Phase transitions in spiked matrix estimation: information-theoretic analysis

Léo Miolane

January 26, 2019
Introduction

Estimating a low-rank object (matrix or tensor) from a noisy observation is a fundamental problem in statistical inference with applications in machine learning, signal processing or information theory. These notes mainly focus on so-called “spiked” models where we observe a signal spike perturbed with some additive noise. We should consider here two popular models.

The first one is often denoted as the spiked Wigner model. One observes

\[ Y = \sqrt{\frac{\lambda}{n}} XX^\top + Z \] \hspace{1cm} (1)

where \( X = (X_1, \ldots, X_n) \sim P_0 \) is the signal vector and \( Z \) is symmetric matrix that account for noise with standard Gaussian entries: \((Z_{i,j})_{i \leq j} \sim \mathcal{N}(0, 1)\). \( \lambda \geq 0 \) is a signal-to-noise ratio.

The second model we consider is the non-symmetric version of (1), sometimes called spiked Wishart\(^1\) or spiked covariance model:

\[ Y = \sqrt{\frac{\lambda}{n}} UV^\top + Z \] \hspace{1cm} (2)

where \( U = (U_1, \ldots, U_n) \sim P_U, V = (V_1, \ldots, V_m) \sim P_V \) are independent. \( Z \) is a noise matrix with standard normal entries: \( Z_{i,j} \sim \mathcal{N}(0, 1) \). \( \lambda > 0 \) captures again the strength of the signal. We are here interested in the regime where \( n, m \to +\infty \), while \( m/n \to \alpha > 0 \). In both models (1-2) the goal of the statistician is to estimate the low-rank signals \((XX^\top \) or \( UV^\top)\) from the observation of \( Y \). This task is often called Principal Component Analysis (PCA) in the literature.

These spiked models have received a lot of attention since their introduction by [37]. From a statistical point of view, there are two main problems linked to the spiked models (1-2).

- The recovery problem: how can we recover the planted signal \( X / U, V \)? Is it possible? Can we do it efficiently?
- The detection problem: is it possible to distinguish between the pure noise case \((\lambda = 0)\) and the case where a spike is present \((\lambda > 0)\)? Is there any efficient test to do this?

We will focus here on the recovery problem. We let the reader refer to [16, 55, 27, 9, 60, 4, 2] and the references therein for a detailed analysis of the detection problem.

The spiked models (1-2) has been extensively studied in random matrix theory. The seminal work of [7] (for the complex spiked Wishart model, and [8] for the real spiked Wishart) established the existence of a phase transition: there exists a critical value of the signal-to-noise ratio \( \lambda \) above which the largest singular value of \( Y/\sqrt{n} \) escapes from the Marchenko-Pastur bulk. The same phenomenon holds for the spiked Wigner model, see [58, 30, 19]. It turns out that for both models the eigenvector (respectively singular vector) corresponding to the largest eigenvalue (respectively singular value) also undergo a phase transition at the same threshold, see [35, 57, 54, 14, 15].

For the spiked Wigner model (1), the main result of interest to us is the following (from [14]). For any probability distribution \( P_0 \) such that \( \mathbb{E}_{P_0}[|X|^2] = 1 \), we have

- if \( \lambda \leq 1 \), the top eigenvalue of \( Y/\sqrt{n} \) converges a.s. to 2 as \( n \to \infty \), and the top eigenvector \( \hat{x} \) (with norm \( ||\hat{x}||^2 = n \)) has asymptotically trivial correlation with \( X: \frac{1}{n}\hat{x}^\top X \to 0 \) a.s.
- if \( \lambda > 1 \), the top eigenvalue of \( Y/\sqrt{n} \) converges a.s. to \( \sqrt{\lambda + 1/\sqrt{\lambda}} > 2 \) and the top eigenvector \( \hat{x} \) (with norm \( ||\hat{x}||^2 = n \)) has asymptotically nontrivial correlation with \( X: \left(\frac{1}{\sqrt{n}}\hat{x}^\top X\right)^2 \to 1 - 1/\lambda \) a.s.

\(^1\)This terminology usually refers to the case where \( V \) is a standard Gaussian vector. We consider here a slightly more general case by allowing the entries of \( V \) to be taken i.i.d. from any probability distribution.
An analog statement for the spiked Wishart model is proved in [15]. These results give us a precise understanding of the performance of the top eigenvectors (or top singular vectors) for recovering the low-rank signals.

However, these naive spectral estimators do not take into account any prior information on the signal. Thus many algorithms have been proposed to exploit additional properties of the signal, such as sparsity [38, 21, 68, 5, 26] or positivity [52].

Another line of works study Approximate Message Passing (AMP) algorithms for the spiked models above, see [61, 25, 44, 53]. Motivated by deep insights from statistical physics, these algorithms are believed (for the models (1-2), when $\lambda$ and the priors $P_0, P_U, P_V$ are known by the statistician) to be optimal among all polynomial-time algorithms. A great property of these algorithms is that their performance can be precisely tracked in the high-dimensional limit by a simple recursion called “state evolution”, see [13, 36]. For a detailed analysis of message-passing algorithms for the models (1-2), see [45].

In the following we will not consider any particular estimator but rather try to compute the best performance achievable by any estimator. We will suppose to be in the so-called “Bayes-optimal” setting, where the statistician knows the prior $P_0$ (or $P_U, P_V$) and the signal-to-noise ratio $\lambda$. In that situation, we will study the posterior distribution of the signal given the observations. As we should see in the sequel, both estimation problems (1-2) can be seen as mean-field spin glass models similar to the Sherrington-Kirkpatrick model, studied in the ground-breaking book of Mézard, Parisi and Virasoro [48]. Therefore, the methods that we will use here come from the mathematical study of spin glasses, namely from the works of Talagrand [63, 64], Guerra [32] and Panchenko [56].

In order to further motivate the study of the models (1-2) let us mention some interesting special cases, depending on the choice of the priors $P_0 / P_U, P_V$.

- **Sparse PCA.** Consider the spiked Wishart model with $P_U = \text{Ber}(\epsilon)$ and $P_V = \mathcal{N}(0, 1)$. In that case, one see that conditionally on $U$ the columns of $Y$ are i.i.d. sampled from $\mathcal{N}(0, I_n + \lambda/n UU^\top)$, which is a sparse spiked covariance model. The spiked Wigner model with $P_0 = \text{Ber}(\epsilon)$ has also been used to study sparse PCA.

- **Submatrix localization.** Take $P_0 = \text{Ber}(p)$ in the spiked Wigner model. The goal of submatrix localization is now to extract a submatrix of $Y$ of size $pn \times pn$ with larger mean.

- **Community Detection in the Stochastic Block Model (SBM).** As shown in [24, 42] recovering two communities of size $pn$ and $(1 - p)n$ in a dense SBM of $n$ vertices is (in some sense) “equivalent” to the spiked Wigner model with prior

$$P_0 = p \delta_{-\frac{1}{\sqrt{p^{-1}}}} + (1 - p) \delta_{-\sqrt{\frac{1}{1-p}}}.$$

- **Z/2 synchronization.** This corresponds to the spiked Wigner model with Rademacher prior $P_0 = \frac{1}{2} \delta_{-1} + \frac{1}{2} \delta_{+1}$.

- **High-dimensional Gaussian mixture clustering.** Consider the multidimensional version of the spiked Wishart model where $U \in \mathbb{R}^{n \times k}$ and $V \in \mathbb{R}^{m \times k}$. If one take $P_V$ (the distribution of the rows of $V$) to be supported by the canonical basis of $\mathbb{R}^k$, the model is equivalent to the clustering of $m$ points (the columns of $Y$) in $n$ dimensions from a Gaussian mixture model. The centers of the clusters are here the columns of $U$. 


## Contents

**Introduction** 1

1 **Bayesian inference in Gaussian noise** 4  
1.1 Definitions and problem setting 4  
1.2 The Nishimori identity 6  
1.3 The I-MMSE relation 6  
1.4 A warm-up: “needle in a haystack” problem 8

2 **A decoupling principle** 10  
2.1 Revealing a small fraction of the planted solution 11  
2.2 Noisy side Gaussian channel 12

3 **Low-rank symmetric matrix estimation** 16  
3.1 Information-theoretic limits 16  
3.2 Information-theoretic and algorithmic phase transitions 18  
3.3 Low-rank symmetric tensor estimation 21  
3.4 Proof of the Replica-Symmetric formula (Theorem 2) 26

4 **Non-symmetric low-rank matrix estimation** 35  
4.1 Information-theoretic limits 35  
4.2 Application to the spiked covariance model 37  
4.3 Proof of the Replica-Symmetric formula (Theorem 4) 39  
4.4 Proof of Theorem 5 46

**Appendix** 51  
A Proofs of some basic properties of the MMSE and the free energy 51  
B Differentiation of a supremum of functions 54

**Bibliography** 55
Chapter 1

Bayesian inference in Gaussian noise

We introduce in this section some general properties of Bayesian inference in presence of additive Gaussian noise, that will be used repeatedly in the sequel.

1.1 Definitions and problem setting

As explained in the introduction, we will be interested in inference problems of the form:

\[ Y = \sqrt{\lambda} X + Z, \]  

where the signal \( X \) is sampled according some probability distribution \( P_X \) over \( \mathbb{R}^n \), and where the noise \( Z = (Z_1, \ldots, Z_n) \overset{i.i.d.}{\sim} \mathcal{N}(0, 1) \) is independent from \( X \). In Sections 3 and 4, \( X \) will typically be a low-rank matrix. The parameter \( \lambda \geq 0 \) plays the role of a signal-to-noise ratio. We assume that \( P_X \) admits a finite second moment: \( \mathbb{E} \| X \|_2^2 < \infty \).

Given the observation channel (1.1), the goal of the statistician is to estimate \( X \) given the observations \( Y \). We will assume to be in the “Bayes-optimal” setting, where the statistician knows all the parameters of the inference model, that is the prior distribution \( P_X \) and the signal-to-noise ratio \( \lambda \). We measure the performance of an estimator \( \hat{\theta} \) (i.e. a measurable function of the observations \( Y \)) by its Mean Squared Error:

\[
\text{MSE}(\hat{\theta}) = \mathbb{E} \left[ \| X - \hat{\theta}(Y) \|_2^2 \right].
\]

One of our main quantity of interest will be the Minimum Mean Squared Error

\[
\text{MMSE}(\lambda) = \min_{\hat{\theta}} \text{MSE}(\hat{\theta}) = \mathbb{E} \left[ \| X - \mathbb{E}[X|Y] \|_2^2 \right],
\]

where the minimum is taken over all measurable function \( \hat{\theta} \) of the observations \( Y \). Since the optimal estimator (in term of Mean Squared Error) is the posterior mean of \( X \) given \( Y \), a natural object to study is the posterior distribution of \( X \).
Definition 1

The posterior distribution of \( X \) given \( Y \) is

\[
dP(x \mid Y) = \frac{1}{Z(\lambda, Y)} e^{H_\lambda,Y(x)} dP_X(x),
\]

where

\[
H_\lambda,Y(x) = \sqrt{\lambda} x^\top Y - \frac{\lambda}{2} \| x \|^2 = \sqrt{\lambda} x^\top Z + \lambda x^\top X - \frac{\lambda}{2} \| x \|^2
\]

is called the Hamiltonian\(^1\) and the normalizing constant

\[
Z(\lambda, Y) = \int dP_X(x) e^{H_\lambda,Y(x)}
\]

is called the partition function.

Expectations with respect the posterior distribution (1.2) will be denoted by the Gibbs brackets \( \langle \cdot \rangle_\lambda \):

\[
\langle f(x) \rangle_\lambda = \mathbb{E}[f(X) \mid Y] = \frac{1}{Z(\lambda, Y)} \int dP_X(x) f(x) e^{H_\lambda,Y(x)},
\]

for any measurable function \( f \) such that \( f(X) \) is integrable.

Definition 2

\( F_\lambda = \mathbb{E} \log Z(\lambda, Y) \) is called the free energy\(^2\). It is related to the mutual information between \( X \) and \( Y \) by

\[
F(\lambda) = \frac{\lambda}{2} \mathbb{E} \| X \|^2 - I(X; Y).
\]

Proof. The mutual information \( I(X; Y) \) is defined as the Kullback-Leibler divergence between \( P_{(X,Y)} \), the joint distribution of \( (X, Y) \) and \( P_X \otimes P_Y \) the product of the marginal distributions of \( X \) and \( Y \). \( P_{(X,Y)} \) is absolutely continuous with respect to \( P_X \otimes P_Y \) with Radon-Nikodym derivative:

\[
\frac{dP_{(X,Y)}}{dP_X \otimes P_Y}(X, Y) = \frac{\exp \left( -\frac{1}{2} \| Y - \sqrt{\lambda} X \|^2 \right)}{\int \exp \left( -\frac{1}{2} \| Y - \sqrt{\lambda} x \|^2 \right) dP_X(x)}.
\]

Therefore

\[
I(X; Y) = \mathbb{E} \log \left( \frac{dP_{(X,Y)}}{dP_X \otimes P_Y}(X, Y) \right) = -\mathbb{E} \log \int dP_X(x) \exp \left( \sqrt{\lambda} x^\top Y - \sqrt{\lambda} x^\top X - \frac{\lambda}{2} \| x \|^2 + \frac{\lambda}{2} \| X \|^2 \right)
\]

\[
= -F(\lambda) + \frac{\lambda}{2} \mathbb{E} \| X \|^2.
\]

We state now two basic properties of the MMSE. A more detailed analysis can be found in \([34,66]\).

Proposition 1

\( \lambda \mapsto \text{MMSE}(\lambda) \) is non-increasing over \( \mathbb{R}_+ \). Moreover

- \( \text{MMSE}(0) = \mathbb{E} \| X - \mathbb{E}[X] \|^2 \),
- \( \text{MMSE}(\lambda) \xrightarrow[\lambda \to +\infty]{} 0 \).

\(^1\)According to the physics convention, this should be minus the Hamiltonian, since a physical system tries to minimize its energy. However, we chose here to remove it for simplicity.

\(^2\)This is in fact minus the free energy, but we chose to remove the minus sign for simplicity.
Proposition 2

\[ \lambda \mapsto \text{MMSE}(\lambda) \text{ is continuous over } \mathbb{R}_+. \]

The proofs of Proposition 1 and 2 can respectively be found in Appendix A.1 and A.2.

1.2 The Nishimori identity

We will often consider i.i.d. samples \( x^{(1)}, \ldots, x^{(k)} \) from the posterior distribution \( P(\cdot | Y) \), independently of everything else. Such samples are called replicas. The (obvious) identity below (which is simply Bayes rule) will be used repeatedly. It states that the planted solution \( X \) behaves like a replica.

Proposition 3 (Nishimori identity)

Let \((X, Y)\) be a couple of random variables on a polish space. Let \( k \geq 1 \) and let \( x^{(1)}, \ldots, x^{(k)} \) be \( k \) i.i.d. samples (given \( Y \)) from the distribution \( P(X = \cdot | Y) \), independently of every other random variables. Let us denote \( \langle \cdot \rangle \) the expectation with respect to \( P(X = \cdot | Y) \) and \( E \) the expectation with respect to \((X, Y)\). Then, for all continuous bounded function \( f \)

\[ E\langle f(Y, x^{(1)}, \ldots, x^{(k)}) \rangle = E\langle f(Y, x^{(1)}, \ldots, x^{(k-1)}, X) \rangle. \]

Proof. It is equivalent to sample the couple \((X, Y)\) according to its joint distribution or to sample first \( Y \) according to its marginal distribution and then to sample \( X \) conditionally to \( Y \) from its conditional distribution \( P(X = \cdot | Y) \). Thus the \((k+1)\)-tuple \((Y, x^{(1)}, \ldots, x^{(k)})\) is equal in law to \((Y, x^{(1)}, \ldots, x^{(k-1)}, X)\). \( \square \)

1.3 The I-MMSE relation

We state now the very useful I-MMSE relation from [33].

Proposition 4

For all \( \lambda \geq 0 \),

\[ \frac{\partial}{\partial \lambda} I(X; Y) = \frac{1}{2} \text{MMSE}(\lambda) \quad \text{and} \quad F'(\lambda) = \frac{1}{2} E\langle x^\top X \rangle_{\lambda} = \frac{1}{2} (E\|X\|^2 - \text{MMSE}(\lambda)). \] (1.4)

\( F \) thus is a convex, differentiable, non-decreasing, and \( \frac{1}{2}E\|X\|^2 \)-Lipschitz function over \( \mathbb{R}_+ \).

We study in this section the simplest model of the form (1.1), namely the additive Gaussian scalar channel:

\[ Y_0 = \sqrt{\lambda}X_0 + Z_0, \] (1.5)

where \( Z_0 \sim \mathcal{N}(0, 1) \) and \( X_0 \) is sampled from a distribution \( P_0 \) over \( \mathbb{R} \). Again, the goal is to recover \( X_0 \) from the observation of \( Y_0 \). The MMSE for this task is

\[ \text{MMSE}_{P_0}(\lambda) = E[(X_0 - E[X_0|Y])^2]. \]

The study of this simple inference channel will be very useful in the following, because we will see that our matrix estimation problems enjoy asymptotically a “decoupling principle” that reduces them to scalar channels like (1.5). The next proposition summarizes the main properties of the free energy \( \psi_{P_0} \) of the channel (1.5).
Proposition 5

Let \( X_0 \sim P_0 \) be a real random variable with finite second moment. For \( \lambda \geq 0 \), let \( Y_0 = \sqrt{\lambda} X_0 + Z_0 \), where \( Z_0 \sim \mathcal{N}(0, 1) \) is independent from \( X_0 \). Then the function

\[
\psi_{P_0} : \lambda \mapsto \mathbb{E} \log \int dP_0(x)e^{\sqrt{\lambda} Y_0 x - \lambda x^2 / 2}
\]  

(1.6)
is convex, continuously differentiable, non-decreasing and \( \frac{1}{2} \mathbb{E}[X_0^2] \)-Lipschitz on \( \mathbb{R}_+ \). Moreover, \( \psi_{P_0} \) is strictly convex, if \( P_0 \) is not a Dirac measure.

Proof. By Proposition 4 it remains only to show that \( \psi_{P_0} \) is strictly convex when \( P_0 \) differs from a Dirac mass. We proceed by truncation and consider the distribution \( P_0^{(N)} \) of \( X_0^{(N)} = X_0 1(-N \leq X_0 \leq N) \). \( \langle \cdot \rangle_{\lambda,N} \) will denote the corresponding posterior distribution. One can compute the second derivative and again, using Gaussian integration by parts and the Nishimori identity one obtain:

\[
\psi''_{P_0}(\lambda) = \frac{1}{2} \mathbb{E} \left[ (\langle x^2 \rangle_{\lambda,N} - \langle x \rangle_{\lambda,N}^2) \right] \geq \frac{1}{2} \text{MMSE}_{P_0^{(N)}}(\lambda)^2,
\]

(1.7)

by Jensen’s inequality. Let now \( 0 < s < t \). We can then find \( 0 < s < t \) such that \( \psi'_{P_0}(s) = \psi'_{P_0}(t) \). By integrating (1.7) we get

\[
\psi'_{P_0^{(N)}}(t) - \psi'_{P_0^{(N)}}(s) \geq \frac{1}{2} \int_s^t \text{MMSE}_{P_0^{(N)}}(\lambda)^2 d\lambda.
\]

(1.8)

The sequence of convex functions \( \langle \psi'_{P_0^{(N)}} \rangle_N \) converges to \( \psi_{P_0} \) which is differentiable. A standard analysis lemma gives then that the derivatives \( \langle \psi'_{P_0^{(N)}} \rangle_N \) converges to \( \psi'_{P_0} \) and \( \text{MMSE}_{P_0^{(N)}} \) converges to \( \text{MMSE}_{P_0} \). Therefore, equation (1.8) gives

\[
\psi'_{P_0}(t) - \psi'_{P_0}(s) \geq \frac{1}{2} \int_s^t \text{MMSE}_{P_0}(\lambda)^2 d\lambda \geq \frac{1}{2} (t-s) \text{MMSE}_{P_0}(t)^2.
\]

If \( P_0 \) is not a Dirac measure, then the last term is strictly positive: this concludes the proof.

Example 1. Let us compute the mutual information and the MMSE for particular choices of prior distributions:

- In the case of Gaussian prior \( P_0 = \mathcal{N}(0, 1) \), one can compute \( \psi_{P_0} \) explicitly: \( \psi_{P_0}(\lambda) = \frac{1}{2} (\lambda - \log(1 + \lambda)) \). We deduce by (1.3) that in that case \( I(X_0; Y_0) = \frac{1}{2} \log(1 + \lambda) \). One deduces by the I-MMSE relation (1.4): \( \text{MMSE}(\lambda) = \frac{1}{1+\lambda} \).

- In the case of Rademacher prior \( P_0 = \frac{1}{2} \delta_{+1} + \frac{1}{2} \delta_{-1} \) we compute \( \psi_{P_0}(\lambda) = \mathbb{E} \log \cosh(\sqrt{\lambda} Z_0 + \lambda) - \frac{\lambda}{2} \) and \( I(X_0; Y_0) = \lambda - \mathbb{E} \log \cosh(\sqrt{\lambda} Z_0 + \lambda) \). The I-MMSE relation gives

\[
\frac{1}{2} \text{MMSE}(\lambda) = \frac{\partial}{\partial \lambda} I(X_0; Y_0) = 1 - \mathbb{E} \left[ \left( \frac{1}{2\sqrt{\lambda}} Z_0 + 1 \right) \tanh(\sqrt{\lambda} Z_0 + \lambda) \right]
\]

\[
= 1 - \mathbb{E} \tanh(\sqrt{\lambda} Z_0 + \lambda) - \frac{1}{2} \mathbb{E} \tanh'(\sqrt{\lambda} Z_0 + \lambda)
\]

\[
= \frac{1}{2} - \mathbb{E} \tanh(\sqrt{\lambda} Z_0 + \lambda) + \frac{1}{2} \mathbb{E} \tanh^2(\sqrt{\lambda} Z_0 + \lambda)
\]

where we used Gaussian integration by parts. Since by the Nishimori property \( \mathbb{E}[x X_0] = \mathbb{E}[x]^2 \), one has \( \mathbb{E} \tanh(\sqrt{\lambda} Z_0 + \lambda) = \mathbb{E} \tanh^2(\sqrt{\lambda} Z_0 + \lambda) \) and therefore \( \text{MMSE}(\lambda) = 1 - \mathbb{E} \tanh(\sqrt{\lambda} Z_0 + \lambda) \).
1.4 A warm-up: “needle in a haystack” problem

In order to illustrate the results seen in the previous sections, we study now a very simple inference model. Let \( (e_1, \ldots, e_{2^n}) \) be the canonical basis of \( \mathbb{R}^{2^n} \). Let \( \sigma_0 \sim \mathcal{U}(\{1, \ldots, 2^n\}) \) and define \( X = e_{\sigma_0} \) (i.e. \( X \) is chosen uniformly over the canonical basis of \( \mathbb{R}^{2^n} \)). Suppose here that we observe:

\[
Y = \sqrt{\lambda n} X + Z,
\]

where \( Z = (Z_1, \ldots, Z_{2^n}) \sim \mathcal{N}(0, 1) \), independently from \( \sigma_0 \). The goal here is to estimate \( X \) or equivalently to find \( \sigma_0 \). The posterior distribution reads:

\[
P(\sigma_0 = \sigma | Y) = \frac{1}{Z_n(\lambda)} 2^{-n} \exp\left( \sqrt{\lambda n} (\sigma \cdot Y) - \frac{\lambda n}{2} ||e_\sigma||^2 \right)
\]

where \( Z_n(\lambda) \) is the partition function

\[
Z_n(\lambda) = \frac{1}{2^n} \sum_{\sigma=1}^{2^n} \exp\left( \sqrt{\lambda n} (\sigma \cdot Y) - \frac{\lambda n}{2} \right).
\]

We will be interested in computing the free energy \( F_n(\lambda) = \frac{1}{n} \mathbb{E} \log Z_n(\lambda) \) in order to deduce then the minimal mean squared error using the I-MMSE relation (1.4) presented in the previous section.

Although its simplicity, this model is interesting for many reasons. First, it is one of the simplest statistical model for which one observes a phase transition. Second it is the “planted” analog of the random energy model (REM) introduced in statistical physics by Derrida \([22, 23]\), for which the free energy reads \( \frac{1}{n} \mathbb{E} \log \sum_{\sigma} \frac{1}{Z_n(\lambda)} \exp\left( \sqrt{\lambda n} \sigma \right) \). Third, as we will see in Section 3.3.1, this model correspond to the “large order limit” of a rank-one tensor estimation model.

We start by computing the limiting free energy:

**Theorem 1**

\[
\lim_{n \to \infty} F_n(\lambda) = \begin{cases} 
0 & \text{if } \lambda \leq 2 \log 2, \\
\frac{\lambda}{2} - \log(2) & \text{if } \lambda \geq 2 \log 2.
\end{cases}
\]

**Proof.** Using Jensen’s inequality

\[
F_n(\lambda) \leq \frac{1}{n} \mathbb{E} \log \mathbb{E} [Z_n(\lambda)|\sigma_0, Z_{\sigma_0}] = \frac{1}{n} \mathbb{E} \log \left( 1 - \frac{1}{2^n} + e^{\sqrt{\lambda n} \sigma_0 + \frac{\lambda n}{2} - \log(2)n} \right)
\]

\[
\leq \frac{1}{n} \mathbb{E} \log \left( 1 + e^{\frac{\lambda n}{2} - \log(2)n} \right) + \frac{\lambda}{n} \xrightarrow{n \to \infty} \begin{cases} 
0 & \text{if } \lambda \leq 2 \log(2), \\
\frac{\lambda}{2} - \log(2) & \text{if } \lambda \geq 2 \log(2).
\end{cases}
\]

\( F_n \) is non-negative since \( F_n(0) = 0 \) and \( F_n \) is non-decreasing. We have therefore \( F_n(\lambda) \xrightarrow{n \to \infty} 0 \) for all \( \lambda \in [0, 2 \log(2)] \). We have also, by only considering the term \( \sigma = \sigma_0 \):

\[
F_n(\lambda) \geq \frac{1}{n} \mathbb{E} \log \left( e^{\sqrt{\lambda n} \sigma_0 + \frac{\lambda n}{2}} \right) = \frac{\lambda}{2} - \log(2).
\]

We obtain therefore that \( F_n(\lambda) \xrightarrow{n \to \infty} \frac{\lambda}{2} - \log(2) \) for \( \lambda \geq 2 \log(2) \). \( \square \)

Using the I-MMSE relation (1.4), we deduce the limit of the minimum mean Squared Error \( \text{MMSE}_n(\lambda) = \min_{\hat{\theta}} \mathbb{E} ||X - \hat{\theta}(Y)||^2 \):

\[
\frac{1}{2} \text{MMSE}_n(\lambda) = \mathbb{E} ||X||^2 - F_n(\lambda) = 1 - F_n(\lambda).
\]
$F_n$ is a convex function of $\lambda$, so its derivative converges to the derivative of its limit at each $\lambda$ at which the limit is differentiable, i.e. for all $\lambda \in (0, +\infty) \setminus \{2 \log(2)\}$. We obtain therefore that for all $\lambda > 0$,

- if $\lambda < 2 \log(2)$, then $\text{MMSE}_n(\lambda) \xrightarrow[n \to \infty]{} 1$: one can not recover $X$ better than a random guess.
- if $\lambda > 2 \log(2)$, then $\text{MMSE}_n(\lambda) \xrightarrow[n \to \infty]{} 0$: one can recover $X$ perfectly.

