\Theta}}_j)^2 \langle \bar{\boldsymbol{\theta}}_j^2 \rangle_{t,h}^{[s]}] \\ &\leq 2\mathbb{E}[\boldsymbol{\Theta}_j^4]^{1/2} \mathbb{E}[\langle (\boldsymbol{\theta}_j - \bar{\boldsymbol{\theta}}_j)^2 \rangle_{t,h}^{[s]2}]^{1/2} + 2\mathbb{E}[(\boldsymbol{\Theta}_j - \bar{\boldsymbol{\Theta}}_j)^4]^{1/2} \mathbb{E}[\langle \bar{\boldsymbol{\theta}}_j^2 \rangle_{t,h}^{[s]2}]^{1/2} \\ &\leq 2\mathbb{E}[\boldsymbol{\Theta}_j^4]^{1/2} \mathbb{E}[(\boldsymbol{\Theta}_j - \bar{\boldsymbol{\Theta}}_j)^4]^{1/2} + 2\mathbb{E}[(\boldsymbol{\Theta}_j - \bar{\boldsymbol{\Theta}}_j)^4]^{1/2} \mathbb{E}[\bar{\boldsymbol{\theta}}_j^4]^{1/2}. \end{aligned} \quad (107)$$

Notice that

$$\begin{aligned} \mathbb{E}[(\boldsymbol{\Theta}_j - \bar{\boldsymbol{\Theta}}_j)^4] &\leq \int_{2K_2\sqrt{\log d}}^{\infty} 4x^3 \mathbb{P}(|\boldsymbol{\Theta}_0| \geq x) \mathrm{d}x \leq 4 \int_{4K_2^2 \log d}^{\infty} y \exp\left(-\frac{y}{K_2^2}\right) \mathrm{d}y \\ &= -4(yK_2^2 + K_2^4) \exp\left(-\frac{y}{K_2^2}\right) \Big|_{4K_2^2 \log d}^{\infty} = \frac{4K_2^4 + 16K_2^4 \log d}{d^4}. \end{aligned} \quad (108)$$

Combining Eqs. (106) to (108), we obtain that $\sup_{t \in [0,1], h \geq 0, S_0 \geq s \geq 0} \left| \frac{\partial}{\partial t} \Phi_{n,t}^{(s)}(h) \right| \rightarrow 0$ as $n, d \rightarrow \infty$, thus completing the proof of the lemma.

F.5 Proof of Lemma D.13

Using Lemma D.12, we have $\lim_{n,d \rightarrow \infty} |\bar{\Phi}_n(s, 0, 0, h) - \Phi_n(s, 0, 0, h)| = 0$. Similar to the proof of Lemma D.8, we can conclude that $\lim_{n,d \rightarrow \infty} |\Phi_n(s, 0, 0, h) - \sup_{q \geq 0} \mathcal{F}(q_{\bar{\mathbf{\Theta}}}^2 s^2 + h, q)| = 0$ as $n, d \rightarrow \infty$. Therefore, in order to prove the lemma, it suffices to show

$$\lim_{n,d \rightarrow \infty} \sup_{a, a' \in [0,10]} |\bar{\Phi}_n(s, a, a', h) - \bar{\Phi}_n(s, 0, 0, h)| = 0.$$

Using Gaussian integration by parts and Nishimori identity (Lemma B.3), we obtain that for all $a, a' \in [0, 10]$,

$$\begin{aligned} \frac{\partial}{\partial \varepsilon_n} \bar{\Phi}_n(s, a, a', h) &= \frac{a^2}{2d} \mathbb{E} \left[\bar{\mathbf{\Theta}}^\top \mathbb{E}[\bar{\mathbf{\Theta}} \mid \bar{\mathbf{A}}(s), \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h)] \right] \leq 50 \mathbb{E}_{\bar{\mathbf{\Theta}}_0 \sim \mu_{\bar{\mathbf{\Theta}}}}[\bar{\mathbf{\Theta}}_0^2], \\ \frac{\partial}{\partial \varepsilon'_n} \bar{\Phi}_n(s, a, a', h) &= \frac{a'^2}{2n} \mathbb{E} \left[\boldsymbol{\Lambda}^\top \mathbb{E}[\boldsymbol{\Lambda} \mid \bar{\mathbf{A}}(s), \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h)] \right] \leq 50 \mathbb{E}_{\boldsymbol{\Lambda}_0 \sim \mu_{\boldsymbol{\Lambda}}}[\boldsymbol{\Lambda}_0^2]. \end{aligned} \quad (109)$$

Notice that if $\varepsilon_n = \varepsilon'_n = 0$, then $\bar{\Phi}_n(s, a, a', h) = \bar{\Phi}_n(s, 0, 0, h)$. Therefore, by Eq. (109), we conclude that as $n, d \rightarrow \infty$,

$$\sup_{a, a' \in [0, 10]} |\bar{\Phi}_n(s, a, a', h) - \bar{\Phi}_n(s, 0, 0, h)| \leq 50(\mathbb{E}_{\Lambda_0 \sim \mu_\Lambda}[\Lambda_0^2] + \mathbb{E}_{\Theta_0 \sim \mu_\Theta}[\bar{\Theta}_0^2])(\varepsilon_n + \varepsilon'_n) \rightarrow 0,$$

thus completing the proof of the lemma.

