Alma Mater Studiorum − Università di Bologna
in cotutela con École Normale Supérieure

DOTTORATO DI RICERCA IN

MATEMATICA

Ciclo XXXV

Settore Concorsuale: 01/A4

Settore Scientifico Disciplinare: MAT/07

NEW PERSPECTIVES IN STATISTICAL MECHANICS AND HIGH-DIMENSIONAL INFERENCE

Presentata da: Francesco Camilli

Coordinatore Dottorato:
Prof. Valeria Simoncini

Supervisore:
Prof. Pierluigi Contucci

Supervisore:
Prof. Marc Mézard

Esame finale anno 2023

Abstract

The main purpose of this thesis is to go beyond two usual assumptions that accompany theoretical analysis in spin-glasses and inference: the *i.i.d.* (independently and identically distributed) hypothesis on the noise elements and the finite rank regime. The first one appears since the early birth of spin-glasses. The second one instead concerns more the inference point of view. Disordered systems and Bayesian inference have a well-established relation, which is evidenced by their continuous cross-fertilization. The thesis makes use of techniques coming both from the rigorous mathematical machinery of spin-glasses, such as the interpolation scheme, and from Statistical Physics, such as the replica method.

The starting point of the work are the Sherrington-Kirkpatrick model and the spiked Wigner model. The first is a mean field spin-glass where the couplings are *i.i.d.* Gaussian random variables. The second instead amounts to establish the information theoretical limits in the reconstruction of a fixed low rank matrix, the "spike", blurred by additive Gaussian noise. Assuming Bayes-optimality, namely to know everything about the generating process of the data at disposal, one can prove that the spiked Wigner model is actually a spin-glass in a sub-region of its phase space called Nishimori line. In such setting a whole set of identities and correlation inequalities hold because of the special phase of the spin-glass or of the inferential setting, and they are sufficient to imply replica symmetry. This in turn leads to finite dimensional variational principles for the free energy.

The analysis of the previous models relies heavily on the *i.i.d.* nature of the noise. To weaken this assumption, we can give the couplings an inhomogenous variance profile, thus breaking an overall permutation symmetry among the particles sites and giving rise to the so-called multi-species models. We study two different types of variance profiles: a convex coupling and a deep coupling. This terminology refers to the possible ways we can couple different particles belonging to different species.

Afterwards, we rigorously study the spiked Wigner model out of the Bayes-optimal setting. Among the several ways to break Bayes-optimality, we focus on the mismatching priors case: the Statistician that wants to reconstruct the spike assumes a wrong prior on its matrix elements. We show that the model can be mapped into a spin-glass out of the Nishimori line, and therefore Nishimori identities and replica symmetry break down. As a step further, we then introduce a spiked model in which the noise is drawn from an orthogonal matrix ensemble, thus breaking the independence assumption. Using the replica method, we derive its information theoretical limits when the noise is drawn from an ensemble with quartic matrix potential. We show how to build an Approximate Message Passing algorithm that saturates such limits.

Finally, we tackle the problem of high rank matrix factorization, providing a new point of view. We abandon Bayes-optimality in favor of a convenient mismatched estimation. Since a symmetric rank P matrix can be represented as a collection of P vectors, we aim at finding those vectors iteratively one after the other. At each step of this procedure, that we call decimation, the resulting inference model can be mapped into a spin-glass model whose Hamiltonian is really similar to that of the Hopfield model, and as such it inherits most of its features. Using

iv

the replica method, we compute the free energy associated to each decimation step and show numerically that it is a viable strategy for matrix factorization in certain ranges of the control parameters. Above all, this shows that matrix factorization is possible.

Sommario

Lo scopo principale di questa tesi è quello di superare due assunzioni che solitamente accompagnano l'analisi teorica degli spin-glasses e l'inferenza: l'ipotesi di elementi di rumore *i.i.d.* (indipendentemente e identicamente distribuiti) e il regime di rango finito. La prima compare fin dalla nascita dei vetri di spin. Il secondo invece concerne più il punto di vista dell'inferenza. I sistemi disordinati e l'inferenza bayesiana hanno una relazione consolidata, testimoniata dal loro reciproco e costante stimolo. La tesi si avvale di tecniche provenienti sia dalla trattazione rigorosa dei vetri di spin, come l'interpolazione, sia dalla Fisica Statistica, come il metodo delle repliche.

Il punto di partenza del lavoro sono il modello di Sherrington-Kirkpatrick e il modello spiked Wigner. Il primo è uno spin-glass a campo medio in cui gli accoppiamenti sono variabili aleatorie gaussiane *i.i.d.*. Il secondo invece consiste nello stabilire i limiti nell'ambito della dell'informazione nella ricostruzione di una matrice di rango basso fissat, lo "spike", sporcata da un rumore gaussiano additivo. Assumendo l'ottimalità bayesiana, cioè di conoscere tutto sul processo di generazione dei dati a disposizione, si può dimostrare che il modello spiked Wigner è in realtà uno spin-glass in una sottoregione del suo spazio di fase chiamata linea di Nishimori. In questo contesto, tutta una serie di identità e disuguaglianze di correlazione sono valide grazie alla fase speciale dello spin-glass o del setting inferenziale, e sono sufficienti a implicare la simmetria di replica. Questo a sua volta porta a principi variazionali in dimensione finita per l'energia libera.

L'analisi dei modelli precedenti si basa fortemenete sulla natura *i.i.d.* del rumore. Per indebolire questa ipotesi, possiamo dare agli accoppiamenti un profilo di varianza non omogeneo, rompendo così una simmetria globale di permutazione tra i siti delle particelle e dando origine ai cosiddetti modelli multispecie. Studiamo qui due diversi tipi di profili di varianza: un accoppiamento convesso e un accoppiamento profondo. Questa terminologia si riferisce ai possibili modi di accoppiare particelle appartenenti a specie diverse.

In seguito, studiamo in modo rigoroso il modello spiked Wigner al di fuori del setting Bayes-ottimale. Tra i vari modi di rompere l'ottimalità, ci concentriamo sul caso di prior non concordanti: il ricevitore che vuole ricostruire lo spike assume un prior sbagliato sugli elementi della sua matrice. Dimostriamo che il modello può essere mappato in uno spin-glass fuori dalla linea di Nishimori, e quindi le identità di Nishimori e la simmetria di replica vengono meno.

Come passo successivo, introduciamo uno spiked model in cui il rumore è estratto da un insieme di matrici random ortogonali, rompendo così l'ipotesi di indipendenza. Utilizzando il metodo delle repliche, ne ricaviamo i limiti statistici quando il rumore proviene da un insieme con potenziale matriciale quartico. Mostriamo come costruire un algoritmo di Approximate Message Passing che satura tali limiti.

Infine, affrontiamo il problema della fattorizzazione di matrici di rango elevato, fornendo un nuovo punto di vista. Abbandoniamo l'ottimalità bayesiana a favore di una stima sub-ottimale conveniente. Poiché una matrice simmetrica di rango P può essere rappresentata come un insieme di P vettori, ci proponiamo di trovare tali vettori iterativamente uno dopo

l'altro. A ogni passo di questa procedura, che chiamiamo decimazione, il modello di inferenza risultante può essere mappato in un vetro di spin, la cui hamiltoniana è simile a quella del modello Hopfield, e come tale ne eredita la maggior parte delle caratteristiche. Utilizzando il metodo delle repliche, calcoliamo l'energia libera associata a ciascun passo di decimazione e dimostriamo numericamente che la decimazione è una strategia valida per la fattorizzazione di matrice in determinati intervalli dei parametri di controllo. Soprattutto, questo dimostra che la fattorizzazione di matrici di rango elevato è possibile.

Résumé

L'objectif principal de cette thèse est d'affaiblir deux hypothèses qui accompagnent habituellement l'analyse théorique dans les verres de spin et l'inférence: l'hypothèse d'éléments de bruit *i.i.d.* (indépendamment et identiquement distribué) et le régime de rang fini. La première est apparue dès la naissance des verres de spin. La seconde concerne plutôt le point de vue de l'inférence. Les systèmes désordonnés et l'inférence bayésienne ont une relation bien établie, qui est mise en évidence par leur constante fertilisation croisée. La thèse fait appel à des techniques développées dans l'étude rigoureuse des verres de spin, comme l'interpolation, et de la physique statistique, comme la méthode des répliques.

La thèse commence par une introduction aux modèles de Sherrington-Kirkpatrick et Wigner spiked. Le premier est un verre de spin à champ moyen avec des couplages *i.i.d.* gaussiennes. Le second revient plutôt à établir les limites statistiques dans la reconstruction d'une matrice de rang fini, le "spike", brouillée par un bruit gaussien additif. En supposant l'optimalité bayésienne, c'est-à-dire en sachant tout sur le processus de génération des données à disposition, on peut prouver que le modèle Wigner spiked est en fait un verre de spin dans une sous-région de son espace de phase appelée ligne de Nishimori. Dans un tel contexte, toute une série d'identités et d'inégalités de corrélation sont valables en raison de cette phase spéciale, et elles sont suffisantes pour forcer la symétrie de réplique qui conduit à des principes variationnels en dimension finie pour l'énergie libre.

Les analyses précédentes reposent fortement sur la nature *i.i.d.* du bruit. Pour affaiblir cette hypothèse, on peut donner aux couplages un profil de variance inhomogène, brisant ainsi une symétrie de permutation globale entre les sites de particules et donnant lieu aux modèles multi-espèces. Nous étudions deux types différents de profils de variance: un couplage convexe et un couplage profond. Cette terminologie fait référence aux manières possibles de coupler différentes particules appartenant à des espèces différentes.

Ensuite, nous étudions rigoureusement le modèle Wigner spiked hors du cadre d'optimalité bayésienne. Parmi les différentes façons de briser l'optimalité de Bayes, nous nous concentrons sur le cas de priors non concordants: le statisticien qui veut reconstruire le spike assume un mauvais prior sur les éléments de sa matrice. Nous montrons que le modèle peut être représenté dans un verre de spin hors de la ligne de Nishimori, et donc les identités de Nishimori et la symétrie de réplique se brisent.

Pour aller plus loin, nous introduisons ensuite un spiked model dans lequel le bruit est extrait d'un ensemble de matrices aléatoire orthogonales, brisant ainsi l'hypothèse d'indépendance. En utilisant la méthode des répliques, nous obtenons ses limites théoriques d'information lorsque le bruit est extrait d'un ensemble avec un potentiel de matrice quartique. Nous montrons comment construire un algorithme approximé de passage de messages qui sature ces limites.

Enfin, nous traitons le problème de la factorisation des matrices de rang élevé sous un nouvel angle. Nous abandonnons l'optimalité bayésienne en faveur d'une estimation sous-optimale pratique. Puisqu'une matrice symétrique de rang P peut être représentée comme une collection de P vecteurs, nous cherchons à les trouver itérativement. À chaque étape de cette procédure,

appelée décimation, le modèle d'inférence résultant peut être mappé dans un verre de spin similaire au modèle de Hopfield, et en tant que tel, il hérite de la plupart de ses caractéristiques. En utilisant la méthode des répliques, nous calculons l'énergie libre associée à chaque étape de décimation et montrons numériquement qu'il s'agit d'une stratégie viable pour la factorisation matricielle dans certains intervalles des paramètres de contrôle. Cela montre surtout que la factorisation de matrices est possible.

Contents

Extended summary

The main purpose of this thesis is to go beyond two usual assumptions that accompany theoretical analysis in spin-glasses and inference: the *i.i.d.* (independently and identically distributed) hypothesis on the noise elements, that can be collected in a matrix for our purposes, and the finite rank regime. The first one appears since the early birth of spin glasses. The second one instead concerns more the inference point of view. Even though they might seem two separate worlds, disordered systems and Bayesian inference have a well established relation, which is evidenced by their continuous cross-fertilization and the number of technical tools they share. The thesis makes use of techniques coming both from the rigorous mathematical machinery of spin glasses, such as the *interpolation scheme*, and from Statistical Physics, such as the *replica method*. We mention that, although the latter is not rigorous, it is widely accepted as an exact method, and it is here employed whenever a rigorous solution is out of reach since it can still give precious hints on the rigorous direction to take.

The prototypical models that are introduced in Chapter 1 are the Sherrington-Kirkpatrick (SK) spin glass model and the spiked Wigner inference problem. The first one is a mean field spin glass where the couplings between the N spins in the system are *i.i.d.* Gaussian random variables, and is defined by the Hamiltonian:

$$H_N^{SK}(\boldsymbol{\sigma}) = -\sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j\,, \quad J_{ij} \overset{\text{iid}}{\sim} \mathcal{N}\Big(\frac{J_0}{2N}, \frac{J^2}{2N}\Big)\,, \ \boldsymbol{\sigma} \equiv (\sigma_i)_{i\leq N} \in \{-1, +1\}^N\,, \qquad (1)$$

with associated Boltzmann-Gibbs measure

$$d\mu_N(\boldsymbol{\sigma}) = \frac{e^{-\beta H_N^{SK}(\boldsymbol{\sigma})}}{\mathscr{Z}_N} \prod_{i=1}^{N} d\Big(\delta_1(\sigma_i) + \delta_{-1}(\sigma_i)\Big)\,. \qquad (2)$$

Within the Statistical Mechanics formalism, the main goal is the computation of the thermodynamic limit of its re-scaled log partition function $\mathscr{Z}_N$, usually called *free entropy*, or *pressure*, per particle

$$p_N(\beta) = \frac{1}{N} \log \mathscr{Z}_N\,. \qquad (3)$$

The variables on the particles sites i can be spins as above, or real variables with an *apriori* distribution as below.

The spiked Wigner model instead amounts to establish the information theoretical limits in the reconstruction of a rank one (or fixed low rank) matrix, called spike, blurred by a Wigner matrix, *i.e.* additive Gaussian noise. More precisely, if $\mathbf{x}^* \equiv (x_i^*)_{i \leq N}$ with $x_i^* \overset{\text{iid}}{\sim} P_X$ for some prior distribution P_X, the observation at disposal are

$$\mathbf{y} = \sqrt{\frac{\mu}{2N}} \mathbf{x}^* \mathbf{x}^{*\mathsf{T}} + \mathbf{z} \in \mathbb{R}^{N \times N}, \quad z_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \mu \geq 0 \tag{4}$$

where $\mathbf{x}^* \mathbf{x}^{*\mathsf{T}}$ is the spike. λ is called signal-to-noise ration (SNR). Assuming Bayes-optimality, namely to know everything about the generating process of the data at disposal, one can prove that the spiked Wigner model is actually a spin-glass in a really privileged sub-region of its phase space, called Nishimori line. Indeed, restricting for simplicity the inference problem (4) to binary spins, the posterior measure for $\mathbf{x}^*$ given the $\mathbf{Y}$ is in the form (2) with $\beta = 1$ and a Hamiltonian (1) where $J_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(\mu/2N, \mu/2N)$. Consequently, one can also define an associate pressure per particle for the spiked Wigner model, that is related to the *mutual information* between the observations $\mathbf{Y}$ and the ground truth $\mathbf{x}^*$, a fundamental quantity to characterize the quality of the spike reconstruction. Hence, the pressure per particle plays a central role for both models.

Thermodynamic limits of quantities like the pressure per particle can be expressed as a variational principle over order parameters whose stationary values have a physical or information theoretical interpretation. In the Wigner-spiked model for instance the order parameter is the projection, or *overlap*, of the Bayes-optimal estimator onto the hidden ground truth signal, and it measures the quality of the reconstruction. Thanks to G. Parisi we know that the free entropy of the SK model instead admits such variational representation but with an infinite number of order parameters, a phenomenon known as *replica symmetry breaking*, and this is a substantial difference with Bayes-optimal inference. The reason behind this discrepancy is that in the Bayes-optimal setting a whole set of identities and correlation inequalities hold, as a consequence of Bayes-rule or, as a dual viewpoint, of the special phase of the associated spin-glass. In particular, it can be shown that, under mild hypothesis, they are sufficient to imply *replica symmetry*, which leads to finite dimensional variational principles for the free entropy. Replica symmetry, and its breaking, are strongly related to the concentration of the mentioned order parameters in the thermodynamic limit around their expected values.

The two previous models have stimulated the production of an impressive volume of research papers, and nowadays we can say we know different ways to compute the main thermodynamic/information theoretic quantities even rigorously. However, most of the related studies rely heavily on the *i.i.d.* nature of the couplings, or of the noise. One way to weaken this assumption is that of giving the couplings an inhomogenous variance profile. In practice, we can divide the N spins into K disjoint sub-groups or *species*, $\{\Lambda_r\}_{r \leq K}$, and assign a variance to the i-j coupling according to which sub-group i and j belong to:

$$J_{ij} \sim \mathcal{N}\left(\frac{J_0}{2N}, \frac{J_{rs}^2}{2N}\right), \quad \text{if } i \in \Lambda_r \text{ and } j \in \Lambda_s. \tag{5}$$

This in particular breaks an overall permutation symmetry of the particle sites of the model, or signal components, that would be there otherwise. In spin glasses literature, the arising model is said to be a *multi-species model*, to stress the different behaviours that the various particle sites acquire. In inference a similar idea, called *spatial coupling*, was introduced but with the purpose of studying the statistical limits of the spiked Wigner model. A multi-species model on the Nishimori line can be directly obtained by (4) taking a couple-dependent SNR, just like J^2_{rs}. In Chapters 2 and 3 we study precisely this situation with two different types of variance profiles: a convex coupling, when the eigenvalues of the variance matrix J^2_{rs} are all non-negative, and a deep coupling respectively, when J^2_{rs} is tridiagonal with zeros on the main diagonal. The latter is clearly not in the first cathegory, since a "deep" J^2_{rs} has eigenvalues of alternated sign. For these two models on the Nishimori line we prove rigorously a finite dimensional variational for their limiting pressure per particle, using an adaptive interpolation technique.

In Chapter 4 we rigorously study the spiked Wigner model out of the Bayes-optimal setting. There are several ways to break Bayes-optimality: *(i)* the Statistician who wants to infer the spike does not know the SNR, or *(ii)* she ignores the nature of the noise involved. Here we focus on a third alternative, that is *(iii)* when the prior is mismatched: the Statistician assumes a wrong prior on the x^*_i's. We show that the inference problem can be mapped into a spin-glass out of the Nishimori line, and as a consequence Nishimori identities and replica symmetry break down. Nevertheless, we are still able to prove that the overlap of the, inevitably sub-optimal, estimator with the ground truth self-averages in the thermodynamic limit. This allows us to write a variational principle for the free entropy per particle, that in general is not finite dimensional. In other words, the mismatch can induce replica symmetry breaking. Furthermore, the mutual information in this setting is no longer well defined, hence we are forced to work with the *cross entropy* between the Statistician's guess of the distribution of the observations $\mathbf{y}$ and the true one, that is still linked to the pressure.

Afterwards, we go back to the *i.i.d.* paradigm. We stress that the multi-species setting preserves the independence of the noise elements. In this respect, in Chapter 5 we study a spiked model just like (4) in which the noise is no longer a Wigner matrix, but we draw it from a wider *orthogonal matrix ensemble*. Orthogonal matrix ensembles are characterized by a joint distribution of matrix elements that is invariant under any orthogonal transformation: $\mathbf{Z} \overset{\mathrm{D}}{=} \mathbf{OZO}^\mathsf{T}$. We can write the associated probability distribution in terms of a matrix potential as follows

$$dP_Z(\mathbf{Z}) \propto d\mathbf{Z} \exp\Big(-\frac{N}{2}\mathrm{Tr}V(\mathbf{Z}) \Big), \quad d\mathbf{Z} = \prod_{i \leq j \leq N} dZ_{ij} \tag{6}$$

where V is a function of a real variable, that is extended to matrices: $V(\mathbf{Z}) = \mathbf{O}V(\boldsymbol{\lambda})\mathbf{O}^\mathsf{T}$, with $\boldsymbol{\lambda}$ the diagonal matrix of eigenvalues of $\mathbf{Z}$, $\mathbf{O}$ the orthogonal matrix diagonalizing $\mathbf{Z}$, and V here is applied component-wise to the diagonal.

The Wigner, or Gaussian Orthogonal Ensemble (GOE), corresponds to a quadratic matrix

potential $V(x) = x^2/2$, but here we choose a quartic potential

$$V(x) = \mu\frac{x^2}{2} + \gamma\frac{x^4}{4}\ .$$

This is sufficient to break the independence among the noise matrix elements as long as $\gamma > 0$. In fact, if we take for simplicity $\mu = 0$ and $\gamma = 1$ then

$$dP_Z(\mathbf{Z}) \propto d\mathbf{Z} \prod_{i,j,k,l} \exp\Big(-\frac{N}{8}Z_{ij}Z_{jk}Z_{kl}Z_{li}\Big)$$

which is clearly not factorizeable.

The techniques used in Chapter 5 generalize to higher order potentials, but we stop at the fourth order for the sake of presentation. Since a rigorous derivation of the free entropy of the associated spin-glass seems to be not at reach for the moment, we had to employ the powerful replica method. We then compared the theoretical prediction of the mean square error of the reconstruction with those of existing algorithms present in the literature, designed to take into account the rotational invariant nature of the noise, realizing they were sub-optimal. As a consequence we propose a modified Approximate Message Passing (AMP) algorithm and derive its state evolution rigorously, showing that it matches the replica prediction. It is worth to stress that, despite the formula for the free entropy, or mutual information, is still not rigorous, it helped us in proving rigorously a gap in performance between our AMP and the ones previously dealt with in the literature.

All the aforementioned spiked inference models deal with finite rank perturbations, the spikes, of large noise matrices, $\mathbf{Z}$. As opposed to that, in the final Chapter 6 the problem of extensive rank matrix factorization under Gaussian noise is studied. Here $\mathbf{x}^*$ is no longer a vector of $\mathbb{R}^N$, but an $N \times P$ matrix. Hence the rank of $\mathbf{x}^*\mathbf{x}^{*\intercal}$ now is P, with $P/N \to \alpha > 0$ when $N \to \infty$. Despite the recent efforts in finding its Bayes-optimal statistical limits, the high rank nature of the problem seems to be an insurmountable obstacle to the production of closed formulae. Hence, we abandon Bayes-optimality in favour of an apparently easier mismatched estimation. In particular, since we can think $\mathbf{x}^*$ as a collection of P vectors $(\mathbf{x}^{*t})^{t \leq P}$, we aim at finding those vectors one after the other. The mismatch is induced because it is as if we were trying to retrieve a rank-one matrix, when the actual inference problem is intrinsically a high rank one. At the first step, the estimation problem can be mapped into a spin glass whose Hamiltonian is really similar to that of Hopfield model, and we show that it inherits most of its features. After the first vector, that can be thought as a *pattern* in neural networks language, is estimated, we insert in the Hamiltonian a repulsive term towards the corresponding direction. This procedure can be iterated till all the patterns are estimated, and we call it *decimation*. Using the replica method, we compute the free entropy associated to each decimation step. From the fixed point equations with a sparse Ising prior $P_X = (1 - \rho)\delta_0 + \frac{\rho}{2}[\delta_{-1/\sqrt{\rho}} + \delta_{1/\sqrt{\rho}}]$ we show numerically that decimation is a viable strategy for matrix factorization in certain ranges of the control parameters. Above all, this shows that matrix factorization is possible.

Sommario esteso

L'obbiettivo di questa tesi è quello di superare due assunzioni che solitamente accompagnano l'analisi teorica dei vetri di spin e l'inferenza: l'ipotesi di elementi di rumore *i.i.d.* (indipendentemente e identicamente distribuiti), che possono essere raccolti in una matrice per i nostri scopi, e il regime di rango finito. La prima compare fin dalla nascita dei vetri di spin. Il secondo invece concerne più il punto di vista inferenziale. Anche se potrebbero sembrare due mondi separati, i sistemi disordinati e l'inferenza bayesiana hanno una relazione ben consolidata, testimoniata dai loro reciproci stimoli e dal numero di strumenti che condividono. La tesi si avvale di tecniche provenienti sia dalla trattazione rigorosa dei vetri di spin, come l'interpolazione, sia dalla Fisica Statistica, come il metodo delle repliche. Si sottolinea che quest'ultimo, pur non essendo rigoroso, è ampiamente accettato come metodo esatto, e viene qui impiegato ogni volta che una soluzione rigorosa non è accessibile, in quanto può comunque fornire preziosi suggerimenti sulla direzione rigorosa da prendere.

I modelli prototipo introdotti nel Capitolo 1 sono il modello di spin-glass di Sherrington-Kirkpatrick (SK) e lo spiked Wigner model. Il primo è un vetro di spin a campo medio in cui gli accoppiamenti tra gli N spin del sistema sono variabili aleatorie gaussiane *i.i.d.*, ed è definito dall'hamiltoniana:

$$H_N^{SK}(\boldsymbol{\sigma}) = -\sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j\,, \quad J_{ij} \overset{\text{iid}}{\sim} \mathcal{N}\Big(\frac{J_0}{2N}, \frac{J^2}{2N}\Big)\,, \; \boldsymbol{\sigma} \equiv (\sigma_i)_{i\leq N} \in \{-1,+1\}^N\,, \qquad (1)$$

con un misura di Boltzmann-Gibbs associata

$$d\mu_N(\boldsymbol{\sigma}) = \frac{e^{-\beta H_N^{SK}(\boldsymbol{\sigma})}}{\mathscr{Z}_N} \prod_{i=1}^{N} d\Big(\delta_1(\sigma_i) + \delta_{-1}(\sigma_i)\Big)\,. \qquad (2)$$

Seguendo il formalismo della Meccanica Statistica, l'obiettivo principale è il calcolo del limite termodinamico del logaritmo della sua funzione di partizione $\mathscr{Z}_N$, normalizzato con N, solitamente chiamato *entropia libera*, o *pressione*, per particella

$$p_N(\beta) = \frac{1}{N} \log \mathscr{Z}_N\,. \qquad (3)$$

Le variabili sui siti di particella i possono essere spin come sopra, o variabili reali con una distribuzione *apriori* come segue.

Il modello spiked Wigner consiste invece nello stabilire i limiti in ambito di teoria dell'informazione nella ricostruzione di una matrice di rango uno (o di rango finito fissato), chiamata spike, sporcata da una matrice di Wigner, cioè da rumore gaussiano additivo. Più precisamente, se $\mathbf{x}^* \equiv (x_i^*)_{i \leq N}$ con $x_i^* \overset{\text{iid}}{\sim} P_X$ per una qualche distribuzione a priori P_X, le osservazioni a disposizione sono

$$\mathbf{y} = \sqrt{\frac{\mu}{2N}}\mathbf{x}^*\mathbf{x}^{*\intercal} + \mathbf{z} \in \mathbb{R}^{N \times N}, \quad z_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \mu \geq 0 \tag{4}$$

dove $\mathbf{x}^*\mathbf{x}^{*\intercal}$ è lo spike. λ è chiamato signal-to-noise ratio (SNR). Assumendo ottimalità bayesiana, cioè di sapere tutto sul processo di generazione dei dati a disposizione, si può dimostrare che lo spiked Wigner model è in realtà un vetro di spin in una sottoregione privilegiata del suo spazio delle fasi, chiamata linea di Nishimori. Infatti, limitando per semplicità il problema inferenziale (4) agli spin binari, la misura a posteriori per $\mathbf{x}^*$ data $\mathbf{Y}$ è nella forma (2) con $\beta = 1$ e un'hamiltoniana (1) dove $J_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(\mu/2N, \mu/2N)$. Di conseguenza, si può anche definire una pressione per particella associata allo spiked Wigner model, che è legata alla mutua informazione tra le osservazioni $\mathbf{Y}$ e il segnale sottostante $\mathbf{x}^*$, una quantità fondamentale per caratterizzare la qualità della ricostruzione dello spike. Pertanto, la pressione per particella svolge un ruolo chiave per entrambi i modelli.

I limiti termodinamici di quantità come la pressione per particella possono essere espressi attraverso un principio variazionale su parametri d'ordine i cui valori stazionari hanno un' interpretazione fisica o in teoria dell'informazione. Nello spiked Wigner model, ad esempio, il parametro d'ordine è la proiezione, o *overlap*, dello stimatore Bayes-ottimale sul segnale $\mathbf{x}^*$ e misura la qualità della ricostruzione. Grazie a G. Parisi è noto che l'entropia libera del modello SK ammette invece tale rappresentazione variazionale, ma con un numero infinito di parametri d'ordine, un fenomeno noto come *rottura della simmetria di replica*, e questa costituisce una differenza sostanziale con l'inferenza in setting Bayes-ottimale. La ragione di questa discrepanza è che nel setting Bayes-ottimale valgono tutta una serie di identità e disuguaglianze di correlazione come conseguenza della regola di Bayes o, da un punto di vista duale, della fase speciale dello spin-glass associato. In particolare, si può dimostrare che, sotto ipotesi blande, queste sono sufficienti a implicare la simmetria di replica, che porta a principi variazionali in dimensione finita per l'entropia libera. La simmetria di replica, e la sua rottura, sono strettamente legate alla concentrazione dei parametri d'ordine citati nel limite termodinamico intorno ai loro valori attesi.

I due modelli precedenti hanno stimolato la produzione di un'impressionante mole di articoli, e ad oggi possiamo dire di conoscere diversi modi per calcolare le principali quantità termodinamiche/in teoria dell'informazione anche in modo rigoroso. Tuttavia, la maggior parte degli studi a tal riguardo si basa fortemente sulla natura *i.i.d.* degli accoppiamenti, o del rumore. Un modo per indebolire questa assunzione è quello di dare agli accoppiamenti un profilo di varianza disomogeneo. In pratica, possiamo dividere gli N spin in K sottogruppi, o *specie*, disgiunti, $\{\Lambda_r\}_{r \leq K}$, e assegnare una varianza all'accoppiamento i-j a seconda del sottogruppo a cui i e j

appartengono:

$$J_{ij} \sim \mathcal{N}\Big(\frac{J_0}{2N}, \frac{J_{rs}^2}{2N}\Big), \quad \text{se } i \in \Lambda_r \text{ e } j \in \Lambda_s. \tag{5}$$

Questo rompe una simmetria di permutazione globale tra siti di particelle del modello, o tra le componenti del segnale, che sarebbe altrimenti presente. Nella letteratura sui vetri di spin, si dice che il modello derivante è un modello *multi-specie*, per sottolineare il diverso comportamento che i vari siti di particelle acquisiscono. In inferenza è stata introdotta un'idea simile, chiamata *accoppiamento spaziale*, ma con lo scopo di studiare i limiti statistici dello spiked Wigner model. Un modello multispecie sulla linea di Nishimori può essere ottenuto direttamente da (4) prendendo un SNR dipendente dai siti accoppiati, come per J_{rs}^2. Nei Capitoli 2 e 3 studiamo proprio questa situazione con due diversi tipi di profili di varianza: un accoppiamento convesso, quando gli autovalori della matrice di varianza J_{rs}^2 sono tutti non negativi, e un accoppiamento profondo rispettivamente, quando J_{rs}^2 è tridiagonale con zeri sulla diagonale principale. Quest'ultimo chiaramente non rientra nella prima categoria, poiché J_{rs}^2 avrebbe autovalori di segno alterno. Per questi due modelli sulla linea di Nishimori dimostriamo in modo rigoroso un principio variazionale finito dimensionale per la loro pressione limite per particella, utilizzando una tecnica di interpolazione adattiva.

Nel Capitolo 4 studiamo in modo rigoroso lo spiked Wigner model fuori dal setting Bayesottimale. Ci sono diversi modi per rompere l'ottimalità: l'osservatore che vuole dedurre lo spike *(i)* non conosce l'SNR, oppure *(ii)* ignora la natura del rumore. In questa sede ci concentriamo su una terza alternativa, ovvero *(iii)* quando l'osservatore assume un prior sbagliato sugli x_i^*. Dimostriamo che il problema inferenziale può essere mappato in uno spin-glass fuori dalla linea di Nishimori, e di conseguenza le identità di Nishimori e la simmetria di replica vengono meno. Ciononostante, siamo ancora in grado di dimostrare che l'overlap dello stimatore, inevitabilmente subottimale, con il segnale sottostante si auto-media nel limite termodinamico. Questo ci permette di scrivere un principio variazionale per l'entropia libera per particella, che in generale non è in dimensione finita. In altre parole, il mismatch può indurre la rottura della simmetria di replica. Inoltre, la mutua informazione in questo contesto non è più ben definita, quindi siamo costretti a lavorare con l'*entropia incrociata* tra la distribuzione delle osservazioni **y** ipotizzata dal ricevitore e quella vera, che è ancora legata alla pressione.

