Proceedings of Machine Learning Research vol 291:1–67, 2025 38th Annual Conference on Learning Theory

Fundamental Limits of Matrix Sensing: Exact Asymptotics, Universality, and Applications

Yizhou Xu YIZHOU.XU@EPFL.CH
Statistical Physics of Computation Laboratory (SPOC), EPFL, Switzerland

Antoine Maillard ANTOINE.MAILLARD@INRIA.FR
Inria Paris, DI ENS, PSL University, Paris, France

Lenka Zdeborová LENKA.ZDEBOROVA@EPFL.CH
Statistical Physics of Computation Laboratory (SPOC), EPFL, Switzerland

Florent Krzakala FLORENT.KRZAKALA@EPFL.CH
Information, Learning, and Physics Laboratory (IDEPHICS), EPFL, Switzerland

Editors: Nika Haghtalab and Ankur Moitra

Abstract

In the matrix sensing problem, one wishes to reconstruct a matrix from (possibly noisy) observations of its linear projections along given directions. We consider this model in the high-dimensional limit: while previous works on this model primarily focused on the recovery of low-rank matrices, we consider in this work more general classes of structured signal matrices with potentially large rank, e.g. a product of two matrices of sizes proportional to the dimension. We provide rigorous asymptotic equations characterizing the Bayes-optimal learning performance from a number of samples which is proportional to the number of entries in the matrix. Our proof is composed of three key ingredients: (i) we prove universality properties to handle structured sensing matrices, related to the "Gaussian equivalence" phenomenon in statistical learning, (ii) we provide a sharp characterization of Bayes-optimal learning in generalized linear models with Gaussian data and structured matrix priors, generalizing previously studied settings, and (iii) we leverage previous works on the problem of matrix denoising. The generality of our results allow for a variety of applications: notably, we mathematically establish predictions obtained via non-rigorous methods from statistical physics in Erba et al. (2024) regarding Bilinear Sequence Regression, a benchmark model for learning from sequences of tokens, and in Maillard et al. (2024) on Bayes-optimal learning in neural networks with quadratic activation function, and width proportional to the dimension.
Keywords: matrix sensing, generalized linear model, universality, sharp asymptotics

1. Introduction

1.1. Setting of the problem

Consider the problem of learning a matrix $S^\star \in \mathbb{R}^{d\times L}$ from noisy and possibly non-linear observations, $\mu = 1, \dots, n$, generated from the following model with $\Phi_\mu \in \mathbb{R}^{L\times d}$:

$$Y_\mu = \varphi(\mathrm{Tr}[\Phi_\mu S^\star], a_\mu) + \sqrt{\Delta} Z_\mu, \tag{1}$$

where $\Delta > 0$, and $\{a_\mu\}_{\mu=1}^n \overset{\text{i.i.d.}}{\sim} P_A$ and $\{Z_\mu\}_{\mu=1}^n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ represent the noise in the observations. We assume that the data samples Φ_μ are drawn from a known probability distribution $\{\Phi_\mu\}_{\mu=1}^n \overset{\text{i.i.d.}}{\sim} P_\Phi$. The signal $S^\star$ is drawn from a rotationally-invariant prior $S^\star \sim P_0$. This setting encompasses cases of particular interest such as Wishart matrices, and products of random matrices. We will consider a quite general data distribution P_Φ (as specified later, see Assumptions 2.1 and

2.4), with two notable motivating examples: Φ as an i.i.d. matrix with standard Gaussian elements, and $\Phi = (xx^\top - \mathrm{I}_d)/\sqrt{d}$, with $x \sim \mathcal{N}(0, \mathrm{I}_d)$, two models which connect to settings studied non-rigorously in previous works Maillard et al. (2024); Erba et al. (2024). Notice that we may also write the output channel of eq. (1) in the equivalent form:

$$Y_\mu \sim P_{\mathrm{out}}(\cdot|\mathrm{Tr}[\Phi_\mu S^\star]), \text{ with } P_{\mathrm{out}}(y|z) \coloneqq \frac{1}{\sqrt{2\pi\Delta}} \int P_A(\mathrm{d}a) \exp\left\{-\frac{1}{2\Delta}(y - \varphi(z,a))^2\right\}. \tag{2}$$

In this paper, we consider the Bayes-optimal, or information theoretic, scenario. The observed dataset $\mathcal{D} = \{Y_\mu, \Phi_\mu\}_{\mu=1}^n$ is given to the statistician who is provided all the information on how the data has been generated, i.e. the value of Δ, the form of the function $\varphi(\cdot,\cdot)$, and the prior distributions P_0, P_A, from which she must learn and provide the best estimates of the values of the weights $S^\star$ (or of the prediction error on the output for a test input sample), by averaging over the posterior distribution. The key interest of such an information-theoretic analysis is to provide a sharp characterization of the best reconstruction performance that *any* algorithm can achieve.

We consider a high-dimensional setting where $L, d \to \infty$, with fixed ratio $d/L \to \beta > 0$. Without loss of generality, we suppose that $L \geq d$ (and thus $\beta \leq 1$): if $\beta > 1$ one can consider instead $\mathrm{Tr}[\Phi_\mu^T (S^\star)^T] = \mathrm{Tr}[\Phi_\mu S^\star]$. While our results apply to more generic prior distributions P_0 of the signal $S^\star$, we will especially focus on the case where $S^\star$ is a product of random matrices of size $d \times m$ and $m \times L$, in which case the width m is such that $m/d \to \kappa > 0$, and quantifies the amount of "structure" in the signal. The number of samples n also satisfies $n \to \infty$, with a fixed ratio $n/(Ld) \to \alpha > 0$: this ensures that the number of samples scales linearly with the number of unknown parameters, and that one can perform non-trivial estimation of $S^\star$.

Our aim is to characterize the Bayes-optimal performance in the stated limit. Our analysis leverages the fact that in many quantities of interest, such as the prediction and estimation error per sample, concentrate as $d \to \infty$ on deterministic values. We derive low-dimensional equations from which these values of interest can be readily extracted.

A symmetric variant – While all our analysis holds for the model of eq. (1), we shall also consider in what follows a "symmetrized" variant of eq. (1), where $L = d$, and both $S^\star$ and Φ_μ are symmetric. We will start the discussion of our results in this symmetric variant and then, in Section 2.4, we come back to the model of eq. (1), and state our main results for it, as well as sketch the straightforward generalization of the proofs in the symmetric model to the non-symmetric case.

1.2. Motivations and related work

The problem we consider connects to multiple aspects of the existing literature; for instance, eq. (1) generalizes classical linear and generalized linear models. In the statistics and computer science literature, this model is often referred to as matrix sensing Recht et al. (2010); Fazel (2002); Candes and Recht (2012); Bhojanapalli et al. (2016); Parker and Schniter (2016); Gunasekar et al. (2017); Li et al. (2018); Romanov and Gavish (2018). Signal recovery using minimum nuclear norm algorithms has been widely studied Recht et al. (2010); Candes and Recht (2012), as well as recovery via gradient descent Bhojanapalli et al. (2016), particularly in relation to overparametrization and the implicit regularization of gradient descent Gunasekar et al. (2017); Li et al. (2018).

Comparing the bulk of literature on the problem to our setting and approach, we trade stronger distributional assumptions for sharper results. While we assume randomness in both the data and the signal, as stated in our setup, we derive results that explicitly include the leading-order constants in the high-dimensional limit. In contrast, most prior works impose much weaker distributional assumptions but, in exchange, can only derive results up to unspecified constants or, more often, up to logarithmic factors in dimensionality. Another key distinction between our work and much of the literature is the scaling of the width/rank parameter m: while most studies consider the regime where $m \ll d$, we instead analyze the case where m and d are proportional.

Only a few works study sharp optimal constants in the high-dimensional limit for the model of eq. (1). Some focus exclusively on the very low-rank case Schülke et al. (2016); Parker and Schniter (2016), while others analyze the nuclear norm minimization algorithm without characterizing the Bayes-optimal performance Donoho et al. (2013). To the best of our knowledge, the only two works that consider special cases of our setting, and characterize sharply the Bayes-optimal performance in the high-dimensional limit, are Maillard et al. (2024) and Erba et al. (2024). However, both of these works rely on analytical non-rigorous methods from statistical physics, and do not provide formal proofs of their results. The main motivation of our work is to fill this gap by establishing mathematically these predictions, while extending them to a more general setup. We note that, although the authors of Maillard et al. (2024) propose a tentative sketch of a rigorous analysis, we take here an approach different from the one proposed in this paper, which suggested using the high-dimensional analysis of an Approximate Message Passing algorithm, while we rather rely on an adaptive interpolation method Barbier and Macris (2019).

The applications studied in Maillard et al. (2024) and Erba et al. (2024) serve as our primary motivation, and we will describe them in more detail in Section 2.5. The authors of Maillard et al. (2024) exploit a known relationship between the matrix sensing problem and a two-layer neural network with quadratic activation to establish sharp results on the optimal prediction error of such networks. Meanwhile, Erba et al. (2024) introduce eq. (1) as the simplest regression model for inputs that are either sequences of length L or d-dimensional tokens. This formulation is inspired by the growing interest in sequence models such as transformers Vaswani et al. (2017). Our findings also extend to matrix compressed sensing Schülke et al. (2016); Parker and Schniter (2016), which typically assumes the signal is a product of two matrices, or a low-rank matrix without any prior distribution.

Methodologically, our proof techniques build on the analysis of high-dimensional generalized linear models from Barbier et al. (2019). Specifically, our model can be seen as a generalization of this model, where the signal is a high-rank rotationally invariant matrix, a non-separable prior far from the i.i.d. signals considered in Barbier et al. (2019). To handle this complex signal structure, we leverage recent advances in the analysis of matrix denoising Maillard et al. (2022); Troiani et al. (2022); Pourkamali et al. (2024); Pourkamali and Macris (2024). Additionally, we build on fundamental Gaussian universality results Goldt et al. (2022); Hu and Lu (2022); Montanari and Saeed (2022); Maillard and Bandeira (2023), enabling us to tackle more general distributions P_Φ of Φ_μ.

1.3. Main contributions

Our contributions are twofold. First, we prove a low-dimensional formula for both the mutual information and the Minimal Mean-Square Error (MMSE) in matrix sensing problems with rotationally-invariant signals, in the high-dimensional limit. Our proof integrates recent advances in Gaussian universality Montanari and Saeed (2022); Maillard and Bandeira (2023), adaptive interpolation Barbier and Macris (2019); Barbier et al. (2019), and matrix denoising Troiani et al. (2022); Pourkamali et al. (2024); Pourkamali and Macris (2024).

Second, we demonstrate how these formulas extend beyond standard signal processing applications Gross et al. (2010); Candes and Plan (2011). In particular, we discuss in Section 2.5 how our results allow us to prove heuristic results relevant to machine learning. Specifically, we address: (a) an open problem posed by Cui et al. (2023) on the learnability of "neural network-like" functions with d^2 parameters from $n = O(d^2)$ samples (at least for a subclass of functions) (b) the Bayes-optimal formula for two-layer neural networks with quadratic activation Maillard et al. (2024); and (c) the information-theoretic limit of bilinear sequence regression problems Erba et al. (2024). Given its generality, we anticipate that our approach will find applications in a variety of other domains.

1.4. Notations

We denote $[d] \coloneqq \{1, \cdots, d\}$ the set of integers from 1 to d, and $\mathcal{S}_d$ the set of $d \times d$ real symmetric matrices. For a function $V : \mathbb{R} \to \mathbb{R}$, and $S \in \mathcal{S}_d$ with eigenvalues $(\lambda_i)_{i=1}^d$, we define $V(S)$ as the matrix with the same eigenvectors as S, and eigenvalues $(V(\lambda_i))_{i=1}^d$. For $S \in \mathcal{S}_d$, we denote $\mu_S \coloneqq (1/d)\sum_{i=1}^d \delta_{\lambda_i}$ the empirical eigenvalue distribution of S. We denote $\mathrm{Tr} S \coloneqq \sum_{i=1}^d \lambda_i$ the trace of S and $\mathrm{tr} S \coloneqq \frac{1}{d}\mathrm{Tr} S$ the normalized trace. $\|\cdot\|$ denotes the ℓ_2 norm of vectors and the Frobenius norm of matrices and tensors. $B_{\mathrm{op}}(M)$ denotes the set of matrices with eigenvalues (or singular values for rectangular matrices) in $[-M, M]$. We refer to Appendix A.1 for some classical definition of rotationally invariant ensembles, specifically the *Gaussian Orthogonal Ensemble* $\mathrm{GOE}(d)$ and the *Wishart Ensemble* $\mathcal{W}_{m,d}$. For X, Y two random variables, we say $X \overset{\mathrm{d}}{=} Y$ if X, Y share the same distribution. $\overset{\mathrm{p}}{\to}$ denotes limit in probability. Finally, we generically denote constants as $C > 0$, whose value may vary from line to line.

2. Main results and applications

2.1. Assumptions

We detail the different assumptions made on the model of eq. (1) in order for our main results to apply. In subsections 2.1–2.3 we consider the *symmetric* variant of the model, where $L = d$, and $S^\star$, Φ_μ are symmetric $d \times d$ matrices, and we assume that $n/d^2 \to \alpha > 0$.

Assumption 2.1 (Assumptions on P_0) *We consider rotationally-invariant priors P_0, more precisely either one of the following two cases holds:*

(i) *P_0 has density:*

$$P_0(S) \propto \exp(-d\,\mathrm{Tr}[V(S)]),$$

for a real and $\mathcal{C}^2$ potential $V : \mathcal{B} \to \mathbb{R}$, where $\mathcal{B} \subseteq \mathbb{R}$ is an interval. We assume that $V''(s) \geq 1/c$ for all $s \in \mathcal{B}$ and some constant $c > 0$.

(ii) *Let $\kappa \in (0, 1]$, and denote $m \coloneqq \lceil \kappa d \rceil$. We assume that P_0 is supported on the space $\mathcal{S}_{d,m}^+$ of positive-semidefinite matrices $S \in \mathcal{S}_d$ with rank m and density:*

$$P_0(S) \propto \exp(-d\,\mathrm{Tr}[V(\Lambda)]),$$

where $\Lambda = \mathrm{Diag}(\{\lambda_i\}_{i=1}^m)$ denotes the non-zero eigenvalues of S. Here $V : \mathcal{B} \to \mathbb{R}$ is a a real and $\mathcal{C}^2$ potential, and $\mathcal{B} \subseteq \mathbb{R}^+$ is an interval. We assume $V''(s) + (1-\kappa)/s^2 \geq 1/c$ for all $s \in \mathcal{B}$ and a constant $c > 0$.

Notice that a mathematically complete description of Assumption 2.1(ii) would require defining the volume measure on $\mathcal{S}_{d,m}^+$: we refer to Uhlig (1994) (Theorem 2) for a precise description of this point. Assumption 2.1 encompasses many natural random matrix distributions, including notably $\mathrm{GOE}(d)$ matrices. It also applies to a large class of Wishart matrices $\mathcal{W}_{m,d}$[1], once one slightly

1. When $m/d \to \kappa \in (0, \infty)\backslash\{1\}$. The case $\kappa = 1$ is not covered by our proof at the moment, due to a technical difficulty in obtaining an equivalent to the Gaussian Poincaré inequality in this setting, see Remark 11.

weakens the assumption on V to $V''(s) \geq 1/c$ only inside a large bounded interval, an assumption under which our proofs generalize directly: we discuss such relaxations of Assumption 2.1, as well as consequences of this assumption, in Appendix A. In particular, by Proposition 13, under Assumption 2.1, the empirical eigenvalue distribution μ_S almost surely weakly converges to a distribution μ_0 with compact support.

Assumptions on P_Φ – The main assumption we make on the data distribution P_Φ is referred to as a *uniform one-dimensional Central Limit Theorem (CLT)*.

Assumption 2.2 (Uniform one-dimensional CLT) *For any $M > 0$, the following holds. For $\Phi \sim P_\Phi$, $G \sim \mathrm{GOE}(d)$, we have $\mathbb{E}[\Phi] = \mathbb{E}[G] = 0$, $\mathbb{E}[\Phi_{ij}\Phi_{kl}] = \mathbb{E}[G_{ij}G_{kl}] = \frac{1}{d}\delta_{ik}\delta_{jl}(1+\delta_{ijkl})$, and*

$$\lim_{d\to\infty} \sup_{S\in B_{\mathrm{op}}(M)} |\mathbb{E}_\Phi[\psi(\mathrm{Tr}[\Phi S])] - \mathbb{E}_G[\psi(\mathrm{Tr}[GS])]| = 0,$$

for any bounded Lipschitz function ψ.

Assumption 2.2 encompasses different distributions P_Φ: two examples used here are matrices with i.i.d. entries (under some bounded moments conditions), as well as centered rank-one matrices $\Phi_\mu = (x_\mu x_\mu^\top - \mathrm{I}_d)/\sqrt{d}$ with $x_\mu \sim \mathcal{N}(0, \mathrm{I}_d)$ for which Assumption 2.2 is proven in (Maillard and Bandeira, 2023, Lemma 4.8). These are related to applications of our results discussed in Section 2.5. We leave a more exhaustive analysis of the distributions satisfying Assumption 2.2 (akin to (Montanari and Saeed, 2022, Section 3) for vector distributions) to future work.

We note that it is actually sufficient to assume Assumption 2.2 for some $M > 0$ large enough, depending only on the choice of P_0, given by Proposition 13-(ii) in Appendix A. We will show that under Assumption 2.2, one can effectively study an equivalent model to eq. (1), with $\{\Phi_\mu\}_{\mu=1}^n$ replaced by $\{G_\mu\}_{\mu=1}^n \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$, following insights of a long line of work Mei and Montanari (2022); Gerace et al. (2020); Goldt et al. (2022); Hu and Lu (2022); Montanari and Saeed (2022); Maillard and Bandeira (2023); Dandi et al. (2024b).

Activation function – Finally, we make the following generic assumption, akin to the one of Barbier et al. (2019), on the activation φ in eq. (1).

Assumption 2.3 (Activation function) *$\varphi(\cdot, a)$ is continuous almost everywhere, almost surely over $a \sim P_A$. Moreover, for $\Phi \sim P_\Phi$, $S^\star \sim P_0$ and $a \sim P_A$, there exists $\gamma > 0$ such that the sequence $\{\mathbb{E}[|\varphi(\mathrm{Tr}[\Phi S^\star], a)|^{2+\gamma}]\}_{d\geq 1}$ is bounded.*

2.2. Limit of the mutual information

In this section, we shall sometimes use a nomenclature originating from *information theory* and *statistical physics* (we refer to Zdeborová and Krzakala (2016) for more details and other applications of this point of view to problems of learning and inference).

Recall that we consider the model of eq. (1), which is stated in an equivalent form in eq. (2). We first define the *partition function* $\mathcal{Z}(Y, \Phi)$, and the expected value of the log-partition function, called the *free entropy* f_d:

$$\mathcal{Z}(Y,\Phi) \coloneqq \mathbb{E}_{S\sim P_0} \prod_{\mu=1}^n P_{\mathrm{out}}(Y_\mu | \mathrm{Tr}[\Phi_\mu S]), \quad \text{and} \quad f_d \coloneqq \frac{1}{d^2}\mathbb{E}_{Y,\Phi} \log \mathcal{Z}(Y,\Phi). \tag{3}$$

Notice that the free entropy is directly related to the *mutual information* $I(S^\star, Y|\Phi)$ of information theory by:

$$\lim_{d\to\infty} \frac{1}{d^2} I(S^\star, Y|\Phi) = -\lim_{d\to\infty} f_d + \alpha \mathbb{E}_V \int_{\mathbb{R}} \mathrm{d}\widetilde{Y} P_{\mathrm{out}}(\widetilde{Y}|\sqrt{2\rho}V) \log P_{\mathrm{out}}(\widetilde{Y}|\sqrt{2\rho}V), \tag{4}$$

where $V \sim \mathcal{N}(0,1)$ and $\rho := \lim_{d\to\infty}(1/d)\mathbb{E}_{S\sim P_0}\mathrm{Tr}[S^2]$ (see more on the definition of ρ in Appendix A.2, eq. (42)). We refer to (Barbier et al., 2019, Corollary 1) for a simple proof of eq. (4), plugging in our CLT (Lemma 46). Notice that the second term in the right-hand side of eq. (4) is a simple function of the one-dimensional noise channel P_{out}.

Let us now introduce the main quantities which will characterize the limit of the mutual information, or free entropy. We define the *replica-symmetric free entropy* as:

$$f_{\mathrm{RS}}(q,r) := \psi_{P_0}(r) + \alpha \Psi_{\mathrm{out}}(q) + \frac{r(\rho - q) + 1}{4}, \tag{5}$$

where $q \in [0,\rho]$, $r \geq 0$, and

$$\begin{cases} \psi_{P_0}(r) & := -\frac{1}{2}\Sigma(\mu_{1/r}) - \frac{1}{4}\log r - \frac{3}{8}, \\ \Psi_{\mathrm{out}}(q) & := \mathbb{E}_{V,W,\widetilde{Y}} \log \int \mathcal{D}w P_{\mathrm{out}}(\widetilde{Y}|\sqrt{2q}V + w\sqrt{2(\rho-q)}). \end{cases} \tag{6}$$

In eq. (6), we used the notation $\widetilde{Y} \sim P_{\mathrm{out}}(\cdot|\sqrt{2q}V + \sqrt{2(\rho-q)}W)$, $W, V \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $\mathcal{D}w := (e^{-w^2/2}/\sqrt{2\pi})\mathrm{d}w$ denotes the standard Gaussian measure. Moreover, $\Sigma(\mu) := \int \mu(\mathrm{d}x)\mu(\mathrm{d}y)\log|x - y|$, and $\mu_t := \mu_0 \boxplus \mu_{s.c.,\sqrt{t}}$ for $t \geq 0$ is the free convolution of μ and a semicircular distribution of variance t: details on these classical definitions are given in Appendix A.1.

Our first main theorem is a proof that the replica-symmetric free entropy f_{RS} is the limit of f_d (or of the mutual information) as $d \to \infty$.

Theorem 1 (Limit of the free entropy and mutual information) *Under Assumptions 2.1, 2.2 and 2.3, we have (recall eq. (4)):*

$$\lim_{d\to\infty} f_d = \sup_{q\in[0,\rho]} \inf_{r\geq 0} f_{\mathrm{RS}}(q,r); \qquad \lim_{d\to\infty} \frac{1}{d^2} I(S^\star, Y|\Phi) = -\sup_{q\in[0,\rho]} \inf_{r\geq 0} f_{\mathrm{RS}}(q,r) + \alpha \Psi_{\mathrm{out}}(\rho). \tag{7}$$

We now investigate the consequences of this result: the strategy of the proof of Theorem 1 is laid out in details in Section 3.

Remark 2 *(Maillard et al., 2024, Theorem 4.3) shows that*

$$\psi_{P_0}(r) = \lim_{d\to\infty} \frac{1}{d^2} \mathbb{E}_{Y'} \log \mathbb{E}_{S\sim P_0}\left[e^{-\frac{d}{4}\mathrm{Tr}[(Y' - \sqrt{r}S)^2]}\right], \tag{8}$$

with $Y' := \sqrt{r}S^\star + Z'$, $S^\star \sim P_0$, and $Z' \sim \mathrm{GOE}(d)$. Informally, $\psi_{P_0}(r)$ corresponds to the free entropy of the problem of denoising *the matrix $S^\star$ from the observation Y'. In this regard, the replica-symmetric free entropy eq. (5) is the natural generalization of the one defined in Barbier et al. (2019), which considers vector signals with i.i.d. prior, and for which the function ψ_{P_0} is the free entropy of the (scalar) problem of denoising from the prior. Similar generalizations for other types of structured prior (e.g. generative) have been developed in the literature Gabrié et al. (2019); Aubin et al. (2019, 2020).*

2.3. Limit of the MMSE

In this section, we state our main result, which sharply characterizes the Bayes-optimal mean squared error of the problem of eqs. (1), (2) in the high-dimensional limit.

Theorem 3 (Limit of the MMSE) *Suppose that assumptions 2.1, 2.2 and 2.3 hold, and that $P_{\rm out}$ is informative*[2]*. Let*

$$\Gamma \coloneqq \left\{q \in [0,\rho], r \geq 0 \,|\, q = \rho + 4\psi'_{P_0}(r), r = 4\alpha\Psi'_{\rm out}(q)\right\}$$

be the set of critical points of the functional $f_{\rm RS}(q,r)$ of eq. (5), and

$$D^\star \coloneqq \{\alpha > 0 \,|\, \inf_{r\geq 0} f_{\rm RS}(q,r) \text{ has a unique maximizer } q^\star(\alpha) \in [0,\rho]\}.$$

Then $D^\star$ is equal to $(0,\infty)$ minus some countable set, and for any $\alpha \in D^\star$:

$$\mathrm{MMSE}_d \coloneqq \frac{1}{d^2}\mathbb{E}\left[||S^\star \otimes S^\star - \mathbb{E}[S\otimes S|Y,\Phi]||^2\right] \xrightarrow{d\to\infty} \rho^2 - q^\star(\alpha)^2, \tag{9}$$

where $S \otimes S$ denotes the tensor product, and $\|\cdot\|$ the Frobenius norm.

The proof of Theorem 3 leverages Theorem 1, using the connection of the free entropy to the mutual information, and the I-MMSE relation of information theory Guo et al. (2005). The proof of the MMSE limit from the mutual information follows the approach developed in (Barbier et al., 2019, Theorem 2). We provide a sketch of how this generalizes to the present context in Appendix D.

Remark 4 *Theorem 3 comes from the fact that*

$$\left|\frac{1}{d}\mathrm{Tr}[SS^\star]\right| = \left|\frac{1}{d}\sum_{i,j=1}^{d} s_{ij}S^\star_{ij}\right| \xrightarrow[d\to\infty]{\mathrm{p}} q^\star(\alpha), \tag{10}$$

where s is sampled from the posterior distribution $P(\cdot|Y,\Phi)$. This is proven in Appendix D, In general the absolute value here cannot be removed, and thus we consider the estimation error on the tensor $(S^\star\otimes S^\star)_{ijkl} \coloneqq S^\star_{ij}S^\star_{kl}$ in eq. (9): this is due to the potential presence of symmetries (i.e., if $\varphi(z,a) = \varphi(-z,a)$), in which case it is only possible to estimate $S^\star$ up to a global sign.

However, notice that under Assumption 2.1(ii), the overlap $\frac{1}{d}\mathrm{Tr}[SS^\star]$ is always nonnegative, and thus we can remove the absolute value in eq. (10), and obtain from it the MMSE on $S^\star$:

$$\mathrm{MMSE}_d \coloneqq \frac{1}{d}\mathbb{E}\left[||S^\star - \mathbb{E}[S|Y,\Phi]||^2\right] \xrightarrow{d\to\infty} \rho - q^\star(\alpha).$$

2.4. Generalization to the rectangular model

In this section we come back to the original model of eqs. (1), (2) with rectangular matrices, and generalize our main results stated above in the symmetric model to this setting. Recall that we consider the large system limit $L, d, n \to \infty$ with $n/(Ld) \to \alpha$ and $d/L \to \beta$, and that without loss of generality we assume $L \geq d$. We can now adapt our main assumptions (see Section 2.1) to this context. We assume that P_0 is a bi-rotationally-invariant distribution, as clarified in the following statement.

2. i.e. there exists $y \in \mathbb{R}$ such that $P_{\rm out}(y|\cdot)$ is not equal almost everywhere to a constant. If $P_{\rm out}$ is not informative, estimation is impossible.

Assumption 2.4 (Assumptions on P_0) *Let $\kappa \in (0,1]$, and denote $m \coloneqq \lceil \kappa d \rceil$. We assume that P_0 is supported on the space of matrices $S \in \mathbb{R}^{d\times L}$ with rank m, and has density of the form:*

$$P_0(S) \propto \exp(-d\,\mathrm{Tr}[V(\Sigma^2)]),$$

where $\Sigma = \mathrm{Diag}(\{\sigma_i\}_{i=1}^m)$ denotes the non-zero singular values of S, i.e. $\{\sigma_i^2\}$ are the non-zero eigenvalues of $SS^\top$. Here $V : \mathcal{B} \to \mathbb{R}$ is a a real and $\mathcal{C}^2$ potential, and $\mathcal{B} \subseteq \mathbb{R}^+$ is an interval. We assume $\frac{\mathrm{d}^2}{\mathrm{d}s^2}V(s^2) + 2(1-\kappa)/s^2 \geq 1/c$ for all $s \in \mathcal{B}$, and a constant $c > 0$.

Remark 5 *Assumption 2.4 is simply Assumption 2.1 applied to the matrix $SS^\top$.*

As in the symmetric model, we make the following uniform one-dimensional CLT assumption, the counterpart to Assumption 2.2.

Assumption 2.5 (Uniform one-dimensional CLT) *For $\Phi \sim P_\Phi$, we have $\mathbb{E}[\Phi] = 0$ and $\mathbb{E}[\Phi_{ij}\Phi_{kl}] = \delta_{ik}\delta_{jl}$ (for all indices i,j,k,l). Moreover, letting $\{G_{ij}\}_{i,j=1}^{L,d} \sim \mathcal{N}(0, 1/\sqrt{Ld})$, we assume that for any $M > 0$ and any bounded Lipschitz function ψ:*

$$\lim_{d\to\infty} \sup_{S \in B_{\mathrm{op}}(M)} |\mathbb{E}_\Phi[\psi(\mathrm{Tr}[\Phi S])] - \mathbb{E}_G[\psi(\mathrm{Tr}[GS])]| = 0.$$

Finally, we assume that the activation φ satisfies Assumption 2.3, as in the symmetric model. Following the same terminology as in the symmetric model, we define the free entropy as

$$f_d^{\mathrm{rec}} \coloneqq \frac{1}{dL}\mathbb{E}_{Y,\Phi} \log \mathbb{E}_S \prod_{\mu=1}^n P_{\mathrm{out}}(Y_\mu | \mathrm{Tr}[\Phi_\mu S]), \tag{11}$$

with Y_μ generated according to eq. (1). The replica-symmetric free entropy functional is defined similarly to eq. (5), with $\rho \coloneqq \lim_{d\to\infty}(1/\sqrt{dL})\mathbb{E}_{S\sim P_0}\mathrm{Tr}[S^2]$:

$$f_{\mathrm{RS}}^{\mathrm{rec}}(q,r) \coloneqq \psi_{P_0}^{\mathrm{rec}}(r) + \alpha \Psi_{\mathrm{out}}^{\mathrm{rec}}(q) + \frac{r(\rho-q)}{2}, \tag{12}$$

$$\Psi_{\mathrm{out}}^{\mathrm{rec}}(q) \coloneqq \mathbb{E}_{V,W,\widetilde{Y}} \int_{\mathbb{R}} \mathcal{D}w P_{\mathrm{out}}(\widetilde{Y}|\sqrt{q}V + \sqrt{\rho-q}w), \tag{13}$$

with $\widetilde{Y} \sim P_{\mathrm{out}}(\cdot|\sqrt{q}V + \sqrt{\rho-q}W)$, and $W, V \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. Finally, $\psi_{P_0}^{\mathrm{rec}}(r)$ is defined as

$$\psi_{P_0}^{\mathrm{rec}}(r) \coloneqq -\frac{1}{2}\log r - (1-\beta)\int \log|x|\widetilde{\mu}_{1/r}(\mathrm{d}x) - \beta\Sigma[\widetilde{\mu}_{1/r}] + C, \tag{14}$$

where $C > 0$ is a constant that is independent of r. $\widetilde{\mu}_t$ (for $t \geq 0$) is a measure akin to the free convolution which appears in eq. (6): its precise definition is given in Appendix A.1, see eq. (48).

We can now state the generalization of Theorems 1 and 3 to the rectangular setting. Their proofs are straightforward transpositions of the ones developed in the symmetric model: we sketch the main steps in Appendix H.