F.6 Proof of Lemma D.14

Since $|\bar{\Theta}_0| \leq 2K_2\sqrt{\log d}$, we then have

$$\begin{aligned} & \left| \mathbb{E}[\langle U(\bar{\Theta}^{(1)}) (\bar{\Theta}^{(1)})^\top \bar{\Theta}^{(2)} / d \rangle_{1, a, a', h}] - \mathbb{E}[\langle (\bar{\Theta}^{(1)})^\top \bar{\Theta}^{(2)} / d \rangle_{1, a, a', h}] \mathbb{E}[\langle U(\bar{\Theta}^{(1)}) \rangle_{1, a, a', h}] \right| \\ & \leq 4K_2^2 \log d \mathbb{E}[\langle |U(\bar{\Theta}) - \mathbb{E}[\langle U(\bar{\Theta}) \rangle_{1, a, a', h}] \rangle_{1, a, a', h}]. \end{aligned} \quad (110)$$

Using Gaussian integration by parts and Nishimori identity (Lemma B.3), we have

$$\mathbb{E}[\langle (\bar{\Theta}^{(1)})^\top \bar{\Theta}^{(2)} / d \rangle_{1, a, a', h}] \mathbb{E}[\langle U(\bar{\Theta}^{(1)}) \rangle_{1, a, a', h}] = a \mathbb{E}[\langle (\bar{\Theta}^{(1)})^\top \bar{\Theta}^{(2)} / d \rangle_{1, a, a', h}]^2, \quad (111)$$

$$\mathbb{E}[\langle U(\bar{\Theta}^{(1)}) (\bar{\Theta}^{(1)})^\top \bar{\Theta}^{(2)} / d \rangle_{1, a, a', h}] = a \mathbb{E}[\langle ((\bar{\Theta}^{(1)})^\top \bar{\Theta}^{(2)} / d)^2 \rangle_{1, a, a', h}]. \quad (112)$$

Next, we combine Eqs. (110) to (112), and conclude that for all $a \in [10^{-1}, 10]$,

$$\begin{aligned} & \mathbb{E}[\langle ((\bar{\Theta}^{(1)})^\top \bar{\Theta}^{(2)} / d - \mathbb{E}[\langle (\bar{\Theta}^{(1)})^\top \bar{\Theta}^{(2)} / d \rangle_{1, a, a', h}])^2 \rangle_{1, a, a', h}] \\ & \leq 40K_2^2 \log d \mathbb{E}[\langle |U(\bar{\Theta}) - \mathbb{E}[\langle U(\bar{\Theta}) \rangle_{1, a, a', h}] \rangle_{1, a, a', h}]. \end{aligned}$$

F.7 Proof of Lemma D.15

One can verify that $\bar{\phi}_n(1, a, a', h)$ is twice differentiable for $a, a' \in (0, 10)$. Using Gaussian integration by parts and Nishimori identity, we can compute its partial derivatives:

$$\frac{\partial}{\partial a} \bar{\phi}_n(1, a, a', h) = \varepsilon_n \langle U(\bar{\Theta}) \rangle_{1, a, a', h}, \quad (113)$$

$$\frac{\partial^2}{\partial a^2} \bar{\phi}_n(1, a, a', h) = n\varepsilon_n^2 \langle (U(\bar{\Theta}) - \langle U(\bar{\Theta}) \rangle_{1, a, a', h})^2 \rangle_{1, a, a', h} + \varepsilon_n \langle 2\bar{\Theta}^\top \bar{\Theta} / d - \bar{\Theta}^\top \bar{\Theta} / d \rangle_{1, a, a', h}. \quad (114)$$

Notice that $|\bar{\Theta}_0| \leq 2K_2\sqrt{\log d}$, then $|\mathbb{E}[\frac{\partial}{\partial a} \bar{\phi}_n(1, a, a', h)]| = \varepsilon_n a \mathbb{E}[\langle \bar{\Theta}^\top \bar{\Theta} / d \rangle_{1, a, a', h}] \leq 40\varepsilon_n K_2^2 \log d$ for all $a, a' \in (0, 10)$. Using these results, we further obtain that

$$\begin{aligned} & \langle (U(\bar{\Theta}) - \langle U(\bar{\Theta}) \rangle_{1, a, a', h})^2 \rangle_{1, a, a', h} \leq \frac{1}{n\varepsilon_n^2} \left(\frac{\partial^2}{\partial a^2} \bar{\phi}_n(1, a, a', h) + 12K_2^2 \varepsilon_n \log d \right), \\ & \int_1^2 \int_1^2 \mathbb{E}[\langle (U(\bar{\Theta}) - \langle U(\bar{\Theta}) \rangle_{1, a, a', h})^2 \rangle_{1, a, a', h}] da da' \\ & \leq \int_1^2 \mathbb{E} \left[\frac{1}{n\varepsilon_n^2} \left(\frac{\partial}{\partial a} \bar{\phi}_n(1, a, a', h) \Big|_{a=2} - \frac{\partial}{\partial a} \bar{\phi}_n(1, a, a', h) \Big|_{a=1} + 12K_2^2 \varepsilon_n \log d \right) \right] da' \\ & \leq CK_2^2 n^{-1} \varepsilon_n^{-1} \log d, \end{aligned} \quad (115)$$

where $C > 0$ is a numerical constant. Leveraging Eq. (114), we conclude that the following two functions are convex for all fixed $a' \in (0, 10)$ and $h \geq 0$:

$$a \mapsto \bar{\phi}_n(1, a, a', h) + 6\varepsilon_n K_2^2 a^2 \log d,$$

$$a \mapsto \mathbb{E}[\bar{\phi}_n(1, a, a', h)] + 6\varepsilon_n K_2^2 a^2 \log d.$$

