Title

Detection limits and fluctuation results in some spiked random matrix models and pooling of discrete data

Author

El Alaoui, Ahmed

Publication Date

2018

Peer reviewed|Thesis/dissertation

Detection limits and fluctuation results in some spiked random matrix models and pooling of discrete data

by

Ahmed El Alaoui El Abidi

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering - Electrical Engineering & Computer Sciences

and the Designated Emphasis

in

Communication, Computation and Statistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael I. Jordan, Chair
Professor Martin J. Wainwright
Professor Nike Sun

Spring 2018

Detection limits and fluctuation results in some spiked random matrix models and pooling of discrete data

Abstract

Detection limits and fluctuation results in some spiked random matrix models and pooling of discrete data

by

Ahmed El Alaoui El Abidi

Doctor of Philosophy in Engineering - Electrical Engineering & Computer Sciences

with Designated Emphasis in Communication, Computation and Statistics

University of California, Berkeley

Professor Michael I. Jordan, Chair

In this thesis we examine the fundamental limits of detecting and recovering a weak signal hidden in a large amount of noise. We will consider two problems. The first one pertains to *principal component analysis*, and is about estimating and detecting the presence of a structured low-rank signal buried inside a large noise matrix. Prominent models of this problem include the so-called *spiked* or *deformed ensembles* from random matrix theory, which are distributions over matrices of the form “signal + noise.” It is known in these models that the top eigenpair of the data matrix becomes indicative of the presence of this signal, or ”spike”, when and only when its strength is above a certain ”spectral” threshold. A natural question is then whether it is possible to identify the spike, or even tell if it’s really present in the data below this threshold. In the first part of this thesis, we will completely characterize the fundamental limits of the detection and estimation of the spike. The analysis leading to this characterization crucially relies on a recently discovered connection of this statistical problem with the mean-field theory of spin glasses. This connection provides the necessary tools to obtain precise control on the behavior of the posterior distribution of the spike as well as the fluctuations of the associated likelihood ratio process.

The second problem we consider is that of *pooling*, and is about recovering a discrete signal, i.e., whose entries only take can a finite number of values, given several observation that are constrained in a combinatorial way. More precisely, in a manner akin to compressed sensing, observations of this signal can be obtained in the form of histograms: pooled measurements counting the occurrence of each symbol across subsets of the variables. We ask what is the minimal number of random measurements of this sort it takes to reconstruct the signal. In the second part of this thesis, we determine sharp upper and lower bounds on the minimal number of measurements for the signal to be essentially unique, thus in principle recoverable from the observed data. We then provide an efficient algorithm to extract it, and show that this strategy is successful only when the number of measurements is much larger than this uniqueness threshold.

To my parents, Mostafa and Mouniya.
To the memory of my grandmother, Halima.

Contents

Acknowledgments

I have been fortunate enough to spend my years of graduate school on the Berkeley campus. Aside from an almost permanent nice weather which is characteristic of the Bay Area, the academic environment is unlike any other; scientific enquiry, and in particular mathematical research, “is in the air” in Berkeley.

I am extremely grateful to my advisor Michael Jordan for guiding me through the ups and downs of graduate school, and most of all, for granting me an enormous intellectual freedom to venture into what seemed at the outset to be quite remote areas. I also thank him for his availability and patience with my ramblings on the white board during every one of our meetings. I have had the pleasure to interact with Martin Wainwright, who has in many respects been a surrogate advisor to me. I learned from Martin the fundamentals of mathematical writing and how to conduct mathematical research. His teaching style, both in class and during research is an example of pedagogy to follow.

I have had the immense fortune to meet Florent Krzakala and Lenka Zdeborová during their semester-long visit to the Simons institute in spring 2016 for the program on counting complexity and phase transitions. I was then in my third year of graduate school and looking for a new research project. Aaditya Ramdas, then a postdoc with Mike and Martin, introduced me to them, and we then quickly started working on what now constitutes the last two chapters of this thesis. This encounter gained me two exceptional—hopefully long term—collaborators and wonderful friends. I am constantly challenged by the breath of their knowledge and the depth of their intuition. Our interactions have heavily influenced my mathematical taste, and will undoubtedly shape the style of my research for the years to come. I am also very thankful to Nike Sun for granting me some of her time to occasionally tell her about my research, and for agreeing to read my thesis. Her constructive feedback has been of great help.

The work leading to this thesis did not happen in the void. I was constantly surrounded with supportive and loving friends and officemates, without whom I would have probably finished this doctorate in one way or another, but the experience would not have been nearly as enjoyable. My special thanks to Aaditya Ramdas, Dimitris Papailiopoulos and Mahdi Soltanolkotabi for their mentoring of me during my first few years, to my officemates in the fourth and fifth floors of Soda Hall: Ross Boczar, Nick Boyd, Sarah Dean, Orianna DeMasi, Nicolas Flammarion, Kevin Jamieson, Lydia Liu, Horia Mania, Philipp Moritz, Alyssa Morrow, Robert Nishihara, Max Rabinovich, Becca Roelofs, Esther Rolf, Vaishaal Shankar, Max Simchowitz, Nilesh Tripuraneni, Stephen Tu, Ashia Wilson and Tijana Zrnic, and to my dear friends Karina Cucchi, Albert Yuen, Benjamin Fildier, Justine Rey and Claire Boët for the memorable moments we shared.

Finally, my thoughts go to my parents Mostafa and Mouniya, my brothers Simo and Taha, my grandmothers, my aunts and uncles in Morocco for their constant support and complaint that I don’t call neither visit them often enough. On this last item, I am totally to blame...

Chapter 1

Introduction

The general theme of this thesis is to investigate the fundamental limits of detecting the presence of a structured signal hidden in a large amount of noise, and when the signal is indeed present, the extend to which it could be reliably estimated. These questions are relevant in a modern context of data analysis in which large amounts of data are gathered in the experimental sciences and industry, and where the Scientist or Engineer wishes to test increasingly complex hypotheses about what this data might entail, or glean information about the faintest signal in it, preferably in a time-efficient way. While the amount of data in one's possession is large, so is the number of parameters or degrees of freedom one wishes to control or estimate. Additionally, in the presence of noise or corruptions, it may be very difficult to extract relevant information, or to even tell if a signal is really there.

We will focus on simple models in which both the signal and the noise that corrupts it have a certain structure, and for which the above detection and estimation problems admit sharp characterizations as the dimension of the problem grows unbounded; this is the high-dimensional, or "big data" regime. The main contribution of this thesis is to develop and deploy the necessary theoretical tools to prove these sharp characterizations in two different settings:

1. The first setting is that of *principal component analysis* (PCA). The goal is roughly as follows: given a set of data points corrupted with noise and living in a high-dimensional Euclidean space $\mathbb{R}^d$, to find out whether there exists a distinguished direction in space along which these data points partially align, or whether these data points are scattered in all directions is a relatively uniform way. Additionally if such a distinguished direction is present, one would like to identify it to the best possible accuracy. We will consider specific models of this problem; the so-called *spiked* or *deformed* ensemble of random matrices in which a signal of low rank structure—or *the spike*—representing the distinguished direction is drowned in a large noise matrix. We provide an almost complete characterization of the limits of detecting and estimating this spike in these models.

2. The second setting is more discrete, and concerns the problem of *pooling.* Consider a

> discrete high-dimensional signal consisting of categorical variables, i.e., the variables can only take a finite number of values. (For example, the blood type of a human, or nucleotides in a string of DNA.) In a manner akin to compressed sensing, observations of this signal can be obtained in the form of "histograms" or "frequency spectra"—pooled measurements counting the occurrence of each category or type across subsets of the variables. In a more concrete way, consider a population of n individuals where each individual belongs to one category among d. An observer repeatedly selects a subset of individuals, computes the histogram of their types (i.e., number of occurrences of each category in that subset), then reveals this histogram along with the individuals in that subset. This gives rise to the inferential problem of determining the category of every individual in the population. We provide tight upper and lower bounds on the minimal number of observations needed for recovery, and ascertain whether this inferential problem can be solved in an efficient manner.

The two problems discussed above share a common characteristic that is worth noting: there exists a certain regime of parameters (strength of the noise, number of samples,...) in which extracting the signal becomes *information-theoretically* possible, but *computationally* challenging. In other words, the data at hand is of sufficiently good quality to at least partially extract some signal, however, all known *efficient* algorithms fail to extract it. In the first problem, one may attempt to test and/or estimate based on the spectrum of the observed matrix. The performance of these spectral tests/estimators has been throughly studied in statistics and random matrix theory and is fairly well understood in the most common situations. We will see that there are situations where it is possible to reliably estimate the spike while the spectrum captures no information about it. The same situation occurs in the second problem. We will provide an efficient algorithm that is only able to recover the signal long after it is information-theoretically identifiable.

1.1 Hypothesis testing and estimation

In this section we put the notions of a test, an estimator and "best" accuracy loosely discussed above on a formal mathematical ground.

The detection problem

The setting we adopt here is that of binary hypothesis testing (Keener, 2011). Let $(\Omega, \mathcal{F}, \rho)$ be a probability space, $(\mathcal{X}, \mathcal{B})$ be a topological space endowed with its Borel σ-algebra $\mathcal{B}$, and Θ a set of "parameters". One observes a random variable $X : \Omega \mapsto \mathcal{X}$ whose law is P_θ where $\theta \in \Theta$, and would like to distinguish between the following two hypotheses:

$$H_0 : \theta = 0 \quad \text{v.s.} \quad H_1 : \theta = \bar{\theta},$$

for a fixed $\bar{\theta} \neq 0$. For the concrete random matrix problems we look at in this thesis, the parameter θ will be real valued and represents the strength of the distinguished direction, or spike, of interest. More precisely, one would like to construct a measurable function (or test) $T : \mathcal{X} \mapsto \{0, 1\}$ that returns "0" for H_0 and "1" for H_1, such that the mis-classification error

$$\mathsf{err}(T) := P_{\bar{\theta}}(T(X) = 0) + P_0(T(X) = 1) \tag{1.1}$$

is minimized among all possible tests T.

In order to make (non)asymptotic statistical statements it is useful to consider not only one such problem, but a sequence of problems indexed by an integer n. This could model the accumulation of data and/or the growth of the number or dimension of parameters θ. In this case, and in addition to the above criterion err, one could also consider a more stringent, asymptotic definition of figure of merit. Namely that a sequence of tests (T_n) must satisfy

$$\lim_{n\to\infty} P_{n,\bar{\theta}}(T_n(X_n) = 0) \vee P_{n,0}(T_n(X_n) = 1) = 0. \tag{1.2}$$

We have made the dependence of P_θ on n explicit. The reader should interpret the above statement as follows: as the amount of data grows, we would like be more and more confident in our guess of where the data came from. Throughout, we refer to the question of existence of a sequence of tests that answers to the requirement (1.2) as the *strong detection* problem, and the question of minimizing the criterion (1.1) as *weak detection*.

Strong detection We would like to understand for what values of $\bar{\theta}$ is strong detection possible. To fix some intuition, let us think of $\bar{\theta}$ as a continuous real-valued parameter that represents the strength of the signal we want to detect. A large $\bar{\theta}$ means an easier detection problem while a small $\bar{\theta}$ means a harder one. We are interested in the smallest value of this parameter such that strong detection is still possible. To this end, Le Cam (1960) defined the notion of *contiguity* between two sequences of probability measures (see, e.g., Van der Vaart, 2000).

Definition 1.1 (Contiguity)**.** *Let (P_n) and (Q_n) be two sequences of probability measures defined on the same sequence of measurable spaces $(\Omega_n, \mathcal{F}_n)$. We will say that (P_n) is contiguous to (Q_n) if $Q_n(A_n) \to 0$ implies that $P_n(A_n) \to 0$ as $n \to \infty$ for every sequence of measurable sets (A_n), $A_n \in \mathcal{F}_n$. We will say that (P_n) and (Q_n) are* mutually *contiguous if (P_n) is contiguous to (Q_n) and vice versa.*

In our case, $P_n = P_{n,\bar{\theta}}$ and $Q_n = P_{n,0}$. It is easy to see that contiguity implies impossibility of strong detection since for instance, if (P_n) is contiguous to (Q_n), then $Q_n(T(X_n) = 1) \to 0$ implies that $P_n(T(X_n) = 0) \to 1$. The most straightforward way to prove contiguity of two sequences of probability measures is via the *second moment method*. We briefly sketch the idea of this method. Assume that P_n is absolutely continuous with respect to Q_n for every n sufficiently large (otherwise contiguity cannot hold). Now define the likelihood ratio or Radon-Nikodym derivative of P_n with respect to Q_n: $L_n \equiv \mathrm{d}P_n/\mathrm{d}Q_n$. For any event $A \in \mathcal{F}_n$,

$$P_n(A) = \int \mathbb{1}_A(\omega)\mathrm{d}P_n(\omega) = \int \mathbb{1}_A(\omega)L_n(\omega)\mathrm{d}Q_n(\omega)$$

$$\leq \sqrt{\int L_n^2 \mathrm{d}Q_n} \cdot \sqrt{\int \mathbb{1}_A \mathrm{d}Q_n} = \sqrt{\mathbb{E}_{Q_n}[L_n^2]} \cdot \sqrt{Q_n(A)}.$$

where the inequality is by Cauchy-Schwarz. Therefore, it suffices to show that

$$\limsup \mathbb{E}_{Q_n}[L_n^2] < \infty, \tag{1.3}$$

in order to show contiguity of (P_n) w.r.t. (Q_n). This is a method of choice for proving impossibility of strong detection in many statistical problems, well beyond what is studied in this thesis (see for example Addario-Berry et al., 2010; Arias-Castro and Verzelen, 2014; Ingster and Suslina, 2012; Verzelen and Arias-Castro, 2015). Its major appeal lies in its simplicity since it only requires to compute the expected square of the likelihood ratio under the null model Q_n. Its drawback on the other hand is that condition (1.3) is in general far from necessary. Two sequences can very well be (mutually) contiguous while their likelihood ratio has an infinite second moment. This problem is alleviated, at least partially, by a conditioning argument: in many cases, the divergence of the second moment in (1.3) can be attributed to the existence of a rare event B_n whose contribution to the overall expectation $\mathbb{E}[L_n^2]$ is comparatively very large so as to make L_n^2 ill-behaved. Then one can simply condition away this bad event for a better control of the second moment of a modified likelihood ratio $\tilde{L}_n = \mathrm{d}\tilde{P}_n/\mathrm{d}Q_n$, where $\tilde{P}_n = P_n(\cdot|\bar{B}_n)$. By rarity of B_n, i.e., $P_n(B_n) \to 0$, one is again able to conclude. However, identifying the right event to condition on can be difficult, and is a matter of a case-by-case deliberation.

A different approach which we take in this thesis is to understand how the likelihood ratio L_n is asymptotically distributed, rather than just controlling its second moment. Then the connection to contiguity is given by Le Cam's first lemma, which we state here.

Lemma 1.2 (Le Cam's first lemma)**.** *The sequence (P_n) is contiguous to (Q_n) if and only if $L_n \rightsquigarrow V$ under Q_n (possibly along a subsequence) implies that $\mathbb{E}[V] = 1$.*

The symbol " $\rightsquigarrow$ " denotes convergence in distribution as $n \to \infty$, and the phrase "under Q_n" means that L_n is seen as a random variable on the probability space $(\Omega_n, \mathcal{F}_n, Q_n)$. Taken together, "$L_n \rightsquigarrow V$ under Q_n" means $\int f(L_n)\mathrm{d}Q_n \to \int f(V)\mathrm{d}\rho$ for every bounded continuous function $f : \mathbb{R} \mapsto \mathbb{R}$. ($\rho$ being the fixed background probability measure.) A notable special case of this lemma is that of asymptotic log-normality, where $\log L_n \rightsquigarrow \mathcal{N}(\mu, \sigma^2)$ under Q_n. In this case, $V = e^Z$, where $Z \sim \mathcal{N}(\mu, \sigma^2)$, so $\mathbb{E}[V] = 1$ if and only if $\mu = -\frac{1}{2}\sigma^2$. The above lemma provides an exact characterization of contiguity in terms of the properties of the asymptotic distributional limit of L_n.

Weak detection. When contiguity holds and testing errors are inevitable, it is natural to weaken one's requirements, and ask to test with accuracy *better than that of a random guess*[1]. More precisely, by the classical Neyman-Pearson lemma (Keener, 2011), the optimal

[1] Strictly speaking, if one only wants to beat random guessing, it is enough to find a sequence of tests that meets the criterion $\limsup \mathsf{err}_n(T_n) < 1$. But we will be interested in tests that minimize err_n.

test minimizing the risk (1.1) is the likelihood ratio test which rejects the null hypothesis H_0 (i.e., returns "1") if $L_n > 1$, and its Type-I and Type-II errors are $Q_n(L_n \geq 1)$ and $P_n(L_n < 1)$ respectively. Once more, if one is able obtain the precise asymptotic distribution of L_n under Q_n (or P_n) then one is able to characterize the performance of the best test, hence determine the fundamental limits of detection.

The estimation problem

The estimation problem can be phrased as follows: given a random variable $X : \Omega \mapsto \mathcal{X}$ defined on $(\Omega, \mathcal{F}, \rho)$ and whose law is P_θ, the goal is to estimate the parameter θ based on the sample X. Here we further assume that the space of parameters Θ is endowed with a metric d, and the quality of an estimator $\widehat{\theta} : \mathcal{X} \mapsto \Theta$ can be measured in terms of its squared distance from θ:

$$\mathbb{E}\left[d(\widehat{\theta}(X), \theta)^2\right].$$

For our purposes, Θ will be the Euclidean space $\mathbb{R}^p$ for some p and we will simply take d to be the ℓ_2 distance. Additionally, if we let $\theta : \Omega \mapsto \Theta$ be a random variable with "prior" distribution μ, then we can explicitly determine the best estimator with respect to the above expected ℓ_2 distance: this is just the posterior mean of θ given X:

$$\widehat{\theta}(X) = \mathbb{E}[\theta|X], \tag{1.4}$$

and the *minimal mean squared error* (MMSE) is

$$\mathbb{E}\left[\left\|\widehat{\theta}(X) - \theta\right\|_{\ell_2}^2\right] = \mathbb{E}\left[\|\theta\|_{\ell_2}^2\right] - \mathbb{E}\left[\left\|\,\mathbb{E}\left[\theta|X\right]\,\right\|_{\ell_2}^2\right]. \tag{1.5}$$

Observe that a simplistic strategy would be to predict θ by its prior mean $\mathbb{E}[\theta]$ without looking at the data X, in which case one incurs an error of $\mathbb{E}\left[\left\|\theta - \mathbb{E}[\theta]\right\|_{\ell_2}^2\right] = \mathrm{var}(\theta)$. Now given a specific statistical model $\theta \sim \mu$, $X \sim P_\theta$, we will pay a particular attention to the following questions:

- What is the value of its MMSE?
- When is it strictly smaller than $\mathrm{var}(\theta)$?
- Can the posterior mean be computed exactly or approximated efficiently?

1.2 Spiked models of random matrices

Now we introduce the specific spiked random matrix models to which the above framework will be applied in the first part of this thesis.

Spiked models are distributions over matrices of the form "signal + noise". They have been a mainstay in the statistical literature since their introduction by Johnstone (2001) as mathematically rich models for the study of high-dimensional principal component analysis (PCA). Their introduction has provided the foundations for a rich theory of PCA, in which the performance of several important tests and estimators is by now well understood (see, e.g., Amini and Wainwright, 2009; Berthet and Rigollet, 2013; Dobriban, 2017; Johnstone and Lu, 2009; Ledoit and Wolf, 2002; Nadler, 2008; Paul, 2007).

In the first three chapters of this thesis, we focus on the following two particular models: The first is the so-called *spiked Wigner model* where one observes a symmetric $N \times N$ matrix of the form

$$\boldsymbol{Y} = \sqrt{\frac{\lambda}{N}} \boldsymbol{x}\boldsymbol{x}^\top + \boldsymbol{W}, \tag{1.6}$$

where $\boldsymbol{x}$ is the spike which represents the direction of interest, $\boldsymbol{W}$ is symmetric noise matrix with entries which we assume Gaussian and independent, up to symmetry. The parameter λ is the strength of the spike and plays the role of a signal-to-noise ratio.

The second model is an asymmetric, rectangular version of the first, where one observes a $N \times M$ matrix

$$\boldsymbol{Y} = \sqrt{\frac{\lambda}{N}} \boldsymbol{u}\boldsymbol{v}^\top + \boldsymbol{W}, \tag{1.7}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ correspond to a low rank factor to be recovered and $\boldsymbol{W}$ is again a Gaussian noise matrix. In particular the model introduced by Johnstone (and generalizations of it) corresponds to the special case $v_j \sim \mathcal{N}(0,1)$. One can see in this case that the M column vectors $\boldsymbol{y}_j \in \mathbb{R}^N$, $1 \leq j \leq M$ of $\boldsymbol{Y}$ are drawn i.i.d. from a centered normal distribution with spiked covariance: $\mathcal{N}(\boldsymbol{0}, \boldsymbol{I} + \frac{\lambda}{N}\boldsymbol{u}\boldsymbol{u}^\top)$. For this reason, model (1.7) and its close relatives are usually referred to as *spiked covariance* models.

The above asymmetric model has more degrees of freedom, namely two independent factors $\boldsymbol{u}$ and $\boldsymbol{v}$, and a additional parameter compared to the symmetric model, which is the aspect ratio M/N of the matrix $\boldsymbol{Y}$.

In this case, we will assume a high-dimensional setting where both M and N grow to infinity while their ratio converges to a finit value α. In both models, the low rank factors are assumed to have independent coordinates drawn from fixed priors on $\mathbb{R}$.

It is known that the spectral norm of the noise matrix $\boldsymbol{W}$ is of order $\sqrt{N}$ as its dimensions grow to infinity (see e.g., Bai and Silverstein, 2010; Tao, 2012). Therefore, scaling the rank-one component by $\sqrt{N}$ puts the magnitudes of the “signal” and “noise” components of the matrix $\boldsymbol{Y}$ on the same scale, and λ allows to control their relative strengths. As we review next, the model is in a *critical regime* where its properties undergo sharp “phase transitions” as λ crosses some finite thresholds.

In these models,

what are the fundamental limits of detection and estimation of the spike?

More precisely,

- for what values of α, λ is strong/weak detection possible?

- what is the performance of the likelihood ratio test?
- what is the performance of the posterior mean in estimating the spike?

Before embarking on the analysis of these questions it is reasonable to first look at tests and estimators that can be easily constructed from the eigenvalues and eigenvectors of $\boldsymbol{Y}$, and understand their performance.

Spectral properties

The spectral properties of these models have been extensively studied, in particular in random matrix theory, where they are known as *deformed ensembles* (Péché, 2014). Landmark investigations in this area have unveiled the existence of sharp transition phenomena in the behavior of the spectrum of the data matrix, where for a spike of strength λ above a certain *spectral* threshold, the top eigenvalue separates from the remaining eigenvalues which are packed together in a "bulk" and thus indicates the presence of the spike. However, below this threshold, the top eigenvalue converges to the edge of the bulk and becomes non-informative about the presence of the spike.

For instance, if the entries of the spike $\boldsymbol{x}$ in model (1.6) are assumed to have unit variance then the value of this spectral threshold is $\lambda = 1$. Similarly, if $\boldsymbol{u}$ and $\boldsymbol{v}$ are independent with entries of unit variance in model (1.7), the spectral threshold, known as the BBP threshold, after Baik, Ben Arous, and Péché (2005) is given by $\alpha\lambda^2 = 1$. Estimation using the top eigenvector undergoes the same transition, where it is known to "lose track" of the spike below the spectral threshold. Moreover, above the spectral threshold, the quality of the overlap of the first eigenvector of the matrix $\boldsymbol{Y}$ with the spike ($\boldsymbol{x}$, $\boldsymbol{u}$ or $\boldsymbol{v}$) is well understood. For more precise results in this direction, we refer to Benaych-Georges and Nadakuditi (2011, 2012); Capitaine, Donati-Martin, and Féral (2009); Féral and Péché (2007); Péché (2006) for results on low-rank deformations of Wigner matrices, and Bai and Yao (2008); Baik, Ben Arous, and Péché (2005); Baik and Silverstein (2006); Johnstone and Lu (2009); Nadler (2008); Paul (2007) for results on spiked covariance models.

1.3 The posterior distribution and spin glasses

These spectral analyses have provided many insights, but they stop short of characterizing the fundamental limits of estimating the spike, or detecting its presence from the observation of a sample matrix. These questions, information-theoretic and statistical in nature, are more naturally approached by looking the posterior law of spike given $\boldsymbol{Y}$ and the associated likelihood ratio process.

The fundamental observation that allows to make progress on these questions is that the posterior distribution of the spike $\boldsymbol{x}$ given $\boldsymbol{Y}$ (taking model (1.6) as an example) is a high-dimensional probability distribution, exactly of the type that has been studied in the statistical physics literature, and later in probability theory under the name of *"the mean-field theory of spin glasses"*. This theory deals with the study of disordered systems of interacting particles: certain systems can be modeled by an energy function that describes certain rules of local interaction between particles. This gives rise to a (random) probability distribution of states of the system (a very high-dimensional configuration space) according to their energy. Let us avoid a further (and almost-certainly inaccurate) general discussion of what spin glass theory is, and refer to Bolthausen and Bovier (2007) for a review on the state of the art of this field (up to 2007). We instead discuss one of its most famous objects of study: the *Sherrington-Kirkpatrick (SK) model*. Let $\boldsymbol{\sigma} \in \{-1,+1\}^N$ represent the "spin" of N particles, which is allowed to be "up" or "down", and that interact in the pairwise fashion according to the energy function or *Hamiltonian*:

$$-H(\boldsymbol{\sigma}) = \frac{1}{\sqrt{N}} \sum_{i<j} g_{ij}\sigma_i\sigma_j. \tag{1.8}$$

The numbers g_{ij} are the "coupling constants" of the interaction between the particles i and j. For a positive parameter β, define the following probability distribution on the N-dimensional hypercube

$$P(\boldsymbol{\sigma}) = \frac{e^{-\beta H(\boldsymbol{\sigma})}}{\sum_{\boldsymbol{\sigma}\in\{\pm 1\}^N} e^{-\beta H(\boldsymbol{\sigma})}}. \tag{1.9}$$

This probability distribution, usually called a *Gibbs measure*, is already interesting if the g_{ij}'s are deterministic constants. If the latter are all equal to $+1$ this gives rise to the *Curie-Weiss model*. More generally, the coupling constants could represent the adjacency structure of a graph on N vertices, in which case the model is known as the *Ising model*. Sherrington and Kirkpatrick (1975) proposed to make the coupling constants random ± 1 or Gaussian independently to model a *disordered*, *conflicting*, or *frustrated* interaction where a particle cannot align with one of its neighbors without paying an energy cost for being mis-aligned with other neighbors. This is gives rise to a random probability distribution whose structure is multiscale and extremely complex. For instance one characteristic quantity of this model is its *free energy*, defined as the logarithm of the denominator in (1.9) (called the partition function and denoted by Z):

$$F_N(\beta) = -\frac{1}{\beta N} \log \sum_{\boldsymbol{\sigma}\in\{\pm 1\}^N} e^{-\beta H(\boldsymbol{\sigma})}. \tag{1.10}$$

One question of interest if whether this quantity has a limit when $N \to \infty$, while β is kept fixed. Standard theorems of concentration of measure—in particular, the Tsirelson-Ibragimov-Sudakov inequality in the case where the "disorder" variables g_{ij} are Gaussian, (see Boucheron, Lugosi, and Massart, 2013)—imply that F_N concentrates very tightly about its expectation, so the above question reduces to the study of the expectation of F_N under the disorder. A precise formula for a limit of this quantity was conjectured by G. Parisi in the late 70's (see Mézard, Parisi, Sourlas, et al., 1984). The problem of understanding the SK model and confirming Parisi's conjecture started attracting mathematicians with the seminal works of Aizenman, Lebowitz, and Ruelle (1987) and Pastur and Shcherbina (1991), and has since spawned an entire field in probability theory.

What does all this have to do with our problem of detecting the spike in a random matrix model? We can see that if the entries of the spike $\boldsymbol{x}$ come i.i.d. from a prior distribution P_{x}, the posterior distribution of $\boldsymbol{x}$ given $\boldsymbol{Y}$ is

$$\mathrm{d}P(\boldsymbol{x}|\boldsymbol{Y}) = \frac{e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})}{\int e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})}, \tag{1.11}$$

where

$$-H(\boldsymbol{x}) = \sum_{i\leq j} \sqrt{\frac{\lambda}{N}} Y_{ij} x_j x_j - \frac{\lambda}{2N} x_i^2 x_j^2. \tag{1.12}$$

There is a clear similarity between the Hamiltonians (1.8) and (1.12). The theory of the SK model becomes all the more relevant when one further notices that under the null distribution, the Y_{ij}'s are Gaussian. The similarity does not stop here. Denoting by $\mathbb{P}_\lambda$ the distribution of the matrix $\boldsymbol{Y}$, the likelihood ratio of $\mathbb{P}_\lambda$ to $\mathbb{P}_0$ is the denominator in (1.11):

$$L(\boldsymbol{Y};\lambda) = \frac{\mathrm{d}\,\mathbb{P}_\lambda}{\mathrm{d}\,\mathbb{P}_0}(\boldsymbol{Y}) = \int e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}). \tag{1.13}$$

This makes the questions of studying the properties of the likelihood ratio and the posterior of the spike amenable to analysis using tools originally invented to study the SK model. Moreover, if $P_{\mathrm{x}} = \frac{1}{2}\delta_{+1} + \frac{1}{2}\delta_{-1}$, then we see that the two problems are the same, possibly up to trivial additive constants.

This connection seems to have been noticed only recently, notably in the work of Lesieur, Krzakala, and Zdeborová (2015a,b, 2017) who then proceed based on plausible but non-rigorous statistical physics arguments to analyze the problem of estimation. Their work has shortly after been made rigorous in a series of papers (Barbier, Dia, et al., 2016; Deshpande, Abbé, and Montanari, 2016; Krzakala, Xu, and Zdeborová, 2016; Lelarge and Miolane, 2017; Miolane, 2017), where the error of the Bayes-optimal estimator has been completely characterized for additive low-rank models with a separable (product) prior on the spike. In particular, these papers confirm an interesting phenomenon discovered by Lesieur, Krzakala, and Zdeborová (2015a,b): for certain priors on the spike, estimation becomes possible—although computationally expensive—below the spectral threshold $\lambda = 1$. More precisely, the posterior mean overlaps with the spike in regions where the top eigenvector is orthogonal to it. Lesieur, Krzakala, and Zdeborová (2017) provides a full account of these phase transitions in a myriad of interesting situations, the majority of which still await rigorous treatment.

1.4 Pooling of discrete data

In the second part of this thesis, we consider a problem of "discrete" nature. Imagine a population of n individuals, each of whom has one among d "types". This could be their blood type, in which case $d = 4$, their age, income or tax bracket group, etc. This information is recorded in a database, and the data analyst is allowed to query it by specifying a random subset of the population, and in response, she observes the *histogram* (a d-dimensional vector of counts) of types of the queried individuals. This measurement scheme is inspired by practical situations where it may only be possible to assay certain summary statistics of the data involving a moderate of large number of participants. This may be done for privacy reasons, or it may be inherent to the data collection process. The latter situation occurs for instance in the analysis of genetic data where, due to experimental constraints, allele measurements across multiple strands of DNA of different individuals are pooled and analyzed together (Heo et al., 2001; Sham et al., 2002). The data then consists of a *frequency spectrum*, or histogram of alleles.

How many measurements of this sort does it take to reconstruct the types of the entire population? This problem falls broadly under the umbrella of the theory of *compressed sensing*, where one is interested in recovering a signal from a few compressed measurements (Donoho, 2006a). It has been understood from the early stages of the development of this theory that the structure of the signal, typically sparsity, plays a key role in the sample complexity, or number of measurements needed for reconstruction (Candés, Romberg, and Tao, 2006; Candés and Tao, 2005; Donoho, 2006b). In this theory, one classically considers a signal that is real-valued, and is compressed by taking random linear combinations of its entries. It is however interesting to move beyond this setting and consider signals that are discrete, where each entry can take a value from a finite alphabet; this is the setting we consider in our work. Then one possible model of compression—since the signal no longer has an additive structure—is to count the occurrence of each symbol in a randomly chosen subset of the signal's entries.

The discrete, combinatorial structure of this reconstruction problem makes it a special kind of a *constraint satisfaction problem* (CSP). These have been the object of intense study in recent years in probability theory, computer science, information theory and statistical physics. For certain families of CSPs, a deep understanding has begun to emerge regarding the number of solutions as a function of problem size, as well as the algorithmic feasibility of finding solutions when they exist (see e.g. Coja-Oghlan and Frieze, 2014; Coja-Oghlan, Haqshenas, and Hetterich, 2016; Coja-Oghlan, Mossel, and Vilenchik, 2009; Coja-Oghlan and Perkins, 2016; Ding, Sly, and Sun, 2015, 2016; Sly, Sun, and Y. Zhang, 2016). Consider in particular a *planted* random constraint satisfaction problem with n variables that take their values in the discrete set $\{1, \cdots, d\}$, with $d \geq 2$. A number of m clauses is drawn uniformly at random under the constraint that they are all satisfied by a pre-specified assignment, which is referred to as *the planted solution*. In our case, the signal is n-dimensional, d is the size of the alphabet, and there are m compressed observations (histograms) of the signal, which represents the planted solution that satisfies all the constraints.

Two questions are of particular importance: *(1) how large should m be so that the planted solution is the unique solution?* and *(2) given that it is unique, how large should m be so that it is recoverable by a "tractable" algorithm?* Significant progress has been made on these questions, often initiated by insights from statistical physics and followed by a growing body of rigorous mathematical investigation. The emerging picture is that in many planted CSPs, when n is sufficiently large, all solutions become highly correlated with the planted one when $m > \kappa_{\mathsf{IT}} \cdot n$, for some "Information-Theoretic" (IT) constant $\kappa_{\mathsf{IT}} > 0$. Furthermore, one of these highly correlated solutions becomes typically recoverable by a random walk or a Belief Propagation (BP)-inspired algorithm when $m > \kappa_{\mathsf{BP}} \cdot n$ for some $\kappa_{\mathsf{BP}} > \kappa_{\mathsf{IT}}$ (Coja-Oghlan and Frieze, 2014; Coja-Oghlan, Mossel, and Vilenchik, 2009; Krzakala, Mézard, and Zdeborová, 2012; Krzakala and Zdeborová, 2009). Interestingly, it is known in many problems, at least heuristically, that these algorithms fail when $\kappa_{\mathsf{IT}} < m/n < \kappa_{\mathsf{BP}}$, and a tractable algorithm that succeeds in this regime is still lacking (Achlioptas and Coja-Oghlan, 2008; Coja-Oghlan, 2009; Coja-Oghlan, Haqshenas, and Hetterich, 2016; Zdeborová and Krzakala, 2016). In other words, there is a non-trivial regime $m/n \in (\kappa_{\mathsf{IT}}, \kappa_{\mathsf{BP}})$ where an essentially unique solution exists, but is hard to recover.

1.5 Overview of the results in this thesis

Detection limits in spiked random matrix models

The first three chapters of this thesis (excluding the introduction) are devoted to the study of the detection and estimation on the spike in the random matrix models (1.6) and (1.7). In Chapter 2 we consider model (1.6) and show that likelihood ratio (1.13) has Gaussian fluctuations for all $\lambda < \lambda_c$ where λ_c is referred to as *the reconstruction threshold*: for all $\lambda < \lambda_c$,

$$\log L(\boldsymbol{Y};\lambda) \rightsquigarrow \mathcal{N}\left(\pm\frac{1}{4}\left(-\log(1-\lambda)-\lambda\right), \frac{1}{2}\left(-\log(1-\lambda)-\lambda\right)\right),$$

where the plus sign holds under the alternative $\boldsymbol{Y} \sim \mathbb{P}_\lambda$ and the minus sign under the null $\boldsymbol{Y} \sim \mathbb{P}_0$. As explained in Section 1.1, this result implies that strong detection is impossible below λ_c, and also pins down the exact performance of the likelihood ratio test. Moreover, we obtain precise formulae for several information-theoretic quantities such as the relative entropy of the planted model to the null model, their total variation, and so on. On the other hand, $\log L$ grows linearly with N under $\mathbb{P}_\lambda$ when $\lambda > \lambda_c$, and this implies the possibility of strong detection in this regime. The reconstruction threshold λ_c thus has an information-theoretic character, independent of any spectral properties of the matrix $\boldsymbol{Y}$. Its definition is intimately related to the limiting value of the normalized log-likelihood $\frac{1}{N}\log L$ under $\mathbb{P}_\lambda$, or equivalently, the Kullback-Leibler divergence between $\mathbb{P}_\lambda$ and $\mathbb{P}_0$. Indeed, it is known that $\frac{1}{N}\mathbb{E}_{\mathbb{P}_\lambda}\log L$ converges to a limiting value; to the so-called *replica-symmetric formula* $\phi_{\mathsf{RS}}(\lambda)$ as $N \to \infty$. And λ_c is defined as the smallest upper bound on the interval on which this limit vanishes. This threshold is generically different from the spectral threshold, and is always no larger than it.

In Chapter 3 we consider the asymmetric model (1.7) and establish a similar result:

$$\log L(\boldsymbol{Y};\lambda) \rightsquigarrow \mathcal{N}\left(\pm\frac{1}{4}\log(1-\alpha\lambda^2), -\frac{1}{2}\log(1-\alpha\lambda^2)\right),$$

as $N, M \to \infty$ and $M/N \to \alpha$, for all $\alpha\lambda^2 < c$ for a constant $c < 1$ that depends on the priors on the entries of $\boldsymbol{u}$ and $\boldsymbol{v}$. While we have been able to reach the optimal threshold in the symmetric model (1.6), this task turns out to be more difficult in the asymmetric case. We also state a conjecture on the maximal region of parameters (α, λ) where asymptotic normality should hold. Similarly to the symmetric model, this is also the region where the limit of the normalized log-likelihood ratio under $\mathbb{P}_\lambda$ vanishes, however as we will explain it, proving this seems to require new ideas.

In Chapter 4 we return to the symmetric model and turn our attention to the estimation problem. The normalized Kullback-Leibler divergence between the planted and null models is known to converge to the replica-symmetric formula $\phi_{\mathsf{RS}}(\lambda)$, the properties of which determine the fundamental limits of estimation in this model. For instance, it is known that the spike could be estimated with non-trivial accuracy if and only if $\lambda > \lambda_c$. However the available proofs of this result and other intimately related ones are quite involved. As a first result, we provide a short and transparent proof of this formula, based on simple executions of Gaussian interpolations and standard concentration-of-measure arguments. Second, we investigate the next-order asymptotics of this convergence: we prove that $D_{\mathsf{KL}}(\mathbb{P}_\lambda, \mathbb{P}_0) - N\phi_{\mathsf{RS}}(\lambda)$ converges to a finite quantity $\psi_{\mathsf{RS}}(\lambda)$ for almost all λ, and with speed $1/\sqrt{N}$. An explicit expression for this quantity is also given. This formula shows that the KL divergence between the null and planted distributions has a non-vanishing *finite size correction.* This correction becomes most important below the reconstruction threshold where $\phi_{\mathsf{RS}}(\lambda) = 0$, in which case we now know that this KL converges to a constant, instead of the mere fact that it is of order $o(N)$.

To study the likelihood ratio and the KL divergence in this setting we build on the technology developed by Aizenman, Guerra, Panchenko, Talagrand, and others, in their study of the Sherrington-Kirkpatrick spin-glass model. Specifically, we make use of Guerra's Gaussian interpolation method and Talagrand's cavity method. An important role is played by the so-called *Franz-Parisi potential* (Franz and Parisi, 1995, 1998): this is the log-partition function of a subset of configurations having a fixed overlap with the spike.

Decoding from pooled data

In the last two chapters of this thesis, we consider the problem of decoding from pooled data, which we call the Histogram Query Problem (HQP), previously described. In Chapter 5, we consider the *dense random* regime where each query involves a linear number of individuals, and where the individuals are chosen uniformly at random. Call $k = \alpha n$ the size of each pool, and m the number of queries. It could be seen by a simple counting argument that less than $m = (1 - o(1))\frac{\log d}{d-1}\frac{n}{\log n}$ queries is insufficient to reconstruct the types of the population. We show that $m = 2(1 + o(1))\frac{\log d}{d-1}\frac{n}{\log n}$ queries or more is on the other hand sufficient. The proof of this result proceeds by viewing the problem as a *planted* random constraint satisfaction problem. We use some sophisticated combinatorics combined with the Laplace method to compute the exponential rate of decay of the expected number of satisfying assignments to this CSP. This quantity is referred to as the *annealed free energy* in the statistical physics literature. Its knowledge implies the upper bound. Observe the gap of a factor of two between these upper and lower bounds. Shortly after our results were made public on the arXiv, Scarlett and Cevher (2017) proved that our upper bound is actually tight. Meaning that it is impossible to recover the type of every individual with less than $2(1-o(1))\frac{\log d}{d-1}\frac{n}{\log n}$ queries, thus establishing a sharp phase transition for the recovery problem at this number.

In Chapter 6, we consider the algorithmic aspect of the problem and design a practical algorithm inspired by *belief propagation* to reconstruct the types of the population. A heuristic analysis of this algorithm exhibits an interesting gap: the success probability of the algorithm undergoes a sharp phase transition at a much higher threshold. The algorithm succeeds at recovering the types of every individual if and only if $m > \kappa n$, with $\kappa \simeq 2\frac{\log d}{d-1}$. This algorithm are heuristically known to be "optimal" in a certain sense, so we expect other classes of algorithms such as linear programming relaxations and random walk-type algorithms to fail as well below $m = \kappa n$. This hints at a possibility of a logarithmic gap between the information-theoretic and algorithmic thresholds.

1.6 Bibliographic notes

The research leading to the results presented in this thesis is in collaboration with my advisor M. Jordan and a few other lovely colleagues. The material presented in Chapters 2 and 4 is a joint work with F. Krzakala and M. Jordan, and appears in the following two papers (El Alaoui and Krzakala, 2018; El Alaoui, Krzakala, and Jordan, 2017). The material presented in Chapter 4 is a joint work with M. Jordan and appears in (El Alaoui and Jordan, 2018). Finally, the material presented in Chapters 5 and 6 is a joint work with A. Ramdas, F. Krzakala, L. Zdeborová and M. Jordan, and appears in (El Alaoui et al., 2016, 2017).

Part I

Detection limits in some spiked random matrix models

Chapter 2

Detection limits in the spiked Wigner model

2.1 Introduction

We focus in this chapter on the *spiked Wigner model*, which is the following symmetric random matrix model

$$\boldsymbol{Y} = \sqrt{\frac{\lambda}{N}}\boldsymbol{x}^*\boldsymbol{x}^{*\top} + \boldsymbol{W}, \tag{2.1}$$

where $W_{ij} = W_{ji} \sim \mathcal{N}(0,1)$ and $W_{ii} \sim \mathcal{N}(0,2)$ are independent for all $1 \leq i \leq j \leq N$. The *spike* vector $\boldsymbol{x}^* \in \mathbb{R}^N$ represents the signal to be recovered, or its presence detected.

We assume that the entries x_i^* of the spike are i.i.d. from a prior distribution P_{x} on $\mathbb{R}$ having *bounded* support. The parameter $\lambda \geq 0$ plays the role of the signal-to-noise ratio, and the scaling by $\sqrt{N}$ is such that the signal and noise components of the observed data are of comparable magnitudes. In this chapter we will be mainly concerned with the detection problem: upon observing $\boldsymbol{Y}$, we want to test wether $\lambda > 0$ or $\lambda = 0$. Meaning, we want to detect the presence of a non-trivial, privileged direction $\boldsymbol{x}^*$ in the data matrix $\boldsymbol{Y}$, without caring about finding out what this specific direction is. We moreover want to understand the performance of the *best* test, i.e., the test that maximizes the probability of a correct guess. By the classical Neyman-Pearson lemma, this test is simply the likelihood ratio test. Therefore the problem reduces to (if one puts aside computational considerations) the study of the behavior of the likelihood ratio of the distributions associated to the two hypotheses of interest ($\lambda = 0$ vs. $\lambda > 0$). The main aim of this chapter and the next two is to analyze the fine-grained behavior of the likelihood ratio of this model, and slight generalizations of it. In Chapter 4 we turn our attention to the estimation problem where we want to estimate $\boldsymbol{x}^*$ with non-trivial accuracy. We will see in particular that under the scaling considered here, estimating $\boldsymbol{x}^*$ with vanishing error as $N \to \infty$ is not possible, but producing an estimator that has *partial* correlation with the spike still is, and we will determine the best correlation any estimator could achieve in a Bayes-optimal sense. As it turns out, the likelihood ratio is still the relevant object to look at in this setting, and several estimation-theoretic quantities can be derived from it.

As a matter of convenience, we discard the diagonal terms Y_{ii} from the observations. Adding the diagonal back does not pose any additional technical difficulties, and our results can be straightforwardly extended to this case. We denote by $\mathbb{P}_\lambda$ the joint probability law of the observations $\boldsymbol{Y} = \{Y_{ij} : 1 \leq i < j \leq N\}$ as per (2.1) and define the likelihood ratio or Radon-Nikodym derivative of $\mathbb{P}_\lambda$ to $\mathbb{P}_0$ as

$$L(\cdot;\lambda) \equiv \frac{\mathrm{d}\mathbb{P}_\lambda}{\mathrm{d}\mathbb{P}_0}. \tag{2.2}$$

For a fixed $\boldsymbol{Y}$, a simple computation based on conditioning on $\boldsymbol{x}^*$ reveals that

$$L(\boldsymbol{Y};\lambda) = \int \exp\Big(\sqrt{\frac{\lambda}{N}}\sum_{i<j} Y_{ij}x_ix_j - \frac{\lambda}{2N}\sum_{i<j} x_i^2x_j^2\Big)\,\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}). \tag{2.3}$$

A first step in studying the behavior of L is to identify the region of parameters λ where L remains of constant order as $N \to \infty$. This will be the region where the testing problem is most non-trivial since a diverging likelihood ratio indicates that $\mathbb{P}_\lambda$ is easily distinguishable from $\mathbb{P}_0$. This could for example be seen in the following way: define the quantity

$$F_N := \frac{1}{N}\,\mathbb{E}_{\mathbb{P}_\lambda}\log L(\boldsymbol{Y};\lambda). \tag{2.4}$$

We see that $F_N = \frac{1}{N}D_{\mathsf{KL}}(\mathbb{P}_\lambda,\mathbb{P}_0)$, where D_{KL} is the Kullback-Leibler divergence between probability measures. By the phenomenon of concentration measure (this will made quantitative later) $\frac{1}{N}\log L$ is unlikely to deviate much from its expectation F_N: for any $\epsilon > 0$,

$$\mathbb{P}_\lambda\left(\big|\frac{1}{N}\log L - F_N\big| \geq \epsilon\right) \longrightarrow 0.$$

Assume now that F_N converges to a limit ℓ as $N \to \infty$. Observe that by the non-negativity of KL (or Jensen's inequality) ℓ must be non-negative. On the other hand, one has the same concentration of $\frac{1}{N}\log L$ under $\mathbb{P}_0$ and moreover, $\mathbb{E}_{\mathbb{P}_0}\log L \leq 0$. Therefore if λ is such that $\ell > 0$, one can distinguish $\mathbb{P}_\lambda$ from $\mathbb{P}_0$ with asymptotic certainty (as N becomes large) by computing $\log L$ and rejecting the null hypothesis $\boldsymbol{Y} \sim \mathbb{P}_0$ if (for instance) $\log L \geq \ell/2$, i.e., this test would have vanishing type-I and type-II errors. On the other hand, in the set of λ where $\ell = 0$ (if it exists), the problem becomes highly non-trivial, since one has to "zoom" into the behavior of the un-normalized log-LR and examine what happens at a constant order.

Of course the above discussion is still valid if one flips the roles of $\mathbb{P}_\lambda$ and $\mathbb{P}_0$, and assumes that $\frac{1}{N}\mathbb{E}_{\mathbb{P}_0}\log L(\boldsymbol{Y};\lambda)$ admits a limit ℓ' instead. (The test would then be to check if $\frac{1}{N}\log L \leq \ell'/2$.) There is no a-priori reason to privilege $\mathbb{P}_\lambda$ over $\mathbb{P}_0$, except that of convenience: in this matrix model, it is far easier to study L under $\mathbb{P}_\lambda$.

At this point, we lay out our plan: we first introduce the limit of F_N, which is referred to as the *replica-symmetric* formula, and its properties. We will see that the above-discussed region where this limit vanishes takes to form of an interval $(0, \lambda_c)$, where the upper bound is referred to as the *reconstruction threshold*. We then move on to study the fluctuations of $\log L$ in this interval. We will prove that in this regime, the log-LR has fluctuations of constant order, and converges asymptotically to a Gaussian with a mean equal to half the variance. This allows for instance to show that $\mathbb{P}_\lambda$ and $\mathbb{P}_0$ are *mutually contiguous* below λ_c. Contiguity (see Definition 1.1) allows to show that *no test* is capable of distinguishing $\mathbb{P}_\lambda$ from $\mathbb{P}_0$ with asymptotic certainty below λ_c, thereby complementing the above discussion and providing a complete answer to the question of asymptotically certain (we will say "strong") detection. The fluctuation result also allows to obtain formulas for the Type-I and Type-II errors for testing, as well as the total variation distance.

The RS formula and the reconstruction threshold

The RS formula. For $r \geq 0$, consider the function

$$\psi(r) := \mathbb{E}_{x^*,z} \log \int \exp\left(\sqrt{r}zx + rxx^* - \frac{r}{2}x^2\right) \mathrm{d}P_{\mathrm{x}}(x), \tag{2.5}$$

where $z \sim \mathcal{N}(0,1)$, and $x^* \sim P_{\mathrm{x}}$. This is the KL divergence between the distributions of the random variables $y = \sqrt{r}x^* + z$ and z. We define the Replica-Symmetric (RS) potential

$$F(\lambda, q) := \psi(\lambda q) - \frac{\lambda q^2}{4}, \tag{2.6}$$

and finally define the RS formula

$$\phi_{\mathsf{RS}}(\lambda) := \sup_{q \geq 0} F(\lambda, q). \tag{2.7}$$

A central result in this context, which was conjectured by Lesieur, Krzakala, and Zdeborová (2015b), and then proved in a sequence of papers (Barbier, Dia, et al., 2016; Deshpande, Abbé, and Montanari, 2016; Krzakala, Xu, and Zdeborová, 2016; Lelarge and Miolane, 2016), is that free energy F_N converges to the RS formula for all $\lambda \geq 0$:

$$F_N \longrightarrow \phi_{\mathsf{RS}}(\lambda).$$

We refer to Lesieur, Krzakala, and Zdeborová (2017) for a derivation of this formula based on non-rigorous statistical physics arguments. In Chapter 4 we analyze this convergence in detail. We provide a short proof of the above result, and study the rate of convergence as well as the next order term.

The values of q that maximize the RS potential and their properties play an important role in the theory. Lelarge and Miolane (2016) proved that the map $q \mapsto F(\lambda, q)$ has a unique maximizer $q^* = q^*(\lambda)$ for all $\lambda \in \mathcal{D}$ where $\mathcal{D}$ is the set of points where the function

$\lambda \mapsto \phi_{\mathsf{RS}}(\lambda)$ is differentiable. By convexity of ϕ_{RS} (see next section), $\mathcal{D} = \mathbb{R}_+ \setminus$ countable set. Moreover, they showed that the map $\lambda \in \mathcal{D} \mapsto q^*(\lambda)$ is non-decreasing, and

$$\lim_{\substack{\lambda \to 0 \\ \lambda \in \mathcal{D}}} q^*(\lambda) = \mathbb{E}_{P_{\mathrm{x}}}[X]^2, \qquad \text{and} \qquad \lim_{\substack{\lambda \to \infty \\ \lambda \in \mathcal{D}}} q^*(\lambda) = \mathbb{E}_{P_{\mathrm{x}}}[X^2]. \tag{2.8}$$

One should interpret the value $q^*(\lambda)$ as the best overlap an estimator $\widehat{\theta}(\boldsymbol{Y})$ based on observing $\boldsymbol{Y}$ can have with the spike $\boldsymbol{x}^*$. Indeed, the overlap $\left|\boldsymbol{x}^\top \boldsymbol{x}^*\right|/N$ between the spike $\boldsymbol{x}^*$ and a random draw $\boldsymbol{x}$ from the posterior $\mathbb{P}_\lambda(\cdot|\boldsymbol{Y})$ should concentrate in the large N limit about $q^*(\lambda)$ (hence the name "replica-symmetry"). A matrix variant of this result (where one estimates $\boldsymbol{x}^*\boldsymbol{x}^{*\top}$) was proved in (Lelarge and Miolane, 2016). In Section 4.3, we prove strong (vector) versions of this result where under mild assumptions, optimal rates of convergence are given.

The reconstruction threshold. The first limit in (2.8) shows that when the prior P_{x} is not centered, it is always possible to have a non-trivial overlap with $\boldsymbol{x}^*$ for any $\lambda > 0$. On the other hand, when the prior has zero mean, and since q^* is a non-decreasing function of λ, it is useful to define the critical value of λ below which estimating $\boldsymbol{x}^*$ becomes impossible:

$$\lambda_c := \sup\left\{\lambda > 0 \;:\; q^*(\lambda) = 0\right\}. \tag{2.9}$$

We refer to λ_c as the *critical* or *reconstruction* threshold. The next lemma establishes a natural bound on λ_c.

Lemma 2.1. *We have*

$$\lambda_c \cdot \left(\mathbb{E}_{P_{\mathrm{x}}}[X^2]\right)^2 \leq 1. \tag{2.10}$$

Proof. Indeed, assume that P_{x} is centered, and let $\lambda > (\mathbb{E}[X^2])^{-2}$. Since $\psi'(0) = \frac{1}{2}\mathbb{E}_{P_{\mathrm{x}}}[X]^2 = 0$ and $\psi''(0) = \frac{1}{2}(\mathbb{E}_{P_{\mathrm{x}}}[X^2])^2$, we see that $\partial_q F(\lambda, 0) = 0$ and $\partial_q^2 F(\lambda, 0) = \frac{\lambda}{2}(\lambda\,\mathbb{E}_{P_{\mathrm{x}}}[X^2]^2 - 1) > 0$. So $q = 0$ cannot be a maximizer of $F(\lambda, \cdot)$. Therefore $q^*(\lambda) > 0$ and $\lambda \geq \lambda_c$. ■

The importance of Lemma 2.1 stems from the fact that the value $\left(\mathbb{E}_{P_{\mathrm{x}}}[X^2]\right)^{-2}$ is the spectral threshold previously discussed. Above this value, the first eigenvalue of the matrix $\boldsymbol{Y}$ leaves the bulk, and is at the edge of the bulk below it (Capitaine, Donati-Martin, and Féral, 2009; Féral and Péché, 2007; Péché, 2006). This value also marks the limit below which the first eigenvector of $\boldsymbol{Y}$ captures no information about the spike $\boldsymbol{x}^*$ (Benaych-Georges and Nadakuditi, 2011). Inequality (2.10) can be strict or turn into equality depending on the prior P_{x}. For instance, there is equality if the prior is Gaussian or Rademacher—so that the first eigenvector overlaps with the spike as soon as estimation becomes possible at all—and strict inequality in the case of the (sufficiently) sparse Rademacher prior $P_{\mathrm{x}} = \frac{\rho}{2}\delta_{-1/\sqrt{\rho}} + (1-\rho)\delta_0 + \frac{\rho}{2}\delta_{+1/\sqrt{\rho}}$. More precisely, there exists a value

$$\rho^* = \inf\left\{\rho \in (0,1) \;:\; \psi'''(0) < 0\right\} \approx 0.092,$$

such that $\lambda_c = 1$ for $\rho \geq \rho^*$, and $\lambda_c < 1$ for $\rho < \rho^*$. In the latter case, the spectral approach to estimating $\boldsymbol{x}^*$ fails for $\lambda \in (\lambda_c, 1)$, and it is believed that no polynomial time algorithm succeeds in this region (Banks, Moore, Vershynin, et al., 2017; Krzakala, Xu, and Zdeborová, 2016; Lesieur, Krzakala, and Zdeborová, 2015b). The "physically plausible" picture one should have in mind here is that (minus) the RS potential $q \mapsto -F(\lambda, q)$ can be interpreted as an "energy landscape" whose global minimum corresponds to the planted spike $\boldsymbol{x}^*$. If λ is small (i.e., $\lambda < \lambda_c$) this potential has a global minimum at $q = 0$ which means that the posterior mean is orthogonal to $\boldsymbol{x}^*$. As λ increases, the global minimum will shift to a strictly positive point q^*. If the prior P_{x} is such that $\lambda_c < 1$, $q = 0$ stays a stable local minimum for all $\lambda \in (\lambda_c, 1)$ while the global minimum is at $q^* > 0$. At the heuristic level, this is the reason of computational hardness of the problem: (iterative) algorithms are trapped in this locally optimal state, and must climb an *energy barrier* in order to fall into the basin of attraction of the global minimum. This latter operation is conjectured to require exponential time.

2.2 Fluctuations below the reconstruction threshold

In this section we are interested in the fluctuations of the log-LR. It can be seen by a standard concentration-of-measure argument that for all $\lambda > 0$, $\log L(\boldsymbol{Y}; \lambda)$ concentrates about its expectation with fluctuations bounded by $\mathcal{O}(\sqrt{N})$. While this bound is likely to be of the right order above λ_c (this is true for the SK model in high temperature and with non-zero external field, see Guerra and F. Toninelli, 2002a), it is very pessimistic below λ_c. Indeed, we will show that the fluctuations are of constant order with a Gaussian limiting law in this regime. This phenomenon was noticed early on in the case of the SK model: Aizenman, Lebowitz, and Ruelle (1987) showed that in the absence of an external field, the log-partition function of this model has (shifted) Gaussian fluctuations about its easily computed "annealed average" in high temperature. We will see in Section 2.4 that their result can be stated as a central limit theorem for $\log L(\boldsymbol{Y}; \lambda)$ under $\mathbb{P}_0$ in the case where the prior P_{x} is Rademacher. Furthermore, a proof by Talagrand (2011b) of their result provided us with a road map for proving a similar result for general P_{x}. Let us now state the fluctuation result along with consequences for hypothesis testing.

Theorem 2.2. *Assume the prior P_{x} is centered and of unit variance. For all $\lambda < \lambda_c$,*

$$\log L(\boldsymbol{Y}; \lambda) \rightsquigarrow \mathcal{N}\left(\pm\frac{1}{4}\left(-\log(1-\lambda) - \lambda\right), \frac{1}{2}\left(-\log(1-\lambda) - \lambda\right)\right),$$

where the plus sign holds under the alternative $\boldsymbol{Y} \sim \mathbb{P}_\lambda$ and the minus sign under the null $\boldsymbol{Y} \sim \mathbb{P}_0$.

The symbol "$\rightsquigarrow$" denotes convergence in distribution as $N \to \infty$.

The sign symmetry between the above two statements is a consequence of Le Cam's third lemma (Van der Vaart, 2000), (or more specifically, the Portmanteau lemma). We will see this in Section 2.5. This result is along the same line of those of Johnstone and Onatski (2015); Onatski, Moreira, and Hallin (2013, 2014), who studied the likelihood ratio of the joint eigenvalue densities under the spiked covariance model with a spherical prior, and showed its asymptotic normality below the spectral threshold. We also point out that similar fluctuation results were recently proved by Baik and Lee (2016, 2017a) for a spherical model where one integrates over the uniform measure on the sphere in the definition of $L(\boldsymbol{Y};\lambda)$. Their model, due to its integrable nature, is amenable to analysis using tools from random matrix theory. The authors are thus able to also analyze a "low temperature" regime (absent from our problem) where the fluctuations are no longer Gaussian but given by the Tracy-Widom distribution. Their techniques seem to be restricted to the spherical case however. Closer to our assumptions is the recent work of Banerjee and Ma (2018), (see also Banerjee, 2018) who use a very precisely conditioned second moment argument to show asymptotic normality of similar log-likelihood ratios. However, this technique (at least in its current flavor) works up to *some* value $\lambda_0 < \lambda_c$, and is not expected to be optimal in the SNR threshold.

Strong and weak detection below λ_c

Consider the problem of deciding whether an array of observations $\boldsymbol{Y} = \{Y_{ij} : 1 \le i < j \le N\}$ is likely to have been generated from $\mathbb{P}_\lambda$ for a fixed $\lambda > 0$ or from $\mathbb{P}_0$. Let us denote by $\boldsymbol{H}_0 : \boldsymbol{Y} \sim \mathbb{P}_0$ the null hypothesis and $\boldsymbol{H}_\lambda : \boldsymbol{Y} \sim \mathbb{P}_\lambda$ the alternative hypothesis. Two formulations of this problem exist: one would like to construct a sequence of measurable tests $T : \mathbb{R}^{N(N-1)/2} \mapsto \{0,1\}$ that returns "0" for $\boldsymbol{H}_0$ and "1" for $\boldsymbol{H}_\lambda$, for which either

$$\lim_{N\to\infty} \max\Big\{\mathbb{P}_\lambda(T(\boldsymbol{Y})=0),\ \mathbb{P}_0(T(\boldsymbol{Y})=1)\Big\} = 0, \tag{2.11}$$

or less stringently, the total mis-classification error, or risk

$$\mathsf{err}(T) := \mathbb{P}_\lambda(T(\boldsymbol{Y})=0) + \mathbb{P}_0(T(\boldsymbol{Y})=1) \tag{2.12}$$

is minimized among all possible tests T.

Strong detection. Using a second moment argument based on the computation of a truncated version of $\mathbb{E}\,L(\boldsymbol{Y};\lambda)^2$, Banks, Moore, Vershynin, et al. (2017) and Perry et al. (2016b) showed that $\mathbb{P}_\lambda$ and $\mathbb{P}_0$ are mutually contiguous when $\lambda < \lambda_0$, where the latter quantity equals λ_c for some priors P_{x} while it is suboptimal for others (e.g., the sparse Rademacher case, see discussion below). It is easy to see that contiguity implies impossibility of strong detection since for instance, if $\mathbb{P}_0(T(\boldsymbol{Y})=1) \to 0$ then $\mathbb{P}_\lambda(T(\boldsymbol{Y})=0) \to 1$. Here we show that Theorem 2.2 provides a more powerful approach to contiguity:

Corollary 2.3. *Assume the prior P_{x} is centered and of unit variance. Then for all $\lambda < \lambda_c$, $\mathbb{P}_\lambda$ and $\mathbb{P}_0$ are mutually contiguous.*

Proof. This is a consequence of either one of the two statements in Theorem 2.2. Indeed, considering the fluctuations under the null, if

$$\frac{\mathrm{d}\,\mathbb{P}_\lambda}{\mathrm{d}\,\mathbb{P}_0} \rightsquigarrow U$$

under $\mathbb{P}_0$ along some subsequence and for some random variable U, then by the continuous mapping theorem we necessarily have

$$U = \exp \mathcal{N}(-\mu, \sigma^2),$$

where $\mu = \frac{1}{4}\left(-\log(1-\lambda) - \lambda\right) = \frac{1}{2}\sigma^2$. We have $\Pr(U > 0) = 1$, and since $\mu = \frac{1}{2}\sigma^2$, we have $\mathbb{E}\,U = 1$. We now conclude using Le Cam's first lemma in both directions (Lemma 6.4 or Example 6.5, Van der Vaart, 2000). ■

This approach allows one to circumvent second moment computations which are not guaranteed to be tight in general, and necessitate careful and prior-specific conditioning that truncates away undesirable events.

We note that in the case of the sparse Rademacher prior $P_{\mathtt{x}} = \frac{\rho}{2}\delta_{-1/\sqrt{\rho}} + (1-\rho)\delta_0 + \frac{\rho}{2}\delta_{+1/\sqrt{\rho}}$, contiguity holds for all $\lambda < 1$ as soon as $\rho \geq \rho^* \approx 0.092$ by the above corollary, thus closing the gaps in the results of Banks, Moore, Vershynin, et al. (2017) and Perry et al. (2016b). Indeed, as argued below Lemma 2.1, the reconstruction and spectral thresholds are equal ($\lambda_c = 1$) for all $\rho \geq \rho^*$, and differ ($\lambda_c < 1$) below ρ^*. This implies that strong detection is impossible for $\lambda < 1$ and possible otherwise when $\rho \geq \rho^*$, while it becomes impossible only below λ_c but possible otherwise when $\rho < \rho^*$.

Weak detection. We have seen that strong detection is possible if and only if $\lambda > \lambda_c$. It is then natural to ask whether weak detection is possible below λ_c, i.e., is it possible to test with accuracy *better than that of a random guess* below the reconstruction threshold? The answer is *yes*, and this is another consequence of Theorem 2.2. More precisely, the optimal test minimizing the risk (2.12) is the likelihood ratio test which rejects the null hypothesis $\boldsymbol{H}_0$ (i.e., returns "1") if $L(\boldsymbol{Y};\lambda) > 1$, and its error is

$$\mathsf{err}^*(\lambda) = \mathbb{P}_\lambda(L(\boldsymbol{Y};\lambda) \leq 1) + \mathbb{P}_0(L(\boldsymbol{Y};\lambda) > 1) = 1 - D_{\mathsf{TV}}(\mathbb{P}_\lambda, \mathbb{P}_0). \tag{2.13}$$

One can readily deduce from Theorem 2.2 the Type-I and Type-II errors of the likelihood ratio test: for all $\lambda < \lambda_c$ the Type-II error is

$$\mathbb{P}_\lambda(\log L(\boldsymbol{Y};\lambda) \leq 0) = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t-\mu)^2/2\sigma^2} \mathrm{d}t + o(1) = \frac{1}{2}\mathsf{erfc}\left(\frac{\sqrt{\mu}}{2}\right) + o(1),$$

and the Type-I error is

$$\mathbb{P}_0(\log L(\boldsymbol{Y};\lambda) > 0) = \int_{0}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(t+\mu)^2/2\sigma^2} \mathrm{d}t + o(1) = \frac{1}{2}\mathsf{erfc}\left(\frac{\sqrt{\mu}}{2}\right) + o(1)$$

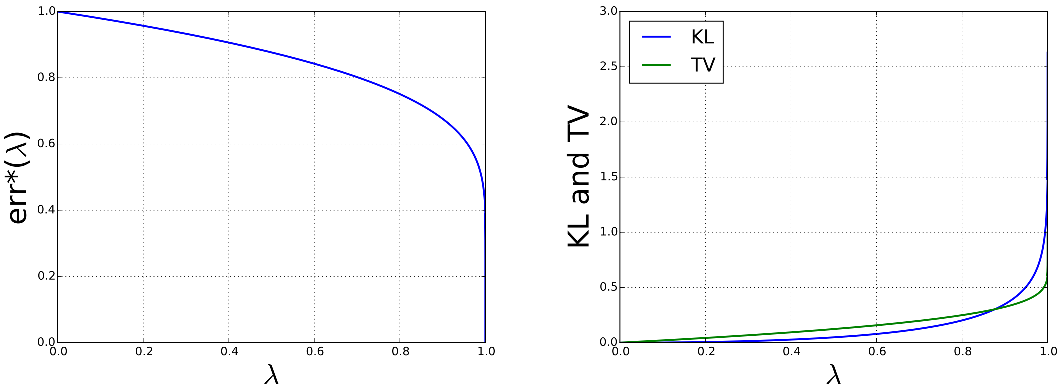

Figure 2.1: Plots of the TV distance and KL divergence between $\mathbb{P}_\lambda$ and $\mathbb{P}_0$. See formulas (2.14) and (2.15).

for all $\lambda < 1$. Here, $\mathsf{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} \mathrm{d}t$ is the complementary error function. These can be combined into a formula for $\mathsf{err}^*(\lambda)$ and the total variation distance between $\mathbb{P}_\lambda$ and $\mathbb{P}_0$ (see plot in Figure 2.1):

Corollary 2.4. *For all $\lambda < \lambda_c$, we have*

$$\lim_{N\to\infty} \mathsf{err}^*(\lambda) = 1 - \lim_{N\to\infty} D_{\mathsf{TV}}(\mathbb{P}_\lambda, \mathbb{P}_0) = \mathsf{erfc}\left(\frac{\sqrt{\mu(\lambda)}}{2}\right). \tag{2.14}$$

Moreover, the proof of Theorem 2.2 allows to obtain a formula for the KL divergence between $\mathbb{P}_\lambda$ and $\mathbb{P}_0$ below the reconstruction threshold λ_c (see plot in Figure 2.1):

Corollary 2.5. *Assume the prior P_{x} is centered and of unit variance. Then for all $\lambda < \lambda_c$,*

$$\lim_{N\to\infty} D_{\mathsf{KL}}(\mathbb{P}_\lambda, \mathbb{P}_0) = \frac{1}{4}\left(-\log\left(1-\lambda\right) - \lambda\right). \tag{2.15}$$

Note that the above formulas are only valid up to λ_c. In the case $\lambda_c < 1$, TV and KL both witness an abrupt discontinuity at λ_c to 1 and ∞ respectively. When $\lambda_c = 1$, then the behavior is more smooth with an asymptote at 1.

2.3 Replicas, overlaps, Gibbs measures and Nishimori

A crucial component of proving our main results is understanding the convergence of the overlap $\boldsymbol{x}^\top \boldsymbol{x}^*/N$, where $\boldsymbol{x}$ is drawn from $\mathbb{P}_\lambda(\cdot|\boldsymbol{Y})$, to its limit $q^*(\lambda)$. By Bayes' rule, we see that

$$\mathrm{d}\,\mathbb{P}_\lambda(\boldsymbol{x}|\boldsymbol{Y}) = \frac{e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})}{\int e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})}, \tag{2.16}$$

where H is the Hamiltonian

$$-H(\boldsymbol{x}) := \sum_{i<j} \sqrt{\frac{\lambda}{N}} Y_{ij} x_i x_j - \frac{\lambda}{2N} x_i^2 x_j^2. \tag{2.17}$$

Of course, when $\boldsymbol{Y} \sim \mathbb{P}_\lambda$, we can write

$$-H(\boldsymbol{x}) = \sum_{i<j} \sqrt{\frac{\lambda}{N}} W_{ij} x_i x_j + \frac{\lambda}{N} x_i x_i^* x_j x_j^* - \frac{\lambda}{2N} x_i^2 x_j^2.$$

From the formulas (2.3) and (5.7), it is straightforward to see that

$$F_N = \frac{1}{N} \mathbb{E}_{\mathbb{P}_\lambda} \log \int e^{-H(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}),$$

This provides another way of interpreting F_N as the expected log-partition function (or normalizing constant) of the posterior $\mathbb{P}_\lambda(\cdot|\boldsymbol{Y})$. For an integer $n \geq 1$ and $f : (\mathbb{R}^N)^{n+1} \mapsto \mathbb{R}$, we define the Gibbs average of f w.r.t. H as

$$\left\langle f(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}, \boldsymbol{x}^*) \right\rangle := \frac{\int f(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}, \boldsymbol{x}^*) \prod_{l=1}^n e^{-H(\boldsymbol{x}^{(l)})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}^{(l)})}{\left(\int e^{-H(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) \right)^n}. \tag{2.18}$$

This is nothing else that the average of f with respect to $\mathbb{P}_\lambda(\cdot|\boldsymbol{Y})^{\otimes n}$. The variables $\boldsymbol{x}^{(l)}, l = 1 \cdots, n$ are called *replicas*, and are interpreted as random variables independently drawn from the posterior. When $n = 1$ we simply write $f(\boldsymbol{x}, \boldsymbol{x}^*)$ instead of $f(\boldsymbol{x}^{(1)}, \boldsymbol{x}^*)$. Throughout the rest of this chapter, we use the following notation: for $l, l' = 1, \cdots, n, *$, we let

$$R_{l,l'} := \boldsymbol{x}^{(l)} \cdot \boldsymbol{x}^{(l')} = \frac{1}{N} \sum_{i=1}^N x_i^{(l)} x_i^{(l')}.$$

The Nishimori property under $\mathbb{P}_\lambda$

The fact that the Gibbs measure $\langle \cdot \rangle$ is a posterior distribution (4.2) has far-reaching consequences. A crucial implication is that the $n+1$-tuples $(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n+1)})$ and $(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}, \boldsymbol{x}^*)$ have the same law under $\mathbb{E}_{\mathbb{P}_\lambda} \langle \cdot \rangle$. To see this, let's perform the following experiment:

1. Construct $\boldsymbol{x}^* \in \mathbb{R}^N$ by independently drawing its coordinates from P_{x}.
2. Construct $\boldsymbol{Y}$ as $Y_{ij} = \sqrt{\frac{\lambda}{N}} x_i^* x_j^* + W_{ij}$, where $W_{ij} \sim \mathcal{N}(0,1)$ are all independent for $i < j$. (Therefore, $\boldsymbol{Y}$ is distributed according to $\mathbb{P}_\lambda$.)
3. Draw $n+1$ independent random vectors $(\boldsymbol{x}^{(l)})_{l=1}^{n+1}$ from $\mathbb{P}_\lambda(\boldsymbol{x} \in \cdot|\boldsymbol{Y})$.

By the tower property of expectations, the following equality of joint laws holds

$$\left(\boldsymbol{Y}, \boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}, \boldsymbol{x}^{(n+1)}\right) \overset{\mathrm{d}}{=} \left(\boldsymbol{Y}, \boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}, \boldsymbol{x}^{*}\right). \tag{2.19}$$

This in particular implies that under the alternative $\mathbb{P}_\lambda$, the overlaps $R_{1,*}$ between a replica and the spike has the same distribution as the overlap $R_{1,2}$ between two replicas. This is a very important property of the planted (spiked) model, which is usually named after Nishimori (2001). It is however worth noting that Nishimori deduced this property from very different considerations of *gauge symmetry* of the SK model with ±1 spins and an appropriately parameterized external field. (We refer to Chapter 4 in his book, or Korada and Macris (2009) for more background on this point of view.) He probably did not have in mind the interpretation of F_N as the log-normalizing constant of the posterior distribution of a spike in a random matrix model, from which the above property is a straightforward consequence. Property (2.19) substantially simplifies important technical arguments that are otherwise very difficult to conduct under the null. A recurring example in our context is the following: to prove the convergence of the overlap between two replicas, $\mathbb{E}\langle R_{1,2}^2\rangle \to 0$, it suffices to prove $\mathbb{E}\langle R_{1,*}^2\rangle \to 0$ since the two quantities are equal. The latter turns out to be a much easier task.

Overlap decay implies super-concentration

Let us now explain how the behavior of the overlaps is related to the fluctuations of $\log L$. For concreteness we consider the null model as an example. Let $\boldsymbol{Y} \sim \mathbb{P}_0$, i.e., $Y_{ij} \sim \mathcal{N}(0,1)$ all independent for $i < j$. The log-likelihood ratio, seen as a function of $\boldsymbol{Y}$, is a differentiable function, and

$$\frac{\mathrm{d}}{\mathrm{d}Y_{ij}} \log L(\boldsymbol{Y}; \beta) = \sqrt{\frac{\beta}{N}} \langle x_i x_j \rangle.$$

By the Gaussian Poincaré inequality, we can bound the variance by the norm of the gradient as

$$\mathbb{E}\left[(\log L - \mathbb{E}\log L)^2\right] \leq \mathbb{E}\left[\|\nabla \log L\|_{\ell_2}^2\right] \leq \frac{\lambda}{2} N\, \mathbb{E}\langle R_{1,2}^2\rangle.$$

(Here $\mathbb{E}$ refers to $\mathbb{E}_{\mathbb{P}_0}$.) The last inequality follows from the fact $\langle x_i x_j\rangle^2 = \langle x_i^{(1)} x_j^{(1)} x_i^{(2)} x_j^{(2)}\rangle$. Since P_{x} has bounded support, $R_{1,2}^2$ is bounded almost surely, and we deduce that the variance is $\mathcal{O}(N)$. Observe now that if the quantity $\mathbb{E}\langle R_{1,2}^2\rangle$ decays, then the much stronger result $\mathrm{var}(\log L) = o(N)$ holds. This behavior of unusually small variance is often referred to as "super-concentration." We refer to Chatterjee (2014) for more on this topic. In our case, not only does $\mathbb{E}\langle R_{1,2}^2\rangle$ decay when $\lambda < \lambda_c$ is sufficiently small, but it does so at a rate of $1/N$ so that $N\,\mathbb{E}\langle R_{1,2}^2\rangle$ converges to a finite limit, and $\mathrm{var}(\log L)$ is constant. This is a first reason why Theorem 2.2 should be expected: if anything, the fluctuations must be of constant order.

2.4 The Aizenman-Lebowitz-Ruelle CLT

We assume in this subsection that $P_{\mathrm{x}} = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}$ and let $\boldsymbol{Y} \sim \mathbb{P}_0$, i.e, $Y_{ij} \sim \mathcal{N}(0,1)$ i.i.d. We then see that the likelihood ratio L is related to the partition function of the Sherrington–Kirkpatrick model via a trivial relation:

$$\begin{aligned}\log L(\boldsymbol{Y};\lambda) &= \log \int \exp\Big(\sqrt{\frac{\lambda}{N}}\sum_{i<j} Y_{ij}x_ix_j - \frac{\lambda}{2N}\sum_{i<j} x_i^2x_j^2\Big)\,\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})\\ &= \log \sum_{\boldsymbol{\sigma}\in\{\pm1\}^N} \exp\Big(\frac{\beta}{\sqrt{N}}\sum_{i<j} Y_{ij}\sigma_i\sigma_j\Big) - N\log 2 - \frac{\beta^2(N-1)}{4}\\ &=: \log Z_N(\beta) - N\log 2 - \frac{\beta^2(N-1)}{4},\end{aligned}$$

where we have let $\beta = \sqrt{\lambda}$. $Z_N(\beta)$ is the partition function of the SK model at inverse temperature $\beta > 0$. It is easy to compute the expectation of $Z_N(\beta)$:

$$\log \mathbb{E}\, Z_N(\beta) = N\log 2 + \frac{\beta^2(N-1)}{4},$$

so that $\mathbb{E}\log L(\boldsymbol{Y};\lambda)$ is the gap between the free energy of the SK model and its *annealed* version. The question of determining the values of the inverse temperature for which this gap is zero (or constant), i.e., *at what temperatures is the free energy given by the annealed computation?* is a central question in statistical physics. Aizenman, Lebowitz, and Ruelle (1987) (ALR) proved that in the high-temperature regime $\beta < 1$, $\log(Z_N(\beta)/\mathbb{E}\,Z_N(\beta))$ converges in distribution the normal law

$$\mathcal{N}\left(\frac{1}{4}(\log(1-\beta^2)+\beta^2), -\frac{1}{2}(\log(1-\beta^2)+\beta^2)\right).$$

In our notation this simply means

$$\log L(\boldsymbol{Y};\lambda) \rightsquigarrow \mathcal{N}(-\mu,\sigma^2)$$

under $\mathbb{P}_0$ where $\mu = \frac{1}{2}\sigma^2 = \frac{1}{4}(-\log(1-\lambda)-\lambda)$. The proof of ALR is combinatorial. It uses the so-called *cluster expansion* technique which expands the partition function

$$Z_N(\beta) = \sum_{\boldsymbol{\sigma}\in\{\pm1\}^N}\prod_{i<j}\exp\Big(\frac{\beta}{\sqrt{N}}Y_{ij}\sigma_i\sigma_j\Big),$$

using the fact

$$\forall\alpha\in\mathbb{R},\quad \exp(\alpha\sigma_i\sigma_j) = \cosh(\alpha) + \sigma_i\sigma_j\sinh(\alpha), \tag{2.20}$$

to write

$$Z_N(\beta) = 2^N\prod_{i<j}\cosh\Big(\frac{\beta}{\sqrt{N}}Y_{ij}\Big)\cdot\sum_G\prod_{(i,j)\in G}\tanh\Big(\frac{\beta}{\sqrt{N}}Y_{ij}\Big),$$

where the sum is over all simple closed graphs G on the vertex set $\{1, \cdots, N\}$. (A graph is simple if no edge is repeated. It is closed if every vertex has even degree.) A very careful argument then shows that for $\beta < 1$, the above sum is almost exclusively due to the contribution of graphs G which are closed paths (where every vertex is of degree two), and these are responsible for the Gaussian fluctuations of $\log Z_N(\beta)$.

Fact (2.20) is true only when the spins σ_i take the values ± 1. It is therefore hard to extend the ALR proof to other types of priors. Alternative proofs were subsequently found by adopting different perspectives on the problem. See, e.g., (Comets and Neveu, 1995) who use connections with stochastic calculus, and Guerra and F. Toninelli (2002a) who use the interpolation and cavity methods. A more recent proof also based on the cavity method is provided by Talagrand in his second book (Talagrand, 2011b, Section 11.4). His method provides an explicit (and optimal) rate of convergence of the moments of the random variable in question to those of the Gaussian. In what follows we use Talagrand's approach to prove a similar central limit theorem for an arbitrary bounded prior $P_{\mathtt{x}}$. In this more general setting, the high temperature region of the model is given by the condition $\lambda < \lambda_c$.

2.5 Proof of the main result

In this section we prove Theorem 2.2. It suffices to prove the fluctuations under one of the hypotheses. Fluctuations under the remaining one comes for free as a consequence of Le Cam's third lemma (or more specifically, the Portmanteau theorem Van der Vaart, 2000, Theorem 6.6). We choose to treat the planted case $\boldsymbol{Y} \sim \mathbb{P}_\lambda$. The reason is that it is easier to deal with the planted model. This is ultimately due to the (generalized) Nishimori property (2.19). Let us first present this argument.

Equivalence of fluctuations under planted and null models

We explain how the fluctuation result under $\mathbb{P}_\lambda$ implies the corresponding fluctuation result under $\mathbb{P}_0$. This is a consequence of the Portmanteau characterization of convergence in distribution. The argument can be made in the other direction as well. Assume that

$$\log L(\boldsymbol{Y};\lambda) \rightsquigarrow \mathcal{N}(\mu, \sigma^2),$$

for $\boldsymbol{Y} \sim \mathbb{P}_\lambda$, where $\mu = \frac{1}{2}\sigma^2$. By the Portmanteau theorem (Van der Vaart, 2000, Lemma 2.2), this is equivalent to the assertion

$$\liminf \mathbb{E}_{\mathbb{P}_\lambda}\left[f(\log L)\right] \geq \mathbb{E}\left[f(Z)\right], \tag{2.21}$$

where $Z \sim \mathcal{N}(\mu, \sigma^2)$ for all nonnegative continuous functions $f : \mathbb{R} \mapsto \mathbb{R}_+$. On the other hand, by a change of measure (and absolute continuity of $\mathbb{P}_0$ w.r.t $\mathbb{P}_\lambda$), we have that for such an f,

$$\mathbb{E}_{\mathbb{P}_0}\left[f(\log L)\right] = \mathbb{E}_{\mathbb{P}_\lambda}\left[\frac{\mathrm{d}\,\mathbb{P}_0}{\mathrm{d}\,\mathbb{P}_\lambda} f(\log L)\right] = \mathbb{E}_{\mathbb{P}_\lambda}\left[e^{-\log L} f(\log L)\right].$$

The function $g : x \mapsto e^{-x} f(x)$ is still nonnegative continuous, so by (2.21), we have

$$\liminf \mathbb{E}_{\mathbb{P}_0}\left[f(\log L)\right] \geq \mathbb{E}\left[e^{-Z} f(Z)\right]. \tag{2.22}$$

Since $\mu = \frac{1}{2}\sigma^2$,

$$\begin{aligned}\mathbb{E}\left[e^{-Z} f(Z)\right] &= \int f(x) e^{-x} e^{-(x-\mu)^2/2\sigma^2} \frac{\mathrm{d}x}{\sqrt{2\pi\sigma^2}} \\ &= \int f(x) e^{-(x+\mu)^2/2\sigma^2} \frac{\mathrm{d}x}{\sqrt{2\pi\sigma^2}} = \mathbb{E}\left[f(Z')\right],\end{aligned}$$

where $Z' \sim \mathcal{N}(-\mu, \sigma^2)$. Since (2.22) is valid for every nonnegative continuous f, the result

$$\log L(\boldsymbol{Y}; \lambda) \rightsquigarrow \mathcal{N}(-\mu, \sigma^2)$$

under $\mathbb{P}_0$ follows.