Theorem 6 (Limit of the free entropy and mutual information) *Under Assumptions 2.3, 2.4, and 2.5, we have:*

$$\lim_{d\to\infty} f_d^{\mathrm{rec}} = \sup_{q\in[0,\rho]} \inf_{r\geq 0} f_{\mathrm{RS}}^{\mathrm{rec}}(q,r); \qquad \lim_{d\to\infty} \frac{1}{d^2} I(S^\star, Y|\Phi) = -\sup_{q\in[0,\rho]} \inf_{r\geq 0} f_{\mathrm{RS}}^{\mathrm{rec}}(q,r) + \alpha\Psi_{\mathrm{out}}^{\mathrm{rec}}(\rho).$$

Again, via the I-MMSE theorem, Theorem 6 leads to a precise characterization of the Bayes-optimal mean squared error.

Theorem 7 (Limit of the MMSE) *Suppose that Assumptions 2.3, 2.4 and 2.5 hold, and that $P_{\rm out}$ is informative. Let*

$$\Gamma := \left\{q \in [0,\rho], r \geq 0 \,|\, q = \rho + 2(\psi_{P_0}^{\rm rec})'(r), r = 2\alpha\Psi'_{\rm out}(q)\right\}$$

be the set of critical points of $f_{\rm RS}^{\rm rec}(q,r)$, and

$$D^\star := \{\alpha > 0 \,|\, \inf_{r\geq 0} f_{\rm RS}(q,r) \text{ has a unique maximizer } q^\star(\alpha) \in [0,\rho]\}.$$

Then $D^\star$ is equal to $(0,\infty)$ minus some countable set, and for any $\alpha \in D^\star$:

$$\mathrm{MMSE}_d := \frac{1}{dL}\mathbb{E}\left[||S^\star \otimes S^\star - \mathbb{E}[S^\star \otimes S^\star|Y,W]||^2\right] \xrightarrow{d\to\infty} \rho^2 - q^\star(\alpha)^2.$$

2.5. Applications

The general results of Theorems 1 and 3 (as well as their counterparts in the rectangular model) can be applied to a variety of settings. We focus now on two such applications, where our results mathematically establish predictions obtained by non-rigorous analytical methods.

Extensive-width neural networks with quadratic activation – Maillard et al. (2024) conjectures the limiting free entropy and Bayes-optimal MMSE in a two-layer neural network with a quadratic activation. Specifically, we consider a dataset of n samples $\mathcal{D} = \{Y_\mu, x_\mu\}_{\mu=1}^n$ where the input data is normal Gaussian of dimension d: $(x_\mu)_{\mu=1}^n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_d)$. We let $m = \kappa d$, with $\kappa > 0$ and $\kappa \neq 1$. We then draw i.i.d. d-dimensional teacher-weight vectors $(w_k^\star)_{k=1}^m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_d)$, and denote $W^\star \in \mathbb{R}^{m\times d}$ the matrix with rows $(w_k^\star)$, and we draw the noise $(z_{\mu,k})_{\mu,k=1}^{n,m} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. We fix a "noise strength" $\Delta > 0$. Finally, the output labels $(Y_\mu)_{\mu=1}^n$ are obtained by a one-hidden layer teacher network with m hidden units and quadratic activation:

$$Y_\mu = f_{W^\star}(x_\mu) := \frac{1}{m}\sum_{k=1}^m \left[\frac{1}{\sqrt{d}}(w_k^\star)^\top x_\mu + \sqrt{\Delta} z_{\mu,k}\right]^2 + \sqrt{\Delta_0}\zeta_\mu. \tag{15}$$

We introduced in eq. (15) a post-activation noise $\{\zeta_\mu\}_{\mu=1}^n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ with $\Delta_0 > 0$: this allows to simplify some parts of the proof, but we believe that this assumption can be relaxed. Following Maillard et al. (2024), we define the generalization error of the Bayes optimal estimator to be

$$\mathrm{MMSE}_d := \frac{m}{2}\mathbb{E}_{W^\star,\mathcal{D}}\mathbb{E}_{Y_{\rm test},x_{\rm test}}\left[\left(Y_{\rm test} - \hat{Y}_{\mathcal{D}}^{\rm BO}(x_{\rm test})\right)^2\right] - \Delta(2+\Delta), \tag{16}$$

where we rescaled the error according to Maillard et al. (2024), and where

$$\hat{Y}_{\mathcal{D}}^{\rm BO}(x_{\rm test}) := \mathbb{E}\left[Y_{\rm test}|x_{\rm test},\mathcal{D}\right] = \int \mathbb{E}_z[f_W(x_{\rm test})]\,\mathbb{P}(W|\mathcal{D})\,\mathrm{d}W\,, \tag{17}$$

with $\mathbb{P}(W|\mathcal{D})$ the posterior distribution of W given the dataset $\mathcal{D}$.

The following theorem rigorously establishes the predictions for the limiting Bayes-optimal MMSE that were obtained in Maillard et al. (2024)[3].

3. Modulo the technical restriction $m/d \to \kappa \in (0,\infty)\backslash\{1\}$, see the discussion around Assumption 2.1.

Theorem 8 (Limit of the generalization MMSE)

$$\lim_{d\to\infty} \mathrm{MMSE}_d = \kappa(\rho - q^\star(\alpha)),$$

for $\alpha \in D^\star$, where $q^\star(\alpha)$ is the unique maximizer of the replica free entropy defined in Theorem 3, with the channel $\varphi(x) = x$, the noise level $\widetilde{\Delta} := \frac{2\Delta(2+\Delta)}{\kappa} + \Delta_0$, and the Wishart prior $\mathcal{W}_{m,d}$.

We sketch the proof strategies of Theorem 8 in Section 3.4, with its details given in Appendix E.

We note that our results are also related to a conjecture of Cui et al. (2023), which suggested that a function defined by a two-layer neural network with a quadratic (i.e. $\mathcal{O}(d^2)$, where d is the input dimension) number of parameters can be learned from $n = \mathcal{O}(d^2)$ samples. For quadratic polynomial functions $f(x) = f_0 + \langle u^\star, x\rangle + x^\top S^\star x +$ noise, we answer by the affirmative. Indeed, a simple plug-in estimator can learn the scalar f_0 and the $d-$dimensional vector $u^\star$ as soon as $n = \omega(d)$ (for instance $\hat{u} := (d/n)\sum_{\mu=1}^n y_\mu x_\mu$ converges to $u_i^\star$ with vanishing error, see (Dandi et al., 2024a, Appendix A.1)), so that the question simply reduces to the case $f(x) = x^\top S^\star x$, for which we provide the asymptotic MMSE, as discussed above.

Bilinear sequence regression – Another model of interest is Bilinear Sequence Regression (BSR), a matrix sensing model studied recently in Erba et al. (2024) as a toy model of learning from sequences of tokens. The observations in this model are generated as

$$Y_\mu \sim P_{\mathrm{out}}\left(\cdot\Big|\frac{1}{r\sqrt[4]{dL}}\sum_{\gamma=1}^{r}\sum_{a,i=1}^{L,d} X_{ia}^\mu U_{i\gamma}^\star V_{\gamma a}^\star\right), \tag{18}$$

where $\{X_{ia}^\mu\}_{\mu,i,a=1}^{n,L,d} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ can be interpreted as n samples of sequences of length L of token embedded in dimension d. The weights $U^\star \in \mathbb{R}^{d\times r}$ and $V^\star \in \mathbb{R}^{r\times L}$ are drawn i.i.d. from $\mathcal{N}(0,1)$. Denoting $S^\star := (1/r)U^\star V^\star$, its distribution satisfies Assumption 2.4. Since the input samples X_{ia}^μ are assumed to be standard Gaussian, the model of eq. (18) falls under our general analysis. More specifically, the results presented in Section 2.4 rigorously establish the predictions for the Bayes-optimal MMSE described in Erba et al. (2024).

3. Sketch of proofs of the main results

The proof of Theorem 1 contains three crucial ingredients.

(i) First, we show that, under Assumption 2.2, one can replace the data matrices $\Phi_\mu \overset{\text{i.i.d.}}{\sim} P_\Phi$ with $G_\mu \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$, without changing the $d \to \infty$ limit of the free entropy f_d. This *universality* argument is inspired by recent results on empirical risk minimization (see e.g. Goldt et al. (2022); Hu and Lu (2022); Montanari and Saeed (2022); Gerace et al. (2024) and references therein) and on the ellipsoid fitting problem Maillard and Bandeira (2023).

(ii) The universality property above implies that we can consider what is essentially a generalized linear model with *Gaussian data* and a rotationally-invariant prior supported on symmetric matrices. We extend results on generalized linear models which were developed for i.i.d. priors Barbier et al. (2019) to thi setting, by adapting an interpolation argument.

(iii) Finally, the analysis of point (ii) yields that the limiting free entropy is expressed as a function of the free entropy of a *matrix denoising* problem. We then use recent results on denoising of symmetric Bun et al. (2016); Maillard et al. (2022); Pourkamali et al. (2024); Semerjian (2024) and non-symmetric Troiani et al. (2022); Pourkamali and Macris (2024) matrices to finish the proof of Theorem 1.

To deduce Theorem 3 from Theorem 1, we establish the limiting free entropy in a so-called "spiked tensor" model, with rotationally invariant signals. We then leverage this result by adding this model as a small side-information to our setting, and use then the I-MMSE theorem to establish the limit of the MMSE. This follows the approach of Barbier et al. (2019), and is sketched in Section 3.3, with details defered to Appendix D. To deduce Theorem 8 from Theorem 3, we need to interpolate between the two-layer neural networks and a matrix sensing model: the argument is sketched in Section 3.4, with details postponed to Appendix E.

3.1. Universality and reduction to Gaussian sensing matrices – ingredient (i)

In this part, we use the representation of the output channel via an explicit random function φ, see eq. (1). Following a strategy introduced in Montanari and Saeed (2022), we define the following "interpolated" model

$$Y_\mu = \varphi(\mathrm{Tr}[U_\mu(t)S^\star], A_\mu) + \sqrt{\Delta} Z_\mu, \tag{19}$$

for $t \in [0, \pi/2]$, where $U_\mu(t) \coloneqq \cos(t)\Phi_\mu + \sin(t)G_\mu$, and $\{G_\mu\}_{\mu=1}^n \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$. Note that $t = 0$ recovers the original model. We naturally define the interpolated free entropy as:

$$F_d(U(t)) \coloneqq \frac{1}{d^2} \log \mathbb{E}_{S\sim P_0}\left[e^{-\sum_{\mu=1}^n \ell(\mathrm{Tr}[U_\mu(t)S], \mathrm{Tr}[U_\mu(t)S^\star], Z_\mu)}\right] - \frac{1}{d^2}\sum_{\mu=1}^n \frac{Z_\mu^2}{2\Delta}, \tag{20}$$

where

$$\ell(x, X, Z) \coloneqq -\log \int dP_A(a) \exp\left(-\frac{(\varphi(x,a) - \varphi(X,A))^2 - 2Z(\varphi(x,a) - \varphi(X,A))}{2\Delta}\right). \tag{21}$$

Note that $F_d(U)$ implicitly relies on the realization of the random variables $Z, A, S^\star$, and that $f_d = \mathbb{E}F_d(U(0))$, where f_d is defined in eq. (3). Our setting essentially corresponds to a "planted" (i.e. with an underlying signal $S^\star$) variant of Maillard and Bandeira (2023): on the other hand, Montanari and Saeed (2022) considers a planted problem as we do, but under stronger assumptions on the prior distribution (see a discussion of this point in (Maillard and Bandeira, 2023, Section 2.2)).

We first state our universality lemma under two assumptions that will be weakened in the end.

Assumption 3.1 *Under Assumption 2.1, we further assume that $\mathcal{B} \subseteq [-M, M]$ for some $M > 0$.*

Assumption 3.2 *φ is a bounded function with bounded first and second derivatives w.r.t. its first argument.*

Essentially, these two assumptions can be weakened to Assumption 2.1 and 2.3, as we can build a bounded $\hat{\varphi}$ (with good regularity properties) such that the output difference $\mathbb{E}[(\varphi(\mathrm{Tr}[\Phi S^\star], a) - \hat{\varphi}(\mathrm{Tr}[\Phi S^\star], a))^2]$ is arbitrarily small, and show that this implies a bound on the difference of the free entropies. To deal with the eigenvalues of $S^\star$, we interpolate between the original potential V in Assumption 2.1 and another potential $\widetilde{V}$ that is infinity outside a bounded interval. See Appendix C.4 for details.

Lemma 9 (Universality) *Under Assumptions 2.2, 3.1 and 3.2, we have*

$$\lim_{d\to\infty} |\mathbb{E}[\psi(F_d(\Phi))] - \mathbb{E}[\psi(F_d(G))]| = 0 \tag{22}$$

for any bounded function ψ with bounded Lipschitz derivative, with free entropy defined in eq. (20).

Let us now briefly sketch the proof of Lemma 9. Clearly, it is sufficient to prove

$$\lim_{d\to\infty} \sup_{\{A_\mu\}_{\mu=1}^n} \sup_{S^\star\in B_{\rm op}(M)} |\mathbb{E}_{\Phi,Z}[\psi(F_d(\Phi))] - \mathbb{E}_{G,Z}[\psi(F_d(G))]| = 0. \tag{23}$$

Using the interpolation path described above and the triangular inequality, we can write:

$$|\mathbb{E}_{\Phi,Z}[\psi(F_d(\Phi))] - \mathbb{E}_{G,Z}[\psi(F_d(G))]| \le \int_0^{\pi/2} \left|\mathbb{E}_{G,\Phi,Z} \frac{\partial \psi(F_d(U(t)))}{\partial t}\right| \mathrm{d}t, \tag{24}$$

and we will prove that the right-hand side of eq. (24) goes to zero uniformly in $S^\star \in B_{\rm op}(M)$ and $\{A_\mu\}_{\mu=1}^n$. This then essentially follows from an adaptation of the arguments of Maillard and Bandeira (2023) on the problem of ellipsoid fitting, that we detail in Appendix B.

3.2. Matrix sensing with Gaussian data – (ii) and (iii)

We sketch here the end of the proof of Theorem 1. Thanks to Lemma 9, we focus on the model

$$Y_\mu \sim P_{\rm out}(\cdot|\mathrm{Tr}[G_\mu S^\star]), \tag{25}$$

where $(G_\mu)_{\mu=1}^n \overset{\text{i.i.d.}}{\sim} \mathrm{GOE}(d)$. We use the adaptive interpolation method Barbier et al. (2019), and introduce:

$$Y_t \sim P_{\rm out}(\cdot|J_t); \qquad Y_t' = \sqrt{R_1(t)}S^\star + Z', \tag{26}$$

with $Z' \sim \mathrm{GOE}(d)$, and where we defined

$$J_{t,\mu} := \sqrt{1-t}\,\mathrm{Tr}[G_\mu S^\star] + \sqrt{2R_2(t)}V_\mu + \sqrt{2\rho t - 2R_2(t) + 2s_d}U_\mu^\star,$$

with $\{V_\mu, U_\mu^\star\}_{\mu=1}^n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. The goal is now to recover *both* $S^\star$ and $U^\star$ from the observations of (Y_t, Y_t'). The choice of the functions $R_1(t), R_2(t)$ will be precised later on, we simply require at the moment that $R_1(t) \ge 0$, $R_2(t) \in [0,\rho]$, and that $R_1(0) = R_2(0) = 0$.

Notice that eq. (26) reduces to (25) for $t = 0$, while for $t = 1$ its free entropy reduces to the replica-symmetric prediction of eq. (5), with $q = R_2(1)$ and $r = R_1(1)$. This last point is a consequence of the analysis of the *matrix denoising* problem, which arises from the observation of Y_t' in eq. (26).

Our goal reduces now to chose the functions $R_1(t), R_2(t)$, so as to control the derivative of the free entropy of the interpolated model with respect to t. Following Barbier et al. (2019), a so-called *adaptive* choice allows to sharply control this derivative, see Lemmas 27 and 28 in Appendix C. The main difference with our setting is that the signal $S^\star$ is a rotationally-invariant matrix (see Assumption 2.1), while Barbier et al. (2019) focuses on simpler i.i.d. priors. Beyond the analysis of the matrix denoising problem, the other key result needed to extend their approach is a Poincaré inequality for rotationally-invariant distributions satisfying Assumption 2.1, see Lemma 45 and Chafaï and Lehec (2020). It essentially reads as:

$$\mathrm{Var}(g(S^\star)) \le \frac{c}{\kappa d}\mathbb{E}\left[\sum_{i=1}^d \left(\frac{\partial g}{\partial \Lambda_i^\star}\right)^2\right], \tag{27}$$

where $\{\Lambda_i^\star\}_{i=1}^d$ are the eigenvalues of $S^\star$, for any "regular enough" function g. Eq. (27) allows us to prove concentration properties for the free entropy of the interpolated model of eq. (26), which in turns makes possible the generalization of the adaptive interpolation approach of Barbier et al. (2019). A detailed treatment of the arguments above, which form the proof of Theorem 1, is given in Appendix C.

3.3. Limit of the MMSE: sketch of proof of Theorem 3

Using the I-MMSE theorem Guo et al. (2005), the MMSE can be related to the free entropy by considering the following observation model, in which we added a "side information":

$$Y_\mu \sim P_{\mathrm{out}}(\cdot|\mathrm{Tr}[G_\mu S^\star]), \ (\mu \in [n]); \qquad Y' = \sqrt{\frac{\lambda}{d^{2p-1}}}(S^\star)^{\otimes 2p} + Z', \tag{28}$$

where $Z' \in \mathcal{S}_d^{\otimes 2p} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ is a tensor, and we chose an integer $p \geq 1$ and a signal-to-noise ratio $\lambda \geq 0$. The additional side information comes from a so-called *spiked tensor* model. In Appendix F we detail the proof of the limiting free entropy for the model of eq. (28) (under Assumptions 2.1, 2.2 and 2.3), which leverages as well the adaptive interpolation method of Barbier and Macris (2019). Finally, we describe in Appendix D how this can be used to end the proof of Theorem 3.

3.4. Reduction of the two-layer network: sketch of proof of Theorem 8

We sketch here the proof, with details given in Appendix E. To prove Theorem 8, we will show that we can replace the model of eq. (15) by a matrix sensing problem in terms of $S^\star := (1/m)\sum_{k=1}^m w_k^\star (w_k^\star)^\top$. This establishes rigorously the heuristic argument in Appendix C.3 of Maillard et al. (2024).

Specifically, we will prove that both models have the same free entropy, and then deduce the equivalence of MMSE using the I-MMSE theorem (by adding a side information channel, whose signal-to-noise ratio eventually goes to 0). We first rewrite the model of eq. (15) as

$$v_\mu = \mathrm{Tr}[\Phi_\mu S^\star] + \sqrt{d}(\mathrm{tr} S^\star - 1) + \sqrt{d}\Delta\left(\frac{\|z_\mu\|^2}{m} - 1\right) + \frac{2\sqrt{\Delta}}{m}\sum_{k=1}^m z_{\mu,k}\left(\frac{x_\mu^\top w_k^\star}{\sqrt{d}}\right) + \sqrt{\Delta_0}\zeta_\mu, \tag{29}$$

where $v_\mu := \sqrt{d}(Y_\mu - \Delta - 1)$ and $\Phi_\mu := (x_\mu x_\mu^\top - \mathrm{I}_d)/\sqrt{d}$. For notational simplicity, we define $\mathrm{tr}(\cdot) := (1/d)\mathrm{Tr}(\cdot)$. We use then an interpolation argument. For $t \in [t_0, 1]$, we define

$$\begin{aligned} v_{t,\mu} &:= \mathrm{Tr}[\Phi_\mu S^\star] + (1-t)\sqrt{d}[\mathrm{tr}\, S^\star - 1] + \sqrt{d(1-t)}\Delta\left(\frac{\|z_\mu\|^2}{m} - 1\right) \\ &+ \frac{2\sqrt{\Delta d(1-t)}}{m}\sum_{k=1}^m z_{\mu,k}\left(\frac{x_\mu^\top w_k^\star}{\sqrt{d}}\right) + \sqrt{\widetilde{\Delta} t}\zeta_\mu, \end{aligned}$$

where $\widetilde{\Delta} := 2\Delta(2+\Delta)/\kappa + \Delta_0$ and $t_0 := \Delta_0/\widetilde{\Delta} < 1$. Notice that, for $t = t_0$, $v_{t_0,\mu} = v_\mu$ is given by eq. (29), while for $t = 1$, $v_{1,\mu} = \mathrm{Tr}[\Phi_\mu S^\star] + \sqrt{\widetilde{\Delta}}\zeta_\mu$ is a "matrix GLM" problem, with Gaussian additive noise. The technical introduction of the post-activation noise $\Delta_0 > 0$ allows to have t lower bounded: while we believe this not to be necessary, it allows for simplifications in part of our proof. We denote $v_{t,\mu} \sim P_{\mathrm{out}}^{(t)}(\cdot|W^\star, x_\mu)$ the distribution of these labels.

Moreover, We introduce a small "side-information" channel to the observations, similarly e.g. to Barbier et al. (2019); Maillard et al. (2020). More precisely, for $\Lambda \geq 0$ we add the observation of $Y' \in \mathbb{R}$, which is built as:

$$Y' = \sqrt{\Lambda} S^{\star}_{12} + \frac{\xi}{\sqrt{d}}, \tag{30}$$

where $\xi \sim \mathcal{N}(0,1)$. Adding this side information channel, we will then be able to leverage the I-MMSE theorem Guo et al. (2005), before eventually taking $\Lambda \downarrow 0$. We define the interpolated free entropy as

$$f_d(t,\Lambda) := \mathbb{E}_{W^*,Y',x} \frac{1}{d^2} \log \int_{\mathbb{R}^{m\times d}} \mathcal{D}W \left[\prod_{\mu=1}^{n} P^{(t)}_{\text{out}}(v_{\mu,t}|W,x_\mu)\right] e^{-\frac{d}{4}(Y'-\sqrt{\Lambda}S_{12})^2}. \tag{31}$$

We denote $\langle\cdot\rangle_{t,\Lambda}$ the average with respect to the Gibbs-like measure with weight proportional to the integrand of eq. (31). By the I-MMSE theorem and the Nishimori identity, we prove that, for any $t \in [t_0, 1]$:

$$\left|\frac{\partial}{\partial\Lambda} f_d(t,0) + \frac{1}{2}\mathbb{E}\,\mathrm{tr}[(S^\star - \langle S\rangle_{t,0})^2]\right| \leq \frac{C(\kappa)}{d}, \tag{32}$$

for some $C(\kappa) > 0$. The MSE on $S^\star$ is connected to the generalization MMSE via

$$\left|\mathrm{MMSE}_d - \kappa\mathbb{E}\mathrm{tr}[(S^\star - \langle S\rangle_{t_0,0})^2\right| \leq \frac{C(\kappa)}{n}. \tag{33}$$

for some $C(\kappa) > 0$: eq. (33) is proven in (Maillard et al., 2024, Lemma D.1). Moreover, by Theorem 3 and Remark 4 we have

$$\mathbb{E}\,\mathrm{tr}[(S^\star - \langle S\rangle_{1,0})^2] \xrightarrow[d\to\infty]{\mathrm{p}} \rho - q^\star(\alpha). \tag{34}$$

It remains then to prove the following lemma.

Lemma 10

$$\lim_{d\to\infty} \left|\frac{\partial}{\partial\Lambda} f_d(1,0) - \frac{\partial}{\partial\Lambda} f_d(t_0,0)\right| = 0. \tag{35}$$

Combining eqs. (32), (33), (34) and (35), we can finish the proof of Theorem 8.

The proofs of eq. (32) and Lemma 10 are given in Appendix E. For Lemma 10, we will utilize the interpolation model, and show that $\frac{\partial}{\partial t} f_d(t,\Lambda)$ goes to zero uniformly in $[t_0, 1]$ and $\Lambda \geq 0$, implying that $|f_d(1,\Lambda) - f_d(t_0,\Lambda)|$ goes to zero uniformly in Λ. Using the convexity of the free entropy as a function of Λ, we then deduce that $\left|\frac{\partial}{\partial\Lambda} f_d(1,0) - \frac{\partial}{\partial\Lambda} f_d(t_0,0)\right|$ also goes to zero.

Acknowledgments

We acknowledge funding from the Swiss National Science Foundation grants SNFS SMArtNet (grant number 212049), OperaGOST (grant number 200021 200390) and DSGIANGO.

References

Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Cambridge university press, 2010.

Benjamin Aubin, Bruno Loureiro, Antoine Baker, Florent Krzakala, and Lenka Zdeborová. Precise asymptotics for phase retrieval and compressed sensing with random generative priors. In *NeurIPS 2019 Workshop on Solving Inverse Problems with Deep Networks*, 2019.

Benjamin Aubin, Bruno Loureiro, Antoine Maillard, Florent Krzakala, and Lenka Zdeborova. The spiked matrix model with generative priors. *IEEE Transactions on Information Theory*, 67(2): 1156–1181, 2020.

Jean Barbier and Nicolas Macris. The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference. *Probability theory and related fields*, 174:1133–1185, 2019.

Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, 116(12):5451–5460, 2019.

Jean Barbier, Nicolas Macris, Mohamad Dia, and Florent Krzakala. Mutual information and optimality of approximate message-passing in random linear estimation. *IEEE Transactions on Information Theory*, 66(7):4270–4303, 2020.

Florent Benaych-Georges. Rectangular random matrices, related convolution. *Probability Theory and Related Fields*, 144:471–515, 2009.

Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro. Global optimality of local search for low rank matrix recovery. *Advances in Neural Information Processing Systems*, 29, 2016.

Joël Bun, Romain Allez, Jean-Philippe Bouchaud, and Marc Potters. Rotational invariant estimator for general noisy matrices. *IEEE Transactions on Information Theory*, 62(12):7475–7490, 2016.

Emmanuel Candes and Benjamin Recht. Exact matrix completion via convex optimization. *Communications of the ACM*, 55(6):111–119, 2012.

Emmanuel J Candes and Yaniv Plan. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory*, 57 (4):2342–2359, 2011.

Djalil Chafaï and Joseph Lehec. On poincaré and logarithmic sobolev inequalities for a class of singular gibbs measures. In *Geometric Aspects of Functional Analysis: Israel Seminar (GAFA) 2017-2019 Volume I*, pages 219–246. Springer, 2020.

Hugo Cui, Florent Krzakala, and Lenka Zdeborová. Bayes-optimal learning of deep random networks of extensive-width. In *International Conference on Machine Learning*, pages 6468–6521. PMLR, 2023.

Yatin Dandi, Florent Krzakala, Bruno Loureiro, Luca Pesce, and Ludovic Stephan. How two-layer neural networks learn, one (giant) step at a time. *Journal of Machine Learning Research*, 25 (349):1–65, 2024a. URL http://jmlr.org/papers/v25/23-1543.html.

Yatin Dandi, Ludovic Stephan, Florent Krzakala, Bruno Loureiro, and Lenka Zdeborová. Universality laws for gaussian mixtures in generalized linear models. *Advances in Neural Information Processing Systems*, 36, 2024b.

David L Donoho, Matan Gavish, and Andrea Montanari. The phase transition of matrix recovery from gaussian measurements matches the minimax mse of matrix denoising. *Proceedings of the National Academy of Sciences*, 110(21):8405–8410, 2013.

Vittorio Erba, Emanuele Troiani, Luca Biggio, Antoine Maillard, and Lenka Zdeborová. Bilinear sequence regression: A model for learning from long sequences of high-dimensional tokens. *arXiv preprint arXiv:2410.18858*, 2024.

Chenjie Fan, Alice Guionnet, Yuqi Song, and Andi Wang. Convergence of eigenvalues to the support of the limiting measure in critical β matrix models. *Random Matrices: Theory and Applications*, 4(03):1550013, 2015.

Maryam Fazel. *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002.

Marylou Gabrié, Andre Manoel, Clément Luneau, Jean Barbier, Nicolas Macris, Florent Krzakala, and Lenka Zdeborová. Entropy and mutual information in models of deep neural networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124014, 2019.

Federica Gerace, Bruno Loureiro, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Generalisation error in learning with random features and the hidden manifold model. In *International Conference on Machine Learning*, pages 3452–3462. PMLR, 2020.

Federica Gerace, Florent Krzakala, Bruno Loureiro, Ludovic Stephan, and Lenka Zdeborová. Gaussian universality of perceptrons with random labels. *Phys. Rev. E*, 109:034305, Mar 2024. doi: 10.1103/PhysRevE.109.034305. URL https://link.aps.org/doi/10.1103/PhysRevE.109.034305.

Sebastian Goldt, Bruno Loureiro, Galen Reeves, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. The gaussian equivalence of generative models for learning with shallow neural networks. In *Mathematical and Scientific Machine Learning*, pages 426–471. PMLR, 2022.

David Gross, Yi-Kai Liu, Steven T Flammia, Stephen Becker, and Jens Eisert. Quantum state tomography via compressed sensing. *Physical review letters*, 105(15):150401, 2010.

Alice Guionnet and Jiaoyang Huang. Large deviations asymptotics of rectangular spherical integral. *Journal of Functional Analysis*, 285(11), 2021.

Suriya Gunasekar, Blake E Woodworth, Srinadh Bhojanapalli, Behnam Neyshabur, and Nati Srebro. Implicit regularization in matrix factorization. *Advances in neural information processing systems*, 30, 2017.

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.

Hong Hu and Yue M Lu. Universality laws for high-dimensional learning with random features. *IEEE Transactions on Information Theory*, 69(3):1932–1964, 2022.

Beatrice Laurent and Pascal Massart. Adaptive estimation of a quadratic functional by model selection. *Annals of statistics*, pages 1302–1338, 2000.

Yuanzhi Li, Tengyu Ma, and Hongyang Zhang. Algorithmic regularization in over-parameterized matrix sensing and neural networks with quadratic activations. In *Conference On Learning Theory*, pages 2–47. PMLR, 2018.

Antoine Maillard and Afonso S Bandeira. Exact threshold for approximate ellipsoid fitting of random points. *arXiv preprint arXiv:2310.05787*, 2023.

Antoine Maillard, Bruno Loureiro, Florent Krzakala, and Lenka Zdeborová. Phase retrieval in high dimensions: Statistical and computational phase transitions. *Advances in Neural Information Processing Systems*, 33:11071–11082, 2020.

Antoine Maillard, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Perturbative construction of mean-field equations in extensive-rank matrix factorization and denoising. *Journal of Statistical Mechanics: Theory and Experiment*, 2022(8):083301, 2022.

Antoine Maillard, Emanuele Troiani, Simon Martin, Florent Krzakala, and Lenka Zdeborova. Bayes-optimal learning of an extensive-width neural network from quadratically many samples. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Song Mei and Andrea Montanari. The generalization error of random features regression: Precise asymptotics and the double descent curve. *Communications on Pure and Applied Mathematics*, 75(4):667–766, 2022.

Andrea Montanari and Basil N Saeed. Universality of empirical risk minimization. In *Conference on Learning Theory*, pages 4310–4312. PMLR, 2022.

Jason T Parker and Philip Schniter. Parametric bilinear generalized approximate message passing. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):795–808, 2016.

Farzad Pourkamali and Nicolas Macris. Rectangular rotational invariant estimator for high-rank matrix estimation. *arXiv preprint arXiv:2403.04615*, 2024.

Farzad Pourkamali, Jean Barbier, and Nicolas Macris. Matrix inference in growing rank regimes. *IEEE Transactions on Information Theory*, 2024.

Benjamin Recht, Maryam Fazel, and Pablo A Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.

Elad Romanov and Matan Gavish. Near-optimal matrix recovery from random linear measurements. *Proceedings of the National Academy of Sciences*, 115(28):7200–7205, 2018.

Christophe Schülke, Philip Schniter, and Lenka Zdeborová. Phase diagram of matrix compressed sensing. *Physical Review E*, 94(6):062136, 2016.

Guilhem Semerjian. Matrix denoising: Bayes-optimal estimators via low-degree polynomials. *Journal of Statistical Physics*, 191(10):139, 2024.

Emanuele Troiani, Vittorio Erba, Florent Krzakala, Antoine Maillard, and Lenka Zdeborová. Optimal denoising of rotationally invariant rectangular matrices. In *Mathematical and Scientific Machine Learning*, pages 97–112. PMLR, 2022.

Harald Uhlig. On singular wishart and singular multivariate beta distributions. *The Annals of Statistics*, pages 395–405, 1994.