Of course, the result we obtain here is (almost) trivial since the maximum likelihood estimator

$$\hat{\sigma}(Y) = \arg \max_{1 \leq \sigma \leq 2^n} Y_{\sigma}$$

of $\sigma_0$ is easy to analyze. Indeed, $\max_{\sigma} Z_{\sigma} \simeq \sqrt{2 \log(2)n}$ with high probability so that the maximum likelihood estimator recovers perfectly the signal for $\lambda > 2 \log(2)$ with high probability.
Chapter 2

A decoupling principle

We present in this section a general “decoupling principle” that will be particularly useful in the study of the spiked models. We consider here the setting where $X = (X_1, \ldots, X_n) \overset{i.i.d.}{\sim} P_0$ for some probability distribution $P_0$ over $\mathbb{R}$ with support $S$. Let $Y \in \mathbb{R}^m$ be another random variable that accounts for noisy observation of $X$. The goal is again to recover the planted solution $X$ from the observations $Y$. We suppose that the distribution of $X$ given $Y$ takes the following form

$$P(X \in A \mid Y) = \frac{1}{Z_n(Y)} \int_{x \in A} dP_0^{\otimes n}(x)e^{H_n(x,Y)}, \quad \text{for all Borel set } A \subset \mathbb{R}^n,$$

where $H_n$ is a measurable function on $\mathbb{R}^n \times \mathbb{R}^m$ that can be equal to $-\infty$ (in which case, we use the convention $\exp(-\infty) = 0$) and $Z_n(Y) = \int_{x \in S^n} dP_0^{\otimes n}(x)e^{H_n(x,Y)}$ is the appropriate normalization.

In the following, we are going to drop the dependency in $Y$ of $H_n(x,Y)$ and simply write $H_n(x)$.

We introduce now a very important notation: the overlap between vectors $u, v \in \mathbb{R}^n$. This is simply the normalized scalar product:

$$u \cdot v = \frac{1}{n} \sum_{i=1}^n u_i v_i.$$

One should really see $x$ as a system of $n$ spins $(x_1, \ldots, x_n)$ interacting through the (random) Hamiltonian $H_n$. Our inference problem should be understood as the study of this spin glass model. A central quantity of interest in spin glass theory is the overlaps $x^{(1)} \cdot x^{(2)}$ between two replicas, i.e. the normalized scalar product between two independent samples $x^{(1)}$ and $x^{(2)}$ from (2.1). Understanding this quantity is fundamental because it allows to deduce the distance between two typical configurations of the system and thus encodes the “geometry” of the “Gibbs measure” (2.1). By the Nishimori identity (Proposition 3) $x^{(1)} \cdot x^{(2)} = \langle x \rangle \cdot x$ in law. Thus the overlap $x^{(1)} \cdot x^{(2)}$ corresponds to the correlation between a typical configuration and the planted configuration. Moreover it is linked to the Minimum Mean Squared Error by

$$\text{MMSE} = \frac{1}{n} \mathbb{E} \left[ \|X - \langle x \rangle\|^2 \right] = \mathbb{E}_{P_0}[X^2] - \mathbb{E}[\langle x \cdot X \rangle],$$

where $\langle \cdot \rangle$ denotes the expectation with respect to $x$ which is sampled from the posterior $P(X = \cdot \mid Y)$ (defined by Equation 2.1), independently of everything else.

In this section we will see a general principle that states that under a small perturbation of the Gibbs distribution (2.1), the overlap $x^{(1)} \cdot x^{(2)}$ between two replicas concentrates around its mean. Such behavior is called “Replica-Symmetric” in statistical physics. It remains to define what “a small perturbation of
where \( \mathbf{x} \) is the Gibbs distribution. In spin glass theory, such perturbations are usually obtained by adding small extra terms to the Hamiltonian. In our context of Bayesian inference a small perturbation will correspond to a small amount of side-information given to the statistician. This extra information will lead to a new posterior distribution. In the following, we will consider two different kind of side-information and we show that the overlaps under the induced posterior concentrate around their mean.

### 2.1 Revealing a small fraction of the planted solution

We suppose here that the support \( S \) of \( P_0 \) is finite. We make this assumption in order to be able to consider the discrete entropy.

In this section, we give extra information to the statistician by revealing a (small) fraction of the coordinates of \( \mathbf{X} \). Let us fix \( \epsilon \in [0, 1] \), and suppose that we have access to the additional observations

\[
Y_i' = \begin{cases} 
X_i & \text{if } L_i = 1, \\
* & \text{if } L_i = 0, 
\end{cases}
\]

for \( 1 \leq i \leq n \),

where \( L_i \overset{i.i.d.}{\sim} \text{Ber}(\epsilon) \) and \( * \) is a value that does not belong to \( S \). The posterior distribution of \( \mathbf{X} \) is now

\[
P(\mathbf{X} = \mathbf{x} \mid \mathbf{Y}, \mathbf{Y}') = \frac{1}{Z_{n,\epsilon}} \left( \prod_{i \mid L_i = 1} 1(x_i = Y_i') \right) \left( \prod_{i \mid L_i = 0} P_0(x_i) \right) e^{H_n(\mathbf{x})},
\]

(2.2)

where \( Z_{n,\epsilon} \) is the appropriate normalization constant. For \( \mathbf{x} \in S^n \) we define the following notation

\[
\bar{\mathbf{x}} = (\bar{x}_1, \ldots, \bar{x}_n) = (L_1 X_1 + (1 - L_1)x_1, \ldots, L_n X_n + (1 - L_n)x_n).
\]

(2.3)

\( \bar{\mathbf{x}} \) is thus obtained by replacing the coordinates of \( \mathbf{x} \) that are revealed by \( \mathbf{Y}' \) by their revealed values. The notation \( \bar{\mathbf{x}} \) allows us to obtain a convenient expression for the free energy of the perturbed model:

\[
F_{n,\epsilon} = \frac{1}{n} \mathbb{E} \log Z_{n,\epsilon} = \frac{1}{n} \mathbb{E} \left[ \log \sum_{\mathbf{x} \in S^n} P_0(\mathbf{x}) \exp(H_n(\bar{\mathbf{x}})) \right].
\]

**Proposition 6**

For all \( n \geq 1 \) and all \( \epsilon \in [0, 1] \), we have

\[
|F_{n,\epsilon} - F_n| \leq H(P_0)\epsilon.
\]

**Proof.** Let us compute

\[
P(\mathbf{Y}' \mid \mathbf{Y}, \mathbf{L}) = \int 1(x_i = Y_i') \text{ for all } i \text{ such that } L_i = 1) dP(\mathbf{x} \mid \mathbf{Y})
\]

\[
= \frac{1}{Z_n} \sum_{\mathbf{x} \in S^n} 1(x_i = Y_i') \text{ for all } i \text{ such that } L_i = 1) e^{H_n(\mathbf{x})} \prod_{i=1}^n P_0(x_i)
\]

\[
= \frac{Z_{n,\epsilon}}{Z_n} \prod_{i \mid L_i = 1} P_0(Y_i') = \frac{Z_{n,\epsilon}}{Z_n} P(\mathbf{Y}' \mid \mathbf{L}).
\]

Therefore, \( nF_{n,\epsilon} - nF_n = H(\mathbf{Y}' \mid \mathbf{L}) - H(\mathbf{Y}' \mid \mathbf{Y}, \mathbf{L}) \) and the proposition follows from the fact that \( 0 \leq H(\mathbf{Y}' \mid \mathbf{Y}, \mathbf{L}) \leq H(\mathbf{Y}' \mid \mathbf{L}) = n\epsilon H(P_0). \)

From now we suppose \( \epsilon_0 \in (0, 1] \) to be fixed and consider \( \epsilon \in [0, \epsilon_0] \). The following lemma comes from [51]. It shows that the extra information \( \mathbf{Y}' \) forces the correlations between the spins under the posterior (2.2) to vanish.
Lemma 1 (Lemma 3.1 from [51])

For all $\epsilon_0 \in [0, 1]$, we have

$$\int_0^{\epsilon_0} d\epsilon \left( \frac{1}{n^2} \sum_{1 \leq i, j \leq n} I(X_i; X_j | Y, Y') \right) \leq \frac{2}{n} H(P_0).$$

Let $\langle \cdot \rangle_{n, \epsilon}$ denote the expectation with respect to two independent samples $x^{(1)}, x^{(2)}$ from the posterior (2.2). Lemma 1 implies that the overlap between these two replicas concentrates:

**Proposition 7**

For all $\epsilon_0 \in [0, 1]$,

$$\int_0^{\epsilon_0} d\epsilon \mathbb{E} \left( \left( \frac{1}{n} \sum_{i=1}^n x_i^{(1)} x_i^{(2)} - \langle \frac{1}{n} \sum_{i=1}^n x_i^{(1)} x_i^{(2)} \rangle_{n, \epsilon} \right)^2 \right)_{n, \epsilon} \xrightarrow{n \to \infty} 0.$$

**Proof.**

$$\langle (x^{(1)} \cdot x^{(2)}) - \langle x^{(1)} \cdot x^{(2)} \rangle_{n, \epsilon} \rangle_{n, \epsilon}^2 = \langle (x^{(1)} \cdot x^{(2)})^2 \rangle_{n, \epsilon} - \langle x^{(1)} \cdot x^{(2)} \rangle_{n, \epsilon}^2$$

$$= \frac{1}{n^2} \sum_{1 \leq i, j \leq n} \langle x_i^{(1)} x_i^{(2)} x_j^{(1)} x_j^{(2)} \rangle_{n, \epsilon} - \langle x_i^{(1)} x_i^{(2)} \rangle_{n, \epsilon} \langle x_j^{(1)} x_j^{(2)} \rangle_{n, \epsilon}$$

$$= \frac{1}{n^2} \sum_{1 \leq i, j \leq n} (x_i x_j)^2_{n, \epsilon} - (x_i)^2_{n, \epsilon} (x_j)^2_{n, \epsilon} \leq C \frac{1}{n^2} \sum_{1 \leq i, j \leq n} \left| \langle x_i x_j \rangle_{n, \epsilon} - \langle x_i \rangle_{n, \epsilon} \langle x_j \rangle_{n, \epsilon} \right|$$

$$\leq C' \frac{1}{n^2} \sum_{1 \leq i, j \leq n} \text{D}_{\text{TV}}(\mathbb{P}(X_i = \cdot, X_j = \cdot | Y, Y') ; \mathbb{P}(X_i = \cdot | Y, Y') \otimes \mathbb{P}(X_j = \cdot | Y, Y'))$$

$$\leq C'' \frac{1}{n^2} \sum_{1 \leq i, j \leq n} \text{D}_{\text{KL}}(\mathbb{P}(X_i = \cdot, X_j = \cdot | Y, Y') ; \mathbb{P}(X_i = \cdot | Y, Y') \otimes \mathbb{P}(X_j = \cdot | Y, Y'))$$

for some constants $C, C', C'' > 0$, where we used Pinsker’s inequality to compare the total variation distance $\text{D}_{\text{TV}}$ with the Kullback-Leibler divergence $\text{D}_{\text{KL}}$. So that:

$$\int_0^{\epsilon_0} d\epsilon \mathbb{E} \left( \left( x^{(1)} \cdot x^{(2)} - \langle x^{(1)} \cdot x^{(2)} \rangle_{n, \epsilon} \right)^2 \right)_{n, \epsilon} \leq C'' \left( \epsilon_0 \int_0^{\epsilon_0} d\epsilon \left( \frac{1}{n^2} \sum_{1 \leq i, j \leq n} I(X_i; X_j | Y, Y') \right) \right) \xrightarrow{n \to \infty} 0.$$

\[\square\]

### 2.2 Noisy side Gaussian channel

We consider in this section of a different kind of side-information: an observation of the signal $X$ perturbed by some Gaussian noise. It was proved in [40] for CDMA systems that such perturbations forces the overlaps to concentrate around their means. The principle here is in fact more general and holds for any observation system, provided some concentration property of the free energy.
We suppose here that the prior \( P_0 \) has a bounded support \( S \subset [-K, K] \), for some \( K > 0 \). Let \( a \in [1/2, 3] \) and \((s_n)_n \in (0, 1)^N\). Let \((Z_i)_{1 \leq i \leq n} i.i.d. N(0, 1)\) independently of everything else. The extra side-information takes now the form

\[
Y'_i = a\sqrt{s_n}X_i + Z_i, \quad \text{for } 1 \leq i \leq n. \tag{2.4}
\]

The posterior distribution of \( X \) given \( Y, Y' \) is now \( \mathbb{P}(x \mid Y, Y') = \frac{1}{Z_{n,a}^{(\text{pert})}} P_0^{\otimes n}(x) \exp(H_{n,a}^{(\text{pert})}(x)) \), where

\[
H_{n,a}^{(\text{pert})}(x) = H_{n,a}(x) + h_{n,a}(x)
\]

\( Z_{n,a}^{(\text{pert})} \) is the appropriate normalization. The corresponding free energy is

\[
F_{n,a}^{(\text{pert})} = \frac{1}{n} \mathbb{E} \log Z_{n,a}^{(\text{pert})} = \frac{1}{n} \mathbb{E} \log \left( \int_{x \in S^n} dP_0^{\otimes n}(x) e^{H_{n,a}^{(\text{pert})}(x)} \right).
\]

**Lemma 2**

Assume that \( s_n \xrightarrow[n \to \infty]{} 0 \). Then \( |F_n - F_{n,a}^{(\text{pert})}| \xrightarrow[n \to \infty]{} 0 \) uniformly in \( a \in [1, 2] \) and consequently

\[
\left| F_n - \int_1^2 F_{n,a}^{(\text{pert})} \, da \right| \xrightarrow[n \to \infty]{} 0.
\]

**Proof.** Let \( \langle \cdot \rangle_n \) denote the measure on \( \mathbb{R}^n \) defined as \( \langle f(x) \rangle_n = \int_{x \in S^n} dP_0^{\otimes n}(x) f(x) \exp(H_{n,a}(x)) \) for every continuous bounded function \( f \). We have

\[
F_{n,a}^{(\text{pert})} - F_n = \frac{1}{n} \mathbb{E} \log \langle e^{h_{n,a}(x)} \rangle_n.
\]

Thus, using Jensen’s inequality twice

\[
\frac{1}{n} \mathbb{E} \langle Z h_{n,a}(x) \rangle_n = \frac{1}{n} \mathbb{E} \langle h_{n,a}(x) \rangle_n \leq F_{n,a}^{(\text{pert})} - F_n \leq \frac{1}{n} \mathbb{E} \log \mathbb{E} Z \langle e^{h_{n,a}(x)} \rangle_n = \frac{1}{n} \mathbb{E} \log \mathbb{E} Z e^{h_{n,a}(x)} \rangle_n,
\]

where \( \mathbb{E} Z \) denotes the expectation with respect to the variables \((Z_i)_{1 \leq i \leq n}\) only. We have, for all \( x \in S^n \),

\[
\frac{1}{n} \mathbb{E} Z h_{n,a}(x) \leq 2nK^2a^2s_n \quad \text{and} \quad |\mathbb{E} Z e^{h_{n,a}(x)}| \leq e^{Kn^2a^2s_n}.
\]

We conclude

\[
-2K^2a^2s_n \leq F_{n,a}^{(\text{pert})} - F_n \leq K^2a^2s_n.
\]

\( \Box \)

Let us define

\[
\phi : a \mapsto \frac{1}{ns_n} \mathbb{E} \log \left( \int_{x \in S} dP_0^{\otimes n}(x) e^{H_{n,a}^{(\text{pert})}(x)} \right).
\]

Define also \( v_n(s_n) = \sup_{1/2 \leq a \leq 3} \mathbb{E} |\phi(a) - \mathbb{E} \phi(a)| \). The following result shows that, in the perturbed system (under some conditions on \( v_n \) and \( s_n \)) the overlap between two replicas concentrates asymptotically around its expected value.

**Proposition 8 (Overlap concentration)**

**Suppose that**

\[
\left\{ \begin{array}{c}
v_n(s_n) \xrightarrow[n \to \infty]{} 0, \\
ns_n \xrightarrow[n \to \infty]{} +\infty.
\end{array} \right.
\]

**Then we have**

\[
\int_1^2 \mathbb{E} \langle (x^{(1)} \cdot x^{(2)}) - \mathbb{E} (x^{(1)} \cdot x^{(2)})_{n,a} \rangle_{n,a}^2 \, da \xrightarrow[n \to \infty]{} 0,
\]

where \( \langle \cdot \rangle_{n,a} \) denotes to the distribution of \( X \) given \( (Y, Y') \). \( x^{(1)} \) and \( x^{(2)} \) are two independent samples from \( \langle \cdot \rangle_{n,a} \), independently of everything else.
Lemma 3

Let \( x \) be a sample from \( \langle \cdot \rangle_{n,a} \), independently of everything else. Under the conditions of Proposition 8, we have

\[
\int_{1}^{2} E \left\langle |U(x) - E(U(x))_{n,a}| \right\rangle_{n,a} \, da \xrightarrow{n \to \infty} 0.
\]

Before proving Lemma 3, let us show how it implies Proposition 8.

Proof of Proposition 8. By the bounded support assumption on \( P_0 \), the overlap between two replicas is bounded by \( K^2 \), thus

\[
\left| E \left\langle U(x^{(1)}) \cdot x^{(2)} \right\rangle_{n,a} - E \left\langle x^{(1)} \cdot x^{(2)} \right\rangle_{n,a} \right| \leq K^2 \left( E \left\langle |U(x) - E(U(x))_{n,a}| \right\rangle_{n,a} \right).
\]

Let us compute the left-hand side of (2.5). By Gaussian integration by parts and using the Nishimori identity (Proposition 3) we get

\[
E \left\langle U(x^{(1)}) \right\rangle_{n,a} = 2a E \left\langle x^{(1)} \cdot x^{(2)} \right\rangle_{n,a}.
\]

Therefore

\[
E \left\langle x^{(1)} \cdot x^{(2)} \right\rangle_{n,a} \leq 2a \left( E \left\langle x^{(1)} \cdot x^{(2)} \right\rangle_{n,a} \right)^2.
\]

Using the same tools, we compute for \( x^{(1)}, x^{(2)}, x^{(3)}, x^{(4)} \overset{i.i.d.}{\sim} \langle \cdot \rangle_{n,a} \), independently of everything else:

\[
E \left\langle (x^{(1)} \cdot x^{(2)})^2 \right\rangle_{n,a} \leq \frac{K^2}{2} E \left\langle |U(x) - E(U(x))_{n,a}| \right\rangle_{n,a},
\]

and we conclude by integrating with respect to \( a \) over \([1,2]\) and using Lemma 3. \( \square \)

Proof of Lemma 3. \( \phi \) is twice differentiable on \([1/2, 3]\), and for \( a \in [1/2, 3] \)

\[
\phi'(a) = \langle U(x) \rangle_{n,a},
\]

\[
\phi''(a) = n s_n \left( \langle U(x)^2 \rangle_{n,a} - \langle U(x) \rangle^2_{n,a} \right) + \frac{1}{n} \sum_{i=1}^{n} \langle 2x_i X_i - x_i^2 \rangle_{n,a}.
\]

Thus \( \langle (U(x) - \langle U(x) \rangle_{n,a})^2 \rangle_{n,a} \leq \frac{1}{ns_n} (\phi''(a) + 2K^2) \) and

\[
\int_{1}^{2} E \langle (U(x) - \langle U(x) \rangle_{n,a})^2 \rangle_{n,a} \, da \leq \frac{1}{ns_n} \left( E \phi''(2) - E \phi'(1) + 2K^2 \right) = O(n^{-1}s_n^{-1}),
\]

because \( E \phi'(a) = 2a E(x \cdot X)_{n,a} \). Hence

\[
\int_{1}^{2} E \langle |U(x) - \langle U(x) \rangle_{n,a}| \rangle_{n,a} \, da \xrightarrow{n \to \infty} 0.
\]

It remains to show that

\[
\int_{1}^{2} E \langle |U(x) - \langle U(x) \rangle_{n,a}| \rangle_{n,a} \, da \xrightarrow{n \to \infty} 0.
\]

We will use the following lemma on convex functions (from [56], Lemma 3.2).
Lemma 4

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

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

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

We apply this lemma to $\lambda \mapsto \phi(\lambda) + \frac{3}{2}K^2\lambda^2$ and $\lambda \mapsto E\phi(\lambda) + \frac{3}{2}K^2\lambda^2$ that are convex because of (2.7) and the bounded support assumption on $P_0$. Therefore, for all $a \in [1, 2]$ and $b \in (0, 1/2)$ we have

$$E|\phi'(a) - E\phi'(a)| \leq E\phi'(a + b) - E\phi'(a - b) + 6K^2b + \frac{3v_n(s_n)}{b}. \quad (2.8)$$

Notice that for all $a \in [1/2, 3]$, $|E\phi'(a)| = |2aE(x \cdot X)_{n,a}| \leq 6K^2$. Therefore, by the mean value theorem

$$\int_1^2 (E\phi'(a + b) - E\phi'(a - b))da = (E\phi(b + 2) - E\phi(2 - b)) + (E\phi(1 - b) - E\phi(1 + b))$$

$$= (E\phi(b + 2) - E\phi(b + 1)) - (E\phi(2 - b) - E\phi(1 - b))$$

$$\leq 24K^2b.$$

Combining this with equation (2.8), we obtain

$$\forall b \in (0, 1/2), \int_1^2 E|\phi'(a) - E\phi'(a)|da \leq C\left(b + \frac{v_n(s_n)}{b}\right). \quad (2.9)$$

for some constant $C > 0$ depending only on $K$. The minimum of the right-hand side is achieved for $b = \sqrt{v_n(s_n)} < 1/2$ for $n$ large enough. Then, (2.9) gives

$$\int_1^2 E|\langle U(x)\rangle_{n,a} - E\langle U(x)\rangle_{n,a}|da = \int_1^2 E|\phi'(a) - E\phi'(a)|da \leq C\sqrt{v_n(s_n)} \xrightarrow{n \to \infty} 0.$$
Chapter 3

Low-rank symmetric matrix estimation

Let us now turn our attention to the spiked Wigner model (1). Let \( P_0 \) be a probability distribution on \( \mathbb{R} \) that admits a finite second moment and consider the following observation channel:

\[
Y_{i,j} = \sqrt{\frac{\lambda}{n}} X_i X_j + Z_{i,j}, \quad \text{for } 1 \leq i < j \leq n,
\]

(3.1)

where \( X_i \overset{i.i.d.}{\sim} P_0 \) and \( Z_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1) \) are independent random variables. Note that we suppose here to only observe the coefficients of \( \sqrt{\lambda/n} XX^\top + Z \) that are above the diagonal. The case where all the coefficients are observed can be directly deduced from this case. In the following, \( \mathbb{E} \) will denote the expectation with respect to the \( X \) and \( Z \) random variables.

Our main quantity of interest is the Minimum Mean Squared Error (MMSE) defined as:

\[
\text{MMSE}_n(\lambda) = \min_{\hat{\theta}} \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \mathbb{E} \left[ (X_i X_j - \hat{\theta}_{i,j}(Y))^2 \right],
\]

where the minimum is taken over all estimators \( \hat{\theta} \) (i.e. measurable functions of the observations \( Y \)). We have the trivial upper-bound

\[
\text{MMSE}_n(\lambda) \leq \text{DMSE} := \mathbb{E}_{P_0}[X^2]^2 - \mathbb{E}_{P_0}[X]^4,
\]

obtained by considering the “dummy” estimator \( \hat{\theta}_{i,j} = \mathbb{E}_{P_0}[X]^2 \). One can also compute the Mean Squared Error achieved by naive PCA. Let \( \hat{x} \) be the top eigenvector of \( Y \) with norm \( \|\hat{x}\|^2 = n \). If we take an estimator proportional to \( \hat{x}_i \hat{x}_j \), i.e. \( \hat{\theta}_{i,j} = \delta \hat{x}_i \hat{x}_j \) for \( \delta \geq 0 \), we can compute explicitly (using the results from [14] presented in the introduction) the resulting MSE as a function of \( \delta \) and minimize it. The optimal value for \( \delta \) depends on \( \lambda \), more precisely if \( \lambda < \mathbb{E}_{P_0}[X^2] - 2 \), then \( \delta = 0 \) while for \( \lambda \geq 1 \), the optimal value for \( \delta \) is \( \mathbb{E}_{P_0}[X^2] - \lambda^{-1}\mathbb{E}_{P_0}[X^2]^{-1} \), resulting in the following MSE for naive PCA:

\[
\text{MSE}_{n,\text{PCA}}(\lambda) \rightarrow \begin{cases} 
\mathbb{E}_{P_0}[X^2]^2 & \text{if } \lambda \leq \mathbb{E}_{P_0}[X^2] - 2, \\
\lambda^{-1}(2 - \lambda^{-1}\mathbb{E}_{P_0}[X^2]^{-2}) & \text{otherwise}. 
\end{cases}
\]

(3.2)

We will see in Section 3.2 that in the particular case of \( P_0 = \mathcal{N}(0, 1) \), PCA is optimal: \( \lim_{n \to \infty} \text{MSE}_{n,\text{PCA}} = \lim_{n \to \infty} \text{MMSE}_n \).

3.1 Information-theoretic limits

In order to formulate our inference problem as a statistical physics problem we introduce the random Hamiltonian

\[
H_n(x) = \sum_{i<j} \sqrt{\frac{\lambda}{n}} x_i x_j Z_{i,j} + \lambda \frac{X_i X_j x_i x_j}{n} - \frac{\lambda}{2n} x_i^2 x_j^2.
\]

16
The posterior distribution of $X$ given $Y$ takes now the form
\[
dP(x \mid Y) = \frac{1}{Z_n(\lambda)} dP^\otimes_n(x) \exp \left( \sum_{i < j} x_i x_j \sqrt{\frac{\lambda}{n}} Y_{i,j} - \frac{\lambda}{2n} x_i^2 x_j^2 \right) = \frac{1}{Z_n(\lambda)} dP^\otimes_n(x) e^{H_n(x)}, \tag{3.3}
\]
where $Z_n(\lambda)$ is the appropriate normalization. The free energy is defined as
\[
F_n(\lambda) = \frac{1}{n} \mathbb{E} \log \int dP^\otimes_n(x) e^{H_n(x)} = \frac{1}{n} \mathbb{E} \log Z_n(\lambda).
\]
We will first compute the limit of the free energy $F_n$ and then deduce the limit of $\text{MMSE}_n$ by an I-MMSE (see Proposition 4) argument. We will express the limit of $F_n$ using the following function
\[
F : (\lambda, q) \mapsto \psi_{P_0}(\lambda q) - \frac{\lambda}{4} q^2 = \mathbb{E} \log \left( \int dP_0(x) \exp \left( \sqrt{\lambda q} Z x + \lambda q x - \frac{\lambda}{2} q x^2 \right) \right) - \frac{\lambda}{4} q^2, \tag{3.4}
\]
where $Z \sim \mathcal{N}(0, 1)$ and $X \sim P_0$ are independent random variables. Recall that $\psi_{P_0}$ denotes the free energy of the scalar channel (1.5) and is studied by Proposition 5. The main result of this section is:

**Theorem 2 (Replica-Symmetric formula for the spiked Wigner model)**

For all $\lambda > 0$,
\[
F_n(\lambda) \xrightarrow{n \to \infty} \sup_{q \geq 0} F(\lambda, q). \tag{3.5}
\]

Theorem 2 is proved in Section 3.4. For the case of Rademacher prior ($P_0 = \frac{1}{2} \delta_{-1} + \frac{1}{2} \delta_{1}$), Theorem 2 was proved in [24]. The expression (3.5) for general priors was conjectured by [43]. For discrete priors $P_0$ for which the map $F(\lambda, \cdot)$ has not more than 3 stationary points, the statement of Theorem 2 was proved in [10]. The full version of Theorem 2 as well as its multidimensional generalization (where $X \in \mathbb{R}^{n \times k}$, $k$ fixed) was proved in [42].