By Lemma B.5, for all $a \in [1, 2]$, $b \in (0, 1/2)$, $a' \in [1/2, 3]$ and $h \geq 0$, we have

$$\begin{aligned} & \mathbb{E} \left[\left| \frac{\partial}{\partial a} \bar{\phi}_n(1, a, a', h) - E \left[\frac{\partial}{\partial a} \bar{\phi}_n(1, a, a', h) \right] \right| \right] \\ & \leq \mathbb{E} \left[\frac{\partial}{\partial a} \bar{\phi}_n(1, a+b, a', h) - \frac{\partial}{\partial a} \bar{\phi}_n(1, a-b, a', h) \right] + 24\varepsilon_n K_2^2 b \log d + \frac{3v_n(h)}{b}. \end{aligned} \quad (116)$$

Again we use the fact that $|\mathbb{E}[\frac{\partial}{\partial a} \bar{\phi}_n(1, a, a', h)]| \leq 40\varepsilon_n K_2^2 \log d$ for all $a, a' \in (0, 10)$, and conclude that

$$\begin{aligned} & \int_1^2 \mathbb{E} \left[\frac{\partial}{\partial a} \bar{\phi}_n(1, a+b, a', h) - \frac{\partial}{\partial a} \bar{\phi}_n(1, a-b, a', h) \right] da \\ & = \mathbb{E} [\bar{\phi}_n(1, b+2, a', h) - \bar{\phi}_n(1, b+1, a', h) - \bar{\phi}_n(1, 2-b, a', h) + \bar{\phi}_n(1, 1-b, a', h)] \\ & \leq C' K_2^2 b \varepsilon_n \log d, \end{aligned} \quad (117)$$

where $C' > 0$ is a numerical constant. Then we combine Eqs. (116) and (117) and obtain that

$$\int_1^2 \int_1^2 \mathbb{E} \left[\left| \frac{\partial}{\partial a} \bar{\phi}_n(1, a, a', h) - E \left[\frac{\partial}{\partial a} \bar{\phi}_n(1, a, a', h) \right] \right| \right] da da' \leq C'' \left(b \varepsilon_n K_2^2 \log d + \frac{v_n(h)}{b} \right), \quad (118)$$

where $C'' > 0$ is another numerical constant. Later in Lemma D.16 we will see that under the current conditions, for n, d large enough we have $v_n(h) < \frac{1}{4} K_2^2 \varepsilon_n \log d$. Since b is arbitrary in $(0, 1/2)$, we can then take $b = \sqrt{v_n(h)/(\varepsilon_n K_2^2 \log d)}$ in Eq. (118) and apply this to Eq. (113), which gives

$$\int_1^2 \int_1^2 \mathbb{E} [|\langle U(\bar{\theta}) \rangle_{1,a,a',h} - \mathbb{E}[\langle U(\bar{\theta}) \rangle_{1,a,a',h}]|] da da' \leq 2C'' K_2 \sqrt{v_n(h) \varepsilon_n^{-1} \log d}. \quad (119)$$

Finally, we combine Hölder's inequality, Eqs. (115) and (119) and concludes the proof of the lemma.

F.8 Proof of Lemma D.16

Conditioning on $(\mathbf{\Lambda}, \bar{\Theta})$, we consider the mapping

$$f : (\mathbf{Z}, \mathbf{g}, \mathbf{g}', \sqrt{n} \mathbf{W}') \mapsto \bar{\phi}_n(1, a, a', h).$$

For n, d large enough, the following inequality holds for all $a, a' \in [0, 10]$.

$$\|\nabla f\|^2 \leq C K_1^2 K_2^2 d^{1/2} n^{-3/2} \log d,$$

where $C > 0$ is a numerical constant. By Gaussian Poincaré inequality [VH14], we conclude that for n, d large enough

$$\mathbb{E}_{\mathbf{Z}, \mathbf{g}, \mathbf{g}', \mathbf{W}'} \left[(\bar{\phi}_n(1, a, a', h) - \mathbb{E}_{\mathbf{Z}, \mathbf{g}, \mathbf{g}', \mathbf{W}'} [\bar{\phi}_n(1, a, a', h)])^2 \right] \leq C K_1^2 K_2^2 d^{1/2} n^{-3/2} \log d. \quad (120)$$

In the above display, the expectations are taken over $(\mathbf{Z}, \mathbf{g}, \mathbf{g}', \mathbf{W}')$.