A Vaswani, N Shazeer, N Parmar, J Uszkoreit, L Jones, A Gomez, L Kaiser, and I Polosukhin. Attention is all you need. In *NIPS*, 2017.

Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.

Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.

Contents

Appendix A. Classical definitions and remarks

A.1. Classical definitions

The following two standard random matrix distributions over the $\mathcal{S}_d$ are of particular interest:

- We say that a random matrix $Y \in \mathcal{S}_d$ is generated from the *Gaussian Orthogonal Ensemble* $\mathrm{GOE}(d)$ if $Y_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, (1 + \delta_{ij})/d)$ for $1 \le i \le j \le d$. It clearly satisfies Assumption 2.1, with $V(x) = x^2/4$.

- For an integer $m \ge 1$, we say that a random matrix $Y \in \mathcal{S}_d$ is generated from the *Wishart Ensemble* $\mathcal{W}_{m,d}$ if $Y = (1/m)W^\top W$, with $W \in \mathbb{R}^{m\times d}$ and $W_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ for $i \in [m], j \in [d]$. Notice that if $m, d \to \infty$ with $m/d \to \kappa \in (0,\infty)\backslash\{1\}$, then Y satisfies Assumption 2.1 (see e.g. Anderson et al. (2010); Uhlig (1994)):
  - If $\kappa > 1$, is satisfies case (i), with[4]
  $$V(x) = \begin{cases} \frac{\kappa}{2}x - \frac{\kappa-1}{2}\log x & \text{if } x > 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{36}$$
  - If $\kappa < 1$, is satisfies case (ii) (Uhlig, 1994, Theorem 6), with $V(x)$ still given by eq. (36).

Notice that in the Wishart setting we do not have $V''(x) = (\kappa-1)/(2x^2) \ge 1/c$ for all $x \in \mathbb{R}$. However for all $M > 0$ we have $V''(x) \ge 1/c$ for $x \in (0, M)$ and some $c = c(M) > 0$, which we show to be sufficient for our analysis, see Remark 15.

4. The probability density actually satisfies $P_0(S) \propto e^{-d\mathrm{Tr}[V_d(S)]}$, with $V_d(x) := V(x) + (1/2d)\log x$. However, this second term only changes the constant c in the Gaussian Poincaré inequality (Lemma 48) by an additive $o_d(1)$, and thus does not influence our results.

Remark 11 (The case $\kappa = 1$) *We notice that the Wishart distribution with $\kappa = 1$ does not satisfy Assumption 2.1, since then V is not strongly convex: $V(x) = x/2$ for $x > 0$. This prevents us from obtaining a Poincaré inequality for this prior (see Lemma 48 and the discussion around this result), which is a key element of our analysis. Since our analysis applies for any $\kappa \neq 1$, we expect this condition to be an artifact of our proof techniques, and we leave a lifting of this restriction for future work.*

The limiting eigenvalue distributions of the two ensembles above are well-known Anderson et al. (2010).

- The semi-circle law
$$\mu_{\text{s.c.}}(x) := \frac{\sqrt{4-x^2}}{2\pi}\mathbb{1}\{|x| \le 2\} \tag{37}$$
is the limiting eigenvalue distribution of a matrix $Y \sim \mathrm{GOE}(d)$.

- If $m/d \to \kappa > 0$ as $d \to \infty$, the Marchenko-Pastur law
$$\mu_{\text{MP}}(x) := \begin{cases} (1-\kappa)\delta(x) + \frac{\kappa\sqrt{(\lambda_+ - x)(x-\lambda_-)}}{2\pi x} & \text{if } \kappa \le 1, \\ \frac{\kappa\sqrt{(\lambda_+ - x)(x-\lambda_-)}}{2\pi x} & \text{if } \kappa \ge 1, \end{cases} \tag{38}$$
where $\lambda_\pm := (1 \pm \kappa^{-1/2})^2$, is the limiting eigenvalue distribution of $Y \sim \mathcal{W}_{m,d}$.

Additive free convolution: symmetric case – We will also consider the (additive) free convolution of measures. Informally, we can interpret the free convolution $\mu \boxplus \nu$ of two measures μ and ν as the limiting spectral measure of $A + B$, where $A, B \in \mathcal{S}_d$ are two random matrices with limiting spectral measures μ and ν respectively, when A and B satisfy a condition called *asymptotic freeness*. We refer to Anderson et al. (2010) for more formal discussions on asymptotic freeness and free convolution. In this paper, we will only use the special case $B \sim \mathrm{GOE}(d)$ independently of A, so that $\nu = \mu_{\text{s.c.}}$. In this case, A, B are asymptotically free (Anderson et al., 2010, Theorem 5.4.5).

A.2. Properties of priors satisfying Assumptions 2.1 and 2.4

A.2.1. THE SYMMETRIC CASE: ASSUMPTION 2.1

Let us first note that the two cases of Assumption 2.1 are essentially equivalent.

Remark 12 *Assumption 2.1(ii) can be seen as a special case of Assumption 2.1(i). Indeed, under assumption 2.1(ii) the joint distribution of the non-zero eigenvalues reads (see (Uhlig, 1994, Theorem 2)):*
$$P(\{\lambda_i\}_{i=1}^m) \propto \Pi_{i<j}|\lambda_i - \lambda_j| e^{-d\sum_{i=1}^m \widetilde{V}(\lambda_i)}, \tag{39}$$
where $\widetilde{V}(s) := V(s) - (1 - \frac{m}{d})\log s$. Eq. (39) is equivalent to the joint eigenvalue distribution under assumption 2.1 (i), replacing d by m and considering the potential $\widetilde{V}$.

We now list important properties of P_0 under Assumption 2.1.

Proposition 13 *Assumption 2.1 implies the following properties.*

(i) *According to (Fan et al., 2015, Remark 1.3), there exists μ_0, a probability measure with compact support such that the empirical eigenvalue distribution μ_S of $S \sim P_0$ a.s. weakly converges to μ_0.*

(ii) *There exists $M, c' > 0$ such that*

$$P_0(\{\lambda_i\}_{i=1}^d \subset [-M, M]) \geq 1 - e^{-c'd}, \tag{40}$$

where $\{\lambda_i\}_{i=1}^d$ are the eigenvalues of $S \sim P_0$.

(iii) *Eq. (40) implies that*

$$P_0\left(\frac{1}{d}\sum_{i=1}^d \lambda_i^2 > M^2\right) \leq e^{-c'd}.$$

By the Borel–Cantelli lemma, this implies that $(1/d)\sum_{i=1}^d \lambda_i^2$ is almost surely bounded, and thus uniformly integrable.

(iv) *By (Villani, 2021, Theorem 7.12-(ii)), as $(1/d)\sum_{i=1}^d \lambda_i^2$ is a.s. bounded and μ_S a.s. weakly converges to μ_0, we have*

$$\lim_{d\to\infty} W_2(\mu_S, \mu_0) = 0, \ a.s., \tag{41}$$

where W_2 denotes the Wasserstein-2 distance. This further implies that

$$\lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^d \lambda_i^2 = \rho := \int \mu_0(\mathrm{d}x)\, x^2, \ a.s. \tag{42}$$

As the left side is uniformly integrable and the (λ_i) have identical marginals, we have

$$\lim_{d\to\infty} \mathbb{E}_{S\sim P_0}[(\lambda_1)^2] = \rho, \tag{43}$$

which will be essential for our proof. Finally, one can show in a similar way that

$$\lim_{d\to\infty} \mathbb{E}_{S\sim P_0}[|\lambda_1|^p] < +\infty,$$

for any $p \geq 1$.

Properties (i), (iii) and (iv) of Proposition 13 are elementary consequences of the mentioned results. Property (ii) on the other hand is a consequence of the following lemma.

Lemma 14 (Boundedness of eigenvalues) *Let P_0 have the density:*

$$P_0(S) \propto \exp(-d\,\mathrm{Tr}[V(S)])$$

for a real continuous potential $V : \mathcal{B} \to \mathbb{R}$, where $\mathcal{B} \subseteq \mathbb{R}$ can be $\mathbb{R}$, an interval, or the union of finitely many disjoint intervals. If $\mathcal{B}$ is unbounded and

$$\liminf_{x\in\mathcal{B}, |x|\to\infty} \frac{V(x)}{2\log|x|} > 1, \tag{44}$$

then there exists $M > 0$ and $c > 0$ such that

$$P_0(\{\lambda_i\}_{i=1}^d \subset [-M, M]) \geq 1 - e^{-c'd}.$$

Proof (Fan et al., 2015, Theorem 1.2) shows that there exists a constant C_V such that

$$J(x) \coloneqq V(x) - 2\int \mu_0(dy) \log|x-y| + C_V > 0$$

for almost every $x \in \mathcal{B}/\mathcal{S}$, where $\mathcal{S}$ is the support of the limit eigenvalue distribution μ_0. Moreover, (Fan et al., 2015, Theorem 1.4) shows that

$$\limsup_{d\to\infty} \frac{1}{d} \log P_0(\exists i, |\lambda_i| > M) \leq -\frac{1}{2} \inf_{x\in\mathcal{B},|x|\geq M} J(x)$$

for any M such that $\mathcal{S} \subset [-M, M]$. Under (44), we have $\liminf_{x\in\mathcal{B},|x|\geq M} J(x) = +\infty$. Therefore, since J is continuous, there exists $M, c' > 0$ such that

$$\limsup_{d\to\infty} \frac{1}{d} \log P_0(\exists i, |\lambda_i| > M) \leq -c',$$

which finishes the proof. ■

It is easy to see that eq. (44) holds under either Assumption 2.1 or 2.4. Moreover, notice that it is sufficient to assume eq. (44) to deduce all properties of Proposition 13 (except).

Remark 15 (Weakening of Assumption 2.1) *For our results to hold (i.e. Theorem 1 and 3), Assumption 2.1-(i) can be weakened by assuming both:*

- *If $\mathcal{B}$ is unbounded:*
$$\liminf_{x\in B,|x|\to\infty} \frac{V(x)}{2\log|x|} > 1,$$
- *$V''(s) \geq 1/c$ for all $s \in \mathcal{B} \cap [-M, M]$, where $M > 0$ is defined by Lemma 14.*

One can similarly define a weakening of Assumption 2.1-(ii) under which our results hold.

Finally, recalling the definition of μ_0 in Property 13-(i) and $\mu_{\text{s.c.}}$ in eq. (37), we denote

$$\mu_t \coloneqq \mu_0 \boxplus \mu_{\text{s.c.},\sqrt{t}}, \tag{45}$$

where $\mu_{\text{s.c.},\sqrt{t}}(x) \coloneqq t^{-1/2}\mu_{\text{s.c.}}(x/\sqrt{t})$.

A.2.2. THE RECTANGULAR CASE: ASSUMPTION 2.4

We directly get the following from Proposition 13 and Remark 5.

Proposition 16 *Assume that $S \in \mathbb{R}^{d\times L}$ with $S \sim P_0$, satisfying Assumption 2.4 (recall in particular that $L \geq d$ and that $d/L \to \beta \leq 1$). Then:*

(i) *There exists $M, c > 0$ such that*

$$P_0(\{\sigma_i\}_{i=1}^d \subset [0, M]) \geq 1 - e^{-cd},$$

where $\{\sigma_i\}_{i=1}^d$ are the singular values of S.

(ii) *For $S \in \mathbb{R}^{d\times L}$, letting $\{\sigma_i\}_{i=1}^d$ be its singular values, we denote*

$$\hat{\mu}_S(x) \coloneqq \frac{1}{2d}\sum_{i=1}^{d}[\delta_{\sigma_i} + \delta_{-\sigma_i}] \tag{46}$$

the symmetrized *empirical singular value distribution of S. Then there exists $\hat{\mu}_0$, a probability measure with compact support, such that, for $S \sim P_0$:*

$$\lim_{d\to\infty} W_2(\hat{\mu}_S, \hat{\mu}_0) = 0, \ a.s..$$

Moreover, we also have

$$\rho \coloneqq \sqrt{\beta} \lim_{d\to\infty} \mathbb{E}_{S\sim P_0}[(\sigma_1)^2] = \sqrt{\beta}\int \hat{\mu}_0(\mathrm{d}x)\, x^2,$$

and

$$\lim_{d\to\infty} \mathbb{E}[|\sigma_1|^p] < +\infty$$

for any $p \geq 1$.

Remark 17 *As the counterpart to Remark 2, Lemma 50 shows that*

$$\psi_{P_0}^{\mathrm{rec}}(r) = \lim_{d\to\infty} \frac{1}{dL}\mathbb{E}_{Y'} \log \int P_0(\mathrm{d}S) e^{-\frac{1}{2}\sqrt{dL}r\mathrm{Tr}[(S)^T S] + \sqrt{dL}r\mathrm{Tr}[(Y')^T S]}, \tag{47}$$

where $Y' \coloneqq \sqrt{r}S^\star + Z'$ with $S^\star \sim P_0$ and $Z' \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/\sqrt{dL})$. The right-hand side of eq. (47) can be regarded as the definition of $\psi_{P_0}^{\mathrm{rec}}$ as well. Similarly to the symmetric setting, it is a natural generalization of the free entropy of the scalar denoising problem in Barbier et al. (2019) to the problem of rectangular matrix denoising Troiani et al. (2022).

Additive free convolution: rectangular case – We refer to Benaych-Georges (2009) for the general definition of free convolution for rectangular matrices. Similarly to the symmetric setting, we will consider the matrix $S + \sqrt{t}Z$, where $S, Z \in \mathbb{R}^{d\times L}$, S satisfies Assumption 2.4 and $Z_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/\sqrt{dL})$, independently of S. S has a well-defined limit symmetrized singular value distribution $\hat{\mu}_0$ (see Proposition 16), and Z also admits a well-defined limit symmetrized singular value distribution $\hat{\mu}_{\mathrm{MP,rec}}$, where the latter is closely related to μ_{MP} of eq. (38), since $Z^\top Z$ is (up to a global scaling) a Wishart matrix. Generalizing eq. (45), we denote the limit symmetrized singular value distribution of $S + \sqrt{t}Z$ as $\widetilde{\mu}_t$, which is given by:

$$\widetilde{\mu}_t \coloneqq \hat{\mu}_0 \boxplus_\beta \hat{\mu}_{\mathrm{MP,rec},\sqrt{t}}, \tag{48}$$

where $\hat{\mu}_{\mathrm{MP,rec},\sqrt{t}}(x) \coloneqq t^{-1/2}\hat{\mu}_{\mathrm{MP,rec}}(x/\sqrt{t})$, and $\boxplus_\beta$ denotes the rectangular free convolution Benaych-Georges (2009).

Appendix B. Universality: proof of Lemma 9

In this section we prove Lemma 9. Recall that Lemma 9 relies on Assumptions 2.2, 3.1, and 3.2. Consequently, all lemmas in this section are also based on these assumptions. Recall that

$$\ell(x, X, Z) := -\log \int \mathrm{d}P_A(a) \exp\left(-\frac{(\varphi(x,a)-\varphi(X,A))^2 - 2Z(\varphi(x,a)-\varphi(X,A))}{2\Delta}\right).$$

One can then check easily that

$$|\ell(x, X, Z)| \leq \frac{2\sup|\varphi|^2 + 2|Z|\sup|\varphi|}{\Delta}, \tag{49}$$

and

$$\begin{aligned}\left|\frac{\partial \ell(x,X,Z)}{\partial x}\right| &\leq \frac{\int \mathrm{d}P_A(a)(4\sup|\varphi|\sup|\varphi'| + 2|Z|\sup|\varphi'|)\exp\left(\frac{4|\sup|\varphi|^2+4|Z|\sup|\varphi|}{2\Delta}\right)}{\int \mathrm{d}P_A(a)\exp\left(-\frac{4|\sup|\varphi|^2+4|Z|\sup|\varphi|}{2\Delta}\right)} \\ &= (4\sup|\varphi|\sup|\varphi'| + 2|Z|\sup|\varphi'|)\exp\left(\frac{4|\sup|\varphi|^2+4|Z|\sup|\varphi|}{\Delta}\right),\end{aligned} \tag{50}$$

where φ' denote the derivative of $\varphi(x, a)$ w.r.t. its first argument. Eqs. (49),(50) imply that, for $X \in \{\Phi, G\}$:

$$\begin{cases} |\ell(X_\mu)| & \leq C_1 + C_2|Z_\mu|, \\ |\partial_1\ell(X_\mu)|, |\partial_2\ell(X_\mu)| & \leq C_3 e^{C_4|Z_\mu|}, \end{cases} \tag{51}$$

where we use the shorthand $\ell(X_\mu) := \ell(\mathrm{Tr}[X_\mu S], \mathrm{Tr}[X_\mu S^\star], Z_\mu)$ and $\partial_1\ell(X_\mu), \partial_2\ell(X_\mu)$ being the derivatives of ℓ w.r.t. its first and second arguments. C_1, C_2, C_3, C_4 are nonnegative constants that only depend on φ and Δ.

More specifically, we will prove that

$$\lim_{d\to\infty} \sup_{\{A_\mu\}_{\mu=1}^n} \sup_{S^\star \in B_{\mathrm{op}}(M)} |\mathbb{E}_{\Phi,Z}[\psi(F_d(\Phi))] - \mathbb{E}_{G,Z}[\psi(F_d(G))]| = 0, \tag{52}$$

which implies eq. (22) by the dominated convergence theorem.

We will consider the following interpolation model, for $t \in [0, \pi/2]$:

$$\begin{cases} U_\mu(t) & := \cos(t)\Phi_\mu + \sin(t)G_\mu, \\ \widetilde{U}_\mu(t) & := -\sin(t)\Phi_\mu + \cos(t)G_\mu. \end{cases}$$

It is worth noticing that the first and second moments of $U_\mu, \widetilde{U}_\mu$ match the ones of $\mathrm{GOE}(d)$ by Assumption 2.2. We start with a first technical lemma.

Lemma 18 *Denote $\Theta^{(1)}$ to be the set of all the variables $\{\{G_\mu, \Phi_\mu\}_{\mu=2}^n, \{A_\mu\}_{\mu=1}^n, S^\star \in B_{\mathrm{op}}(M)\}$. Then*

$$\sup_{d\geq 1} \sup_{t\in[0,\pi/2]} \sup_{\Theta^{(1)}} \mathbb{E}_{G_1,\Phi_1,Z_1}\left[\frac{\langle e^{-\ell(U_1)}|\mathrm{Tr}[S\widetilde{U}_1(t)]\partial_1\ell(U_1(t)) + \mathrm{Tr}[S^\star\widetilde{U}_1(t)]\partial_2\ell(U_1(t))|\rangle_1}{\langle e^{-\ell(U_1)}\rangle_1}\right] < +\infty,$$

where we denoted

$$\langle\cdot\rangle_\mu := \frac{\int P_0(\mathrm{d}S)e^{-\sum_{\nu\neq\mu}\ell(U_\nu(t))}(\cdot)}{\int P_0(\mathrm{d}S)e^{-\sum_{\nu\neq\mu}^n\ell(U_\nu(t))}}.$$

Proof [Proof of Lemma 18] We have:

$$
\begin{aligned}
&\mathbb{E}_{G_1,\Phi_1,Z_1} \frac{\langle e^{-\ell(U_1)} |\mathrm{Tr}[S\widetilde{U}_1(t)]\partial_1\ell(U_1(t)) + \mathrm{Tr}[S^\star \widetilde{U}_1(t)]\partial_2\ell(U_1(t))|\rangle_1}{\langle e^{-\ell(U_1)}\rangle_1} \\
&\leq \mathbb{E}_{Z_1} C_3 e^{C_1+(C_2+C_4)|Z_1|} \mathbb{E}_{G_1,\Phi_1} \langle (|\mathrm{Tr}[S\widetilde{U}_1(t)]| + |\mathrm{Tr}[S^\star \widetilde{U}_1(t)]|\rangle_1 \\
&\leq C\langle (\mathbb{E}_{G_1,\Phi_1}[\mathrm{Tr}[S\widetilde{U}_1(t)]]^2)^{1/2} + (\mathbb{E}_{G_1,\Phi_1}[\mathrm{Tr}[S^\star \widetilde{U}_1(t)]]^2)^{1/2}\rangle_1 \\
&= C\langle (\mathrm{Tr}[S^2]/d)^{1/2} + (\mathrm{Tr}[(S^\star)^2]/d)^{1/2}\rangle_1 \\
&\leq CM,
\end{aligned}
$$

where C only depends on φ, Δ. We used eq. (51) for the first inequality, the Cauchy-Schwarz inequality, and the Jensen inequality for the second inequality, Assumption 2.2 for the third equality, and Assumption 3.1 for the last inequality. ■

We will need the following domination lemma to control eq. (24).

Lemma 19 (Domination)

$$
\int_0^{\pi/2} \sup_{d\geq 1} \sup_{\{A_\mu\}_{\mu=1}^n} \sup_{S^\star \in B_{\mathrm{op}}(M)} \left| \mathbb{E}_{G,W,Z} \frac{\partial \psi(F_d(U(t)))}{\partial t} \right| \mathrm{d}t < \infty.
$$

Proof [Proof of Lemma 19] By definition, we have

$$
\begin{aligned}
&\frac{\partial \psi(F_d(U(t)))}{\partial t} = -\frac{\psi'(F_d(U(t)))}{d^2} \times \\
&\sum_{\mu=1}^n \frac{\int P_0(\mathrm{d}S) e^{-\sum_{\nu=1}^n \ell(U_\nu(t))} (\mathrm{Tr}[S\widetilde{U}_\mu(t)]\partial_1\ell(U_\mu(t)) + \mathrm{Tr}[S^\star \widetilde{U}_\mu(t)]\partial_2\ell(U_\mu(t)))}{\int P_0(\mathrm{d}S) e^{-\sum_{\nu=1}^n \ell(U_\nu(t))}},
\end{aligned}
$$

which gives

$$
\begin{aligned}
&\mathbb{E}_{G,\Phi,Z}\left[\frac{\partial \psi(F_d(U(t)))}{\partial t}\right] \\
&= -\frac{n}{d^2}\mathbb{E}\left[\psi'(F_d(U(t))) \frac{\langle e^{-\ell(U_1)} (\mathrm{Tr}[S\widetilde{U}_1(t)]\partial_1\ell(U_1(t)) + \mathrm{Tr}[S^\star \widetilde{U}_1(t)]\partial_2\ell(U_1(t)))\rangle_1}{\langle e^{-\ell(U_1)}\rangle_1}\right].
\end{aligned}
$$

Then we have

$$
\begin{aligned}
&\sup_{d\geq 1} \sup_{t\in[0,\pi/2]} \sup_{\{A_\mu\}_{\mu=1}^n} \sup_{S^\star \in B_{\mathrm{op}}(M)} \left| \mathbb{E}_{G,\Phi,Z}\left[\frac{\partial \psi(F_d(U(t)))}{\partial t}\right]\right| \\
&\leq \sup_{d\geq 1} \sup_{t\in[0,\pi/2]} \sup_{\{A_\mu\}_{\mu=1}^n} \sup_{S^\star \in B_{\mathrm{op}}(M)} \frac{n}{d^2} ||\psi'||_\infty \\
&\qquad \times \mathbb{E}_{G,\Phi,Z} \frac{\langle e^{-\ell(U_1)} |\mathrm{Tr}[S\widetilde{U}_1(t)]\partial_1\ell(U_1(t)) + \mathrm{Tr}[S^\star \widetilde{U}_1(t)]\partial_2\ell(U_1(t))|\rangle_1}{\langle e^{-\ell(U_1)}\rangle_1} < +\infty,
\end{aligned}
$$

which finishes the proof. The last inequality follows from Lemma 18. ■

The following lemma allows us to control the expectation of the derivative of the free entropy along the interpolation path. It is a straightforward combination of (Maillard and Bandeira, 2023, Lemma D.3) and (Montanari and Saeed, 2022, Lemma 2)[5], and forms the core of the interpolation argument, where Assumption 2.2 is primarily used. We refer to the two works Montanari and Saeed (2022); Maillard and Bandeira (2023) for more details.

Lemma 20 *Recall that we denote $\Theta^{(1)}$ to be the set of all the variables $\{\{G_\mu, \Phi_\mu\}_{\mu=2}^n, \{A_\mu\}_{\mu=1}^n, S^\star \in B_{\rm op}(M)\}$. Then:*

$$\lim_{d\to\infty} \sup_{\Theta^{(1)}} \left\langle \mathbb{E}_{G_1,\Phi_1,Z_1} \frac{e^{-\ell(U_1)}(\mathrm{Tr}[S\widetilde{U}_1(t)]\partial_1 \ell(U_1(t)) + \mathrm{Tr}[S^\star \widetilde{U}_1(t)]\partial_2 \ell(U_1(t)))}{\langle e^{-\ell(U_1)}\rangle_1} \right\rangle_1 = 0.$$

Proof [Proof of Lemma 20] Following (Maillard and Bandeira, 2023, Lemma D.3) and (Montanari and Saeed, 2022, Lemma 2), we have:

$$\begin{aligned} &\lim_{d\to\infty} \sup_{\Theta^{(1)}} \left\langle \mathbb{E}_{G_1,\Phi_1,Z_1} \frac{e^{-\ell(U_1)}(\mathrm{Tr}[S\widetilde{U}_1(t)]\partial_1 \ell(U_1(t)) + \mathrm{Tr}[S^\star \widetilde{U}_1(t)]\partial_2 \ell(U_1(t)))}{\langle e^{-\ell(U_1)}\rangle_1} \right\rangle_1 \qquad (53) \\ &= \lim_{d\to\infty} \sup_{\Theta^{(1)}} \left\langle \mathbb{E}_{G_1,\tilde{G}_1,Z_1} \frac{e^{-\ell(V_1)}(\mathrm{Tr}[S\widetilde{V}_1(t)]\partial_1 \ell(V_1(t)) + \mathrm{Tr}[S^\star \widetilde{V}_1(t)]\partial_2 \ell(V_1(t)))}{\langle e^{-\ell(V_1)}\rangle_1} \right\rangle_1 = 0, \end{aligned}$$

where $V_1 := \cos(t)\tilde{G}_1 + \sin(t)G_1, \widetilde{V}_1 := -\sin(t)\tilde{G}_1 + \cos(t)G_1$ with $\tilde{G}_1 \sim \mathrm{GOE(d)}$ independent of G_1. Informally, this is a consequence of Assumption 2.2, which allows to replace in the left-hand side of eq. (53) the matrix Φ_1 by a $\mathrm{GOE}(d)$ matrix $\widetilde{G}_1$. In detail, the proof follows directly from the arguments of (Montanari and Saeed, 2022, Lemma 2), to condition on the event $|Z_1| \le B$ for an arbitrary $B > 0$, and then directly using the proof arguments of (Maillard and Bandeira, 2023, Lemma D.3) (see also (Montanari and Saeed, 2022, Lemma 3)) under this event. Taking the limit $B \to \infty$ allows then to end the proof. The last equality uses that $\mathbb{E}[\mathrm{Tr}[S\widetilde{V}_1(t)]] = \mathbb{E}[\mathrm{Tr}[S^\star \widetilde{V}_1(t)]] = 0$ since $V_1(t)$ and $\widetilde{V}_1(t)$ are now *independent* $\mathrm{GOE}(d)$ matrices. ■

A direct consequence of Lemma 20 is a control of the limit of the time derivative in eq. (24).

Lemma 21

$$\lim_{d\to\infty} \sup_{\{A_\mu\}_{\mu=1}^n} \sup_{S^\star \in B_{\rm op}(M)} \mathbb{E}_{G,\Phi,Z} \frac{\partial \psi(F_d(U(t)))}{\partial t} = 0.$$

Proof [Proof of Lemma 21] We have

$$\left| \mathbb{E}_{G,\Phi,Z} \frac{\partial \psi(F_d(U(t)))}{\partial t} \right| \le \frac{n}{d^2}(I_1 + I_2),$$

5. (Maillard and Bandeira, 2023, Lemma D.3) deals with rotationally-invariant matrix priors like us, but in a non-planted model, while (Montanari and Saeed, 2022, Lemma 3) deals with the additional planted signal and randomness. Note that our loss function ℓ is locally Lipschitz, satisfying the assumption used in Montanari and Saeed (2022).

The first term is

$$I_1 := \left|\mathbb{E}_{G,\Phi,Z}\left[(\psi'(F_d(U)) - \psi'(F_d(U^{(1)})))\times \left(\frac{\langle e^{-\ell(U_1)}(\mathrm{Tr}[S\widetilde{U}_1(t)]\partial_1\ell(U_1(t)) + \mathrm{Tr}[S^\star\widetilde{U}_1(t)]\partial_2\ell(U_1(t)))\rangle_1}{\langle e^{-\ell(U_1)}\rangle_1}\right)\right]\right|,$$

where $U^{(1)}$ denotes $\{\circ, U_2, \cdots, U_n\}$, the symbol $\circ$ denoting that we remove the terms corresponding to U_1 and Z_1 in eq. (20). Concretely:

$$F_d(U) - F_d(U^{(1)}) = \frac{1}{d^2}\log\left\langle e^{-\ell(\mathrm{Tr}[G_1S],\mathrm{Tr}[G_1S^\star],Z_1)}\right\rangle_1 - \frac{Z_1^2}{2\Delta d^2}.$$

Repeating the proof of the bound of Lemma 18 with the additional upper bound (which uses again eq. (51)):

$$\begin{aligned}|\psi'(F_d(U)) - \psi'(F_d(U^{(1)}))| &\le \frac{||\psi''||_\infty}{d^2}\left|\log\left\langle e^{-\ell(\mathrm{Tr}[G_1S],\mathrm{Tr}[G_1S^\star],Z_1)}\right\rangle_1 + \frac{Z_1^2}{2\Delta}\right|, \\ &\le \frac{||\psi''||_\infty}{d^2}\left(e^{C_1+C_2|Z_1|} + \frac{Z_1^2}{2\Delta}\right),\end{aligned}$$

we reach $I_1 \to 0$, uniformly in $S^\star$, $\{A_\mu\}_{\mu=1}^n$. The second term is (since $F_d(U^{(1)})$ is now independent of G_1, Φ_1, Z_1):

$$\begin{aligned}I_2 &:= \left|\mathbb{E}_{G,\Phi,Z}\left[\psi'(F_d(U^{(1)}))\left(\frac{\langle e^{-\ell(U_1)}(\mathrm{Tr}[S\widetilde{U}_1(t)]\partial_1\ell(U_1(t)) + \mathrm{Tr}[S^\star\widetilde{U}_1(t)]\partial_2\ell(U_1(t)))\rangle_1}{\langle e^{-\ell(U_1)}\rangle_1}\right)\right]\right|, \\ &\le \mathbb{E}_{\{G_\mu,\Phi_\mu,Z_\mu\}_{\mu=2}^n}\left[|\psi'(F_d(U^{(1)}))|\left|\mathbb{E}_{G_1,\Phi_1,Z_1}\left(\frac{\langle e^{-\ell(U_1)}(\mathrm{Tr}[S\widetilde{U}_1(t)]\partial_1\ell(U_1(t)) + \mathrm{Tr}[S^\star\widetilde{U}_1(t)]\partial_2\ell(U_1(t)))\rangle_1}{\langle e^{-\ell(U_1)}\rangle_1}\right)\right|\right], \\ &\to 0\end{aligned}$$

by Lemma 20, where the limit is also uniform in $S^\star$, $\{A_\mu\}_{\mu=1}^n$. This finishes the proof of Lemma 21. ■

Finally, as

$$|\mathbb{E}_{\Phi,Z}[\psi(F(\Phi))] - \mathbb{E}_{G,Z}[\psi(F(G))]| \le \int_0^{\pi/2}\left|\mathbb{E}_{G,\Phi,Z}\frac{\partial\psi(F(U(t)))}{\partial t}\right|\mathrm{d}t,$$

we prove eq. (23) by combining Lemma 19, Lemma 21, and using the dominated convergence theorem. This ends our proof of Lemma 9.

Appendix C. Proof of Theorem 1

In this section we prove Theorem 1. As mentioned in the main text, our proof uses an adaptive interpolation argument, as we generalize the proof approach of Barbier et al. (2019) to our more general setting of structured matrix priors. Nevertheless, a significant portion of our proof arguments follow from direct transpositions of the arguments of Barbier et al. (2019), and we will refer to them when necessary.