The Replica-Symmetric formula allows us to compute the limit of the mutual information between the signal $X$ and the observations $Y$. Indeed, by using (1.3):

**Corollary 1**

\[
\lim_{n \to \infty} \frac{1}{n} I(X; Y) = \frac{\lambda \mathbb{E}_{P_0}(X^2)^2}{4} - \sup_{q \geq 0} F(\lambda, q).
\]

We will now use Theorem 2 to obtain the limit of the Minimum Mean Squared Error $\text{MMSE}_n$ by the I-MMSE relation of Proposition 4. Let us define
\[
D = \{ \lambda > 0 \mid F(\lambda, \cdot) \text{ has a unique maximizer } q^*(\lambda) \}.
\]

We start by computing the derivative of $\lim_{n \to \infty} F_n(\lambda)$ with respect to $\lambda$.

**Proposition 9**

\[
D \text{ is equal to } \mathbb{R}^*_+ \text{ minus some countable set and is precisely the set of } \lambda > 0 \text{ at which the function } \\
\phi : \lambda \mapsto \sup_{q \geq 0} F(\lambda, q) \text{ is differentiable. Moreover, for all } \lambda \in D \\\n\phi'(\lambda) = \frac{q^*(\lambda)^2}{4}.
\]

**Proof.** Let $\lambda > 0$ and compute
\[
\frac{\partial}{\partial q} F(\lambda, q) = \lambda \psi_{P_0}'(\lambda q) - \frac{\lambda q}{2} \leq \frac{\lambda}{2} (\mathbb{E}_{P_0}[X^2] - q),
\]

because \( \psi_{P_0} \) is \( \frac{1}{4} \mathbb{E}_{P_0}[X^2] \)-Lipschitz by Proposition 5. Consequently, the maximum of \( F(\lambda, \cdot) \) is achieved on \([0, \mathbb{E}_{P_0}[X^2]]\). For all maximizer \( q^* \) of \( F(\lambda, \cdot) \), the optimality condition gives \( q^* = 2\psi_{P_0}(\lambda q^*) \). Consequently, for all maximizer \( q^* \)

\[
\frac{\partial}{\partial \lambda} F(\lambda, q^*) = q^* \psi_{P_0}'(\lambda q^*) - \frac{\lambda (q^*)^2}{4} = \frac{(q^*)^2}{4}.
\]

Now, Proposition 21 in Appendix B gives that the \( \lambda > 0 \) at which \( \phi \) is differentiable is exactly the \( \lambda > 0 \) for which

\[
\left\{ \frac{\partial}{\partial \lambda} F(\lambda, q^*) = \frac{1}{4} (q^*)^2 \right\} q^* \text{ maximizer of } F(\lambda, \cdot)
\]

is a singleton. These \( \lambda \) are precisely the elements of \( D \). Moreover, Proposition 21 gives also that for all \( \lambda \in D \), \( \phi'(\lambda) = \frac{q^*(\lambda)^2}{4} \). This concludes the proof.

We deduce then the limit of \( \text{MMSE}_n \):

**Corollary 2**

\[
\text{MMSE}_n(\lambda) \xrightarrow{n \to \infty} (\mathbb{E}_{P_0}[X^2])^2 - q^*(\lambda)^2.
\]

**Proof.** By Proposition 4, \((F_n)_{n \geq 1}\) is a sequence of differentiable convex functions that converges pointwise on \( \mathbb{R}_+^* \) to \( \phi \). A standard lemma of convex analysis tells us that \( F_n'(\lambda) \xrightarrow{n \to \infty} \phi'(\lambda) \) for every \( \lambda > 0 \) at which \( \phi \) is differentiable, that is for all \( \lambda \in D \). We conclude using the I-MMSE relation (1.4):

\[
\frac{n - 1}{4n} (\mathbb{E}_{P_0}[X^2]^2 - \text{MMSE}_n(\lambda)) = F_n'(\lambda) \xrightarrow{n \to \infty} \phi'(\lambda) = \frac{q^*(\lambda)^2}{4}.
\]

Let us now define the information-theoretic threshold

\[
\lambda_c = \inf \left\{ \lambda \in D \left| q^*(\lambda) > (\mathbb{E}_{P_0}[X^2])^2 \right. \right\}.
\]

If the above set is empty, we define \( \lambda_c = 0 \). By Corollary 2 we obtain that

- if \( \lambda > \lambda_c \), then \( \lim_{n \to \infty} \text{MMSE}_n < \text{DMSE} \): one can estimate the signal better than a random guess.
- if \( \lambda < \lambda_c \), then \( \lim_{n \to \infty} \text{MMSE}_n = \text{DMSE} \): one can not estimate the signal better than a random guess.

Thus, there is no hope for reconstructing the signal below \( \lambda_c \). Interestingly, if one denotes by \( Q_\lambda \) the distribution of \( Y \) given by (3.1), the work from [4] shows that for \( \lambda < \lambda_c \) one can not asymptotically distinguish between \( Q_\lambda \) and \( Q_0 \): both distributions are contiguous.

### 3.2 Information-theoretic and algorithmic phase transitions

#### 3.2.1 Approximate Message Passing (AMP) algorithms

Approximate Message Passing (AMP) algorithms, introduced in [29] for compressed sensing, have been widely used to study the matrix factorization problem (4.1). Rigorous properties of AMP algorithms have then be established in [13, 36]. In the context of low-rank matrix estimation an AMP algorithm has been proposed by [61] for the rank-one case and then by [47] for finite-rank matrix estimation. For detailed review and developments about matrix factorization with message-passing algorithms, see [45]. We will not give a precise description of AMP here, we let the reader refer to [61, 25, 43, 53]. The only thing we
will need to know about the AMP algorithm is:

There exists an iterative polynomial-time algorithm algorithm called AMP that produces iterates \( \bar{x}^0, \ldots, \bar{x}^t \) such that for all \( t \geq 0 \) we have almost-surely

\[
|\bar{x}^t \cdot X| \xrightarrow{n \to \infty} q_t \quad \text{and} \quad \frac{1}{n} \mathbb{E} \| XX^\top - \bar{x}^t(\bar{x}^t)^\top \|^2 \xrightarrow{n \to \infty} \mathbb{E} P_0 [X^2]^2 - (q_t)^2,
\]

where \( (q_t)_{t \geq 0} \) is given by the recursion (called “state evolution")

\[
\begin{cases}
q_0 = 0, \\
q_t = 2 \psi'_{P_0}(\lambda q_{t-1}).
\end{cases}
\]

The sequence \( (q_t)_{t \geq 0} \) converges therefore to a critical point of \( F(\lambda, \cdot) \) and in the case where \( q_t \xrightarrow{t \to \infty} q^*(\lambda) \), AMP achieves the MMSE. An important issue here is that \( q = 0 \) can be a fixed point of the state evolution (this is the case when \( \mathbb{E} P_0 X = 0 \)). In that case, AMP has a trivial performance. To remedy to is, AMP can use spectral initialization techniques (see \([53]\)) in order to initialize the state evolution recursion with \( q_0 = \epsilon > 0 \). Even in the case where it does not reach the maximizer of \( F(\lambda, \cdot) \), AMP (with spectral initialization) is conjectured (see for instance \([67, 6]\)) to be optimal among polynomial-time algorithms, i.e.

\[
\text{MSE}_{\text{AMP}} = \mathbb{E} P_0 [X^2]^2 - \lim_{\epsilon \to 0} \lim_{t \to \infty} q_t^2_{|q_0=\epsilon}.
\]

\[ (3.9) \]

is conjectured to be the best Mean Squared Error achievable by any polynomial-time algorithm.

### 3.2.2 Examples of phase transitions

We give here some illustrations and interpretations of the results presented in the previous sections. Let us first study the case where \( P_0 = \mathcal{N}(0,1) \). In the case the formulas \((3.5)\) and \((3.6)\) can be evaluated explicitly. Indeed, as see at the end of Section 1.3, \( \psi_{\mathcal{N}(0,1)}(q) = \frac{1}{2} (q - \log(1 + q)) \), we can then compute \( q^*(\lambda) = (1 - \lambda^{-1})^+ \) which gives

\[
\lim_{n \to \infty} \text{MMSE}_n(\lambda) = \begin{cases}
0 & \text{if } \lambda \leq 1, \\
\frac{1}{\lambda} (2 - \frac{1}{\lambda}) & \text{if } \lambda \geq 1.
\end{cases}
\]

Comparing the limit above with the performance of (naive) PCA given by \((3.2)\) we see that in the case \( P_0 = \mathcal{N}(0,1) \), PCA is information-theoretically optimal.

However, as we see on \((3.2)\), the MSE of PCA only depends on the second moment of \( P_0 \): naive PCA is not able to exploit additional properties of the signal. We compare on Figure 3.1 the asymptotic performance of the naive PCA \((3.2)\) and the Approximate Message Passing (AMP) algorithm \((3.9)\) to the asymptotic Minimum Mean Squared Error for the prior

\[
P_0 = p \delta \sqrt{\frac{1-p}{p}} + (1-p) \delta - \sqrt{\frac{1-p}{p}}.
\]

\[ (3.10) \]

where \( p \in (0,1) \). This is a two-points distribution with zero mean and unit variance. It is of particular interest because it is related with the community detection problem in the (dense) Stochastic Block Model, see \([24, 42]\). We see on Figure 3.1 that the MMSE is equal to 1 for \( \lambda \) below the information-theoretic threshold \( \lambda_c \simeq 0.6 \). One can not asymptotically recover the signal better than a random guess in this region: we call this region the “impossible” phase. For \( \lambda > 1 \) we see that spectral methods and AMP perform better than random guessing. This region is therefore called the “easy” phase, because non-trivial estimation is here possible using efficient algorithms. Notice also that AMP achieves the Minimum Mean Squared Error for \( \lambda > 1 \), as proved in \([53]\). The region \( \lambda_c < \lambda < 1 \) is more intriguing. It is still possible to build a non-trivial estimator (for instance by computing the posterior mean), but our two polynomial-time algorithms fail. This region is thus denoted as the “hard” phase because it is conjectured here that
only exponential-time algorithms can provide non-trivial estimates.

Quite surprisingly, one can guess in which phase (easy-hard-impossible) we are, simply by plotting the “potential” \( q \mapsto -F(\lambda, q) \). This is done in Figure 3.2. By Corollary 2 we know that the limit of the

\[ MSE \approx \begin{cases} q^* & \text{easy phase} \\ \infty & \text{impossible phase} \end{cases} \]

MMSE is equal to \( 1 - q^*(\lambda)^2 \) where \( q^*(\lambda) \) is the minimizer of \( -F(\lambda, \cdot) \). Thus when \( -F(\lambda, \cdot) \) is minimal at \( q = 0 \), we are in the impossible phase. However, when \( q^*(\lambda) > 0 \), the shape of \( -F(\lambda, \cdot) \) indicates whether we are in the easy or hard phase. If the \( q = 0 \) is a local maximum, then we are in the easy phase, whereas when it is a local minimum we are in a hard phase. \( -F(\lambda, \cdot) \) could be interpreted as a simplified “free energy landscape”: the hard phase appears when the “informative” minimum \( q^*(\lambda) > 0 \) is separated from the non-informative critical point \( q = 0 \) by a “free energy barrier” as in Figure 3.2 (b).

The phase diagram from Figure 3.3 displays the three phases on the \((p, \lambda)\)-plane. One observes that the hard phase only appears when the prior is sufficiently asymmetric, i.e. for \( p < p^* = \frac{1}{2} - \frac{1}{2\sqrt{3}} \), as computed in [10, 18]. For a more detailed analysis of the phase transitions in the spiked Wigner model, see [45] where many other priors are considered.

**Figure 3.1:** Mean Squared Errors for the Spiked Wigner model with prior \( P_0 \) given by (3.10) with \( p = 0.05 \).

**Figure 3.2:** Plots of \( q \mapsto -F(\lambda, q) \) for different values of \( \lambda \) and \( P_0 \) given by (3.10) with \( p = 0.05 \).
3.3 Low-rank symmetric tensor estimation

We consider in this section the tensor-analog of the spiked Wigner model (1), namely the spiked tensor model introduced in [62]. Let \( p \geq 2 \) and consider

\[
Y = \sqrt{\frac{\lambda}{n^{p-1}}} X^\otimes p + Z. \tag{3.11}
\]

More precisely one supposes to observe

\[
Y_{i_1, \ldots, i_p} = \sqrt{\frac{\lambda}{n^{p-1}}} X_{i_1} \ldots X_{i_p} + Z_{i_1, \ldots, i_p} \quad \text{for } 1 \leq i_1 < \cdots < i_p \leq n, \tag{3.12}
\]

where \( X = (X_1, \ldots, X_n) \overset{iid}{\sim} P_0 \) and \((Z_{i_1, \ldots, i_p})_{i_1 < \cdots < i_p} \overset{iid}{\sim} \mathcal{N}(0, 1)\) are independent. We define the Hamiltonian

\[
H_n(x) = \sum_{i_1 < \cdots < i_p} \sqrt{\frac{\lambda}{n^{p-1}}} Y_{i_1, \ldots, i_p} x_{i_1} \ldots x_{i_p} - \frac{\lambda}{2n^{p-1}} (x_{i_1} \cdots x_{i_p})^2, \tag{3.13}
\]

for \( x = (x_1, \ldots, x_n) \in \mathbb{R}^n \). The posterior distribution of \( X \) given \( Y \) reads then:

\[
dP(X = x | Y) = \frac{1}{Z_n} dP_0(x) e^{H_n(x)}, \tag{3.14}
\]

where \( Z_n \) is the appropriate normalizing factor. The free energy is thus \( F_n(\lambda) = \frac{1}{n} \mathbb{E} \log Z_n \). Again, we will express the limit of \( F_n(\lambda) \) using the following "potential":

\[
\mathcal{F} : (\lambda, q) \mapsto \psi_{P_0} \left( \frac{\lambda^{q^p-1}}{(p-1)!} \right) - \frac{p - 1}{2p!} \lambda q^p, \tag{3.15}
\]

where \( \psi_{P_0} \) is the free energy of the scalar channel (1.5), defined by (1.6).

**Theorem 3 (Replica-Symmetric formula for the spiked tensor model)**

Let \( P_0 \) be a probability distribution over \( \mathbb{R} \), with finite second moment. Then, for all \( \lambda > 0 \)

\[
F_n(\lambda) \xrightarrow{n \to \infty} \sup_{q \geq 0} \mathcal{F}(\lambda, q). \tag{3.16}
\]
Theorem 3 was proved in [46] by the same arguments used for Theorem 2. By the I-MMSE relation (1.4) we deduce from Theorem 3 the limit of the Minimum Mean Squared Error:

$$\text{MMSE}_n(\lambda) = \inf_{\hat{\theta}} \left\{ \frac{p!}{n^p} \sum_{i_1 < \ldots < i_p} \left( X_{i_1} \ldots X_{i_p} - \hat{\theta}(Y)_{i_1 \ldots i_p} \right)^2 \right\},$$

where the infimum is taken over all measurable functions $\hat{\theta}$ of the observations $Y$.

**Corollary 3**

For almost every $\lambda > 0$, $\mathcal{F}(\lambda, \cdot)$ admits a unique maximizer $q^*(\lambda)$ over $\mathbb{R}_+$ and

$$\text{MMSE}_n(\lambda) \xrightarrow{n \to \infty} \left( \mathbb{E}_{P_0}X^2 \right)^p - q^*(\lambda)^p.$$

The proof is exactly the same than for the matrix case, see the proof of Corollary 2.

The information-theoretic threshold $\lambda_c$ is the minimal value of $\lambda$ such that $\lim \text{MMSE}_n < \left( \mathbb{E}_{P_0}X^2 \right)^p - \mathbb{E}_{P_0}[X]^{2p}$ (which is the performance achieved by random guess). We obtain thus the precise location of the information-theoretic threshold:

$$\lambda_c = \inf \left\{ \lambda > 0 \mid q^*(\lambda) > \mathbb{E}_{P_0}[X]^2 \right\}.$$

For Rademacher prior $P_0 = \frac{1}{2} \delta_+ + \frac{1}{2} \delta_-$, the value of $\lambda_c$ was also shown (see [20]) to be the “detection threshold”: for $\lambda > \lambda_c$, $\text{DTV}(Q_0, Q_{\lambda}) \xrightarrow{n \to \infty} 1$, whereas for $\lambda < \lambda_c$, $\text{DTV}(Q_0, Q_{\lambda}) \xrightarrow{n \to \infty} 0$, where $Q_{\lambda}$ denotes here the law of $Y$ given by (3.12).

### 3.3.1 Large order limit

We will now make a simple analysis of the symmetric tensor estimation model (3.12) with Rademacher prior $P_0 = \frac{1}{2} \delta_+ + \frac{1}{2} \delta_-$ in the limit $p \to \infty$. This large $p$ scenario has been studied in [59], where the detection problem is also investigated. For scaling and convenience we will suppose to observe

$$Y_{i_1, \ldots, i_p} = \sqrt{\frac{\lambda}{n^{p-1}}} X_{i_1} \ldots X_{i_p} + Z_{i_1, \ldots, i_p}, \quad (3.17)$$

for all $(i_1, \ldots, i_p) \in \{1, \ldots, n\}^p$. The $Z_{i_1, \ldots, i_p}$ are i.i.d. standard Gaussian, independent from $X_1, \ldots, X_n$ i.i.d. $\frac{1}{2} \delta_+ + \frac{1}{2} \delta_-$. Note that we suppose here to observe all the entries of the tensor $Y$, this correspond therefore to the observation model (3.12) with signal-to-noise ratio equal to $\lambda p!$. We will denote by $F_n^{(p)}(\lambda)$ and $\text{MMSE}_n^{(p)}(\lambda)$ the corresponding free energy and MMSE.

We will see that simple arguments (that does not require the knowledge of the exact formulas of Theorem 3 and Corollary 3 above) give that for large values of $p$ we have

- if $\lambda \leq 2 \log(2) - O((\log(p)^2/p^3)$ then $\lim_{n \to \infty} \text{MMSE}_n^{(p)}(\lambda) = 1$,
- if $\lambda \geq 2 \log(2)$ then $\lim_{n \to \infty} \text{MMSE}_n^{(p)}(\lambda) = O((\log(p)^2/p^3)$.

We start with the study of the free energy. For $\lambda_0 \in \mathbb{R}$ we define

$$f_{\lambda_0}(\lambda) = \begin{cases} 0 & \text{if } \lambda \leq \lambda_0, \\ \frac{1}{2}(\lambda - \lambda_0) & \text{if } \lambda \geq \lambda_0. \end{cases}$$
Proposition 10

There exists an increasing sequence \((\lambda_p^+)_p \geq 2\) such that \(\lambda_2^+ = 1/2\) and \(\lambda_p^+ = 2\log(2) - O(\log(p)^2/p^2)\) and for all \(p, n, \lambda \geq 0\)

\[
f_{2\log(2)}(\lambda) \leq F_n^{(p)}(\lambda) \leq f_{\lambda_p^+}(\lambda) + \sqrt{\frac{\lambda}{n} + \frac{\log(2n + 1)}{n}}.
\]

Proof. We start with the lower bound. For all \(\lambda \geq 0\) we have \(F_n^{(p)}(\lambda) \geq 0\) and

\[
F_n^{(p)}(\lambda) = \frac{1}{n} \log \left( \sum_{x \in \{-1,1\}^n} \frac{1}{2^n} \exp \left( \sum_{i_1, \ldots, i_p} \frac{\lambda}{n_p-1} Y_{i_1, \ldots, i_p} x_{i_1} \cdots x_{i_p} - \frac{\lambda n}{2} \right) \right).
\]

Let us now prove the upper-bound. For \(x \in \{-1,1\}^n\) let us write \(Z(x) = n^{-p/2} \sum_{i_1, \ldots, i_p} Z_{i_1, \ldots, i_p} x_{i_1} \cdots x_{i_p}\).

Using this notation, the Hamiltonian reads:

\[
H_n(x) = \sqrt{\lambda n} Z(x) + \lambda n (x \cdot X)^p - \frac{\lambda n}{2}.
\]

By Jensen’s inequality, we have

\[
F_n^{(p)}(\lambda) \leq \frac{1}{n} \log \left( \sum_{x \in \{-1,1\}^n} \frac{1}{2^n} \mathbb{E} \left[ \exp \left( \sqrt{\lambda n} Z(x) + \lambda n (x \cdot X)^p - \frac{\lambda n}{2} \right) \middle| X, Z(X) \right] \right).
\]

For \(x^{(1)}, x^{(2)} \in \{-1,1\}^n\) fixed, the covariance between \(Z(x^{(1)})\) and \(Z(x^{(2)})\) is \(\mathbb{E}[Z(x^{(1)})Z(x^{(2)})] = (x^{(1)} \cdot x^{(2)})^p\). Consequently, for \(x \in \{-1,1\}^n\) the law of \(Z(x)\) conditionally on \(Z(X)\) is \(\mathcal{N}((x \cdot X)^p, 1 - (x \cdot X)^{2p})\) and therefore

\[
\mathbb{E} \left[ \exp \left( \sqrt{\lambda n} Z(x) \right) \middle| X, Z(X) \right] = \exp \left( \sqrt{\lambda n} (x \cdot X)^p Z(X) + \frac{\lambda n}{2} (1 - (x \cdot X)^{2p}) \right).
\]

We obtain thus:

\[
F_n^{(p)}(\lambda) \leq \frac{1}{n} \log \left( \sum_{x \in \{-1,1\}^n} \frac{1}{2^n} \exp \left( \sqrt{\lambda n} |Z(X)| + \frac{\lambda n}{2} (2(x \cdot X)^p - (x \cdot X)^{2p}) \right) \right)
\]

\[
\leq \frac{1}{n} \log \left( \sum_{x \in \{-1,1\}^n} \frac{1}{2^n} \exp \left( \frac{\lambda n}{2} (2(x \cdot X)^p - (x \cdot X)^{2p}) \right) \right) + \sqrt{\frac{\lambda}{n}}.
\]

Now, for \(k \in \{-n, \ldots, n\}\), we have

\[
\# \left\{ x \in \{-1,1\}^n \middle| \sum_{i=1}^n x_i X_i = k \right\} \leq 2^n \exp \left( -\frac{n}{2} I(k/n) \right),
\]

where \(I(t) = (1 + t) \log(1 + t) + (1 - t) \log(1 - t)\). This gives

\[
F_n^{(p)}(\lambda) \leq \frac{1}{n} \log \left( \sum_{k=-n}^n \exp \left( \frac{n}{2} (2\lambda(k/n)^p - \lambda(k/n)^{2p} - I(k/n)) \right) \right) + \sqrt{\frac{\lambda}{n}}
\]

\[
\leq \frac{1}{n} \log \left( (2n + 1) \exp \left( \frac{n}{2} \max_{t \in [0,1]} \{2\lambda t^p - \lambda t^{2p} - I(t)\} \right) \right) + \sqrt{\frac{\lambda}{n}}
\]

\[
= \frac{1}{2} \max_{t \in [0,1]} \left\{ \lambda(2t^p - t^{2p}) - I(t) \right\} + \log(2n + 1) + \sqrt{\frac{\lambda}{n}}.
\]
The function \( \lambda \mapsto \max_{t \in [0,1]} \{ \lambda(2t^p - t^{2p}) - I(t) \} \) is continuous, 1-Lipschitz and equal to 0 for \( \lambda < \lambda_p^\sim \) where
\[
\lambda_p^\sim = \inf_{t \in (0,1)} \frac{I(t)}{2t^p - t^{2p}}.
\]
Therefore, for \( \lambda \geq \lambda_p^\sim \), \( \max_{t \in [0,1]} \{ \lambda(2t^p - t^{2p}) - I(t) \} \leq \lambda - \lambda_p^\sim \).

It remains to show \( \lambda_p^\sim = 1/2 \) and \( 2 \log(2) - O(\log(p)^2/p^2) \leq \lambda_p^\sim \leq 2 \log(2) \). Let us start with the case \( p = 2 \). The maximizer \( t \mapsto \lambda(2t^2 - t^4) - I(t) \) satisfies
\[
4\lambda (t - t^3) = I'(t) = \log \left( \frac{1 + t}{1 - t} \right) = 2 \text{artanh}(t).
\]
Therefore \( t = \tanh(2\lambda(t - t^3)) \). For \( \lambda \leq 1/2 \), this equation admits a unique solution \( t^0 \), whereas of \( \lambda > 1/2 \) it admits a second solution \( t^1 \). If \( \lambda > 2 \log(2) \), \( \lambda = 1/2 \) and \( \lambda_p^\sim = 1/2 \).

Let now \( p \geq 3 \). Let \( t_p \) be the largest minimizer of \( h_p : t \mapsto I(t)/(2t^p - t^{2p}) \) on \((0,1]\). One have \( t_p \in (0,1) \). For \( t \in (0,1) \), \( h_p'(t) \) has the same sign as
\[
\frac{t(\log(1 + t) - \log(1 - t))}{I(t)} - \frac{2p(1 - t^p)}{2 - t^p},
\]
which is decreasing in \( p \). This gives that \( (t_p)_{p \geq 3} \) is increasing and converges to 1, which is the only possible limit because \( t_p \) cancels (3.18). One has also \( t_p^p \to 1 \). Define \( u_p = 1 - t_p \). Since \( t_p^p \to 1 \), one has \( pu_p \to 0 \). Then \( t_p^p = \exp(p \log(1 - u_p)) = 1 - pu_p + O(pu_p^2) \). We get
\[
- \log(u_p) + o(1) = p \log(2) \frac{pu_p + O(u_p^2)}{1 + o(1)} = \log(2)p^2 u_p + o(1).
\]
Therefore
\[
2 \log(p) = \log(2)p^2 u_p + \log(p^2 u_p) + o(1) \sim \log(2)p^2 u_p.
\]
We deduce that \( 1 - t_p = u_p \sim \frac{2 \log(p)}{\log(2)p^2} \). We have then \( 2t_p^p - t_p^{2p} = 1 - p^2 u_p^2 + o(2u_p^2) \) and \( I(t_p) = 2 \log(2) + O(u_p) + u_p \log(u_p) \). We conclude:
\[
\lambda_p^\sim = h_p(t_p) = 2 \log(2) - O(\log(p)^2/p^2).
\]

Proof. By the I-MMSE relation (1.4), we have \( \text{MMSE}_n^{(p)}(\lambda) = 1 - 2F_n^{(p)}(\lambda) \). If \( \lambda < \lambda_p^\sim \) then by convexity and the fact that the free energy is non-decreasing:
\[
0 \leq F_n^{(p)}(\lambda) \leq \frac{F_n^{(p)}(\lambda_p^\sim) - F_n^{(p)}(\lambda)}{\lambda_p^\sim - \lambda} \xrightarrow{n \to \infty} 0,
\]
Figure 3.4: Minimal Mean Squared Errors $\text{MMSE}_n^{(p)}$ for tensor estimation (3.17) with Rademacher prior, for $p = 2, 3, 4, 6$, as given by Corollary 3.

by Proposition 10. This proves the result for $\lambda < \lambda_p^\sim$. For $\lambda > 2 \log(2)$ again, by convexity and Proposition 10 we have

$$\liminf_{n \to \infty} F_n^{(p)}(\lambda) \geq \liminf_{n \to \infty} \frac{F_n^{(p)}(\lambda) - F_n^{(p)}(2 \log(2))}{\lambda - 2 \log(2)} \geq \frac{f_{2 \log(2)}(\lambda) - f_{2 \log(2)}(2 \log(2))}{\lambda - 2 \log(2)} = \frac{1}{2} - \frac{2 \log(2) - \lambda_p^\sim}{2(\lambda - 2 \log(2))},$$

which concludes the proof.