In seguito, torniamo al paradigma *i.i.d.*. Sottolineiamo che il setting multispecie preserva l'indipendenza degli elementi di rumore. A questo proposito, nel Capitolo 5 studiamo uno spiked model, come (4), in cui il rumore non è più una matrice di Wigner, ma è estratto da un più ampio ensemble matriciale. Gli ensemble su cui ci concentriamo sono caratterizzati da una distribuzione congiunta degli elementi di matrice che è invariante per trasformazioni ortogonali: $\mathbf{Z} \stackrel{\mathrm{D}}{=} \mathbf{OZO}^\intercal$. Possiamo scrivere la distribuzione di probabilità associata in termini di un potenziale matriciale come segue

$$dP_Z(\mathbf{Z}) \propto d\mathbf{Z}\exp\Big(-\frac{N}{2}\mathrm{Tr}V(\mathbf{Z})\Big), \quad d\mathbf{Z} = \prod_{i \leq j \leq N} dZ_{ij} \tag{6}$$

dove V è una funzione di una variabile reale, estesa a variabili matriciali: $V(\mathbf{Z}) = \mathbf{O}V(\boldsymbol{\lambda})\mathbf{O}^{\mathsf{T}}$, con $\boldsymbol{\lambda}$ la matrice diagonale degli autovalori di $\mathbf{Z}$, $\mathbf{O}$ la matrice ortogonale che diagonalizza $\mathbf{Z}$, e V applicata componente per componente alla diagonale.

L'ensemble di Wigner, o Gaussiano Ortogonale (GOE), corrisponde a un potenziale di matrice quadratico $V(x) = x^2/2$; qui ci concentriamo invece su un potenziale quartico

$$V(x) = \mu \frac{x^2}{2} + \gamma \frac{x^4}{4}\,.$$

Questo è sufficiente a rompere l'indipendenza tra gli elementi della matrice di rumore a patto che $\gamma > 0$. Infatti, se prendiamo per semplicità $\mu = 0$ e $\gamma = 1$ allora

$$dP_Z(\mathbf{Z}) \propto d\mathbf{Z} \prod_{i,j,k,l} \exp\Big(-\frac{N}{8} Z_{ij} Z_{jk} Z_{kl} Z_{li}\Big)$$

che chiaramente non è fattorizzabile.

Le tecniche utilizzate nel Capitolo 5 sono generalizzabili a potenziali di ordine superiore, ma per comodità di presentazione ci fermiamo al quarto ordine. Poiché una derivazione rigorosa dell'entropia libera dello spin-glass associato non sembra essere accessibile al momento, ricorriamo al versatile metodo delle repliche. Confrontiamo successivamente la previsione teorica dell'errore quadratico medio della ricostruzione con quella di algoritmi esistenti in letteratura, progettati per tenere conto dell'invarianza rotazionale del rumore, realizzando che sono sub-ottimali. Di conseguenza, proponiamo un algoritmo di Approximate Message Passing (AMP) modificato e deriviamo la sua evoluzione di stato in modo rigoroso, dimostrando il suo accordo con la previsione ottenuta dal calcolo con le repliche. Vale la pena sottolineare che, nonostante la formula per l'entropia libera, o mutua informazione, non sia ancora rigorosa, essa ci ha aiutato a dimostrare in modo rigoroso un gap di prestazioni tra il nostro AMP e quelli precedentemente trattati in letteratura.

Tutti gli spiked models menzionati in precedenza trattano perturbazioni di rango finito, gli spike, di grandi matrici di rumore, $\mathbf{Z}$. Al contrario, nel Capitolo 6 ci occupiamo del problema della fattorizzazione di matrici di rango estensivo in presenza di rumore gaussiano. In questo caso, $\mathbf{x}^*$ non è più un vettore in $\mathbb{R}^N$, ma una matrice $N \times P$. Quindi il rango di $\mathbf{x}^* \mathbf{x}^{*\mathsf{T}}$ è ora P, con $P/N \to \alpha > 0$ quando $N \to \infty$. Nonostante i recenti sforzi per trovare i limiti statistici, la presenza di un rango estensivo nel problema sembra essere un ostacolo insormontabile alla produzione di formule chiuse. Per questo motivo, abbandoniamo l'ottimalità bayesiana a favore di una stima subottimale apparentemente più semplice. In particolare, dal momento che possiamo pensare $\mathbf{x}^*$ come a un insieme di P vettori $(\mathbf{x}^{*t})^{t \leq P}$, puntiamo a trovare questi vettori uno dopo l'altro, iterativamente. Al primo passo, il problema inferenziale può essere mappato in un vetro di spin la cui hamiltoniana è simile a quella del modello di Hopfield, e mostriamo che ne eredita la maggior parte delle caratteristiche. Dopo aver stimato il primo vettore, che nel linguaggio delle reti neurali può essere considerato come un *pattern*, inseriamo nell'hamiltoniana un termine repulsivo verso la direzione corrispondente. Questa procedura, chiamata *decimazione*,

può essere iterata fino a quando tutti i pattern sono stati stimati. Utilizzando il metodo delle repliche, calcoliamo l'entropia libera associata a ciascun passo di decimazione. Dalle equazioni di punto fisso con un prior Ising con sparsità $P_X = (1 - \rho)\delta_0 + \frac{\rho}{2}[\delta_{-1/\sqrt{\rho}} + \delta_{1/\sqrt{\rho}}]$ mostriamo numericamente che la decimazione è una strategia valida per la fattorizzazione di matrice in determinati intervalli dei parametri di controllo. Soprattutto, ciò mostra che la fattorizzazione di matrice è possibile.

Résumé détaillé

L'objectif principal de cette thèse est d'affaiblir deux hypothèses qui accompagnent habituellement l'analyse théorique dans les verres de spin et l'inférence: l'hypothèse d'éléments de bruit *i.i.d.* (indépendamment et identiquement distribué), qui peuvent être rassemblés dans une matrice pour nos besoins, et le régime de rang fini. La première hypothèse apparaît dès la naissance des verres de spin. La seconde concerne plutôt le point de vue de l'inférence. Bien qu'ils puissent sembler deux mondes séparés, les systèmes désordonnés et l'inférence bayésienne ont une relation bien établie, ce qui est mis en évidence par leur constante fertilisation croisée et le nombre d'outils techniques qu'ils partagent. La thèse fait appel à des techniques développées dans l'étude rigoureuse des verres de spin, comme l'*interpolation*, et de la physique statistique, comme la *méthode des répliques*. Nous mentionnons que, bien que cette dernière ne soit pas rigoureuse, elle est largement acceptée comme une méthode exacte, et elle est employée ici chaque fois qu'une solution rigoureuse est hors de portée, car elle peut toujours donner des indications précieuses sur la direction rigoureuse à prendre.

Les modèles prototypiques présentés dans le Chapitre 1 sont le modèle de verre de spin de Sherrington-Kirkpatrick (SK) et le problème d'inférence spiked Wigner. Le premier est un verre de spin à champ moyen où les couplages entre les N spins du système sont des variables aléatoires gaussiennes *i.i.d.*, et est défini par l'hamiltonien:

$$H_N^{SK}(\boldsymbol{\sigma}) = -\sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j \,, \quad J_{ij} \overset{\text{iid}}{\sim} \mathcal{N}\Big(\frac{J_0}{2N}, \frac{J^2}{2N}\Big) \,, \ \boldsymbol{\sigma} \equiv (\sigma_i)_{i\leq N} \in \{-1,+1\}^N \,, \tag{1}$$

avec la mesure de Boltzmann-Gibbs associée

$$d\mu_N(\boldsymbol{\sigma}) = \frac{e^{-\beta H_N^{SK}(\boldsymbol{\sigma})}}{\mathscr{Z}_N} \prod_{i=1}^{N} d\Big(\delta_1(\sigma_i) + \delta_{-1}(\sigma_i)\Big) \,. \tag{2}$$

Dans le cadre du formalisme de la mécanique statistique, l'objectif principal est le calcul de la limite thermodynamique du logarithme de la fonction de partition $\mathscr{Z}_N$, généralement appelée *entropie libre*, ou *pression*, par particule:

$$p_N(\beta) = \frac{1}{N}\log \mathscr{Z}_N \,. \tag{3}$$

Les variables sur les sites i des particules peuvent être des spins comme ci-dessus, ou des variables réelles avec une distribution *apriori* comme ci-dessous.

Le modèle spiked Wigner revient plutôt à établir les limites de la théorie de l'information dans la reconstruction d'une matrice de rang un (ou de rang faible fixe), appelée spike, brouillée par une matrice de Wigner, *i.e.* bruit gaussien additif. Plus précisément, si $\mathbf{x}^* \equiv (x_i^*)_{i \le N}$ avec $x_i^* \overset{\text{iid}}{\sim} P_X$ pour une certaine distribution *apriori* P_X, les observations dont on dispose sont les suivantes

$$\mathbf{y} = \sqrt{\frac{\mu}{2N}}\mathbf{x}^*\mathbf{x}^{*\intercal} + \mathbf{z} \in \mathbb{R}^{N \times N}, \quad z_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \mu \ge 0 \tag{4}$$

où $\mathbf{x}^*\mathbf{x}^{*\intercal}$ est le spike. On appelle λ le rapport signal/bruit (signal-to-noise ratio, SNR). En supposant l'optimalité bayésienne, c'est-à-dire en sachant tout sur le processus de génération des données à disposition, on peut prouver que le modèle spiked Wigner est en fait un verre de spin dans une sous-région vraiment privilégiée de son espace de phase, appelée ligne de Nishimori. En effet, en restreignant pour des raisons de simplicité le problème d'inférence (4) aux spins binaires, la mesure de probabilité *a posteriori* pour $\mathbf{x}^*$ étant donné le $\mathbf{Y}$ est de la forme (2) avec $\beta = 1$ et un hamiltonien (1) où $J_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(\mu/2N, \mu/2N)$. Par conséquent, on peut également définir une pression par particule associée au modèle spiked Wigner, qui est liée à la *information mutuelle* entre les observations $\mathbf{Y}$ et le signal $\mathbf{x}^*$, une quantité fondamentale pour caractériser la qualité de la reconstruction des spikes. Ainsi, la pression par particule joue un rôle central pour les deux modèles.

Les limites thermodynamiques de quantités comme la pression par particule peuvent être exprimées par un principe variationnel sur des paramètres d'ordre dont les valeurs stationnaires ont une interprétation physique ou dans la théorie de l'information. Dans le modèle spiked Wigner, par exemple, le paramètre d'ordre est la projection, ou *overlap*, de l'estimateur bayésien optimal sur le signal caché $\mathbf{x}^*$, et il mesure la qualité de la reconstruction. Grâce à G. Parisi, nous savons que l'entropie libre du modèle SK admet une telle représentation variationnelle mais avec un nombre infini de paramètres d'ordre, un phénomène connu sous le nom de *replica symmetry breaking*, et ceci constitue une différence substantielle avec l'inférence Bayes-optimale. La raison de cette différence est que dans le cadre de l'inférence Bayes-optimale, tout un ensemble d'identités et d'inégalités de corrélation sont valables, en conséquence de la règle de Bayes ou, d'un point de vue dual, de la phase spéciale du verre de spin associé. En particulier, on peut montrer que, sous des hypothèses faibles, elles sont suffisantes pour impliquer la *symétrie de réplique*, qui conduit à des principes variationnels en dimension finie pour l'entropie libre. La symétrie de réplique, et sa rupture, sont fortement liées à la concentration des paramètres d'ordre mentionnés dans la limite thermodynamique autour de leurs valeurs moyennes.

Les deux modèles précédents ont stimulé la production d'un volume impressionnant d'articles de recherche, et aujourd'hui nous pouvons dire que nous connaissons différentes façons de calculer les principales quantités thermodynamiques, même de façon rigoureuse. Cependant, la plupart des études connexes reposent fortement sur la nature *i.i.d.* des couplages, ou du bruit.

Une façon d'affaiblir cette hypothèse est de donner aux couplages un profil de variance inhomogène. En pratique, nous pouvons diviser les N spins en K sous-groupes, ou *espèces*, disjoints, $\{\Lambda_r\}_{r \leq K}$, et attribuer une variance au couplage i-j selon le sous-groupe auquel i et j appartiennent:

$$J_{ij} \sim \mathcal{N}\Big(\frac{J_0}{2N}, \frac{J_{rs}^2}{2N}\Big), \quad \text{si } i \in \Lambda_r \text{ et } j \in \Lambda_s. \tag{5}$$

Cela brise en particulier une symétrie de permutation globale entre les sites de particules du modèle, qui existerait autrement. Dans la littérature sur les verres de spin, on dit que le modèle qui en résulte est un *modèle multi-espèces*, pour souligner les différents comportements que les divers sites de particules acquièrent. En inférence, une idée similaire, appelée *couplage spatial*, a été introduite mais dans le but d'étudier les limites statistiques du modèle spiked Wigner. Un modèle multi-espèces sur la ligne de Nishimori peut être directement obtenu en (4) prenant un SNR dépendant des sites couplés, tout comme J_{rs}^2. Dans les Chapitres 2 et 3, nous étudions précisément cette situation avec deux types différents de profils de variance: un couplage convexe, lorsque les valeurs propres de la matrice de variance J_{rs}^2 sont toutes non négatives, et un couplage profond respectivement, lorsque J_{rs}^2 est tridiagonale avec des zéros sur la diagonale principale. Ce dernier n'est clairement pas dans la première catégorie, puisqu'un J_{rs}^2 profonde a des valeurs propres de signe alterné. Pour ces deux modèles sur la ligne de Nishimori, nous prouvons rigoureusement un principe variationnel en dimension finie pour leur pression limite par particule, en utilisant une technique d'interpolation adaptive.

Dans le Chapitre 4, nous étudions rigoureusement le modèle spiked Wigner hors du cadre de l'optimalité bayésienne. Il existe plusieurs façons de briser l'optimalité bayésienne: *(i)* le statisticien qui veut déduire le spike ne connaît pas le SNR, ou *(ii)* il ignore la nature du bruit. Nous nous concentrons ici sur une troisième alternative, c'est-à-dire *(iii)* lorsque le prior est mal adapté : le statisticien assume un prior erroné sur les x_i^*. Nous montrons que le problème d'inférence peut être représenté par une verre de spin hors de la ligne de Nishimori, et par conséquent, les identités de Nishimori et la symétrie des répliques ne sont plus valables. Néanmoins, nous sommes toujours capables de prouver que l'overlap de l'estimateur, inévitablement sous-optimal, avec le signal se concentre dans la limite thermodynamique. Cela nous permet d'écrire un principe variationnel pour l'entropie libre par particule, qui en général n'est pas de dimension finie. En d'autres termes, le *mismatch* peut induire une rupture de symétrie des répliques. De plus, l'information mutuelle dans ce contexte n'est plus bien définie, ce qui nous oblige à travailler avec la *entropie croisée* entre la supposition du statisticien de la distribution des observations $\mathbf{y}$ et la vraie, qui est toujours liée à la pression.

Ensuite, nous revenons au paradigme *i.i.d.*. Nous soulignons que le cadre multi-espèces préserve l'indépendance des éléments de bruit. À cet égard, dans le Chapitre 5, nous étudions un spiked modèle tout comme (4) dans lequel le bruit n'est plus une matrice de Wigner, mais il est extrait d'un *ensemble de matrices orthogonales*. Les ensembles de matrices orthogonales sont caractérisés par une distribution conjointe des éléments de matrice qui est invariante sous transformation orthogonale: $\mathbf{Z} \overset{\text{D}}{=} \mathbf{O}\mathbf{Z}\mathbf{O}^{\mathsf{T}}$. Nous pouvons écrire la distribution de probabilité

associée en termes de potentiel matriciel comme suit

$$dP_Z(\mathbf{Z}) \propto d\mathbf{Z} \exp\Big(-\frac{N}{2}\mathrm{Tr}V(\mathbf{Z})\Big), \quad d\mathbf{Z} = \prod_{i \le j \le N} dZ_{ij} \tag{6}$$

où V est une fonction d'une variable réelle, que l'on étend aux matrices: $V(\mathbf{Z}) = \mathbf{O}V(\boldsymbol{\lambda})\mathbf{O}^{\intercal}$, avec $\boldsymbol{\lambda}$ la matrice diagonale des valeurs propres de $\mathbf{Z}$, $\mathbf{O}$ la matrice de vecteurs propres de $\mathbf{Z}$, et V est ici appliqué élément par élément à la diagonale.

L'ensemble orthogonal de Wigner, ou ensemble orthogonal gaussien (GOE), correspond à un potentiel matriciel quadratique $V(x) = x^2/2$, mais nous choisissons ici un potentiel quartique

$$V(x) = \mu\frac{x^2}{2} + \gamma\frac{x^4}{4}.$$

Ceci est suffisant pour briser l'indépendance entre les éléments de la matrice de bruit tant que $\gamma > 0$. En fait, si nous prenons pour la simplicité $\mu = 0$ et $\gamma = 1$ alors

$$dP_Z(\mathbf{Z}) \propto d\mathbf{Z} \prod_{i,j,k,l} \exp\Big(-\frac{N}{8}Z_{ij}Z_{jk}Z_{kl}Z_{li}\Big)$$

qui est manifestement non factorisable.

Les techniques utilisées dans le Chapitre 5 se généralisent aux potentiels d'ordre supérieur, mais nous nous arrêtons au quatrième ordre pour les besoins de la présentation. Puisqu'une dérivation rigoureuse de l'entropie libre du spin-glass associé semble ne pas être à portée de main pour le moment, nous avons dû employer la méthode des répliques. Nous avons ensuite comparé la prédiction théorique de l'erreur quadratique moyenne de la reconstruction (mean square error, MSE) avec celles d'algorithmes existants présents dans la littérature, conçus pour prendre en compte l'invariance rotationnelle du bruit, réalisant qu'ils étaient sous-optimaux. En conséquence, nous proposons un algorithme AMP (Approximate Message Passing) modifié et dérivons rigoureusement son évolution d'état, en montrant qu'il correspond à la prédiction des répliques. Il convient de souligner que, bien que la formule de l'entropie libre, ou information mutuelle, ne soit pas encore rigoureuse, elle nous a aidés à prouver de manière rigoureuse un écart de performance entre notre AMP et les algorithmes précédemment traités dans la littérature.

Tous les spiked modèles d'inférence mentionnés ci-dessus traitent des perturbations de rang fini, les spikes, de grandes matrices de bruit, $\mathbf{Z}$. Par opposition, le dernier Chapitre 6 traite du problème de la factorisation des matrices de rang extensif sous un bruit gaussien. Ici, $\mathbf{x}^*$ n'est plus un vecteur en $\mathbb{R}^N$, mais une matrice $N \times P$. Par conséquent, le rang de $\mathbf{x}^*\mathbf{x}^{*\intercal}$ est maintenant P, avec $P/N \to \alpha > 0$ lorsque $N \to \infty$. Malgré les efforts récents pour trouver ses limites statistiques Bayes-optimales, la nature de rang élevé du problème semble être un obstacle insurmontable. Par conséquent, nous abandonnons l'optimalité bayésienne en faveur d'une estimation sous-optimale pratique. En particulier, puisque nous pouvons considérer $\mathbf{x}^*$ comme une collection de P vecteurs $(\mathbf{x}^{*t})^{t \le P}$, nous cherchons à trouver ces vecteurs l'un après

l'autre. Lors de la première étape, le problème d'estimation peut être représenté par un verre de spin dont l'hamiltonien est très similaire à celui du modèle de Hopfield, et nous montrons qu'il hérite de la plupart de ses caractéristiques. Après l'estimation du premier vecteur, qui peut être considéré comme un *pattern* dans le langage des réseaux neuronaux, nous insérons dans le hamiltonien un terme répulsif vers la direction correspondante. Cette procédure peut être itérée jusqu'à tous les patterns soient estimés, et nous l'appelons *décimation*. En utilisant la méthode des répliques, nous calculons l'entropie libre associée à chaque étape de décimation. À partir des équations du point fixe avec un prior d'Ising avec sparsité $P_X = (1-\rho)\delta_0 + \frac{\rho}{2}[\delta_{-1/\sqrt{\rho}} + \delta_{1/\sqrt{\rho}}]$, nous montrons numériquement que la décimation est une stratégie viable pour la factorisation matricielle dans certains intervalles des paramètres de contrôle. Cela montre surtout que la factorisation de matrices est possible.

Chapter 1

Basic notions

In this chapter we give the basic definitions needed throughout the thesis. The starting point is the celebrated Sherrington-Kirkpatrick (SK) spin glass model [1]. Given the volume of research papers published on this model in the last 40 years, both on the Physics and the Mathematics side, it will be impossible to be exhaustive about this topic. Hence, we shall restrict to the mathematical tools and physical ansatz's that will be useful to our analysis.

SK is a mean field version of the Edwards-Anderson spin glass model [2], which can be thought as nearest neighbour interaction model and was in turn first introduced to explain the appearance of cusps in the susceptibility of metallic alloys [3]. Sherrington and Kirkpatrick provided themselves a candidate free energy, the so-called *replica symmetric* solution. However, they were aware that their solution had a flaw: in the low temperature limit it yields a negative entropy, which is not allowed. Giorgio Parisi proposed instead a *replica symmetry breaking* solution [4, 5] via the *replica method* [6], that needs the introduction of a collection of infinite order parameters. The replica method is a really powerful tool that will be used in the thesis whenever a rigorous alternative is out of reach. Though not rigorous strictly speaking, it is widely accepted as an exact method. Nevertheless, a formal proof of the variational principle for SK was achieved nearly 30 years later thanks to Francesco Guerra [7] and Michel Talagrand [8], that were able to find matching bounds for the free energy from below and above respectively. The proof, that requires a remarkable technical effort, was further simplified by Panchenko, who was able to connect the Ghirlanda-Guerra identities in spin glasses [9, 10, 11] to ultrametricity in the SK model [12].

Secondly, an introduction to high dimensional Bayesian Inference follows. Here we define the main Information Theoretic quantities in order to analyze the statistical limits to the reconstruction of some signals blurred by additive Gaussian noise (at least in this introductory section). The classic example of such inference tasks is the spiked Wigner model, that is largely studied in (not only) recent literature. We will then show how to map this problem into a disordered Statistical Mechanics model similar to SK. This will allow us to borrow some tools from the mathematical machinery of disordered systems to establish its statistical limits. The key idea will be the so-called *adaptive interpolation* [13, 14] (see also [15, 16, 17]), an adaptive

version of that of Guerra's.

The two previous models in some sense represent two paradigms that this thesis aims to break, but at the same time they provide us with mechanisms and ideas that pervade modern Statistical Physics.

1.1 Statistical Mechanics formalism

Before we get started, some initial definitions are in order. For our purposes, each Statistical Mechanics model will be defined once its *Hamiltonian*, or *Energy* or *Cost function*, is given. The role of the Hamiltonian is to assign a given energy to each possible configuration state of a system of many particles, whose number is denoted by N. It can depend on some parameters and on the degrees of freedom of said particles. For our analysis we are not interested in translational or rotational degrees of freedom of these particles, so the Hamiltonian will depend only on their internal degrees of freedom: typically a real number drawn from an *apriori* distribution P, or simply ± 1 in case of Ising spins. A generic Hamiltonian will thus be denoted by $H_N(\mathbf{x}; a)$ where a is a collection of possibly random parameters and $\mathbf{x}$ are the particles degrees of freedom.

The set of possible configurations of the system $\mathcal{X}_N$ is endowed with the (possibly random) Boltzmann-Gibbs probability measure:

$$d\mu_N(\mathbf{x}) = \frac{\exp\left[-\beta H_N(\mathbf{x}; a)\right]}{Z_N(\beta, a)} \prod_{i=1}^{N} dP(x_i)\,, \tag{1.1}$$

where the normalization

$$Z_N(\beta, a) = \int \prod_{i=1}^{N} dP(x_i) \exp\left[-\beta H_N(\mathbf{x}; a)\right] \tag{1.2}$$

is called *partition function* and β is the inverse absolute temperature. The (random) expectations w.r.t. (1.1) are denoted by $\langle \cdots \rangle_{\beta, a, N}$, and subscripts could be omitted to lighten the notation. With a slight abuse the same notation is adopted for the *replicated* Boltzmann-Gibbs measure $\mu_N^{\otimes \infty}$, used to average functions of different independent samples drawn from μ_N. However, we stress that the potential disorder introduced by the parameters a, as well as β, is the same for each replica of the system. These expectations can be further averaged over the disorder a if any and this is indicated with $\mathbb{E}$.

A quantity of particular interest will be the (random) *pressure*, or *free entropy*, per particle

$$p_N(\beta, a) = \frac{1}{N} \log Z_N(\beta, a) = \frac{P_N(\beta, a)}{N}\,. \tag{1.3}$$

From a probabilistic perspective, the extensive pressure Np_N is the moment generating function of the Hamiltonian in the measure (1.1). Furthermore, as we shall see in the following, derivatives

of p_N w.r.t. the parameters a typically yield expectations of macroscopic quantities that one can measure in a system, such as the magnetization in a ferromagnet, or the global overlap of an estimator with a signal we want to recover, needed to establish theoretical limits in its reconstruction.

1.2 The Sherrington-Kirkpatrick model

Consider the $N^2 + N$ random variables $J_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$, $h_i \overset{\text{iid}}{\sim} P_h$. The mean field spin glass SK model is a two body, inifinte-range interaction model defined by the hamiltonian:

$$H_N(\boldsymbol{\sigma}; J, h) = H_N^{SK}(\boldsymbol{\sigma}; J) - \sum_{i=1}^{N} h_i \sigma_i, \quad H_N^{SK}(\boldsymbol{\sigma}; J) = -\frac{1}{\sqrt{2N}} \sum_{i,j=1}^{N} J_{ij} \sigma_i \sigma_j, \qquad (1.4)$$

with $\sigma_i = \pm 1 \,\forall i$. The SK model is usually introduced with non random external magnetic fields but we will need them for our purposes and they do not introduce great complications. If one wants non-random external fields it is sufficient to choose $P_h = \delta_h$ for some real h. Our goal is the computation of the pressure

$$p_N(\beta, J, h) = \frac{1}{N} \log \sum_{\boldsymbol{\sigma} \in \{-1,1\}^N} \exp\left(-\beta H_N(\boldsymbol{\sigma}; J, h)\right). \qquad (1.5)$$

The previous is a random quantity that is known to concentrate in the thermodynamic limit $N \to \infty$. In particular it can be proved that it converges almost surely to the limit of its expectations sequence thanks to the exponential concentration inequality [18, 19, 20]

$$\mathbb{P}\left(|p_N(\beta, J, h) - \mathbb{E}p_N(\beta, J, h)| \geq x\right) \leq 2 \exp\left(-\frac{Nx^2}{2C\beta^2}\right), \qquad (1.6)$$

with some positive constant C. A tail integration also yields a bound on the variance

$$\mathbb{E}\left[(p_N(\beta, J, h) - \mathbb{E}p_N(\beta, J, h))^2\right] \leq \frac{\bar{C}}{N} \qquad (1.7)$$

for some positive constant $\bar{C}$. Since the r.h.s. of (1.6) is summable in N by Borel-Cantelli Lemma almost sure convergence is guaranteed. This means in particular that the specific realization of the noise is irrelevant when the system is really large. In some cases it will suffice to have the L^2-like convergence (1.7).

Therefore, from this moment on we will focus on the *quenched* free entropy per particle, namely

$$\bar{p}_N(\beta, h) = \mathbb{E}p_N(\beta, J, h) = \frac{\bar{P}_N(\beta, h)}{N} \qquad (1.8)$$

where the average is taken w.r.t. the disorder $J = \{J_{ij}\}_{i,j \leq N}$ and the external biases $h = \{h_i\}_{i \leq N}$. We stress that there is an abuse of notation: the averaged pressure on the l.h.s. depends only on the distribution of the h's. From this point of view, the SK model could have been defined in a totally equivalent way taking as Hamiltonian a 2^N-dimensional Gaussian process, labeled by spin configurations, with covariance

$$\mathbb{E}H_N^{SK}(\boldsymbol{\sigma})H_N^{SK}(\boldsymbol{\tau}) = \frac{N}{2}q_N^2(\boldsymbol{\sigma}, \boldsymbol{\tau}), \quad q_N(\boldsymbol{\sigma}, \boldsymbol{\tau}) := \frac{1}{N}\sum_{i=1}^{N}\sigma_i\tau_i \tag{1.9}$$

and then add back the one body terms $-\sum_i h_i\sigma_i$. q_N is called *overlap*, and we will soon realize it plays a central role both for the SK model and for the spiked Wigner model.

1.2.1 The replica trick

In order to compute the limiting free entropy (1.8) we use the following identity

$$\frac{1}{Nn}\log \mathbb{E}Z(\beta, J, h)^n \xrightarrow{n \to 0} \frac{1}{N}\mathbb{E}\log Z(\beta, J, h). \tag{1.10}$$

In principle this does not seem to help much, since n is only a real number going to 0. However, if it were a simple integer we would have a Gaussian expectation of the partition function, that we now how to compute since it amounts only to the computation of the moment generating of the Gaussian process H_N^{SK}. Hence, in what follows we will assume we are able to consider n as an integer up to the moment when we will need to send it back to 0.

This strategy, known as *replica trick*, turns out to be really versatile and can be applied to many different contexts. However, unfortunately at the present day it has not been proved to be mathematically rigorous. As we shall briefly overview later, a formal proof of the limiting free entropy conjectured with replicas requires the introduction of brand new rigorous tools. Let us now start from the replicated partition function

$$\mathbb{E}Z^n = \mathbb{E}\prod_{a=1}^{n}\sum_{\boldsymbol{\sigma}^a \in \{-1,1\}^N}\exp\left[-\sum_{a=1}^{n}H_N^{SK}(\boldsymbol{\sigma}^a; J) - \sum_{a=1}^{n}\sum_{i=1}^{N}h_i\sigma_i^a\right]. \tag{1.11}$$

As mentioned above we now perform a Gaussian integration over the Hamiltonian process, obtaining

$$\mathbb{E}Z^n = \mathbb{E}\prod_{a=1}^{n}\sum_{\boldsymbol{\sigma}^a \in \{-1,1\}^N}\exp\left[\frac{N\beta^2}{4}\sum_{a,b=1}^{n}q_N^2(\boldsymbol{\sigma}^a, \boldsymbol{\sigma}^b) + \beta\sum_{a=1}^{n}\sum_{i=1}^{N}h_i\sigma_i^a\right] \tag{1.12}$$

where the remaining $\mathbb{E}$ averages over the h_i's. Now the aim is to decouple the argument of the exponential in the particle indices i. To do this we perform a Hubbard-Stratonovič transforma-

tion on the q_N^2 term.

$$\mathbb{E}Z^n = \mathbb{E}\prod_{a=1}^{n}\sum_{\boldsymbol{\sigma}^a\in\{-1,1\}^N}\int\prod_{a\leq b=1}^{n}dq_{ab}\sqrt{\frac{\beta^2 N}{2\pi(1+\delta_{a,b})}}$$

$$\exp\left[-\frac{N\beta^2}{4}\sum_{a,b=1}^{n}q_{ab}^2 + \frac{N\beta^2}{2}\sum_{a,b=1}^{n}q_{ab}\,q_N(\boldsymbol{\sigma}^a,\boldsymbol{\sigma}^b) + \beta\sum_{a=1}^{n}\sum_{i=1}^{N}h_i\sigma_i^a\right] =$$

$$= \mathbb{E}\prod_{a=1}^{n}\sum_{\boldsymbol{\sigma}^a\in\{-1,1\}^N}\int\prod_{a\leq b=1}^{n}dq_{ab}\sqrt{\frac{\beta^2 N}{2\pi(1+\delta_{a,b})}} \tag{1.13}$$

$$\exp\left[-\frac{N\beta^2}{4}\sum_{a,b=1}^{n}q_{ab}^2 + \frac{\beta^2}{2}\sum_{i=1}^{N}\sum_{a,b=1}^{n}q_{ab}\,\sigma_i^a\sigma_i^b + \beta\sum_{i=1}^{N}\sum_{a=1}^{n}h_i\sigma_i^a\right].$$

Now that we have fully decoupled particles we can also factorize the spin sums over the particle indices. What remains is system that is still coupled only within replicas. Furthermore, when the system becomes really large only extremal values of the exponent will matter, hence we are allowed to use a saddle point approximation. These considerations lead to

$$\frac{1}{Nn}\log\mathbb{E}Z^n \xrightarrow{N\to\infty} \text{Extr}_{q_{a,b}}\left\{-\frac{\beta^2}{4n}\sum_{a,b}q_{a,b}^2 + \frac{1}{n}\log\mathbb{E}\prod_{a=1}^{n}\sum_{\sigma^a}\exp\left[\frac{\beta^2}{2}\sum_{a,b=1}^{n}q_{ab}\,\sigma^a\sigma^b + \beta h\sum_{a=1}^{n}\sigma^a\right]\right\} \tag{1.14}$$

where h is a copy of the h_i's. It is easy to realize that at saddle point $q_{aa}=1$.