Next, we show that $\mathbb{E}_{\mathbf{Z}, \mathbf{g}, \mathbf{g}', \mathbf{W}'} [\bar{\phi}_n(1, a, a', h)]$ (as a function of $(\mathbf{\Lambda}, \bar{\Theta})$), concentrates around its expectation. Notice that for n, d large enough, for all $i \in [n], j \in [d]$ we have

$$\left| \frac{\partial}{\partial \mathbf{\Lambda}_i} \mathbb{E}_{\mathbf{Z}, \mathbf{g}, \mathbf{g}', \mathbf{W}'} [\bar{\phi}_n(1, a, a', h)] \right| \leq C' K_1 K_2^2 d^{1/2} n^{-3/2} \log d,$$

$$\left| \frac{\partial}{\partial \Theta_j} \mathbb{E}_{\mathbf{Z}, \mathbf{g}, \mathbf{g}', \mathbf{W}'} [\bar{\phi}_n(1, a, a', h)] \right| \leq C'' K_1^2 K_2 d^{-1/2} n^{-1/2} (\log d)^{1/2},$$

where $C', C'' > 0$ are numerical constants. By Efron-Stein inequality [VH14], we see that there exists a numerical constant $C''' > 0$, such that for n, d large enough

$$\mathbb{E} \left[\left(\mathbb{E}_{\mathbf{Z}, \mathbf{g}, \mathbf{g}', \mathbf{W}'} [\bar{\phi}_n(1, a, a', h)] - \mathbb{E} [\bar{\phi}_n(1, a, a', h)] \right)^2 \right] \leq C''' K_1^4 K_2^4 d n^{-2} (\log d)^2. \quad (121)$$

Finally, we combine Eqs. (120) and (121) and conclude that for n, d large enough, there exists a numerical constant $C_1 > 0$, such that for all $a, a' \in [0, 10]$

$$\mathbb{E} \left[\left(\bar{\phi}_n(1, a, a', h) - \mathbb{E} [\bar{\phi}_n(1, a, a', h)] \right)^2 \right] \leq C_1^2 K_1^4 K_2^4 d n^{-2} (\log d)^2,$$

which concludes the proof of the lemma using Cauchy-Schwarz inequality.

F.9 Proof of the first claim of Lemma D.17

Invoking Lemma D.13, as $n, d \rightarrow \infty$ we have

$$\sup_{a \in [0, 2a_*], a' \in [0, 2a'_*]} \left| \bar{\Phi}_n(1, a, a', h) - \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q) \right| = o_n(1). \quad (122)$$

By Jensen's inequality, for $a \sim \text{Unif}[a_*/2, a_*]$ and $a' \sim \text{Unif}[a'_*/2, a'_*]$ we have

$$\begin{aligned} & \frac{1}{n^2} \mathbb{E} \left[\left\| \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}(1), \mathbf{Y}'(h)] - \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}(1), \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h)] \right\|_F^2 \right] \\ & \leq \frac{1}{n^2} \mathbb{E} \left[\left\| \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}(1), \mathbf{Y}'(h)] - \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}(1), \mathbf{x}'(a'_*), \bar{\mathbf{x}}(a_*), \mathbf{Y}'(h)] \right\|_F^2 \right]. \end{aligned} \quad (123)$$

Notice that the mapping $h \mapsto \bar{\Phi}_n(1, a_*, a'_*, h)$ is convex and differentiable, and $h \mapsto \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q)$ is differentiable for all $q_\Theta^2 + h \in D$. Therefore, using Gaussian integration by parts, Lemmas B.3 and B.4, we conclude that for $h + q_\Theta^2 \in D$, as $n, d \rightarrow \infty$ the right hand side of Eq. (123) converges to 0 as $n, d \rightarrow \infty$. Furthermore,

$$\frac{\partial}{\partial a'} \mathbb{E} [\|\mathbf{\Lambda}_{1, a, a', h}\|^2] = \varepsilon'_n a' \mathbb{E} \left[\left\| \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}(1), \bar{\mathbf{x}}(a), \mathbf{x}'(a'), \mathbf{Y}'(h)] - \mathbf{\Lambda}_{1, a, a', h} \mathbf{\Lambda}_{1, a, a', h}^\top \right\|_F^2 \right] \leq 4K_1^4 n^2 \varepsilon'_n a'.$$

Therefore, for $a \sim \text{Unif}[a_*/2, a_*]$ and $a' \sim \text{Unif}[a'_*/2, a'_*]$, using the above equation we have

$$\begin{aligned} & \frac{1}{n^2} \mathbb{E} \left[\left\| \mathbb{E}[\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}(1), \mathbf{x}'(a'), \bar{\mathbf{x}}(a), \mathbf{Y}'(h)] - \mathbf{\Lambda}_{1, a, a', h} \mathbf{\Lambda}_{1, a, a', h}^\top \right\|_F^2 \right] \\ & \leq \frac{8}{n^2 a_* (a'_*)^2 \varepsilon'_n} \int_{a_*/2}^{a_*} \int_{a'_*/2}^{a'_*} \frac{\partial}{\partial a'} \mathbb{E} [\|\mathbf{\Lambda}_{1, a, a', h}\|^2] da' da \\ & \leq \frac{4K_1^2}{n \varepsilon'_n (a'_*)^2}, \end{aligned} \quad (124)$$

which converges to zero as $n, d \rightarrow \infty$ under the current assumptions. Eqs. (123) and (124) imply

$$\frac{1}{n^2} \mathbb{E} [\|\mathbf{\Lambda} \mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}(1), \mathbf{Y}'(h)\|_F^2] = \frac{1}{n^2} \|\mathbf{\Lambda}_{1, a, a', h}\|^4 + o_P(1).$$