The “abrupt” phase transition at $\lambda = 2 \log(2)$ that we see on Figure 3.4 reminds of the phase transition for the “needle in a haystack” problem seen in Section 1.4. This is not surprising, and this has been known for a long time in statistical physics: the Random Energy Model (which is the non-planted analog of the needle in a haystack problem) can be seen as the $p \to \infty$ limit of the $p$-spin model (which corresponds to the spiked tensor model (3.17)), see [22].

3.3.2 Hardness of low-rank tensor estimation

The brutal jump of the minimal mean squared error on Figure 3.4 and the fact that tensor estimation is closed to the “needle in a haystack” problem of Section 1.4 seems to indicate that the low-rank tensor estimation problem (3.11) for $p \geq 3$ is computationally hard. The study of [62] supports this picture and show that unless the signal-to-noise ratio $\lambda$ goes to infinity with $n$, popular algorithms such as power iteration, tensor unfolding or message passing fail to recover the signal $X$, when $X$ is uniformly distributed on the sphere of radius $\sqrt{n}$.

This suggests that we would be in a “hard regime” (where polynomial time algorithms can only achieve trivial performance) for all (finite) values of the signal-to-noise ratio $\lambda$. The work [46] provides a more optimistic vision that can be summarized as:

- If the distribution of the signal has zero mean (i.e. $\mathbb{E}_{P_0}X = 0$) we are indeed in a hard phase for all values of $\lambda$.

- However, if $\mathbb{E}_{P_0}X \neq 0$ then polynomial-time algorithms (as AMP) can achieve a non-trivial performance and can even be optimal if $\mathbb{E}_{P_0}X$ is not too small.
We will now give some intuition about these points. As we have seen in Section 3.2 the presence of a “hard regime” is characterized by the fact that \( q = 0 \) is a local minimum of the potential \( q \mapsto -\mathcal{F}(\lambda, q) \). We thus expand around \( q = 0 \)

\[
-\mathcal{F}(\lambda, q) = -\frac{\lambda q^{p-1}}{2(p-1)!} E_{P_0}[X]^2 + \frac{\lambda(p-1)}{2p!} q^p + O(q^{2(p-1)}) ,
\]

because \( \psi_{P_0}(0) = 0 \) and \( \psi'_{P_0} = \frac{1}{2} E_{P_0}[X]^2 \). Consequently, if \( E_{P_0} X = 0 \) and \( p \geq 3 \), then \( q = 0 \) is a local minimum of \(-\mathcal{F}(\lambda, \cdot)\) and we are in a “hard regime”. But if the prior \( P_0 \) has a non-zero mean, then \( q = 0 \) is not a local minimum anymore and it is possible (with AMP for instance) to estimate the signal \( X \) with efficient algorithms.

![Graphs](image)

**Figure 3.5:** Plots of \( q \mapsto -\mathcal{F}(\lambda, q) \) for \( p = 3 \), \( \lambda = 10 \) and \( P_0 = \mathcal{N}(\mu, 1) \), for different values of \( \mu \).

The plots of Figure 3.5 confirm this picture. On the first plot (where \( P_0 = \mathcal{N}(\mu = 0, 1) \)) we observe that the local minimum \( q = 0 \) is separated from the global minimum by a barrier, which indicates hardness (see the discussion in Section 3.2.2). Since \( X = (X_1, \ldots, X_n) \overset{i.i.d.}\sim \mathcal{N}(0, 1) \) is up to a normalization uniformly distributed on the sphere of radius \( \sqrt{n} \) this is coherent with the above mentioned results of [62]. On the second plot, where the prior has a small mean \( \mu = 0.15 \), the local minimum at \( q = 0 \) disappears and is replaced by another local minimum at \( q_0 \), close to 0. It is possible in this situation to achieve non-trivial performance by efficient algorithms (as AMP), but it is again conjectured that there correlation with the planted solution \( X \) will be at most equal to \( q_0 \), which is quite small compared to the optimal overlap \( q^*(\lambda) \). Polynomial-time algorithms can thus have non-trivial performance but are still far from optimal. On the third and fourth plot, we see that for larger means, the local minimum around zero disappears completely. It is now possible (using for instance AMP) to achieve the optimal performance in polynomial time.

### 3.4 Proof of the Replica-Symmetric formula (Theorem 2)

We prove Theorem 2 in this section, following [42]. We have to mention that other proofs of Theorem 2 has appeared since then: see [11, 3].

Because of an approximation argument presented in Section 3.4.7 it suffices to prove Theorem 2 for priors \( P_0 \) with finite (and thus bounded) support \( S \subset [-K, K] \), for some \( K > 0 \). From now, we assume to be in that case.

#### 3.4.1 The lower bound: Guerra’s interpolation method

The following result comes from [41]. It adapts arguments from the study of the gauge symmetric p-spin glass model of [39] to the inference model (3.1). It is based on Guerra’s interpolation technique for the Sherrington-Kirkpatrick model, see [32]. We reproduce the proof for completeness.

**Proposition 11**

\[
\liminf_{n \to \infty} F_n(\lambda) \geq \sup_{q \geq 0} \mathcal{F}(\lambda, q) .
\]
Proof. Let \( q \geq 0 \). For \( t \in [0, 1] \) we define
\[
H_{n,t}(x) = \sum_{i<j} \frac{\lambda t}{n} Z_{i,j} x_i x_j + \frac{\lambda}{n} x_i x_j X_j - \frac{\lambda t}{2n} x_i^2 x_j^2 + \sum_{i=1}^{n} \sqrt{(1-t)\lambda q Z'_i x_i + (1-t)\lambda q x_i X_i - \frac{1-t}{2} \lambda q x_i^2}.
\]

Let \( \langle \cdot \rangle_{n,t} \) denote the Gibbs measure associated with the Hamiltonian \( H_{n,t}(x) \):
\[
\langle f(x) \rangle_{n,t} = \frac{\sum_{x \in S^n} P^{\otimes n}_{0}(x) f(x) e^{H_{n,t}(x)}}{\sum_{x \in S^n} P^{\otimes n}_{0}(x) e^{H_{n,t}(x)}},
\]
for any function \( f \) on \( S^n \). The Gibbs measure \( \langle \cdot \rangle_{n,t} \) corresponds to the distribution of \( X \) given \( Y \) and \( Y' \) in the following inference channel:
\[
\begin{align*}
Y_{i,j} &= \sqrt{\frac{\lambda t}{n}} X_i X_j + Z_{i,j} & \text{for } 1 \leq i < j \leq n, \\
Y'_i &= \sqrt{(1-t)\lambda q X_i + Z'_i} & \text{for } 1 \leq i \leq n,
\end{align*}
\]
where \( X_i \overset{i.i.d.}{\sim} P_0 \) and \( Z_{i,j}, Z'_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1) \) are independent random variables. We will therefore be able to apply the Nishimori property (Proposition 3) to the Gibbs measure \( \langle \cdot \rangle_{n,t} \). Let us define
\[
\psi : t \in [0, 1] \mapsto \frac{1}{n} \mathbb{E} \log \sum_{x \in S^n} P^{\otimes n}_{0}(x) e^{H_{n,t}(x)}.
\]
We have \( \psi(1) = F_n(\lambda) \) and
\[
\begin{align*}
\psi(0) &= \frac{1}{n} \mathbb{E} \log \sum_{x \in S^n} P^{\otimes n}_{0}(x) \exp \left( \sum_{i=1}^{n} \sqrt{\lambda q Z'_i x_i + \lambda q x_i X_i - \frac{\lambda q x_i^2}{2}} \right) \\
&= \frac{1}{n} \mathbb{E} \log \prod_{i=1}^{n} \left( \sum_{x_i \in S} P_0(x_i) \exp \left( \sqrt{\lambda q Z'_i x_i + \lambda q x_i X_i - \frac{\lambda q x_i^2}{2}} \right) \right) \\
&= F(\lambda, q) + \frac{\lambda q^2}{4}.
\end{align*}
\]
\( \psi \) is continuous on \([0, 1]\), differentiable on \((0, 1)\). For \( 0 < t < 1 \),
\[
\psi'(t) = \frac{1}{n} \mathbb{E} \left( \sum_{i<j} \frac{\sqrt{\lambda}}{2\sqrt{nt}} Z_{i,j} x_i x_j + \frac{\lambda}{n} x_i x_j X_j - \frac{\lambda}{2n} x_i^2 x_j^2 - \sum_{i=1}^{n} \frac{\sqrt{\lambda q}}{2\sqrt{1-t}} Z'_i x_i - \lambda q x_i X_i + \frac{\lambda q}{2} x_i^2 \right)_{n,t}.
\]
For \( 1 \leq i < j \leq n \) we have, by Gaussian integration by parts and by the Nishimori property
\[
\mathbb{E} \left[ Z_{i,j} \left( \frac{\sqrt{\lambda}}{2\sqrt{nt}} x_i x_j \right)_{n,t} \right] = \frac{\lambda}{2n} \left( \mathbb{E} \langle x_i^2 x_j^2 \rangle_{n,t} - \mathbb{E} \langle x_i x_j \rangle_{n,t}^2 \right) = \frac{\lambda}{2n} \left( \mathbb{E} \langle x_i^2 x_j^2 \rangle_{n,t} - \mathbb{E} \langle x_i(1) x_j(1) x_i(2) x_j(2) \rangle_{n,t} \right)
\]
\[
= \frac{\lambda}{2n} \left( \mathbb{E} \langle x_i^2 x_j^2 \rangle_{n,t} - \mathbb{E} \langle x_i x_j X_i \rangle_{n,t} \right).
\]
Similarly, we have for \( 1 \leq i \leq n \)
\[
\mathbb{E} \left( \frac{\sqrt{\lambda q}}{2\sqrt{1-t}} Z'_i x_i \right)_{n,t} = \frac{\lambda q}{2} \left( \mathbb{E} \langle x_i^2 \rangle_{n,t} - \mathbb{E} \langle x_i X_i \rangle_{n,t} \right).
\]
Therefore (3.20) simplifies
\[
\psi'(t) = \frac{1}{n} \mathbb{E} \left( \sum_{i<j} \frac{\lambda}{2n} x_i x_j X_j - \frac{\lambda q}{2} x_i X_i \right)_{n,t} = \frac{\lambda}{4} \mathbb{E} \langle (\mathbf{x} \cdot \mathbf{X})^2 - 2q \mathbf{x} \cdot \mathbf{X} \rangle_{n,t} + o_n(1)
\]
\[
= \frac{\lambda}{4} \mathbb{E} \langle (\mathbf{x} \cdot \mathbf{X} - q)^2 \rangle_{n,t} - \frac{\lambda q^2}{4} + o_n(1) \geq -\frac{\lambda q^2}{4} + o_n(1),
\]
(3.21)
where \( o_n(1) \) denotes a quantity that goes to 0 uniformly in \( t \in (0, 1) \). Then
\[
F_n(\lambda) - F(\lambda, q) - \frac{\lambda}{4} q^2 = \psi(1) - \psi(0) = \int_0^1 \psi'(t) dt \geq -\frac{\lambda}{4} q^2 + o_n(1).
\]
Thus \( \liminf_{n \to \infty} F_n(\lambda) \geq F(\lambda, q) \), for all \( q \geq 0 \).

### 3.4.2 Adding a small perturbation

It remains to prove the converse bound of (3.19). For this purpose, we will need that the overlap \( x \cdot X \) concentrates around its mean. To obtain such a result, we follow the ideas of Section 2.1 that states that giving a small amount of side information to the statistician forces the overlap to concentrate, while keeping the free energy almost unchanged.

Let us fix \( \epsilon \in [0, 1] \), and suppose we have access to the additional information, for \( 1 \leq i \leq n \)
\[
Y_i' = \begin{cases} 
X_i & \text{if } L_i = 1, \\
* & \text{if } L_i = 0,
\end{cases}
\]
where \( L_i \overset{i.i.d.}\sim \text{Ber}(\epsilon) \) and \( * \) is a value that does not belong to \( S \). Recall the free energy that corresponds to this perturbed inference channel is
\[
F_{n,\epsilon} = \frac{1}{n} \mathbb{E} \left[ \log \sum_{x \in S^n} P_0^{\otimes n}(x) \exp(H_n(\bar{x})) \right],
\]
where
\[
\bar{x} = (\bar{x}_1, \ldots, \bar{x}_n) = (L_1 X_1 + (1 - L_1) x_1, \ldots, L_n X_n + (1 - L_n) x_n).
\]
From now we suppose \( \epsilon_0 \in (0, 1) \) to be fixed and consider \( \epsilon \in [0, \epsilon_0] \). We will compute the limit of \( F_{n,\epsilon} \) as \( n \to \infty \) and then let \( \epsilon \to 0 \) to deduce the limit of \( F_n \), because by Proposition 6
\[
|F_{n,\epsilon} - F_n| \leq H(P_0) \epsilon.
\]

### 3.4.3 Aizenman-Sims-Starr scheme

The Aizenman-Sims-Starr scheme was introduced in [1] in the context of the SK model. This is what physicists call a “cavity computation”: one compare the system with \( n + 1 \) variables to the system with \( n \) variables and see what happen to the \((n+1)\)th variable we add.

With the convention \( F_{0,\epsilon} = 0 \), we have \( F_{n,\epsilon} = \frac{1}{n} \sum_{i=0}^{n-1} A_{k,\epsilon}^{(i)} \) where
\[
A_{k,\epsilon}^{(i)} = (k + 1) F_{k+1,\epsilon} - k F_{k,\epsilon} = \mathbb{E} \left[ \log(Z_{k+1,\epsilon}) \right] - \mathbb{E} \left[ \log(Z_{k,\epsilon}) \right].
\]
We recall that \( Z_{n,\epsilon} = \sum_{x \in S^n} P_0^{\otimes n}(x) \exp(H_n(\bar{x})) \) where the notation \( \bar{x} \) is defined by equation (3.23). Consequently
\[
\limsup_{n \to \infty} \int_0^{\epsilon_0} dF_{n,\epsilon} \leq \limsup_{n \to \infty} \int_0^{\epsilon_0} dA_{n,\epsilon}^{(i)}.
\]
We now compare \( H_{n+1} \) with \( H_n \). Let \( x \in S^n \) and \( \sigma \in S \). \( \sigma \) plays the role of the \((n+1)\)th variable. We decompose \( H_{n+1}(x, \sigma) = H'_n(x) + \sigma z_0(x) + \sigma^2 s_0(x) \), where
\[
H'_n(x) = \sum_{1 \leq i < j \leq n} \frac{\lambda}{n+1} Z_{i,j} x_i x_j + \frac{\lambda}{n+1} X_i X_j x_i x_j - \frac{\lambda}{2(n+1)} x_i^2 x_j^2,
\]
\[
z_0(x) = \sum_{i=1}^{n} \sqrt{\frac{\lambda}{n+1}} Z_{i,n+1} x_i + \frac{\lambda}{n+1} X_i X_{n+1} x_i,
\]
\[
s_0(x) = -\frac{\lambda}{2(n+1)} \sum_{i=1}^{n} x_i^2.
\]
Let \((Z_{i,j})_{1 \leq i < j \leq n}\) be independent, standard Gaussian random variables, independent of all other random variables. We have then \(H_n(x) = H'_n(x) + y_0(x)\) in law, where

\[
y_0(x) = \sum_{1 \leq i < j \leq n} \sqrt{\frac{\lambda}{n(n+1)}} Z_{i,j} x_i x_j + \frac{\lambda}{n(n+1)} X_i X_j x_i x_j - \frac{\lambda}{2(n+1)n} x_i^2 x_j^2.
\]

We define the Gibbs measure \(\langle \cdot \rangle_{n,\epsilon}\) by

\[
\langle f(x) \rangle_{n,\epsilon} = \frac{1}{Z_{n,\epsilon}} \sum_{x \in S^n} P_0(x) f(\bar{x}) \exp(H'_n(\bar{x})),
\]

for any function \(f\) on \(S^n\). Let us define \(\bar{\sigma} = (1 - L_{n+1})\sigma + L_{n+1} X_{n+1}\). We can rewrite \(Z_{n+1,\epsilon} = \sum_{x \in S^n} P_0^n(x) e^{H'_n(\bar{x})} \left( \sum_{\sigma \in S} P_0(\sigma) \exp(\sigma z_0(\bar{x}) + \bar{\sigma}^2 s_0(\bar{x})) \right)\). Thus

\[
A_{n,\epsilon}^{(0)} = \mathbb{E} \log \left( \sum_{\sigma \in S} P_0(\sigma) \exp(\sigma z_0(x) + \bar{\sigma}^2 s_0(x)) \right)_{n,\epsilon} - \mathbb{E} \log \left( \exp(y_0(x)) \right)_{n,\epsilon}.
\]

In the sequel, it will be more convenient to use “simplified” versions of \(z_0, s_0\) and \(y_0\) in order to obtain “nice expressions”. We define

\[
z(x) = \sum_{i=1}^{n} \sqrt{\frac{\lambda}{n}} Z_{i,n+1} x_i + \frac{\lambda}{n} X_i X_{n+1} x_i = \sqrt{\frac{\lambda}{n}} \sum_{i=1}^{n} x_i Z_{i,n+1} + \lambda(x \cdot X) X_{n+1},
\]

\[
s(x) = -\frac{\lambda}{2n} \sum_{i=1}^{n} x_i^2 = -\frac{\lambda}{2} x \cdot x,
\]

\[
y(x) = \frac{\lambda}{\sqrt{2n}} \sum_{i=1}^{n} Z''_{i} x_i^2 + \frac{\lambda}{2n} \sum_{i=1}^{n} \left( x_i^2 x_i^2 - \frac{x_i^4}{2} \right) + \frac{\lambda}{n} \sum_{1 \leq i < j \leq n} x_i x_j \left( \bar{Z}_{i,j} + \frac{\lambda}{n} X_i X_j \right) - \frac{\lambda}{2n^2} x_i^2 x_j^2
\]

\[
= \frac{\lambda}{\sqrt{2n}} \sum_{i=1}^{n} Z''_{i} x_i^2 + \frac{\lambda}{n} \sum_{1 \leq i < j \leq n} x_i x_j \bar{Z}_{i,j} + \frac{\lambda}{2} \left( x \cdot X \right)^2 - \frac{1}{2} (x \cdot x)^2,
\]

where \(Z''_{i} \overset{i.i.d.}{\sim} N(0, 1)\) independently of any other random variables. Define now

\[
A_{n,\epsilon} = \mathbb{E} \log \left( \sum_{\sigma \in S} P_0(\sigma) \exp(\bar{\sigma}^2 z(x)) \right)_{n,\epsilon} - \mathbb{E} \log \left( \exp(y(x)) \right)_{n,\epsilon}.
\]

Using Gaussian interpolation techniques, it is not difficult to show that \(\int_0^\epsilon d\epsilon (A_{n,\epsilon} - A_{n,\epsilon}^{(0)}) \xrightarrow{n \to \infty} 0\) because the modifications made in \(z_0, s_0\) and \(y_0\) are of negligible order. Using (3.24) we conclude

\[
\limsup_{n \to \infty} \int_0^\epsilon d\epsilon F_{n,\epsilon} \leq \limsup_{n \to \infty} \int_0^\epsilon d\epsilon A_{n,\epsilon}.
\]

### 3.4.4 Overlap concentration

Proposition 7 implies that the overlap between two replicas, i.e. two independent samples \(x^{(1)}\) and \(x^{(2)}\) from the Gibbs distribution \(\langle \cdot \rangle_{n,\epsilon}\), concentrates. Let us define the random variables

\[
Q = \langle \frac{1}{n} \sum_{i=1}^{n} x_i^{(1)} x_i^{(2)} \rangle_{n,\epsilon} \quad \text{and} \quad b_i = \langle x_i \rangle_{n,\epsilon}.
\]

Notice that \(Q = \frac{1}{n} \sum_i b_i^2 \geq 0\). By Proposition 7 we know that

\[
\int_0^\epsilon d\epsilon \mathbb{E} \langle (x^{(1)} \cdot x^{(2)} - Q)^2 \rangle_{n,\epsilon} \xrightarrow{n \to \infty} 0.
\]

Thus, using the Nishimori property (Proposition 3) we deduce:

\[
\int_0^\epsilon d\epsilon \mathbb{E} \langle (x \cdot x - Q)^2 \rangle_{n,\epsilon} \xrightarrow{n \to \infty} 0 \quad \text{and} \quad \int_0^\epsilon d\epsilon \mathbb{E} \langle (x \cdot b - Q)^2 \rangle_{n,\epsilon} \xrightarrow{n \to \infty} 0.
\]
3.4.5 The main estimate

Let us denote, for \( \epsilon \in [0, 1] \),

\[
\mathcal{F}_\epsilon : (\lambda, q) \mapsto -\frac{\lambda}{4} q^2 + \epsilon (\mathbb{E} P_0 X^2) \frac{\lambda q}{2} + (1 - \epsilon) \mathbb{E} \left[ \log \sum_{x \in S} P_0(x) \exp \left( \sqrt{\lambda q} Z x + \lambda q x X - \frac{\lambda}{2} q x^2 \right) \right]
\]

where the expectation \( \mathbb{E} \) is taken with respect to the independent random variables \( X \sim P_0 \) and \( Z \sim \mathcal{N}(0, 1) \). The following proposition is one of the key steps of the proof.

**Proposition 12**

For all \( \epsilon_0 \in [0, 1] \),

\[
\int_{0}^{\epsilon_0} d \epsilon \left( A_{n, \epsilon} - \mathbb{E} \left[ \mathcal{F}_\epsilon (\lambda, Q) \right] \right) \xrightarrow{n \to \infty} 0.
\]

The proof of Proposition 12 is reported to Section 3.4.6. We deduce here Theorem 2 from Proposition 12 and the results of the previous sections. Because of Proposition 11, we only have to show that

\[
\limsup_{n \to \infty} F_n \leq \sup_{q \geq 0} \mathcal{F}(\lambda, q).
\]

By Proposition 6 we have

\[
\epsilon_0 F_n \leq \int_{0}^{\epsilon_0} d \epsilon F_{n, \epsilon} + \frac{1}{2} H(P_0) \epsilon_0^2.
\]

Therefore by equation (3.26) and Proposition 12

\[
\epsilon_0 \limsup_{n \to \infty} F_n \leq \limsup_{n \to \infty} \int_{0}^{\epsilon_0} d \epsilon A_{n, \epsilon} + \frac{1}{2} H(P_0) \epsilon_0^2 \leq \limsup_{n \to \infty} \int_{0}^{\epsilon_0} d \epsilon \mathbb{E} \mathcal{F}_\epsilon (\lambda, Q) + \frac{1}{2} H(P_0) \epsilon_0^2. \tag{3.29}
\]

It remains then to show that

\[
\limsup_{n \to \infty} \int d \epsilon \mathbb{E} \mathcal{F}_\epsilon (\lambda, Q) \leq \epsilon_0 \sup_{q \geq 0} \mathcal{F}(\lambda, q) + O(\epsilon_0^2). \]

We have for \( \epsilon \in [0, 1] \),

\[
\sup_{q \in [0, K^2]} |\mathcal{F}_\epsilon (\lambda, q) - \mathcal{F}(\lambda, q)| \leq \epsilon \sup_{q \in [0, K^2]} \left\{ \frac{\lambda q}{2} \mathbb{E} P_0 [X^2] + \mathbb{E} \log \sum_{x \in S} P_0(x) \exp(\sqrt{\lambda q} Z x + \lambda q x X - \frac{\lambda}{2} q x^2) \right\}
\]

\[
\leq C \epsilon,
\]

for some constant \( C \) that only depends on \( \lambda \) and \( P_0 \). Noticing that \( Q \in [0, K^2] \) a.s., we have then

\[
|\mathbb{E} \mathcal{F}_\epsilon (\lambda, Q) - \mathbb{E} \mathcal{F}(\lambda, Q)| \leq C \epsilon_0,
\]

for all \( \epsilon \in [0, \epsilon_0] \) and therefore

\[
\int_{0}^{\epsilon_0} d \epsilon \mathbb{E} \mathcal{F}_\epsilon (\lambda, Q) \leq \epsilon_0 \sup_{q \geq 0} \mathcal{F}(\lambda, q) + \frac{1}{2} C \epsilon_0^2.
\]

Combined with (3.29), this implies

\[
\limsup_{n \to \infty} F_n \leq \sup_{q \geq 0} \mathcal{F}(\lambda, q) + \frac{1}{2} H(P_0) \epsilon_0 + \frac{1}{2} C \epsilon_0, \text{ for all } \epsilon_0 \in (0, 1].
\]

Theorem 2 is proved.

3.4.6 Proof of Proposition 12

In this section, we prove Proposition 12 which is a consequence of Lemmas 5 and 6 below. In order to lighten the formulas, we will use the following notations

\[
X' = X_{n+1} \quad \text{and} \quad Z_i' = Z_{i,n+1}.
\]

Recall

\[
A_{n, \epsilon} = \mathbb{E} \log \left( \sum_{\sigma \in S} P_0(\sigma) \exp(\bar{\sigma} z(x) + \bar{\sigma}^2 s(x)) \right)_{n, \epsilon} - \mathbb{E} \log \left( \exp(y(x)) \right)_{n, \epsilon}, \tag{3.30}
\]

where for \( \sigma \in S, \bar{\sigma} = (1 - L_{n+1}) \sigma + L_{n+1} X' \). The computations here are closely related to the cavity computations in the SK model, see for instance [63].
The next lemma follows from the simple fact that for $\lambda Z_0 \sim \mathcal{N}(0, 1)$ is independent of all other random variables.

**Lemma 6**

\[
\int_0^\infty d\epsilon \left| \mathbb{E} \log \left( \sum_{\sigma \in S} P_0(\sigma) \exp(\sigma z(x) + \sigma^2 s(x)) \right) \right|_{n, \epsilon} - \left( \epsilon (\mathbb{E} P_0 X^2) \mathbb{E} \frac{\lambda Q}{2} + (1 - \epsilon) \mathbb{E} \log \sum_{\sigma \in S} P_0(\sigma) \exp \left( \sqrt{\lambda Q} \sigma Z_0 + \lambda Q \sigma X' - \frac{\lambda Q^2}{2} Q \right) \right) \rightarrow 0, \\
\] where $Z_0 \sim \mathcal{N}(0, 1)$ is independent of all other random variables.

We will only prove Lemma 5 here since Lemma 6 follows from the same kind of arguments (the complete proof can be found in [42]). The remaining of the section is thus devoted to the proof of Lemma 5.