Fluctuations under $\mathbb{P}_\lambda$: a planted version of the ALR CLT

In this section we prove a slightly stronger result than convergence in distribution. We prove the convergence of all moments with an explicit rate of $\mathcal{O}(N^{-1/2})$. Let $\lambda < \lambda_c$, and $\boldsymbol{Y} \sim \mathbb{P}_\lambda$. We define the random variable

$$X(\lambda) = \log L(\boldsymbol{Y}) - \mu(\lambda),$$

where

$$\mu(\lambda) = \frac{1}{4}(-\log(1-\lambda) - \lambda), \quad \text{and} \quad b(\lambda) = \sigma^2(\lambda) = 2\mu(\lambda).$$

We will prove that the integer moments of $X(\lambda)$ converge to those of the Gaussian with variance $b(\lambda)$. This is a sufficient condition for convergence in distribution to hold, since the Gaussian is uniquely determined by its moments (see section 3.3.5, Durrett, 2010).

Theorem 2.6. *For all $\lambda < \lambda_c$ and integers k, there exists a constant $K(\lambda, k) \geq 0$ such that*

$$\left|\mathbb{E}_{\mathbb{P}_\lambda}\left[X(\lambda)^k\right] - m(k) b(\lambda)^{k/2}\right| \leq \frac{K(\lambda, k)}{\sqrt{N}},$$

where $m(k) = \mathbb{E}[g^k]$ is the k-th moment of the standard Gaussian $g \sim \mathcal{N}(0,1)$.

This theorem mirrors Theorem 11.4.1 in (Talagrand, 2011b), and our approach is inspired by his. We define the function

$$f(\lambda) := \mathbb{E}_{\mathbb{P}_\lambda}\left[X(\lambda)^k\right].$$

Lemma 2.7. *For all $\lambda < \lambda_c$,*

$$\begin{aligned} f'(\lambda) = \frac{k}{4} \mathbb{E}\left[\left(N\langle R_{1,2}^2\rangle - \langle x_N^2\rangle^2\right) X(\lambda)^{k-1}\right] - k\mu'(\lambda)\,\mathbb{E}\left[X(\lambda)^{k-1}\right] \\ + \frac{k(k-1)}{4} \mathbb{E}\left[\left(N\langle R_{1,2}^2\rangle - \langle x_N^2\rangle^2\right) X(\lambda)^{k-2}\right] \end{aligned} \tag{2.23}$$

Proof. By simple differentiation and regrouping of terms, we obtain

$$\begin{aligned} f'(\lambda) = &-\frac{k}{4}\,\mathbb{E}\left[\left(N\langle R_{1,2}^2\rangle - \langle x_N^2\rangle^2\right)X(\lambda)^{k-1}\right] + \frac{k}{2}\,\mathbb{E}\left[\left(N\langle R_{1,*}^2\rangle - \langle x_N^2 x_N^{*2}\rangle\right)X(\lambda)^{k-1}\right] \\ &- k\mu'(\lambda)\,\mathbb{E}\left[X(\lambda)^{k-1}\right] + \frac{k(k-1)}{4}\,\mathbb{E}\left[\left(N\langle R_{1,2}^2\rangle - \langle x_N^2\rangle^2\right)X(\lambda)^{k-2}\right]. \end{aligned} \tag{2.24}$$

Since we are under the planted model $\mathbb{P}_\lambda$ and $X(\lambda)$ only depends on $\boldsymbol{Y}$, we can use the generalized Nishimori property (2.19) to deduce that

$$\mathbb{E}\left[\left(N\langle R_{1,*}^2\rangle - \langle x_N^2 x_N^{*2}\rangle\right)X(\lambda)^{k-1}\right] = \mathbb{E}\left[\left(N\langle R_{1,2}^2\rangle - \langle x_N^2\rangle^2\right)X(\lambda)^{k-1}\right],$$

and this concludes the proof. ■

The derivative involves averages of the form

$$\mathbb{E}\left[\left(N\langle R_{1,2}^2\rangle - \langle x_N^2\rangle^2\right)X(\lambda)^{k}\right],$$

In the first line of (2.24), we see that the planted term $l = *$ has a pre-factor twice as big the that of the replica term $l = 2$. This is the reason the mean of the limiting Gaussian is μ and not $-\mu$ in the planted case. A crucial step in the argument is to show that $X(\lambda)^k$ and its pre-factor in the above expression are asymptotically uncorrelated, so that one can split the expectation of the product into the product of the expectations. More precisely, one should expect the quantities $N\langle R_{1,2}^2\rangle$ and $\langle x_N^2\rangle^2$ to tightly concentrate about some deterministic values when $\lambda < \lambda_c$, in such way that the above expectation is a multiple of $\mathbb{E}[X(\lambda)^k]$. This is what is usually referred to as *replica-symmetry* in the statistical physics literature.

Proposition 2.8. *For all $\lambda < \lambda_c$ and integers $k \geq 1$, we have*

$$\mathbb{E}\left[\left(N\langle R_{1,2}^2\rangle - \langle x_N^2\rangle^2\right)X(\lambda)^{k}\right] = \frac{\lambda}{1-\lambda}\,\mathbb{E}\left[X(\lambda)^k\right] + \delta,$$

where $|\delta| \leq K(k,\lambda)/\sqrt{N}$.

From here, we can prove the convergence of the moments of $\log L$ by integrating the differential equation given in Lemma 2.7.

***Proof** of Theorem 2.6.* Plugging the result of Proposition 2.8 into Lemma 2.7 yields

$$f'(\lambda) = k\left(\frac{1}{4}\frac{\lambda}{1-\lambda} - \mu'(\lambda)\right)\mathbb{E}\left[X(\lambda)^{k-1}\right] + \frac{k(k-1)}{4}\frac{\lambda}{1-\lambda}\,\mathbb{E}\left[X(\lambda)^{k-2}\right] + \delta.$$

Notice that with our choice of the function μ, the first term on the right-hand side vanishes. (Setting this term to zero provides another way of discovering the function μ.) Now we let $b(\lambda) = 2\mu(\lambda)$. We have for all λ and all $k \geq 1$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\,\mathbb{E}\left[X(\lambda)^k\right] = \frac{k(k-1)}{2}b'(\lambda)\,\mathbb{E}\left[X(\lambda)^{k-2}\right] + \delta. \tag{2.25}$$

By induction, and since $X(0) = 0$, we see that for all $k \geq 1$

$$\mathbb{E}\left[X(\lambda)^k\right] = m(k)b(\lambda)^{k/2} + \mathcal{O}\left(\frac{K(k,\lambda)}{\sqrt{N}}\right),$$

where $m(k) = (k-1)m(k-2)$ and $m(0) = 1$. The last recursion defines the sequence of Gaussian moments. ■

Now it remains to prove Proposition 2.8. This will require the deployment of a number of ideas from the theory of mean-field spin glasses.

Sketch of proof of Proposition 2.8

The idea is to show self-consistency relations among the quantities of interest. Namely, we will prove that for all $\lambda < 1$,

$$N\,\mathbb{E}\left[\langle R_{1,2}^2\rangle X(\lambda)^k\right] = \frac{1}{1-\lambda}\,\mathbb{E}\left[\langle x_N^2\rangle^2 X(\lambda)^k\right] + \delta, \tag{2.26}$$

and

$$\mathbb{E}\left[\langle x_N^2\rangle^2 X(\lambda)^k\right] = \mathbb{E}\left[X(\lambda)^k\right] + \delta, \tag{2.27}$$

where in both cases

$$|\delta| \leq K(k,\lambda)N\left(\mathbb{E}\left\langle R_{1,2}^4\right\rangle\right)^{3/4}.$$

Next, we need to prove the convergence of fourth moment of the overlap $R_{1,2}$ under $\mathbb{E}\langle\cdot\rangle$ at an optimal rate of $\mathcal{O}(1/N^2)$:

Theorem 2.9. *For all $\lambda < \lambda_c$, there exist a constant $K = K(\lambda) < \infty$ such that*

$$\mathbb{E}\langle R_{1,2}^4\rangle \leq \frac{K}{N^2}.$$

This will allow us to conclude that $|\delta| \leq K(k,\lambda)/\sqrt{N}$. It is interesting to note that while the self-consistent (or cavity) equations (2.26) and (2.27) hold for all $\lambda < 1$, overlap convergence is only true up to λ_c.

2.6 Proof of asymptotic decoupling

In this section we prove Proposition 2.8. As explained earlier, the argument is in two stages. We first prove that

$$N\,\mathbb{E}\left[\langle R_{1,2}^2\rangle X(\lambda)^k\right] = \frac{1}{1-\lambda}\,\mathbb{E}\left[\langle x_N^2\rangle^2 X(\lambda)^k\right] + \delta,$$

and then

$$\mathbb{E}\left[\langle x_N^2\rangle^2 X(\lambda)^k\right] = \mathbb{E}\left[X(\lambda)^k\right] + \delta,$$

where in both cases $|\delta| \leq K(k,\lambda)/\sqrt{N}$.

Preliminary bounds

We make repeated use of interpolation arguments in our proofs. We state here a few elementary lemmas we will invoke several times. We denote the overlaps between replicas where the last variable x_N is deleted by a superscript "-":

$$R^-_{l,l'} = \frac{1}{N}\sum_{i=1}^{N-1} x_i^{(l)}x_i^{(l')}.$$

Let $\{H_t : t \in [0,1]\}$ be a family of interpolating Hamiltonians. We let $\langle\cdot\rangle_t$ denote the corresponding Gibbs average, similarly to (4.15). Following Talagrand's notation, we write

$$\nu_t(f) := \mathbb{E}\langle f\rangle_t,$$

for a generic function f of n replicas $\boldsymbol{x}^{(l)}$, $l = 1,\cdots,n$. And abbreviate ν_1 by ν. The main tool we use is the following interpolation that isolates the last variable x_N from the rest of the system:

$$\begin{aligned}-H_t(\boldsymbol{x}) := &\sum_{1\le i<j\le N-1} \sqrt{\frac{\lambda}{N}}W_{ij}x_ix_j + \frac{\lambda}{N}x_ix_i^*x_jx_j^* - \frac{\lambda}{2N}x_i^2x_j^2 \\ &+ \sum_{i=1}^{N-1}\sqrt{\frac{\lambda t}{N}}W_{iN}x_ix_N + \frac{\lambda t}{N}x_ix_i^*x_Nx_N^* - \frac{\lambda t}{2N}x_i^2x_N^2.\end{aligned} \tag{2.28}$$

At $t = 1$ we have $H_t = H$, and at $t = 0$ the variable x_N decouples from the rest of the variables.

Lemma 2.10. *let f be a function of n replicas $\boldsymbol{x}^{(l)}$, $l = 1,\cdots,n$ and $\boldsymbol{x}^*$. Then*

$$\begin{aligned}\nu_t'(f) = &\frac{\lambda}{2}\sum_{1\le l\ne l'\le n}\nu_t(R^-_{l,l'}y^{(l)}y^{(l')}f) - \lambda n\sum_{l=1}^{n}\nu_t(R^-_{l,n+1}y^{(l)}y^{(n+1)}f) \\ &+ \lambda n\sum_{l=1}^{n}\nu_t(R^-_{l,*}y^{(l)}y^*f) - \lambda n\nu_s(R^-_{n+1,*}y^{(n+1)}y^*f) \\ &+ \lambda\frac{n(n+1)}{2}\nu_t(R^-_{n+1,n+2}y^{(n+1)}y^{(n+2)}f),\end{aligned}$$

where we have written $y = x_N$.

Proof. The computation relies on Gaussian integration by parts. See Talagrand (2011a, Lemma 1.6.3) for the details of a similar computation. ■

Lemma 2.11. *If f is a bounded non-negative function, then for all $t \in [0,1]$,*

$$\nu_t(f) \le K(\lambda, n)\nu(f).$$

Proof. Since the variables and the overlaps are all bounded, using Lemma 4.18 we have for all $t \in [0,1]$

$$|\nu_t'(f)| \leq K(\lambda, n)\nu_t(f).$$

Then we conclude using Grönwall's lemma. ■

Executing the cavity method

In its essence, the cavity method amounts to isolating one variable from the system and analyzing the influence of the rest of the variables on it. It was initially introduced as an analytic tool, alternative to the replica method, to solve certain models of spin glasses (Mézard, Parisi, and Virasoro, 1990), and has since been tremendously successful in predicting the behavior of many mean-field models. The underlying principle is known as the *leave-one-out* method in statistics. In our setting, this principle is materialized in the form of an interpolation method that separates the last variable from the rest. Let

$$Y(t) := \log \int e^{-H_t(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) - \mu(\lambda),$$

where H_t is defined in (2.28). We have $Y(1) = X(\lambda)$. We consider the quantity

$$\mathbb{E}\left[\left(N\langle R_{1,l}^2\rangle - \langle (x_N^{(1)} x_N^{(2)})^2\rangle\right) X(\lambda)^k\right].$$

By symmetry between sites, the above is equal to

$$N\,\mathbb{E}\left[\left\langle x_N^{(1)} x_N^{(2)} R_{1,2}^-\right\rangle X(\lambda)^k\right].$$

Now we consider the function

$$\varphi(t) := N\,\mathbb{E}\left[\left\langle x_N^{(1)} x_N^{(2)} R_{1,2}^-\right\rangle_t Y(t)^k\right].$$

Our strategy is approximate $\varphi(t)$ by $\varphi(0) + \varphi'(0)$. The approach is very similar to the one used to prove optimal rates of convergence of the overlaps. Notice that since the last variables decouple from the rest of the system at $t = 0$, we have

$$\begin{aligned}\varphi(0) &= N\,\mathbb{E}\left[\langle x_N^{(1)} x_N^{(2)}\rangle_0\right] \cdot \mathbb{E}\left[\left\langle R_{1,2}^-\right\rangle_0 Y(0)^k\right]\\ &= N\,\mathbb{E}_{P_{\mathrm{x}}}[X]^2 \cdot \mathbb{E}\left[\left\langle R_{1,2}^-\right\rangle_0 Y(0)^k\right] = 0.\end{aligned}$$

The expressions of the derivatives are a bit cumbersome so we do not display them, but we will describe their main features. The derivative $\varphi'(t)$ will be a sum of different terms, all of the form

$$\lambda N c(k)\,\mathbb{E}\left[\left\langle R_{1,2}^- R_{a,b}^- x_N^{(1)} x_N^{(2)} x_N^{(a)} x_N^{(b)}\right\rangle_t Y(t)^n\right], \tag{2.29}$$

where $n \in \{k-2, k-1, k\}$ and $(a,b) \in \{(1,2),(1,3),(3,4),(1,*),(3,*)\}$, and $c(k)$ is a polynomial of degree ≤ 2 in k. We see that at $t = 0$, if the above expression involves a variable x_N of degree 1 then this term vanishes. Therefore the only remaining term is the one where $(a,b) = (1,2)$. One can verify that $c(k) = 1$ for this term. Therefore

$$\begin{aligned}\varphi'(0) &= \lambda N \,\mathbb{E}\left[\langle {x_N^{(1)}}^2 {x_N^{(2)}}^2 \rangle_0\right] \cdot \mathbb{E}\left[\langle (R_{1,2}^-)^2\rangle_0 Y(0)^k\right] \\ &= \lambda N \,\mathbb{E}_{P_\mathrm{x}}\left[X^2\right]^2 \cdot \mathbb{E}\left[\langle (R_{1,2}^-)^2\rangle_0 Y(0)^k\right] \\ &= \lambda N \,\mathbb{E}\left[\langle (R_{1,2}^-)^2\rangle_0 Y(0)^k\right]. \end{aligned} \tag{2.30}$$

Now we turn to $\varphi''(t)$. Taking another derivative generates monomials of degree three in the overlaps and the last variable, so $\varphi''(t)$ is a sum of terms of the form

$$\lambda^2 N c'(k)\,\mathbb{E}\left[\left\langle R_{1,2}^- R_{a,b}^- R_{c,d}^- x_N^{(1)} x_N^{(2)} x_N^{(a)} x_N^{(b)} x_N^{(c)} x_N^{(d)} \right\rangle_t Y(t)^n\right], \tag{2.31}$$

where $c'(k)$ is a polynomial of degree ≤ 3 in k, and $n \in \{k-3, k-2, k-1, k\}$. Our goal is to bound the second derivative independently of t, so that we are able to use the Taylor approximation

$$|\varphi(1) - \varphi(0) - \varphi'(0)| \leq \sup_{0 \leq t \leq 1} |\varphi''(t)|\,. \tag{2.32}$$

Since prior P_x has bounded support, Hölder's inequality implies that (2.31) is bounded by

$$\begin{aligned} &N K(k,\lambda)\,\mathbb{E}\left[\left\langle \left|R_{1,2}^- R_{a,b}^- R_{c,d}^-\right|\right\rangle_t^p\right]^{1/p} \mathbb{E}\left[|Y(t)|^{nq}\right]^{1/q} \\ &\qquad \leq N K(k,\lambda)\,\mathbb{E}\left[\left\langle |R_{1,2}^-|^{3p}\right\rangle_t\right]^{1/p} \mathbb{E}\left[|Y(t)|^{nq}\right]^{1/q}, \end{aligned}$$

where $1/p + 1/q = 1$. The last bound follows from Jensen's inequality (since $p \geq 1$) and another application of Hölder's inequality. We let $p = 4/3$ and $q = 4$. Using Lemma 4.19 and the convergence of the fourth moment, Theorem 2.9, we have

$$\mathbb{E}\left\langle (R_{1,2}^-)^4\right\rangle_t \leq K(\lambda)\,\mathbb{E}\left\langle (R_{1,2}^-)^4\right\rangle \leq \frac{K(\lambda)}{N^2}.$$

We use the following lemma to bound the moments of $Y(t)$:

Lemma 2.12. *For all $\lambda < \lambda_c$ and integers k, there exists a constant $K(k,\lambda) \geq 0$ such that for all $t \in [0,1]$*

$$\mathbb{E}\left[Y(t)^{2k}\right] \leq K(k,\lambda).$$

Proof. Taking a derivative w.r.t. time, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\left[Y(t)^k\right] = -\frac{\lambda k}{2}\,\mathbb{E}\left[\left\langle x_N^{(1)} x_N^{(2)} R_{1,2}^-\right\rangle_t Y(t)^{k-1}\right] + \lambda k\,\mathbb{E}\left[\left\langle x_N^{(1)} x_N^{*} R_{1,*}^-\right\rangle_t Y(t)^{k-1}\right]$$

$$+\frac{\lambda k(k-1)}{2}\,\mathbb{E}\left[\left\langle x_N^{(1)}x_N^{(2)}R^-_{1,2}\right\rangle_t Y(t)^{k-2}\right].$$

By Hölder's inequality and boundedness of the variables and overlaps,

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\left[Y(t)^k\right]\right| \leq K(k,\lambda)\left(\mathbb{E}\left[|Y(t)|^k\right]^{1-1/k}+\mathbb{E}\left[|Y(t)|^k\right]^{1-2/k}\right).$$

The first term is generated by the terms involving $Y(t)^{k-1}$ in the derivative, and the second term comes from the one involving $Y(t)^{k-2}$. Since k is even, we drop the absolute values on the right-hand side. Next, use the fact $x^a \leq 1+x$ for all $x \geq 0$ and $0 \leq a \leq 1$, then we use Grönwall's lemma to conclude. ■

Therefore by the above estimates we have

$$\sup_{0\leq t\leq 1}|\varphi''(t)| \leq \frac{K(k,\lambda)}{\sqrt{N}}. \tag{2.33}$$

Now, our next goal is to prove

$$\left|\varphi'(0)-\lambda N\,\mathbb{E}\left[\langle R_{1,2}^2\rangle X(\lambda)^k\right]\right| \leq \frac{K(k,\lambda)}{\sqrt{N}}. \tag{2.34}$$

We consider the function

$$\psi(t) := \lambda N\,\mathbb{E}\left[\left\langle (R^-_{1,2})^2\right\rangle_t Y(t)^k\right].$$

Observe that (2.30) tells us $\psi(0)=\varphi'(0)$. On the other hand,

$$\begin{aligned}\left|\psi(1)-\lambda N\,\mathbb{E}\left[\langle R_{1,2}^2\rangle X(\lambda)^k\right]\right| &\leq 2\lambda\,\mathbb{E}\left[\left\langle\left|R^-_{1,2}x_N^{(1)}x_N^{(2)}\right|\right\rangle|X(\lambda)|^k\right]\\ &\quad+\frac{\lambda}{N}\,\mathbb{E}\left[\left\langle (x_N^{(1)}x_N^{(2)})^2\right\rangle|X(\lambda)|^k\right].\end{aligned}$$

Using Lemma 2.12 and Hölder's inequality, the first term is bounded by

$$K(k,\lambda)(\mathbb{E}\langle (R^-_{1,2})^2\rangle)^{1/2} \leq K(k,\lambda)/\sqrt{N},$$

and the second term is bounded by $K(k,\lambda)/N$. So it suffices to show that

$$\sup_{0\leq t\leq 1}|\psi'(t)| \leq \frac{K(k,\lambda)}{\sqrt{N}}.$$

Similarly to φ, the derivative of ψ is a sum of terms of the form

$$\lambda^2 N c(k)\,\mathbb{E}\left[\left\langle (R^-_{1,2})^2 R^-_{a,b}x_N^{(a)}x_N^{(b)}\right\rangle_t Y(t)^n\right].$$

It is clear that the same method used to bound φ'' (the generic term of which is (2.31)) also works in this case, so we obtain the desired bound on ψ'. Finally, using (2.32), (2.33) and (2.34), we obtain

$$N\,\mathbb{E}\left[\langle R_{1,2}^2\rangle X(\lambda)^k\right] - \mathbb{E}\left[\left\langle (x_N^{(1)}x_N^{(2)})^2\right\rangle X(\lambda)^k\right] = \lambda N\,\mathbb{E}\left[\langle R_{1,2}^2\rangle X(\lambda)^k\right] + \delta,$$

where $|\delta| \leq K(k,\lambda)/\sqrt{N}$. This is equivalent to (2.26) and closes the first stage of the argument. Now we need to show that

$$\mathbb{E}\left[\left\langle (x_N^{(1)}x_N^{(2)})^2\right\rangle X(\lambda)^k\right] = \mathbb{E}\left[X(\lambda)^k\right] + \delta.$$

We similarly consider the function

$$\psi(t) = \mathbb{E}\left[\left\langle (x_N^{(1)}x_N^{(2)})^2\right\rangle_t Y(t)^k\right].$$

We have

$$\psi(0) = \mathbb{E}\left[\left\langle (x_N^{(1)}x_N^{(2)})^2\right\rangle_0\right] \cdot \mathbb{E}\left[Y(0)^k\right] = \mathbb{E}_{P_\mathrm{x}}[X^2]^2 \cdot \mathbb{E}\left[Y(0)^k\right] = \mathbb{E}\left[Y(0)^k\right].$$

The derivative of ψ is a sum of term of the form

$$\lambda c(k)\,\mathbb{E}\left[\left\langle (x_N^{(1)}x_N^{(2)})^2 R_{a,b}^{-} x_N^{(a)} x_N^{(b)}\right\rangle_t Y(t)^n\right].$$

By our earlier argument, $|\psi'(t)| \leq K(k,\lambda)/\sqrt{N}$ for all t. We similarly argue that $\left|\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}[Y(t)^k]\right| \leq K(k,\lambda)/\sqrt{N}$ for all t, so that

$$\left|\psi(1) - \mathbb{E}\left[Y(1)^k\right]\right| \leq \frac{K(k,\lambda)}{\sqrt{N}}.$$

This yields (2.27) and thus concludes the proof.

2.7 Overlap convergence

In this section we prove Theorem 2.9 on the convergence of the overlaps to 0 under $\mathbb{P}_\lambda$, and below λ_c. At a high level, we will first prove that the overlap $R_{1,2}$ converges in probability to zero under $\mathbb{E}\langle\cdot\rangle$: for all $\epsilon > 0$,

$$\mathbb{E}\langle \mathbb{1}\{|R_{1,2}| \geq \epsilon\}\rangle \leq Ke^{-cN}.$$

This will be achieved via an interpolation bound at fixed overlap, combined with a concentration-of-measure argument. Next, this crude bound is boosted to a statement of convergence of the second moment:

$$\mathbb{E}\langle R_{1,2}^2\rangle \leq \frac{K}{N},$$

which is in turn boosted to a statement of convergence of the fourth moment:

$$\mathbb{E}\langle R_{1,2}^4\rangle \leq \frac{K}{N^2}.$$

The apparent recursive nature of this argument is a feature of the cavity method: one can control higher order quantities once one knows how to control low order ones.

Notice that all the statements above are about overlaps between two replicas. By the Nishimori property, everything could be equivalently stated in terms of the overlap of one replica with the spike $\boldsymbol{x}^*$. This is important since we only know how to control $R_{1,*}$ in the arguments to come.

Interpolation bound at fixed overlap

We find it iseful to introduce the so-called *Franz-Parisi (FP) potential* (Franz and Parisi, 1995, 1998). For $\boldsymbol{x}^* \in \mathbb{R}^N$ fixed, $m \in \mathbb{R} \setminus \{0\}$ and $\epsilon > 0$ define the set

$$A = \begin{cases} R_{1,*} \in [m, m+\epsilon) & \text{if } m > 0, \\ R_{1,*} \in (m-\epsilon, m] & \text{if } m < 0. \end{cases}$$

Now define the FP potential as

$$\Phi_\epsilon(m, \boldsymbol{x}^*) := \frac{1}{N}\,\mathbb{E}_{\boldsymbol{W}} \log \int \mathbb{1}\{\boldsymbol{x} \in A\} e^{-H(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}), \tag{2.35}$$

where the expectation is only over the Gaussian disorder $\boldsymbol{W}$. This is the free energy of a subsystem of configurations having an overlap close to a fixed value m with the planted signal $\boldsymbol{x}^*$. It is clear that $\Phi_\epsilon(m; \boldsymbol{x}^*) \leq \frac{1}{N}\mathbb{E}_{\boldsymbol{W}} \log L(\boldsymbol{Y}, \lambda)$, where $\boldsymbol{Y}$ is the upper triangular part of the matrix $\sqrt{\frac{\lambda}{N}}\boldsymbol{x}^*\boldsymbol{x}^{*\top} + \boldsymbol{W}$. We will prove that when $|m| > 0$, then there is a sizable gap between $\Phi_\epsilon(m; \boldsymbol{x}^*)$ and $\frac{1}{N}\mathbb{E}_{\boldsymbol{W}} \log L(\boldsymbol{Y}, \lambda)$. This estimate is a main ingredient is our proof of overlap concentration.

For $r \geq 0$ and $s \in \mathbb{R}$, we let

$$\widehat{\psi}(r, s) := \mathbb{E}_z \log \int \exp\left(\sqrt{r}zx + sx - \frac{r}{2}x^2\right) \mathrm{d}P_{\mathrm{x}}(x). \tag{2.36}$$

and

$$\begin{aligned} \bar{\psi}(r, s) &:= \mathbb{E}_{x^*}\, \widehat{\psi}(r, sx^*) \\ &= \mathbb{E}_{x^*, z} \log \int \exp\left(\sqrt{r}zx + sxx^* - \frac{r}{2}x^2\right) \mathrm{d}P_{\mathrm{x}}(x). \end{aligned} \tag{2.37}$$

We see that $\bar{\psi}(r, r) = \psi(r)$, but unlike ψ, the function $\bar{\psi}$ does not have an interpretation as the KL between two distributions. The next lemma states a key property of this function that will be useful later on:

Lemma 2.13. *For all $r \geq 0$, $\bar{\psi}(r,-r) \leq \bar{\psi}(r,r)$.*

Proof. This is a special case of the more general Lemma 4.15, proved in Chapter 4. ■

Aditionally, for $\boldsymbol{x}^* \in \mathbb{R}^N$ fixed, we define the function

$$\widehat{F}(\lambda, m, q) := \frac{1}{N}\sum_{i=1}^{N} \widehat{\psi}(\lambda q, \lambda m x_i^*) - \frac{\lambda}{2}m^2 + \frac{\lambda}{4}q^2.$$

Recall that $\mathbb{E}_{x^*}\,\widehat{F}(\lambda, q, q)$ is the RS potential $F(\lambda, q)$ from (2.6).

Proposition 2.14. *Fix $m \in \mathbb{R}$, $\epsilon > 0$ and $\lambda \geq 0$. There exist constants $K = K(\lambda) > 0$ such that*

$$\Phi_\epsilon(m; \boldsymbol{x}^*) \leq \widehat{F}\big(\lambda, |m|, m\big) + \frac{\lambda\epsilon^2}{2} + \frac{K}{N}.$$

Proof. To obtain a bound on $\Phi_\epsilon(m; \boldsymbol{x}^*)$ we use the interpolation method with Hamiltonian

$$\begin{aligned}-H_t(\boldsymbol{x}) := &\sum_{i<j} \sqrt{\frac{t\lambda}{N}} W_{ij} x_i x_j + \frac{t\lambda}{N} x_i x_i^* x_j x_j^* - \frac{t\lambda}{2N} x_i^2 x_j^2 \\ &+ \sum_{i=1}^{N} \sqrt{(1-t)\lambda|m|}\, z_i x_i + (1-t)\lambda m x_i x_i^* - \frac{(1-t)\lambda|m|}{2} x_i^2.\end{aligned}$$

by varying $t \in [0,1]$. The r.v.'s W, z are all i.i.d. standard Gaussians independent of everything else. We define

$$\varphi(t) := \frac{1}{N}\,\mathbb{E}_{\boldsymbol{W},\boldsymbol{z}} \log \int \mathbb{1}\{\boldsymbol{x} \in A\} e^{-H_t(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}).$$

We compute the derivative w.r.t. t, and use Gaussian integration by prts to obtain

$$\begin{aligned}\varphi'(t) = &-\frac{\lambda}{4}\,\mathbb{E}\left\langle (R_{1,2} - |m|)^2 \right\rangle_t + \frac{\lambda t}{4}|m|^2 + \frac{\lambda}{4N^2}\sum_{i=1}^{N} \mathbb{E}\left\langle {x_i^{(1)}}^2 {x_i^{(2)}}^2 \right\rangle_t \\ &+ \frac{\lambda}{2}\,\mathbb{E}\left\langle (R_{1,*} - m)^2 \right\rangle_t - \frac{\lambda}{2}m^2 - \frac{\lambda}{2N^2}\sum_{i=1}^{N} \mathbb{E}\left\langle {x_i}^2 {x_i^*}^2 \right\rangle_t,\end{aligned}$$

where $\langle\cdot\rangle_t$ is the Gibbs average w.r.t. the Hamiltonian $-H_t(\boldsymbol{x}) + \log \mathbb{1}\{\boldsymbol{x} \in A\}$. A few things now happen. Notice that the planted term (first term in the second line) is trivially smaller than $\lambda\epsilon^2/2$ due to the overlap restriction. Moreover, the last terms in both lines are of order $1/N$ since the variables x_i are bounded. The first term in the first line, which involves the overlap between two replicas, is more challenging. What makes this term difficult to control is that the Gibbs measure $\langle\cdot\rangle_t$ no longer satisfies the Nishimori property due to the overlap restriction, so the overlap between two replicas no longer has the same distribution as the overlap of one replica with the planted spike. Fortunately, this term is always non-positive so we can ignore it altogether and obtain an upper bound:

$$\varphi'(t) \le -\frac{\lambda}{4}m^2 + \frac{\lambda\epsilon^2}{2} + \frac{\lambda K}{N}.$$

Integrating over t, we get

$$\Phi_\epsilon(m;\boldsymbol{x}^*) \le \varphi(0) - \frac{\lambda}{4}m^2 + \frac{\lambda\epsilon^2}{2} + \frac{\lambda K}{N}.$$

Finally, by dropping the indicator, we have

$$\begin{aligned}\varphi(0) &= \frac{1}{N}\mathbb{E}_{\boldsymbol{z}} \log \int \mathbb{1}\{\boldsymbol{x} \in A\} e^{-H_0(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) \\ &\le \frac{1}{N}\mathbb{E}_{\boldsymbol{z}} \log \int e^{-H_0(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) \\ &= \frac{1}{N}\sum_{i=1}^N \mathbb{E}_z \log \int \exp\left(\sqrt{\lambda|m|}zx_i + \lambda m x x_i^* - \frac{\lambda|m|}{2}x^2\right) \mathrm{d}P_{\mathrm{x}}(x) \\ &= \frac{1}{N}\sum_{i=1}^N \widehat{\psi}(\lambda|m|, \lambda m x_i^*).\end{aligned}$$

■

Convergence in probability of the overlaps

A consequence of the above proposition is the convergence in probability of the overlaps:

Proposition 2.15. *For all $\lambda < \lambda_c$ and $\epsilon > 0$, there exist constants $K = K(\lambda,\epsilon) \ge 0,\ c = c(\lambda,\epsilon,P_{\mathrm{x}}) \ge 0$ such that*

$$\mathbb{E}\left\langle \mathbb{1}\left\{|R_{1,*}| \ge \epsilon\right\}\right\rangle \le Ke^{-cN}.$$

To prove the above proposition, we first show that the partition function of the model enjoys sub-Gaussian concentration in logarithmic scale. This is an elementary consequence of two classical concentration-of-measure results: concentration of Lipschitz functions of Gaussian random variables, and concentration of *convex* Lipschitz function of *bounded* random variables.

Lemma 2.16. *Fix $\boldsymbol{x}^* \in \mathbb{R}^N$ and let A be a Borel subset of $\mathbb{R}^N$. Define the random variable*

$$Z := \int_A e^{-H(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}),$$

where the randomness comes from the Gaussian disorder $\boldsymbol{W}$. There exist a constant $K > 0$ depending on λ and P_{x} such that for all $u \ge 0$,

$$\Pr\left(|\log Z - \mathbb{E}\log Z| \ge Nu\right) \le 2e^{-Nu^2/K}.$$

Proof. We notice that the map $\boldsymbol{W} \mapsto \frac{1}{N}\log Z$ is Lipschitz with constant $K\sqrt{\frac{\lambda}{N}}$ for every $\boldsymbol{x}^* \in \mathbb{R}^N$. Then we invoke the Borell-Tsirelson-Ibragimov-Sudakov inequality of concentration of Lipschitz functions of Gaussian r.v.'s. See Boucheron, Lugosi, and Massart (2013). ■

Lemma 2.17. *Define the random variable*

$$\mathtt{f} := \frac{1}{N}\,\mathbb{E}_{\boldsymbol{W}} \log \int e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}),$$

where the randomness comes from the planted vector $\boldsymbol{x}^*$. There exist a constant $K > 0$ depending on λ and P_{x} such that for all $u \geq 0$,

$$\Pr\left(|\mathtt{f} - \mathbb{E}\,\mathtt{f}| \geq u\right) \leq 2e^{-Nu^2/K}.$$

Proof. We notice that the map $\boldsymbol{x}^* \mapsto \mathtt{f}$ is Lipschitz with constant $K\frac{\lambda}{\sqrt{N}}$ and convex. Moreover, the coordinates x_i^* are bounded. We then invoke Talagrand's inequality on the concentration convex Lipschitz functions of bounded r.v.'s. See Boucheron, Lugosi, and Massart (2013). ■

Lemma 2.18. *There exist a constant $K > 0$ depending on λ, m and P_{x} such that for all $u \geq 0$,*

$$\Pr\left(\left|\sum_{i=1}^N \widehat{\psi}(\lambda|m|, \lambda m x_i^*) - \bar{\psi}(\lambda|m|, \lambda m)\right| \geq Nu\right) \leq 2e^{-Nu^2/K}.$$

Proof. Since $\left|\partial_s\widehat{\psi}(r, sx^*)\right| \leq K^2$, $\left|\partial_r\widehat{\psi}(r, sx^*)\right| \leq K^2/2$ and $\widehat{\psi}(0,0) = 0$, where K is a bound on the radius of the support of $P_{\boldsymbol{x}}$, we $\left|\widehat{\psi}(r, sx^*)\right| \leq K^2(r/2 + s)$. The claim now follows from Hoeffding's inequality. ■

***Proof** of Proposition 2.15.* For $\epsilon, \epsilon' > 0$, we can write the decomposition

$$\begin{aligned}\mathbb{E}\left\langle \mathbb{1}\left\{|R_{1,*}| \geq \epsilon\right\}\right\rangle = &\sum_{l\geq 0} \mathbb{E}\left\langle \mathbb{1}\big\{R_{1,*} - \epsilon \in [l\epsilon', (l+1)\epsilon')\big\}\right\rangle \\ &+ \sum_{l \geq 0} \mathbb{E}\left\langle \mathbb{1}\big\{-R_{1,*} - \epsilon \in [l\epsilon', (l+1)\epsilon')\big\}\right\rangle,\end{aligned}$$

where the integer index l ranges over a finite set of size $\leq K/\epsilon'$ since the prior P_{x} has bounded support. We will only treat the first sum in the above expression since the argument extends trivially to the second sum. Let $A = \big\{R_{1,*} - \epsilon \in [l\epsilon', (l+1)\epsilon')\big\}$ and write

$$\mathbb{E}\left\langle \mathbb{1}\{\boldsymbol{x} \in A\}\right\rangle = \mathbb{E}_{\boldsymbol{x}^*}\,\mathbb{E}_{\boldsymbol{W}}\left[\frac{\int_A e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})}{\int e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})}\right]. \tag{2.38}$$

In virtue of Lemma 2.16 the two quantities in the above fraction enjoy sub-Gaussian concentration in logarithmic scale over the Gaussian disorder. For any given l and $u \geq 0$, we simultaneously have

$$\frac{1}{N}\log\int e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) \geq \frac{1}{N}\mathbb{E}_{\boldsymbol{W}}\log\int e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) - u.$$

and

$$\begin{aligned}\frac{1}{N}\log\int_A e^{-H_t(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) &\leq \frac{1}{N}\mathbb{E}_{\boldsymbol{W}}\log\int_A e^{-H_t(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) + u\\ &= \Phi_{\epsilon'}(\epsilon + l\epsilon'; \boldsymbol{x}^*) + u,\end{aligned}$$

with probability at least $1 - 2e^{-Nu^2/K}$. On the complement of this event, we simply bound the fraction in (2.38) by 1. Combining the above bounds we obtain

$$\mathbb{E}\left\langle \mathbb{1}\{\boldsymbol{x}\in A\}\right\rangle \leq 2e^{-Nu^2/K} + \mathbb{E}_{\boldsymbol{x}^*}\left[e^{N(\Delta+2u)}\right],$$

where

$$\Delta = \Phi_{\epsilon'}(m;\boldsymbol{x}^*) - \frac{1}{N}\mathbb{E}_{\boldsymbol{W}}\log\int e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}),$$

with $m = \epsilon + l\epsilon'$. By Proposition 2.14, $\Phi_{\epsilon'}$ is upper-bounded by a quantity that concentrates over the randomness of $\boldsymbol{x}^*$. We use Lemma 2.17 and Lemma 2.18 in the same way we used Lemma 2.16: for $u' \geq 0$, we simultaneously have

$$\Phi_{\epsilon'}(m;\boldsymbol{x}^*) \leq F(\lambda,|m|,m) + \frac{\lambda\epsilon^2}{2} + \frac{\lambda K}{N} + u',$$

and

$$\frac{1}{N}\mathbb{E}_{\boldsymbol{W}}\log\int e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) \geq F_N - u'$$

with probability at least $1 - 4e^{-Nu'^2/K}$, where

$$F_N = \mathbb{E}_{\boldsymbol{W},\boldsymbol{x}^*}\log\int e^{-H}\mathrm{d}P^{\otimes N} = \mathbb{E}_{\mathbb{P}_\lambda}\log L(\boldsymbol{Y};\lambda).$$

Moreover, by Lemma 2.13, we have $F(\lambda,|m|,m) \leq F(\lambda,|m|,|m|) \equiv F(\lambda,m)$. Therefore

$$\mathbb{E}_{\boldsymbol{x}^*}\left[e^{N\Delta}\right] \leq \exp\left(N(F(\lambda,|m|) - F_N + 2u')\right) + 4e^{-Nu'^2/K}.$$

The second term is obtained on the complement low probability event and noting that $\Delta \leq 0$.

Now we know that $F_N \simeq \sup_q F(\lambda,q)$. More precisely, we use a result of Krzakala, Xu, and Zdeborová (2016) stating that

$$F_N \geq \sup_{q\geq 0} F(\lambda,q) - \frac{K}{N} = \phi_{\mathsf{RS}}(\lambda) - \frac{K}{N}.$$

(In Chapter 4, we give a relatively short proof of the fact $F_N \to \sup_q F(\lambda, q)$, including the above bound.)

When $\lambda < \lambda_c$, $q = 0$ is the unique maximizer of the RS potential. Therefore $F(\lambda, |m|) - F_N < -c(\epsilon) < 0$ for all $|m| > \epsilon$. We therefore obtain

$$\mathbb{E}\left\langle \mathbb{1}\{\boldsymbol{x} \in A\}\right\rangle \leq 2e^{-Nu^2/K} + 4e^{-Nu'^2/K+2Nu} + e^{N(-c(\epsilon)+2u+2u')}.$$

Finally, adjusting the parameters u, u' yields the desired result (e.g., $u' = c(\epsilon)/3$ and $u = c(\epsilon)^2/36 \wedge c(\epsilon)/9$). ■

Convergence of the second moment

In this subsection we prove the convergence of the second moment of the overlaps: $\mathbb{E}\langle R_{1,*}^2\rangle \leq \frac{K}{N}$. The following lemma will be useful.

Lemma 2.19. *For all $t \in [0,1]$, and all $\tau_1, \tau_2 > 0$ such that $1/\tau_1 + 1/\tau_2 = 1$,*

$$|\nu_t(f) - \nu_0(f)| \leq K(\lambda, n)\nu\left(\left|R_{1,*}^{-}\right|^{\tau_1}\right)^{1/\tau_1} \cdot \nu\left(|f|^{\tau_2}\right)^{1/\tau_2} \tag{2.39}$$

$$|\nu_t(f) - \nu_0(f) - \nu_0'(f)| \leq K(\lambda, n)\nu\left(\left|R_{1,*}^{-}\right|^{\tau_1}\right)^{1/\tau_1} \cdot \nu\left(|f|^{\tau_2}\right)^{1/\tau_2}. \tag{2.40}$$

Proof. We use Taylor's approximations

$$|\nu_t(f) - \nu_0(f)| \leq \sup_{0\leq s\leq 1} |\nu_s'(f)|,$$
$$|\nu_s(f) - \nu_0(f) - \nu_0'(f)| \leq \sup_{0\leq s\leq 1} |\nu_s''(f)|,$$

then Lemma 4.18 and the triangle inequality to bound the right hand sides, then Hölder's inequality to bound each term in the derivative, and then we apply Lemma 4.19. (To compute the second derivative, one need to use Lemma 4.18 recursively.) ■

By symmetry between sites, we have

$$\mathbb{E}\langle R_{1,*}^2\rangle = \mathbb{E}\langle x_N x_N^* R_{1,*}\rangle = \frac{1}{N}\mathbb{E}\langle (x_N x_N^*)^2\rangle + \mathbb{E}\langle x_N x_N^* R_{1,*}^{-}\rangle.$$

By the first bound (4.32) of Lemma 4.20 with $\tau_1 = 1$, $\tau_2 = \infty$, we have

$$\nu((x_N x_N^*)^2) = \nu_0((x_N x_N^*)^2) + \delta = \mathbb{E}_{P_{\mathrm{x}}}[X^2]^2 + \delta = 1 + \delta,$$

with $|\delta| \leq K(\lambda)\nu(|R_{1,*}^{-}|)$. On the other hand, by the second bound (4.33) with $\tau_1 = 1$, $\tau_2 = \infty$, we get

$$\nu(R_{1,*}^{-} x_N x_N^*) = \nu_0'(R_{1,*}^{-} x_N x_N^*) + \delta.$$

This is because $\nu_0(R^-_{1,*}x_N x^*_N) = 0$, since last variable x_N decouples from the remaining $N-1$ variables under the measure ν_0. Now, we use Lemma 4.18 with $n = 1$, to evaluate the above derivative at $t = 0$. We still write $y^{(l)} = x^{(l)}_N$.

$$\begin{aligned}\nu_0'(R^-_{1,*}x_N x^*_N) &= -\lambda\nu_0(y^{(1)}y^{(2)}y^{(1)}y^* R^-_{1,*}R^-_{1,2}) + \lambda\nu_0(y^{(1)}y^*y^{(1)}y^*R^{-2}_{1,*})\\ &\quad - \lambda\nu_0(y^{(2)}y^*y^{(1)}y^*R^-_{1,*}R^-_{2,*}) + \lambda\nu_0(y^{(2)}y^{(3)}y^{(1)}y^*R^-_{1,*}R^-_{2,3})\\ &= \lambda\nu_0(R^{-2}_{1,*}).\end{aligned}$$

In the above, the only term that survived is the second one since all variables y appearing in it are squared. We now use Lemma 4.20 to argue that $\nu_0(R^{-2}_{1,*}) \simeq \nu_1(R^{-2}_{1,*})$. We apply the estimate (4.32) with $t = 1$, $\tau_1 = 3$ and $\tau_2 = 3/2$ to obtain

$$\nu_0(R^{-2}_{1,*}) = \nu(R^{-2}_{1,*}) + \delta$$

with $|\delta| \le K(\lambda)\nu(|R^-_{1,*}|^3)$. Moreover,

$$\nu(R^{-2}_{1,*}) = \nu((R_{1,*} - \frac{1}{N}yy^*)^2) = \nu(R^2_{1,*}) - \frac{2}{N}\nu(yy^*R_{1,*}) + \frac{1}{N^2}\nu((yy^*)^2).$$

The third term is of order $1/N^2$, and the second term is bounded by $\frac{1}{N}\nu(|R_{1,*}|)$. Therefore

$$\nu_0((R^-_{1,*})^2) = \nu(R^2_{1,*}) + \delta',$$

with

$$|\delta'| \le K(\lambda)\left(\frac{1}{N}\nu(|R^-_{1,*}|) + \nu(|R^-_{1,*}|^3) + \frac{1}{N^2}\right).$$

Putting things together, we have proved that

$$\nu(R^2_{1,*}) = \frac{1}{N} + \lambda\nu(R^2_{1,*}) + \delta, \tag{2.41}$$

where

$$|\delta| \le K(\lambda)\left(\frac{1}{N}\nu(|R^-_{1,*}|) + \nu(|R^-_{1,*}|^3) + \frac{1}{N^2}\right). \tag{2.42}$$

Now we need to control the error term δ. By elementary manipulations,

$$\nu(|R^-_{1,*}|) \le \nu(|R_{1,*}|) + \frac{K}{N},$$

and

$$\nu(|R^-_{1,*}|^3) \le \nu(|R_{1,*}|^3) + \frac{K}{N}\nu(R^2_{1,*}) + \frac{K}{N^2}\nu(|R_{1,*}|) + \frac{K}{N^3}.$$

Therefore, from (4.39) we have

$$|\delta| \le K\left(\nu(|R_{1,*}|^3) + \frac{1}{N}\nu(R^2_{1,*}) + \frac{1}{N}\nu(|R_{1,*}|) + \frac{1}{N^2}\right). \tag{2.43}$$

At this point, the apriori knowledge that $R_{1,*}$ is small with high probability is useful. It implies that $\nu(|R_{1,*}|) \ll 1$ and $\nu(|R_{1,*}|^3) \ll \nu(R_{1,*}^2)$. With Proposition 2.15 we have for $\epsilon > 0$

$$\nu(|R_{1,*}|) \leq \epsilon + K(\epsilon)e^{-c(\epsilon)N},$$

and

$$\nu(|R_{1,*}|^3) \leq \epsilon\nu(R_{1,*}^2) + K(\epsilon)e^{-c(\epsilon)N}.$$

Combining the above two bounds with (4.42), and then injecting in (2.41), we get

$$\nu(R_{1,*}^2) \leq \frac{1}{N} + (\lambda + \frac{K}{N} + K\epsilon)\nu(R_{1,*}^2) + \frac{K}{N^2} + Ke^{-cN}.$$

We choose ϵ small enough and N large enough that $K(\epsilon + \frac{1}{N}) < 1 - \lambda$. We obtain

$$\nu\left(R_{1,*}^2\right) \leq \frac{K}{N} + \frac{K}{N^2} + Ke^{-cN}.$$

Convergence of the fourth moment

In this subsection we prove the convergence of the fourth moment: $\mathbb{E}\langle R_{1,*}^4\rangle \leq \frac{K}{N^2}$. We adopt the same technique based on the cavity method, with the extra knowledge that the second moment converges. Many parts of the argument are exactly the same so we will only highlight the main novelties in the proof. Let By symmetry between sites,

$$\begin{aligned}\nu(R_{1,*}^4) &= \nu\left(R_{1,*}^3 x_N x_N^*\right)\\ &= \nu((R_{1,*}^-)^3 x_N x_N^*) + \frac{3}{N}\nu((R_{1,*}^-)^2(x_N x_N^*)^2) + \frac{3}{N^2}\nu(R_{1,*}^-(x_N x_N^*)^3) + \frac{1}{N^3}\nu((x_N x_N^*)^4).\end{aligned}$$

The quadratic term is bounded as

$$\nu((R_{1,*}^-)^2(x_N x_N^*)^2) \leq K\nu((R_{1,*}^-)^2) \leq \frac{K}{N}.$$

The last inequality is using our extra knowledge about the convergence of the second moment. The last two terms are also bounded by K/N^2 and K/N^3 respectively. Now we must deal with the cubic term, and here, we apply the exact same technique used to deal with the term $\nu(R_{1,*}^- x_N x_N^*)$ in the previous proof. The argument applies verbatim and we obtain

$$\nu(R_{1,*}^4) \leq \frac{K}{N^2} + \lambda\nu(R_{1,*}^4) + K\nu(|R_{1,*}^-|^5) + K\nu(|R_{1,*}^-|^3)$$

Using Proposition 2.15, we have for $\epsilon > 0$,

$$\nu(|R_{1,*}|^5) \leq \epsilon\nu(R_{1,*}^4) + K(\epsilon)e^{-c(\epsilon)N},$$

$$\nu(|R_{1,*}|^3) \leq \epsilon\nu(R_{1,*}^2) + K(\epsilon)e^{-c(\epsilon)N}.$$

Now, we finish the argument in the same way, by choosing ϵ sufficiently small. This concludes the proof of Theorem 2.9.

Chapter 3

Detection limits in the spiked rectangular model

In this chapter we consider an asymmetric version of the model previously discussed. Concretely we consider the observation of an $N \times M$ matrix of the form

$$\boldsymbol{Y} = \sqrt{\frac{\beta}{N}}\boldsymbol{u}\boldsymbol{v}^\top + \boldsymbol{W}, \tag{3.1}$$

where $\boldsymbol{u}$ and $\boldsymbol{v}$ are unknown factors and $\boldsymbol{W}$ is a matrix with i.i.d. noise entries. We will assume as before that the noise is standard Gaussian. The parameter β represents the strength of the spike. Notice that the model has a new degree of freedom that was not present before: the aspect ratio of the matrix $\boldsymbol{Y}$. We assume a high-dimensional setting where $M/N \to \alpha$. When the factors are independent, model (3.1) can be viewed as a linear model with additive noise and scalar random design:

$$\boldsymbol{y}_j = \bar{\beta} v_j \boldsymbol{u} + \boldsymbol{w}_j,$$

with $1 \le j \le M$, $\bar{\beta} = \sqrt{\beta/N}$. Assuming v_j has zero mean and unit variance, this is a model of *spiked covariance*: the mean of the empirical covariance matrix $\widehat{\boldsymbol{\Sigma}} = \frac{1}{M}\sum_{j=1}^{M} \boldsymbol{y}_j \boldsymbol{y}_j^\top$ is a rank-one perturbation of the identity: $\boldsymbol{I}_N + \frac{\beta}{N}\boldsymbol{u}\boldsymbol{u}^\top$.

As before, we want to test whether $\beta > 0$ or $\beta = 0$, so we will look at the behavior of the likelihood ratio.

3.1 Background and related work

The introduction of a particular spiked covariance model by Johnstone (2001)—one corresponding to the special case $v_j \sim \mathcal{N}(0,1)$—has provided the foundations for a rich theory of Principal Component Analysis (PCA), in which the performance of several important tests and estimators is by now well understood (see, e.g., Amini and Wainwright, 2009; Berthet and Rigollet, 2013; Dobriban, 2017; Johnstone and Lu, 2009; Ledoit and Wolf, 2002; Nadler, 2008; Paul, 2007). Parallel developments in random matrix theory have unveiled the existence of sharp transition phenomena in the behavior of the spectrum of the data matrix, where for a spike of strength above a certain *spectral* threshold, the top eigenvalue separates from the remaining eigenvalues which are packed together in a "bulk" and thus indicates the presence of the spike; below this threshold, the top eigenvalue converges to the edge of the bulk. See Benaych-Georges and Nadakuditi (2011, 2012); Capitaine, Donati-Martin, and Féral (2009); Féral and Péché (2007); Péché (2006) for results on low-rank deformations of Wigner matrices, and Bai and Yao (2008, 2012); Baik, Ben Arous, and Péché (2005); Baik and Silverstein (2006) for results on spiked covariance models. More recently, an intense research effort has been undertaken to pin down the fundamental limits for both estimating and detecting the spike.

In a series of papers (Barbier, Dia, et al., 2016; Deshpande, Abbé, and Montanari, 2016; Korada and Macris, 2009; Krzakala, Xu, and Zdeborová, 2016; Lelarge and Miolane, 2017; Miolane, 2017), the error of the Bayes-optimal estimator has been completely characterized for additive low-rank models with a separable (product) prior on the spike. In particular, these papers confirm an interesting phenomenon discovered by Lesieur, Krzakala, and Zdeborová (2015a,b), based on plausible but non-rigorous arguments: for certain priors on the spike, estimation becomes possible—although computationally expensive—below the spectral threshold $\beta = 1$. More precisely, the posterior mean overlaps with the spike in regions where the top eigenvector is orthogonal to it. Lesieur, Krzakala, and Zdeborová (2017) provides a full account of these phase transitions in a myriad of interesting situations, the majority of which still await rigorous treatment. As for the testing problem, Onatski, Moreira, and Hallin (2013, 2014) and Johnstone and Onatski (2015) considered the spiked covariance model for a uniformly distributed unit norm spike, and studied the asymptotics of the likelihood ratio (LR) of a spiked alternative against a spherical null. They showed that the log-LR is asymptotically Gaussian below the spectral threshold $\alpha\beta^2 = 1$ (which in this setting is known as the BBP threshold, after Baik, Ben Arous, and Péché, 2005), while it is divergent above it.

However their proof is intrinsically tied to the assumption of a spherical prior. Indeed, by rotational symmetry of the model, the LR depends only on the spectrum, the joint distribution of which is available in closed form. A representation of the LR in terms of a contour integral is then possible (in the single spike case), which can then be analyzed via the method of steepest descent. In a similar but unrelated effort, Baik and Lee (2016, 2017a,b) studied the fluctuations of the free energy of spherical, symmetric and bipartite versions of the Sherrington–Kirkpatrick (SK) model. This free energy coincides with the log-LR associated with the model (3.1) for a choice of parameters. The sphericity assumption is again key to their analysis, and both approaches require the execution of very delicate asymptotics and appeal to advanced results from random matrix theory.

3.2 Fluctuations below the BBP threshold

We consider the case of separable priors: we assume that the entries of $\boldsymbol{u}$ and $\boldsymbol{v}$ are independent and identically distributed from base priors $P_{\mathtt{u}}$ and $P_{\mathtt{v}}$, respectively, both having bounded support (boundedness is required for technical reasons. This unfortunately rules out the case where one factor is Gaussian.) We prove fluctuation results for the log-LR in this setting with entirely different methods than used for spherical priors. In particular, our proof is more probabilistic and operates through general principles. The tools we use come from the mathematical theory of spin glasses (see Talagrand, 2011a,b).

We assume that the priors $P_{\mathtt{u}}$ and $P_{\mathtt{v}}$ have zero mean, unit variance, and (as in the previous chapter) supports bounded in radius by $K_{\mathtt{u}}$ and $K_{\mathtt{v}}$ respectively. Let $\mathbb{P}_\beta$ be the probability distribution of the matrix $\boldsymbol{Y}$ as per (3.1). Define $L(\cdot;\beta)$ to be the likelihood ratio, or Radon-Nikodym derivative of $\mathbb{P}_\beta$ with respect to $\mathbb{P}_0$:

$$L(\cdot;\beta) \equiv \frac{\mathrm{d}\,\mathbb{P}_\beta}{\mathrm{d}\,\mathbb{P}_0}.$$

For a fixed $\boldsymbol{Y} \in \mathbb{R}^{N\times M}$, by conditioning on $\boldsymbol{u}$ and $\boldsymbol{v}$, we can write

$$L(\boldsymbol{Y};\beta) = \int \exp\Big(\sum_{i,j}\sqrt{\frac{\beta}{N}}Y_{ij}u_iv_j - \frac{\beta}{2N}u_i^2v_j^2\Big)\mathrm{d}P_{\mathtt{u}}^{\otimes N}(\boldsymbol{u})\mathrm{d}P_{\mathtt{v}}^{\otimes M}(\boldsymbol{v}).$$

The main result of this chapter is the following asymptotic distributional result.

Theorem 3.1. *Let $\alpha,\beta \geq 0$ such that $K_{\mathtt{u}}^4K_{\mathtt{v}}^4\alpha\beta^2 < 1$. Then in the limit $N \to \infty$ and $M/N \to \alpha$,*

$$\log L(\boldsymbol{Y};\beta) \rightsquigarrow \mathcal{N}\left(\pm\frac{1}{4}\log\left(1-\alpha\beta^2\right), -\frac{1}{2}\log\left(1-\alpha\beta^2\right)\right),$$

where "$\rightsquigarrow$" denotes convergence in distribution. The sign of the mean is $+$ under the null $\boldsymbol{Y}\sim\mathbb{P}_0$ and $-$ under the alternative $\boldsymbol{Y}\sim\mathbb{P}_\beta$.

We see that the region of parameters (α,β) we are able to cover with our proof method is optimal when (and only when) $P_{\mathtt{u}}$ and $P_{\mathtt{v}}$ are both symmetric Rademacher. This is in charp contrast with the symmetric case where we've been able to prove Gaussian fluctuations up to the optimal threshold λ_c for abritrary priors. This is intrinsically due to the bipartite nature of the problem, for which certain positivity arguments are not available. In Section 3.6 we formulate a conjecture on the *maximal* region in which the log-LR has asymptotically Gaussian fluctuations. This region is of course below the BBP threshold, but does *not* extend up to it in general.

A consequence of either one of the above statements and Le Cam's first lemma (Van der Vaart, 2000, Lemma 6.4) is the mutual contiguity (see Definition 1.1) between the null and the spiked alternative:

Corollary 3.2. *For $K_{\mathtt{u}}^4 K_{\mathtt{v}}^4 \alpha\beta^2 < 1$, the families of distributions $\mathbb{P}_0$ and $\mathbb{P}_\beta$ (indexed by M, N) are mutually contiguous in the limit $N \to \infty$, $M/N \to \alpha$.*

Contiguity implies impossibility of strong detection: there exists no test that, upon observing a random matrix $\boldsymbol{Y}$ with the promise that it is sampled either from $\mathbb{P}_0$ or $\mathbb{P}_\beta$, can tell which is the case with asymptotic certainty in this regime. We also mention that contiguity can be proved through the second-moment method and its conditional variants, as was done by Banks, Moore, Vershynin, et al. (2017); Montanari, Reichman, and Zeitouni (2015); Perry et al. (2016a) for closely related models. However, identifying the right event on which to condition in order to tame the second moment of L is a matter of a case-by-case deliberation. Study of the fluctuations of the log-LR appears to provide a more systematic route: the logarithm has a smoothing effect that kills the wild (but rare) events that otherwise dominate in the second moment. This being said, our result is optimal only in one special case:

When $P_{\mathtt{u}}$ and $P_{\mathtt{v}}$ are symmetric Rademacher, $K_{\mathtt{u}} = K_{\mathtt{v}} = 1$, and Theorem 3.1 covers the entire (α, β) region where such fluctuations hold. Indeed, for $\alpha\beta^2 > 1$, one can distinguish $\mathbb{P}_\beta$ from $\mathbb{P}_0$ by looking at the top eigenvalue of the empirical covariance matrix $\boldsymbol{Y}\boldsymbol{Y}^\top$ (Benaych-Georges and Nadakuditi, 2012). So the conclusion of Theorem 3.1 cannot hold in light of the above contiguity argument. Beyond this special case, our result is not expected to be optimal.

Limits of weak detection Since contiguity implies that testing errors are inevitable, it is natural to aim for tests $T : \mathbb{R}^{N\times M} \mapsto \{0, 1\}$ that minimize the sum of the Type-I and Type-II errors:

$$\mathsf{err}(T) = \mathbb{P}_0\left(T(\boldsymbol{Y}) = 1\right) + \mathbb{P}_\beta\left(T(\boldsymbol{Y}) = 0\right).$$

By the Neyman-Pearson lemma, the test minimizing the above error is the likelihood ratio test that rejects the null iff $L(\boldsymbol{Y}; \beta) > 1$. The optimal error is thus

$$\mathsf{err}^*_{M,N}(\beta) = \mathbb{P}_0\left(\log L(\boldsymbol{Y}; \beta) > 0\right) + \mathbb{P}_\beta\left(\log L(\boldsymbol{Y}; \beta) \le 0\right) = 1 - D_{\mathsf{TV}}(\mathbb{P}_\beta, \mathbb{P}_0).$$

The symmetry of the means under the null and the alternative in Theorem 3.1 implies that the above Type-I and Type-II errors are equal, and that the total error has a limit:

Corollary 3.3. *For $\alpha, \beta \ge 0$ such that $K_{\mathtt{u}}^4 K_{\mathtt{v}}^4 \alpha\beta^2 < 1$,*

$$\lim_{\substack{N\to\infty\\ M/N\to\alpha}} \mathsf{err}^*_{M,N}(\beta) = 1 - \lim_{\substack{N\to\infty\\ M/N\to\alpha}} D_{\mathsf{TV}}(\mathbb{P}_\beta, \mathbb{P}_0) = \mathsf{erfc}\left(\frac{1}{4}\sqrt{-\log\left(1 - \alpha\beta^2\right)}\right),$$

where $\mathsf{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-t^2}\mathrm{d}t$ is the complementary error function.

Furthermore, our proof of Theorem 3.1 allows us obtain the convergence of the mean (actually, all moments of $\log L$) under $\mathbb{P}_\beta$, which corresponds to the Kullback-Liebler divergence of $\mathbb{P}_\beta$ to $\mathbb{P}_0$:

Proposition 3.4. *For all $\alpha, \beta \geq 0$ such that $K_{\mathtt{u}}^4 K_{\mathtt{v}}^4 \alpha\beta^2 < 1$,*

$$\lim_{\substack{N\to\infty \\ M/N\to\alpha}} D_{\mathsf{KL}}(\mathbb{P}_\beta, \mathbb{P}_0) = -\frac{1}{4}\log\left(1-\alpha\beta^2\right).$$

3.3 Replicas, overlaps, Gibbs measures and Nishimori

To embark on the argument, we introduce similar notation and terminology as in Chapter 2. Let $H : \mathbb{R}^{N+M} \to \mathbb{R}$ be the (random) function, which we refer to as a *Hamiltonian*, defined as

$$-H(\boldsymbol{u},\boldsymbol{v}) = \sum_{i,j} \sqrt{\frac{\beta}{N}} Y_{ij} u_i v_j - \frac{\beta}{2N} u_i^2 v_j^2, \tag{3.2}$$

where $\boldsymbol{Y} = (Y_{ij})$ comes from $\mathbb{P}_\beta$ or $\mathbb{P}_0$. Letting ρ denote the product measure $P_{\mathtt{u}}^{\otimes N} \otimes P_{\mathtt{v}}^{\otimes M}$, we have

$$L(\boldsymbol{Y};\beta) = \int e^{-H(\boldsymbol{u},\boldsymbol{v})} \mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v}).$$

Let us define the Gibbs average of a function $f : (\mathbb{R}^{N+M})^n \mapsto \mathbb{R}$ of n replica pairs $(\boldsymbol{u}^{(l)}, \boldsymbol{v}^{(l)})_{l=1}^n$ with respect to the Hamiltonian H as

$$\langle f \rangle = \frac{\int f \prod_{l=1}^n e^{-H(\boldsymbol{u}^{(l)},\boldsymbol{v}^{(l)})} \mathrm{d}\rho(\boldsymbol{u}^{(l)},\boldsymbol{v}^{(l)})}{\left(\int e^{-H(\boldsymbol{u},\boldsymbol{v})} \mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v})\right)^n}. \tag{3.3}$$

This is the mean of f with respect to the posterior distribution of $(\boldsymbol{u},\boldsymbol{v})$ given $\boldsymbol{Y}$: $\mathbb{P}_\beta(\cdot|\boldsymbol{Y})^{\otimes n}$. We interpret the replicas as random and independent draws from this posterior. When $\boldsymbol{Y} \sim \mathbb{P}_\beta$ we also allow f to depend on the spike pair $(\boldsymbol{u}^*, \boldsymbol{v}^*)$. For two different replicas $(\boldsymbol{u}^{(l)}, \boldsymbol{v}^{(l)})$ and $(\boldsymbol{u}^{(l')}, \boldsymbol{v}^{(l')})$ (l' is allowed to take the value $*$) we denote the overlaps of the $\mathtt{u}$ and $\mathtt{v}$ parts, both normalized by N, as

$$R_{l,l'}^{\mathtt{u}} = \frac{1}{N}\sum_{i=1}^N u_i^{(l)} u_i^{(l')} \quad \text{and} \quad R_{l,l'}^{\mathtt{v}} = \frac{1}{N}\sum_{j=1}^M v_j^{(l)} v_j^{(l')}.$$

The Nishimori property under $\mathbb{P}_\beta$

The Nishimori property is deduced in the same way as in the symmetric case, this time we have two latent factors to consider:

1. Construct $\boldsymbol{u}^* \in \mathbb{R}^N$ and $\boldsymbol{v}^* \in \mathbb{R}^M$ by independently drawing their coordinates from $P_{\mathtt{u}}$ and $P_{\mathtt{v}}$ respectively.

2. Construct $\boldsymbol{Y} = \sqrt{\frac{\beta}{N}}\boldsymbol{u}^*\boldsymbol{v}^{*\top} + \boldsymbol{W}$, where $W_{ij} \sim \mathcal{N}(0,1)$ are all independent. ($\boldsymbol{Y}$ is distributed according to $\mathbb{P}_\beta$.)

3. Draw $n+1$ independent random vector pairs, $(\boldsymbol{u}^{(l)}, \boldsymbol{v}^{(l)})_{l=1}^{n+1}$, from $\mathbb{P}_\beta((\boldsymbol{u},\boldsymbol{v}) \in \cdot|\boldsymbol{Y})$.

By the tower property of expectations, the following equality of joint laws holds

$$\big(\boldsymbol{Y}, (\boldsymbol{u}^{(1)}, \boldsymbol{v}^{(1)}), \cdots, (\boldsymbol{u}^{(n+1)}, \boldsymbol{v}^{(n+1)})\big) \overset{\mathrm{d}}{=} \big(\boldsymbol{Y}, (\boldsymbol{u}^{(1)}, \boldsymbol{v}^{(1)}), \cdots, (\boldsymbol{u}^{(n)}, \boldsymbol{v}^{(n)}), (\boldsymbol{u}^{*}, \boldsymbol{v}^{*})\big)\,. \tag{3.4}$$

This in particular implies that under $\mathbb{P}_\beta$, the overlaps $(R^{\mathtt{u}}_{1,*}, R^{\mathtt{v}}_{1,*})$ between replica and spike pairs have the same distribution as the overlaps $(R^{\mathtt{u}}_{1,2}, R^{\mathtt{v}}_{1,2})$ between two replica pairs.

For the same reasons as in the symmetric case, overlap decay implies super-concentration. Indeed, we see that for $\boldsymbol{Y} \sim \mathbb{P}_0$, the Gaussian Poincaré inequality similarly implies

$$\mathbb{E}\left[(\log L - \mathbb{E}\log L)^2\right] \le \mathbb{E}\left[\|\nabla \log L\|_{\ell_2}^2\right] = \beta N\, \mathbb{E}\langle R^{\mathtt{u}}_{1,2} R^{\mathtt{v}}_{1,2}\rangle.$$

We deduce that the variance is $\mathcal{O}(N)$ since our priors have bounded support and $M/N = \mathcal{O}(1)$.

3.4 Proof of the main result

In this section we prove Theorem 3.1. As we've seen in Chapter 2, it suffices to treat the planted case $\boldsymbol{Y} \sim \mathbb{P}_\beta$. In this model, we are able to achieve control on the overlaps and show their concentration under the alternative in a wider region of parameters (α, β) than under the null.

This time, instead of showing the convergence of the moments (which could be done in pretty much the same way), we will show the convergence of the characteristic function of $\log L$ to that of a Gaussian. Let $\mu = -\frac{1}{4}\log(1-\alpha\beta^2)$, $\sigma^2 = -\frac{1}{2}\log(1-\alpha\beta^2)$, and let ϕ be the characteristic function of the Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$: for $s \in \mathbb{R}$ and $\mathbf{i}^2 = -1$, let $\phi(s) = \exp\{\mathbf{i}s\mu - \frac{\sigma^2}{2}s^2\}$. The following is a more quantitative convergence result that implies Theroem 3.1.

Theorem 3.5. *Let $s \in \mathbb{R}$ and $\alpha, \beta \ge 0$. There exists $K = K(s, \alpha, \beta, K_{\mathtt{u}}, K_{\mathtt{v}}) < \infty$ such that for M, N sufficiently large and $M = \alpha N + \mathcal{O}(\sqrt{N})$, the following holds. If $\alpha\beta^2 K_{\mathtt{u}}^4 K_{\mathtt{v}}^4 < 1$, then*

$$\left|\mathbb{E}_{\mathbb{P}_\beta}\left[e^{\mathbf{i}s\log L(\boldsymbol{Y};\beta)}\right] - \phi(s)\right| \le \frac{K}{\sqrt{N}}.$$

Our approach is to show that the function

$$\phi_N(\beta) = \mathbb{E}_{\mathbb{P}_\beta}\left[e^{\mathbf{i}s\log L(\boldsymbol{Y};\beta)}\right]$$

(for $s \in \mathbb{R}$ fixed) is an approximate solution to a differential equation whose solution is the characteristic function of the Gaussian.

Lemma 3.6. *For all $\beta \ge 0$, it holds that*

$$\frac{\mathrm{d}}{\mathrm{d}\beta}\phi_N(\beta) = \frac{\mathbf{i}s - s^2}{2} N\, \mathbb{E}\left[\langle R^{\mathtt{u}}_{1,2} R^{\mathtt{v}}_{1,2}\rangle\, e^{\mathbf{i}s\log L}\right]. \tag{3.5}$$

Proof. Since $\boldsymbol{Y} \sim \mathbb{P}_\beta$, we can rewrite the Hamiltonian (3.2) as

$$\begin{aligned} -H(\boldsymbol{u}, \boldsymbol{v}) &= \sum_{i,j} \sqrt{\frac{\beta}{N}} Y_{ij} u_i v_j - \frac{\beta}{2N} u_i^2 v_j^2, \\ &= \sum_{i,j} \sqrt{\frac{\beta}{N}} W_{ij} u_i v_j + \frac{\beta}{N} u_i v_j u_i^* v_j^* - \frac{\beta}{2N} u_i^2 v_j^2. \end{aligned}$$

We take a derivative with respect to β:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\beta}\phi_N(\beta) &= \mathrm{i}s\,\mathbb{E}\left[\left\langle -\frac{\mathrm{d}H}{\mathrm{d}\beta}\right\rangle e^{\mathrm{i}s\log L}\right] \\ &= \mathrm{i}s \sum_{i,j}\left(\frac{1}{2\sqrt{\beta N}}\,\mathbb{E}\left[W_{ij}\left\langle u_i v_j\right\rangle e^{\mathrm{i}s\log L}\right] - \frac{1}{2N}\,\mathbb{E}\left[\left\langle u_i^2 v_j^2\right\rangle e^{\mathrm{i}s\log L}\right]\right) \\ &\quad + \mathrm{i}s\frac{1}{N}\sum_{i,j}\mathbb{E}\left[\left\langle u_i v_j u_i^* v_j^*\right\rangle e^{\mathrm{i}s\log L}\right]. \end{aligned}$$

The last term is equal to $\mathrm{i}sN\,\mathbb{E}[\langle R^{\mathtt{u}}_{1,*}R^{\mathtt{v}}_{1,*}\rangle e^{\mathrm{i}s\log L}]$. As for the first term, since $W_{ij} \overset{\text{ind.}}{\sim} \mathcal{N}(0,1)$, we use Gaussian integration by parts to obtain

$$\begin{aligned} \mathbb{E}\left[W_{ij}\left\langle u_i v_j\right\rangle e^{\mathrm{i}s\log L}\right] &= \mathbb{E}\left[\frac{\mathrm{d}}{\mathrm{d}W_{ij}}\Big(\left\langle u_i v_j\right\rangle e^{\mathrm{i}s\log L}\Big)\right] \\ &= \sqrt{\frac{\beta}{N}}\left(\mathbb{E}\left[\left\langle u_i^2 v_j^2\right\rangle e^{\mathrm{i}s\log L}\right] - \mathbb{E}\left[\left\langle u_i v_j\right\rangle^2 e^{\mathrm{i}s\log L}\right] + \mathrm{i}s\,\mathbb{E}\left[\left\langle u_i v_j\right\rangle^2 e^{\mathrm{i}s\log L}\right]\right). \end{aligned}$$

Regrouping terms, we get

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\beta}\phi_N(\beta) &= -\mathrm{i}s\frac{N}{2}\,\mathbb{E}\left[\left\langle R^{\mathtt{u}}_{1,2}R^{\mathtt{v}}_{1,2}\right\rangle e^{\mathrm{i}s\log L}\right] + \mathrm{i}sN\,\mathbb{E}\left[\left\langle R^{\mathtt{u}}_{1,*}R^{\mathtt{v}}_{1,*}\right\rangle e^{\mathrm{i}s\log L}\right] \\ &\quad + (\mathrm{i}s)^2\frac{N}{2}\,\mathbb{E}\left[\left\langle R^{\mathtt{u}}_{1,2}R^{\mathtt{v}}_{1,2}\right\rangle e^{\mathrm{i}s\log L}\right]. \end{aligned} \tag{3.6}$$

The first and third terms in (3.6) contain overlaps between two replicas while the middle term contains an overlap between one replica and the spike vectors. By the Nishimori property (3.4), we can replace the spike by a second replica in the overlaps appearing in the middle term, and this finishes the proof. ■

A heuristic argument Let us now heuristically consider what should happen. A rigorous argument will be presented shortly. If the quantity $N\langle R^{\mathtt{u}}_{1,2}R^{\mathtt{v}}_{1,2}\rangle$ concentrates very strongly about some deterministic value $\theta = \theta(\alpha, \beta)$, we would expect that the Gibbs averages in (3.5) would behave approximately independently from $\log L$, and we would obtain the following differential equation

$$\frac{\mathrm{d}}{\mathrm{d}\beta}\phi_N(\beta) \simeq \frac{1}{2}\left(\mathrm{i}s - s^2\right)\theta\phi_N(\beta).$$

Since $\phi_N(0) = 1$, one obtains $\phi_N(\beta) \simeq \exp\{\frac{1}{2}(\mathrm{i}s - s^2)\int_0^\beta \theta \mathrm{d}\beta'\}$ by integrating over β, and the result would follow. The concentration assumption we used is commonly referred to as *replica-symmetry* or *the replica-symmetric ansatz* in the statistical physics literature. Most of the difficulty of the proof lies in showing rigorously that replica symmetry indeed holds.

Sign symmetry between $\mathbb{P}_\beta$ and $\mathbb{P}_0$ One can execute the same argument under the null model. Since there is no planted term in the Hamiltonian, the analogue of (3.6) one obtains does not contain the middle term. Hence the differential equation one obtains is

$$\frac{\mathrm{d}}{\mathrm{d}\beta}\phi_N(\beta) \simeq \frac{1}{2}\left(-\mathrm{i}s - s^2\right)\theta\phi_N(\beta).$$

This is one way to interpret the sign symmetry of the means of the limiting Gaussians under the null and the alternative: the interaction of one replica with the planted spike under the planted model accounts for twice the contribution of the interaction between two independent replicas, and this flips the sign of the mean.

We now replace the above heuristic with a rigorous statement. Recall that $\boldsymbol{Y} \sim \mathbb{P}_\beta$.

Proposition 3.7. *For $s \in \mathbb{R}$ and $\alpha, \beta \geq 0$ such that $\alpha\beta^2 K_{\mathtt{u}}^4 K_{\mathtt{v}}^4 < 1$, there exist a constant $K = K(s, \alpha, \beta, K_{\mathtt{u}}, K_{\mathtt{v}}) < \infty$ such that*

$$N\,\mathbb{E}\left[\left\langle R_{1,2}^{\mathtt{u}} R_{1,2}^{\mathtt{v}}\right\rangle e^{\mathrm{i}s\log L}\right] = \frac{\alpha\beta}{1-\alpha\beta^2}\,\mathbb{E}\left[e^{\mathrm{i}s\log L}\right] + \delta,$$

where $|\delta| \leq K/\sqrt{N}$. Moreover, K, seen as a function of β, is bounded on any interval $[0, \beta']$ when $\alpha\beta'^2 K_{\mathtt{u}}^4 K_{\mathtt{v}}^4 < 1$.

Taking $s = 0$, we see that $\theta = \frac{\alpha\beta}{1-\alpha\beta^2}$. Proposition 3.7 vindicates replica symmetry, and its proof occupies the majority of the rest of the manuscript.

Proof of Theorem 3.5. Plugging the results of Proposition 3.7 in the derivative computed in Lemma 3.6, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\beta}\phi_N(\beta) = \left(\frac{\mathrm{i}s - s^2}{2}\frac{\alpha\beta}{1-\alpha\beta^2}\right)\phi_N(\beta) + \delta,$$

where $|\delta| \leq \frac{K}{\sqrt{N}}\max\{|s|, s^2\}$, and K is the constant from Proposition 3.7. Integrating w.r.t. β we obtain

$$|\phi_N(\beta) - \phi(s)| \leq \frac{K'}{\sqrt{N}},$$

where K' depends on α, β, s and $K_{\mathtt{u}}, K_{\mathtt{v}}$, and $K' < \infty$ as long as $\alpha\beta^2 K_{\mathtt{u}}^4 K_{\mathtt{v}}^4 < 1$. ■

Let us prove in passing the convergence of the KL divergence between the null and alternative.

Proof of Proposition 3.4. Similarly to the computation of the derivative of ϕ_N, we can obtain

$$\frac{\mathrm{d}}{\mathrm{d}\beta}\,\mathbb{E}_{\mathbb{P}_\beta}\log L(\boldsymbol{Y};\beta) = -\frac{N}{2}\,\mathbb{E}\left\langle R^{\mathrm{u}}_{1,2}R^{\mathrm{v}}_{1,2}\right\rangle + N\,\mathbb{E}\left\langle R^{\mathrm{u}}_{1,*}R^{\mathrm{v}}_{1,*}\right\rangle = \frac{N}{2}\,\mathbb{E}\left\langle R^{\mathrm{u}}_{1,2}R^{\mathrm{v}}_{1,2}\right\rangle,$$

where we used the Nishimori property. By Proposition 3.7 with $s=0$, this derivative is $K/\sqrt{N}$ away from $\frac{1}{2}\frac{\alpha\beta}{1-\alpha\beta^2}$. Integration and boundedness of K finishes the proof. ■

3.5 Overlap convergence

The question of overlap convergence is purely a spin glass problem. We will use the machinery developed by Talagrand to solve it. In particular, a crucial use is made of the cavity method and Guerra's interpolation scheme. In this section, we present the main underlying ideas. The arguments are conceptually the same as the ones used for the symmetric case, with slight further technical complications. We delay their full execution to a later section. We refer to Talagrand (2007) for a leisurely high-level introduction to these ideas.

Sketch of proof of Proposition 3.7

The basic idea is to show that the quantities of interest approximately obey a self-consistent (or self-bounding) property, the error terms of which can be controlled. This approach will be used at different stages of the proof. We will show that

$$N\,\mathbb{E}\left[\left\langle R^{\mathrm{u}}_{1,2}R^{\mathrm{v}}_{1,2}\right\rangle e^{\mathrm{i}s\log L}\right] = \alpha\beta\,\mathbb{E}\left[e^{\mathrm{i}s\log L}\right] + \alpha\beta^2 N\,\mathbb{E}\left[\left\langle R^{\mathrm{u}}_{1,2}R^{\mathrm{v}}_{1,2}\right\rangle e^{\mathrm{i}s\log L}\right] + \delta,$$

where δ is the error term. This will be achieved in two steps. We first prove

$$N\,\mathbb{E}\left[\left\langle R^{\mathrm{u}}_{1,2}R^{\mathrm{v}}_{1,2}\right\rangle e^{\mathrm{i}s\log L}\right] = N\beta\,\mathbb{E}\left[\left\langle (R^{\mathrm{v}}_{1,2})^2\right\rangle e^{\mathrm{i}s\log L}\right] + \delta, \tag{3.7}$$

via a cavity on N, i.e., by isolating the effect of the last variable u_N on the rest of the variables. We then show

$$N\,\mathbb{E}\left[\left\langle (R^{\mathrm{v}}_{1,2})^2\right\rangle e^{\mathrm{i}s\log L}\right] = \frac{M}{N}\,\mathbb{E}\left[e^{\mathrm{i}s\log L}\right] + M\beta\,\mathbb{E}\left[\left\langle R^{\mathrm{u}}_{1,2}R^{\mathrm{v}}_{1,2}\right\rangle e^{\mathrm{i}s\log L}\right] + \delta, \tag{3.8}$$

via a cavity on M, i.e., isolating the effect of v_M. In the arguments leading to (3.7) and (3.8), we accumulate error terms that are proportional to the third moments of the overlaps:

$$\delta \lesssim N\,\mathbb{E}\left\langle |R^{\mathrm{u}}_{1,2}|^3\right\rangle + N\,\mathbb{E}\left\langle |R^{\mathrm{v}}_{1,2}|^3\right\rangle, \tag{3.9}$$

where we hide constants depending on α and β. These cavity equations impose only a mild restriction on the parameters so that our bounds go in the right direction, namely that $\alpha\beta^2 < 1$. This is about to change. We prove that $\delta = \mathcal{O}(1/\sqrt{N})$ with methods that impose the stronger restrictions on (α,β) that ultimately appear in the final result.

Convergence in the planted model: from crude estimates to optimal rates

We prove overlap convergence under the alternative. Let $\boldsymbol{Y} \sim \mathbb{P}_\beta$.