C.1. Proof of Theorem 1 via the adaptive interpolation method

We will first prove Theorem 1 under Assumptions 3.1 and 3.2 and for Gaussian data $\{G_\mu\}_{\mu=1}^n \overset{\text{i.i.d.}}{\sim}$ $\mathrm{GOE}(d)$. All lemmas in Sections C.1, C.2 and C.3 are under such assumptions. Thanks to Lemma 9, Gaussian data can be replaced by more general data in the end. Assumptions 3.1 and 3.2 can then be relaxed as detailed in Section C.4, by leveraging Lemmas 32 and 33.

Let us consider the following interpolation model

$$\begin{cases} Y_{t,\mu} \sim P_{\text{out}}(\cdot|J_{t,\mu}), \ \mu = 1, \cdots, n \\ Y'_t = \sqrt{d}\left(\sqrt{R_1(t)}S^\star + Z'\right) \in \mathcal{S}_d, \end{cases} \tag{54}$$

with

$$J_{t,\mu} \coloneqq \sqrt{1-t}\mathrm{Tr}[G_\mu S^\star] + \sqrt{2R_2(t)}V_\mu + \sqrt{2\rho t - 2R_2(t) + 2\iota_d}U^\star_\mu,$$

$Z' \sim \mathrm{GOE}(d)$ and $\{V_\mu, U^\star_\mu\}_{\mu=1}^n \overset{iid}{\sim} \mathcal{N}(0,1)$. The problem is to recover $S^\star, U^\star$ from the observations $\{Y_{t,\mu}\}_{\mu=1}^n, Y'_t$ given the knowledge of $\{G_\mu\}_{\mu=1}^n$ and $\{V_\mu\}_{\mu=1}^n$. Thus the $\sqrt{d}$ scaling in (54) is only for simplicity and will not influence the free entropy. The interpolating functions are given by

$$R_1(t) \coloneqq \epsilon_1 + \int_0^t r(v)\mathrm{d}v, \ R_2(t) \coloneqq \epsilon_2 + \int_0^t q(v)\mathrm{d}v,$$

where $(\epsilon_1, \epsilon_2) \in B_d \coloneqq [\iota_d, 2\iota_d] \otimes [\iota_d, 2\iota_d]$ are two small quantities and $q : [0,1] \to [0,\rho]$, $r : [0,1] \to [0, r_{\max}]$ are two continuous functions (that might depend on $\epsilon \coloneqq (\epsilon_1, \epsilon_2)$ as well) with $r_{\max} \coloneqq 4\alpha \sup_{q\in[0,\rho]} \Psi'_{\text{out}}(q) = 4\alpha\Psi'_{\text{out}}(\rho)$ (by the convexity of Ψ_{out} shown in Lemma 51). We will choose $\iota_d \to 0$ as $d \to \infty$ slowly enough, it remains arbitrary for now.

The interpolating free entropy reads

$$f_{d,\epsilon}(t) \coloneqq \frac{1}{d^2}\mathbb{E}\log \mathcal{Z}_{t,\epsilon}(Y, Y', G, V) \coloneqq \frac{1}{d^2}\mathbb{E}\log \int P_0(\mathrm{d}s)\mathcal{D}u\, e^{-H_{t,\epsilon}(s,u,Y,Y',G,V)},$$

where $\mathcal{D}u \coloneqq (2\pi)^{-n/2}e^{-\sum_{\mu=1}^n u_\mu^2/2}\mathrm{d}u$ is the standard Gaussian measure, and

$$H_{t,\epsilon}(s,u,Y,Y',G,V) \coloneqq \frac{1}{2}\sum_{\mu=1}^n u_{Y_{t,\mu}}(j_{t,\mu}) + \frac{1}{4}\sum_{i,j=1}^d (Y'_{ij} - \sqrt{dR_1(t)}s_{ij})^2$$

is the Hamitonian. Moreover, we have defined

$$u_{Y_{t,\mu}}(j_t) \coloneqq \log P_{\text{out}}(Y_{t,\mu}|j_{t,\mu})$$

with $j_{t,\mu} \coloneqq \sqrt{1-t}\mathrm{Tr}[G_\mu s] + \sqrt{2R_2(t)}V_\mu + \sqrt{2\rho t - 2R_2(t) + 2\iota_d}u_\mu$. Accordingly, the Gibbs bracket is defined as

$$\langle g(s,u)\rangle \coloneqq \frac{1}{\mathcal{Z}_{t,\epsilon}(Y, Y', G, V)}\int P_0(\mathrm{d}s)\mathcal{D}u\, g(s,u)e^{-H_{t,\epsilon}(s,u,Y,Y',G,V)}.$$

The following lemma connects the interpolation model to the original model.

Lemma 22

$$f_{d,\epsilon}(0) = f_d - \frac{1}{4} + O(\iota_d),$$
$$f_{d,\epsilon}(1) = \psi_{P_0}\left(\int_0^1 r(t)\mathrm{d}t\right) + \alpha\Psi_{\mathrm{out}}\left(\int_0^1 q(t)\mathrm{d}t\right) + O(\iota_d).$$

Proof [Proof of Lemma 22] We have

$$\left|\frac{\mathrm{d}f_{d,\epsilon}(0)}{\mathrm{d}\epsilon_1}\right| = \frac{1}{4d}\left|\mathbb{E}\sum_{i,j=1}^{d}\langle\epsilon_1(S_{ij}-s_{ij}) + \sqrt{\frac{1}{\epsilon_1}}(S_{ij}-s_{ij})Z'_{ij}\rangle\right|$$
$$= \frac{1}{2d}\left|\mathbb{E}\sum_{i,j=1}((S^\star_{ij})^2 - \langle s_{ij}\rangle^2)\right| \le M^2$$

where we use the Nishimori identity (Proposition 43) and the fact that $\sum_{i,j=1}^d (S^\star_{ij})^2 = \mathrm{Tr}[(S^\star)^2] \le dM^2$, and the same for $\langle s\rangle$ (according to Assumption 3.1). Moreover,

$$\left|\frac{df_{d,\epsilon}(0)}{d\epsilon_2}\right| = \frac{1}{2d^2}\sum_{\mu=1}^{n}|\mathbb{E}[u'_{Y_{0,\mu}}(J_{t,\mu})\langle u'_{Y_{0,\mu}}(j_{t,\mu})\rangle]| \le C(\varphi, M, \alpha),$$

where the inequality follows from Assumption 3.2 (see (Barbier et al., 2019, Section A.6)) and $C(\varphi, M, \alpha)$ is a generic non-negative constant depending only on φ, M and α. Therefore, we have $|f_{d,\epsilon}(0) - f_{d,0}(0)| \le C(\varphi, M, \alpha)\iota_d$. As $f_{d,0}(0) = f_d - \frac{1}{4}$, we obtain the first equality of Lemma 22. The second equality is obtained through

$$f_{d,\epsilon}(1) = \psi_{P_0}(R_1(1)) + \alpha\Psi_{\mathrm{out}}(R_2(1) + \iota_d)$$

and using the Lipschitz property of ψ_{P_0} and Ψ_{out} (Lemmas 49, 51). ■

Two important concentration lemmas are presented below, where we use $C(\varphi, M, \alpha, \kappa)$ for a generic non-negative constant dependent only on φ, M, α, and κ.

Lemma 23

$$\mathbb{E}\left[\left(\frac{1}{d^2}\log\mathcal{Z}_{t,\epsilon} - f_{d,\epsilon}(t)\right)^2\right] \le \frac{C(\varphi, M, \alpha, \kappa)}{d^2}.$$

Lemma 23 is proved in Section C.2.

Lemma 24

$$\frac{1}{\iota_d^2}\int_{B_d}\mathrm{d}\epsilon\int_0^1 \mathrm{d}t\mathbb{E}\langle(Q - \mathbb{E}\langle Q\rangle)^2\rangle \le C(\varphi, M, \alpha, \kappa)\cdot o_d(1),$$

where $Q := \mathrm{tr}[sS^\star] = (1/d)\sum_{i,j=1}^d s_{ij}S^\star_{ij}$.

Lemma 24 is proven in Section C.3. We now prove the following result on the derivative of the free entropy along the interpolation path:

Lemma 25 (Derivative of $f_{d,\epsilon}(t)$) *We have, uniformly in $t \in [0,1]$:*

$$\frac{\mathrm{d}f_{d,\epsilon}(t)}{\mathrm{d}t} = -\frac{1}{2}\mathbb{E}\left\langle \left(\frac{1}{d^2}\sum_{\mu=1}^{n} u'_{Y_{t,\mu}}(J_{t,\mu})u'_{Y_{t,\mu}}(j_{t,\mu}) - r(t)\right)(Q - q(t))\right\rangle + \frac{r(t)}{4}(q(t)-\rho) + o_d(1). \tag{55}$$

Proof [Proof of Lemma 25] By definition, we have

$$\frac{\mathrm{d}f_{d,\epsilon}(t)}{\mathrm{d}t} = -\frac{1}{d^2}\mathbb{E}[H'_{t,\epsilon}\log \mathcal{Z}_{t,\epsilon}] - \frac{1}{d^2}\mathbb{E}\langle H'_{t,\epsilon}\rangle.$$

Following (Barbier et al., 2019, Section A.5.1)[6], we have

$$\frac{\mathrm{d}f_{d,\epsilon}(t)}{\mathrm{d}t} = \mathbb{E}\left[\frac{1}{d^2}\frac{P''_{\mathrm{out}}(Y_{t,\mu}|J_{t,\mu})}{P_{\mathrm{out}}(Y_{t,\mu}|J_{t,\mu})}\left(\frac{1}{d}\sum_{i,j=1}^{d}(S^\star_{ij})^2 - \rho\right)\log \mathcal{Z}_{t,\epsilon}\right] + \frac{r(t)}{4}(q(t)-\rho)$$
$$+ \mathbb{E}\left\langle\left(\frac{1}{d}\sum_{i,j=1}^{d} S^\star_{ij}s_{ij} - q(t)\right)\left(u'_{Y_{t,\mu}}(J_{t,\mu})u'_{Y_{t,\mu}}(j_{t,\mu}) - \frac{r(t)}{4}\right)\right\rangle + o_d(1).$$

The $o_d(1)$ term is uniform in t. Now we denote the first term to be $A_{d,\epsilon}$ and we only need to show that it goes to zero uniformly as $d \to \infty$. We have

$$\mathbb{E}\left[\frac{P''_{\mathrm{out}}(Y_{t,\mu}|J_{t,\mu})}{P_{\mathrm{out}}(Y_{t,\mu}|J_{t,\mu})}\middle| S^\star, J_t\right] = \int \mathrm{d}Y_{t,\mu}P''_{\mathrm{out}}(Y_{t,\mu}|J_{t,\mu}) = 0, \tag{56}$$

since $P_{\mathrm{out}}(Y|J)$ is a probability distribution, and thus by the law of total expectation:

$$\mathbb{E}\left[\frac{1}{d^2}\frac{P''_{\mathrm{out}}(Y_{t,\mu}|J_{t,\mu})}{P_{\mathrm{out}}(Y_{t,\mu}|J_{t,\mu})}\left(\frac{1}{d}\sum_{i,j=1}^{d}(S^\star_{ij})^2 - \rho\right)\right] = 0.$$

Consequently we have

$$|A_{d,\epsilon}| = \left|\mathbb{E}\left[\sum_{\mu=1}^{n}\frac{P''_{\mathrm{out}}(Y_{t,\mu}|J_{t,\mu})}{P_{\mathrm{out}}(Y_{t,\mu}|J_{t,\mu})}\left(\frac{1}{d}\sum_{i,j=1}^{d}(S^\star_{ij})^2 - \rho\right)\left(\frac{1}{d^2}\log \mathcal{Z}_{t,\epsilon} - f_{d,\epsilon}(t)\right)\right]\right|$$
$$\leq \mathbb{E}\left[\left(\sum_{\mu=1}^{n}\frac{P''_{\mathrm{out}}(Y_{t,\mu}|J_{t,\mu})}{P_{\mathrm{out}}(Y_{t,\mu}|J_{t,\mu})}\right)^2\left(\mathrm{tr}[(S^\star)^2] - \rho\right)^2\right]^{1/2}\mathbb{E}\left[\left(\frac{1}{d^2}\log \mathcal{Z}_{t,\epsilon} - f_{d,\epsilon}(t)\right)^2\right]^{1/2}. \tag{57}$$

6. We use the Gaussian integration by parts property shown in Lemma 47.

Recall that $\mathrm{tr}(S^2) = (1/d)\mathrm{Tr}[S^2]$. As (conditionally on J_t) $\{\frac{P''_{\text{out}}(Y_{t,\mu}|J_{t,\mu})}{P_{\text{out}}(Y_{t,\mu}|J_{t,\mu})}\}_{\mu=1,2,\cdots,n}$ are i.i.d. and centered (see (56)) random variables, we have

$$\mathbb{E}\left[\left(\sum_{\mu=1}^{n}\frac{P''_{\text{out}}(Y_{t,\mu}|J_{t,\mu})}{P_{\text{out}}(Y_{t,\mu}|J_{t,\mu})}\right)^2 \middle| J_{t,\mu}, S^\star\right] \leq C(\varphi)n, \tag{58}$$

where we use the boundedness of $\varphi, \varphi', \varphi''$. Moreover, we have

$$\mathbb{E}\left[\left(\mathrm{tr}[(S^\star)^2] - \rho\right)^2\right] = o(1) \tag{59}$$

as a consequence of Lemma 46-(c). Taking eqs. (58) and (59) into eq. (57), using Lemma 23, we have

$$|A_{d,\epsilon}| \leq C(\varphi, M, \alpha, \kappa) \cdot o(1),$$

which goes to zero uniformly in t. ■

The combination of the following two lemmas give us a lower bound on the free entropy of the original problem.

Lemma 26 *If $q(t) = \mathbb{E}\langle Q\rangle$, then*

$$f_d = \frac{1}{\iota_d^2}\int_{B_d} \mathrm{d}\epsilon \left(\psi_{P_0}\left(\int_0^1 r(t)\mathrm{d}t\right) + \alpha\Psi_{\text{out}}\left(\int_0^1 q(t)\mathrm{d}t\right) - \frac{1}{4}\int_0^1 r(t)(\rho - q(t))\mathrm{d}t\right) \\ + \frac{1}{4} + o_d(1).$$

Proof [Proof of Lemma 26] We begin from controlling the first term in eq. (55). We have

$$\left(\frac{1}{\iota_d^2}\int_{B_d}\mathrm{d}\epsilon\int_0^1 \mathrm{d}t\mathbb{E}\left\langle\left(\frac{1}{d^2}\sum_{\mu=1}^{n} u'_{Y_{t,\mu}}(J_{t,\mu})u'_{Y_{t,\mu}}(j_{t,\mu}) - r(t)\right)(Q - q(t))\right\rangle\right)^2$$
$$\leq \left(\frac{1}{\iota_d^2}\int_{B_d}\mathrm{d}\epsilon\int_0^1 \mathrm{d}t\mathbb{E}\left\langle\left(\frac{1}{d^2}\sum_{\mu=1}^{n} u'_{Y_{t,\mu}}(J_{t,\mu})u'_{Y_{t,\mu}}(j_{t,\mu}) - r(t)\right)^2\right\rangle\right) \times$$
$$\left(\frac{1}{\iota_d^2}\int_{B_d}\mathrm{d}\epsilon\int_0^1 \mathrm{d}t\mathbb{E}\left\langle(Q - q(t))^2\right\rangle\right),$$

where we use the Cauchy-Schwarz inequality. By (Barbier et al., 2019, Section A.6) we have

$$\mathbb{E}\left\langle\left(\frac{1}{d^2}\sum_{\mu=1}^{n} u'_{Y_{t,\mu}}(J_{t,\mu})u'_{Y_{t,\mu}}(j_{t,\mu}) - r(t)\right)^2\right\rangle \leq C(\alpha, \varphi).$$

Then by Lemma 24, we have

$$\left|\frac{1}{\iota_d^2}\int_{B_d}\mathrm{d}\epsilon\int_0^1 \mathrm{d}t\mathbb{E}\left\langle\left(\frac{1}{d^2}\sum_{\mu=1}^{n} u'_{Y_{t,\mu}}(J_{t,\mu})u'_{Y_{t,\mu}}(j_{t,\mu}) - r(t)\right)(Q - \mathbb{E}\langle Q\rangle)\right\rangle\right| = o_d(1).$$

Using now Lemma 25, we reach

$$\frac{1}{\iota_d^2}\int_{B_d} \mathrm{d}\epsilon \int_0^1 \mathrm{d}t \frac{\mathrm{d}f_{d,\epsilon}(t)}{\mathrm{d}t} = \frac{1}{4}\int_0^1 r(t)(q(t)-\rho)\mathrm{d}t + o_d(1).$$

We finish the proof by combining this result with Lemma 22. ■

We can now state the aforementioned lower bound.

Lemma 27

$$\liminf_{d\to\infty} f_d \geq \sup_{r\geq 0}\inf_{q\in[0,\rho]} f_{\mathrm{RS}}(q,r).$$

Proof [Proof of Lemma 27] Following (Barbier et al., 2019, Proposition 7), we can choose $r(t) = r$ and $q(t) = \mathbb{E}\langle Q\rangle$. Then Lemma 26 gives

$$\liminf_{d\to\infty} f_d = \liminf_{d\to\infty} \frac{1}{\iota_d^2}\int_{B_d} \mathrm{d}\epsilon f_{\mathrm{RS}}\left(\int_0^1 q(t)\mathrm{d}t, r\right) \geq \inf_{q\in[0,\rho]} f_{\mathrm{RS}}(q,r),$$

and thus (since this holds for any $r \geq 0$)

$$\liminf_{d\to\infty} f_d \geq \sup_{r\in[0,r_{\max}]}\inf_{q\in[0,\rho]} f_{\mathrm{RS}}(q,r).$$

Recall that we chose $r_{\max} = 4\alpha\Psi'_{\mathrm{out}}(\rho) \geq 4\alpha\Psi'_{\mathrm{out}}(q)$, so for $r \geq r_{\max}$ and $q \in [0,\rho]$ we have $\frac{\partial}{\partial q} f_{\mathrm{RS}}(q,r) = \alpha\Psi'_{\mathrm{out}}(q) - \frac{1}{4}r \leq 0$. Therefore for $r > r_{\max}$,

$$\frac{\partial}{\partial r}\inf_{q\in[0,\rho]} f_{\mathrm{RS}}(q;r) = \frac{\partial}{\partial r} f_{\mathrm{RS}}(\rho;r) = \psi'_{P_0}(r) \leq 0,$$

where we use Lemma 49. Therefore:

$$\liminf_{d\to\infty} f_d \geq \sup_{r\in[0,r_{\max}]}\inf_{q\in[0,\rho]} f_{\mathrm{RS}}(q,r) = \sup_{r\geq 0}\inf_{q\in[0,\rho]} f_{\mathrm{RS}}(q,r).$$

■

The following lemma gives the corresponding upper bound.

Lemma 28

$$\limsup_{d\to\infty} f_d \leq \sup_{r\geq 0}\inf_{q\in[0,\rho]} f_{\mathrm{RS}}(q,r).$$

Proof [Proof of Lemma 28] Following (Barbier et al., 2019, Proposition 8), we can choose $q(t) = \mathbb{E}\langle Q\rangle$ and $r(t) = 4\alpha\Psi'_{\mathrm{out}}(q(t)) \leq r_{\max}$. As in Lemma 27, we have

$$\begin{aligned}\limsup_{d\to\infty} f_d &= \limsup_{d\to\infty} \frac{1}{\iota_d^2}\int_{B_d} \mathrm{d}\epsilon \left(\psi_{P_0}\left(\int_0^1 r(t)\mathrm{d}t\right) + \alpha\Psi_{\mathrm{out}}\left(\int_0^1 q(t)\mathrm{d}t\right)\right. \\ &\qquad \left. -\frac{1}{4}\int_0^1 r(t)(\rho - q(t))\mathrm{d}t\right) + \frac{1}{4}, \\ &\leq \limsup_{d\to\infty} \frac{1}{\iota_d^2}\int_{B_d} \mathrm{d}\epsilon \int_0^1 \mathrm{d}t f_{\mathrm{RS}}(q(t), r(t)),\end{aligned}$$

where we use Jensen's inequality because $\psi_{P_0}, \Psi_{\rm out}$ are convex according to Lemmas 49 and 51. As $\alpha\Psi_{\rm out}(q) - \frac{1}{4}r(t)q$ is convex with respect to q by Lemma 51, and $r(t) = 4\alpha\Psi'_{\rm out}(q(t))$, we have

$$\alpha\Psi_{\rm out}(q(t)) - \frac{1}{4}r(t)q(t) = \inf_{q\in[0,\rho]}\left[\alpha\Psi_{\rm out}(q) - \frac{1}{4}r(t)q\right],$$

which gives

$$\limsup_{d\to\infty} f_d \le \limsup_{d\to\infty} \frac{1}{\iota_d^2}\int_{B_d} \mathrm{d}\epsilon \int_0^1 \mathrm{d}t \inf_{q\in[0,\rho]} f_{\rm RS}(q, r(t)) \le \sup_{r\ge 0}\inf_{q\in[0,\rho]} f_{\rm RS}(q,r).$$

■

Proof [Proof of Theorem 1] We note that $\sup_{r\ge0}\inf_{q\in[0,\rho]} f_{\rm RS}(q,r) = \sup_{q\in[0,\rho]}\inf_{r\ge0} f_{\rm RS}(q,r)$ by (Barbier et al., 2019, Corollary 8), combined with our Lemmas 49 and 51. Combining Lemmas 27 and 28, we have

$$\lim_{d\to\infty} f_d = \sup_{q\in[0,\rho]}\inf_{r\ge0} f_{\rm RS}(q,r), \tag{60}$$

which gives Theorem 1 for Gaussian data, under the Assumptions 3.1 and 3.2.

Lemma 23 together with eq. (60) shows that $F_d(G)$ converges in probability to $\sup_{q\in[0,\rho]}\inf_{r\ge0} f_{\rm RS}(q,r)$ under Assumptions 3.1 and 3.2. Combining it with Lemma 9, it is classical to show that $F_d(\Phi)$ also converges in probability (see e.g. (Montanari and Saeed, 2022, Section A.1.3)) to $\sup_{q\in[0,\rho]}\inf_{r\ge0} f_{\rm RS}(q,r)$. Moreover, we have

$$|F_d(\Phi)| \le \frac{1}{d^2}\sum_{\mu=1}^{n} \frac{(2\sup|\varphi| + |Z_\mu|)^2}{2\Delta},$$

so $F_d(\Phi)$ is uniformly integrable, which gives

$$\lim_{d\to\infty} \mathbb{E}[F_d(\Phi)] = \sup_{q\in[0,\rho]}\inf_{r\ge0} f_{\rm RS}(q,r).$$

Recall that $f_d := \mathbb{E}[F_d(\Phi)]$, and thus we finish the proof under Assumptions 3.1 and 3.2. The arguments in Section C.4 imply then that one can relax Assumptions 3.1 and 3.2 to Assumptions 2.1 and 2.3. ■

C.2. Concentration of the free entropy

In this section we will prove Lemma 23. Recall that we are using Assumptions 3.1, 3.2 and Gaussian data. We can rewrite the free entropy as

$$\frac{1}{d^2}\log \mathcal{Z}_{t,\epsilon} = \frac{1}{d^2}\log \hat{\mathcal{Z}}_{t,\epsilon} - \frac{1}{2d^2}\sum_{\mu=1}^{n} Z_\mu^2 - \frac{1}{4d}\sum_{i,j=1}^{d}(Z'_{ij})^2 - \frac{\alpha}{2}\log(2\pi), \tag{61}$$

where

$$\frac{1}{d^2}\log\hat{\mathcal{Z}}_{t,\epsilon} := \frac{1}{d^2}\log\int P_0(\mathrm{d}s)P_A(\mathrm{d}a)\mathcal{D}u e^{-\hat{H}_t(s,u,a)},$$

$$\hat{H}_t(s,u,a) := \frac{1}{2}\sum_{\mu=1}^{n}(\Gamma_{t,\mu}(s,u_\mu,a_\mu)^2 + 2Z_\mu\Gamma_{t,\mu}(s,u_\mu,a_\mu)) \\ + \frac{d}{4}\sum_{i,j=1}^{d}(R_1(t)(S^\star_{ij}-s_{ij})^2 + 2Z'_{ij}\sqrt{R_1(t)}(S^\star_{ij}-s_{ij}))$$

and

$$\Gamma_{t,\mu}(s,u_\mu,a_\mu) := \varphi(\sqrt{1-t}\mathrm{Tr}[G_\mu S^\star] + k_1(t)V_\mu + k_2(t)U^\star_\mu, A_\mu) \\ - \varphi(\sqrt{1-t}\mathrm{Tr}[G_\mu s] + k_1(t)V_\mu + k_2(t)u_\mu, a_\mu).$$

Here $A_\mu \overset{\text{i.i.d.}}{\sim} P_A$, $k_1(t) := \sqrt{2R_2(t)}$ and $k_2(t) := \sqrt{2\rho t - 2R_2(t) + 2\iota_d}$. Since all the terms but $(1/d^2)\log\hat{\mathcal{Z}}_{t,\epsilon}$ in eq. (61) are easily shown to have variance $\mathcal{O}(1/d^2)$ – as they are sums of i.i.d. Gaussian variables – Lemma 23 results from the following lemma.

Lemma 29

$$\mathit{Var}\left(\frac{1}{d^2}\log\hat{\mathcal{Z}}_{t,\epsilon}\right) \le \frac{C(\varphi,M,\alpha,\kappa)}{d^2}.$$

Proof [Proof of Lemma 29] We first consider $g := \log\hat{\mathcal{Z}}_{t,\epsilon}/d^2$ as a function of Z, Z'. We have

$$\left|\frac{\partial g}{\partial Z_\mu}\right| = d^{-2}|\langle\Gamma_{t,\mu}\rangle| \le 2d^{-2}\sup|\varphi|,$$

and

$$\sum_{i,j=1}^{d}\left(\frac{\partial g}{\partial Z'_{ij}}\right)^2 = d^{-2}R_1(t)\sum_{i,j=1}^{d}(S^\star_{ij}-\langle s_{ij}\rangle)^2 = d^{-2}R_1(t)\mathrm{Tr}(S^\star-\langle s\rangle)^2 \\ \le 2d^{-2}R_1(t)\mathrm{Tr}((S^\star)^2+\langle s\rangle^2) \le 4d^{-1}KM^2,$$

where $K := 1+\max(\rho, r_{\max})$ upper bounds R_1, R_2 and we use $\mathrm{Tr}(S^\star)^2, \mathrm{Tr}\langle s\rangle^2 \le dM^2$, because under Assumption 3.1, the eigenvalues of $S^\star, \langle s\rangle$ are bounded by M. By Lemma 48 we obtain

$$\mathbb{E}\left[\left(\frac{1}{d^2}\log\hat{\mathcal{Z}}_{t,\epsilon} - \frac{1}{d^2}\mathbb{E}_{Z,Z'}\log\hat{\mathcal{Z}}_{t,\epsilon}\right)^2\right] \le \frac{C(\varphi,M,\alpha)}{d^2}. \tag{62}$$

Following (Barbier et al., 2019, Lemma 27), we also have

$$\mathbb{E}\left[\mathbb{E}_{Z,Z'}\left(\frac{1}{d^2}\log\hat{\mathcal{Z}}_{t,\epsilon} - \frac{1}{d^2}\mathbb{E}_{Z,Z',A}\log\hat{\mathcal{Z}}_{t,\epsilon}\right)^2\right] \le \frac{C(\varphi,M,\alpha)}{d^2}$$

by the bounded difference inequality w.r.t. $A = \{A_\mu\}_{\mu=1}^n$. Then we consider $g := \mathbb{E}_{Z,Z',A}\log\hat{\mathcal{Z}}_{t,\epsilon}/d^2$ as a function of $V, U^\star, G$. We have

$$\left|\frac{\partial g}{\partial V_\mu}\right| = d^{-2}\left|\mathbb{E}_{Z,Z',A}\left\langle(\Gamma_{t,\mu}+Z_\mu)\frac{\partial\Gamma_{t,\mu}}{\partial V_\mu}\right\rangle\right| \le d^{-2}\left(2\sup|\varphi| + \sqrt{\frac{2}{\pi}}\right)2\sqrt{K}\sup|\varphi'|.$$

The same inequality holds for $|\frac{\partial g}{\partial U_i^\star}|$. Moreover, we have

$$
\begin{aligned}
\sum_{i,j=1}^{d}\left(\frac{\partial g}{\partial(G_{\mu,ij})}\right)^2 &= d^{-5}\sum_{i,j=1}^{d}\left(\mathbb{E}_{Z,Z',A}\left\langle(\Gamma_{t,\mu}+Z_\mu)\frac{\partial\Gamma_{t,\mu}}{\partial(G_{\mu,ij})}\right\rangle\right)^2 \\
&\le d^{-4}\mathbb{E}_{Z,Z',A}\sum_{i,j=1}^{d}\left\langle(\Gamma_{t,\mu}+Z_\mu)\frac{\partial\Gamma_{t,\mu}}{\partial(G_{\mu,ij})}\right\rangle^2 \\
&\le d^{-4}\mathbb{E}_{Z,Z',A}(2\sup|\varphi|+|Z_\mu|)^2\sum_{i,j=1}^{d}\langle|S^\star_{ij}\varphi'(J_{t,\mu})-s_{ij}\varphi'(j_{t,\mu})|\rangle^2 \\
&\le 2d^{-4}\mathbb{E}_{Z,Z',A}(2\sup|\varphi|+|Z_\mu|)^2|\sup|\varphi'|^2\sum_{i,j=1}^{d}((S^\star_{ij})^2+\langle s_{ij}^2\rangle) \\
&\le 4d^{-3}\left(2\sup|\varphi|+\sqrt{\frac{2}{\pi}}\right)^2 M^2\sup|\varphi'|^2,
\end{aligned}
$$

where the first inequality follows from Jensen's inequality. By Lemma 48 we have

$$
\mathbb{E}\left[\left(\frac{1}{d^2}\mathbb{E}_{Z,Z',A}\log\hat{\mathcal{Z}}_{t,\epsilon}-\frac{1}{d^2}\mathbb{E}_\Theta\log\hat{\mathcal{Z}}_{t,\epsilon}\right)^2\right]\le\frac{C(\varphi,M,\alpha)}{d^2}, \tag{63}
$$

where Θ represents all quenched variables except $S^\star$. Next we write $S^\star = O\Lambda^\star O^T$ with O drawn from the uniform (Haar) measure on the orthogonal group $\mathcal{O}(d)$, and $\Lambda^\star$ the diagonal matrix of eigenvalues of $S^\star$. We denote $\tilde{G}_\mu := O^T G_\mu O$, $\tilde{s} := O^T s O$ and $\tilde{Z}' = O^T\tilde{Z}'O$. In this way the Hamiltonian reads

$$
\begin{aligned}
\hat{H}_t(\tilde{s},u,a) := &\frac{1}{2}\sum_{\mu=1}^{n}(\Gamma_{t,\mu}(\tilde{s},u_\mu,a_\mu)^2+2Z_\mu\Gamma_{t,\mu}(\tilde{s},u_\mu,a_\mu)) \\
&+\frac{1}{4}\sum_{i,j=1}^{d}(dR_1(t)(\Lambda^\star_{ij}-\tilde{s}_{ij})^2+2\tilde{Z}'_{ij}\sqrt{dR_1(t)}(\Lambda^\star_{ij}-\tilde{s}_{ij})),
\end{aligned} \tag{64}
$$

where

$$
\Gamma_{t,\mu}(\tilde{s},u_\mu) := \varphi(\sqrt{1-t}\mathrm{Tr}[\tilde{G}_\mu\Lambda^\star]+k_1(t)V_\mu+k_2(t)U^\star_\mu,A_\mu)-\varphi(\sqrt{1-t}\mathrm{Tr}[\tilde{G}_\mu\tilde{s}]+k_1(t)V_\mu+k_2(t)u^\star_\mu,a_\mu).
$$