Let us write $f(z, s) = \sum_{\sigma \in S} P_0(\sigma) e^{\sigma z + \sigma^2 s}$ and we define:

\[
U = \langle f(z(x), s(x)) \rangle_{n, \epsilon}, \\
V = \sum_{\sigma \in S} P_0(\sigma) \exp \left( \sigma \sqrt{\frac{1}{n} \sum_{i=1}^n b_i Z_i' + \lambda Q X' - \frac{\lambda Q}{2} \sigma^2} \right).
\]

**Lemma 7**

\[
\int_0^\infty d\epsilon \mathbb{E} [(U - V)^2] \rightarrow 0.
\]

**Proof.** It suffices to show that $\int d\epsilon |\mathbb{E} U^2 - \mathbb{E} V^2| \rightarrow 0$ and $\int d\epsilon |\mathbb{E} UV - \mathbb{E} V^2| \rightarrow 0$. Let $\mathbb{E}_{Z'}$ denote the expectation with respect to $Z' = (Z_{i, n+1})_{1 \leq i \leq n}$ only. Compute

\[
\mathbb{E}_{Z'} V^2 = \mathbb{E}_{Z'} \sum_{\sigma_1, \sigma_2 \in S} P_0(\sigma_1, \sigma_2) \exp \left( (\sigma_1 + \sigma_2) \sqrt{\frac{1}{n} \sum_{i=1}^n b_i Z_i' + \lambda Q X' (\sigma_1 + \sigma_2) - \frac{\lambda Q}{2} (\sigma_1^2 + \sigma_2^2)} \right)
\]

\[
= \sum_{\sigma_1, \sigma_2 \in S} P_0(\sigma_1, \sigma_2) \exp \left( (\sigma_1 + \sigma_2)^2 \frac{\lambda Q}{2} + \lambda Q X' (\sigma_1 + \sigma_2) - \frac{\lambda Q}{2} (\sigma_1^2 + \sigma_2^2) \right)
\]

\[
= \sum_{\sigma_1, \sigma_2 \in S} P_0(\sigma_1, \sigma_2) \exp \left( \sigma_1 \sigma_2 \lambda Q + \lambda Q X' (\sigma_1 + \sigma_2) \right)
\]

(3.31)

where we write for $i = 1, 2$, $\sigma_i = (1 - L_{n+1}) \sigma_i + L_{n+1} X'$, as before.

Let us show that $\int d\epsilon |\mathbb{E} U^2 - \mathbb{E} V^2| \rightarrow 0$.

\[
\mathbb{E}_{Z'} U^2 = \mathbb{E}_{Z'} \langle f(z(x), s(x)) \rangle_{n, \epsilon}^2
\]

\[
= \mathbb{E}_{Z'} \langle f(z(x^{(1)}), s(x^{(1)})) f(z(x^{(2)}), s(x^{(2)})) \rangle_{n, \epsilon} (x^{(1)}$ and $x^{(2)}$ are indep. samples from $\langle \cdot \rangle_{n, \epsilon})
\]

\[
= \left( \mathbb{E}_{Z'} f(z(x^{(1)}), s(x^{(1)})) f(z(x^{(2)}), s(x^{(2)})) \right)_{n, \epsilon}
\]

\[
= \left( \sum_{\sigma_1, \sigma_2 \in S} P_0(\sigma_1, \sigma_2) \mathbb{E}_{Z'} \exp \left( \sigma_1 z(x^{(1)}) + \sigma_1^2 s(x^{(1)}) + \sigma_2 z(x^{(2)}) + \sigma_2^2 s(x^{(2)}) \right) \right)_{n, \epsilon}.
\]

The next lemma follows from the simple fact that for $N \sim \mathcal{N}(0, 1)$ and $t \in \mathbb{R}$, $e^{tN} = \exp\left(\frac{t^2}{2}\right)$. 

31
Lemma 8

Let \( x^{(1)}, x^{(2)} \in S^n \) and \( \sigma_1, \sigma_2 \in S \) be fixed. Then

\[
\mathbb{E}_Z \exp \left( \sigma_1 \frac{1}{n} \sum_{i=1}^{n} x_i^{(1)} Z_i' + \sigma_2 \frac{1}{n} \sum_{i=1}^{n} x_i^{(2)} Z_i' \right) = \exp \left( \lambda \sigma_1 \sigma_2 \mathbf{x}^{(1)} \cdot \mathbf{x}^{(2)} + \frac{\lambda \sigma_1^2}{2n} \| \mathbf{x}^{(1)} \|^2 + \frac{\lambda \sigma_2^2}{2n} \| \mathbf{x}^{(2)} \|^2 \right).
\]

Thus, for all \( x^{(1)}, x^{(2)} \in S^n \) and \( \sigma_1, \sigma_2 \in S \)

\[
\mathbb{E}_Z e^{\tilde{a}_1 z(x^{(1)}) + \tilde{a}_2 z(x^{(2)})} = e^{\lambda \tilde{a}_1 \tilde{a}_2 x^{(1)} \cdot x^{(2)} + \lambda X' (\tilde{a}_1 (x^{(1)} \cdot \mathbf{x}) + \tilde{a}_2 (x^{(2)} \cdot \mathbf{x}))},
\]

where we used the fact that \( s(x) = -\frac{\lambda}{2n} \| x \|^2 \) for all \( x \in S^n \). We have therefore

\[
\mathbb{E}_Z U^2 = \left\langle \sum_{\sigma_1, \sigma_2 \in S} P_0(\sigma_1, \sigma_2) \exp \left( \lambda \tilde{a}_1 \tilde{a}_2 x^{(1)} \cdot x^{(2)} + \lambda X' (\tilde{a}_1 (x^{(1)} \cdot \mathbf{x}) + \tilde{a}_2 (x^{(2)} \cdot \mathbf{x})) \right) \right\rangle_{n, \epsilon}.
\]

Define

\[
g : (s, r_1, r_2) \in [-K^2, K^2]^3 \mapsto \sum_{\sigma_1, \sigma_2 \in S} P_0(\sigma_1, \sigma_2) \exp \left( \lambda \tilde{a}_1 \tilde{a}_2 s + \lambda X' (\tilde{a}_1 r_1 + \tilde{a}_2 r_2) \right).
\]

We have \( \mathbb{E}_Z U^2 = \langle g(x^{(1)} \cdot x^{(2)}, x^{(1)} \cdot \mathbf{x}, x^{(2)} \cdot \mathbf{x}) \rangle_{n, \epsilon} \) and by (3.31), \( \mathbb{E}_Z V^2 = g(Q, Q, Q). \)

Lemma 9

There exists a constant \( M \) that only depends on \( \lambda \) and \( K \), such that \( g \) is almost surely \( M \)-Lipschitz.

Proof. \( g \) is a random function that depends only on the random variables \( X' \) and \( L_{n+1} \) (because of \( \tilde{a}_1 \) and \( \tilde{a}_2 \)). \( g \) is \( C^1 \) on the compact \([-K^2, K^2]^3\). An easy computation show that

\[
\forall (s, r_1, r_2) \in [-K^2, K^2]^3, \| \nabla g(s, r_1, r_2) \| \leq 3 \lambda K^4 \exp(3 \lambda K^4).
\]

\( g \) is thus \( M \)-Lipschitz with \( M = 3 \lambda K^4 \exp(3 \lambda K^4) \).

Using Lemma 9 we obtain

\[
\langle |g(x^{(1)} \cdot x^{(2)}, x^{(1)} \cdot \mathbf{x}, x^{(2)} \cdot \mathbf{x}) - g(Q, Q, Q)| \rangle_{n, \epsilon} \leq M \left\langle \sqrt{(x^{(1)} \cdot x^{(2)} - Q)^2 + (x^{(1)} \cdot \mathbf{x} - Q)^2 + (x^{(2)} \cdot \mathbf{x} - Q)^2} \right\rangle_{n, \epsilon}.
\]

We recall relation (3.31) to notice that \( g(Q, Q, Q) = \mathbb{E}_Z V^2 \). Thus, using (3.27) and (3.28)

\[
\int_0^\infty d\epsilon \mathbb{E}_Z U^2 - \mathbb{E}_Z V^2 \leq M \int_0^\infty d\epsilon \mathbb{E} \left\langle \sqrt{(x^{(1)} \cdot x^{(2)} - Q)^2 + (x^{(1)} \cdot \mathbf{x} - Q)^2 + (x^{(2)} \cdot \mathbf{x} - Q)^2} \right\rangle_{n, \epsilon},
\]

and the right-hand side goes to 0 by (3.27-3.28).

Showing that \( \int d\epsilon |\mathbb{E}_Z U^2 - \mathbb{E}_Z V^2| \to 0 \) goes exactly the same way. We thus omit this part here for the sake of brevity, but the reader can refer to [42] where all details are presented.

Using the fact that \( |\log U - \log V| \leq \max(U^{-1}, V^{-1})|U - V| \) and the Cauchy-Schwarz inequality, we have

\[
\mathbb{E} |\log U - \log V| \leq \sqrt{\mathbb{E} U^{-2} + \mathbb{E} V^{-2}} \sqrt{\mathbb{E} (U - V)^2}.
\]
Lemma 10

There exists a constant $C$ that depends only on $\lambda$ and $K$ such that

$$EU^{-2} + EV^{-2} \leq C.$$ 

Proof. Using Jensen inequality, we have $U \geq f(\langle z(x) \rangle_{n,\epsilon}, \langle s(x) \rangle_{n,\epsilon})$. Then

$$U^{-2} \leq f(\langle z(x) \rangle_{n,\epsilon}, \langle s(x) \rangle_{n,\epsilon})^{-2} \leq \sum_{\sigma \in S} P_0(\sigma) \exp \left( -2\bar{\sigma} \langle z(x) \rangle_{n,\epsilon} - 2\bar{\sigma}^2 \langle s(x) \rangle_{n,\epsilon} \right).$$

It remains to bound $E \exp(-2\bar{\sigma} \langle z(x) \rangle_{n,\epsilon} - 2\bar{\sigma}^2 \langle s(x) \rangle_{n,\epsilon})$. $P_0$ has a bounded support, therefore

$$E \exp(-2\bar{\sigma} \langle z(x) \rangle_{n,\epsilon} - 2\bar{\sigma}^2 \langle s(x) \rangle_{n,\epsilon}) \leq C_0 E \exp \left( -2\bar{\sigma} \sum_{i=1}^n \frac{\lambda}{n} (x_i)_{n,\epsilon} Z_i \right) = C_0 E \exp(2\lambda Q^2) \leq C_1,$$

for some constant $C_0, C_1$ depending only on $\lambda$ and $K$. Similar arguments show that $EV^{-2}$ is upper-bounded by a constant.

Using the previous lemma we obtain $\int_0^\epsilon d\epsilon E |\log U - \log V| \to 0$. We now compute $E \log V$ explicitly.

Lemma 11

$$E \log V = \epsilon(E P_0 X^2) E \frac{\lambda Q}{2} + (1 - \epsilon) E \log \sum_{\sigma \in S} P_0(\sigma) \exp \left( \sigma \sqrt{\frac{\lambda}{n} \sum_{i=1}^n b_i Z'_i + \lambda Q X' - \frac{\lambda^2}{2} Q} \right).$$

Proof. It suffices to distinguish the cases $L_{n+1} = 0$ and $L_{n+1} = 1$. If $L_{n+1} = 1$ then for all $\sigma \in S$, $\tilde{\sigma} = X'$ and

$$\log V = \log \left( \exp \left( X' \sqrt{\frac{\lambda}{n} \sum_{i=1}^n b_i Z'_i + \lambda Q X' - \frac{\lambda^2}{2} Q} \right) \right) = X' \sqrt{\frac{\lambda}{n} \sum_{i=1}^n b_i Z'_i + \frac{\lambda Q X'}{2}}.$$

$L_{n+1}$ is independent of all other random variables, thus

$$E \left[ 1(L_{n+1} = 1) \log V \right] = \epsilon(E P_0 X^2) \frac{\lambda}{2} E Q,$$

because the $Z'_i$ are centered, independent from $X'$ and because $X'$ is independent from $Q$. The case $L_{n+1} = 0$ is obvious.

The variables $(b_i)_{1 \leq i \leq n}$ and $(Z'_i)_{1 \leq i \leq n}$ are independent. Recall that $Q = \frac{1}{n} \sum_{i=1}^n b_i^2$. Therefore,

$$\left( Q, \sqrt{\frac{1}{n} \sum_{i=1}^n b_i Z'_i} \right) = \left( Q, \sqrt{Q} Z_0 \right) \text{ in law},$$

where $Z_0 \sim \mathcal{N}(0, 1)$ is independent of all other random variables. The expression of $E \log V$ from Lemma 11 simplifies

$$E \log V = \epsilon(E P_0 X^2) E \frac{\lambda Q}{2} + (1 - \epsilon) E \log \sum_{\sigma \in S} P_0(\sigma) \exp \left( \sqrt{\lambda Q} \sigma Z_0 + \lambda Q X' - \frac{\lambda^2}{2} Q \right),$$

thus

$$\int_0^\epsilon d\epsilon \left| E \log U - \left( \epsilon(E P_0 X^2) E \frac{\lambda Q}{2} + (1 - \epsilon) E \log \sum_{\sigma \in S} P_0(\sigma) \exp \left( \sqrt{\lambda Q} \sigma Z_0 + \lambda Q X' - \frac{\lambda^2}{2} Q \right) \right) \right| \to 0,$$

which is precisely the statement of Lemma 5.
3.4.7 Reduction to distribution with finite support

We will show in this section that it suffices to prove Theorems 2 for input distribution $P_0$ with finite support.

Suppose the Theorem 2 holds for all prior distributions over $\mathbb{R}$ with finite support. Let $P_0$ be a probability distribution that admits a finite second moment: $\mathbb{E}_{P_0} X^2 < \infty$. We are going to approach $P_0$ with distributions with finite supports.

Let $0 < \epsilon \leq 1$. Let $K > 0$ such that $P_0([-K, K]) > 1 - \epsilon^2$. Let $m \in \mathbb{N}$ such that $\frac{K}{m} \leq \epsilon$. For $x \in \mathbb{R}$ we will use the notation

$$\bar{x} = \begin{cases} \frac{x m}{K} & \text{if } x \in [-K, K], \\ 0 & \text{otherwise.} \end{cases}$$

Consequently if $x \in [-K, K]$, $x \leq x < \bar{x} + \frac{K}{m} \leq \bar{x} + \epsilon$. We define $\bar{P}_0$ the image distribution of $P_0$ through the application $x \mapsto \bar{x}$. Let $n \geq 1$. We will note $F_n$, the free energy corresponding to the distribution $\bar{P}_0$ and $\bar{F}$ the function $F$ from (3.4) corresponding to the distribution $\bar{P}_0$. $\bar{P}_0$ has a finite support, we have then by assumptions

$$F_n(\lambda) \xrightarrow{n \to \infty} \sup_{q \geq 0} \bar{F}(\lambda, q) .$$

By construction, there exists a constant $C$ that only depend on $P_0$ such that

$$\mathbb{E} \| (X_i X_j)_{i<j} - (\bar{X}_i \bar{X}_j)_{i<j} \|^2 \leq C \epsilon^2 .$$

Consequently, by “pseudo-Lipschitz” continuity of the free energy with respect to the Wasserstein metric (see Proposition 19 in Appendix A.4) there exists a constant $C > 0$ depending only on $\bar{P}_0$, such that, for all $n \geq 1$ and all $\lambda \geq 0$,

$$|F_n(\lambda) - F_{\bar{P}_0}(\lambda)| \leq \lambda C \epsilon .$$

Lemma 12

| There exists a constant $C' > 0$ that depends only on $P_0$, such that |

| $| \sup_{q \geq 0} \bar{F}(\lambda, q) - \sup_{q \geq 0} \bar{F}(\lambda, q) | \leq \lambda C' \epsilon . $ |

Proof. First notice that both suprema are achieved over a common compact set $[0, \mathbb{E}_{\bar{P}_0}[X^2 + \bar{X}^2]]$.

Indeed, for $q \geq 0$,

$$\frac{\partial}{\partial q} F(\lambda, q) = \lambda \psi \frac{\partial}{\partial q} p_0(\lambda q) - \frac{\lambda q}{2} \leq \frac{\lambda}{2} (\mathbb{E}_{\bar{P}_0}[X^2] - q)$$

because $\psi_{P_0}$ is $\frac{1}{2} \mathbb{E}_{P_0} [X^2]$-Lipschitz by Proposition 5. Consequently, the maximum of $F(\lambda, \cdot)$ is achieved on $[0, \mathbb{E}_{P_0}[X^2]]$ and similarly the supremum of $F(\lambda, \cdot)$ is achieved over $[0, \mathbb{E}_{\bar{P}_0}[X^2]]$. Using Proposition 19 in Appendix A.4, we obtain that there exists a constant $C'$ depending only on $P_0$ such that $\forall q \in [0, \mathbb{E}_{P_0}[X^2 + \bar{X}^2]]$, $|F(\lambda, q) - \bar{F}(\lambda, q)| \leq \lambda C' \epsilon$. The lemma follows.

Combining Equation 3.32 and 3.33 and Lemma 12, we obtain that there exists $n_0 \geq 1$ such that for all $n \geq n_0$,

$$|F_n - \sup_{q \geq 0} F(\lambda, q)| \leq \lambda (C + C' + 1) \epsilon ,$$

where $C$ and $C'$ are two constants that only depend on $P_0$. This proves Theorem 2.
Chapter 4

Non-symmetric low-rank matrix estimation

We consider now the spiked Wishart model (2). Let $P_U$ and $P_V$ be two probability distributions on $\mathbb{R}$ with finite second moment. We assume that $\text{Var}_{P_U}(U), \text{Var}_{P_V}(V) > 0$. Let $n, m \geq 1, \lambda > 0$ and consider $U = (U_1, \ldots, U_n) \overset{\text{i.i.d.}}{\sim} P_U$ and $V = (V_1, \ldots, V_m) \overset{\text{i.i.d.}}{\sim} P_V$, independent from each other.

Suppose that we observe $Y_{i,j} = \sqrt{\lambda/n} U_i V_j + Z_{i,j}$, for $1 \leq i \leq n$ and $1 \leq j \leq m$, (4.1)

where $(Z_{i,j})_{i,j}$ are i.i.d. standard normal random variables, independent from $U$ and $V$. In the following, $\mathbb{E}$ will denote the expectation with respect to the variables $(U, V)$ and $Z$. Let us define the Minimum Mean Squared Error (MMSE) for the estimation of the matrix $UV^\top$ given the observation of the matrix $Y$:

$$\text{MMSE}_n(\lambda) = \min_{\theta} \left\{ \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \mathbb{E} \left[ (U_i V_j - \hat{\theta}_{i,j}(Y))^2 \right] \right\} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \mathbb{E} \left[ (U_i V_j - \mathbb{E}[U_i V_j|Y])^2 \right],$$

where the minimum is taken over all estimators $\hat{\theta}$ (i.e. measurable functions of the observations $Y$). In order to get an upper bound on the MMSE, let us consider the “dummy estimator” $\hat{\theta}_{i,j} = \mathbb{E}[U_i V_j]$ for all $i, j$ which achieves a “dummy” matrix Mean Squared Error of:

$$\text{DMSE} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m \mathbb{E} \left[ (U_i V_j - \mathbb{E}[U_i V_j])^2 \right] = \mathbb{E}[U^2]\mathbb{E}[V^2] - (\mathbb{E}U)^2(\mathbb{E}V)^2.$$

4.1 Information-theoretic limits

We define the Hamiltonian

$$H_n(u, v) = \sum_{n,m} \sqrt{\lambda/n} u_i v_j Z_{i,j} + \frac{\lambda}{n} u_i v_j U_i V_j - \lambda \frac{1}{2n} u_i^2 v_j^2, \quad \text{for } (u, v) \in \mathbb{R}^n \times \mathbb{R}^m. \quad (4.2)$$

The posterior distribution of $(U, V)$ given $Y$ is then

$$dP(u, v | Y) = \frac{1}{Z_n(\lambda)} e^{H_n(u, v)} dP_U^{\otimes n}(u) dP_V^{\otimes m}(v), \quad (4.3)$$

where $Z_n(\lambda) = \int e^{H_n(u, v)} dP_U^{\otimes n}(u) dP_V^{\otimes m}(v)$ is the appropriate normalization. The corresponding free energy is

$$F_n(\lambda) = \frac{1}{n} \mathbb{E} \log Z_n(\lambda) = \frac{1}{n} \mathbb{E} \log \left( \int e^{H_n(u, v)} dP_U^{\otimes n}(u) dP_V^{\otimes m}(v) \right).$$
We consider here the high-dimensional limit where \( n, m \to \infty \), while \( m/n \to \alpha > 0 \). We will be interested in the following fixed point equations, sometimes called “state evolution equations”.

**Definition 3**

We define the set \( \Gamma(\lambda, \alpha) \) as

\[
\Gamma(\lambda, \alpha) = \left\{ (q_u, q_v) \in \mathbb{R}^2_+ \mid q_u = 2\psi_{P_U}(\lambda \alpha q_v) \text{ and } q_v = 2\psi_{P_V}(\lambda q_u) \right\} . \tag{4.4}
\]

First notice that \( \Gamma(\lambda, \alpha) \) is not empty. The function \( f : (q_u, q_v) \mapsto (2\psi_{P_U}(\lambda \alpha q_v), 2\psi_{P_V}(\lambda q_u)) \) is continuous from the convex compact set \([0, \mathbb{E}U^2] \times [0, \mathbb{E}V^2]\) into itself (see Proposition 5). Brouwer’s Theorem gives the existence of a fixed point of \( f \): \( \Gamma(\lambda, \alpha) \neq \emptyset \).

We will express the limit of \( F_n \) using the following function

\[
\mathcal{F} : (\lambda, \alpha, q_u, q_v) \mapsto \psi_{P_U}(\lambda \alpha q_v) + \alpha \psi_{P_V}(\lambda q_u) - \frac{\lambda \alpha}{2} q_u q_v . \tag{4.5}
\]

Recall that \( \psi_{P_U} \) and \( \psi_{P_V} \), defined by (1.6), are the free energies of additive Gaussian scalar channels (1.5) with priors \( P_U \) and \( P_V \). The Replica-Symmetric formula states that the free energy \( F_n \) converges to the supremum of \( \mathcal{F} \) over \( \Gamma(\lambda, \alpha) \).

**Theorem 4 (Replica-Symmetric formula for the spiked Wishart model)**

\[
F_n(\lambda) \xrightarrow{n \to \infty} \sup_{(q_u, q_v) \in \Gamma(\lambda, \alpha)} \mathcal{F}(\lambda, \alpha, q_u, q_v) = \sup_{q_u \geq 0, q_v \geq 0} \inf_{\lambda, \alpha} \mathcal{F}(\lambda, \alpha, q_u, q_v) . \tag{4.6}
\]

Moreover, these extrema are achieved over the same couples \((q_u, q_v) \in \Gamma(\lambda, \alpha)\).

This result from [50] proves a conjecture from [43], in particular \( \mathcal{F} \) corresponds to the “Bethe free energy” ([43], Equation 47). Theorem 4 is proved in Section 4.3. For the rank-\( k \) case (where \( P_U \) and \( P_V \) are probability distributions over \( \mathbb{R}^k \)), see [50]. As before, the Replica-Symmetric formula (Theorem 2) allows also to compute the limit of the MMSE.

**Proposition 13 (Limit of the MMSE)**

Let

\[
D_\alpha = \left\{ \lambda > 0 \mid \mathcal{F}(\lambda, \alpha, \cdot, \cdot) \text{ has a unique maximizer } (q_u^*(\lambda, \alpha), q_v^*(\lambda, \alpha)) \text{ over } \Gamma(\lambda, \alpha) \right\} .
\]

Then \( D_\alpha \) is equal to \((0, +\infty)\) minus a countable set and for all \( \lambda \in D_\alpha \) (and thus almost every \( \lambda > 0 \))

\[
\text{MMSE}_n(\lambda) \xrightarrow{n \to \infty} \mathbb{E}[U^2] \mathbb{E}[V^2] - q_u^*(\lambda, \alpha) q_v^*(\lambda, \alpha) . \tag{4.7}
\]

Again, this was conjectured in [43]: the performance of the Bayes-optimal estimator (i.e. the MMSE) corresponds to the fixed point of the state-evolution equations (4.4) which has the greatest Bethe free energy \( \mathcal{F} \). Proposition 13 follows from the same kind of arguments than Corollary 2 so we omit its proof for the sake of brevity.

Proposition 13 allows to locate the information-theoretic threshold for our matrix estimation problem. Let us define

\[
\lambda_c(\alpha) = \inf \left\{ \lambda \in D_\alpha \mid q_u^*(\lambda, \alpha) q_v^*(\lambda, \alpha) > (\mathbb{E}U)^2 (\mathbb{E}V)^2 \right\} . \tag{4.8}
\]

If the set of the left-hand side is empty, one define \( \lambda_c(\alpha) = 0 \). Proposition 13 gives that \( \lambda_c(\alpha) \) is the information-theoretic threshold for the estimation of \( UV^\top \) given \( Y \).
• If \( \lambda < \lambda_c(\alpha) \), then \( \text{MMSE}_n(\lambda) \rightarrow \text{DMSE} \). It is not possible to reconstruct the signal \( \mathbf{U} \mathbf{V}^\top \) better than a “dummy” estimator.

• If \( \lambda > \lambda_c(\alpha) \), then \( \lim_{n \to \infty} \text{MMSE}_n(\lambda) < \text{DMSE} \). It is possible to reconstruct the signal \( \mathbf{U} \mathbf{V}^\top \) better than a “dummy” estimator.

4.2 Application to the spiked covariance model

Proposition 13 gives us the limit of the MMSE for the estimation of the matrix \( \mathbf{U} \mathbf{V}^\top \), but does not gives us the minimal error for the estimation of \( \mathbf{U} \) or \( \mathbf{V} \) only. As we will see below in the case of the spiked covariance model, one can be interested in estimating \( \mathbf{U} \mathbf{U}^\top \) or \( \mathbf{V} \mathbf{V}^\top \), only. Let us define:

\[
\text{MMSE}_n^{(u)}(\lambda) = \frac{1}{n^2} \mathbb{E} \left[ \sum_{1 \leq i, j \leq n} (U_i U_j - \mathbb{E}[U_i U_j | \mathbf{Y}])^2 \right],
\]

\[
\text{MMSE}_n^{(v)}(\lambda) = \frac{1}{m^2} \mathbb{E} \left[ \sum_{1 \leq i, j \leq m} (V_i V_j - \mathbb{E}[V_i V_j | \mathbf{Y}])^2 \right].
\]

**Theorem 5**

For all \( \alpha > 0 \) and all \( \lambda \in D_\alpha \)

\[
\text{MMSE}_n^{(u)}(\lambda) \stackrel{n \to \infty}{\longrightarrow} \mathbb{E}_{P_U}[U^2]^2 - q^*_u(\lambda, \alpha)^2 \quad \text{and} \quad \text{MMSE}_n^{(v)}(\lambda) \stackrel{n \to \infty}{\longrightarrow} \mathbb{E}_{P_V}[V^2]^2 - q^*_v(\lambda, \alpha)^2.
\]

Theorem 5 is proved in Section 4.4. Let us consider now the so-called spiked covariance model. Let \( \mathbf{U} = (U_1, \ldots, U_n) \overset{i.i.d.}{\sim} P_U \), where \( P_U \) is a distribution over \( \mathbb{R} \) with finite second moment. Define the “spiked covariance matrix”

\[
\Sigma = \mathbf{I}_n + \frac{\lambda}{n} \mathbf{UU}^\top,
\]

(4.9)

and suppose that we observe \( \mathbf{Y}_1, \ldots, \mathbf{Y}_m \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \Sigma) \). Given the matrix \( \mathbf{Y} = (\mathbf{Y}_1 | \cdots | \mathbf{Y}_m) \), one would like to estimate the “spike” \( \mathbf{UU}^\top \). We deduce from Theorem 5 above the minimal mean squared error for this task, in the asymptotic regime where \( n, m \to +\infty \) and \( m/n \to \alpha > 0 \).