From here, we can proceed in two ways. The simplest one is the so-called *replica symmetric ansatz*, namely assuming that at stationarity one has

$$q_{ab} = \begin{cases} 1 & a=b \\ q & a\neq q \end{cases}. \tag{1.15}$$

The extremization remains only over q. With this ansatz the variational expression in (1.14) simplifies considerably into

$$-\frac{\beta^2}{4}(1+(n-1)q^2) + \frac{1}{n}\log\mathbb{E}\prod_{a=1}^{n}\sum_{\sigma^a}\exp\left[\frac{\beta^2}{2}n(1-q) + \frac{\beta^2 q}{2}\left(\sum_{a=1}^{n}\sigma^a\right)^2 + \beta h\sum_{a=1}^{n}\sigma^a\right]$$

$$= \frac{\beta^2}{4}(1-2q+(n-1)q^2) + \frac{1}{n}\log\mathbb{E}\mathbb{E}_Z\left\{\sum_{\sigma=\pm1}\exp\left[\beta(Z\sqrt{q}+h)\sigma\right]\right\}^n \tag{1.16}$$

where Z is a standard Gaussian r.v. If we finally let $n\to0$ we recover the replica symmetric free entropy for the SK model

$$\frac{1}{Nn}\log\mathbb{E}Z^n \xrightarrow[n\to0]{N\to\infty} \text{Extr}_q\left\{\frac{\beta^2}{4}(1-q)^2 + \mathbb{E}\log 2\cosh[\beta(Z\sqrt{q}+h)]\right\} =: p_{RS}^{SK}(\beta). \tag{1.17}$$

A similar computation could have been carried out with different apriori measure on the spins. The only difference is that the diagonal elements of the overlap matrix q_{ab} are no longer one, but are themselves parameters over which extremization is needed. Extremization over q in (1.17) yields the fixed point equation

$$q = \mathbb{E}\tanh^2\left[\beta(Z\sqrt{q} + h)\right].\tag{1.18}$$

It is immediate to see that if $\beta < 1$ and $h \sim \delta_0$ the unique solution is $q = 0$. Indeed

$$q = \mathbb{E}\tanh^2\left[\beta(Z\sqrt{q})\right] \leq \mathbb{E}\beta^2 Z^2 q = \beta^2 q$$

which is satisfied only if $q = 0$. $q = 0$ is actually always a solution of (1.18) for $h = 0$. However, it can be shown that it becomes unstable for $\beta > 1$ and another stable solution $q > 0$ appears. In particular the solution with positive overlap is automatically selected first inserting an external magnetic field and then sending it to 0 (see [20, Proposition 1.3.8]). If $\mathbb{E}h^2 > 0$ instead (any non trivial random variable), the solution is always unique.

As anticipated this ansatz has a major issue in the low temperature limit, as stated in the following

Proposition 1.1. *Consider the quantity*

$$s_{RS}^{SK} := p_{RS}^{SK} - \beta\frac{\partial p_{RS}^{SK}}{\partial\beta},\tag{1.19}$$

with $h = 0$. The following limit holds

$$\lim_{\beta\to\infty} s_{RS}^{SK} = -\frac{1}{2\pi}.\tag{1.20}$$

Proof. Writing (1.19) explicitly after an integration by parts one gets

$$s^{RS} = -\frac{\beta^2}{4}(1 - \bar{q})^2 + \mathbb{E}_\zeta \log 2\cosh(\beta\sqrt{\bar{q}}Z) - \beta^2\bar{q}(1 - \bar{q}).\tag{1.21}$$

From (1.18) we see that $\lim_{\beta\to\infty} \bar{q} = 1$, where by $\bar{q}$ we denote the stable solution of the consistency equation here and below. Hence we have to carefully estimate the products of β and $1 - \bar{q}$:

$$\beta\sqrt{\bar{q}}(1 - \bar{q}) = \beta\sqrt{\bar{q}}\int_\mathbb{R}\frac{dz}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\left[1 - \tanh^2\left(\beta z\sqrt{\bar{q}}\right)\right] = \int_\mathbb{R}\frac{dz}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}\frac{d}{dz}\tanh\left(\beta z\sqrt{\bar{q}}\right)$$

$$= \int_\mathbb{R}\frac{dz}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}z\tanh\left(\beta z\sqrt{\bar{q}}\right) = 2\int_0^\infty\frac{dz}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}z\tanh\left(\beta z\sqrt{\bar{q}}\right) =$$

$$= 2\int_0^\infty\frac{dz}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}z + 2\int_0^\infty\frac{dz}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}z\left[\frac{1 - e^{-2\beta z\sqrt{\bar{q}}}}{1 + e^{-2\beta z\sqrt{\bar{q}}}} - 1\right] = (z\to z/\beta\sqrt{\bar{q}}) =$$

$$= \sqrt{\frac{2}{\pi}} - \frac{4}{\beta^2\bar{q}}\int_0^\infty\frac{dz}{\sqrt{2\pi}}e^{-\frac{z^2}{2\beta^2\bar{q}}}z\frac{e^{-2z}}{1 + e^{-2z}}.\tag{1.22}$$

The last integral is convergence and positive, and dominated by

$$\int_0^\infty \frac{dz}{\sqrt{2\pi}} z e^{-2z} = \frac{1}{4\sqrt{2\pi}} \ . \tag{1.23}$$

Hence, since $\bar{q} \to 1$, we can conclude that as $\beta \to \infty$

$$\beta\sqrt{\bar{q}}(1 - \bar{q}) = \sqrt{\frac{2}{\pi}} + O\left(\frac{1}{\beta^2}\right) , \tag{1.24}$$

$$\beta(1 - \bar{q}) = \sqrt{\frac{2}{\pi}} + O\left(\frac{1}{\beta^2}\right) . \tag{1.25}$$

Moreover,

$$\mathbb{E}_\zeta \log 2 \cosh(\beta\sqrt{\bar{q}}\zeta) = 2\int_0^\infty \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \beta\sqrt{\bar{q}} z + 2\int_0^\infty \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \log\left(1 + e^{-2\beta z\sqrt{\bar{q}}}\right) \tag{1.26}$$

where the last integral is of order $1/\beta$. To see it it suffices to perform the change of variables $z \mapsto z/\beta$. Therefore

$$\mathbb{E}_\zeta \log 2 \cosh(\beta\sqrt{\bar{q}}\zeta) = \beta\sqrt{\bar{q}}\sqrt{\frac{2}{\pi}} + O\left(\frac{1}{\beta}\right) . \tag{1.27}$$

Plugging the previous estimates into the formula for s^{RS} we finally get

$$s^{RS} = -\frac{1}{2\pi} + \beta\sqrt{\bar{q}}\left[\sqrt{\frac{2}{\pi}} + O\left(\frac{1}{\beta^2}\right) - \beta\sqrt{\bar{q}}(1 - \bar{q})\right] =$$

$$= -\frac{1}{2\pi} + \beta\sqrt{\bar{q}}\left[\sqrt{\frac{2}{\pi}} - \sqrt{\frac{2}{\pi}} + O\left(\frac{1}{\beta^2}\right)\right] = -\frac{1}{2\pi} + O\left(\frac{1}{\beta}\right) . \tag{1.28}$$

$$\square$$

If p_{RS}^{SK} were a well behaved solution, then s_{RS}^{SK} would be its entropy. Since in this thesis we will not need to make any replica symmetry breaking ansatz, we present it directly from a rigorous point of view in the following.

1.2.2 Interpolation and existence of the thermodynamic limit

The proof of existence of the thermodynamic limit for the SK model arrived only in 2002 [21] and needed the introduction of the so-called *interpolation technique*. The idea is to define an auxiliary Hamiltonian depending on an interpolating "time" parameter t that varies continuously, typically in $[0, 1]$. This allows to compare the models related to the Hamiltonians obtained

at the extremal times, 0 and 1, through a control on the t-derivatives of the interpolating model for intermediate times. This technique is really rich, easy to handle, and can be used both to prove existence of thermodynamic limits and to actually compute them via upper and lower bounds.

Theorem 1.2 (Guerra-Toninelli [21]). *The thermodynamic limit of the SK model exists and*

$$\lim_{N \to \infty} \bar{p}_N(\beta) = \sup_N \frac{\bar{P}_N}{N} < \infty \,. \tag{1.29}$$

Proof. First off, we prove that the pressure is upper bounded a constant. Thanks to Jensen's inequality, and to the fact that H_N^{SK} is Gaussian with variance $N/2$ we have:

$$\bar{p}_N(\beta, h) = \frac{1}{N} \mathbb{E}_{J,h} \log \sum_{\boldsymbol{\sigma} \in \{-1,1\}^N} e^{-\beta H_N(\boldsymbol{\sigma}; J, h)} \le \frac{1}{N} \mathbb{E}_h \log \sum_{\boldsymbol{\sigma} \in \{-1,1\}^N} e^{\frac{N\beta^2}{4} + \beta \sum_{i=1}^N h_i \sigma_i} \tag{1.30}$$

that leads to

$$\bar{p}_N \le \mathbb{E} \log 2 \cosh \beta h + \frac{\beta^2}{4} \,. \tag{1.31}$$

where h is a copy of the h_i's.

Secondly, we proceed with a super-additivity argument and then apply Fekete's lemma. The proof of super-addivity of the is carried out by interpolation. Let $t \in [0,1]$ and N_1, N_2 two integers such that $N_1 + N_2 = N$. Define now the interpolating scheme:

$$H_t(\boldsymbol{\sigma}; J, \tilde{J}, h) = \sqrt{t} H_N^{SK}(\boldsymbol{\sigma}; J) + \sqrt{1-t} \left[H_{N_1}^{SK}(\boldsymbol{\sigma}; J) + H_{N_2}^{SK}(\boldsymbol{\sigma}; J) \right] - \sum_{i=1}^N h_i \sigma_i \tag{1.32}$$

$$H_{N_1}^{SK}(\boldsymbol{\sigma}; \tilde{J}) = \frac{1}{\sqrt{2N_1}} \sum_{i,j=1}^{N_1} \tilde{J}_{ij} \sigma_i \sigma_j \,, \quad H_{N_2}^{SK}(\boldsymbol{\sigma}; \tilde{J}) = \frac{1}{\sqrt{2N_2}} \sum_{i,j=N_1+1}^{N} \tilde{J}_{ij} \sigma_i \sigma_j \tag{1.33}$$

$$\bar{P}_N(t) \equiv N \bar{p}_N(t) = \mathbb{E} \log Z_N(t) = \mathbb{E} \log \sum_{\boldsymbol{\sigma} \in \{-1,1\}^N} e^{-\beta H_t(\boldsymbol{\sigma}; J, \tilde{J}, h)} \,, \tag{1.34}$$

where $\tilde{J}_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ are independent of all the rest. This entails in particular:

$$\mathbb{E}[H_N(\sigma) H_{N_1}(\tau)] = \mathbb{E}[H_N(\sigma) H_{N_2}(\tau)] = [H_{N_1}(\sigma) H_{N_2}(\tau)] = 0 \,. \tag{1.35}$$

The first t-derivative yields

$$\dot{P}_N(t) = -\beta \mathbb{E} \Big\langle \frac{dH_t}{dt} \Big\rangle_{N,t} = -\beta \mathbb{E} \Big\langle \frac{1}{2\sqrt{t}} H_N^{SK} - \frac{1}{2\sqrt{1-t}} [H_{N_1}^{SK} + H_{N_2}^{SK}] \Big\rangle_{N,t} \,. \tag{1.36}$$

We stress that the Gibbs measure depends on t through the corresponding Hamiltonian H_t and we keep track of this dependency in the subscripts.

Thanks to (1.35) we can focus on the three terms appearing in the expectation separately. Using Gaussian integration by parts we obtain

$$-\beta \mathbb{E}\Big\langle \frac{1}{2\sqrt{t}} H_N \Big\rangle_{N,t} = N\frac{\beta^2}{4}\left[1 - \mathbb{E}\langle q_N^2 \rangle_{N,t}\right] \tag{1.37}$$

$$\beta \mathbb{E}\Big\langle \frac{1}{2\sqrt{1-t}} H_{N_1} \Big\rangle_{N,t} = -N_1 \frac{\beta^2}{4}[1 - \mathbb{E}\langle q_{N_1}^2 \rangle_{N,t}], \quad q_{N_1}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \frac{1}{N_1}\sum_{i,j=1}^{N_1} \sigma_i \tau_i \tag{1.38}$$

$$\beta \mathbb{E}\Big\langle \frac{1}{2\sqrt{1-t}} H_{N_2} \Big\rangle_{N,t} = -N_2 \frac{\beta^2}{4}[1 - \mathbb{E}\langle q_{N_2}^2 \rangle_{N,t}], \quad q_{N_2}(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \frac{1}{N_1}\sum_{i,j=N_1+1}^{N} \sigma_i \tau_i\,. \tag{1.39}$$

Now we can plug the previous contribution back into $\dot{P}_N(t)$:

$$\dot{P}_N(t) = N\frac{\beta^2}{4}\mathbb{E}\left[1 - \langle q_N^2 \rangle_{N,t} - \frac{N_1}{N} + \frac{N_1}{N}\langle q_{N_1}^2 \rangle_{N,t} - \frac{N_2}{N} + \frac{N_2}{N}\langle q_{N_2}^2 \rangle_{N,t}\right] =$$

$$= N\frac{\beta^2}{4}\mathbb{E}\left[\frac{N_1}{N}\langle q_{N_1}^2 \rangle_{N,t} + \frac{N_2}{N}\langle q_{N_2}^2 \rangle_{N,t} - \langle q_N^2 \rangle_{N,t}\right]\,. \tag{1.40}$$

Observe that

$$Nq_N(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \sum_{i=1}^{N} \sigma_i \tau_i = \sum_{i=1}^{N_1} \sigma_i \tau_i + \sum_{i=N_1+1}^{N} \sigma_i \tau_i = N_1 q_{N_1}(\boldsymbol{\sigma}, \boldsymbol{\tau}) + N_2 q_{N_2}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \tag{1.41}$$

$$q_N(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \frac{N_1}{N}q_{N_1}(\boldsymbol{\sigma}, \boldsymbol{\tau}) + \frac{N_2}{N}q_{N_2}(\boldsymbol{\sigma}, \boldsymbol{\tau}) \tag{1.42}$$

namely the total overlap is a convex combination of the overlaps of the two subsystems we are interpolating with. Since the square is a convex function one gets:

$$\frac{N_1}{N}q_{N_1}^2 + \frac{N_2}{N}q_{N_2}^2 - q_N^2 \geq 0 \tag{1.43}$$

and finally

$$\dot{P}_N(t) \geq 0 \quad \Rightarrow \quad \bar{P}_N(1) = \bar{P}_N \geq \bar{P}_N(0) = \bar{P}_{N_1} + \bar{P}_{N_2}\,, \tag{1.44}$$

which proves super-additivity. The rest of the statement follows directly from Fekete's lemma.
$\square$

As a corollary one has also the existence of the ground state.

Corollary 1.3 (Ground state energy, SK model). *The thermodynamic limit of the ground state energy of the SK model*

$$e_N := \inf_{\boldsymbol{\sigma} \in \{-1,1\}^N} H_N(\boldsymbol{\sigma}; J, h) \tag{1.45}$$

exists, converges almost surely to the limit of its expectations sequence $\bar{e}_N := \mathbb{E}e_N$, and

$$\lim_{N \to \infty} \bar{e}_N = \inf_N \bar{e}_N . \tag{1.46}$$

Proof. We start with the simple bounds:

$$\frac{1}{N\beta} \log \sum_{\sigma \in \Sigma_N} e^{-\beta H_N(\sigma)} \le \frac{1}{N\beta} \log 2^N e^{-N\beta e_N(J)} = \frac{\log 2}{\beta} - e_N \tag{1.47}$$

$$\frac{1}{N\beta} \log \sum_{\sigma \in \Sigma_N} e^{-\beta H_N(\sigma)} \ge \frac{1}{N\beta} \log e^{-N\beta e_N(J)} = -e_N . \tag{1.48}$$

The previous inequalities can be rewritten as

$$e_N - \frac{\log 2}{\beta} \le -\frac{1}{N\beta} \log \sum_{\sigma \in \Sigma_N} e^{-\beta H_N(\sigma)} \le e_N \tag{1.49}$$

that in turn entails

$$e_N = \lim_{\beta \to \infty} -\frac{1}{N\beta} \log \sum_{\sigma \in \Sigma_N} e^{-\beta H_N(\sigma)} . \tag{1.50}$$

Taking instead the expectation first in (1.49) we get the relation

$$\bar{e}_N = \mathbb{E}e_N(J) = \lim_{\beta \to \infty} -\frac{\bar{p}_N(\beta)}{\beta} = \frac{1}{N} \lim_{\beta \to \infty} -\frac{P_N(\beta)}{\beta} . \tag{1.51}$$

From the previous one we also see that both the sub-additivity of $\bar{E}_N = N\bar{e}_N$ sub-additivity, and the almost sure convergence of $\bar{e}_N$ are inherited from the pressure. $\square$

1.2.3   The replica symmetry breaking solution

Let us introduce two non-decreasing sequences:

$$0 = m_0 \le m_1 \le \cdots \le m_k \le m_{k+1} = 1 \tag{1.52}$$
$$0 = q_0 \le q_1 \le \cdots \le q_{k-1} \le q_k = 1 \tag{1.53}$$

where k is a positive integer. We can associate a discrete probability distribution χ (with at most k atoms, or $k + 1$ if we include 0) to the triple $(k, \mathbf{m}, \mathbf{q})$, as follows:

$$\chi([0, q]) = \sum_{l=0}^{k} (m_{l+1} - m_l)\theta(q - q_l) \tag{1.54}$$

where the Heaviside step function is taken to be continuous from the right. The space of atomic probability distributions with at most k atoms on $[0, 1]$ is denoted by $\mathcal{M}_{[0,1]}^k$. Consider now the recursion

$$Z_{l-1}^{m_l} = \mathbb{E}_l \left(Z_l^{m_l} \right) \quad \mathbb{E}_l[\cdot] = \int_{\mathbb{R}} \frac{d\eta_l}{\sqrt{2\pi}} e^{-\eta_l^2/2}(\cdot) \tag{1.55}$$

$$Z_k = \cosh \left[\beta \left(h + \sum_{l=1}^{k} \eta_l \sqrt{q - q_{l-1}} \right) \right] \quad \eta_l \overset{\text{iid}}{\sim} \mathcal{N}(0, 1). \tag{1.56}$$

The Parisi functional is defined as follows.

Definition 1.1 (Parisi Functional-1). Given the triple $(k, \mathbf{m}, \mathbf{q})$ as above, the Parisi functional is:

$$\mathcal{P}(\chi; \beta) = \log 2 + \mathbb{E} \log Z_0 - \frac{\beta^2}{2} \int_0^1 q\chi([0, q]) \, dq. \tag{1.57}$$

where the expectation of $\log Z_0$ averages over the possibly random variable h, that is a copy of the h_i's.

There is an equivalent way to express the nested term $\log Z_0$ by means of an anti-parabolic PDE. To each $\chi \in \mathcal{M}_{[0,1]}^k$ we associate a function $\Phi_\chi(\cdot, \cdot; \beta)$ that is the solution to the final value problem

$$\partial_s \Phi_\chi(s, y; \beta) = -\frac{\beta^2}{2} \left(\partial_y^2 \Phi_\chi(s, y; \beta) + \chi([0, s])(\partial_y \Phi_\chi(s, y; \beta))^2 \right)$$
$$\Phi_\chi(1, y; \beta) = \log \cosh y. \tag{1.58}$$

It is then not difficult to see that [22]

$$\mathbb{E} \log Z_0 = \mathbb{E}_h \Phi_\chi(0, h; \beta) =: \bar{\Phi}_\chi(0; \beta). \tag{1.59}$$

Therefore, for future convenience we give the equivalent

Definition 1.2 (Parisi Functional-2). For any $\chi \in \mathcal{M}_{[0,1]}^k$ as above, the Parisi functional is

$$\chi \in \mathcal{M}_{[0,1]}^k \longmapsto \mathcal{P}(\chi; \beta) = \log 2 + \bar{\Phi}_\chi(0; \beta) - \frac{\beta^2}{2} \int_0^1 dq \, q\chi([0, q]). \tag{1.60}$$

It is well known [7, 23] that for any $\chi_1, \chi_2 \in \mathcal{M}^k_{[0,1]}$

$$|\Phi_{\chi_1}(s, y; \beta) - \Phi_{\chi_2}(s, y; \beta)| \leq \frac{\beta^2}{2} \int_s^1 dq\, |\chi_1([0, q]) - \chi_2([0, q])| \tag{1.61}$$

namely $\chi \mapsto \Phi_\chi$ is Lipschitz in the $L^1([s, 1], dq)$ norm. This allows us to extend the functional Φ_χ to all the probability measures $\mathcal{M}_{[0,1]}$ with the prescription

$$\Phi_\chi := \lim_{n \to \infty} \Phi_{\chi_n} \tag{1.62}$$

for any sequence $(\chi_n)_{n \geq 1}$ in $\mathcal{M}^d_{[0,1]}$ such that $\chi_n \longrightarrow \chi \in \mathcal{M}_{[0,1]}$ weakly. We hereby collect the continuity and differentiability properties of $\bar{\Phi}_\chi$ and $\mathcal{P}(\chi; \cdot)$. Without loss of generality we can assume $h_i = h\xi_i$ with $h \in \mathbb{R}$ and $\xi_i \stackrel{iid}{\sim} P_\xi$, since it amounts only to a re-scaling of P_h. We thus consider $\mathcal{P}$ as a function of $(\chi; \beta, h)$ in the following proposition.

Proposition 1.4. *Let $a := \mathbb{E}\xi_1^2$. The following hold:*

i) $\bar{\Phi}_\chi$ (and $\mathcal{P}(\chi; \cdot, \cdot)$) can be continuously extended to $\mathcal{M}_{[0,1]}$ w.r.t. the weak convergence and

$$\bar{\Phi}_\chi(s, h; \beta) := \lim_{n \to \infty} \mathbb{E}\bar{\Phi}_{\chi_n}(s, h\xi; \beta) = \mathbb{E}\Phi_\chi(s, h\xi; \beta) \tag{1.63}$$

for any sequence $(\chi_n)_{n \geq 1}$ in $\mathcal{M}^d_{[0,1]}$ such that $\chi_n \longrightarrow \chi \in \mathcal{M}_{[0,1]}$ weakly.

ii) $\chi \mapsto \bar{\Phi}_\chi$ is convex in $\mathcal{M}_{[0,1]}$.

iii) $\bar{\Phi}_\chi$ is twice h-differentible for any $\chi \in \mathcal{M}_{[0,1]}$ and

$$|\partial_h \bar{\Phi}_\chi(s, h; \beta)| \leq \sqrt{a}, \quad 0 < \partial_h^2 \bar{\Phi}_\chi(s, h; \beta) \leq a. \tag{1.64}$$

In particular it is convex in h.

iv) Consider a sequence $(\chi_n)_{n \geq 1}$ in $\mathcal{M}_{[0,1]}$ such that $\chi_n \longrightarrow \chi \in \mathcal{M}_{[0,1]}$ weakly. Then

$$\partial_h \bar{\Phi}_{\chi_n} \longrightarrow \partial_h \bar{\Phi}_\chi. \tag{1.65}$$

v) The function $\mathcal{P}(\beta, h) = \inf_{\chi \in \mathcal{M}_{[0,1]}} \mathcal{P}(\chi; \beta, h)$ is h-differentiable at any $h \in \mathbb{R}$ and

$$\partial_h \mathcal{P}(\beta, h) = \partial_h \mathcal{P}(\chi^*(\beta, h); \beta, h) = \partial_h \bar{\Phi}_{\chi^*(\beta, h)}(0, h; \beta) \tag{1.66}$$

where $\chi^*(\beta, h)$ is the unique distribution at which the infimum is attained and only the explicit dependence on h is taken into account.

Proof. i). Consider $\chi_1, \chi_2 \in \mathcal{M}_{[0,1]}^k$. By (1.61)

$$|\bar{\Phi}_{\chi_1}(s, h; \beta) - \bar{\Phi}_{\chi_2}(s, h; \beta)| \leq \frac{\beta^2}{2} \int_s^1 dq \, |\chi_1([0, q]) - \chi_2([0, q])| \tag{1.67}$$

namely $\chi \mapsto \bar{\Phi}_\chi$ is Lipschitz too on $\mathcal{M}_{[0,1]}^k$. Therefore we perform a continuous extension to $\mathcal{M}_{[0,1]}$ obtaining a continuous functional with respect to the weak convergence. Furthermore, given a sequence $(\chi_n)_{n \geq 1}$ converging to $\chi \in \mathcal{M}_{[0,1]}$ weakly we have

$$|\bar{\Phi}_{\chi_n}(s, h; \beta) - \mathbb{E}\Phi_\chi(s, h\xi; \beta)| \leq \frac{\beta^2}{2} \int_s^1 dq \, |\chi_n([0, q]) - \chi([0, q])| \longrightarrow 0 \tag{1.68}$$

by dominated convergence.

ii). The thesis immediately follows from *i)* and the main result in [24] that asserts the convexity of Φ_χ.

iii). By Proposition 2 in [24] the first two y-derivatives of Φ_χ exist and are continuous, with $|\partial_y \Phi_\chi(s, y; \beta)| \leq 1$, $C/\cosh^2 y \leq \partial_y^2 \Phi_\chi(s, y; \beta) \leq 1$ for some $C > 0$. Then, using Lagrange's mean value theorem and dominated convergence one can show that

$$\partial_h \bar{\Phi}_\chi(s, h; \beta) = \mathbb{E}\left[\xi \partial_y \Phi_\chi(s, h\xi; \beta)\right], \quad \partial_h^2 \bar{\Phi}_\chi(s, h; \beta) = \mathbb{E}\left[\xi^2 \partial_y^2 \Phi_\chi(s, h\xi; \beta)\right] \tag{1.69}$$

which implies (1.64) and in turn the convexity of $\bar{\Phi}_\chi$ in h.

iv). Since $\bar{\Phi}_\eta$ is convex in h for any $\eta \in \mathcal{M}_{[0,1]}$, $\bar{\Phi}_{\chi_n}$ is a sequence of convex functions. Therefore, thanks to points and *i), ii)* and *iii)*

$$\lim_{n \to \infty} \partial_h \bar{\Phi}_{\chi_n} = \partial_h(\lim_{n \to \infty} \bar{\Phi}_{\chi_n}) = \partial_h \bar{\Phi}_\chi. \tag{1.70}$$

v). $\mathcal{P}(\beta, h)$ is convex in h because it is the limit of a sequence of convex functions. Hence it is sufficient to prove that at any $h \in \mathbb{R}$ the sub-differential is single valued (as done for instance in [25]). For any fixed $\delta > 0$ and b in the sub-differential the following holds

$$\frac{\mathcal{P}(\beta, h) - \mathcal{P}(\beta, h - \delta)}{\delta} \leq b \leq \frac{\mathcal{P}(\beta, h + \delta) - \mathcal{P}(\beta, h)}{\delta}. \tag{1.71}$$

Now, thanks to point *i)* and *ii)*, $\mathcal{P}(\chi; \beta, h)$ is also χ-convex, thus it has a unique minimizer χ^*, and it is continuous w.r.t. the weak convergence. Hence we can find a sequence of measures such that $\chi_n \longrightarrow \chi^*$ weakly and

$$\mathcal{P}(\chi_n; \beta, h) \leq \mathcal{P}(\chi^*; \beta, h) + \frac{1}{n} = \mathcal{P}(\beta, h) + \frac{1}{n} \tag{1.72}$$

whilst it is obvious that $\mathcal{P}(\chi_n; \beta, h) \geq \mathcal{P}(\beta, h)$. Inserting these inequalities in (1.71) produces

$$-\frac{1}{n\delta} + \frac{\mathcal{P}(\chi_n; \beta, h) - \mathcal{P}(\chi_n; \beta, h-\delta)}{\delta} \leq b \leq \frac{1}{n\delta} + \frac{\mathcal{P}(\chi_n; \beta, h+\delta) - \mathcal{P}(\chi_n; \beta, h)}{\delta} . \qquad (1.73)$$

Notice that $\partial_h^{1,2} \mathcal{P}(\chi_n; \beta, h) = \partial_h^{1,2} \bar{\Phi}_{\chi_n}(0, h; \beta)$ hence we can expand the Parisi functional up to the second order obtaining

$$-\frac{1}{n\delta} + \partial_h \mathcal{P}(\chi_n; \beta, h) - \frac{a\delta}{2} \leq b \leq \frac{1}{n\delta} + \partial_h \mathcal{P}(\chi_n; \beta, h) + \frac{a\delta}{2} , \qquad (1.74)$$

where we have used (1.64). Choose now $\delta = n^{-1/2}$ and then send $n \to \infty$. Finally, applying point *iv)* we conclude that the unique possible value for b is $\partial_h \bar{\Phi}_{\chi^*}(0, h; \beta)$. $\qquad \square$

We are now ready to state the theorem containing the thermodynamic limit of the SK model. The proof of it lies outside of the scopes of the thesis. However, it is worth mentioning that the upper bound was obtained by Guerra in [7] by means of interpolation. The main idea is to interpolate between the SK model and a one-body (Gaussian) interaction model, that is completely decoupled, and thus directly integrable. The difficulty here is to give this model the right correlation structure in order to reproduce the complicated nested term $\log Z_0$ or Φ_χ. One way of doing it, is by means of Ruelle probability cascades [18], that thanks to their invariance properties naturally reproduce $\log Z_0$.

Theorem 1.5 (Guerra [7], Talagrand [8], [26])**.** *The limiting pressure of the SK model fulfills the functional variational principle*

$$\bar{p}_N(\beta, h) \xrightarrow{N \to \infty} \inf_{\chi \in \mathcal{M}_{[0,1]}} \mathcal{P}(\chi; \beta, h) \qquad (1.75)$$

where the infimum point $\chi^*(\beta, h)$ is unique.