Proposition 3.8. *For all $\alpha, \beta \geq 0$ such that $K_{\mathtt{u}}^4 K_{\mathtt{v}}^4 \alpha\beta^2 < 1$, there exists $K = K(\alpha,\beta) < \infty$ such that*

$$\mathbb{E}\left\langle (R_{1,2}^{\mathtt{u}})^4 \right\rangle \vee \mathbb{E}\left\langle (R_{1,2}^{\mathtt{v}})^4 \right\rangle \leq \frac{K}{N^2}.$$

The proof proceeds as follows. We use the cavity method to show the following self-consistency equations:

$$\mathbb{E}\left\langle (R_{1,2}^{\mathtt{u}})^4 \right\rangle = \alpha\beta^2\, \mathbb{E}\left\langle (R_{1,2}^{\mathtt{u}})^4 \right\rangle + \overline{M}_{\mathtt{u}} + \delta_{\mathtt{u}}, \tag{3.10}$$

$$\mathbb{E}\left\langle (R_{1,2}^{\mathtt{v}})^4 \right\rangle = \alpha\beta^2\, \mathbb{E}\left\langle (R_{1,2}^{\mathtt{v}})^4 \right\rangle + \overline{M}_{\mathtt{v}} + \delta_{\mathtt{v}}, \tag{3.11}$$

where $|\overline{M}_{\mathtt{u}}|, |\overline{M}_{\mathtt{v}}|$ are bounded by sums of expectations of monomials of degree five in the overlaps $R^{\mathtt{u}}$ and $R^{\mathtt{v}}$:

$$|\overline{M}_{\mathtt{u}}| \lesssim \sum_{a,b,c,d} \mathbb{E}\left\langle \left|(R_{1,2}^{\mathtt{u}})^3 R_{a,b}^{\mathtt{u}} R_{c,d}^{\mathtt{u}}\right| \right\rangle + \mathbb{E}\left\langle \left|(R_{1,2}^{\mathtt{u}})^3 R_{a,b}^{\mathtt{v}} R_{c,d}^{\mathtt{v}}\right| \right\rangle,$$

$$|\overline{M}_{\mathtt{v}}| \lesssim \sum_{a,b,c,d} \mathbb{E}\left\langle \left|(R_{1,2}^{\mathtt{v}})^3 R_{a,b}^{\mathtt{v}} R_{c,d}^{\mathtt{v}}\right| \right\rangle + \mathbb{E}\left\langle \left|(R_{1,2}^{\mathtt{v}})^3 R_{a,b}^{\mathtt{u}} R_{c,d}^{\mathtt{u}}\right| \right\rangle,$$

where the sum is over a finite number of combinations (a,b,c,d), and

$$\delta_{\mathtt{u}} \lesssim \frac{1}{N}\,\mathbb{E}\left\langle (R_{1,2}^{\mathtt{u}})^2 \right\rangle + \mathcal{O}\Big(\frac{1}{N^2}\Big), \qquad \delta_{\mathtt{v}} \lesssim \frac{1}{N}\,\mathbb{E}\left\langle (R_{1,2}^{\mathtt{v}})^2 \right\rangle + \mathcal{O}\Big(\frac{1}{N^2}\Big).$$

These results hold for *all* $\alpha, \beta \geq 0$. From here, further progress is unlikely unless one has *a priori* knowledge that the overlaps are unlikely to be large, so that the fifth-order terms do not overwhelm the main terms. More precisely, suppose that we are able to prove the following crude bound on the overlaps: for $\epsilon > 0$, there is $K = K(\epsilon, \alpha, \beta) > 0$ such that

$$\mathbb{E}\left\langle \mathbb{1}\left\{ \left|R_{1,2}^{\mathtt{u}}\right| \geq \epsilon \right\} \right\rangle \vee \mathbb{E}\left\langle \mathbb{1}\left\{ \left|R_{1,2}^{\mathtt{v}}\right| \geq \epsilon \right\} \right\rangle \leq K e^{-N/K}. \tag{3.12}$$

Then the fifth-order terms can be controlled by fourth-order terms as follows:

$$\begin{aligned} \mathbb{E}\left\langle \left|(R_{1,2}^{\mathtt{u}})^3 R_{a,b}^{\mathtt{v}} R_{c,d}^{\mathtt{v}}\right| \right\rangle &\leq \epsilon\, \mathbb{E}\left\langle \left|(R_{1,2}^{\mathtt{u}})^3 R_{a,b}^{\mathtt{v}}\right| \right\rangle + K_{\mathtt{u}}^6 K_{\mathtt{v}}^4 K e^{-N/K} \\ &\leq \epsilon M + K e^{-N/K}, \end{aligned}$$

where $M = \mathbb{E}\langle (R_{1,2}^{\mathtt{u}})^4 \rangle \vee \mathbb{E}\langle (R_{1,2}^{\mathtt{v}})^4 \rangle$, and the last step is by Hölder's inequality. This way, $\overline{M}_{\mathtt{u}}$ and $\overline{M}_{\mathtt{v}}$ are controlled. Now it remains to control $\delta_{\mathtt{u}}$ and $\delta_{\mathtt{v}}$. We could re-execute the cavity argument on the second moment instead of the fourth, and this would allow us to obtain $\mathbb{E}\langle (R_{1,2}^{\mathtt{u}})^2 \rangle \vee \mathbb{E}\langle (R_{1,2}^{\mathtt{v}})^2 \rangle \leq K/N$. We instead use a shorter argument based on an elegant *quadratic replica coupling* technique of Guerra and F. Toninelli (2002b) to prove this. This is presented in Section 3.9. Plugging these estimates into (3.10) and (3.11), we obtain

$$\mathbb{E}\left\langle (R^{\mathtt{u}}_{1,2})^4 \right\rangle \le \alpha\beta^2\, \mathbb{E}\left\langle (R^{\mathtt{u}}_{1,2})^4 \right\rangle + K\epsilon M + \delta',$$
$$\mathbb{E}\left\langle (R^{\mathtt{v}}_{1,2})^4 \right\rangle \le \alpha\beta^2\, \mathbb{E}\left\langle (R^{\mathtt{v}}_{1,2})^4 \right\rangle + K\epsilon M + \delta',$$

where $\delta' \le K/N^2 + Ke^{-N/K}$, and this implies the desired result for ϵ sufficiently small.

The a priori bound (3.12) is proved via an interpolation argument at fixed overlap, combined with concentration of measure, and is presented in Section 3.9. These arguments impose a restriction on the parameters (α, β) that shows up in the final result. Finally, Proposition 3.8 allow us to conclude (via Jensen's inequality) that the error term δ displayed in (3.9) is bounded by $K/\sqrt{N}$.

3.6 A conjecture

The limiting factor in our approach to prove LR fluctuations is the need for precise non-asymptotic control of moments of the overlaps $R^{\mathtt{u}}_{1,2}$ and $R^{\mathtt{v}}_{1,2}$ under the expected Gibbs measure $\mathbb{E}\langle\cdot\rangle$. We were able to reach this level of control only in a restricted regime. This is due to the failure of our approach to prove the crude estimate (3.12) in a larger region. In this section, we formulate a conjecture on the largest region where these fluctuations and overlap decay should occur. In one sentence, this should be the entire *annealed* or *paramagnetic* region of the model, as dictated by the vanishing of its replica-symmetric (RS) formula. We shall now be more precise.

Let $z \sim \mathcal{N}(0,1)$, $u^* \sim P_{\mathtt{u}}$ and $v^* \sim P_{\mathtt{v}}$ all independent. Define

$$\psi_{\mathtt{u}}(r) := \mathbb{E}_{u^*,z} \log \int \exp\left(\sqrt{r}zu + ruu^* - \frac{r}{2}u^2\right) \mathrm{d}P_{\mathtt{u}}(u),$$
$$\psi_{\mathtt{v}}(r) := \mathbb{E}_{v^*,z} \log \int \exp\left(\sqrt{r}zv + rvv^* - \frac{r}{2}v^2\right) \mathrm{d}P_{\mathtt{v}}(v).$$

Moreover, define the RS potential as

$$F(\alpha, \beta, q_{\mathtt{u}}, q_{\mathtt{v}}) := \psi_{\mathtt{u}}(\beta q_{\mathtt{v}}) + \alpha\psi_{\mathtt{v}}(\beta q_{\mathtt{u}}) - \frac{\beta q_{\mathtt{u}} q_{\mathtt{v}}}{2}.$$

and finally define the RS formula as

$$\phi_{\mathsf{RS}}(\alpha, \beta) := \sup_{q_{\mathtt{v}} \ge 0} \inf_{q_{\mathtt{u}} \ge 0} F(\alpha, \beta, q_{\mathtt{u}}, q_{\mathtt{v}}).$$

It was argued by Lesieur, Krzakala, and Zdeborová (2015a) based on the plausibility of the replica-symmetric ansatz, and then proved by Miolane (2017), that in the limit $N \to \infty, M/N \to \alpha$, $\frac{1}{N}\mathbb{E}_{\mathbb{P}_\beta} \log L(\boldsymbol{Y};\beta) \to \phi_{\mathsf{RS}}(\alpha,\beta)$ for all $\alpha, \beta \ge 0$. (See also Barbier, Krzakala, et al., 2017, for results in a more general setup.) Of course, by change of measure and Jensen's inequality,

$$\mathbb{E}_{\mathbb{P}_\beta} \log L(\boldsymbol{Y};\beta) = \mathbb{E}_{\mathbb{P}_0} L(\boldsymbol{Y};\beta) \log L(\boldsymbol{Y};\beta) \geq 0,$$

for all M, N; therefore ϕ_{RS} is always nonnegative. Let

$$\Gamma = \left\{(\alpha,\beta) \in \mathbb{R}_+ \ : \ \phi_{\mathsf{RS}}(\alpha,\beta) = 0\right\}.$$

It is not hard to prove the following lemma by analyzing the stability of $(0,0)$ as a stationary point of the RS potential:

Lemma 3.9. $\Gamma \subseteq \{(\alpha,\beta) \in \mathbb{R}_+ \ : \ \alpha\beta^2 \leq 1\}$.

This lemma tells us (unsurprisingly) that Γ is entirely below the BBP threshold. The inclusion may or may not be strict depending on the priors P_{u} and P_{v}. For instance, there is equality of the above sets if P_{u} and P_{v} are symmetric Rademacher and/or Gaussian respectively. One case of strict inclusion is when P_{v} is Gaussian $\mathcal{N}(0,1)$ and P_{u} is a sparse Rademacher prior, $\frac{\rho}{2}\delta_{1/\sqrt{\rho}} + (1-\rho)\delta_0 + \frac{\rho}{2}\delta_{-1/\sqrt{\rho}}$, for sufficiently small ρ (e.g., $\rho = .04$). This is a canonical model for sparse principal component analysis. In this case, there is a region of parameters below the BBP threshold where the posterior mean $\mathbb{E}[\boldsymbol{u}^*|\boldsymbol{Y}]$ ($= \langle \boldsymbol{u} \rangle$ in our notation) has a non-trivial overlap with the spike $\boldsymbol{u}^*$, while the top eigenvector of the empirical covariance matrix $\boldsymbol{Y}\boldsymbol{Y}^\top$ is orthogonal to it. Estimation becomes impossible only in the region Γ, so the following conjecture is highly plausible:

Conjecture 1. *Let Γ' be the interior of Γ. For all $(\alpha,\beta) \in \Gamma'$,*

$$\log L(\boldsymbol{Y},\beta) \rightsquigarrow \mathcal{N}\left(\pm\frac{1}{4}\log(1-\alpha\beta^2), -\frac{1}{2}\log(1-\alpha\beta^2)\right),$$

where the plus sign holds under the null $\mathbb{P}_0$ and the minus sign under the alternative $\mathbb{P}_\beta$.

Our conjecture is formulated only in the interior of Γ; this is not a superfluous condition since diverging behavior may appear at the boundary. Moreover, this conjecture is about the *maximal* region in which such fluctuations can take place. This is not difficult to show. By (sub-Gaussian) concentration of the normalized likelihood ratio, we have for $\epsilon > 0$

$$\mathbb{P}_\beta\left(\frac{1}{N}\log L(\boldsymbol{Y};\beta) - \phi_{\mathsf{RS}}(\alpha,\beta) \leq -\epsilon\right) \longrightarrow 0,$$

where $K = K(\alpha,\beta) < \infty$. This already shows that $\log L$ must grow with N under the alternative if $\phi_{\mathsf{RS}} > 0$. As for the behavior under the null, the same sub-Gaussian concentration holds, although the expectation is not known (see Question 1):

$$\mathbb{P}_0\left(\frac{1}{N}\log L(\boldsymbol{Y};\beta) - \frac{1}{N}\mathbb{E}_{\mathbb{P}_0}\log L(\boldsymbol{Y};\beta) \geq \epsilon\right) \longrightarrow 0.$$

We do know that the above expectation is non-positive, by Jensen's inequality. Therefore if (α, β) are such that $\phi_{\mathsf{RS}} > 0$, one can distinguish $\mathbb{P}_\beta$ from $\mathbb{P}_0$ with asymptotic certainty by testing whether $\frac{1}{N}\log L(\boldsymbol{Y};\beta)$ is above or below (say) $\frac{1}{2}\phi_{\mathsf{RS}}(\alpha,\beta)$. This implies that $\mathbb{P}_\beta$ and $\mathbb{P}_0$ are not contiguous outside Γ. This—short of proving that $\log L$ grows in the negative direction with N—shows that the fluctuations cannot be of the above form under the null, since this would contradict Le Cam's first lemma.

The difficulty we encountered in our attempts to prove the above conjecture is a loss of control over the overlaps $R^{\mathsf{u}}_{1,2}$ and $R^{\mathsf{v}}_{1,2}$ near the boundary of the set Γ. The interpolation bound at fixed overlap (between a replica and the spike) we used under the alternative $\mathbb{P}_\beta$ is vacuous beyond the region $\alpha\beta^2 < (K_{\mathsf{u}}K_{\mathsf{v}})^{-4}$. It is possible that the latter bound could be marginally improved by more careful analysis, but this is unlikely to yield the optimal result since no information about ϕ_{RS} is used in the proof. One can imagine refining this technique by constraining two replicas and using an interpolation with broken replica-symmetry, in the spirit of the "2D" Guerra-Talagrand bound (Guerra, 2003; Talagrand, 2011b). Although this strategy is successful in the symmetric model where $\boldsymbol{u} = \boldsymbol{v}$ it is not at all obvious why such an interpolation bound should be true in the bipartite case: in the analysis, certain terms that are hard to control have a sign in the symmetric case, hence they can be dropped to obtain a bound. This is no longer true (or at least not obviously so) in the bipartite case.

Another interesting question concerns the LR asymptotics under the null, outside Γ. While under the alternative $\mathbb{P}_\beta$, the normalized log-likelihood ratio converges to the RS formula ϕ_{RS} for all (α,β), no such simple formula is expected to hold under the null. Even the existence of a limit seems to be unknown.

Question 1. *Does $\frac{1}{N}\mathbb{E}_{\mathbb{P}_0}\log L(\boldsymbol{Y};\beta)$ have a limit for all (α,β)? If so, what is its value?*

We refer to Barra, Galluzzi, et al. (2014); Barra, Genovese, and Guerra (2011) and Auffinger and Chen (2014) for some progress on the replica-symmetric phase, and Panchenko (2015) for progress on the related problem of the "multispecies" SK model at all temperatures.

3.7 Notation and useful lemmas

We make repeated use of interpolation arguments in our proofs. In this section, we state a few elementary lemmas we subsequently invoke several times. We denote the overlaps between replicas when the last variables are deleted by a superscript " $-$ " :

$$R^{\mathsf{u}-}_{l,l'} = \frac{1}{N}\sum_{i=1}^{N-1} u_i^{(l)}u_i^{(l')} \quad \text{and} \quad R^{\mathsf{v}-}_{l,l'} = \frac{1}{N}\sum_{j=1}^{M-1} v_j^{(l)}v_j^{(l')}.$$

If $\{H_t : t \in [0,1]\}$ is a generic family of random Hamiltonians, we let $\langle\cdot\rangle_t$ be the corresponding Gibbs average, and $\nu_t(f) = \mathbb{E}\langle f\rangle_t$, where the expectation is over the randomness of H_t. We will often write ν for ν_1.

In our executions of the cavity method, we use interpolations that isolate one last variable (either u_N or v_M) from the rest of the system. Taking the first case an example, we consider

$$-H_t(\boldsymbol{u},\boldsymbol{v}) = \sum_{i=1}^{N-1}\sum_{j=1}^{M}\sqrt{\frac{\beta}{N}}W_{ij}u_iv_j + \frac{\beta}{N}u_iu_i^*v_jv_j^* - \frac{\beta}{2N}u_i^2v_j^2 + \sum_{j=1}^{M}\sqrt{\frac{\beta t}{N}}W_{Nj}u_Nv_j + \frac{\beta t}{N}u_Nu_N^*v_jv_j^* - \frac{\beta t}{2N}u_N^2v_j^2.$$

Lemma 3.10. *Let f be a function of n replicas $(\boldsymbol{u}^{(l)},\boldsymbol{v}^{(l)})_{1\le l\le n}$. Then*

$$\frac{\mathrm{d}}{\mathrm{d}t}\nu_t(f) = \frac{\beta}{2}\sum_{1\le l\neq l'\le n}\nu_t(R^{\mathrm{v}}_{l,l'}u^{(l)}u^{(l')}f) - \frac{\beta}{2}n\sum_{l=1}^{n}\nu_t(R^{\mathrm{v}}_{l,n+1}u^{(l)}u^{(n+1)}f) + \beta n\sum_{l=1}^{n}\nu_t(R^{\mathrm{v}}_{l,*}u^{(l)}u^*f) - \beta n\nu_t(R^{\mathrm{v}}_{n+1,*}u^{(n+1)}u^*f) + \beta\frac{n(n+1)}{2}\nu_t(R^{\mathrm{v}}_{n+1,n+2}u^{(n+1)}u^{(n+2)}f).$$

Proof. This is a simple computation based on Gaussian integration by parts, similarly to Lemma 3.5. ■

The next lemma allows us to control interpolated averages by averages at time 1.

Lemma 3.11. *Let f be a nonnegative function of n replicas $(\boldsymbol{u}^{(l)},\boldsymbol{v}^{(l)})_{1\le l\le n}$. Then for all $t\in[0,1]$*

$$\nu_t(f)\le K(n,\alpha,\beta)\nu(f).$$

Proof. This is a consequence of Lemma 3.10, boundedness of the variables u_i and v_j, and Grönwall's lemma. ■

It is clear that Lemma 3.11 also holds if we switch the roles of $\boldsymbol{u}$ and $\boldsymbol{v}$ and extract v_M instead (so that ν_t is defined accordingly).

3.8 Proof of Proposition 3.7

We make use of two interpolation arguments; the first one extracts the last variable u_N from the system, and the second one extracts v_M. This allows to establish the self-consistency equations (3.7) and (3.8). We will assume decay of the forth moments of the overlaps, i.e., we assume Proposition 3.8 (which we prove in Section 3.9), and this allows us the prove that the error terms emerging from the cavity method converge to zero. Recall that the Nishimori property implies

$$\mathbb{E}\left[\left\langle R^{\mathrm{u}}_{1,2}R^{\mathrm{v}}_{1,2}\right\rangle e^{\mathrm{i}s\log L}\right] = \mathbb{E}\left[\left\langle R^{\mathrm{u}}_{1,*}R^{\mathrm{v}}_{1,*}\right\rangle e^{\mathrm{i}s\log L}\right].$$

As it turns out, it is more convenient to work with the right-hand side.

Cavity on N

By symmetry of the $\mathtt{u}$ variables, we have

$$\mathbb{E}\left[\left\langle R^{\mathtt{u}}_{1,*}R^{\mathtt{v}}_{1,*}\right\rangle e^{\mathrm{i}s\log L}\right] = \mathbb{E}\left[\left\langle u^{(1)}_N u^*_N R^{\mathtt{v}}_{1,*}\right\rangle e^{\mathrm{i}s\log L}\right].$$

Now we consider the interpolating Hamiltonian

$$\begin{aligned}-H_t(\boldsymbol{u},\boldsymbol{v}) = &\sum_{i=1}^{N-1}\sum_{j=1}^{M}\sqrt{\frac{\beta}{N}}W_{ij}u_iv_j + \frac{\beta}{N}u_iu_i^*v_jv_j^* - \frac{\beta}{2N}u_i^2v_j^2\\ &+\sum_{j=1}^{M}\sqrt{\frac{\beta t}{N}}W_{Nj}u_Nv_j + \frac{\beta t}{N}u_Nu_N^*v_jv_j^* - \frac{\beta t}{2N}u_N^2v_j^2,\end{aligned}$$

and let $\langle\cdot\rangle_t$ be the associated Gibbs average. We let

$$X(t) = \exp\Big(\mathrm{i}s\log\int e^{-H_t(\boldsymbol{u},\boldsymbol{v})}\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v})\Big),$$

and

$$\varphi(t) = N\,\mathbb{E}\left[\left\langle u^{(1)}_N u^*_N R^{\mathtt{v}}_{1,*}\right\rangle_t X(t)\right].$$

Observe that $\varphi(1)$ is the quantity we seek to analyze. We will use the following error bound on Taylor's expansion:

$$|\varphi(1)-\varphi(0)-\varphi'(0)| \leq \sup_{0\leq t\leq 1}|\varphi''(t)|,$$

to approximate $\varphi(1)$ by $\varphi(0)+\varphi'(0)$. Since $P_{\mathtt{u}}$ is centered, we have $\varphi(0)=0$. With a computation similar to the one leading to Lemma 3.6, the time derivative $\varphi'(t)$ is a sum of terms of the form

$$N\beta\,\mathbb{E}\left[\left\langle u^{(1)}_N u^*_N u^{(a)}_N u^{(b)}_N R^{\mathtt{v}}_{1,*}R^{\mathtt{v}}_{a,b}\right\rangle_t X(t)\right],$$

for $(a,b)\in\{(1,*),(2,*),(1,2),(2,3)\}$. At $t=0$ all terms vanish expect when $(a,b)=(1,*)$ and we get

$$\varphi'(0) = N\beta\,\mathbb{E}\left[\left\langle (R^{\mathtt{v}}_{1,*})^2\right\rangle_0 X(0)\right].$$

Now we wish to replace the time index $t=0$ in the above quantity by the time index $t=1$. Similarly to φ, the derivative of the function $t\mapsto N\beta\,\mathbb{E}[\langle (R^{\mathtt{v}}_{1,*})^2\rangle_t X(t)]$, is a sum of terms of the form

$$N\beta^2\,\mathbb{E}\left[\left\langle u^{(a)}_N u^{(b)}_N (R^{\mathtt{v}}_{1,*})^2 R^{\mathtt{v}}_{a,b}\right\rangle_t X(t)\right].$$

By boundedness of the $\mathtt{u}$ variables and Hölder's inequality, this is bounded by

$$\begin{aligned}N\beta^2K_{\mathtt{u}}^4\,\mathbb{E}\left[\left\langle\left|(R^{\mathtt{v}}_{1,*})^2R^{\mathtt{v}}_{a,b}\right|\right\rangle_t\right] &\leq N\beta^2K_{\mathtt{u}}^4\,\mathbb{E}\left[\left\langle|R^{\mathtt{v}}_{1,*}|^3\right\rangle_t\right]\\ &\leq N\beta^2K_{\mathtt{u}}^4K\,\mathbb{E}\left[\left\langle|R^{\mathtt{v}}_{1,*}|^3\right\rangle\right]\end{aligned}$$

$$\leq \frac{K\beta^2}{\sqrt{N}},$$

where the second bound is by Lemma 3.11, and the last bound is a consequence of Proposition 3.8 (and Jensen's inequality). Therefore

$$\left|\varphi'(0) - N\beta\,\mathbb{E}\left[\left\langle (R^{\mathtt{v}}_{1,*})^2\right\rangle X(1)\right]\right| \leq \frac{K}{\sqrt{N}}.$$

Similarly, we control the second derivative φ''. This can be written as a finite sum of terms of the form

$$N\beta^2\,\mathbb{E}\left[\left\langle u_N^{(1)}u_N^* u_N^{(a)}u_N^{(b)}u_N^{(c)}u_N^{(d)}R^{\mathtt{v}}_{1,*}R^{\mathtt{v}}_{a,b}R^{\mathtt{v}}_{c,d}\right\rangle_t X(t)\right],$$

which are bounded in the same way by

$$N\beta^2 K_{\mathtt{u}}^6\,\mathbb{E}\left[\left\langle\left|R^{\mathtt{v}}_{1,*}R^{\mathtt{v}}_{a,b}R^{\mathtt{v}}_{c,d}\right|\right\rangle_t\right] \leq \beta^2 K_{\mathtt{u}}^6\frac{K}{\sqrt{N}}.$$

Therefore $|\varphi''| \leq K/\sqrt{N}$. We end up with

$$N\,\mathbb{E}\left[\left\langle R^{\mathtt{u}}_{1,*}R^{\mathtt{v}}_{1,*}\right\rangle e^{\mathrm{i}s\log L}\right] = N\beta\,\mathbb{E}\left[\left\langle (R^{\mathtt{v}}_{1,*})^2\right\rangle e^{\mathrm{i}s\log L}\right] + \delta, \tag{3.13}$$

where $|\delta| \leq K/\sqrt{N}$ whenever (α,β) satisfy the conditions of Proposition 3.8.

Cavity on M

By symmetry of the $\mathtt{v}$ variables,

$$\begin{aligned}N\,\mathbb{E}\left[\left\langle (R^{\mathtt{v}}_{1,*})^2\right\rangle e^{\mathrm{i}s\log L}\right] &= M\,\mathbb{E}\left[\left\langle v_M^{(1)}v_M^* R^{\mathtt{v}}_{1,*}\right\rangle e^{\mathrm{i}s\log L}\right]\\ &= \frac{M}{N}\,\mathbb{E}\left[\left\langle (v_M^{(1)}v_M^*)^2\right\rangle e^{\mathrm{i}s\log L}\right] + M\,\mathbb{E}\left[\left\langle v_M^{(1)}v_M^* R^{\mathtt{v}-}_{1,*}\right\rangle e^{\mathrm{i}s\log L}\right].\end{aligned}$$

Now we execute the same argument as above with the roles of $\mathtt{u}$ and $\mathtt{v}$ flipped to prove that

$$\mathbb{E}\left[\left\langle (v_M^{(1)}v_M^*)^2\right\rangle e^{\mathrm{i}s\log L}\right] = \mathbb{E}\left[e^{\mathrm{i}s\log L}\right] + \delta,$$

and

$$M\,\mathbb{E}\left[\left\langle v_M^{(1)}v_M^* R^{\mathtt{v}-}_{1,*}\right\rangle e^{\mathrm{i}s\log L}\right] = M\beta\,\mathbb{E}\left[\left\langle R^{\mathtt{u}}_{1,*}R^{\mathtt{v}}_{1,*}\right\rangle e^{\mathrm{i}s\log L}\right] + \delta,$$

where $|\delta| \leq K(M/N^{3/2} \vee 1/\sqrt{N})$. Here we use the interpolating Hamiltonian

$$-H_t(\boldsymbol{u},\boldsymbol{v}) = \sum_{j=1}^{M-1}\sum_{i=1}^{N}\sqrt{\frac{\beta}{N}}W_{ij}u_iv_j + \frac{\beta}{N}u_iu_i^*v_jv_j^* - \frac{\beta}{2N}u_i^2v_j^2$$

$$+\sum_{i=1}^{N}\sqrt{\frac{\beta t}{N}}W_{iM}u_iv_M+\frac{\beta t}{N}u_iu_i^*v_Mv_M^*-\frac{\beta t}{2N}u_i^2v_M^2,$$

and similarly define the random variable $X(t)=\exp\big(\mathrm{i}s\log\int e^{-H_t(\boldsymbol{u},\boldsymbol{v})}\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v})\big)$. After executing the argument, we obtain

$$N\,\mathbb{E}\left[\left\langle(R_{1,*}^{\mathtt{v}})^2\right\rangle e^{\mathrm{i}s\log L}\right]=\frac{M}{N}\,\mathbb{E}\left[e^{\mathrm{i}s\log L}\right]+M\beta\,\mathbb{E}\left[\left\langle R_{1,*}^{\mathtt{u}}R_{1,*}^{\mathtt{v}}\right\rangle e^{\mathrm{i}s\log L}\right]+\delta. \tag{3.14}$$

From (3.13) and (3.14), we obtain

$$N\,\mathbb{E}\left[\left\langle R_{1,*}^{\mathtt{u}}R_{1,*}^{\mathtt{v}}\right\rangle e^{\mathrm{i}s\log L}\right]=\frac{M}{N}\beta\,\mathbb{E}\left[e^{\mathrm{i}s\log L}\right]+M\beta^2\,\mathbb{E}\left[\left\langle R_{1,*}^{\mathtt{u}}R_{1,*}^{\mathtt{v}}\right\rangle e^{\mathrm{i}s\log L}\right]+\delta,$$

where $|\delta|\le K(M/N^{3/2}\vee 1/\sqrt{N})$. For $M=\alpha N+\mathcal{O}(\sqrt{N})$, we arrive at

$$N\,\mathbb{E}\left[\left\langle R_{1,*}^{\mathtt{u}}R_{1,*}^{\mathtt{v}}\right\rangle e^{\mathrm{i}s\log L}\right]=\frac{\alpha\beta}{1-\alpha\beta^2}\,\mathbb{E}\left[e^{\mathrm{i}s\log L}\right]+\delta,$$

with $|\delta|\le K/\sqrt{N}$, and this finishes the proof.

3.9 Proof of Proposition 3.8

This section is about overlap convergence in the planted model. As explained in the main text, the proof is in several steps. We first present a proof of convergence of the second moment of the overlaps that does not rely on the cavity method, but on a *quadratic replica coupling* scheme of Guerra and F. Toninelli (2002b). Then we present the interpolation argument as a fixed overlap that will allow us to prove the crude convergence bound (3.12). Finally we execute a round of the cavity method to prove convergence of the fourth moment.

Convergence of the second moment

Proposition 3.12. *For all α,β such that $K_{\mathtt{u}}^4K_{\mathtt{v}}^4\alpha\beta^2<1$, there exists $K=K(\alpha,\beta)<\infty$ such that*

$$\mathbb{E}\left\langle(R_{1,*}^{\mathtt{u}})^2\right\rangle\vee\mathbb{E}\left\langle(R_{1,*}^{\mathtt{v}})^2\right\rangle\le\frac{K}{N^2}.$$

Of course, by the Nishimori property, this is also a statement about the overlaps between two independent replicas.

Proof. Let $\sigma_{\mathtt{u}}$ and $\sigma_{\mathtt{v}}$ be the sub-Gaussian parameters of $P_{\mathtt{u}}$ and $P_{\mathtt{v}}$ respectively. We since $P_{\mathtt{u}}$ and $P_{\mathtt{v}}$ have unit variance, we have $1\le\sigma_{\mathtt{u}}^2\le K_{\mathtt{u}}^2$ and similarly for $P_{\mathtt{v}}$.

We start with the $\mathtt{u}$-overlap. Let us define the function

$$\Phi_{\mathtt{u}}(\lambda)=\frac{1}{N}\,\mathbb{E}\log\int\exp\left(-H(\boldsymbol{u},\boldsymbol{v})+\frac{\lambda}{2}N(R_{1,*}^{\mathtt{u}})^2\right)\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v}).$$

The outer expectation is on $\boldsymbol{Y} \sim \mathbb{P}_\beta$ (or equivalently on $\boldsymbol{u}^*$, $\boldsymbol{v}^*$ and $\boldsymbol{W}$ independently). A simple inspection shows that the above function is convex and increasing in λ, and

$$\Phi_{\mathtt{u}}'(0) = \frac{1}{2}\,\mathbb{E}\left\langle (R_{1,*}^{\mathtt{u}})^2 \right\rangle.$$

The convexity then implies for all $\lambda \geq 0$,

$$\frac{\lambda}{2}\,\mathbb{E}\left\langle (R_{1,*}^{\mathtt{u}})^2 \right\rangle \leq \Phi_{\mathtt{u}}(\lambda) - \Phi_{\mathtt{u}}(0).$$

Of course $\Phi_{\mathtt{u}}(0) = \frac{1}{N}\,\mathbb{E}_{\mathbb{P}_\beta} \log L(\boldsymbol{Y};\beta) \geq 0$ by Jensen's inequality, so it remains to upper bound $\Phi_{\mathtt{u}}(\lambda)$. To this end we consider the interpolation

$$\Phi_{\mathtt{u}}(\lambda, t) = \frac{1}{N}\,\mathbb{E}\log \int \exp\left(-H_t(\boldsymbol{u},\boldsymbol{v}) + \frac{\lambda}{2} N (R_{1,*}^{\mathtt{u}})^2\right) \mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v}),$$

where

$$-H_t(\boldsymbol{u},\boldsymbol{v}) = \sum_{i,j} \sqrt{\frac{\beta t}{N}} W_{ij} u_i v_j + \frac{\beta}{N} u_i u_i^* v_j v_j^* - \frac{\beta t}{2N} u_i^2 v_j^2.$$

Notice that the planted (middle) term in the Hamiltonian is left unaltered. The time derivative is

$$\partial_t \Phi_{\mathtt{u}}(\lambda, t) = -\frac{\beta}{2}\,\mathbb{E}\left\langle (R_{1,2}^{\mathtt{u}})^2 \right\rangle_{\lambda,t} \leq 0,$$

where $\langle\cdot\rangle_{\lambda,t}$ is the Gibbs average w.r.t $-H_t(\boldsymbol{u},\boldsymbol{v}) + \frac{\lambda}{2} N (R_{1,*}^{\mathtt{u}})^2$. Therefore

$$\begin{aligned}\Phi_{\mathtt{u}}(\lambda) \leq \Phi_{\mathtt{u}}(\lambda, 0) &= \frac{1}{N}\,\mathbb{E}\log \int \exp\left(\beta N R_{1,*}^{\mathtt{u}} R_{1,*}^{\mathtt{v}} + \frac{\lambda}{2} N (R_{1,*}^{\mathtt{u}})^2\right) \mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v}) \\ &\leq \frac{1}{N}\,\mathbb{E}\log \int \exp\left(\frac{\alpha\beta^2\sigma_{\mathtt{v}}^2 \widehat{v} + \lambda}{2} N (R_{1,*}^{\mathtt{u}})^2\right) \mathrm{d}P_{\mathtt{u}}^{\otimes N}(\boldsymbol{u}),\end{aligned}$$

where we have used the sub-Gaussianity of $P_{\mathtt{v}}$, and let $\widehat{v} = \frac{1}{M}\sum_{j=1}^{M} v_j^{*2}$. (Here, we have abused notation and let $\alpha = \frac{M}{N}$. This will not cause any problems.) Next we introduce an independent r.v. $g \sim \mathcal{N}(0,1)$, exchange integrals by Fubini's theorem, and continue:

$$\begin{aligned}&\frac{1}{N}\,\mathbb{E}\log \mathbb{E}_g \left[\int \exp\left(\sqrt{(\alpha\beta^2\sigma_{\mathtt{v}}^2\widehat{v} + \lambda)N} R_{1,*}^{\mathtt{u}} g\right) \mathrm{d}P_{\mathtt{u}}^{\otimes N}(\boldsymbol{u})\right] \\ &\leq \frac{1}{N}\,\mathbb{E}\log \mathbb{E}_g \left[\exp\left(\frac{\alpha\beta^2\sigma_{\mathtt{v}}^2\widehat{v} + \lambda}{2} \sigma_{\mathtt{u}}^2 \widehat{u} g^2\right)\right],\end{aligned}$$

where we use the sub-Gaussianity of $P_{\mathtt{u}}$, and let $\widehat{u} = \frac{1}{N}\sum_{i=1}^{N} u_i^{*2}$. We bound $\widehat{u}$ and $\widehat{v}$ by $K_{\mathtt{u}}^2$ and $K_{\mathtt{v}}^2$ respectively and integrate on g to obtain the upper bound

$$\Phi_{\mathtt{u}}(\lambda) \leq -\frac{1}{2N} \log\left(1 - (\alpha\beta^2\sigma_{\mathtt{v}}^2 K_{\mathtt{v}}^2 + \lambda)\sigma_{\mathtt{u}}^2 K_{\mathtt{u}}^2\right),$$

valid as long as $(\alpha\beta^2\sigma_{\mathtt{v}}^2K_{\mathtt{v}}^2+\lambda)\sigma_{\mathtt{u}}^2K_{\mathtt{u}}^2<1$. Letting $\lambda=(1-\alpha\beta^2\sigma_{\mathtt{v}}^2K_{\mathtt{v}}^2\sigma_{\mathtt{u}}^2K_{\mathtt{u}}^2)/(2\sigma_{\mathtt{u}}^2K_{\mathtt{u}}^2)>0$, we obtain

$$\mathbb{E}\left\langle(R_{1,*}^{\mathtt{u}})^2\right\rangle\leq\frac{K(\alpha,\beta)}{N},$$

with $K(\alpha,\beta)=\frac{2\sigma_{\mathtt{u}}^2K_{\mathtt{u}}^2\log((1-\alpha\beta^2\sigma_{\mathtt{v}}^2K_{\mathtt{v}}^2\sigma_{\mathtt{u}}^2K_{\mathtt{u}}^2)/2)}{(1-\alpha\beta^2\sigma_{\mathtt{v}}^2K_{\mathtt{v}}^2\sigma_{\mathtt{u}}^2K_{\mathtt{u}}^2)}$.

We use the exact same argument for the $\mathtt{v}$-overlaps. We define $\Phi_{\mathtt{v}}(\lambda)$ in the same way by replacing the quadratic term $\frac{\lambda}{2}N(R_{1,*}^{\mathtt{u}})^2$ by $\frac{\lambda}{2}N(R_{1,*}^{\mathtt{v}})^2$ and obtain

$$\Phi_{\mathtt{v}}(\lambda)\leq-\frac{1}{2N}\log\left(1-(\beta^2\sigma_{\mathtt{u}}^2K_{\mathtt{u}}^2+\lambda)\alpha\sigma_{\mathtt{v}}^2K_{\mathtt{v}}^2\right).$$

We choose $\lambda=(1-\alpha\beta^2\sigma_{\mathtt{v}}^2K_{\mathtt{v}}^2\sigma_{\mathtt{u}}^2K_{\mathtt{u}}^2)/(2\alpha\sigma_{\mathtt{v}}^2K_{\mathtt{v}}^2)$ and use the same convexity argument to obtain

$$\mathbb{E}\left\langle(R_{1,*}^{\mathtt{v}})^2\right\rangle\leq\frac{K'(\alpha,\beta)}{N},$$

with $K'(\alpha,\beta)=\frac{2\alpha\sigma_{\mathtt{v}}^2K_{\mathtt{v}}^2\log((1-\alpha\beta^2\sigma_{\mathtt{v}}^2K_{\mathtt{v}}^2\sigma_{\mathtt{u}}^2K_{\mathtt{u}}^2)/2)}{(1-\alpha\beta^2\sigma_{\mathtt{v}}^2K_{\mathtt{v}}^2\sigma_{\mathtt{u}}^2K_{\mathtt{u}}^2)}$. ■

Interpolation bound at fixed overlap

In this section we present and prove an interpolation bound on the free energy of a subpopulation of configurations having a fixed overlap with the planted spike $(\boldsymbol{u}^*,\boldsymbol{v}^*)$. This is a key step in proving the crude bound (3.12).

Proposition 3.13. *Fix $\boldsymbol{u}^*\in\mathbb{R}^N,\boldsymbol{v}^*\in\mathbb{R}^M$ with $\|\boldsymbol{u}^*\|_{\ell_2}^2/N\leq K_{\mathtt{u}}^2$ and $\|\boldsymbol{v}^*\|_{\ell_2}^2/M\leq K_{\mathtt{v}}^2$. Let $\alpha=\frac{M}{N}$ and $\Delta=\alpha\beta^2\sigma_{\mathtt{u}}^2\sigma_{\mathtt{v}}^2K_{\mathtt{u}}^2K_{\mathtt{v}}^2-1$. For $m\in\mathbb{R}\setminus\{0\}$, $\epsilon\geq0$, let $A_{\mathtt{u}}$ be the event*

$$A_{\mathtt{u}}=\begin{cases}R_{1,*}^{\mathtt{u}}\in[m,m+\epsilon) & \text{if } m>0,\\ R_{1,*}^{\mathtt{u}}\in(m-\epsilon,m] & \text{if } m<0.\end{cases}$$

Define $A_{\mathtt{v}}$ similarly. We have

$$\frac{1}{N}\mathbb{E}\log\int\mathbb{1}(A_{\mathtt{u}})e^{-H(\boldsymbol{u},\boldsymbol{v})}\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v})\leq\frac{\Delta}{2\sigma_{\mathtt{u}}^2K_{\mathtt{u}}^2}m^2+\alpha\beta K_{\mathtt{v}}^2\epsilon,\tag{3.15}$$

and

$$\frac{1}{N}\mathbb{E}\log\int\mathbb{1}(A_{\mathtt{v}})e^{-H(\boldsymbol{u},\boldsymbol{v})}\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v})\leq\frac{\Delta}{2\alpha\sigma_{\mathtt{v}}^2K_{\mathtt{v}}^2}m^2+\beta K_{\mathtt{u}}^2\epsilon.\tag{3.16}$$

The expectation $\mathbb{E}$ is over the Gaussian disorder $\boldsymbol{W}$.

These are version of the Franz-Parisi potential discussed in Chapter 2.

Proof. We only prove (3.15). The bound (3.16) follows by flipping the roles of $\boldsymbol{u}$ and $\boldsymbol{v}$. We consider the interpolating Hamiltonian

$$-H_t(\boldsymbol{u},\boldsymbol{v}) = \sum_{i,j} \sqrt{\frac{\beta t}{N}} W_{ij} u_i v_j + \frac{\beta t}{N} u_i u_i^* v_j v_j^* - \frac{\beta t}{2N} u_i^2 v_j^2 + \sum_{j=1}^{M} (1-t)\beta m v_j v_j^*,$$

and let

$$\varphi(t) = \frac{1}{N}\mathbb{E}\log\int \mathbb{1}\{R^{\mathsf{u}}_{1,*} \in [m, m+\epsilon)\} e^{-H_t(\boldsymbol{u},\boldsymbol{v})} \mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v}).$$

We have

$$\varphi'(t) = -\frac{\beta}{2}\mathbb{E}\left\langle R^{\mathsf{u}}_{1,2} R^{\mathsf{v}}_{1,2}\right\rangle_t + \beta\,\mathbb{E}\left\langle R^{\mathsf{u}}_{1,*} R^{\mathsf{v}}_{1,*}\right\rangle_t - \beta m\,\mathbb{E}\left\langle R^{\mathsf{v}}_{1,*}\right\rangle_t.$$

The first term in the above expression is ≤ 0, and since the overlap $R^{\mathsf{u}}_{1,*}$ is constrained to be close to m we have $\left|\mathbb{E}\left\langle (R^{\mathsf{u}}_{1,*} - m) R^{\mathsf{v}}_{1,*}\right\rangle_t\right| \leq \alpha K_{\mathsf{v}}^2 \epsilon$. So $\varphi'(t) \leq \alpha K_{\mathsf{v}}^2\epsilon$. Moreover, the variables $\boldsymbol{u}$ and $\boldsymbol{v}$ decouple at $t=0$ and one can write

$$\varphi(1) \leq \frac{1}{N}\log\Pr\left(A_{\mathsf{u}}\right) + \frac{1}{N}\sum_{j=1}^{M}\log\mathbb{E}_v\left[e^{\beta m v v_j^*}\right] + K_{\mathsf{v}}^2\epsilon.$$

By sub-Gaussianity of the prior P_{v} we have $\mathbb{E}_v\left[e^{\beta m v v_j^*}\right] \leq e^{\beta^2\sigma_{\mathsf{v}}^2 m^2 v_j^{*2}/2}$. On the other hand, for a fixed parameter γ of the same sign as m, we have

$$\frac{1}{N}\log\Pr\left(A_{\mathsf{u}}\right) \leq -\gamma m + \frac{1}{N}\sum_{i=1}^{N}\log\mathbb{E}_u[e^{\gamma u u_i^*}] \leq -\gamma m + \frac{1}{2N}\sum_{i=1}^{N} u_i^{*2}\sigma_{\mathsf{u}}^2\gamma^2.$$

The last inequality uses sub-Gaussianity of P_{u}. We minimize this quadratic w.r.t γ and obtain

$$\varphi(1) \leq -\frac{m^2}{2\sigma_{\mathsf{u}}^2\widehat{u}} + \frac{M}{2N}\beta^2\sigma_{\mathsf{v}}^2\widehat{v} m^2 + \alpha K_{\mathsf{v}}^2\epsilon,$$

where $\widehat{u} = \frac{1}{N}\sum_{i=1}^{N} u_i^{*2}$ and $\widehat{v} = \frac{1}{M}\sum_{j=1}^{M} v_j^{*2}$. We upper bound the latter two numbers by K_{u}^2 and K_{v}^2 respectively. ■

Overlap concentration (proof of (3.12))

Here we prove convergence of the overlaps to zero in probability. We first state a useful and standard result of concentration of measure.

Lemma 3.14. *Let $\boldsymbol{Y} = \sqrt{\frac{\beta}{N}}\boldsymbol{u}^*\boldsymbol{v}^{*\top} + \boldsymbol{W}$, where the planted vectors $\boldsymbol{u}^*$ and $\boldsymbol{v}^*$ are fixed, and $W_{ij} \sim \mathcal{N}(0,1)$. For a Borel set $A \subset \mathbb{R}^{M+N}$, let*

$$Z = \int_A e^{-H(\boldsymbol{u},\boldsymbol{v})}\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v}).$$

We have for every $t \geq 0$,

$$\Pr\left(|\log Z - \mathbb{E}\log Z| \geq Nt\right) \leq 2e^{-\frac{Nt^2}{2\beta K_{\mathtt{u}}^2 K_{\mathtt{v}}^2}}.$$

(Here $\Pr$ *and* $\mathbb{E}$ *are conditional on* $\boldsymbol{u}^*$ *and* $\boldsymbol{v}^*$*.)*

Proof. We simply observe that the function $\boldsymbol{W} \mapsto \log Z$ is Lipschitz with constant $\sqrt{N\beta\alpha K_{\mathtt{u}}^2 K_{\mathtt{v}}^2}$. The result follows from concentration of Lipschitz functions of Gaussian r.v.'s (this is the Borell-Tsirelson-Ibragimov-Sudakov inequality; see Boucheron, Lugosi, and Massart, 2013, Theorem 5.6). ■

Proposition 3.15. *Let* α, β *such that* $\alpha\beta^2\sigma_{\mathtt{u}}^2\sigma_{\mathtt{v}}^2 K_{\mathtt{u}}^2 K_{\mathtt{v}}^2 < 1$*, and* $\epsilon > 0$*. There exist constants* $c = c(\epsilon, \alpha, \beta, K_{\mathtt{u}}, K_{\mathtt{v}}) > 0$ *and* $K = K(K_{\mathtt{u}}, K_{\mathtt{v}}) > 0$ *such that*

$$\mathbb{E}\left\langle \mathbb{1}\{|R^{\mathtt{u}}_{1,*}| \geq \epsilon\}\right\rangle \vee \mathbb{E}\left\langle \mathbb{1}\{|R^{\mathtt{v}}_{1,*}| \geq \epsilon\}\right\rangle \leq \frac{K}{\epsilon^2}e^{-cN}.$$

Proof. We only prove the assertion for the $\mathtt{u}$-overlap since the argument is strictly the same for the $\mathtt{v}$-overlap.

For $\epsilon, \epsilon' > 0$, we can write the decomposition

$$\begin{aligned}\mathbb{E}\left\langle \mathbb{1}\{\left|R^{\mathtt{u}}_{1,*}\right| \geq \epsilon'\}\right\rangle = &\sum_{l\geq 0}\mathbb{E}\left\langle \mathbb{1}\{R^{\mathtt{u}}_{1,*} - \epsilon' \in [l\epsilon, (l+1)\epsilon)\}\right\rangle \\ &+ \sum_{l\geq 0}\mathbb{E}\left\langle \mathbb{1}\{-R^{\mathtt{u}}_{1,*} + \epsilon' \in [l\epsilon, (l+1)\epsilon)\}\right\rangle,\end{aligned}$$

where the integer index l ranges over a finite set of size $\leq K/\epsilon$. We only treat the generic term in the first sum; the second sum can be handled similarly. Fix $m > 0, \epsilon > 0$. We have

$$\mathbb{E}\left\langle \mathbb{1}\{R^{\mathtt{u}}_{1,*} \in [m, m+\epsilon)\}\right\rangle = \mathbb{E}\left[\frac{\int \mathbb{1}\{R^{\mathtt{u}}_{1,*} \in [m, m+\epsilon)\}e^{-H(\boldsymbol{u},\boldsymbol{v})}\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v})}{\int e^{-H(\boldsymbol{u},\boldsymbol{v})}\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v})}\right]. \tag{3.17}$$

Let

$$A = \frac{1}{N}\mathbb{E}_{\boldsymbol{W}}\log\int \mathbb{1}\{R^{\mathtt{u}}_{1,*} \in [m, m+\epsilon)\}e^{-H(\boldsymbol{u},\boldsymbol{v})}\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v}),$$

and

$$B = \frac{1}{N}\mathbb{E}_{\boldsymbol{W}}\log\int e^{-H(\boldsymbol{u},\boldsymbol{v})}\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v}).$$

By concentration over the Gaussian disorder, Lemma 3.14, for any $u \geq 0$, we simultaneously have

$$\frac{1}{N}\log\int \mathbb{1}\{R^{\mathtt{u}}_{1,*} \in [m, m+\epsilon)\}e^{-H(\boldsymbol{u},\boldsymbol{v})}\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v}) - A \leq u,$$

and

$$\frac{1}{N}\log\int e^{-H(\boldsymbol{u},\boldsymbol{v})}\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v}) - B \geq -u,$$

with probability at least $1-4e^{-Nu^2/(2\beta K_{\mathrm{u}}^2K_{\mathrm{v}}^2)}$. On the complement event we simply upper bound the fraction (3.17) by 1. Therefore, we have

$$\mathbb{E}\left\langle \mathbb{1}\{R_{1,*}^{\mathrm{u}} \in [m,m+\epsilon)\}\right\rangle \le \mathbb{E}_{\boldsymbol{u}^*,\boldsymbol{v}^*}\left[e^{N(A-B+2u)}\right] + 4e^{-Nu^2/(2\beta K_{\mathrm{u}}^2K_{\mathrm{v}}^2)}.$$

By Proposition 3.13 we have $A \le \frac{\Delta}{2\sigma_{\mathrm{u}}^2K_{\mathrm{u}}^2}m^2 + \alpha\beta K_{\mathrm{v}}^2\epsilon$ deterministically over $\boldsymbol{u}^*$ and $\boldsymbol{v}^*$. Now it remains to control $\mathbb{E}_{\boldsymbol{u}^*,\boldsymbol{v}^*}\left[e^{-NB}\right]$.

Lemma 3.16. *We have* $\mathbb{E}_{\boldsymbol{u}^*,\boldsymbol{v}^*}\left[e^{-NB}\right] \le 2e^{-N\,\mathbb{E}_{\boldsymbol{u}^*,\boldsymbol{v}^*}[B]}$.

Moreover, observe that

$$\begin{aligned}\mathbb{E}_{\boldsymbol{u}^*,\boldsymbol{v}^*}[B] &= \frac{1}{N}\,\mathbb{E}\log\int e^{-H(\boldsymbol{u},\boldsymbol{v})}\mathrm{d}\rho(\boldsymbol{u},\boldsymbol{v})\\ &= \frac{1}{N}\,\mathbb{E}_{\mathbb{P}_\beta}\log L(\boldsymbol{Y};\beta)\\ &= \frac{1}{N}\,\mathbb{E}_{\mathbb{P}_0}L(\boldsymbol{Y};\beta)\log L(\boldsymbol{Y};\beta) \ge 0.\end{aligned}$$

Positivity is obtained by Jensen's inequality and convexity of $x \mapsto x\log x$. In view of the above, Lemma 3.16 means that the random variable B is "essentially" positive. Therefore,

$$\mathbb{E}\left\langle \mathbb{1}\{R_{1,*}^{\mathrm{u}} \in [m,m+\epsilon)\}\right\rangle \le 2e^{N(\delta+2u)} + 4e^{-Nu^2/(2\beta K_{\mathrm{u}}^2K_{\mathrm{v}}^2)},$$

where $\delta = \frac{\Delta}{2\sigma_{\mathrm{u}}^2K_{\mathrm{u}}^2}m^2 + \alpha\beta K_{\mathrm{v}}^2\epsilon$. We let $u = -\delta/3 \ge 0$, and $m = \epsilon' + l\epsilon$. Since $\Delta < 0$, $\Delta m^2 \le \Delta\epsilon'^2$. Now we let $\epsilon = -\frac{\Delta}{4\alpha\beta\sigma_{\mathrm{u}}^2K_{\mathrm{u}}^2K_{\mathrm{v}}^2}\epsilon'^2$ so that $\delta \le \frac{3\Delta}{4\sigma_{\mathrm{u}}^2K_{\mathrm{u}}^2}\epsilon'^2 < 0$. ■

***Proof** of Lemma 3.16.* We abbreviate $\mathbb{E}_{\boldsymbol{u}^*,\boldsymbol{v}^*}$ by $\mathbb{E}$. We have

$$\mathbb{E}\left[e^{N(\mathbb{E}[B]-B)}\right] = \int_{-\infty}^{+\infty} e^t\,\Pr\left(N(\mathbb{E}[B]-B) \ge t\right)\mathrm{d}t \le 1 + \int_0^{+\infty} e^t\,\Pr\left(N(\mathbb{E}[B]-B)\ge t\right)\mathrm{d}t.$$

Now we bound the lower tail probability. The r.v. B, seen as a function of the vector $[\boldsymbol{u}^*|\boldsymbol{v}^*] \in \mathbb{R}^{N+M}$ is jointly convex (the Hessian can be easily shown to be positive semi-definite), and Lipschitz with constant $\beta K_{\mathrm{u}}K_{\mathrm{v}}\sqrt{\frac{\alpha K_{\mathrm{u}}^2+\alpha^2K_{\mathrm{v}}^2}{N}}$ with respect to the ℓ_2 norm. Under the above conditions, a bound on the lower tail of deviation of B is available; this is (one side of) Talagrand's inequality (see Boucheron, Lugosi, and Massart, 2013, Theorem 7.12). Therefore, we have for all $t \ge 0$

$$\Pr\left(B - \mathbb{E}[B] \le -t\right) \le e^{-Nt^2/2K^2},$$

where $K^2 = \alpha\beta^2K_{\mathrm{u}}^2K_{\mathrm{v}}^2(K_{\mathrm{u}}^2+\alpha K_{\mathrm{v}}^2)$. Thus,

$$\mathbb{E}\left[e^{N(\mathbb{E}[B]-B)}\right] \le 1 + \int_0^{+\infty} e^te^{-t^2/(2NK^2)}\mathrm{d}t$$

$$
\begin{aligned}
&= 1 + K\sqrt{N} e^{NK^2/2} \int_{K\sqrt{N}}^{+\infty} e^{-t^2/2} \mathrm{d}t \\
&\leq 2.
\end{aligned}
$$

The last inequality is a restatement of the fact $\Pr(g \geq t) \leq \frac{e^{-t^2/2}}{\sqrt{2\pi} t}$ where $g \sim \mathcal{N}(0,1)$. ■

Convergence of the fourth moment

In this section we prove that for all α, β such that $\alpha\beta^2\sigma_{\mathtt{u}}^2\sigma_{\mathtt{v}}^2 K_{\mathtt{u}}^2 K_{\mathtt{v}}^2 < 1$, we have

$$
\mathbb{E}\left\langle (R_{1,2}^{\mathtt{u}})^4 \right\rangle \vee \mathbb{E}\left\langle (R_{1,2}^{\mathtt{v}})^4 \right\rangle \leq \frac{K(\alpha,\beta)}{N^2}.
$$

We proceed as follows. Let

$$
M = \max\left\{ \mathbb{E}\left\langle (R_{1,2}^{\mathtt{u}})^4 \right\rangle, \mathbb{E}\left\langle (R_{1,2}^{\mathtt{v}})^4 \right\rangle \right\}.
$$

We prove that for $\epsilon > 0$, the following self-boundedness properties hold:

$$
\mathbb{E}\left\langle (R_{1,2}^{\mathtt{u}})^4 \right\rangle \leq \alpha\beta^2\, \mathbb{E}\left\langle (R_{1,2}^{\mathtt{u}})^4 \right\rangle + K\epsilon M + \delta, \tag{3.18}
$$

$$
\mathbb{E}\left\langle (R_{1,2}^{\mathtt{v}})^4 \right\rangle \leq \alpha\beta^2\, \mathbb{E}\left\langle (R_{1,2}^{\mathtt{v}})^4 \right\rangle + K\epsilon M + \delta, \tag{3.19}
$$

where $\delta \leq K/N^2 + K/\epsilon^2 e^{-c(\epsilon)N}$. This implies the desired result by letting ϵ be sufficiently small (e.g., $\epsilon = (1-\alpha\beta^2)/2$). We prove (3.18) and (3.19) using the cavity method, i.e. by isolating the effect of the last variables u_N and v_M, one at a time. We prove (3.18) in full detail, then briefly highlight how (3.19) is obtained in a similar way.

By symmetry between the $\mathtt{u}$ variables, we have

$$
\begin{aligned}
\mathbb{E}\left\langle (R_{1,*}^{\mathtt{u}})^4 \right\rangle &= \mathbb{E}\left\langle u_N^{(1)} u_N^* (R_{1,*}^{\mathtt{u}})^3 \right\rangle \\
&= \mathbb{E}\left\langle u_N^{(1)} u_N^* \left(R_{1,*}^{\mathtt{u}-} + \frac{1}{N} u_N^{(1)} u_N^* \right)^3 \right\rangle.
\end{aligned}
$$

Expanding the term $\left(R_{1,*}^{\mathtt{u}-} + \frac{1}{N} u_N^{(1)} u_N^*\right)^3$ we obtain

$$
\mathbb{E}\left\langle (R_{1,*}^{\mathtt{u}})^4 \right\rangle \leq \mathbb{E}\left\langle u_N^{(1)} u_N^* (R_{1,*}^{\mathtt{u}-})^3 \right\rangle + \frac{K_{\mathtt{u}}^4}{N} \mathbb{E}\left\langle (R_{1,*}^{\mathtt{u}-})^2 \right\rangle + \frac{K_{\mathtt{u}}^6}{N^2} \mathbb{E}\left\langle |R_{1,*}^{\mathtt{u}-}| \right\rangle + \frac{K_{\mathtt{u}}^8}{N^3}. \tag{3.20}
$$

We have already proved convergence of the second moment (Proposition 3.12), hence $\mathbb{E}\langle (R_{1,*}^{\mathtt{u}-})^2 \rangle \leq K/N$ and $\mathbb{E}\langle |R_{1,*}^{\mathtt{u}-}| \rangle \leq K/\sqrt{N}$. Now we need to control the leading term involving $(R_{1,*}^{\mathtt{u}-})^3$. The next proposition shows that this quantity can be related back to $(R_{1,*}^{\mathtt{u}})^4$, plus additional higher-order terms. This is is achieved through the cavity method.

Proposition 3.17. *For $\alpha, \beta \geq 0$, there exists a constant $K = K(\alpha, \beta, K_{\mathtt{u}}, K_{\mathtt{v}}) > 0$ such that*

$$\mathbb{E}\left\langle u_N^{(1)} u_N^* (R^{\mathtt{u}-}_{1,*})^3 \right\rangle = \beta\, \mathbb{E}\left\langle (R^{\mathtt{u}}_{1,*})^3 R^{\mathtt{v}}_{1,*} \right\rangle + \delta_1, \tag{3.21}$$

where

$$|\delta_1| \leq K \sum_{a,b,c,d} \mathbb{E}\left\langle \left| (R^{\mathtt{u}-}_{1,*})^3 R^{\mathtt{v}}_{a,b} R^{\mathtt{v}}_{c,d} \right| \right\rangle .$$

Moreover,

$$\mathbb{E}\left\langle (R^{\mathtt{u}}_{1,*})^3 R^{\mathtt{v}}_{1,*} \right\rangle = \alpha\beta\, \mathbb{E}\left\langle (R^{\mathtt{u}}_{1,*})^4 \right\rangle + \delta_2, \tag{3.22}$$

where

$$|\delta_2| \leq K \sum_{a,b,c,d} \mathbb{E}\left\langle \left| (R^{\mathtt{u}}_{1,*})^3 R^{\mathtt{u}}_{a,b} R^{\mathtt{u}}_{c,d} \right| \right\rangle .$$

From Proposition 3.17 we deduce

$$\mathbb{E}\left\langle u_N^{(1)} u_N^* (R^{\mathtt{u}-}_{1,*})^3 \right\rangle = \alpha\beta^2\, \mathbb{E}\left\langle (R^{\mathtt{u}}_{1,*})^4 \right\rangle + \delta,$$

where $\delta = \delta_1 + \delta_2$. Plugging into (3.20), we obtain

$$\mathbb{E}\left\langle (R^{\mathtt{u}}_{1,*})^4 \right\rangle \leq \alpha\beta^2\, \mathbb{E}\left\langle (R^{\mathtt{u}}_{1,*})^4 \right\rangle + \frac{K}{N^2} + \delta.$$

Now we need to control the error term δ, which involves monomials of degree 5 in the overlaps $R^{\mathtt{u}}$ and $R^{\mathtt{v}}$. This is where the a priori bound on the convergence of the overlaps, Proposition 4.12, is useful. Since the overlaps are bounded, we can write for any $\epsilon > 0$,

$$\begin{aligned}
\mathbb{E}\left\langle \left| (R^{\mathtt{u}}_{1,*})^3 R^{\mathtt{v}}_{a,b} R^{\mathtt{v}}_{c,d} \right| \right\rangle &\leq \epsilon\, \mathbb{E}\left\langle \left| (R^{\mathtt{u}}_{1,*})^3 R^{\mathtt{v}}_{a,b} \right| \right\rangle + K_{\mathtt{u}}^6 K_{\mathtt{v}}^4\, \mathbb{E}\left\langle \mathbb{1}\{ \left| R^{\mathtt{v}}_{c,d} \right| \geq \epsilon \} \right\rangle \\
&= \epsilon\, \mathbb{E}\left\langle \left| (R^{\mathtt{u}}_{1,*})^3 R^{\mathtt{v}}_{a,b} \right| \right\rangle + K_{\mathtt{u}}^6 K_{\mathtt{v}}^4\, \mathbb{E}\left\langle \mathbb{1}\{ \left| R^{\mathtt{v}}_{1,*} \right| \geq \epsilon \} \right\rangle ,
\end{aligned}$$

where the last line is a consequence of the Nishimori property. Now we use Hölder's inequality on the first term:

$$\begin{aligned}
\mathbb{E}\left\langle \left| (R^{\mathtt{u}}_{1,*})^3 R^{\mathtt{v}}_{a,b} \right| \right\rangle &\leq \left(\mathbb{E}\left\langle \left| (R^{\mathtt{u}}_{1,*})^4 \right| \right\rangle \right)^{3/4} \left(\mathbb{E}\left\langle \left| (R^{\mathtt{v}}_{a,b} \right|)^4 \right\rangle \right)^{1/4} \\
&= \left(\mathbb{E}\left\langle \left| (R^{\mathtt{u}}_{1,*})^4 \right| \right\rangle \right)^{3/4} \left(\mathbb{E}\left\langle \left| (R^{\mathtt{v}}_{1,*} \right|)^4 \right\rangle \right)^{1/4} \\
&\leq M.
\end{aligned}$$

Using Proposition 4.12, we have $\mathbb{E}\langle \mathbb{1}\{ |R^{\mathtt{v}}_{1,*}| \geq \epsilon \} \rangle \leq K e^{-cN}/\epsilon^2$. Therefore,

$$|\delta_1| \leq K\epsilon M + \frac{K}{\epsilon^2} e^{-cN}.$$

It is clear that we can use the same argument to bound δ_2, so we end up with

$$\mathbb{E}\left\langle (R^{\mathtt{u}}_{1,*})^4 \right\rangle \leq \alpha\beta^2\, \mathbb{E}\left\langle (R^{\mathtt{u}}_{1,*})^4 \right\rangle + \frac{K}{N^2} + K\epsilon M + \frac{K}{\epsilon^2} e^{-cN},$$

thereby proving (3.18). To prove (3.19) we use the same approach. We write

$$\begin{aligned}\mathbb{E}\left\langle (R^{\mathtt{v}}_{1,*})^4\right\rangle &= \frac{M}{N}\,\mathbb{E}\left\langle v^{(1)}_M v^*_M (R^{\mathtt{v}}_{1,*})^3\right\rangle\\ &= \alpha\,\mathbb{E}\left\langle v^{(1)}_M v^*_M\Big(R^{\mathtt{v}-}_{1,*}+\frac{1}{N}v^{(1)}_M v^*_M\Big)^3\right\rangle.\end{aligned}$$

Then use an equivalent of Proposition 3.17 in this case, which is obtained by flipping the role of the $\mathtt{u}$ and $\mathtt{v}$ variables:

$$\mathbb{E}\left\langle v^{(1)}_N v^*_N (R^{\mathtt{v}-}_{1,*})^3\right\rangle = \beta\,\mathbb{E}\left\langle (R^{\mathtt{v}}_{1,*})^3 R^{\mathtt{u}}_{1,*}\right\rangle + \delta_1,$$

and

$$\mathbb{E}\left\langle (R^{\mathtt{v}}_{1,*})^3 R^{\mathtt{u}}_{1,*}\right\rangle = \beta\,\mathbb{E}\left\langle (R^{\mathtt{v}}_{1,*})^4\right\rangle + \delta_2,$$

where δ_1 and δ_2 are similarly bounded by expectations of monomials of degree 5 in the overlaps $R^{\mathtt{u}}$ and $R^{\mathtt{v}}$. These two quantities are then bounded in exactly the same way.

Proof of Proposition 3.17. The proof uses two interpolations; the first one decouples the variable u_N from the rest of the system and allows to obtain (3.21), and the second one decouples the variable v_M and allows to obtain (3.22). We start with the former.

Proof of (3.21)**.** Consider the interpolating Hamiltonian

$$\begin{aligned}-H_t(\boldsymbol{u},\boldsymbol{v}) = &\sum_{i=1}^{N-1}\sum_{j=1}^{M}\sqrt{\frac{\beta}{N}}W_{ij}u_iv_j + \frac{\beta}{N}u_iu_i^*v_jv_j^* - \frac{\beta}{2N}u_i^2v_j^2\\ &+\sum_{j=1}^{M}\sqrt{\frac{\beta t}{N}}W_{Nj}u_Nv_j + \frac{\beta t}{N}u_Nu_N^*v_jv_j^* - \frac{\beta t}{2N}u_N^2v_j^2,\end{aligned}$$

and let $\langle\cdot\rangle_t$ be the associated Gibbs average and $\nu_t(\cdot) = \mathbb{E}\langle\cdot\rangle_t$. The idea is to approximate $\nu_1(f)$ where $f \equiv u^{(1)}_N u^*_N (R^{\mathtt{u}-}_{1,*})^3$ by $\nu_0(f)+\nu_0'(f)$. Of course one then has to control the second derivative, as dictated by Taylor's approximation

$$|\nu_1(f)-\nu_0(f)-\nu_0'(f)| \le \sup_{0\le t\le 1}|\nu_t''(f)|\,. \tag{3.23}$$

We see that at time $t=0$, the variables u_N and u_N^* decouple the Hamiltonian, so

$$\nu_0(u^{(1)}_N u^*_N (R^{\mathtt{u}-}_{1,*})^3) = \mathbb{E}[u_N]\,\mathbb{E}[u_N^*]\nu_0((R^{\mathtt{u}-}_{1,*})^3) = 0. \tag{3.24}$$

On the other hand, by applying Lemma 3.10 with $n=1$, we see that $\nu_0'(u^{(1)}_N u^*_N(R^{\mathtt{u}-}_{1,*})^3)$ is a sum of a few terms of the form

$$\nu_0\big(u^{(1)}_N u^*_N u^{(a)}_N u^{(b)}_N (R^{\mathtt{u}-}_{1,*})^3 R^{\mathtt{v}}_{a,b}\big).$$

Since $P_{\mathtt{u}}$ has zero mean, all terms in which a variable $u_N^{(a)}$ (for any a) appears with degree 1 vanish. We are thus left with one term where $a = 1, b = *$, and we get

$$\nu_0'(u_N^{(1)} u_N^*(R_{1,*}^{\mathtt{u}-})^3) = \beta\, \mathbb{E}[(u_N^{(1)})^2]\, \mathbb{E}[(u_N^*)^2] \nu_0((R_{1,*}^{\mathtt{u}-})^3 R_{1,*}^{\mathtt{v}}) = \beta \nu_0((R_{1,*}^{\mathtt{u}-})^3 R_{1,*}^{\mathtt{v}}). \tag{3.25}$$

Moreover, we see that $\nu_0((R_{1,*}^{\mathtt{u}-})^3 R_{1,*}^{\mathtt{v}}) = \nu_0((R_{1,*}^{\mathtt{u}})^3 R_{1,*}^{\mathtt{v}})$ since the last variable u_N has no contribution under ν_0. Now we are tempted to replace the average at time $t = 0$ by an average at time $t = 1$ in the last quantity. We use Lemmas 3.10 and 3.11 to justify this. Indeed these lemmas and boundedness of the variables u_N imply

$$\left|\nu_0((R_{1,*}^{\mathtt{u}})^3 R_{1,*}^{\mathtt{v}}) - \nu_1((R_{1,*}^{\mathtt{u}})^3 R_{1,*}^{\mathtt{v}})\right| \le K(\alpha,\beta) \sum_{a,b} \nu(\left|(R_{1,*}^{\mathtt{u}})^3 R_{1,*}^{\mathtt{v}} R_{a,b}^{\mathtt{v}})\right|), \tag{3.26}$$

where $(a, b) \in \{(1, 2), (1, *), (2, *), (2, 3)\}$. Now we control the second derivative $\sup_t \nu_t''(\cdot)$. In view of Lemma 3.10, we see that taking two derivative of $\nu_t(u_N^{(1)} u_N^* (R_{1,*}^{\mathtt{u}-})^3)$ creates terms of the form

$$\nu_t\left(u_N^{(1)} u_N^* u_N^{(a)} u_N^{(b)} u_N^{(c)} u_N^{(d)} (R_{1,*}^{\mathtt{u}-})^3 R_{a,b}^{\mathtt{v}} R_{c,d}^{\mathtt{v}}\right),$$

with a larger (but finite) set of combinations (a, b, c, d). We use Lemma 3.11 to replace ν_t by ν_1 and use boundedness of variables u_N to obtain the bound

$$\left|\sup_{0\le t\le 1} \nu_t''\Big(u_N^{(1)} u_N^* (R_{1,*}^{\mathtt{u}-})^3)\Big)\right| \le K(\alpha,\beta) \sum_{a,b,c,d} \nu\Big(\Big|(R_{1,*}^{\mathtt{u}-})^3 R_{a,b}^{\mathtt{v}} R_{c,d}^{\mathtt{v}}\Big|\Big). \tag{3.27}$$

Now putting the bounds and estimates (3.23), (3.24), (3.25), (3.26), and (3.27), we obtain the desired bound (3.21):

$$\left|\nu\big(u_N^{(1)} u_N^* (R_{1,*}^{\mathtt{u}-})^3\big) - \beta \nu((R_{1,*}^{\mathtt{u}})^3 R_{1,*}^{\mathtt{v}})\right| \le K(\alpha,\beta) \sum_{a,b,c,d} \nu\Big(\Big|(R_{1,*}^{\mathtt{u}-})^3 R_{a,b}^{\mathtt{v}} R_{c,d}^{\mathtt{v}}\Big|\Big).$$

Proof of (3.22)**.** By symmetry of the $\mathtt{v}$ variables we have

$$\mathbb{E}\left\langle (R_{1,*}^{\mathtt{u}})^3 R_{1,*}^{\mathtt{v}} \right\rangle = \frac{M}{N}\, \mathbb{E}\left\langle (R_{1,*}^{\mathtt{u}})^3 v_M^{(1)} v_M^* \right\rangle.$$

Now we apply the same machinery. Consider the interpolating Hamiltonian

$$\begin{aligned} -H_t(\boldsymbol{u},\boldsymbol{v}) &= \sum_{j=1}^{M-1} \sum_{i=1}^{N} \sqrt{\frac{\beta}{N}} W_{ij} u_i v_j + \frac{\beta}{N} u_i u_i^* v_j v_j^* - \frac{\beta}{2N} u_i^2 v_j^2 \\ &\quad + \sum_{i=1}^{N} \sqrt{\frac{\beta t}{N}} W_{iM} u_i v_M + \frac{\beta t}{N} u_i u_i^* v_M v_M^* - \frac{\beta t}{2N} u_i^2 v_M^2, \end{aligned}$$

and let $\langle\cdot\rangle_t$ be the associated Gibbs average and $\nu_t(\cdot) = \mathbb{E}\langle\cdot\rangle_t$. The exact same argument goes through with the roles of $\mathtt{u}$ and $\mathtt{v}$ flipped. For instance, when one takes time derivatives, terms of the form $v_M^{(a)} v_M^{(b)} R^{\mathtt{u}}_{a,b}$ arise from the Hamiltonian, and one sees that

$$\nu_0'\big((R^{\mathtt{u}}_{1,*})^3 v_M^{(1)} v_M^*\big) = \beta\nu_0\big((R^{\mathtt{u}}_{1,*})^4\big).$$

Thus we similarly obtain

$$\Big|\nu\big((R^{\mathtt{u}}_{1,*})^3 v_M^{(1)} v_M^*\big) - \beta\nu\big((R^{\mathtt{u}}_{1,*})^4\big)\Big| \le K(\alpha,\beta)\sum_{a,b,c,d}\nu\Big(\Big|(R^{\mathtt{u}}_{1,*})^3 R^{\mathtt{u}}_{a,b} R^{\mathtt{u}}_{c,d}\Big|\Big).$$

■

3.10 A supplement: Latała's argument

In this section we present a proof of convergence of the overlaps under the null model $\mathbb{P}_0$. As alluded to earlier, the region (α,β) where this convergence is obtained is much smaller than under the alternative. The argument, discovered (but not published) by R. Latała, is nevertheless worth presenting as it is short and quite natural. This can also be applied to the symmetric model of Chapter 2, but in this case the argument is the same as for the SK model, and we refer to Talagrand (2011a) for its execution.

Overlap convergence under the null

Under the null model, we prove overlap convergence by bounding their moment generating function. Let $\boldsymbol{Y} \sim \mathbb{P}_0$.

Proposition 3.18. *For all $\alpha,\beta \ge 0$ such that $16K_{\mathtt{u}}^4\beta < 1$ and $16K_{\mathtt{v}}^4\alpha\beta < 1$, there exists $K = K(\alpha,\beta) < \infty$ such that*

$$\mathbb{E}\left\langle (R^{\mathtt{u}}_{1,2})^4\right\rangle \vee \mathbb{E}\left\langle (R^{\mathtt{v}}_{1,2})^4\right\rangle \le \frac{K}{N^2}.$$

Proof. Consider a family of interpolating Hamiltonians

$$-H_t(\boldsymbol{u},\boldsymbol{v}) = \sum_{i,j}\sqrt{\frac{\beta t}{N}}Y_{ij}u_iv_j - \frac{\beta t}{2N}u_i^2v_j^2,$$

for $t \in [0,1]$. Let $\langle\cdot\rangle_t$ be the associated Gibbs average, and let $\nu_t(\cdot) = \mathbb{E}\langle\cdot\rangle_t$. Let $C_{1,2} = |R^{\mathtt{u}}_{1,2}| \vee |R^{\mathtt{v}}_{1,2}|$, and for $\gamma \ge 0$ consider the function

$$\psi(t,\gamma) = \nu_t(e^{\gamma N C_{1,2}^2}).$$

We will relate the value of ψ at time 1 to its behavior at time 0, which is easier to control.

Lemma 3.19. *For all $\gamma \geq 0$, we have*

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\psi(t,\gamma)\right| \leq 8\beta N \nu_t\left(C_{1,2}^2 e^{\gamma N C_{1,2}^2}\right).$$

Consequently, the function $t \mapsto \psi(t, \gamma - 8\beta t)$ is non-increasing for all $\gamma \geq 8\beta$. Indeed,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\psi(t,\gamma-8\beta t) &= \partial_x\psi(t,\gamma-8\beta t) - 8\beta\partial_y\psi(t,\gamma-8\beta t)\\ &= \partial_x\psi(t,\gamma-8\beta t) - 8\beta N\nu_t\left(C_{1,2}^2 e^{\gamma N C_{1,2}^2}\right) \leq 0.\end{aligned}$$

where the last line follows from Lemma 3.19. Therefore $\psi(1,\gamma-8\beta) \leq \psi(0,\gamma)$, i.e.,

$$\mathbb{E}\left\langle e^{(\gamma-8\beta)NC_{1,2}^2}\right\rangle \leq \mathbb{E}\left\langle e^{\gamma N C_{1,2}^2}\right\rangle_0 = \int e^{\gamma N C_{1,2}^2}\mathrm{d}\rho(\boldsymbol{u}^{(1)},\boldsymbol{v}^{(1)})\mathrm{d}\rho(\boldsymbol{u}^{(2)},\boldsymbol{v}^{(2)}).$$

Lemma 3.20. *For all $\gamma \geq 0$, we have $\mathbb{E}\left\langle e^{\gamma N C_{1,2}^2}\right\rangle_0 \leq K_1(\gamma)K_2(\gamma)$, where $K_1(\gamma) = (1 - 2\gamma K_{\mathtt{u}}^4)^{-1/2}$, and $K_2(\gamma) = (1-2\gamma\frac{M}{N}K_{\mathtt{v}}^4)^{-1/2}$.*

From here we deduce that for all $\gamma \geq 0$

$$\frac{\gamma^2}{2}N^2\,\mathbb{E}\left\langle C_{1,2}^4\right\rangle \leq \mathbb{E}\left\langle e^{\gamma N C_{1,2}^2}\right\rangle \leq K_1(\gamma+8\beta)K_2(\gamma+8\beta).$$

We finish the proof by letting $\gamma = \frac{1-16\beta K_{\mathtt{u}}^4}{4K_{\mathtt{u}}^4} \vee \frac{1-16\alpha\beta K_{\mathtt{v}}^4}{4\alpha K_{\mathtt{v}}^4}$. ■

It remains to prove Lemma 3.19 and Lemma 3.20.

***Proof** of Lemma 3.19.* A short calculation shows that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\psi(t,\gamma) = \beta N\nu_t&\left((R_{1,2}^{\mathtt{u}}R_{1,2}^{\mathtt{v}} - 2R_{1,3}^{\mathtt{u}}R_{1,3}^{\mathtt{v}} - 2R_{2,3}^{\mathtt{u}}R_{2,3}^{\mathtt{v}})e^{\gamma N C_{1,2}^2}\right)\\ &+ 3\beta N\nu_t\left(R_{3,4}^{\mathtt{u}}R_{3,4}^{\mathtt{v}}e^{\gamma N C_{1,2}^2}\right).\end{aligned}$$

By the triangle inequality and the upper bound $\left|R_{a,b}^{\mathtt{u}}R_{a,b}^{\mathtt{v}}\right| \leq C_{a,b}^2$, we obtain

$$\left|\frac{\mathrm{d}}{\mathrm{d}t}\psi(t,\gamma)\right| = \beta N\nu_t\left((C_{1,2}^2 + 2C_{1,3}^2 + 2C_{2,3}^2 + 3C_{3,4}^2)e^{\gamma N C_{1,2}^2}\right).$$

Let us examine a generic term of the form

$$\nu_t\left(C_{a,b}^2 e^{\gamma N C_{c,d}^2}\right),$$

for $a,b,c,d \in \{1,2,3,4\}$. By expanding the Taylor series of the exponential, and monotone convergence, we have

$$\nu_t\left(C_{a,b}^2 e^{\gamma N C_{c,d}^2}\right) = \sum_{k=0}^{\infty}\frac{(\gamma N)^k}{k!}\nu_t\left(C_{a,b}^2 C_{c,d}^{2k}\right).$$

We use Hölder's inequality on the inner terms with $p = k+1$ and $1/p + 1/q = 1$ to obtain

$$\nu_t\left(C_{a,b}^2 C_{c,d}^{2k}\right) \leq \nu_t\left(C_{a,b}^{2p}\right)^{1/p} \cdot \nu_t\left(C_{c,d}^{2kq}\right)^{1/q} = \nu_t\left(C_{a,b}^{2(k+1)}\right).$$

Now we sum the series and finally obtain a "symmetrization" bound

$$\nu_t\left(C_{a,b}^2 e^{\gamma N C_{c,d}^2}\right) \leq \nu_t\left(C_{a,b}^2 e^{\gamma N C_{a,b}^2}\right).$$

The conclusion then follows. ■

***Proof** of Lemma 3.20.* We have

$$\mathbb{E}\left\langle e^{\gamma N C_{1,2}^2}\right\rangle_0 = \int e^{\gamma N C_{1,2}^2} \mathrm{d}P_{\mathtt{u}}^{\otimes N}(\boldsymbol{u}^{(1)}) \mathrm{d}P_{\mathtt{u}}^{\otimes N}(\boldsymbol{u}^{(2)}) \mathrm{d}P_{\mathtt{v}}^{\otimes M}(\boldsymbol{v}^{(1)}) \mathrm{d}P_{\mathtt{v}}^{\otimes M}(\boldsymbol{v}^{(2)}).$$

Since $C_{1,2}^2 \leq (R_{1,2}^{\mathtt{u}})^2 + (R_{1,2}^{\mathtt{v}})^2$, we only need to control each overlap separately. We introduce an independent Gaussian r.v. g and write

$$\int e^{\gamma N (R_{1,2}^{\mathtt{u}})^2} \mathrm{d}P_{\mathtt{u}}^{\otimes N}(\boldsymbol{u}^{(1)}) \mathrm{d}P_{\mathtt{u}}^{\otimes N}(\boldsymbol{u}^{(2)}) = \mathbb{E}_g \int e^{\sqrt{2\gamma N} g R_{1,2}^{\mathtt{u}}} \mathrm{d}P_{\mathtt{u}}^{\otimes N}(\boldsymbol{u}^{(1)}) \mathrm{d}P_{\mathtt{u}}^{\otimes N}(\boldsymbol{u}^{(2)}).$$

Sub-Gaussianity of $P_{\mathtt{u}}$ (with parameter $\sigma_{\mathtt{u}}$) implies that the above is bounded by

$$\begin{aligned} \mathbb{E}_g \int e^{\frac{\gamma}{N} g \sum_{i=1}^N u_i^{(2)2} \sigma_{\mathtt{u}}^2} \mathrm{d}P_{\mathtt{u}}^{\otimes N}(\boldsymbol{u}^{(2)}) &\leq \mathbb{E}_g\left[e^{\gamma g \sigma_{\mathtt{u}}^2 K_{\mathtt{u}}^2}\right] \\ &\leq \mathbb{E}_g\left[e^{\gamma g K_{\mathtt{u}}^4}\right] \\ &= \left(1 - 2\gamma K_{\mathtt{u}}^4\right)^{-1/2} =: K_1(\gamma). \end{aligned}$$

By the same argument, we have

$$\int e^{\gamma N (R_{1,2}^{\mathtt{v}})^2} \mathrm{d}P_{\mathtt{v}}^{\otimes M}(\boldsymbol{v}^{(1)}) \mathrm{d}P_{\mathtt{v}}^{\otimes M}(\boldsymbol{v}^{(2)}) \leq \left(1 - 2\gamma \frac{M}{N} K_{\mathtt{v}}^4\right)^{-1/2} =: K_2(\gamma).$$

Therefore

$$\mathbb{E}\left\langle e^{\gamma N C_{1,2}^2}\right\rangle_0 \leq K_1(\gamma) K_2(\gamma).$$

■

Chapter 4

The replica-symmetric formula and its finite-size corrections

In this chapter we return to the case of the symmetric model, i.e., the spiked Wigner model,

$$\boldsymbol{Y} = \sqrt{\frac{\lambda}{N}}\boldsymbol{x}^*\boldsymbol{x}^{*\top} + \boldsymbol{W}, \tag{4.1}$$

where $W_{ij} = W_{ji} \sim \mathcal{N}(0,1)$ and $W_{ii} \sim \mathcal{N}(0,2)$ are independent for all $1 \leq i \leq j \leq N$. The entries of $\boldsymbol{x}^*$ are drawn i.i.d. from a (Borel) prior P_{x} on $\mathbb{R}$ with bounded support, so that the scaling in the above model puts the problem in a high-noise regime where only partial recovery of the spike is possible.