By Corollary D.1 we have

$$\frac{1}{n} \|\mathbf{\Lambda}_{1,a,a',h}\|^2 = \frac{1}{n} \mathbb{E} [\|\mathbf{\Lambda}_{1,a,a',h}\|^2 \mid a, a'] + o_P(1).$$

By Jensen's inequality, for all $a \in [a_*/2, a_*]$ and $a' \in [a'_*/2, a'_*]$ we have

$$\mathbb{E} [\|\mathbf{\Lambda}_{1,a,a',h}\|^2 \mid a, a'] \leq \mathbb{E} [\|\mathbf{\Lambda}_{1,a_*,a'_*,h}\|^2].$$

Next, we combine the above equations and obtain that as $n, d \rightarrow \infty$

$$\frac{1}{n^2} \|\mathbb{E}[\mathbf{\Lambda}\mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}(1), \mathbf{Y}'(h)]\|_F^2 \leq \frac{1}{n^2} \mathbb{E} [\|\mathbf{\Lambda}_{1,a_*,a'_*,h}\|^2]^2 + o_P(1).$$

Similarly, if we consider $a \sim \text{Unif}[a_*, 2a_*]$ and $a' \sim \text{Unif}[a'_*, 2a'_*]$, then we have

$$\frac{1}{n^2} \|\mathbb{E}[\mathbf{\Lambda}\mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}(1), \mathbf{Y}'(h)]\|_F^2 \geq \frac{1}{n^2} \mathbb{E} [\|\mathbf{\Lambda}_{1,a_*,a'_*,h}\|^2]^2 + o_P(1).$$

Since $\text{support}(\mathbf{\Lambda}_0) \subseteq [-K_1, K_1]$, by Lebesgue dominated convergence theorem we have

$$\lim_{n,d \rightarrow \infty} \frac{1}{n^2} \mathbb{E} [\|\mathbf{\Lambda}_{1,a_*,a'_*,h}\|^2]^2 = \lim_{n,d \rightarrow \infty} \frac{1}{n^2} \mathbb{E} [\|\mathbb{E}[\mathbf{\Lambda}\mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}(1), \mathbf{Y}'(h)]\|_F^2]. \quad (125)$$

By Lemma D.13, as $n, d \rightarrow \infty$ we have $\bar{\Phi}_n(1, 0, 0, h) \rightarrow \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q)$. Furthermore, the mapping $h \mapsto \bar{\Phi}_n(1, 0, 0, h)$ is convex and differentiable. Therefore, if $q_\Theta^2 + h \in D$, then by Lemma B.4 and Gaussian integration by parts we have

$$\lim_{n,d \rightarrow \infty} \frac{1}{4n^2} \mathbb{E} [\|\mathbb{E}[\mathbf{\Lambda}\mathbf{\Lambda}^\top \mid \bar{\mathbf{A}}(1), \mathbf{Y}'(h)]\|_F^2] = \frac{\partial}{\partial h} \sup_{q \geq 0} \mathcal{F}(q_\Theta^2 + h, q). \quad (126)$$

The proof of the first claim follows immediately from Eqs. (125) and (126).

F.10 Proof of the second claim of Lemma D.17

We let $a \sim \text{Unif}[a_*/2, a_*]$ and $a' \sim \text{Unif}[a'_*/2, a'_*]$, then by Corollary D.1, we have

$$\begin{aligned} \frac{1}{n} \|\mathbf{\Lambda}_{1,a,a',h}\|^2 &= \frac{1}{n} \mathbb{E} [\|\mathbf{\Lambda}_{1,a,a',h}\|^2 \mid a, a'] + \delta_{n,1}, \\ \frac{1}{\sqrt{nd}} \|\bar{\mathbf{\Theta}}_{1,a,a',h}\|^2 &= \frac{1}{\sqrt{nd}} \mathbb{E} [\|\bar{\mathbf{\Theta}}_{1,a,a',h}\|^2 \mid a, a'] + \delta_{n,2}, \end{aligned}$$

where $\mathbb{E}[\delta_{n,1}^2]$ and $\mathbb{E}[\delta_{n,2}^2]$ are random variables that converge to 0 as $n, d \rightarrow \infty$. By Eq. (57) we have

$$\frac{1}{2n\sqrt{nd}} \|\mathbf{M}_{1,0,0,h}\|_F^2 = \frac{1}{2n\sqrt{nd}} \|\mathbf{\Lambda}_{1,a,a',h}\|^2 \|\bar{\mathbf{\Theta}}_{1,a,a',h}\|^2 + \delta_{n,0}.$$

In the above equation, $\delta_{n,0}$ is a random variable satisfying $\mathbb{E}[\delta_{n,0}] \rightarrow 0$ as $n, d \rightarrow \infty$. Therefore, for all $a \in [a_*/2, a_*]$ and $a' \in [a'_*/2, a'_*]$

$$\begin{aligned} &\frac{1}{2n\sqrt{nd}} \|\mathbf{M}_{1,0,0,h}\|_F^2 \\ &\leq \frac{1}{2n\sqrt{nd}} \mathbb{E} [\|\mathbf{\Lambda}_{1,a,a',h}\|^2] \mathbb{E} [\|\bar{\mathbf{\Theta}}_{1,a,a',h}\|^2] + \frac{1}{2} |\delta_{n,1}| |\delta_{n,2}| \end{aligned} \quad (127)$$

$$+ \frac{|\delta_{n,1}|}{2\sqrt{nd}} \mathbb{E} [\|\bar{\Theta}_{1,a,a',h}\|^2] + \frac{|\delta_{n,2}|}{2n} \mathbb{E} [\|\mathbf{\Lambda}_{1,a,a',h}\|^2] + \delta_{n,0}, \quad (128)$$