1.2.4   The replica symmetric bound and the Almeida-Thouless line

In Section 1.2.1 we saw how the replica symmetric ansatz looks like, and we showed it cannot be the correct solution in the low temperature region. However, one can still wonder if it is correct for high enough temperatures (low β). In particular, if we consider $h_i = h\xi_i$ with $\xi_i \overset{\text{iid}}{\sim} P_\xi$ without loss of generality, we expect a transition line in the phase plane β, h. This line is thought to be the celebrated Almeida-Thouless line [27]:

$$\beta^2 \mathbb{E} \cosh^{-4} \left[\beta \left(Z\sqrt{\bar{q}} + h\xi \right) \right] = 1 \qquad (1.76)$$

with $\bar{q}$ solving (1.18). For too large β the l.h.s. exceeds 1 and one should use the RSB solution for the SK model. On the contrary, it is conjectured that when the l.h.s. is below 1 the replica symmetric ansatz (1.17) is exact. From a rigorous viewpoint we can prove much less. To begin with we prove that the replica symmetric pressure bounds the real one from above:

Theorem 1.6 (Replica symmetric bound)**.** *The pressure of the SK model is uniformly bounded from above by the replica symmetric pressure:*

$$\bar{p}_N(\beta, h) \leq p_{RS}^{SK}(\beta, h; q), \quad \forall\, q \in [0, 1] \tag{1.77}$$

$$p_{RS}^{SK}(\beta, h; q) = \frac{\beta^2}{4}(1-q)^2 + \mathbb{E} \log 2 \cosh\left[\beta\left(h\xi + \sqrt{q}z\right)\right]. \tag{1.78}$$

Proof. Consider the one-body Hamiltonian

$$\tilde{H}_N(\boldsymbol{\sigma}; \tilde{J}, q) = -\sqrt{q} \sum_{i=1}^{N} \tilde{J}_i \sigma_i \quad \tilde{J}_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \tag{1.79}$$

$$\tilde{p}_N = \frac{1}{N} \mathbb{E} \log \sum_{\boldsymbol{\sigma} \in \{-1,1\}^N} e^{-\beta \tilde{H}_N(\sigma) + \beta h \sum_{i=1}^{N} \xi_i \sigma_i} = \mathbb{E} \log 2 \cosh\left[\beta\left(h\xi + z\sqrt{q}\right)\right] \tag{1.80}$$

Define then the interpolating hamiltonian:

$$H_t(\boldsymbol{\sigma}; J, \tilde{J}, h) = \sqrt{t} H_N^{SK}(\boldsymbol{\sigma}; J) + \sqrt{1-t}\, \tilde{H}_N(\boldsymbol{\sigma}; \tilde{J}, q) - h \sum_{i=1}^{N} \xi_i \sigma_i \tag{1.81}$$

where J and $\tilde{J}$ are independent disorders. $\tilde{H}$ can be seen again as a Gaussian family with covariance

$$\mathbb{E}[\tilde{H}_N(\boldsymbol{\sigma}; \tilde{J}, q)\tilde{H}_N(\boldsymbol{\tau}; \tilde{J}, q)] = N q q_N(\boldsymbol{\sigma}, \boldsymbol{\tau}) \tag{1.82}$$

In order to obtain a bound on the pressure we aim to control the first derivative of the interpolating pressure:

$$p_N(t) := \frac{1}{N} \mathbb{E} \log \sum_{\boldsymbol{\sigma} \in \{-1,1\}^N} e^{-\beta H_N(\boldsymbol{\sigma}; J, \tilde{J}, h)}. \tag{1.83}$$

A direct computation using integration by parts yields

$$\dot{p}_N(t) = -\frac{\beta}{N} \mathbb{E} \left\langle \frac{1}{2\sqrt{t}} H_N^{SK}(\boldsymbol{\sigma}; J, h) - \frac{1}{2\sqrt{1-t}} \tilde{H}_N(\boldsymbol{\sigma}; \tilde{J}, q) \right\rangle_{N,t} =$$

$$= \frac{\beta^2}{4} \mathbb{E} \left\langle 1 - q_N^2(\boldsymbol{\sigma}, \boldsymbol{\tau}) - 2q + 2q q_N(\boldsymbol{\sigma}, \boldsymbol{\tau}) \right\rangle_{N,t} =$$

$$= -\frac{\beta^2}{4} \mathbb{E} \left\langle (q_N(\boldsymbol{\sigma}, \boldsymbol{\tau}) - q)^2 \right\rangle_{N,t} + \frac{\beta^2}{4}(1-q)^2. \tag{1.84}$$

Hence, by the theorem of integral calculus:

$$p_N^{SK} = \tilde{p}_N + \frac{\beta^2}{2}(1-q)^2 - \frac{\beta^2}{4} \int_0^1 dt\, \mathbb{E} \left\langle (q_N(\boldsymbol{\sigma}, \boldsymbol{\tau}) - q)^2 \right\rangle_{N,t} \leq \tilde{p}_N + \frac{\beta^2}{2}(1-q)^2. \tag{1.85}$$

After plugging $\tilde{p}_N$ in the previous inequality it is clear that the r.h.s. is exactly p_{RS}. $\qquad \square$

The only rather general result on the Almeida-Thouless line is the following negative state-
ment [28].

Theorem 1.7. *If*

$$\beta^2 \mathbb{E} \cosh^{-4} \left[\beta \left(Z \sqrt{\bar{q}} + h\xi \right) \right] > 1 \tag{1.86}$$

then the pressure of the SK model is strictly smaller than the replica symmetric pressure, namely:

$$\lim_{N \to \infty} \bar{p}_N(\beta, h) < \inf_{q \in [0,1]} p_{RS}^{SK}(\beta, h; q) \tag{1.87}$$

Proof. From Guerra's replica symmetry breaking bound [7] we know that:

$$\bar{p}_N(\beta, h) \leq \inf_{\chi \in \mathcal{M}_{[0,1]}} \mathcal{P}(\chi; \beta, h) \tag{1.88}$$

We just need to provide a measure $\tilde{\chi}$ with more than one atom, hence different from a replica
symmetric one, such that $\mathcal{P}(\tilde{\chi}; \beta, h) < p_{RS}^{SK}(\beta, h; \bar{q})$ when (1.86) is fulfilled. Take for instance:

$$\chi([0, q]) = \begin{cases} 0 & \text{if } q \in [0, \bar{q}] \\ m & \text{if } q \in [\bar{q}, r] \\ 1 & \text{if } q \in [r, 1] \end{cases} \tag{1.89}$$

where $m \in [0, 1]$ and $r \in [\bar{q}, 1]$. With these sequences the Parisi functional becomes the 1-step
RSB functional

$$\mathcal{P}(\beta, h; m, r) = \frac{\beta^2}{4} \left(1 + m\bar{q}^2 + (1 - m)r^2 - 2r \right) + \log 2 +$$
$$\frac{1}{m} \mathbb{E}_\xi \mathbb{E}_{Z'} \left[\log \mathbb{E}_Z \cosh^m \beta \left(\sqrt{(r - \bar{q})} Z + \sqrt{\bar{q}} Z' + h\xi \right) \right] \tag{1.90}$$

When we take $m = 1$, and $r = \bar{q}$ and we go back to replica symmetric pressure, namely:
$\mathcal{P}(\beta, h; 1, \bar{q}) = p_{RS}^{SK}(\beta, h)$, because we are at the zero-th step of the replica symmetry breaking.
If $\mathcal{P}(\beta, h; m, r)$ touches $m = 1$ with positive m-derivative for some r the proof is finished.

$$K(\beta, h; r) := \left. \frac{\partial \mathcal{P}(\beta, h; m, r)}{\partial m} \right|_{m=1} = -\frac{\beta^2}{4} (r^2 - \bar{q}^2) -$$
$$- \mathbb{E}_\xi \mathbb{E}_{Z'} \left[\log \mathbb{E}_Z \cosh^m \beta \left(\sqrt{(r - \bar{q})} Z + \sqrt{\bar{q}} Z' + h\xi \right) \right] +$$
$$+ \mathbb{E}_\xi \mathbb{E}_{Z'} \left[\frac{\mathbb{E}_Z \cosh \beta \left(\sqrt{(r - \bar{q})} Z + \sqrt{\bar{q}} Z' + h\xi \right) \log \cosh \beta \left(\sqrt{(r - \bar{q})} Z + \sqrt{\bar{q}} Z' + h\xi \right)}{\mathbb{E}_Z \cosh \beta \left(\sqrt{(r - \bar{q})} Z + \sqrt{\bar{q}} Z' + h\xi \right)} \right] \tag{1.91}$$

In order to establish its sign in a neighborhood of $r = \bar{q}$ we perform a Taylor expansion that yields:

$$K(\beta, h; \bar{q}) = 0 \tag{1.92}$$

$$\left.\frac{\partial K(\beta, h; r)}{\partial r}\right|_{r=\bar{q}} = 0 \tag{1.93}$$

$$\left.\frac{\partial^2 K(\beta, h; r)}{\partial r^2}\right|_{r=\bar{q}} = -\frac{\beta^2}{2}\left(1 - \beta^2 \mathbb{E}_Z \cosh^{-4}\beta\left(\sqrt{\bar{q}}z + h\xi\right)\right) \tag{1.94}$$

The latter is positive for r in a neighborhood of $\bar{q}$ by (1.86). This concludes the proof. $\square$

Nevertheless, if $\xi_i \overset{iid}{\sim} \mathcal{N}(0, a)$ the RSB region can be fully characterized in terms of (1.76). Only the statement of the following theorem will be useful in our analysis, so we omit the proof. The interested reader can directly look at the original reference.

Theorem 1.8 (Chen [22]). *Consider the SK model with centered Gaussian external field. For any $\beta > 0$ and $h > 0$, the Parisi formula exhibits the replica symmetric solution if and only if (β, h) lies inside the AT line, i.e.,*

$$\beta^2 \mathbb{E} \cosh^{-4}\left[\beta\left(Z\sqrt{\bar{q}} + h\xi\right)\right] \leq 1. \tag{1.95}$$

As a concluding remark, we stress that Replica Symmetry and its breaking are connected to the concentration of the overlap, which is the key order parameter in the SK model. A first signal of this connection can be found in (1.85). In fact, following the first equality, if one could prove that the quadratic deviation $\mathbb{E}\left\langle(q_N(\boldsymbol{\sigma}, \boldsymbol{\tau}) - q)^2\right\rangle_{N,t} \xrightarrow{N\to\infty} 0$ with some t-uniformity condition, then the RS pressure would be exact. For the SK model this is not the case, as it can be seen from Theorem 1.7. In the SK model the overlap fluctuates even in the thermodynamic limit and, at least for our purposes[1], $\chi^*(\beta, h)$ might be interpreted as its asymptotic distribution. However, writing a so-called *sum rule* like the one in (1.85) for other models, such as Curie-Weiss, can be a winning strategy.

1.3 High dimensional Inference and Statistical Mechanics

In this section we start by highlighting the correspondence between the Statistical Mechanics formalism already outlined above, and the Information theoretic one. The goal of Inference is generally to retrieve some signal, called *ground truth*, and usually denoted by $\mathbf{x}^*$ from a set of noisy observations that depend on it, denoted by $\mathbf{y}$. The typical situation also tackled in

---

[1]It can be shown that the first moments of the overlap are asymtptotically computable through χ^*. See [18] Theorem 3.7 and [25] for precise statements.

undergraduate courses is that of low dimensional inference, where the parameters to estimate (say the components of $\mathbf{x}^*$) are finite in number, whereas one assumes to have a growing number of observations at disposal. An example is trying to estimate mean and variance of a Gaussian given N samples from it. In this case, it is well known that the max-likelihood approach gives a good answer to the problem. More precisely, it allows the observer to estimate the lacking parameters exactly when the number of observations becomes really large, ideally infinity. Notice that this kind of approach is Bayesian: one needs to know the probability of the observations *given* the parameters sought.

In this thesis however, we deal with high dimensional inference problems, where both the number of observations and *ground truth* signal components grow. Typically, if the size of the second grows with N, the first grow with N^k for some integer k. This makes the problem much more complicated and it is known that the max-likelihood approach is no longer the optimal one [29]. We stress that introducing such scalings in the inference task could also imply the presence of phase transitions in some of the control parameters of the problem.

1.3.1 Basic definitions and Bayes Optimal Setting (BOS)

The goal of the inference tasks treated hereby is the reconstruction of non-negligible fraction of components of a ground truth signal $\mathbf{x}^*$ drawn from some *prior* distribution P_X. In some cases we will assume factorization over the components: $X_i^* \overset{\text{iid}}{\sim} P_X$. We denote with a $*$ the ground truth, namely those signal vectors that are directly drawn from the prior. The observations can be modeled as a random function, the randomness being in some kind of noise $\mathbf{Z}$, of the ground truth: $\mathbf{Y} = \mathbf{F_Z}(\mathbf{X}^*)$. This, from a probabilistic perspective, translates into a *likelihood* distribution density for the observations given an instance $\mathbf{x}$ sampled from P_X: $p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}$. We stress that the form of this likelihood is strongly affected by the nature of the noise, since for a fixed realization $\mathbf{x}$ the randomness in $\mathbf{Y}$ is solely inherited by the noise.

For our purposes we can assume to be able to write the posterior measure of the process that plays a central role:

$$dP_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x}) = \frac{p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) \, dP_X(\mathbf{x})}{Z(\mathbf{y})} \,, \quad Z(\mathbf{y}) = \int p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y}) \, dP_X(\mathbf{x}) \,. \tag{1.96}$$

$Z(\mathbf{y})$ is a normalization probability usually called *evidence*. It is important to mention that inference is performed by Statisticians that can have different features, Bayes-optimality being one of them:

Definition 1.3 (Bayes-optimality)**.** A Statistician is said to be Bayes-optimal, or in the Bayes optimal setting (BOS), if they know everything of the data generating process, namely P_X and $\mathbf{F_Z}(\cdot)$, or equivalently they know the correct posterior measure in (1.96).

As intuition suggests, a Bayes-optimal Statistician is able to reconstruct the signal in the best possible way in a suitable error metric: the L^2 norm.

Proposition 1.9. *Define the mean square error as the* a posteriori *expected quadratic deviation of an estimator* $\hat{\mathbf{x}}(\mathbf{Y})$:

$$MSE(\mathbf{y}) = \int \|\mathbf{x} - \hat{\mathbf{x}}(\mathbf{y})\|^2 \, dP_X(\mathbf{x}) \,. \tag{1.97}$$

The MSE for a given set of observations $\mathbf{y}$ *is minimized by the estimator*

$$\hat{x}(\mathbf{y}) = \int \mathbf{x} \, dP_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x}) =: \langle \mathbf{X} \rangle_{\mathbf{y}} \,. \tag{1.98}$$

Proof. The MSE is convex in $\hat{\mathbf{x}}(\mathbf{y})$. It is thus sufficient to impose that the gradient w.r.t. vanishes. $\qquad\square$

The previous proposition also motivates our interest in the posterior distribution (1.96). One may decide to average also the possible $\mathbf{Y}$-outcomes to have a general average expected error. This quantity is called *minimum* (expected) *mean square error*:

$$\mathrm{MMSE} := \mathbb{E}_{\mathbf{Y}} \left\langle \|\mathbf{X} - \langle \mathbf{X} \rangle_{\mathbf{Y}}\|^2 \right\rangle_{\mathbf{Y}} = \mathbb{E}_{\mathbf{Y}} \mathbb{E}_{\mathbf{X}|\mathbf{Y}} \|\mathbf{X} - \mathbb{E}_{\mathbf{X}|\mathbf{Y}} \mathbf{X}\|^2 \,. \tag{1.99}$$

One can rewrite (1.99) expanding the square and using the total probability rule of conditional expectations

$$\mathrm{MMSE} = \mathbb{E}\|\mathbf{X}\|^2 - 2\mathbb{E}_{\mathbf{Y}} \mathbb{E}_{\mathbf{X}|\mathbf{Y}}[\mathbf{X} \cdot \mathbb{E}_{\mathbf{X}|\mathbf{Y}} \mathbf{X}] + \mathbb{E}_{\mathbf{Y}} \|\mathbb{E}_{\mathbf{X}|\mathbf{Y}} \mathbf{X}\|^2 = \mathbb{E}\|\mathbf{X}\|^2 - \mathbb{E}_{\mathbf{Y}} \mathbb{E}_{\mathbf{X}|\mathbf{Y}}[\mathbf{X} \cdot \mathbb{E}_{\mathbf{X}|\mathbf{Y}} \mathbf{X}] =$$
$$= \mathbb{E}\|\mathbf{X}\|^2 - \mathbb{E}_{\mathbf{X}^*,\mathbf{Y}}[\mathbf{X}^* \cdot \mathbb{E}_{\mathbf{X}|\mathbf{Y}} \mathbf{X}] = \mathbb{E}_{\mathbf{X}^*,\mathbf{Y}} \|\mathbf{X}^* - \langle \mathbf{X} \rangle_{\mathbf{Y}}\|^2 \,. \tag{1.100}$$

Therefore, we see that the MMSE is really a measure of the quadratic deviation from the ground truth, and hence a good measure of error. (1.100) is first example of the application of a Nishimori identity: we have replaced a sample $\mathbf{X}$ from the posterior with a sample drawn directly from the ground truth. This is actually a rather general fact stated in the following

Proposition 1.10 (Nishimori identities)**.** *Let f be a bounded function of the observations $\mathbf{Y}$, the ground truth $\mathbf{X}^*$ and $n-1$ replicas $(\mathbf{X}^{(k)})_{k=2}^n$ drawn independently from the posterior (1.96). Then*

$$\mathbb{E}_{\mathbf{X}^*,\mathbf{Y}} \langle f(\mathbf{Y}; \mathbf{X}^*, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(n)}) \rangle_{\mathbf{Y}} = \mathbb{E}_{\mathbf{Y}} \langle f(\mathbf{Y}; \mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(n)}) \rangle_{\mathbf{Y}} \tag{1.101}$$

where now $\mathbf{X}^{(1)} \sim P_{\mathbf{X}|\mathbf{Y}}$.

Proof. We omit the function in the proof and focus only on the chain of expectations on the r.h.s. of (1.101):

$$\mathbb{E}_{\mathbf{Y}} \prod_{k=1}^n \mathbb{E}_{\mathbf{X}^{(k)}|\mathbf{Y}}(\cdots) = \mathbb{E}_{\mathbf{Y}} \mathbb{E}_{\mathbf{X}^{(1)}|\mathbf{Y}} \prod_{k=2}^n \mathbb{E}_{\mathbf{X}^{(k)}|\mathbf{Y}}(\cdots) = \mathbb{E}_{\mathbf{X}^*,\mathbf{Y}} \prod_{k=2}^n \mathbb{E}_{\mathbf{X}^{(k)}|\mathbf{Y}}(\cdots) \,. \tag{1.102}$$

The previous completes the proof. $\qquad\square$

It is important to stress here that the validity of these identities is implied by Bayes-optimality. A sub-optimal Statistician (see Chapter 4) cannot access the Nishimori identities for reasons that will be clarified later. Despite being an almost immediate consequence of the total probability rule, and hence of the optimal setting, the Nishimori identities are strongly related to replica symmetry.

The reader will have noticed at this point some similarities with the Statistical Mechanics formalism. In particular, $Z(\mathbf{y})$ in (1.96) can be interpreted as a partition function corresponding to a Hamiltonian $-\log p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{y})$ and unit inverse absolute temperature. The associated quenched free entropy per signal component, with a total of N components, is then

$$\bar{p}_N = \frac{1}{N}\mathbb{E}_{\mathbf{Y}}\log Z(\mathbf{Y}) = \frac{1}{N}\mathbb{E}_{\mathbf{Y}}\log \int p_{\mathbf{Y}|\mathbf{X}=\mathbf{x}}(\mathbf{Y})\,dP_X(\mathbf{x}) = -\frac{1}{N}\mathcal{H}_N(\mathbf{Y})\,, \tag{1.103}$$

namely the Shannon entropy of the observations up to a sign. The latter can depend also on some other parameters of the problem that we neglected for the moment to lighten the notation. The main information theoretic quantity of interest is the *mutual information* per signal component between the ground truth and the observations:

$$\frac{1}{N}I_N(\mathbf{X}^*,\mathbf{Y}) = \frac{1}{N}\mathcal{H}_N(\mathbf{Y}) - \frac{1}{N}\mathcal{H}_N(\mathbf{Y}|\mathbf{X}^*)\,. \tag{1.104}$$

The last contribution in (1.104) is the conditional Shannon entropy of $\mathbf{Y}$ given $\mathbf{X}^*$ and, since $\mathbf{Y}$ given a realization of $\mathbf{X}^*$ is basically noise, $\mathcal{H}_N(\mathbf{Y}|\mathbf{X}^*)$ can be considered as a pure noise contribution.

A fundamental choice for $\mathbf{F}_{\mathbf{Z}}(\mathbf{X}^*)$ is the *Gaussian channel*, namely

$$\mathbf{Y} = \sqrt{\mu}\mathbf{X}^* + \mathbf{Z} \tag{1.105}$$

where $Z_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ and $\mu > 0$ is called *signal to noise ratio* (SNR). Then the posterior measure rewrites as

$$dP_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x}) = \frac{1}{Z(\mathbf{y})}\frac{\exp\left(-\frac{1}{2}\|\mathbf{y} - \sqrt{\mu}\mathbf{x}\|^2\right)}{(2\pi)^{N/2}}\,dP_X(\mathbf{x}) \tag{1.106}$$

and the corresponding Shannon entropy is

$$\mathcal{H}_N(\mathbf{Y}) = -\mathbb{E}_{\mathbf{Y}}\log \int \frac{\exp\left(-\frac{1}{2}\|\mathbf{Y} - \sqrt{\mu}\mathbf{x}\|^2\right)}{(2\pi)^{N/2}}\,dP_X(\mathbf{x})\,. \tag{1.107}$$

Inserting the change of variables (1.105) into the previous equation one can fully exploit the Gaussian nature of the noise and prove the following

Proposition 1.11 (I-MMSE formula). *For the Gaussian channel* (1.105) *it holds that*

$$\frac{d}{d\mu}I_N(\mathbf{X}^*,\mathbf{Y}) = \frac{1}{2}MMSE = \frac{1}{2}\mathbb{E}\|\mathbf{X}^* - \langle\mathbf{X}\rangle_{\mathbf{Y}}\|^2\,. \tag{1.108}$$

Proof. A direct calculation yields

$$\frac{d}{d\mu}I_N(\mathbf{X}^*,\mathbf{Y}) = \frac{d}{d\mu}\mathcal{H}_N(\mathbf{Y}) =$$

$$-\frac{d}{d\mu}\int dP_X(\mathbf{x}^*)\int \frac{d^N\mathbf{z}}{(2\pi)^{N/2}}e^{-\frac{\|\mathbf{z}\|^2}{2}}\log\int dP_X(\mathbf{x})\frac{\exp\left(-\frac{1}{2}\|\mathbf{z}-\sqrt{\mu}(\mathbf{x}-\mathbf{x}^*)\|^2\right)}{(2\pi)^{N/2}} =$$

$$\frac{1}{2\sqrt{\mu}}\mathbb{E}_{\mathbf{Z},\mathbf{X}^*}\langle(\mathbf{Z}-\sqrt{\mu}(\mathbf{X}-\mathbf{X}^*))\cdot(\mathbf{X}-\mathbf{X}^*)\rangle_{\mathbf{Z},\mathbf{X}^*}. \quad (1.109)$$

Using Gaussian integration by parts over $\mathbf{Z}$ one gets

$$\frac{d}{d\mu}I_N(\mathbf{X}^*,\mathbf{Y}) = \frac{1}{2}\mathbb{E}_{\mathbf{Z},\mathbf{X}^*}\langle\|\mathbf{X}-\mathbf{X}^*\|^2\rangle_{\mathbf{Z},\mathbf{X}^*} - \frac{1}{2\sqrt{\mu}}\mathbb{E}_{\mathbf{Z},\mathbf{X}^*}\langle\mathbf{Z}-\sqrt{\mu}(\mathbf{X}-\mathbf{X}^*)\rangle\cdot(\mathbf{X}-\mathbf{X}^*)\rangle_{\mathbf{Z},\mathbf{X}^*}$$

$$+\frac{1}{2\sqrt{\mu}}\mathbb{E}_{\mathbf{Z},\mathbf{X}^*}\langle\mathbf{Z}-\sqrt{\mu}(\mathbf{X}-\mathbf{X}^*)))\rangle_{\mathbf{Z},\mathbf{X}^*}\cdot(\mathbf{X}-\mathbf{X}^*)\rangle_{\mathbf{Z},\mathbf{X}^*} = \frac{1}{2}\mathbb{E}_{\mathbf{Z},\mathbf{X}^*}\|\mathbf{X}^*-\langle\mathbf{X}\rangle_{\mathbf{Z},\mathbf{X}^*}\|^2. \quad (1.110)$$

Finally, one replaces back $\mathbb{E}_{\mathbf{Z},\mathbf{X}^*}$ with $\mathbb{E}_{\mathbf{Y}}$ and $\langle\cdot\rangle_{\mathbf{Z},\mathbf{X}^*}$ with $\langle\cdot\rangle_{\mathbf{Y}}$, and the statement is proved. □

Notice that I_N is related to the MMSE only through a Nishimori identity. In fact, the final result in (1.110) would not match the definition of MMSE in absence of the Nishimori identities, but it would still be a measure of divergence from the ground truth one wants to estimate.

1.4 The spiked Wigner model

The spiked Wigner model (WSM) was first introduced in [30] as a model for Principal Component Analysis (PCA), and it was widely studied in recent literature. Without pretension of being exhaustive, we refer the interested reader to [31, 32, 17, 14, 13, 33, 34], that are some of the key papers for this thesis[2]. The WSM features an observation channel of the form

$$y_{ij} = \sqrt{\frac{\mu}{N}}x_i^*x_j^* + z_{ij}, \quad i \le j \le N \quad (1.111)$$

where $x_i^* \stackrel{\text{iid}}{\sim} P_X$ and $z_{ij} \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1+\delta_{ij})$, and by convention we also take $z_{ij} = z_{ji}$. With these symmetry constraints one may rewrite the observations as

$$\mathbf{y} = \sqrt{\frac{\mu}{N}}\mathbf{x}^*\mathbf{x}^{*T} + \mathbf{z}, \quad (1.112)$$

meaning that $\mathbf{F_Z}(\mathbf{X}^*) = \sqrt{\frac{\mu}{N}}\mathbf{X}^*\mathbf{X}^{*T} + \mathbf{Z}$. The noise matrix $\mathbf{z} = (z_{ij})_{i,j\le N}$ is also called *Wigner matrix*.

---

[2]See also the introduction of Chapter 5 for other references.
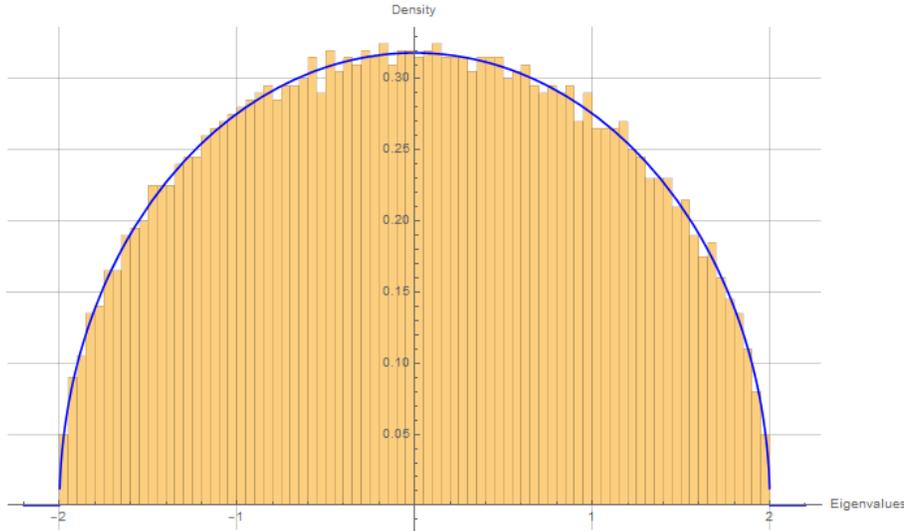
Figure 1.1: Histogram of eigenvalues against Wigner's semicircle law in blue.

At its birth, the WSM was studied from the point of view of its spectral properties. It is known that the density of eigenvalues of a properly re-scaled Wigner matrix, such as $\mathbf{Z}/\sqrt{N}$, forms a semicircle of radius 2 centered at the origin [35, 36]. However, when a rank one perturbation is added to it like in (1.111), one can rigorously prove that, if the perturbation is "strong" enough, namely if the SNR parameter μ is large enough, one eigenvalue pops out of the Wigner semicircular bulk, see Figure 1.2. This phenomenon is also known as *BBP transition* [37], and seems to suggest that the retrieval of the given instance of the signal $\mathbf{x}^*$ is possible as long as this "special" eigenvalue has popped out of the Wigner sea.

The result can be generalized for finite rank perturbations of extensive (namely growing with the size of the system/matrix) rank matrices drawn from other random matrix ensembles [38]. Furthermore, it is possible to evaluate the average overlap between the ground truth $\mathbf{x}^*$ and the eigenvector(s) corresponding to the mentioned leading eigenvalue(s) coming out of the bulk, that usually serves as an estimator for $\mathbf{x}^*$. The higher this overlap is, the better the reconstruction.

Notice though, that if the Statistician uses only this procedure to retrieve the signal then she is missing an important piece of information that is actually at hand: the prior distribution. As a matter of fact, the previous principal component analysis strategy does not take into account the prior from which $\mathbf{x}^*$ is drawn, that is known in a Bayes-optimal setting. Nevertheless, it can be proved, using techniques we shall see later, that for the Gaussian prior PCA is actually also Bayes-optimal, and we can give an intuitive reason for that: the Gaussian distribution is the one that contains least information (has the largest entropy) for fixed mean and variance, that are usually two parameters under control. Hence, there is no other information than the one contained in them to exploit.
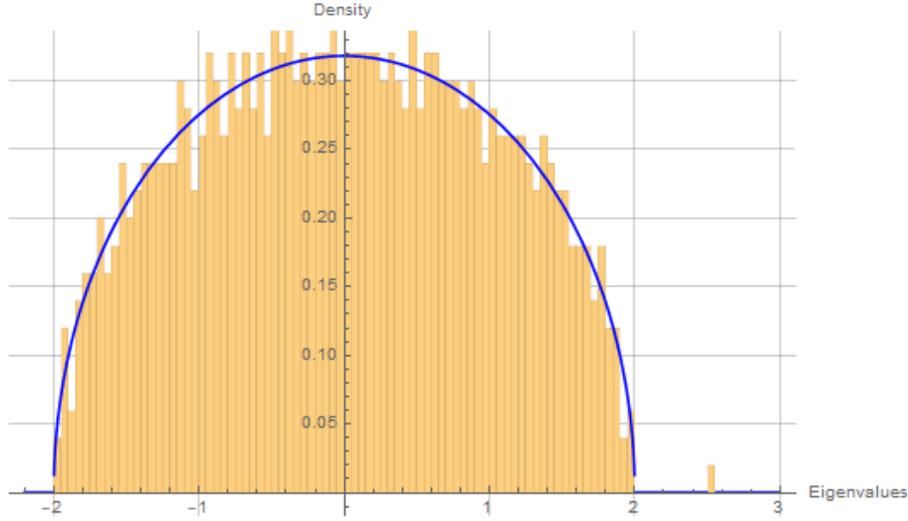
Figure 1.2: BBP transition. Spectrum of the observations (1.111) for $x_i^* \overset{\text{iid}}{\sim} (\delta_1 + \delta_{-1})/2$, $\mu = 4$. Wigner's semicircular law in blue as before.

1.4.1 Free entropy in the BOS via adaptive interpolation

For our purposes, it will suffice to restrict ourselves to the Rademacher prior, that is $P_X = (\delta_1 + \delta_{-1})/2$. The computation will simplify considerably, though maintaining most of the important features, and it will allow us to show the parallelism with the gauge theory of spin glasses, where the Nishimori identities were first found. Another convenient choice to match the notation used in spin glasses and that follows in the next chapters, is to use spins denoted by σ's, instead of the $x's$ and take the asymmetric observations

$$y_{ij} = \sqrt{\frac{\mu}{2N}}\sigma_i^*\sigma_j^* + z_{ij}, \quad 1 \leq i,j \leq N \tag{1.113}$$

with $z_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. The probabilistic properties of the original channel (1.111) can be re-obtained simply by symmetrizing the previous one: $(y_{ij} + y_{ji})/\sqrt{2}$.

(1.113) is a type of Gaussian channel. Hence, writing the related posterior measure and Statistical Mechanics Hamiltonian is not difficult:

$$dP_{\boldsymbol{\sigma}|\mathbf{Y}}(\boldsymbol{\sigma}) = \frac{1}{Z(\mathbf{y})}\prod_{i,j=1}^N \exp\left[-\frac{1}{2}\sum_{i,j=1}^N\left(y_{ij} - \sqrt{\frac{\mu}{2N}}\sigma_i\sigma_j\right)^2\right]\prod_{i=1}^N\left(\frac{\delta_{-1}(\sigma_i) + \delta_1(\sigma_i)}{2}\right). \tag{1.114}$$

In order to lighten the notation, instead of carrying all the delta products, we just sum over the sigma's keeping in mind they're just binary variables. Now we expand the square in the exponential taking into account that $\sigma_i^2 = 1$ and neglecting terms not depending on $\boldsymbol{\sigma}$, i.e.

sums of y_{ij}^2:

$$\mu_N(\boldsymbol{\sigma}) = \frac{1}{Z(\mathbf{y})} \exp\left[\sum_{i,j=1}^{N} \sqrt{\frac{\mu}{2N}} z_{ij}\sigma_i\sigma_j + \frac{\mu}{2N}\sigma_i\sigma_i^*\sigma_j\sigma_j^*\right], \tag{1.115}$$

from which we clearly recognize the two-body interaction Hamiltonian:

$$-H_N(\boldsymbol{\sigma};\mathbf{z},\boldsymbol{\sigma}^*) = \sum_{i,j=1}^{N} \left(\sqrt{\frac{\mu}{2N}} z_{ij} + \frac{\mu}{2N}\sigma_i^*\sigma_j^*\right)\sigma_i\sigma_j. \tag{1.116}$$

We denote the corresponding thermodynamic pressure per particle as

$$\bar{p}_N(\mu) = \frac{1}{N}\mathbb{E}\log\sum_{\boldsymbol{\sigma}\in\{-1,+1\}^N} \exp\left(-H_N(\boldsymbol{\sigma};\mathbf{z},\boldsymbol{\sigma}^*)\right) \tag{1.117}$$

where the average is over all the disorder. Now we look for an interpolation strategy to decouple problem in such a way that we can then sum over the spins. However, since we wish to maintain optimality (and in particular the Nishimori identities), we must identify another channel such that for any interpolation time t the model comes from an inference problem in the BOS.