While the previous two chapters dealt with the question of detection of the spike, in this chapter we pay attention to the estimation problem: *for what values of the SNR λ is it possible to estimate the spike $\boldsymbol{x}^*$ with non-trivial accuracy?* We recall that spectral methods, or more precisely, estimation using the top eigenvector of $\boldsymbol{Y}$, are known to succeed above a *spectral threshold* and fail below (Benaych-Georges and Nadakuditi, 2011). Since the posterior mean is the estimator with minimal mean squared error, this question boils down to the study of the posterior distribution of $\boldsymbol{x}^*$ given $\boldsymbol{Y}$, which by Bayes' rule, can be written as

$$\mathrm{d}\,\mathbb{P}_\lambda(\boldsymbol{x}|\boldsymbol{Y}) = \frac{e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})}{\int e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})}, \tag{4.2}$$

where H is the Hamiltonian

$$\begin{aligned} -H(\boldsymbol{x}) &:= \sum_{i<j} \sqrt{\frac{\lambda}{N}} Y_{ij}x_i x_j - \frac{\lambda}{2N}x_i^2 x_j^2 \qquad (4.3) \\ &= \sum_{i<j} \sqrt{\frac{\lambda}{N}} W_{ij}x_i x_j + \frac{\lambda}{N}x_i x_j x_i^* x_j^* - \frac{\lambda}{2N}x_i^2 x_j^2. \end{aligned}$$

The free energy of the model is defined as the expected log-partition function (i.e., normalizing constant) of the posterior $\mathbb{P}_\lambda(\cdot|\boldsymbol{Y})$:

$$F_N = \frac{1}{N}\,\mathbb{E}_{\mathbb{P}_\lambda} \log \int e^{-H(\boldsymbol{x})} \mathrm{d}P_{\mathsf{x}}^{\otimes N}(\boldsymbol{x}), \tag{4.4}$$

Recall that this is equal to $\frac{1}{N}\,\mathbb{E}_{\mathbb{P}_\lambda} \log L(\boldsymbol{Y};\lambda)$ with the notation of Chapter 2. It was initially argued via heuristic replica and cavity computations (Lesieur, Krzakala, and Zdeborová, 2015b, 2017) that F_N converges to a limit $\phi_{\mathsf{RS}}(\lambda)$, which is referred to as the *replica-symmetric formula*, defined as follows: For $r \in \mathbb{R}_+$, let

$$\psi(r) := \mathbb{E}_{x^*,z} \log \int \exp\left(\sqrt{r}zx + rxx^* - \frac{r}{2}x^2\right) \mathrm{d}P_{\mathsf{x}}(x),$$

where $z \sim \mathcal{N}(0,1)$, and $x^* \sim P_{\mathsf{x}}$. Now define the RS potential

$$F(\lambda, q) := \psi(\lambda q) - \frac{\lambda q^2}{4},$$

and the RS formula

$$\phi_{\mathsf{RS}}(\lambda) := \sup_{q \geq 0} F(\lambda, q).$$

This formula, variational in nature, encodes in principle a full characterization of the limits of estimating the spike with non-trivial accuracy. Indeed, and as we will see, various formulae for other information-theoretic quantities can be deduced from it, including the mutual information between $\boldsymbol{x}^*$ and $\boldsymbol{Y}$, the minimal mean squared error of estimating $\boldsymbol{x}^*$ based on $\boldsymbol{Y}$, and the overlap $|\boldsymbol{x}^\top \boldsymbol{x}^*|/N$ of a draw $\boldsymbol{x}$ from the posterior $\mathbb{P}_\lambda(\cdot|\boldsymbol{Y})$ with the spike $\boldsymbol{x}^*$. Most of these claims have subsequently been proved rigorously in a series of papers (Barbier, Dia, et al., 2016; Deshpande, Abbé, and Montanari, 2016; Deshpande and Montanari, 2014; Krzakala, Xu, and Zdeborová, 2016; Lelarge and Miolane, 2016) under various assumptions on the prior.

Theorem 4.1 (Barbier, Dia, et al. (2016); Deshpande, Abbé, and Montanari (2016); Korada and Macris (2009); Krzakala, Xu, and Zdeborová (2016); Lelarge and Miolane (2016))**.** *For all $\lambda \geq 0$,*

$$\lim_{N\to\infty} F_N = \phi_{\mathsf{RS}}(\lambda).$$

The above statement contains precious statistical information. It can be written in at least two other equivalent ways, in terms of the mutual information between $\boldsymbol{x}^*$ and $\boldsymbol{Y}$:

$$\lim_{N\to\infty} \frac{1}{N} I(\boldsymbol{Y}, \boldsymbol{x}^*) \;=\; \frac{\lambda}{4}\left(\mathbb{E}_{P_{\mathsf{x}}}[X^2]\right)^2 - \phi_{\mathsf{RS}}(\lambda),$$

or in terms of the Kullback-Liebler divergence between $\mathbb{P}_\lambda$ and $\mathbb{P}_0$:

$$\lim_{N\to\infty} \frac{1}{N} D_{\mathsf{KL}}(\mathbb{P}_\lambda, \mathbb{P}_0) \;=\; \phi_{\mathsf{RS}}(\lambda).$$

Furthermore, the point $q^*(\lambda)$ achieving the maximum in the RS formula (which can be shown to be unique and finite for almost every λ) can be interpreted as the best overlap any estimator $\widehat{\theta}(\boldsymbol{Y})$ can have with the spike $\boldsymbol{x}^*$. Indeed, we will show that the overlap of a draw $\boldsymbol{x}$ from the posterior $\mathbb{P}_\lambda(\cdot|\boldsymbol{Y})$ with $\boldsymbol{x}^*$ concentrates about $q^*(\lambda)$.

Our main goal in this chapter is to gain a fine understanding of the asymptotic behavior of the log-likelihood ratio $\log L(\boldsymbol{Y};\lambda)$ as N becomes large. We first provide an almost elementary proof of Theorem 4.1, based on simple Gaussian interpolations and straightforward applications of concentration of measure arguments. Second, we determine *the finite-size correction* of F_N to its limit $\phi_{\mathsf{RS}}(\lambda)$: we prove under mild conditions on P_{x} that $N(F_N - \phi_{\mathsf{RS}}(\lambda))$ converges to a limit $\psi_{\mathsf{RS}}(\lambda)$ with rate $\mathcal{O}(1/\sqrt{N})$. Besides providing an explicit rate of convergence of F_N to its limit, this result translates into a formula for the Kullback-Leibler divergence D_{KL} valid for (almost) all $\lambda \geq 0$. We will see that while $D_{\mathsf{KL}}(\mathbb{P}_\lambda, \mathbb{P}_0)$ is extensive in the size of the system above the reconstruction threshold λ_c defined in Chapter 2, it brutally ceases to be so as we cross λ_c, to converge to a constant value in accordance with the formula of Proposition 2.5.

Comment on the existing proofs of Theorem 4.1

The proof of the lower bound $\liminf F_N \geq \phi_{\mathsf{RS}}(\lambda)$ relies on an application of Guerra's interpolation method, and is fairly short and transparent Krzakala, Xu, and Zdeborová (2016). Available proofs of the converse bound $\limsup F_N \leq \phi_{\mathsf{RS}}(\lambda)$ (and overlap concentration) are on the other hand highly involved. Barbier, Dia, et al. (2016) and Deshpande, Abbé, and Montanari (2016) adopt an algorithmic approach: they analyze an Approximate Message Passing (AMP) procedure and show that the produced estimator asymptotically achieves an overlap of $q^*(\lambda)$ with the spike. Thus the posterior mean, being the optimal estimator, must also achieve the same overlap. This allows to prove overlap convergence and thus show the converse bound. A difficulty one has to overcome with this method is that AMP (and supposedly any other algorithm) may fail to achieve the optimal overlap in the presence of first-order phase transitions, which trap the algorithm in a bad local optimum of the RS potential. *Spatial coupling*, an idea from coding theory, is used in Barbier, Dia, et al. (2016) to overcome this problem. Lelarge and Miolane (2016) on the other hand use the Aizenman-Sims-Starr scheme (Aizenman, Sims, and Starr, 2003), a relative of the cavity method developed within spin-glass theory, to prove the upper bound. Barbier and Macris (2017) prove the upper bound via a *adaptive* version of the interpolation method that proceeds via a sequence of intermediate interpolation steps. All the current approaches (perhaps to a lesser extent for (Barbier and Macris, 2017)) require the execution of long and technical arguments.

In this chapter, we show that the upper bound in Theorem 2.7 admits a fairly simple proof based on the same interpolation idea that yielded the lower bound, combined with an application of the Laplace method and concentration of measure. The main idea is to use the Franz-Parisi potential (i.e., a version of the free energy (5.7) of a subsystem of configurations $\boldsymbol{x}$ having a *fixed* overlap with the spike $\boldsymbol{x}^*$). We then proceed by applying the Guerra bound and optimize over this free parameter to obtain an upper bound in the form of a saddle (max-min) formula. A small extra effort is needed to show that the latter is another representation of the RS formula. Our proof thus hinges on a upper bound on this potential, which is may be of independent interest. We first start by presenting the proof of the lower bound, which is a starting point for our argument. As in Chapter 2, we present the proof in the case where we omit the diagonal terms of $\boldsymbol{Y}$. This is only done to keep the displays concise; recovering the general case is straightforward since the diagonal has vanishing contribution. Finally, we mention that the method can be easily generalized to all spiked tensor models of even order Richard and Montanari (2014), thus recovering the main results of Lesieur, Miolane, et al. (2017).

4.1 A short proof of the RS formula

We use the interpolation method of Guerra (2001), which is already used in the previous two chapters as a device supporting the cavity method. The idea now is to construct a continuous interpolation path between the Hamiltonian H and a simpler Hamiltonian that decouples all the variables, and analyze the incremental change in the free energy along the path. We present two versions of this method.

Let $t \in [0,1]$ and consider an interpolating Hamiltonian

$$\begin{aligned} -H_t(\boldsymbol{x}) := \sum_{i<j} \sqrt{\frac{t\lambda}{N}} W_{ij} x_i x_j + \frac{t\lambda}{N} x_i x_i^* x_j x_j^* - \frac{t\lambda}{2N} x_i^2 x_j^2 \\ + \sum_{i=1}^{N} \sqrt{(1-t)r} z_i x_i + (1-t) r x_i x_i^* - \frac{(1-t)r}{2} x_i^2, \end{aligned} \tag{4.5}$$

where the z_i's are i.i.d. standard Gaussian r.v.'s independent of everything else. For $f : (\mathbb{R}^N)^{n+1} \mapsto \mathbb{R}$, we define the Gibbs average of f as

$$\left\langle f(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}, \boldsymbol{x}^*) \right\rangle_t := \frac{\int f(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}, \boldsymbol{x}^*) \prod_{l=1}^n e^{-H_t(\boldsymbol{x}^{(l)})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}^{(l)})}{\int \prod_{l=1}^n e^{-H_t(\boldsymbol{x}^{(l)})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}^{(l)})}. \tag{4.6}$$

This is the average of f with respect to the posterior distribution of n copies $\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}$ of $\boldsymbol{x}^*$ given the augmented set of observations

$$\begin{cases} Y_{ij} &= \sqrt{\frac{t\lambda}{N}} x_i^* x_j^* + W_{ij}, \quad 1 \le i \le j \le N, \\ y_i &= \sqrt{(1-t)r} x_i^* + z_i, \quad 1 \le i \le N. \end{cases} \tag{4.7}$$

The variables $\boldsymbol{x}^{(l)}, l = 1 \cdots, n$ are called *replicas*, and are interpreted as random variables independently drawn from the posterior. When $n = 1$ we simply write $f(\boldsymbol{x}, \boldsymbol{x}^*)$ instead of $f(\boldsymbol{x}^{(1)}, \boldsymbol{x}^*)$. We shall denote the overlaps between two replicas as follows: for $l, l' = 1, \cdots, n, *$, we let

$$R_{l,l'} := \boldsymbol{x}^{(l)} \cdot \boldsymbol{x}^{(l')} = \frac{1}{N}\sum_{i=1}^{N} x_i^{(l)} x_i^{(l')}.$$

The lower bound

Reproducing the argument of Krzakala, Xu, and Zdeborová (2016), we prove that

$$F_N \geq \phi_{\mathsf{RS}}(\lambda) - \mathcal{O}\Big(\frac{1}{N}\Big).$$

We let $r = \lambda q$ in the definition of H_t and let

$$\varphi(t) := \frac{1}{N}\,\mathbb{E}\log\int e^{-H_t(\boldsymbol{x})}\mathrm{d}P_{\mathsf{x}}^{\otimes N}(\boldsymbol{x}).$$

A short calculation based on Gaussian integration by parts shows that

$$\begin{aligned}\varphi'(t) = &-\frac{\lambda}{4}\,\mathbb{E}\left\langle (R_{1,2}-q)^2\right\rangle_t + \frac{\lambda}{4}q^2 + \frac{\lambda}{4N^2}\sum_{i=1}^{N}\mathbb{E}\left\langle x_i^{(1)^2} x_i^{(2)^2}\right\rangle_t \\ &+ \frac{\lambda}{2}\,\mathbb{E}\left\langle (R_{1,*}-q)^2\right\rangle_t - \frac{\lambda}{2}q^2 - \frac{\lambda}{2N^2}\sum_{i=1}^{N}\mathbb{E}\left\langle {x_i}^2 x_i^{*2}\right\rangle_t,\end{aligned}$$

By the Nishimori property (2.19), the expressions involving the pairs $(\boldsymbol{x}, \boldsymbol{x}^*)$ on the one hand and $(\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)})$ on the other in the brackets are equal. We then obtain

$$\varphi'(t) = \frac{\lambda}{4}\,\mathbb{E}\left\langle (R_{1,*}-q)^2\right\rangle_t - \frac{\lambda}{4}q^2 - \frac{\lambda}{4N}\,\mathbb{E}\left\langle {x_N}^2 {x_N^*}^2\right\rangle_t.$$

Observe that the last term is $\mathcal{O}(1/N)$ since the variables x_N are bounded. Moreover, the first term is always non-negative so we obtain

$$\varphi'(t) \geq -\frac{\lambda}{4}q^2 - \frac{K}{N}.$$

Since $\varphi(1) = F_N$ and $\varphi(0) = \psi(\lambda q)$, integrating over t, we obtain for all $q \geq 0$

$$F_N \geq F(\lambda, q) - \frac{K}{N},$$

and this yields the lower bound.

The upper bound

We prove the converse bound

$$F_N \le \phi_{\mathsf{RS}}(\lambda) + \mathcal{O}\Big(\frac{\log N}{\sqrt{N}}\Big).$$

Recall the Franz-Parisi potential (Franz and Parisi, 1995, 1998) from Chapter 2. For $\boldsymbol{x}^* \in \mathbb{R}^N$ fixed, $m \in \mathbb{R}$ and $\epsilon > 0$ we define

$$\Phi_\epsilon(m, \boldsymbol{x}^*) := \frac{1}{N}\,\mathbb{E}\log\int \mathbb{1}\{R_{1,*} \in [m, m+\epsilon)\} e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathsf{x}}^{\otimes N}(\boldsymbol{x}),$$

where the expectation is over $\boldsymbol{W}$. This is the free energy of a subsystem of configurations having an overlap close to a fixed value m with a planted signal $\boldsymbol{x}^*$. It is clear that $\mathbb{E}_{\boldsymbol{x}^*}\,\Phi_\epsilon(m, \boldsymbol{x}^*) \le F_N$. We will argue via the Laplace method and concentration of measure that $\sup_{m\in\mathbb{R}} \mathbb{E}_{\boldsymbol{x}^*}\,\Phi_\epsilon(m, \boldsymbol{x}^*) \approx F_N$, then use Guerra's interpolation to *upper* bound $\Phi_\epsilon(m, \boldsymbol{x}^*)$ (notice that this method yielded a *lower* bound on F_N due to the Nishimori property). Let us define a bit of more notation. For $r \in \mathbb{R}_+, s \in \mathbb{R}$, let

$$\widehat{\psi}(r, s) := \mathbb{E}_z \log\int \exp\Big(\sqrt{r}zx + sx - \frac{r}{2}x^2\Big)\,\mathrm{d}P_{\mathsf{x}}(x),$$

where $z \sim \mathcal{N}(0,1)$, and $\bar{\psi}(r, s) = \mathbb{E}_{x^*}\,\widehat{\psi}(r, sx^*)$ where $x^* \sim P_{\mathsf{x}}$. Moreover, let

$$\widehat{F}(\lambda, m, q, \boldsymbol{x}^*) := \frac{1}{N}\sum_{i=1}^N \widehat{\psi}(\lambda q, \lambda m x_i^*) - \frac{\lambda m^2}{2} + \frac{\lambda q^2}{4},$$

and similarly define $\bar{F}(\lambda, m, q) = \mathbb{E}_{\boldsymbol{x}^*}\,\widehat{F}(\lambda, m, q, \boldsymbol{x}^*) = \bar{\psi}(\lambda q, \lambda m) - \frac{\lambda m^2}{2} + \frac{\lambda q^2}{4}$.

Proposition 4.2. *There exist $K > 0$ such that for all $\epsilon > 0$, we have*

$$F_N \le \mathbb{E}_{\boldsymbol{x}^*}\Big[\max_{l\in\mathbb{Z}, |l|\le K/\epsilon} \Phi_\epsilon(l\epsilon, \boldsymbol{x}^*)\Big] + \frac{\log(K/\epsilon)}{\sqrt{N}}.$$

Now we upper bound Φ_ϵ in terms of $\widehat{F}$:

Proposition 4.3 (Interpolation upper bound)**.** *There exist $K > 0$ depending on $\lambda \ge 0$ such that for all $m \in \mathbb{R}$ and $\epsilon > 0$ we have*

$$\Phi_\epsilon(m, \boldsymbol{x}^*) \le \inf_{q\ge 0} \widehat{F}(\lambda, m, q, \boldsymbol{x}^*) + \frac{\lambda}{2}\epsilon^2 + \frac{K}{N}.$$

Remark: This simple upper bound on the Franz-Parisi potential—which may be of independent interest—can be straightforwardly generalized to spiked tensor models of even order. Indeed, as will be apparent from the proof in the present matrix case, a crucial step in obtaining the inequality is the positivity of a certain hard-to-control remainder[1] term. Tensor models of even order enjoy a convexity property that ensures the positivity of this remainder.

From Propositions 4.2 and 4.11, an upper bound on F_N in the form of a saddle formula begins to emerge. For a fixed $m \in \mathbb{R}$ let $\bar{q} = \bar{q}(\lambda, m)$ be any minimizer of $q \mapsto \overline{F}(\lambda, m, q)$ on $\mathbb{R}_+$. (By differentiating $\overline{F}$, we can check that $\bar{q}$ is bounded uniformly in m.) Then we have

$$F_N \leq \mathbb{E}_{\boldsymbol{x}^*}\Big[\max_{\substack{m=l\epsilon \\ |l|\leq K/\epsilon}} \widehat{F}(\lambda, m, \bar{q}(\lambda, m), \boldsymbol{x}^*)\Big] + \frac{\lambda}{2}\epsilon^2 + \frac{\log(K/\epsilon)}{\sqrt{N}}. \tag{4.8}$$

At this point we need to push the expectation inside the supremum. This will be done using a concentration argument.

Lemma 4.4. *There exists $K > 0$ such that for all $\lambda \geq 0$, $m \in \mathbb{R}$, $q \geq 0$ and $t \geq 0$,*

$$\Pr_{\boldsymbol{x}^*}\left(\left|\widehat{F}(\lambda, m, q, \boldsymbol{x}^*) - \overline{F}(\lambda, m, q)\right| \geq t\right) \leq 2e^{-\frac{Nt^2}{\lambda^2 K(|m|+q)^2}}.$$

It is a routine computation to deduce from Lemma 4.4 (and boundedness of both m and q) that the expected supremum is bounded by the supremum of the expectation plus a small entropy term (the full details of a similar argument are given in the proof of Proposition 4.2):

$$\mathbb{E} \sup_{\substack{m=l\epsilon \\ |l|\leq K/\epsilon}} \widehat{F}(\lambda, m, \bar{q}(\lambda, m), \boldsymbol{x}^*) \leq \sup_{\substack{m=l\epsilon \\ |l|\leq K/\epsilon}} \overline{F}(\lambda, m, \bar{q}(\lambda, m)) + \frac{K\log(K/\epsilon)}{\sqrt{N}}.$$

Since $\bar{q}$ is a minimizer of $\overline{F}$, it follows from (4.8) that

$$F_N \leq \sup_{m\in\mathbb{R}} \inf_{q\geq 0} \overline{F}(\lambda, m, q) + \frac{\lambda}{2}\epsilon^2 + \frac{K\log(K/\epsilon)}{\sqrt{N}}. \tag{4.9}$$

We now let $\epsilon = N^{-1/4}$, and conclude by noticing that the above saddle formula is another expression for ϕ_{RS}:

Proposition 4.5.

$$\phi_{\mathsf{RS}}(\lambda) = \sup_{m\in\mathbb{R}} \inf_{q\geq 0} \overline{F}(\lambda, m, q).$$

[1] We note that the adaptive interpolation method of Barbier and Macris Barbier and Macris (2017) is able to bypass this issue of positivity of the remainder term along the interpolation path, as long as this interpolation "stays on the Nishimori line", i.e., the partition function must correspond to an inference problem for every t (this is however not true in the case of the FP potential.) They are thus able to compute the free entropy of (asymmetric) spiked tensor models of odd order. See Barbier, Krzakala, et al. (2017); Barbier, Macris, and Miolane (2017).

Proof. One inequality follows from (4.9) and the lower bound $F_N \geq \phi_{\mathsf{RS}}(\lambda) - o_N(1)$. For the converse inequality, we notice that for all $m \in \mathbb{R}$

$$\inf_{q\geq 0} \bar{F}(\lambda, m, q) \leq \bar{F}(\lambda, m, |m|) = \bar{\psi}(\lambda|m|, \lambda m) - \frac{\lambda}{4}|m|^2.$$

Now we use Lemma 2.13 which says that the function $\bar{\psi}$ is largest when its second arguments is positive: for all $r \geq 0$, $\bar{\psi}(r, -r) \leq \bar{\psi}(r, r)$. This implies

$$\inf_{q\geq 0} \bar{F}(\lambda, m, q) \leq F(\lambda, |m|) - \frac{\lambda}{4}|m|^2.$$

Taking the supremum over m yields the converse bound. ■

***Proof** of Proposition 4.2.* Let $\epsilon > 0$. Since the prior P_{x} has bounded support, we can grid the set of the overlap values $R_{1,*}$ by $2K/\epsilon$ many intervals of size ϵ for some $K > 0$. This allows the following discretization, where l runs over the finite range $\{-K/\epsilon, \cdots, K/\epsilon - 1\}$:

$$\begin{aligned} F_N &= \frac{1}{N}\,\mathbb{E}\log\sum_l \int \mathbb{1}\{R_{1,*} \in [l\epsilon, (l+1)\epsilon)\} e^{-H(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) \\ &\leq \frac{1}{N}\,\mathbb{E}\log\frac{2K}{\epsilon}\max_l \int \mathbb{1}\{R_{1,*} \in [l\epsilon, (l+1)\epsilon)\} e^{-H(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) \\ &= \frac{1}{N}\,\mathbb{E}\max_l \log \int \mathbb{1}\{R_{1,*} \in [l\epsilon, (l+1)\epsilon)\} e^{-H(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) + \frac{\log(2K/\epsilon)}{N}. \end{aligned} \tag{4.10}$$

In the above, $\mathbb{E}$ is w.r.t. both $\boldsymbol{W}$ and $\boldsymbol{x}^*$. We use concentration of measure to push the expectation over $\boldsymbol{W}$ to the left of the maximum. Let

$$Z_l := \int \mathbb{1}\{R_{1,*} \in [l\epsilon, (l+1)\epsilon)\} e^{-H(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}).$$

We show that each term $X_l = \frac{1}{N}\log Z_l$ concentrates about its expectation (in the randomness of $\boldsymbol{W}$). Let $\mathbb{E}'$ denote the expectation w.r.t. $\boldsymbol{W}$.

Lemma 4.6. *There exists a constant $K > 0$ such that for all $\gamma \geq 0$ and all l,*

$$\mathbb{E}'\, e^{\gamma(X_l - \mathbb{E}'[X_l])} \leq \frac{K\gamma}{\sqrt{N}} e^{K\gamma^2/N}.$$

Therefore, the expectation of the maximum concentrates as well:

$$\begin{aligned} \mathbb{E}'\max_l(X_l - \mathbb{E}'[X_l]) &\leq \frac{1}{\gamma}\log\mathbb{E}'\exp\Big(\gamma\max_l(X_l - \mathbb{E}'[X_l])\Big) \\ &= \frac{1}{\gamma}\log\mathbb{E}'\max_l e^{\gamma(X_l - \mathbb{E}'[X_l])} \end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{\gamma}\log \mathbb{E}' \sum_l e^{\gamma(X_l-\mathbb{E}'[X_l])}\\
&\leq \frac{1}{\gamma}\log\left(\frac{2K}{\epsilon}\frac{\gamma K}{\sqrt{N}}e^{\gamma^2 K/N}\right)\\
&= \frac{\log(2K/\epsilon)}{\gamma} + \frac{1}{\gamma}\log\frac{\gamma K}{\sqrt{N}} + \frac{\gamma K}{N}.
\end{aligned}$$

We set $\gamma = \sqrt{N}$ and obtain

$$\mathbb{E}'\max_l (X_l - \mathbb{E}'[X_l]) \leq \frac{\log(K/\epsilon)}{\sqrt{N}}.$$

Therefore, plugging the above estimates into (4.10), we obtain

$$\begin{aligned}
F_N &\leq \mathbb{E}_{\boldsymbol{x}^*}\max_l \mathbb{E}'\, X_l + \frac{\log(K/\epsilon)}{\sqrt{N}} + \frac{\log(K/\epsilon)}{N}\\
&\leq \mathbb{E}_{\boldsymbol{x}^*}\max_l \Phi_\epsilon(l\epsilon,\boldsymbol{x}^*) + 2\frac{\log(K/\epsilon)}{\sqrt{N}}.
\end{aligned}$$

■

***Proof** of Proposition 4.11.* Let $t\in[0,1]$ and consider a slightly modified interpolating Hamiltonian that has two parameters $r=\lambda q\geq 0$ and $s=\lambda m\in\mathbb{R}$:

$$\begin{aligned}
-H_t(\boldsymbol{x}) := &\sum_{i<j}\sqrt{\frac{t\lambda}{N}}W_{ij}x_ix_j + \frac{t\lambda}{N}x_ix_i^*x_jx_j^* - \frac{t\lambda}{2N}x_i^2x_j^2 \qquad (4.11)\\
&+\sum_{i=1}^N \sqrt{(1-t)r}z_ix_i + (1-t)sx_ix_i^* - \frac{(1-t)r}{2}x_i^2,
\end{aligned}$$

where the z_i's are i.i.d. standard Gaussian r.v.'s independent of everything else. Let

$$\varphi(t) := \frac{1}{N}\mathbb{E}\log\int \mathbb{1}\{R_{1,*}\in[m,m+\epsilon)\}e^{-H_t(\boldsymbol{x})}\mathrm{d}P_{\mathbf{x}}^{\otimes N}(\boldsymbol{x}),$$

where $\mathbb{E}$ is over the Gaussian disorder $\boldsymbol{W}$ and $\boldsymbol{z}$ ($\boldsymbol{x}^*$ is fixed). Let $\langle\cdot\rangle_t$ be the corresponding Gibbs average, similarly to (4.15). By differentiation and Gaussian integration by parts,

$$\begin{aligned}
\varphi'(t) = &-\frac{\lambda}{4}\mathbb{E}\left\langle (R_{1,2}-q)^2\right\rangle_t + \frac{\lambda}{4}q^2 + \frac{\lambda}{4N^2}\sum_{i=1}^N \mathbb{E}\left\langle (x_i^{(1)}x_i^{(2)})^2\right\rangle_t\\
&+\frac{\lambda}{2}\mathbb{E}\left\langle (R_{1,*}-m)^2\right\rangle_t - \frac{\lambda}{2}m^2 - \frac{\lambda}{2N^2}\sum_{i=1}^N \mathbb{E}\left\langle (x_ix_i^*)^2\right\rangle_t,
\end{aligned}$$

Notice that by the overlap restriction, $\mathbb{E}\left\langle (R_{1,*}-m)^2\right\rangle_t \le \epsilon^2$. Moreover, the last terms in the first and second lines in the above are of order $1/N$ since the variables x_i are bounded. Next, since $\mathbb{E}\left\langle (R_{1,2}-q)^2\right\rangle_t$ has non-negative sign, we can ignore it and obtain an upper bound:

$$\varphi'(t) \le -\frac{\lambda}{2}m^2 + \frac{\lambda}{4}q^2 + \frac{\lambda}{2}\epsilon^2 + \frac{K}{N}.$$

Integrating over t, we obtain

$$\Phi_\epsilon(m, \boldsymbol{x}^*) \le -\frac{\lambda}{2}m^2 + \frac{\lambda}{4}q^2 + \frac{\lambda}{2}\epsilon^2 + \varphi(0) + \frac{K}{N}.$$

Now we use a trivial upper bound on $\varphi(0)$:

$$\begin{aligned}\varphi(0) &= \frac{1}{N}\,\mathbb{E}\log\int \mathbb{1}\{R_{1,*}\in[m, m+\epsilon)\}e^{-H_0(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})\\ &\le \frac{1}{N}\,\mathbb{E}\log\int e^{-H_0(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})\\ &= \frac{1}{N}\sum_{i=1}^{N}\widehat{\psi}(\lambda q, \lambda m x_i^*).\end{aligned}$$

Hence,

$$\Phi_\epsilon(m, \boldsymbol{x}^*) \le \widehat{F}(\lambda, m, q, \boldsymbol{x}^*) + \frac{\lambda}{2}\epsilon^2 + \frac{K}{N}.$$

■

***Proof** of lemma 4.4.* The random part of $\widehat{F}(\lambda, m, \bar{q}, \boldsymbol{x}^*)$ is the average of i.i.d. terms $\widehat{\psi}(\lambda q, \lambda m x_i^*)$. Since $\left|\partial_s\widehat{\psi}(r, sx^*)\right| \le K^2$, $\left|\partial_r\widehat{\psi}(r, sx^*)\right| \le K^2/2$ and $\widehat{\psi}(0,0) = 0$, where K is a bound on the support of P_{x}, we have $\left|\widehat{\psi}(r, sx^*)\right| \le K^2(r/2 + |s|)$. For bounded r and s, the claim follows from concentration of the average of i.i.d. bounded r.v.'s. ■

***Proof** of Lemma 4.6.* We notice that X_l seen as a function of $\boldsymbol{W}$ is Lipschitz with constant $K\sqrt{\frac{\lambda}{N}}$. By Gaussian concentration of Lipschitz functions, there exist a constant K depending only on λ such that for all $t \ge 0$,

$$\Pr\left(X_l - \mathbb{E}'\,X_l \ge t\right) \le e^{-Nt^2/K}.$$

Then we conclude by means of the identity

$$\mathbb{E}'\,e^{\gamma(X_l - \mathbb{E}'[X_l])} = \gamma\int_{-\infty}^{+\infty}\Pr(X_l - \mathbb{E}'[X_l] \ge t)\;e^{\gamma t}\mathrm{d}t,$$

and integrate the tail. ■

4.2 Finite size corrections to the RS formula

In this section we determine the finite-size correction of F_N to its limit $\phi_{\mathsf{RS}}(\lambda)$, i.e., the constant-order term in the asymptotic expansion of $\mathbb{E}_{\mathbb{P}_\lambda} \log L(\boldsymbol{Y};\lambda)$ in N, for large N. Recall from Chapter 2 that uniqueness of q^* (the maximizer of the RS potential F) only needs first differentiability of the RS formula. In contrast, the results we are about to present hold under a slightly stronger condition: we will need a second derivative to exist. In the physics parlance, our results do not hold at values of λ at which a particular kind of *first-order phase transitions* occur, namely, those in which the order parameter q^* is not differentiable. The presence of these transitions depends on the prior P_{x}. For the Gaussian and Rademacher prior, there are no such transitions, while for the sparse Rademacher prior $P_{\mathsf{x}} = \frac{\rho}{2}\delta_{-1/\sqrt{\rho}} + (1-\rho)\delta_0 + \frac{\rho}{2}\delta_{+1/\sqrt{\rho}}$, there is *one* first-order transition where ${q^*}'$ is not defined for every $\rho < \rho^* \approx 0.092$. (See Chapter 2.) Thus we define the set

$$\mathcal{A} = \big\{\lambda > 0 \;:\; \phi_{\mathsf{RS}} \text{ is twice differentiable at } \lambda.\big\}.$$

Since ϕ_{RS} is the point-wise limit of a sequence (F_N) of convex functions, it is also convex. Then by Alexandrov's theorem (Aleskandrov, 1939), the set $\mathcal{A}$ is of full Lebesgue measure in $\mathbb{R}_+$. Moreover, we can see that $(0,\lambda_c) \subset \mathcal{A}$, since if $\lambda \in \mathcal{A} \cap (0,\lambda_c)$, we have $q^*(\lambda) = 0$, therefore $\phi_{\mathsf{RS}}(\lambda) = 0$. By continuity, ϕ_{RS} vanishes on the entire interval $(0,\lambda_c)$. Our first main result is to establish the existence of a function $\lambda \mapsto \psi_{\mathsf{RS}}(\lambda)$ defined on $\mathcal{A}$ such that either below λ_c or above it when the prior P_{x} is not symmetric about the origin, we have

$$N(F_N - \phi_{\mathsf{RS}}(\lambda)) \longrightarrow \psi_{\mathsf{RS}}(\lambda).$$

An explicit formula for ψ_{RS} will be given. But first we need to introduce some notation. Let $\lambda \in \mathcal{A}$ and consider the quantities

$$a(0) = \mathbb{E}\left[\langle x^2\rangle_r^2\right] - q^{*2}(\lambda), \quad a(1) = \mathbb{E}\left[\langle x^2\rangle_r \langle x\rangle_r^2\right] - q^{*2}(\lambda), \quad a(2) = \mathbb{E}\left[\langle x\rangle_r^4\right] - q^{*2}(\lambda), \tag{4.12}$$

where

$$\langle\cdot\rangle_r = \frac{\int \cdot \exp\left(\sqrt{r}zx + rxx^* - \frac{r}{2}x^2\right) \mathrm{d}P_{\mathsf{x}}(x)}{\int \exp\left(\sqrt{r}zx + rxx^* - \frac{r}{2}x^2\right) \mathrm{d}P_{\mathsf{x}}(x)},$$

with $r = \lambda q^*(\lambda)$ and the expectation operator $\mathbb{E}$ is w.r.t. $x^* \sim P_{\mathsf{x}}$ and $z \sim \mathcal{N}(0,1)$. The Gibbs measure $\langle\cdot\rangle_r$ can be interpreted as the posterior distribution of x^* given the observation $y = \sqrt{r}x^* + z$. (More on this point of view in Section 4.3.) Now let

$$\begin{aligned} \mu_1(\lambda) &= \lambda(a(0) - 2a(1) + a(2)), \\ \mu_2(\lambda) &= \lambda(a(0) - 3a(1) + 2a(2)), \end{aligned} \tag{4.13}$$

and finally define

$$\psi_{\mathsf{RS}}(\lambda) := \frac{1}{4}\left(\log(1-\mu_1) - 2\log(1-\mu_2) + \lambda\frac{4a(1) - 3a(2)}{1-\mu_1} - \lambda a(0)\right). \tag{4.14}$$

We will prove (Lemma 4.21) that $\mu_2 \le \mu_1 < 1$ for all $\lambda \in \mathcal{A}$ so that this function is well defined on $\mathcal{A}$.

Theorem 4.7. *For $\lambda \in \mathcal{A}$, if either $\lambda < \lambda_c$, or $\lambda > \lambda_c$ and the prior P_{x} is not symmetric about the origin, then*

$$N\big(F_N - \phi_{\mathsf{RS}}(\lambda)\big) = \psi_{\mathsf{RS}}(\lambda) + \mathcal{O}\Big(\frac{1}{\sqrt{N}}\Big),$$

or equivalently, $D_{\mathsf{KL}}(\mathbb{P}_\lambda, \mathbb{P}_0) = N\phi_{\mathsf{RS}}(\lambda) + \psi_{\mathsf{RS}}(\lambda) + \mathcal{O}(1/\sqrt{N})$.

The theorem asserts that either below the reconstruction threshold, or above it when the prior P_{x} is not symmetric, the free energy F_N has a finite-size correction of order $1/N$ to its limit ϕ_{RS} and a subsequent term of order $N^{-3/2}$ in the expansion. In the case $\lambda > \lambda_c$ with symmetric prior, the problem is invariant under a sign flip of the spike, so the overlap $\boldsymbol{x}^\top \boldsymbol{x}^*/N$ has a symmetric distribution, and hence concentrates equiprobably about *two distinct* values $\pm q^*(\lambda)$. Our techniques do not survive this symmetry, and resolving this case seems to require a new approach.

We see that $D_{\mathsf{KL}}(\mathbb{P}_\lambda, \mathbb{P}_0)$ is an extensive quantity in N whenever $\phi_{\mathsf{RS}}(\lambda) > 0$, or equivalently, $\lambda > \lambda_c$. On the other hand, this KL is of constant order below λ_c:

Centered prior. Let us consider the case where the prior P_{x} has zero mean, and unit variance (the latter can be assumed without loss of generality by rescaling λ), so that Lemma 2.1 reads $\lambda_c \le 1$. If $\lambda < \lambda_c$, we have $q^*(\lambda) = 0$, $\phi_{\mathsf{RS}}(\lambda) = 0$, and one can check that in this case

$$a(0) = (\mathbb{E}_{P_{\mathsf{x}}}[X^2])^2 = 1, \quad a(1) = \mathbb{E}_{P_{\mathsf{x}}}[X^2]\,\mathbb{E}_{P_{\mathsf{x}}}[X]^2 = 0, \quad a(2) = \mathbb{E}_{P_{\mathsf{x}}}[X]^4 = 0.$$

Therefore, expression (4.14) simplifies to

$$\psi_{\mathsf{RS}}(\lambda) = \frac{1}{4}\left(-\log\left(1-\lambda\right) - \lambda\right),$$

and Theorem 4.7 reduces the formula of Proposition 2.5.

More information on ψ_{RS}. Expression (4.14) looks mysterious at first sight. Let us briefly explain its origin. A slightly less processed expression for ψ_{RS} is the following

$$\psi_{\mathsf{RS}}(\lambda) = \frac{1}{4}\int_0^1 \left(-\frac{\mu_1}{1-t\mu_1} + \frac{2\mu_2}{1-t\mu_2} + \lambda\frac{4a(1)-3a(2)}{(1-t\mu_1)^2}\right)\mathrm{d}t - \frac{\lambda}{4}a(0),$$

after which (4.14) follows by simple integration. The integrand in the above expression is obtained, as we will show, as the first entry $z(0)$ of the solution $\boldsymbol{z} = [z(0), z(1), z(2)]^\top$ of the 3×3 linear system

$$(\boldsymbol{I} - t\boldsymbol{A})\boldsymbol{z} = \boldsymbol{a},$$

where $\boldsymbol{a} = [a(0), a(1), a(2)]^\top$ and $\boldsymbol{A}$ is the "cavity" matrix

$$\boldsymbol{A} := \lambda \cdot \begin{bmatrix} a(0) & -2a(1) & a(2) \\ a(1) & a(0) - a(1) - 2a(2) & -2a(1) + 3a(2) \\ a(2) & 4a(1) - 6a(2) & a(0) - 6a(1) + 6a(2) \end{bmatrix}.$$

The above matrix happens to have two eigenvalues which are exactly μ_1 and μ_2. The matrix $\boldsymbol{A}$ and the above linear system will emerge naturally as a result of the cavity method. On the other hand, the integral over the time parameter t is along Guerra's interpolation path (see also Guerra and F. Toninelli, 2002c), and the integrand can be interpreted as the asymptotic variance in a central limit theorem satisfied by the overlap between two replicas under the law induced by a certain interpolating Gibbs measure. A definition of these notions with the corresponding results can be found in Sections 4.3 and 4.4. The full execution of the cavity method is relegated to Section 4.5.

4.3 Overlap convergence: optimal rates

We recall a couple of definitions from previous chapters and sections. The posterior of $\boldsymbol{x}^*$ given $\boldsymbol{Y}$ is

$$\mathrm{d}\,\mathbb{P}_\lambda(\boldsymbol{x}|\boldsymbol{Y}) = \frac{e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})}{\int e^{-H(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})},$$

where H is the Hamiltonian

$$\begin{aligned} -H(\boldsymbol{x}) &:= \sum_{i<j} \sqrt{\frac{\lambda}{N}} Y_{ij} x_i x_j - \frac{\lambda}{2N} x_i^2 x_j^2 \\ &= \sum_{i<j} \sqrt{\frac{\lambda}{N}} W_{ij} x_i x_j + \frac{\lambda}{N} x_i x_i^* x_j x_j^* - \frac{\lambda}{2N} x_i^2 x_j^2. \end{aligned}$$

For an integer $n \geq 1$ and $f : (\mathbb{R}^N)^{n+1} \mapsto \mathbb{R}$, we define the Gibbs average of f w.r.t. H as

$$\big\langle f(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}, \boldsymbol{x}^*) \big\rangle := \frac{\int f(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}, \boldsymbol{x}^*) \prod_{l=1}^n e^{-H(\boldsymbol{x}^{(l)})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}^{(l)})}{\int \prod_{l=1}^n e^{-H(\boldsymbol{x}^{(l)})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}^{(l)})}. \tag{4.15}$$

Recall also the definition of the overlap between replicas: for $l, l' = 1, \cdots, n, *$, we let

$$R_{l,l'} := \boldsymbol{x}^{(l)} \cdot \boldsymbol{x}^{(l')} = \frac{1}{N} \sum_{i=1}^N x_i^{(l)} x_i^{(l')}.$$

In this section we show the convergence of the first four moments of the overlap at optimal rates under some conditions: if either the prior P_{x} is not symmetric about the origin or the Hamiltonian H is "perturbed" in the following sense. Let $t \in [0,1]$ and consider the interpolating Hamiltonian

$$\begin{aligned} -H_t(\boldsymbol{x}) := &-\frac{t\lambda}{2N}\sum_{i<j} x_i^2 x_j^2 + \sqrt{\frac{t\lambda}{N}}\sum_{i<j} W_{ij}x_i x_j + \frac{t\lambda}{N}\sum_{i<j} x_i x_i^* x_j x_j^* \\ &- \frac{(1-t)r}{2}\sum_{i=1}^N x_i^2 + \sqrt{(1-t)r}\sum_{i=1}^N z_i x_i + (1-t)r\sum_{i=1}^N x_i x_i^*, \end{aligned} \tag{4.16}$$

where the z_i's are i.i.d. standard Gaussian r.v.'s independent of everything else, and $r = \lambda q^*(\lambda)$. We similarly define the Gibbs average $\langle\cdot\rangle_t$ as in (4.15) where H is replaced by H_t. We now state a fundamental property satisfied by both $\langle\cdot\rangle$ and $\langle\cdot\rangle_t$.

The Nishimori property. As explained in Chapter 2, the fact that the Gibbs measure $\langle\cdot\rangle$ is a posterior distribution (4.2) has important consequences. The $n+1$-tuples $(\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(n+1)})$ and $(\boldsymbol{x}^{(1)},\cdots,\boldsymbol{x}^{(n)},\boldsymbol{x}^*)$ have the same law under $\mathbb{E}\langle\cdot\rangle$. Moreover, this property is preserved under the interpolating Gibbs measure $\langle\cdot\rangle_t$ for all $t \in [0,1]$. Indeed, the interpolation is constructed in such a way that $\langle\cdot\rangle_t$ is the posterior distribution of the signal $\boldsymbol{x}^*$ given the augmented set of observations

$$\begin{cases} Y_{ij} &= \sqrt{\frac{t\lambda}{N}}x_i^* x_j^* + W_{ij}, \quad 1 \le i < j \le N, \\ y_i &= \sqrt{(1-t)r}x_i^* + z_i, \quad 1 \le i \le N, \end{cases} \tag{4.17}$$

where one receives side information about $\boldsymbol{x}^*$ from a scalar Gaussian channel, $r = \lambda q^*(\lambda)$, and the signal-to-noise ratios of the two channels are altered in a time dependent way. Now we state our concentration result.

Theorem 4.8. *For all $\lambda \in \mathcal{A}$ and all $t \in [0,1]$, there exist constants $K(\lambda) \ge 0$ and $c(t) \ge 0$ such that*

$$\mathbb{E}\left\langle (R_{1,*} - q^*)^4 \right\rangle_t \le K(\lambda)\Big(\frac{1}{N^2} + e^{-c(t)N}\Big). \tag{4.18}$$

Moreover, $c(t) > 0$ on $[0,1)$, and if either $\lambda < \lambda_c$ or P_{x} is not symmetric about the origin, then $c(t) \ge c_0$ for some constant $c_0 = c_0(\lambda) > 0$. Otherwise, $c(t) \sim c_0(1-t)^2$ as $t \to 1$.

If P_{x} is symmetric about the origin then the distribution of $R_{1,*}$ under $\mathbb{E}\langle\cdot\rangle$ is also symmetric, so $\mathbb{E}\langle R_{1,*}\rangle = 0$. If moreover $q^*(\lambda) > 0$ (i.e., $\lambda > \lambda_c$) then (4.18) becomes trivial at $t = 1$ since both sides are constant. On the other hand, if either $t < 1$ or P_{x} is asymmetric, the sign symmetry of the spike is broken. This forces the overlap to be positive and hence concentrate about $q^*(\lambda)$. Finally, if $\lambda < \lambda_c$, $q^*(\lambda) = 0$ and the sign symmetry becomes irrelevant since the overlap converges to zero regardless. Let us mention that in the symmetric unperturbed case ($t = 1$), we expect a variant of (4.18) to hold where $R_{1,*}$ is replaced by its absolute value in the statement, and the upper bound would be K/N^2. Unfortunately, our methods do not allow us to prove such a statement, but we are able to prove a weaker result (see Lemma 4.12): for all $\epsilon > 0$,

$$\mathbb{E}\left\langle \mathbb{1}\left\{\big||R_{1,*}| - q^*\big| \geq \epsilon\right\}\right\rangle \longrightarrow 0. \tag{4.19}$$

Although this a minor technical point, we also point out that the estimate $c(t) \sim c_0(1-t)^2$ in the statement is suboptimal. A heuristic argument allows us to get $c(t) \sim c_0(1-t)$ as $t \to 1$, but we are currently unable to rigorously justify it.

MMSE. The bound (4.18) can be used to deduce the optimal error of estimating $\boldsymbol{x}^*$ based on the observations (4.17). The posterior mean $\langle \boldsymbol{x} \rangle_t$ is the estimator with Minimal Mean Squared Error (MMSE) among all estimators $\widehat{\theta}(\boldsymbol{Y}, \boldsymbol{y}) \in \mathbb{R}^N$, and the MMSE is

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\left[(x_i^* - \langle x_i \rangle_t)^2\right] &= \mathbb{E}_{P_{\mathrm{x}}}[X^2] - \frac{2}{N}\sum_{i=1}^{N}\mathbb{E}\langle x_i x_i^* \rangle_t + \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\langle x_i \rangle_t^2 \\ &= \mathbb{E}_{P_{\mathrm{x}}}[X^2] - \mathbb{E}\langle R_{1,*} \rangle_t.\end{aligned}$$

The last line follows from the Nishimori property, since $\mathbb{E}\langle x \rangle_t^2 = \mathbb{E}\langle x^{(1)} x^{(2)} \rangle_t = \mathbb{E}\langle x x^* \rangle_t$. Theorem 4.8 implies in particular (under the conditions of its validity) that $\mathbb{E}\langle R_{1,*} \rangle_t \to q^*(\lambda)$, yielding the value of the MMSE. It is in particular possible to estimate the spike $\boldsymbol{x}^*$ from the observations (4.17) with non-trivial accuracy if and only if $\lambda > \lambda_c$. Note that at $t = 1$ (no side information) the result still holds below λ_c or when the prior is not symmetric. Otherwise, as mentioned before, the problem is invariant under a sign flip of $\boldsymbol{x}^*$ so one has to change the measure of performance. Beside the result (4.19), we are unable to say much in this situation.

Asymptotic variance. By Jensen's inequality we deduce from (4.18) the convergence of the second moment:

$$\mathbb{E}\left\langle (R_{1,*} - q^*)^2 \right\rangle_t \leq K(\lambda)\Big(\frac{1}{N} + e^{-c(t)N}\Big). \tag{4.20}$$

To establish our finite-size correction result (Theorem 4.7) we need to prove a result stronger than (4.20), namely that $N \cdot \mathbb{E}\left\langle (R_{1,*} - q^*)^2 \right\rangle_t$ converges to a limit. For $t \in [0,1]$ and $\lambda \in \mathcal{A}$, we let

$$\Delta_{\mathsf{RS}}(\lambda; t) := \frac{1}{\lambda}\left(-\frac{\mu_1}{1 - t\mu_1} + \frac{2\mu_2}{1 - t\mu_2} + \lambda\frac{4a(1) - 3a(2)}{(1 - t\mu_1)^2}\right), \tag{4.21}$$

where μ_1 and μ_2 are defined in (4.13).

Theorem 4.9. *For all $\lambda \in \mathcal{A}$ and all $t \in [0,1]$, there exist constants $K(\lambda) \geq 0$ and $c(t) \geq 0$ such that*

$$\left|N \cdot \mathbb{E}\left\langle (R_{1,*} - q^*)^2 \right\rangle_t - \Delta_{\mathsf{RS}}(\lambda; t)\right| \leq K(\lambda)\left(\frac{1}{\sqrt{N}} + Ne^{-c(t)N}\right).$$

Moreover, $c(t) > 0$ on $[0,1)$, and if either $\lambda < \lambda_c$ or P_{x} is not symmetric about the origin, then $c(t) \geq c_0$ for some constant $c_0 = c_0(\lambda) > 0$. Otherwise, $c(t) \sim c_0(1-t)^2$ as $t \to 1$.

The proofs of Theorems 4.8 and 4.9 rely on the cavity method, and will be presented in Section 4.5. Finally, the techniques we use could be easily extended to prove convergence of all the moments at optimal rates: for all integers k,

$$\mathbb{E}\left\langle (R_{1,*} - q^*)^{2k} \right\rangle_t \leq \frac{K(k)}{N^k} + K(k)e^{-c(k,t)N},$$

but we will not need this stronger statement.

4.4 The interpolation method

In this section we apply once more the interpolation method of Guerra (2001) to prove Theorem 4.7. This time we keep track of the lower order terms.

The Guerra interpolation

Our interpolating Hamiltonian is H_t from (4.16) with $r = \lambda q$ for some $q \geq 0$. Now we consider the interpolating free energy

$$\varphi(t) := \frac{1}{N}\,\mathbb{E}\log\int e^{-H_t(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}). \tag{4.22}$$

We see that $\varphi(1) = F_N$ and $\varphi(0) = \psi(\lambda q)$. This function is moreover differentiable in t, and by differentiation, we have

$$\begin{aligned}
\varphi'(t) &= \frac{1}{N}\,\mathbb{E}\left\langle -\frac{\mathrm{d}H_t(\boldsymbol{x})}{\mathrm{d}t} \right\rangle_t \\
&= \frac{1}{N}\,\mathbb{E}\left\langle -\frac{\lambda}{2N}\sum_{i<j} x_i^2 x_j^2 + \frac{1}{2}\sqrt{\frac{\lambda}{tN}}\sum_{i<j} W_{ij}x_i x_j + \frac{\lambda}{N}\sum_{i<j} x_i x_i^* x_j x_j^* \right\rangle_t \\
&\quad + \frac{1}{N}\,\mathbb{E}\left\langle \frac{\lambda q}{2}\sum_{i=1}^N x_i^2 - \frac{1}{2}\sqrt{\frac{\lambda q}{1-t}}\sum_{i=1}^N z_i x_i - \lambda q \sum_{i=1}^N x_i x_i^* \right\rangle_t .
\end{aligned}$$

Now we use Gaussian integration by parts to eliminate the variables W_{ij} and z_i. The details of this computation are explained extensively in many sources. See (Krzakala, Xu, and Zdeborová, 2016; Lelarge and Miolane, 2016; Talagrand, 2011a). We get

$$\begin{aligned}
\varphi'(t) = &-\frac{\lambda}{2N^2}\,\mathbb{E}\left\langle \sum_{i<j} x_i^{(1)}x_j^{(1)}x_i^{(2)}x_j^{(2)} \right\rangle_t + \frac{\lambda}{N^2}\,\mathbb{E}\left\langle \sum_{i<j} x_i x_i^* x_j x_j^* \right\rangle_t \\
&+ \frac{\lambda q}{2N}\,\mathbb{E}\left\langle \sum_{i=1}^N x_i^{(1)}x_i^{(2)} \right\rangle_t - \frac{\lambda q}{N}\,\mathbb{E}\left\langle \sum_{i=1}^N x_i x_i^* \right\rangle_t .
\end{aligned}$$

Completing the squares yields

$$\varphi'(t) = -\frac{\lambda}{4}\,\mathbb{E}\left\langle(\boldsymbol{x}^{(1)}\cdot\boldsymbol{x}^{(2)}-q)^2\right\rangle_t + \frac{\lambda}{4}q^2 + \frac{\lambda}{4N^2}\sum_{i=1}^{N}\mathbb{E}\left\langle {x_i^{(1)}}^2 {x_i^{(2)}}^2\right\rangle_t \tag{4.23}$$
$$+\frac{\lambda}{2}\,\mathbb{E}\left\langle(\boldsymbol{x}\cdot\boldsymbol{x}^{*}-q)^2\right\rangle_t - \frac{\lambda}{2}q^2 - \frac{\lambda}{2N^2}\sum_{i=1}^{N}\mathbb{E}\left\langle {x_i}^2 {x_i^{*}}^2\right\rangle_t .$$

The first line in the above expression involves overlaps between two independent replicas, while the second one involves overlaps between one replica and the planted solution. Using the Nishimori property, the derivative of φ can be written as

$$\varphi'(t) = \frac{\lambda}{4}\,\mathbb{E}\left\langle(R_{1,*}-q)^2\right\rangle_t - \frac{\lambda}{4}q^2 - \frac{\lambda}{4N}\,\mathbb{E}\left\langle {x_N}^2 {x_N^{*}}^2\right\rangle_t . \tag{4.24}$$

The last term follows by symmetry between sites. Now, integrating over t, the difference between the free energy and the RS potential $F(\lambda,q)$ can be written in the form of a sum rule:

$$F_N - F(\lambda,q) = \frac{\lambda}{4}\int_0^1\Big(\mathbb{E}\left\langle(R_{1,*}-q)^2\right\rangle_t - \frac{1}{N}\,\mathbb{E}\left\langle {x_N}^2 {x_N^{*}}^2\right\rangle_t\Big)\mathrm{d}t. \tag{4.25}$$

We see from (4.25) that F_N converges to $F(\lambda,q)$ if and only if the overlap $R_{1,*}$ concentrates about q. This happens only for a value of q that maximizes the RS potential $F(\lambda,\cdot)$. Using Theorem 4.8 one can already prove the $1/N$ optimal rate below λ_c or above it when the prior is not symmetric. Indeed since $c(t)$ is lower-bounded by a positive constant in this case, the bound (4.20) yields $\int_0^1\mathbb{E}\langle(R_{1,*}-q^*)^2\rangle_t\mathrm{d}t \le K(\lambda)/N$. Also, the second integrand in (4.25) is bounded by K/N for some constant $K\ge 0$, so we have for all $\lambda\in\mathcal{A}$, $F_N = \phi_{\mathsf{RS}}(\lambda)+\mathcal{O}(1/N)$. If $\lambda>\lambda_c$ and the prior is symmetric then we are only able to prove a rate of $1/\sqrt{N}$ due to the fact $c(t)\sim c_0(1-t)^2$ as $t\to 1$. The $1/N$ rate would follow immediately in this case if one is able to improve the latter estimate to $c(t)\sim c_0(1-t)$. To go further, we use Theorem 4.9, and the additional fact that $\mathbb{E}\langle {x_N}^2 {x_N^{*}}^2\rangle_t$ has a limit:

Lemma 4.10. *For all $\lambda\in\mathcal{A}$ and for all $t\in[0,1)$, there exist constants $K(\lambda)\ge 0$ and $c(t)\ge 0$ such that*

$$\left|\mathbb{E}\left\langle {x_N}^2 {x_N^{*}}^2\right\rangle_t - a(0)\right| \le K(\lambda)\left(\frac{1}{\sqrt{N}} + e^{-c(t)N}\right).$$

Moreover, $c(t)>0$ on $[0,1)$, and if either $\lambda<\lambda_c$ or P_{x} is not symmetric about the origin, then $c(t)\ge c_0$ for some constant $c_0=c_0(\lambda)>0$. Otherwise, $c(t)\sim c_0(1-t)^2$ as $t\to 1$.

The proof of Lemma 4.10 relies on the cavity method, and will be presented in the Section 4.5. Now we are ready to prove Theorem 4.7.

Proof of Theorem 4.7. By formula (4.25) with the choice $q=q^*(\lambda)$, we have

$$\left|N(F_N-\phi_{\mathsf{RS}}(\lambda)) - \frac{\lambda}{4}\Big(\int_0^1\Delta_{\mathsf{RS}}(\lambda;t)\mathrm{d}t - a(0)\Big)\right| \le \frac{\lambda}{4}\int_0^1\left|N\,\mathbb{E}\left\langle(R_{1,*}-q^*)^2\right\rangle_t - \Delta_{\mathsf{RS}}(\lambda;t)\right|\mathrm{d}t$$

$$+ \frac{\lambda}{4}\int_0^1 \left|\mathbb{E}\left\langle {x_N}^2 {x_N^*}^2\right\rangle_t - a(0)\right| \mathrm{d}t.$$

By Theorem 4.9 and Lemma 4.10, the integrands on the right-hand side are bounded by $K/\sqrt{N} + KNe^{-c(t)N}$ where $c(t) > c_0 > 0$ for all t in the cases $\lambda < \lambda_c$ or P_{x} not symmetric about the origin, so the convergence follows. The function $\psi_{\mathsf{RS}}(\lambda)$ is the second term in the left-hand side. Formula (4.14) follows by integration. ■

The main estimate: energy gap at suboptimal overlap

Recall the interpolating Hamiltonian H_t from (4.16) with $r = \lambda q^*(\lambda)$. Let us now consider a version of the Franz-Parisi potential defined on H_t. For $\boldsymbol{x}^* \in \mathbb{R}^N$ fixed, $m \in \mathbb{R} \setminus \{0\}$ and $\epsilon > 0$ define the set

$$A = \begin{cases} R_{1,*} \in [m, m+\epsilon) & \text{if } m > 0,\\ R_{1,*} \in (m-\epsilon, m] & \text{if } m < 0.\end{cases}$$

Now define the FP potential as

$$\Phi_\epsilon(m,t) := \frac{1}{N}\mathbb{E}_{\boldsymbol{W}} \log \int \mathbb{1}\{\boldsymbol{x} \in A\} e^{-H_t(\boldsymbol{x})} \mathrm{d}P_{\mathsf{x}}^{\otimes N}(\boldsymbol{x}),$$

where the expectation is only over the Gaussian disorder $\boldsymbol{W}$. The dependence of Φ_ϵ on $\boldsymbol{x}^*$ is kept implicit. We will adopt the same approach as the ine used in Chapter 2: An upper bound on the FP potential is a main ingredient is our proof of overlap concentration. To prove this we will need the auxiliary function

$$\phi_{\mathsf{RS}}(\lambda;t) = \sup_{q \geq 0}\left\{\psi(\lambda q) - \frac{t\lambda q^2}{4}\right\}.$$

One can show that the above formula is the limit of $\varphi(t)$ as $N \to \infty$, by either using the approach of Section 4.1, or with the argument of Lelarge and Miolane (2016). For our purpose we will only need the inequality

$$\varphi(t) \geq \phi_{\mathsf{RS}}(\lambda;t) - \frac{K\lambda t}{N}, \tag{4.26}$$

which can be proved using the interpolation method presented in the previous section and dropping the non-negative term $\mathbb{E}\langle(R_{1,*} - q^*)^2\rangle$ from the expression analogous to (4.24) in this case. Now it suffices to compare $\Phi_\epsilon(m;t)$ to $\phi_{\mathsf{RS}}(\lambda;t)$. The result is given in Proposition 4.14, and we finish this subsection by Proposition 4.12 showing convergence in probability of the overlaps as a straightforward consequence.

Let us recall some notation from Chapter 2. For $r \geq 0$ and $s \in \mathbb{R}$, we let

$$\widehat{\psi}(r,s) := \mathbb{E}_z \log \int \exp\left(\sqrt{r}zx + sx - \frac{r}{2}x^2\right) \mathrm{d}P_{\mathsf{x}}(x).$$

and

$$\begin{aligned}\bar{\psi}(r,s) &:= \mathbb{E}_{x^*}\,\widehat{\psi}(r,sx^*)\\ &= \mathbb{E}_{x^*,z}\log\int\exp\left(\sqrt{r}zx+sxx^*-\frac{r}{2}x^2\right)\mathrm{d}P_{\mathrm{x}}(x).\end{aligned}$$

We now state a useful interpolation bound on $\Phi_\epsilon(m;t)$. This is a simpler version of the Guerra-Talagrand 1RSB interpolation bound at fixed overlap, a key invention that ultimately paved the way towards a proof of the Parisi formula (Guerra, 2003; Talagrand, 2006). In some sense, since we are dealing with a planted model, we only need a replica-symmetric version of this bound.

Proposition 4.11. *Fix $\boldsymbol{x}^*\in\mathbb{R}^N$, $m\in\mathbb{R}$, $\epsilon>0$, $t\in[0,1]$ and $\lambda\geq 0$. Let $r=(1-t)\lambda q^*+t\lambda|m|$, $\bar{r}=(1-t)\lambda q^*+t\lambda m$. There exist a constant $K=K(P_{\mathrm{x}})>0$ such that*

$$\Phi_\epsilon(m;t)\leq\inf_{h\in\mathbb{R}}\left\{\frac{1}{N}\sum_{i=1}^N\widehat{\psi}\big(r,(\bar{r}+h)x_i^*\big)-hm\right\}-\frac{t\lambda m^2}{4}+\frac{\lambda t\epsilon^2}{2}+\frac{\lambda K}{N}.$$

Proof. This is proved in exactly the same way as Proposition 2.14. The interpolation we consider here is

$$\begin{aligned}-H_{t,s}(\boldsymbol{x}) := \sum_{i<j} -\frac{ts\lambda}{2N}x_i^2x_j^2+\sqrt{\frac{ts\lambda}{N}}W_{ij}x_ix_j+\frac{ts\lambda}{N}x_ix_i^*x_jx_j^*\\ +\sum_{i=1}^N-\frac{(1-t)\lambda q^*}{2}x_i^2+\sqrt{(1-t)\lambda q^*}z_ix_i+(1-t)\lambda q^*x_ix_i^*\\ +\sum_{i=1}^N-\frac{(1-s)t\lambda|m|}{2}x_i^2+\sqrt{(1-s)t\lambda|m|}z_i'x_i+(1-s)t\lambda m x_ix_i^*,\end{aligned}$$

where t is fixed and $s\in[0,1]$ is varying. The r.v.'s W,z,z' are all i.i.d. standard Gaussians independent of everything else. We define

$$\varphi(t,s):=\frac{1}{N}\mathbb{E}_{\boldsymbol{W},\boldsymbol{z},\boldsymbol{z}'}\log\int\mathbb{1}\{\boldsymbol{x}\in A\}e^{-H_{t,s}(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}).$$

We compute the derivative w.r.t. s. The same algebraic manipulations conducted in the computation of φ' up to (4.23) apply here, and we get

$$\begin{aligned}\partial_s\varphi(t,s) = &-\frac{\lambda t}{4}\mathbb{E}\left\langle(\boldsymbol{x}^{(1)}\cdot\boldsymbol{x}^{(2)}-|m|)^2\right\rangle_{t,s}+\frac{\lambda t}{4}|m|^2+\frac{\lambda t}{4N^2}\sum_{i=1}^N\mathbb{E}\left\langle {x_i^{(1)}}^2{x_i^{(2)}}^2\right\rangle_{t,s}\\ &+\frac{\lambda t}{2}\mathbb{E}\left\langle(\boldsymbol{x}\cdot\boldsymbol{x}^*-m)^2\right\rangle_{t,s}-\frac{\lambda t}{2}m^2-\frac{\lambda t}{2N^2}\sum_{i=1}^N\mathbb{E}\left\langle {x_i}^2{x_i^*}^2\right\rangle_{t,s},\end{aligned}$$

where $\langle\cdot\rangle_{t,s}$ is the Gibbs average w.r.t. the Hamiltonian $-H_{t,s}(\boldsymbol{x}) + \log \mathbb{1}\{\boldsymbol{x} \in A\}$. The planted term is trivially smaller than $t\lambda\epsilon^2/2$ due to the overlap restriction. Moreover, the last terms in both lines are of order $1/N$ since the variables x_i are bounded. The first term in the first line, which involves the overlap between two replicas, is always non-positive so we can ignore it and obtain an upper bound:

$$\partial_s \varphi(t,s) \leq -\frac{\lambda t}{4}m^2 + \frac{\lambda t\epsilon^2}{2} + \frac{\lambda K}{N}.$$

Integrating over s, we get

$$\Phi_\epsilon(m;t) \leq \varphi(t,0) - \frac{\lambda t}{4}m^2 + \frac{\lambda t\epsilon^2}{2} + \frac{\lambda K}{N}.$$

Finally, for h of the same sign as m, we have

$$\begin{aligned}\varphi(t,0) &= \frac{1}{N}\mathbb{E}\log\int \mathbb{1}\{\boldsymbol{x} \in A\}e^{-H_{t,0}(\boldsymbol{x})}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})\\ &\leq -hm + \frac{1}{N}\mathbb{E}\log\int \mathbb{1}\{\boldsymbol{x} \in A\}e^{-H_{t,0}(\boldsymbol{x})+hNR_{1,*}}\mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})\\ &= -hm + \frac{1}{N}\sum_{i=1}^N \widehat{\psi}(r,(\bar{r}+h)x_i^*),\end{aligned}$$

where the last line follows by dropping the indicator from the integral. ■

A consequence of the above bound is the convergence in probability of the overlaps:

Proposition 4.12. *For all $\lambda \in \mathcal{A}$, all $\epsilon > 0$ and all $t \in [0,1]$, there exist constants $K = K(\lambda,\epsilon) \geq 0$, $c = c(\lambda,\epsilon,t,P_{\mathrm{x}}) \geq 0$ such that*

$$\mathbb{E}\left\langle \mathbb{1}\left\{\left|R_{1,*} - q^*(\lambda)\right| \geq \epsilon\right\}\right\rangle_t \leq Ke^{-cN}.$$

The map $t \mapsto c(t)$ has the following properties:

- *If $t < 1$ then $c(t) > 0$.*
- *If either $\lambda < \lambda_c$ or P_{x} is not symmetric about the origin then $\inf_{t\in[0,1]} c(t) > 0$.*
- *Conversely, if $\lambda > \lambda_c$ and P_{x} is symmetric about the origin then $c(t) \sim c_0(1-t)^2$ as $t \to 1$, for some $c_0 = c_0(\lambda,\epsilon,P_{\mathrm{x}}) > 0$.*

Moreover, if $\lambda > \lambda_c$, P_{x} is symmetric and $t = 1$ then one still has

$$\mathbb{E}\left\langle \mathbb{1}\left\{\left||R_{1,*}| - q^*(\lambda)\right| \geq \epsilon\right\}\right\rangle \leq Ke^{-cN},$$

with $c = c(\lambda,\epsilon,P_{\mathrm{x}}) > 0$.

Proof of Proposition 4.12

The proof is similar to that of Proposition 2.15, with additional technical complications when $\lambda > \lambda_c$. For $\epsilon, \epsilon' > 0,\ t \in [0,1)$, we can write the decomposition

$$\begin{aligned}\mathbb{E}\left\langle \mathbb{1}\left\{|R_{1,*} - q^*(\lambda)| \geq \epsilon\right\}\right\rangle_t = &\sum_{l\geq 0} \mathbb{E}\left\langle \mathbb{1}\big\{R_{1,*} - q^* - \epsilon \in [l\epsilon', (l+1)\epsilon')\big\}\right\rangle_t \\ &+ \sum_{l\geq 0} \mathbb{E}\left\langle \mathbb{1}\big\{ - R_{1,*} + q^* - \epsilon \in [l\epsilon, (l+1)\epsilon')\big\}\right\rangle_t ,\end{aligned}$$

where the integer index l ranges over a finite set of size $\leq K/\epsilon'$ since the prior P_{x} has bounded support. We will only treat the first sum in the above expression since the argument extends trivially to the second sum. Let $A = \big\{R_{1,*} - q^* - \epsilon \in [l\epsilon, (l+1)\epsilon')\big\}$ and write

$$\mathbb{E}\left\langle \mathbb{1}(A)\right\rangle_t = \mathbb{E}_{\boldsymbol{x}^*}\, \mathbb{E}_{\boldsymbol{W},\boldsymbol{z}} \left[\frac{\int_A e^{-H_t(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})}{\int e^{-H_t(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})}\right]. \tag{4.27}$$

By concentration over the Gaussian disorder $\boldsymbol{W}, \boldsymbol{z}$ (Lemma 2.16), for any given l and $u \geq 0$, we simultaneously have

$$\frac{1}{N}\log \int e^{-H_t(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) \geq \frac{1}{N}\, \mathbb{E}_{\boldsymbol{W},\boldsymbol{z}} \log \int e^{-H_t(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) - u$$

and

$$\begin{aligned}\frac{1}{N}\log \int \mathbb{1}\{\boldsymbol{x} \in A\} e^{-H_t(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) &\leq \frac{1}{N}\, \mathbb{E}_{\boldsymbol{W},\boldsymbol{z}} \log \int_A e^{-H_t(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}) + u \\ &= \Phi_{\epsilon'}(q^* + \epsilon + l\epsilon'; t) + u,\end{aligned}$$

with probability at least $1 - 4e^{-Nu^2/K}$. On the complement of this event event, we simply bound the fraction in (4.27) by 1. Combining the above bounds we have

$$\mathbb{E}\left\langle \mathbb{1}(A)\right\rangle_t \leq 2e^{-Nu^2/K} + \mathbb{E}_{\boldsymbol{x}^*}\left[e^{N(\Delta + 2u)}\right], \tag{4.28}$$

where

$$\Delta = \Phi_{\epsilon'}(m;t) - \frac{1}{N}\, \mathbb{E}_{\boldsymbol{W},\boldsymbol{z}} \log \int_A e^{-H_t(\boldsymbol{x})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x})\ (\leq 0),$$

with $m = q^* + \epsilon + l\epsilon'$. Proposition 4.11 implies that

$$\begin{aligned}\Delta \leq \inf_{h\in\mathbb{R}} \left\{\frac{1}{N}\sum_{i=1}^{N} \widehat{\psi}(r, (\bar{r} + h)x_i^*) - hm\right\} - \frac{\lambda t m^2}{4} \\ - \frac{1}{N}\, \mathbb{E}_{\boldsymbol{W},\boldsymbol{z}} \log \int_A e^{-H_t(\boldsymbol{x})} + \frac{\lambda t \epsilon^2}{2} + \frac{\lambda K}{N}.\end{aligned}$$

Concentration over the randomness of $\boldsymbol{x}^*$ (Lemmas 2.17 and 2.18) implies that for $u' \geq 0$, we have

$$\begin{aligned}\Delta &\leq \inf_{h\in\mathbb{R}} \left\{ \bar{\psi}(r, \bar{r}+h) - hm \right\} - \frac{\lambda t m^2}{4} - \varphi(t) + 2u' + \frac{\lambda\epsilon^2}{2} + \frac{\lambda K}{N} \\ &\leq \inf_{h\in\mathbb{R}} \left\{ \bar{\psi}(r, \bar{r}+h) - hm \right\} - \frac{\lambda t m^2}{4} - \phi_{\mathsf{RS}}(\lambda; t) + 2u' + \frac{\lambda\epsilon^2}{2} + \frac{\lambda K}{N}.\end{aligned}$$

with probability at least $1 - 4e^{-Nu'^2/K}$, where the last inequality come from (4.26).

Lemma 4.13. *There exists $K = K(P_{\mathrm{x}}) > 0$ such that*

$$\inf_{h\in\mathbb{R}} \left\{ \bar{\psi}(r, \bar{r}+h) - hm \right\} \leq \bar{\psi}(r, \bar{r}) - \frac{1}{2K^2}\big(m - \partial_s\bar{\psi}(r, \bar{r})\big)^2.$$

Therefore, with probability at least $1 - 4e^{-Nu'^2/K}$, we have

$$\Delta \leq \delta\phi + 2u' + \frac{\lambda\epsilon^2}{2} + \frac{\lambda K}{N}, \tag{4.29}$$

with

$$\delta\phi = \bar{\psi}(r, \bar{r}) - \frac{\lambda t m^2}{4} - \phi_{\mathsf{RS}}(\lambda; t) - \frac{1}{2K^2}\big(m - \partial_s\bar{\psi}(r, \bar{r})\big)^2. \tag{4.30}$$

Now, the crucial observation is that $\delta\phi$ is strictly negative when m is far from q^*:

Proposition 4.14. *For all $\lambda \in \mathcal{A}$, all $\epsilon > 0$ and all $t \in [0,1]$, there exist constants $c = c(\lambda, \epsilon, t, P_{\mathrm{x}}) \geq 0$ such that*

$$\forall m \in \mathbb{R} \qquad |m - q^*(\lambda)| \geq \epsilon \qquad \Longrightarrow \qquad \delta\phi \leq -c.$$

Moreover, if $t < 1$ then $c > 0$. If either $\lambda < \lambda_c$ or P_{x} is not symmetric about the origin $\inf_{t\in[0,1]} c(t) > 0$. *Lastly, if $\lambda > \lambda_c$ and P_{x} is symmetric, then $c(t) \sim c_0(1-t)$ as $t \to 1$, for some $c_0 = c_0(\lambda, \epsilon, P_{\mathrm{x}}) > 0$.*

We can now finish the general using the above Proposition. Since $m = q^* + \epsilon + l\epsilon'$, $|m - q^*| \geq \epsilon$, and by Proposition 4.14, and (4.29) we have $\Delta \leq -c + 2u' + \frac{\lambda\epsilon^2}{2} + \frac{\lambda K}{N}$ with probability at least $1 - 4e^{-Nu'^2/K}$; otherwise $\Delta \leq 0$. Plugging this into (4.28), we obtain

$$\mathbb{E}\left\langle \mathbb{1}(A) \right\rangle_t \leq 2e^{-Nu^2/K} + 4e^{-Nu'^2/K + 2Nu} + 2e^{N(-c+2u+2u'+\lambda\epsilon^2/2)+\lambda K}.$$

We now adjust the parameters ϵ, u, u', so that the right-hand side is exponentially small in N: $\mathbb{E}\left\langle \mathbb{1}(A) \right\rangle_t \leq Ke^{-cN}$. Finally, if P_{x} is symmetric and $t = 1$, then it suffices to consider non-negative values of m in the above argument to prove the corresponding statement. ■

To prove Propopsition 4.14, we will need to following useful lemma:

Lemma 4.15. *For all $r \geq 0$, it holds that*

- *The function $s \mapsto \bar{\psi}(r, s)$ is strictly convex, hence strongly convex on any compact.*
- *There exist a constant $c = c(r, P_{\mathtt{x}}) \geq 0$ such that $\bar{\psi}(r, -r) \leq \bar{\psi}(r, r) - c$. If $r > 0$ then $c > 0$* unless *the prior $P_{\mathtt{x}}$ is symmetric about the origin (in which case $\bar{\psi}(r, -r) = \bar{\psi}(r, r)$).*
- *The map $r \mapsto c(r, P_{\mathtt{x}})$ is increasing on $\mathbb{R}_+$.*

Proof of Propopsition 4.14. The bound we seek to prove will come from different sources, depending on whether t is small or not and whether m positive and negative. Recall that $r = (1-t)\lambda q^* + t\lambda|m|$ and $\bar{r} = (1-t)\lambda q^* + t\lambda m$.

Large t. Assume $t \geq t_0$ to be determined later. For $m \geq 0$, $\bar{\psi}(r, \bar{r}) = \bar{\psi}(r, r) = \psi(r)$. Since ψ is a convex function we have

$$\begin{aligned}\psi(r) - \frac{t\lambda m^2}{4} &\leq (1-t)\psi(\lambda q^*) + t\psi(\lambda m) - \frac{t\lambda m^2}{4} \\ &= (1-t)\psi(\lambda q^*) + tF(\lambda, m). \qquad (4.31)\end{aligned}$$

Since $q^*(\lambda)$ is the unique maximizer of $m \mapsto F(\lambda, m)$, $|m - q^*| \geq \epsilon > 0$ implies that $F(\lambda, m) \leq F(\lambda, q^*) - c(\epsilon)$ for some $c(\epsilon) > 0$. This in turn implies

$$\begin{aligned}\delta\phi &\leq (1-t)\psi(\lambda q^*) + t\big(\psi(\lambda q^*) - \frac{\lambda q^{*2}}{4} - c(\epsilon)\big) - \phi_{\mathsf{RS}}(\lambda; t) \\ &= \psi(\lambda q^*) - \frac{t\lambda q^{*2}}{4} - \phi_{\mathsf{RS}}(\lambda; t) - tc(\epsilon) \\ &\leq -t_0 c(\epsilon).\end{aligned}$$

The conclusion is reached for $m \geq 0$. Now we would like to prove the same bound for negative overlaps. Assume that $m < 0$. Although m is far from q^*, the issue is that it could still be close to $-q^*$. This case will need special care.

If $\lambda < \lambda_c$ then $q^*(\lambda) = 0$, and by Proposition 4.15 we have $\bar{\psi}(-t\lambda m, t\lambda m) - \frac{t\lambda m^2}{4} \leq \psi(-t\lambda m) - \frac{t\lambda m^2}{4} \leq t_0 F(\lambda, -m)$, and we finish the argument as in the case of positive overlap. Now we deal with the case $\lambda > \lambda_c$,i.e., $q^* > 0$.

- Suppose $|m + q^*| \geq \epsilon$. We let $\alpha = \frac{(1-t)q^*}{(1-t)q^* - tm}$.

$$\begin{aligned}\bar{\psi}(r, \bar{r}) &= \bar{\psi}(r, \alpha r - (1-\alpha)r) \\ &\leq \alpha\bar{\psi}(r, r) + (1-\alpha)\bar{\psi}(r, -r) \\ &\leq \psi(r)\end{aligned}$$

  where the last two lines follow from Proposition 4.15. Since $|m + q^*| \geq \epsilon$ we finish once again as in the positive overlap case, starting from line (4.31).

- Suppose $|m+q^*| \leq \epsilon$. Then $1-\alpha \geq t_0\frac{q^*-\epsilon}{q^*+\epsilon}$ and $|r-\lambda q^*| \leq \lambda\epsilon$. If P_{x} is asymmetric we use the bounds of Proposition 4.15:

$$\begin{aligned}\bar{\psi}(r,\bar{r}) &\leq \alpha\bar{\psi}(r,r) + (1-\alpha)\bar{\psi}(r,-r)\\ &\leq \psi(r) - (1-\alpha)c(r)\\ &\leq \psi(r) - t_0\frac{q^*-\epsilon}{q^*+\epsilon}c(\lambda(q^*-\epsilon)).\end{aligned}$$

  The last line follows since $r \mapsto c(r)$ is increasing. Then we finish the argument as in (4.31).

Small t. Assume now that $t \leq t_0$. In this situation, we draw the gap from the term $(m-\partial_s\bar{\psi}(r,\bar{r}))^2$ (so far unused) in (4.30). The functions $\bar{\psi}(r,\cdot)$ and $\bar{\psi}(\cdot,\cdot) = \psi(\cdot)$ have bounded second derivatives so

$$\max\ \left\{ \left|\partial_s\bar{\psi}(r,\bar{r}) - \partial_s\bar{\psi}(r,r)\right| \ \ , \ \ \left|\partial_s\bar{\psi}(r,r) - \partial_s\bar{\psi}(q^*,q^*)\right| \right\} \ \leq K\lambda t_0.$$

Moreover,

$$(m-q^*)^2 \leq 2(m-\partial_s\bar{\psi}(r,\bar{r}))^2 + 2(q^*-\partial_s\bar{\psi}(r,\bar{r}))^2.$$

Since $\partial_s\bar{\psi}(q^*,q^*) = q^*$ we have

$$(m-\partial_s\bar{\psi}(r,\bar{r}))^2 \geq \frac{1}{2}(m-q^*)^2 - K\lambda^2 t_0^2 \geq \frac{\epsilon^2}{2} - K\lambda^2 t_0^2,$$

and here we choose t_0 to be accordingly small, and we finish the argument.

Note that the assumption that P_{x} is not symmetric about the origin is used only in the case where the (negative) overlap m is close to $-q^*$. Consequently, the gap is independent of t in all cases. Alternatively, without this asymmetry assumption (and when $q^* > 0$), we see that there is no hope of a gap independent of t since the potential $\mathbb{E}_{\boldsymbol{x}^*}\Phi_{\epsilon'}(m;t)$ is closer and closer to being even as $t \to 1$. But we can still obtain a gap that depends on $t(1-t)$ via a strong convexity argument.

The function $s \mapsto \bar{\psi}(r,s)$ is strongly convex on any interval, and for all $r \geq 0$. Therefore, for $m \geq 0$ recalling $r = (1-t)\lambda q^* - t\lambda m$ and $\alpha = \frac{(1-t)q^*}{(1-t)q^*-tm}$, there exists a constant $c > 0$ depending only on λ and P_{x} (this constant is a bound on r) such that

$$\begin{aligned}\bar{\psi}(r,\bar{r}) &= \bar{\psi}(r,\alpha r - (1-\alpha)r)\\ &\leq \alpha\bar{\psi}(r,r) + (1-\alpha)\bar{\psi}(r,-r) - \frac{c}{2}\alpha(1-\alpha)(2r)^2\\ &= \alpha\bar{\psi}(r,r) + (1-\alpha)\bar{\psi}(r,-r) - 2ct(1-t)\lambda^2 q^*|m|\\ &\leq \bar{\psi}(r,r) - 2ct(1-t)\lambda^2 q^*(q^*-\epsilon),\end{aligned}$$

where the last bounds follows from $\bar{\psi}(r,-r) \leq \bar{\psi}(r,r)$ and $|m+q^*| \leq \epsilon$ (recall that this is the only case where such an argument is needed). ■

Proof of Lemma 4.13. This is a consequence of a classical symmetrization argument. If we define $\langle\cdot\rangle$ as the Gibbs average associated to the random Hamiltonian $x \mapsto \sqrt{r}zx + \bar{r}xx^* - rx^2/2$, we have

$$\bar{\psi}(r, \bar{r} + h) = \bar{\psi}(r, \bar{r}) + \mathbb{E}_{z,x^*} \log \left\langle e^{hxx^*} \right\rangle .$$

Now, for fixed h, x^* and z, we have by Jensen's inequality

$$\left\langle e^{hx^*(x-\langle x\rangle)} \right\rangle \leq \left\langle e^{hx^*(x-x')} \right\rangle ,$$

where x' is an indepdenent copy of x (i.e. distributed according to $\langle\cdot\rangle$). Since the random variable $x - x'$ is symmetric, its odd moments vanish, and we have

$$\begin{aligned}\left\langle e^{hx^*(x-x')} \right\rangle &= \sum_{k\geq 0} \frac{(hx^*)^k}{k!} \left\langle (x - x')^k \right\rangle \\ &= \sum_{k\geq 0} \frac{(hx^*)^{2k}}{(2k)!} \left\langle (x - x')^{2k} \right\rangle .\end{aligned}$$

Since $|x| \leq K$ a.s. $|x - x'| \leq K' = 2K$, and $(2k)! \geq 2^k k!$, it follows that

$$\begin{aligned}\langle e^{hx^*(x-x')} \rangle &\leq \sum_{k\geq 0} \frac{(hx^*)^{2k}}{2^k k!} K'^{2k} \\ &= e^{h^2 x^{*2} K'^2/2}.\end{aligned}$$

From the above we obtain

$$\mathbb{E}_{z,x^*} \log \left\langle e^{hxx^*} \right\rangle \leq h\,\mathbb{E}\langle xx^* \rangle + \frac{1}{2}h^2\,\mathbb{E}[x^{*2}]K'^2 \leq h\,\mathbb{E}\langle xx^* \rangle + \frac{1}{2}h^2 K''^2.$$

Now,

$$\begin{aligned}\min_h\{\bar{\psi}(r, \bar{r} + h) - hm\} &\leq \bar{\psi}(r, \bar{r}) + \min_h\{h(\mathbb{E}\langle xx^* \rangle - m) + h^2 K''^2/2\} \\ &= \bar{\psi}(r, \bar{r}) - \frac{1}{2K''^2}(\mathbb{E}\langle xx^* \rangle - m)^2.\end{aligned}$$

■

Now we prove Lemma 4.15. A straightforward calculation reveals that

$$\frac{\partial}{\partial s}\bar{\psi}(r, s) = \mathbb{E}\left[\langle xx^* \rangle\right], \quad \text{and} \quad \frac{\partial^2}{\partial s^2}\bar{\psi}(r, s) = \mathbb{E}\left[x^{*2}(\langle x^2 \rangle - \langle x \rangle^2)\right] > 0,$$

so that $s \mapsto \frac{\partial}{\partial s}\bar{\psi}(r, s)$ is Lipschitz and strongly convex on any interval, and for all $r \geq 0$.