**Corollary 5**

For all \( \alpha > 0 \), the function

\[
q \mapsto \left\{ q_{P_U}(\lambda, \alpha q) + \frac{\alpha}{2} (q + \log(1 - q)) \right\}
\]

admits for almost all \( \lambda > 0 \) a unique maximizer \( q^*(\lambda, \alpha) \) on \( [0, 1) \) and

\[
\text{MMSE}_n^{(u)}(\lambda) = \frac{1}{n^2} \mathbb{E} \left[ \| \mathbf{U} \mathbf{U}^\top - \mathbb{E}[\mathbf{U} \mathbf{U}^\top | \mathbf{Y}] \|^2 \right] \stackrel{n \to \infty}{\longrightarrow} \mathbb{E}_{P_U}[U^2]^2 - \left( q^*(\lambda, \alpha) \right)^2.
\]

**Proof.** There exists independent Gaussian random variables \( \mathbf{V} = (V_1, \ldots, V_m) \overset{i.i.d.}{\sim} \mathcal{N}(0, 1) \) and \( Z_{i,j} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1) \), independent from \( \mathbf{U} \) such that

\[
\mathbf{Y} = (\mathbf{Y}_1 | \cdots | \mathbf{Y}_m) = \sqrt{\frac{\lambda}{n}} \mathbf{U} \mathbf{V}^\top + \mathbf{Z}.
\]

Therefore, the limit of the MMSE for the estimation of \( \mathbf{U} \mathbf{V}^\top \) is given by Theorem 5 above. It remains only to evaluate the formulas of Theorems 4 and 5 in the case \( P_V = \mathcal{N}(0, 1) \). As computed at the end of
Section 1.3, $\psi_{N(0,1)}(q) = \frac{1}{2}(q - \log(1 + q))$. Thus, the limit of the free energy $\psi_{u}$ becomes:

$$
\sup_{q_v \in [0,1]} \left\{ \psi_{P_U}(\lambda \alpha q_v) + \frac{\alpha}{2} (q_v + \log(1 - q_v)) \right\}.
$$

By Theorem 5 for all $\alpha > 0$ and almost all $\lambda > 0$ this supremum admits a unique maximizer $q_u^*(\lambda, \alpha)$ and $\text{MMSE}^{(u)}_{n}(\lambda) \to \mathbb{E}_{P_U}[U^2] - q_u^2(\lambda, \alpha)^2$ where

$$
q_u^*(\lambda, \alpha) = 2 \psi_{N(0,1)}(\lambda q_u^*(\lambda, \alpha)) = \frac{\lambda q_u^*(\lambda, \alpha)}{1 + \lambda q_u^*(\lambda, \alpha)}.
$$

We deduce from the equation above that $q_u^*(\lambda, \alpha) = \frac{q_u^2(\lambda, \alpha)}{\lambda (1 + q_u^2(\lambda, \alpha))}$, which concludes the proof.

We will now compare the MMSE given by Corollary 5 of the mean squared errors achieved by PCA and Approximate Message Passing (AMP). We consider a case where the signal is sparse:

$$
P_U = s\mathcal{N}(0, 1/s) + (1 - s)\delta_0,
$$

for some $s \in (0, 1]$, so that $\mathbb{E}_{P_U}[U^2] = 1$.

![Figure 4.1: Mean Squared Errors for the spiked covariance model, where the spike is generated by (4.10) with $s = 0.15$, $\lambda = 1$. The right-hand side panel is a zoom of the left-hand side panel around $\alpha = 1$.](image)

Let $\tilde{u}$ be a singular vector of $Y/\sqrt{n}$ associated with $\sigma_1$, the top singular value of $Y/\sqrt{n}$, such that $\|\tilde{u}\| = \sqrt{n}$. Then results from [15, 28] give that almost surely:

$$
\lim_{n \to \infty} (\tilde{u} \cdot U)^2 = \begin{cases} 
\frac{\lambda^2 \alpha - 1}{\lambda (\lambda \alpha + 1)} & \text{if } \lambda^2 \alpha \geq 1, \\
0 & \text{otherwise,}
\end{cases}
$$

and

$$
\lim_{n \to \infty} \sigma_1 = \begin{cases} 
\sqrt{\frac{1 + \lambda}{\lambda (\alpha^{-1} + 1)}} & \text{if } \lambda^2 \alpha \geq 1, \\
1 + 1/\sqrt{\alpha} & \text{otherwise.}
\end{cases}
$$

We are then going to estimate $UU^\top$ using $\hat{\theta}^{\text{PCA}} = \hat{\delta} \hat{\mu}^\top$, where $\hat{\delta}$ is chosen in order to minimize the mean squared error. The optimal choice of $\hat{\delta}$ is $\delta^* = \frac{2}{\lambda (\lambda \alpha + 1)}$, note that $\delta^*$ can be estimated using $\sigma_1$.

We obtain the mean squared error of the spectral estimator $\hat{\theta}^{\text{PCA}}$:

$$
\lim_{n \to \infty} \text{MSE}^{\text{PCA}}_{n} = \begin{cases} 
\frac{1 + \lambda}{\lambda (\lambda \alpha + 1)} \left(2 - \frac{1 + \lambda}{\lambda (\lambda \alpha + 1)}\right) & \text{if } \lambda^2 \alpha \geq 1, \\
1 & \text{otherwise.}
\end{cases}
$$

As in the symmetric case (see Section 3.2.1) one can define an Approximate Message Passing (AMP) algorithm to estimate $UU^\top$. For a precise description of the algorithm, see [61, 26, 44]. The MSE achieved by AMP after $t$ iterations is:

$$
\lim_{n \to \infty} \text{MSE}^{\text{AMP}}_{n} = 1 - (q_u^t)^2,
$$

38
where \( q^t_u \) is given by the recursion:

\[
\begin{align*}
q^{t+1}_u &= 2\psi'_P(\lambda\alpha q^t_u) \\
q^{t+1}_v &= 2\psi'_P(\lambda q^t_v),
\end{align*}
\]

(4.11)

with initialization \((q^0_u, q^0_v) = (0, 0)\). If \((q^t_u, q^t_v) \to (q^*_u(\lambda, \alpha), q^*_v(\lambda, \alpha))\), then AMP is information-theoretically optimal.

Let us now comment the plots of Figure 4.1. \( \lambda = 1 \) so the “spectral threshold” (the minimal value of \( \alpha \) for which PCA performs better than a random guess) it at \( \alpha = 1 \) (green dashed line). This threshold corresponds also to the threshold for AMP: \( \text{MSE}^{\text{AMP}} = 1 \) for \( \alpha < 1 \) while \( \text{MSE}^{\text{AMP}} < 1 \) for \( \alpha > 1 \). The information-theoretic threshold \( \alpha_{\text{IT}} \) is however strictly less than 1. For \( \alpha \in (\alpha_{\text{IT}}, 1) \) inference is “hard”: it is information-theoretically possible to achieve a MSE strictly less than 1, but PCA and AMP fail (and it is conjectured that any polynomial-time algorithm will also fail).

However, even for \( \alpha > 1 \), AMP does not always succeed to reach the MMSE. For \( \alpha \in (1, \alpha_{\text{Alg}}) \), \( \text{MSE}^{\text{AMP}} \) is strictly less than 1 but is still very bad. So, the region \( \alpha \in (1, \alpha_{\text{Alg}}) \) is also a “hard region” in the sense that achieving the MMSE seems impossible for polynomial-time algorithms (under the conjecture that AMP is optimal among polynomial-time algorithms). The scenario presented on Figure 4.1 is not the only one possible: various cases have been studied in great details in [45]. See in particular Figure 6 from [45] and the phase diagrams of Figure 7 and 8.

### 4.3 Proof of the Replica-Symmetric formula (Theorem 4)

#### 4.3.1 Proof ideas

The proof of the Replica formula for the non-symmetric case is a little bit more involved compared to the symmetric case, because one can not use the convexity argument of Proposition 11 to obtain the lower bound. Indeed, a key step in the proof of Proposition 11 was the inequality (3.21) that was obtained by saying that

\[
\mathbb{E}\langle (x \cdot X - q)^2 \rangle \geq 0,
\]

(4.12)

for every \( q \geq 0 \) (we omit the notation’s details here in order to focus on the main ideas). However, if one follows the strategy of Proposition 11 in the non-symmetric case, one obtain

\[
\mathbb{E}\langle (u \cdot U - q_u)(v \cdot V - q_v) \rangle
\]

(4.13)

instead of (4.12). Now, it not obvious anymore that (4.13) is non-negative. In order to prove it, one has to investigate further the distributions of the overlaps \( u \cdot U \) and \( v \cdot V \) under the posterior distribution \( \langle \cdot \rangle \). By following the approach used by Talagrand in [63] to prove the TAP equations (discovered by Thouless, Anderson and Palmer in [65]) for the Sherrington-Kirkpatrick model, one can show that the overlaps asymptotically approximately satisfy

\[
\begin{align*}
u \cdot V &\simeq 2\psi'_P(\lambda v \cdot V) \\
u \cdot U &\simeq 2\psi'_P(\lambda u \cdot U).
\end{align*}
\]

This is precisely the fixed point equations verified by \((q_u, q_v) \in \Gamma(\lambda, \alpha)\). Thus one has

\[
\mathbb{E}\langle (u \cdot U - q_u)(v \cdot V - q_v) \rangle \simeq \mathbb{E}\langle 2\psi'_P(\lambda u \cdot V) - 2\psi'_P(\lambda v \cdot U) \rangle \geq 0,
\]

(4.14)

because by Proposition 5, \( \psi'_P \) is non-decreasing. One obtain thus the analog of the lower-bound of Proposition 11 for the non-symmetric case. The converse upper-bound is proved following the Aizenman-Sims-Starr scheme, as in the symmetric case.

In the following sections we will not, however, follow the proof strategy that we just described. This was done in [50]. We will instead provide a more straightforward proof from [12] that uses an evolution of Guerra’s interpolation technique, see [11].
4.3.2 Interpolating inference model

We prove Theorem 4 in this section. First, notice that it suffices to prove Theorem 4 for $\lambda = 1$, because the dependency in $\lambda$ can be "incorporated" in the prior $P_U$. We will thus consider in this section that $\lambda = 1$ and consequently alleviate the notations by removing the dependencies in $\lambda$. Second, it suffices to prove that

$$F_n \xrightarrow{n \to \infty} \sup_{q_v \geq 0, n \geq 0} \inf F(\alpha, q_u, q_v)$$

(4.15)

because the equality with $\sup_{(q_u, q_v) \in \Gamma(\lambda, \alpha)} F(\alpha, q_u, q_v)$ follows then from simple convex analysis arguments (Proposition 17) presented in Section 4.3.5.

Third, by a straightforward adaptation of the approximation argument of Section 3.4.7 to the non-symmetric case, it suffices now to prove (4.15) in the case where the priors $P_U$ and $P_V$ have a bounded support. We suppose now that the above conditions are verified and we will prove that (4.15) holds.

Let $q_v \geq 0$ and $q_u : [0, 1] \to \mathbb{R}_+$ be a differentiable function such that $q_u(0) = 0$. For $0 \leq t \leq 1$ we consider the following observation channel

$$
\begin{align*}
Y_t &= \sqrt{(1 - t)/n} U V^\top + Z \\
Y_t^{(u)} &= \sqrt{t \alpha q_u} U + Z^{(u)} \\
Y_t^{(v)} &= \sqrt{q_v(t)} V + Z^{(v)},
\end{align*}
$$

(4.16)

where $Z_i^{(u)}, Z_j^{(v)} \overset{i.i.d.}{\sim} N(0, 1)$, are independent from everything else. The observation channel (4.16) interpolates between the initial matrix estimation problem (2) ($t = 0$), and two decoupled inference channels on $U$ and $V$ ($t = 1$). We define the interpolating Hamiltonian as:

$$H_{n,t}(u, v) = \sum_{i,j} \sqrt{\frac{(1 - t)}{n}} u_i v_j Z_{i,j} + \frac{(1 - t)}{n} u_i v_j U_j - \frac{(1 - t)}{2n} u_i^2 v_j^2 + \sum_{i=1}^{n} \sqrt{t \alpha q_u} u_i Z_i^{(u)} + t \alpha q_u u_i U_i - \frac{t \alpha q_u}{2} u_i^2 + \sum_{j=1}^{m} \sqrt{q_v(t)} v_j Z_j^{(u)} + q_v(t) v_j V_j - \frac{1}{2} q_v(t) v_j^2.
$$

The posterior distribution of $(U, V)$ given $(Y_t, Y_t^{(u)}, Y_t^{(v)})$ is then

$$dP(u, v \mid Y_t, Y_t^{(u)}, Y_t^{(v)}) = \frac{1}{Z_{n,t}} e^{H_{n,t}(u, v)} dP_U^n(u) dP_V^m(v),$$

(4.17)

where $Z_{n,t}$ is the appropriate normalization. The Gibbs measure $(\cdot)_{n,t}$ (which denotes the expectation with respect to samples $(u, v)$ from the posterior (4.17)) is defined as

$$\langle f(u, v) \rangle_{n,t} = \frac{1}{Z_{n,t}} \int f(u, v) e^{H_{n,t}(u, v)} dP_U^n(u) dP_V^m(v),$$

(4.18)

for all function $f$ such that the right-hand side is well defined. The interpolating free energy is then

$$F_{n,t} = \frac{1}{n} \mathbb{E} \log Z_{n,t} = \frac{1}{n} \mathbb{E} \log \left( \int e^{H_{n,t}(u, v)} dP_U^n(u) dP_V^m(v) \right).$$

(4.19)

Notice that

$$
\begin{align*}
F_{n,0} &= F_n \\
F_{n,1} &= \psi_{P_U}(\alpha q_u) + \frac{m}{n} \psi_{P_V}(q_u(1)).
\end{align*}
$$

(4.20)

$F_{n,1}$ looks similar to the limiting expression $F$ defined by (4.5). We would therefore like to compare $F_{n,1}$ and $F_n = F_{n,0}$. We thus compute the derivative with respect to $t$:
Lemma 13

For all \( t \in (0, 1) \),

\[
\frac{\partial}{\partial t} F_{n,t} = \frac{\alpha}{2} q'_u(t) q_v - \frac{1}{2} \mathbb{E} \left\langle \left( u \cdot U - q'_u(t) \right) \left( \frac{m}{n} v \cdot V - \alpha q_v \right) \right\rangle_{n,t}.
\]  

(4.21)

Proof. Let \( t \in (0, 1) \). Compute

\[
\frac{\partial}{\partial t} F_{n,t} = \frac{1}{n} \mathbb{E} \left\langle \frac{\partial}{\partial t} H_{n,t}(u, v) \right\rangle_{n,t}.
\]

Using Gaussian integration by parts and the Nishimori property (Proposition 3) as in the proof of Proposition 4, one obtain:

\[
\frac{1}{n} \mathbb{E} \left\langle \frac{\partial}{\partial t} H_{n,t}(u, v) \right\rangle_{n,t} = \frac{1}{2} \alpha q_u \mathbb{E} \left\langle u \cdot U \right\rangle_{n,t} + \frac{1}{2} q'_u(t) \mathbb{E} \left\langle \frac{m}{n} v \cdot V \right\rangle_{n,t} - \frac{1}{2} \mathbb{E} \left\langle (u \cdot U)(\frac{m}{n} v \cdot V) \right\rangle_{n,t},
\]

which leads to (4.21).

Our goal now is to show that the expectation of the Gibbs measure in (4.21) vanishes. If this is the case, the relation \( F_n = F_{n,0} = F_{n,1} - \int_0^1 \frac{\partial}{\partial t} F_{n,t} dt \) would give us almost the formula that we want to prove. The arguments to show that can be summarized as follows:

• First, we show that the overlap \( u \cdot U \) concentrates around its mean \( \mathbb{E}(u \cdot U)_{n,t} \).

• Then, we chose \( q_u \) to be solution to the differential equation \( q'_u(t) = \mathbb{E}(u \cdot U)_{n,t} \) in order to cancel the Gibbs average in (4.21).

4.3.3 Overlap concentration

By Section 2.2, we know that adding some side information about the variables we want to infer forces the corresponding overlaps to concentrate. More precisely, suppose that we observe, in addition to the observations (4.16):

\[
\begin{align*}
(Y_{\text{pert}}^{(u)} &= a_u \sqrt{s_n} U + Z_{\text{pert}}^{(u)}) \\
(Y_{\text{pert}}^{(v)} &= a_v \sqrt{s_n} V + Z_{\text{pert}}^{(v)})
\end{align*}
\]

(4.22)

where \( a_u, a_v \in [1, 2], Z_{\text{pert}}^{(u)} = (Z_{\text{pert},1}^{(u)}, \ldots, Z_{\text{pert},n}^{(u)}) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1), Z_{\text{pert}}^{(v)} = (Z_{\text{pert},1}^{(v)}, \ldots, Z_{\text{pert},m}^{(v)}) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1) \) are independent random variables, independently of everything else.

These new observations induce additional terms in the Hamiltonian \( H_{n,t} \) which becomes:

\[
H_{n,t}^{\text{pert}}(u, v) = H_{n,t}(u, v) + \sum_{i=1}^n a_u \sqrt{s_n} Z_{\text{pert},i}^{(u)} u_i + a_u^2 s_n U_i u_i - \frac{a_u^2 s_n}{2} u_i^2
\]
\[
+ \sum_{j=1}^m a_v \sqrt{s_n} Z_{\text{pert},j}^{(v)} v_j + a_v^2 s_n V_j v_j - \frac{a_v^2 s_n}{2} v_j^2.
\]

Again, one can associate to this perturbed Hamiltonian a perturbed free energy \( F_{n,t}^{\text{pert}} \) and a perturbed Gibbs measure \( \langle \cdot \rangle_{n,t}^{\text{pert}} \). Define, similarly as in Section 2.2

\[
\phi_t : (a_u, a_v) \mapsto \frac{1}{n s_n} \log \left( \int_{u,v} dP_{U}^{\otimes n}(u) dP_{V}^{\otimes m}(v) e^{H_{n,t}^{\text{pert}}(u,v)} \right)
\]

and \( v_n(s_n) = \sup_{t \in [0,1]} \sup_{1/2 \leq a_u, a_v \leq 3} \mathbb{E} |\phi_t(a_u, a_v) - \mathbb{E} \phi_t(a_u, a_v)| \). A straightforward extension of Lemma 2 and Proposition 8 gives
Proposition 14

Suppose that \( s_n \xrightarrow{n \to \infty} 0 \). Then

\[
\left| F_{n,t} - \int_1^2 \int_1^2 F^\text{pert}_{n,t} \, da_u da_v \right| \xrightarrow{n \to \infty} 0 ,
\]  

uniformly in \( t \in [0, 1] \). Moreover, if \( \begin{cases} 
  v_n(s_n) \xrightarrow{n \to \infty} 0 \\
  ns_n \xrightarrow{n \to \infty} +\infty 
\end{cases} \), then we have uniformly in \( t \in [0, 1] \)

\[
\int_1^2 \int_1^2 \mathbb{E} \left( (u^{(1)} \cdot u^{(2)} - \mathbb{E}(u^{(1)} \cdot u^{(2)})_{n,t})^2 \right)_{n,t} \, da_u da_v \xrightarrow{n \to \infty} 0 ,
\]

\[
\int_1^2 \int_1^2 \mathbb{E} \left( (v^{(1)} \cdot v^{(2)} - \mathbb{E}(v^{(1)} \cdot v^{(2)})_{n,t})^2 \right)_{n,t} \, da_u da_v \xrightarrow{n \to \infty} 0 .
\]

The next lemma states that the hypotheses of Proposition 14 are verified for \( s_n = n^{-1/4} \):

Lemma 14

\[ v_n(s_n) = O(n^{-1/2} s_n^{-1}) . \]

We delay the proof of Lemma 14 to Section 4.3.5.

For simplicity, we will now assume that (4.24) and (4.25) remains true without perturbation, i.e.

\[
\mathbb{E} \left( (u^{(1)} \cdot u^{(2)} - \mathbb{E}(u^{(1)} \cdot u^{(2)})_{n,t})^2 \right)_{n,t} \xrightarrow{n \to \infty} 0 ,
\]

\[
\mathbb{E} \left( (v^{(1)} \cdot v^{(2)} - \mathbb{E}(v^{(1)} \cdot v^{(2)})_{n,t})^2 \right)_{n,t} \xrightarrow{n \to \infty} 0 .
\]

uniformly in \( t \in [0, 1] \). We know by (4.23) that considering averages over small perturbations (4.22) does not affect the limiting free energy. For simplicity, we will in the following assume (4.26) to hold, since the presence of the perturbation terms does not make any change in the proof.

4.3.4 Lower and upper bounds

Now that we know that we can assume that (4.26) holds, we can go back to our computations of the derivative of the free energy (4.21) along the interpolation path. One can verify easily that the mapping

\[
(q_u(t), t, q_v) \mapsto \mathbb{E}(u \cdot U)_{n,t} \in [0, \mathbb{E}_{P_U}[U^2]]
\]

is continuous and is Lipschitz with respect to its first coordinate. Consequently the equation

\[
q'_u(t) = \mathbb{E}(u \cdot U)_{n,t}
\]

is an order one differential equation on \( q_u \), parametrized by \( q_v \) to which we can apply the parametric Cauchy-Lipschitz Theorem. There exists a (unique) function \( q_{u,n}(\cdot, q_v) : [0, 1] \to [0, \mathbb{E}_{P_U}[U^2]] \) that takes the value 0 at 0 and is solution of (4.28). Moreover the mapping

\[
Q_n : \mathbb{R}_+ \to [0, \mathbb{E}_{P_U}[U^2]],
\]

\[
q_v \mapsto q_{u,n}(1, q_v),
\]

is continuous.

Lemma 15

For all bounded sequence \( (q_{v,n}) \in \mathbb{R}_+^N \), if one choose \( q_v = q_{v,n} \) and \( q_u = q_{u,n}(\cdot, q_{v,n}) \) in the definition of \( F_{n,t} \), we have

\[
F_n = \mathcal{F}(\lambda = 1, \alpha, Q_n(q_{v,n}), q_{v,n}) + o_n(1)
\]

where \( o_n(1) \) is a quantity that goes to 0 as \( n \to \infty \).
Proof. By Lemma 13 we have for all \( t \in [0, 1] \):

\[
\frac{\partial}{\partial t} F_{n,t}(\lambda) = \frac{\alpha}{2} q_{u,n}'(t, q_{v,n}) q_{v,n} - \frac{1}{2} \mathbb{E} \left( \left( \mathbf{u} \cdot \mathbf{U} - q_{u,n}'(t, q_{v,n}) \right) \left( \frac{m}{n} \mathbf{v} \cdot \mathbf{V} - \alpha q_{v,n} \right) \right)_{n,t}.
\] (4.31)

By Cauchy-Schwartz inequality

\[
\left| \mathbb{E} \left( \left( \mathbf{u} \cdot \mathbf{U} - q_{u,n}'(t, q_{v,n}) \right) \left( \frac{m}{n} \mathbf{v} \cdot \mathbf{V} - \alpha q_{v,n} \right) \right)_{n,t} \right| \leq \left( \mathbb{E} \left( \left( \mathbf{u} \cdot \mathbf{U} - q_{u,n}'(t, q_{v,n}) \right)^2 \right)_{n,t} \right)^{1/2} \left( \mathbb{E} \left( \left( \frac{m}{n} \mathbf{v} \cdot \mathbf{V} - \alpha q_{v,n} \right)^2 \right)_{n,t} \right)^{1/2}.
\]

The right-hand side above goes to 0 uniformly in \( t \) because:

- \( \mathbb{E} \left( \left( \frac{m}{n} \mathbf{v} \cdot \mathbf{V} - \alpha q_{v,n} \right)^2 \right)_{n,t} \) is bounded because we assumed that \( P_V \) had a bounded support, \( (q_{v,n}) \)
  is bounded and \( m/n \to \alpha \).

- \( \mathbb{E} \left( \left( \mathbf{u} \cdot \mathbf{U} - q_{u,n}'(t, q_{v,n}) \right)^2 \right)_{n,t} \) goes to 0 by (4.26), uniformly in \( t \in [0, 1] \).

The integration of (4.31) with respect to \( t \in [0, 1] \) (combined with (4.20)) gives the result. \( \square \)

For simplicity we will now omit the dependencies on \( \lambda \) and \( \alpha \) in \( \mathcal{F} \). The proof of (4.15) will now follow from the two matching lower- and upper-bounds below.

\section*{Lower bound}

From (4.30) one deduces easily the following lower bound:

\textbf{Proposition 15}

\[
\liminf_{n \to \infty} F_n \geq \sup_{q_v \geq 0} \inf_{q_u \geq 0} \mathcal{F}(q_u, q_v).
\]

\textbf{Proof.} Let \( q_v \geq 0 \). For all \( n \geq 1 \) we have \( \mathcal{F}(Q_n(q_v), q_v) \geq \inf_{q_v \geq 0} \mathcal{F}(q_u, q_v) \). Thus, by (4.30) we deduce that for all \( q_v \geq 0 \):

\[
\liminf_{n \to \infty} F_n \geq \inf_{q_v \geq 0} \mathcal{F}(q_u, q_v).
\]

This inequality is true for all \( q_v \geq 0 \), hence the result. \( \square \)

\section*{Upper bound}

We will now deduce the converse upper bound.

\textbf{Proposition 16}

\[
\limsup_{n \to \infty} F_n \leq \sup_{q_v \geq 0} \inf_{q_u \geq 0} \mathcal{F}(q_u, q_v).
\]

\textbf{Proof.} By Proposition 5 and the continuity of \( Q_n \), the function \( L_n = 2\psi_P V \circ Q_n \) is continuous from \([0, \mathbb{E}P_V V^2]\) to \([0, \mathbb{E}P_V V^2]\). \( L_n \) admits therefore a fixed point \( q_{v,n} = L_n(q_{v,n}) \) in \([0, \mathbb{E}P_V V^2]\). We now notice that

\[
\mathcal{F}(Q_n(q_{v,n}), q_{v,n}) = \inf_{q_v \geq 0} \mathcal{F}(q_u, q_v).
\] (4.32)

Indeed the function \( g_n : q_u \to \mathcal{F}(q_u, q_{v,n}) = \psi_P \left( \alpha q_{v,n} \right) + \alpha \psi_P(q_u) - \frac{\alpha}{2} q_u q_{v,n} \) is convex and differentiable by Proposition 5 with derivative

\[
g_n'(q_u) = \frac{\alpha}{2} \left( 2\psi_P(q_u) - q_{v,n} \right).
\]
Thus \( g_n' (Q_n(q,v,n)) = \frac{\partial}{\partial n} (L_n(q,v,n) - q_v,v) = 0 \), by definition of \( q_v,v \). This proves (4.32). From (4.32) we now deduce that

\[
\mathcal{F}(Q_n(q,v,n),q,v) = \inf_{q_v \geq 0} \mathcal{F}(q,v), \quad \text{for some constant } C.
\]

It is not difficult to verify that \( \mathcal{F}(Q_n(q,v,n),q,v) \) satisfies a "bounded difference property" (see [17], Section 3.2) because the components of \( Q \) depend only on \( K > 0 \). We conclude the proof using Jensen’s inequality. \( \square \)

4.3.5 Technical lemmas

Concentration of the free energy

In this section, we prove Lemma 14 which states that

\[
v_n(s_n) = O(n^{-1/2}s_n^{-1}).
\]

We recall that \( v_n(s_n) = \sup_{t \in [0,1]} \sup_{1/2 \leq u, v \leq 3} \mathbb{E}[|\phi_t(u,v) - \mathbb{E}_z(\phi_t(u,v))|] \), where

\[
\phi_t : (a_u, a_v) \mapsto \frac{1}{n^{s_n}} \log \left( \int_{u, v} dP_U^\otimes n(u) dP_V^\otimes m(v) e^{\mathcal{F}(n,t)} \right).
\]

We have to prove that the perturbed free energy concentrates around its mean, uniformly in the perturbation. Lemma 14 follows from Lemma 16 and Lemma 17 below. Let \( \mathbb{E}_z \) denote the expectation with respect to the Gaussian random variables \( Z, Z^{(u)}, Z^{(v)}, Z^{(u)}_{\text{pert}}, Z^{(v)}_{\text{pert}} \).