Definition 1.4 (Adaptive interpolation for WSM). Let $t \in [0,1]$, $\epsilon \sim \text{Uniform}_{[s_N, 2s_N]}$ with $s_N \propto N^{-\frac{1}{16}}$. Define

$$Q_\epsilon(t) = \epsilon + \mu \int_0^t ds\, q_\epsilon(s). \tag{1.118}$$

with $q_\epsilon(s) \geq 0$ any non negative function of s to be chosen later.

The interpolating inference problem is defined by the observations

$$y_{ij} = \sqrt{\frac{(1-t)\mu}{2N}}\sigma_i^*\sigma_j^* + z_{ij} \tag{1.119}$$

$$y_i = \sqrt{Q_\epsilon(t)}\sigma_i^* + \tilde{z}_i \tag{1.120}$$

with $z_i \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ and independent on any other random variable. The related Hamiltonian is

$$-H_t(\boldsymbol{\sigma};\mathbf{z},\tilde{\mathbf{z}},\boldsymbol{\sigma}^*) = \sum_{i,j=1}^{N} \left(\sqrt{\frac{(1-t)\mu}{2N}} z_{ij} + \frac{(1-t)\mu}{2N}\sigma_i^*\sigma_j^*\right)\sigma_i\sigma_j + \sum_{i=1}^{N} \left(\sqrt{Q_\epsilon(t)}\tilde{z}_i + Q_\epsilon(t)\sigma_i^*\right)\sigma_i. \tag{1.121}$$

(1.121) induces an interpolating Gibbs measure denoted by $\langle\cdot\rangle_{N,\epsilon,t}$.

The ϵ-perturbation in the Hamiltonian has been introduced to "avoid" criticalities along the interpolation. Let us now check the corresponding interpolating pressure at the two ends.

Lemma 1.12. *Denote*

$$\bar{p}_N(t) = \frac{1}{N}\mathbb{E}\log \sum_{\boldsymbol{\sigma}\in\{-1,+1\}^N} \exp\left(-H_t(\boldsymbol{\sigma};\mathbf{z},\tilde{\mathbf{z}},\boldsymbol{\sigma}^*)\right) \tag{1.122}$$

where $\mathbb{E}$ is over all the quenched disorder, expect ϵ, and

$$\psi(q) := \mathbb{E}\log 2\cosh(Z\sqrt{q} + q\sigma^*), \quad Z \sim \mathcal{N}(0,1), \sigma^* \sim \frac{\delta_1 + \delta_{-1}}{2}. \tag{1.123}$$

Then

$$\bar{p}_N(t=0) = \bar{p}_N(\mu) + O(s_N) \tag{1.124}$$

$$\bar{p}_N(t=1) = \psi\left(Q_\epsilon(1)\right) = \psi\left(\mu\int_0^1 dt\, q_\epsilon(t)\right) + O(s_N) \tag{1.125}$$

with $Z \sim \mathcal{N}(0,1)$.

Proof. Let us begin with $\bar{p}_N(t=0)$. $\bar{p}_N(t=0)$ is Lipschitz in ϵ. Indeed, using integration by parts we can show that

$$\frac{\partial \bar{p}_N(t=0)}{\partial \epsilon} = \frac{1}{2N}\sum_{i=1}^N \mathbb{E}[1 - \langle \sigma_i^{(1)}\sigma_i^{(2)}\rangle_{N,\epsilon,t=0} + \langle \sigma_i\sigma_i^*\rangle_{N,\epsilon,t=0}],$$

from which, using a Nishimori identity on the second term in the square brackets, we get

$$\frac{\partial \bar{p}_N(t=0)}{\partial \epsilon} = \frac{1}{2N}\sum_{i=1}^N [1 + \mathbb{E}\langle \sigma_i\sigma_i^*\rangle_{N,\epsilon,t=0}] \in [0,1]. \tag{1.126}$$

Hence, (1.124) follows straightforwardly by a first order Taylor expansion in ϵ.

The proof for (1.125) follows the same lines. In particular, it is sufficient to prove that ψ is Lipschitz. This involves the remarkable identity

$$\mathbb{E}\sigma^* \tanh\left(z\sqrt{q} + q\sigma^*\right) = \mathbb{E}\tanh^2\left(z\sqrt{q} + q\sigma^*\right) \tag{1.127}$$

that is itself a Nishimori identity of a one body, or side information, channel. We have that

$$\frac{\partial \psi}{\partial q}(q) = \frac{1 + \mathbb{E}\sigma^* \tanh\left(z\sqrt{q} + q\sigma^*\right)}{2} \in [0,1]. \tag{1.128}$$

(1.125) then follows again from a first order Taylor expansion. $\square$

We have seen that derivatives of the pressure make renormalized spin sums appear. It is thus convenient to introduce the following standard quantity.

Definition 1.5 (Mattis magnetization (or overlap)). Given a sample from the posterior measure (1.115), $\boldsymbol{\sigma}$, and an instance of the ground truth $\boldsymbol{\sigma}^*$, the Mattis magnetization is the quantity

$$m_N(\boldsymbol{\sigma}, \boldsymbol{\sigma}^*) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i \sigma_i^* . \tag{1.129}$$

The definition extends to generic priors, different than Rademacher.

The Mattis magnetization is an overlap, just like the one defined for SK, but instead of being between two replicas of the system, it is between the ground truth and a replica of the system. Thanks to the Nishimori identities, we know that actually there is no difference between the two quantities in law, in fact

$$\mathbb{E}\langle \sigma_i \sigma_i^* \rangle_N = \mathbb{E}\langle \sigma_i^{(1)} \sigma_i^{(2)} \rangle_N \tag{1.130}$$

whenever $\mathbb{E}\langle \cdot \rangle_N$ is the quenched posterior average of an inference problem in the BOS. We are now ready to state a sum rule for the adaptive interpolation:

Proposition 1.13 (Sum rule for WSM, Ising spins). *For any choice of the (non-negative) function $q_\epsilon(t)$, the pressure of the WSM (with Ising spins) (1.117) obeys the following sum rule:*

$$\bar{p}_N(\mu) = O(s_N) + \psi(Q_\epsilon(1)) + \frac{\mu}{4} \int_0^1 dt \, \left[(1 - q_\epsilon(t))^2 - 2q_\epsilon^2(t) \right] + \frac{\mu}{4} \int_0^1 dt \, R_\epsilon(t) , \tag{1.131}$$

with a remainder

$$R_\epsilon(t) := \mathbb{E}\langle (m_N(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) - q_\epsilon(t))^2 \rangle_{N,\epsilon,t} . \tag{1.132}$$

Proof. Thanks to the previous Lemma, in particular to (1.124) and (1.125), it is sufficient to compute the derivative of the interpolating pressure and to re-integrate it to evaluate the finite increment to subtract from $\bar{p}_N(1)$.

Using integration by parts over the Gaussian disorder one gets

$$\dot{p}_N(t) = -\frac{\mu}{4} \left[1 - \mathbb{E}\langle q_N^2(\boldsymbol{\sigma}, \boldsymbol{\tau}) \rangle_{N,\epsilon,t} \right] + \frac{\mu}{2} q_\epsilon(t) \left[1 - \mathbb{E}\langle q_N(\boldsymbol{\sigma}, \boldsymbol{\tau}) \rangle_{N,\epsilon,t} \right]$$
$$- \frac{\mu}{2} \mathbb{E}\langle m_N^2(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) \rangle_{N,\epsilon,t} + \mu q_\epsilon(t) \mathbb{E}\langle m_N(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) \rangle_{N,\epsilon,t} . \tag{1.133}$$

Since the interpolating model is build to be on the Nishimori line for any t we are allowed to use the Nishimori identities, that in particular imply $\mathbb{E}\langle m_N^k(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) \rangle_{N,\epsilon,t} = \mathbb{E}\langle q_N^k(\boldsymbol{\sigma}, \boldsymbol{\tau}) \rangle_{N,\epsilon,t}$ for any k. These yields

$$\dot{p}_N(t) = \frac{\mu}{2} q_\epsilon^2(t) - \frac{\mu}{4}(1 - q_\epsilon(t))^2 - \frac{\mu}{4} \mathbb{E}\langle (m_N(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) - q_\epsilon(t))^2 \rangle_{N,\epsilon,t} . \tag{1.134}$$

The result is then directly implied by the fundamental theorem of integral calculus. □

Notice that, exactly as for the SK model, the sum rule presents a remainder whose vanishing is in direct connection with the replica symmetry (breaking) of the model. In this case though, thanks to Bayes-optimality and the Nishimori identities one can show that the overlap (Mattis magnetization) does indeed concentrate in a specific measure, for a specific choice of $q_\epsilon(t)$.

The one interesting for our purposes is made according to the following ODE

$$\dot{Q}_\epsilon(t) = \mu \mathbb{E} \langle m_N(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) \rangle_{N,\epsilon,t}, \quad Q_\epsilon(0) = \epsilon. \tag{1.135}$$

Notice that we cannot choose since the very beginning $q_\epsilon(t) = \mathbb{E} \langle m_N(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) \rangle_{N,\epsilon,t}$, since the brackets themselves depend on $q_\epsilon(t)$. Hence, we have to formalize the choice using an ODE as done above. It can be proved that the previous ODE has a unique solution on $[0, 1]$. In fact, the velocity field on the r.h.s. is smooth, Lipschitz and non-decreasing in $Q_\epsilon(t)$ for any fixed N. In order to keep the this introduction light, we defer the proof of this fact into Appendix A.1, where it is done in a more generic setting. All this is to say that the choice $q_\epsilon(t) = \mathbb{E} \langle m_N(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) \rangle_{N,\epsilon,t}$ is indeed licit.

Lemma 1.14 (Concentration WSM, Ising spins)**.** *Denote* $\mathbb{E}_\epsilon(\cdot) := \frac{1}{s_N} \int_{s_N}^{2s_N} d\epsilon\,(\cdot)$, *and choose the interpolation functions according to* (1.135). *Then it holds that*

$$\mathbb{E}_\epsilon \mathbb{E} \Big\langle \big(m_N(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) - \mathbb{E} \langle m_N(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) \rangle_{N,\epsilon,t} \big)^2 \Big\rangle_{N,\epsilon,t} \xrightarrow{N \to \infty} 0. \tag{1.136}$$

The proof of this fact is important but rather technical. We refer again the reader to Appendix A.1 for it. For the moment, we only mention that this result can be proved in general Inference problems in the Bayes optimal setting [39, 40], and that an L^2 type of concentration for the pressure is needed.

Lemma 1.15 (L^2-concentration of WSM pressure, Ising spins)**.** *There exists a positive constant* $C > 0$ *such that*

$$\mathbb{E} \left[\left(\frac{1}{N} \log \sum_{\boldsymbol{\sigma} \in \{-1,+1\}^N} \exp\left(-H_t(\boldsymbol{\sigma}; \mathbf{z}, \tilde{\mathbf{z}}, \boldsymbol{\sigma}^*) \right) - \bar{p}_N(t) \right)^2 \right] \leq \frac{C}{N}. \tag{1.137}$$

The previous is a simple consequence of Efron-Stein concentration inequality. See the proof of Lemma 4.8 for an exhaustive example. Notice moreover, that (1.136) is expressed in ϵ-average, similarly to what happens in classical thermodynamic stability results [41], precisely to "tame" the behaviour of the fluctuations close to critical points.

With the Lemmas above, we can rigorously prove the replica symmetric variational formula for the asymptotic pressure of the spiked Wigner model. As it is clear from the proof, the concentration (1.136) is sufficient to imply a finite dimensional variational principle, in contrast to what happens in the SK model, that has no underlying planted signal $\mathbf{x}^*$.

Theorem 1.16 (Thermodynamic limit for WSM, Ising spins). *In the thermodynamic limit, the pressure of the WSM satisfies the replica symmetric variational principle*

$$\bar{p}_N(\mu) \xrightarrow{N\to\infty} \sup_{x\geq 0} p(\mu; x) \,, \tag{1.138}$$

where the variational potential p is

$$p(\mu; x) := \frac{\mu}{4}(1-x)^2 - \frac{\mu x^2}{2} + \psi(\mu x) \,. \tag{1.139}$$

Proof. The proof consists in finding two bounds matching in the thermodynamic limit.

Lower Bound: choose $q_\epsilon(t) = x \in \mathbb{R}_{\geq 0}$. The sum rule (1.131) with this choice simplifies to

$$\bar{p}_N(t) = O(s_N) + \psi(\mu x) + \frac{\mu}{4}(1-x)^2 - \frac{\mu x^2}{2} +$$
$$+ \frac{\mu}{4}\int_0^1 \mathbb{E}\langle (x - m_N(\boldsymbol{\sigma}, \boldsymbol{\sigma}^*))^2 \rangle_{N,\epsilon,t} \geq p(\mu; x) + O(s_N) \tag{1.140}$$

where we used also the second equality in (1.125). The previous bound is uniform in x. So we just let $N \to \infty$ to get rid of $O(s_N)$ and we optimize w.r.t. $x \geq 0$.

Upper bound: We start by proving that ψ is convex in its argument, that will be needed later. Observe that the following remarkable identity holds:

$$\mathbb{E}\sigma^* \tanh^3(z\sqrt{q} + q\sigma^*) = \mathbb{E}\tanh^4(z\sqrt{q} + q\sigma^*) \,, \tag{1.141}$$

again because it is a Nishimori identity of a side information Gaussian channel. Then, using this and Gaussian integration by parts one can prove that

$$\frac{d^2\psi}{dq^2}(q) = \frac{1}{2}\mathbb{E}\left[(1 - \tanh^2(z\sqrt{q} + q\sigma^*))^2\right] > 0 \,. \tag{1.142}$$

Once the convexity is proved, we use Jensen's inequality in the sum rule to extract the integral inside ψ in (1.125) at the expense of an inequality:

$$\bar{p}_N(\mu) \leq O(s_N) + \int_0^1 dt \left[\psi(\mu q_\epsilon(t)) + \frac{\mu}{4}(1 - q_\epsilon(t))^2 - \frac{\mu}{2}q_\epsilon^2(t)\right] + \int_0^1 dt \, R_\epsilon(t) =$$
$$= O(s_N) + \int_0^1 dt \, p(\mu; q_\epsilon(t)) + \int_0^1 dt \, R_\epsilon(t) \leq O(s_N) + \sup_{x\geq 0} p(\mu; x) + \int_0^1 dt \, R_\epsilon(t) \,. \tag{1.143}$$

What remains to do now is to choose the $q_\epsilon(t)$ accurately, in such a way to make $R_\epsilon(t)$ disappear. The choice is made according to (1.135). The remainder thus acquires the familiar form:

$$\mathbb{E}\left\langle \left(m_N(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) - \mathbb{E}\langle m_N(\boldsymbol{\sigma}^*, \boldsymbol{\sigma})\rangle_{N,\epsilon,t}\right)^2 \right\rangle_{N,\epsilon,t} \,. \tag{1.144}$$

We then average over ϵ on both sides of (1.143) and by Fubini's Theorem we exchange $\mathbb{E}_\epsilon$ with the t-integral of the remainder in order to be able to use (1.136):

$$\bar{p}_N(\mu) \leq O(s_N) + \sup_{x \geq 0} p(\mu; x) + \int_0^1 dt\, \mathbb{E}_\epsilon \mathbb{E} \Big\langle \big(m_N(\boldsymbol{\sigma}^*, \boldsymbol{\sigma}) - \mathbb{E}\langle m_N(\boldsymbol{\sigma}^*, \boldsymbol{\sigma})\rangle_{N,\epsilon,t}\big)^2 \Big\rangle_{N,\epsilon,t}. \quad (1.145)$$

Finally, by dominated convergence the last term vanishes when $N \to \infty$ and this completes the proof. $\qquad\square$

Now we have to reconnect the statistical pressure to the mutual information of the problem. Revising our steps that led to the definition of the pressure, it is not difficult to see what are the terms to re-add:

$$\frac{1}{N} I_N(\mathbf{X}^*, \mathbf{Y}) \xrightarrow{N \to \infty} i(\mu) := \log 2 + \frac{\mu}{2} - \sup_{x \geq 0} p(\mu; x). \quad (1.146)$$

1.4.2 Spin glasses on the Nishimori line

The WSM free entropy was computed omitting on purpose the fact that the entire problem can be made independent of the ground truth when the prior is Rademacher, as a consequence of a simple gauge transformation. We preferred to present the proofs as above because they generalize easier to the case of generic priors and the information channels are more evident.

The mentioned gauge transformation leaves the statistical properties of the model totally unchanged. More specifically, considering the Hamiltonian (1.121) for example, without changing the laws of the quenched random variables one can operate the transformation

$$z_{ij} \mapsto \sigma_i^* \sigma_j^* z_{ij}, \quad \sigma_i \mapsto \sigma_i \sigma_i^*, \quad \tilde{z}_i \mapsto \sigma_i^* \tilde{z}_i. \quad (1.147)$$

The same operation can be made on the original Hamiltonian of course, and on ψ with $Z \mapsto Z\sigma^*$ and using the parity of the hyperbolic cosine:

$$\psi(q) = \mathbb{E} \log 2 \cosh(Z\sqrt{q} + q). \quad (1.148)$$

Doing so, the interpolating Hamiltonian becomes

$$-H_t(\boldsymbol{\sigma}; \mathbf{z}, \tilde{\mathbf{z}}, \boldsymbol{\sigma}^*) = \sum_{i,j=1}^N J_{ij}(t) \sigma_i \sigma_j + \sum_{i=1}^N h_i(t) \sigma_i \quad (1.149)$$

with $J_{ij}(t) \overset{\text{iid}}{\sim} \mathcal{N}\left(\frac{(1-t)\mu}{2N}, \frac{(1-t)\mu}{2N}\right)$, $h_i(t) \overset{\text{iid}}{\sim} \mathcal{N}(Q_\epsilon(t), Q_\epsilon(t))$. We thus get a spin-glass model on the Nishimori line, namely with all the coupling parameters with mean equal to their variance [29, 19]. The very same gauge symmetry has also other consequences, among which

$$\mathbb{E}\langle \sigma_i \rangle_{N,\epsilon,t}^2 = \mathbb{E}\langle \sigma_i \tau_i \rangle_{N,\epsilon,t} = \mathbb{E}\langle \sigma_i \sigma_i^* \rangle_{N,\epsilon,t} = \mathbb{E}\langle \sigma_i \rangle_{N,\epsilon,t,\boldsymbol{\sigma}^*=\mathbf{1}} = \mathbb{E}\langle \sigma_i \rangle_{N,\epsilon,t,\boldsymbol{\sigma}^*=\mathbf{1}}^2. \quad (1.150)$$

On the Nishimori line, a spin-glass enjoys a whole family of identities as the one above, and correlation inequalities [42] that will be displayed in the following chapter.

Chapter 2

Convex multi-species spin glass on the Nishimori line

In this chapter we investigate the properties of the elliptic multi-species Sherrington-Kirkpatrick model along the Nishimori line [43], namely the sub-manifold of the phase space in which mean and variance of the random parameters, interactions and magnetic fields, coincide. The multi-species version of a mean field model is simply obtained by relaxing the full invariance under the symmetric group into the weaker one of the product of the symmetric groups on a given partition of the system. The ratios of the sizes of the sets in the partition over the size of the entire system, also called form factors, are kept fixed in the thermodynamic limit. The ellipticity condition provides the positivity and monotonicity properties that allow to study the system with interpolation methods [7, 21, 18] and obtain a Parisi like solution for Gaussian centered interactions and deterministic magnetic fields [44, 45] (see also [46] for a case with a ferromagnetic mean of the interactions).

The choice to study the model on the Nishimori line [29] reflects the importance of this sub-manifold of the phase space due to its ubiquitous appearance in error correcting codes [29], signal processing and inference problems [47, 17].

The main results, Theorem 2.6 and Lemma 2.5 in Section 2.3, are the proof of the variational expression for the pressure per particle in the thermodynamic limit and the self-averaging of the magnetization per particle. The techniques we use to prove them are obtained by merging methods whose origins belong both to statistical mechanics and high dimensional inference [48, 49, 13, 11, 19, 9, 7, 42, 50], and among them the Guerra interpolation scheme and the adaptive interpolation introduced in the previous chapter have a particular relevance.

In Section 2.1 we introduce the multi-species setting and give the definition of the model together with its main properties, such as the self-averaging of the pressure and the Nishimori identities. In Section 2.2 we extend the adaptive interpolation method our multi-dimensional model and we use it to compute the exact solution in Section 2.3, by writing the pressure in the thermodynamic limit in terms of a finite-dimensional variational principle. Finally we study the main properties of the extremizers of our variational expression. The conclusions summarise the

results and specify the connection of our model with the spatially coupled spiked model [51, 52, 47, 33]. In the Appendix A.1 the reader can find the details of the proof of the concentration of the magnetization in the thermodynamic limit, which ultimately leads to replica symmetry. For completeness the properties of the mono-species case (SK) on the Nishimori line, already introduced in the previous chapter, are studied in Appendix A.2.

2.1 Definitions and basic properties

Consider a set Λ of indices with cardinality $|\Lambda| = N$. Let us partition Λ in K disjoint subsets:

$$\Lambda = \bigcup_{r=1}^{K} \Lambda_r, \quad \Lambda_r \cap \Lambda_s = \emptyset \; \forall r \neq s, \quad |\Lambda_r| =: N_r, \quad \alpha_r := \frac{N_r}{N} \in (0,1) \tag{2.1}$$

Each subset will be called *species* from now on. The model is defined by the following Gaussian Hamiltonian:

$$H_N(\boldsymbol{\sigma}) := -\sum_{r,s=1}^{K} \sum_{(i,j)\in\Lambda_r\times\Lambda_s} \tilde{J}_{ij}^{rs}\sigma_i\sigma_j - \sum_{r=1}^{K}\sum_{i\in\Lambda_r} \tilde{h}_i^r\sigma_i, \quad \boldsymbol{\sigma} \in \{+1,-1\}^N =: \Sigma_N \tag{2.2}$$

$$\tilde{J}_{ij}^{rs} \overset{\text{iid}}{\sim} \mathcal{N}\left(\frac{\mu_{rs}}{2N}, \frac{\mu_{rs}}{2N}\right), \quad \tilde{h}_i^r \overset{\text{iid}}{\sim} \mathcal{N}(h_r, h_r) \tag{2.3}$$

where μ_{rs} and h_r are positive real numbers, and the $K \times K$ matrix $\mu = (\mu_{rs})_{r,s=1,\dots,K}$ can be assumed to be symmetric without loss of generality. Throughout this Chapter, as can be seen from the previous definitions, the family of Gaussian variables (2.3) are assumed to be in a special line where mean values and variances are tied to be identical. One can see that this condition, in the context of statistical mechanics, is known as Nishimori line and was introduced in [53] for the SK model with Bernoulli couplings. For the Gaussian SK at inverse temperature β and random couplings $J_{ij} \overset{\text{iid}}{\sim} \mathcal{N}\left(\frac{J_0}{2N}, \frac{J}{2N}\right)$ the Nishimori line is defined by $\beta J = J_0$ (see Paragraph 4.3 in [29]) which is equivalent to (2.3) when $K = 1$, which explains also why we set $\beta = 1$ throughout this chapter, and the next one, without loss of generality.

It is also convenient to rewrite the Hamiltonian (2.2) in terms of centered Gaussians. To do that we introduce the following notation for species magnetizations and overlaps that will be used throughout:

$$m_r(\boldsymbol{\sigma}) := \frac{1}{N_r}\sum_{i\in\Lambda_r} \sigma_i, \quad q_r(\boldsymbol{\sigma},\boldsymbol{\tau}) := \frac{1}{N_r}\sum_{i\in\Lambda_r}\sigma_i\tau_i \tag{2.4}$$

$$\mathbf{m}(\boldsymbol{\sigma}) := (m_r(\boldsymbol{\sigma}))_{r=1,\dots,K}, \quad \mathbf{q}(\boldsymbol{\sigma},\boldsymbol{\tau}) := (q_r(\boldsymbol{\sigma},\boldsymbol{\tau}))_{r=1,\dots,K} \tag{2.5}$$

where bold characters here and below stand for vectors and $\boldsymbol{\sigma},\boldsymbol{\tau} \in \Sigma_N := \{-1,1\}^N$. We also set:

$$\Delta := (\alpha_r\mu_{rs}\alpha_s)_{r,s=1,\dots,K}, \quad \hat{\alpha} := \text{diag}(\alpha_1,\alpha_2,\dots,\alpha_K), \quad \mathbf{h} := (h_r)_{r=1,\dots,K}. \tag{2.6}$$

We will call Δ the *effective interaction matrix* because it encodes the interactions and relative sizes of the species in our model and we notice that it is positive definite if and only if μ is. See Figure2.1 for a scheme.
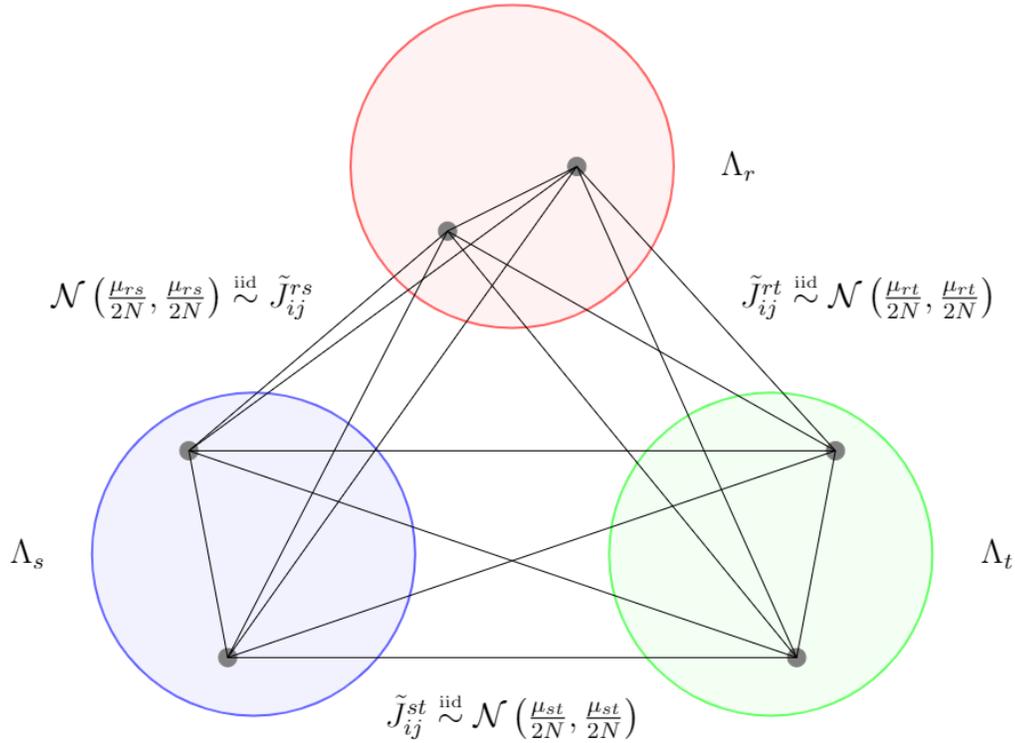
Figure 2.1: Scheme of the structure of the interactions.

With these notations we can write a Hamiltonian in terms of centered Gaussian variables which is equivalent in distribution to the one in (2.2):

$$H_N(\boldsymbol{\sigma}) = -\frac{1}{\sqrt{2N}} \sum_{r,s=1}^{K} \sum_{(i,j)\in\Lambda_r\times\Lambda_s} J_{ij}^{rs}\sigma_i\sigma_j - \sum_{r=1}^{K}\sum_{i\in\Lambda_r} h_i^r\sigma_i +$$

$$-\frac{N}{2}(\mathbf{m},\Delta\mathbf{m}) - N(\hat{\alpha}\mathbf{h},\mathbf{m}), \quad J_{ij}^{rs} \overset{\text{iid}}{\sim} \mathcal{N}\left(0,\mu_{rs}\right), \quad h_i^r \overset{\text{iid}}{\sim} \mathcal{N}(0,h_r). \quad (2.7)$$

The last expression allows us to identify the model with a multi-species Sherrington-Kirkpatrick model (SK) with the addition of a ferromagnetic interaction and a positive external field whose intensity coincide with the variances of the random terms.

Now we define the main quantity under investigation, the random and average quenched

pressure densities:

$$p_N := \frac{1}{N} \log \sum_{\boldsymbol{\sigma} \in \Sigma_N} \exp\left(-H_N(\boldsymbol{\sigma})\right) \tag{2.8}$$

$$\bar{p}_N(\mu, \mathbf{h}) := \mathbb{E} p_N \tag{2.9}$$

where we emphasize the dependence of the quenched pressure on the mean parameters μ_{rs}, $\mathbf{h}$ and the symbol $\mathbb{E}$ stands for the Gaussian expectation with respect to the disorder. We denote the Boltzmann-Gibbs expectation as:

$$\langle \cdot \rangle_N := \frac{\sum_{\boldsymbol{\sigma} \in \Sigma_N} e^{-H_N(\boldsymbol{\sigma})}(\cdot)}{Z_N}, \quad Z_N := \sum_{\boldsymbol{\sigma} \in \Sigma_N} e^{-H_N(\boldsymbol{\sigma})} . \tag{2.10}$$

In this setting the self-averaging of the pressure can be proved in its strongest form, as anticipated. We recall that the proof of replica symmetry relies on this property and on the Nishimori identities as well.

Remark 2.1. While throughout this chapter and the next one we keep the form factors α_r's constant as $N \to \infty$, all the results hold also under the weaker hypothesis that $N_r/N \to \alpha_r \in (0, 1)$. Indeed any vanishing correction to α_r doesn't change the thermodynamic limit of the quenched pressure density (2.9). This can be seen proving by interpolation method that at given N the quenched pressure is a Lipschitz function of Δ w.r.t. the entrywise matrix norm $\sum_{r,s \leq K} |\Delta_{r,s}|$.