Note that $\tilde{G}_\mu$ and $\tilde{Z}'$ are still independent GOE(d) matrices, and the distribution of $\tilde{s}$ is the same as that of s, so we have

$$
\mathbb{E}_{O,\Theta}\frac{1}{d^2}\log\hat{\mathcal{Z}}_{t,\epsilon} = \mathbb{E}_{\tilde{G},\tilde{Z}',Z,A,V,U^\star}\frac{1}{d^2}\log\int P_A(\mathrm{d}a)P_0(\mathrm{d}\tilde{s})\mathcal{D}u e^{-\hat{H}_t(\tilde{s},u,a)}, \tag{65}
$$

where both sides are only a function of $\Lambda^\star$, which gives

$$\begin{aligned}
&\mathbb{E}\left[\left(\frac{1}{d^2}\mathbb{E}_\Theta \log \hat{\mathcal{Z}}_{t,\epsilon} - \frac{1}{d^2}\mathbb{E}_{O,\Theta}\log \hat{\mathcal{Z}}_{t,\epsilon}\right)^2\right] \\
&\leq \mathbb{E}\left[\frac{1}{d^2}\left(\mathbb{E}_{\tilde{G},\tilde{Z}',Z,A,V,U^\star}\log\int P_A(\mathrm{d}a)P_0(\mathrm{d}\tilde{s})\mathcal{D}ue^{-\hat{H}_t(\tilde{s},u,a)} - \frac{1}{d^2}\log\int P_A(\mathrm{d}a)P_0(\mathrm{d}\tilde{s})\mathcal{D}ue^{-\hat{H}_t(\tilde{s},u,a)}\right)^2\right] \\
&\leq \frac{C(\varphi,M,\alpha)}{d^2}.
\end{aligned} \tag{66}$$

The last inequality is from our previous calculation. Finally we consider $g := \mathbb{E}_{O,\Theta}\frac{1}{d^2}\log\hat{\mathcal{Z}}_{t,\epsilon}$ as a function of $\Lambda^\star$. We have (see eq. (64)):

$$\begin{aligned}
\sum_{i=1}^d\left(\frac{\partial g}{\partial \Lambda_i^\star}\right)^2 &= d^{-4}\sum_{i=1}^d\left(\mathbb{E}_{O,\Theta}\left\langle\frac{\partial \hat{H}_t}{\partial\Lambda_i}\right\rangle\right)^2 \\
&= d^{-4}\sum_{i=1}^d\left(\mathbb{E}_{O,\Theta}\left\langle\sum_{\mu=1}^n(\Gamma_{t,\mu}+Z_\mu)\frac{\partial\Gamma_{t,\mu}}{\partial(\Lambda_i^\star)} + \frac{dR_1(t)}{2}(\Lambda_i^\star - \tilde{s}_{ii}) + 2\tilde{Z}'_{ii}\sqrt{dR_1(t)}\right\rangle\right)^2, \\
&\leq 3(I_1+I_2+I_3),
\end{aligned}$$

where we use

$$\left(\sum_{i=1}^p v_i\right)^2 \leq p\sum_{i=1}^p v_i^2 \tag{67}$$

for any integer $p \geq 1$. The first term is

$$\begin{aligned}
I_1 &:= d^{-4}\sum_{i=1}^d\left(\mathbb{E}_{O,\Theta}\left\langle\sum_{\mu=1}^n(\Gamma_{t,\mu}+Z_\mu)\frac{\partial\Gamma_{t,\mu}}{\partial(\Lambda_i^\star)}\right\rangle\right)^2 \\
&= d^{-4}\sum_{i=1}^d\left(\mathbb{E}_{O,\Theta}\left\langle\sum_{\mu=1}^n(\Gamma_{t,\mu}+Z_\mu)\sqrt{1-t}\tilde{G}_{\mu,ii}\varphi'(J_{t,\mu})\right\rangle\right)^2 \\
&\overset{(a)}{\leq} 16d^{-6}\sum_{i=1}^d\left(\mathbb{E}_{O,\Theta}\left\langle\sum_{\mu=1}^n\varphi'(J_{t,\mu})(\Lambda_i^\star\varphi'(J_{t,\mu}) - \tilde{s}_{ii}\varphi'(j_{t,\mu}))\right\rangle\right)^2 + \left(\mathbb{E}_{O,\Theta}\left\langle\Lambda_i^\star(\Gamma_{t,\mu}+Z_\mu)\varphi''(J_{t,\mu})\right\rangle\right)^2 \\
&\quad + \left(\mathbb{E}_{O,\Theta}\langle(\Gamma_{t,\mu}+Z_\mu)^2\varphi'(J_{t,\mu})(\Lambda_i^\star\varphi'(J_{t,\mu}) - \tilde{s}_{ii}\varphi'(j_{t,\mu}))\rangle\right)^2 \\
&\quad + \left(\mathbb{E}_{O,\Theta}\langle(\Gamma_{t,\mu}+Z_\mu)\varphi'(J_{t,\mu})\rangle\langle(\Gamma_{t,\mu}+Z_\mu)(\Lambda_i^\star\varphi'(J_{t,\mu}) - \tilde{s}_{ii}\varphi'(j_{t,\mu}))\rangle\right)^2 \\
&\leq 16d^{-2}\sum_{i=1}^d 2\sup|\varphi'|^2\mathbb{E}_{O,\Theta}((\Lambda_i^\star)^2 + \langle\tilde{s}_{ii}\rangle^2) + \sup|\varphi''|^2(\mathbb{E}_{O,\Theta}(\sup|\varphi| + |Z_\mu|)\Lambda_i^\star)^2 \\
&\quad + 4\sup|\varphi'|^2(\mathbb{E}_{O,\Theta}(\sup|\varphi| + |Z_\mu|)^2(\Lambda_i^\star)^2 + \langle\tilde{s}_{ii}\rangle^2))^2 \\
&\overset{(b)}{\leq} \frac{C(\varphi,M,\alpha)}{d}.
\end{aligned}$$

In (a) we use Gaussian integration by parts w.r.t. $\tilde{G}_{\mu,ii}$, the fact that $J_{t,\mu} = \varphi(\sqrt{1-t}\mathrm{Tr}[\tilde{G}_\mu \Lambda^\star] + k_1(t)V_\mu + k_2(t)U^\star_\mu$, and eq. (67). In (b) we use $\sum_{i=1}^d \langle \tilde{s}_{ii}\rangle^2 = \mathrm{Tr}\langle \tilde{s}\rangle^2 \le dM^2$ by the boundedness of its eigenvalues for the last inequality. The second term is

$$\begin{aligned} I_2 &\coloneqq d^{-4}\sum_{i=1}^d \left(\mathbb{E}_{O,\Theta}\left\langle \frac{dR_1(t)}{2}(\Lambda_i^\star - \tilde{s}_{ii})\right\rangle\right)^2 \\ &\le \frac{d^{-2}K^2}{2}\sum_{i=1}^d \mathbb{E}_{O,\Theta}\left[(\Lambda_i^\star)^2 + \langle \tilde{s}_{ii}\rangle^2\right] \le d^{-1}K^2M^2, \end{aligned} \tag{68}$$

where we use Jensen's inequality for the first inequality. Finally, the third term is

$$I_3 \coloneqq d^{-4}\sum_{i=1}^d \left(\mathbb{E}_{O,\Theta}\left\langle 2\tilde{Z}'_{ii}\sqrt{dR_1(t)}\right\rangle\right)^2 = d^{-3}R_1(t)\sum_{i=1}^d (\mathbb{E}_{O,\Theta}|\tilde{Z}'_{ii}|)^2 \le \frac{C(\varphi, M, \alpha)}{d^2}. \tag{69}$$

Combining these three terms, we have

$$\mathbb{E}\left[\left(\frac{1}{d^2}\mathbb{E}_\Theta \log \hat{\mathcal{Z}}_{t,\epsilon} - \frac{1}{d^2}\mathbb{E}\log \hat{\mathcal{Z}}_{t,\epsilon}\right)^2\right] \le \frac{C(\varphi, M, \alpha, \kappa)}{d^2} \tag{70}$$

by using the Poincaré inequality for rotationally-invariant priors (Lemma 45 in Appendix G). We finish the proof of Lemma 29 by combining eqs. (62), (63), (66) and (70). ∎

C.3. Concentration of the overlap

In this section we prove Lemma 24. Recall that we are using Assumptions 3.1, 3.2 and Gaussian data. Concentration of the overlap results from the concentration of

$$\mathcal{L} \coloneqq \frac{1}{d^2}\frac{\mathrm{d}H_{t,\epsilon}}{\mathrm{d}R_1} = \frac{1}{2d}\sum_{i,j=1}^d \left(\frac{s_{ij}^2}{2} - s_{ij}S^\star_{ij} - \frac{s_{ij}Z'_{ij}}{2\sqrt{R_1}}\right)$$

because, following (Barbier and Macris, 2019, Section 6), we have

$$\begin{aligned} \mathbb{E}\langle(\mathcal{L} - \mathbb{E}\langle\mathcal{L}\rangle)^2\rangle &= \frac{1}{16d^2}\sum_{i,j,k,l=1}^d (\mathbb{E}[\langle s_{ij}s_{kl}\rangle^2] - \mathbb{E}[\langle s_{ij}\rangle^2]\mathbb{E}[\langle s_{kl}\rangle^2]) \\ &\quad + \frac{1}{8d^2}\sum_{i,j,k,l=1}^d (\mathbb{E}[\langle s_{ij}s_{kl}\rangle^2] - \mathbb{E}[\langle s_{ij}s_{kl}\rangle\langle s_{ij}\rangle\langle s_{kl}\rangle]) + \frac{1}{16d^2R_1}\sum_{i,j=1}^d \mathbb{E}[\langle s_{ij}^2\rangle], \end{aligned}$$

which leads to

$$\mathbb{E}\langle(Q - \mathbb{E}\langle Q\rangle)^2\rangle \le 16\mathbb{E}\langle(\mathcal{L} - \mathbb{E}\langle\mathcal{L}\rangle)^2\rangle, \tag{71}$$

where we recall that $Q \coloneqq \frac{1}{d}\sum_{i,j=1}^d s_{ij}S^\star_{ij}$. We refer to (Barbier and Macris, 2019, Section 6) for the proof of eq. (71). Then we can bound the fluctuation of the overlap by the fluctuation of $\mathcal{L}$. The following lemma is a direct transposition of (Barbier et al., 2019, Lemma 29) to our setting: we refer to Barbier et al. (2019) for its proof, as it does not use the prior structure and thus directly applies in our context.

Lemma 30 *For any $\iota_d \leq 1/2$:*

$$\int_{B_d} \mathrm{d}\epsilon \, \mathbb{E}\langle(\mathcal{L} - \langle\mathcal{L}\rangle)^2\rangle \leq \frac{\rho(1+\rho)}{d^2}.$$

We now prove:

Lemma 31

$$\int_{B_d} \mathrm{d}\epsilon \mathbb{E}[(\langle\mathcal{L}\rangle - \mathbb{E}\langle\mathcal{L}\rangle)^2] \leq C(\varphi, M, \alpha, \kappa) \cdot o_d(1).$$

Proof [Proof of Lemma 31] $\mathcal{L}$ is connected to the free entropy through

$$\frac{\mathrm{d}F_{d,\epsilon}(t)}{\mathrm{d}R_1} = -\langle\mathcal{L}\rangle - \frac{1}{4d}\sum_{i,j=1}^{d}((S^\star_{ij})^2 + \frac{1}{\sqrt{R_1}} S^\star_{ij} Z'_{ij}), \tag{72}$$

$$\frac{1}{d^2}\frac{\mathrm{d}^2 F_{d,\epsilon}(t)}{\mathrm{d}R_1^2} = \langle\mathcal{L}^2\rangle - \langle\mathcal{L}\rangle^2 - \frac{1}{8d^3 R_1^{3/2}}\sum_{i,j=1}^{d}\langle s_{ij}\rangle Z'_{ij},$$

where $F_{d,\epsilon}(t) := \frac{1}{d^2}\log \mathcal{Z}_{t,\epsilon}(Y, Y', G, V)$. Now we suppose that the eigenvalues of Z' are $\{\Lambda_{Z,i}\}_{i=1}^d$, and further define

$$\tilde{F}(R_1) := F_{d,\epsilon}(t) - \frac{\sqrt{R_1}}{2d} M \sum_{i=1}^{d} |\Lambda_{Z,i}|, \; \tilde{f}(R_1) := \mathbb{E}\tilde{F}(R_1) = f_{d,\epsilon}(t) - \frac{\sqrt{R_1}}{2d} M \sum_{i=1}^{d} \mathbb{E}|\Lambda_{Z,i}|,$$

where we recall that $f_{d,\epsilon}(t) = \mathbb{E}F_{d,\epsilon}(t)$. $\tilde{F}(R_1)$ is a convex function because

$$\frac{1}{d^2}\frac{\mathrm{d}^2\tilde{F}(t)}{\mathrm{d}R_1^2} = \langle\mathcal{L}^2\rangle - \langle\mathcal{L}\rangle^2 + \frac{1}{8dR_1^{3/2}}\left(\sum_{i=1}^{d} M|\Lambda_{Z,i}| - \sum_{i,j=1}^{d}\langle s_{ij}\rangle Z'_{ij}\right) \geq 0,$$

where the last inequality is a consequence of $|\mathrm{Tr}[AB]| \leq \|A\|_{\mathrm{op}} \sum_{i=1}^{d} |\lambda_{B,i}|$. Moreover, $\tilde{f}(R_1)$ is also convex because $f_{d,\epsilon}(t)$ is convex in R_1 (its derivative is the MMSE, which decreases with increasing R_1).

By definition and eq. (72), we have

$$\tilde{F}'(R_1) - \tilde{f}'(R_1) = \mathbb{E}\langle\mathcal{L}\rangle - \langle\mathcal{L}\rangle + \frac{\rho}{4} - \frac{1}{4d}\sum_{i,j=1}^{d}((S^\star_{ij})^2 + \frac{1}{\sqrt{R_1}} S^\star_{ij} Z'_{ij}) - \frac{1}{4\sqrt{R_1}} MA,$$

where

$$A := \frac{1}{d}\sum_{i=1}^{d}(|\Lambda_{Z,i}| - \mathbb{E}|\Lambda_{Z,i}|)$$

satisfies $\mathbb{E}[A^2] \leq ad^{-2}$ (by repeating the proof arguments of Lemma 46-(b), i.e. directly using the Poincaré inequality of Lemma 45 for the $\mathrm{GOE}(d)$ distribution).

Due to the convexity of $\tilde{F}$ and $\tilde{f}$, we have (Barbier et al., 2019, Lemma 31)

$$|\langle\mathcal{L}\rangle - \mathbb{E}\langle\mathcal{L}\rangle| \leq \delta^{-1} \sum_{u\in\{R_1-\delta,R_1,R_1+\delta\}} \left(|F_{d,\epsilon}(t, R_1 = u) - f_{d,\epsilon}(t, R_1 = u)| + \frac{1}{2}M|A|\sqrt{u}\right)$$
$$+ C_\delta^+(R_1) + C_\delta^-(R_1) + \frac{M|A|}{4\sqrt{\epsilon_1}} + \left|\frac{\rho}{4} - \frac{1}{4d}\sum_{i,j=1}^{d}\left(\left(S_{ij}^\star\right)^2 + \frac{1}{\sqrt{R_1}}S_{ij}^\star Z_{ij}'\right)\right|,$$

where $C_\delta(R_1)^+ := \tilde{f}'(R_1+\delta) - \tilde{f}'(R_1) \geq 0$, $C_\delta(R_1)^- := \tilde{f}'(R_1) - \tilde{f}'(R_1-\delta) \geq 0$. Recall that $R_1 \geq \varepsilon_1$. Denote the last term as $|B|$. We then have

$$B := \frac{\rho}{4} - \frac{1}{4d}\sum_{i=1}^{d}\left((\Lambda_i^\star)^2 + \frac{1}{\sqrt{R_1}}\Lambda_{ii}^\star \tilde{Z}_{ii}'\right),$$

where we denote $S^\star := O\Lambda^\star O^T$ and $\tilde{Z}' := O^T Z' O$. $\Lambda^\star$ is independent of O, so $\tilde{Z}' \sim \mathrm{GOE}(d)$ is independent of $\Lambda^\star$. By Lemma 46 we have:

$$\mathbb{E}[B^2] \leq \frac{1}{8}\mathbb{E}\left[\left(\rho - \frac{1}{d}\sum_{i=1}^{d}(\Lambda_i^\star)^2\right)^2\right] + \frac{1}{8R_1}\mathbb{E}\left[\left(\frac{1}{d}\sum_{i=1}^{d}\Lambda_{ii}^\star\tilde{Z}_{ii}'\right)^2\right] \leq o(1) + b\iota_d^{-1}d^{-1},$$

for a constant b. Then we have

$$\frac{1}{10}\mathbb{E}[(\langle\mathcal{L}\rangle - \mathbb{E}\langle\mathcal{L}\rangle)^2] \leq 3\delta^{-2}(C(\varphi, M, \alpha, \kappa) + aM^2 d)d^{-2} + C_\delta^+(R_1)^2 + C_\delta^-(R_1)^2 + \frac{M^2 a}{4\epsilon_1 d^2} + \frac{b}{\iota_d d} + o(1),$$

where we use Lemma 29. As

$$\left|\frac{\mathrm{d}f_{d,\epsilon}(t)}{\mathrm{d}R_1}\right| = \left|\frac{1}{4d}\sum_{i,j=1}^{d}\mathbb{E}[\langle s_{ij}\rangle^2] - \frac{\rho}{4}\right| \leq \frac{\rho}{2}$$

and $R_1 \geq \epsilon_1$, we have

$$|\tilde{f}'(R_1)| \leq \frac{1}{2}(\rho + \frac{M}{\sqrt{\epsilon_1}}).$$

Following steps in (Barbier et al., 2019, Lemma 30), we have

$$\int_{B_d} d\epsilon(C_\delta^+(R_1)^2 + C_\delta^-(R_1)^2) \leq \frac{1}{2}\delta(\iota_d + \rho)\left(\rho + \frac{M}{\sqrt{\iota_d}}\right)^2,$$

and thus

$$\mathbb{E}[(\langle\mathcal{L}\rangle - \mathbb{E}\langle\mathcal{L}\rangle)^2] \leq 30\delta^{-2}(C(\varphi, M, \alpha, \kappa) + aM^2 d)d^{-2} + 5\delta(\iota_d + \rho)\left(\rho + \frac{M}{\sqrt{\iota_d}}\right)^2$$
$$+ 5M^2 a\frac{\log 2}{2}\frac{\iota_d}{d} + \frac{b}{\iota_d d} + o(1).$$

We now finish the proof by choosing $\delta = \iota_d d^{-1/4}$ (and $\iota_d = \frac{1}{2}d^{-1/8}$). ∎

Finally, we prove Lemma 24 by combining eq. (71), Lemma 30 and Lemma 31.

C.4. Relaxation of Assumptions 3.1 and 3.2

Lemma 32 *Denote $f_d(V)$ to be the free entropy of eq. (3) corresponding to the prior P_0 with potential V, and denote $\widetilde{V}_M$ the truncation of V to $[-M, M]$, with $\widetilde{V}_M(x) = +\infty$ if $|x| > M$. Notice that then the prior with potential $\widetilde{V}$ satisfies Assumption 3.1. Suppose that V satisfies Assumption 2.1 and φ satisfies Assumption 3.2. Then, for $M > 0$ large enough, we have*

$$\lim_{d\to\infty} |f_d(V) - f_d(\widetilde{V}_M)| = 0.$$

Proof [Proof of Lemma 32] We consider the following interpolation

$$\begin{aligned} Y_{1,\mu} &= \sqrt{t}\varphi(\mathrm{Tr}[\Phi_\mu S_1^\star], A_\mu) + \sqrt{\Delta} Z_{1,\mu}, \\ Y_{2,\mu} &= \sqrt{1-t}\varphi(\mathrm{Tr}[\Phi_\mu S_2^\star], A_\mu) + \sqrt{\Delta} Z_{2,\mu}, \end{aligned} \tag{73}$$

where $\{A_\mu\}_{\mu=1}^n \overset{\text{i.i.d.}}{\sim} P_A, \{Z_{1,\mu}\}_{\mu=1}^n, \{Z_{2,\mu}\}_{\mu=1}^n \overset{iid}{\sim} \mathcal{N}(0,1)$ are independent of $S_1^\star, S_2^\star$. $S_1^\star$ is sampled from the prior with potential V and $S_2^\star$ is sampled from the prior with potential $\widetilde{V}_M$, and are coupled as follows. The joint distribution of $S_1^\star, S_2^\star$ is given by $S_1^\star \coloneqq O\Lambda_1^\star O^T, S_2^\star \coloneqq O\Lambda_2^\star O^T$, where O is Haar distributed, and the coupling between $\Lambda_1^\star$ and $\Lambda_2^\star$ remains arbitrary for now.

We can define the mutual information $I(t) \coloneqq I(S_1^\star, S_2^\star; Y_1, Y_2|\Phi)$. Following (Barbier et al., 2019, Proposition 14) and using the I-MMSE theorem (Barbier et al., 2020, Lemma 4.5), we have

$$I'(t) = \frac{1}{2}\alpha\mathbb{E}\left[\sum_{\mu=1}^n (\varphi(\mathrm{Tr}[\Phi_\mu S_1^\star], A_\mu) - \varphi(\mathrm{Tr}[\Phi_\mu\langle s_1\rangle], A_\mu))^2 - (\varphi(\mathrm{Tr}[\Phi_\mu S_2^\star], A_\mu) - \varphi(\mathrm{Tr}[\Phi_\mu\langle s_2\rangle], A_\mu))^2\right],$$

where the Gibbs bracket $\langle\cdot\rangle$ denotes the expectation with respect to the posterior distribution $\mathbb{P}(s_1, s_2|Y_1, Y_2, \Phi)$ of the model (eq. (73)).

For simplicity, we denote $z_{1,\mu} \coloneqq \mathrm{Tr}[\Phi_\mu S_1^\star], z_{2,\mu} \coloneqq \mathrm{Tr}[\Phi_\mu S_2^\star], e_{1,\mu} \coloneqq \mathrm{Tr}[\Phi_\mu\langle s_1\rangle], e_{2,\mu} \coloneqq \mathrm{Tr}[\Phi_\mu\langle s_2\rangle]$ and thus

$$\begin{aligned} |I'(t)| &= \frac{\alpha}{2}|\mathbb{E}[||\varphi(z_1, A) - \varphi(e_1, A)||^2 - ||\varphi(z_2, A) - \varphi(e_2, A)||^2]| \\ &= \frac{\alpha}{2}\mathbb{E}[(||\varphi(z_1, A) - \varphi(e_1, A)|| + ||\varphi(z_2, A) - \varphi(e_2, A)||)(||\varphi(z_1, A) - \varphi(e_1, A)|| - ||\varphi(z_2, A) - \varphi(e_2, A)||)] \\ &\le 2\alpha\sqrt{n}\sup|\varphi|\mathbb{E}[||\varphi(z_1, A) - \varphi(e_1, A) - \varphi(z_2, A) + \varphi(e_2, A)||] \\ &\le 2\alpha\sqrt{n}\sup|\varphi|\mathbb{E}[||\varphi(z_1, A) - \varphi(z_2, A)|| + ||\varphi(e_1, A) - \varphi(e_2, A)||] \\ &\overset{(a)}{=} 4\alpha\sqrt{n}\sup|\varphi|\mathbb{E}[||\varphi(z_1, A) - \varphi(z_2, A)||] \\ &\le 4\alpha\sqrt{n}\sup|\varphi|\sup|\varphi'|\mathbb{E}[||z_1 - z_2||^2]^{1/2} \\ &\overset{(b)}{=} 4\alpha\sqrt{nd}\sup|\varphi|\sup|\varphi'|\mathbb{E}[||\Lambda_1^\star - \Lambda_2^\star||^2]^{1/2}. \end{aligned}$$

We use the Nishimori identity in equation (a), see Proposition 43, and Assumption 2.2 in equation (b). We also used the triangle inequality, the mean value inequality, and Jensen's inequality.

Recall that the free entropy is related to the mutual information as (see eq. (4)):

$$\frac{1}{d^2}I(1) = -f_d(V) + \alpha\Psi_{\text{out}}(\rho_{S_1}) + o_d(1), \ \frac{1}{d^2}I(0) = -f_d(\tilde{V}) + \alpha\Psi_{\text{out}}(\rho_{S_2}) + o_d(1),$$

where we denote $\rho_{S_1} \coloneqq \lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d}(\Lambda_1^\star)^2$ and $\rho_{S_2} \coloneqq \lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d}(\Lambda_2^\star)^2$. This gives

$$|f_d(V)-f_d(\tilde{V})| \leq 4\alpha\sqrt{\alpha}\sup|\varphi|\sup|\varphi'|\mathbb{E}[W_2(\mu_{S_1},\mu_{S_2})^2]^{1/2}+\alpha|\Psi_{\text{out}}(\rho_{S_1})-\Psi_{\text{out}}(\rho_{S_2})|+o_d(1), \tag{74}$$

where $W_2(\mu_{S_1},\mu_{S_2}) \coloneqq \sqrt{\inf \frac{1}{d}||\Lambda_1^\star - \Lambda_2^\star||^2}$ denotes the Wasserstein-2 distance between the empirical distributions of $\Lambda_1^\star$ and $\Lambda_2^\star$. Note that the infimum is w.r.t. the coupling between the random variables $\Lambda_1^\star$ and $\Lambda_2^\star$. Given that our analysis was made for an arbitrary choice of this coupling, we chose the coupling between $\Lambda_1^\star$ and $\Lambda_2^\star$ such that $\frac{1}{d}||\Lambda_1^\star - \Lambda_2^\star||^2$ is $W_2(\mu_{S_1},\mu_{S_2})^2+\epsilon$, before taking $\epsilon \to 0$: this gave rise to eq. (74)

Lemma 44 in Appendix G shows that μ_{S_1},μ_{S_2} both weakly converges to μ_0 almost surely for $M > 0$ large enough. Thus from Proposition 13(iv) we have $\rho_{S_1} = \rho_{S_2}$, $\lim_{d\to\infty} W_2(\mu_{S_1},\mu_0) = 0, a.s.$ and $\lim_{d\to\infty} W_2(\mu_{S_2},\mu_0) = 0, a.s.$, which gives

$$\lim_{d\to\infty} W_2(\mu_{S_1},\mu_{S_2}) \leq \lim_{d\to\infty} W_2(\mu_{S_1},\mu_0) + \lim_{d\to\infty} W_2(\mu_{S_2},\mu_0) = 0, a.s.$$

Moreover, by definition we have

$$W_2(\mu_{S_1},\mu_{S_2})^2 \leq 2\left(\int \mu_{S_1}(\mathrm{d}x)x^2 + \int \mu_{S_2}(\mathrm{d}x)x^2\right).$$

By Proposition 13(iii), both terms are uniformly integrable, so we have

$$\lim_{d\to\infty} \mathbb{E}[W_2(\mu_{S_1},\mu_{S_2})^2] = 0$$

by using dominated convergence. Finally, plugging all these results into eq. (74), we get

$$\lim_{d\to\infty} |f_d(V) - f_d(\tilde{V})| = 0,$$

which finishes the proof. ∎

The following lemma states that one can relax Assumption 3.2. It directly follows from (Barbier et al., 2019, Proposition 25), plugging in the central limit theorem result of Lemma 46.

Lemma 33 *Suppose that the potential satisfies Assumption 2.1 and φ satisfies Assumption 2.3. Then for all $\varepsilon > 0$, there exists a $\hat{\varphi}$ satisfying Assumption 3.2 such that $|f_d(\varphi) - f_d(\hat{\varphi})| < \varepsilon$ for d large enough, where $f_d(\varphi)$ denotes the free entropy corresponding to φ.*

Lemmas 32 and 33 show that for any P_0 satisfying Assumption 2.1 and φ satisfying Assumption 2.3, there exists $\hat{P}_0$ satisfying Assumption 3.1 and $\hat{\varphi}$ satisfying Assumption 3.2, such that the free entropies for (P_0,φ) and $(\hat{P}_0,\hat{\varphi})$ are arbitrarily close as $d \to \infty$: this allows us to relax Assumptions 3.1 and 3.2 to Assumptions 2.1 and 2.3 in the proof of Theorem 1.

Appendix D. Proof of Theorem 3

The proof of Theorem 3 follows from (Barbier et al., 2019, Theorem 2), so we will only describe its sketch. To obtain an upper bound on the overlap $Q_d \coloneqq (1/d)\mathrm{Tr}[sS^\star]$, we consider the following model

$$\begin{cases} Y \sim P_{\text{out}}(\cdot|\mathrm{Tr}[GS^*]) \\ Y' = \sqrt{\dfrac{\lambda}{d^{2p-1}}}(S^*)^{\otimes 2p} + Z', \end{cases} \tag{75}$$

where $Z' \in \mathcal{S}_d^{\otimes 2p} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$. The additional side information comes from a so-called *spiked tensor model*, which is considered in Appendix F. A combination of Theorems 1 and 37 gives its free entropy (see also (Barbier et al., 2019, Section 5.3) for details):

$$\lim_{d\to\infty} f_d = \sup_{q\in[0,\rho]} \inf_{r\geq 0} \psi_{P_0}(r+4p\lambda q^{2p-1}) + \alpha\Psi_{\text{out}}(q) + \frac{1}{4} + p\lambda\rho q^{2p-1} + \frac{1}{4}r(\rho - q) - \frac{\lambda(2p-1)q^{2p}}{2}.$$

Similarly to (Barbier et al., 2019, Section 5.3), we can use the I-MMSE theorem with respect to the signal-to-noise ration λ to obtain

$$\frac{1}{d^{2p}}\text{MMSE}((S^*)^{\otimes 2p}|Y, Y', G) \to \rho^{2p} - q^*(\alpha, \lambda)^{2p},$$

and thus (since the side-information channel in eq. (75) can only reduce the MMSE with respect to the original observation model):

$$\liminf_{d\to\infty} \frac{1}{d^{2p}}\text{MMSE}((S^*)^{\otimes 2p}|Y, G) \geq \rho^{2p} - q^*(\alpha)^{2p}.$$

As the left side of this last equation is equal $\rho^{2p} - \mathbb{E}[Q_d^{2p}]$ by the Nishimori identity (Proposition 43), we obtain an upper bound on the overlap:

$$\limsup_{d\to\infty} \mathbb{E}[Q_d^{2p}] \leq q^*(\alpha)^{2p},$$

for any $p \geq 1$. This implies

$$\lim_{d\to\infty} \mathbb{P}(|Q_d| \geq q^*(\alpha) + \epsilon) = 0 \tag{76}$$

for all $\epsilon > 0$. The lower bound on the overlap

$$\lim_{d\to\infty} \mathbb{P}(|Q_d| \leq q^*(\alpha) - \epsilon) = 0 \tag{77}$$

is obtained in the exact same way as (Barbier et al., 2019, Section 5.3.2), thus we omit its proof. Combining eqs. (76) and (77), we obtain the convergence of the overlap, and thus the MMSE.

Finally, we note that all the above arguments still hold if we replace G_μ with Φ_μ under Assumption 2.2, because a direct generalization of Theorem 1 to the model of eq. (75) shows that the limiting free entropy is unchanged upon replacing G_μ by Φ_μ.

Appendix E. Proof of Theorem 8

According to Section 3.4, it remains to prove eq. (32) and Lemma 10.