Let $\nu = P_{\mathtt{x}}$, and let μ be the symmetric part of $P_{\mathtt{x}}$, i.e., $\mu(A) = (P_{\mathtt{x}}(A) + P_{\mathtt{x}}(-A))/2$ for all Borel $A \subseteq \mathbb{R}$. We observe that ν is absolutely continuous with respect to μ, so that the Radon-Nikodym derivative $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ is a well-defined measurable function from $\mathbb{R}$ to $\mathbb{R}_+$ that integrates to one.

Proposition 4.16. *For all $r \geq 0$, we have*

$$\bar{\psi}(r,r) - \bar{\psi}(r,-r) \geq 2\,\mathbb{E}\left[\left\langle \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) - 1 \right\rangle_{\mu,r}^2\right],$$

where $\langle\cdot\rangle_{\mu,r}$ is the average w.r.t. to the Gibbs measure corresponding to the Gaussian channel $y = \sqrt{r}x^* + z$, $x^* \sim \mu$ and $z \sim \mathcal{N}(0,1)$. Moreover, if $r > 0$, the right-hand side of the above inequality is zero if and only if *$\mu = \nu$, i.e., the prior P_{x} is symmetric.*

Finally, the last statement is given here.

Lemma 4.17. *The map $r \mapsto \mathbb{E}\left[\left\langle \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) - 1 \right\rangle_{\mu,r}^2\right]$ is increasing on $\mathbb{R}_+$.*

Proof. This is a matter of showing that the derivative of the above function is non-negative. By standard manipulations (Gaussian integration by parts, Nishimori property), the derivative can be written as

$$\mathbb{E}\left[\left\langle x\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) - 1\right) \right\rangle_{\mu,r}^2\right].$$

■

***Proof** of Proposition 4.16.* The argument relies on a smooth interpolation method between the two measures μ and ν. Let $t \in [0,1]$ and let $\rho_t = (1-t)\mu + t\nu$. Further, let $r, s \geq 0$ be fixed, and

$$\bar{\psi}(r,s;t) := \mathbb{E}_z \int \left(\log \int \exp\left(\sqrt{r}zx + sxx^* - \frac{r}{2}x^2\right) \mathrm{d}\rho_t(x)\right) \mathrm{d}\rho_t(x^*),$$

where $z \sim \mathcal{N}(0,1)$. Now let

$$\phi(t) = \bar{\psi}(r,r;t) - \bar{\psi}(r,-r;t).$$

We have $\phi(1) = \bar{\psi}(r,r) - \bar{\psi}(r,-r)$ on the one hand, and since μ is a symmetric distribution, $\phi(0) = 0$ on the other. We will show that ϕ is a convex increasing function on the interval $[0,1]$, strictly so if $\mu \neq \nu$, and that $\phi'(0) = 0$. Then we deduce that $\phi(1) \geq \frac{\phi''(0)}{2}$, allowing us to conclude. First, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\bar{\psi}(r,r;t) &= \mathbb{E}_z \int \log \int e^{\sqrt{r}zx + rxx^* - \frac{r}{2}x^2} \mathrm{d}\rho_t(x)\ \mathrm{d}(\nu-\mu)(x^*) \\ &\quad + \mathbb{E}_z \int \frac{\int e^{\sqrt{r}zx + rxx^* - \frac{r}{2}x^2} \mathrm{d}(\nu-\mu)(x)}{\int e^{\sqrt{r}zx + rxx^* - \frac{r}{2}x^2} \mathrm{d}\rho_t(x)}\ \mathrm{d}\rho_t(x^*),\end{aligned}$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\bar{\psi}(r,r;t) = 2\,\mathbb{E}_z \int \frac{\int e^{\sqrt{r}zx + rxx^* - \frac{r}{2}x^2} \mathrm{d}(\nu-\mu)(x)}{\int e^{\sqrt{r}zx + rxx^* - \frac{r}{2}x^2} \mathrm{d}\rho_t(x)}\ \mathrm{d}(\nu-\mu)(x^*)$$

$$- 2\,\mathbb{E}_z \int \left(\frac{\int e^{\sqrt{r}zx+rxx^*-\frac{r}{2}x^2} \mathrm{d}(\nu-\mu)(x)}{\int e^{\sqrt{r}zx+rxx^*-\frac{r}{2}x^2} \mathrm{d}\rho_t(x)} \right)^2 \mathrm{d}\rho_t(x^*).$$

Similar expressions holds for $\bar{\psi}(r,-r;t)$ where x^* is replaced by $-x^*$ inside the exponentials. We see from the expression of the first derivative at $t=0$ that $\bar{\psi}(r,r;0)' = \bar{\psi}(r,-r;0)'$. This is because $\rho_0 = \mu$ is symmetric about the origin, so a sign change (of x for the first term, and x^* for the second term in the expression) does not affect the value of the integrals. Hence $\phi'(0) = 0$. Now, we focus on the second derivative. Observe that since μ is the symmetric part of ν, $\nu - \mu$ is anti-symmetric. This implies that the first term in the expression of the second derivative changes sign under a sign change in x^*, and keeps the same modulus. As for the second term, a sign change in x^* induces integration against $\mathrm{d}\rho_t(-x^*)$. Hence we can write the difference $(\bar{\psi}(r,r;t) - \bar{\psi}(r,-r;t))''$ as

$$\begin{aligned}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\phi(t) = 4\,\mathbb{E}_z \int \frac{\int e^{\sqrt{r}zx+rxx^*-\frac{r}{2}x^2} \mathrm{d}(\nu-\mu)(x)}{\int e^{\sqrt{r}zx+rxx^*-\frac{r}{2}x^2} \mathrm{d}\rho_t(x)}\, \mathrm{d}(\nu-\mu)(x^*) \\ - 2\,\mathbb{E}_z \int \left(\frac{\int e^{\sqrt{r}zx+rxx^*-\frac{r}{2}x^2} \mathrm{d}(\nu-\mu)(x)}{\int e^{\sqrt{r}zx+rxx^*-\frac{r}{2}x^2} \mathrm{d}\rho_t(x)} \right)^2 (\mathrm{d}\rho_t(x^*) - \mathrm{d}\rho_t(-x^*)).\end{aligned}$$

For any Borel A, we have $\rho_t(A) - \rho_t(-A) = (1-t)(\mu(A) - \mu(-A)) + t(\nu(A) - \nu(-A)) = 2t(\nu-\mu)(A)$. Therefore the second term in the above expression becomes

$$-4t\,\mathbb{E}_z \int \left(\frac{\int e^{\sqrt{r}zx+rxx^*-\frac{r}{2}x^2} \mathrm{d}(\nu-\mu)(x)}{\int e^{\sqrt{r}zx+rxx^*-\frac{r}{2}x^2} \mathrm{d}\rho_t(x)} \right)^2 \mathrm{d}(\nu-\mu)(x^*).$$

Since both μ and ν are absolutely continuous with respect to ρ_t for all $0 \le t < 1$ we write

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\phi(t) = 4\,\mathbb{E}_{z,x^*} \left\langle \frac{\mathrm{d}(\nu-\mu)}{\mathrm{d}\rho_t}(x) \frac{\mathrm{d}(\nu-\mu)}{\mathrm{d}\rho_t}(x^*) \right\rangle - 4t\,\mathbb{E}_{z,x^*} \left\langle \frac{\mathrm{d}(\nu-\mu)}{\mathrm{d}\rho_t}(x) \right\rangle^2,$$

where the Gibbs average is with respect to the posterior of x given z, x^* under the Gaussian channel $y = \sqrt{r}x^* + z$, and the expectation is under $x^* \sim \rho_t$ and $z \sim \mathcal{N}(0,1)$. By the Nishimori property, we simplify the above expression to

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\phi(t) = 4(1-t)\,\mathbb{E}\left[\left\langle \frac{\mathrm{d}(\nu-\mu)}{\mathrm{d}\rho_t}(x) \right\rangle^2 \right],$$

where the expression is valid for all $0 \le t < 1$. From here we see that the function ϕ is convex on $[0,1]$ (where we have closed the right end of the interval by continuity). Since $\phi(0) = \phi'(0) = 0$, ϕ is also increasing on $[0,1]$. Therefore we have

$$\phi(1) \ge \frac{1}{2}\phi''(0) = 2\,\mathbb{E}\left[\left\langle \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) - 1 \right\rangle_{\mu,r}^2 \right].$$

Now it remains to show that if $\bar{\psi}(r,r) = \bar{\psi}(r,-r)$ for some $r > 0$ then $\mu = \nu$. By the lower bound we have shown, equality of $\bar{\psi}(r,r)$ and $\bar{\psi}(r,-r)$ would imply

$$\left\langle \frac{\mathrm{d}\nu}{\mathrm{d}\mu}(x) \right\rangle_{\mu,r} = \frac{\int e^{\sqrt{r}zx + rxx^* - \frac{r}{2}x^2} \mathrm{d}\nu(x)}{\int e^{\sqrt{r}zx + rxx^* - \frac{r}{2}x^2} \mathrm{d}\mu(x)} = 1$$

for (Lebesgue-)almost all z and P_{x}-almost all x^*. We make the change of variable $z \mapsto \sqrt{r}(z - x^*)$ and complete the squares, then the above is equivalent to

$$\int e^{-\frac{r}{2}(x-z)^2} \mathrm{d}\nu(x) = \int e^{-\frac{r}{2}(x-z)^2} \mathrm{d}\mu(x)$$

for almost all z. The above expressions are convolutions of the measures ν and μ against the Gaussian kernel. By taking the Fourier transform on both sides and using Fubini's theorem, we get equality of the characteristic functions of μ and ν: for all $\xi \in \mathbb{R}$,

$$\int e^{\mathrm{i}\xi x} \mathrm{d}\nu(x) = \int e^{\mathrm{i}\xi x} \mathrm{d}\mu(x).$$

This is because the Fourier transform of the Gaussian (another Gaussian) vanishes nowhere on the real line, thus it can be simplified on both sides. This of course implies that $\nu = \mu$, and concludes our proof. ■

4.5 The cavity method

Now that we have established the convergence in probability of $R_{1,*}$ to $q^*(\lambda)$ under $\mathbb{E}\langle\cdot\rangle_t$ in Lemma 4.12, we use the cavity method to prove the convergence of the moments of the overlap.

Our proofs of Theorems 4.8 and 4.9 are interlaced. The skeleton of the argument is as follows:

1. We first prove convergence of the second moment: $\mathbb{E}\left\langle (R_{1,*} - q^*)^2 \right\rangle_t \leq \mathcal{O}(1/N + e^{-c(t)N})$.

2. We then deduce from 1. the convergence of the fourth moment via an inductive argument: $\mathbb{E}\left\langle (R_{1,*} - q^*)^4 \right\rangle_t \leq \mathcal{O}(1/N^2 + e^{-c(t)N})$. This finishes the proof of Theorem 4.8.

3. Using 2., we revisit our proof of 1., and refine the estimates in order to obtain the sharper result $N \cdot \mathbb{E}\left\langle (R_{1,*} - q^*)^2 \right\rangle_t \to \Delta_{\mathsf{RS}}(\lambda; t)$. This finishes the proof of Theorem 4.9.

We will start by defining our interpolating Hamiltonian and state some preliminary bounds and properties. Then we will move on to the execution of the cavity computations.

Preliminary bounds

In this section the parameter $t \in [0,1]$ is fixed and treated as a constant. We consider the Hamiltonian

$$\begin{aligned}-H_t^-(\boldsymbol{x}) :=& \sum_{i<j\leq N-1} -\frac{\lambda t}{2N} x_i^2 x_j^2 + \sqrt{\frac{\lambda t}{N}} W_{ij} x_i x_j + \frac{\lambda t}{N} x_i x_i^* x_j x_j^* \\ &+ \sum_{i=1}^{N-1} -\frac{(1-t)r}{2} x_i^2 + \sqrt{(1-t)r} z_i x_i + (1-t) r x_i x_i^*,\end{aligned}$$

where we have striped away the contribution of the variable x_N from H_t (equ. (4.16)). This contribution is considered separately: for $t' \in [0,1]$, we let

$$\begin{aligned}-h_{t'}(\boldsymbol{x}) :=& \sum_{i=1}^{N-1} -\frac{\lambda t'}{2N} x_i^2 x_N^2 + \sqrt{\frac{\lambda t'}{N}} W_{iN} x_i x_N + \frac{\lambda t'}{N} x_i x_i^* x_N x_N^* \\ &- \frac{(1-t')r}{2} x_N^2 + \sqrt{(1-t')r} z_N x_N + (1-t') r x_N x_N^*.\end{aligned}$$

We let $r = \lambda q^*(\lambda)$ and let our interpolation, with the time parameter $s \in [0,1]$, be

$$H_{t,s}(\boldsymbol{x}) := H_t^-(\boldsymbol{x}) + h_{ts}(\boldsymbol{x}).$$

At $s = 1$ we have $H_{t,s} = H_t$, and at $s = 0$ the variable x_N decouples from the rest of the variables. For an integer $n \geq 1$ and $f : (\mathbb{R}^N)^{n+1} \mapsto \mathbb{R}$, we define

$$\left\langle f(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}, \boldsymbol{x}^*) \right\rangle_{t,s} := \frac{\int f(\boldsymbol{x}^{(1)}, \cdots, \boldsymbol{x}^{(n)}, \boldsymbol{x}^*) \prod_{l=1}^n e^{-H_{t,s}(\boldsymbol{x}^{(l)})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}^{(l)})}{\int \prod_{l=1}^n e^{-H_{t,s}(\boldsymbol{x}^{(l)})} \mathrm{d}P_{\mathrm{x}}^{\otimes N}(\boldsymbol{x}^{(l)})},$$

similarly to (4.15). Following Talagrand's notation, we write

$$R_{l,l'}^- = \frac{1}{N} \sum_{i=1}^{N-1} x_i^{(l)} x_i^{(l')}, \quad \text{and} \quad \nu_s(f) = \mathbb{E}\langle f \rangle_{t,s}.$$

In our last notation, we have only emphasized the dependence of the average on s; the parameter t will henceforth remain fixed. Moreover, we write $\nu(f)$ for $\nu_1(f)$. The following three lemmas are variants of Lemma 1.6.3, Lemma 1.6.4 and Proposition 1.8.1 respectively in (Talagrand, 2011a).

Lemma 4.18. *For all $n \geq 1$,*

$$\frac{\mathrm{d}}{\mathrm{d}s}\nu_s(f) = \frac{\lambda t}{2} \sum_{1\leq l\neq l'\leq n} \nu_s((R_{l,l'}^- - q) y^{(l)} y^{(l')} f) - \lambda t n \sum_{l=1}^{n} \nu_s(R_{l,n+1}^- - q) y^{(l)} y^{(n+1)} f)$$

$$+ \lambda tn \sum_{l=1}^{n} \nu_s((R^-_{l,*} - q)y^{(l)}y^* f) - \lambda tn \nu_s((R^-_{n+1,*} - q)y^{(n+1)}y^* f)$$
$$+ \lambda t \frac{n(n+1)}{2} \nu_s((R^-_{n+1,n+2} - q)y^{(n+1)}y^{(n+2)} f),$$

where we have written $y = x_N$.

Proof. The computation relies on Gaussian integration by parts. See Talagrand (2011a, Lemma 1.6.3) for the details of a similar computation. ■

Lemma 4.19. *If f is a bounded non-negative function, then for all $s \in [0,1]$,*

$$\nu_s(f) \le K(\lambda, n)\nu(f).$$

Proof. Since the variables and the overlaps are all bounded, and $t \le 1$, using Lemma 4.18 we have for all $s \in [0,1]$

$$|\nu_s'(f)| \le K(\lambda, n)\nu_s(f).$$

Then we conclude using Grönwall's lemma. ■

Lemma 4.20. *For all $s \in [0,1]$, and all $\tau_1, \tau_2 > 0$ such that $1/\tau_1 + 1/\tau_2 = 1$,*

$$|\nu_s(f) - \nu_0(f)| \le K(\lambda, n)\nu\left(\left|R^-_{1,*} - q\right|^{\tau_1}\right)^{1/\tau_1} \cdot \nu\left(|f|^{\tau_2}\right)^{1/\tau_2} \tag{4.32}$$
$$|\nu_s(f) - \nu_0(f) - \nu_0'(f)| \le K(\lambda, n)\nu\left(\left|R^-_{1,*} - q\right|^{\tau_1}\right)^{1/\tau_1} \cdot \nu\left(|f|^{\tau_2}\right)^{1/\tau_2}. \tag{4.33}$$

Proof. We use Taylor's approximations

$$|\nu_s(f) - \nu_0(f)| \le \sup_{0 \le s \le 1} |\nu_s'(f)|,$$
$$|\nu_s(f) - \nu_0(f) - \nu_0'(f)| \le \sup_{0 \le s \le 1} |\nu_s''(f)|,$$

then Lemma 4.18 and the triangle inequality to bound the right hand sides, then Hölder's inequality to bound each term in the derivative, and then we apply Lemma 4.19. (To compute the second derivative, one need to use Lemma 4.18 recursively.) ■

The cavity matrix

Recall the parameters $a(0)$, $a(1)$ and $a(2)$ from (4.12):

$$a(0) = \mathbb{E}\left[\langle x^2\rangle_r^2\right] - q^{*2}(\lambda), \quad a(1) = \mathbb{E}\left[\langle x^2\rangle_r \langle x\rangle_r^2\right] - q^{*2}(\lambda), \quad a(2) = \mathbb{E}\left[\langle x\rangle_r^4\right] - q^{*2}(\lambda),$$

where $r = \lambda q^*(\lambda)$. Now let

$$\boldsymbol{A} := \lambda \cdot \begin{bmatrix} a(0) & -2a(1) & a(2) \\ a(1) & a(0) - a(1) - 2a(2) & -2a(1) + 3a(2) \\ a(2) & 4a(1) - 6a(2) & a(0) - 6a(1) + 6a(2) \end{bmatrix}. \tag{4.34}$$

One can easily check that the transpose of this matrix has two eigenvalues μ_1 and μ_2 with expressions

$$\begin{aligned} \mu_1(\lambda) &= \lambda(a(0) - 2a(1) + a(2)), \\ \mu_2(\lambda) &= \lambda(a(0) - 3a(1) + 2a(2)), \end{aligned} \tag{4.35}$$

and associated eigenvectors $(1, -2, 1)$ and $(2, -3, 2)$, and of multiplicities two and one respectively. (The first eigenvalue appears in a 2×2 Jordan block.) We will need to control the largest eigenvalue of $\boldsymbol{A}^\top$. This matrix is the "planted" analogue of the one displayed in (Talagrand, 2011a, equ. (1.234)) for the SK model. By Cauchy-Schwarz, $\mu_1 - \mu_2 = \lambda(a(1) - a(2)) \geq 0$. As will be clear from the next subsection, the cavity computations we are about present are only informative when $\mu_1 < 1$. Interestingly, this is true for all values of λ where the RS formula ϕ_{RS} has two derivatives:

Lemma 4.21. *For all $\lambda \in \mathcal{A}$, $\mu_1(\lambda) < 1$.*

Proof. First, if $\lambda < \lambda_c$, then $q^*(\lambda) = 0$, and $\mu_1(\lambda) = \lambda(\mathbb{E}_{P_\mathsf{x}}[X^2])^2$. By Lemma 2.1, $\mu_1(\lambda) < 1$. Now we assume $\lambda \in \mathcal{A} \cap (\lambda_c, +\infty)$. Recall

$$\psi(r) = \mathbb{E}_{x^*,z} \log \int \exp\left(\sqrt{r}zx + rxx^* - \frac{r}{2}x^2\right) \mathrm{d}P_\mathsf{x}(x),$$

and the RS potential

$$F(\lambda, q) = \psi(\lambda q) - \frac{\lambda q^2}{4}.$$

It is a straightforward exercise to compute the first and the second derivatives of ψ using Gaussian integration by parts and the Nishimori property:

$$\psi'(r) = \frac{1}{2}\,\mathbb{E}\langle xx^*\rangle_r,$$

$$\psi''(r) = \frac{1}{2}\left(\mathbb{E}\langle x^2 x^{*2}\rangle_r - 2\,\mathbb{E}\langle {x^{(1)}}^2 x^{(2)} x^*\rangle_r + \mathbb{E}\langle x^{(1)}x^{(2)}x^{(3)}x^*\rangle_r\right).$$

With the choice $r = \lambda q^*(\lambda)$, we see that $\mu_1(\lambda) = 2\lambda\psi''(r)$. Now we observe that

$$\frac{\partial^2 F}{\partial q^2}(\lambda, q) = \frac{\lambda}{2}(2\lambda\psi''(\lambda q) - 1).$$

Since $q^*(\lambda)$ is a maximizer of the smooth function $F(\lambda, \cdot)$, and lies in the interior of its domain ($q^*(\lambda) > 0$ for $\lambda > \lambda_c$), then it must be a first-order stationary point: $\frac{\partial F}{\partial q}(\lambda, q^*) = 0$. Hence $\frac{\partial^2 F}{\partial q^2}(\lambda, q^*) \leq 0$, i.e., $\mu_1(\lambda) \leq 1$ for all $\lambda > \lambda_c$. Now we claim that the inequality must be strict for $\lambda \in \mathcal{A}$. Indeed, Lelarge and Miolane (2016, Proposition 15) show that whenever ϕ_{RS} is differentiable at λ, then the maximizer of $F(\lambda, \cdot)$ is unique and

$$\phi'_{\mathsf{RS}}(\lambda) = \frac{q^{*2}(\lambda)}{4}.$$

Therefore, *twice* differentiability of ϕ_{RS} implies first differentiability of $\lambda \mapsto q^*(\lambda)$ whenever $q^*(\lambda) > 0$ (i.e., $\lambda > \lambda_c$). Now we take advantage of first-order optimality: $\frac{\partial F}{\partial q}(\lambda, q^*) = 0$ is the same as

$$\psi'(\lambda q^*(\lambda)) = \frac{q^*(\lambda)}{2}.$$

The above can be seen as an equality of functions (of λ) defined almost everywhere. Taking one derivative yields

$$q^*(\lambda)\psi''(\lambda q^*(\lambda)) = \frac{1}{2}\big(1 - 2\lambda\psi''(\lambda q^*(\lambda))\big){q^*}'(\lambda).$$

Since $q^*(\lambda)$ and $\psi''(\lambda q^*(\lambda))$ are both positive, the right-hand side cannot vanish. This concludes the proof. ■

Cavity computations for the second moment

In this subsection we prove the convergence of the second moment of the overlaps:

$$\nu((R_{1,*} - q^*)^2) \leq \frac{K}{N} + Ke^{-c(t)N},$$

with $c(t) \sim c_0(1-t)^2$ as $t \to 1$ when $\lambda > \lambda_c$ and P_{x} is symmetric about the origin, and uniformly lower-bounded by a positive constant otherwise. To lighten the notation in the calculations to come, $q^*(\lambda)$ will be denoted simply by q, and we recall the notation $\nu(\cdot) = \mathbb{E}\langle\cdot\rangle_{t,1}$. Let

$$A = \nu\left((R_{1,*} - q)^2\right), \quad B = \nu\left((R_{1,*} - q)(R_{2,*} - q)\right), \quad C = \nu\left((R_{1,*} - q)(R_{2,3} - q)\right).$$

By symmetry between sites,

$$A = \nu\left((R_{1,*} - q)(x_N x_N^* - q)\right) = \frac{1}{N}\nu\left(x_N x_N^*(x_N x_N^* - q)\right) + \nu((R_{1,*}^- - q)(x_N x_N^* - q)).$$

By the first bound (4.32) of Lemma 4.20 with $\tau_1 = 1$, $\tau_2 = \infty$, we get

$$\nu(x_N x_N^*(x_N x_N^* - q)) = \nu_0(x_N x_N^*(x_N x_N^* - q)) + \delta = a(0) + \delta,$$

with $|\delta| \leq K(\lambda)\nu(|R_{1,*}^- - q|)$. On the other hand, by the second bound (4.33) with $\tau_1 = 1$, $\tau_2 = \infty$, we get

$$\nu((R_{1,*}^- - q)(x_N x_N^* - q)) = \nu_0'((R_{1,*}^- - q)(x_N x_N^* - q)) + \delta.$$

This is because $\nu_0((R_{1,*}^- - q)(x_N x_N^* - q)) = 0$, since last variable x_N decouples from the remaining $N-1$ variables under the measure ν_0. Now, we use Lemma 4.18 with $n = 1$, to evaluate the above derivative at $t = 0$. We still write $y^{(l)} = x_N^{(l)}$.

$$\nu_0'((R_{1,*}^- - q)(x_N x_N^* - q)) = -\lambda t \nu_0(y^{(1)}y^{(2)}(y^{(1)}y^* - q)(R_{1,*}^- - q)(R_{1,2}^- - q))$$

$$
\begin{aligned}
&+ \lambda t\nu_0(y^{(1)}y^*(y^{(1)}y^* - q)(R^-_{1,*} - q)^2)\\
&- \lambda t\nu_0(y^{(2)}y^*(y^{(1)}y^* - q)(R^-_{1,*} - q)(R^-_{2,*} - q))\\
&+ \lambda t\nu_0(y^{(2)}y^{(3)}(y^{(1)}y^* - q)(R^-_{1,*} - q)(R^-_{2,3} - q)).
\end{aligned}
$$

We extract the average on the y-variables from the rest of the expression as pre-factors, so that the above is equal to

$$
\begin{aligned}
&- \lambda t a(1)\nu_0((R^-_{1,*} - q)(R^-_{1,2} - q)) + \lambda t a(0)\nu_0((R^-_{1,*} - q)^2)\\
&- \lambda t a(1)\nu_0((R^-_{1,*} - q)(R^-_{2,*} - q)) + \lambda t a(2)\nu_0((R^-_{1,*} - q)(R^-_{2,3} - q)).
\end{aligned}
$$

We notice that by the Nishimori property that

$$\nu_0((R^-_{1,*} - q)(R^-_{1,2} - q)) = \nu_0((R^-_{1,*} - q)(R^-_{2,*} - q)).$$

Now we observe that $\nu_0'((R^-_{1,*} - q)(x_N x^*_N - q))$ is a linear combination of terms that resemble, but are not quite equal to A, B and C. We are nevertheless tempted to make the substitution since we expect them to be close. We use Lemma 4.20 to justify this. Taking $\nu_0((R^-_{1,*} - q)^2)$ as an example, we apply the estimate (4.32) with $t = 1$, $\tau_1 = 3$ and $\tau_2 = 3/2$. We get

$$\nu_0((R^-_{1,*} - q)^2) = \nu((R^-_{1,*} - q)^2) + \delta$$

with $|\delta| \leq K(\lambda)\nu(|R^-_{1,*} - q|^3)$. Moreover,

$$\nu((R^-_{1,*} - q)^2) = \nu((R_{1,*} - \frac{1}{N}yy^* - q)^2) = \nu((R_{1,*} - q)^2) - \frac{2}{N}\nu(yy^*(R_{1,*} - q)) + \frac{1}{N^2}\nu(y^2 y^{*2}).$$

The third term is of order $1/N^2$, and the second term is bounded by $\frac{1}{N}\nu_0(|R_{1,*} - q|)$. Therefore

$$\nu_0((R^-_{1,*} - q)^2) = \nu((R_{1,*} - q)^2) + \delta',$$

with

$$|\delta'| \leq K(\lambda)\left(\frac{1}{N}\nu(|R^-_{1,*} - q|) + \nu(|R^-_{1,*} - q|^3) + \frac{1}{N^2}\right).$$

This argument applies equally to the remaining terms $\nu_0((R^-_{1,*} - q)(R^-_{2,*} - q))$ and $\nu_0((R^-_{1,*} - q)(R^-_{2,3} - q))$. We then end up with the identity

$$A = \frac{a(0)}{N} + \lambda' a(0)A - 2\lambda' a(1)B + \lambda' a(2)C + \delta(0), \tag{4.36}$$

where $\lambda' = t\lambda$, and $|\delta(0)|$ is bounded by the same quantity as $|\delta'|$.

Next, we apply the same reasoning to B and C as well, (e.g., Lemma 4.18 needs to applied with $n = 2$ for B and $n = 3$ for C) we get

$$B = \frac{a(1)}{N} + \lambda' a(1)A + \lambda'(a(0) - a(1) - 2a(2))B + \lambda'(-2a(1) + 3a(2))C + \delta(1), \tag{4.37}$$

$$C = \frac{a(2)}{N} + \lambda' a(2) A + \lambda'(4a(1) - 6a(2))B + \lambda'(a(0) - 6a(1) + 6a(2))C + \delta(2), \tag{4.38}$$

where for $i = 0, 1, 2$,

$$|\delta(i)| \leq K(\lambda) \left(\frac{1}{N} \nu(|R^-_{1,*} - q|) + \nu(|R^-_{1,*} - q|^3) + \frac{1}{N^2} \right). \tag{4.39}$$

We have ended up with a linear system in the quantities A, B and C. Let $\boldsymbol{z} = [A, B, C]^\top$ and $\boldsymbol{\delta} = [\delta(0), \delta(1), \delta(2)]^\top$. Then the equations (4.36), (4.37) and (4.38) can be written as

$$\boldsymbol{z} = \frac{1}{N}\boldsymbol{a} + t\boldsymbol{A}\boldsymbol{z} + \boldsymbol{\delta}, \tag{4.40}$$

where $\boldsymbol{a} = [a(0), a(1), a(2)]^\top$, and the matrix $\boldsymbol{A}$ are defined in (4.34). The above system implies useful bounds on the coefficients of the vector $\boldsymbol{z}$ only if the largest eigenvalue of the matrix $t\boldsymbol{A}$ is smaller than 1. This is insured by Lemma 4.21 when $\lambda \in \mathcal{A}$ (independently of t). Now we can invert the linear system and extract $\boldsymbol{z}$:

$$\boldsymbol{z} = \frac{1}{N}(\boldsymbol{I} - t\boldsymbol{A})^{-1}\boldsymbol{a} + (\boldsymbol{I} - t\boldsymbol{A})^{-1}\boldsymbol{\delta}. \tag{4.41}$$

Now we need to control the entries of $\boldsymbol{\delta}$. By elementary manipulations,

$$\nu(|R^-_{1,*} - q|) \leq \nu(|R_{1,*} - q|) + \frac{K}{N},$$

and

$$\nu(|R^-_{1,*} - q|^3) \leq \nu(|R_{1,*} - q|^3) + \frac{K}{N}\nu((R_{1,*} - q)^2) + \frac{K}{N^2}\nu(|R_{1,*} - q|) + \frac{K}{N^3}.$$

Therefore, from (4.39) we have for all $i = 0, 1, 2$,

$$|\delta(i)| \leq K \left(\nu(|R_{1,*} - q|^3) + \frac{1}{N}\nu((R_{1,*} - q)^2) + \frac{1}{N}\nu(|R_{1,*} - q|) + \frac{1}{N^2} \right). \tag{4.42}$$

Now we will argue that $\nu(|R_{1,*} - q|) \ll 1$ and $\nu(|R_{1,*} - q|^3) \ll \nu((R_{1,*} - q)^2)$. With Lemma 4.12 we have for $\epsilon > 0$

$$\nu(|R_{1,*} - q|) \leq \epsilon + K(\epsilon)e^{-cN},$$

and

$$\nu(|R_{1,*} - q|^3) \leq \epsilon\nu((R_{1,*} - q)^2) + K(\epsilon)e^{-cN}.$$

Combining the above two bounds with (4.42), and then injecting in (4.41), we get

$$\begin{aligned}\nu((R_{1,*} - q)^2) = z(0) \leq \|z\|_{\ell_2} &\leq \left\| \frac{1}{N}(\boldsymbol{I} - t\boldsymbol{A})^{-1}\boldsymbol{a} \right\|_{\ell_2} + \left\|(\boldsymbol{I} - t\boldsymbol{A})^{-1}\right\|_{\mathrm{op}} \|\boldsymbol{\delta}\|_{\ell_2} \\ &\leq \frac{\|\boldsymbol{c}\|_{\ell_2}}{N} + K(\epsilon + \frac{1}{N})\nu((R_{1,*} - q)^2) + K(\epsilon)e^{-cN}.\end{aligned}$$

The symbols $\|\cdot\|_{\ell_2}$ and $\|\cdot\|_{\mathrm{op}}$ refer to the ℓ_2 norm of a vector and the matrix operator norm respectively. Here, $\boldsymbol{c} = (\boldsymbol{I} - t\boldsymbol{A})^{-1}\boldsymbol{a}$. Note that the matrix inverses are bounded even as $t \to 1$ since $\mu_1 < 1$ for $\lambda \in \mathcal{A}$. We choose ϵ small enough and N large enough that $K(\epsilon + \frac{1}{N}) < 1$. We therefore get

$$\nu\left((R_{1,*} - q)^2\right) \le \frac{K(\lambda)}{N} + K(\lambda)e^{-c(t)N}.$$

Cavity computations for the fourth moment

In this subsection we prove the convergence of the fourth moment:

$$\nu((R_{1,*} - q^*)^4) \le \frac{K}{N^2} + Ke^{-c(t)N},$$

where $c(t)$ is of the same type as before. We adopt the same technique based on the cavity method, with the extra knowledge that the second moment converges. Many parts of the argument are exactly the same so we will only highlight the main novelties in the proof. Let

$$A = \nu\left((R_{1,*} - q)^4\right), \quad B = \nu\left((R_{1,*} - q)^3(R_{2,*} - q)\right), \quad C = \nu\left((R_{1,*} - q)^3(R_{2,3} - q)\right).$$

By symmetry between sites,

$$\begin{aligned} A &= \nu\left((R_{1,*} - q)^3(x_N x_N^* - q)\right) \\ &= \nu((R_{1,*}^- - q)^3(x_N x_N^* - q)) + \frac{3}{N}\nu((R_{1,*}^- - q)^2 x_N x_N^*(x_N x_N^* - q)) \\ &\quad + \frac{3}{N^2}\nu((R_{1,*}^- - q){x_N}^2 {x_N^*}^2(x_N x_N^* - q)) + \frac{1}{N^3}\nu({x_N}^3 {x_N^*}^3(x_N x_N^* - q)). \end{aligned}$$

The quadratic term is bounded as

$$\nu((R_{1,*}^- - q)^2 x_N x_N^*(x_N x_N^* - q)) \le K\nu((R_{1,*}^- - q)^2) \le \frac{K}{N} + Ke^{-cN}.$$

The last inequality is using our extra knowledge about the convergence of the second moment. The last two terms are also bounded by K/N^2 and K/N^3 respectively. Now we must deal with the cubic term, and here, we apply the exact same technique used to deal with the term $\nu((R_{1,*}^- - q)(x_N x_N^* - q))$ in the previous proof. The argument goes verbatim. Then we equally treat the terms B and C. We end up with a similar linear system relating A, B and C:

$$\boldsymbol{z} = \frac{1}{N^2}\boldsymbol{d} + t\boldsymbol{A}\boldsymbol{z} + \boldsymbol{\delta},$$

where $\boldsymbol{z} = [A, B, C]^\top$. The differences with the earlier linear system (4.40) are in the vector of coefficients $\boldsymbol{d}$ (that could be determined from the recursions) and the error terms $\delta(i)$, which are now bounded as

$$|\delta(i)| \le K\nu(|R_{1,*}^- - q|^5) + K\sum_{l=1}^{3} \frac{1}{N^{3-l}}\nu(|R_{1,*}^- - q|^l).$$

Crucially, the matrix $\boldsymbol{A}$ remains the same. Using Lemma 4.12, we have for $\epsilon > 0$,

$$\nu(|R_{1,*} - q|^5) \le \epsilon \nu((R_{1,*} - q)^4) + K(\epsilon)e^{-cN},$$
$$\nu(|R_{1,*} - q|^3) \le \epsilon \nu((R_{1,*} - q)^2) + K(\epsilon)e^{-cN}.$$

With the bound we already have on $\nu((R_{1,*} - q)^2)$, we finish the argument in the same way, by choosing ϵ sufficiently small. This concludes the proof of Theorem 4.8.

Sharper results: the asymptotic variance

Finally, given the convergence of the fourth moment, we can refine the convergence result of the second moment. Indeed we are now able to compute the limit of $N \cdot \nu((R_{1,*} - q)^2)$. Using Jensen's inequality on the second and fourth moment, we have

$$\nu(|R_{1,*} - q|) \le \frac{K}{\sqrt{N}} + Ke^{-c(t)N}, \quad \text{and} \quad \nu(|R_{1,*} - q|^3) \le \frac{K}{N^{3/2}} + Ke^{-c'(t)N}.$$

Looking back at (4.42), we can now assert that

$$|\delta(i)| \le \frac{K}{N^{3/2}} + Ke^{-c(t)N}.$$

We plug this new knowledge in (4.41), and obtain

$$\left\| N\boldsymbol{z} - (\boldsymbol{I} - t\boldsymbol{A})^{-1}\boldsymbol{a} \right\|_{\ell_2} \le N \left\| (\boldsymbol{I} - t\boldsymbol{A})^{-1} \right\|_{\text{op}} \|\boldsymbol{\delta}\|_{\ell_2} \le K\Big(\frac{1}{\sqrt{N}} + Ne^{-c(t)N}\Big).$$

The last line follows since $\sup_t \|(\boldsymbol{I} - t\boldsymbol{A})^{-1}\|_{\text{op}} \le K(\lambda)$ for $\lambda \in \mathcal{A}$. We have just proved that

$$\begin{aligned}
\nu((R_{1,*} - q)^2) &= \frac{c(0)}{N} + K(\lambda)\Big(\frac{1}{N^{3/2}} + e^{-c(t)N}\Big),\\
\nu((R_{1,*} - q)(R_{2,*} - q)) &= \frac{c(1)}{N} + K(\lambda)\Big(\frac{1}{N^{3/2}} + e^{-c(t)N}\Big),\\
\nu((R_{1,*} - q)(R_{2,3} - q)) &= \frac{c(2)}{N} + K(\lambda)\Big(\frac{1}{N^{3/2}} + e^{-c(t)N}\Big),
\end{aligned}$$

where $\boldsymbol{c} = (\boldsymbol{I} - t\boldsymbol{A})^{-1}\boldsymbol{a}$. One can solve this linear system explicitly and obtain the expression of the coordinates of $\boldsymbol{c}$:

$$\begin{aligned}
c(0) &= \frac{1}{\lambda t}\left(-1 + \frac{2}{1 - t\mu_2} + \frac{2}{1 - t\mu_1} + \frac{-3 + 3\lambda t a(0) - 2\lambda t a(1)}{(1 - t\mu_1)^2}\right),\\
c(1) &= \frac{1}{\lambda t}\left(\frac{-3 + 3\lambda t a(0) - 2\lambda t a(1)}{(1 - t\mu_1)^2} + \frac{3}{1 - t\mu_2}\right),\\
c(2) &= \frac{4\lambda t a(1)^2 + (1 - \lambda t a(0) - 5\lambda t a(1))a(2) + 2\lambda t a(2)^2}{(1 - t\mu_1)^2(1 - t\mu_2)}.
\end{aligned}$$

The expression of the first coordinate defines $\Delta_{\mathsf{RS}}(\lambda; t)$, equ.(4.21). This concludes the proof of Theorem 4.9.

Proof of Lemma 4.10

Let $f(x, x^*) = x^2 x^{*2}$. We have $\nu_0(f) = a(0)$. We use the first assertion of Lemma 4.20 with $\tau_1 = 1$ and $\tau_2 = \infty$ to get

$$|\nu(f) - \nu_0(f)| \leq K(\lambda)\nu(|R_{1,*}^{-} - q^*|) \leq \frac{K}{\sqrt{N}} + Ke^{-c(t)N},$$

where the last bound follows from Theorem 4.8 and Jensen's inequality.

Part II

Decoding a discrete signal from pooled data

Chapter 5

Decoding from pooled Data: sharp information-theoretic bounds

The theory of compressed sensing (Donoho, 2006a), where one is interested in recovering a signal from a few compressed measurements of it has grown into a rich field of investigation and found many applications. From the inception of this theory, it has been understood that the structure of the signal, typically sparsity, plays a key role in the sample complexity, or number of measurements needed for reconstruction (Candés, Romberg, and Tao, 2006; Candés and Tao, 2005; Donoho, 2006b). Here one usually considers a signal that is real-valued, and is compressed by taking random linear combinations of its entries. It is however interesting to move beyond this setting and consider signals that are discrete, where each entry can take a value from a finite alphabet. Then one possible model of compression—since the signal no longer has an additive structure—is to count the occurrence of each symbol in a randomly chosen subset of the signal's entries. Therefore, the measurements we consider are histograms of pooled subsets of the signal. This model is motivated by applications such as pooling of genetic data, on which we expand later on.

The discrete, combinatorial structure of this reconstruction problem makes it a special kind of a *constraint satisfaction problem* (CSP). These have been the object of intense study in recent years in probability theory, computer science, information theory and statistical physics. For certain families of CSPs, a deep understanding has begun to emerge regarding the number of solutions as a function of problem size, as well as the algorithmic feasibility of finding solutions when they exist (see e.g. Coja-Oghlan and Frieze, 2014; Coja-Oghlan, Haqshenas, and Hetterich, 2016; Coja-Oghlan, Mossel, and Vilenchik, 2009; Coja-Oghlan and Perkins, 2016; Ding, Sly, and Sun, 2015, 2016; Sly, Sun, and Y. Zhang, 2016)). Consider in particular a *planted* random constraint satisfaction problem with n variables that take their values in the discrete set $\{1, \cdots, d\}$, with $d \geq 2$. A number of m clauses is drawn uniformly at random under the constraint that they are all satisfied by a pre-specified assignment, which is referred to as *the planted solution*. In our case, the signal is n-dimensional, d is the size of the alphabet, and there are m compressed observations (histograms) of the signal, which represents the planted solution that satisfies all the constraints.

Two questions are of particular importance: *(1) how large should m be so that the planted solution is the unique solution?* and *(2) given that it is unique, how large should m be so that it is recoverable by a "tractable" algorithm?* Significant progress has been made on these questions, often initiated by insights from statistical physics and followed by a growing body of rigorous mathematical investigation. The emerging picture is that in many planted CSPs, when n is sufficiently large, all solutions become highly correlated with the planted one when $m > \kappa_{\mathsf{IT}} \cdot n$, for some "Information-Theoretic" (IT) constant $\kappa_{\mathsf{IT}} > 0$. Furthermore, one of these highly correlated solutions becomes typically recoverable by a random walk or a Belief Propagation (BP)-inspired algorithm when $m > \kappa_{\mathsf{BP}} \cdot n$ for some $\kappa_{\mathsf{BP}} > \kappa_{\mathsf{IT}}$ (Coja-Oghlan and Frieze, 2014; Coja-Oghlan, Mossel, and Vilenchik, 2009; Krzakala, Mézard, and Zdeborová, 2012; Krzakala and Zdeborová, 2009). Interestingly, it is known in many problems, at least heuristically, that these algorithms fail when $\kappa_{\mathsf{IT}} < m/n < \kappa_{\mathsf{BP}}$, and a tractable algorithm that succeeds in this regime is still lacking (Achlioptas and Coja-Oghlan, 2008; Coja-Oghlan, 2009; Coja-Oghlan, Haqshenas, and Hetterich, 2016; Zdeborová and Krzakala, 2016). In other words, there is a non-trivial regime $m/n \in (\kappa_{\mathsf{IT}}, \kappa_{\mathsf{BP}})$ where an essentially unique solution exists, but is hard to recover.

For the random CSP we consider in this chapter and the next, which we call the Histogram Query Problem (HQP), we undertake a detailed information-theoretic analysis which shows that the planted solution becomes unique at as soon as $m > \gamma^* n/\log n$ with high probability as $n \to \infty$ for an explicit constant $\gamma^* = \gamma^*(d) > 0$. In the next chapter, we consider the algorithmic aspect of the problem and provide a BP-based algorithm that recovers the planted assignment if $m \geq \kappa^* \cdot n$ for a specific threshold κ^* and fails otherwise. This leaves a logarithmic gap between the information-theoretic threshold and the point at which our algorithm succeeds.

5.1 Problem and motivation

The setting Let $\{\boldsymbol{h}_a\}_{1\leq a\leq m}$ be a collection of d-dimensional arrays with non-negative integer entries. For an assignment $\tau : \{1, \cdots, n\} \mapsto \{1, \cdots, d\}$ of the n variables, and given a realization of m random subsets $S_a \subset \{1, \cdots, n\}$, the constraints of the HQP are given by $\boldsymbol{h}_a = \boldsymbol{h}_a(\tau)$ for all $1 \leq a \leq m$ with

$$\boldsymbol{h}_a(\tau) := \left(\left|\tau^{-1}(1) \cap S_a\right|, \cdots, \left|\tau^{-1}(d) \cap S_a\right|\right) \in \mathbb{Z}_+^d.$$

We let $\tau^* : \{1, \cdots, n\} \mapsto \{1, \cdots, d\}$ be a planted assignment; i.e., we set $\boldsymbol{h}_a := \boldsymbol{h}_a(\tau^*)$ for all a for some realization of the sets $\{S_a\}$, and consider the problem of recovering the map τ^* given the observation of the arrays $\{\boldsymbol{h}_a\}_{1\leq a\leq m}$.

This problem can be viewed informally as that of decoding a discrete high-dimensional signal consisting of categorical variables from a set of measurements formed by pooling together the variables belonging to a subset of the signal. It is useful to think of the n variables as each describing the type or category of an individual in a population of size n, where each individual has exactly one type among d. For instance the categories may represent blood types or some other discrete feature such as ethnicity or age group. Then, the observation $\boldsymbol{h}_a$ is the histogram of types of a subpopulation S_a. We let $\boldsymbol{\pi} = \frac{1}{n}\left(\left|\tau^{*-1}(1)\right|, \cdots, \left|\tau^{*-1}(d)\right|\right)$ denote the vector of proportions of assigned values; i.e., the empirical distribution of categories.

We consider here a model in which each variable participates in a given constraint independently and with probability $\alpha \in (0,1)$. Thus, the sets $\{S_a\}_{1\leq a\leq m}$ are independent draws of a random set S where $\Pr(i \in S) = \alpha$ independently for each $i \in \{1,\dots,n\}$. We are thus in the "dense regime" where $\mathbb{E}[|S|] = \alpha n$; i.e., the number of variables participating in each constraint (the degree of each factor in the CSP) is linear in n.

Motivation This model is inspired by practical problems in which a data analyst can only assay certain summary statistics involving a moderate or large number of participants. This may be done for privacy reasons, or it may be inherent in the data-collection process (see e.g. (Heo et al., 2001; Sham et al., 2002)). For example, in DNA assays, the pooling of allele measurements across multiple strands of DNA is necessary given the impracticality of separately analyzing individual strands. Thus the data consists of a frequency spectrum of alleles; a "histogram" in our language. In the privacy-related situation, one may take the viewpoint of an attacker whose goal is to gain a granular knowledge of the database from coarse measurements, or that of a guard who wishes to prevent this scenario from happening. It is then natural to ask how many histogram queries it takes to exactly determine the category of each individual.

Related problems Note that the case $d = 2$ of HQP can be seen as a compressed sensing problem with a binary sensing matrix and binary signal. While the bulk of the literature in the field of compressed sensing is devoted to the case in which both the signal of interest and the sensing matrix are real-valued, the binary case has also been considered, notably in relation to Code Division Multiple Access (CDMA) (Tanaka, 2002; Zigangirov, 2004), and Group Testing (Du and Hwang, 2006; Mézard and C. Toninelli, 2011): in the latter, one observes the logical "OR" of subsets of the entries of the signal. In the case of categorical variables with $d \geq 3$ categories, it is natural to consider measurements consisting of histograms of the categories in the pooled sub-population. In the literature on compressed sensing one commonly considers the setting where the sensing matrices have i.i.d. entries with finite second moment, and the signal has an arbitrary empirical distribution of its entries. It has been established that, under the scaling $m = \kappa n$, whereas the success of message-passing algorithms requires $\kappa > \kappa_{\mathsf{BP}}$ (Bayati, Lelarge, and Montanari, 2015), the information-theoretic threshold is $\kappa_{\mathsf{IT}} = 0$ in the discrete signal case (Donoho, Javanmard, and Montanari, 2013; Wu and Verdú, 2009), indicating that uniqueness of the solution happens at a finer scale $m = o(n)$. Here we consider the HQP with arbitrary d, for which the exact scaling for investigating uniqueness is $m = \gamma\frac{n}{\log n}$ with finite $\gamma > 0$, and provide tight bounds on the information-theoretic threshold.

Prior work on HQP The study of this problem for generic values of d was initiated in (Wang et al., 2016) in the two settings where the sets $\{S_a\}$ are deterministic and random. They showed in both these cases with a simple counting argument that under the condition that $\boldsymbol{\pi}$ is the uniform distribution, if $m < \frac{\log d}{d-1} \frac{n}{\log n}$ then the set of collected histograms does not uniquely determine the planted assignment τ^* (with high probability in the random case). On the other hand, for the deterministic setting, they provided a querying strategy that recovers τ^* provided that $m > c_0 \frac{n}{\log n}$, where c_0 is an absolute constant independent of d. For the random setting and under the condition that the sets S_a are of average size $n/2$, they proved via a first moment bound that $m > c_1 \frac{n}{\log n}$ with c_1 also constant and independent of d, suffices to uniquely determine τ^*, although no algorithm was proposed in this setting.

In the above results, there is a gap that is both information-theoretic and algorithmic depending on the dimension d between the upper and lower bounds. Intuitively, the upper bounds should also depend on d since the decoding problem becomes easier (or at least, it is no harder) for large d, for the simple reason that if it is possible to determine the categories of the population for $d = 2$, then one can proceed by dichotomy for larger d by merging the d groups into two super-groups, identifying which individuals belong to each of the two super-groups, and then recurse. We attempt to fill the information-theoretic gap in the random setting by providing tighter upper and lower bounds on the number of queries m necessary and sufficient to uniquely determine the planted assignment τ^* with high probability, which depend on the dimension d and $\boldsymbol{\pi}$ along with explicit constants. In the next chapter, we consider the algorithmic aspect of the problem and provide a Belief Propagation-based algorithm that recovers the planted assignment if $m \geq \kappa^*(\boldsymbol{\pi}, d) \cdot n$ for a specific threshold $\kappa^*(\boldsymbol{\pi}, d)$ and fails otherwise, indicating the putative existence of a statistical-computational gap in the random setting.

5.2 The uniqueness threshold

Let Δ^{d-1} be the $d-1$-dimensional simplex and $H(\boldsymbol{x}) = -\sum_{r=1}^{d} x_r \log x_r$ for $\boldsymbol{x} \in \Delta^{d-1}$ be the Shannon entropy function. We write $\tau \sim \boldsymbol{\pi}$ to indicate that τ is a random assignment drawn from the uniform distribution over maps $\tau : \{1, \cdots, n\} \mapsto \{1, \cdots, d\}$ such that $\frac{1}{n} \left(|\tau^{-1}(1)|, \cdots, |\tau^{-1}(d)|\right) = \boldsymbol{\pi}$.

Theorem 5.1. *For $n \geq 2$ integer, $m = \gamma \frac{n}{\log n}$, $\gamma > 0$, $\alpha \in (0,1)$, and $\boldsymbol{\pi} \in \Delta^{d-1}$ with entries bounded away from 0 and 1. Let $\mathcal{E}$ be the event that τ^* is* not *the unique satisfying assignment to* HQP*:*

$$\mathcal{E} = \left\{\exists \tau \in \{1, \cdots, d\}^n \ : \ \tau \neq \tau^*, \ \boldsymbol{h}_a(\tau) = \boldsymbol{h}_a(\tau^*) \ \forall a \in \{1, \cdots, m\}\right\}.$$

(i) If

$$\gamma < \gamma_{low} := \frac{H(\boldsymbol{\pi})}{d-1},$$

then

$$\lim_{n\to\infty} \mathbb{E}_{\tau^*\sim\boldsymbol{\pi}} \Pr(\mathcal{E}) = 1.$$

(ii) On the other hand, let $\boldsymbol{\pi}_{[\cdot]}$ *be the vector of order statistics of* $\boldsymbol{\pi}$*:* $\pi_{[1]} \geq \pi_{[2]} \geq \cdots \geq \pi_{[d]}$*. For* $1 \leq k \leq d-1$*, let* $\boldsymbol{\pi}^{(k)} \in \Delta^{k-1}$ *be defined as* $\pi_1^{(k)} = \sum_{r=1}^{d-k+1} \pi_{[r]}$ *and* $\pi_l^{(k)} = \pi_{[d-k+l]}$ *for all* $2 \leq l \leq k$ *(if* $k \geq 2$*). If*

$$\gamma > \gamma_{up} := 2 \max_{1\leq k\leq d-1} \frac{H(\boldsymbol{\pi}) - H(\boldsymbol{\pi}^{(k)})}{d-k},$$

then

$$\lim_{n\to\infty} \mathbb{E}_{\tau^*\sim\boldsymbol{\pi}} \Pr(\mathcal{E}) = 0.$$

Remarks and special cases:

- For $d = 2$, $\gamma_{\text{up}} = 2H(\boldsymbol{\pi}) = 2\gamma_{\text{low}}$.
- If $\boldsymbol{\pi} = (\frac{1}{d}, \cdots, \frac{1}{d})$, or more generally, if $\boldsymbol{\pi}$ is such that $k = 1$ maximizes the expression defining γ_{up} then $\gamma_{\text{up}} = 2\frac{H(\boldsymbol{\pi})}{d-1} = 2\gamma_{\text{low}}$.
- The resulting bounds do not depend on α as long as it is fixed and bounded away from 0 and 1. Its contribution in the problem is sub-dominant and vanishes as $n \to \infty$ under the scaling considered here.
- The number k in the expression of γ_{up} can be interpreted as the number of connected components of a graph on d vertices that depends on the overlap structure of the two assignments τ and τ^*, and induces “maximum confusion” between them. This will become clear in latter sections.

After this result appeared on the preprint server arXiv in Nov. 2016, Scarlett and Cevher (2017) showed that the upper bound (ii) is actually tight, in the sense that one can now replace γ_{low} by γ_{up} in the statement of the lower bound (i). Consequently, the uniqueness of the planted assignment τ^* undergoes a sharp phase transition exactly at $\gamma = \gamma_{\text{up}}$. Their result combined with ours show that HQP has a rather unusual feature; namely, it is an example of a planted CSP where a plain first moment method identifies the exact satisfiability threshold with no conditioning needed.

The proof of the above Theorem occupies the rest of this chapter.

Main ideas of the proof

Our main contribution is the second part of Theorem 5.1, which establishes an upper bound on the uniqueness threshold of the random CSP with histogram constraints HQP. The proof uses the first moment method to upper bound the probability of existence of a non-planted solution. Since we are in a planted model, the analysis of the first moment ends up bearing many similarities with a second moment computation in a purely random (non-planted) model. Although second moment computations often require approximations, for the HQP it turns out that we are able to compute the exact annealed free energy of the model in the thermodynamic limit. That is, letting $\mathcal{Z}$ be the number of solutions of the CSP, we show that the limit

$$\mathfrak{F}(\gamma) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[\mathcal{Z} - 1\right]$$

exists and we compute its value exactly. Then the value of the threshold γ_{up} is obtained by locating the first point at which $\mathfrak{F}$ becomes negative:

$$\gamma_{\text{up}} = \inf\left\{\gamma > 0 \;:\; \mathfrak{F}(\gamma) < 0\right\}.$$

Together with the fact that $\mathfrak{F}$ is a monotone function, which will become clear once $\mathfrak{F}$ is computed, it is clear that for any $\gamma > \gamma_{\text{up}}$, $\mathbb{E}[\mathcal{Z} - 1]$ decays exponentially with n when the latter is sufficiently large.

This general strategy has been successfully pursued for a range of CSPs, such as K-SAT, NAE-SAT, and Independent Set, most of which are Boolean. For larger domain sizes, in order to carry out the second moment method one needs fine control of the overlap structure between the planted and a candidate solution. This control is at the core of the difficulty that arises in any second moment computation. To obtain such control, researchers have often imposed additional assumptions, at a cost of a weakening of the resulting bounds. For example, existing proofs for Graph Coloring and similar problems assume certain balancedness conditions (the overlap matrix needs to be close to doubly stochastic.) without which the annealed free energy cannot be computed (Achlioptas and Moore, 2004; Achlioptas and Naor, 2005; Banks, Moore, Neeman, et al., 2016; Bapst et al., 2016; Coja-Oghlan, Efthymiou, and Hetterich, 2016); this yields results that fall somewhat short of the bounds that the second moment method could achieve in principle (Dani, Moore, and Olson, 2012). In the present problem, due its rich combinatorial structure, we are able to obtain unconditional control of the overlap structure, for any domain size d, and compute the exact annealed free energy.

Concretely, computing the function $\mathfrak{F}$ requires tight control of the "collision probability" of two non-equal assignments τ_1 and τ_2. This is the probability that the random histograms $\boldsymbol{h}(\tau_1) = (|{\tau_1}^{-1}(1) \cap S|, \cdots, |{\tau_1}^{-1}(d) \cap S|)$ and $\boldsymbol{h}(\tau_2) = (|{\tau_2}^{-1}(1) \cap S|, \cdots, |{\tau_2}^{-1}(d) \cap S|)$ generated from a random draw of a pool S coincide. The collision probability roughly measures the correlation strength between the two assignments. Specifically, we will be interested in the collision probabilities of the pairs (τ^*, τ) where τ^* is the planted assignment and τ is any candidate assignment. Its decay reveals how long an assignment τ "survives" as a satisfying assignment to HQP as $n \to \infty$. The study of these collision probabilities requires the evaluation of certain Gaussian integrals over the space of *Eulerian flows* of a weighted graph on d vertices that is defined based on the overlap structure of τ and τ^*. We prove a family of identities that relate these integrals to some combinatorial polynomials in the weights of the graph: the spanning tree and spanning forest polynomials. We believe that these identities are of independent interest beyond the problem presently studied. Once these collision probabilities are controlled, the computation of $\mathfrak{F}(\gamma)$ per se requires the analysis of a certain sequence of optimization problems. We show that the sequence of maximum values converges to a finite limit that yields the value of the annealed free energy.

On the other hand, the proof of the first part of Theorem 5.1 is straightforward—it is an extension of a standard counting argument used in (P. Zhang et al., 2013) and (Wang et al., 2016). The argument goes as follows: if m is too small then the number of possible histograms one could potentially observe is exponentially smaller than the number of assignments of n variables that agree with $\boldsymbol{\pi}$. Therefore when the planted assignment τ^* is drawn at random, there will exist at least one $\tau \neq \tau^*$ that satisfies the constraints of the CSP with overwhelming probability. We begin with this argument in the next section and then turn to the more challenging computation of the upper bound.

5.3 Proof of the main result

In this section we prove Theorem 5.1.

Notation We denote vectors in $\mathbb{R}^d$ in bold lower case letters, e.g., $\boldsymbol{x}$, and matrices in $\mathbb{R}^{d\times d}$ will be written in bold lower case underlined letters, e.g., $\boldsymbol{x}$. We denote the coordinates of such vectors and matrices as x_r and x_{rs} respectively. Matrices that act either as linear operators on the space $\mathbb{R}^{d\times d}$ or that are functions of elements in this space are written in bold upper case letters, e.g., $\boldsymbol{M}\boldsymbol{x}$ and $\boldsymbol{L}(\boldsymbol{x})$, for $\boldsymbol{x} \in \mathbb{R}^{d\times d}$. These choices will be clear from the context. We may write $\boldsymbol{x}/\boldsymbol{y}$ to indicate coordinate-wise division. Additionally, for two $d \times d$ matrices $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{d\times d}$, $\boldsymbol{a} \odot \boldsymbol{b} \in \mathbb{R}^{d\times d}$ is their Hadamard product. We let $\mathbf{1} \in \mathbb{R}^d$ be the all-ones vector.

The first part of Theorem 5.1: the lower bound

Let $m = \gamma \frac{n}{\log n}$ with $\gamma > 0$. The number of potential histograms one could possibly observe in a single query with pool size $|S| = k$ is $f(k,d) := \binom{d+k-1}{d-1} \leq (k+1)^{d-1}$. Since the queries are independent, the number of collections of histograms $\{\boldsymbol{h}_a\}_{1\leq a\leq m}$ one could potentially observe in m queries is $\prod_{a=1}^{m} f(|S_a|, d)$. On the other hand, the number of possible assignments $\tau : \{1, \cdots, n\} \mapsto \{1, \cdots, d\}$ satisfying the constraint $\boldsymbol{\pi} = \frac{1}{n}\left(\left|\tau^{*-1}(1)\right|, \cdots, \left|\tau^{*-1}(d)\right|\right)$ is $\binom{n}{n\boldsymbol{\pi}} = \binom{n}{n\pi_1, \cdots, n\pi_d} \geq C(\boldsymbol{\pi}) n^{-(d-1)/2} \exp(H(\boldsymbol{\pi})n)$, for some constant $C(\boldsymbol{\pi}) > 0$ depending on $\boldsymbol{\pi}$.

Now, the probability that τ^* is the unique satisfying assignment of the CSP with constraints given by the random histograms $\{\boldsymbol{h}_a(\tau^*)\}_{1\leq a\leq m}$, averaged over the random choice of $\tau^* \sim \boldsymbol{\pi}$, is

$$\begin{aligned}
&\mathbb{E}_{\tau^*\sim\boldsymbol{\pi}}\, \mathbb{E}_{\{S_a\}}\Big[\mathbb{1}\{\forall \tau \in \{1, \cdots, d\}^n \; : \; \boldsymbol{h}_a(\tau) = \boldsymbol{h}_a(\tau^*) \; \forall a \in \{1, \cdots, m\} \quad \Longrightarrow \quad \tau = \tau^*\}\Big] \\
&\qquad \leq \binom{n}{n\boldsymbol{\pi}}^{-1} \cdot \mathbb{E}_S\left[f(|S|, d)\right]^m
\end{aligned}$$

$$\begin{aligned}
&\le \binom{n}{n\boldsymbol{\pi}}^{-1} \cdot \mathbb{E}_S \left[(|S|+1)^{d-1}\right]^m \\
&\le C(\boldsymbol{\pi})\ n^{(d-1)/2} \cdot \exp\Big(-H(\boldsymbol{\pi})n\Big) \cdot (n+1)^{m(d-1)} \\
&\le C(\boldsymbol{\pi})\ n^{(d-1)/2} \cdot \exp\Big((\gamma(d-1) - H(\boldsymbol{\pi}))n\Big).
\end{aligned}$$

If $\gamma < \gamma_{\text{low}}$ the last quantity tends to 0 as $n \to \infty$. This concludes the proof of the first assertion of the theorem.

The second part of Theorem 5.1 : the upper bound

We use a first moment method to show that when γ is greater than γ_{up}, the only assignment satisfying HQP is τ^* with high probability. Let $\mathcal{Z}$ be the number of satisfying assignments to HQP:

$$\mathcal{Z} := \Big|\{\tau \in \{1, \cdots, d\}^n \ :\ \boldsymbol{h}_a(\tau) = \boldsymbol{h}_a(\tau^*)\ \forall a \in \{1, \cdots, m\}\}\Big|. \tag{5.1}$$

The planted assignment τ^* is obviously a solution, so we always have $\mathcal{Z} \ge 1$. Recall the definition of the annealed free energy

$$\mathfrak{F}(\gamma) := \lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\left[\mathcal{Z} - 1\right]. \tag{5.2}$$

Also, recall that for $1 \le k \le d-1$, $\boldsymbol{\pi}^{(k)} \in \Delta^{k-1}$ be defined as $\pi_1^{(k)} = \sum_{r=1}^{d-k+1} \pi_{[r]}$ and $\pi_l^{(k)} = \pi_{[d-k+l]}$ for all $2 \le l \le k$ (if $k \ge 2$).

Theorem 5.2. *Let $m = \gamma \frac{n}{\log n}$ with $\gamma > 0$. The limit (5.2) exists for all $\gamma > 0$ and its value is*

$$\mathfrak{F}(\gamma) = \max_{1\le k\le d-1} \Big\{H(\boldsymbol{\pi}) - H(\boldsymbol{\pi}^{(k)}) - \frac{\gamma}{2}(d-k)\Big\}. \tag{5.3}$$

We can deduce from Theorem 5.2 the smallest value of γ past which $\mathfrak{F}(\gamma)$ becomes negative. In particular, we see that $\mathfrak{F}$ is a decreasing function of γ that crosses the horizontal axis at

$$\gamma_{\text{up}} = 2 \max_{1\le k\le d-1} \frac{H(\boldsymbol{\pi}) - H(\boldsymbol{\pi}^{(k)})}{d-k}.$$

From this result it is easy to prove the second assertion of Theorem 5.1. By averaging over τ^* and applying Markov's inequality, we have:

$$\begin{aligned}
\mathbb{E}_{\tau^*\sim\boldsymbol{\pi}} \Pr\left(\exists \tau \in \{1, \cdots, d\}^n \ :\ \tau \ne \tau^*, \boldsymbol{h}_a(\tau) = \boldsymbol{h}_a(\tau^*)\ \forall a \in \{1, \cdots, m\}\right) \\
= \mathbb{E}_{\tau^*\sim\boldsymbol{\pi}} \Pr\left(\mathcal{Z} \ge 2\right) \le \mathbb{E}[\mathcal{Z} - 1].
\end{aligned}$$

For $\gamma > \gamma_{\text{up}}$, it is clear that $\mathfrak{F}(\gamma) < 0$. Let $0 < \epsilon < |\mathfrak{F}(\gamma)|/2$; then there is an integer $n_0(\epsilon) \ge 0$ such that for all $n \ge n_0(\epsilon)$,

$$\mathbb{E}_{\tau^*\sim\boldsymbol{\pi}} \Pr\left(\exists \tau \in \{1, \cdots, d\}^n \ :\ \tau \ne \tau^*, \boldsymbol{h}_a(\tau) = \boldsymbol{h}_a(\tau^*)\ \forall a \in \{1, \cdots, m\}\right) \le \exp n\ \left(\mathfrak{F}(\gamma) + \epsilon\right),$$

$$\leq \exp n\ \mathfrak{F}(\gamma)/2, \\ \xrightarrow[n\to\infty]{} 0.$$

Now it remains to prove Theorem 5.2, and this represents the main technical thrust of our approach.

Collisions, overlaps, and the first moment

Preliminaries We begin by presenting the main quantities to be analyzed in our application of the first moment method. We have

$$\begin{aligned}\mathbb{E}_{\tau^*\sim\boldsymbol{\pi}}\,\mathbb{E}_{\{S_a\}}[\mathcal{Z}-1] &= \mathbb{E}_{\tau^*\sim\boldsymbol{\pi}}\left[\sum_{\substack{\tau\in\{1,\cdots,d\}^n\\ \tau\neq\tau^*}} \Pr\left(\boldsymbol{h}_a(\tau)=\boldsymbol{h}_a(\tau^*)\ \forall a\in\{1,\cdots,m\}\right)\right]\\ &= (d^n-1)\Pr_{\tau,\tau^*,\{S_a\}}\left(\boldsymbol{h}_a(\tau)=\boldsymbol{h}_a(\tau^*)\ \forall a\in\{1,\cdots,m\}\right),\end{aligned}$$

where $\tau^*\sim\boldsymbol{\pi}$, $\tau\sim\mathrm{Unif}(\{1,\cdots,d\}^n\backslash\{\tau^*\})$. By conditional independence,

$$\begin{aligned}\Pr_{\tau,\tau^*,\{S_a\}}\left(\boldsymbol{h}_a(\tau)=\boldsymbol{h}_a(\tau^*)\ \forall a\in\{1,\cdots,m\}\right) &= \mathbb{E}_{\tau,\tau^*}\left[\Pr_{\{S_a\}}\left(\boldsymbol{h}_a(\tau)=\boldsymbol{h}_a(\tau^*)\ \forall a\in\{1,\cdots,m\}\right)\right]\\ &= \mathbb{E}_{\tau,\tau^*}\left[\Pr_S\left(\boldsymbol{h}(\tau)=\boldsymbol{h}(\tau^*)\right)^m\right].\end{aligned}$$

Next, we write the *collision probability*, $\Pr_S\left(\boldsymbol{h}(\tau)=\boldsymbol{h}(\tau^*)\right)$, for fixed τ and τ^* in a convenient form. Let us first define the *overlap matrix*, $\boldsymbol{\mu}(\tau,\tau^*)=(\mu_{rs})_{1\leq r,s\leq d}\in\mathbb{Z}_+^{d\times d}$, of τ and τ^*, by

$$\mu_{rs}=\left|\tau^{-1}(r)\cap\tau^{*-1}(s)\right| \quad \text{for all } r,s=1,\cdots,d. \tag{5.4}$$

Remark that $\boldsymbol{h}(\tau)=\boldsymbol{h}(\tau^*)$ if and only if $|S\cap\tau^{-1}(r)|=\left|S\cap\tau^{*-1}(r)\right|$ for all $r\in\{1,\cdots,d\}$. Since the collection of sets $\{\tau^{-1}(r)\}_{1\leq r\leq d}$ forms a partition of $\{1,\cdots,n\}$, and similarly with τ^*, the event $\{\boldsymbol{h}(\tau)=\boldsymbol{h}(\tau^*)\}$ is the same as

$$\left\{\sum_{s=1}^d\left|S\cap\tau^{-1}(r)\cap\tau^{*-1}(s)\right|=\sum_{s=1}^d\left|S\cap\tau^{-1}(s)\cap\tau^{*-1}(r)\right|,\ \forall r\in\{1,\cdots,d\}\right\}.$$

Therefore, the probability that two assignments τ and τ^* collide on a random pool S—meaning that their histograms formed on the pool S coincide—is

$$\Pr_S\left(\boldsymbol{h}(\tau)=\boldsymbol{h}(\tau^*)\right)=\sum_{\boldsymbol{\nu}}\left(\prod_{r,s=1}^d\binom{\mu_{rs}}{\nu_{rs}}\alpha^{\nu_{rs}}(1-\alpha)^{\mu_{rs}-\nu_{rs}}\right)\mathbb{1}\left\{\sum_{s=1}^d\nu_{rs}=\sum_{s=1}^d\nu_{sr}\ ,\ \forall r\in[d]\right\}, \tag{5.5}$$

where the outer sum is over all arrays of integer numbers $\boldsymbol{\nu} = (\nu_{rs})_{1\leq r,s\leq d}$ such that $0 \leq \nu_{rs} \leq \mu_{rs}$ for all r, s. We see from the above expression that the collision probability of τ and τ^* only depends on the overlap matrix $\boldsymbol{\mu}(\tau, \tau^*)$. We henceforth denote the probability in equation (5.5) by $q(\boldsymbol{\mu})$, where we dropped the dependency on τ and τ^*. Remark that $\tau = \tau^*$ if and only if their overlap matrix $\boldsymbol{\mu}$ is diagonal. Thus, we can rewrite the expected number of solutions as

$$\mathbb{E}[\mathcal{Z} - 1] = \binom{n}{n\boldsymbol{\pi}}^{-1} \cdot \sum_{\boldsymbol{\mu}} \binom{n}{\boldsymbol{\mu}} q(\boldsymbol{\mu})^m \, \mathbb{1}\left\{\sum_{r=1}^{d} \mu_{rs} = n\pi_s, \; s \in \{1, \cdots, d\}\right\}, \tag{5.6}$$

where the sum is over all non-diagonal arrays $\boldsymbol{\mu} = (\mu_{rs})_{1\leq r,s\leq d}$ with non-negative integer entries that sum to n, and $\binom{n}{\boldsymbol{\mu}} = \frac{n!}{\prod_{r,s}\mu_{rs}!}$.

The rest of the proof From here, the proof of Theorem 5.2 roughly breaks into three parts:

(i) One needs to have tight asymptotic control on the collision probability $q(\boldsymbol{\mu})$ when any subset of the entries of $\boldsymbol{\mu}$ becomes large. This will be achieved via the Laplace method (see, e.g., (De Bruijn, 1970)). The outcome of this analysis is an asymptotic estimate that exhibits two different speeds of decay, polynomial or exponential, depending on the "balancedness" of $\boldsymbol{\mu}$ as its entries become large. This notion of balancedness, namely that $\boldsymbol{\mu}$ must have equal row- and column-sums[1], is specific to the histogram setting and departs from the usual "double stochasticity" that arises in other more classical problems such as Graph Coloring, and Community Detection under the stochastic block model (Achlioptas and Moore, 2004; Achlioptas and Naor, 2005; Banks, Moore, Neeman, et al., 2016; Bapst et al., 2016; Coja-Oghlan, Efthymiou, and Hetterich, 2016). As we will explain in the next section, configurations (τ, τ^*) with an unbalanced overlap matrix have an exponentially decaying collision probability, i.e., they exhibit weak correlation, and disappear very early on as $n \to \infty$ under the scaling $m = \gamma\frac{n}{\log n}$. On the other hand, those configurations with balanced overlap exhibit a slow decay of correlation: their collision probability decays only polynomially, and these are the last surviving configurations in expression (5.6) as $n \to \infty$.

(ii) Understanding the above-mentioned polynomial decay of $q(\boldsymbol{\mu})$ requires the evaluation of a multivariate Gaussian integral (which is a product of the above analysis) over the space of constraints of the array $\boldsymbol{\nu}$ in (5.5); the latter being the space of *Eulerian flows* on the graph on d vertices whose edges are weighted by the (large) entries of $\boldsymbol{\mu}$. We show that this integral, properly normalized, evaluates to *the inverse square root of the spanning tree (or forest) polynomial* of this graph. This identity seems to be new, to the best of our knowledge, and may be of independent interest. We therefore provide two very different proofs of it, each highlighting different combinatorial aspects.

(iii) Lastly, armed with these estimates, we show the existence of, and compute the exact value of, the annealed free energy of the model in the thermodynamic limit, thereby

[1] These are exactly the constraints on $\boldsymbol{\nu}$ showing up in (5.5).

completing the proof of Theorem 5.2. This last part requires the analysis of a certain optimization problem involving an entropy term and an "energy" term accounting for the correlations discussed above. Here we can exactly characterize the maximizing configurations for large n, and this allows the computation of the value of $\mathfrak{F}(\gamma)$. We note once more that this situation contrasts with the more traditional case of Graph Coloring, where we lack a rigorous understanding of the maximizing configurations of the second moment, except when certain additional constraints are imposed on their overlap matrix.

5.4 Bounding the collision probabilities

Here we provide tight asymptotic bounds on the collision probabilities $q(\boldsymbol{\mu})$ defined in (5.5). Consider the following subspace of $\mathbb{R}^{d\times d}$, which will play a key role in the analysis:

$$\mathcal{F} := \left\{ \boldsymbol{x} \in \mathbb{R}^{d\times d} \ : \ \sum_{s=1}^{d} x_{rs} = \sum_{s=1}^{d} x_{sr} \ , \ \forall r \in \{1, \cdots, d\} \right\}. \tag{5.7}$$

This is a linear subspace of dimension $(d-1)^2 + d$ in $\mathbb{R}^{d\times d}$. For $p, q \in (0,1)$, let $D(p \parallel q) = p\log(p/q) + (1-p)\log((1-p)/(1-q))$ be the Kullback-Leibler divergence. Let $G = (V, E)$ be an undirected graph on d vertices where we allow up to two parallel edges between each pair of vertices, i.e., $V = \{1, \cdots, d\}$, and $E \subseteq \{(r,s) \ : \ r, s \in V, \ r \neq s\}$. For $\boldsymbol{\nu}, \boldsymbol{\mu} \in \mathbb{R}_+^{d\times d}$, $\boldsymbol{x} \in [0,1]^{d\times d}$ let

$$\varphi_{\boldsymbol{\mu}}(\boldsymbol{x}) := \sum_{(r,s)\in E} \mu_{rs} D(x_{rs} \parallel \alpha). \tag{5.8}$$

and recalling that $\odot$ represents the Hadamard product, we let

$$\vartheta(\boldsymbol{\nu}, \boldsymbol{\mu}) := \min_{\substack{\boldsymbol{x}\in[0,1]^{d\times d} \\ \boldsymbol{M}_G(\boldsymbol{x}\odot\boldsymbol{\mu}, \boldsymbol{\nu})\in\mathcal{F}}} \sum_{(r,s)\in E} \mu_{rs} D(x_{rs} \parallel \alpha), \tag{5.9}$$

where for two $d \times d$ matrices $\boldsymbol{a}, \boldsymbol{b}$, $\boldsymbol{M}_G(\boldsymbol{a}, \boldsymbol{b})$ is the $d \times d$ matrix with entries a_{rs} if $(r,s) \in E$ and b_{rs} otherwise. By strong duality (see, e.g., (Boyd and Vandenberghe, 2004; Rockafellar, 1970)), the function (5.9) can be written in the more transparent form

$$\begin{aligned} \vartheta(\boldsymbol{\nu}, \boldsymbol{\mu}) &= \sup_{\boldsymbol{\lambda}\in\mathbb{R}^d} \left\{ \sum_{(r,s)\notin E} \nu_{rs}(\lambda_r - \lambda_s) + \sum_{(r,s)\in E} \mu_{rs} \log\left(\frac{e^{\lambda_r - \lambda_s}}{\alpha + (1-\alpha)e^{\lambda_r - \lambda_s}} \right) \right\}, \\ &= \phi_{\boldsymbol{\mu}}^*(\boldsymbol{\nu}\mathbf{1} - \boldsymbol{\nu}^\intercal\mathbf{1}), \end{aligned}$$

where $\phi_{\boldsymbol{\mu}}^*$ is the Legendre-Fenchel transform of the (convex) function

$$\phi_{\boldsymbol{\mu}}(\boldsymbol{\lambda}) := -\sum_{(r,s)\in E} \mu_{rs} \log\left(\frac{e^{\lambda_r - \lambda_s}}{\alpha + (1-\alpha)e^{\lambda_r - \lambda_s}} \right).$$

We may note that since $\phi^*_{\boldsymbol{\mu}}$ is convex on $\mathbb{R}^d$, ϑ is a continuous function of its first argument. Before we state our bounds on the collision probability, we recall the following concept from algebraic graph theory. Define *the spanning tree polynomial* of G as

$$T_G(\boldsymbol{z}) := \frac{1}{\mathsf{nst}(G)} \sum_T \prod_{(r,s)\in T} z_{rs},$$

for $\boldsymbol{z} \in \mathbb{R}_+^{d\times d}$, where the sum is over all spanning trees of G, and $\mathsf{nst}(G)$ is the number of spanning trees of G. In cases where G is not connected, we define the following polynomial

$$P_G := \prod_{l=1}^{\mathsf{ncc}(G)} T_{G_l},$$

where G_l is the lth connected component of G, and we denote by $\mathsf{ncc}(G)$ the number of connected components of G. This polynomial may be interpreted as the generating polynomial of *spanning forests* of G having exactly $\mathsf{ncc}(G)$ trees. The polynomials T_G and P_G are multi-affine, homogenous of degree $d-1$ for T_G (when G is connected) and $d - \mathsf{ncc}(G)$ for P_G, and do not depend on the diagonal entries $\{z_{rr} : 1 \le r \le d\}$. Furthermore, letting $z_{rs} = 1$ for all $r \neq s$, we have $P_G(\boldsymbol{z}) = T_G(\boldsymbol{z}) = 1$. We now provide tight asymptotic bounds on the collision probability $q(\boldsymbol{\mu})$ when a subset E of the entries of $\boldsymbol{\mu}$ become large.

Theorem 5.3. *Let $G = (V, E)$ with $V = \{1, \cdots, d\}$, $E = \{(r,s) \in V^2 \ : \ r \neq s\}$, and $\epsilon \in (0,1)$. There exist two constants $0 < c_u < c_l$ depending on ϵ, d and α such that for all n sufficiently large, and all $\boldsymbol{\mu} \in \{0, \cdots, n\}^{d\times d}$ with $\mu_{rs} \ge \epsilon n$ if and only if $(r,s) \in E$, we have*

$$c_l \frac{e^{-\vartheta_l(\boldsymbol{\mu})}}{P_G(\boldsymbol{\mu})^{1/2}} \le q(\boldsymbol{\mu}) \le c_u \frac{e^{-\vartheta_u(\boldsymbol{\mu})}}{P_G(\boldsymbol{\mu})^{1/2}}.$$

with

$$\vartheta_u(\boldsymbol{\mu}) = \inf_{\boldsymbol{\nu}} \{\vartheta(\boldsymbol{\nu}, \boldsymbol{\mu}) : 0 \le \nu_{rs} \le \mu_{rs} \ \forall (r,s) \notin E\},$$

and

$$\vartheta_l(\boldsymbol{\mu}) = \sup_{\boldsymbol{\nu}} \{\vartheta(\boldsymbol{\nu}, \boldsymbol{\mu}) : 0 \le \nu_{rs} \le \mu_{rs} \ \forall (r,s) \notin E\}.$$

Let us now expand on the above result and derive some special cases and corollaries. First, we see that the collision probabilities can decay at two different speeds—polynomial or exponential—in the entries of the overlap matrix $\boldsymbol{\mu}$, depending on whether $\vartheta_u(\boldsymbol{\mu})$ (and/or $\vartheta_l(\boldsymbol{\mu})$) is zero or strictly negative. Second, the apparent gap in the exponential decay of $q(\boldsymbol{\mu})$ in the above characterization is artificial; one can make ϑ_u and ϑ_l equal by taking $\mu_{rs} = 0$ for all $(r,s) \notin E$. Alternatively, they could be made arbitrarily close to each other under an appropriate limit: Assume for simplicity that $\mu_{rs} = nw_{rs} > 0$ for all $(r,s) \in E$ for some $\boldsymbol{w} \in [0,1]^{d\times d}$. We have

$$\vartheta(\boldsymbol{\nu}, \boldsymbol{\mu}) = n\vartheta(\boldsymbol{\nu}/n, \boldsymbol{w}).$$

For $(r,s) \notin E$, we have $\mu_{rs} < \epsilon n$, therefore

$$\vartheta_u(\boldsymbol{\mu})/n \leq \inf_{\boldsymbol{x}}\{\vartheta(\boldsymbol{x},\boldsymbol{w}) : 0 \leq x_{rs} \leq \epsilon \ \forall (r,s) \notin E\} \xrightarrow[\epsilon \to 0]{} \vartheta(\boldsymbol{0},\boldsymbol{w}).$$

The last step is justified by the continuity of $\vartheta(\cdot,\boldsymbol{w})$. The same argument holds for $v_l(\boldsymbol{\mu})$. Denoting the limiting function under this operation as $\vartheta(\boldsymbol{w})$, we obtain:

$$\vartheta(\boldsymbol{w}) = \sup_{\boldsymbol{\lambda}\in\mathbb{R}^d} \sum_{(r,s)\in E} w_{rs} \log\left(\frac{e^{\lambda_r-\lambda_s}}{\alpha+(1-\alpha)e^{\lambda_r-\lambda_s}}\right) = \min_{\substack{\boldsymbol{x}\in[0,1]^{d\times d}\\ \boldsymbol{w}\odot\boldsymbol{x}\in\mathcal{F}}} \varphi_{\boldsymbol{w}}(\boldsymbol{x}).$$

The function ϑ can be seen as the exponential rate of decay of $q(\boldsymbol{\mu})$. The reason ϑ_u and ϑ_l cannot (in general) be replaced by ϑ in Theorem 5.3 is that all control on the constants c_u and c_l is lost when $\epsilon \to 0$. Next, we identify the cases where this exponential decay is non-vacuous.