Notice that

$$\limsup_{n,d \rightarrow \infty} \frac{1}{\sqrt{nd}} \mathbb{E} [\|\bar{\Theta}_{1,a_*,a'_*,h}\|^2] < \infty, \quad \limsup_{n,d \rightarrow \infty} \frac{1}{n} \mathbb{E} [\|\mathbf{\Lambda}_{1,a_*,a'_*,h}\|^2] < \infty. \quad (129)$$

We plug Eq. (129) into Eq. (128) then take the expectation, which implies that as $n, d \rightarrow \infty$ we have

$$\begin{aligned} & \liminf_{n,d \rightarrow \infty} \frac{1}{2n\sqrt{nd}} \mathbb{E} [\|\mathbf{\Lambda}_{1,a_*,a'_*,h}\|^2] \mathbb{E} [\|\bar{\Theta}_{1,a_*,a'_*,h}\|^2] \\ & \geq \lim_{n,d \rightarrow \infty} \frac{1}{2n\sqrt{nd}} \mathbb{E} [\|\mathbf{M}_{1,0,0,h}\|^2] = D_{\Theta}(h). \end{aligned} \quad (130)$$

Similarly, if we let $a \sim \text{Unif}[a_*, 2a_*]$ and $a' \sim \text{Unif}[a'_*, 2a'_*]$, then we can conclude that

$$\begin{aligned} & \limsup_{n,d \rightarrow \infty} \frac{1}{2n\sqrt{nd}} \mathbb{E} [\|\mathbf{\Lambda}_{1,a_*,a'_*,h}\|^2] \mathbb{E} [\|\bar{\Theta}_{1,a_*,a'_*,h}\|^2] \\ & \leq \lim_{n,d \rightarrow \infty} \frac{1}{2n\sqrt{nd}} \mathbb{E} [\|\mathbf{M}_{1,0,0,h}\|^2] = D_{\Theta}(h). \end{aligned} \quad (131)$$

Notice that $D_{\Theta}(h) = 2q_{\Theta}^2 \frac{\partial}{\partial h} \sup_{q \geq 0} \mathcal{F}(q_{\Theta}^2 + h, q)$, then the proof of the second claim follows from Eq. (130), Eq. (131) and the first claim.

Appendix G Proofs for the Gaussian mixture clustering example

G.1 Proof of Proposition 5.1 claim (a)

We define the pairwise overlap achieved by estimator $\hat{\mathbf{\Lambda}}$ as

$$\text{PairOverlap}_n := \frac{2}{n^2} \sum_{i < j} \mathbb{1} \left\{ \mathbb{1}\{\mathbf{\Lambda}_i = \mathbf{\Lambda}_j\} = \mathbb{1}\{\hat{\mathbf{\Lambda}}_i = \hat{\mathbf{\Lambda}}_j\} \right\}.$$

We notice that

$$\begin{aligned} \text{PairOverlap}_n &= \frac{2}{n^2} \sum_{i < j} \left(\mathbb{1}\{\mathbf{\Lambda}_i = \hat{\mathbf{\Lambda}}_i\} \mathbb{1}\{\mathbf{\Lambda}_j = \hat{\mathbf{\Lambda}}_j\} + (1 - \mathbb{1}\{\mathbf{\Lambda}_i = \hat{\mathbf{\Lambda}}_i\})(1 - \mathbb{1}\{\mathbf{\Lambda}_j = \hat{\mathbf{\Lambda}}_j\}) \right) \\ &= 2\text{Overlap}_n^2 + 1 - 2\text{Overlap}_n + o_n(1). \end{aligned} \quad (132)$$

According to [LM19, Section 2.3], under the symmetric model (10), if $q_{\Theta} \leq 1$, then we have $\lim_{n \rightarrow \infty} \text{MMSE}_n^{\text{symm}}(\mu_{\Lambda}; q_{\Theta}) = 1$. This is also the mean square error achieved by the constant estimator $\mathbf{0}_{n \times n}$. For an estimate of the labels $\hat{\mathbf{\Lambda}} \in \{-1, +1\}^n$ and $a \in (0, 1)$, we define the rescaled vector $\hat{\mathbf{\Lambda}}_a := \sqrt{a} \hat{\mathbf{\Lambda}} \in \{-\sqrt{a}, +\sqrt{a}\}^n$. Then by Theorem 4.4, we have

$$\begin{aligned} \frac{1}{n^2} \mathbb{E} \left[\left\| \mathbf{\Lambda} \mathbf{\Lambda}^{\top} - \hat{\mathbf{\Lambda}}_a \hat{\mathbf{\Lambda}}_a^{\top} \right\|_F^2 \right] &= \mathbb{E} \left[(1-a)^2 (\text{PairOverlap}_n + n^{-1}) + (1+a)^2 (1 - n^{-1} - \text{PairOverlap}_n) \right] \\ &\geq \text{MMSE}_n^{\text{asym}}(\mu_{\Lambda}, \mu_{\Theta}). \end{aligned}$$

Theorem 4.4 implies that $\lim_{n,d \rightarrow \infty} \text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta) = 1$, thus

$$\limsup_{n,d \rightarrow \infty} \mathbb{E} [\text{PairOverlap}_n] \leq \frac{a^2 + 2a}{4a},$$

which holds for every $a \in (0, 1)$. Let $a \rightarrow 0^+$, we then conclude that $\limsup_{n,d \rightarrow \infty} \mathbb{E} [\text{PairOverlap}_n] \leq 1/2$. Next, we plug this result into Eq. (132) then apply dominated convergence theorem, which gives $\text{Overlap}_n \xrightarrow{P} 1/2$. In summary, partial recovery of component identity is impossible in the current setting.