E.1. Proof of eq. (32)

The I-MMSE theorem Guo et al. (2005) implies that (this can also easily be re-derived from eq. (31) using the Nishimori identity, see Proposition 43), for any $\Lambda \geq 0$:

$$\frac{\partial}{\partial\Lambda} f_d(\Lambda) = -\frac{d}{2}\mathbb{E}\left[(S^\star_{12} - \langle S_{12}\rangle_{t,\Lambda})^2\right]. \tag{78}$$

Moreover, for $\Lambda = 0$, by permutation invariance of the law of S, we have

$$\begin{aligned} \frac{\partial}{\partial \Lambda} f_d(0) &= -\frac{1}{2(d-1)} \sum_{i \neq j} \mathbb{E}\left[(S^\star_{ij} - \langle S_{ij} \rangle_{t,0})^2\right], \\ &= -\frac{1}{2(d-1)} \left[d\, \mathbb{E}\,\mathrm{tr}[(S^\star - \langle S \rangle_{t,0})^2] - \sum_{i=1}^d \mathbb{E}[(S^\star_{ii} - \langle S_{ii} \rangle_{t,0})^2] \right]. \end{aligned} \tag{79}$$

In particular, this implies that (using again permutation invariance and the Nishimori identity)

$$\begin{aligned} \left| \frac{\partial}{\partial \Lambda} f_d(0) + \frac{1}{2} \mathbb{E}\,\mathrm{tr}[(S^\star - \langle S \rangle_{t,0})^2] \right| &\leq \frac{d}{2(d-1)} \mathbb{E}[(S^\star_{11})^2 - \langle S_{11} \rangle_{t,0}^2] + \frac{1}{2(d-1)} \mathbb{E}\,\mathrm{tr}[(S^\star)^2 - \langle S \rangle_{t,0}^2], \\ &\leq \frac{d}{2(d-1)} \mathbb{E}[(S^\star_{11} - 1)^2 - \langle S_{11} - 1 \rangle_{t,0}^2] + \frac{1}{2(d-1)} \mathbb{E}\,\mathrm{tr}[(S^\star)^2 - \langle S \rangle_{t,0}^2], \\ &\leq \frac{d}{2(d-1)} \mathbb{E}[(S^\star_{11} - 1)^2] + \frac{1}{2(d-1)} \mathbb{E}\,\mathrm{tr}[(S^\star)^2], \\ &\leq \frac{C(\kappa)}{d}, \end{aligned} \tag{80}$$

for some $C(\kappa) > 0$ (we used the form of $S^\star = (1/m) \sum_{k=1}^m w_k w_k^\top$ for $(w_k)_{k=1}^m \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \mathrm{I}_d)$), which gives (32).

E.2. Proof of MMSE equivalence: Lemma 10

Lemma 10 relies on a main lemma, which establishes that the derivative of the interpolating free entropy goes to zero uniformly. It is proven in Appendix E.3.

Lemma 34

$$\lim_{d \to \infty} \sup_{\Lambda \geq 0} \left| \frac{\partial f_d(t, \Lambda)}{\partial t} \right| = 0,$$

uniformly in $t \in [t_0, 1]$.

By the fundamental theorem of calculus and Jensen's inequality, we have

$$\begin{aligned} \sup_{\Lambda \geq 0} |f_d(1, \Lambda) - f_d(t_0, \Lambda)| &= \sup_{\Lambda \geq 0} \left| \int_0^1 \frac{\partial f_d(t, \Lambda)}{\partial t} \, \mathrm{d}t \right|, \\ &\leq \int_{t_0}^1 \sup_{\Lambda \geq 0} \left| \frac{\partial f_d(t, \Lambda)}{\partial t} \right| \mathrm{d}t. \end{aligned}$$

By the uniform convergence proven in Lemma 34, this implies

$$\sup_{\Lambda \geq 0} |f_d(1, \Lambda) - f_d(t_0, \Lambda)| \leq h_d, \tag{81}$$

for some $h_d \to 0$ as $d \to \infty$. Notice that by the I-MMSE theorem, for any $t \in [t_0, 1]$, $f_d(t, \Lambda)$ is a convex function of Λ because

$$\frac{\partial^2 f_d(t, \Lambda)}{\partial \Lambda^2} = \frac{d^2}{2} \mathbb{E}\left[\left(\langle S_{12}^2 \rangle_{t,\Lambda} - \langle S_{12} \rangle_{t,\Lambda}^2 \right)^2 \right] \geq 0. \tag{82}$$

By said convexity we have, for any $\Lambda > 0$:

$$\begin{aligned}\left(\frac{\partial f_d(t_0,\Lambda)}{\partial \Lambda}\right)_{\Lambda=0} &\leq \frac{f_d(t_0,\Lambda) - f_d(t_0,0)}{\Lambda}, \\ &\overset{\text{(a)}}{\leq} \frac{2h_d}{\Lambda} + \frac{f_d(1,\Lambda) - f_d(1,0)}{\Lambda}, \\ &\overset{\text{(b)}}{\leq} \frac{2h_d}{\Lambda} + \left(\frac{\partial f_d(1,\Lambda)}{\partial \Lambda}\right)_{\Lambda}. \end{aligned} \tag{83}$$

where we used eq. (81) in (a), and again convexity in (b). By a symmetric argument, we get

$$\left(\frac{\partial f_d(1,\Lambda)}{\partial \Lambda}\right)_{\Lambda=0} \leq \frac{2h_d}{\Lambda} + \left(\frac{\partial f_d(t_0,\Lambda)}{\partial \Lambda}\right)_{\Lambda}. \tag{84}$$

By the fundamental theorem of analysis and eqs. (83) and (84), we get that for all $\Lambda > 0$:

$$\begin{aligned}\left|\left(\frac{\partial f_d(t_0,\Lambda)}{\partial \Lambda}\right)_{\Lambda=0} - \left(\frac{\partial f_d(1,\Lambda)}{\partial \Lambda}\right)_{\Lambda=0}\right| &\leq \frac{2h_d}{\Lambda} + \max_{t\in\{t_0,1\}} \int_0^{\Lambda} \frac{\partial^2 f_d(t,u)}{\partial u^2}\,\mathrm{d}u, \\ &\leq \frac{2h_d}{\Lambda} + \sum_{t\in\{t_0,1\}} \int_0^{\Lambda} \frac{\partial^2 f_d(t,u)}{\partial u^2}\,\mathrm{d}u, \end{aligned} \tag{85}$$

where we also used that $\partial_u^2 f_d(t,u) \geq 0$ by convexity. By eq. (82) we have for any $\Lambda \geq 0$:

$$\frac{\partial^2 f_d(t,\Lambda)}{\partial \Lambda^2} \leq \frac{d^2}{2}\mathbb{E}[\langle S_{12}^2\rangle_{t,\Lambda}^2] \leq \frac{d^2}{2}\mathbb{E}[(S_{12}^\star)^4] = \frac{3}{2\kappa^2} + \frac{3}{\kappa^4 d^2}, \tag{86}$$

where we used the Cauchy-Schwartz inequality and the Nishimori identity (Proposition 43). By taking taking $\Lambda = \sqrt{h_d}$, we obtain an upper bound of the right hand side of eq. (85)

$$\inf_{\Lambda>0}\left\{\frac{2h_d}{\Lambda} + \sum_{t\in\{0,1\}}\int_0^{\Lambda} \frac{\partial^2 f_d(t,u)}{\partial u^2}\,\mathrm{d}u\right\} \leq C(\kappa)\sqrt{h_d}, \tag{87}$$

for some $C(\kappa) > 0$. We finish the proof of Lemma 10 by combining eqs. (85) and (87).

E.3. Proof of free entropy equivalence: Lemma 34

Recall that the interpolating model is

$$\begin{aligned} v_{t,\mu} \coloneqq \mathrm{Tr}[\Phi_\mu S^\star] + (1-t)\sqrt{d}[\mathrm{tr}\, S^\star - 1] + \sqrt{d(1-t)}\Delta\left(\frac{\|z_\mu\|^2}{m} - 1\right) \\ + \frac{2\sqrt{\Delta d(1-t)}}{m}\sum_{k=1}^{m} z_{\mu,k}\left(\frac{x_\mu^\top w_k^\star}{\sqrt{d}}\right) + \sqrt{\widetilde{\Delta} t}\zeta_\mu, \end{aligned} \tag{88}$$

with a side information channel

$$Y' = \sqrt{\Lambda}S_{12}^\star + \frac{\xi}{\sqrt{d}}. \tag{89}$$

Its free entropy is given by

$$f_d(t,\Lambda) = \mathbb{E}_{\{x_\mu\}_{\mu=1}^n} \frac{1}{d^2} \int \mathcal{D}W^\star \int dY' \left[\prod_{\mu=1}^n dv_\mu P_{\mathrm{out}}^{(t)}(v_\mu|W^\star, x_\mu)\right] e^{-\frac{d}{4}(Y'-\sqrt{\Lambda}S_{12}^\star)^2}$$
$$\log \int \mathcal{D}W \left[\prod_{\mu=1}^n P_{\mathrm{out}}^{(t)}(v_\mu|W, x_\mu)\right] e^{-\frac{d}{4}(Y'-\sqrt{\Lambda}S_{12})^2}.$$

Note that the output channel can be written as

$$P_{\mathrm{out}}^{(t)}(v_\mu|W^\star, x_\mu) = \mathbb{E}_{z,\zeta}\delta(v_\mu - \widetilde{v}(t, x_\mu, W^\star, z_\mu, \zeta_\mu)),$$

where

$$\widetilde{v}(t, x_\mu, W^\star, z_\mu, \zeta_\mu) := \mathrm{Tr}[\Phi_\mu S^\star] + (1-t)\sqrt{d}[\mathrm{tr}\, S^\star - 1] + \sqrt{d(1-t)}\Delta\left(\frac{\|z_\mu\|^2}{m} - 1\right)$$
$$+ \frac{2\sqrt{\Delta d(1-t)}}{m}\sum_{k=1}^m z_{\mu,k}\left(\frac{x_\mu^\top w_k^\star}{\sqrt{d}}\right) + \sqrt{\widetilde{\Delta} t}\zeta_\mu.$$

Thus we have (with $z = \{z_\mu\}$, $x = \{x_\mu\}$, $\zeta = \{\zeta_\mu\}$):

$$f_d(t,\Lambda) = \frac{1}{d^2}\mathbb{E}_{x,W^\star,z,\zeta,Y'} \log \int \mathcal{D}W \left[\prod_{\mu=1}^n P_{\mathrm{out}}^{(t)}(\widetilde{v}(t, x_\mu, W^\star, z_\mu, \zeta_\mu)|W, x_\mu)\right] e^{-\frac{d}{4}(Y'-\sqrt{\Lambda}S_{12})^2}.$$

Its derivative can be computed as

$$\frac{\partial}{\partial t} f_d(t,\Lambda) = \frac{1}{d^2}\mathbb{E}_{x,W^\star,z,\zeta,Y'} \sum_{\mu=1}^n \left\langle \frac{\partial}{\partial t} \log P_{\mathrm{out}}^{(t)}(\widetilde{v}(t, x_\mu, W^\star, z_\mu, \zeta_\mu)|W, x_\mu)\right\rangle, \tag{90}$$

where the Gibbs bracket is defined as:

$$\langle g(W)\rangle := \frac{\int \mathcal{D}W g(W) \left[\prod_{\mu=1}^n P_{\mathrm{out}}^{(t)}(\widetilde{v}(t, x_\mu, W^\star, z_\mu, \zeta_\mu)|W, x_\mu)\right] e^{-\frac{d}{4}(Y'-\sqrt{\Lambda}S_{12})^2}}{\int \mathcal{D}W \left[\prod_{\mu=1}^n P_{\mathrm{out}}^{(t)}(\widetilde{v}(t, x_\mu, W^\star, z_\mu, \zeta_\mu)|W, x_\mu)\right] e^{-\frac{d}{4}(Y'-\sqrt{\Lambda}S_{12})^2}}.$$

Notice that this Gibbs bracket depends on the realization of $x, W^\star, z, \zeta, Y'$. The following two lemmas give a bound of the right hand side of eq. (90). They are proven in Appendices E.4 and E.5.

Lemma 35 *There exists $\delta_0(\Delta,\kappa,t_0) > 0$ such that for any $\delta < \delta_0$ and $d > d_0(\delta,\Delta,\kappa,t_0)$, as long as*

$$\left|\frac{x_\mu^\top S x_\mu}{d} - 1\right| \le \frac{\delta}{\sqrt[4]{d}},\ |\sqrt{d}(\mathrm{tr}S - 1)| \le \frac{\delta}{\sqrt[4]{d}}, |\widetilde{v}_\mu - \mathrm{Tr}[\Phi_\mu S^\star]|^2 < 2t_0\widetilde{\Delta}^2\sqrt[4]{d}, \tag{91}$$

we have

$$\left|\frac{\partial}{\partial t}\log P_{\mathrm{out}}^{(t)}(\widetilde{v}_\mu|W, x_\mu)\right| \le C(\Delta,\kappa,t_0)\delta,$$

for some $C(\Delta,\kappa,t_0) > 0$.

Lemma 36 *For any $\mu = 1, \cdots, n$, we have*

$$\mathbb{E}_{x,W^\star,z,\zeta,Y'} \left\langle \left(\frac{\partial}{\partial t} \log P_{\mathrm{out}}^{(t)}(\widetilde{v}(t, x_\mu, W^\star, z_\mu, \zeta_\mu)|W, x_\mu) \right)^2 \right\rangle \leq \frac{\mathrm{poly(d)}}{t_0^{12}},$$

where $\mathrm{poly(d)}$ *represents a polynomial of* d*.*

In the following, we will call $\delta < \delta_0$ "sufficiently small δ" and $d > d_0$ "sufficiently large d".

We denote the three events in eq. (91) as $A_{1,\mu}(S, x_\mu), A_2(S), A_{3,\mu}(S^\star, x_\mu, z_\mu, \zeta_\mu)$. With $A_\mu \coloneqq A_{1,\mu} \cap A_2 \cap A_{3,\mu}$, and we have

$$\begin{aligned}
\left| \frac{\partial}{\partial t} f_{n,d}(t, \Lambda) \right| &= \left| \frac{1}{d^2} \mathbb{E}_{x,W^\star,z,\zeta,Y'} \sum_{i=1}^{n} \left\langle \frac{\partial}{\partial t} \log P_{\mathrm{out}}^{(t)}(\tilde{v}_\mu(t, x_\mu, W^\star, z_\mu, \zeta_\mu)|W, x_i)(\mathbb{1}\{A_\mu\} + \mathbb{1}\{A_\mu^\complement\}) \right\rangle \right| \\
&\leq \alpha C(\Delta, \kappa, t_0)\delta + \frac{1}{d^2} \sum_{\mu=1}^{n} \left[\mathbb{E}_{x,W^\star,z,\zeta,Y'} \left\langle \mathbb{1}\{A_\mu^\complement\} \right\rangle \right]^{1/2} \\
&\qquad \left[\mathbb{E}_{x,W^\star,z,\zeta,Y'} \left\langle \left(\frac{\partial}{\partial t} \log P_{\mathrm{out}}^{(t)}(\tilde{v}_\mu(t, x_\mu, W^\star, z_\mu, \zeta_\mu)|W, x_\mu) \right)^2 \right\rangle \right]^{1/2},
\end{aligned}$$

using Lemma 35 and the Cauchy-Schwartz inequality. By the union bound and the Nishimori identity, we have

$$\begin{aligned}
&\mathbb{E}_{x,W^\star,z,\zeta,Y'} \left\langle \mathbb{1}\{A_\mu^\complement\} \right\rangle \\
&\leq \mathbb{E}_{x,W^\star,z,\zeta,Y'} \left\langle \mathbb{1}\{A_{1,\mu}^\complement(S, x_\mu)\} + \mathbb{1}\{A_2^\complement(S)\} + \mathbb{1}\{A_{3,\mu}^\complement(S^\star, x_\mu, z_\mu, \zeta_\mu)\} \right\rangle \\
&= \mathbb{P}(A_{1,\mu}^\complement(S^\star, x_\mu)) + \mathbb{P}(A_2^\complement(S^\star)) + \mathbb{P}(A_{3,\mu}^\complement(S^\star, x_\mu, z_\mu, \zeta_\mu)).
\end{aligned} \tag{92}$$

Now let us control the three terms on the right hand side. By Bernstein's inequality and the Hanson-Wright inequality for any $0 < \delta < 1$, it is easy to show that with probability at least $1 - 2e^{-C(\kappa)d\delta^2}$:

$$\left| \frac{x_\mu^\top S^\star x_\mu}{d} - 1 \right| \leq \delta, \ |\sqrt{d}(\mathrm{tr} S^\star - 1)| \leq \delta,$$

Details on this classical derivation can be found in (Maillard et al., 2024, Appendix D.5). Thus:

$$\mathbb{P}(A_{1,\mu}^\complement), \mathbb{P}(A_2^\complement) \leq 2e^{-C(\kappa)\sqrt{d}\delta^2}. \tag{93}$$

for sufficiently small δ. To control $\mathbb{P}(A_{3,\mu}^\complement)$, we recall that

$$\begin{aligned}
\tilde{v}_\mu - \mathrm{Tr}[\Phi_\mu S^\star] \coloneqq &(1-t)\sqrt{d}[\mathrm{tr}\, S^\star - 1] + \sqrt{d(1-t)}\Delta \left(\frac{\|z_\mu\|^2}{m} - 1 \right) \\
&+ \frac{2\sqrt{\Delta d(1-t)}}{m} \sum_{k=1}^{m} z_{\mu,k} \left(\frac{x_\mu^\top w_k^\star}{\sqrt{d}} \right) + \sqrt{\widetilde{\Delta} t}\zeta_\mu.
\end{aligned} \tag{94}$$

For the first term, we have

$$\mathbb{P}(|(1-t)\sqrt{d}[\operatorname{tr} S^\star - 1]| > \varepsilon) \leq \mathbb{P}(|\sqrt{d}[\operatorname{tr} S^\star - 1]| > \varepsilon) \leq 2e^{-Cd\varepsilon^2}, \tag{95}$$

for any $\varepsilon < 1$, by Bernstein's inequality. For the second term, the classical Laurent-Massart bound of Laurent and Massart (2000) shows that

$$\mathbb{P}(|\|z_\mu\|^2 - m| > 2\sqrt{m}\varepsilon + 2\varepsilon^2) \leq 2e^{-\varepsilon^2},$$

and thus

$$\mathbb{P}\left(\sqrt{d(1-t)}\Delta\left|\frac{\|z_\mu\|^2}{m} - 1\right| > \varepsilon\right) = \mathbb{P}\left(|\|z_\mu\|^2 - m| > \frac{\kappa\sqrt{d}\varepsilon}{\Delta\sqrt{1-t}}\right) \leq 2e^{-C(\kappa,\Delta,t_0)\varepsilon^2}. \tag{96}$$

For the third term, we have

$$\frac{2\sqrt{\Delta d(1-t)}}{m}\sum_{k=1}^m z_{\mu,k}\left(\frac{x_\mu^\top w_k^\star}{\sqrt{d}}\right)\Big|W^\star, x_\mu \sim \mathcal{N}\left(0, \frac{4\Delta^2}{\kappa}\frac{x_\mu^\top S^\star x_\mu}{d}\right),$$

and thus

$$\mathbb{P}\left(\left|\frac{2\sqrt{\Delta d(1-t)}}{m}\sum_{k=1}^m z_k\left(\frac{x_\mu^\top w_k^\star}{\sqrt{d}}\right)\right| > \varepsilon\Big|W^\star, x_\mu\right) \leq 2\exp\left\{-\frac{\kappa}{8\Delta^2}\frac{d}{x_\mu^\top S^\star x_\mu}\varepsilon^2\right\},$$

which gives

$$\begin{aligned}
&\mathbb{P}\left(\left|\frac{2\sqrt{\Delta d(1-t)}}{m}\sum_{k=1}^m z_{\mu,k}\left(\frac{x_\mu^\top w_k^\star}{\sqrt{d}}\right)\right| > \varepsilon\right) \\
&\leq \mathbb{P}\left(\left|\frac{2\sqrt{\Delta d(1-t)}}{m}\sum_{k=1}^m z_{\mu,k}\left(\frac{x_\mu^\top w_k^\star}{\sqrt{d}}\right)\right| > \varepsilon\Big|\frac{x_\mu^\top S^\star x_\mu}{d} < 2\right) + \mathbb{P}\left(\frac{x_\mu^\top S^\star x_\mu}{d} > 2\right) \\
&\leq 2e^{-\frac{\kappa\varepsilon^2}{16\Delta^2}} + e^{-C(\kappa)\sqrt{d}},
\end{aligned} \tag{97}$$

using again eq. (93). For the last term, we have

$$\mathbb{P}(|\sqrt{\widetilde{\Delta}t_0}\zeta_\mu| > \varepsilon) \leq 2e^{-\frac{\varepsilon^2}{2\widetilde{\Delta}t_0}}. \tag{98}$$

Combining all the bounds of eqs. (95),(96),(97),(98), and choosing $\varepsilon := \frac{1}{2}\widetilde{\Delta}\sqrt{t_0}\sqrt[8]{d}$, we have

$$\begin{aligned}
\mathbb{P}(A_{3,\mu}^\complement) &\leq \mathbb{P}(|(1-t)\sqrt{d}[\operatorname{tr} S^\star - 1]| > \varepsilon) + \mathbb{P}\left(\sqrt{d(1-t)}\Delta\left|\frac{\|z_\mu\|^2}{m} - 1\right| > \varepsilon\right) \\
&\quad + \mathbb{P}\left(\left|\frac{2\sqrt{\Delta d(1-t)}}{m}\sum_{k=1}^m z_{\mu,k}\left(\frac{x_\mu^\top w_k^\star}{\sqrt{d}}\right)\right| > \varepsilon\right) + \mathbb{P}(|\sqrt{\widetilde{\Delta}t_0}\zeta_\mu| > \varepsilon) \\
&\leq e^{-C(\kappa,\Delta,t_0)\sqrt[4]{d}}
\end{aligned}$$

for some constant $C(\kappa, \Delta, t_0) > 0$. Therefore, we have

$$\mathbb{E}_{x,W^\star,z,\zeta,Y'} \left\langle \mathbb{1}\{A_\mu^\complement\} \right\rangle \leq e^{-C(\kappa,\Delta,t_0)\sqrt[4]{d}}$$

according to eq. (92), which gives

$$\left| \frac{\partial}{\partial t} f_{n,d}(t, \Lambda) \right| \leq \alpha C(\Delta, \kappa, t_0)\delta + \alpha e^{-\sqrt{C(\Delta,\kappa,t_0)}\sqrt[8]{d}} \frac{\mathrm{poly}(d)}{t_0^6}, \tag{99}$$

where we use Lemma 36. Notice that all these bounds did not depend on Λ and are valid uniformly in Λ. Therefore, we have

$$\limsup_{d\to\infty} \sup_{\Lambda \geq 0} \left| \frac{\partial}{\partial t} f_{n,d}(t, \Lambda) \right| \leq \alpha C(\Delta, \kappa, t_0)\delta,$$

which proves Lemma 34 by taking the limit $\delta \downarrow 0$.

E.4. Proof of Lemma 35

From now on we drop the index μ for notational simplicity. Recall that we have assumed

$$\left| \frac{x^\top S x}{d} - 1 \right| \leq \frac{\delta}{\sqrt[4]{d}}, \; |\sqrt{d}(\mathrm{tr} S - 1)| \leq \frac{\delta}{\sqrt[4]{d}}, \; |\bar{v}|^2 < 2t_0 \tilde{\Delta}^2 \sqrt[4]{d},$$

where $\bar{v} := \tilde{v} - \mathrm{Tr}[\Phi S^\star]$. Note that $P_{\mathrm{out}}^{(t)}(v|W, x)$ can also be written as

$$\begin{aligned} P_{\mathrm{out}}^{(t)}(v|W, x) := \frac{1}{\sqrt{2\pi\tilde{\Delta}t}} \mathbb{E}_z \exp\Bigg\{ -\frac{1}{2\tilde{\Delta}t} \Bigg(& v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr} S - 1] \\ & + \sqrt{d(1-t)}\Delta \left(\frac{\|z\|^2}{m} - 1 \right) + \frac{2\sqrt{\Delta d(1-t)}}{m} \sum_{k=1}^m z_k \left(\frac{x^\top w_k}{\sqrt{d}} \right) \Bigg)^2 \Bigg\}. \end{aligned}$$

By using the identity

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-z^2/2\sigma} = \frac{1}{2\pi} \int dp e^{-\sigma^2 p^2/2 + ipz},$$

we have

$$\begin{aligned} P_{\mathrm{out}}^{(t)}(v|W, x) &= \frac{1}{2\pi} \int dp \exp\left\{ -\frac{\tilde{\Delta}tp^2}{2} + ip\left(v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr} S - 1] \right) \right\} \\ &\quad \prod_{k=1}^m \mathbb{E}_z \exp\left\{ ip\left(\sqrt{d(1-t)}\Delta \left(\frac{z^2}{m} - 1 \right) + \frac{2\sqrt{\Delta d(1-t)}}{m} z \left(\frac{x^\top w_k^\star}{\sqrt{d}} \right) \right) \right\} \\ &= \frac{1}{2\pi} \int dp \exp\left\{ -\frac{\tilde{\Delta}tp^2}{2} + ip\left(v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr} S - 1] \right) \right\} \\ &\quad \exp\left\{ -\frac{\kappa d}{2} \log\left(1 - \frac{2ip\Delta\sqrt{1-t}}{\kappa\sqrt{d}} \right) - ip\sqrt{d(1-t)}\Delta - \frac{2p^2\Delta(1-t)x^\top S x}{\kappa d\left(1 - \frac{2ip\Delta\sqrt{1-t}}{\kappa\sqrt{d}} \right)} \right\}, \end{aligned}$$

where we explicitly calculate the expectation w.r.t. $z \sim \mathcal{N}(0,1)$, and use that $m = \kappa d$. An important limit to notice is that

$$
\begin{aligned}
g_d(t,p) &:= -\frac{\widetilde{\Delta} t p^2}{2} - \frac{\kappa d}{2}\log\left(1 - \frac{2ip\Delta\sqrt{1-t}}{\kappa\sqrt{d}}\right) - ip\sqrt{d(1-t)}\Delta - \frac{2p^2\Delta(1-t)}{\kappa\left(1-\frac{2ip\Delta\sqrt{1-t}}{\kappa\sqrt{d}}\right)} \\
&= -\frac{\widetilde{\Delta}p^2}{2} + \mathcal{O}\left(\frac{p^3}{\sqrt{d}}\right),
\end{aligned}
$$

by direct calculation. Therefore, we will truncate the integral on $|p| \le \sqrt[8]{d}$, so that that $g_d(t,p) + \frac{\widetilde{\Delta}p^2}{2} = o\left(\frac{1}{\sqrt[4]{d}}\right)$ uniformly for $|p| \le \sqrt[8]{d}$. We estimate the truncation error by the domination function

$$
\begin{aligned}
&\Bigg|\exp\left\{-\frac{\tilde{\Delta} t p^2}{2} + ip\left(v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr}S - 1]\right)\right\} \\
&\qquad \exp\left\{-\frac{\kappa d}{2}\log\left(1 - \frac{2ip\Delta\sqrt{1-t}}{\kappa\sqrt{d}}\right) - ip\sqrt{d(1-t)}\Delta - \frac{2p^2\Delta(1-t)x^\top S x}{\kappa d\left(1 - \frac{2ip\Delta\sqrt{1-t}}{\kappa\sqrt{d}}\right)}\right\}\Bigg| \\
&\le \exp\left\{-\frac{\tilde{\Delta} t p^2}{2}\right\}.
\end{aligned} \tag{100}
$$

We have

$$
\begin{aligned}
&|P_{\text{out}}^{(t)}(v|W,x)\sqrt{2\pi\tilde{\Delta}}e^{\frac{1}{2\tilde{\Delta}}(v-\mathrm{Tr}[\Phi S])^2} - 1| \\
&:= \left|\sqrt{\frac{\widetilde{\Delta}}{2\pi}}\int dp \exp\left\{ip(1-t)\sqrt{d}[\mathrm{tr}S - 1] - \frac{2p^2\Delta(1-t)}{\kappa\left(1 - \frac{2ip\Delta\sqrt{1-t}}{\kappa\sqrt{d}}\right)}\left(\frac{x^\top S x}{d} - 1\right) + g_d(t,p) + ip\left(v - \mathrm{Tr}[\Phi S]\right) + \frac{1}{2\tilde{\Delta}}(v - \mathrm{Tr}[\Phi S])^2\right\} - 1\right| \\
&\overset{(a)}{\le} \sup_{|a_1|,|a_2|,|a_3|,|a_4| \le \delta/\sqrt[4]{d}}\left|\sqrt{2\pi\tilde{\Delta}}\int_{|p|<\sqrt[8]{d}} \mathrm{d}p \exp\left\{ipa_1 + \frac{2p^2\Delta}{\kappa}a_2(1+a_3) - \frac{\widetilde{\Delta}}{2}\left(p - \frac{i}{\widetilde{\Delta}}\left(v - \mathrm{Tr}[\Phi S]\right)\right)^2 + a_4 p^2\right\} - 1\right| \\
&+ \sqrt{2\pi\tilde{\Delta}}\int_{|p|>\sqrt[8]{d}} dp e^{-\widetilde{\Delta} t p^2/2} e^{\frac{1}{2\widetilde{\Delta}}(v - \mathrm{Tr}[\Phi S])^2} \\
&\le \sup_{|a_1|,|a_2|,|a_3|,|a_4| \le \delta/\sqrt[4]{d}}\left|\sqrt{2\pi\tilde{\Delta}}\int \mathrm{d}p \exp\left\{ipa_1 + \frac{2p^2\Delta}{\kappa}a_2(1+a_3) - \frac{\widetilde{\Delta}}{2}\left(p - \frac{i}{\widetilde{\Delta}}\left(v - \mathrm{Tr}[\Phi S]\right)\right)^2 + a_4 p^2\right\} - 1\right| \\
&+ \sqrt{2\pi\tilde{\Delta}}\int_{|p|>\sqrt[8]{d}} \mathrm{d}p \left|\exp\left\{ipa_1 + \frac{2p^2\Delta}{\kappa}a_2(1+a_3) - \frac{\widetilde{\Delta}}{2}\left(p - \frac{i}{\widetilde{\Delta}}\left(v - \mathrm{Tr}[\Phi S]\right)\right)^2 + a_4 p^2\right\}\right| + \sqrt{2\pi\tilde{\Delta}}\int_{|p|>\sqrt[8]{d}} \mathrm{d}p e^{-\widetilde{\Delta} t p^2/2} e^{\frac{1}{2\widetilde{\Delta}}(v - \mathrm{Tr}[\Phi S])^2}, \\
&\overset{(b)}{\le} \sup_{|a_1|,|a_2|,|a_3|,|a_4| \le \delta/\sqrt[4]{d}}\left|\sqrt{2\pi\tilde{\Delta}}\int \mathrm{d}p \exp\left\{ipa_1 + \frac{2p^2\Delta}{\kappa}a_2(1+a_3) - \frac{\widetilde{\Delta}}{2}\left(p - \frac{i}{\widetilde{\Delta}}\left(v - \mathrm{Tr}[\Phi S]\right)\right)^2 + a_4 p^2\right\} - 1\right| \\
&+ \sqrt{2\pi\tilde{\Delta}}\int_{|p|>\sqrt[8]{d}} (1 + e^{\delta p^2}) \mathrm{d}p e^{-\widetilde{\Delta} t p^2/2} e^{\frac{1}{2\widetilde{\Delta}}(v - \mathrm{Tr}[\Phi S])^2}.
\end{aligned} \tag{101}
$$