Proposition 2.1. *There exists $C = C(\mu, h) > 0$ such that for every $x > 0$*

$$\mathbb{P}\left(|p_N - \bar{p}_N(\mu, \mathbf{h})| \geq x\right) \leq 2 \exp\left(-\frac{Nx^2}{4C}\right) . \tag{2.11}$$

As a consequence

$$\mathbb{E}[(p_N - \bar{p}_N(\mu, \mathbf{h}))^2] \leq \frac{8C}{N} . \tag{2.12}$$

Proof. The random pressure p_N is a Lipschitz function of the independent standard Gaussian variables $\hat{J} = (J_{ij}^{rs}/\sqrt{\mu_{rs}})_{i,j,r,s}$, $\hat{h} = (h_i^r/\sqrt{h_r})_{i,r}$. Indeed:

$$N^2 \|\nabla_{\hat{J}, \hat{h}} \, p_N\|^2 \leq N \left(\frac{(\mathbf{1}, \Delta \mathbf{1})}{2} + (\hat{\alpha} \mathbf{h}, \mathbf{1})\right) \equiv C \, N \tag{2.13}$$

The inequality (2.11) then follows by a standard concentration property of the Gaussian measure (see Theorem 1.3.4 in [20]). A tail integration finally leads to (2.12). $\qquad\square$

2.1.1 Nishimori identities and correlation inequalities

Here we will list some identities and inequalities on the Nishimori line. The identities were introduced in the original work by H. Nishimori [53], while the inequalities were noticed and proved much later [42, 50]. The proof of the Nishimori identities that is most suitable for our model can be found in Paragraph 2.6 of [19]. In particular, for our purposes, we will need

$$\mathbb{E}[\langle\sigma_i\rangle_N^{2n}] = \mathbb{E}[\langle\sigma_i\rangle_N^{2n-1}]\,, \quad n = 1, 2, 3, \ldots \tag{2.14}$$

$$\mathbb{E}[\langle\sigma_i\sigma_j\rangle_N^2] = \mathbb{E}[\langle\sigma_i\sigma_j\rangle_N] \tag{2.15}$$

$$\mathbb{E}[\langle\sigma_i\rangle_N\langle\sigma_i\sigma_j\rangle_N] = \mathbb{E}[\langle\sigma_i\rangle_N\langle\sigma_j\rangle_N] \tag{2.16}$$

for all $i, j \in \Lambda$. In particular they imply that:

$$\mathbb{E}[\langle q_s\rangle_N] = \sum_{i\in\Lambda_s}\frac{1}{N_s}\mathbb{E}[\langle\sigma_i\rangle_N\langle\tau_i\rangle_N] = \sum_{i\in\Lambda_s}\frac{1}{N_s}\mathbb{E}[\langle\sigma_i\rangle_N^2] = \sum_{i\in\Lambda_s}\frac{1}{N_s}\mathbb{E}[\langle\sigma_i\rangle_N] = \mathbb{E}[\langle m_s\rangle_N] \tag{2.17}$$

$$\mathbb{E}[\langle q_r q_s\rangle_N] = \sum_{(i,j)\in\Lambda_r\times\Lambda_s}\frac{\mathbb{E}[\langle\sigma_i\sigma_j\rangle_N\langle\tau_i\tau_j\rangle_N]}{N_r N_s} = \sum_{(i,j)\in\Lambda_r\times\Lambda_s}\frac{\mathbb{E}[\langle\sigma_i\sigma_j\rangle_N^2]}{N_r N_s} = \mathbb{E}[\langle m_r m_s\rangle_N] \tag{2.18}$$

and finally:

$$\mathbb{E}\Big\langle(\mathbf{q}, \Delta\mathbf{q})\Big\rangle_N = \mathbb{E}\Big\langle(\mathbf{m}, \Delta\mathbf{m})\Big\rangle_N\,. \tag{2.19}$$

The previous identities strongly suggest that the model has a unique order parameter, that can be regarded as a magnetization or equivalently an overlap. We will choose the first point of view. This intuitive statement will acquire a precise meaning when we will write the sum rule for the quenched pressure.

Following [54, 42, 50] (see Theorem 2.18 in [19] for a straightforward proof) we obtain the so-called type I and II correlation inequalities respectively:

$$\frac{\partial\bar{p}_N}{\partial h_r} = \frac{1}{2N}\sum_{i\in\Lambda_r}\mathbb{E}[1 + \langle\sigma_i\rangle_N] = \frac{\alpha_r}{2}[1 + \mathbb{E}\langle m_r\rangle_N] \geq 0 \tag{2.20}$$

$$\frac{\partial^2\bar{p}_N}{\partial h_r\partial h_s} = \frac{\alpha_r}{2}\frac{\partial}{\partial h_s}\mathbb{E}\langle m_r\rangle_N = \frac{1}{2N}\sum_{(i,j)\in\Lambda_r\times\Lambda_s}\mathbb{E}[(\langle\sigma_i\sigma_j\rangle_N - \langle\sigma_i\rangle_N\langle\sigma_j\rangle_N)^2] \geq 0\,. \tag{2.21}$$

Analogous identities and inequalities hold for the first and second derivatives w.r.t. μ_{rs}. The pressure and the first moment are monotonically increasing with respect to the Nishimori parameters μ_{rs}, h_r. In particular the magnetization is always increasing w.r.t. the external field mean:

$$\frac{\partial\mathbb{E}\langle m_r\rangle_N}{\partial h_s} \geq 0\,. \tag{2.22}$$

This monotonicity will be a key ingredient to prove replica symmetry.

2.2   Adaptive interpolation and sum rule

In this section we build up an interpolating model with some specific features. The method here employed is an extension of the standard Guerra-Toninelli interpolation [21], also called *adaptive interpolation technique*, developed in [13] by J. Barbier and N. Macris.

Definition 2.1 (Interpolating model)**.** Let $t \in [0, 1]$. The hamiltonian of the interpolating model is:

$$H_t(\boldsymbol{\sigma}) := -\frac{\sqrt{1-t}}{\sqrt{2N}} \sum_{r,s=1}^{K} \sum_{(i,j)\in\Lambda_r\times\Lambda_s} J_{ij}^{rs}\sigma_i\sigma_j - (1-t)\frac{N}{2}(\mathbf{m}, \Delta\mathbf{m})+$$
$$-\sum_{r=1}^{K}\sum_{i\in\Lambda_r}\left(\sqrt{Q_{\epsilon,r}(t)}J_i^r + Q_{\epsilon,r}(t)\right)\sigma_i - \sum_{r=1}^{K}\sum_{i\in\Lambda_r}h_i^r\sigma_i - N(\hat{\alpha}\mathbf{h},\mathbf{m}) \quad (2.23)$$

with $J_i^r \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ independent of all the other Gaussian random variables, and

$$\mathbf{Q}_\epsilon(t) := \boldsymbol{\epsilon} + \hat{\alpha}^{-1}\Delta \int_0^t \mathbf{q}_\epsilon(s)\,ds, \quad \epsilon_r \in [s_N, 2s_N],\ s_N \propto N^{-\frac{1}{16K}}\ .$$

Here $\mathbf{Q}_\epsilon =: (Q_{\epsilon,r})_{r=1,\ldots,K}$, while $\mathbf{q}_\epsilon := (q_{\epsilon,r})_{r=1,\ldots,K}$ denotes a vector of K non-negative functions that will be suitably chosen in the following.

Remark 2.2. We notice that the interpolating model is on the Nishimori line for any $t \in [0,1]$. Indeed (2.23) equals in distribution the following Hamiltonian

$$\tilde{H}_\sigma(t) = -\sum_{r,s=1}^{K}\sum_{(i,j)\in\Lambda_r\times\Lambda_s} \tilde{J}_{ij}^{rs}(t)\sigma_i\sigma_j - \sum_{r=1}^{K}\sum_{i\in\Lambda_r}\tilde{J}_i^{\epsilon,r}(t)\sigma_i - \sum_{r=1}^{K}\sum_{i\in\Lambda_r}\tilde{h}_i^r\sigma_i \quad (2.24)$$

where
$$\tilde{J}_{ij}^{rs}(t) \overset{\text{iid}}{\sim} \mathcal{N}\left(\frac{(1-t)\mu_{rs}}{2N}, \frac{(1-t)\mu_{rs}}{2N}\right), \quad \tilde{J}_i^{\epsilon,r}(t) \overset{\text{iid}}{\sim} \mathcal{N}(Q_{\epsilon,r}(t), Q_{\epsilon,r}(t)) \quad (2.25)$$

and $\tilde{h}_i^r$ is defined in (2.3). Given $t \in [0,1]$, $\tilde{H}_\sigma(t)$ is a linear combination of independent non-centered Gaussian random variables where mean equals variance. Therefore the Nishimori identities (2.14), (2.15) and (2.19) can be used by replacing $\langle\cdot\rangle$ with the Gibbs measure induced by the interpolating hamiltonian (2.23), that is $\langle\cdot\rangle_{N,t}^{(\epsilon)}$. Notice also that the role played by the functions $\mathbf{Q}_\epsilon(t)$ is that of an external magnetic field.

The corresponding interpolating pressure will be denoted as

$$\bar{p}_{N,\epsilon}(t) := \frac{1}{N}\mathbb{E}\log\sum_{\boldsymbol{\sigma}} e^{-H_t(\boldsymbol{\sigma})}\ . \quad (2.26)$$

In the previous equation and in the following we drop the explicit dependence on $\mathbf{q}_\epsilon(t)$ to lighten the notation. Let us check the interpolating pressure at extremal values:

Lemma 2.2 (Interpolating pressure at $t = 0, 1$). *Setting*

$$\psi(Q) := \mathbb{E}_z \log 2 \cosh \left[z\sqrt{Q} + Q \right], \quad z \sim \mathcal{N}(0,1) \tag{2.27}$$

we have the following:

$$\bar{p}_{N,\epsilon}(1) = \sum_{r=1}^K \alpha_r \, \psi(Q_{\epsilon,r}(1) + h_r) =$$

$$= \mathcal{O}(s_N) + \sum_{r=1}^K \alpha_r \psi \left(\left(\hat{\alpha}^{-1} \Delta \int_0^1 \mathbf{q}_\epsilon(t) \, dt + \mathbf{h} \right)_r \right) \tag{2.28}$$

$$\bar{p}_{N,\epsilon}(0) = \mathcal{O}(s_N) + \bar{p}_N(\mu, \mathbf{h}). \tag{2.29}$$

Proof. Each ϵ_r can be regarded as the mean (or variance) of a small magnetic field.

At $t = 1$ the system is *free*, non interacting. Its pressure can be explicitly computed. Take $z_i^r \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$. Then:

$$\bar{p}_{N,\epsilon}(1) = \frac{1}{N} \mathbb{E} \log \prod_{r=1}^K \sum_{\sigma \in \Sigma_{N_r}} \exp \left(\sum_{i \in \Lambda_r} \left(\sqrt{Q_{\epsilon,r}(1)} J_i^r + Q_{\epsilon,r}(1) \right) \sigma_i + \right.$$

$$\left. + \sum_{i \in \Lambda_r} (\sqrt{h_r} z_i^r + h_r) \sigma_i \right) =$$

$$= \sum_{r=1}^K \frac{\alpha_r}{N_r} \mathbb{E} \log \sum_{\sigma \in \Sigma_{N_r}} \exp \left(\sum_{i \in \Lambda_r} \left(J_i^r \sqrt{Q_{\epsilon,r}(1) + h_r} + Q_{\epsilon,r}(1) + h_r \right) \sigma_i \right)$$

where the last equality follows from the fact that J_i^r and z_i^r are independent standard Gaussian random variables. Finally:

$$\bar{p}_{N,\epsilon}(1) = \sum_{r=1}^K \alpha_r \mathbb{E}_z \log 2 \cosh \left[z\sqrt{Q_{\epsilon,r}(1) + h_r} + Q_{\epsilon,r}(1) + h_r \right], \quad z \sim \mathcal{N}(0,1).$$

By (2.20) the derivatives of the pressure w.r.t. magnetic fields are bounded by α_r and then we can get rid of the explicit dependence on ϵ_r at the expense of a term $\mathcal{O}(s_N)$, thus getting (2.28).

Analogously, by setting $t = 0$, the interpolating Hamiltonian simply reduces to the original one (2.7) except for the ϵ_r's that can be neglected again at the expense of terms $\mathcal{O}(s_N)$. □

Proposition 2.3 (Sum rule). *For any choice of the function* $\mathbf{q}_\epsilon(t)$, *the quenched pressure of the model* (2.9) *obeys to the following sum rule:*

$$\bar{p}_N(\mu, \mathbf{h}) = \mathcal{O}(s_N) + \sum_{r=1}^K \alpha_r \psi(Q_{\epsilon,r}(1) + h_r) +$$

$$+ \int_0^1 dt \left[\frac{(\mathbf{1} - \mathbf{q}_\epsilon(t), \Delta(\mathbf{1} - \mathbf{q}_\epsilon(t)))}{4} - \frac{(\mathbf{q}_\epsilon(t), \Delta\mathbf{q}_\epsilon(t))}{2} \right] + \frac{1}{4} \int_0^1 dt \, R_\epsilon(t, \mu, h) \tag{2.30}$$

where the remainder is:

$$R_\epsilon(t, \mu, h) = \mathbb{E}\Big\langle (\mathbf{m} - \mathbf{q}_\epsilon(t), \Delta(\mathbf{m} - \mathbf{q}_\epsilon(t))) \Big\rangle_{N,t}^{(\epsilon)}. \tag{2.31}$$

Proof. The proof consists in computing the first derivative by using Gaussian integration by parts for the terms containing the disorder.

$$\dot{p}_{N,\epsilon}(t) = -\frac{1}{4}\mathbb{E}\Big\langle (\mathbf{1}, \Delta\mathbf{1}) - (\mathbf{q}, \Delta\mathbf{q}) \Big\rangle_{N,t}^{(\epsilon)} - \frac{1}{2}\mathbb{E}\Big\langle (\mathbf{m} - \mathbf{q}_\epsilon(t), \Delta(\mathbf{m} - \mathbf{q}_\epsilon(t))) \Big\rangle_{N,t}^{(\epsilon)} +$$
$$+ \frac{1}{2}(\mathbf{q}_\epsilon(t), \Delta\mathbf{q}_\epsilon(t)) + \frac{1}{2}\mathbb{E}\Big\langle (\mathbf{1}, \Delta\mathbf{q}_\epsilon(t)) - (\mathbf{q}_\epsilon(t), \Delta\mathbf{q}) \Big\rangle_{N,t}^{(\epsilon)} =$$
$$= -\frac{1}{4}(\mathbf{1} - \mathbf{q}_\epsilon(t), \Delta(\mathbf{1} - \mathbf{q}_\epsilon(t))) + \frac{1}{2}(\mathbf{q}_\epsilon(t), \Delta\mathbf{q}_\epsilon(t)) +$$
$$+ \frac{1}{4}\mathbb{E}\Big\langle (\mathbf{q} - \mathbf{q}_\epsilon(t), \Delta(\mathbf{q} - \mathbf{q}_\epsilon(t))) \Big\rangle_{N,t}^{(\epsilon)} - \frac{1}{2}\mathbb{E}\Big\langle (\mathbf{m} - \mathbf{q}_\epsilon(t), \Delta(\mathbf{m} - \mathbf{q}_\epsilon(t))) \Big\rangle_{N,t}^{(\epsilon)}$$

Using the Nishimori identities (2.14) and (2.15) we can sum the last two terms together:

$$\dot{p}_{N,\epsilon}(t) = -\frac{1}{4}(\mathbf{1} - \mathbf{q}_\epsilon(t), \Delta(\mathbf{1} - \mathbf{q}_\epsilon(t))) + \frac{1}{2}(\mathbf{q}_\epsilon(t), \Delta\mathbf{q}_\epsilon(t)) +$$
$$-\frac{1}{4}\underbrace{\mathbb{E}\Big\langle (\mathbf{m} - \mathbf{q}_\epsilon(t), \Delta(\mathbf{m} - \mathbf{q}_\epsilon(t))) \Big\rangle_{N,t}^{(\epsilon)}}_{R_\epsilon(t,\mu,h)}. \tag{2.32}$$

The sum rule then follows from a simple application of the Fundamental Theorem of Calculus and the previous Lemma:

$$\bar{p}_{N,\epsilon}(0) = \mathcal{O}(s_N) + \bar{p}_N(\mu, \mathbf{h}) = \bar{p}_{N,\epsilon}(1) - \int_0^1 dt\, \dot{p}_{N,\epsilon}(t). \tag{2.33}$$

$\square$

2.3   Solution of the model

In this section we present the rigorous derivation of the thermodynamic limit of the model under the hypothesis of a positive semi-definite effective interaction matrix: $\Delta \geq 0$. First, we need a couple of lemmas listed below.

Lemma 2.4 (Liouville's formula)**.** *Consider two matrices whose elements depend on a real parameter:* $\Phi(t)$, $A(t)$. *Suppose that* Φ *satisfies:*

$$\dot{\Phi}(t) = A(t)\Phi(t) \tag{2.34}$$
$$\Phi(0) = \Phi_0 \tag{2.35}$$

Then:

$$\det(\Phi(t)) = \det(\Phi_0) \exp\left\{ \int_0^t ds \, \mathrm{Tr}(A(s)) \right\} \tag{2.36}$$

Definition 2.2 (Regularity of $\epsilon \longmapsto \mathbf{Q}_\epsilon(\cdot)$). We will say that the map $\epsilon \longmapsto \mathbf{Q}_\epsilon(\cdot)$ is *regular* if

$$\det\left(\frac{\partial \mathbf{Q}_\epsilon(t)}{\partial \boldsymbol{\epsilon}} \right) \geq 1 \quad \forall t \in [0,1] \tag{2.37}$$

Remark 2.3. Choosing $\mathbf{Q}_\epsilon$ as the solution of the following ODE:

$$\dot{\mathbf{Q}}_\epsilon(t) = \hat{\alpha}^{-1} \Delta \, \mathbb{E}\langle \mathbf{m} \rangle_{N,t}^{(\epsilon)}, \qquad \mathbf{Q}_\epsilon(0) = \boldsymbol{\epsilon} \tag{2.38}$$

the map $\epsilon \longmapsto \mathbf{Q}_\epsilon(\cdot)$ turns out to be regular. Indeed

$$\frac{d}{dt} \frac{\partial \mathbf{Q}_\epsilon(t)}{\partial \boldsymbol{\epsilon}} = \underbrace{\frac{\partial}{\partial \mathbf{Q}_\epsilon(t)} \hat{\alpha}^{-1} \Delta \mathbb{E}\langle \mathbf{m} \rangle_{N,t}^{(\epsilon)}}_{=:A(t)} \frac{\partial \mathbf{Q}_\epsilon(t)}{\partial \boldsymbol{\epsilon}} \; ; \tag{2.39}$$

since $Q_{\epsilon,r}(t)$ can be regarded as the variance of a magnetic field on the Nishimori line in (2.24) and the entries of $\hat{\alpha}^{-1}$ and Δ are non-negative we have:

$$\mathrm{Tr} A(t) \geq 0 \,, \tag{2.40}$$

by the correlation inequalities of type II (2.21), (2.22). Finally using Liouville's formula we get:

$$\det\left(\frac{\partial \mathbf{Q}_\epsilon(t)}{\partial \boldsymbol{\epsilon}} \right) = \underbrace{\det\left(\frac{\partial \mathbf{Q}_\epsilon(0)}{\partial \boldsymbol{\epsilon}} \right)}_{=1} \exp\left\{ \int_0^t ds \, \mathrm{Tr}(A(s)) \right\} \geq 1 \tag{2.41}$$

We stress that the sign of Δ plays no role yet, since we have used only the positivity of its entries so far.

Lemma 2.5 (Concentration). *Suppose $\epsilon \longmapsto \mathbf{Q}_\epsilon(\cdot)$ is a regular map. Consider the quantity:*

$$\mathcal{L}_r := \frac{1}{N_r} \sum_{i \in \Lambda_r} \left(\sigma_i + \frac{J_i^r \sigma_i}{2\sqrt{Q_{\epsilon,r}(t)}} \right), \qquad J_i^r \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \tag{2.42}$$

and introduce the $\boldsymbol{\epsilon}$-average:

$$\mathbb{E}_{\boldsymbol{\epsilon}}[\cdot] = \prod_{r=1}^K \left(\frac{1}{s_N} \int_{s_N}^{2s_N} d\epsilon_r \right) (\cdot) \,. \tag{2.43}$$

We have:

$$\mathbb{E}_{\boldsymbol{\epsilon}}\mathbb{E}\Big\langle \Big(\mathcal{L}_r - \mathbb{E}\langle\mathcal{L}_r\rangle_{N,t}^{(\epsilon)}\Big)^2 \Big\rangle_{N,t}^{(\epsilon)} \longrightarrow 0 \;, \quad when \; N \to \infty \tag{2.44}$$

and

$$\mathbb{E}\Big\langle \Big(\mathcal{L}_r - \mathbb{E}\langle\mathcal{L}_r\rangle_{N,t}^{(\epsilon)}\Big)^2 \Big\rangle_{N,t}^{(\epsilon)} \geq \frac{1}{4}\mathbb{E}\Big\langle \Big(m_r - \mathbb{E}\langle m_r\rangle_{N,t}^{(\epsilon)}\Big)^2 \Big\rangle_{N,t}^{(\epsilon)} \tag{2.45}$$

therefore the magnetization (or the overlap) concentrates in $\boldsymbol{\epsilon}$-average.

The proof, simple but lengthy (see Appendix A.1), is based on controlling the thermal and disorder-related fluctuations of $\mathcal{L}_r$. This implies the control of the fluctuations of the magnetization thus ensuring the replica symmetry of the model. We stress that the only hypothesis used up to now is the positivity of the matrix elements of Δ.

The role of $\boldsymbol{\epsilon}$ is that of a regularizing perturbation and it is crucial for the proof. Its introduction intuitively allows to avoid critical points where the limiting pressure presents singularities and concentration may not occur, thus helping us to select always the stable state of the system. Indeed, for vanishing external magnetic fields $\mathbf{h} = 0$ and in absence of $\boldsymbol{\epsilon}$, the system can remain stuck in a vanishing average magnetization state because of the resulting spin flip symmetry in the Hamiltonian. However, as we shall see, in the appropriate range of parameters the latter is thermodynamically unstable, meaning that any arbitrarily small magnetic field would bring the magnetization to positive values.

We have laid the ground for our main result: the computation of the quenched pressure in the thermodynamic limit in form of a finite dimensional (due to the concentration lemma) variational principle.

Theorem 2.6 (Thermodynamic limit). *On the Nishimori line, when $\Delta \geq 0$, the thermodynamic limit of the pressure $\bar{p}(\mu, h) := \lim_{N\to\infty} \bar{p}_N(\mu, \mathbf{h})$ exists and:*

$$\bar{p}(\mu, h) = \sup_{\mathbf{x}\in\mathbb{R}_{\geq 0}^K} \bar{p}(\mu, h; \mathbf{x}) \tag{2.46}$$

where

$$\bar{p}(\mu, h; \mathbf{x}) := \frac{(\mathbf{1} - \mathbf{x}, \Delta(\mathbf{1} - \mathbf{x}))}{4} - \frac{(\mathbf{x}, \Delta\mathbf{x})}{2} + \sum_{r=1}^{K} \alpha_r \psi((\hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h})_r) \tag{2.47}$$

with the following stationary condition:

$$\mathbf{x} - \mathbb{E}_z \tanh\left(z\sqrt{\hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h}} + \hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h}\right) \in Ker\Delta\,, \quad z \sim \mathcal{N}(0, 1) \tag{2.48}$$

Proof. Let us divide the proof in two steps.

Lower Bound: We initially fix $\mathbf{q}_\epsilon(t) = \mathbf{x} \in \mathbb{R}_{\geq 0}^K$ in (2.30). Up to orders $\mathcal{O}(s_N)$ we find:

$$\bar{p}_N(\mu, \mathbf{h}) = \mathcal{O}(s_N) + \frac{(\mathbf{1} - \mathbf{x}, \Delta(\mathbf{1} - \mathbf{x}))}{4} - \frac{(\mathbf{x}, \Delta\mathbf{x})}{2} +$$

$$+ \sum_{r=1}^K \alpha_r \mathbb{E}_z \log 2 \cosh\left(z\sqrt{(\hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h})_r} + (\hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h})_r \right) +$$

$$+ \frac{1}{4} \int_0^1 dt\, R_\epsilon(t, \mu, h) \quad (2.49)$$

We have exploited the result in Lemma 2.2. Being Δ positive semi-definite, the remainder has a positive sign, for it is a quadratic form exactly with matrix Δ.

Hence:

$$\bar{p}_N(\mu, \mathbf{h}) \geq \mathcal{O}(s_N) + \frac{(\mathbf{1} - \mathbf{x}, \Delta(\mathbf{1} - \mathbf{x}))}{4} - \frac{(\mathbf{x}, \Delta\mathbf{x})}{2} + \sum_{r=1}^K \alpha_r \psi((\hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h})_r)$$

Then, taking the $\liminf_{N\to\infty}$ on both sides and optimizing with $\sup_{\mathbf{x}}$ we get the first bound:

$$\liminf_{N\to\infty} \bar{p}_N \geq \sup_{\mathbf{x}} \left\{ \frac{(\mathbf{1} - \mathbf{x}, \Delta(\mathbf{1} - \mathbf{x}))}{4} - \frac{(\mathbf{x}, \Delta\mathbf{x})}{2} + \sum_{r=1}^K \alpha_r \psi((\hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h})_r) \right\}. \quad (2.50)$$

Upper Bound: We recall a key observation: $\psi(\cdot)$ is a convex function. With the tools we have introduced, this can finally be seen as a consequence of the correlation inequalities (2.20) (2.21) on the Nishimori line. In fact, $\psi(Q)$ can be recast in the following way:

$$\psi(Q) = \mathbb{E}_z \log \sum_{\sigma=\pm 1} e^{\sigma(z\sqrt{Q}+Q)} = \mathbb{E}_{z(Q)} \log \sum_{\sigma=\pm 1} e^{\sigma z(Q)}, \ z(Q) \sim \mathcal{N}(Q, Q).$$

This is a simple 1-particle, free system on the Nishimori line. For this model we have:

$$\frac{\partial \psi}{\partial Q} = \frac{1}{2}\mathbb{E}_z[1 + \langle\sigma\rangle], \quad \frac{\partial^2 \psi}{\partial Q^2} = \frac{1}{2}\mathbb{E}[(1 - \langle\sigma\rangle^2)^2)], \quad \langle\sigma\rangle = \frac{\sum_{\sigma=\pm 1} e^{\sigma z(Q)}\sigma}{\sum_{\sigma=\pm 1} e^{\sigma z(Q)}} = \tanh z(Q). \quad (2.51)$$

This allows us to use Jensen's inequality to extract the integral in $Q_{\epsilon,r}(1)$ from the terms containing ψ in (2.28) (in Lemma 2.2). More explicitly

$$\sum_{r=1}^K \alpha_r \psi\left(\left(\hat{\alpha}^{-1}\Delta \int_0^1 \mathbf{q}_\epsilon(t)\, dt + \mathbf{h}\right)_r\right) \leq \sum_{r=1}^K \alpha_r \int_0^1 \psi\left((\hat{\alpha}^{-1}\Delta\mathbf{q}_\epsilon(t) + \mathbf{h})_r\right)\, dt.$$

By inserting the previous inequality in the sum rule (2.30) we have that:

$$\bar{p}_N(\mu, \mathbf{h}) \le \mathcal{O}(s_N) + \int_0^1 dt \left[\frac{(\mathbf{1} - \mathbf{q}_\epsilon(t), \Delta(\mathbf{1} - \mathbf{q}_\epsilon(t)))}{4} - \frac{(\mathbf{q}_\epsilon(t), \Delta \mathbf{q}_\epsilon(t))}{2} + \right.$$

$$\left. + \sum_{r=1}^K \alpha_r \psi((\hat{\alpha}^{-1} \Delta \mathbf{q}_\epsilon(t) + \mathbf{h})_r) \right] + \frac{1}{4} \int_0^1 dt\, R_\epsilon(t, \mu, h) \le$$

$$\le \mathcal{O}(s_N) + \sup_{\mathbf{x}} \left\{ \frac{(\mathbf{1} - \mathbf{x}, \Delta(\mathbf{1} - \mathbf{x}))}{4} - \frac{(\mathbf{x}, \Delta \mathbf{x})}{2} + \sum_{r=1}^K \alpha_r \psi((\hat{\alpha}^{-1} \Delta \mathbf{x} + \mathbf{h})_r) \right\} +$$

$$+ \frac{1}{4} \int_0^1 dt\, R_\epsilon(t, \mu, h) \quad (2.52)$$

If we finally take the expectation $\mathbb{E}_\epsilon$ on both sides of the previous inequality we get:

$$\bar{p}_N(\mu, \mathbf{h}) \le \mathcal{O}(s_N) + \sup_{\mathbf{x}} \left\{ \frac{(\mathbf{1} - \mathbf{x}, \Delta(\mathbf{1} - \mathbf{x}))}{2} - (\mathbf{x}, \Delta \mathbf{x}) + \right.$$

$$\left. + \sum_{r=1}^K \alpha_r \psi((\hat{\alpha}^{-1} \Delta \mathbf{x} + \mathbf{h})_r) \right\} + \frac{1}{2} \mathbb{E}_\epsilon \int_0^1 dt\, R_\epsilon(t, \mu, h) \quad (2.53)$$

Recall that the remainder, defined in (2.31), depends on the functions $\mathbf{q}_\epsilon(t)$. This time we choose $\mathbf{q}_\epsilon(t)$ according to a different criterion. We would like to have: $\Delta \mathbf{q}_\epsilon(t) = \Delta \mathbb{E}\langle \mathbf{m} \rangle_{N,t}^{(\epsilon)}$. In this way we could use the concentration Lemma 2.5. This can be achieved through the following ODE:

$$\dot{\mathbf{Q}}_\epsilon(t) = \hat{\alpha}^{-1} \Delta \mathbb{E}\langle \mathbf{m} \rangle_{N,t}^{(\epsilon)} =: \mathbf{F}(t, \mathbf{Q}_\epsilon(t)), \quad \mathbf{Q}_\epsilon(0) = \boldsymbol{\epsilon} \quad (2.54)$$

As seen in (2.21), the derivatives of $\mathbf{F}$ are positive and bounded for any fixed N. This guarantees the existence of a unique solution over $[0, 1]$.

Then, exchanging the two integrals by Fubini's theorem in (2.53), and applying Lemma 2.5 we get:

$$\limsup_{N \to \infty} \bar{p}_N \le \sup_{\mathbf{x}} \left\{ \frac{(\mathbf{1} - \mathbf{x}, \Delta(\mathbf{1} - \mathbf{x}))}{4} - \frac{(\mathbf{x}, \Delta \mathbf{x})}{2} + \sum_{r=1}^K \alpha_r \psi((\hat{\alpha}^{-1} \Delta \mathbf{x} + \mathbf{h})_r) \right\} .$$

The two bounds match and this proves (2.46). Moreover, using the properties (2.51) the gradient of (2.46) is:

$$\nabla_{\mathbf{x}} \bar{p}(\mu, h; \mathbf{x}) = -\frac{\Delta}{2}(\mathbf{1} - \mathbf{x}) - \Delta \mathbf{x} + \frac{\Delta}{2}\mathbf{1} +$$

$$+ \frac{\Delta}{2} \mathbb{E}_z \tanh\left(z\sqrt{\hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h}} + \hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h}\right) =$$

$$= \frac{\Delta}{2}\left[-\mathbf{x} + \mathbb{E}_z \tanh\left(z\sqrt{\hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h}} + \hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h}\right)\right] \quad (2.55)$$

and it vanishes exactly when (2.48) holds. □

Remark 2.4. Notice that the positive definiteness of Δ is used to ensure the positivity of the remainder in (2.49), that ultimately leads to the lower bound. It is evident that if the sign of Δ is not definite the technique used does not produce any bound. In that case indeed, there is a direction along which the quadratic form in (2.47) can blow up to infinity. Thus, one should expect a $\min - \max$ principle as happens for bipartite systems, *e.g.* the Wishart model [17] and the bipartite SK in its replica symmetric phase [55, 56, 57, 58], that notoriously have a non-elliptic interaction structure.

Proposition 2.7. *Let Δ be strictly positive definite in* (2.47)*. Denote by $\rho(A)$ the spectral radius of a matrix A and by $\mathscr{H}_{\mathbf{x}}\bar{p}$ the Hessian matrix of $\bar{p}$. The following implication holds:*

$$\rho(\hat{\alpha}^{-1}\Delta) < 1 \quad \Rightarrow \quad \mathscr{H}_{\mathbf{x}}\bar{p}(\mu, h; \mathbf{x}) < 0, \ \forall \mathbf{x} \in \mathbb{R}^K_{\geq 0} \tag{2.56}$$

or equivalently $\bar{p}(\mu, h; \mathbf{x})$ is strictly concave w.r.t. $\mathbf{x}$.