Lemma 5.4. *Let $\alpha \in (0,1)$, and $\boldsymbol{\mu} \in \mathbb{R}_+^{d\times d}$. We have*

(i) $\vartheta(\boldsymbol{\mu}) = 0$ *if and only if* $\boldsymbol{\mu} \in \mathcal{F}$,

(ii) $\vartheta_u(\boldsymbol{\mu}) = 0$ *if and only if* $\boldsymbol{M}_G(\alpha\boldsymbol{\mu},\boldsymbol{\nu}) \in \mathcal{F}$ *for* some $\boldsymbol{\nu} \in \mathbb{R}_+^{d\times d}$ *such that* $0 \leq \nu_{rs} \leq \mu_{rs}$ *for all* $(r,s) \notin E$.

Now we specialize Theorem 5.3 to the case where the entries of the overlap matrix are either zero or grow proportionally to n. From Theorem 5.3 and Lemma 5.4, we deduce a key corollary on the convergence of the properly rescaled logarithm of the collision probabilities.

Corollary 5.5. *Given a graph $G = (V,E)$, let $\boldsymbol{w} \in [0,1]^{d\times d}$ be such that $w_{rs} > 0$ if and only if $(r,s) \in E$. If $\boldsymbol{w} \in \mathcal{F}$ then*

$$\lim_{n\to\infty} \frac{\log q(n\boldsymbol{w})}{\log n} = -\frac{d - \mathsf{ncc}(G)}{2}.$$

Otherwise if $\boldsymbol{w} \notin \mathcal{F}$, then

$$\lim_{n\to\infty} \frac{\log q(n\boldsymbol{w})}{n} = -\vartheta(\boldsymbol{w}).$$

We see that the assignments τ such that $\boldsymbol{\mu}(\tau,\tau^*) \in \mathcal{F}$ exhibit a much stronger correlation to τ^* than those for which this overlap matrix does not belong to $\mathcal{F}$, and will hence survive much longer as $n \to \infty$.

***Proof** of Lemma 5.4.* Let $\boldsymbol{\mu},\boldsymbol{\nu} \in \mathbb{R}_+^{d\times d}$ with $\boldsymbol{\mu} \neq \boldsymbol{0}$. Let $\alpha \in (0,1)$, and let $G = (V,E)$ denote a graph on d vertices. The function $\varphi_{\boldsymbol{\mu}}$ defined in (5.8) is strictly convex on the support of $\boldsymbol{\mu}$, i.e., on the subspace induced by the non-zero coordinates of $\boldsymbol{\mu}$, so it admits a unique minimizer on the closed convex set $\{\boldsymbol{x} \in [0,1]^{d\times d} : \boldsymbol{M}_G(\boldsymbol{x}^* \odot \boldsymbol{\mu},\boldsymbol{\nu}) \in \mathcal{F}\}$ intersected with that subspace. Let $\boldsymbol{x}^*$ be this minimizer. By differentiating the associated Lagrangian, the entries of $\boldsymbol{x}^*$ admit the expressions

$$x^*_{rs} = \frac{\alpha}{\alpha + (1-\alpha)e^{\lambda_r - \lambda_s}},$$

for all $(r,s) \in E$ (recall that $\mu_{rs} > 0$ for all such (r,s)), and where the vector $\boldsymbol{\lambda} \in \mathbb{R}^d$ is the unique solution up to global shifts of the system of equations: $\forall r \in \{1, \cdots, d\}$

$$\sum_{s:(r,s)\in E} \frac{\alpha\mu_{rs}}{\alpha + (1-\alpha)e^{\lambda_r - \lambda_s}} + \sum_{s:(r,s)\notin E} \nu_{rs} = \sum_{s:(r,s)\in E} \frac{\alpha\mu_{sr}}{\alpha + (1-\alpha)e^{\lambda_s - \lambda_r}} + \sum_{s:(r,s)\notin E} \nu_{sr}. \tag{5.10}$$

The claims of the lemma follow directly from the system of equations (5.10) and the fact that the non-negative function $\varphi_{\boldsymbol{\mu}}$ vanishes if and only if $x^*_{rs} = \alpha$ for all $(r,s) \in E$: to show (i), we take $\boldsymbol{\nu} = \mathbf{0}$. It is clear from the equations that $\boldsymbol{\mu} \in \mathcal{F}$ if and only if $\boldsymbol{\lambda} = c\mathbf{1}$, $c \in \mathbb{R}$, is a solution to the above equations; and this is equivalent to $x^*_{rs} = \alpha$ whenever $\mu_{rs} > 0$. This is in turn equivalent to $\vartheta(\boldsymbol{\mu}) = \varphi_{\boldsymbol{\mu}}(\boldsymbol{x}^*) = 0$. The same strategy is employed to show (ii), in conjunction with the continuity of the function $\boldsymbol{\nu} \mapsto \vartheta(\boldsymbol{\nu}, \boldsymbol{\mu})$ over a compact domain (the infimum defining ϑ_u is attained). ■

***Proof** of Corollary 5.5.* Fix $G = (V, E)$, let $\boldsymbol{w} \in (0,1)^{d\times d}$ with $w_{rs} > 0$ if and only if $(r,s) \in E$, and let n be an integer. For simplicity, assume that for $n\boldsymbol{w}$ is an array of integer entries. The non-integer part introduces easily manageable error terms. Applying Theorem 5.3 with $\epsilon = \min_{(r,s)\in E} w_{rs}$, we have for n large

$$c_l P_G(n\boldsymbol{w})^{-1/2} \exp -\vartheta_l(n\boldsymbol{w}) \le q(n\boldsymbol{w}) \le c_u P_G(n\boldsymbol{w})^{-1/2} \exp -\vartheta_u(n\boldsymbol{w}).$$

Moreover, since $w_{rs} = 0$ for $(r,s) \notin E$, we have

$$\vartheta_u(n\boldsymbol{w}) = \vartheta_l(n\boldsymbol{w}) = n\vartheta(\boldsymbol{w}).$$

On the other hand, by homogeneity of the polynomial P_G, $P_G(n\boldsymbol{w}) = n^{d-\mathsf{ncc}(G)} P_G(\boldsymbol{w})$. Applying Lemma 5.4 yields the desired result: If $\boldsymbol{w} \in \mathcal{F}$ then

$$\lim_{n\to\infty} \frac{\log q(n\boldsymbol{w})}{\log n} = -\frac{d - \mathsf{ncc}(G)}{2}.$$

Otherwise,

$$\lim_{n\to\infty} \frac{\log q(n\boldsymbol{w})}{n} = -\vartheta(\boldsymbol{w}).$$

■

A Gaussian integral

One important step in proving Theorem 5.3 (specifically for obtaining the polynomial decay part of $q(\boldsymbol{\mu})$) is the following identity relating the Gaussian integral on a linear space $\mathcal{F}(G)$ defined based on a graph G to the spanning tree/forest polynomial of G. We denote by K_d the complete graph on d vertices where every pair of distinct vertices is connected by *two parallel edges.*

Proposition 5.6. *Let $G = (V, E)$ be a graph on d vertices, where self-loops and up to two parallel edges are allowed: $V = \{1, \cdots, d\}$, $E \subseteq V \times V$. Further, let*

$$\mathcal{F}(G) = \Big\{\boldsymbol{x} \in \mathcal{F} \ : \ x_{rs} = 0 \textit{ for } (r, s) \notin E\Big\}.$$

For any array of positive real numbers $(w_{rs})_{(r,s)\in E}$, we have

$$\int_{\mathcal{F}(G)} e^{-\sum_{rs} x_{rs}^2/2w_{rs}} \,\mathrm{d}\boldsymbol{x} = \left((2\pi)^{\dim(\mathcal{F}(G))} \frac{\prod_{r,s} w_{rs}}{P_G(\boldsymbol{w})}\right)^{1/2}.$$

In the case where G is the complete graph K_d, $\mathcal{F}(G) = \mathcal{F}$, $\dim(\mathcal{F}) = (d-1)^2 + d$, and $P_G = T_G = (2^{d-1}d^{d-2})^{-1}\sum_T \prod_{(r,s)\in T} w_{rs}$ where the sum is over all spanning trees of K_d. The pre-factor in the last expression comes from Cayley's formula for the number of spanning trees of the complete graph. We will show that it suffices to prove Proposition 5.6 in the case where $G = K_d$ in order to establish it for any graph G. We were not able to locate this identity in the literature. To illuminate the combinatorial mechanisms behind it, we provide what appear to be two very different proofs of it. A first "direct" and purely combinatorial proof views $\mathcal{F}(G)$ as the space of *Eulerian flows* of the graph G. A second, slightly indirect proof which is mainly analytic, and relates the above Gaussian integral to the characteristic polynomial of the Laplacian matrix of G then invokes the Principal Minors Matrix Tree theorem (see, e.g., (Chaiken, 1982)).

5.5 Computing the annealed free energy

In this section we establish the existence of $\mathfrak{F}(\gamma)$, and compute its value for all $\gamma > 0$. For $1 \le k \le d$ let $\mathcal{D}_k$ denote the set of binary matrices $\boldsymbol{X} \in \{0,1\}^{k\times d}$ such that each column of $\boldsymbol{X}$ contains *exactly* one non-zero entry and each row contains *at least* one non-zero entry. The elements of $\mathcal{D}_k$ represent partitions of the set $\{1, \cdots, d\}$ into k non-empty subsets.

Proposition 5.7. *Let $m = \gamma\frac{n}{\log n}$ with $\gamma > 0$ fixed for all $n \ge 2$. We have*

$$\mathfrak{F}(\gamma) = \max_{1\le k\le d-1}\left\{H(\boldsymbol{\pi}) - \min_{\boldsymbol{X}\in\mathcal{D}_k} H(\boldsymbol{X}\boldsymbol{\pi}) - \frac{\gamma}{2}(d-k)\right\}.$$

Moreover, the inner minimization problem in the above expression can be solved explicitly:

Lemma 5.8. *Let $\boldsymbol{\pi}_{[\cdot]}$ be a permutation of the vector $\boldsymbol{\pi}$ such that $\pi_{[1]} \geq \pi_{[2]} \geq \cdots \geq \pi_{[d]}$. And for $1 \leq k \leq d-1$, let $\boldsymbol{\pi}^{(k)} \in \Delta^{k-1}$ defined as $\pi_1^{(k)} = \sum_{r=1}^{d-k+1} \pi_{[r]}$ and $\pi_l^{(k)} = \pi_{[d-k+l]}$ for all $2 \leq l \leq k$ (if $k \geq 2$). Then*

$$\min_{\boldsymbol{X} \in \mathcal{D}_k} H(\boldsymbol{X}\boldsymbol{\pi}) = H(\boldsymbol{\pi}^{(k)}).$$

Theorem 5.2 follows from Proposition 5.7 and Lemma 5.8. We begin with the proof of the latter and devote the next subsection to the lengthier proof of the former.

***Proof** of Lemma 5.8.* We start with an arbitrary partition of $\boldsymbol{\pi}$ into k groups, and define a sequence of operations on the set of k-partitions of $\boldsymbol{\pi}$ that strictly decreases $H(\boldsymbol{X}\boldsymbol{\pi})$ at each step, and, irrespective of the starting point, always converges to $\boldsymbol{\pi}^{(k)}$. Starting with an arbitrary k-partition, write down the groups from left to right in decreasing order of total weight of each group. Initially, every group is marked *incomplete*. Then we perform the following operations:

1. Start with the rightmost incomplete group.

2. If it has more than one element, transfer the largest element to the leftmost group. This strictly decreases the entropy, since the heaviest group gets heavier and the lightest group gets lighter. Repeat this step until the rightmost group has exactly one element, and then move to the next step.

3. Consider this (now singleton) group. If there is no element to its left that is lighter than it, mark the group as complete. Else, swap this element with the lightest element to its left, and then mark it complete. Then go back to step 1.

■

Proof of Proposition 5.7

Let $m = \gamma \frac{n}{\log n}$. Recall from equation (5.6) that

$$\mathbb{E}[\mathcal{Z} - 1] = \binom{n}{n\boldsymbol{\pi}}^{-1} \cdot \sum_{\boldsymbol{\mu}} \binom{n}{\boldsymbol{\mu}} q(\boldsymbol{\mu})^m \, \mathbb{1}\left\{\boldsymbol{\mu}^\intercal \mathbf{1} = n\boldsymbol{\pi}\right\},$$

where the sum is over all arrays $\boldsymbol{\mu} \in \mathbb{Z}_+^{d \times d}$ such that $\mathbf{1}^\intercal \boldsymbol{\mu} \mathbf{1} = n$, $1 \leq \sum_{r \neq s} \mu_{rs}$. Since the sum defining $\mathbb{E}[\mathcal{Z} - 1]$ is larger than its maximum term and smaller than the maximum term times $(n+1)^{d^2}$, we only need to understand the convergence of the sequence

$$\mathfrak{F}_n := \frac{1}{n} \log \left(\max_{\substack{\boldsymbol{\mu} \in \{0, \cdots, n\}^{d \times d} \\ \text{non-diagonal}}} \binom{n}{\boldsymbol{\mu}} q(\boldsymbol{\mu})^m \, \mathbb{1}\left\{\boldsymbol{\mu}^\intercal \mathbf{1} = n\boldsymbol{\pi}\right\} \right)$$

$$= \max\left\{\frac{1}{n}\log\binom{n}{\boldsymbol{\mu}} + \gamma\frac{\log q(\boldsymbol{\mu})}{\log n} \;:\; \boldsymbol{\mu}\in\{0,\cdots,n\}^{d\times d}, \sum_{r\neq s}\mu_{rs}\geq 1, \boldsymbol{\mu}^\intercal\mathbf{1} = n\boldsymbol{\pi}\right\}.$$

If this sequence converges, we would have

$$\mathfrak{F}(\gamma) = -H(\boldsymbol{\pi}) + \lim_{n\to\infty}\mathfrak{F}_n, \tag{5.11}$$

since $\frac{1}{n}\log\binom{n}{n\boldsymbol{\pi}} \to H(\boldsymbol{\pi})$ by Stirling's formula. Next, we show that the above limit indeed exists. Let

$$\psi_n(\boldsymbol{w}) := \frac{1}{n}\log\binom{n}{n\boldsymbol{w}} + \gamma\frac{\log q(n\boldsymbol{w})}{\log n}. \tag{5.12}$$

By Corollary 5.5, the function

$$\psi(\boldsymbol{w}) := \begin{cases} H(\boldsymbol{w}) - \frac{\gamma}{2}(d - \mathsf{ncc}(\boldsymbol{w})) & \text{if } \boldsymbol{w}\in\mathcal{F},\\ -\infty & \text{otherwise,}\end{cases} \tag{5.13}$$

is the point-wise limit of the sequence of functions $\{\psi_n\}_{n\geq 2}$ on $\Delta^{d\times d-1}$. Next, we use the following lemma which states that any non-diagonal sequence of maximizers $\{\boldsymbol{\mu}^{(n)}\}_n$ of ψ_n is such that $\sum_{r\neq s}\mu_{rs}^{(n)}$ grows proportionally to n.

Lemma 5.9. *For all $n\geq 2$, let*

$$\boldsymbol{\mu}^{(n)} \in \arg\max\left\{\psi_n(\boldsymbol{\mu}/n) \;:\; \boldsymbol{\mu}\in\{0,\cdots,n\}^{d\times d},\; 1\leq\sum_{r\neq s}\mu_{rs}\leq n,\; \boldsymbol{\mu}^\intercal\mathbf{1} = n\boldsymbol{\pi}\right\}.$$

It holds that

$$\liminf_{n\to\infty}\frac{\sum_{r\neq s}\mu_{rs}^{(n)}}{n} > 0.$$

By Lemma 5.9, which we prove at the end of the current argument, we can safely restrict the set of candidate maximizers to those $\boldsymbol{\mu}$ such that $\sum_{r\neq s}\mu_{rs}\geq c_0 n$ for some fixed but small $c_0 > 0$. From here, and by a change of variables $\boldsymbol{\mu} = n\boldsymbol{w}$, mere point-wise convergence suffices to interchange lim inf and sup:

$$\begin{aligned}\liminf_{n\to\infty}\mathfrak{F}_n &\geq \liminf_{n\to\infty}\;\sup\left\{\psi_n(\boldsymbol{w}) \;:\; \boldsymbol{w}\in\{i/n : 0\leq i\leq n\}^{d\times d},\; c_0\leq\sum_{r\neq s}w_{rs}\leq 1,\; \boldsymbol{w}^\intercal\mathbf{1} = \boldsymbol{\pi}\right\}\\ &\geq \sup\left\{\psi(\boldsymbol{w}) \;:\; \boldsymbol{w}\in[0,1]^{d\times d}\cap\mathcal{F},\; c_0\leq\sum_{r\neq s}w_{rs}\leq 1,\; \boldsymbol{w}^\intercal\mathbf{1} = \boldsymbol{\pi}\right\}. \end{aligned}\tag{5.14}$$

Now we present a matching upper bound for $\limsup\mathfrak{F}_n$. For $\epsilon > 0$, let $G_n = (\{1,\cdots,d\}, E_n)$ be defined such that $(r,s)\in E_n$ if and only if $w_{rs}^{(n)}\geq\epsilon$. Let $(G_l)_{l=1}^k$ denote the connected components of the graph G_n, $k = \mathsf{ncc}(G_n)$. Also, for $\boldsymbol{w}$ an array for positive entries, let $\mathsf{ncc}^\epsilon(\boldsymbol{w})$ denote the number of connected components of the graph $G(\boldsymbol{w}, \epsilon) = (V, E(\boldsymbol{w}, \epsilon))$, $V = \{1, \cdots, d\}$, $E(\boldsymbol{w}, \epsilon) = \{(r, s) \ : \ r \neq s, \ w_{rs} > \epsilon\}$, and let

$$\vartheta^\epsilon(\boldsymbol{w}) := \inf_{\boldsymbol{x}} \{\vartheta(\boldsymbol{x}, \boldsymbol{w}) : 0 \leq x_{rs} \leq \epsilon \ \forall (r, s) \notin E(\boldsymbol{w}, \epsilon)\}.$$

We will also write $\mathsf{ncc}(\boldsymbol{w})$ for $\mathsf{ncc}^0(\boldsymbol{w})$. Let $\boldsymbol{w}^{(n)} = \boldsymbol{\mu}^{(n)}/n$ for all $n \geq 2$, where $\boldsymbol{\mu}^{(n)}$ is defined in Lemma 5.9. By Theorem 5.3, we have for n sufficiently large

$$q(n\boldsymbol{w}^{(n)}) \leq c_u(\epsilon, d, \alpha) P_{G_n}(n\boldsymbol{w}^{(n)})^{-1/2} \exp -\vartheta^\epsilon(n\boldsymbol{w}^{(n)}).$$

Since $w_{rs}^{(n)} \geq \epsilon$ of all the edges (r, s) of G_n^ϵ, $\prod_l T_{G_l}(\boldsymbol{w}^{(n)})$ is bounded below by ϵ^d independently of n. Therefore, for n sufficiently large,

$$\begin{aligned}
\psi_n(\boldsymbol{w}^{(n)}) &= \frac{1}{n} \log \binom{n}{n\boldsymbol{w}^{(n)}} + \gamma \frac{\log q(n\boldsymbol{w}^{(n)})}{\log n} \\
&\leq \frac{1}{n} \log \binom{n}{n\boldsymbol{w}^{(n)}} - \frac{\gamma}{2}(d - \mathsf{ncc}^\epsilon(\boldsymbol{w}^{(n)})) - \frac{\gamma n}{\log n} \vartheta^\epsilon(\boldsymbol{w}^{(n)}) \\
&\quad + \mathcal{O}\left(\frac{\log c_u(\epsilon, d, \alpha) + d \log(1/\epsilon)}{\log n}\right) \\
&\leq \sup \Bigg\{ H(\boldsymbol{w}) - \frac{\gamma}{2}(d - \mathsf{ncc}^\epsilon(\boldsymbol{w})) - \frac{\gamma n}{\log n} \vartheta^\epsilon(\boldsymbol{w}) \ : \\
&\qquad\qquad \boldsymbol{w} \in [0, 1]^{d \times d}, c_0 \leq \sum_{r \neq s} w_{rs} \leq 1, \ \boldsymbol{w}^\intercal \mathbf{1} = \boldsymbol{\pi} \Bigg\} \\
&\quad + \mathcal{O}\left(\frac{\log c_u(\epsilon, d, \alpha) + d \log(1/\epsilon)}{\log n}\right),
\end{aligned}$$

where the last inequality is obtained by Stirling's formula and taking a supremum over all $\boldsymbol{w}$. By Lemma 5.4, $\vartheta^\epsilon(\boldsymbol{w}) = 0$ if and only if $\boldsymbol{M}_G(\alpha\boldsymbol{w}, \boldsymbol{x}) \in \mathcal{F}$ for some $\boldsymbol{x} \in [0, 1]^{d \times d}$ such that $0 \leq x_{rs} \leq \epsilon$ for all $(r, s) \notin E$, $G = (V, E)$ being the graph whose edges are $(r, s) : w_{rs} \geq \epsilon$. This constrains the supremum to be achieved in the space of such $\boldsymbol{w}$ for n sufficiently large. Moreover, this condition implies in particular that

$$\|\boldsymbol{w}\mathbf{1} - \boldsymbol{w}^\intercal \mathbf{1}\|_{\ell_\infty} \leq 2d\alpha^{-1}\epsilon,$$

where $\|\cdot\|_{\ell_\infty}$ is the ℓ_∞ norm of a vector in $\mathbb{R}^d$. Consequently, this yields the following upper bound as $n \to \infty$,

$$\limsup_{n \to \infty} \mathfrak{F}_n \leq \sup \left\{ H(\boldsymbol{w}) - \frac{\gamma}{2}(d - \mathsf{ncc}^\epsilon(\boldsymbol{w})) \ : \ \begin{array}{c} \boldsymbol{w} \in [0, 1]^{d \times d}, \|\boldsymbol{w}\mathbf{1} - \boldsymbol{w}^\intercal \mathbf{1}\|_{\ell_\infty} \leq 2d\alpha^{-1}\epsilon, \\ c_0 \leq \sum_{r \neq s} w_{rs} \leq 1, \ \boldsymbol{w}^\intercal \mathbf{1} = \boldsymbol{\pi} \end{array} \right\}, \tag{5.15}$$

for all $\epsilon > 0$. Next, we argue that as $\epsilon \to 0$, the right-hand side of the above inequality converges to

$$\sup\left\{H(\boldsymbol{w}) - \frac{\gamma}{2}(d - \mathsf{ncc}(\boldsymbol{w})) \;:\; \boldsymbol{w} \in [0,1]^{d\times d} \cap \mathcal{F}, c_0 \le \sum_{r\neq s} w_{rs} \le 1,\; \boldsymbol{w}^\intercal \mathbf{1} = \boldsymbol{\pi}\right\},$$

thereby establishing the existence of the limit $\lim \mathfrak{F}_n$ along with its precise value. Since the function $\epsilon \to \mathsf{ncc}^\epsilon(\boldsymbol{w})$ is non-decreasing for any fixed $\boldsymbol{w}$, the limit of the right-hand side of (5.15) as $\epsilon \to 0$ exists by monotone convergence. The limit can be decomposed as

$$\begin{aligned}&\lim_{\epsilon\to 0}\sup\left\{H(\boldsymbol{w}) - \frac{\gamma}{2}(d - \mathsf{ncc}^\epsilon(\boldsymbol{w})) : \begin{array}{c}\boldsymbol{w} \in [0,1]^{d\times d}, \|\boldsymbol{w}\mathbf{1} - \boldsymbol{w}^\intercal\mathbf{1}\|_{\ell_\infty} \le 2d\alpha^{-1}\epsilon,\\ c_0 \le \sum_{r\neq s} w_{rs} \le 1,\; \boldsymbol{w}^\intercal\mathbf{1} = \boldsymbol{\pi}\end{array}\right\}\\ &= \max_{1\le k\le d}\max_{\{V_l\}_{l=1}^k}\lim_{\epsilon\to 0}\;\sup\left\{H(\boldsymbol{w}) - \frac{\gamma}{2}(d-k) \;:\; \text{such that (5.16) holds}\right\},\end{aligned}$$

$$\left\{\begin{array}{c}\boldsymbol{w} \in [0,1]^{d\times d},\; \|\boldsymbol{w}\mathbf{1} - \boldsymbol{w}^\intercal\mathbf{1}\|_{\ell_\infty} \le 2d\alpha^{-1}\epsilon,\\ w_{rs} \le \epsilon\; \forall (r,s) \in V_l \times V_{l'},\; l \neq l',\\ G_l(\boldsymbol{w}) \text{ is connected},\; \forall l,\; c_0 \le \sum_{r\neq s} w_{rs} \le 1,\; \boldsymbol{w}^\intercal\mathbf{1} = \boldsymbol{\pi},\end{array}\right. \tag{5.16}$$

where $\{V_l\}_{l=1}^k$ ranges over a partitions of the set $\{1,\cdots,d\}$ with k non-empty subsets, and $G_l(\boldsymbol{w}) = (V_l, \{(r,s) \in V_l \times V_l \;:\; w_{rs} > \epsilon\})$ for all $1 \le l \le k$. Letting $\epsilon < c_0$, the range of the outer-most maximum becomes $1 \le k \le d-1$. By concavity of the entropy, the constraint that the graphs $G_l(\boldsymbol{w})$ must be connected can be safely removed from the maximization problem without changing its maximum value since it will be automatically satisfied. Thus, the inner-most optimization problem is that of a continuous function on a closed and bounded domain that shrinks with ϵ. Its value is therefore a continuous function of ϵ. Hence, by sending ϵ to 0, in conjunction with the lower bound (5.14), we conclude that

$$\lim_{n\to\infty}\mathfrak{F}_n = \sup\left\{\psi(\boldsymbol{w}) \;:\; \boldsymbol{w} \in [0,1]^{d\times d},\; c_0 \le \sum_{r\neq s} w_{rs} \le 1,\; \boldsymbol{w}\mathbf{1} = \boldsymbol{w}^\intercal\mathbf{1} = \boldsymbol{\pi}\right\}. \tag{5.17}$$

As a final step, we make the above expression a bit more explicit. As argued previously, the supremum in (5.17) can be decomposed such that one first takes the maximum of $\psi(\boldsymbol{w})$ over all $\boldsymbol{w}$ such that $w_{rs} = 0$ for all $(r,s) \in V_l \times V_{l'},\; l \neq l'$ where $\{V_l\}_{1\le l\le k}$ is a fixed partition of $\{1,\cdots,d\}$ into non-empty subsets, then maximize over all such partitions, then over all $1 \le k \le d-1$. The first optimization problem has a value

$$\sup\left\{H(\boldsymbol{w}) - \frac{\gamma}{2}(d-k) \;:\; \boldsymbol{w} \in [0,1]^{d\times d},\; w_{rs} = 0,\; (r,s) \in V_l \times V_{l'},\; l \neq l',\; \boldsymbol{w}\mathbf{1} = \boldsymbol{w}^\intercal\mathbf{1} = \boldsymbol{\pi}\right\},$$

where the constraint $c_0 \le \sum_{r\neq s} w_{rs} \le 1$ is not active for c_0 small enough, hence can be removed. Let $\boldsymbol{w}$ be in the above constraint set. Then $H(\boldsymbol{w}) = -\sum_{l=1}^k \sum_{(r,s)\in V_l\times V_l} w_{rs}\log w_{rs}$, and this is maximized at

$$w^*_{rs} = \begin{cases}(\pi_r\pi_s)/\sum_{r'\in V_l}\pi_{r'} & \text{if } (r,s) \in V_l \times V_l,\; l \in \{1,\cdots,k\},\\ 0 & \text{otherwise,}\end{cases} \tag{5.18}$$

with maximum value

$$\begin{aligned} H(\boldsymbol{w}^*) &= 2H(\boldsymbol{\pi}) + \sum_{l=1}^{k} \left(\sum_{r\in V_l} \pi_r\right) \log \left(\sum_{r\in V_l} \pi_r\right), \\ &= 2H(\boldsymbol{\pi}) - H(\boldsymbol{X}\boldsymbol{\pi}), \end{aligned} \tag{5.19}$$

where $\boldsymbol{X} \in \{0,1\}^{k\times d}$, $X_{l,r} = 1$ if and only if $r \in V_l$. Note that $\mathcal{D}_k$ is the set of all such matrices (each one corresponding to a partition $\{V_l\}$ of $\{1,\cdots,d\}$). Finally, by maximizing over all possible partitions, and using (5.11) we get

$$\mathfrak{F}(\gamma) = \max_{1\le k\le d-1} \left\{ H(\boldsymbol{\pi}) - \min_{\boldsymbol{X}\in\mathcal{D}_k} H(\boldsymbol{X}\boldsymbol{\pi}) - \frac{\gamma}{2}(d-k) \right\}.$$

This completes the proof of Proposition 5.7, except for the proof Lemma 5.9, which we provide below.

***P**roof of Lemma 5.9.* Let

$$\boldsymbol{\mu}^{(n)} \in \arg\max \left\{ \psi_n(\boldsymbol{\mu}/n) \;:\; \boldsymbol{\mu} \in \{0,\cdots,n\}^{d\times d},\ 1 \le \sum_{r\neq s} \mu_{rs},\ \boldsymbol{\mu}^\intercal \mathbf{1} = n\boldsymbol{\pi} \right\}.$$

We show that

$$\liminf_{n\to\infty}\ n^{-1} \sum_{r\neq s} \mu_{rs}^{(n)} > 0.$$

Let us first show that

$$\frac{(\log n)^3}{n} \sum_{r\neq s} \mu_{rs}^{(n)} \longrightarrow \infty,$$

and then remove the logarithmic factor. We proceed by contradiction, by showing that if the above statement is not true, then the expected number of non-planted solutions $\mathbb{E}[\mathcal{Z}-1]$ vanishes as $n\to\infty$ for any $\gamma > 0$, which contradicts our lower bound of Theorem 5.1. We have

$$\mathbb{E}\left[\mathcal{Z}-1\right] \le \binom{n}{n\boldsymbol{\pi}}^{-1} \cdot (n+1)^{d^2} \cdot \binom{n}{\boldsymbol{\mu}^{(n)}} \cdot q_{\max}^{\gamma n/\log n},$$

with $q_{\max} = \max\left\{ q(\boldsymbol{\mu}) \;:\; 1 \le \sum_{r\neq s} \mu_{rs} \right\} < 1$. Moreover,

$$\binom{n}{\boldsymbol{\mu}^{(n)}} = \binom{n}{n\boldsymbol{\pi}} \prod_{r=1}^{d} \frac{(n\pi_r)!}{\prod_{s\neq r} \mu_{sr}!(n\pi_r - \sum_{s\neq r}\mu_{sr})!} \le \binom{n}{n\boldsymbol{\pi}} \prod_{r=1}^{d} (n\pi_r)^{\sum_{s\neq r}\mu_{sr}}.$$

If $\sum_{r\neq s} \mu_{rs}^{(n)} \le Cn/(\log n)^3$ for some constant $C > 0$, then

$$\mathbb{E}\left[\mathcal{Z}-1\right] \le (n+1)^{d^2} \cdot n^{Cn/(\log n)^3} \cdot q_{\max}^{\gamma n/\log n} \xrightarrow[n\to\infty]{} 0,$$

for all $\gamma > 0$, and this contradicts the fact that below γ_{low} there are exponentially many distinct satisfying assignments.

Now let us assume that $\frac{(\log n)^3}{n}\sum_{r\neq s}\mu_{rs}^{(n)} \to \infty$ but $\liminf n^{-1}\sum_{r\neq s}\mu_{rs}^{(n)} = 0$. We proceed by contradiction once more, and construct a sequence of points that have a higher objective value than $\boldsymbol{\mu}^{(n)}$. Instead of working with convergent subsequences, we may as well assume that $\{\boldsymbol{\mu}^{(n)}\}$ is convergent. Let

$$E_n = \left\{(r,s)\ :\ r \neq s,\ \mu_{rs}^{(n)} > \epsilon \sum_{r\neq s}\mu_{rs}^{(n)}\right\},$$

and

$$E_\infty = \left\{(r,s)\ :\ r \neq s,\ \liminf_{n\to\infty}\frac{\mu_{rs}^{(n)}}{\sum_{r\neq s}\mu_{rs}^{(n)}} > 0\right\},$$

for all n and some $\epsilon > 0$ sufficiently small. Let $k_n = \mathsf{ncc}(G_n)$ be the number of connected components of the graph $G_n = (\{1,\cdots,d\}, E_n)$, and similarly, let $k_\infty = \mathsf{ncc}(G_\infty)$ with $G_\infty = (\{1,\cdots,d\}, E_\infty)$. Observe that E_∞ and E_n are both non-empty sets, hence $k_\infty, k_n \le d-1$ for all n.

Now we consider an arbitrary partition of the set of vertices $\{1,\cdots,d\}$ into k_∞ subsets $\{V_l\}_{1\le l\le k_\infty}$, and let G be the graph on d vertices with edge set $\cup_{l=1}^{k_\infty} V_l \times V_l$; i.e., G is the union of k_∞ *complete* connected components. Finally, let $\boldsymbol{v}^{(n)} := n\boldsymbol{w}$ for all n, with

$$w_{rs} = \begin{cases}(\pi_r\pi_s)/\sum_{r'\in V_l}\pi_{r'} & \text{if } (r,s)\in V_l\times V_l,\ l\in\{1,\cdots,k_\infty\},\\ 0 & \text{otherwise,}\end{cases}$$

Recall that this construction provides one of the candidate maximizers of the annealed free energy (see (5.18)). Observe that $\boldsymbol{v}^{(n)}$ satisfies all the constraints satisfied by $\boldsymbol{\mu}^{(n)}$, and additionally, $\boldsymbol{v}^{(n)} \in \mathcal{F}$. Therefore, by Corollary 5.5, we have

$$\psi_n(\boldsymbol{v}^{(n)}/n) = H(\boldsymbol{w}) - \frac{\gamma}{2}(d-k_\infty) + o_n(1).$$

Recall that the function ψ_n is defined in (5.12). On the other hand, to study the asymptotics of $\psi_n(\boldsymbol{\mu}^{(n)}/n)$, we apply Theorem 5.3 with n replaced by $\sum_{r\neq s}\mu_{rs}^{(n)}$ (which grows to infinity), and we get

$$\psi_n(\boldsymbol{\mu}^{(n)}/n) \le H(\boldsymbol{\pi}) - \frac{\gamma}{2}(d-k_n)\left(1 - 3\frac{\log\log n}{\log n}\right) - \frac{\vartheta_u(\boldsymbol{\mu}^{(n)})}{\log n} + o_n(1).$$

The term in the right-hand side follows from Stirling's formula and the fact that $\mu_{rs}^{(n)}/n \to 0$ for all $r \neq s$. The second term follows from the fact that

$$P_{G_n}(\boldsymbol{\mu}^{(n)}) \ge \left(\epsilon\sum_{r\neq s}\mu_{rs}^{(n)}\right)^{d-k_n} \gg \left(\frac{n}{(\log n)^3}\right)^{d-k_n}.$$

Next, we argue based on these estimates that $\psi_n(\boldsymbol{v}^{(n)}/n) > \psi_n(\boldsymbol{\mu}^{(n)}/n)$ for all n large enough. First, the term involving ϑ_u in the upper bound on $\psi_n(\boldsymbol{\mu}^{(n)}/n)$ can be dropped since it is always non-negative. By direct computation (we already showed this in (5.19)), we have

$$H(\boldsymbol{w}) - H(\boldsymbol{\pi}) = H(\boldsymbol{\pi}) - H(\boldsymbol{p}),$$

with $\boldsymbol{p} \in \Delta^{k_\infty - 1}$ with $p_l = \sum_{r \in V_l} \pi_r$ for all $1 \leq l \leq k_\infty$. We show that the right-hand side of this equality is strictly positive:

$$\begin{aligned}
H(\boldsymbol{\pi}) - H(\boldsymbol{p}) &= -\sum_{r=1}^{d} \pi_r \log \pi_r + \sum_{l=1}^{k_\infty} \left(\sum_{r \in V_l} \pi_r \right) \log \left(\sum_{r \in V_l} \pi_r \right) \\
&= -\sum_{l=1}^{k_\infty} \sum_{r \in V_l} \pi_r \log \left(\frac{\pi_r}{p_l} \right) \\
&= -\sum_{l=1}^{k_\infty} p_l \sum_{r \in V_l} \frac{\pi_r}{p_l} \log \left(\frac{\pi_r}{p_l} \right) \\
&\geq -\sum_{l=1}^{k_\infty} p_l \log \left(\frac{\sum_{r \in V_l} \pi_r^2}{p_l^2} \right), \\
&\geq 0.
\end{aligned}$$

We used Jensen's inequality on the concave function $x \mapsto \log x$, and the fact that $\sum_{r \in V_l} \pi_r^2 \leq p_l \sum_{r \in V_l} \pi_r = p_l^2$ for all l. Moreover, since all coordinates of $\boldsymbol{\pi}$ are strictly positive, equality holds if and only if $\pi_r = p_l$ for all l and $r \in V_l$ which implies that the partition must be trivial; i.e., $k_\infty = d$. Recall that this does not happen since E_∞ is non-empty.

On the other hand, by setting ϵ sufficiently small (smaller than all the limits in the definition of E_∞), any edge in E_∞ will eventually (and permanently from then on) be in E_n. Therefore the number of connected components of G_n does not exceed that of G_∞: $k_n \leq k_\infty$ for n sufficiently large. We conclude that $\psi_n(\boldsymbol{v}^{(n)}/n) > \psi_n(\boldsymbol{\mu}^{(n)}/n)$ for all n large enough. Therefore $\boldsymbol{\mu}^{(n)}$ is not always a maximizer of ψ_n, and this leads to a contradiction. ■

5.6 Proof of Theorem 5.3

Our proof is based on the method of Laplace from asymptotic analysis: when the entries of $\boldsymbol{\mu}$ are large, the sum defining $q(\boldsymbol{\mu})$ is dominated by its largest term corrected by a sub-exponential term which is represented by a Gaussian integral (see, e.g., (De Bruijn, 1970) for the univariate case). Since we are in a multivariate situation, the asymptotics of q depend on which subset of the entries of $\boldsymbol{\mu}$ are large. Our approach is inspired by (Achlioptas and

Naor, 2005). We recall that for $\boldsymbol{\mu} \in \mathbb{Z}_+^{d\times d}$,

$$q(\boldsymbol{\mu}) = \sum_{\substack{\boldsymbol{\nu}\in\mathbb{Z}_+^{d\times d}\cap\mathcal{F} \\ 0\leq\nu_{rs}\leq\mu_{rs}}} \left(\prod_{r,s=1}^{d} \binom{\mu_{rs}}{\nu_{rs}} \alpha^{\nu_{rs}}(1-\alpha)^{\mu_{rs}-\nu_{rs}} \right).$$

Let $G = (V, E)$ with $V = \{1, \cdots, d\}$ and $E = \{(r, s) \in V^2 \ : \ r \neq s\}$. The graph G will be used to store information about which entries of $\boldsymbol{\mu}$ are going to infinity linearly in n, and which entries are not. We can split the sum defining q into a double sum, one involving the large terms (A in subsequent notation), and the rest:

$$q(\boldsymbol{\mu}) = \sum_{\substack{0\leq\nu'_{rs}\leq\mu_{rs} \\ (r,s)\notin E}} \prod_{(r,s)\notin E} \binom{\mu_{rs}}{\nu'_{rs}} \alpha^{\nu'_{rs}}(1-\alpha)^{\mu_{rs}-\nu'_{rs}} A(\boldsymbol{\nu}', \boldsymbol{\mu}),$$

with

$$A(\boldsymbol{\nu}', \boldsymbol{\mu}) = \sum_{\substack{0\leq\nu_{rs}\leq\mu_{rs} \\ (r,s)\in E}} \prod_{(r,s)\in E} \binom{\mu_{rs}}{\nu_{rs}} \alpha^{\nu_{rs}}(1-\alpha)^{\mu_{rs}-\nu_{rs}} \mathbb{1}\left\{\boldsymbol{M}_G(\boldsymbol{\nu}, \boldsymbol{\nu}') \in \mathcal{F}\right\},$$

where for two $d \times d$ matrices $\boldsymbol{a}, \boldsymbol{b}$, $\boldsymbol{M}_G(\boldsymbol{a}, \boldsymbol{b})$ is the $d \times d$ matrix with entries a_{rs} if $(r, s) \in E$ and b_{rs} otherwise. The quantity A will be approximated using the Laplace method. Recall from the expressions (5.8) and (5.9) that

$$\varphi_{\boldsymbol{\mu}}(\boldsymbol{x}) = \sum_{(r,s)\in E} \mu_{rs} D(x_{rs} \parallel \alpha),$$

and

$$\vartheta(\boldsymbol{\nu}, \boldsymbol{\mu}) = \min_{\substack{\boldsymbol{x}\in[0,1]^{d\times d} \\ \boldsymbol{M}_G(\boldsymbol{x}\odot\boldsymbol{\mu},\boldsymbol{\nu})\in\mathcal{F}}} \varphi_{\boldsymbol{\mu}}(\boldsymbol{x}).$$

Let $\boldsymbol{x}^*(\boldsymbol{\nu}, \boldsymbol{\mu})$ be the optimal solution of the above optimization problem.

Before stating our asymptotic approximation result for A, we state an important lemma on the boundedness of the entries of $\boldsymbol{x}^*(\boldsymbol{\nu}, \boldsymbol{\mu})$, where the bounds depend only on ϵ and α.

Lemma 5.10. *Let G be fixed as above, $\alpha \in (0, 1)$ and $\epsilon \in (0, 1)$. There exist two constants $0 < c_l \leq c_u < 1$ depending only on d, α and ϵ such that the following is true: For all integers $n \geq 1$, and $\boldsymbol{\mu} \in \{0, \cdots, n\}^{d\times d}$ such that $\mu_{rs} \geq \epsilon n$ iff $(r, s) \in E$. For all $\boldsymbol{\nu}' \in \{0, \cdots, n\}^{\bar{E}}$ such that $0 \leq \nu'_{rs} \leq \mu_{rs}$ for all $(r, s) \notin E$, we have*

$$c_l \leq \min_{(r,s)\in E} x^*_{rs} \leq \max_{(r,s)\in E} x^*_{rs} \leq c_u.$$

Therefore, the entries of $\boldsymbol{x}^*$ can effectively be treated as constants throughout the rest of the proof. Now we state our asymptotic estimate for A.

Proposition 5.11. *Let G be fixed as above, and $\epsilon > 0$. For all n sufficiently large, all $\boldsymbol{\mu} \in \{0, \cdots, n\}^{d\times d}$ with $\mu_{rs} \geq \epsilon n$ iff $(r,s) \in E$, and all $\boldsymbol{\nu}' \in \{0, \cdots, n\}^{\bar{E}}$ such that $0 \leq \nu'_{rs} \leq \mu_{rs}$ for all $(r,s) \notin E$, we have*

$$A(\boldsymbol{\nu}', \boldsymbol{\mu}) \quad \asymp_{G,d,\epsilon,\alpha} \frac{e^{-\vartheta(\boldsymbol{\nu}', \boldsymbol{\mu})}}{P_G(\boldsymbol{\mu})^{1/2}}.$$

Here, the symbol "$\asymp_{G,d,\epsilon,\alpha}$" means that the ratio is upper- and lower-bounded by constants depending only on G, d, ϵ and α.

By the above proposition, we have

$$q(\boldsymbol{\mu}) \quad \asymp_{G,d,\epsilon,\alpha} \sum_{\substack{\boldsymbol{\nu} \in \mathbb{Z}_+^{\bar{E}} \\ 0 \leq \nu_{rs} \leq \mu_{rs}}} \left(\prod_{(r,s)\notin E} \binom{\mu_{rs}}{\nu_{rs}} \alpha^{\nu_{rs}} (1-\alpha)^{\mu_{rs}-\nu_{rs}} \right) \frac{e^{-\vartheta(\boldsymbol{\nu}, \boldsymbol{\mu})}}{P_G(\boldsymbol{\mu})^{1/2}}.$$

The estimate above (ignoring the term $P_G(\boldsymbol{\mu})$) can be interpreted as the expected value of the function $e^{-\vartheta(\boldsymbol{\nu},\boldsymbol{\mu})}$ under the law of the random variable $\boldsymbol{\nu}$ where each entry ν_{rs} for $(r,s) \notin E$ is independently binomial with parameters α and μ_{rs}. From here, the bounds claimed in Theorem 5.3 follow immediately.

Proof of Proposition 5.11. We will show that

$$A(\boldsymbol{\nu}', \boldsymbol{\mu}) \quad \asymp_{G,d,\epsilon,\alpha} \quad e^{-\vartheta(\boldsymbol{\nu}', \boldsymbol{\mu})} \prod_{(r,s)\in E} \mu_{rs}^{-1/2} \int_{\mathcal{F}(G)} e^{-\sum_{(r,s)\in E} z_{rs}^2/2\mu_{rs}} \, \mathrm{d}\boldsymbol{z}.$$

Then the result follows by applying Proposition 5.6 to evaluate the Gaussian integral. We proceed by showing the upper and lower bounds separately.

The upper bound For $\boldsymbol{\nu}' \in \mathbb{Z}_+^{\bar{E}}, \boldsymbol{\mu} \in \mathbb{Z}_+^{d\times d}$ fixed and some parameter $C(\boldsymbol{\mu}) > 0$ to be adjusted, let

$$\Omega = \left\{ \boldsymbol{\nu} \in \mathbb{Z}_+^E \ : \ \boldsymbol{M}_G(\boldsymbol{\nu}, \boldsymbol{\nu}') \in \mathcal{F}, \ 0 \leq \nu_{rs} \leq \mu_{rs}, \ \sum_{(r,s)\in E} \frac{(\nu_{rs} - x_{rs}^*\mu_{rs})^2}{x_{rs}^*(1-x_{rs}^*)\mu_{rs}} \leq C(\boldsymbol{\mu})^2 \right\}.$$

For $\boldsymbol{\nu} \in \mathbb{Z}_+^E$, we let $\boldsymbol{x} \in [0,1]^E$ defined by $x_{rs} = \nu_{rs}/\mu_{rs}$ for all $(r,s) \in E$. We upper bound the binomial coefficients $\binom{\mu_{rs}}{\nu_{rs}}$ based on whether $\boldsymbol{\nu}$ is in Ω or not:

- If $\boldsymbol{\nu} \in \Omega$ we use the upper bound $\binom{\mu_{rs}}{\nu_{rs}} \leq \left(2\pi\mu_{rs}x_{rs}(1-x_{rs})\right)^{-1/2} \exp \mu_{rs} H(x_{rs})$.
- Otherwise, we use the upper bound $\binom{\mu_{rs}}{\nu_{rs}} \leq 3\sqrt{\mu_{rs}} \exp \mu_{rs} H(x_{rs})$.

Here, $H(x_{rs}) = -x_{rs}\log x_{rs} - (1-x_{rs})\log(1-x_{rs})$. Thus, the summand in $A(\boldsymbol{\nu}', \boldsymbol{\mu})$ is bounded by

$$\prod_{(r,s)\in E} (2\pi\mu_{rs}x_{rs}(1-x_{rs}))^{-1/2} \exp \mu_{rs} D(x_{rs} \parallel \alpha) = \prod_{(r,s)\in E} (2\pi\mu_{rs}x_{rs}(1-x_{rs}))^{-1/2} \exp(-\varphi_{\boldsymbol{\mu}}(\boldsymbol{x}))$$

if $\boldsymbol{\nu} \in \Omega$, and

$$\prod_{(r,s)\in E} 3\mu_{rs}^{1/2} \exp(-\varphi_{\boldsymbol{\mu}}(\boldsymbol{x}))$$

if not. The function $\varphi_{\boldsymbol{\mu}}$ is smooth, and we have $\frac{\mathrm{d}\varphi_{\boldsymbol{\mu}}}{\mathrm{d}x_{rs}}(\boldsymbol{x}) = \mu_{rs}\log\left(\frac{x_{rs}(1-\alpha)}{\alpha(1-x_{rs})}\right)$, and $\frac{\mathrm{d}^2\varphi_{\boldsymbol{\mu}}}{\mathrm{d}x_{rs}^2}(\boldsymbol{x}) = \frac{\mu_{rs}}{x_{rs}(1-x_{rs})} \geq 0$. Therefore, by convexity,

$$\varphi_{\boldsymbol{\mu}}(\boldsymbol{x}) \geq \varphi_{\boldsymbol{\mu}}(\boldsymbol{x}^*) + \frac{1}{2}\sum_{(r,s)\in E}^{d} \frac{\mu_{rs}}{x_{rs}^*(1-x_{rs}^*)}(x_{rs}-x_{rs}^*)^2.$$

Let

$$\ell_{\boldsymbol{\mu}}(\boldsymbol{\nu}) = \sum_{(r,s)\in E} \frac{(\nu_{rs} - \mu_{rs}x_{rs}^*)^2}{x_{rs}^*(1-x_{rs}^*)\mu_{rs}}. \tag{5.20}$$

Based on Lemma 5.10, all the entries of $\boldsymbol{x}^*$ will be treated as constants. If $\boldsymbol{\nu} \in \Omega$ then $\frac{(\nu_{rs}-\mu_{rs}x_{rs}^*)^2}{x_{rs}^*(1-x_{rs}^*)\mu_{rs}} \leq \ell_{\boldsymbol{\mu}}(\boldsymbol{\nu}) \leq C(\boldsymbol{\mu})^2$, and

$$\nu_{rs} \in \left[\mu_{rs}x_{rs}^* \pm C(\boldsymbol{\mu})\sqrt{x_{rs}^*(1-x_{rs}^*)\mu_{rs}}\right].$$

Now let us assume that $C(\boldsymbol{\mu}) = o(\mu_{rs}^{1/2})$ for all $(r,s) \in E$. Then we have

$$\frac{\nu_{rs}}{\mu_{rs}}(1 - \frac{\nu_{rs}}{\mu_{rs}}) \geq x_{rs}^*(1-x_{rs}^*) - o_n(1).$$

If $\boldsymbol{\nu} \notin \Omega$ then $\ell(\boldsymbol{\nu}) \geq C(\boldsymbol{\mu})^2$, therefore A is bounded by

$$\left(\prod_{(r,s)\in E} (2\pi\mu_{rs}x_{rs}^*(1-x_{rs}^*))^{-1/2} \sum_{\boldsymbol{\nu}\in\Omega} e^{-\ell_{\boldsymbol{\mu}}(\boldsymbol{\nu})/2} + \prod_{(r,s)\in E} 3\mu_{rs}^{1/2} \sum_{\substack{\boldsymbol{\nu}\notin\Omega \\ 0\leq\nu_{rs}\leq\mu_{rs}}} e^{-C(\boldsymbol{\mu})^2/2}\right) \cdot \exp(-\varphi_{\boldsymbol{\mu}}(\boldsymbol{x}^*)). \tag{5.21}$$

The second term in the sum above is bounded by $c^{d^2}(\prod_{(r,s)\in E}\mu_{rs}^{3/2})e^{-C(\boldsymbol{\mu})^2/2}$ for some constant $c > 0$. Taking $C(\boldsymbol{\mu})^2 = 5\log\prod_{(r,s)\in E}\mu_{rs}$, this term is $c^{d^2}\prod_{(r,s)\in E}\mu_{rs}^{-1} = \mathcal{O}(n^{-|E|})$. Moreover, for all $(r,s) \in E$, $\mu_{rs} > \epsilon n$, therefore

$$C(\boldsymbol{\mu})^2 \leq 4d^2 \log n \ll \mu_{rs};$$

this choice satisfies the condition $C(\boldsymbol{\mu}) = o(\mu_{rs}^{1/2})$ for $(r,s) \in E$. On the other hand, let

$$S(\boldsymbol{\mu}) = \sum_{\boldsymbol{\nu} \in \mathbb{Z}^E} \mathbb{1}\{\boldsymbol{M}_G(\boldsymbol{\nu}, \boldsymbol{\nu}') \in \mathcal{F}\} \exp(-\ell_{\boldsymbol{\mu}}(\boldsymbol{\nu})/2), \tag{5.22}$$

with $\ell_{\boldsymbol{\mu}}$ defined in (5.20). We ignored the dependence of S on $\boldsymbol{\nu}'$ in the notation on purpose: this dependence is inessential. The first sum in (5.21) is upper bounded by $S(\boldsymbol{\mu})$. Therefore

$$A(\boldsymbol{\nu}', \boldsymbol{\mu}) \leq c_u \prod_{(r,s)\in E} \mu_{rs}^{-1/2} \, \left(S(\boldsymbol{\mu}) + \varepsilon_n\right) \, e^{-\vartheta(\boldsymbol{\nu}', \boldsymbol{\mu})}, \tag{5.23}$$

for some c_u depending on ϵ and d, and $\epsilon_n \to 0$ as $n \to \infty$. We now turn our attention to the lower bound, deferring the analysis of the Gaussian sum $S(\boldsymbol{\mu})$ to a subsequent paragraph.

The lower bound We have

$$A(\boldsymbol{\nu}', \boldsymbol{\mu}) \geq \sum_{\boldsymbol{\nu} \in \Omega} \left(\prod_{(r,s)\in E} \binom{\mu_{rs}}{\nu_{rs}} \alpha^{\nu_{rs}} (1-\alpha)^{\mu_{rs} - \nu_{rs}} \right).$$

Using $\binom{\mu_{rs}}{\nu_{rs}} \geq (8\pi\nu_{rs}(1-\nu_{rs}/\mu_{rs}))^{-1/2} e^{H(\nu_{rs}/\mu_{rs})}$ for all $(r,s) \in E$, we get

$$A(\boldsymbol{\nu}', \boldsymbol{\mu}) \geq \sum_{\boldsymbol{\nu} \in \Omega} \left(\prod_{(r,s)\in E} 8\pi\nu_{rs}(1 - \frac{\nu_{rs}}{\mu_{rs}}) \right)^{-1/2} \cdot \exp\left(-\varphi_{\boldsymbol{\mu}}\left(\boldsymbol{\nu}/\boldsymbol{\mu}\right)\right).$$

For $\boldsymbol{\nu} \in \Omega$, we have $\frac{\nu_{rs}}{\mu_{rs}}(1 - \frac{\nu_{rs}}{\mu_{rs}}) \leq x_{rs}^*(1 - x_{rs}^*) + o_n(1)$ for all r, s, and since $\varphi_{\boldsymbol{\mu}}$ is a smooth function, a Taylor expansion yields

$$\varphi_{\boldsymbol{\mu}}(\boldsymbol{\nu}/\boldsymbol{\mu}) = \varphi_{\boldsymbol{\mu}}(\boldsymbol{x}^*) + \frac{1}{2} \sum_{(r,s)\in E} \frac{(\nu_{rs} - x_{rs}^*\mu_{rs})^2}{x_{rs}^*(1 - x_{rs}^*)\mu_{rs}} + o_n(1).$$

Therefore,

$$\begin{aligned} A(\boldsymbol{\nu}', \boldsymbol{\mu}) &\geq c_l e^{-\varphi_{\boldsymbol{\mu}}(\boldsymbol{x}^*)} \prod_{(r,s)\in E} \mu_{r,s}^{-1/2} \cdot \sum_{\boldsymbol{\nu} \in \Omega} \exp\left(-\frac{1}{2} \sum_{rs} \frac{(\nu_{rs} - x_{rs}^*\mu_{rs})^2}{x_{rs}^*(1 - x_{rs}^*)\mu_{rs}} \right) \\ &= c_l e^{-\vartheta(\boldsymbol{\nu}', \boldsymbol{\mu})} \prod_{(r,s)\in E} \mu_{r,s}^{-1/2} \cdot \left(S(\boldsymbol{\mu}) - \varepsilon_n\right), \end{aligned}$$

where $c_l = c_l(\epsilon, d)$, $S(\boldsymbol{\mu})$ is defined in (5.22) and

$$\varepsilon_n := \prod_{(r,s)\in E} (\mu_{rs} + 1) e^{-C(\boldsymbol{\mu})^2/2} + \sum_{\substack{\boldsymbol{\nu} \in \mathbb{Z}_+^E \\ \nu_{rs} \geq \mu_{rs}+1}} \exp -\ell_{\boldsymbol{\mu}}(\boldsymbol{\nu})/2 + \sum_{\boldsymbol{\nu} \in \mathbb{Z}_-^E} \exp -\ell_{\boldsymbol{\mu}}(\boldsymbol{\nu})/2.$$

We take $C(\boldsymbol{\mu})^2 = 4\log\prod_{(r,s)\in E}\mu_{rs}$. This makes the first term in ε_n bounded by $\prod_{r,s}\mu_{rs}^{-1} = \mathcal{O}(n^{-|E|})$. On the other hand, the remaining tail sums are easily bounded by the tail probability function of a normal random variable (i.e., the error function):

$$\sum_{\substack{\boldsymbol{\nu}\in\mathbb{Z}_+^E\\ \nu_{rs}\geq\mu_{rs}+1}} \exp -\ell_{\boldsymbol{\mu}}(\boldsymbol{\nu})/2 \leq \prod_{(r,s)\in E}\mu_{rs}^{1/2}\mathsf{erfc}\left(\sqrt{\frac{x^*_{rs}}{1-x^*_{rs}}\mu_{rs}}\right),$$

$$\sum_{\boldsymbol{\nu}\in\mathbb{Z}_-^E} \exp -\ell_{\boldsymbol{\mu}}(\boldsymbol{\nu})/2 \leq \prod_{(r,s)\in E}\mu_{rs}^{1/2}\mathsf{erfc}\left(\sqrt{\frac{1-x^*_{rs}}{x^*_{rs}}\mu_{rs}} - \frac{1}{\sqrt{x^*_{rs}(1-x^*_{rs})\mu_{rs}}}\right),$$

with $\mathsf{erfc}(x) = \int_x^\infty e^{-t^2/2}dt$. Since $\mathsf{erfc}(x) \leq e^{-x^2/2}/x$ for all $x > 0$, these two terms decay in a sub-Gaussian way in n.

Bounding the Gaussian sum. Here we approximate S by a continuous Gaussian integral. We prove that

$$S(\boldsymbol{\mu}) \asymp \int_{\mathcal{F}(G)} \exp\left(-\sum_{(r,s)\in E}\frac{\mu_{rs}^{-1}}{2x^*_{rs}(1-x^*_{rs})}z_{rs}^2\right)\mathrm{d}\boldsymbol{z},$$

where the symbol " $\asymp$ " hides constants depending on G, ϵ, d and α as $n \to \infty$. For $\boldsymbol{\nu} \in \mathcal{F}(G)$ an array of integer numbers such that $0 \leq \nu_{rs} \leq \mu_{rs}$, let $T(\boldsymbol{\nu}) = \boldsymbol{\nu} + \mathcal{C}\cap\mathcal{F}(G)$ where $\mathcal{C} = [-1/2, 1/2]^E$. The sum is understood in the Minkowski sense. $T(\boldsymbol{\nu})$ is a "tile" of side 1 centered around $\boldsymbol{\nu}$. Two crucial facts are (i) : $T(\boldsymbol{\nu})$ and $T(\boldsymbol{\nu}')$ are of disjoint interiors when $\boldsymbol{\nu} \neq \boldsymbol{\nu}'$ and (ii) : $T(\boldsymbol{\nu}) \subset \mathcal{F}(G)$. Now for a fixed $\boldsymbol{\nu}$, let $\boldsymbol{z} \in T(\boldsymbol{\nu})$. For $r, s \in E$, we have $\nu_{rs} - 1/2 \leq z_{rs} \leq \nu_{rs} + 1/2$ and $\frac{\nu_{rs}-1/2}{\mu_{rs}} - x^*_{rs} \leq \frac{z_{rs}}{\mu_{rs}} - x^*_{rs} \leq \frac{\nu_{rs}+1/2}{\mu_{rs}} - x^*_{rs}$. Thus

$$\left(\frac{z_{rs}}{\mu_{rs}} - x^*_{rs}\right)^2 \leq \max\left\{\left(\frac{\nu_{rs}-1/2}{\mu_{rs}} - x^*_{rs}\right)^2, \left(\frac{\nu_{rs}+1/2}{\mu_{rs}} - x^*_{rs}\right)^2\right\}.$$

Using the fact $\max\{(x-1/2)^2, (x+1/2)^2\} \leq 2x^2 + 1$ for all $x \in \mathbb{R}$, we get

$$\exp\left(-\sum_{(r,s)\in E}\frac{\mu_{rs}^{-1}}{4x^*_{rs}(1-x^*_{rs})}\left(2\left(\nu_{rs} - \mu_{rs}x^*_{rs}\right)^2 + 1\right)\right)$$
$$\leq \exp\left(-\sum_{(r,s)\in E}\frac{\mu_{rs}^{-1}}{4x^*_{rs}(1-x^*_{rs})}\left(z_{rs} - \mu_{rs}x^*_{rs}\right)^2\right).$$

By integrating both sides of the above inequality on $T(\boldsymbol{\nu})$ in the variable $\boldsymbol{z}$, and summing over all $\boldsymbol{\nu}$ with integer entries such that $\boldsymbol{M}_G(\boldsymbol{\nu}, \boldsymbol{\nu}') \in \mathcal{F}$, we get

$$\mathrm{vol}(\mathcal{C}\cap\mathcal{F}(G))e^{-\sum_{(r,s)\in E}\frac{\mu_{rs}^{-1}}{4x^*_{rs}(1-x^*_{rs})}}\ S(\boldsymbol{\mu})$$

$$\leq \sum_{\substack{\boldsymbol{\nu}\in\mathbb{Z}^E\\ \boldsymbol{M}_G(\boldsymbol{\nu},\boldsymbol{\nu}')\in\mathcal{F}}} \int_{T(\boldsymbol{\nu})} \exp\left(-\sum_{(r,s)\in E} \frac{\mu_{rs}^{-1}\left(z_{rs}-\mu_{rs}x^*_{rs}\right)^2}{4x^*_{rs}(1-x^*_{rs})}\right)\mathrm{d}\boldsymbol{z},$$

$$= \sum_{\substack{\boldsymbol{\nu}\in\mathbb{Z}^E\\ \boldsymbol{M}_G(\boldsymbol{\nu},\boldsymbol{\nu}')\in\mathcal{F}}} \int_{T(\boldsymbol{\nu}-\boldsymbol{x}^*\odot\boldsymbol{\mu})} \exp\left(-\sum_{(r,s)\in E} \frac{\mu_{rs}^{-1}}{4x^*_{rs}(1-x^*_{rs})}z_{rs}^2\right)\mathrm{d}\boldsymbol{z},$$

where vol is the volume according to the $\dim(\mathcal{F}(G))$-dimensional Lebesgue measure. Since $\boldsymbol{M}_G(\boldsymbol{x}^*\odot\boldsymbol{\mu},\boldsymbol{\nu}')\in\mathcal{F}$, we have $\boldsymbol{\nu}-\boldsymbol{x}^*\odot\boldsymbol{\mu}\in\mathcal{F}$ for all $\boldsymbol{\nu}$ we are summing over. Moreover, since the tiles $T(\boldsymbol{\nu})$ are of mutually disjoint interiors, and given that their union is in $\mathcal{F}(G)$, the left-hand side is upper bounded by (there is actually equality)

$$\int_{\mathcal{F}(G)} \exp\left(-\sum_{(r,s)\in E} \frac{\mu_{rs}^{-1}}{4x^*_{rs}(1-x^*_{rs})}z_{rs}^2\right)\mathrm{d}\boldsymbol{z}.$$

Here, to get sharper constants, one could apply a theorem by Vaaler (Vaaler, 1979) which states that the volume of any linear subspace intersected with the cube $\mathcal{C}$ is at least 1; i.e., $\mathrm{vol}(\mathcal{C}\cap\mathcal{F}(G))\geq 1$. This yields

$$S(\boldsymbol{\mu}) \leq e^{\sum_{(r,s)\in E}\frac{\mu_{rs}^{-1}}{4x^*_{rs}(1-x^*_{rs})}}\cdot \int_{\mathcal{F}(G)} \exp\left(-\sum_{(r,s)\in E} \frac{\mu_{rs}^{-1}}{4x^*_{rs}(1-x^*_{rs})}z_{rs}^2\right)\mathrm{d}\boldsymbol{z}.$$

As for the reverse inequality, slightly more care is needed in constructing the approximation. For a given $\boldsymbol{\nu}$, let $\Omega^+=\{(r,s)\ :\ \nu_{rs}\geq x^*_{rs}\mu_{rs}+1/2\}$ and $\Omega^-=\{(r,s)\ :\ \nu_{rs}\leq x^*_{rs}\mu_{rs}-1/2\}$. For $\boldsymbol{z}\in T(\boldsymbol{\nu})$, we have $(z_{rs}-x^*_{rs}\mu_{rs})^2\geq(\nu_{rs}-x^*_{rs}\mu_{rs}-1/2)^2$ if $(r,s)\in\Omega^+$ and $(z_{rs}-x^*_{rs}\mu_{rs})^2\geq(\nu_{rs}-x^*_{rs}\mu_{rs}+1/2)^2$ if $(r,s)\in\Omega^-$. Otherwise, for $(r,s)\notin\Omega^+\cup\Omega^-$, we have $|\nu_{rs}-x^*_{rs}\mu_{rs}|<1/2$ and $|(z_{rs}-x^*_{rs}\mu_{rs})^2-(\nu_{rs}-x^*_{rs}\mu_{rs})^2|<1/2(1+1/2)=3/4$. Therefore

$$\sum_{(r,s)\in\Omega^+}\frac{(\nu_{rs}-x^*_{rs}\mu_{rs}+1/2)^2}{\mu_{rs}x^*_{rs}(1-x^*_{rs})}+\sum_{(r,s)\in\Omega^-}\frac{(\nu_{rs}-x^*_{rs}\mu_{rs}-1/2)^2}{\mu_{rs}x^*_{rs}(1-x^*_{rs})}+\sum_{(r,s)\notin\Omega^+\cup\Omega^-}\frac{(\nu_{rs}-x^*_{rs}\mu_{rs})^2}{\mu_{rs}x^*_{rs}(1-x^*_{rs})}$$
$$\leq \sum_{(r,s)\in E}\frac{(z_{rs}-x^*_{rs}\mu_{rs})^2}{\mu_{rs}x^*_{rs}(1-x^*_{rs})}+\sum_{(r,s)\in E}\frac{3\mu_{rs}^{-1}}{4x^*_{rs}(1-x^*_{rs})}.$$

On the other hand, $(\nu_{rs}-x^*_{rs}\mu_{rs})^2\leq(\nu_{rs}-x^*_{rs}\mu_{rs}+1/2)^2$ when $(r,s)\in\Omega^+$ and $(\nu_{rs}-x^*_{rs}\mu_{rs})^2\leq(\nu_{rs}-x^*_{rs}\mu_{rs}-1/2)^2$ when $(r,s)\in\Omega^-$. After integrating on $T(\boldsymbol{\nu})$ and summing over all $\boldsymbol{\nu}\in\mathbb{Z}^E$ such that $\boldsymbol{\nu}-\boldsymbol{x}^*\odot\boldsymbol{\mu}\in\mathcal{F}$, we obtain

$$\mathrm{vol}(\mathcal{C}\cap\mathcal{F}(G))S(\boldsymbol{\mu}) \geq e^{-\sum_{(r,s)\in E}\frac{3\mu_{rs}^{-1}}{8x^*_{rs}(1-x^*_{rs})}}$$

$$\cdot \sum_{\substack{\boldsymbol{\nu}\in\mathbb{Z}^E \\ \boldsymbol{\nu}-\boldsymbol{x}^*\odot\boldsymbol{\mu}\in\mathcal{F}}} \int_{T(\boldsymbol{\nu}-\boldsymbol{x}^*\odot\boldsymbol{\mu})} \exp\left(-\sum_{(r,s)\in E} \frac{\mu_{rs}^{-1}}{2x_{rs}^*(1-x_{rs}^*)} z_{rs}^2\right) \mathrm{d}\boldsymbol{z},$$

and the last sum is equal to

$$\sum_{\boldsymbol{\nu}\in(\mathbb{Z}^E+\boldsymbol{x}^*\odot\boldsymbol{\mu})\cap\mathcal{F}(G)} \int_{T(\boldsymbol{\nu})} \exp\left(-\sum_{(r,s)\in E} \frac{\mu_{rs}^{-1}}{2x_{rs}^*(1-x_{rs}^*)} z_{rs}^2\right) \mathrm{d}\boldsymbol{z}$$
$$= \int_{\mathcal{F}(G)} \exp\left(-\sum_{(r,s)\in E} \frac{\mu_{rs}^{-1}}{2x_{rs}^*(1-x_{rs}^*)} z_{rs}^2\right) \mathrm{d}\boldsymbol{z}.$$

Finally,

$$S(\boldsymbol{\mu}) \geq c(G,d) e^{-\sum_{(r,s)\in E} \frac{3\mu_{rs}^{-1}}{2x_{rs}^*(1-x_{rs}^*)}} \cdot \int_{\mathcal{F}(G)} \exp\left(-\sum_{(r,s)\in E} \frac{\mu_{rs}^{-1}}{2x_{rs}^*(1-x_{rs}^*)} z_{rs}^2\right) \mathrm{d}\boldsymbol{z}.$$

■

***Proof** of Lemma 5.10.* Recall that $\boldsymbol{x}^*$ is the unique minimizer of the function

$$\varphi_{\boldsymbol{\mu}} = \sum_{(r,s)\in E} \mu_{rs} D(x_{rs} \parallel \alpha)$$

on $[0,1]^{d\times d}$ subject to $\boldsymbol{M}_G(\boldsymbol{x}^*\odot\boldsymbol{\mu},\boldsymbol{\nu}) \in \mathcal{F}$. Recall also that the entries of $\boldsymbol{x}^*$ admit the expressions

$$x_{rs}^* = \frac{\alpha}{\alpha+(1-\alpha)e^{\lambda_r^*-\lambda_s^*}},$$

for all $(r,s)\in E$. The vector $\boldsymbol{\lambda}^*\in\mathbb{R}^d$ is the unique solution up to global shifts to the dual optimization problem (strong duality holds here (Boyd and Vandenberghe, 2004; Rockafellar, 1970))

$$\sup_{\boldsymbol{\lambda}\in\mathbb{R}^d} \left\{ \sum_{(r,s)\notin E} \nu_{rs}(\lambda_r-\lambda_s) + \sum_{(r,s)\in E} \mu_{rs} \log\left(\frac{e^{\lambda_r-\lambda_s}}{\alpha+(1-\alpha)e^{\lambda_r-\lambda_s}}\right) \right\}. \tag{5.24}$$

Our claim reduces to the boundedness of the differences $|\lambda_r^*-\lambda_s^*|$ for all $(r,s)\in E$ independently of $n,\boldsymbol{\mu},\boldsymbol{\nu}$ and r,s. We will shortly prove the following inequality

$$\sum_{(r,s)\in E} \mu_{rs}(\lambda_r^*-\lambda_s^*)^2 \leq \kappa(\alpha) \sum_{(r,s)\in E} \mu_{rs}, \tag{5.25}$$

where $\kappa(\alpha) = \frac{1}{\alpha^2} + \frac{1}{(1-\alpha)^2}$. Assuming the above is true, by the Cauchy-Schwarz inequality, we would have

$$\sum_{(r,s)\in E} |\lambda_r^* - \lambda_s^*| \leq \left(\sum_{(r,s)\in E} \mu_{rs}^{-1}\right)^{1/2} \left(\kappa(\alpha) \sum_{(r,s)\in E} \mu_{rs}\right)^{1/2} \leq d^2(\kappa(\alpha)/\epsilon)^{1/2},$$

since $\epsilon n \leq \mu_{rs} \leq n$ for all $(r,s) \in E$. We would then be done. Now, the inequality (5.25) follows from convexity considerations. We let ϕ be the function being maximized in (5.24). By concavity of ϕ, we have

$$\phi(\boldsymbol{\lambda}^*) - \phi(\mathbf{0}) \leq \boldsymbol{\lambda}^{*\intercal}\nabla\phi(\mathbf{0}) + \frac{1}{2}\boldsymbol{\lambda}^{*\intercal}\nabla^2\phi(\mathbf{0})\boldsymbol{\lambda}^*. \tag{5.26}$$

The gradient and the Hessian of ϕ are

$$[\nabla\phi(\boldsymbol{\lambda})]_r = \sum_{s:(r,s)\in E} \frac{\alpha\mu_{rs}}{\alpha + (1-\alpha)e^{\lambda_r - \lambda_s}} - \frac{\alpha\mu_{sr}}{\alpha + (1-\alpha)e^{\lambda_s - \lambda_r}} + \sum_{s:(r,s)\notin E} \nu_{rs} - \nu_{sr}, \quad r \in \{1, \cdots, d\},$$

$$\nabla^2\phi(\boldsymbol{\lambda}) = -\alpha(1-\alpha) \sum_{(r,s)\in E} w_{rs}(\boldsymbol{\lambda})(\boldsymbol{e}_r - \boldsymbol{e}_s)(\boldsymbol{e}_r - \boldsymbol{e}_s)^\intercal,$$

with

$$w_{rs}(\boldsymbol{\lambda}) = \frac{\mu_{rs}e^{\lambda_r - \lambda_s}}{(\alpha + (1-\alpha)e^{\lambda_r - \lambda_s})^2} + \frac{\mu_{sr}e^{\lambda_s - \lambda_r}}{(\alpha + (1-\alpha)e^{\lambda_s - \lambda_r})^2},$$

and $\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d$ being the standard unit vectors in $\mathbb{R}^d$. The concavity inequality (5.26) becomes

$$\phi(\boldsymbol{\lambda}^*) \leq \alpha \sum_{(r,s)\in E} \mu_{rs}(\lambda_r^* - \lambda_s^*) + \sum_{(r,s)\notin E} \nu_{rs}(\lambda_r^* - \lambda_s^*) - \alpha(1-\alpha) \sum_{(r,s)\in E} \mu_{rs}(\lambda_r^* - \lambda_s^*)^2.$$

Substituting in the expression of $\phi(\boldsymbol{\lambda}^*)$, the term $\sum_{(r,s)\notin E} \nu_{rs}(\lambda_r^* - \lambda_s^*)$ cancels out on both sides and we get

$$\sum_{(r,s)\in E} \mu_{rs} \log\left(\frac{e^{\lambda_r^* - \lambda_s^*}}{\alpha + (1-\alpha)e^{\lambda_r^* - \lambda_s^*}}\right) \leq \alpha \sum_{(r,s)\in E} \mu_{rs}(\lambda_r^* - \lambda_s^*) - \alpha(1-\alpha) \sum_{(r,s)\in E} \mu_{rs}(\lambda_r^* - \lambda_s^*)^2,$$

which can be written as

$$\sum_{(r,s)\in E} \mu_{rs} \left(\alpha(1-\alpha)(\lambda_r^* - \lambda_s^*)^2 + (1-\alpha)(\lambda_r^* - \lambda_s^*) - \log\left(\alpha + (1-\alpha)e^{\lambda_r^* - \lambda_s^*}\right)\right) \leq 0. \tag{5.27}$$

Now we approximate the logarithm by the positive part: $\log(\alpha + (1-\alpha)e^x) \leq x_+ = \max\{0, x\}$ for all $x \in \mathbb{R}$ and $\alpha \in (0,1)$, so that we almost get a quadratic polynomial inequality. We make this a genuine quadratic inequality by applying the additional approximation that for all $x \in \mathbb{R}$ and $\alpha \in (0,1)$:

$$\alpha(1-\alpha)x^2 + (1-\alpha)x - x_+ \geq \frac{\alpha(1-\alpha)}{2}x^2 - \frac{1-\alpha}{2\alpha} - \frac{\alpha}{2(1-\alpha)}.$$

This is easy to check by verifying that the discriminants of the resulting quadratics (one for $x \geq 0$ and one for $x < 0$) are negative. Now, inequality (5.27) implies

$$\frac{\alpha(1-\alpha)}{2} \sum_{(r,s)\in E} \mu_{rs}(\lambda_r^* - \lambda_s^*)^2 \leq \left(\frac{1-\alpha}{2\alpha} + \frac{\alpha}{2(1-\alpha)}\right) \sum_{(r,s)\in E} \mu_{rs}.$$

In other words,

$$\sum_{(r,s)\in E} \mu_{rs}(\lambda_r^* - \lambda_s^*)^2 \leq \kappa(\alpha) \sum_{(r,s)\in E} \mu_{rs}.$$

■

5.7 Two proofs of Proposition 5.6

We first reduce the proof to the case where $G = K_d$ by a limiting argument. Let $G = (V, E)$ be a graph on d vertices. If G is not connected then the constraints defining the space $\mathcal{F}(G)$ decouple across the connected components of G and so does the integrand $\exp -\frac{1}{2}\sum_{(r,s)\in E} x_{rs}^2/w_{rs}$, therefore the Gaussian integral factors across the connected components of G. Hence, we may assume that G is connected. Now, if

$$\int_{\mathcal{F}} e^{-\frac{1}{2}\sum_{rs} x_{rs}^2/w_{rs}} \,\mathrm{d}\boldsymbol{x} = (2\pi)^{((d-1)^2+d)/2} \left(\frac{\prod_{r,s} w_{rs}}{T(\boldsymbol{w})}\right)^{1/2},$$

for all $\boldsymbol{w} \in \mathbb{R}_+^{d\times d}$ where $T = T_{K_d}$, then taking a limit $w_{rs} \to 0$ for all $(r,s) \notin E$, we get

$$\frac{1}{\left(\prod_{(r,s)\notin E} w_{rs}\right)^{1/2}} \int_{\mathcal{F}} e^{-\frac{1}{2}\sum_{rs} x_{rs}^2/w_{rs}} \,\mathrm{d}\boldsymbol{x} \longrightarrow c(G) \int_{\mathcal{F}(G)} e^{-\frac{1}{2}\sum_{(r,s)\in E} x_{rs}^2/w_{rs}} \,\mathrm{d}\boldsymbol{x},$$

where $c(G) > 0$ is a constant that only depends on G. On the other hand

$$T(\boldsymbol{w}) \longrightarrow \frac{\mathsf{nst}(G)}{2^{d-1}d^{d-2}} T_G(\boldsymbol{w}).$$

Therefore

$$c(G) \int_{\mathcal{F}(G)} e^{-\frac{1}{2}\sum_{(r,s)\in E} x_{rs}^2/w_{rs}} \,\mathrm{d}\boldsymbol{x} = (2\pi)^{((d-1)^2+d)/2} \left(\frac{2^{d-1}d^{d-2}}{\mathsf{nst}(G)} \frac{\prod_{(r,s)\in E} w_{rs}}{T_G(\boldsymbol{w})}\right)^{1/2}.$$

Now we set $w_{rs} = 1$ for all $(r, s) \in E$ to clear out the constants. Since $\int_{\mathcal{F}(G)} e^{-\frac{1}{2}\sum_{(r,s)\in E} x_{rs}^2} \,\mathrm{d}\boldsymbol{x} = (2\pi)^{\dim(\mathcal{F}(G))/2}$, we get

$$\int_{\mathcal{F}(G)} e^{-\frac{1}{2}\sum_{(r,s)\in E} x_{rs}^2/w_{rs}} \,\mathrm{d}\boldsymbol{x} = (2\pi)^{\dim(\mathcal{F}(G))/2} \left(\frac{\prod_{(r,s)\in E} w_{rs}}{T_G(\boldsymbol{w})}\right)^{1/2}.$$

Now it remains to prove the proposition for the complete graph.

A combinatorial proof

We proceed by adopting a combinatorial view on the structure of the space $\mathcal{F}$. This will lead us to consider a very special basis of $\mathcal{F}$ in which the computations become tractable. (Background on the concepts used in this construction can be found in (Biggs, 1997).) We first orient K_d in such a way that every pair of distinct vertices is connected by two parallel edges pointing in opposite directions. There are $d(d-1)$ (oriented) edges in total. Then, the subgraphs whose edges are weighted by an array $\boldsymbol{x} \in \mathcal{F}$ are called *Eulerian*: the total weight of the incoming edges is equal to that of the outgoing edges on each vertex. An important property of Eulerian graphs is that they can be decomposed into a superposition of cycles. In particular, fix a spanning tree T^* of K_d (the tree uses only one edge, if any, between each pair of vertices, and ignores their orientation). Every edge $e \notin T^*$ can be identified with the oriented cycle C_e in the graph which consists of the oriented edge e and the unique path between the endpoints of e in the tree T^* (where the direction of the edges on the path are flipped if necessary). Let $\boldsymbol{\chi}_e \in \{0, \pm 1\}^{d(d-1)}$ be the indicator vector of the cycle C_e[2]. Since a cycle is Eulerian, the vector $\boldsymbol{\chi}_e$—when folded into a $d \times d$ matrix—belongs to $\mathcal{F}$. Furthermore, the family $\{\boldsymbol{\chi}_e \ :\ e \notin T^*\}$ is linearly independent since a cycle C_e contains at least one edge—namely e—that is not contained in any other cycle $C_{e'}$, $e' \neq e$. There are exactly $d(d-1) - (d-1) = (d-1)^2$ off-tree edges in K_d, and this number coincides with the dimension of $\mathcal{F}$. Therefore $\mathcal{B} = \{\boldsymbol{\chi}_e \ :\ e \notin T^*\}$ is a basis of $\mathcal{F}$, that we henceforth call a *fundamental cycle basis* of $\mathcal{F}$.

Let $\boldsymbol{P} \in \{0, \pm 1\}^{(d-1)^2 \times d(d-1)}$ be the matrix where the rows are indexed by the off-tree edges of the graph, and whose eth row is equal to $\boldsymbol{\chi}_e$. The matrix $\boldsymbol{P}$ can be regarded as the *cycle-edge incidence matrix* of the graph K_d: an entry (e, e') is non-zero if and only if $e' \in C_e$.

Let $\boldsymbol{M} \in \mathbb{R}^{d(d-1)\times d(d-1)}$ be the diagonal matrix with entries w_{rs}, $r \neq s$ on the diagonal. Then by a change of variables

$$\begin{aligned}\int_{\mathcal{F}} e^{-\sum_{rs} x_{rs}^2/2w_{rs}} \mathrm{d}\boldsymbol{x} &= Det(\boldsymbol{P}\boldsymbol{P}^\intercal)^{1/2} \int_{\mathbb{R}^{(d-1)^2}} e^{-\boldsymbol{z}^\intercal(\boldsymbol{P}\boldsymbol{M}^{-1}\boldsymbol{P}^\intercal)\boldsymbol{z}/2} \mathrm{d}\boldsymbol{z} \\ &= (2\pi)^{(d-1)^2/2}\ Det(\boldsymbol{P}\boldsymbol{P}^\intercal)^{1/2} Det(\boldsymbol{P}\boldsymbol{M}^{-1}\boldsymbol{P}^\intercal)^{-1/2}.\end{aligned}$$

[2] Each non-zero entry in the vector corresponds to an edge present in the cycle, and the non-zero value is $+1$ if the cycle flows along the orientation of that edge, and -1 if the flow is in the opposite direction. In particular, the eth coordinate of $\boldsymbol{\chi}_e$ is always $+1$.

Now it remains to show that $Det(\boldsymbol{PM}^{-1}\boldsymbol{P}^{\intercal}) = \sum_T \prod_{(r,s)\notin T} w_{rs}^{-1}$ where the sum is over all spanning trees of K_d. This will finish the proof since we would then have $Det(\boldsymbol{PP}^{\intercal}) = \mathsf{nst}(K_d) = 2^{d-1}d^{d-2}$ by Cayley's formula on the number of spanning trees in the complete graph.

We expand the determinant using the Cauchy-Binet formula. Let $\boldsymbol{D} = \boldsymbol{M}^{-1/2}$, and let E be the set of edges in K_d. For a matrix $\boldsymbol{A}$ of size $n \times m$, $I \subseteq \{1, \cdots, n\}$, $J \subseteq \{1, \cdots, m\}$, we denote by $\boldsymbol{A}[I, J]$ the matrix of size $|I| \times |J|$ whose rows and columns are indexed by I and J respectively. If $I = \{1, \cdots, n\}$, then we write $\boldsymbol{A}[\ :\ ,\ J]$, and likewise for the column indices. Then, we have

$$Det(\boldsymbol{PM}^{-1}\boldsymbol{P}^{\intercal}) = \sum_{\substack{S\subseteq E \\ |S|=(d-1)^2}} Det(\boldsymbol{PD}[\ :\ ,\ S\ ])^2. \tag{5.28}$$

Now we use the following key lemma that we prove later.