In (a) we used the domination function of eq. (100), and the fact that

$$
\left|\frac{1}{1 - \frac{2ip\Delta\sqrt{1-t}}{\kappa\sqrt{d}}} - 1\right| \le \frac{\delta}{\sqrt[4]{d}}, \ |g_d(t,p) + \frac{\widetilde{\Delta}p^2}{2}| \le \frac{\delta}{\sqrt[4]{d}}p^2
$$

for all $p \leq \sqrt[8]{d}$ and d large enough. In (b) we use

$$\begin{aligned}
&\left|\exp\left\{ipa_1 + \frac{2p^2\Delta}{\kappa}a_2(1+a_3) - \frac{\widetilde{\Delta}}{2}\left(p - \frac{i}{\widetilde{\Delta}}\left(v - \mathrm{Tr}[\Phi S]\right)\right)^2 + a_4p^2\right\}\right| \\
&= \exp\left\{\frac{2p^2\Delta}{\kappa}a_2(1+a_3) - \frac{\widetilde{\Delta}}{2}p^2 + \frac{1}{2\widetilde{\Delta}}\left(v - \mathrm{Tr}[\Phi S]\right)^2 + a_4p^2\right\} \\
&\leq \exp\left\{-\frac{\widetilde{\Delta}}{2}p^2 + \frac{1}{2\widetilde{\Delta}}\left(v - \mathrm{Tr}[\Phi S]\right)^2 + \delta p^2\right\}
\end{aligned}$$

for large enough d. The first term on the right hand side of (101) can be controlled because

$$\begin{aligned}
&\sup_{|a_1|,|a_2|,|a_3|,|a_4| \leq \delta/\sqrt[4]{d}}\left|\sqrt{\frac{\tilde{\Delta}}{2\pi}}\int \mathrm{d}p \exp\left\{ipa_1 + \frac{2p^2\Delta}{\kappa}a_2(1+a_3) - \frac{\widetilde{\Delta}}{2}\left(p - \frac{i}{\widetilde{\Delta}}\left(v - \mathrm{Tr}[\Phi S]\right)\right)^2 + a_4p^2\right\} - 1\right| \\
&\leq \sup_{|a_1| \leq \delta/\sqrt[4]{d}, |a_2| \leq (1+4\Delta/\kappa)\delta/\sqrt[4]{d}}\left|\sqrt{\frac{\tilde{\Delta}}{2\pi}}\int \mathrm{d}p \exp\left\{ipa_1 + a_2p^2 - \frac{\widetilde{\Delta}}{2}\left(p - \frac{i}{\widetilde{\Delta}}\left(v - \mathrm{Tr}[\Phi S]\right)\right)^2\right\} - 1\right| \\
&= \sup_{|a_1| \leq \delta/\sqrt[4]{d}, |a_2| \leq (1+4\Delta/\kappa)\delta/\sqrt[4]{d}}\left|\exp\left\{\frac{\bar{v}^2}{2\widetilde{\Delta}} - \frac{(\bar{v}+a_1)^2}{2\widetilde{\Delta}(1+a_2)}\right\} - 1\right| \\
&\leq \left|e^{C(\Delta,\kappa)\delta\bar{v}^2/\sqrt[4]{d} + C(\Delta,\kappa)\delta} - 1\right| \qquad (102)
\end{aligned}$$

for some $C(\Delta, \kappa) > 0$. The second term on the right hand side of (101) can be controlled by

$$\int_{|p| > \sqrt[8]{d}} \mathrm{d}p(1 + e^{\delta p^2})e^{-\widetilde{\Delta}tp^2/2}e^{\bar{v}^2/2\widetilde{\Delta}} \leq \frac{e^{-\widetilde{\Delta}t\sqrt[4]{d}/4}}{\widetilde{\Delta}t\sqrt[4]{d}}e^{\widetilde{\Delta}t\sqrt[8]{d}} \leq C(\Delta, \kappa, t_0)\delta$$

for d large enough, where we recall that $|\bar{v}|^2 \leq 2t\widetilde{\Delta}^2\sqrt[4]{d}$. This finally gives the bound

$$|P_{\text{out}}^{(t)}(v|W, x)\sqrt{2\pi\widetilde{\Delta}}e^{\frac{1}{2\widetilde{\Delta}}(v - \mathrm{Tr}[\Phi S])^2} - 1| \leq C(\Delta, \kappa, t_0)\delta,$$

where we use $|e^x - 1| \leq 2|x|$ for x sufficiently small. Similarly, we have

$$\begin{aligned}
&\left|\frac{\partial}{\partial t}P_{\text{out}}^{(t)}(v|W, x)\right|\sqrt{2\pi\tilde{\Delta}}e^{\frac{1}{2\tilde{\Delta}}(v - \mathrm{Tr}[\Phi S])^2} \\
&= \left|\frac{1}{2\pi}\int dp \exp\left\{ip(1-t)\sqrt{d}[\mathrm{tr}S - 1] - \frac{2p^2\Delta(1-t)}{\kappa\left(1 - \frac{2ip\Delta\sqrt{1-t}}{\kappa\sqrt{d}}\right)}\left(\frac{x^\top Sx}{d} - 1\right) + g_d(t,p) + ip\left(v - \mathrm{Tr}[\Phi S]\right) + \left(v - \mathrm{Tr}[\Phi S]\right)^2/2\Delta\right\}\right. \\
&\left.\left(-p\sqrt{d}[\mathrm{tr}S - 1] - \frac{\partial}{\partial t}\frac{2p^2\Delta(1-t)}{\kappa\left(1 - \frac{2ip\Delta\sqrt{1-t}}{\kappa\sqrt{d}}\right)}\left(\frac{x^\top Sx}{d} - 1\right) + \frac{\partial}{\partial t}g_d(t,p)\right)\right| \\
&\leq C(\Delta, \kappa, t_0)\delta \qquad (103) \\
&+ \sup_{|a_i| < \delta/\sqrt[4]{d},\, i=1,\cdots,7}\left|\sqrt{2\pi\tilde{\Delta}}\int dp \exp\left\{ipa_1 + \frac{2p^2\Delta}{\kappa}a_2(1+a_3) - \frac{\widetilde{\Delta}}{2}\left(p - \frac{i}{\widetilde{\Delta}}\left(v - \mathrm{Tr}[\Phi S]\right)\right)^2 + a_4p^2\right\}\left(-pa_1 - \frac{2p^2\Delta}{\kappa}a_4(1+a_5) + a_6 + a_7p^2\right)\right|,
\end{aligned}$$

where we truncate the integral again on $|p| \leq \sqrt[8]{d}$, and estimate the truncation error in the same way. The first term on the right side of eq. (103) corresponds to the truncation error. In eq. (103)

we also use the fact that after truncation, $\left|\frac{\partial}{\partial t} g_d(t,p)\right| \leq \frac{\delta}{\sqrt[4]{d}}$ for $p \leq \sqrt[8]{d}$ and d sufficiently large. Treating the integral on the right hand side of eq. (103) in exactly the same way as above (see eq. (102)), we reach:

$$\left|\frac{\partial}{\partial t} P_{\text{out}}^{(t)}(v|W,x)\right| \sqrt{2\pi\tilde{\Delta}} e^{\frac{1}{2\tilde{\Delta}}(v-\mathrm{Tr}[\Phi S])^2} \leq C(\Delta,\kappa,t_0)\delta$$

for some $C(\Delta,\kappa,t_0) > 0$, which gives

$$\left|\frac{\partial}{\partial t} \log P_{out}^{(t)}(v|W,x)\right| = \frac{\left|\frac{\partial}{\partial t} P_{\text{out}}^{(t)}(v|W,x)\right| \sqrt{2\pi\tilde{\Delta}} e^{\frac{1}{2\tilde{\Delta}}(v-\mathrm{Tr}[\Phi S])^2}}{|P_{\text{out}}^{(t)}(v|W,x)\sqrt{2\pi\tilde{\Delta}} e^{\frac{1}{2\tilde{\Delta}}(v-\mathrm{Tr}[\Phi S])^2}|} \leq C(\Delta,\kappa,t_0)\delta$$

for some $C(\Delta,\kappa,t_0) > 0$ and sufficiently small δ. This proves Lemma 35.

E.5. Proof of Lemma 36

Starting from

$$\begin{aligned} P_{\text{out}}^{(t)}(v|W,x) \coloneqq \frac{1}{\sqrt{2\pi\tilde{\Delta}t}} \mathbb{E}_z \exp\Bigg\{-\frac{1}{2\tilde{\Delta}t}\Bigg(v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr}S - 1] \\ +\sqrt{d(1-t)}\Delta\left(\frac{\|z\|^2}{m} - 1\right) + \frac{2\sqrt{\Delta d(1-t)}}{m}\sum_{k=1}^m z_k\left(\frac{x^\top w_k}{\sqrt{d}}\right)\Bigg)^2\Bigg\}, \end{aligned}$$

we have

$$\begin{aligned} &\frac{\partial}{\partial t}\log P_{\text{out}}^{(t)}(v|W,x) \\ &= -\frac{1}{2t} + \mathbb{E}_{z\sim P_z}\left[\frac{1}{2\tilde{\Delta}t^2}\left(v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr}S-1] + \sqrt{d(1-t)}\Delta\left(\frac{\|z\|^2}{m}-1\right) + \frac{2\sqrt{\Delta d(1-t)}}{m}\sum_{k=1}^m z_k\left(\frac{x^\top w_k}{\sqrt{d}}\right)\right)^2\right] \\ &+ \mathbb{E}_{z\sim P_z}\Bigg[\frac{1}{2\tilde{\Delta}t}\left(v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr}S-1] \quad + \sqrt{d(1-t)}\Delta\left(\frac{\|z\|^2}{m}-1\right) + \frac{2\sqrt{\Delta d(1-t)}}{m}\sum_{k=1}^m z_k\left(\frac{x^\top w_k}{\sqrt{d}}\right)\right) \\ &\left(-\sqrt{d}[\mathrm{tr}S-1] - \frac{1}{2}\sqrt{\frac{d}{1-t}}\Delta\left(\frac{\|z\|^2}{m}-1\right) - \frac{1}{m}\sqrt{\frac{\Delta d}{(1-t)}}\sum_{k=1}^m z_k\left(\frac{x^\top w_k}{\sqrt{d}}\right)\right)\Bigg] \end{aligned} \tag{104}$$

where the distribution of z is given by

$$\begin{aligned} P_z(z) = \mathcal{C}\exp\Bigg\{-\frac{1}{2}z^2 - \frac{1}{2\tilde{\Delta}t}\Bigg(v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr}S-1] \\ +\sqrt{d(1-t)}\Delta\left(\frac{\|z\|^2}{m}-1\right) + \frac{2\sqrt{\Delta d(1-t)}}{m}\sum_{k=1}^m z_k\left(\frac{x^\top w_k}{\sqrt{d}}\right)\Bigg)^2\Bigg\}, \end{aligned}$$

with $\mathcal{C}$ a normalization factor. Notie that we have

$$\sup_{\|u\|=1} P_z(|u^\top z| > \xi) \leq \min\{2\mathcal{C}e^{-\xi^2/2}, 1\} \leq \min\{2e^{-\xi^2/2\log(2\mathcal{C})}, 1\}.$$

We also have, by Jensen's inequality:

$$\begin{aligned}
\log \mathcal{C} &:= -\log \int Dz \exp\left\{ -\frac{1}{2\tilde{\Delta}t} \left(v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr}S - 1] + \sqrt{d(1-t)}\Delta \left(\frac{\|z\|^2}{m} - 1 \right) + \frac{2\sqrt{\Delta d(1-t)}}{m} \sum_{k=1}^m z_k \left(\frac{x^\top w_k}{\sqrt{d}} \right) \right)^2 \right\} \\
&\leq \frac{1}{2\tilde{\Delta}t} \mathbb{E}_{z\sim\mathcal{N}(0,1)} \left[\left(v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr}S - 1] + \sqrt{d(1-t)}\Delta \left(\frac{\|z\|^2}{m} - 1 \right) + \frac{2\sqrt{\Delta d(1-t)}}{m} \sum_{k=1}^m z_k \left(\frac{x^\top w_k}{\sqrt{d}} \right) \right)^2 \right] \\
&= \frac{1}{2\tilde{\Delta}t} \left((v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr}S - 1])^2 + \frac{2(1-t)\Delta^2}{\kappa} + \frac{4(1-t)\Delta}{\kappa} \frac{x^\top S x}{d} \right).
\end{aligned}$$

Therefore P_z is a sub-Gaussian distribution with variance proxy $\log(2\mathcal{C})$. Its moments are thus bounded by

$$\mathbb{E}_{z\sim P_z}[\|z\|^p] \leq m(\log(2\mathcal{C}))^p p^{p/2} =: mM_p.$$

Taking it into eq. (104), we have

$$\begin{aligned}
&\left(\frac{\partial}{\partial t} \log P_{\mathrm{out}}^{(t)}(v|W,x) \right)^2 \\
&\leq \frac{9}{4t^2} + \frac{81}{4\tilde{\Delta}^2 t^4} \left((v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr}S - 1])^2 + d(1-t)\Delta^2 \left(\frac{M_4}{m} + 2M_2 + 1 \right) + 4\Delta d(1-t) m M_2 \sum_{k=1}^m \left(\frac{x^\top w_k}{\sqrt{d}} \right)^2 \right)^2 \\
&+ \frac{9}{4\tilde{\Delta}^2 t^2} (v - \mathrm{Tr}[\Phi S] + (1-t)\sqrt{d}[\mathrm{tr}S - 1])^2 d(\mathrm{tr}S - 1)^2 + \frac{9}{\tilde{\Delta}^2 t^2} \left(d(\mathrm{tr}S - 1)^2 + d\Delta^2 \left(\frac{M_4}{m} + 2M_2 + 1 \right) \right)^2 \\
&+ \frac{9}{4\tilde{\Delta}^2 t^2} \left(v - \mathrm{Tr}[\Phi S] + \frac{3}{2}(1-t)\sqrt{d}[\mathrm{tr}S - 1] \right)^2 \left(\sqrt{\frac{d}{1-t}} \Delta (M_2 + 1) + 2\frac{1}{m} \sqrt{\frac{\Delta d}{1-t}} M_1 \sum_{k=1}^m \left(\frac{x^\top w_k}{\sqrt{d}} \right) \right)^2 \\
&\leq \frac{\mathrm{poly}\left(v, \mathrm{Tr}[\Phi S], \mathrm{tr}S, x^\top S x, \{x^\top w_k\}_{k=1}^m \right)}{t^{12}},
\end{aligned}$$

where we uses eq. (67). An important observation is that

$$\mathbb{E}_{x,W^\star,z,\zeta,Y'} \left\langle \mathrm{poly}\left(v, \mathrm{Tr}[\Phi S], \mathrm{tr}S, x^\top S x, \{x^\top w_k\}_{k=1}^m \right) \right\rangle$$

can be bounded by polynomials of d. To see this, consider variables $(A, A^\star, B, B^\star)$ (e.g. $(x^\top S x, x^\top S^\star x)$). We have by the Cauchy-Schwarz inequality and the Nishimori identity (Proposition 43):

$$\left| \mathbb{E} \left\langle \sum_{i,j=1}^K a_{ij} A^i B^j \right\rangle \right| \leq \sum_{i,j=1}^K |a_{ij}| \sqrt{\mathbb{E}\langle A^{2i} \rangle \mathbb{E}\langle B^{2j} \rangle} = \sum_{i,j=1}^K |a_{ij}| \sqrt{\mathbb{E}[(A^\star)^{2i}] \mathbb{E}[(B^\star)^{2j}]}$$

Therefore, the expectation of polynomials of A, B is bounded by polynomials of d if the moments of A^*, B^* are bounded by polynomials of d. As the moments of $v, \mathrm{Tr}[\Phi S^\star], x^\top S^\star x, \{x^\top w_k^\star\}_{k=1}^m$ are bounded by polynomials of d, we have

$$\begin{aligned}
\mathbb{E}_{x,W^\star,z,\zeta,Y'} \left\langle \left(\frac{\partial}{\partial t} \log P_{\mathrm{out}}^{(t)}(\tilde{v}(t,x,W^\star,z,\zeta)|W,x_i) \right)^2 \right\rangle &\leq \mathbb{E}_{x,W^\star,z,\zeta,Y'} \left\langle \frac{\mathrm{poly}(\tilde{v}, W, x)}{t^{12}} \right\rangle \\
&\leq \frac{\mathrm{poly}(d)}{t_0^{12}},
\end{aligned}$$

which finishes the proof of Lemma 36.

Appendix F. Spiked tensor model

F.1. Main results

This section presents the results concerning the following "spiked tensor" model

$$Y = \sqrt{\frac{\lambda}{d^{p-2}}}(S^\star)^{\otimes p} + Z,$$

where $S \sim P_0$ is a symmetric matrix and $Z \in \mathcal{S}_d^{\otimes p}$ has i.i.d. $\mathcal{N}(0,1)$ elements. $p \geq 2$ is an integer. This model is primarily used for the proof of Theorem 3 in our paper, but it might be of independent interest.

Its free entropy is given by

$$f_d^{\text{spike}} := \frac{1}{d^2}\mathbb{E}_{S^\star,Z}\left[\log \int P_0(\mathrm{d}S)e^{-H(S;S^\star,Z)}\right], \tag{105}$$

where

$$H(S;S^\star,Z) := \frac{1}{2}\sum_{i_1,j_1,\cdots,i_p,j_p}\left(\frac{\lambda}{2d^{p-2}}S_{i_1,j_1}^2\cdots S_{i_p,j_p}^2 - \frac{\lambda}{d^{p-2}}S_{i_1,j_1}S_{i_1,j_1}^\star\cdots S_{i_p,j_p}S_{i_p,j_p}^\star \right.$$
$$\left. -\sqrt{\frac{\lambda}{d^{p-2}}}Z_{i_1,j_1,\cdots,i_p,j_p}S_{i_1,j_1}\cdots S_{i_p,j_p}\right).$$

The RS potential is defined as

$$f_{\text{RS}}^{\text{spike}}(q) := -\frac{1}{2}\lambda(p-1)q^p + \psi_{P_0}(2p\lambda q^{p-1}) + \frac{1}{4} + \frac{1}{2}p\lambda\rho q^{p-1}, \tag{106}$$

where ψ_{P_0} is defined in eq. (6). The following theorem states our main results regarding the asymptotics of the free entropy.

Theorem 37 *Under Assumption 2.1,*

$$\lim_{d\to\infty} f_d^{\textit{spike}} = \inf_{q\in[0,\rho]} f_{RS}^{\textit{spike}}(q).$$

We notice that Theorem 37 can be generalized to the rectangular setting (i.e. $S \in \mathbb{R}^{d\times L}$) in a straightforward manner. We now focus on the proof of Theorem 37.

F.2. Interpolation model

For notational simplicity, we consider here $p = 2$, but it is easy to generalize the proof to any $p \geq 2$ (see (Barbier and Macris, 2019, Section 3)). We will first prove Theorem 37 under Assumption 3.1. All lemmas in Sections F.2 and F.3 are under Assumptions 3.1, and we then relax this assumption to Assumption 2.1 in Section F.4.

The interpolation model reads

$$Y = \sqrt{\lambda(1-t)}S^\star \otimes S^\star + Z$$
$$Y' = \sqrt{d}(\sqrt{R_d(t,\epsilon)}S^\star + Z'),$$

with $Z' \sim \mathrm{GOE}(d)$ independent of Z. The interpolating free entropy reads

$$f_{d,\epsilon}(t) \coloneqq \frac{1}{d^2}\mathbb{E}\log \mathcal{Z}_{t,\epsilon}(Y,Y') \coloneqq \frac{1}{d^2}\mathbb{E}\log \int P_0(ds)e^{-H_{t,\epsilon}(s,Y,Y')},$$

where

$$\begin{aligned} H_{t,\epsilon}(s,Y,Y') \coloneqq \sum_{i_1,j_1,i_2,j_2=1}^{d} &(1-t)\lambda\frac{s_{i_1j_1}^2 s_{i_2j_2}^2}{2} - \sqrt{(1-t)\lambda}s_{i_1j_1}s_{i_2j_2}Y_{i_1j_1i_2j_2} \\ &+ \frac{1}{4}\sum_{i,j=1}^{d}(Y'_{ij} - \sqrt{dR_d(t,\epsilon)}s_{ij})^2. \end{aligned} \tag{107}$$

and

$$R_d(t,\epsilon) \coloneqq \epsilon + 4\lambda\int_0^t \mathrm{d}u\, q_d(u,\epsilon),$$

with $\epsilon \in [\iota_d, 2\iota_d]$ and $\iota_d \coloneqq \frac{1}{2}d^{-1/8}$. Accordingly, the Gibbs bracket is defined as

$$\langle g(s)\rangle \coloneqq \frac{1}{\mathcal{Z}_{t,\epsilon}(Y,Y')}\int P_0(\mathrm{d}s)g(s)e^{-H_{t,\epsilon}(s,Y,Y')}.$$

Lemma 38

$$f_{d,\epsilon}(0) = f_d - \frac{1}{4} + O(\iota_d).$$

$$f_{d,\epsilon}(1) = \psi_{P_0}\left(4\lambda\int_0^1 q(t)\mathrm{d}t\right) + O(\iota_d).$$

Proof [Proof of Lemma 38] It is analogous to Lemma 22. We obtain the first equation from

$$\left|\frac{\mathrm{d}f_{d,\epsilon}(0)}{\mathrm{d}\epsilon}\right| = \frac{1}{4d}\left|\mathbb{E}\sum_{i,j=1}^{d}\langle\epsilon(S^\star_{ij} - s_{ij}) + \sqrt{\frac{1}{\epsilon}}(S^\star_{ij} - s_{ij})Z'_{ij}\rangle\right| \le M^2$$

and $f_{d,0}(0) = f_d - \frac{1}{4}$. We obtain the second equality from

$$f_{d,\epsilon}(1) = \psi_{P_0}(R_d(1,\epsilon))$$

and the Lipschitz property of ψ_{P_0} (Lemma 49). ■

Lemma 39 *Denote* $f_{\mathrm{RS}} = f_{\mathrm{RS}}^{\textit{spike}}$ *given in eq.* (106). *Then:*

$$f_d = f_{\mathrm{RS}}\left(\int_0^1 q(t)\mathrm{d}t\right) - \frac{\lambda}{2}(\mathcal{R}_1 - \mathcal{R}_2 - \mathcal{R}_3) + O(\iota_d),$$

where

$$\mathcal{R}_1 \coloneqq \int_0^1 \mathrm{d}t\left(q(t) - \int_0^1 \mathrm{d}t' q(t')\right)^2,$$

$$\mathcal{R}_2 \coloneqq \int_0^1 \mathrm{d}t\mathbb{E}\langle(Q - \mathbb{E}\langle Q\rangle)^2\rangle,$$

$$\mathcal{R}_3 \coloneqq \int_0^1 \mathrm{d}t(\mathbb{E}\langle Q\rangle - q(t))^2.$$

Proof [Proof of Lemma 39] Taking the derivative w.r.t t and integrating by parts, we have

$$\frac{\mathrm{d}f_{d,\epsilon}(t,\epsilon)}{\mathrm{d}t} = -\frac{\lambda}{2d^2}\sum_{i_1,j_1,i_2,j_2}\mathbb{E}[S^\star_{i_1j_1}S^\star_{i_2j_2}\langle s_{i_1j_1}s_{i_2j_2}\rangle] - \frac{\lambda q(t)}{d}\mathbb{E}\sum_{i,j=1}^{d}(S^\star_{ij}-\langle s_{ij}\rangle)^2.$$

Using the Nishimori identity (Proposition 43), we have

$$\frac{\mathrm{d}f_{d,\epsilon}(t,\epsilon)}{\mathrm{d}t} = -\frac{\lambda}{2}\mathbb{E}\langle Q^2\rangle - \lambda q(t)(\rho-\mathbb{E}\langle Q\rangle).$$

By Lemma 38, we have

$$\begin{aligned} f_d &= \psi_{P_0}\left(4\lambda\int_0^1 q(t)\mathrm{d}t\right) + \frac{1}{4} + \int_0^1 \mathrm{d}t\left[\frac{\lambda}{2}\mathbb{E}\langle Q^2\rangle + \lambda q(t)(\rho-\mathbb{E}\langle Q\rangle)\right] + O(\iota_d), \\ &= f_{\mathrm{RS}}\left(\int_0^1 q(t)\mathrm{d}t\right) + \frac{\lambda}{2}\left(\int_0^1 \mathrm{d}t q(t)\right)^2 + \frac{\lambda}{2}\int_0^1 \mathrm{d}t\mathbb{E}\langle Q^2\rangle - \lambda\int_0^1 \mathrm{d}t q(t)\mathbb{E}\langle Q\rangle + O(\iota_d), \end{aligned}$$

which gives the desired result. ■

We will need the following concentration lemma, which will be proven in Section F.3.

Lemma 40 *There exists a constant $C(M,\kappa)>0$ such that*

$$\frac{1}{\iota_d}\int_{\iota_d}^{2\iota_d}\mathrm{d}\epsilon\int_0^1 \mathrm{d}t\mathbb{E}\langle(Q-\mathbb{E}\langle Q\rangle)^2\rangle \le \frac{C(M,\kappa)}{d^{1/4}}.$$

Combining Lemmas 39 and 40, we can obtain Theorem 37 under Assumption 3.1, following the standard choice of interpolation (Barbier and Macris (2019), i.e. $q(t)=q$ and $q(t)=\mathbb{E}\langle Q\rangle$), by repeating the argument used in the proof of Theorem 1. Lemma 42 allows then to relax Assumption 3.1.

F.3. Concentration of the overlap: proof of Lemma 40

We first prove the concentration of the free entropy. We can rewrite the free entropy as

$$\frac{1}{d^2}\log\mathcal{Z}_{t,\epsilon} = \frac{1}{d^2}\log\hat{\mathcal{Z}}_{t,\epsilon} - \frac{1}{4d}\sum_{i,j=1}^{d}(Z'_{ij})^2,$$

where

$$\frac{1}{d^2}\log\hat{\mathcal{Z}}_{t,\epsilon} := \frac{1}{d^2}\log\int P_0(\mathrm{d}s)e^{-\hat{H}_t(s)},$$

and

$$\begin{aligned} \hat{H}_t(s) := \frac{1}{4}\sum_{i_1,j_1,i_2,j_2=1}^{d} &(1-t)\lambda\frac{s^2_{i_1j_1}s^2_{i_2j_2}}{2} - (1-t)\lambda s_{i_1j_1}s_{i_2j_2}S^\star_{i_1j_1}S^\star_{i_2j_2} \\ &- \sqrt{(1-t)\lambda}s_{i_1j_1}s_{i_2j_2}Z_{i_1j_1i_2j_2} + \frac{d}{4}\sum_{i,j=1}^{d}\left[R_1(t)(S^\star_{ij}-s_{ij})^2 + 2Z'_{ij}\sqrt{R_1(t)}(S^\star_{ij}-s_{ij})\right]. \end{aligned}$$

The concentration of the free entropy results from the following lemma.

Lemma 41

$$\textit{Var}\left(\frac{1}{d^2}\log\hat{\mathcal{Z}}_{t,\epsilon}\right) \leq \frac{C(\varphi, M, \alpha, \kappa)}{d^2}.$$

Proof [Proof of Lemma 41] We first consider $g := \log\hat{\mathcal{Z}}_{t,\epsilon}/d^2$ as a function of Z, Z'. We have

$$\begin{aligned}\sum_{i_1,j_1,i_2,j_2=1}^{d}\left(\frac{\partial g}{\partial Z_{i_1j_1i_2j_2}}\right)^2 &= d^{-4}(1-t)\lambda\sum_{i_1,j_1,i_2,j_2=1}^{d}\langle s_{i_1j_1}s_{i_2j_2}\rangle^2 \\ &\leq d^{-4}(1-t)\lambda\sum_{i_1,j_1,i_2,j_2=1}^{d}\langle s_{i_1j_1}^2 s_{i_2j_2}^2\rangle \leq d^{-2}M^4,\end{aligned}$$

where the first inequality is from Jensen's inequality, and the last inequality is a consequence of

$$\sum_{i_1,j_1,i_2,j_2=1}^{d} s_{i_1j_1}^2 s_{i_2j_2}^2 = \left(\sum_{i,j=1}^{d} s_{ij}^2\right)^2 \leq d^2M^4.$$

We also have

$$\sum_{i,j=1}^{d}\left(\frac{\partial g}{\partial Z'_{ij}}\right)^2 = d^{-2}R(t)\sum_{i,j=1}^{d}(S^\star_{ij} - \langle s_{ij}\rangle)^2 \leq 4d^{-1}KM^2,$$

where $K := 1 + \lambda\rho$ upper bounds $R(t)$. By Lemma 48 we obtain

$$\mathbb{E}\left[\left(\frac{1}{d^2}\log\hat{\mathcal{Z}}_{t,\epsilon} - \frac{1}{d^2}\mathbb{E}_{Z,Z'}\log\hat{\mathcal{Z}}_{t,\epsilon}\right)^2\right] \leq \frac{C(M)}{d^2}. \tag{108}$$

Next we write $S^\star = O\Lambda^\star O^T$ with O drawn from the Haar measure on $\mathcal{O}(d)$, independently of $\Lambda^\star$. We denote $\tilde{s} := O^T sO$ and $\widetilde{Z} = O^T ZO$, $\widetilde{Z}' = O^T\widetilde{Z}'O$. Notice that we have $\Lambda^\star = \mathrm{Diag}(\{\Lambda^\star_i\}_{i=1}^d)$: we will abuse a bit notations and denote $\Lambda^\star_{ij} := \Lambda^\star_i\delta_{ij}$. In this way the Hamitonian reads

$$\begin{aligned}\hat{H}_t(\tilde{s}) := \frac{1}{4}\sum_{i_1,j_1,i_2,j_2=1}^{d}\Bigg[&(1-t)\lambda\frac{\tilde{s}_{i_1j_1}^2\tilde{s}_{i_2j_2}^2}{2} - (1-t)\lambda\tilde{s}_{i_1j_1}\tilde{s}_{i_2j_2}\Lambda^\star_{i_1j_1}\Lambda^\star_{i_2j_2} \\ &-\sqrt{(1-t)\lambda}\tilde{s}_{i_1j_1}\tilde{s}_{i_2j_2}\widetilde{Z}_{i_1j_1i_2j_2}\Bigg] + \frac{d}{4}\sum_{i,j=1}^{d}(R_1(t)(\Lambda^\star_{ij} - \tilde{s}_{ij})^2 + 2\widetilde{Z}'_{ij}\sqrt{R_1(t)}(\Lambda^\star_{ij} - \tilde{s}_{ij})).\end{aligned}$$

Note that the distribution of $\tilde{s}$ is the same as that of s (by rotation invariance of P_0), so we have

$$\mathbb{E}_{O,Z,Z'}\frac{1}{d^2}\log\hat{\mathcal{Z}}_{t,\epsilon} = \mathbb{E}_{\widetilde{Z},Z'}\frac{1}{d^2}\log\int P_0(\mathrm{d}\tilde{s})Dwe^{-\hat{H}_t(\tilde{s},w)}.$$

We can now use a very similar argument to the one detailed around eq. (65): as $(Z, Z') \stackrel{\mathrm{d}}{=} (\widetilde{Z}, \widetilde{Z}')$ (meaning equality in law), we have

$$\begin{aligned} &\mathbb{E}\left[\left(\frac{1}{d^2}\mathbb{E}_{O,Z,Z'} \log \hat{\mathcal{Z}}_{t,\epsilon} - \frac{1}{d^2}\mathbb{E}_{Z,Z'} \log \hat{\mathcal{Z}}_{t,\epsilon}\right)^2\right] \\ &\leq \mathbb{E}\left[\left(\frac{1}{d^2}\mathbb{E}_{\widetilde{Z},\widetilde{Z}'} \log \int P_0(\mathrm{d}\tilde{s}) Dw e^{-\hat{H}_t(\tilde{s},w)} - \frac{1}{d^2} \log \int P_0(\mathrm{d}\tilde{s}) Dw e^{-\hat{H}_t(\tilde{s},w)}\right)^2\right], \\ &\leq \frac{C(M)}{d^2}. \end{aligned} \tag{109}$$

We used eq. (108) in the last inequality. Finally we consider $g := \mathbb{E}_{O,Z,Z'} \frac{1}{d^2} \log \hat{\mathcal{Z}}_{t,\epsilon}$ as a function of $\Lambda^\star$. We have

$$\begin{aligned} \sum_{i=1}^d \left(\frac{\partial g}{\partial \Lambda_i^\star}\right)^2 &= d^{-4} \sum_{i=1}^d \left(\mathbb{E}_{O,Z,Z'} \left\langle \frac{\partial \hat{H}_t}{\partial \Lambda_i^\star} \right\rangle\right)^2, \\ &= d^{-4} \sum_{i=1}^d \left(\mathbb{E}_{O,Z,Z'} \left\langle -\frac{(1-t)\lambda}{2} \sum_{j=1}^n \tilde{s}_{ii}\tilde{s}_{jj}\Lambda_j^\star + \frac{dR_1(t)}{2}(\Lambda_i^\star - \tilde{s}_{ii}) + 2\widetilde{Z}'_{ii}\sqrt{dR_1(t)} \right\rangle\right)^2, \\ &\leq 9(I_1 + I_2 + I_3). \end{aligned}$$

The first term is

$$\begin{aligned} I_1 &:= \frac{\lambda^2}{4} d^{-4} \sum_{i=1}^d \left(\mathbb{E}_{O,Z,Z'} \left\langle \sum_{j=1}^d \tilde{s}_{ii}\tilde{s}_{jj}\Lambda_j^\star \right\rangle\right)^2 \\ &\leq \frac{\lambda^2}{4} d^{-2} M^4 \sum_{i=1}^d \left(\mathbb{E}_{O,Z,Z'} \langle \tilde{s}_{ii} \rangle\right)^2 \leq \frac{\lambda^2}{4} d^{-2} M^4 \mathbb{E}_{O,Z,Z'} \left\langle \sum_{i=1}^d \tilde{s}_{ii}^2 \right\rangle \leq \frac{\lambda^2}{4} d^{-1} M^6, \end{aligned}$$

where we use $\left|\sum_{j=1}^d \tilde{s}_{jj}\Lambda_{jj}^\star\right| \leq \sqrt{\sum_{j=1}^d (\Lambda_{jj}^\star)^2}\sqrt{\sum_{j=1}^d \tilde{s}_{jj}^2} \leq dM^2$ for the first inequality, and Jensen's inequality for the second inequality. The second and third terms I_2, I_3 are the same as in Lemma 29, see eqs. (68), (69). Combining these three terms, we have

$$\mathbb{E}\left[\left(\frac{1}{d^2}\mathbb{E}_{O,Z,Z'} \log \hat{\mathcal{Z}}_{t,\epsilon} - \frac{1}{d^2}\mathbb{E} \log \hat{\mathcal{Z}}_{t,\epsilon}\right)^2\right] \leq \frac{C(M,\kappa)}{d^2} \tag{110}$$

by Lemma 45. We finish the proof by combining eqs. (108), (109) and (110). ■

We now sketch how we can end the proof of Lemma 40 from Lemma 41. Let (recall eq. (107)):

$$\mathcal{L} := \frac{1}{d^2}\frac{dH_{t,\epsilon}}{dR} = \frac{1}{2d}\sum_{i,j=1}^d \left(\frac{s_{ij}^2}{2} - s_{ij}S_{ij}^\star - \frac{s_{ij}Z'_{ij}}{2\sqrt{R}}\right).$$

We can now observe that $\mathcal{L}$ is exactly the same quantity analyzed in Section C.3: directly repeating the analysis given there to this setting yields Lemma 40.