G.2 Proof of Theorem 5.1 part (a)

Let $\hat{\mathbf{A}} \in \mathbb{R}^{n \times k}$ be any estimator of the cluster assignments constructed based on data matrix \mathbf{A} . For $a > 0$, we define $\hat{\mathbf{A}}_a := \sqrt{a} \hat{\mathbf{A}}$. Under the current conditions, as $n, d \rightarrow \infty$ we have

$$\begin{aligned} & \text{MMSE}_n^{\text{asym}}(\mu_\Lambda, \mu_\Theta) \\ &= k^{-2}(k-1) + o_n(1) \\ &\leq \frac{1}{n^2} \mathbb{E} \left[\left\| \mathbf{A} \mathbf{A}^\top - k^{-1} \mathbf{1}_{n \times n} - \hat{\mathbf{A}}_a \hat{\mathbf{A}}_a^\top \right\|_F^2 \right] \\ &= k^{-2}(k-1) + \frac{a^2}{n^2} \sum_{i,j \in [n]} \mathbb{E} \left[\mathbb{1}\{\hat{\mathbf{A}}_i = \hat{\mathbf{A}}_j\} \right] - \frac{2a}{n^2} \sum_{i,j \in [n]} \mathbb{E} \left[\mathbb{1}\{\hat{\mathbf{A}}_i = \hat{\mathbf{A}}_j\} (\mathbb{1}\{\mathbf{A}_i = \mathbf{A}_j\} - k^{-1}) \right] + o_n(1) \\ &\leq k^{-2}(k-1) + a^2 - \frac{2a}{n^2} \sum_{i,j \in [n]} \mathbb{E} \left[\mathbb{1}\{\hat{\mathbf{A}}_i = \hat{\mathbf{A}}_j\} (\mathbb{1}\{\mathbf{A}_i = \mathbf{A}_j\} - k^{-1}) \right] + o_n(1). \end{aligned}$$

Using the above equation we can conclude that

$$\limsup_{n,d \rightarrow \infty} \frac{1}{n^2} \sum_{i,j \in [n]} \mathbb{E} \left[\mathbb{1}\{\hat{\mathbf{A}}_i = \hat{\mathbf{A}}_j\} (\mathbb{1}\{\mathbf{A}_i = \mathbf{A}_j\} - k^{-1}) \right] \leq \frac{a}{2}.$$

Since $a > 0$ is arbitrary, we then have

$$\limsup_{n,d \rightarrow \infty} \frac{1}{n^2} \sum_{i,j \in [n]} \mathbb{E} \left[\mathbb{1}\{\hat{\mathbf{A}}_i = \hat{\mathbf{A}}_j\} (\mathbb{1}\{\mathbf{A}_i = \mathbf{A}_j\} - k^{-1}) \right] \leq 0. \quad (133)$$

For $s, r \in [k]$, we define

$$C_{sr} := \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{\hat{\mathbf{A}}_i = \mathbf{e}_s, \mathbf{A}_i = \mathbf{e}_r\}. \quad (134)$$

We immediately see that $C_{sr} \geq 0$ and $\sum_{s \in [k]} \sum_{r \in [k]} C_{sr} = 1$. Furthermore, notice that

$$\begin{aligned} \frac{1}{n^2} \sum_{i,j \in [n]} \mathbb{1}\{\hat{\mathbf{A}}_i = \hat{\mathbf{A}}_j, \mathbf{A}_i = \mathbf{A}_j\} &= \frac{1}{n^2} \sum_{i,j \in [n]} \sum_{s,r \in [k]} \mathbb{1}\{\hat{\mathbf{A}}_i = \hat{\mathbf{A}}_j = \mathbf{e}_s, \mathbf{A}_i = \mathbf{A}_j = \mathbf{e}_r\} \\ &= \sum_{s,r \in [k]} \left(\frac{1}{n} \sum_{i \in [n]} \mathbb{1}\{\hat{\mathbf{A}}_i = \mathbf{e}_s, \mathbf{A}_i = \mathbf{e}_r\} \right)^2 \\ &= \sum_{s,r \in [k]} C_{sr}^2, \end{aligned} \quad (135)$$

$$\begin{aligned}
\frac{1}{kn^2} \sum_{i,j \in [n]} \mathbb{1}\{\hat{\Lambda}_i = \hat{\Lambda}_j\} &= \frac{1}{kn^2} \sum_{i,j \in [n]} \sum_{s \in [k]} \mathbb{1}\{\hat{\Lambda}_i = \mathbf{e}_s\} \mathbb{1}\{\hat{\Lambda}_j = \mathbf{e}_s\} \\
&= \frac{1}{k} \sum_{s \in [k]} \left(\frac{1}{n} \sum_{i \in [n]} \sum_{r \in [k]} \mathbb{1}\{\hat{\Lambda}_i = \mathbf{e}_s, \Lambda_i = \mathbf{e}_r\} \right)^2 \\
&= \frac{1}{r} \sum_{s \in [k]} (C_{s1} + \dots + C_{sk})^2.
\end{aligned} \tag{136}$$