Proof. The Hessian matrix can be computed starting from the gradient (2.55) and using properties (2.51):

$$\mathscr{H}_{\mathbf{x}}\bar{p}(\mu, h; \mathbf{x}) = -\frac{\Delta}{2} + \frac{1}{2}\Delta\mathcal{D}(\mathbf{x}, \mathbf{h})\hat{\alpha}^{-1}\Delta =$$

$$= \frac{1}{2}\Delta^{1/2}\left[-\mathbb{1} + \Delta^{1/2}\hat{\alpha}^{-1}\mathcal{D}(\mathbf{x}, \mathbf{h})\Delta^{1/2}\right]\Delta^{1/2} \tag{2.57}$$

$$\mathcal{D}(\mathbf{x}, \mathbf{h}) := \mathrm{diag}\left\{\mathbb{E}_z\left(1 - \tanh^2\left(z\sqrt{\hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h}} + \hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h}\right)_r\right)^2\right\}_{r=1,\ldots,K}. \tag{2.58}$$

Since similar matrices have the same spectral radius we have:

$$\rho\left(\Delta^{1/2}\hat{\alpha}^{-1}\mathcal{D}(\mathbf{x}, \mathbf{h})\Delta^{1/2}\right) = \rho(\hat{\alpha}^{-1}\mathcal{D}(\mathbf{x}, \mathbf{h})\Delta) =$$

$$= \rho\left(\mathcal{D}^{1/2}(\mathbf{x}, \mathbf{h})\hat{\alpha}^{-1/2}\Delta\hat{\alpha}^{-1/2}\mathcal{D}^{1/2}(\mathbf{x}, \mathbf{h})\right). \tag{2.59}$$

Now we use the fact that, for symmetric matrices, the spectral radius coincides with the matrix norm induced by the Euclidean norm:

$$\rho\left(\Delta^{1/2}\hat{\alpha}^{-1}\mathcal{D}(\mathbf{x}, \mathbf{h})\Delta^{1/2}\right) = \|\mathcal{D}^{1/2}(\mathbf{x}, \mathbf{h})\hat{\alpha}^{-1/2}\Delta\hat{\alpha}^{-1/2}\mathcal{D}^{1/2}(\mathbf{x}, \mathbf{h})\| \leq$$

$$\leq \|\mathcal{D}(\mathbf{x}, \mathbf{h})\|\|\hat{\alpha}^{-1/2}\Delta\hat{\alpha}^{-1/2}\| = \|\mathcal{D}(\mathbf{x}, \mathbf{h})\|\rho\left(\hat{\alpha}^{-1/2}\Delta\hat{\alpha}^{-1/2}\right) \leq$$

$$\leq \rho\left(\hat{\alpha}^{-1/2}\Delta\hat{\alpha}^{-1/2}\right). \tag{2.60}$$

Finally, exploiting again matrix similarity:

$$\rho\left(\Delta^{1/2}\hat{\alpha}^{-1}\mathcal{D}(\mathbf{x}, \mathbf{h})\Delta^{1/2}\right) \leq \rho\left(\hat{\alpha}^{-1}\Delta\right) < 1 \tag{2.61}$$

by hypothesis. The previous one implies that:

$$-\mathbb{1} + \Delta^{1/2}\hat{\alpha}^{-1}\mathcal{D}(\mathbf{x}, \mathbf{h})\Delta^{1/2} < 0 \tag{2.62}$$

whence, for any test vector $\mathbf{v}$:

$$\left(\mathbf{v}, \Delta^{1/2}\left[-\mathbb{1} + \Delta^{1/2}\hat{\alpha}^{-1}\mathcal{D}(\mathbf{x}, \mathbf{h})\Delta^{1/2}\right]\Delta^{1/2}\mathbf{v}\right) =$$
$$= \left(\Delta^{1/2}\mathbf{v}, \left[-\mathbb{1} + \Delta^{1/2}\hat{\alpha}^{-1}\mathcal{D}(\mathbf{x}, \mathbf{h})\Delta^{1/2}\right](\Delta^{1/2}\mathbf{v})\right) < 0 . \tag{2.63}$$

$\square$

Remark 2.5. The previous proposition implies that, whenever Δ is invertible, $\mathbf{h} = 0$ and $\rho(\hat{\alpha}^{-1}\Delta) < 1$, the point $\mathbf{x} = 0$ is the unique maximizer of (2.47). On the contrary, when $\rho(\hat{\alpha}^{-1}\Delta) > 1$ we have

$$\mathscr{H}_{\mathbf{x}}\bar{p}(\mu, 0; 0) = \frac{1}{2}\Delta^{1/2}\left[-\mathbb{1} + \Delta^{1/2}\hat{\alpha}^{-1}\Delta^{1/2}\right]\Delta^{1/2}$$

and the matrix in square brackets has at least one positive eigenvalue, therefore $\mathbf{x} = 0$ becomes an unstable saddle point for the variational pressure, thus signalling a phase transition. Notice that this instability can be generated both varying the parameters Δ_{rs} and the form factors α_r.

Remark 2.6. If Δ is non singular, our variational pressure (2.47) goes to $-\infty$ as $\|\mathbf{x}\| \to \infty$, because the concave quadratic form always dominates the sum of the terms containing ψ, which is Lispchitz with $\mathrm{Lip}(\psi) \le 1$ (again by (2.51)). This, together with the regularity of $\bar{p}$ ensures that there is a global maximum satisfying the fixed point equation:

$$\mathbf{x} = \mathbb{E}_z \tanh\left(z\sqrt{\hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h}} + \hat{\alpha}^{-1}\Delta\mathbf{x} + \mathbf{h}\right) =: \mathbf{T}(\mathbf{x}; \mathbf{h}) . \tag{2.64}$$

The Jacobian matrix of $\mathbf{T}(\cdot; \mathbf{h})$ is:

$$\mathrm{D}\mathbf{T}(\mathbf{x}; \mathbf{h}) = \mathcal{D}(\mathbf{x}, \mathbf{h})\hat{\alpha}^{-1}\Delta \tag{2.65}$$

and satisfies:

$$\rho(\mathrm{D}\mathbf{T}(\mathbf{x}; \mathbf{h})) = \rho(\mathcal{D}(\mathbf{x}, \mathbf{h})\hat{\alpha}^{-1}\Delta) \le \rho(\hat{\alpha}^{-1}\Delta) \tag{2.66}$$

as proved in Proposition 2.7. Equality holds at $\mathbf{h} = 0$ and $\mathbf{x} = 0$. Hence when $\rho(\hat{\alpha}^{-1}\Delta) < 1$ the iteration of $\mathbf{T}(\cdot; \mathbf{h})$ converges to a fixed point. If this does not hold, we still have that at one local maximum point, say $\mathbf{x}^*$:

$$\mathscr{H}_{\mathbf{x}}\bar{p}(\mu, h; \mathbf{x}^*) < 0 \quad \text{or} \quad \rho(\Delta^{1/2}\hat{\alpha}^{-1}\mathcal{D}(\mathbf{x}^*, \mathbf{h})\Delta^{1/2}) = \rho(\mathrm{D}\mathbf{T}(\mathbf{x}^*; \mathbf{h})) < 1 . \tag{2.67}$$

The latter implies that the iteration $\mathbf{x}_{n+1} = \mathbf{T}(\mathbf{x}_n; \mathbf{h})$ converges to $\mathbf{x}^*$ (locally) provided that $\|\mathbf{x}_0 - \mathbf{x}^*\| < \delta$ with δ sufficiently small.

Remark 2.7. Our parameters lie in $\mathbb{R}_{\geq 0}^K$, thus the vanishing gradient condition *a priori* allows us only to find maximizers of (2.47) in the interior, namely when $x_r > 0 \,\forall\, r = 1, \ldots, K$. More rigorously, the necessary conditions for a point $\bar{\mathbf{x}} \in \mathbb{R}_{\geq 0}^K$ to be a maximizer are:

$$
\begin{cases}
\partial_{x_r} \bar{p}(\mu, h; \bar{\mathbf{x}}) = \frac{1}{2} \left[\Delta(-\bar{\mathbf{x}} + \mathbf{T}(\bar{\mathbf{x}}; \mathbf{h})) \right]_r \leq 0 \\
\bar{x}_r \partial_{x_r} \bar{p}(\mu, h; \bar{\mathbf{x}}) = \frac{1}{2} x_r \left[\Delta(-\bar{\mathbf{x}} + \mathbf{T}(\bar{\mathbf{x}}; \mathbf{h})) \right]_r = 0
\end{cases}
\tag{2.68}
$$

If we notice that $T_r(\mathbf{x}; \mathbf{h}) \geq 0$ these conditions imply:

$$
\begin{cases}
(\mathbf{T}(\bar{\mathbf{x}}; \mathbf{h}), \Delta(-\bar{\mathbf{x}} + \mathbf{T}(\bar{\mathbf{x}}; \mathbf{h}))) \leq 0 \\
(\bar{\mathbf{x}}, \Delta(-\bar{\mathbf{x}} + \mathbf{T}(\bar{\mathbf{x}}; \mathbf{h}))) = 0
\end{cases}
\quad \Rightarrow \quad (-\bar{\mathbf{x}} + \mathbf{T}(\bar{\mathbf{x}}; \mathbf{h}), \Delta(-\bar{\mathbf{x}} + \mathbf{T}(\bar{\mathbf{x}}; \mathbf{h}))) \leq 0 \ . \tag{2.69}
$$

However, since $\Delta > 0$ we must necessarily have:

$$
(-\bar{\mathbf{x}} + \mathbf{T}(\bar{\mathbf{x}}; \mathbf{h}), \Delta(-\bar{\mathbf{x}} + \mathbf{T}(\bar{\mathbf{x}}; \mathbf{h}))) = 0 \quad \Leftrightarrow \quad -\bar{\mathbf{x}} + \mathbf{T}(\bar{\mathbf{x}}; \mathbf{h}) = 0 \ . \tag{2.70}
$$

From the previous we can see that the consistency equation (2.64) is necessarily satisfied also by maximizers on the boundary.

2.4 Concluding remarks

Although it was presented as a spin-glass, the multi-species model dealt with above admits an inferential interpretation. We saw in Chapter 1 that our case $K = 1$, *i.e.* SK on the Nishimori line, is equivalent, thanks to the $\mathbb{Z}_2$ gauge symmetry, to the spiked Wigner model with Rademacher prior $\rho = 1/2(\delta_1 + \delta_{-1})$. For generic K instead, the corresponding inference problem originates from the Gaussian channel

$$
y_{ij}(\mu_{rs}) = \sqrt{\frac{\mu_{rs}}{2N}} \sigma_i^* \sigma_j^* + z_{ij} \ , \quad z_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \ , \ (i,j) \in \Lambda_r \times \Lambda_s \tag{2.71}
$$

where $(\mu_{rs})_{r,s=1,\ldots,K}$ are non-negative numbers, $\sigma^* \in \{\pm 1\}^N$ (the ground truth) is the signal we want to recover through the observations y_{ij}. Here μ_{rs} play the role of an index dependent signal-to-noise ratio, which gives the model, and in particular the noise, more structure. However, the noise elements remain independent. Up to constants, the Gibbs measure associated to our Hamiltonian (2.2) corresponds to the posterior distribution in the Bayes-optimal setting and the pressure corresponds to the mutual information, exactly as we have seen in the previous chapter.

We mention that model we take into account was studied under some specific assumptions on the μ_{rs}, listed in Paragraph 2.3 in [33] (see also [47]). The thermodynamic properties the authors focus on are obtained by first considering the infinite volume limit of each block and then sending the number of blocks to infinity thus recovering the limiting mutual information

of the spiked Wigner model, i.e. the case with homogeneous ($\mu_{rs} = \mu$) signal-to-noise ratio. In the present work instead, in the case of a Rademacher planted signal and the only positive definiteness assumption on the matrix μ, the model is studied and solved for arbitrary number of species and form factors. The positivity assumption on μ rules out some interesting non-elliptic structures such as restricted Boltzmann machines. We will show in the next chapter how to deal with these peculiar class of non-convexities, proving a replica symmetric variational formula for the pressure of the deep Boltzmann Machine on the Nishimori line [59].

Chapter 3

The deep Boltzmann machine on the Nishimori line

The direct problem of a deep (restricted) Boltzmann machine can be considered as a special case of the mean field multi-species spin glass model [44, 60, 45]. Specifically the set of spins is arranged into a geometry made of consecutive layers and only interactions among spins belonging to adjacent layers are allowed. In particular intra-layer interactions are forbidden. Such architectural assumption makes it impossible to fulfill the positivity hypothesis under which the results of [44, 45] and the previous chapter were obtained. In fact, the positivity property requires dominant intra-group interactions with respect to inter-group ones. While the general deep (restricted) Boltzmann machine is still an unsolved problem (see nevertheless [55, 56, 26, 61, 57, 58, 62, 63] for centered Gaussian interactions), we present here its exact and rigorous solution on the Nishimori line [59]. In the previous chapter we have fully solved the elliptic multi-species model on the Nishimori line, where the property of replica symmetry, *i.e.* the concentration of the overlap, was shown to hold. Such property indeed is fully general on the Nishimori line, see [40] on this respect, and does not rely on any positivity assumption of the interactions. While the positivity properties carry with them the typical bounds of Guerra's method [7, 21], here the technical support to control and solve the model is based only on the Nishimori identities [54, 53, 29], among which (2.14)-(2.16), and correlation inequalities [42], such as (2.20)-(2.21). We hereby provide the first exact solution of a disordered Statistical Mechanics model in a deep architecture and describes how the relative size of the layers influences the phase transition. Furthermore, the model admits an inferential counter part as usual: it can be seen as a deep spatially coupled spiked Wigner model [47, 33] with K layers, which in the case $K = 2$ coincides with the Wishart model (rank-one non-symmetric matrix estimation [17]).

The chapter is organized as follows. In Section 3.1 we introduce the model and we present the main results in three theorems. In section 3.2.1 we decline the adaptive interpolation sum rule (2.30) in the multi-species setting to the one needed for the DBM. The proofs are contained in Section 3.2 and Section 3.3 collects some conclusions and perspectives.

3.1   Definitions and results

The setting is that of the multi-species model. We refer the reader to Section 2.1 for the details. In this chapter we call the disjoint sets of the partition *layers*, and denote them by $\{L_r\}_{r=1,\dots,K}$ with cardinality $|L_r| = N_r$, such that $\sum_{r=1}^{K} N_r = N$. We recall the form of the Hamiltonian for the reader's convenience:

$$H_N(\sigma) := -\sum_{r,s=1}^{K} \sum_{(i,j)\in L_r\times L_s} \tilde{J}_{ij}^{rs}\sigma_i\sigma_j - \sum_{r=1}^{K}\sum_{i\in L_r} \tilde{h}_i^r\sigma_i \tag{3.1}$$

where the interaction coefficients and the external fields are independent Gaussian random variables distributed as follows

$$\tilde{J}_{ij}^{rs} \overset{\text{iid}}{\sim} \mathcal{N}\left(\frac{\mu_{rs}}{2N}, \frac{\mu_{rs}}{2N}\right), \quad \tilde{h}_i^r \overset{\text{iid}}{\sim} \mathcal{N}(h_r, h_r). \tag{3.2}$$

The peculiarity of the DBM is that μ has the following tridiagonal structure:

$$\mu = \begin{pmatrix} 0 & \mu_{12} & 0 & \cdots & 0 \\ \mu_{21} & 0 & \mu_{23} & \cdots & 0 \\ 0 & \mu_{32} & 0 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \mu_{K-1,K} \\ 0 & 0 & 0 & \mu_{K,K-1} & 0 \end{pmatrix} \tag{3.3}$$

and is assumed to be symmetric without loss of generality. The geometrical architecture of the model is illustrated in Figure3.1. Besides the effective interaction matrix Δ (2.6) we introduce

$$M := (\mu_{rs}\alpha_s)_{r,s=1,\dots,K} \tag{3.4}$$

Notice that Δ and M are tridiagonal matrices too.

The first result of this chapter is the computation of the random pressure (2.8) in the thermodynamic limit.

Theorem 3.1 (Solution of the model). *The random pressure* (2.8) *of a K-layer deep Boltzmann machine on the Nishimori line converges almost surely in the thermodynamic limit and its value is given by a K-dimensional variational principle:*

$$\lim_{N\to\infty} p_N \overset{\text{a.s.}}{=} \lim_{N\to\infty} \bar{p}_N(\mu, \mathbf{h}) = \sup_{\mathbf{x}_o}\inf_{\mathbf{x}_e} p_{var}(\mathbf{x}; \mu, \mathbf{h}), \tag{3.5}$$

where $\mathbf{x}_o$ and $\mathbf{x}_e$ denote the vectors of the odd and even components of the order parameter $\mathbf{x} \in [0,1)^K$ respectively,

$$p_{var}(\mathbf{x}; \mu, \mathbf{h}) := \sum_{r=1}^{K} \alpha_r\,\psi\left((M\mathbf{x})_r + h_r\right) + \sum_{r=1}^{K} \frac{\Delta_{r,r+1}}{2}\left[(1-x_r)(1-x_{r+1}) - 2x_r x_{r+1}\right] \tag{3.6}$$
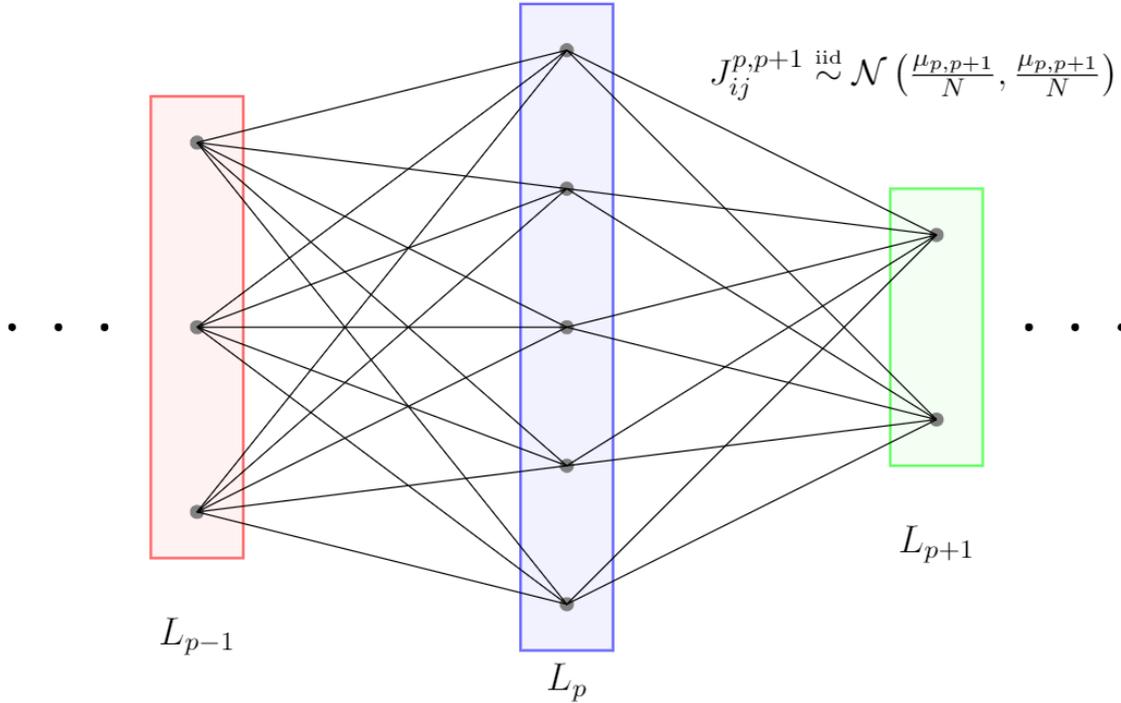
Figure 3.1: Graph of the interactions between layers.

and for any $x \geq 0$

$$\psi(x) := \mathbb{E}_z \log 2 \cosh \left(z\sqrt{x} + x \right) , \quad z \sim \mathcal{N}(0,1) . \tag{3.7}$$

Moreover, defining $\bar{\mathbf{x}}$ as the solution of the variational problem (3.5), we have

$$\lim_{N \to \infty} \mathbb{E} \langle q_r \rangle_N = \lim_{N \to \infty} \mathbb{E} \langle m_r \rangle_N = \bar{x}_r \tag{3.8}$$

for every $r = 1, \ldots, K$ and for all the points of the phase space $(\mu, \hat{\alpha}, \mathbf{h})$ where $\bar{x}$ is $\mathbf{h}$-differentiable and the matrix Δ is invertible.

The proof of Theorem 3.1 relies on the adaptive interpolation method [13] combined with the concentration result Lemma 2.5. The main difference with the convex case presented in the previous chapter is that the matrix Δ is not definite, indeed its eigenvalues have alternating signs. This entails that the remainder identified by interpolation has not a definite sign and cannot be discarded a priori at the expense of an inequality. Moreover, the concentration of the overlap strongly depends on the notion of regularity in Definition 2.2 of the path followed by the adaptive interpolation. Hence one has to carefully choose a path that is regular and allows also to exploit the convexities of the two sums involved in the functional (3.6).

Secondly, we focus on the properties of the consistency equation obtained from the optimization problem (3.5) when the matrix Δ is invertible, that is when K is even. The stability of the optimizers of (3.5) is a more delicate problem with respect to the convex multi-species case, due to the min-max nature of the variational principle. In the following, given a square matrix A we denote by $\rho(A)$ its spectral radius and by $A^{(eo)}$ the submatrix of A obtained by keeping only even rows and odd columns of A. An analogous definition is given for $A^{(oe)}, A^{(oo)}, A^{(ee)}$. Notice that, when K is even, $\Delta^{(eo)}$ is an upper triangular $K/2 \times K/2$ square matrix with non-zero diagonal elements and therefore it is invertible. Similar considerations hold for the sub-matrix $\Delta^{(oe)} = [\Delta^{(eo)}]^T$. We prove the following

Theorem 3.2. *Let K be even and $\mathbf{h} = 0$. If $\rho([M^2]^{(oo)}) < 1$ then $\mathbf{x} = 0$ is the unique solution to the variational problem (3.5). Conversely, if $\rho([M^2]^{(oo)}) > 1$ then the solution of (3.5) is a vector $\mathbf{x} = \bar{\mathbf{x}}(M)$ with strictly positive components satisfying the consistency equation:*

$$x_r = \mathbb{E}_z \tanh\left(z\sqrt{(M\mathbf{x})_r} + (M\mathbf{x})_r\right) \quad \forall\, r = 1, \ldots, K \tag{3.9}$$

where z denotes a standard Gaussian random variable.

The proof of Theorem 3.2 amounts to the computation of the Hessian matrix of an auxiliary function introduced later and in a check of its eigenvalues. The peculiar form of the consistency equations due to the structure (3.3) plays a central role. Theorem 3.2 implies the existence of a phase transition in our model localized at zero magnetic field and unitary spectral radius as discussed in Remark 3.1 below. The following Proposition further clarifies the structure of the phase transition and how the system's geometry, encoded in the form factors α_r's, can influence it.

Proposition 3.3. *For any given interaction matrix μ, we have*

$$\sup_{\alpha_1,\ldots,\alpha_K} \rho\left([M^2]^{(oo)}\right) = \frac{1}{4}\max_r \mu_{r,r+1}^2 \tag{3.10}$$

where the sup on the l.h.s. is taken over the form factors $\alpha_1, \ldots, \alpha_K \geq 0$, $\sum_{r=1}^{K} \alpha_r = 1$ and the max on the r.h.s. is taken over $r = 1, \ldots, K - 1$. Furthermore the sup on the l.h.s. of (3.10) is attained if and only if one of the following conditions is verified:

a) there exists $r^* \in \{1, \ldots, K - 1\}$ such that

$$\alpha_{r^*} = \alpha_{r^*+1} = \frac{1}{2} \quad, \quad \mu_{r^*,r^*+1} = \max_r \mu_{r,r+1} \,; \tag{3.11}$$

b) there exists $r^* \in \{2, \ldots, K - 1\}$ such that

$$\alpha_{r^*} = \alpha_{r^*-1} + \alpha_{r^*+1} = \frac{1}{2} \quad, \quad \mu_{r^*-1,r^*} = \mu_{r^*,r^*+1} = \max_r \mu_{r,r+1}\,. \tag{3.12}$$

Remark 3.1. For even K, Proposition 3.3 together with Theorems 3.1 and 3.2 show that if the interaction strengths $\mu_{r,r+1} < 2$ for all $r = 1, \ldots, K-1$, then the magnetisations and the overlaps vanish as $N \to \infty$ for every choice of the form factors $(\alpha_1, \ldots, \alpha_K) \in (0,1)^K$. By Theorem 3.2 $\bar{\mathbf{x}}$ is not identically zero on the space of parameters $(\mu, \hat{\alpha})$, hence the limiting quenched pressure (3.5) cannot be an analytic function.

Proposition 3.3 also shows that as soon as $\mu_{r,r+1} > 2$ for some $r = 1, \ldots, K-1$, then, by suitably localizing only two extensive layers near the maximal interaction (condition (3.11)), their magnetisations and overlaps turn out to be positive in the limit $N \to \infty$.

Finally, we prove a uniqueness result that holds for arbitrary spectral radius.

Theorem 3.4. *Let $h_r > 0 \ \forall\, r = 1, \ldots, K$. The consistency equation*

$$x_r = \mathbb{E}_z \tanh\left(z\sqrt{(M\mathbf{x})_r + h_r} + (M\mathbf{x})_r + h_r\right) \quad \forall\, r = 1, \ldots, K \tag{3.13}$$

admits a unique solution $\mathbf{x} = \bar{\mathbf{x}}(M, \mathbf{h}) \in (0,1)^K$.

3.2   Proofs

We start with some simple lemmas. The convexity of ψ will be crucial in the proof of Theorem 3.1. In particular, we recall that

Lemma 3.5. *The function*

$$f(\mathbf{x}) := \sum_{r=1}^{K} \alpha_r\, \psi((M\mathbf{x})_r) \tag{3.14}$$

is convex for $\mathbf{x}$ such that $M\mathbf{x} \geq 0$ component-wise.

Proof. ψ is convex on $\mathbb{R}_{\geq 0}$ by equation (2.51). Then, using the linearity of $(M\mathbf{x})_r$, it is easy to verify that for any $\lambda \in [0,1]$ and $\mathbf{x}_1, \mathbf{x}_2 \in A$ we have:

$$f(\lambda \mathbf{x}_1 + (1-\lambda)\mathbf{x}_2) \leq \lambda f(\mathbf{x}_1) + (1-\lambda)f(\mathbf{x}_2) . \tag{3.15}$$

$\square$

In addition, for Theorem 3.4 we also need:

Lemma 3.6. *Let z be a standard Gaussian random variable. The function*

$$F(h) := \mathbb{E} \tanh\left(z\sqrt{h} + h\right) \tag{3.16}$$

is strictly positive, increasing and concave for $h > 0$.

Proof. Since $F = 2\psi' - 1$, positivity and monotonicity follow from equations (2.51). The concavity instead follows from the sign of the third derivative of ψ can be obtained avoiding Gaussian integration by parts. Indeed by setting $y = z\sqrt{h} + h$, replacing $\frac{z}{2\sqrt{h}} + 1 = \frac{y+h}{2h}$ in the computations and using the identities (2.14) for $n = 2, 3$ and (2.51) one finds:

$$\psi'''(h) = -\frac{1}{h} \mathbb{E}\left[\left(1 - \tanh^2 y\right)^2 y \tanh y\right] - \mathbb{E}\left[\left(1 - \tanh^2 y\right)^2 \tanh^2 y\right] < 0 . \tag{3.17}$$

$$\square$$

Remark 3.2. As a consequence the function F is invertible on $[0, \infty)$. Its inverse F^{-1} is non negative and increasing on $[0, 1)$. Moreover one has

$$\lim_{x \to 1^-} F^{-1}(x) = +\infty . \tag{3.18}$$

3.2.1 Specializing the interpolation

The interpolating model is the same as the one in Definition 2.1, and yields exactly the same sum rule (2.30). However, the tridiagonal form of Δ allows us to specialize the latter as follows:

$$\bar{p}_N(\mu, \mathbf{h}) = \mathcal{O}(s_N) + \sum_{r=1}^{K} \alpha_r \psi\left(Q_{\epsilon,r}(1) + h_r\right) +$$

$$+ \sum_{r=1}^{K} \frac{\Delta_{r,r+1}}{2} \int_0^1 dt \, [(1 - q_{\epsilon,r}(t))(1 - q_{\epsilon,r+1}(t)) - 2q_{\epsilon,r}(t)q_{\epsilon,r+1}(t)] + \tag{3.19}$$

$$+ \sum_{r=1}^{K} \frac{\Delta_{r,r+1}}{2} \int_0^1 dt \, \mathbb{E}\big\langle (m_r - q_{\epsilon,r}(t)) \, (m_{r+1} - q_{\epsilon,r+1}(t)) \big\rangle_{N,t}^{(\epsilon)} ,$$

or better, using the notation introduced for Theorem 3.2,

$$\bar{p}_N(\mu, \mathbf{h}) = \mathcal{O}(s_N) + \sum_{r=1}^{K} \alpha_r \psi\left(Q_{\epsilon,r}(1) + h_r\right) +$$

$$+ \frac{1}{2} \int_0^1 dt \, \left[(\mathbf{1}_o - \mathbf{q}_{\epsilon,o}(t), \Delta^{(oe)}(\mathbf{1}_e - \mathbf{q}_{\epsilon,e}(t))) - 2(\mathbf{q}_{\epsilon,o}(t), \Delta^{(oe)}\mathbf{q}_{\epsilon,e}(t))\right] + \tag{3.20}$$

$$+ \frac{1}{2} \int_0^1 dt \, \mathbb{E}\big\langle (\mathbf{m}_o - \mathbf{q}_{\epsilon,o}(t), \Delta^{(oe)}(\mathbf{m}_e - \mathbf{q}_{\epsilon,e}(t))) \big\rangle_{N,t}^{(\epsilon)} ,$$

where again the subscripts o, e denote the odd or even components of a vector, $\mathbf{1} := (1)_{r=1,\dots,K}$. We also denote

$$\mathbf{Q}_{\epsilon,o}(t) = \boldsymbol{\epsilon}_o + M^{(oe)} \int_0^t \mathbf{q}_{\epsilon,e}(s) \, ds , \quad \mathbf{Q}_{\epsilon,e}(t) = \boldsymbol{\epsilon}_e + M^{(eo)} \int_0^t \mathbf{q}_{\epsilon,o}(s) \, ds . \tag{3.21}$$

The sum rules (3.19), (3.20) motivate the definition of the variational pressure (3.6) that for future convenience can be rewritten as:

$$p_{var}(\mathbf{x}; \mu, \mathbf{h}) = \sum_{r=1}^{K} \alpha_r \, \psi\left((M\mathbf{x})_r + h_r\right) + \frac{(\mathbf{1}_o - \mathbf{x}_o, \Delta^{(oe)}(\mathbf{1}_e - \mathbf{x}_e))}{2} - (\mathbf{x}_o, \Delta^{(oe)}\mathbf{x}_e). \quad (3.22)$$

Remark 3.3. The variational function p_{var} is convex in the even components $\mathbf{x}_e$ and the odd components $\mathbf{x}_o$ separately. This is due to the fact that the two bilinear forms in (3.22) have vanishing second derivatives w.r.t. pure odd or even components, while the terms containing ψ are convex by Lemma 3.5.

3.2.2 Proof of Theorem 3.1

The almost sure equality in (3.5) holds regardless of the convexity of the problem, thanks to Proposition 2.1.

Lower bound. We select a path contained in $[0,1)^K$ by means of the following coupled ODEs

$$\dot{\mathbf{Q}}_{\epsilon,e}(t) = M^{(eo)}\mathbf{x}_o =: \mathbf{f}_e(t, \mathbf{Q}_\epsilon(t)), \quad \mathbf{Q}_{\epsilon,e}(0) = \boldsymbol{\epsilon}_e \quad (3.23)$$

$$\dot{\mathbf{Q}}_{\epsilon,o}(t) = M^{(oe)}\mathbb{E}\langle\mathbf{m}_e\rangle_t =: \mathbf{f}_o(t, \mathbf{Q}_\epsilon(t)), \quad \mathbf{Q}_{\epsilon,o}(0) = \boldsymbol{\epsilon}_o, \quad (3.24)$$

where $\mathbf{f}(t, \mathbf{Q})$ is the velocity field of the ODE. The perturbation is here introduced as an initial condition in order to have the interpolating functions in the form (3.21). Notice that $\mathbf{f}_e$ is constant, while $\mathbf{f}_o$ is a positive Lipschitz function of $\mathbf{Q}_\epsilon(t) \in (0, \infty)^K$ thanks to identity (2.21) (where N is fixed). Therefore, by Cauchy-Lipschitz's theorem, the system of ODEs (3.23)-(3.24) has a unique global solution $\mathbf{Q}_\epsilon(t)$, $t \in [0, 1]$, whose components are positive.