**Lemma 16**

There exists a constant \( C > 0 \) such that for all \( t \in [0, 1] \) and \( (a_u, a_v) \in [1/2, 3]^2 \),

\[
\mathbb{E} |\phi_t(a_u, a_v) - \mathbb{E}_z(\phi_t(a_u, a_v))| \leq C n^{-1/2} s_n^{-1}.
\]

**Proof.** Let \( (a_u, a_v) \in [1/2, 3] \) and consider \( U \) and \( V \) to be fixed (i.e. we first work conditionally on \( U, V \)). Consider the function

\[
f : (Z, Z^{(u)}, Z^{(v)}, Z^{(u)}_{\text{pert}}, Z^{(v)}_{\text{pert}}) \mapsto \phi_t(a_u, a_v).
\]

It is not difficult to verify that

\[
\|\nabla f\|^2 \leq C' n^{-1} s_n^{-2} (1 + s_n) \leq 2 C' n^{-1} s_n^{-2},
\]

for some constant \( C' > 0 \) that depends only on \( K \) and \( \alpha \). The Gaussian Poincaré inequality (see [17] Chapter 3) gives then

\[
\mathbb{E}_z (\phi_t(a_u, a_v) - \mathbb{E}_z(\phi_t(a_u, a_v))^2 \leq 2 C' n^{-1} s_n^{-2}.
\]

We obtain the lemma by integration over \( U, V \) and Jensen’s inequality. \( \square \)

**Lemma 17**

There exists a constant \( C > 0 \) such that for all \( t \in [0, 1] \) and \( (a_u, a_v) \in [1/2, 3]^2 \),

\[
\mathbb{E} |\mathbb{E}_z(\phi_t(a_u, a_v) - \mathbb{E}_z(\phi_t(a_u, a_v))| \leq C n^{-1/2} s_n^{-1}.
\]

**Proof.** It is not difficult to verify that the function

\[
g : (U, V) \mapsto \mathbb{E}_z(\phi_t(a_u, a_v)
\]

verifies a “bounded difference property” (see [17], Section 3.2) because the components of \( U \) and \( V \) are bounded by a constant \( K > 0 \). Then Corollary 3.2 from [17] (which is a corollary from the Efron-Stein inequality) implies that for all \( t \in [0, 1] \) and \( a_u, a_v \in [1/2, 3] \)

\[
\mathbb{E} (\mathbb{E}_z(\phi_t(a_u, a_v) - \mathbb{E}_z(\phi_t(a_u, a_v))^2 \leq C' n^{-1} s_n^{-2}.
\]

for some constant \( C' \) depending only on \( K \) and \( \alpha \). We conclude the proof using Jensen’s inequality. \( \square \)

44
A sup-inf formula

Proposition 17
Let \( f \) be a convex, strictly increasing, differentiable function on \( \mathbb{R}_+ \) and \( g \) be a strictly convex, increasing, Lipschitz, differentiable function on \( \mathbb{R}_+ \). For \((q_1, q_2) \in \mathbb{R}_+\) we define \( \varphi(q_1, q_2) = f(q_1) + g(q_2) - q_1q_2 \). Then
\[
\sup_{q_1 \geq 0} \inf_{q_2 \geq 0} \varphi(q_1, q_2) = \sup_{q_1 = g'(q_2)} \varphi(q_1, q_2). \quad (43.3)
\]
Moreover, both extrema are achieved and precisely at the same couples \((q_1, q_2)\).

**Proof.** Let \( q_1, q_2 \geq 0 \) such that \( q_1 = g'(q_2) \). The function \( \varphi(q_1, \cdot) \) is convex, and its derivative at \( q_2 \) is equal to zero. Thus \( \varphi(q_1, q_2) = \inf_{q_2 \geq 0} \varphi(q_1, q_2) \) and consequently:
\[
\sup_{q_1 = g'(q_2)} \varphi(q_1, q_2) \leq \sup_{q_1 \geq 0} \inf_{q_2 \geq 0} \varphi(q_1, q_2).
\]
Let us now prove the converse inequality. Let \( L_g = \|g'|_{\infty} < +\infty \) because \( g \) is Lipschitz. Let \( q_1 \geq 0 \). We have \( \inf_{q_2 \geq 0} \varphi(q_1, q_2) = f(q_1) - g^*_*(q_1) \) where
\[
g^*_*(q_1) = \sup_{q_2 \geq 0} q_1q_2 - g(q_2)
\]
is the Fenchel-Legendre conjugate of \( g \). \( g^*_* \) is convex (as a supremum of affine functions) and one verify easily that \( g^*_*(q_1) \) is finite for \( q_1 < L_g \) whereas \( g^*_*(q_1) = +\infty \) for \( q_1 > L_g \). The next lemma follows from standard convex arguments.

Lemma 18
\( g^* \) is differentiable over \([0, L_g)\) and for all \( q_1 \in [0, L_g) \),
\[
g^*_*(q_1) = q^*_2(q_1) = (g')^{-1}(q_1),
\]
where \( q^*_2(q_1) \geq 0 \) is the unique maximizer of \( g^*_2(q_2) = 0 \mapsto q_1q_2 - g(q_2) \).

**Proof.** Let \( 0 \leq q_1 < L_g \). Then \( q_1q_2 - g(q_2) \xrightarrow[q_2 \to \infty]{} -\infty \). Thus, by strict convexity of \( g \), the supremum \( \sup_{q_2 \geq 0} q_1q_2 - g(q_2) \) is achieved at a unique \( q^*_2(q_1) = (g')^{-1}(q_1) \). The Lemma follows then from Proposition 21 in Appendix B. \( \square \)

Let us distinguish two cases:

- \( g^*_*(L_g) = +\infty \). In that case, it is not difficult to verify that \( g^*_*(q_1) \xrightarrow[q_1 \to L_g]{} +\infty \). \( f - g^* \) is thus continuous on \([0, L_g)\) and tend to \(-\infty \) at \( L_g \); it achieves its supremum at some \( q^*_1 \in [0, L_g) \).

- \( g^*_*(L_g) \) is finite. In that case \( f - g^* \) is continuous on \([0, L_g)\) (by convexity of \( f \) and \( g^*_* \)) and reaches therefore its maximum at some \( q^*_1 \). However, \( g^*_*(q_1) = (g')^{-1}(q_1) \) \xrightarrow[q_1 \to L_g]{} +\infty \), hence \( q^*_1 < L_g \).

In both cases \( \inf_{q_2 \geq 0} \varphi(\cdot, q_2) \) achieves its supremum at some \( q^*_1 \in [0, L_g) \). Let us show that \( g^*(q^*_2(q^*_1)) = q^*_1 \). The optimality condition of \( q^*_2 := q^*_2(q^*_1) \) gives
\[
g^*(q^*_2) \geq q^*_1.
\]
Let us suppose that \( g^*(q^*_2) > q^*_1 \). This is only the case if \( q^*_2 = 0 \). The minimum of \( q_2 \mapsto g(q_2) - g^*(0)q_2 \) is achieved (by convexity) at \( q_2 = 0 \). Thus \( \inf_{q_2 \geq 0} g(q_2) - g^*(0)q_2 = g(0) = \inf_{q_2 \geq 0} g(q_2) - q^*_1q_2 \). But \( f \) is increasing, so \( f(g'(0)) > f(q^*_1) \) and consequently
\[
\inf_{q_2 \geq 0} \varphi(g'(0), q_2) > \inf_{q_2 \geq 0} \varphi(q^*_1, q_2),
\]
whereas
which contradicts the optimality of $q^*_1$. We obtain that $g'(q^*_2) = q^*_1$.

Let us now show that $q^*_2 = f'(q^*_1)$. The optimality condition at $q^*_1$ gives

$$0 \geq f'(q^*_1) - g'_1(q^*_1) = f'(q^*_2) - q^*_2.$$

Suppose that $f'(q^*_1) > q^*_2$. This is only the case for $q^*_1 = 0$. $g$ is increasing so

$$\varphi(0, f'(0)) < \varphi(0, q^*_2) = \inf_{q_2 \geq 0} \varphi(0, q_2).$$

This is absurd: $q^*_2 = f'(q^*_1)$. We conclude that

$$\sup_{q_1 \geq 0} \inf_{q_2 \geq 0} \varphi(q_1, q_2) = \varphi(q^*_1, q^*_2) \leq \sup_{q_1 = g'(q_2)} \varphi(q_1, q_2),$$

which finishes the proof of (4.33).

From the proof above, we see that a couple $(q^*_1, q^*_2)$ that achieves the “sup-inf” verifies $q^*_1 = g'(q^*_2)$ and $q^*_2 = f'(q^*_1)$ and thus achieves the supremum of the right-hand side of (4.33). Conversely, if $(q^*_1, q^*_2)$ achieves the supremum of the right-hand side of (4.33), then as noticed at the beginning of the proof

$$\varphi(q^*_1, q^*_2) = \inf_{q_2 \geq 0} \varphi(q^*_1, q_2)$$

and therefore, using (4.33), $(q^*_1, q^*_2)$ achieves the “sup-inf”. Therefore both extrema in (4.33) are achieved and precisely at the same couples $(q_1, q_2)$. □

### 4.4 Proof of Theorem 5

In order to prove Theorem 5, we are going to use Proposition 13 in order to get an upper bound on the MMSE, and consider the following model with side information to obtain a lower bound. Suppose that we observe for $\epsilon \geq 0$

$$\begin{align*}
Y &= \sqrt{\frac{n}{2}} UV^\top + Z \\
Y'_\epsilon &= \sqrt{\frac{n}{2}} UU^\top + Z'
\end{align*}$$

where $Z'_{i,j} \sim \mathcal{N}(0, 1)$ are independent from everything else. Define the corresponding free energy

$$F_n(\epsilon) = \frac{1}{n} \mathbb{E} \log \int dP^{\otimes n}_U(u) dP^{\otimes m}_V(v) \exp \left( \sum_{1 \leq i,j \leq n} \sqrt{\frac{\epsilon}{n}} Y'_{i,j} u_i u_j - \frac{\epsilon u_i^2 u_j^2}{2} - \sum_{i,j} \frac{\lambda Y_{i,j} u_i u_j}{2n} \right).$$

In order to simplify the notations, we will remove the dependencies in $\lambda$ and $\alpha$.

**Proposition 18**

Let $\rho_u \geq \mathbb{E}_{P_U}\left[|U|^2\right] + 1$ and $\rho_v \geq \mathbb{E}_{P_V}\left[|V|^2\right] + 1$. Define $\epsilon_{\text{max}} = \alpha \min \left( \lambda^{-1} \rho_u^{-1}, \frac{1}{2} \text{MMSE}_{P_U}(\lambda \rho_u)^2 \right)$. Then for all $0 \leq \epsilon \leq \epsilon_{\text{max}}$, we have

$$\lim_{n \to \infty} F_n(\epsilon) = \sup_{0 \leq q_u \leq \rho_u} \inf_{0 \leq q_v \leq \rho_v} \left\{ \psi_{P_U}(\alpha \lambda q_v + 2\epsilon q_u) + \alpha \psi_{P_V}(\lambda q_u) - \frac{\epsilon q_u^2}{2} - \frac{\alpha \lambda q_u q_v}{2} \right\}.$$

Proposition 18 is proved at the end of this section. Let us now deduce Theorem 5 from Proposition 18. Let us fix $\lambda \in D_\alpha$. Let us define

$$\text{MMSE}^{(u)}_{n, \epsilon} = \frac{1}{n^2} \mathbb{E} \left[ \sum_{1 \leq i,j \leq n} (U_i U_j - \mathbb{E}[U_i U_j | Y, Y'_\epsilon])^2 \right].$$
Let \( \epsilon \in [0, \epsilon_{\text{max}}] \). Obviously, \( \text{MMSE}^{(u)}_{n, 0} \geq \text{MMSE}^{(u)}_{n, \epsilon} \). \( F_n(\epsilon) \) is a convex function of \( \epsilon \in [0, \epsilon_{\text{max}}] \); so is \( f(\epsilon) \), its pointwise limit. Consequently for almost all \( \epsilon \in [0, \epsilon_{\text{max}}] \)

\[
F_n'(\epsilon) = \lim_{n \to \infty} f'(\epsilon).
\]

By the I-MMSE relation (1.4) we have \( F_n'(\epsilon) = \frac{1}{2} \left( \mathbb{E}_{P_v}[U^2]^2 - \text{MMSE}^{(u)}_{n, \epsilon} \right) \), which then give

\[
\text{MMSE}^{(u)}_{n, \epsilon} \underset{n \to \infty}{\longrightarrow} \mathbb{E}_{P_v}[U^2]^2 - 2f'(\epsilon).
\]

We deduce that for almost all \( \epsilon \in [0, \epsilon_{\text{max}}] \),

\[
\liminf_{n \to \infty} \text{MMSE}^{(u)}_{n}(\lambda, \epsilon) \geq \mathbb{E}_{P_v}[U^2]^2 - 2f'(\epsilon). \tag{4.34}
\]

Let \( q_u \in [0, \rho_u] \). The function

\[
\phi_{q_u}(\epsilon) = \inf_{0 \leq q_v \leq \rho_v} \left\{ \psi_{P_v}(\alpha \lambda q_v + 2\epsilon q_u) - \frac{\epsilon q_u^2}{2} + \alpha \psi_{P_v}(\lambda q_u) - \frac{\alpha \lambda q_u q_v}{2} \right\}
\]

is an infimum over a compact set of differentiable function of \( \epsilon \). By strict convexity of \( \psi_{P_v} \) (see Proposition 5) the infimum is achieved at a unique \( q_v = q_v(\epsilon, \lambda, \alpha) \). By the Proposition 21 presented in Appendix B, \( \phi_{q_u} \) is differentiable on \( [0, \epsilon_{\text{max}}] \), with derivative:

\[
\phi_{q_u}'(\epsilon) = 2q_u \psi_{P_v}'(\alpha \lambda q_v(\epsilon, \lambda, \alpha) + 2\epsilon q_u) - \frac{q_u^2}{2}.
\]

Now, \( f(\epsilon) = \sup_{q_u \in [0, \rho_u]} \phi_{q_u}(\epsilon) \). Since \( \lambda \in D_\alpha \), we know that for \( \epsilon = 0 \), the supremum in \( q_u \) is achieved at a unique point \( q^*_u = q^*_u(\lambda, \alpha) \). Therefore, by Proposition 21 in Appendix B, \( f \) is right-differentiable at \( \epsilon = 0 \) and

\[
f'(0^+) = \phi_{q^*_u}'(0) = 2q^*_u \psi_{P_v}'(\alpha \lambda q^*_v) - \frac{(q^*_u)^2}{2} = \frac{(q^*_u)^2}{2},
\]

because, by Theorem 4, \( 2\psi_{P_v}'(\alpha \lambda q^*_v) = q^*_v \). \( f \) is convex, therefore \( f'(\epsilon) \to f'(0^+) \) as \( \epsilon \to 0 \) in (4.34) leads to

\[
\liminf_{n \to \infty} \text{MMSE}^{(u)}_{n, 0} \geq \mathbb{E}_{P_v}[U^2]^2 - (q^*_u)^2.
\]

Let \( \langle \cdot \rangle_n \) denotes the expectation with respect to \( (u, v) \) sampled from the posterior distribution of \( (U, V) \) given \( Y \), independently of everything else. Then \( \text{MMSE}^{(u)}_{n, 0} = \mathbb{E}_{P_v}[U^2]^2 - \mathbb{E}\langle (u \cdot U)^2 \rangle_n \). This gives (the corresponding result for \( V \) is obtained by symmetry):

\[
\limsup_{n \to \infty} \mathbb{E}\langle (u \cdot U)^2 \rangle_n \leq (q^*_u)^2 \quad \text{and} \quad \limsup_{n \to \infty} \mathbb{E}\langle (v \cdot V)^2 \rangle_n \leq (q^*_v)^2. \tag{4.35}
\]

Now, we know by Proposition 13 that

\[
\mathbb{E}_{P_v}[U^2]\mathbb{E}_{P_v}[V^2] - \mathbb{E}\langle (u \cdot U)(v \cdot V) \rangle_n = \text{MMSE}_n(\lambda) \longrightarrow_{n \to \infty} \mathbb{E}_{P_v}[U^2]\mathbb{E}_{P_v}[V^2] - q^*_u q^*_v,
\]

which gives \( \mathbb{E}\langle (u \cdot U)(v \cdot V) \rangle_n \rightarrow q^*_u q^*_v \). By Cauchy-Schwarz inequality we have

\[
\mathbb{E} \left[ \langle (u \cdot U)(v \cdot V) \rangle_n \right]^2 \leq \mathbb{E}\langle (u \cdot U)^2 \rangle_n \mathbb{E}\langle (v \cdot V)^2 \rangle_n
\]

which gives, by taking the liminf:

\[
(q^*_u q^*_v)^2 \leq \left( \liminf_{n \to \infty} \mathbb{E}\langle (u \cdot U)^2 \rangle_n \right) \left( \liminf_{n \to \infty} \mathbb{E}\langle (v \cdot V)^2 \rangle_n \right).
\]

47
Combining this with (4.35) gives the result.

**Proof of Proposition 18.** It suffices to prove the result in the case where $P_U$ and $P_V$ have bounded support, because we can then proceed by approximation as in Section 3.4.7. From now, we suppose to be in that case. Since the dependency in $\lambda$ can be incorporated in the prior $P_U$, we only have to prove Proposition 18 in the case $\lambda = 1$.

We will follow the same steps as in Section 4.3: the only difference is that we have here some extra observations of $U^U$. We will therefore only present the main steps. Define
\[
\mathcal{F}_c(q_u, q_v) = \psi_{P_U}(\alpha \lambda q_v + 2\epsilon q_u) + \alpha \psi_{P_V}(\lambda q_u) - \frac{c q_u^2}{2} - \frac{\alpha \lambda q_u q_v}{2}.
\]
Let $q_u, q_v : [0, 1] \to \mathbb{R}_+$ be two differentiable functions such that $q_u(0) = q_v(0) = 0$. For $0 \leq t \leq 1$ we consider the following observation channel
\[
\begin{aligned}
Y_t &= \sqrt{(1-t)/n} \ U V^\top + Z, \\
Y'_t &= \sqrt{\epsilon(1-t)/n} \ U U^\top + Z', \\
Y^{(u)}_t &= \sqrt{\alpha q_u(t) + 2\epsilon q_u(t)} \ U + Z^{(u)}, \\
Y^{(v)}_t &= \sqrt{q_u(t)} \ V + Z^{(v)},
\end{aligned}
\]
where $Z_t^{(u)}, Z_t^{(v)} \sim N(0, 1)$ are independent of everything else. We will denote (analogously to (4.19)) by $F_{n,t}$ the interpolating free energy and by $\langle \cdot \rangle_{n,t}$ (analogously to (4.18)) corresponding Gibbs measure. We have
\[
\begin{cases}
F_{n,0} = F_n(\epsilon), \\
F_{n,1} = F_c(q_u(1), q_v(1)) + \epsilon \frac{q_u(1)^2}{2} + \alpha \frac{q_u(1)q_v(1)}{2}.
\end{cases}
\]
We have the analog of Lemma 13:
\[
\frac{\partial}{\partial t} F_{n,t} = \frac{\alpha}{2} q_u'(t) q_u'(t) + \frac{\epsilon}{2} q_u'(t)^2 - \frac{1}{2} \mathbb{E} \left( \langle u \cdot U - q_u'(t) \rangle \langle v \cdot V - q_v'(t) \rangle \right)_{n,t} - \frac{1}{2} \mathbb{E} \left( \langle u \cdot U - q_u'(t) \rangle^2 \right)_{n,t} + o_n(1),
\]
where $o_n(1) \to 0$, uniformly in $t \in [0, 1]$. As for the proof of Theorem 4, we are only interested in the limit of the free energy, we can assume (see 4.3.3) that the overlaps concentrates around their expectations:
\[
\mathbb{E} \left( \langle u^{(1)} \cdot u^{(2)} - \mathbb{E} \langle u^{(1)} \cdot u^{(2)} \rangle_{n,t} \rangle_{n,t} \right)^2_{n,t} , \quad \mathbb{E} \left( \langle v^{(1)} \cdot v^{(2)} - \mathbb{E} \langle v^{(1)} \cdot v^{(2)} \rangle_{n,t} \rangle_{n,t} \right)^2_{n,t} \to 0, \quad t \to \infty,
\]
uniformly in $t \in [0, 1]$.

**Lemma 19**
\[
\liminf_{n \to \infty} F_n(\epsilon) \geq \sup_{q_u \in [0, \rho_u]} \inf_{q_v \in [0, \rho_v]} F_c(q_u, q_v).
\]

**Proof.** Let us now chose $q_u(t) = q_u t$ for some $q_u \geq 0$ and $q_v$ to be solution of the Cauchy problem:
\[
q_v(0) = 0, \quad \text{and} \quad q_v'(t) = \mathbb{E} \langle v \cdot V \rangle_{n,t} \in [0, \mathbb{E} \langle V^2 \rangle].
\]

We obtain, using the relation $F_n(\epsilon) = F_{n,1} - \int_0^1 \frac{\partial F_{n,t}}{\partial t} dt$:
\[
F_n(\epsilon) = F_c(q_u, q_v(1)) + \int_0^1 \left( \frac{1}{2} \mathbb{E} \left( \langle u \cdot U - q_u \rangle \langle v \cdot V - q_v'(t) \rangle \right)_{n,t} + \frac{1}{2} \mathbb{E} \left( \langle u \cdot U - q_u \rangle^2 \right)_{n,t} \right) dt + o_n(1)
\]
and thus, using (4.37): $F_n(\epsilon) = F_c(q_u, q_v(1)) + o_n(1) \geq \inf_{q_v \in [0, \rho_v]} F_c(q_u, q_v) + o_n(1)$ for all $q_u \geq 0$, so that
\[
\liminf_{n \to \infty} F_n(\epsilon) \geq \sup_{q_u \in [0, \rho_u]} \inf_{q_v \in [0, \rho_v]} F_c(q_u, q_v).
\]
Lemma 20

Let $0 \leq \epsilon \leq \frac{4}{9} \text{MMSE}_{P_{\nu}}(\rho_u)^2$. Then the function $h : q_u \mapsto \alpha \psi'_{P_{\nu}}(q_u) - \frac{\epsilon q_u^2}{2}$ is strictly convex on $[0, \rho_u]$.

**Proof.** $P_{\nu}$ has bounded support, thus by (1.7) we have for all $q \geq 0$, $\psi''_{P_{\nu}}(q) \geq \frac{1}{2} \text{MMSE}_{P_{\nu}}(q)^2$. We have thus for all $q_u \in [0, \rho_u]$

$$h''(q_u) \geq \alpha \psi''_{P_{\nu}}(q_u) - \epsilon > 0$$

because for all $q_u \leq \rho_u$, MMSE$_{P_{\nu}}(q_u) \geq$ MMSE$_{P_{\nu}}(\rho_u)$ and therefore $\alpha \psi''_{P_{\nu}}(q_u) > \epsilon$ for all $\epsilon \leq \frac{4}{9} \text{MMSE}_{P_{\nu}}(\rho_u)^2$. \hfill \Box

Lemma 21

For all $0 \leq \epsilon \leq \frac{1}{2} \alpha \rho_u^{-1}$, all $q_u \in [0, \rho_u]$ and all $q_v \geq 0$, we have

$$g(q_u, q_v) := \frac{4\epsilon}{\alpha} \psi'_{P_{\nu}}(\alpha q_v + 2\epsilon q_u) + 2\psi'_{P_{\nu}}(q_u) - \frac{2\epsilon}{\alpha} q_u \in [0, \rho_v].$$

**Proof.** $q_u \mapsto \psi'_{P_{\nu}}(q_u) - \frac{\epsilon q_u^2}{2}$ is convex on $[0, \rho_u]$ by Lemma 20, therefore $q_u \mapsto \psi'_{P_{\nu}}(q_u) - \epsilon q_u/\alpha$ is increasing and thus non-negative since $\psi'_{P_{\nu}}(0) \geq 0$. Hence for all $q_u \in [0, \rho_u]$ and all $q_v \geq 0$, $g(q_u, q_v) \geq 0$. For the upper bound, we have $2\psi'_{P_{\nu}} \leq \mathbb{E}_{P_{\nu}}[U^2]$ and $2\psi'_{P_{\nu}} \leq \mathbb{E}_{P_{\nu}}[V^2]$ by Proposition 5. Consequently, for $\epsilon \leq \frac{1}{2} \alpha \rho_u^{-1}$ and all $q_u \in [0, \rho_u]$ and $q_v \geq 0$ we have $g(q_u, q_v) \leq \mathbb{E}_{P_{\nu}}[V^2] + 1 \leq \rho_v$. \hfill \Box

Lemma 22

Under the hypotheses of Proposition 18 we have

$$\limsup_{n \to \infty} F_n(\epsilon) \leq \sup_{q_v \in [0, \rho_v]} \inf_{q_u \in [0, \rho_u]} \mathcal{F}_\epsilon(q_u, q_v).$$

**Proof.** Let us consider again the interpolation scheme (4.36) with $q_v(t) = q_v t$ and $q_u = q_{u,n}(\cdot, q_v)$, the solution of the Cauchy problem

$$q_u(0) = 0 \quad \text{and} \quad q_u'(t) = \mathbb{E}(u \cdot U)_{n,t} \in [0, \mathbb{E}_{P_{\nu}}[U^2]].$$

Since $u \cdot U$ concentrates around its mean, see (4.37), we obtain:

$$F_n(\epsilon) \leq \mathcal{F}_\epsilon(q_{u,n}(1, q_v), q_v) + \frac{\epsilon}{2} \left(q_{u,n}(1, q_v)^2 - \int_0^1 q_u'(t, q_v)^2 dt\right) + o_n(1)$$

$$\leq \mathcal{F}_\epsilon(q_{u,n}(1, q_v), q_v) + o_n(1).$$

By the parametric Cauchy-Lipschitz Theorem, $q_v \mapsto q_{u,n}(1, q_v)$ is continuous. Therefore, the function

$$q_v \mapsto g(q_{u,n}(1, q_v), q_v)$$

is continuous from $[0, \rho_v]$ to $[0, \rho_v]$: it admits a fixed point $q_{v,n}^*$. If we now chose $q_v = q_{v,n}^*$, we have

$$\mathcal{F}_\epsilon(q_{u,n}(1, q_{v,n}^*), q_{v,n}^*) = \inf_{q_u \in [0, \rho_u]} \mathcal{F}_\epsilon(q_u, q_{v,n}^*) \leq \sup_{q_v \in [0, \rho_v]} \inf_{q_u \in [0, \rho_u]} \mathcal{F}_\epsilon(q_u, q_v).$$

The equality above comes from the fact that the function $q_u \mapsto \mathcal{F}_\epsilon(q_u, q_{v,n}^*)$ is convex (by Lemma 20 above) and $q_{u,n}(1, q_{v,n}^*)$ cancels its derivative because $q_{v,n}^*$ is a fixed point of (4.39). Plugging this inequality into (4.38), we obtain the lemma by taking the limsup. \hfill \Box