Lemma 5.12. *Assuming the diagonal entries of the (diagonal) matrix $\boldsymbol{D}$ are positive, the matrix $\boldsymbol{PD}[\ :\ ,\ S\ ]$ is singular if and only if the graph spanned by the complement $\bar{S} = E \backslash S$ of S in K_d contains a cycle.*

Since there are exactly $(d-1)$ edges left unchosen by S, this lemma implies that they must form a spanning tree in order for the corresponding term to contribute to the sum in identity (5.28). Hence

$$Det(\boldsymbol{PM}^{-1}\boldsymbol{P}^{\intercal}) = \sum_{T\ :\ \text{spanning tree}} Det(\boldsymbol{PD}[\ :\ ,\ \bar{T}\ ])^2.$$

Fix a spanning tree T of K_d. Observe that if $T = T^*$ then the edges that generate the cycles in the fundamental cycle basis $\mathcal{B}$ are exactly the ones that are selected in $\bar{T}$. In other words, each row and each column of $\boldsymbol{PD}[\ :\ ,\ \bar{T}\ ]$ contain exactly one non-zero entry, (i.e., $\boldsymbol{P}[\ :\ ,\ \bar{T}\ ]$ is a permutation matrix), hence $Det(\boldsymbol{PD}[\ :\ ,\ \bar{T}\ ]) = \pm \prod_{(r,s)\notin T} w_{rs}^{-1/2}$. If $T \neq T^*$ then we split the set of edges in $\bar{T}$ into those that belong to T^* and those that do not. Each column in $\boldsymbol{PD}[\ :\ ,\ \bar{T}\ ]$ corresponding to an edge in $\bar{T} \cap \overline{T^*}$ contains only one non-zero entry (since this edge is contained in only one cycle in $\mathcal{B}$). Therefore all such edges (columns) along with the corresponding cycles (rows of the non-zero entry) can be successively eliminated from the determinant, yielding

$$Det(\boldsymbol{PD}[\ :\ ,\ \bar{T}\ ]) = \pm \left(\prod_{(r,s)\in\bar{T}\cap\bar{T}^*} w_{rs}^{-1/2} \right) \cdot Det\left(\boldsymbol{PD}[\ T \cap \overline{T^*}\ ,\ \bar{T} \cap T^*\ ]\right). \tag{5.29}$$

Notice that this operation has drastically reduced the size of the problem; the common size k of the sets $T \cap \overline{T^*}$ and $\bar{T} \cap T^*$ is anywhere between 0 and $d-1$ at most. Now we will show that

$$Det\left(\boldsymbol{PD}[\ T \cap \overline{T^*}\ ,\ \bar{T} \cap T^*\ ]\right) = \pm \prod_{(r,s)\in\bar{T}\cap T^*} w_{rs}^{-1/2},$$

using a peeling argument slightly more delicate than the one previously applied. Observe that $\boldsymbol{PD}[\, T \cap \overline{T^*}\ ,\ \bar{T} \cap T^*\,]$ is the $k \times k$ cycle-edge incidence matrix with k edges $T \cap \overline{T^*}$ indexing the rows and k edges in $\bar{T} \cap T^*$ indexing the columns, such that a row indexed by e indicates the edges $e' \in \bar{T} \cap T^*$ that participate in the cycle C_e.

So far, the spanning tree T^* was arbitrary. To continue, we choose T^* to be the *star tree* rooted at vertex 1 (see Figure 5.1, left). This choice simplifies the combinatorial argument to come, because the fundamental cycle basis $\mathcal{B}$ is now composed of triangles rooted at vertex 1. Crucially, this is where the assumption $G = K_d$ is needed; to ensure the existence of a star spanning tree. Figure 5.1 (right) illustrates the remaining edges after the first elimination procedure.

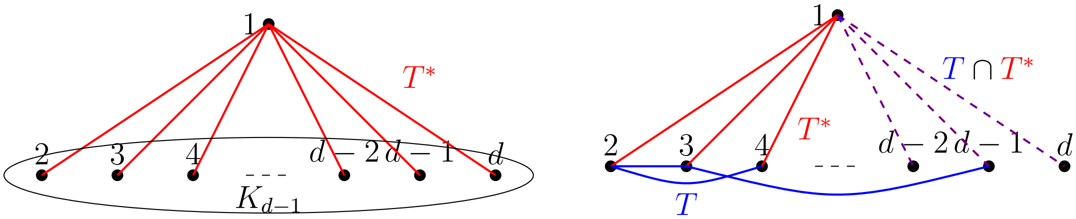

Figure 5.1: Left: the graph K_d where the star tree T^* is highlighted in red. Right: remaining edges in red and blue after the first elimination procedure (violet edges were removed).

Since T is a tree, by Lemma 9, each row and column of the matrix $\boldsymbol{PD}[\, T \cap \overline{T^*}\ ,\ \bar{T} \cap T^*\,]$ contains at least one non-zero entry. Furthermore, T^* being the star graph, each cycle $C_e \in \mathcal{B}$ is a triangle rooted at vertex 1, thus each row of the above matrix contains at most two non-zero entries. This is simply because one of the three edges that compose the triangle C_e—namely e—is not selected by the set $\bar{T} \cap T^*$ that indexes the columns of the matrix. See Figure 5.1, right (any blue edge has at most two adjacent red edges).

Furthermore, if all the rows contain exactly two non-zero entries then by the pigeonhole principle (since $|T \cap \overline{T^*}| = |\bar{T} \cap T^*|$), there will exist three edges in $T \cap \overline{T^*}$ that form a cycle C (see Figure 5.2, left). However, we assumed that T is a tree so this cannot happen. Therefore there must exist at least one row in the matrix with exactly one non-zero entry (i.e., there must exist an edge $e \in T \cap \overline{T^*}$ such that $C_e = \{e, e_1, e_2\} \in \mathcal{B}$ with $e_1 \in \bar{T} \cap T^*$ and $e_2 \in T \cap T^*$). See Figure 5.2.

Hence, we can eliminate this row and its corresponding column from the determinant. This corresponds to eliminating (dashing) the edges e and e_1 in the right figure above. Applying this argument iteratively allows us to peel all the edges and the cycles they belong to (see Figure 5.3), so that we obtain

$$Det\left(\boldsymbol{PD}[\, T \cap \overline{T^*}\ ,\ \bar{T} \cap T^*\,]\right) = \pm \prod_{(r,s) \in \bar{T} \cap T^*} w_{rs}^{-1/2}.$$

This completes the proof.

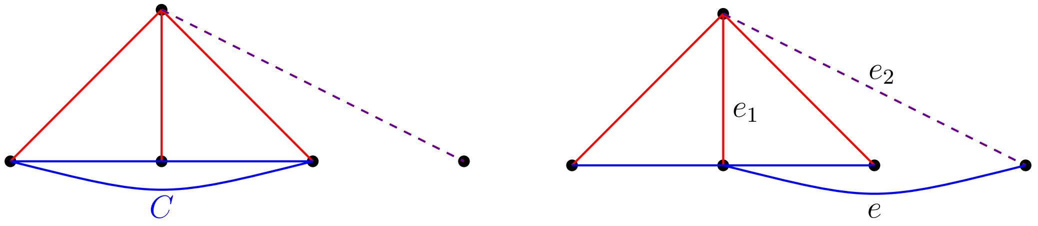

Figure 5.2: Left: an impossible situation where there remains a cycle C where no edge was eliminated in the first step. Right: a logical situation where there exist a fundamental cycle $C_e = (e, e_1, e_2)$ with one edge in T^* only, one edge in T only, and one edge in their intersection.

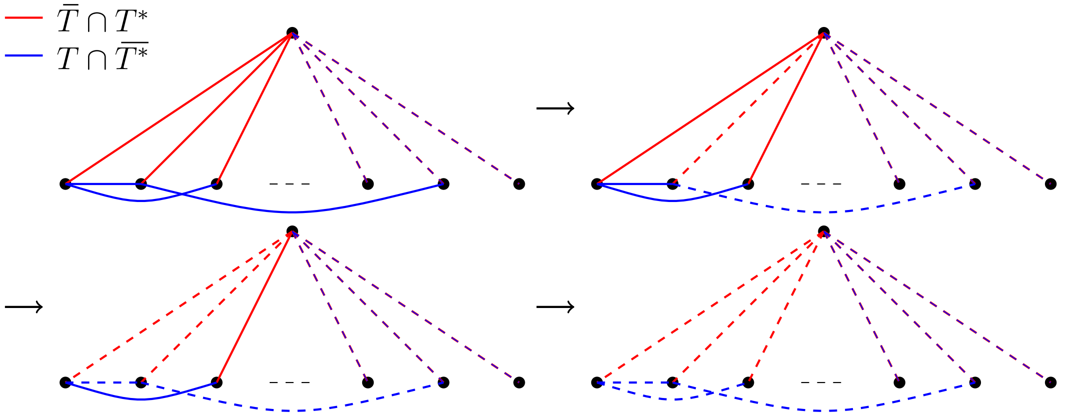

Figure 5.3: An illustration of the peeling process. "Wedges" with one edge in T only and the other in T^* only are eliminated successively until no edges remain. Violet edges were eliminated in the first step.

Proof of Lemma 5.12. Since we assumed the entries of the diagonal matrix $\boldsymbol{D}$ are strictly positive, we assume without loss of generality that $\boldsymbol{D}$ is the identity matrix. Assume now that the complement of S contains a cycle whose indicator vector is $\boldsymbol{\xi} \in \{0, \pm 1\}^{d(d-1)}$. Since $\mathcal{B}$ is a fundamental cycle basis, there exists $\boldsymbol{x} \in \mathbb{R}^{(d-1)^2} \backslash \{0\}$ such that $\boldsymbol{\xi} = \sum_{e \notin T^*} x_e \boldsymbol{\chi}_e = \boldsymbol{P}^\intercal \boldsymbol{x}$. Since S selects no edges in the cycle indicated by $\boldsymbol{\xi}$, it is clear that $\boldsymbol{x}^\intercal \boldsymbol{P}[\ :\ ,\ S\ ] = 0$, and this settles one direction. As for the other direction, let $\boldsymbol{x} \in \mathbb{R}^{(d-1)^2} \backslash \{0\}$ lie in the null space of $(\boldsymbol{P}[\ :\ ,\ S\ ])^\intercal$. The vector $\boldsymbol{P}^\intercal \boldsymbol{x}$ indicates the weights of a Eulerian subgraph in K_d (this vector belong to $\mathcal{F}$ when written in the form of a $d \times d$ matrix). The condition $(\boldsymbol{P}[\ :\ ,\ S\ ])^\intercal \boldsymbol{x} = 0$ implies that this Eulerian subgraph involves no edges from S. In particular, any cycle from this subgraph (there always exists one) is in the complement of E. This completes the proof. ∎

An analytic proof

This proof contrasts with the previous purely combinatorial approach in that it is mainly analytic. The approach relies on an interpolation argument that involves expressing the Gaussian integral over $\mathcal{F}$ as the *limit* of another parameterized Gaussian integral, when the parameter tends to zero. This latter integral can on the other hand be written in closed form, by relating it to the characteristic polynomial of a Laplacian matrix. Then the Principal Minors Matrix-Tree Theorem is invoked to finish the argument. This final step is the only place where combinatorics appear. (This proof approach was suggested to us by Andrea Sportiello.) Incidentally, this proof can be carried out with an arbitrary graph G; there is no need to reduce to the complete case. For $\delta > 0$ let

$$I(\delta) = \frac{1}{(2\pi\delta^2)^{(d-1)/2}} \int_{\mathbb{R}^{d\times d}} e^{-\frac{1}{2}\sum_{rs} x_{rs}^2/w_{rs}} \, e^{-\frac{1}{2\delta^2}\|(\boldsymbol{x}-\boldsymbol{x}^\intercal)\boldsymbol{1}\|_{\ell_2}^2} \, \mathrm{d}\boldsymbol{x}.$$

The additional Gaussian term in $I(\delta)$ gradually concentrates the mass of the integral on $\mathcal{F}$ as δ becomes small, and we have the following limiting statement:

Lemma 5.13. *We have*

$$\lim_{\delta\to 0} I(\delta) = c_d \int_{\mathcal{F}} e^{-\frac{1}{2}\sum_{rs} x_{rs}^2/2w_{rs}} \, \mathrm{d}\boldsymbol{x},$$

with

$$c_d = \frac{1}{(2\pi)^{(d-1)/2}} \int_{\mathcal{F}^\perp} e^{-2\|\boldsymbol{z}\boldsymbol{1}\|_{\ell_2}^2} \, \mathrm{d}\boldsymbol{z} = (2d)^{-(d-1)/2}.$$

On the other hand, a straightforward computation allows us to write $I(\delta)$ in closed form:

Lemma 5.14. *Let $G = (V, E)$ be a weighted graph with $V = \{1, \cdots, d\}$, $E = \{(r,s) \in V \times V, r \neq s\}$ where the edges are weighted by the array $\boldsymbol{w} \in \mathbb{R}_+^{d\times d}$. Let $\boldsymbol{L}(\boldsymbol{w}) \in \mathbb{R}^{d\times d}$ be the Laplacian matrix of G. For all $\delta > 0$, it holds that*

$$I(\delta) = (2\pi)^{((d-1)^2+d)/2} \left(\prod_{r,s} w_{rs}\right)^{1/2} \frac{\delta}{\mathit{Det}\left(\delta^2\boldsymbol{I} + \boldsymbol{L}(\boldsymbol{w})\right)^{1/2}}.$$

Now, by the Principal Minors Matrix-Tree Theorem (see, e.g., (Chaiken, 1982)), the characteristic polynomial of the Laplacian matrix of a graph admits the following expansion

$$\mathit{Det}\left(x\boldsymbol{I} + \boldsymbol{L}(\boldsymbol{w})\right) = \sum_{F} x^{|\mathrm{roots}(F)|} \prod_{(r,s)\in F} w_{rs},$$

where the sum is over all rooted spanning forests F of the graph. We finish the argument by taking a limit in δ:

$$\delta^{2(d-1)}\mathit{Det}\left(\boldsymbol{I} + \delta^{-2}\boldsymbol{L}(\boldsymbol{w})\right) = \delta^{-2}\mathit{Det}\left(\delta^2\boldsymbol{I} + \boldsymbol{L}(\boldsymbol{w})\right) \xrightarrow[\delta\to 0]{} d\sum_{T} \prod_{(r,s)\in T} w_{rs} = (2d)^{d-1}T(\boldsymbol{w}),$$

since the above limit singles out the rooted spanning forests with exactly one root—i.e., rooted spanning trees—from the characteristic polynomial, and there are d ways of choosing the root of a spanning tree. This exactly leads to the desired identity

$$\int_{\mathcal{F}} e^{-\frac{1}{2}\sum_{rs} x_{rs}^2/2w_{rs}} \,\mathrm{d}\boldsymbol{x} = (2\pi)^{((d-1)^2+d)/2} \left(\frac{\prod_{r,s} w_{rs}}{T(\boldsymbol{w})}\right)^{1/2}.$$

***Proof** of Lemma 5.13.* We decompose $\mathbb{R}^{d\times d}$ into the direct sum $\mathcal{F} \oplus \mathcal{F}^{\perp}$. It is easy to see that $\mathcal{F}^{\perp} = \{\boldsymbol{z} = \boldsymbol{\lambda}\mathbf{1}^{\intercal} - \mathbf{1}\boldsymbol{\lambda}^{\intercal},\ \boldsymbol{\lambda} \in \mathbb{R}^d\}$ which is a $(d-1)$-dimensional space. For $\boldsymbol{x} \in \mathbb{R}^{d\times d}$, let $\boldsymbol{y} \in \mathbb{R}^{d\times d}$ be its orthogonal projection on $\mathcal{F}$, and $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y}$. Therefore $(\boldsymbol{x} - \boldsymbol{x}^{\intercal})\mathbf{1} = (\boldsymbol{z} - \boldsymbol{z}^{\intercal})\mathbf{1} = 2\boldsymbol{z}\mathbf{1} = 2(d\boldsymbol{\lambda} - (\mathbf{1}^{\intercal}\boldsymbol{\lambda})\mathbf{1})$. For $\delta > 0$, we have

$$I(\delta) = \frac{1}{(2\pi\delta^2)^{(d-1)/2}} \int_{F\times F^{\perp}} e^{-\frac{1}{2}\sum_{r,s}(y_{rs}+z_{rs})^2/w_{rs}} \; e^{-\frac{2}{\delta^2}\|\boldsymbol{z}\mathbf{1}\|_{\ell_2}^2} \,\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{z}.$$

We make the change of variables $\boldsymbol{z}' = \boldsymbol{z}/\delta$:

$$I(\delta) = \frac{1}{(2\pi)^{(d-1)/2}} \int_{\mathcal{F}\times\mathcal{F}^{\perp}} e^{-\frac{1}{2}\sum_{r,s}(y_{rs}+\delta z'_{rs})^2/w_{rs}} \; e^{-2\|\boldsymbol{z}'\mathbf{1}\|_{\ell_2}^2} \,\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{z}'.$$

By dominated convergence,

$$\begin{aligned}\lim_{\delta\to 0} I(\delta) &= \frac{1}{(2\pi)^{(d-1)/2}} \int_{\mathcal{F}\times\mathcal{F}^{\perp}} e^{-\frac{1}{2}\sum_{r,s} y_{rs}^2/w_{rs}} \; e^{-2\|\boldsymbol{z}\mathbf{1}\|_{\ell_2}^2} \,\mathrm{d}\boldsymbol{y}\mathrm{d}\boldsymbol{z} \\ &= \frac{1}{(2\pi)^{(d-1)/2}} \int_{\mathcal{F}} e^{-\frac{1}{2}\sum_{r,s} y_{rs}^2/w_{rs}} \mathrm{d}\boldsymbol{y} \int_{\mathcal{F}^{\perp}} e^{-2\|\boldsymbol{z}\mathbf{1}\|_{\ell_2}^2} \mathrm{d}\boldsymbol{z}.\end{aligned}$$

Moreover,

$$\int_{\mathcal{F}^{\perp}} e^{-2\|\boldsymbol{z}\mathbf{1}\|_{\ell_2}^2} \mathrm{d}\boldsymbol{z} = (2d)^{(d-1)/2} \int_{\{\boldsymbol{\lambda}\in\mathbb{R}^d, \mathbf{1}^{\intercal}\boldsymbol{\lambda}=0\}} e^{-2d^2\|\boldsymbol{\lambda}\|_{\ell_2}^2} \,\mathrm{d}\boldsymbol{\lambda} = (2\pi)^{(d-1)/2}(2d)^{-(d-1)/2},$$

where the pre-factor in the first equality comes from the fact that $\|\boldsymbol{z}\|_F^2 = 2d\,\|\boldsymbol{\lambda}\|_{\ell_2}^2$ for $\boldsymbol{z} = \boldsymbol{\lambda}\mathbf{1}^{\intercal} - \mathbf{1}\boldsymbol{\lambda}^{\intercal},\ \boldsymbol{\lambda} \in \mathbb{R}^d,\ \mathbf{1}^{\intercal}\boldsymbol{\lambda} = 0$. ■

***Proof** of Lemma 5.14.* Let $\delta > 0$. We linearize the quadratic term $\|(\boldsymbol{x} - \boldsymbol{x}^{\intercal})\mathbf{1}\|_{\ell_2}^2$ in $I(\delta)$ by writing the corresponding Gaussian as the Fourier transform of another Gaussian: $\forall \boldsymbol{x} \in \mathbb{R}^{d\times d}$,

$$e^{-\frac{1}{2\delta^2}\|(\boldsymbol{x}-\boldsymbol{x}^{\intercal})\mathbf{1}\|_{\ell_2}^2} = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^d} e^{-\mathrm{i}\delta^{-1}\boldsymbol{y}^{\intercal}(\boldsymbol{x}-\boldsymbol{x}^{\intercal})\mathbf{1} - \frac{1}{2}\|\boldsymbol{y}\|_{\ell_2}^2} \,\mathrm{d}\boldsymbol{y},$$

where $\mathrm{i}^2 = -1$. Then

$$I(\delta) = \frac{1}{(2\pi\delta^2)^{(d-1)/2}} \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{R}^{d\times d}} \int_{\mathbb{R}^d} e^{-\frac{1}{2}\sum_{rs} x_{rs}^2/w_{rs}} \; e^{-\mathrm{i}\delta^{-1}\boldsymbol{y}^{\intercal}(\boldsymbol{x}-\boldsymbol{x}^{\intercal})\mathbf{1} - \frac{1}{2}\|\boldsymbol{y}\|_{\ell_2}^2} \,\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}.$$

We complete the square involving x_{rs} in the exponentiated expression:

$$-\frac{1}{2}\sum_{r,s} x_{rs}^2/w_{rs} - \mathbf{i}\delta^{-1}\boldsymbol{y}^\intercal(\boldsymbol{x}-\boldsymbol{x}^\intercal)\mathbf{1} = -\frac{1}{2}\sum_{r,s}\frac{1}{w_{rs}}\left(\left(x_{rs}+\mathbf{i}\frac{w_{rs}}{\delta}(y_r-y_s)\right)^2+\frac{w_{rs}^2}{\delta^2}(y_r-y_s)^2\right).$$

Then by Fubini's theorem,

$$I(\delta) = \frac{1}{(2\pi\delta^2)^{(d-1)/2}}\frac{1}{(2\pi)^{d/2}}\int_{\mathbb{R}^d} e^{-\frac{1}{2}\|\boldsymbol{y}\|_{\ell_2}^2-\frac{1}{2}\sum_{rs}\frac{w_{rs}}{\delta^2}(y_r-y_s)^2}\int_{\mathbb{R}^{d\times d}} e^{-\frac{1}{2}\sum_{rs}\frac{1}{w_{rs}}\left(x_{rs}+\mathrm{i}\frac{w_{rs}}{\delta}(y_r-y_s)\right)^2}\,\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}.$$

The inner integral evaluates to $\left(\prod_{r,s} 2\pi w_{rs}\right)^{1/2}$. Hence

$$\begin{aligned} I(\delta) &= \frac{(2\pi)^{(d-1)^2/2}}{\delta^{d-1}}\left(\prod_{r,s} w_{rs}\right)^{1/2}\int_{\mathbb{R}^d} e^{-\frac{1}{2}\|\boldsymbol{y}\|_{\ell_2}^2-\frac{1}{2}\sum_{r,s}\frac{w_{rs}}{\delta^2}(y_r-y_s)^2}\,\mathrm{d}\boldsymbol{y} \\ &= \frac{(2\pi)^{((d-1)^2+d)/2}}{\delta^{d-1}}\left(\prod_{r,s} w_{rs}\right)^{1/2} Det\left(\boldsymbol{I}+\delta^{-2}\boldsymbol{L}(\boldsymbol{w})\right)^{-1/2}, \end{aligned}$$

where $\boldsymbol{L}(\boldsymbol{w}) \in \mathbb{R}^{d\times d}$ is the Laplacian matrix of the weighted graph G. ■

5.8 Discussion

Our main result, Theorem 5.1, leaves a gap of essentially a factor of two between γ_{low} and γ_{up}. This is a limitation of the methods employed. In particular, it is plausible that the upper bound is loose due to a possible lack of concentration of the random variable $\mathcal{Z}$ about its mean, and this translates to the possibility of existence of a non-trivial interval inside $[\gamma_{\text{low}}, \gamma_{\text{up}}]$ where $\mathcal{Z}$ is typically close to 1 while its expectation is exponentially large. This is a standard issue in the use of the first (or second) moment method encountered in many random CSPs. Surprisingly enough—and as mentioned below 5.1—this is not the case in HQP, as it was shown by Scarlett and Cevher (2017), after a preliminary version of this result was made public, that γ_{up} is the sharp threshold. Therefore the first moment method does identify the phase transition in this problem.

Beyond our setting, the "sparse" regime where the sets S_a are of constant size k (exactly or on average) could also be of interest. Here, the relevant scaling is one where m is proportional to n. The lower bound argument could be easily extended and yields a bound of $\frac{H(\boldsymbol{\pi})}{(d-1)\log k}$. As for the upper bound, one could in principle follow the same first moment strategy, but our analysis breaks in a quite serious fashion, in that none of our asymptotic estimates hold true in this regime.

Chapter 6

Decoding from pooled Data: phase transitions of massage-passing

We recall the setting of the *Histogram Query Problem*: let $\tau^* : \{1, \cdots, n\} \mapsto \{1, \cdots, d\}$ be an assignment of n variables to d categories. We denote the queried subpopulations by $S_a \subset \{1, \cdots, n\}$, $1 \leq a \leq m$. Given m subsets S_a, the histogram of categories of the pooled subpopulation S_a is denoted by $\boldsymbol{h}_a \in \mathbb{Z}_+^d$, i.e., for all $1 \leq a \leq m$,

$$\boldsymbol{h}_a := \left(\left|\tau^{*-1}(1) \cap S_a\right|, \cdots, \left|\tau^{*-1}(d) \cap S_a\right|\right). \tag{6.1}$$

We let $\boldsymbol{\pi} = \frac{1}{n}\left(\left|\tau^{*-1}(1)\right|, \cdots, \left|\tau^{*-1}(d)\right|\right)$ denote the vector of proportions of assigned values; i.e., the empirical distribution of categories. We place ourselves in a random dense regime in which the sets $\{S_a\}_{1 \leq a \leq m}$ are independent draws of a random set S where $\Pr(i \in S) = \alpha$ independently for each $i \in \{1, \ldots, n\}$, for some fixed $\alpha \in (0, 1)$. Meaning, at each query, the size of the pool is proportional to the size of the population: $\mathbb{E}[|S|] = \alpha n$.

Here we adopt a linear-algebraic formulation which will be more convenient for the presentation of the algorithm. We can represent the map τ^*, which we refer to as *the planted solution*, as a set of vectors $\boldsymbol{x}_i^* = \boldsymbol{e}_{\tau^*(i)} \in \mathbb{R}^d$, for $1 \leq i \leq n$. Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ represent the sensing matrix: $A_{ai} = \mathbb{1}\{i \in S_a\}$, for all $1 \leq i \leq n, 1 \leq a \leq m$. The histogram equations (6.1) can be written in the form of a linear system of m equations:

$$\boldsymbol{h}_a = \sum_{i=1}^{n} A_{ai} \boldsymbol{x}_i^*, \qquad a \in \{1, \cdots, m\}. \tag{6.2}$$

Our goal can thus be rephrased as that of inverting the linear system (6.2). Note that the problem becomes trivial if $m = n$, since the square random matrix $\boldsymbol{A}$ will be invertible with high probability. However, as we have seen in the previous chapter, a detailed information-theoretic analysis of the problem shows that the planted solution is uniquely determined by the above linear system for $m = \gamma \frac{n}{\log n}$, $\gamma > 0$. In this chapter we study the algorithmic problem in the regime $m = \kappa n$, $\kappa < 1$.

We note that in the deterministic setting, where one is allowed to design the sensing matrix $\boldsymbol{A}$, i.e. choose the pools S_a at each query, Wang et al. (2016) provided a querying strategy that recovers τ^* provided that $m > c_1 \frac{n}{\log n}$, where c_1 is an absolute constant. Ignoring the dependence on d, this almost matches the information-theoretic limit. The random setting has not been treated so far.

We present an Approximate Message Passing (AMP) algorithm for the *random dense* setting, where each query involves a random subset of individuals of size proportional to n. We characterize the exact asymptotic behavior of the algorithm in the limit of large number of individuals n and a proportionally large number of queries m, i.e. $m/n \to \kappa$. This is done by heuristically deriving the corresponding State Evolution (SE) equations corresponding to the AMP algorithm. Then, a rigorous analysis of the SE dynamics reveals a rich and interesting behavior; namely the existence of phase transition phenomena in the parameters $\kappa, d, \boldsymbol{\pi}$ of the problem, due to which the behavior of AMP changes radically, from exact recovery to very weak correlation with the planted solution. We exactly locate these phase transitions in simple situations, such as the binary case $d = 2$, the symmetric case $\boldsymbol{\pi} = (\frac{1}{d}, \cdots, \frac{1}{d})$, and the general case under the condition that the SE iteration is initialized from a special point. The latter exhibits an intriguing phenomenon: the existence of not one, but an entire sequence of thresholds in the parameter κ that rules the behavior of the SE dynamics. These thresholds correspond to sharp changes in the structure of the covariance matrix of the estimates output by AMP. We expect this phenomenon to be generic beyond the special initialization case studied here. Beyond the precise characterization of the phase transition thresholds in these special cases, we initiate the study of State Evolution in a multivariate setting by proving the convergence of the full-dimensional SE iteration, when initialized from a "far enough" point, to a fixed point, and show further properties of the iterate sequence.

6.1 Approximate message passing and state evolution

In this section we present the Approximate Message Passing (AMP) algorithm and the corresponding State Evolution (SE) equations.

The AMP algorithm

The AMP algorithm of Donoho, Maleki, and Montanari (2009), known as the TAP equations, after Thouless, Anderson, and Palmer (1977), can be derived from Belief Propagation (BP) on the factor graph modeling the recovery problem. The latter is a bipartite graph of $n+m$ vertices. The variables $\{\boldsymbol{x}_i : 1 \leq i \leq n\}$ constitute one side of the bipartition, and the observations $\{\boldsymbol{h}_a : 1 \leq a \leq m\}$ constitute the other side. The adjacency structure is encoded in the sensing matrix $\boldsymbol{A}$. Endowing each edge (i, a) with two messages $\boldsymbol{m}_{i\to a}, \boldsymbol{m}_{a\to i} \in \Delta^{d-1}$, Δ^{d-1} being the probability simplex, one can write the self-consistency equations for the messages at each node by enforcing the histogram constraints at each observation (or check) node while treating the incoming messages as probabilistically independent in the marginalization operation. The iterative version of these self-consistency equations is the BP algorithm. BP is further simplified to AMP by exploiting the fact that the factor graph is random and dense, i.e. one only needs to track the average of the messages incoming to each node. This reduces the number of passed messages from $m \times n$ to $m+n$. For the present d-variate problem, the algorithm we present is a special case of Hybrid-GAMP of Rangan et al. (2012). We let $\bar{\boldsymbol{h}}_a = (\boldsymbol{h}_a - \alpha n \boldsymbol{\pi})/\sqrt{n}$ and $\bar{\boldsymbol{A}} = (\boldsymbol{A} - \alpha \mathbf{1}_m \mathbf{1}_n^\intercal)/\sqrt{n}$ be the centered and rescaled data, and assume that the parameters α and $\boldsymbol{\pi}$ are known to the algorithm. The AMP algorithm reads as follows: At iteration $t = 1, 2, \dots,$ we update the check nodes $a = 1, \cdots, m$ as

$$\begin{aligned}
\boldsymbol{\omega}_a^t &= \sum_{j \in \partial a} \bar{A}_{aj} \hat{\boldsymbol{x}}_j^t - \boldsymbol{V}_a^t \left(\boldsymbol{V}_a^{t-1}\right)^{-1} (\bar{\boldsymbol{h}}_a - \boldsymbol{\omega}_a^{t-1}), \\
\boldsymbol{V}_a^t &= \sum_{j \in \partial a} \bar{A}_{aj}^2 \boldsymbol{B}_j^t,
\end{aligned}$$

and then update the variable nodes $i = 1, \cdots, n$ as

$$\begin{aligned}
\boldsymbol{z}_i^t &= \hat{\boldsymbol{x}}_i^t + \boldsymbol{\Sigma}_i^t \cdot \sum_{b \in \partial i} \bar{A}_{bi} \left(\boldsymbol{V}_b^t\right)^{-1} (\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_b^t), \\
\boldsymbol{\Sigma}_i^t &= \left(\sum_{b \in \partial i} \bar{A}_{bi}^2 \left(\boldsymbol{V}_b^t\right)^{-1} \right)^{-1}, \\
\hat{\boldsymbol{x}}_i^{t+1} &= \boldsymbol{\eta}(\boldsymbol{z}_i^t, \boldsymbol{\Sigma}_i^t), \\
\boldsymbol{B}_i^{t+1} &= \mathrm{Diag}(\hat{\boldsymbol{x}}_i^{t+1}) - \hat{\boldsymbol{x}}_i^{t+1} \cdot \hat{\boldsymbol{x}}_i^{t+1\intercal},
\end{aligned}$$

with

$$\boldsymbol{\eta}(\boldsymbol{z}, \boldsymbol{\Sigma}) := \sum_{r=1}^{d} \pi_r \boldsymbol{e}_r \frac{e^{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{e}_r)^\intercal \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{e}_r)}}{Z(\boldsymbol{z}, \boldsymbol{\Sigma})} \in \mathbb{R}^d, \tag{6.3}$$

where $Z(\boldsymbol{z}, \boldsymbol{\Sigma}) = \sum_{r=1}^{d} \pi_r e^{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{e}_r)^\intercal \boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{e}_r)}$ is a normalization factor so that the entries of $\boldsymbol{\eta}$ sum to one. The map $\boldsymbol{\eta}$ plays the role of a "thresholding function" with a matrix parameter $\boldsymbol{\Sigma}$ that is adaptively tuned by the algorithm. One should compare this situation to the case of sparse estimation (Donoho, Maleki, and Montanari, 2009) where the soft thresholding function is used. Here, the form taken by $\boldsymbol{\eta}$ is adapted to the structure of the signal we seek to recover. The variables $\boldsymbol{\omega}_a$ and $\boldsymbol{V}_a$ represent estimates of the histogram $\boldsymbol{h}_a$ and their variances. The variables $\boldsymbol{z}_i$ and $\boldsymbol{\Sigma}_i$ are estimators of the planted solution $\boldsymbol{x}_i^*$ and their variances before thresholding, while $\hat{\boldsymbol{x}}_i \in \Delta^{d-1}$ and $\boldsymbol{B}_i$ are the posterior estimates of $\boldsymbol{x}_i^*$ and its variance, i.e., after thresholding. The algorithm can be initialized in a "non-informative" way by setting $\hat{\boldsymbol{x}}_i^0 = \boldsymbol{\pi}, \boldsymbol{B}_i^0 = \mathrm{Diag}(\boldsymbol{\pi}) - \boldsymbol{\pi}\boldsymbol{\pi}^\intercal$ for all $i = 1, \dots, n$, and $\boldsymbol{\omega}_a^{-1} = \mathbf{0}$ and $\boldsymbol{V}_a^{-1} = \boldsymbol{I}$ for all $a = 1, \cdots, m$ for example.s We defer the details of the derivation to Section 6.5.

State evolution

State Evolution (SE) (Bayati, Lelarge, and Montanari, 2012; Donoho, Maleki, and Montanari, 2009), a version of the cavity method of statistical physics (Mézard, Parisi, and Virasoro, 1990), allows us to exactly characterize the asymptotic behavior of AMP at each time step t, by tracking the evolution in time of the relevant *order parameters* of the algorithm. More precisely, let

$$\boldsymbol{M}_{t,n} := \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{x}}_i^t \boldsymbol{x}_i^{*\intercal}, \qquad \text{and} \qquad \boldsymbol{Q}_{t,n} := \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{x}}_i^t \hat{\boldsymbol{x}}_i^{t\intercal}.$$

The matrix $\boldsymbol{M}_{t,n}$ tracks the average alignment of the estimates with the true solution, and $\boldsymbol{Q}_{t,n}$ their average covariance structure. The SE equations relate the values of these order parameters at $t{+}1$ to those at time t in the limit $n \to \infty$, $m/n \to \kappa$. We let $\boldsymbol{M}_t$ and $\boldsymbol{Q}_t$ denote the respective limits of $\boldsymbol{M}_{t,n}$ and $\boldsymbol{Q}_{t,n}$, which we assume exist in this "replica-symmetric" regime, and let $\boldsymbol{D} = \mathrm{Diag}(\boldsymbol{\pi})$. The SE equations read

$$\begin{aligned}
\boldsymbol{M}_{t+1} &= \sum_{r=1}^{d} \pi_r \, \mathbb{E}_{\boldsymbol{g}} \left[\boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}_t^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R}_t)\right] \cdot \boldsymbol{e}_r^\intercal, \\
\boldsymbol{Q}_{t+1} &= \sum_{r=1}^{d} \pi_r \, \mathbb{E}_{\boldsymbol{g}} \left[\boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}_t^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R}_t) \cdot \boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}_t^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R}_t)^\intercal\right], \\
\boldsymbol{X}_t &= \kappa^{-1}(\boldsymbol{D} - \boldsymbol{M}_t - \boldsymbol{M}_t^\intercal + \boldsymbol{Q}_t), \\
\boldsymbol{R}_t &= \mathrm{Diag}(\boldsymbol{Q}_t \mathbf{1}) - \boldsymbol{Q}_t,
\end{aligned}$$

with $\boldsymbol{g} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. The matrix $\kappa\boldsymbol{X}_t$ is the covariance matrix of the error of the estimates output by AMP at time t, and $\boldsymbol{R}_t$ can be interpreted as the average covariance matrix of the estimates themselves. Note that the parameter α has disappeared from the characterization by the SE equations, just as in the information theoretic study.

The full derivation of these equations can be found in Section 6.6. The main hypothesis behind the derivation, which we *do not* rigorously verify, is that the variables $\boldsymbol{z}_i^t$ are asymptotically Gaussian, centered about $\boldsymbol{x}_i^*$ and with covariance $\boldsymbol{X}_t$: the measure $\frac{1}{n}\sum_{i=1}^n \delta_{\boldsymbol{z}_i^t - \boldsymbol{x}_i^*}$ converges weakly to $\mathcal{N}(\mathbf{0}, \boldsymbol{X}_t)$. We refer to Bayati, Lelarge, and Montanari (2012); Bayati and Montanari (2011) for rigorous results, the assumptions of which do not apply to this setting. It is an interesting problem to prove the exactness of the SE equations in this setting.

Simplification of SE

Here we simplify the system of SE equations above to a single iteration. This crucially relies on the following Proposition:

Proposition 6.1. *If $\boldsymbol{M}_0 = \boldsymbol{Q}_0$, then for all t we have* *(i)* $\boldsymbol{M}_t = \boldsymbol{Q}_t$*. In particular,* $\boldsymbol{M}_t$ *is a symmetric PSD matrix, and* $\boldsymbol{M}_t\boldsymbol{1} = \boldsymbol{\pi}$*.*

(ii) $\boldsymbol{R}_t = \kappa\boldsymbol{X}_t = \boldsymbol{D} - \boldsymbol{M}_t$*.*

The proof of the above proposition is deferred to Section 6.4. We pause to make a few remarks. The assumption of Proposition 6.1 could be enforced for example by setting the initial estimates of AMP as $\hat{\boldsymbol{x}}_i^0 = \boldsymbol{\pi}$ for all i. This yields $\boldsymbol{M}_0 = \boldsymbol{Q}_0 = \boldsymbol{\pi}\boldsymbol{\pi}^\intercal$, and hence $\boldsymbol{X}_0 = \kappa^{-1}(\boldsymbol{D} - \boldsymbol{\pi}\boldsymbol{\pi}^\intercal)$. The statements in this proposition can be seen as "dynamic" versions of the Nishimori property discussed in the first three chapters (see Zdeborová and Krzakala, 2016). They simplify the SE equations to a single iteration on $\boldsymbol{X}_t$. To succinctly present this simplification, for $r \in \{1, \cdots, d\}$, and $\boldsymbol{X} \succeq \boldsymbol{0}$, we let

$$\boldsymbol{\eta}_r(\boldsymbol{X}) := \boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}^{\frac{1}{2}}\boldsymbol{g}, \boldsymbol{X}) \in \Delta^{d-1}.$$

Then, the SE equations can be seen to boil down to the single equation

$$\boldsymbol{X}_{t+1} = \kappa^{-1} f(\boldsymbol{X}_t), \tag{6.4}$$

where, recalling that $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$, we define

$$\begin{aligned} f(\boldsymbol{X}) &:= \boldsymbol{D} - \sum_{r=1}^{d} \pi_r \, \mathbb{E}_{\boldsymbol{g}}\left[\boldsymbol{\eta}_r(\boldsymbol{X})\boldsymbol{\eta}_r(\boldsymbol{X})^\intercal\right] &(6.5)\\ &= \boldsymbol{D} - \sum_{r=1}^{d} \pi_r \, \mathbb{E}_{\boldsymbol{g}}\left[\boldsymbol{\eta}_r(\boldsymbol{X})\right] \cdot \boldsymbol{e}_r^\intercal, &(6.6)\\ &= \sum_{r=1}^{d} \pi_r \, \mathbb{E}_{\boldsymbol{g}}\left[(\boldsymbol{e}_r - \boldsymbol{\eta}_r(\boldsymbol{X})) \cdot (\boldsymbol{e}_r - \boldsymbol{\eta}_r(\boldsymbol{X}))^\intercal\right], &(6.7) \end{aligned}$$

where equations (6.5) and (6.6) correspond to substituting the value of $\boldsymbol{Q}_t$ and $\boldsymbol{M}_t$ into statement (ii) of the above proposition, while the last equality (6.7) is just a consequence of the first two, (6.5) and (6.6). Furthermore, via elementary algebra, the coordinates of the vector $\boldsymbol{\eta}_r(\boldsymbol{X})$ can written as

$$(\boldsymbol{\eta}_r(\boldsymbol{X}))_s = \frac{\pi_s \exp\left(-\boldsymbol{g}^\intercal \boldsymbol{X}^{-\frac{1}{2}}(\boldsymbol{e}_r - \boldsymbol{e}_s) - \frac{1}{2}\left\|\boldsymbol{X}^{-\frac{1}{2}}(\boldsymbol{e}_r - \boldsymbol{e}_s)\right\|_{\ell_2}^2\right)}{Z_r(\boldsymbol{X})}, \tag{6.8}$$

with

$$Z_r(\boldsymbol{X}) := \sum_{s=1}^{d} \pi_s \exp\left(-\boldsymbol{g}^\intercal \boldsymbol{X}^{-\frac{1}{2}}(\boldsymbol{e}_r - \boldsymbol{e}_s) - \frac{1}{2}\left\|\boldsymbol{X}^{-\frac{1}{2}}(\boldsymbol{e}_r - \boldsymbol{e}_s)\right\|_{\ell_2}^2\right).$$

The mean squared & 0-1 errors

We can measure the performance of AMP by the mean squared error of the estimates $\{\hat{\boldsymbol{x}}_i^t\}_{i=1}^n$:

$$\mathsf{MSE}_{t,n} = \frac{1}{n}\sum_{i=1}^n \left\|\hat{\boldsymbol{x}}_i^t - \boldsymbol{x}_i^*\right\|_{\ell_2}^2 .$$

Since $\hat{\boldsymbol{x}}_i^t \in \Delta^{d-1}$, an alternative measure of performance would be the expected 0-1 distance between a random category drawn from the multinomial $\hat{\boldsymbol{x}}_i$ and the true category $\boldsymbol{x}_i^*$, then averaged over $i = 1, \cdots, n$. This error would be written as

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n\sum_{r=1}^d \hat{x}_{ir}^t(1 - \boldsymbol{e}_r^\intercal \boldsymbol{x}_i^*) &= 1 - \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{x}}_i^{t\intercal}\boldsymbol{x}_i^* \\ &= 1 - \mathrm{trace}(\boldsymbol{M}_{t,n}) = \mathrm{trace}\left(\boldsymbol{D} - \boldsymbol{M}_{t,n}\right). \end{aligned}$$

On the other hand, the MSE in the large n limit reads

$$\begin{aligned} \mathsf{MSE}_t := \lim_{n\to\infty} \mathsf{MSE}_{t,n} &= \mathrm{trace}\left(\boldsymbol{Q}_t - \boldsymbol{M}_t - \boldsymbol{M}_t^\intercal + \boldsymbol{D}\right), \\ &= \mathrm{trace}\left(\boldsymbol{D} - \boldsymbol{M}_t\right), \end{aligned}$$

so the two notions of error coincide in the limit. Note that the MSE at each step t can be deduced from SE iterate at time t: $\mathsf{MSE}_t = \kappa\,\mathrm{trace}(\boldsymbol{X}_t)$.

6.2 Analysis of the state evolution dynamics

In this section we present our main results on the convergence of the SE iteration (6.4) to a fixed point, and the location of the phase transition thresholds in three special cases. We start by analyzing the SE map f and present some important generic results.

Analysis of the SE map f

From expression (6.7), we see that the map f sends the positive semi-definite (PSD) cone $\mathbb{S}_+^{d\times d}$ to itself. As written, f is only defined for invertible matrices $\boldsymbol{X}$, but it could be extended by continuity to singular matrices: if $\boldsymbol{e}_r - \boldsymbol{e}_s$ is in the null space of $\boldsymbol{X}$, we declare that $\exp(-\frac{1}{2}\|\boldsymbol{X}^{-\frac{1}{2}}(\boldsymbol{e}_r - \boldsymbol{e}_s)\|^2) = 0$. This convention is consistent with the limiting value of a sequence $\left\{\exp(-\frac{1}{2}\|\boldsymbol{X}_n^{-\frac{1}{2}}(\boldsymbol{e}_r - \boldsymbol{e}_s)\|^2)\right\}_{n\geq 0}$ where $\left\{\boldsymbol{X}_n\right\}_{n\geq 0}$ is a sequence of invertible matrices approaching $\boldsymbol{X}$. This also has an interpretation based on an analogy with electrical circuits, which we discuss shortly. This extension will be also denoted by f. It is continuous over the $\mathbb{S}_+^{d\times d}$, and we have $f(\mathbf{0}) = \mathbf{0}$. Now, we state an important property of f, namely that it is monotone:

Proposition 6.2. *The map f is order-preserving on $\mathbb{S}_+^{d\times d}$; i.e., for all $\boldsymbol{X}, \boldsymbol{Y} \succeq \boldsymbol{0}$, if $\boldsymbol{X} \preceq \boldsymbol{Y}$ then $f(\boldsymbol{X}) \preceq f(\boldsymbol{Y})$.*

The proof of this Proposition is conceptually simple but technical, and is thus deferred to Section 6.4. Next, we adopt a combinatorial view of the structure of the SE dynamics. This will help us identity subspaces of $\mathbb{S}_+^{d\times d}$ that are left invariant by f. Note that the definition of f involves $\boldsymbol{X}^{-\frac{1}{2}}$ acting on $\mathrm{span}(\boldsymbol{1})^{\perp}$. Additionally, it is easy to verify that for all $\boldsymbol{X} \in \mathbb{S}_+^{d\times d}$, $f(\boldsymbol{X})\boldsymbol{1} = \boldsymbol{0}$, and $f(\boldsymbol{X})_{rs} \le 0$ for all $r \neq s$. Therefore, without loss of generality, we can restrict the study of the state evolution iteration to the set

$$\mathcal{A} := \left\{ \boldsymbol{X} \in \mathbb{S}_+^{d\times d},\ \boldsymbol{X}\boldsymbol{1} = \boldsymbol{0},\ X_{rs} \le 0\ \forall (r,s) \text{ s.t. } r \neq s \right\},$$

since it is invariant under the dynamics. The set $\mathcal{A}$ can be seen as the set of Laplacian matrices of weighted graphs on d vertices (every edge (r,s) is weighted by $-X_{rs}$ for $\boldsymbol{X} \in \mathcal{A}$). Hence f can be seen as a transformation on weighted graphs. This transformation enjoys the following invariance property:

Proposition 6.3. *For all $\boldsymbol{X} \in \mathcal{A}$, f preserves the connected component structure of the graph represented by $\boldsymbol{X}$; i.e, two distinct connected components of the graph whose Laplacian matrix is $\boldsymbol{X}$ remain distinct when transformed by f.*

Proof. The proof relies on the concept of *effective resistance*. One can view a graph of Laplacian $\boldsymbol{X} \in \mathcal{A}$ as a network of resistors with resistances $1/(-X_{rs})$. The effective resistance of an edge (r,s) is the resistance of the entire network when one unit of current is injected at r and collected at s (or vice-versa). Its expression is a simple consequence of Kirchhoff's law, and is equal to $R_{rs} := \left\|\boldsymbol{X}^{-1/2}(\boldsymbol{e}_r - \boldsymbol{e}_s)\right\|_{\ell_2}^2$ (see e.g. Spielman, n.d.). It is clear that the effective resistance of an edge is finite if and only if both its endpoints belong to the same connected component of the graph, otherwise $R_{rs} = +\infty$, and $(\boldsymbol{\eta}_r(\boldsymbol{X}))_s = 0$. This causes f to "factor" across connected components, and thus acts on them independently. ■

Next, let us look at the limit of $f(t\boldsymbol{X})$ for large t. For $\boldsymbol{X} \in \mathcal{A}$ invertible on $\mathrm{span}(\boldsymbol{1})^{\perp}$, we have $\lim_{t\to\infty} f(t\boldsymbol{X}) = \boldsymbol{D} - \boldsymbol{\pi}\boldsymbol{\pi}^{\intercal}$, since $\boldsymbol{\eta}_r(t\boldsymbol{X}) \to \boldsymbol{\pi}$ almost surely. More generally, if $\boldsymbol{X}$ represents a graph with $\{V_k\}_{1\le k\le K}$ connected components, $(\boldsymbol{\eta}_r(t\boldsymbol{X}))_s \neq 0$ only if r, s are in the same component. Hence, $\boldsymbol{\eta}_r(t\boldsymbol{X}) \to \frac{\boldsymbol{P}_k\boldsymbol{\pi}}{\boldsymbol{1}^{\intercal}\boldsymbol{P}_k\boldsymbol{\pi}}$, where $\boldsymbol{P}_k$ is the orthogonal projector onto the span of the coordinates in V_k where $r \in V_k$, and we have

$$\lim_{t\to\infty} f(t\boldsymbol{X}) = \boldsymbol{D} - \sum_{k=1}^{K} \frac{\boldsymbol{P}_k\boldsymbol{\pi}\boldsymbol{\pi}^{\intercal}\boldsymbol{P}_k}{\boldsymbol{1}^{\intercal}\boldsymbol{P}_k\boldsymbol{\pi}} =: \boldsymbol{L}_K. \tag{6.9}$$

By Propositions 6.2 and 6.3 and the limit calculation (6.9), we deduce that for any partition $\{V_k\}_{1\le k\le K}$ of $\{1,\cdots,d\}$, and all Laplacian matrices $\boldsymbol{X} \succeq \boldsymbol{0}$ of graphs with connected components $V_1,\cdots,V_K$, we have

$$f(\boldsymbol{X}) \preceq \boldsymbol{L}_K. \tag{6.10}$$

Indeed, since $\boldsymbol{X} \preceq t\boldsymbol{X}$ for all $t \geq 1$, we have $f(\boldsymbol{X}) \preceq f(t\boldsymbol{X})$ by monotonicity of f. Letting $t \to \infty$ settles the claim. In particular, with $K = 1$, $\boldsymbol{L}_1 = \boldsymbol{D} - \boldsymbol{\pi}\boldsymbol{\pi}^\intercal$, and for all $\boldsymbol{X} \in \mathcal{A}$ representing a connected graph (i.e. $\mathrm{rank}(\boldsymbol{X}) = d - 1$), we have $f(\boldsymbol{X}) \preceq \boldsymbol{D} - \boldsymbol{\pi}\boldsymbol{\pi}^\intercal$. We are now ready to state the main result of this subsection.

Theorem 6.4. *Let $\{V_k\}_{1\leq k\leq K}$ be a partition of $\{1, \cdots, d\}$, and $\boldsymbol{L}_K$ defined as in (6.9). Let $\boldsymbol{X}_0 \in \mathcal{A}$ with connected components $V_1, \cdots, V_K$, and such that $\boldsymbol{X}_0 \succeq \kappa^{-1}\boldsymbol{L}_K$. If the SE iteration (6.4) is initialized from $\boldsymbol{X}_0$, then the sequence $\{\boldsymbol{X}_t\}_{t\geq 0}$ is decreasing in the PSD order, i.e., $\boldsymbol{X}_t \preceq \boldsymbol{X}_{t-1}$ for all $t \geq 1$, and converges to a fixed point $\boldsymbol{X}^*$, i.e., $\boldsymbol{X}^* = \kappa^{-1}f(\boldsymbol{X}^*)$.*

Proof. Let $\boldsymbol{X}_0$ satisfy the conditions of the Theorem. Using $\boldsymbol{X}_0 \succeq \kappa^{-1}\boldsymbol{L}_K$ and observation (6.10), we have $\boldsymbol{X}_1 = \kappa^{-1}f(\boldsymbol{X}_0) \preceq \boldsymbol{X}_0$. By monotonicity of f, we deduce that the SE iterates form a monotone sequence: $\boldsymbol{X}_{t+1} \preceq \boldsymbol{X}_t$ for all $t \geq 0$. Since $\boldsymbol{X}_t \succeq \boldsymbol{0}$ for all t, then this sequence must have a limit[1] $\boldsymbol{X}^* \succeq \boldsymbol{0}$. By continuity of f, this limit must satisfy $\boldsymbol{X}^* = \kappa^{-1}f(\boldsymbol{X}^*)$. ■

We expect that for κ large enough, $\boldsymbol{X}^* = \boldsymbol{0}$, meaning that $\lim \boldsymbol{M}_t = \boldsymbol{D}$ and $\lim \mathsf{MSE}_t = 0$. This situation corresponds to perfect recovery of the planted solution $\{\boldsymbol{x}_i^*\}_{i=1}^n$ by AMP. We can easily show that this is the case for

$$\kappa > \kappa^* := \sup_{\boldsymbol{X}\in\mathcal{A}} \frac{\lambda_{\max}(f(\boldsymbol{X}))}{\lambda_{\max}(\boldsymbol{X})}. \tag{6.11}$$

Indeed,

$$\lambda_{\max}(\boldsymbol{X}_{t+1}) = \kappa^{-1}\lambda_{\max}(f(\boldsymbol{X}_t)) \leq \frac{\kappa^*}{\kappa}\lambda_{\max}(\boldsymbol{X}_t).$$

If $\kappa > \kappa^*$ then the SE iterates converge to $\boldsymbol{0}$ for *every* initial point. It is currently unclear to us whether this condition is also necessary. Instead, we consider three special cases and exactly locate the phase transitions thresholds.

The binary case

In this section we treat the case $d = 2$, which is akin to a noiseless version of the CDMA problem (Guo and Verdú, 2005) or the problem of compressed sensing with binary prior. In this case, the SE iteration becomes one-dimensional. Indeed, we have $\mathcal{A} = \{x\boldsymbol{u}\boldsymbol{u}^\intercal, x \geq 0\}$, with $\boldsymbol{u} = (1, -1)^\intercal$. And since this space is invariant under f, the latter can be parameterized by one scalar function $x \mapsto \varphi(x)$, defined by

$$f(x\boldsymbol{u}\boldsymbol{u}^\intercal) = \varphi(x)\boldsymbol{u}\boldsymbol{u}^\intercal, \quad \forall x \geq 0.$$

[1] One can see this by observing that $\{\boldsymbol{z}^\intercal\boldsymbol{X}_t\boldsymbol{z}\}_{t\geq 0}$ is a non-negative monotonically decreasing sequence for all $\boldsymbol{z} \in \mathbb{R}^d$; hence it must have a (non-negative) limit. Then, via the identity $\boldsymbol{y}^\intercal\boldsymbol{X}_t\boldsymbol{z} = \frac{1}{2}((\boldsymbol{y}+\boldsymbol{z})^\intercal\boldsymbol{X}_t(\boldsymbol{y}+\boldsymbol{z}) - (\boldsymbol{y}-\boldsymbol{z})^\intercal\boldsymbol{X}_t(\boldsymbol{y}-\boldsymbol{z}))$, one deduces that $\{\boldsymbol{y}^\intercal\boldsymbol{X}_t\boldsymbol{z}\}_{t\geq 0}$ has a limit for all $\boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^d$. These limits define a bi-linear operator which is $(\boldsymbol{y}, \boldsymbol{z}) \mapsto \boldsymbol{y}^\intercal\boldsymbol{X}^*\boldsymbol{z}$.

Next, we compute φ. For $\boldsymbol{X} = x\boldsymbol{u}\boldsymbol{u}^\intercal$, we have $\boldsymbol{X}^{-\frac{1}{2}}\boldsymbol{u} = \frac{1}{\sqrt{2x}}\boldsymbol{u}$. Then, letting $\boldsymbol{\pi} = (p, 1-p)^\intercal$, using (6.6) we have

$$\begin{aligned}\varphi(x) = f(x\boldsymbol{u}\boldsymbol{u}^\intercal)_{1,1} &= p - p\,\mathbb{E}_{\boldsymbol{g}}\left[\frac{p}{p+(1-p)e^{-\boldsymbol{g}^\intercal\boldsymbol{u}/\sqrt{2x}-1/2x}}\right],\\ &= \mathbb{E}_{\boldsymbol{g}}\left[\frac{p(1-p)}{1-p+pe^{\boldsymbol{g}^\intercal\boldsymbol{u}/\sqrt{2x}+1/2x}}\right],\\ &= \mathbb{E}_g\left[\frac{p(1-p)}{1-p+pe^{g/\sqrt{x}+1/2x}}\right]. \qquad (6.12)\end{aligned}$$

Letting $\boldsymbol{X}_t = a_t\boldsymbol{u}\boldsymbol{u}^\intercal$, for all $t \geq 0$, the SE reduces to

$$a_{t+1} = \kappa^{-1}\varphi(a_t). \qquad (6.13)$$

The function φ is continuous, increasing on $\mathbb{R}_+$, and bounded (since $\varphi(\infty) = p(1-p) < \infty$). Moreover, $\varphi(0) = 0$. Therefore, the sequence (6.13) converges to zero for all initial conditions $a_0 > 0$ if and only if $\kappa^{-1}\varphi(x) < x$ for all $x > 0$, i.e.

$$\kappa > \kappa^*_{\text{binary}}(p) := \sup_{x>0}\ \mathbb{E}_g\left[\frac{p(1-p)x^2}{1-p+p\exp\left(gx+x^2/2\right)}\right].$$

By a change of variables $g + x/2 \to g$, one can also write this threshold as

$$\kappa^*_{\text{binary}}(p) = \sup_{x>0}\ \mathbb{E}_g\left[\frac{p(1-p)x^2e^{-x^2/8}}{pe^{gx/2}+(1-p)e^{-gx/2}}\right]. \qquad (6.14)$$

If $\kappa < \kappa^*_{\text{binary}}(p)$, then a new stable fixed point $a^* > 0$ appears and the sequence $\{a_t\}_{t\geq 0}$ converges to it for all initial conditions $a_0 \geq a^*$, and the asymptotic MSE of the AMP algorithm is $\lim_{t\to\infty}\mathsf{MSE}_t = a^*\,\mathrm{trace}(\boldsymbol{u}\boldsymbol{u}^\intercal) = 2a^*$.

Figure 6.1 demonstrates the accuracy of the above theoretical predictions — the predicted MSE by State Evolution matches the empirical MSE of AMP on a random instance with $n = 2000$, across the whole range of p and κ.

The symmetric case

In this section we treat the symmetric case where all types have equal proportions: $\boldsymbol{\pi} = (\frac{1}{d}, \cdots, \frac{1}{d})$, and analyze the SE dynamics. In this situation, the half-line $\{x(\boldsymbol{D} - \boldsymbol{\pi}\boldsymbol{\pi}^\intercal)\ ,\ x \geq 0\}$ is stable under the application of the map f, and the dynamics becomes one-dimensional if initialized on this half-line.

Lemma 6.5. *Assume $\boldsymbol{\pi} = (\frac{1}{d}, \cdots, \frac{1}{d})$. For all $x > 0$, we have*

$$f\left(x(\boldsymbol{I} - \frac{1}{d}\mathbf{1}\mathbf{1}^\intercal)\right) = \varphi(x)\left(\boldsymbol{I} - \frac{1}{d}\mathbf{1}\mathbf{1}^\intercal\right),$$

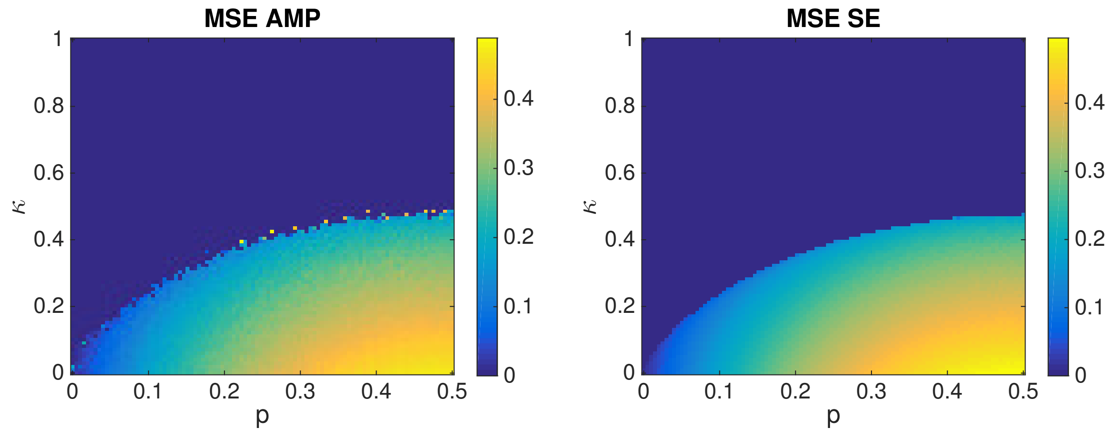

Figure 6.1: MSE of AMP on a random instance with $n = 2000$ in the binary case (left), and predicted MSE by State Evolution (right) as a function of $p = \pi_1$ and κ. The blue region corresponds to exact recovery. The boundary of this region is traced by the curve $p \mapsto \kappa^*_{\text{binary}}(p)$ in equation (6.14).

with

$$\varphi(x) = \mathbb{E}_{\boldsymbol{g}}\left[\frac{\exp(g_2/\sqrt{x})}{\exp(g_1/\sqrt{x} + 1/x) + \sum_{r=2}^{d} \exp(g_r/\sqrt{x})}\right].$$

Proof. Let $\boldsymbol{P} = (\boldsymbol{I} - \frac{1}{d}\mathbf{1}\mathbf{1}^\intercal)$, and $\boldsymbol{X} = x\boldsymbol{P}$ with $x > 0$. The matrix $\boldsymbol{P}$ is the orthogonal projector on $\text{span}(\mathbf{1})^\perp$. Therefore, we have

$$\boldsymbol{X}^{-1/2}(\boldsymbol{e}_r - \boldsymbol{e}_s) = (\boldsymbol{e}_r - \boldsymbol{e}_s)/\sqrt{x}.$$

Therefore for all $r \neq s$,

$$f(\boldsymbol{X})_{rs} = -\frac{1}{d}\ \mathbb{E}_{\boldsymbol{g}}\left[\frac{\exp\left(-\boldsymbol{g}^\intercal(\boldsymbol{e}_r - \boldsymbol{e}_s)/\sqrt{x} - 1/x\right)}{1 + \sum_{l\neq r} \exp\left(-\boldsymbol{g}^\intercal(\boldsymbol{e}_r - \boldsymbol{e}_l)/\sqrt{x} - 1/x\right)}\right].$$

By permutation-invariance of the Gaussian distribution, we see that $f(\boldsymbol{X})$ is constant on its off-diagonal entries, hence on its diagonal entries as well since $f(\boldsymbol{X})\mathbf{1} = \mathbf{0}$. Writing $f(\boldsymbol{X}) = \frac{\alpha}{d}\boldsymbol{I} - \frac{\beta}{d}(\mathbf{1}\mathbf{1}^\intercal - \boldsymbol{I})$, we have $(\alpha + \beta) = d\beta$. Hence, $f(\boldsymbol{X}) = \beta(\boldsymbol{I} - \frac{1}{d}\mathbf{1}\mathbf{1}^\intercal)$ with

$$\begin{aligned}\beta &= \mathbb{E}_{\boldsymbol{g}}\left[\frac{\exp\left(-\boldsymbol{g}^\intercal(\boldsymbol{e}_1 - \boldsymbol{e}_2)/\sqrt{x} - 1/x\right)}{1 + \sum_{l\neq r} \exp\left(-\boldsymbol{g}^\intercal(\boldsymbol{e}_1 - \boldsymbol{e}_l)/\sqrt{x} - 1/x\right)}\right], \\ &= \mathbb{E}_{\boldsymbol{g}}\left[\frac{\exp(g_2/\sqrt{x})}{\exp(g_1/\sqrt{x} + 1/x) + \sum_{r=2}^{d} \exp(g_r/\sqrt{x})}\right],\end{aligned}$$

as claimed. ■

Therefore, if the SE iteration is initialized on this half-line: $\boldsymbol{X}_0 = a_0(\boldsymbol{I} - \frac{1}{d}\mathbf{1}\mathbf{1}^\intercal)$, with $a_0 > 0$, then $\boldsymbol{X}_t = a_t(\boldsymbol{I} - \frac{1}{d}\mathbf{1}\mathbf{1}^\intercal)$ for all t, with

$$a_{t+1} = \kappa^{-1}\varphi(a_t).$$

Just as in the binary case, the function φ is continuous, increasing and bounded with $\varphi(0) = 0$. Hence, we have convergence to zero for all initial condition $a_0 > 0$ if and only if $\kappa^{-1}\varphi(x) < x$ for all $x > 0$, i.e.

$$\kappa > \kappa^*_{\mathrm{sym}}(d) := \sup_{x>0} \ \mathbb{E}_{\boldsymbol{g}}\left[\frac{x^2\exp(g_2 x)}{\exp(g_1 x + x^2) + \sum_{r=2}^d \exp(g_r x)}\right]. \tag{6.15}$$

Otherwise, it converges to a non-zero value a^* for all initial conditions $a_0 > a^*$, and the asymptotic MSE of the AMP algorithm is $a^*\,\mathrm{trace}(\boldsymbol{I} - \frac{1}{d}\mathbf{1}\mathbf{1}^\intercal) = (d-1)a^*$. Using the change of variables $g_1 + x \to g_1$, one can also write this threshold as

$$\kappa^*_{\mathrm{sym}}(d) = \sup_{x>0} \ \mathbb{E}_{\boldsymbol{g}}\left[\frac{x^2 e^{-x^2/2}\exp((g_1+g_2)x)}{\sum_{r=1}^d \exp(g_r x)}\right].$$

It is not straightforward to read off the magnitude of $\kappa^*_{\mathrm{sym}}(d)$ from the above expression. We provide a table of approximate values for several small values of d:

d	2	3	4	5	6	7	8	9	10
κ^*_{sym}	.47	.39	.34	.30	.27	.24	.22	.21	.20

For larger d, an asymptotic expression for this threshold may be desirable. We prove the following in Section 6.4:

Proposition 6.6. *There exist two constants $0 < c_l < c_u$ such that when d is large enough,*

$$c_l \frac{\log d}{d} \le \kappa^*_{\mathrm{sym}}(d) \le c_u \frac{\log d}{d},$$

Furthermore, one can take $c_l = 1 - o_d(1)$, and $c_u = 2 + o_d(1)$.

The general case initialized with a matching

Here we consider the SE iteration in arbitrary dimension and with arbitrary proportions of types $\boldsymbol{\pi}$, but we initialize the dynamics from a special point $\boldsymbol{X}_0$ that corresponds to a matching of the vertices $\{1, \cdots, d\}$: each edge present in the matching corresponds to its own connected component. This case reveals an interesting behavior which we suspect is generic regardless of the initialization: the existence of a sequence of thresholds $\kappa^*_1, \kappa^*_2, \cdots$ ruling the behavior of the SE dynamics. Let $\mathcal{M} = \{(i_1, i_2), (i_3, i_4), \cdots, (i_{K-1}, i_K)\}$ be a matching on the set of vertices $\{1, \cdots, d\}$ (not all vertices are necessarily part of the matching), and let

$\boldsymbol{X}_0$ be its Laplacian matrix, where edges are weighted by arbitrary positive numbers. By Proposition 6.3, f "factors" across connected components, thus each edge in the matching will follow its own dynamics independently of the other edges. The edges not initially present in the matching remain inactive forever. For $(r,s) \in \mathcal{M}$, we have $(X_t)_{rr} = (X_t)_{ss} = -(X_t)_{rs} = -(X_t)_{sr}$, and

$$\boldsymbol{X}_t^{-\frac{1}{2}}(\boldsymbol{e}_r - \boldsymbol{e}_s) = \frac{1}{\sqrt{2(X_t)_{rr}}}(\boldsymbol{e}_r - \boldsymbol{e}_s),$$

and therefore, using expression (6.6) and letting $x = (X_t)_{rr}$,

$$\begin{aligned} f(\boldsymbol{X}_t)_{rr} &= \pi_r - \mathbb{E}_{\boldsymbol{g}}\left[(\boldsymbol{\eta}_r(\boldsymbol{X}_t))_r\right], \\ &= \pi_r\,\mathbb{E}_{\boldsymbol{g}}\left[\frac{\pi_s}{\pi_r e^{(g_r-g_s)/\sqrt{2x}+1/2x} + \pi_s}\right], \\ &= \pi_r\,\mathbb{E}_g\left[\frac{\pi_s}{\pi_r e^{g/\sqrt{x}+1/2x} + \pi_s}\right]. \end{aligned}$$

Therefore, the SE iteration reduces to

$$(X_{t+1})_{rr} = \kappa^{-1}\,\mathbb{E}_g\left[\frac{\pi_r\pi_s}{\pi_r e^{g/\sqrt{(X_t)_{rr}}+1/2(X_t)_{rr}} + \pi_s}\right],$$

for all vertices $(r,s) \in \mathcal{M}$. Note that this iteration is essentially the same as the one in the binary case (6.12)-(6.13), where p becomes π_r and $1-p$ becomes π_s. For each $(r,s) \in \mathcal{M}$, the above iteration converges to the fixed point zero for every initial point if and only if

$$\kappa > \kappa^*_{rs} := \sup_{x>0}\ \mathbb{E}_g\left[\frac{\pi_r\pi_s x^2 e^{-x^2/8}}{\pi_r e^{gx/2} + \pi_s e^{-gx/2}}\right]. \tag{6.16}$$

Here, we symmetrized the expression just as in the binary case (6.14). Arranging these thresholds as $\kappa^*_1 > \kappa^*_2 > \cdots$ from largest to smallest we see that the fixed point of the SE iteration gains one non-zero edge at each κ^*_i as κ decreases from some large value to zero. Equivalently, $\boldsymbol{X}^*$ gains a rank one component corresponding to the connected component constituted by that edge. It is an interesting problem to determine the behavior of the SE iteration and locate these thresholds, if they exist, beyond this simple matching case.

6.3 Summary of results

We presented an algorithm for decoding categorical variables of a signal from randomly pooled observations of it, and characterized its performance it terms of a state evolution equation. The analysis of this evolution revealed phase transition phenomena in the parameters of the problem that happen in the linear regime $m/n \to \kappa$. These algorithmic results, combined with information-theoretic ones leave a large region in parameter space ($\gamma\frac{n}{\log n} < m < \kappa n$) where the signal is identifiable but AMP fails at recovering it, hinting at a possible computational hardness in this structured signal recovery problem. This could have interesting applications in privacy-related considerations. Further, we proved the convergence of the SE dynamics to a fixed point. The analysis of the properties of this fixed point as a function of the parameters $\kappa, \boldsymbol{\pi}$ in the general case, together with rigorous proof of the exactness of the state evolution equations for this problem are interesting open problems.

6.4 Technical proofs

Proof of Proposition 6.1

We proceed by induction. Now assume that $\boldsymbol{M}_{t-1} = \boldsymbol{Q}_{t-1}$ and that $\boldsymbol{R}_{t-1} = \kappa\boldsymbol{X}_{t-1}$. We prove that $\boldsymbol{R}_t = \kappa\boldsymbol{X}_t$ and then that $\boldsymbol{M}_t$ is symmetric and $\boldsymbol{M}_t = \boldsymbol{Q}_t$.

The first step is to show that $\boldsymbol{Q}_t\mathbf{1} = \boldsymbol{\pi}$. By assumption, $\boldsymbol{X}_{t-1} = \kappa^{-1}(\boldsymbol{D} - \boldsymbol{Q}_{t-1}) = \kappa^{-1}\boldsymbol{R}_{t-1}$. Let us define,

$$\eta_{rs} := \boldsymbol{\eta}_r(\boldsymbol{X})_s = \frac{\pi_s \exp\left(-\boldsymbol{g}^\intercal\boldsymbol{X}^{-1/2}(\boldsymbol{e}_r - \boldsymbol{e}_s) - \frac{1}{2}\left\|\boldsymbol{X}^{-1/2}(\boldsymbol{e}_r - \boldsymbol{e}_s)\right\|_{\ell_2}^2\right)}{Z_r(\boldsymbol{X})}. \tag{6.17}$$

The sth coordinate of $\boldsymbol{Q}_t\mathbf{1}$ is

$$(\boldsymbol{Q}_t\mathbf{1})_s = \sum_{r=1}^{d} \pi_r\,\mathbb{E}_{\boldsymbol{g}}\left[\left(\boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}_t^{1/2}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R}_t)\right)_s\right] = \sum_{r=1}^{d} \pi_r\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\right].$$

Moreover, letting $\boldsymbol{\delta}_{rs} = \boldsymbol{X}_{t-1}^{-1/2}(\boldsymbol{e}_r - \boldsymbol{e}_s)$, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\right] &= \int \frac{\pi_s \exp(-\frac{1}{2}\left\|\boldsymbol{g} + \boldsymbol{\delta}_{rs}\right\|_{\ell_2}^2)}{\sum_{l=1}^d \pi_l \exp(-\frac{1}{2}\left\|\boldsymbol{g} + \boldsymbol{\delta}_{rl}\right\|_{\ell_2}^2)} \frac{e^{-\frac{1}{2}\|\boldsymbol{g}\|_{\ell_2}^2}}{(2\pi)^{d/2}}\mathrm{d}\boldsymbol{g}, \\
&\overset{(i)}{=} \int \frac{\exp(-\frac{1}{2}\left\|\boldsymbol{g}\right\|_{\ell_2}^2)}{\sum_{l=1}^d \pi_l \exp(-\frac{1}{2}\left\|\boldsymbol{g} + \boldsymbol{\delta}_{rl}\right\|_{\ell_2}^2)} \frac{\pi_s e^{-\frac{1}{2}\|\boldsymbol{g}-\boldsymbol{\delta}_{rs}\|_{\ell_2}^2}}{(2\pi)^{d/2}}\mathrm{d}\boldsymbol{g}, \\
&= \int \frac{\pi_s \exp(-\frac{1}{2}\left\|\boldsymbol{g} + \boldsymbol{\delta}_{sr}\right\|_{\ell_2}^2)}{\sum_{l=1}^d \pi_l \exp(-\frac{1}{2}\left\|\boldsymbol{g} + \boldsymbol{\delta}_{sl}\right\|_{\ell_2}^2)} \frac{e^{-\frac{1}{2}\|\boldsymbol{g}\|_{\ell_2}^2}}{(2\pi)^{d/2}}\mathrm{d}\boldsymbol{g}.
\end{aligned}$$

The only non-trivial equality is (i) and it was obtained through a simple change of variable $\boldsymbol{g} + \boldsymbol{\delta}_{rs} \to \boldsymbol{g}$. Therefore,

$$(\boldsymbol{Q}_t\mathbf{1})_s = \pi_s \sum_{r=1}^{d} \mathbb{E}_{\boldsymbol{g}}\left[\frac{\pi_r \exp(-\frac{1}{2}\left\|\boldsymbol{g} + \boldsymbol{\delta}_{sr}\right\|_{\ell_2}^2)}{\sum_{l=1}^d \pi_l \exp(-\frac{1}{2}\left\|\boldsymbol{g} + \boldsymbol{\delta}_{sl}\right\|_{\ell_2}^2)}\right] = \pi_s.$$

In addition, the above argument also shows that $\boldsymbol{M}_t$ is symmetric since, for $r, s \in \{1, \cdots, d\}$,

$$(\boldsymbol{M}_t)_{rs} = \pi_s \, \mathbb{E}_{\boldsymbol{g}}[\eta_{sr}].$$

Now we have that $\boldsymbol{R}_t = \boldsymbol{D} - \boldsymbol{Q}_t$, and by symmetry of $\boldsymbol{M}_t$, $\boldsymbol{X}_t = \kappa^{-1}(\boldsymbol{D} - 2\boldsymbol{M}_t + \boldsymbol{Q}_t)$. To complete the proof, it remains to show that $\boldsymbol{M}_t = \boldsymbol{Q}_t$. For $r, s \in \{1, \cdots, d\}$ we have

$$(\boldsymbol{Q}_t)_{rs} = \sum_{l=1}^{d} \pi_l \;\; \mathbb{E}_{\boldsymbol{g}}\left[\eta_{lr}\eta_{ls}\right].$$

Once again, we make the change of variable $\boldsymbol{g} + \boldsymbol{\delta}_{lr} \to \boldsymbol{g}$:

$$\begin{aligned}
(\boldsymbol{Q}_t)_{rs} &= \pi_r \pi_s \sum_{l=1}^{d} \pi_l \;\; \int \frac{\exp(-\frac{1}{2}\left\|\boldsymbol{g}\right\|_{\ell_2}^2) \;\; \exp(-\frac{1}{2}\left\|\boldsymbol{g} + \boldsymbol{\delta}_{rs}\right\|_{\ell_2}^2)}{\left(\sum_{l'=1}^{d} \pi_{l'} \exp(-\frac{1}{2}\left\|\boldsymbol{g} + \boldsymbol{\delta}_{rl'}\right\|_{\ell_2}^2)\right)^2} \frac{e^{-\frac{1}{2}\left\|\boldsymbol{g} - \boldsymbol{\delta}_{lr}\right\|_{\ell_2}^2}}{(2\pi)^{d/2}} \mathrm{d}\boldsymbol{g}, \\
&= \pi_r \pi_s \, \mathbb{E}_{\boldsymbol{g}}\left[\frac{\exp(-\frac{1}{2}\left\|\boldsymbol{g} + \boldsymbol{\delta}_{rs}\right\|_{\ell_2}^2)}{\sum_{r'=1}^{d} \pi_{r'} \exp(-\frac{1}{2}\left\|\boldsymbol{g} + \boldsymbol{\delta}_{rl'}\right\|_{\ell_2}^2)}\right], \\
&= (\boldsymbol{M}_t)_{sr}\,.
\end{aligned}$$

Proof of Proposition 6.2

The map f is differentiable at every $\boldsymbol{X} \succeq \boldsymbol{0}$ invertible on $\mathrm{span}(\boldsymbol{1})^{\perp}$. Let $\boldsymbol{0} \preceq \boldsymbol{X} \preceq \boldsymbol{Y}$, and $\boldsymbol{W} : [0, 1] \to \mathbb{S}_{+}^{d\times d}$ defined by $\boldsymbol{W}(t) = (1 - t)\boldsymbol{X} + t\boldsymbol{Y}$. We will show that $\frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{W}(t)) \succeq \boldsymbol{0}$ for all $t \in [0, 1]$ and conclude with the fundamental theorem of calculus

$$f(\boldsymbol{Y}) - f(\boldsymbol{X}) = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{W}(t))\mathrm{d}t.$$

We start by computing the derivative of each entry of $f(\boldsymbol{W}(t))$. Let $r, s \in \{1, \ldots, d\}$. We have

$$\frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{W}(t))_{rs} = -\frac{\mathrm{d}}{\mathrm{d}t} \pi_r \, \mathbb{E}\left[(\boldsymbol{\eta}_r(\boldsymbol{W}(t)))_s\right].$$

To prepare for further calculations, let us write

$$\boldsymbol{A}(t) := \boldsymbol{W}(t)^{-1/2} \frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{W}(t)^{-1/2}\right),$$

and

$$\boldsymbol{B}(t) := \frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{W}(t)^{-1}\right) = -\boldsymbol{W}(t)^{-1} \cdot \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{W}(t) \cdot \boldsymbol{W}(t)^{-1}.$$

We observe by the chain rule of differentiation that

$$\boldsymbol{A}(t) + \boldsymbol{A}(t)^{\intercal} = \boldsymbol{B}(t). \tag{6.18}$$

This identity will be used several times. Now we start computing the derivative. Let

$$\begin{aligned} D_{rs} :&= \pi_s \frac{\mathrm{d}}{\mathrm{d}t} \exp\left(-\boldsymbol{g}^\intercal \boldsymbol{W}(t)^{-1/2}(\boldsymbol{e}_r - \boldsymbol{e}_s) - \frac{1}{2} \left\| \boldsymbol{W}(t)^{-1/2}(\boldsymbol{e}_r - \boldsymbol{e}_s) \right\|_{\ell_2}^2 \right) \\ &= \pi_s \left(-\boldsymbol{g}^\intercal \frac{\mathrm{d}}{\mathrm{d}t} \left(\boldsymbol{W}(t)^{-1/2} \right) (\boldsymbol{e}_r - \boldsymbol{e}_s) - \frac{1}{2} (\boldsymbol{e}_r - \boldsymbol{e}_s)^\intercal \boldsymbol{B}(t) (\boldsymbol{e}_r - \boldsymbol{e}_s) \right) \\ &\qquad \times \exp\left(-\boldsymbol{g}^\intercal \boldsymbol{W}(t)^{-1/2}(\boldsymbol{e}_r - \boldsymbol{e}_s) - \frac{1}{2} \left\| \boldsymbol{W}(t)^{-1/2}(\boldsymbol{e}_r - \boldsymbol{e}_s) \right\|_{\ell_2}^2 \right). \end{aligned}$$

Then,

$$\frac{\mathrm{d}}{\mathrm{d}t} \boldsymbol{\eta}_r(\boldsymbol{W}(t))_s = \frac{D_{rs}}{Z_r(\boldsymbol{W}(t))} - \boldsymbol{\eta}_r(\boldsymbol{W}(t))_s \times \sum_{l=1}^{d} \frac{D_{rl}}{Z_r(\boldsymbol{W}(t))}. \tag{6.19}$$

Now, by differentiating under the expectation sign, we are lead to process expressions of the form

$$\mathbb{E}_{\boldsymbol{g}} \left[\frac{D_{rs}}{Z_r(\boldsymbol{W}(t))} \right] \quad \text{and} \quad \mathbb{E}_{\boldsymbol{g}} \left[\boldsymbol{\eta}_r(\boldsymbol{W}(t))_s \frac{D_{rl}}{Z_r(\boldsymbol{W}(t))} \right].$$

Here, the Gaussian integration by parts formula

$$\mathbb{E}_g \left[g h(g) \right] = \mathbb{E}_g \left[h'(g) \right]$$

for all univariate differentiable functions h with moderate growth (say polynomial) at infinity, will be used multiple times. Recalling

$$\eta_{rs} = \boldsymbol{\eta}_r(\boldsymbol{W}(t))_s = \frac{\pi_s \exp\left(-\boldsymbol{g}^\intercal \boldsymbol{W}(t)^{-1/2}(\boldsymbol{e}_r - \boldsymbol{e}_s) - \frac{1}{2} \left\| \boldsymbol{W}(t)^{-1/2}(\boldsymbol{e}_r - \boldsymbol{e}_s) \right\|_{\ell_2}^2 \right)}{Z_r(\boldsymbol{W}(t))},$$

from (6.17), we have

$$\begin{aligned} \mathbb{E}_{\boldsymbol{g}} \left[\boldsymbol{g}^\intercal \frac{\mathrm{d}}{\mathrm{d}t} \left(\boldsymbol{W}(t)^{-1/2} \right) (\boldsymbol{e}_r - \boldsymbol{e}_s) \; \eta_{rs} \right] &= \mathbb{E}_{\boldsymbol{g}} \left[(\nabla_{\boldsymbol{g}} \eta_{rs})^\intercal \frac{\mathrm{d}}{\mathrm{d}t} \left(\boldsymbol{W}(t)^{-1/2} \right) (\boldsymbol{e}_r - \boldsymbol{e}_s) \right] \\ &= -(\boldsymbol{e}_r - \boldsymbol{e}_s)^\intercal \boldsymbol{A}(t) (\boldsymbol{e}_r - \boldsymbol{e}_s) \; \mathbb{E}_{\boldsymbol{g}} \left[\eta_{rs} \right] \\ &\quad + \sum_{l=1}^{d} (\boldsymbol{e}_r - \boldsymbol{e}_l)^\intercal \boldsymbol{A}(t) (\boldsymbol{e}_r - \boldsymbol{e}_s) \; \mathbb{E}_{\boldsymbol{g}} \left[\eta_{rs} \eta_{rl} \right], \end{aligned}$$

and similarly,

$$\begin{aligned} \mathbb{E}_{\boldsymbol{g}} \left[\boldsymbol{g}^\intercal \frac{\mathrm{d}}{\mathrm{d}t} \left(\boldsymbol{W}(t)^{-1/2} \right) (\boldsymbol{e}_r - \boldsymbol{e}_l) \; \eta_{rs} \eta_{rl} \right] &= -\big((\boldsymbol{e}_r - \boldsymbol{e}_l)^\intercal \boldsymbol{A}(t) (\boldsymbol{e}_r - \boldsymbol{e}_l) \\ &\quad + (\boldsymbol{e}_r - \boldsymbol{e}_s)^\intercal \boldsymbol{A}(t) (\boldsymbol{e}_r - \boldsymbol{e}_l) \big) \, \mathbb{E}_{\boldsymbol{g}} \left[\eta_{rs} \eta_{rl} \right] \end{aligned}$$

$$+ 2\sum_{r'=1}^{d}(\boldsymbol{e}_r - \boldsymbol{e}_{r'})^\intercal \boldsymbol{A}(t)(\boldsymbol{e}_r - \boldsymbol{e}_l)\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\eta_{rr'}\right].$$

Therefore,

$$\mathbb{E}_{\boldsymbol{g}}\left[\frac{D_{rs}}{Z_r(\boldsymbol{W}(t))}\right] = (\boldsymbol{e}_r - \boldsymbol{e}_s)^\intercal \boldsymbol{A}(t)(\boldsymbol{e}_r - \boldsymbol{e}_s)\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\right] - \frac{1}{2}(\boldsymbol{e}_r - \boldsymbol{e}_s)^\intercal \boldsymbol{B}(t)(\boldsymbol{e}_r - \boldsymbol{e}_s)\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\right]$$
$$- \sum_{l=1}^{d}(\boldsymbol{e}_r - \boldsymbol{e}_l)^\intercal \boldsymbol{A}(t)(\boldsymbol{e}_r - \boldsymbol{e}_s)\ \mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\right].$$

Since $\boldsymbol{A}(t) + \boldsymbol{A}(t)^\intercal = \boldsymbol{B}(t)$ (identity (6.18)), the first two terms in the above expression cancel each other, and we are left with

$$\mathbb{E}_{\boldsymbol{g}}\left[\frac{D_{rs}}{Z_r(\boldsymbol{W}(t))}\right] = -\sum_{l=1}^{d}(\boldsymbol{e}_r - \boldsymbol{e}_l)^\intercal \boldsymbol{A}(t)(\boldsymbol{e}_r - \boldsymbol{e}_s)\ \mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\right].$$

On the other hand, using the identity (6.18) again,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{g}}\left[\boldsymbol{\eta}_r(\boldsymbol{W}(t))_s\frac{D_{rl}}{Z_r(\boldsymbol{W}(t))}\right] &= \big((\boldsymbol{e}_r - \boldsymbol{e}_l)^\intercal \boldsymbol{A}(t)(\boldsymbol{e}_r - \boldsymbol{e}_l) + (\boldsymbol{e}_r - \boldsymbol{e}_s)^\intercal \boldsymbol{A}(t)(\boldsymbol{e}_r - \boldsymbol{e}_l)\big)\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\right]\\ &\quad - \frac{1}{2}(\boldsymbol{e}_r - \boldsymbol{e}_l)^\intercal \boldsymbol{B}(t)(\boldsymbol{e}_r - \boldsymbol{e}_l)\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\right]\\ &\quad - 2\sum_{r'=1}^{d}(\boldsymbol{e}_r - \boldsymbol{e}_{r'})^\intercal \boldsymbol{A}(t)(\boldsymbol{e}_r - \boldsymbol{e}_l)\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\eta_{rr'}\right]\\ &= (\boldsymbol{e}_r - \boldsymbol{e}_s)^\intercal \boldsymbol{A}(t)(\boldsymbol{e}_r - \boldsymbol{e}_l)\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\right]\\ &\quad - 2\sum_{r'=1}^{d}(\boldsymbol{e}_r - \boldsymbol{e}_{r'})^\intercal \boldsymbol{A}(t)(\boldsymbol{e}_r - \boldsymbol{e}_l)\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\eta_{rr'}\right].\end{aligned}$$

Now, using the above two formulas, and recalling (6.19), we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{E}\left[\boldsymbol{\eta}_r(\boldsymbol{W}(t))_s\right] &= -\sum_{l=1}^{d}(\boldsymbol{e}_r - \boldsymbol{e}_l)^\intercal \boldsymbol{A}(t)(\boldsymbol{e}_r - \boldsymbol{e}_s)\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\right]\\ &\quad - \sum_{l=1}^{d}(\boldsymbol{e}_r - \boldsymbol{e}_s)^\intercal \boldsymbol{A}(t)(\boldsymbol{e}_r - \boldsymbol{e}_l)\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\right]\\ &\quad + 2\sum_{l=1}^{d}\sum_{r'=1}^{d}(\boldsymbol{e}_r - \boldsymbol{e}_{r'})^\intercal \boldsymbol{A}(t)(\boldsymbol{e}_r - \boldsymbol{e}_l)\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\eta_{rr'}\right].\end{aligned}$$

Using identity (6.18), the sum of the first two terms in the above expression is

$$-\sum_{l=1}^{d}(\boldsymbol{e}_r - \boldsymbol{e}_s)^\intercal \boldsymbol{B}(t)(\boldsymbol{e}_r - \boldsymbol{e}_l)\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\right],$$

$$= - \sum_{l,r'=1}^{d} (\boldsymbol{e}_r - \boldsymbol{e}_s)^\intercal \boldsymbol{B}(t)(\boldsymbol{e}_r - \boldsymbol{e}_l)\, \mathbb{E}_{\boldsymbol{g}} \left[\eta_{rs}\eta_{rl}\eta_{rr'}\right],$$

where we used the fact $\sum_{r'} \eta_{rr'} = 1$ in the last expression. Similarly, the third term is equal to

$$\sum_{l,r'=1}^{d} (\boldsymbol{e}_r - \boldsymbol{e}_{r'})^\intercal \boldsymbol{B}(t)(\boldsymbol{e}_r - \boldsymbol{e}_l)\, \mathbb{E}_{\boldsymbol{g}} \left[\eta_{rs}\eta_{rl}\eta_{rr'}\right].$$

Therefore we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[\boldsymbol{\eta}_r(\boldsymbol{W}(t))_s\right] = \sum_{l,r'=1}^{d} (\boldsymbol{e}_r - \boldsymbol{e}_{r'})^\intercal \boldsymbol{B}(t)(\boldsymbol{e}_s - \boldsymbol{e}_l)\, \mathbb{E}_{\boldsymbol{g}} \left[\eta_{rs}\eta_{rl}\eta_{rr'}\right].$$

The expression we just obtained does not appear to be symmetric in the indices (r, s), but it does become symmetric when multiplied by π_r, thanks to the following identity:

Lemma 6.7. *Recall the definition of η_{rs} from (6.17). For all $r, s, l \in \{1, \cdots, d\}$ we have*

$$\pi_r \, \mathbb{E}_{\boldsymbol{g}} \left[\eta_{rs}\eta_{rl}\right] = \sum_{l'=1}^{d} \pi_{l'} \, \mathbb{E}_{\boldsymbol{g}} \left[\eta_{l'r}\eta_{l's}\eta_{l'l}\right].$$

Using the above, we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{W}(t))_{rs} &= -\pi_r \frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}\left[\boldsymbol{\eta}_r(\boldsymbol{W}(t))_s\right], \\
&= - \sum_{l,l',r'=1}^{d} \pi_{l'} (\boldsymbol{e}_r - \boldsymbol{e}_{r'})^\intercal \boldsymbol{B}(t)(\boldsymbol{e}_s - \boldsymbol{e}_l)\, \mathbb{E}_{\boldsymbol{g}} \left[\eta_{l'r}\eta_{l's}\eta_{l'l}\eta_{l'r'}\right], \\
&= - \sum_{l'=1}^{d} \pi_{l'} \, \mathbb{E}_{\boldsymbol{g}} \left[\eta_{l'r}\eta_{l's} \cdot (\boldsymbol{e}_r - \boldsymbol{\eta}_{l'})^\intercal \boldsymbol{B}(t)(\boldsymbol{e}_s - \boldsymbol{\eta}_{l'})\right].
\end{aligned}$$

This implies that for all $\boldsymbol{z} \in \mathbb{R}^d$

$$\begin{aligned}
\boldsymbol{z}^\intercal \frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{W}(t)) \boldsymbol{z} &= \sum_{r,s=1}^{d} \frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{W}(t))_{rs} z_r z_s, \\
&= - \sum_{l'=1}^{d} \pi_{l'} \, \mathbb{E}_{\boldsymbol{g}} \left[(\boldsymbol{z} \odot \boldsymbol{\eta}_{l'} - (\boldsymbol{z}^\intercal \boldsymbol{\eta}_{l'})\boldsymbol{\eta}_{l'})^\intercal \boldsymbol{B}(t)(\boldsymbol{z} \odot \boldsymbol{\eta}_{l'} - (\boldsymbol{z}^\intercal \boldsymbol{\eta}_{l'})\boldsymbol{\eta}_{l'})\right],
\end{aligned}$$

where $\odot$ denote the entry-wise product of two vectors. Since $\boldsymbol{B}(t) = -\boldsymbol{W}(t)^{-1} \cdot \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{W}(t) \cdot \boldsymbol{W}(t)^{-1}$ and $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{W}(t) = \boldsymbol{Y} - \boldsymbol{X} \succeq \boldsymbol{0}$, we see that

$$\boldsymbol{z}^\intercal \frac{\mathrm{d}}{\mathrm{d}t} f(\boldsymbol{W}(t)) \boldsymbol{z} = \sum_{l'=1}^{d} \pi_{l'} \, \mathbb{E}_{\boldsymbol{g}} \left[\left\| (\boldsymbol{Y} - \boldsymbol{X})^{\frac{1}{2}} \boldsymbol{W}(t)^{-1} \left(\boldsymbol{z} \odot \boldsymbol{\eta}_{l'} - (\boldsymbol{z}^\intercal \boldsymbol{\eta}_{l'})\boldsymbol{\eta}_{l'}\right) \right\|_{\ell_2}^2 \right] \geq 0,$$

hence concluding the general argument. It now remains to prove Lemma 6.7.