F.4. Relaxation of Assumption 3.1

We now show how to relax Assumption 3.1 to Assumption 2.1 in the proof of Theorem 37.

Lemma 42 *Denote $f_d^{spike}(V)$ to be the free entropy of eq. (105) corresponding to the prior P_0 with potential V, and denote $\widetilde{V}_M$ the truncation of V to $[-M, M]$, with $\widetilde{V}_M(x) = +\infty$ if $|x| > M$. Notice that then the prior with potential $\widetilde{V}$ satisfies Assumption 3.1. Suppose that V satisfies Assumption 2.1. Then, for $M > 0$ large enough, we have*

$$\lim_{d\to\infty} |f_d^{spike}(V) - f_d^{spike}(\widetilde{V}_M)| = 0.$$

Proof [Proof of Lemma 42] Following Lemma 32, we consider the interpolation

$$\begin{aligned} Y_1 &= \sqrt{\lambda t} S_1^\star \otimes S_1^\star + Z_1, \\ Y_2 &= \sqrt{\lambda(1-t)} S_2^\star \otimes S_2^\star + Z_2, \end{aligned}$$

The joint distribution of $S_1^\star, S_2^\star$ is the same as in Lemma 32. Specially, $S_1^\star$ satisfies Assumption 2.1 and $S_2^\star$ satisfies Assumption 3.1. Its free entropy is defined as

$$f_d(t) := \frac{1}{d^2}\mathbb{E}\log \mathcal{Z}_t(Y_1, Y_2) := \frac{1}{d^2}\mathbb{E}\log \int P_0(\mathrm{d}s_1, \mathrm{d}s_2) e^{-H_t(s_1,s_2,Y_1,Y_2)},$$

where

$$\begin{aligned} H_t(s_1, s_2, Y_1, Y_2) := &\frac{1}{4}\sum_{i_1,j_1,i_2,j_2=1}^{\mathrm{d}} t\lambda \frac{s_{1,i_1j_1}^2 s_{1,i_2j_2}^2}{2} - \sqrt{t\lambda} s_{1,i_1j_1} s_{1,i_2j_2} Y_{1,i_1j_1i_2j_2} \\ &+ \frac{1}{4}\sum_{i_1,j_1,i_2,j_2=1}^{d} (1-t)\lambda \frac{s_{2,i_1j_1}^2 s_{2,i_2j_2}^2}{2} - \sqrt{(1-t)\lambda} s_{2,i_1j_1} s_{2,i_2j_2} Y_{2,i_1j_1i_2j_2}. \end{aligned}$$

Thus we have $f_d^{\text{spike}}(V) = f(1)$ and $f_d^{\text{spike}}(\widetilde{V}_M) = f(0)$. Accordingly, the Gibbs bracket is defined as

$$\langle g(s_1, s_2)\rangle := \frac{1}{\mathcal{Z}_t(Y_1, Y_2)} \int P_0(\mathrm{d}s_1, \mathrm{d}s_2) g(s_1, s_2) e^{-H_t(s_1,s_2,Y_1,Y_2)}.$$

Taking the derivative and integrating by part, we obtain

$$f_d'(t) = -\frac{\lambda}{4d^2}\mathbb{E}\left[\sum_{i_1,j_1,i_2,j_2=1}^{d} S_{1,i_1j_1}^\star S_{1,i_2j_2}^\star \langle s_{1,i_1j_1} s_{1,i_2j_2}\rangle - S_{2,i_1j_1}^\star S_{2,i_2j_2}^\star \langle s_{2,i_1j_1} s_{2,i_2j_2}\rangle\right],$$

and thus

$$
\begin{aligned}
|f_d'(t)| &\leq \frac{\lambda}{4d^2}\mathbb{E}\left[\left|\sum_{i_1,j_1,i_2,j_2=1}^{d} S^\star_{1,i_1j_1}S^\star_{1,i_2j_2}\langle s_{1,i_1j_1}s_{1,i_2j_2}\rangle - S^\star_{1,i_1j_1}S^\star_{1,i_2j_2}\langle s_{2,i_1j_1}s_{2,i_2j_2}\rangle\right|\right. \\
&\qquad \left.+\left|\sum_{i_1,j_1,i_2,j_2=1}^{d} S^\star_{1,i_1j_1}S^\star_{1,i_2j_2}\langle s_{2,i_1j_1}s_{2,i_2j_2}\rangle - S^\star_{2,i_1j_1}S^\star_{2,i_2j_2}\langle s_{2,i_1j_1}s_{2,i_2j_2}\rangle\right|\right] \\
&\leq \frac{\lambda}{4d^2}\left(\mathbb{E}\|S_1^\star \otimes S_1^\star\|^2\mathbb{E}\langle\|s_1\otimes s_1 - s_2\otimes s_2\|^2\rangle\right)^{1/2} \\
&\qquad + \frac{\lambda}{4d^2}\left(\mathbb{E}\|S_1^\star \otimes S_1^\star - S_2^\star\otimes S_2^\star\|^2\mathbb{E}\langle\|s_2\otimes s_2\|^2\rangle\right)^{1/2} \\
&= \frac{\lambda}{4d^2}((\mathbb{E}[\|\Lambda_1^\star\|^4])^{1/2} + (\mathbb{E}[\|\Lambda_2^\star\|^4])^{1/2})\left(\mathbb{E}\|S_1^\star \otimes S_1^\star - S_2^\star\otimes S_2^\star\|^2\right)^{1/2},
\end{aligned}
$$

where we used the Cauchy-Schwarz inequality and the Nishimori identity (Proposition 43). Recall that $\Lambda_1^\star$, $\Lambda_2^\star$ are the diagonal matrices of eigenvalues of $S_1^\star, S_2^\star$. The tensor product can be expanded as

$$
\begin{aligned}
\|S_1^\star \otimes S_1^\star - S_2^\star\otimes S_2^\star\|^2 &= \|(O\otimes O)(\Lambda_1^\star\otimes\Lambda_1^\star - \Lambda_2^\star\otimes\Lambda_2^\star)(O^T\otimes O^T)\|^2 \\
&= \|\Lambda_1^\star\otimes\Lambda_1^\star - \Lambda_2^\star\otimes\Lambda_2^\star\|^2 \\
&\leq 2(\|\Lambda_1^\star\|^2 + \|\Lambda_2^\star\|^2)\|\Lambda_1^\star - \Lambda_2^\star\|^2,
\end{aligned}
$$

which gives (using Proposition 13 and the Cauchy-Schwarz inequality)

$$
|f_d'(t)| \leq C(P_0)\lambda\left(\frac{1}{d^2}\mathbb{E}\|\Lambda_1^\star - \Lambda_2^\star\|^4\right)^{1/4}.
$$

Notice that we used Proposition 13(iv) that gives that $\mathbb{E}[\|\Lambda^\star\|^{2p}] \leq C(P_0,p)\cdot d^p$. Following the arguments in Lemma 32, and taking the notations from this paragraph, we obtain for a well-chosen coupling of $(S_1^\star, S_2^\star)$:

$$
|f_d'(t)| \leq C(P_0)\lambda\left(\mathbb{E}[W_2(\mu_{S_1},\mu_{S_2})^4]\right)^{1/4}.
$$

The arguments detailed in the proof of Lemma 32 can be repeated to obtain

$$
\lim_{d\to\infty}\mathbb{E}[W_2(\mu_{S_1},\mu_{S_2})^4] = 0,
$$

which gives then $\lim_{d\to\infty}|f_d(0)-f_d(1)| = 0$, and proves Lemma 42. ■

Finally it is worth noticing that for a general p, we will need to upper bound $\mathbb{E}\|\Lambda^\star\|^{2p}$: this bound is given by Proposition 13(iv) as we mentioned, so that the proof arguments sketched above directly generalize to any $p \geq 3$.

Appendix G. Technical lemmas

This section contains several technical lemmas, which are used throughout the proofs of our main results.

G.1. Nishimori identity

We state here the Nishimori identity, a classical consequence of Bayes optimality.

Proposition 43 (Nishimori identity) *Let (X, Y) be random variables on a Polish space E. Let $k \in \mathbb{N}^\star$ and $(X_1, \cdots, X_k)$ i.i.d. random variables sampled from the conditional distribution $\mathbb{P}(X|Y)$. We denote $\langle \cdot \rangle_Y$ the average with respect to $\mathbb{P}(X|Y)$, and $\mathbb{E}[\cdot]$ the average with respect to the joint law of (X, Y). Then, for all $f : E^{k+1} \to \mathbb{K}$ continuous and bounded:*

$$\mathbb{E}[\langle f(Y, X_1, \cdots, X_k)\rangle_Y] = \mathbb{E}[\langle (Y, X_1, \cdots, X_{k-1}, X)\rangle_Y]. \tag{111}$$

Proof [Proof of Proposition 43] The proposition arises as a trivial consequence of Bayes' formula:

$$\begin{aligned}\mathbb{E}[\langle f(Y, X_1, \cdots, X_{k-1}, X)\rangle_Y] &= \mathbb{E}_Y \mathbb{E}_{X|Y}[\langle f(Y, X_1, \cdots, X_{k-1}, X)\rangle_Y], \\ &= \mathbb{E}_Y[\langle f(Y, X_1, \cdots, X_k)\rangle_Y].\end{aligned}$$

■

G.2. Potential truncation

Lemma 44 (Potential Truncation) *Let the potential V satisfy Assumption 2.1, and $\widetilde{V}_M$ be the truncation of V to $[-M, M]$, according to Assumption 3.1, i.e. $\widetilde{V}_M(x) = +\infty$ for $|x| > M$. There is $M > 0$ large enough (depending only on V) such that if S is sampled according to a prior P_0 with potential $\widetilde{V}_M$, then μ_S almost surely weakly converges to μ_0.*

Proof [Proof of Lemma 44] An important observation is that

$$\mathbb{P}_{\widetilde{V}}(\{\lambda_i\}_{i=1}^d) \propto \mathbb{P}_V(\{\lambda_i\}_{i=1}^d) \mathbb{1}\{\{\lambda_i\}_{i=1}^d \subset [-M, M]\},$$

where $\mathbb{P}_{\widetilde{V}}, \mathbb{P}_V$ are the joint eigenvalue distributions under the priors with potential $\widetilde{V}_M, V$, respectively. Let dist denote a distance that metrizes the weak topology, then (Anderson et al., 2010, Theorem 2.6.1) shows that

$$\mathbb{P}_V(\text{dist}(\mu_S, \mu_0) > \epsilon) \leq e^{-C(\epsilon)d^2},$$

for any $\epsilon > 0$ and some $C(\epsilon) > 0$. Thus,

$$\mathbb{P}_{\widetilde{V}}(\text{dist}(\mu_S, \mu_0) > \epsilon) = \frac{\mathbb{P}_V(\text{dist}(\mu_S, \mu_0) > \epsilon \ \ \& \ \ \{\lambda_i\}_{i=1}^d \subset [-M, M])}{\mathbb{P}_V(\{\lambda_i\}_{i=1}^d \subset [-M, M])} \leq 2e^{-C(\epsilon)d^2}$$

for d large enough and a large enough constant $M > 0$, where we used Lemma 14. By the Borel–Cantelli lemma, we get

$$\mathbb{P}_{\widetilde{V}}(\limsup_{d\to\infty} \text{dist}(\mu_S, \mu_0) > \epsilon) = 0,$$

which finishes the proof. ■

G.3. Poincaré inequality for rotionally invariant priors

The following result is critical in our analysis, and is a direct consequence of existing results Anderson et al. (2010); Chafaï and Lehec (2020).

Lemma 45 (Poincaré Inequality for Rotionally Invariant Priors) *Under Assumption 2.1, denote the eigenvalues of S to be $\{\lambda_i\}_{i=1}^d$ and their joint distribution to be P_S. Let $g : \mathbb{R}^d \to \mathbb{R}$ a function in $L^2(P_S)$ whose weak derivative (in the sense of distributions) belongs to $L^2(P_S)$. Then:*

$$\mathit{Var}(g(\{\lambda_i\}_{i=1}^d)) \leq \frac{c}{\kappa d}\mathbb{E}\left[\sum_{i=1}^d \left(\frac{\partial g}{\partial \lambda_i}\right)^2\right], \tag{112}$$

where c is the constant given in Assumption 2.1 (note that one should take $\kappa = 1$ for the first case of Assumption 2.1). The same holds under Assumption 2.4, replacing the eigenvalues $\{\lambda_i\}_{i=1}^d$ by the singular values $\{\sigma_i\}_{i=1}^d$.

Proof [Proof of Lemma 45] For the first case of Assumption 2.1, the joint eigenvalue distribution can be written as

$$P_S(\{\lambda_i\}_{i=1}^d) \propto e^{-d\sum_{i=1}^d V(\lambda_i)} \Pi_{i<j}(\lambda_i - \lambda_j),$$

where $\lambda_1 > \lambda_2 > \cdots > \lambda_d$. By (Chafaï and Lehec, 2020, Theorem 1.1) we obtain the desired result. A brief explanation is that we can rewrite the joint distribution as

$$P_S(\{\lambda_i\}_{i=1}^d) \propto e^{-\left(d\sum_{i=1}^d V(\lambda_i) - \sum_{i<j}\log(\lambda_i - \lambda_j)\right)},$$

where the term inside the bracket is strictly convex in $\mathcal{I}$, so that it satisfies the Bakry-Emery condition (see e.g. (Anderson et al., 2010, Definition 4.4.16)). Such a condition is known to imply strong concentration inequalities, notably log-Sobolev inequalities, but also the Poincaré inequality that we require here.

Tackling the second case in Assumption 2.1 is similar, because the joint eigenvalue distribution can be written as

$$P_S(\{\lambda_i\}_{i=1}^{\lceil \kappa d\rceil}) \propto e^{-\left(d\sum_{i=1}^{\lceil \kappa d\rceil} \widetilde{V}(\lambda_i) - \sum_{i<j}\log(\lambda_i - \lambda_j)\right)},$$

where $\widetilde{V}(s) := V(s) - (1 - \frac{\lceil \kappa d\rceil}{d})\log s$. The term inside the bracket is strictly convex in $\mathcal{B}$ under Assumption 2.1, so we obtain again the Poincaré inequality from (Chafaï and Lehec, 2020, Theorem 1.1).

In the rectangular case, the joint singular value distribution reads Anderson et al. (2010)

$$P_S(\{\sigma_i\}_{i=1}^{\lceil \kappa d\rceil}) \propto e^{-\left(d\sum_{i=1}^{\lceil \kappa d\rceil} \widetilde{V}(\sigma_i^2) - \sum_{i<j}\log(\sigma_i^2 - \sigma_j^2)\right)}, \tag{113}$$

where $\widetilde{V}(s) := V(s) - 2(1 - \frac{\lceil \kappa d\rceil}{d})\log s$. Under Assumption 2.4, the term inside the bracket in eq. (113) is strictly convex in $\{\sigma_i\}$, so again we obtain the desired result. ■

G.4. Central limit theorem for simple statistics

The following lemma clarifies the behavior of some simple statistics of S and Φ under our assumptions.

Lemma 46 (Central Limit Theorem) *We have (recall* $\mathrm{tr}(\cdot) := (1/d)\mathrm{Tr}[\cdot]$*):*

(a) *Under Assumption 2.1,* $\mathbb{E}[\mathrm{tr} S^2] \to \rho$ *as* $d \to \infty$.

(b) *Under Assumptions 2.1 and 2.2,* $\mathrm{Tr}[\Phi S]$ *converges in distribution to* $\mathcal{N}(0, 2\rho)$ *for* $\Phi \sim P_\Phi$ *and* $S \sim P_0$.

(c) *Under Assumption 2.1, for* $S \sim P_0$*:*

$$\mathbb{E}[(\mathrm{Tr}[S^2] - \mathbb{E}[\mathrm{Tr} S^2])^2] = \mathcal{O}_d(1).$$

In particular, $\mathbb{E}[(\mathrm{tr}[S^2] - \rho)^2] = o(1)$.

Proof [Proof of Lemma 46] Point (a) was already proven in Proposition 13. We now show (b). For any bounded Lipschitz ψ and any $M \geq 0$, we have, with $G \sim \mathrm{GOE}(d)$:

$$\begin{aligned} &|\mathbb{E}[\psi(\mathrm{Tr}[\Phi S])] - \mathbb{E}[\psi(\mathrm{Tr}[GS])]| \\ &\leq 2\|\psi\|_\infty \mathbb{P}(\|S\|_{\mathrm{op}} > M) + |\mathbb{E}[\mathbb{1}\{\|S\|_{\mathrm{op}} \leq M\}(\psi(\mathrm{Tr}[\Phi S]) - \psi(\mathrm{Tr}[GS]))]|. \end{aligned}$$

By Proposition 13(ii), $\mathbb{P}(\|S\|_{\mathrm{op}} > M) \to 0$ as $d \to \infty$ for all $M > 0$ large enough. Furthermore:

$$\begin{aligned} |\mathbb{E}[\mathbb{1}\{\|S\|_{\mathrm{op}} \leq M\}(\psi(\mathrm{Tr}[\Phi S]) - \psi(\mathrm{Tr}[GS]))]| &\leq \mathbb{E}[\mathbb{1}\{\|S\|_{\mathrm{op}} \leq M\}|\mathbb{E}_{\Phi,G}(\psi(\mathrm{Tr}[\Phi S]) - \psi(\mathrm{Tr}[GS]))|], \\ &\leq \sup_{\|S\|_{\mathrm{op}} \leq M} |\mathbb{E}_{\Phi,G}(\psi(\mathrm{Tr}[\Phi S]) - \psi(\mathrm{Tr}[GS]))|. \end{aligned}$$

The right-hand side of this last inequality goes to 0 as $d \to \infty$ by Assumption 2.2. In the end:

$$\lim_{d\to\infty} |\mathbb{E}[\psi(\mathrm{Tr}[\Phi S])] - \mathbb{E}[\psi(\mathrm{Tr}[GS])]| = 0.$$

Since $\mathrm{Tr}[GS] \sim \mathcal{N}(0, 2\mathrm{tr}[S^2])$, we get point (b) using (a). Point (c) is a consequence the Poincaré inequality of Lemma 45: using it with $g(\{\lambda_i\}) = \sum_i \lambda_i^2$, we get:

$$\mathrm{Var}[\mathrm{Tr}[S^2]] \leq 4\,\frac{c}{d}\,\mathbb{E}\left[\sum_{i=1}^d \lambda_i^2\right] = \mathcal{O}(1),$$

which ends the proof using point (a). ■

G.5. Properties of Gaussian matrices

The following result is a classical property of Gaussian random variables, here adapted to $\mathrm{GOE}(d)$ matrices.

Lemma 47 (Gaussian Integral by Part) *Let $G \sim GOE(d)$ and $g : \mathbb{R}^{d\times d} \to \mathbb{R}^{d\times d}$ to be a continuous and differentiable function with $g_{ij} = g_{ji}$ and $\frac{\partial g_{ij}}{\partial G_{ij}} = \frac{\partial g_{ij}}{\partial G_{ji}}$ for $1 \le i, j \le d$, then we have*

$$\mathbb{E}_G[\mathrm{Tr}[Gg(G)]] = \frac{2}{d}\mathbb{E}_G\left[\sum_{i,j=1}^{d} \frac{\partial g_{ij}}{\partial G_{ij}}\right].$$

Proof [Proof of Lemma 47] We have

$$\begin{aligned}
\mathbb{E}_G[\mathrm{Tr}[Gg(G)]] &= \sum_{i=1}^{d} \mathbb{E}_G[G_{ii}g_{ii}(G)] + 2\sum_{1\le i<j\le d} \mathbb{E}_G[G_{ij}g_{ij}(G)] \\
&= \frac{2}{d}\sum_{i=1}^{d} \mathbb{E}_G\left[\frac{\partial g_{ii}}{\partial G_{ii}}\right] + \frac{2}{d}\sum_{1\le i<j\le d} \mathbb{E}_G\left[\frac{\partial g_{ij}}{\partial G_{ij}} + \frac{\partial g_{ij}}{\partial G_{ji}}\right] \\
&= \frac{2}{d}\mathbb{E}_G\left[\sum_{i,j=1}^{d} \frac{\partial g_{ij}}{\partial G_{ij}}\right],
\end{aligned}$$

where we use the standard Gaussian integral by part (Stein's lemma) with respect to $\sqrt{\frac{d}{2}}G_{ii}$ and $\sqrt{d}G_{ij}$. ■

Similarly, we can adapt the Gaussian Poincaré inequality to the matrix setting.

Lemma 48 (Gaussian Poincaré Inequality) *Let $Z \in \mathbb{R}^n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$, $Z' \sim GOE(d)$ independent of Z and g to a continuous and differentiable function with $\frac{\partial g}{\partial Z'_{ij}} = \frac{\partial g}{\partial Z'_{ji}}$ for $1 \le i, j \le d$. Then we have*

$$\mathit{Var}(g(Z, Z')) \le \mathbb{E}\left[\sum_{\mu=1}^{n}\left(\frac{\partial g}{\partial Z_\mu}\right)^2 + \frac{2}{d}\sum_{i,j=1}^{d}\left(\frac{\partial g}{\partial Z'_{ij}}\right)^2\right]$$

Proof [Proof of Lemma 48] By the standard Gaussian Poincaré inequality, we have

$$\begin{aligned}
\mathrm{Var}(g(Z, Z')) &\le \mathbb{E}\left[\sum_{\mu=1}^{n}\left(\frac{\partial g}{\partial Z_\mu}\right)^2 + \frac{1}{d}\sum_{1\le i<j\le d}\left(2\frac{\partial g}{\partial Z'_{ij}}\right)^2 + \frac{2}{d}\sum_{i=1}^{d}\left(\frac{\partial g}{\partial Z'_{ii}}\right)^2\right] \\
&= \mathbb{E}\left[\sum_{\mu=1}^{n}\left(\frac{\partial g}{\partial Z_\mu}\right)^2 + \frac{2}{d}\sum_{i,j=1}^{d}\left(\frac{\partial g}{\partial Z'_{ij}}\right)^2\right],
\end{aligned}$$

which finishes the proof. ■

G.6. Properties of two auxiliary channels

We present here the properties of two auxiliary channels. The first one is the matrix denoising problem:

$$Y' = \sqrt{r}S + Z,$$

with $Z \sim \mathrm{GOE}(d)$. Its free entropy is given by

$$\tilde{f}_d(r) = \frac{1}{d^2}\mathbb{E}_{Y'} \log \int P_0(\mathrm{d}S) e^{-\frac{d}{4}\mathrm{Tr}[(Y'-\sqrt{r}S)^2]},$$

and has the following property.

Lemma 49 *Under Assumption 2.1,*

$$\lim_{d\to\infty} \tilde{f}_d(r) = \psi_{P_0}(r).$$

Moreover, ψ_{P_0} is a non-increasing, $\frac{\rho}{4}$-Lipschitz and convex function.

Proof [Proof of Lemma 49] By the I-MMSE theorem Guo et al. (2005), we have

$$\tilde{f}_d'(r) = \frac{d}{4}\mathrm{MMSE}(S^*|Y) - \frac{d}{4}\mathbb{E}[\mathrm{Tr}(S^*)^2],$$

which gives $\tilde{f}_d'(r) \leq 0$, $\liminf_{d\to\infty} \tilde{f}_d'(r) \geq -\frac{\rho}{4}$ and that $\tilde{f}_d(r)$ is convex. Therefore, $\{\tilde{f}_d(r)\}_{d\geq 1}$ are non-increasing, uniformly Lipschitz and strictly convex. According to (Maillard et al., 2024, Theorem 4.3), $\lim_{d\to\infty} \tilde{f}_d(r) = \psi_{P_0}(r)$, and thus $\psi_{P_0}(r)$ is non-increasing, $\frac{\rho}{4}$-Lipschitz, and convex. ■

In the rectangular model, for $S \in \mathbb{R}^{d\times L}$, and $\{Z'_{ij}\}_{i,j=1}^{d,L} \overset{iid}{\sim} \mathcal{N}(0, \frac{1}{\sqrt{dL}})$, the free entropy of the denoising problem is given by

$$\tilde{f}_d^{\mathrm{rec}}(r) := \frac{1}{dL}\mathbb{E}_{Y'} \log \int P_0(\mathrm{d}S) e^{-\frac{1}{2}\sqrt{dL}r\mathrm{Tr}[S^TS] + \sqrt{dLr}\mathrm{Tr}[(Y')^TS]},$$

with $Y' = \sqrt{r}S^\star + Z$, and satisfies the following property. Recall the definition of ρ in Proposition 16.

Lemma 50 *Under Assumption 2.4, we have*

$$\lim_{d\to\infty} \tilde{f}_d^{rec}(r) = \psi_{P_0}^{rec}(r).$$

Moreover $\psi_{P_0}^{rec}$ is a non-increasing, $\frac{\rho}{2}$-Lipschitz, and convex function.

Proof [Proof of Lemma 50] By (Pourkamali and Macris, 2024, Theorem 4), we have

$$\lim_{d\to\infty} \tilde{f}_d^{\mathrm{rec}}(r) = -\frac{\rho r}{2} + \beta J[\hat{\mu}, \hat{\nu}],$$

where $\hat{\mu}, \hat{\nu}$ refer to the limiting symmetrized singular value distributions of $A := \sqrt{r/\sqrt{\beta}}S$, $B := \sqrt{r/\sqrt{\beta}}S + Z'/\sqrt[4]{\beta}$, and

$$J[\hat{\mu}, \hat{\nu}] := \lim_{d\to\infty} \frac{1}{d^2} \log \int\int \mathcal{D}U\mathcal{D}V\, e^{d\mathrm{Tr}[A^\top UBV^\top]}$$

is a so-called rectangular spherical integral Guionnet and Huang (2021) (here $\mathcal{D}$ is the Haar measure over the orthogonal group). Notice that $Z'' := Z'/\sqrt[4]{\beta}$ has i.i.d. elements with variance $1/d$. By Guionnet and Huang (2021), we have

$$J[\mu,\nu] = -(\beta^{-1}-1)\int \log|x|\hat{\nu}(x)dx - \Sigma[\hat{\nu}] + \frac{1}{2}\left[\int x^2\hat{\mu}(x)dx + \int x^2\hat{\nu}(x)dx\right] - I_{\hat{\mu}}(\hat{\nu}) + const,$$

where $I_{\hat{\mu}}(\hat{\nu})$ is the large deviations rate function of the symmetrized empirical singular value distribution of the matrix $A+Z''$, over the law of the matrix Z'' which has i.i.d. $\mathcal{N}(0,1/d)$ elements, and in the scale d^2. Recall as well that $\Sigma(\mu) := \int \mu(\mathrm{d}x)\mu(\mathrm{d}y)\log|x-y|$. As $B = A+Z''$ has limiting symmetrized singular value distribution given by $\hat{\nu}$, by definition (with dist a distance metrizing the weak convergence):

$$\begin{aligned} I_{\hat{\mu}}(\hat{\nu}) &:= -\lim_{\delta\to 0}\lim_{d\to\infty}\frac{1}{d^2}\log\mathbb{P}(\mathrm{dist}(\hat{\mu}_B,\hat{\nu})\le\delta), \\ &= 0. \end{aligned}$$

Moreover, by definition, we have $\int x^2\hat{\mu}(x)dx + \int x^2\hat{\nu}(x)dx = 2r\rho/\beta + const$. Notably, this establishes eq. (14). Repeating the arguments of the proof of Lemma 49, we obtain similarly that $\psi^{\mathrm{rec}}_{P_0}(r)$ is non-increasing, $\frac{\rho}{2}$-Lipschitz, and convex. ∎

The second model we consider is a scalar channel involving P_{out}. We obtain the following property from Barbier and Macris (2019).

Lemma 51 ((Barbier et al., 2019, Proposition 18)) *Suppose that $P_{out}(\cdot|x)$ is the law of $\varphi(x,a)+\sqrt{\Delta}Z$, where φ is a bounded function with bounded first and second derivatives, then Ψ_{out} is convex, C^2 and non-decreasing on $[0,\rho]$, and thus it is also Lipschitz on $[0,\rho]$.*

Appendix H. Sketch of the Generalization to the Rectangular Setting

We very briefly discuss how our main results transpose straightforwardly to the rectangular setting.

- The properties of the spiked model (Appendix F), as well as the universality results (Appendix B), do not rely on whether the matrix is symmetric or rectangular, and thus can be immediately stated for this case.
- The interpolation studied in Appendix C reads in the rectangular setting:

$$\begin{cases} Y_t \sim P_{\mathrm{out}}(\cdot|J_t) \\ Y'_t = \sqrt{R_1(t)}S^* + Z' \end{cases} \tag{114}$$

  with
$$J_{t,\mu} := \sqrt{1-t}\mathrm{Tr}[G_\mu^\top S^*] + \sqrt{2R_2(t)}V_\mu + \sqrt{2\rho t - 2R_2(t) + 2s_d}W^*_\mu,$$

  $Z'_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\frac{1}{\sqrt{dL}})$ for $i\in[d], j\in[L]$ and $\{V_\mu, W^*_\mu\}_{\mu=1}^n \overset{iid}{\sim} \mathcal{N}(0,1)$. The free entropy of the first channel (the first equation of (114)) remains the same as in the symmetric setting, and the free entropy of the second channel (the second equation of (114)) is given by Lemma 50. The adaptive interpolation argument, the concentration inequalities, and the relaxation of the assumptions exactly follow Appendix C, with all the key components: potential truncation (Lemma 44), central limit theorem (Lemma 46), the Poincaré inequality (Lemma 45), the Gaussian integral by part (Lemma 47) and the Gaussian Poincaré Inequality (Lemma 48), all extending straightforwardly to the rectangular case. More generally, all the properties of symmetric matrices that we use in this paper also extend to the rectangular case when considering singular values (Proposition 16).