In order to prove Proposition 18, it remains to show that $\sup_{q_v} \inf_{q_u} \mathcal{F}_\epsilon(q_u, q_v) \geq \sup_{q_v} \inf_{q_u} \mathcal{F}_\epsilon(q_u, q_v)$. By Lemma 20, for all $q_v \in [0, \rho_v]$, the function $q_u \mapsto \mathcal{F}_\epsilon(q_u, q_v)$ is strictly convex and admits thus a
unique minimizer $q_u^*(q_v)$. The minimum is taken over a compact set, therefore Proposition 21 gives that the function

$$\phi : q_v \mapsto \inf_{q_u \in [0, \rho_u]} \mathcal{F}_\epsilon(q_u, q_v)$$

is differentiable on $[0, \rho_v]$ with derivative $\phi'(q_v) = \alpha'(q_v) \alpha q_v + 2\epsilon q_u^*(q_v) - \frac{q_u^*(q_v)}{2}$. Let $q_v^*$ be a maximizer of $\phi$ over $[0, \rho_v]$. We have thus

$$\sup_{q_v} \inf_{q_u} \mathcal{F}_\epsilon(q_u, q_v) = \mathcal{F}_\epsilon(q_u^*(q_v^*), q_v^*)$$

It is not difficult to verify (by playing with the optimality conditions) that $\phi'(q_v^*) = 0$. $q_v^*$ is thus the minimizer of

$$q_v \mapsto \mathcal{F}_\epsilon(q_u^*(q_v^*), q_v)$$

because $q_v^*$ cancels the derivative of this convex function. We get

$$\sup_{q_v \in [0, \rho_v]} \inf_{q_u \in [0, \rho_u]} \mathcal{F}_\epsilon(q_u, q_v) = \mathcal{F}_\epsilon(q_u^*(q_v^*), q_v^*) = \inf_{q_v \in [0, \rho_v]} \mathcal{F}_\epsilon(q_u^*(q_v^*), q_v) \leq \sup_{q_u \in [0, \rho_u]} \inf_{q_v \in [0, \rho_v]} \mathcal{F}_\epsilon(q_u, q_v),$$

which concludes the proof. \qed
Appendix

A Proofs of some basic properties of the MMSE and the free energy

A.1 Proof of Proposition 1

Let $0 < \lambda_2 \leq \lambda_1$. Define $\Delta_1 = \lambda_1^{-1}$, $\Delta_2 = \lambda_2^{-1}$ and
\[
\begin{cases}
Y_1 = X + \sqrt{\Delta_1} Z_1 \\
Y_2 = X + \sqrt{\Delta_1} Z_1 + \sqrt{\Delta_2 - \Delta_1} Z_2,
\end{cases}
\]
where $X \sim P_X$ is independent from $Z_1, Z_2 \overset{i.i.d.}{\sim} \mathcal{N}(0, I_n)$. Now, by independence between $(X, Y_1)$ and $Z_2$ we have
\[
\operatorname{MMSE}(\lambda_1) = \mathbb{E} \|X - \mathbb{E}[X|Y_1]\|^2 = \mathbb{E} \|X - \mathbb{E}[X|Y_1, Z_2]\|^2 = \mathbb{E} \|X - \mathbb{E}[X|Y_1, Y_2]\|^2 \\
\leq \mathbb{E} \|X - \mathbb{E}[X|Y_2]\|^2 = \operatorname{MMSE}(\lambda_2).
\]
Next, notice that
\[
\operatorname{MMSE}(\lambda_1) = \mathbb{E} \|X - \mathbb{E}[X|Y_1]\|^2 \leq \mathbb{E} \|X - \mathbb{E}[X]\|^2 = \operatorname{MMSE}(0).
\] (40)
This shows that the MMSE is non-increasing on $\mathbb{R}_+$. It remains to prove the last point:
\[
0 \leq \operatorname{MMSE}(\lambda) = \mathbb{E} \|X - \mathbb{E}[X|Y]\|^2 \leq \mathbb{E} \|X - \frac{1}{\sqrt{\lambda}} Y\|^2 = \frac{n}{\lambda} \xrightarrow[\lambda \to +\infty]{} 0.
\]

A.2 Proof of Proposition 2

We start by proving that MMSE is continuous at $\lambda = 0$. Let $\lambda \geq 0$ and consider $Y, X, Z$ as given by (1.1). Let $c > 0$ such that $\mathbb{P}(\|X\| \leq c) > 0$ and define the random variable $B_c = 1(\|X\| \leq c)$.
\[
\operatorname{MMSE}(\lambda) \geq \mathbb{E} \|X - \mathbb{E}[X|Y, B_c]\|^2 \geq \mathbb{E} \left[ B_c \|X - \mathbb{E}[X|Y, B_c]\|^2 \right].
\] (41)
By dominated convergence one has almost surely that
\[
B_c \mathbb{E}[X|Y, B_c] = B_c \frac{\int_{\|x\| \leq c} dP_X(x) e^{H_X, Y(x)}}{\int_{\|x\| \leq c} dP_X(x) e^{H_X, Y(x)}} \xrightarrow[\lambda \to 0]{} B_c \mathbb{P}(\|X\| \leq c) \mathbb{P}(1(\|X\| \leq c)X),
\]
and therefore another application of the dominated convergence theorem gives:
\[
\mathbb{E} \left[ B_c \|X - \mathbb{E}[X|Y, B_c]\|^2 \right] \xrightarrow[\lambda \to 0]{} \mathbb{E} \left[ (1(\|X\| \leq c) \|X - \mathbb{P}(\|X\| \leq c)X\|^2 \right].
\]
Again, by the dominated convergence theorem, the right-hand side converges to $\mathbb{E} \|X - \mathbb{E}[X]\|^2$ as $c \rightarrow +\infty$. Going back to (41), we obtain that
\[
\liminf_{\lambda \to 0} \operatorname{MMSE}(\lambda) \geq \mathbb{E} \|X - \mathbb{E}[X]\|^2,
\]
which combined with the bound $\operatorname{MMSE}(\lambda) \leq \mathbb{E} \|X - \mathbb{E}[X]\|^2$, gives $\operatorname{MMSE}(\lambda) \xrightarrow[\lambda \to 0]{} \mathbb{E} \|X - \mathbb{E}[X]\|^2$. This proves that the MMSE is continuous at $\lambda = 0$.

Let us now prove that the MMSE is continuous on $\mathbb{R}_+$. We need here a technical lemma:
Lemma 23

For all $\lambda > 0$, $p \geq 1$

\[ \mathbb{E}\|X - \langle x\rangle_\lambda\|^2p \leq \frac{2p(2p!)^n}{\lambda^p} \eta_p^{p+1}. \]

Proof. We reproduce here the proof from [34], Proposition 5. We start with the equality

\[ \sqrt{\lambda}(X - \langle x\rangle_\lambda) = \sqrt{\lambda}X - \mathbb{E}[\sqrt{\lambda}X|Y] = Y - Z - \mathbb{E}[Y - Z|Y] = \mathbb{E}[Z|Y] - Z. \]

We have therefore

\[ \mathbb{E}\|X - \langle x\rangle_\lambda\|^2p = \frac{1}{\lambda^p} \mathbb{E}\|E[Z|Y] - Z\|^2p \leq \frac{2^{2p-1}}{\lambda^p} \mathbb{E}\|[E[Z|Y]]^2p + \|Z\|^2p \leq \frac{2^{2p}}{\lambda^p} \mathbb{E}\|Z\|^2p. \]

It remains to bound

\[ \mathbb{E}\|Z\|^2p \leq n^p \mathbb{E}\left[ \sum_{i=1}^n Z_i^{2p} \right] = n^{p+1} \frac{(2p)!}{2p!}. \]

□

Let $\lambda_0 > 0$. The family of random variables $\{\|X - \langle x\rangle_\lambda\|^2\}_{\lambda \geq \lambda_0}$ is bounded in $L^2$ by Lemma 23 and is therefore uniformly integrable. The function $\lambda \mapsto \|X - \langle x\rangle_\lambda\|^2$ is continuous on $[\lambda_0, +\infty)$, the uniform integrability ensures that MMSE : $\lambda \mapsto \mathbb{E}\|X - \langle x\rangle_\lambda\|^2$ is continuous over $[\lambda_0, +\infty)$. This is valid for all $\lambda_0 > 0$: we conclude that MMSE is continuous over $(0, +\infty)$.

A.3 Proof of the I-MMSE relation: Proposition 4

\[ \text{MMSE}(\lambda) = \mathbb{E}\|X - \langle x\rangle_\lambda\|^2 = \mathbb{E}\|X\|^2 + \mathbb{E}\|\langle x\rangle_\lambda\|^2 - 2\mathbb{E}(x^\top X)_\lambda \]

Now, by the Nishimori property $\mathbb{E}\|\langle x\rangle_\lambda\|^2 = \mathbb{E}(\langle x(1)^{\top} x(2) \rangle_\lambda = \mathbb{E}(x^\top X)_\lambda$. Thus

\[ \text{MMSE}(\lambda) = \mathbb{E}\|X\|^2 - \mathbb{E}(x^\top X)_\lambda. \]

(42)

By (42) and (1.3), it suffices now to prove the second equality in (1.4). This will follow from the lemmas below.

Lemma 24

The free energy $F$ is continuous at $\lambda = 0$.

Proof. For all $\lambda \geq 0$,

\[ F(\lambda) = \mathbb{E}\log \int dP_X(x) e^{-\frac{1}{2}\|Y - \sqrt{\lambda}x\|^2 + \frac{1}{2}\|Y\|^2} = \mathbb{E}\log \int dP_X(x) e^{-\frac{1}{2}\|\sqrt{\lambda}X - \sqrt{\lambda}x + Z\|^2 + \lambda \mathbb{E}\|X\|^2} + n. \]

By dominated convergence $\int dP_X(x) e^{-\frac{1}{2}\|\sqrt{\lambda}X - \sqrt{\lambda}x + Z\|^2} \xrightarrow{\lambda \to 0} e^{-\frac{1}{2}\|Z\|^2}$. Jensen’s inequality gives

\[ \left| \log \int dP_X(x) e^{-\frac{1}{2}\|\sqrt{\lambda}X - \sqrt{\lambda}x + Z\|^2} \right| = - \log \int dP_X(x) e^{-\frac{1}{2}\|\sqrt{\lambda}X - \sqrt{\lambda}x + Z\|^2} \leq \frac{1}{2} \int dP_X(x) \|\sqrt{\lambda}X - \sqrt{\lambda}x + Z\|^2 \leq \frac{3}{2} \left\| X \right\|^2 + \mathbb{E}\|X\|^2 + \left\| Z \right\|^2, \]

for all $\lambda \in [0, 1]$. One can thus apply the dominated convergence theorem again to obtain that $F$ is continuous at $\lambda = 0$. □
Lemma 25

For all \( \lambda \geq 0 \),

\[
F(\lambda) - F(0) = \frac{1}{2} \int_{0}^{\lambda} \mathbb{E}(x^\top x) \, d\gamma.
\]

Proof. Compute for \( \lambda > 0 \)

\[
\frac{\partial}{\partial \lambda} \log \mathcal{Z}(\lambda, \mathbf{Y}) = \left\langle \frac{1}{2\sqrt{\lambda}} x^\top Z + x^\top X - \frac{1}{2} \|x\|^2 \right\rangle_\lambda.
\]

Since \( \mathbb{E}\|X\|^2 < \infty \), the right-hand side is integrable and one can apply Fubini’s theorem to obtain

\[
F(\lambda_2) - F(\lambda_1) = \int_{\lambda_1}^{\lambda_2} \mathbb{E}\left\langle \frac{1}{2\sqrt{\lambda}} x^\top Z + x^\top X - \frac{1}{2} \|x\|^2 \right\rangle_\lambda \, d\lambda.
\]

By Gaussian integration by parts, we have for all \( i \in \{1, \ldots, n\} \) and \( \lambda > 0 \)

\[
E Z_i(x_i) = E \frac{\partial}{\partial Z_i} \langle x_i \rangle_\lambda = E \left[ (\sqrt{\lambda} x_i^2)_\lambda - \sqrt{\lambda} \langle x_i \rangle_\lambda^2 \right] = \sqrt{\lambda} E \left[ \langle x_i^2 \rangle_\lambda - \langle x_i X_i \rangle_\lambda \right],
\]

where the last equality comes from the Nishimori property (Proposition 3). We have therefore

\[
F(\lambda_2) - F(\lambda_1) = \frac{1}{2} \int_{\lambda_1}^{\lambda_2} \mathbb{E}\langle x^\top X \rangle_\lambda \, d\lambda.
\]

By Lemma 24, \( F \) is continuous at 0 so we can take the limit \( \lambda_1 \to 0 \) to obtain the result. \( \square \)

A.4 Pseudo-Lipschitz continuity of the free energy with respect to the Wasserstein distance

Let \( P_1 \) and \( P_2 \) be two probability distributions on \( \mathbb{R}^n \), that admits a finite second moment. We denote by \( W_2(P_1, P_2) \) the 2nd Wasserstein distance between \( P_1 \) and \( P_2 \). For \( i = 1, 2 \) the free energy is defined as

\[
F_{P_i}(\lambda) = \mathbb{E} \log \int dP_i(x) \exp \left( \sqrt{\lambda} x^\top Z + \lambda x^\top X - \frac{\lambda}{2} \|x\|^2 \right),
\]

where the expectation is with respect to \( (X, Z) \sim P_i \otimes \mathcal{N}(0, I_n) \).

Proposition 19

For all \( \lambda \geq 0 \),

\[
|F_{P_1}(\lambda) - F_{P_2}(\lambda)| \leq \frac{\lambda}{2} (\mathbb{E}P_1\|X\|^2 + \sqrt{\mathbb{E}P_2\|X\|^2}) W_2(P_1, P_2).
\]

A similar result was proved in [66] but with a weaker bound for the \( W_2 \) distance.

Proof. Let \( \epsilon > 0 \). Let us fix a coupling of \( X_1 \sim P_1 \) and \( X_2 \sim P_2 \) such that

\[
(\mathbb{E}\|X_1 - X_2\|^2)^{1/2} \leq W_2(P_1, P_2) + \epsilon.
\]

Let us consider for \( t \in [0, 1] \) the observation model

\[
\begin{cases}
Y_1^{(t)} = \sqrt{t} X_1 + Z_1, \\
Y_2^{(t)} = \sqrt{(1 - t)} X_2 + Z_2,
\end{cases}
\]

53
where $Z_1, Z_2 \sim \mathcal{N}(0, I_n)$ are independent from $(X_1, X_2)$. Define
\[
f(t) = \mathbb{E} \log \int dP_1(x_1) dP_2(x_2) \exp \left( \sqrt{\lambda x_1} Y_1^{(t)} - \frac{\lambda t}{2} \|x_1\|^2 + \sqrt{\lambda(1-t)} x_2 Y_2^{(t)} - \frac{(1-t)}{2} \|x_2\|^2 \right).
\]
We have $f(0) = F_{P_2}(\lambda)$ and $f(1) = F_{P_1}(\lambda)$. By an easy extension of the I-MMSE relation (1.4) we have for all $t \in [0, 1]$
\[
f'(t) = \frac{\lambda}{2} \mathbb{E} \langle X_1^\top x_1 - X_2^\top x_2 \rangle_t,
\]
where $\langle \cdot \rangle_t$ denotes the expectation with respect to $(x_1, x_2)$ sampled from the posterior distribution of $(X_1, X_2)$ given $Y_1^{(t)}, Y_2^{(t)}$, independently of everything else. We have then
\[
\left| \frac{2}{\lambda} f'(t) \right| = \left| \mathbb{E} \langle X_1^\top (x_1 - x_2) - (X_2 - X_1)^\top x_2 \rangle_t \right |
\]
\[
\leq \left( \mathbb{E} \|X_1\|^2 \mathbb{E} \|x_1 - x_2\|^2 \right)^{1/2} + \left( \mathbb{E} \|x_2\|^2 \mathbb{E} \|X_2 - X_1\|^2 \right)^{1/2}
\]
\[
= \left( \mathbb{E} \|X_1\|^2 \mathbb{E} \|x_1 - x_2\|^2 \right)^{1/2} + \left( \mathbb{E} \|X_2\|^2 \mathbb{E} \|X_2 - X_1\|^2 \right)^{1/2}
\]
\[
\leq \left( \mathbb{E} \|X_1\|^2 + \mathbb{E} \|X_2\|^2 \right)^{1/2} (W_2(P_1, P_2) + \epsilon),
\]
where we used successively the Cauchy-Schwarz inequality and the Nishimori property (Proposition 3). We then let $\epsilon \to 0$ to obtain the result. \qed

\section*{B Differentiation of a supremum of functions}

We recall in this section two results about the differentiation of a supremum of functions from Milgrom and Segal [49]. Let $X$ be a set of parameters and consider a function $f : X \times [0, 1] \to \mathbb{R}$. Define, for $t \in [0, 1]$
\[
V(t) = \sup_{x \in X} f(x, t),
\]
\[
X^*(t) = \{ x \in X \mid f(x, t) = V(t) \}.
\]

\begin{proposition}[Theorem 1 from [49]]
Let $t \in [0, 1]$ such that $X^*(t) \neq \emptyset$. Let $x^* \in X^*(t)$ and suppose that $f(x^*, \cdot)$ is differentiable at $t$, with derivative $f_t(x^*, t)$.

- If $t > 0$ and if $V$ is left-hand differentiable at $t$, then $V'(t^-) \leq f_t(x^*, t)$.
- If $t < 0$ and if $V$ is right-hand differentiable at $t$, then $V'(t^+) \geq f_t(x^*, t)$.
- If $t \in (0, 1)$ and if $V$ is differentiable at $t$, then $V'(t) = f_t(x^*, t)$.
\end{proposition}

\begin{proposition}[Corollary 4 from [49]]
Suppose that $X$ is nonempty and compact. Suppose that for all $t \in [0, 1], f(\cdot, t)$ is continuous. Suppose also that $f$ admits a partial derivative $f_t$ with respect to $t$ that is continuous in $(x, t)$ over $X \times [0, 1]$. Then

- $V'(t^+) = \max_{x^* \in X^*(t)} f_t(x^*, t)$ for all $t \in [0, 1]$ and $V'(t^-) = \min_{x^* \in X^*(t)} f_t(x^*, t)$ for all $t \in (0, 1]$.

- $V$ is differentiable at $t \in (0, 1)$ is and only if $\left\{ f_t(x^*, t) \mid x^* \in X^*(t) \right\}$ is a singleton. In that case $V'(t) = f_t(x^*, t)$ for all $x^* \in X^*(t)$.
\end{proposition}
[1] Michael Aizenman, Robert Sims, and Shannon L Starr. Extended variational principle for the sherrington-kirkpatrick spin-glass model. *Physical Review B*, 68(21):214403, 2003.

[2] Ahmed El Alaoui and Michael I Jordan. Detection limits in the high-dimensional spiked rectangular model. *arXiv preprint arXiv:1802.07309*, 2018.

[3] Ahmed El Alaoui and Florent Krzakala. Estimation in the spiked wigner model: A short proof of the replica formula. *arXiv preprint arXiv:1801.01593*, 2018.

[4] Ahmed El Alaoui, Florent Krzakala, and Michael I Jordan. Finite size corrections and likelihood ratio fluctuations in the spiked wigner model. *arXiv preprint arXiv:1710.02903*, 2017.

[5] Arash A Amini and Martin J Wainwright. High-dimensional analysis of semidefinite relaxations for sparse principal components. In *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pages 2454–2458. IEEE, 2008.

[6] Fabrizio Antenucci, Silvio Franz, Pierfrancesco Urbani, and Lenka Zdeborová. On the glassy nature of the hard phase in inference problems. *arXiv preprint arXiv:1805.05857*, 2018.

[7] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697, 2005.

[8] Jinho Baik and Jack W Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.

[9] Jess Banks, Cristopher Moore, Roman Vershynin, Nicolas Verzelen, and Jiaming Xu. Information-theoretic bounds and phase transitions in clustering, sparse pca, and submatrix localization. In *Information Theory (ISIT), 2017 IEEE International Symposium on*, pages 1137–1141. IEEE, 2017.

[10] Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, and Lenka Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems*, pages 424–432, 2016.

[11] Jean Barbier and Nicolas Macris. The stochastic interpolation method: A simple scheme to prove replica formulas in bayesian inference. *arXiv preprint arXiv:1705.02780*, 2017.

[12] Jean Barbier, Nicolas Macris, and Léo Miolane. The layered structure of tensor estimation and its mutual information. *arXiv preprint arXiv:1709.10368*, 2017.

[13] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

[14] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.
Florent Benaych-Georges and Raj Rao Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012.

Quentin Berthet, Philippe Rigollet, et al. Optimal detection of sparse principal components in high dimension. *The Annals of Statistics*, 41(4):1780–1815, 2013.

Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

Francesco Caltagirone, Marc Lelarge, and Léo Miolane. Recovering asymmetric communities in the stochastic block model. *IEEE Transactions on Network Science and Engineering*, 2017.

Mireille Capitaine, Catherine Donati-Martin, and Delphine Féral. The largest eigenvalues of finite rank deformation of large wigner matrices: convergence and nonuniversality of the fluctuations. *The Annals of Probability*, pages 1–47, 2009.

Wei-Kuo Chen. Phase transition in the spiked random tensor with rademacher prior. *arXiv preprint arXiv:1712.01777*, 2017.

Alexandre d’Aspremont, Laurent E Ghaoui, Michael I Jordan, and Gert R Lanckriet. A direct formulation for sparse pca using semidefinite programming. In *Advances in neural information processing systems*, pages 41–48, 2005.

Bernard Derrida. Random-energy model: Limit of a family of disordered models. *Physical Review Letters*, 45(2):79, 1980.

Bernard Derrida. Random-energy model: An exactly solvable model of disordered systems. *Physical Review B*, 24(5):2613, 1981.

Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the balanced binary stochastic block model. *Information and Inference: A Journal of the IMA*, 6(2):125–170, 2016.

Yash Deshpande and Andrea Montanari. Information-theoretically optimal sparse pca. In 2014 *IEEE International Symposium on Information Theory*, pages 2197–2201. IEEE, 2014.

Yash Deshpande and Andrea Montanari. Sparse pca via covariance thresholding. In *Advances in Neural Information Processing Systems*, pages 334–342, 2014.

Edgar Dobriban et al. Sharp detection in pca under correlations: all eigenvalues matter. *The Annals of Statistics*, 45(4):1810–1833, 2017.

Edgar Dobriban, William Leeb, and Amit Singer. Pca from noisy, linearly reduced data: the diagonal case. *arXiv preprint arXiv:1611.10333*, 2016.

David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

Delphine Féral and Sandrine Péché. The largest eigenvalue of rank one deformation of large wigner matrices. *Communications in mathematical physics*, 272(1):185–228, 2007.

Stefano Ghirlanda and Francesco Guerra. General properties of overlap probability distributions in disordered spin systems. towards parisi ultrametricity. *Journal of Physics A: Mathematical and General*, 31(46):9149, 1998.

Francesco Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Communications in mathematical physics*, 233(1):1–12, 2003.
[33] 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.

[34] Dongning Guo, Yihong Wu, Shlomo S Shitz, and Sergio Verdú. Estimation in gaussian noise: Properties of the minimum mean-square error. *IEEE Transactions on Information Theory*, 57(4):2371–2385, 2011.

[35] David C Hoyle and Magnus Rattray. Statistical mechanics of learning multiple orthogonal signals: asymptotic theory and fluctuation effects. *Physical review E*, 75(1):016101, 2007.

[36] Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference*, page iat004, 2013.

[37] Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, pages 295–327, 2001.

[38] Iain M Johnstone and Arthur Yu Lu. Sparse principal components analysis. *Unpublished manuscript*, 7, 2004.

[39] Satish Babu Korada and Nicolas Macris. Exact solution of the gauge symmetric p-spin glass model on a complete graph. *Journal of Statistical Physics*, 136(2):205–230, 2009.

[40] Satish Babu Korada and Nicolas Macris. Tight bounds on the capacity of binary input random cdma systems. *IEEE Transactions on Information Theory*, 56(11):5590–5613, 2010.

[41] Florent Krzakala, Jiaming Xu, and Lenka Zdeborová. Mutual information in rank-one matrix estimation. In *Information Theory Workshop (ITW), 2016 IEEE*, pages 71–75. IEEE, 2016.

[42] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, Apr 2018.

[43] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel. In 53rd *Annual Allerton Conference on Communication, Control, and Computing, Allerton 2015, Allerton Park & Retreat Center, Monticello, IL, USA, September 29 - October 2, 2015*, pages 680–687, 2015.

[44] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Phase transitions in sparse PCA. In *IEEE International Symposium on Information Theory, ISIT 2015, Hong Kong, China, June 14-19, 2015*, pages 1635–1639. IEEE, 2015.

[45] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Constrained low-rank matrix estimation: phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(7):073403, 2017.

[46] Thibault Lesieur, Léo Miolane, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. 2017 IEEE *International Symposium on Information Theory (ISIT)*, pages 511–515, 2017.

[47] Ryosuke Matsushita and Toshiyuki Tanaka. Low-rank matrix reconstruction and clustering via approximate message passing. In *Advances in Neural Information Processing Systems*, pages 917–925, 2013.

[48] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Co Inc, 1987.

[49] Paul Milgrom and Ilya Segal. Envelope theorems for arbitrary choice sets. *Econometrica*, 70(2):583–601, 2002.
[50] Léon Miolane. Fundamental limits of low-rank matrix estimation: the non-symmetric case. *arXiv preprint arXiv:1702.00473*, 2017.

[51] Andrea Montanari. Estimating random variables from random sparse observations. *European Transactions on Telecommunications*, 19(4):385–403, 2008.

[52] Andrea Montanari and Emile Richard. Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Transactions on Information Theory*, 62(3):1458–1484, 2016.

[53] Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. *arXiv preprint arXiv:1711.01682*, 2017.

[54] Boaz Nadler. Finite sample approximation results for principal component analysis: A matrix perturbation approach. *The Annals of Statistics*, pages 2791–2817, 2008.

[55] Alexei Onatski, Marcelo J Moreira, Marc Hallin, et al. Asymptotic power of sphericity tests for high-dimensional data. *The Annals of Statistics*, 41(3):1204–1231, 2013.

[56] Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer Science & Business Media, 2013.

[57] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

[58] Sandrine Péché. The largest eigenvalue of small rank perturbations of hermitian random matrices. *Probability Theory and Related Fields*, 134(1):127–173, 2006.

[59] Amelia Perry, Alexander S Wein, and Afonso S Bandeira. Statistical limits of spiked tensor models. *arXiv preprint arXiv:1612.07728*, 2016.

[60] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Optimality and sub-optimality of pca for spiked random matrices and synchronization. *arXiv preprint arXiv:1609.05573*, 2016.

[61] Sundeep Rangan and Alyson K Fletcher. Iterative estimation of constrained rank-one matrices in noise. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 1246–1250. IEEE, 2012.

[62] Emile Richard and Andrea Montanari. A statistical model for tensor pca. In *Advances in Neural Information Processing Systems*, pages 2897–2905, 2014.

[63] Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science & Business Media, 2010.

[64] Michel Talagrand. *Mean field models for spin glasses: Volume II: Advanced Replica-Symmetry and Low Temperature*, volume 55. Springer Science & Business Media, 2011.

[65] David J Thouless, Philip W Anderson, and Robert G Palmer. Solution of ‘solvable model of a spin glass’. *Philosophical Magazine*, 35(3):593–601, 1977.

[66] Yihong Wu and Sergio Verdú. Functional properties of minimum mean-square error and mutual information. *IEEE Transactions on Information Theory*, 58(3):1289–1301, 2012.

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

[68] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2):265–286, 2006.