Proof of Lemma 6.7. The proof relies on a simple change of variables in the expectation. Using (6.17), and letting $\boldsymbol{\delta}_{rs} = \boldsymbol{W}(t)^{-1/2}(\boldsymbol{e}_r - \boldsymbol{e}_s)$ for all r, s, we have

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{g}}\left[\eta_{l'r}\eta_{l's}\eta_{l'l}\right] &= \pi_r\pi_s\pi_l\,\mathbb{E}_{\boldsymbol{g}}\left[\frac{e^{-\boldsymbol{g}^\intercal(\boldsymbol{\delta}_{l'r}+\boldsymbol{\delta}_{l's}+\boldsymbol{\delta}_{l'l})-\frac{1}{2}\|\boldsymbol{\delta}_{l'r}\|_{\ell_2}^2-\frac{1}{2}\|\boldsymbol{\delta}_{l's}\|_{\ell_2}^2-\frac{1}{2}\|\boldsymbol{\delta}_{l'l}\|_{\ell_2}^2}}{\left(\sum_{r'=1}^d \pi_{r'}e^{-\boldsymbol{g}^\intercal\boldsymbol{\delta}_{l'r'}-\frac{1}{2}\|\boldsymbol{\delta}_{l'r'}\|_{\ell_2}^2}\right)^3}\right]\\
&= \pi_r\pi_s\pi_l\,\mathbb{E}_{\boldsymbol{g}}\left[\frac{e^{-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{l'r}\|_{\ell_2}^2-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{l's}\|_{\ell_2}^2-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{l'l}\|_{\ell_2}^2}}{\left(\sum_{r'=1}^d \pi_{r'}e^{-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{l'r'}\|_{\ell_2}^2}\right)^3}\right]\\
&= \pi_r\pi_s\pi_l\int_{\mathbb{R}^d}\frac{e^{-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{l'r}\|_{\ell_2}^2-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{l's}\|_{\ell_2}^2-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{l'l}\|_{\ell_2}^2}}{\left(\sum_{r'=1}^d \pi_{r'}e^{-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{l'r'}\|_{\ell_2}^2}\right)^3}\,\frac{e^{-\frac{1}{2}\|\boldsymbol{g}\|_{\ell_2}^2}}{(2\pi)^{d/2}}\,\mathrm{d}\boldsymbol{g}.
\end{aligned}$$

We make the change of variables $\boldsymbol{g} + \boldsymbol{\delta}_{l'r} \to \boldsymbol{g}$. The term $\|\boldsymbol{g}+\boldsymbol{\delta}_{l'r}\|_{\ell_2}^2$ becomes $\|\boldsymbol{g}\|_{\ell_2}^2$, $\|\boldsymbol{g}+\boldsymbol{\delta}_{l's}\|_{\ell_2}^2$ becomes $\|\boldsymbol{g}+\boldsymbol{\delta}_{rs}\|_{\ell_2}^2$, $\|\boldsymbol{g}+\boldsymbol{\delta}_{l'l}\|_{\ell_2}^2$ becomes $\|\boldsymbol{g}+\boldsymbol{\delta}_{rl}\|_{\ell_2}^2$, $\|\boldsymbol{g}\|_{\ell_2}^2$ becomes $\|\boldsymbol{g}+\boldsymbol{\delta}_{rl'}\|_{\ell_2}^2$, and $\|\boldsymbol{g}+\boldsymbol{\delta}_{l'r'}\|_{\ell_2}^2$ becomes $\|\boldsymbol{g}+\boldsymbol{\delta}_{rr'}\|_{\ell_2}^2$ in the denominator. The first term will assume the role of the Gaussian density, and we rewrite the above as an expectation under the Gaussian distribution:

$$\pi_r\pi_s\pi_l\,\mathbb{E}_{\boldsymbol{g}}\left[\frac{e^{-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{rs}\|_{\ell_2}^2-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{rl}\|_{\ell_2}^2-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{rl'}\|_{\ell_2}^2}}{\left(\sum_{r'=1}^d \pi_{r'}e^{-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{rr'}\|_{\ell_2}^2}\right)^3}\right].$$

If the above expression is multiplied by $\pi_{l'}$ and summed over all l', the third term in the numerator cancels with one power of the denominator, and the result is

$$\pi_r\pi_s\pi_l\,\mathbb{E}_{\boldsymbol{g}}\left[\frac{e^{-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{rs}\|_{\ell_2}^2-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{rl}\|_{\ell_2}^2}}{\left(\sum_{r'=1}^d \pi_{r'}e^{-\frac{1}{2}\|\boldsymbol{g}+\boldsymbol{\delta}_{rr'}\|_{\ell_2}^2}\right)^2}\right] = \pi_r\,\mathbb{E}_{\boldsymbol{g}}\left[\eta_{rs}\eta_{rl}\right].$$

■

Proof of Proposition 6.6

For $x > 0$, we let

$$\phi_d(x) := \mathbb{E}_{\boldsymbol{g}}\left[\frac{x^2\sum_{r=2}^d e^{g_r\sqrt{\log(d-1)}x}}{e^{g_1\sqrt{\log(d-1)}x}\cdot(d-1)^{x^2}+\sum_{r=2}^d e^{g_r\sqrt{\log(d-1)}x}}\right].$$

By symmetry in the variables g_r, $r \geq 2$, we can see that

$$\phi_d\left(\frac{x}{\sqrt{\log(d-1)}}\right) = \frac{d-1}{\log(d-1)}\,\mathbb{E}_{\boldsymbol{g}}\left[\frac{x^2\exp(g_2x)}{\exp(g_1x+x^2)+\sum_{r=2}^d\exp(g_rx)}\right].$$

Our claim reduces to exhibiting upper and lower bounds on $\sup_{x>0}\phi_d(x)$ which are asymptotically independent of d. We start with the upper bound. Since, the function $x \to \frac{x}{1+x}$ is concave on $\mathbb{R}_+$, we have by Jensen's inequality,

$$\begin{aligned}\phi_d(x) &\leq \mathbb{E}_{g_1}\left[\frac{x^2\sum_{r=2}^{d}\mathbb{E}_{g_r:r\geq 2}\left[e^{g_r\sqrt{\log(d-1)}x}\right]}{e^{g_1\sqrt{\log(d-1)}x}\cdot(d-1)^{x^2}+\sum_{r=2}^{d}\mathbb{E}_{g_r:r\geq 2}\left[e^{g_r\sqrt{\log(d-1)}x}\right]}\right],\\ &= \mathbb{E}_{g_1}\left[\frac{x^2(d-1)^{1+x^2/2}}{e^{g_1\sqrt{\log(d-1)}x}\cdot(d-1)^{x^2}+(d-1)^{1+x^2/2}}\right].\end{aligned}$$

We split the analysis into two cases: $x \leq \sqrt{2}+\epsilon$, $x > \sqrt{2}+\epsilon$ for some $\epsilon > 0$. We see that $\phi_d(x) \leq x^2$ for all $x > 0$. If $x \leq \sqrt{2}+\epsilon$, then $\phi_d(x) \leq (\sqrt{2}+\epsilon)^2$. For the remaining case, let $\alpha = \alpha(\epsilon) > 0$ such that $x^2/2 - \alpha x - 1 > 0$ for all $x > \sqrt{2}+\epsilon$. One can find such an α as the solution to the equation $\alpha + \sqrt{\alpha^2+2} = \sqrt{2}+\epsilon$. Next, we let $\mathcal{E}$ be the event that $g_1 \leq \frac{1-x^2/2+\alpha x}{x}\sqrt{\log(d-1)}$, and write

$$\begin{aligned}\phi_d(x) \leq\ & \mathbb{E}_{g_1}\left[\frac{x^2}{(d-1)^{x^2/2-1}\cdot e^{g_1\sqrt{\log(d-1)}x}+1}\middle|\bar{\mathcal{E}}\right]\Pr(\bar{\mathcal{E}})\\ &+\mathbb{E}_{g_1}\left[\frac{x^2}{(d-1)^{x^2/2-1}\cdot e^{g_1\sqrt{\log(d-1)}x}+1}\middle|\mathcal{E}\right]\Pr(\mathcal{E}),\end{aligned}$$

Under $\bar{\mathcal{E}}$, we have $-x^2/2+1-g_1x\sqrt{\log(d-1)} \leq -\alpha x$, and the first term in the above expression is upper bounded by

$$x^2(d-1)^{-\alpha x}.$$

On the other hand, we upper bound the conditional expectation in the second term by x^2, and use the fact that $\Pr(\mathcal{E}) \leq (d-1)^{-(1-x^2/2+\alpha x)^2/(2x^2)}$. We obtain the upper bound

$$\phi_d(x) \leq x^2\left((d-1)^{-\alpha x}+(d-1)^{-(1-x^2/2+\alpha x)^2/(2x^2)}\right),$$

which decays to 0 as $d \to \infty$ $\underline{\text{uniformly in } x \geq \sqrt{2}+\epsilon}$. This proves that

$$\sup_{x>0}\phi_d(x) \leq (\sqrt{2}+\epsilon)^2$$

for all d sufficiently large.

Now we turn our attention on the lower bound. Since the function $x \to \frac{x}{1+x}$ is increasing, we have

$$\phi_d(x) \geq \mathbb{E}_{\boldsymbol{g}}\left[\frac{x^2e^{\max_{r\geq 2}g_r\sqrt{\log(d-1)}x}}{e^{g_1\sqrt{\log(d-1)}x}\cdot(d-1)^{x^2}+e^{\max_{r\geq 2}g_r\sqrt{\log(d-1)}x}}\right].$$

The maximum of finitely many Gaussians concentrates in a sub-Gaussian way: for all $t \geq 0$,

$$\Pr\left(\max_{r\geq 2} g_r - \mathbb{E}[\max_{r\geq 2} g_r] \leq -t\right) \leq e^{-t^2/2}.$$

We write $\mathbb{E}[\max_{r\geq 2} g_r] = c_d\sqrt{\log(d-1)}$; it is known that $c_d = \sqrt{2}(1-o_d(1))$. Letting $t = \epsilon c_d\sqrt{\log(d-1)}$ for some $\epsilon > 0$, we have

$$\phi_d(x) \geq \mathbb{E}_{g_1}\left[\frac{x^2(d-1)^{(1-\epsilon)c_d x}}{e^{g_1\sqrt{\log(d-1)}x}\cdot(d-1)^{x^2} + (d-1)^{(1-\epsilon)c_d x}}\right]\cdot\left(1-(d-1)^{-\epsilon^2 c_d^2/2}\right).$$

We plug the value $x = (1-\epsilon)c_d$ in the right hand side, and deduce

$$\sup_{x>0}\phi_d(x) \geq \mathbb{E}_{g_1}\left[\frac{(1-\epsilon)^2 c_d^2}{e^{g_1(1-\epsilon)c_d\sqrt{\log(d-1)}}+1}\right]\cdot\left(1-(d-1)^{-\epsilon^2 c_d^2/2}\right).$$

We see that the above converges to the value $(1-\epsilon)^2$ as $d \to \infty$.

6.5 Deriving the approximate message passing equations

We divide the derivation of the AMP equations into two parts. First, we write down the Belief Propagation (BP) equations, and simplify them to the "relaxed" BP equations. Then, we show how to transform the relaxed BP equations into the AMP iteration.

From Belief Propagation (BP) to Relaxed BP

The factor graph G of our model consists of a bipartite graph with the variables $\{\boldsymbol{x}_i,\ 1 \leq i \leq n\}$ on one side of the bipartition and the measurements $\{\boldsymbol{h}_a,\ 1 \leq a \leq m\}$ on the other side. A measurement (or check) node $\boldsymbol{h}_a$ is connected to $k = \alpha n$ variables nodes in expectation chosen uniformly at random (i.e. those such that $A_{ai} = 1$) from all the variable nodes.

We rescale the elements of the sensing matrix $\boldsymbol{A}$ such that A_{ai} has expectation 0 and variance $\frac{\alpha(1-\alpha)}{n}$. This can be done by subtracting the vector $\alpha n\boldsymbol{\pi}$ from each observation $\boldsymbol{h}_a$ and dividing everything by $\sqrt{n}$. Hence, we let

$$\bar{\boldsymbol{h}}_a := (\boldsymbol{h}_a - \alpha n\boldsymbol{\pi})/\sqrt{n},$$

and

$$\bar{\boldsymbol{A}} = (\boldsymbol{A} - \alpha\mathbf{1}_m\mathbf{1}_n^{\intercal})/\sqrt{n}.$$

The linear system $\boldsymbol{h}_a = \sum_{j=1}^n A_{aj}\boldsymbol{x}_j^*$ is equivalent to $\bar{\boldsymbol{h}}_a = \sum_{j=1}^n \bar{A}_{aj}\boldsymbol{x}_j^*$.

We now write the messages of the Belief Propagation algorithm. Let $\vec{E}$ be the set of directed edges of the factor graph with all possible directions, i.e., each edge (i, a) is endowed with both directions $i \to a$ and $a \to i$. Note that $|\vec{E}| = 2km$. The message passing procedure consists of iterating a map $\mathsf{BP} : \left(\Delta^{d-1}\right)^{\vec{E}} \to \left(\Delta^{d-1}\right)^{\vec{E}}$ from some initial guess until (possible) convergence. For convenience, for all $r \in \{1, \cdots, d\}$, any set of messages $\boldsymbol{m} = \{\boldsymbol{m}_{i\to a}\ ,\ \boldsymbol{m}_{a\to i}\ :\ A_{ai} = 1\} \in \left(\Delta^{d-1}\right)^{\vec{E}}$ on G, and any directed edge $a \to i$, we denote the rth coordinate of the d-dimensional message $\boldsymbol{m}_{a\to i}$ by $\boldsymbol{m}_{a\to i}(\boldsymbol{e}_r)$ instead of $(\boldsymbol{m}_{a\to i})_r$, and similarly for the coordinates of $\boldsymbol{m}_{a\to i}$. With this notation in hand, the map BP is defined as follows: We consider a prior distribution on the messages that agrees with the category proportions in the planted solution τ^*, i.e., for every i and r,

$$P(\boldsymbol{x}_i = \boldsymbol{e}_r) = \pi_r$$

This is our “uninformative” prior: under lack of any further information, the algorithm predicts that $\boldsymbol{x}_i = \boldsymbol{e}_r$ with probability π_r for all i and r. Then for all $\boldsymbol{x} \in \{\boldsymbol{e}_1, \cdots, \boldsymbol{e}_d\}$,

$$\mathsf{BP}(\boldsymbol{m})_{i\to a}(\boldsymbol{x}) := \frac{1}{Z_{i\to a}(\boldsymbol{m})} P(\boldsymbol{x}) \prod_{b\in\partial i\setminus a} \boldsymbol{m}_{b\to i}(\boldsymbol{x}), \tag{6.20}$$

$$\mathsf{BP}(\boldsymbol{m})_{a\to i}(\boldsymbol{x}) := \frac{1}{Z_{a\to i}(\boldsymbol{m})} \sum_{\substack{\boldsymbol{x}_j\in\{\boldsymbol{e}_1,\cdots,\boldsymbol{e}_d\}\\ j\in\partial a\setminus i}} \mathbb{1}\left\{\bar{\boldsymbol{h}}_a = \bar{A}_{ai}\boldsymbol{x} + \sum_{j\neq i} \bar{A}_{aj}\boldsymbol{x}_j\right\} \prod_{j\in\partial a\setminus i} \boldsymbol{m}_{j\to a}(\boldsymbol{x}_j), \tag{6.21}$$

with $Z_{i\to a}(\boldsymbol{m})$ and $Z_{a\to i}(\boldsymbol{m})$ are the normalizing factors such that $\sum_{r=1}^d \mathsf{BP}(\boldsymbol{m})_{i\to a}(\boldsymbol{e}_r) = \sum_{r=1}^d \mathsf{BP}(\boldsymbol{m})_{a\to i}(\boldsymbol{e}_r) = 1$. If G was a tree, the map BP would compute the exact posterior distribution of the category assignments $\{\boldsymbol{x}_i\ :\ 1 \le i \le n\}$ given the observations $\{\boldsymbol{h}_a\ :\ 1 \le a \le m\}$. In our case this will only be true when m/n is large enough.

We see that the second equation above has a sum involving d^{k-1} terms, which makes the execution of the BP algorithm intractable. We derive a set of *relaxed* Belief Propagation messages from the above that only require linear-algebraic computations of size polynomial in n and m. Later, we further simplify these equations by leveraging the fact that our factor graph is random and dense, to finally arrive at the Approximate Message Passing iteration.

We now proceed by replacing the indicator in (6.21) by a Gaussian with small variance $\sigma > 0$, which we then linearize by writing it as the Fourier transform of the standard Gaussian measure (this is also known as the Hubbard-Stratonovich transformation):

$$\begin{aligned}\mathsf{BP}_\sigma(\boldsymbol{m})_{a\to i}(\boldsymbol{x}) &:= \frac{1}{Z_{a\to i}(\boldsymbol{m})} \sum_{\substack{\boldsymbol{x}_j\in\{\boldsymbol{e}_1,\cdots,\boldsymbol{e}_d\}\\ j\in\partial a\setminus i}} \exp\left(-\left\|\bar{\boldsymbol{h}}_a - \sum_{j=1}^n \bar{A}_{aj}\boldsymbol{x}_j\right\|_{\ell_2}^2 \Big/ 2\sigma^2\right) \prod_{j\in\partial a\setminus i} \boldsymbol{m}_{j\to a}(\boldsymbol{x}_{\boldsymbol{j}}),\\ &\propto \sum_{\substack{\boldsymbol{x}_j\in\{\boldsymbol{e}_1,\cdots,\boldsymbol{e}_d\}\\ j\in\partial a\setminus i}} \int_{\mathbb{R}^d} \exp\left(\mathrm{i}\sigma^{-1}\boldsymbol{g}^\intercal\left(\bar{\boldsymbol{h}}_a - \sum_{j=1}^n \bar{A}_{aj}\boldsymbol{x}_j\right)\right) \prod_{j\in\partial a\setminus i} \boldsymbol{m}_{j\to a}(\boldsymbol{x}_j)\gamma_d(\mathrm{d}\boldsymbol{g}),\end{aligned}$$

where we let γ_d refer to the standard d-dimensional Gaussian measure.

$$\propto \int_{\mathbb{R}^d} \exp\Big(\mathrm{i}\sigma^{-1}\boldsymbol{g}^\intercal\big(\bar{\boldsymbol{h}}_a - \overline{A}_{ai}\boldsymbol{x}\big)\Big) \times \prod_{j\in\partial a\setminus i}\Big[\sum_{\boldsymbol{x}_j\in\{\boldsymbol{e}_1,\cdots,\boldsymbol{e}_d\}} \exp\big(-\mathrm{i}\sigma^{-1}\overline{A}_{aj}\boldsymbol{g}^\intercal\boldsymbol{x}_j\big)\ \boldsymbol{m}_{j\to a}(\boldsymbol{x}_j)\Big]\gamma_d(\mathrm{d}\boldsymbol{g}).$$

Now, observe that the exponentials in the sum above involve the terms $\overline{A}_{aj}$ which are of order $1/\sqrt{n}$. By expanding the Taylor series of the exponential, one can show

$$\begin{aligned}\sum_{\boldsymbol{x}_j\in\{\boldsymbol{e}_1,\cdots,\boldsymbol{e}_d\}} \exp(-\mathrm{i}\sigma^{-1}\overline{A}_{aj}\boldsymbol{g}^\intercal\boldsymbol{x}_j)\ \boldsymbol{m}_{j\to a}(\boldsymbol{x}_j) &= \sum_{r=1}^{d}\exp(-\mathrm{i}\sigma^{-1}\overline{A}_{aj}\boldsymbol{g}^\intercal\boldsymbol{e}_r)\ \boldsymbol{m}_{j\to a}(\boldsymbol{e}_r)\\ &= \exp\Big(-\mathrm{i}\sigma^{-1}\overline{A}_{aj}\boldsymbol{g}^\intercal\boldsymbol{m}_{j\to a} - \frac{1}{2}\sigma^{-2}\overline{A}_{aj}^2\boldsymbol{g}^\intercal\boldsymbol{B}_{j\to a}\boldsymbol{g}\Big)\\ &\quad + \mathcal{O}(1/n^{3/2}),\end{aligned}$$

where

$$\boldsymbol{B}_{j\to a} = \mathrm{Diag}(\boldsymbol{m}_{j\to a}) - \boldsymbol{m}_{j\to a}\boldsymbol{m}_{j\to a}^\intercal. \tag{6.22}$$

Plugging the above expression into the message, we get

$$\begin{aligned}\mathsf{BP}_\sigma(\boldsymbol{m})_{a\to i}(\boldsymbol{x}) &\approx \frac{1}{Z_{a\to i}(\boldsymbol{m})}\int_{\mathbb{R}^d}\exp\Big(\mathrm{i}\sigma^{-1}\boldsymbol{g}^\intercal\big(\bar{\boldsymbol{h}}_a - \overline{A}_{ai}\boldsymbol{x}\big)\Big)\\ &\qquad\times\prod_{j\in\partial a\setminus i}\exp\Big(-\mathrm{i}\sigma^{-1}\overline{A}_{aj}\boldsymbol{g}^\intercal\boldsymbol{m}_{j\to a} - \frac{1}{2}\sigma^{-2}\overline{A}_{aj}^2\boldsymbol{g}^\intercal\boldsymbol{B}_{j\to a}\boldsymbol{g}\Big)\gamma_d(\mathrm{d}\boldsymbol{g}),\\ &= \frac{1}{Z_{a\to i}(\boldsymbol{m})}\int_{\mathbb{R}^d}\exp\Big(\mathrm{i}\sigma^{-1}\boldsymbol{g}^\intercal\big(\bar{\boldsymbol{h}}_a - \overline{A}_{ai}\boldsymbol{x}\big)\\ &\qquad - \sum_{j\in\partial a\setminus i}\mathrm{i}\sigma^{-1}\overline{A}_{aj}\boldsymbol{g}^\intercal\boldsymbol{m}_{j\to a} - \frac{1}{2}\sum_{j\in\partial a\setminus i}\sigma^{-2}\overline{A}_{aj}^2\boldsymbol{g}^\intercal\boldsymbol{B}_{j\to a}\boldsymbol{g}\Big)\gamma_d(\mathrm{d}\boldsymbol{g}).\end{aligned}$$

We denote the "average" message and variance that appear in the formula above by

$$\boldsymbol{\omega}_{a\to i} := \sum_{j\in\partial a\setminus i}\overline{A}_{aj}\boldsymbol{m}_{j\to a}, \tag{6.23}$$

$$\boldsymbol{V}_{a\to i} := \sum_{j\in\partial a\setminus i}\overline{A}_{aj}^2\boldsymbol{B}_{j\to a}. \tag{6.24}$$

The exponentiated term in the integrand, when combined with the contribution of the Gaussian density, becomes

$$\mathrm{i}\sigma^{-1}\boldsymbol{g}^\intercal\big(\bar{\boldsymbol{h}}_a - \overline{A}_{ai}\boldsymbol{x} - \boldsymbol{\omega}_{a\to i}\big) - \frac{1}{2}\sigma^{-2}\boldsymbol{g}^\intercal\boldsymbol{V}_{a\to i}\boldsymbol{g} - \frac{1}{2}\left\|\boldsymbol{g}\right\|_{\ell_2}^2$$

$$= \mathrm{i}\sigma^{-1}\boldsymbol{g}^\intercal\left(\bar{\boldsymbol{h}}_a - \bar{A}_{ai}\boldsymbol{x} - \boldsymbol{\omega}_{a\to i}\right) - \frac{1}{2}\boldsymbol{g}^\intercal(\sigma^{-2}\boldsymbol{V}_{a\to i} + \boldsymbol{I})\boldsymbol{g}.$$

Now, computing the integral yields

$$\mathsf{BP}_\sigma(\boldsymbol{m})_{a\to i}(\boldsymbol{x}) \propto \exp\left(-\frac{1}{2\sigma^2}\left\|(\sigma^{-2}\boldsymbol{V}_{a\to i} + \boldsymbol{I})^{-\frac{1}{2}}\left(\bar{\boldsymbol{h}}_a - \bar{A}_{ai}\boldsymbol{x} - \boldsymbol{\omega}_{a\to i}\right)\right\|_{\ell_2}^2\right),$$

and letting $\sigma \to 0$ yields

$$\mathsf{BP}(\boldsymbol{m})_{a\to i}(\boldsymbol{x}) \propto \exp\left(-\frac{1}{2}\left\|\boldsymbol{V}_{a\to i}^{-\frac{1}{2}}(\bar{\boldsymbol{h}}_a - \bar{A}_{ai}\boldsymbol{x} - \boldsymbol{\omega}_{a\to i})\right\|_{\ell_2}^2\right).$$

On the other hand, by injecting the above formula into the messages from-variable-to-check node (6.20), the latter can be written as

$$\begin{aligned}
\mathsf{BP}(\boldsymbol{m})_{i\to a}(\boldsymbol{x}) &\propto P(\boldsymbol{x})\exp\left(\sum_{b\in\partial i\setminus a} -\frac{1}{2}\left\|\boldsymbol{V}_{b\to i}^{-1/2}(\bar{\boldsymbol{h}}_b - \bar{A}_{bi}\boldsymbol{x} - \boldsymbol{\omega}_{b\to i})\right\|_{\ell_2}^2\right), \\
&\propto P(\boldsymbol{x})\exp\left(-\frac{1}{2}\boldsymbol{x}^\intercal\left(\sum_{b\in\partial i\setminus a}\bar{A}_{bi}^2\boldsymbol{V}_{b\to i}^{-1}\right)\boldsymbol{x} + \boldsymbol{x}^\intercal\left(\sum_{b\in\partial i\setminus a}\bar{A}_{bi}\boldsymbol{V}_{b\to i}^{-1}(\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_{b\to i})\right)\right), \\
&\propto P(\boldsymbol{x})\exp(-(\boldsymbol{x} - \boldsymbol{z}_{i\to a})^\intercal\boldsymbol{\Sigma}_{i\to a}^{-1}(\boldsymbol{x} - \boldsymbol{z}_{i\to a})/2), \qquad (6.25)
\end{aligned}$$

where we denoted the average message and variance by

$$\boldsymbol{z}_{i\to a} = \boldsymbol{\Sigma}_{i\to a}\sum_{b\in\partial i\setminus a}\bar{A}_{bi}\boldsymbol{V}_{b\to i}^{-1}(\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_{b\to i}), \qquad (6.26)$$

$$\boldsymbol{\Sigma}_{i\to a}^{-1} := \sum_{b\in\partial i\setminus a}\bar{A}_{bi}^2\boldsymbol{V}_{b\to i}^{-1}. \qquad (6.27)$$

The combination of the equations (6.22-6.27) forms the set of *Relaxed Belief Propagation* (RBP) equations:

$$\begin{aligned}
\boldsymbol{m}_{i\to a} &= \boldsymbol{\eta}(\boldsymbol{z}_{i\to a}, \boldsymbol{\Sigma}_{i\to a}), \\
\boldsymbol{B}_{i\to a} &= \mathrm{Diag}(\boldsymbol{m}_{i\to a}) - \boldsymbol{m}_{i\to a}\boldsymbol{m}_{i\to a}^\intercal, \\
\boldsymbol{z}_{i\to a} &= \boldsymbol{\Sigma}_{i\to a}\sum_{b\in\partial i\setminus a}\bar{A}_{bi}\boldsymbol{V}_{b\to i}^{-1}(\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_{b\to i}), \\
\boldsymbol{\Sigma}_{i\to a}^{-1} &= \sum_{b\in\partial i\setminus a}\bar{A}_{bi}^2\boldsymbol{V}_{b\to i}^{-1}, \qquad (6.28) \\
\boldsymbol{\omega}_{a\to i} &= \sum_{j\in\partial a\setminus i}\bar{A}_{aj}\boldsymbol{m}_{j\to a}, \\
\boldsymbol{V}_{a\to i} &= \sum_{j\in\partial a\setminus i}\bar{A}_{aj}^2\boldsymbol{B}_{j\to a},
\end{aligned}$$

with

$$\boldsymbol{\eta}(\boldsymbol{z},\boldsymbol{\Sigma}) := \frac{1}{Z(\boldsymbol{z},\boldsymbol{\Sigma})}\sum_{r=1}^{d}\pi_r \boldsymbol{e}_r \exp\left(-\frac{1}{2}(\boldsymbol{e}_r-\boldsymbol{z})^\intercal\boldsymbol{\Sigma}^{-1}(\boldsymbol{e}_r-\boldsymbol{z})\right), \tag{6.29}$$

where $Z(\boldsymbol{z},\boldsymbol{\Sigma})$ is the normalization constant so that $\mathbf{1}^\intercal\boldsymbol{\eta}(\boldsymbol{z},\boldsymbol{\Sigma}) = 1$. The complexity of the iterative version of these equations is of order at most $\mathcal{O}(d^3nm)$ which is essentially quadratic in n. Next, we further reduce the complexity of the iteration to $\mathcal{O}(d^3(n+m))$ by showing that it suffices to track the average of the incoming messages at each node. This is due to the fact that the factor graph is dense and its edges are independent.

From Relaxed BP to AMP

Let us now derive the equations of the (more efficient) AMP algorithm. We will define a notion of "total messages" $\boldsymbol{m}_i, \boldsymbol{B}_i, \boldsymbol{z}_i, \boldsymbol{\Sigma}_i, \boldsymbol{\omega}_a, \boldsymbol{V}_a$ and relate them to one another. The expressions (6.23), (6.24), (6.26), and (6.27) defining $\boldsymbol{\omega}_{a\to i}$, $\boldsymbol{V}_{a\to i}$, $\boldsymbol{z}_{i\to a}$ and $\boldsymbol{\Sigma}_{i\to a}$ respectively involve sums over all the neighbors of the node sending the message except the node receiving the message. We first define $\boldsymbol{\omega}_a$, $\boldsymbol{V}_a$ and $\boldsymbol{\Sigma}_i$ by adding this last term:

$$\begin{aligned}
\boldsymbol{\omega}_a^t &:= \sum_{j\in\partial a}\bar{A}_{aj}\boldsymbol{m}_{j\to a}^t = \boldsymbol{\omega}_{a\to i}^t + \bar{A}_{ai}\boldsymbol{m}_{i\to a}^t,\\
\boldsymbol{V}_a^t &:= \sum_{j\in\partial a}\bar{A}_{aj}^2\boldsymbol{B}_{j\to a}^t = \boldsymbol{V}_{a\to i}^t + \bar{A}_{ai}^2\boldsymbol{B}_{i\to a}^t,\\
\left(\boldsymbol{\Sigma}_i^t\right)^{-1} &:= \sum_{b\in\partial i}\bar{A}_{bi}^2\left(\boldsymbol{V}_b^t\right)^{-1}.
\end{aligned}$$

where we introduced a time index t to track the iteration count. Now we attempt to find a notion of total message $\boldsymbol{z}_i^t$ for $\boldsymbol{z}_{i\to a}^t$ such that the obtained set of equations becomes self consistent. Once $\boldsymbol{z}_i^t$ is found, then we define $\boldsymbol{m}_i^{t+1}$ and $\boldsymbol{B}_i^{t+1}$ as $\boldsymbol{\eta}(\boldsymbol{z}_i^t,\boldsymbol{\Sigma}_i^t)$ and $\mathrm{Diag}(\boldsymbol{\eta}(\boldsymbol{z}_i^t,\boldsymbol{\Sigma}_i^t)) - \boldsymbol{\eta}(\boldsymbol{z}_i^t,\boldsymbol{\Sigma}_i^t)\boldsymbol{\eta}(\boldsymbol{z}_i^t,\boldsymbol{\Sigma}_i^t)^\intercal$, respectively. Since $\boldsymbol{\Sigma}_{i\to a}^t - \boldsymbol{\Sigma}_i^t = \mathcal{O}(1/n)$ and $\boldsymbol{V}_{a\to i}^t - \boldsymbol{V}_a^t = \mathcal{O}(1/n)$, we have using (6.26)

$$\begin{aligned}
\boldsymbol{z}_{i\to a}^t &= \boldsymbol{\Sigma}_{i\to a}^t\cdot\sum_{b\in\partial i\setminus a}\bar{A}_{bi}\left(\boldsymbol{V}_{b\to i}^t\right)^{-1}(\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_{b\to i}^t),\\
&\simeq \boldsymbol{\Sigma}_i^t\cdot\sum_{b\in\partial i\setminus a}\bar{A}_{bi}\left(\boldsymbol{V}_b^t\right)^{-1}(\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_{b\to i}^t).
\end{aligned}$$

Substituting the expression $\boldsymbol{\omega}_{a\to i}^t = \boldsymbol{\omega}_a^t - \bar{A}_{ai}\boldsymbol{m}_{i\to a}^t$ in the above, we get

$$\begin{aligned}
\boldsymbol{z}_{i\to a}^t &= \boldsymbol{\Sigma}_i^t\cdot\sum_{b\in\partial i\setminus a}\bar{A}_{bi}\left(\boldsymbol{V}_b^t\right)^{-1}(\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_b^t) + \boldsymbol{\Sigma}_i^t\cdot\sum_{b\in\partial i\setminus a}\bar{A}_{bi}^2\left(\boldsymbol{V}_b^t\right)^{-1}\boldsymbol{m}_{i\to b}^t\\
&\simeq \boldsymbol{\Sigma}_i^t\cdot\sum_{b\in\partial i}\bar{A}_{bi}\left(\boldsymbol{V}_b^t\right)^{-1}(\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_b^t) + \boldsymbol{\Sigma}_i^t\cdot\sum_{b\in\partial i}\bar{A}_{bi}^2\left(\boldsymbol{V}_b^t\right)^{-1}\boldsymbol{m}_{i\to b}^t,
\end{aligned}$$

where we also allowed the above sums to run over all neighbors of i since the additional terms are of order $1/\sqrt{n}$ compared to the entire sum which is of order 1. Now we make the assumption that the messages $\boldsymbol{m}_{i\to b}^t$ are approximately equal for all $b \in \partial i$ to a common value $\boldsymbol{m}_i^t$, up to error $1/\sqrt{n}$. This assumption is justified by the fact that the graph is dense with equally strong edge weights, so the messages outgoing from every node are equal, up to first order. This simplifies the second term:

$$\boldsymbol{\Sigma}_i^t \cdot \sum_{b\in\partial i} \bar{A}_{bi}^2 \left(\boldsymbol{V}_b^t\right)^{-1} \boldsymbol{m}_{i\to b}^t \simeq \boldsymbol{\Sigma}_i^t \cdot \sum_{b\in\partial i} \bar{A}_{bi}^2 \left(\boldsymbol{V}_b^t\right)^{-1} \boldsymbol{m}_i^t = \boldsymbol{m}_i^t.$$

Based on these approximations, we define

$$\boldsymbol{z}_i^t := \boldsymbol{\Sigma}_i^t \cdot \sum_{b\in\partial i} \bar{A}_{bi} \left(\boldsymbol{V}_b^t\right)^{-1} (\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_b^t) + \boldsymbol{m}_i^t.$$

Now we treat $\boldsymbol{\omega}_a^t$. Recall $\boldsymbol{\omega}_a^t = \sum_{j\in\partial a} \bar{A}_{aj} \boldsymbol{m}_{j\to a}^t$, and $\boldsymbol{m}_{j\to a}^t = \boldsymbol{\eta}(\boldsymbol{z}_{j\to a}^{t-1}, \boldsymbol{\Sigma}_{j\to a}^{t-1})$. We write

$$\begin{aligned}\boldsymbol{z}_{j\to a}^{t-1} &= \boldsymbol{\Sigma}_{j\to a}^{t-1} \cdot \sum_{b\in\partial j} \bar{A}_{bj} \left(\boldsymbol{V}_{b\to j}^{t-1}\right)^{-1} (\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_{b\to j}^{t-1}) - \boldsymbol{\Sigma}_{j\to a}^{t-1} \cdot \bar{A}_{aj} \left(\boldsymbol{V}_{a\to j}^{t-1}\right)^{-1} (\bar{\boldsymbol{h}}_a - \boldsymbol{\omega}_{a\to j}^{t-1}),\\ &\simeq \boldsymbol{z}_j^{t-1} - \boldsymbol{\Sigma}_{j\to a}^{t-1} \cdot \bar{A}_{aj} \left(\boldsymbol{V}_a^{t-1}\right)^{-1} (\bar{\boldsymbol{h}}_a - \boldsymbol{\omega}_a^{t-1}).\end{aligned}$$

The second term is negligible compared to the first one, so we develop a first order Taylor approximation of the function $\boldsymbol{\eta}$ in the second term, and obtain

$$\begin{aligned}\boldsymbol{\omega}_a^t &= \sum_{j\in\partial a} \bar{A}_{aj} \boldsymbol{\eta}(\boldsymbol{z}_{j\to a}^{t-1}, \boldsymbol{\Sigma}_{j\to a}^{t-1}),\\ &\simeq \sum_{j\in\partial a} \bar{A}_{aj} \left(\boldsymbol{\eta}(\boldsymbol{z}_j^{t-1}, \boldsymbol{\Sigma}_j^{t-1}) - \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\boldsymbol{z}}(\boldsymbol{z}_{j\to a}^{t-1}, \boldsymbol{\Sigma}_{j\to a}^{t-1}) \cdot \boldsymbol{\Sigma}_{j\to a}^{t-1} \cdot \bar{A}_{aj} (\boldsymbol{V}_a^{t-1})^{-1} (\bar{\boldsymbol{h}}_a - \boldsymbol{\omega}_a^{t-1})\right),\\ &= \sum_{j\in\partial a} \bar{A}_{aj} \boldsymbol{m}_j^t - \left(\sum_{j\in\partial a} \bar{A}_{aj}^2 \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\boldsymbol{z}}(\boldsymbol{z}_{j\to a}^{t-1}, \boldsymbol{\Sigma}_{j\to a}^{t-1}) \cdot \boldsymbol{\Sigma}_{j\to a}^{t-1}\right) (\boldsymbol{V}_a^{t-1})^{-1} (\bar{\boldsymbol{h}}_a - \boldsymbol{\omega}_{\boldsymbol{a}}^{t-1}).\end{aligned}$$

Based on the expression (6.29) of $\boldsymbol{\eta}$, one can easily check that

$$\frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\boldsymbol{z}} (\boldsymbol{z}, \boldsymbol{\Sigma}) = (\mathrm{Diag}(\boldsymbol{\eta}(\boldsymbol{z}, \boldsymbol{\Sigma})) - \boldsymbol{\eta}(\boldsymbol{z}, \boldsymbol{\Sigma}) \boldsymbol{\eta}(\boldsymbol{z}, \boldsymbol{\Sigma})^\intercal) \cdot \boldsymbol{\Sigma}^{-1},$$

hence

$$\sum_{j\in\partial a} \bar{A}_{aj}^2 \frac{\mathrm{d}\boldsymbol{\eta}}{\mathrm{d}\boldsymbol{z}} (\boldsymbol{z}_{j\to a}^{t-1}, \boldsymbol{\Sigma}_{j\to a}^{t-1}) \cdot \boldsymbol{\Sigma}_{j\to a}^{t-1} = \sum_{j\in\partial a} \bar{A}_{aj}^2 \boldsymbol{B}_{j\to a}^t = \boldsymbol{V}_a^t.$$

We therefore end up with the following approximate message passing procedure:

$$\begin{aligned}
\boldsymbol{m}_i^{t+1} &= \boldsymbol{\eta}(\boldsymbol{z}_i^t, \boldsymbol{\Sigma}_i^t),\\
\boldsymbol{B}_i^{t+1} &= \mathrm{Diag}(\boldsymbol{\eta}(\boldsymbol{z}_i^t, \boldsymbol{\Sigma}_i^t)) - \boldsymbol{\eta}(\boldsymbol{z}_i^t, \boldsymbol{\Sigma}_i^t)\boldsymbol{\eta}(\boldsymbol{z}_i^t, \boldsymbol{\Sigma}_i^t)^\intercal,\\
\boldsymbol{\Sigma}_i^t &= \left(\sum_{b\in\partial i} \bar{A}_{bi}^2 \left(\boldsymbol{V}_b^t\right)^{-1}\right)^{-1},\\
\boldsymbol{z}_i^t &= \boldsymbol{m}_i^t + \boldsymbol{\Sigma}_i^t \cdot \sum_{b\in\partial i} \bar{A}_{bi} \left(\boldsymbol{V}_b^t\right)^{-1} (\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_b^t),\\
\boldsymbol{\omega}_a^t &= \sum_{j\in\partial a} \bar{A}_{aj}\boldsymbol{m}_j^t - \boldsymbol{V}_a^t \left(\boldsymbol{V}_a^{t-1}\right)^{-1} (\bar{\boldsymbol{h}}_a - \boldsymbol{\omega}_a^{t-1}),\\
\boldsymbol{V}_a^t &= \sum_{j\in\partial a} \bar{A}_{aj}^2 \boldsymbol{B}_j^t.
\end{aligned}$$

This is rearranged to the AMP algorithm displayed in Section 6.1, with the notation $\hat{\boldsymbol{x}}_i^t$ replacing $\boldsymbol{m}_i^t$.

6.6 State evolution equations

We derive the state evolution equations from the Relaxed Belief Propagation (RBP) equations (6.28). Let $\boldsymbol{M}_t = \frac{1}{n}\sum_{i=1}^n \boldsymbol{m}_i^t \boldsymbol{x}_i^{*\intercal}$ and $\boldsymbol{Q}_t = \frac{1}{n}\sum_{i=1}^n \boldsymbol{m}_i^t \boldsymbol{m}_i^{t\intercal}$. As we argued in the previous section, we can redefine $\boldsymbol{M}_t$ and $\boldsymbol{Q}_t$ by substituting $\boldsymbol{m}_i^t$ by $\boldsymbol{m}_{i\to a}^t$ at the cost of an asymptotically vanishing error. In this section, we drop the time indices to lighten the notation. We expect the variance parameters $\boldsymbol{V}_{a\to i}$ in RBP to be concentrated about a constant:

$$\mathbb{E}[\boldsymbol{V}_{a\to i}] \simeq \sum_{j\neq i} \mathbb{E}[\bar{A}_{aj}^2]\boldsymbol{B}_{j\to a} = \frac{1}{n}\alpha(1-\alpha)\sum_{j\neq i} \boldsymbol{B}_{j\to a} = \alpha(1-\alpha)\boldsymbol{R},$$

with $\boldsymbol{R} := \frac{1}{n}\sum_j \boldsymbol{B}_{j\to a}$. A calculation of the second moment of $\boldsymbol{V}_{a\to i}$ reveals that it is equal to the expectation of $\boldsymbol{V}_{a\to i}$ plus a lower order term. Therefore we can safely assume that the quantities $\boldsymbol{V}_{a\to i}$ are essentially constant and equal to $\alpha(1-\alpha)\boldsymbol{R}$. Next, we deal with $\boldsymbol{\Sigma}_{i\to a}$. By assuming approximate independence of $\bar{A}_{bi}$ and $\boldsymbol{V}_{b\to i}$, we get

$$\mathbb{E}\left[\boldsymbol{\Sigma}_{i\to a}^{-1}\right] = \sum_{b\neq a} \mathbb{E}\left[\bar{A}_{bi}^2\right]\mathbb{E}\left[\boldsymbol{V}_{b\to i}^{-1}\right] = \frac{1}{n}\alpha(1-\alpha)\sum_{b\neq a}\frac{\boldsymbol{R}^{-1}}{\alpha(1-\alpha)} \simeq \kappa\boldsymbol{R}^{-1}.$$

We then make the approximation $\boldsymbol{\Sigma}_{i\to a}^{-1} \simeq \mathbb{E}[\boldsymbol{\Sigma}_{i\to a}^{-1}]$, i.e. $\boldsymbol{\Sigma}_{i\to a} \simeq \kappa^{-1}\boldsymbol{R}$. Next, we turn our attention to $\boldsymbol{z}_{i\to a}$:

$$\boldsymbol{z}_{i\to a} = \boldsymbol{\Sigma}_{i\to a} \cdot \sum_{b\neq a} \bar{A}_{bi}\boldsymbol{V}_{b\to i}^{-1}(\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_{b\to i})$$

$$\simeq \frac{1}{\kappa\alpha(1-\alpha)} \sum_{b \neq a} \bar{A}_{bi}(\bar{\boldsymbol{h}}_b - \boldsymbol{\omega}_{b\to i}).$$

Now using $\boldsymbol{\omega}_{b\to i} = \sum_{j\neq i} \bar{A}_{bj}\boldsymbol{m}_{j\to b}$ and $\bar{\boldsymbol{h}}_b = \sum_{j=1}^{n} \bar{A}_{bj}\boldsymbol{x}_j^*$, we get

$$\boldsymbol{z}_{i\to a} \simeq \frac{1}{\kappa\alpha(1-\alpha)} \sum_{b\neq a} \bar{A}_{bi} \left(\sum_{j \neq i} \bar{A}_{bj}(\boldsymbol{x}_j^* - \boldsymbol{m}_{j\to a}) + \bar{A}_{bi}\boldsymbol{x}_i^* \right).$$

The inner sum in the above expression involves n weakly independent terms, so we expect a central limit theorem to hold. Therefore the only relevant quantities are the expectation and the variance of $\boldsymbol{z}$: $\mathbb{E}[\boldsymbol{z}_{i\to a}] = \boldsymbol{x}_i^*$, and

$$\begin{aligned}
\mathbb{E}[(\boldsymbol{z}_{i\to a} - \boldsymbol{x}_i^*)(\boldsymbol{z}_{i\to a} - \boldsymbol{x}_i^*)^\intercal] &= \frac{1}{(\kappa\alpha(1-\alpha))^2} \sum_{b\neq a}\sum_{j\neq i}\sum_{b'\neq a}\sum_{j'\neq i} \mathbb{E}[\bar{A}_{bi}\bar{A}_{b'i}]\,\mathbb{E}[\bar{A}_{bj}\bar{A}_{bj'}] \\
&\qquad\qquad \times (\boldsymbol{x}_j^* - \boldsymbol{m}_{j\to a})(\boldsymbol{x}_j^* - \boldsymbol{m}_{j\to a})^\intercal \\
&= \frac{1}{(\kappa\alpha(1-\alpha))^2} \sum_{b\neq a}\sum_{j\neq i} \frac{(\alpha(1-\alpha))^2}{n^2} (\boldsymbol{x}_j^* - \boldsymbol{m}_{j\to a})(\boldsymbol{x}_j^* - \boldsymbol{m}_{j\to a})^\intercal \\
&= \kappa^{-2}\frac{m}{n}\frac{1}{m}\sum_{b\neq a}\frac{1}{n}\sum_{j\neq i}(\boldsymbol{x}_j^* - \boldsymbol{m}_{j\to a})(\boldsymbol{x}_j^* - \boldsymbol{m}_{j\to a})^\intercal \\
&\simeq \kappa^{-1}(\boldsymbol{D} - \boldsymbol{M} - \boldsymbol{M}^\intercal + \boldsymbol{Q}),
\end{aligned}$$

with $\boldsymbol{D} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i^*\boldsymbol{x}_i^{*\intercal} = \mathrm{Diag}(\boldsymbol{\pi})$. Hence, we define

$$\boldsymbol{X} := \kappa^{-1}(\boldsymbol{D} - \boldsymbol{M} - \boldsymbol{M}^\intercal + \boldsymbol{Q}).$$

Therefore we have made the assumption that $\boldsymbol{z}_{i\to a} \sim \mathcal{N}(\boldsymbol{x}_i^*, \boldsymbol{X})$. Next, we assume that the $\boldsymbol{z}_{i\to a}$ are "independent enough" that a law of large numbers holds in limit $n\to\infty$, $m/n\to\kappa$:

$$\frac{1}{n}\sum_{i:\boldsymbol{x}_i^*=\boldsymbol{e}_r} \boldsymbol{m}_{i\to a} = \frac{1}{n}\sum_{i:\boldsymbol{x}_i^*=\boldsymbol{e}_r} \boldsymbol{\eta}(\boldsymbol{z}_{i\to a}, \boldsymbol{\Sigma}_{i\to a}) \simeq \pi_r\,\mathbb{E}_{\boldsymbol{g}}\left[\boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R})\right],$$

and

$$\frac{1}{n}\sum_{i:\boldsymbol{x}_i^*=\boldsymbol{e}_r} \boldsymbol{m}_{i\to a}\boldsymbol{m}_{i\to a}^\intercal \simeq \pi_r\,\mathbb{E}_{\boldsymbol{g}}\left[\boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R})\cdot\boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R})^\intercal\right],$$

for all $r\in\{1,\cdots,d\}$, with $\boldsymbol{g}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{I})$. Plugging the above into $\boldsymbol{M}$ and $\boldsymbol{Q}$ yields

$$\begin{aligned}
\boldsymbol{M} &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\eta}(\boldsymbol{x}_i^* + \boldsymbol{X}^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R})\boldsymbol{x}_i^{*\intercal}, \\
&\simeq \sum_{r=1}^{d} \pi_r\,\mathbb{E}_{\boldsymbol{g}}\left[\boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R})\right]\boldsymbol{e}_r^\intercal,
\end{aligned}$$

$$\begin{aligned}\boldsymbol{Q} &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\eta}(\boldsymbol{x}_i^* + \boldsymbol{X}^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R}) \cdot \boldsymbol{\eta}(\boldsymbol{x}_i^* + \boldsymbol{X}^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R})^\intercal, \\ &\simeq \sum_{r=1}^{d}\pi_r\,\mathbb{E}_{\boldsymbol{g}}\left[\boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R}) \cdot \boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R})^\intercal\right].\end{aligned}$$

Finally, it remains to find an expression for $\boldsymbol{R}$. Recall $\boldsymbol{B}_{i\to a} = \mathrm{Diag}(\boldsymbol{m}_{i\to a}) - \boldsymbol{m}_{i\to a}\boldsymbol{m}_{i\to a}^\intercal$. Averaging over i and using the assumed concentration of the messages $\boldsymbol{m}_{i\to a}$ yields

$$\begin{aligned}\boldsymbol{R} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{B}_{i\to a} &\simeq \mathrm{Diag}\left(\sum_{r=1}^{d}\pi_r\,\mathbb{E}_{\boldsymbol{g}}\left[\boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R})\right]\right) - \boldsymbol{Q}, \\ &= \mathrm{Diag}(\boldsymbol{Q}\boldsymbol{1}) - \boldsymbol{Q}.\end{aligned}$$

To sum up, we get a system of self-consistent equations in $\boldsymbol{M}_t$, $\boldsymbol{Q}_t$, $\boldsymbol{X}_t$ and $\boldsymbol{R}_t$:

$$\begin{aligned}\boldsymbol{M}_{t+1} &= \sum_{r=1}^{d}\pi_r\,\mathbb{E}_{\boldsymbol{g}}\left[\boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}_t^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R}_t)\right] \cdot \boldsymbol{e}_r^\intercal, \\ \boldsymbol{Q}_{t+1} &= \sum_{r=1}^{d}\pi_r\,\mathbb{E}_{\boldsymbol{g}}\left[\boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}_t^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R}_t) \cdot \boldsymbol{\eta}(\boldsymbol{e}_r + \boldsymbol{X}_t^{\frac{1}{2}}\boldsymbol{g}, \kappa^{-1}\boldsymbol{R}_t)^\intercal\right], \\ \boldsymbol{X}_t &= \kappa^{-1}(\boldsymbol{D} - \boldsymbol{M}_t - \boldsymbol{M}_t^\intercal + \boldsymbol{Q}_t), \\ \boldsymbol{R}_t &= \mathrm{Diag}(\boldsymbol{Q}_t\boldsymbol{1}) - \boldsymbol{Q}_t.\end{aligned}$$

This set of equations constitute the state evolution equations.

Bibliography

Achlioptas, D. and Coja-Oghlan, A. (2008). “Algorithmic barriers from phase transitions”. In: *Foundations of Computer Science, 2008. FOCS’08. IEEE 49th Annual IEEE Symposium on.* IEEE, pp. 793–802.

Achlioptas, D. and Moore, C. (2004). “The chromatic number of random regular graphs”. In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques.* Springer, pp. 219–228.

Achlioptas, D. and Naor, A. (2005). “The Two Possible Values of the Chromatic Number of a Random Graph”. In: *Annals of Mathematics* 162.3, pp. 1335–1351.

Addario-Berry, L. et al. (2010). “On combinatorial testing problems”. In: *The Annals of Statistics* 38.5, pp. 3063–3092.

Aizenman, M., Lebowitz, J. L., and Ruelle, D. (1987). “Some rigorous results on the Sherrington–Kirkpatrick spin glass model”. In: *Communications in Mathematical Physics* 112.1, pp. 3–20.

Aizenman, M., Sims, R., and Starr, S. L. (2003). “Extended variational principle for the Sherrington-Kirkpatrick spin-glass model”. In: *Physical Review B* 68.21, p. 214403.

Aleskandrov, A. (1939). “Almost everywhere existence of the second differential of a convex function and some properties of convex functions”. In: *Leningrad Univ. Ann.* 37, pp. 3–35.

Amini, A. A. and Wainwright, M. J. (2009). “High-dimensional analysis of semidefinite relaxations for sparse principal components”. In: *Annals of Statistics* 37.5B, pp. 2877–2921.

Arias-Castro, E. and Verzelen, N. (2014). “Community detection in dense random networks”. In: *Ann. Statist.* 42.3, pp. 940–969. DOI: 10.1214/14-AOS1208. URL: https://doi.org/10.1214/14-AOS1208.

Auffinger, A. and Chen, W.-K. (2014). “Free Energy and Complexity of Spherical Bipartite Models”. In: *Journal of Statistical Physics* 157.1, pp. 40–59.

Bai, Z. and Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices.* Vol. 20. Springer.

Bai, Z. and Yao, J. f. (2008). “Central limit theorems for eigenvalues in a spiked population model”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 44.3, pp. 447–474.

Bai, Z. and Yao, J. f. (2012). "On sample eigenvalues in a generalized spiked population model". In: *Journal of Multivariate Analysis* 106, pp. 167–177.

Baik, J., Ben Arous, G., and Péché, S. (2005). "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices". In: *Annals of Probability* 33.5, pp. 1643–1697.

Baik, J. and Lee, J. O. (2016). "Fluctuations of the free energy of the spherical Sherrington–Kirkpatrick model". In: *Journal of Statistical Physics* 165.2, pp. 185–224.

— (2017a). "Fluctuations of the Free Energy of the Spherical Sherrington–Kirkpatrick Model with Ferromagnetic Interaction". In: *Annales Henri Poincaré*. Vol. 18. Springer, pp. 1867–1917.

— (2017b). "Free energy of bipartite spherical Sherrington–Kirkpatrick model". In: *arXiv preprint arXiv:1711.06364*.

Baik, J. and Silverstein, J. W. (2006). "Eigenvalues of large sample covariance matrices of spiked population models". In: *Journal of Multivariate Analysis* 97.6, pp. 1382–1408.

Banerjee, D. (2018). "Contiguity and non-reconstruction results for planted partition models: the dense case". In: *Electron. J. Probab.* 23, 28 pp. DOI: 10.1214/17-EJP128. URL: https://doi.org/10.1214/17-EJP128.

Banerjee, D. and Ma, Z. (2018). "Asymptotic normality and analysis of variance of log-likelihood ratios in spiked random matrix models". In: *arXiv preprint arXiv:1804.00567*.

Banks, J., Moore, C., Neeman, J., et al. (2016). "Information-theoretic thresholds for community detection in sparse networks". In: *29th Annual Conference on Learning Theory*. Vol. 49, pp. 383–416.

Banks, J., Moore, C., Vershynin, R., et al. (2017). "Information-theoretic bounds and phase transitions in clustering, sparse PCA, and submatrix localization". In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE, pp. 1137–1141.

Bapst, V. et al. (2016). "The condensation phase transition in random graph coloring". In: *Communications in Mathematical Physics* 341.2, pp. 543–606.

Barbier, J., Dia, M., et al. (2016). "Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula". In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 424–432.

Barbier, J., Krzakala, F., et al. (2017). "Phase transitions, optimal errors and optimality of message-passing in generalized linear models". In: *arXiv preprint arXiv:1708.03395*.

Barbier, J. and Macris, N. (2017). "The stochastic interpolation method: A simple scheme to prove replica formulas in Bayesian inference". In: *arXiv preprint arXiv:1705.02780*.

Barbier, J., Macris, N., and Miolane, L. (2017). "The layered structure of tensor estimation and its mutual information". In: *arXiv preprint arXiv:1709.10368*.

Barra, A., Galluzzi, A., et al. (2014). "Mean field bipartite spin models treated with mechanical techniques". In: *The European Physical Journal B* 87.3, p. 74.

Barra, A., Genovese, G., and Guerra, F. (2011). "Equilibrium statistical mechanics of bipartite spin systems". In: *Journal of Physics A: Mathematical and Theoretical* 44.24, p. 245002.

Bayati, M., Lelarge, M., and Montanari, A. (2012). "Universality in polytope phase transitions and iterative algorithms". In: *IEEE International Symposium on Information Theory Proceedings (ISIT)*. IEEE, pp. 1643–1647.

— (2015). "Universality in polytope phase transitions and message passing algorithms". In: *Annals of Applied Probability* 25.2, pp. 753–822.

Bayati, M. and Montanari, A. (2011). "The dynamics of message passing on dense graphs, with applications to compressed sensing". In: *IEEE Transactions on Information Theory* 57.2, pp. 764–785.

Benaych-Georges, F. and Nadakuditi, R. R. (2011). "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices". In: *Advances in Mathematics* 227.1, pp. 494–521.

— (2012). "The singular values and vectors of low rank perturbations of large rectangular random matrices". In: *Journal of Multivariate Analysis* 111, pp. 120–135.

Berthet, Q. and Rigollet, P. (2013). "Optimal detection of sparse principal components in high dimension". In: *Annals of Statistics* 41.4, pp. 1780–1815.

Biggs, N. (1997). "Algebraic potential theory on graphs". In: *Bulletin of the London Mathematical Society* 29.6, pp. 641–682.

Bolthausen, E. and Bovier, A. (2007). *Spin glasses, volume 1900 of Lecture Notes in Mathematics*. Springer.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge, UK: Cambridge University Press.

Candés, E., Romberg, J., and Tao, T. (2006). "Stable signal recovery from incomplete and inaccurate measurements". In: *Communications on Pure and Applied Mathematics* 59.8, pp. 1207–1223.

Candés, E. and Tao, T. (2005). "Decoding by linear programming". In: *IEEE Transactions on Information Theory* 51.12, pp. 4203–4215.

Capitaine, M., Donati-Martin, C., and Féral, D. (2009). "The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations". In: *Annals of Probability*, pp. 1–47.

Chaiken, S. (1982). "A combinatorial proof of the all minors matrix tree theorem". In: *SIAM Journal on Algebraic Discrete Methods* 3.3, pp. 319–329.

Chatterjee, S. (2014). *Superconcentration and Related Topics*. Springer.

Coja-Oghlan, A. (2009). "Random constraint satisfaction problems". In: *arXiv preprint arXiv:0911.2322*.

Coja-Oghlan, A., Efthymiou, C., and Hetterich, S. (2016). "On the chromatic number of random regular graphs". In: *Journal of Combinatorial Theory, Series B* 116, pp. 367–439.

Coja-Oghlan, A. and Frieze, A. (2014). "Analyzing Walksat on random formulas". In: *SIAM Journal on Computing* 43.4, pp. 1456–1485.

Coja-Oghlan, A., Haqshenas, A., and Hetterich, S. (2016). "Walksat stalls well below the satisfiability threshold". In: *arXiv preprint arXiv:1608.00346*.

Coja-Oghlan, A., Mossel, E., and Vilenchik, D. (2009). "A spectral approach to analysing belief propagation for 3-colouring". In: *Combinatorics, Probability and Computing* 18.6, pp. 881–912.

Coja-Oghlan, A. and Perkins, W. (2016). "Belief Propagation on replica symmetric random factor graph models". In: *arXiv preprint arXiv:1603.08191*.

Comets, F. and Neveu, J. (1995). "The Sherrington-Kirkpatrick model of spin glasses and stochastic calculus: the high temperature case". In: *Communications in Mathematical Physics* 166.3, pp. 549–564.

Dani, V., Moore, C., and Olson, A. (2012). "Tight bounds on the threshold for permuted k-colorability". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Springer, pp. 505–516.

De Bruijn, N. G. (1970). *Asymptotic Methods in Analysis*. Dover Publications.

Deshpande, Y., Abbé, E., and Montanari, A. (2016). "Asymptotic mutual information for the binary stochastic block model". In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE, pp. 185–189.

Deshpande, Y. and Montanari, A. (2014). "Information-theoretically optimal sparse PCA". In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE, pp. 2197–2201.

Ding, J., Sly, A., and Sun, N. (2015). "Proof of the satisfiability conjecture for large k". In: *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*. ACM, pp. 59–68.

— (2016). "Satisfiability threshold for random regular NAE-SAT". In: *Communications in Mathematical Physics* 341.2, pp. 435–489.

Dobriban, E. (2017). "Sharp detection in PCA under correlations: all eigenvalues matter". In: *Annals of Statistics* 45.4, pp. 1810–1833.

Donoho, D. L. (2006a). "Compressed sensing". In: *IEEE Transactions on Information Theory* 52.4, pp. 1289–1306.

— (2006b). "For Most Large Underdetermined Systems of Linear Equations, The Minimal ℓ_1-Norm Solution Is Also The Sparsest Solution". In: *Communications on Pure and Applied Mathematics* 59.6, pp. 797–829.

Donoho, D. L., Javanmard, A., and Montanari, A. (2013). "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing". In: *IEEE Transactions on Information Theory* 59.11, pp. 7434–7464.

Donoho, D. L., Maleki, A., and Montanari, A. (2009). "Message-passing algorithms for compressed sensing". In: *Proceedings of the National Academy of Sciences* 106.45, pp. 18914–18919.

Du, D. and Hwang, F. (2006). *Pooling Designs and Nonadaptive Group Testing: Important Tools for DNA Sequencing*. Vol. 18. World Scientific Publishing Company.

Durrett, R. (2010). *Probability: theory and examples*. Cambridge university press.

El Alaoui, A. and Jordan, M. I. (2018). "Detection limits in the high-dimensional spiked rectangular model". In: *arXiv preprint arXiv:1802.07309*.

El Alaoui, A. and Krzakala, F. (2018). "Estimation in the Spiked Wigner Model: A Short Proof of the Replica Formula". In: *arXiv preprint arXiv:1801.01593*.

El Alaoui, A., Krzakala, F., and Jordan, M. I. (2017). "Finite Size Corrections and Likelihood Ratio Fluctuations in the Spiked Wigner Model". In: *arXiv preprint arXiv:1710.02903*.

El Alaoui, A. et al. (2016). "Decoding from pooled data: Sharp information-theoretic bounds". In: *arXiv preprint arXiv:1611.09981*.

— (2017). "Decoding from pooled data: Phase transitions of message passing". In: *Information Theory (ISIT), 2017 IEEE International Symposium on*. IEEE, pp. 2780–2784.

Féral, D. and Péché, S. (2007). "The largest eigenvalue of rank one deformation of large Wigner matrices". In: *Communications in Mathematical Physics* 272.1, pp. 185–228.

Franz, S. and Parisi, G. (1995). "Recipes for metastable states in spin glasses". In: *Journal de Physique I* 5.11, pp. 1401–1415.

— (1998). "Effective potential in glassy systems: theory and simulations". In: *Physica A: Statistical Mechanics and its Applications* 261.3-4, pp. 317–339.

Guerra, F. (2001). "Sum rules for the free energy in the mean field spin glass model". In: *Fields Institute Communications* 30, pp. 161–170.

— (2003). "Broken replica symmetry bounds in the mean field spin glass model". In: *Communications in Mathematical Physics* 233.1, pp. 1–12.

Guerra, F. and Toninelli, F. (2002a). "Central limit theorem for fluctuations in the high temperature region of the Sherrington–Kirkpatrick spin glass model". In: *Journal of Mathematical Physics* 43.12, pp. 6224–6237.

— (2002b). "Quadratic replica coupling in the Sherrington–Kirkpatrick mean field spin glass model". In: *Journal of Mathematical Physics* 43.7, pp. 3704–3716.

— (2002c). "The thermodynamic limit in mean field spin glass models". In: *Communications in Mathematical Physics* 230.1, pp. 71–79.

Guo, D. and Verdú, S. (2005). "Randomly spread CDMA: Asymptotics via statistical physics". In: *IEEE Transactions on Information Theory* 51.6, pp. 1983–2010.

Heo, M. et al. (2001). "Pooling Analysis of Genetic Data: The Association of Leptin Receptor (LEPR) Polymorphisms With Variables Related to Human Adiposity". In: *Genetics* 159.3, pp. 1163–1178. ISSN: 0016-6731.

Ingster, Y. and Suslina, I. A. (2012). *Nonparametric goodness-of-fit testing under Gaussian models*. Vol. 169. Springer Science & Business Media.

Johnstone, I. M. (2001). "On the distribution of the largest eigenvalue in principal components analysis". In: *Annals of Statistics*, pp. 295–327.

Johnstone, I. M. and Lu, A. Y. (2009). "On consistency and sparsity for principal components analysis in high dimensions". In: *Journal of the American Statistical Association* 104.486, pp. 682–693.

Johnstone, I. M. and Onatski, A. (2015). "Testing in high-dimensional spiked models". In: *arXiv preprint arXiv:1509.07269*.

Keener, R. W. (2011). *Theoretical statistics: Topics for a core course*. Springer.

Korada, S. B. and Macris, N. (2009). "Exact solution of the gauge symmetric p-spin glass model on a complete graph". In: *Journal of Statistical Physics* 136.2, pp. 205–230.

Krzakala, F., Mézard, M., and Zdeborová, L. (2012). "Reweighted belief propagation and quiet planting for random K-SAT". In: *arXiv preprint arXiv:1203.5521.*

Krzakala, F., Xu, J., and Zdeborová, L. (2016). "Mutual information in rank-one matrix estimation". In: *Information Theory Workshop (ITW).* IEEE, pp. 71–75.

Krzakala, F. and Zdeborová, L. (2009). "Hiding quiet solutions in random constraint satisfaction problems". In: *Physical Review Letters* 102.23, p. 238701.

Le Cam, L. (1960). *Locally Asymptotically Normal Families of Distributions. Certain Approximations to Families of Distributions and Their Use in the Theory of Estimation and Testing Hypotheses.* Berkeley & Los Angeles.

Ledoit, O. and Wolf, M. (2002). "Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size". In: *Annals of Statistics*, pp. 1081–1102.

Lelarge, M. and Miolane, L. (2016). "Fundamental limits of symmetric low-rank matrix estimation". In: *arXiv preprint arXiv:1611.03888.*

— (2017). "Fundamental limits of symmetric low-rank matrix estimation". In: *Proceedings of the 2017 Conference on Learning Theory.* Ed. by S. Kale and O. Shamir. Vol. 65. Proceedings of Machine Learning Research. Amsterdam, Netherlands: PMLR (arXiv:1611.03888), pp. 1297–1301.

Lesieur, T., Krzakala, F., and Zdeborová, L. (2015a). "MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel". In: *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on.* IEEE, pp. 680–687.

— (2015b). "Phase transitions in sparse PCA". In: *IEEE International Symposium on Information Theory (ISIT).* IEEE, pp. 1635–1639.

— (2017). "Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications". In: *Journal of Statistical Mechanics: Theory and Experiment* 2017.7.

Lesieur, T., Miolane, L., et al. (2017). "Statistical and computational phase transitions in spiked tensor estimation". In: *IEEE International Symposium on Information Theory (ISIT).* IEEE, pp. 511–515.

Mézard, M., Parisi, G., Sourlas, N., et al. (1984). "Replica symmetry breaking and the nature of the spin glass phase". In: *Journal de Physique* 45.5, pp. 843–854.

Mézard, M., Parisi, G., and Virasoro, M.-A. (1990). *Spin glass theory and beyond.* World Scientific Publishing.

Mézard, M. and Toninelli, C. (2011). "Group testing with random pools: Optimal two-stage algorithms". In: *IEEE Transactions on Information Theory* 57.3, pp. 1736–1745.

Miolane, L. (2017). "Fundamental limits of low-rank matrix estimation". In: *arXiv preprint arXiv:1702.00473.*

Montanari, A., Reichman, D., and Zeitouni, O. (2015). "On the limitation of spectral methods: From the gaussian hidden clique problem to rank-one perturbations of gaussian tensors". In: *Advances in Neural Information Processing Systems*, pp. 217–225.

Nadler, B. (2008). "Finite sample approximation results for principal component analysis: A matrix perturbation approach". In: *Annals of Statistics*, pp. 2791–2817.

Nishimori, H. (2001). *Statistical physics of spin glasses and information processing: an introduction*. Vol. 111. Clarendon Press.

Onatski, A., Moreira, M. J., and Hallin, M. (2013). "Asymptotic power of sphericity tests for high-dimensional data". In: *Annals of Statistics* 41.3, pp. 1204–1231.

— (2014). "Signal detection in high dimension: The multispiked case". In: *Annals of Statistics* 42.1, pp. 225–254.

Panchenko, D. (2015). "The free energy in a multi-species Sherrington–Kirkpatrick model". In: *Annals of Probability* 43.6, pp. 3494–3513.

Pastur, L. and Shcherbina, M. (1991). "Absence of self-averaging of the order parameter in the Sherrington-Kirkpatrick model". In: *Journal of Statistical Physics* 62.1-2, pp. 1–19.

Paul, D. (2007). "Asymptotics of sample eigenstructure for a large dimensional spiked covariance model". In: *Statistica Sinica*, pp. 1617–1642.

Péché, S. (2006). "The largest eigenvalue of small rank perturbations of Hermitian random matrices". In: *Probability Theory and Related Fields* 134.1, pp. 127–173.

— (2014). "Deformed ensembles of random matrices". In: *Proceedings of the International Congress of Mathematicians, Seoul*. Vol. III. ICM, pp. 1059–1174.

Perry, A. et al. (2016a). "On the optimality and sub-optimality of PCA for spiked random matrix models". In: *Annals of Statistics (to appear). arXiv preprint arXiv:1609.05573*.

— (2016b). "Optimality and sub-optimality of PCA for spiked random matrices and synchronization". In: *arXiv preprint arXiv:1609.05573*.

Rangan, S. et al. (2012). "Hybrid generalized approximate message passing with applications to structured sparsity". In: *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*. IEEE, pp. 1236–1240.

Richard, E. and Montanari, A. (2014). "A statistical model for tensor PCA". In: *Advances in Neural Information Processing Systems (NIPS)*, pp. 2897–2905.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton: Princeton University Press.

Scarlett, J. and Cevher, V. (2017). "Phase transitions in the pooled data problem". In: *Conference on Neural Information Processing Systems (NIPS)*. Long Beach, California.

Sham, P. et al. (2002). "DNA pooling: a tool for large-scale association studies". In: *Nature Reviews Genetics* 3.11, pp. 862–871.

Sherrington, D. and Kirkpatrick, S. (1975). "Solvable model of a spin-glass". In: *Physical Review Letters* 35.26, pp. 1792–1796.

Sly, A., Sun, N., and Zhang, Y. (2016). "The number of solutions for random regular NAE-SAT". In: *arXiv preprint arXiv:1604.08546*.

Spielman, D. *Spectral Graph Theory*. http://www.cs.yale.edu/homes/spielman/561.

Talagrand, M. (2006). "The Parisi formula". In: *Annals of Mathematics*, pp. 221–263.

— (2007). "Mean field models for spin glasses: some obnoxious problems". In: *Spin Glasses*. Springer, pp. 63–80.

— (2011a). *Mean field models for spin glasses. Volume I: Basic examples*. Vol. 54. Springer Science & Business Media.

— (2011b). *Mean field models for spin glasses. Volume II: Advanced replica-symmetry and low temperature*. Vol. 55. Springer Science & Business Media.

Tanaka, T. (2002). "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors". In: *IEEE Transactions on Information theory* 48.11, pp. 2888–2910.

Tao, T. (2012). *Topics in random matrix theory.* Vol. 132. American Mathematical Soc.

Thouless, D. J., Anderson, P. W., and Palmer, R. G. (1977). "Solution of 'solvable model of a spin glass'". In: *Philosophical Magazine* 35.3, pp. 593–601.

Vaaler, J. D. (1979). "A geometric inequality with applications to linear forms." In: *Pacific Journal of Mathematics* 83.2, pp. 543–553.

Van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge University Press.

Verzelen, N. and Arias-Castro, E. (2015). "Community detection in sparse random networks". In: *Ann. Appl. Probab.* 25.6, pp. 3465–3510. DOI: 10.1214/14-AAP1080. URL: https://doi.org/10.1214/14-AAP1080.

Wang, I.-H. et al. (2016). "Data extraction via histogram and arithmetic mean queries: Fundamental limits and algorithms". In: *2016 IEEE International Symposium on Information Theory (ISIT).* IEEE, pp. 1386–1390.

Wu, Y. and Verdú, S. (2009). "Fundamental limits of almost lossless analog compression". In: *2009 IEEE International Symposium on Information Theory.* IEEE, pp. 359–363.

Zdeborová, L. and Krzakala, F. (2016). "Statistical physics of inference: Thresholds and algorithms". In: *Advances in Physics* 65.5, pp. 453–552.

Zhang, P. et al. (2013). "Non-adaptive pooling strategies for detection of rare faulty items". In: *2013 IEEE International Conference on Communications Workshops (ICC).* IEEE, pp. 1409–1414.

Zigangirov, K. S. (2004). *Theory of Code Division Multiple Access Communication.* Vol. 6. John Wiley & Sons.