Next, we subtract Eq. (135) by Eq. (136) and apply Eq. (133), which gives

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\sum_{s \in [k]} \sum_{1 \leq r_1 < r_2 \leq k} (C_{sr_1} - C_{sr_2})^2 \right] = 0.$$

Note that for all $s \in [k]$, there exists $r_s \in [k]$, such that $[k] = \{r_s : s \in [k]\}$ and $\text{Overlap}_n = \sum_{s \in [k]} C_{sr_s}$. Combining the above results, we conclude that $\text{Overlap}_n = k^{-1} + o_P(1)$, thus completing the proof of part (a).

G.3 Proof of Theorem 5.1 part (b)

Suppose the statement is not true, then for any $\hat{\Lambda}$ and any subsequence of \mathbb{N}_+ , there further exists a subsequence $\{n_i\}_{i \in \mathbb{N}_+} \subseteq \mathbb{N}_+$ of the previous subsequence, such that $n_i < n_{i+1}$ and $\lim_{i \rightarrow \infty} \mathbb{E}[\text{Overlap}_{n_i}] = k^{-1}$. Therefore, $\text{Overlap}_{n_i} \xrightarrow{P} k^{-1}$. In the following parts of the proof, we will restrict to this subsequence $\{n_i\}_{i \in \mathbb{N}_+}$.

We assume $\hat{\Lambda} \in \mathbb{R}^{n \times k}$ such that $\hat{\Lambda} \stackrel{d}{=} \mu(\Lambda = \cdot \mid \mathbf{A})$. Recall that for $s, r \in [k]$, C_{sr} is defined in Eq. (134). Furthermore, notice that $\hat{\Lambda} \stackrel{d}{=} \Lambda$, then by the law of large numbers, for all $s, r \in [k]$ we have

$$C_{s1} + C_{s2} + \dots + C_{sk} \xrightarrow{P} \frac{1}{k}, \quad C_{1r} + C_{2r} + \dots + C_{kr} \xrightarrow{P} \frac{1}{k}. \tag{137}$$

For $\delta > 0$, if $C_{11} > k^{-2} + \delta$, then by Eq. (137) $\sum_{2 \leq s, r \leq k} C_{sr} > (1 - k^{-1})^2 + \delta + o_P(1)$. As a result, we conclude that there exists a permutation π of $\{2, 3, \dots, k\}$, such that $C_{2\pi(2)} + C_{3\pi(3)} + \dots + C_{k\pi(k)} \geq k^{-2}(k-1) + (r-1)^{-1}\delta + o_P(1)$. Therefore, $\text{Overlap}_n \geq C_{11} + C_{2\pi(2)} + C_{3\pi(3)} + \dots + C_{k\pi(k)} > k^{-1} + k(k-1)^{-1}\delta + o_P(1)$. For $s, r \in [k]$, we define the set $S_{sr}^\delta := \{C_{sr} > k^{-2} + \delta\}$. Since $\text{Overlap}_{n_i} \xrightarrow{P} k^{-1}$, using the above analysis we derive that $\lim_{i \rightarrow \infty} \mathbb{P}(S_{11}^\delta) = 0$. Indeed, we can repeat such analysis for all $s, r \in [k]$ and conclude that $\lim_{i \rightarrow \infty} \mathbb{P}(S_{sr}^\delta) = 0$, thus $C_{sr} \leq k^{-2} + \delta + o_P(1)$ along the subsequence $\{n_i\}_{i \in \mathbb{N}_+}$. Since $\sum_{s, r \in [k]} C_{sr} = 1$ and δ is arbitrary, we deduce that $C_{sr} \xrightarrow{P} k^{-2}$ along the subsequence $\{n_i\}_{i \in \mathbb{N}_+}$. This further implies that $C_{sr} \xrightarrow{P} k^{-2}$.

However, according to [BMV⁺18, Theorem 2] and Theorem 4.5, we see that under the conditions of this part

$$\liminf_{n, d \rightarrow \infty} \frac{1}{n^2} \sum_{i, j \in [n]} \mathbb{E} \left[\mathbb{1}\{\hat{\Lambda}_i = \hat{\Lambda}_j\} \mathbb{1}\{\Lambda_i = \Lambda_j\} \right] = \liminf_{n \rightarrow \infty} \frac{1}{n^2} \sum_{i, j \in [n]} \mathbb{E} [\mathbb{E}[\mathbb{1}\{\Lambda_i = \Lambda_j\} \mid \mathbf{A}]^2] > k^{-2}.$$

Finally, we plug Eq. (135) into the formula above, which leads to $\liminf_{n, d \rightarrow \infty} \mathbb{E}[\sum_{s, k \in [r]} C_{sk}^2] > k^{-2}$.

This is in contradiction with the previously established claim that $C_{sr} \xrightarrow{P} k^{-2}$ for all $s, r \in [k]$, thus completing the proof of part (b).