By (3.23)-(3.24) we have $\Delta^{(eo)}\mathbf{q}_{\epsilon,o}(t) = \Delta^{(eo)}\mathbf{x}_o$ and $\Delta^{(oe)}\mathbf{q}_{\epsilon,e}(t) = \Delta^{(oe)}\mathbb{E}\langle\mathbf{m}_e\rangle_t$, hence:

$$\int_0^1 dt \, (\mathbf{1}_o - \mathbf{q}_{\epsilon,o}(t), \Delta^{(oe)}(\mathbf{1}_e - \mathbf{q}_{\epsilon,e}(t))) = \left(\mathbf{1}_o - \mathbf{x}_o, \Delta^{(oe)}\left(\mathbf{1}_e - \int_0^1 dt \, \mathbb{E}\langle\mathbf{m}_e\rangle_t\right)\right) \quad (3.25)$$

and reasoning in a similar way for the other t-integrations appearing in the sum rule (3.20) we obtain:

$$\bar{p}_N = \mathcal{O}(s_N) + p_{var}\left(\mathbf{x}_o, \int_0^1 dt \, \mathbb{E}\langle\mathbf{m}_e\rangle_t\right) + \int_0^1 dt \, R_\epsilon(t) \geq$$
$$\geq \mathcal{O}(s_N) + \inf_{\mathbf{x}_e} p_{var}(\mathbf{x}_o, \mathbf{x}_e) + \int_0^1 dt \, R_\epsilon(t), \quad (3.26)$$

where the reminder is

$$R_\epsilon(t) = \frac{1}{2}\mathbb{E}\langle\left((\mathbf{m}_o - \mathbf{x}_o), \Delta^{(oe)}(\mathbf{m}_e - \mathbb{E}\langle\mathbf{m}_e\rangle_t)\right)\rangle_t. \quad (3.27)$$

Using Cauchy-Schwartz's inequality,

$$|R_\epsilon(t)| \leq \frac{1}{2} \|\mu^{(oe)}\| \, \mathbb{E}^{1/2} \langle |\hat{\alpha}^{(oo)}(\mathbf{m}_o - \mathbf{x}_o)|^2 \rangle_t \, \mathbb{E}^{1/2} \langle |\hat{\alpha}^{(ee)}(\mathbf{m}_e - \mathbb{E}\langle \mathbf{m}_e \rangle)\rangle_t|^2 \rangle_t \, , \qquad (3.28)$$

thus, provided that the map $\boldsymbol{\epsilon} \mapsto \mathbf{Q}_\epsilon(t)$ is regular, the remainder $R_\epsilon(t)$ vanishes in $\boldsymbol{\epsilon}$-average as $N \to \infty$ by Lemma 2.5. To show that $\mathbf{Q}_\epsilon$ is regular we introduce the following matrix fields:

$$\Phi_\epsilon(t) := \frac{\partial \mathbf{Q}_\epsilon(t)}{\partial \boldsymbol{\epsilon}} \, , \quad A_\epsilon(t) := \frac{\partial \mathbf{f}(t, \mathbf{Q}_\epsilon(t))}{\partial \mathbf{Q}_\epsilon(t)} \qquad (3.29)$$

Applying the chain rule we have:

$$\dot{\Phi}_\epsilon(t) = \frac{\partial \dot{\mathbf{Q}}_\epsilon(t)}{\partial \boldsymbol{\epsilon}} = A_\epsilon(t) \, \Phi_\epsilon(t) \, , \quad \Phi_\epsilon(0) = \mathbb{1} \, , \qquad (3.30)$$

hence, by Liouville's formula (2.4) the Jacobian $\det(\Phi_\epsilon(t))$ is

$$\det\left(\frac{\partial \mathbf{Q}_\epsilon}{\partial \boldsymbol{\epsilon}}(t)\right) = \exp\left\{\int_0^t ds \, \mathrm{Tr}\big(A_\epsilon(s)\big)\right\} . \qquad (3.31)$$

Now, using equations (3.23)-(3.24) one can compute:

$$\begin{aligned}
\mathrm{Tr}\big(A_\epsilon(t)\big) &= \sum_{r=1}^K \big(A_\epsilon(t)\big)_{r,r} = \sum_{r \text{ odd}} \frac{\partial\big(M^{(oe)} \mathbb{E}\langle \mathbf{m}_e \rangle_t\big)_r}{\partial Q_{\epsilon,r}(t)} = \\
&= \sum_{r \text{ odd}} \sum_{r' \text{even}} M_{rr'} \frac{\partial \mathbb{E}\langle m_{r'} \rangle_t}{\partial Q_{\epsilon,r}(t)} \geq 0
\end{aligned} \qquad (3.32)$$

where non-negativity is a consequence of the correlation inequality (2.21), since $Q_{\epsilon,r}(t)$ can be seen as the variance of an external field on the Nishimori line in the interpolating Hamiltonian (2.23). Combining (3.31) and (3.32), it follows that $\mathbf{Q}_\epsilon$ is regular, as desired (see Definition 2.2).

Now, averaging on $\boldsymbol{\epsilon}$ (we consider it uniform over $[s_N, 2s_N]^K$) and tanking the $\liminf_{N\to\infty}$ in inequality (3.26) we have

$$\liminf_{N\to\infty} \bar{p}_N \geq \inf_{\mathbf{x}_e} p_{var}(\mathbf{x}_o, \mathbf{x}_e) + \liminf_{N\to\infty} \mathbb{E}_\epsilon \int_0^1 dt \, R_\epsilon(t) . \qquad (3.33)$$

The last term vanishes by Fubini's theorem, dominated convergence and Lemma 2.5. Finally, optimizing w.r.t. $\mathbf{x}_o$ we get:

$$\liminf_{N\to\infty} \bar{p}_N \geq \sup_{\mathbf{x}_o} \inf_{\mathbf{x}_e} p_{var}(\mathbf{x}_o, \mathbf{x}_e) . \qquad (3.34)$$

Upper bound.   Now, we set

$$\dot{\mathbf{Q}}_{\epsilon,e}(t) = M^{(eo)}F(M^{(oe)}\mathbb{E}\langle\mathbf{m}_e\rangle_t + \mathbf{h}_o), \quad \mathbf{Q}_{\epsilon,e}(0) = \boldsymbol{\epsilon}_e \tag{3.35}$$

$$\dot{\mathbf{Q}}_{\epsilon,o}(t) = M^{(oe)}\mathbb{E}\langle\mathbf{m}_e\rangle_t, \quad \mathbf{Q}_{\epsilon,o}(0) = \boldsymbol{\epsilon}_o. \tag{3.36}$$

In (3.35) the application of F, defined in (3.16), to the vector $M^{(oe)}\mathbb{E}\langle\mathbf{m}_e\rangle_t + \mathbf{h}_o$ has to be understood as component-wise. For future convenience let us set

$$\mathcal{D}(\mathbf{x},\mathbf{h}) := \operatorname{diag}\left\{F'\left((M\mathbf{x})_r + h_r\right)\right\}_{r=1,\ldots,K}. \tag{3.37}$$

With a slight abuse of notation we will stress the dependence of $\mathcal{D}^{(oo)}(\mathbf{x},\mathbf{h})$ and $\mathcal{D}^{(ee)}(\mathbf{x},\mathbf{h})$ on the even and odd components of $\mathbf{x}$ respectively as follows

$$\mathcal{D}^{(oo)}(\mathbf{x},\mathbf{h}) \equiv \mathcal{D}^{(oo)}(\mathbf{x}_e,\mathbf{h}), \quad \mathcal{D}^{(ee)}(\mathbf{x},\mathbf{h}) \equiv \mathcal{D}^{(ee)}(\mathbf{x}_o,\mathbf{h}). \tag{3.38}$$

$M^{(eo)}F(M^{(oe)}\mathbb{E}\langle\mathbf{m}_e\rangle_t + \mathbf{h}_o)$ is a positive function of $\mathbf{Q}_\epsilon(t)$ with bounded derivatives for fixed N thanks to Lemma 3.6, indeed

$$\frac{\partial}{\partial Q_{\epsilon,r}}F(M^{(oe)}\mathbb{E}\langle\mathbf{m}_e\rangle_t + \mathbf{h}_o) = \mathcal{D}(\mathbb{E}\langle\mathbf{m}_e\rangle_t,\mathbf{h})^{(oo)}M^{(oe)}\frac{\partial\mathbb{E}\langle\mathbf{m}_e\rangle_t}{\partial Q_{\epsilon,r}}. \tag{3.39}$$

This ensures the existence of a unique global solution over $[0,1]$ to the system of ODEs (3.35)-(3.36). Moreover, the latter implies also that the map $\boldsymbol{\epsilon} \longmapsto \mathbf{Q}_\epsilon(\cdot)$ is still regular, because F' is positive as proved in Lemma 3.6 and $\frac{\partial\mathbb{E}\langle\mathbf{m}_e\rangle_t}{\partial Q_{\epsilon,r}} \geq 0$ thanks again to (2.21). This guarantees the positivity of the trace in (3.31) and forces the remainder R_ϵ in $\boldsymbol{\epsilon}$-average to vanish by Lemma 2.5. Using Jensen's inequality, by the convexity of ψ we have

$$\sum_{r=1}^{K}\alpha_r\psi\left(\left(M\int_0^1\mathbf{q}_\epsilon(t)\,dt + \mathbf{h}\right)_r\right) \leq \sum_{r=1}^{K}\alpha_r\int_0^1\psi\left((M\mathbf{q}_\epsilon(t) + \mathbf{h})_r\right)dt \tag{3.40}$$

and inserting it into the sum rule (3.20) we get

$$\bar{p}_N \leq \mathcal{O}(s_N) + \int_0^1 dt\, p_{var}(\mathbf{F}_{\epsilon,o}(t),\mathbb{E}\langle\mathbf{m}_e\rangle_t) + \int_0^1 R_\epsilon(t)\,dt =$$

$$= \mathcal{O}(s_N) + \int_0^1 dt\,\inf_{\mathbf{x}_e} p_{var}(\mathbf{F}_{\epsilon,o}(t),\mathbf{x}_e) + \int_0^1 R_\epsilon(t)\,dt, \tag{3.41}$$

where $\mathbf{F}_{\epsilon,o}(t) := F(M^{(oe)}\mathbb{E}\langle\mathbf{m}_e\rangle_t + \mathbf{h}_o)$ for brevity. As far as the last equality is concerned, we used the following:

$$\inf_{\mathbf{x}_e} p_{var}(\mathbf{F}_{\epsilon,o}(t),\mathbf{x}_e) = p_{var}(\mathbf{F}_{\epsilon,o}(t),\mathbb{E}\langle\mathbf{m}_e\rangle_t). \tag{3.42}$$

This is a consequence of the convexity of p_{var} in $\mathbf{x}_e$ (see Remark 3.3). In fact, a computation of the gradient of p_{var} w.r.t. $\mathbf{x}_e$ evaluated at $\mathbb{E}\langle \mathbf{m}_e \rangle_t$ yields:

$$\frac{\partial p_{var}}{\partial \mathbf{x}_e}(\mathbf{F}_{\epsilon,o}(t), \mathbf{x}_e)\Big|_{\mathbb{E}\langle \mathbf{m}_e \rangle_t} = \frac{\Delta^{(eo)}}{2}[\mathbf{1}_o + \mathbf{F}_{\epsilon,o}(t)]+$$

$$+ \frac{\Delta^{(eo)}}{2}[-\mathbf{1}_o + \mathbf{F}_{\epsilon,o}(t)] - \Delta^{(eo)}\mathbf{F}_{\epsilon,o}(t) = 0 , \quad (3.43)$$

where we explicitly notice that the first term comes from the derivative of ψ (2.51). Then, taking the sup of p_{var} over the odd components and the $\boldsymbol{\epsilon}$-average we get:

$$\bar{p}_N \leq \mathcal{O}(s_N) + \sup_{\mathbf{x}_o} \inf_{\mathbf{x}_e} p_{var}(\mathbf{x}_o, \mathbf{x}_e) + \mathbb{E}_{\boldsymbol{\epsilon}} \int_0^1 R_{\epsilon}(t)\, dt . \quad (3.44)$$

Applying Lemma 2.5, Fubini's theorem and dominated convergence the two bounds match after sending $N \to \infty$.

Proof of (3.8). Equations (2.20) and (2.21) imply that the quenched pressure is convex in each h_r. Hence it is possible to exchange the derivative w.r.t. h_r in (2.20) with the $N \to \infty$ limit where $\bar{\mathbf{x}}$ is differentiable in h_r (see Lemma IV.6.3 in [64]). Since for invertible Δ the optimal order parameter must be a critical point of p_{var} (see Proposition 3.7 below) by (2.51) and (3.6) we have that:

$$\lim_{N\to\infty} \frac{\partial \bar{p}_N}{\partial h_r} = \frac{\partial}{\partial h_r} p_{var}(\bar{\mathbf{x}}(M, \mathbf{h}); \mu, \mathbf{h}) = \frac{\partial p_{var}}{\partial \mathbf{x}}\Big|_{\bar{\mathbf{x}}(M,\mathbf{h})} \frac{\partial \bar{\mathbf{x}}(M,\mathbf{h})}{\partial h_r} + \frac{\partial p_{var}}{\partial h_r} = \frac{\partial p_{var}}{\partial h_r} =$$

$$= \alpha_r \, \psi'((M\bar{\mathbf{x}}(M, \mathbf{h}))_r + h_r) = \frac{\alpha_r}{2}\left[1 + \mathbb{E}_z \tanh\left(z\sqrt{(M\bar{\mathbf{x}})_r + h_r} + (M\bar{\mathbf{x}})_r + h_r\right)\right] =$$

$$= \frac{\alpha_r}{2}[1 + \bar{x}_r] . \quad (3.45)$$

A comparison with (2.20) and the Nishimori identity (2.14) lead to the identification:

$$\lim_{N\to\infty} \mathbb{E}\langle q_r \rangle_N = \lim_{N\to\infty} \mathbb{E}\langle m_r \rangle_N = \bar{x}_r . \quad (3.46)$$

Remark 3.4. Assume for now that K is even. Observe that the entire proof could have been carried out also by computing all the $\inf_{\mathbf{x}_e}$ over the convex set:

$$A := \{\mathbf{x}_e \mid M^{(oe)}\mathbf{x}_e + \mathbf{h}_o \geq 0 \text{ component-wise}\} \supseteq [0,1)^{K/2} , \quad (3.47)$$

on which all the functions involved are still real and well defined. This freedom is essentially due to the convexity of p_{var} in $\mathbf{x}_e$. Indeed, p_{var} has always a critical point in the domain A for

any fixed $\mathbf{x}_o \in [0, 1)^{K/2}$, that must coincide with its minimum point by convexity as can be seen by direct inspection

$$\frac{\partial p_{var}}{\partial \mathbf{x}_e}\bigg|_{\bar{\mathbf{x}}_e} = \frac{\Delta^{(eo)}}{2}\left[-\mathbf{x}_o + F(M^{(oe)}\bar{\mathbf{x}}_e + \mathbf{h}_o)\right] = 0 \iff \bar{\mathbf{x}}_e = [M^{(oe)}]^{-1}(F^{-1}(\mathbf{x}_o) - \mathbf{h}_o). \quad (3.48)$$

The inequality (3.26), that leads to the lower bound, clearly holds also for $\mathbf{x}_e \in A \supseteq [0, 1)^{K/2}$. The validity of (3.42) is less trivial and is due to the special choice $\mathbf{F}_{\epsilon,o}(t)$. In this case in fact, the critical point falls inside $[0, 1)^{K/2}$ and this lets us extend the domain of $\mathbf{x}_e$ to A without any loss of generality thanks to the mentioned convexity in $\mathbf{x}_e$. We will see later that even with this extension the point that realizes the sup inf lies inside the cube $[0, 1)^K$.

3.2.3   Proof of Theorem 3.2

For this proof we rely on Remark 3.4, this will ease our computations. Let us write the gradient of (3.6)

$$\frac{\partial p_{var}(\mathbf{x}; \mu, \mathbf{h})}{\partial x_r} = \left(\frac{\Delta}{2}(-\mathbf{x} + F(M\mathbf{x} + \mathbf{h}))\right)_r =$$

$$= \frac{\Delta_{r,r+1}}{2}\left[-x_{r+1} + F((M\mathbf{x})_{r+1} + h_{r+1})\right] + \frac{\Delta_{r,r-1}}{2}\left[-x_{r-1} + F((M\mathbf{x})_{r-1} + h_{r-1})\right]. \quad (3.49)$$

where we have used (2.14) and (2.51). In absence of external magnetic field ($\mathbf{h} = 0$) $\mathbf{x} = 0$ is a critical point for p_{var}, namely a solution to the consistency equation obtained by equating (3.49) to 0.

First of all, by Remark 3.3 and Remark 3.4 we infer that the optimization w.r.t. the even components $\mathbf{x}_e$ is always stable, in the sense that there is always one optimizer once the odd components $\mathbf{x}_o$ are fixed and it belongs to A. Define now the auxiliary function:

$$\pi(\mathbf{x}_o; \mu, \mathbf{h}) := \inf_{\mathbf{x}_e \in A} p_{var}(\mathbf{x}_o, \mathbf{x}_e; \mu, \mathbf{h}) = p_{var}(\mathbf{x}_o, \bar{\mathbf{x}}_e; \mu, \mathbf{h}), \quad (3.50)$$

with $\bar{\mathbf{x}}_e$ defined in (3.48). The following proposition investigates the possibility to have boundary solutions to the variational problem.

Proposition 3.7. *Let K be even. The points $\mathbf{x}_o$ at which the $\sup_{\mathbf{x}_o} \pi(\mathbf{x}_o; \mu, \mathbf{h})$ is attained fulfill the consistency equation:*

$$\bar{\mathbf{x}}_e = F(M^{(eo)}\mathbf{x}_o + \mathbf{h}_e). \quad (3.51)$$

As a consequence the necessary condition for $\mathbf{x}$ to realize the $\sup_{\mathbf{x}_o} \inf_{\mathbf{x}_e} p_{var}(\mathbf{x}_o, \mathbf{x}_e; \mu, \mathbf{h})$ is to be a critical point, namely to satisfy (3.51).

Proof. Using (3.48), the gradient of π is:

$$
\frac{\partial \pi(\mathbf{x}_o; \mu, \mathbf{h})}{\partial \mathbf{x}_o} = \frac{\partial p_{var}}{\partial \mathbf{x}_o} + \frac{\partial p_{var}}{\partial \mathbf{x}_e}\bigg|_{\bar{\mathbf{x}}_e} \frac{\partial \bar{\mathbf{x}}_e}{\partial \mathbf{x}_o} = \frac{\Delta^{(oe)}}{2} \left[-\bar{\mathbf{x}}_e + F(M^{(eo)}\mathbf{x}_o + \mathbf{h}_e) \right] =
$$

$$
= \frac{\hat{\alpha}^{(oo)}}{2} \left[-F^{-1}(\mathbf{x}_o) + \mathbf{h}_o + M^{(oe)} F(M^{(eo)}\mathbf{x}_o + \mathbf{h}_e) \right] . \quad (3.52)
$$

We start by considering the case $\mathbf{h} = 0$. One can immediately rule out the possibility that the sup is attained at the right border, *i.e.* $x_{2l-1} \to 1^-$ for some l, because thanks to (3.18) $\partial_{x_{2l-1}}\pi \to -\infty$. Then, the necessary condition for a point $\mathbf{x}_o \in [0,1)^{K/2}$ to realize the sup is that:

$$
-F^{-1}(\mathbf{x}_o) + M^{(oe)} F(M^{(eo)}\mathbf{x}_o) \leq 0 , \quad (3.53)
$$

component-wise, where equality holds for those components for which $x_{2l-1} > 0$. If we set $M_{0,1} = M_{K,K+1} = 0$, the generic $2l-1$ component of the previous is given by

$$
- F^{-1}(x_{2l-1}) + M_{2l-1,2l-2}F(M_{2l-2,2l-3}x_{2l-3} + M_{2l-2,2l-1}x_{2l-1})+
$$
$$
+ M_{2l-1,2l}F(M_{2l,2l-1}x_{2l-1} + M_{2l,2l+1}x_{2l+1}) \quad (3.54)
$$

whence we understand that if $x_{2l-1} = 0$ the only chance for the previous to be non positive is to have also $x_{2l-3} = x_{2l+1} = 0$ because F is positive and monotonic. On the contrary, if $x_{2l-1} > 0$ first the corresponding gradient component must vanish; second by looking at the $2l+1$ component for instance

$$
- F^{-1}(x_{2l+1}) + M_{2l+1,2l+2}F(M_{2l+2,2l+3}x_{2l+3} + M_{2l+2,2l+1}x_{2l+1})+
$$
$$
+ M_{2l+1,2l}F(M_{2l,2l+1}x_{2l+1} + M_{2l,2l-1}x_{2l-1}) \quad (3.55)
$$

we see that the last term is strictly positive. Necessarily, x_{2l+1} must be strictly positive too with the corresponding gradient component that vanishes, and so on. Similar considerations hold for x_{2l-3}. Finally, iterating these arguments, we infer that the supremum is attained at a point $\mathbf{x}_o$ such that:

$$
\mathbf{x}_o = 0 \quad \text{or} \quad \bar{\mathbf{x}}_e = F(M^{(eo)}\mathbf{x}_o) . \quad (3.56)
$$

The first in particular implies that also $\bar{\mathbf{x}}_e = 0$. In both cases we can say that (3.51) is satisfied.

When any h_r is strictly positive it is immediate to see that there is a component of (3.52) with a positive contribution, the corresponding component of $\mathbf{x}_o$ must then be positive. Therefore one iterates the same arguments as above obtaining again (3.51). In any case, by (3.48) the sup inf is attained at critical points of p_{var}. $\qquad\square$

The Jacobian matrix of $F(M\mathbf{x} + \mathbf{h})$ is

$$DF(M\mathbf{x} + \mathbf{h}) = \mathcal{D}(\mathbf{x}, \mathbf{h})M . \tag{3.57}$$

Thanks to the convexity of ψ, $\mathcal{D}(\mathbf{x}, \mathbf{h})$, defined in (3.37), is diagonal, positive definite, invertible and its spectral radius is bounded by 1. From (3.52), an application of the Inverse Function Theorem leads to the Hessian matrix

$$\mathscr{H}_{\mathbf{x}_o}\pi = \frac{\Delta^{(oe)}}{2}\left[-\frac{\partial \bar{\mathbf{x}}_e}{\partial \mathbf{x}_o} + \frac{\partial}{\partial \mathbf{x}_o}F(M^{(eo)}\mathbf{x}_o + \mathbf{h_e}) \right] =$$
$$\frac{\Delta^{(oe)}}{2}\left[-[M^{(oe)}]^{-1}[\mathcal{D}^{(oo)}(\bar{\mathbf{x}}_e, \mathbf{h})]^{-1} + \mathcal{D}^{(ee)}(\mathbf{x}_o, \mathbf{h})M^{(eo)} \right] . \tag{3.58}$$

Thanks to the peculiar tridiagonal form of M we also have that

$$[\mathcal{D}(\mathbf{x}, \mathbf{h})M]^{(oe)}[\mathcal{D}(\mathbf{x}, \mathbf{h})M]^{(eo)} = [(\mathcal{D}(\mathbf{x}, \mathbf{h})M)^2]^{(oo)} , \tag{3.59}$$

from which by a simple rearrangement we can write the Hessian in its final form

$$\mathscr{H}_{\mathbf{x}_o}\pi = \frac{\hat{\alpha}^{(oo)}[\mathcal{D}(\bar{\mathbf{x}}_e, \mathbf{h})^{(oo)}]^{-1}}{2}\left[-\mathbb{1} + (\mathcal{D}(\mathbf{x}, \mathbf{h})M)^2 \right]^{(oo)} =$$
$$= \frac{[\alpha^{(oo)}]^{1/2}[\mathcal{D}^{(oo)}]^{-1/2}}{2}\left[-\mathbb{1} + S^{(oo)} \right][\alpha^{(oo)}]^{1/2}[\mathcal{D}^{(oo)}]^{-1/2} \tag{3.60}$$

with

$$S^{(oo)} := [\mathcal{D}^{(oo)}]^{1/2}[\hat{\alpha}^{(oo)}]^{-1/2}\Delta^{(oe)}\mathcal{D}^{(ee)}[\hat{\alpha}^{(ee)}]^{-1}\Delta^{(eo)}[\hat{\alpha}^{(oo)}]^{-1/2}[\mathcal{D}^{(oo)}]^{1/2} \tag{3.61}$$

where for brevity we have neglected all the dependencies after the second equality in (3.60) and used (3.59). (3.60) uses only symmetric matrices in order to make manifest the global sign of the Hessian. It remains to show that the spectral radius of $S^{(oo)}$ is controlled by that of $[M^2]^{(oo)}$. $S^{(oo)}$ is symmetric because $\Delta^{(oe)} = [\Delta^{(eo)}]^T$, thus its spectral radius coincides with the matrix norm induced by the Euclidean scalar product. Then by norms sub-multiplicativity and matrix similarity one easily gets

$$\rho\left(S^{(oo)}\right) \leq \rho\left(\mathcal{D}^{(oo)}\right)\rho\left([\hat{\alpha}^{(oo)}]^{-1/2}\Delta^{(oe)}\mathcal{D}^{(ee)}[\hat{\alpha}^{(ee)}]^{-1}\Delta^{(eo)}[\hat{\alpha}^{(oo)}]^{-1/2}\right) \leq$$
$$\leq \rho\left(M^{(oe)}\mathcal{D}^{(ee)}M^{(eo)}\right) = \rho\left(\mathcal{D}^{(ee)}M^{(eo)}M^{(oe)}\right) =$$
$$= \rho\left([\mathcal{D}^{(ee)}]^{1/2}[\hat{\alpha}^{(ee)}]^{-1/2}\Delta^{(eo)}\hat{\alpha}^{(oo)-1}\Delta^{(oe)}[\hat{\alpha}^{(ee)}]^{-1/2}[\mathcal{D}^{(ee)}]^{1/2}\right) . \tag{3.62}$$

Iterating the same arguments we get to

$$\rho\left(S^{(oo)}\right) \leq \rho\left(M^{(eo)}M^{(oe)}\right) = \rho\left(M^{(oe)}M^{(eo)}\right) < 1 , \tag{3.63}$$

where the last equality follows again by matrix similarity. The previous implies that the matrix $[-\mathbb{1} + S]^{(oo)}$ in (3.60) is negative definite, making $\mathcal{H}_{\mathbf{x}_o}\pi$ negative definite too, and hence π is concave under the hypothesis $\rho([M^2]^{(oo)}) < 1$. In turn, this ensures uniqueness of the solution to the consistency equation (3.51) and to the variational problem (3.5). In particular when $\mathbf{h} = 0$, $\mathbf{x} = 0$ is the unique solution.

Conversely, for $\mathbf{h} = 0$ and $\rho([M^2]^{(oo)}) > 1$ the Hessian has at least one positive eigenvalue at the origin $\mathbf{x}_o = 0$, but this is in general not enough to ensure $\mathbf{x}_o = 0$ does not realize the sup anymore. One has to check that there is a direction of increment of π that intersects the cube $[0, 1)^{K/2}$, otherwise the system could remain stuck on the border at $\mathbf{x}_o = 0$ due to the positivity constraints on the variables.

It is easy to see that $[M^2]^{(oo)}$ is irreducible, because its associated graph is strongly connected, and it has non negative entries. Hence, by Perron-Frobenius Theorem the eigenvector $\mathbf{v}$ relative to the largest eigenvalue $\rho([M^2]^{(oo)})$ is component-wise strictly positive, thus pointing inside the cube, and by a Taylor expansion around $\mathbf{x}_o = 0$ we have:

$$\pi(\epsilon\mathbf{v}; \mu, 0) - \pi(0; \mu, 0) = \frac{\epsilon^2}{2}\left(\mathbf{v}, \frac{\hat{\alpha}^{(oo)}}{2}\mathbf{v}\right)\left[-1 + \rho([M^2]^{(oo)})\right] + o(\epsilon^2) \tag{3.64}$$

that is positive form small enough $\epsilon > 0$. Finally, by Proposition 3.7 the solution shifts in favour of a point $\mathbf{x} = \bar{\mathbf{x}}(M) \in (0, 1)^K$.

3.2.4   Proof of Proposition 3.3

Proposition 3.3 relies on an algebraic lemma, which we write here for convenience. Its proof can be found in [56] (see Lemma 1 therein).

Lemma 3.8. *Let $P \geq 2$, $x_1, \ldots, x_P \geq 0$ and $b_1, \ldots, b_{P-1} \geq 0$. Set $S \equiv \sum_{p=1}^{P} x_p$ and $B \equiv \max_{p=1,\ldots,P-1} b_p$. Then:*

$$\sum_{p=1}^{P-1} b_p\, x_p\, x_{p+1} \leq \frac{B\, S^2}{4}\,. \tag{3.65}$$

Moreover we have equality in (3.65) if and only if one of the following conditions is verified:

a) there exists $p^* \in \{1, \ldots, P-1\}$ such that

$$x_{p^*} = x_{p^*+1} = \frac{S}{2}\quad,\quad b_{p^*} = B\,; \tag{3.66}$$

b) there exists $p^* \in \{2, \ldots, P-1\}$ such that

$$x_{p^*} = x_{p^*-1} + x_{p^*+1} = \frac{S}{2}\quad,\quad b_{p^*-1} = b_{p^*} = B\,. \tag{3.67}$$

Proof of Proposition 3.3. Denote by ρ the spectral radius of the matrix $[M^2]^{(oo)}$. We have

$$\rho \leq \left\| [M^2]^{(oo)} \right\|_\infty . \tag{3.68}$$

As $[M^2]^{(oo)}$ is a tridiagonal matrix, its ∞-norm can be easily computed leading to

$$\left\| [M^2]^{(oo)} \right\|_\infty = \max_r \sum_s (M^2)_{2r-1,2s-1} = \max_r \sum_{p=2r-3}^{2r} b_p^{(r)} \, \alpha_p \, \alpha_{p+1} \leq \frac{\widehat{\mu}^2}{4} , \tag{3.69}$$

where we set $\widehat{\mu}^2 \equiv \max_r \mu_{r,r+1}^2$ and for every r, p

$$\begin{aligned}
b_p^{(r)} \equiv \;& \delta_{p,\,2r-3} \, \mu_{2r-3,\,2r-2} \, \mu_{2r-2,\,2r-1} + \delta_{p,\,2r-2} \, \mu_{2r-2,\,2r-1}^2 + \\
& + \delta_{p,\,2r-1} \, \mu_{2r-1,\,2r}^2 + \delta_{p,\,2r} \, \mu_{2r-1,\,2r} \, \mu_{2r,\,2r+1} .
\end{aligned} \tag{3.70}$$

For convenience we set $\alpha_p \equiv 0$ for $p \notin \{1,\ldots,K\}$ and $\mu_{p,p+1} \equiv 0$ for $p \notin \{1,\ldots,K-1\}$. The last inequality in (3.69) follows by Lemma 3.8, since $b_p^{(r)} \leq \widehat{\mu}^2$ and $\sum_p \alpha_p = 1$. Therefore $\rho \leq \frac{\widehat{\mu}^2}{4}$ combining inequalities (3.68), (3.69).

Now assume that $\rho = \frac{\widehat{\mu}^2}{4}$. In particular the inequality in (3.69) must be saturated, hence there exists r such that

$$\sum_{p=2r-3}^{2r} b_p^{(r)} \, \alpha_p \, \alpha_{p+1} = \frac{\widehat{\mu}^2}{4} . \tag{3.71}$$

Then by Lemma 3.8, condition (3.11) or (3.12) must be verified.

Vice-versa assume that condition (3.11) or (3.12) holds true. In this case notice that many of the α_r's are zero, since $\sum_{r=1}^K \alpha_r = 1$. Thus the matrix $[M^2]^{(oo)}$ notably simplifies and one can check directly that $\frac{\widehat{\mu}^2}{4}$ is (the only non-zero) eigenvalue. This proves $\rho = \frac{\widehat{\mu}^2}{4}$. □

Remark 3.5. It is not difficult to realize that Theorem 3.1 holds also when $\alpha_r \to 0$ for some r. Indeed, by (3.28) and Lemma 2.5 we see that it is sufficient to require:

$$\alpha_r^2 \, \mathbb{E}\left\langle \left(\mathcal{L}_r - \mathbb{E}\langle \mathcal{L}_r \rangle_{N,t}^{(\epsilon)} \right)^2 \right\rangle_{N,t}^{(\epsilon)} \longrightarrow 0 \quad \text{as } N \to \infty . \tag{3.72}$$

The proof of Lemma 2.5 consists in showing that (see inequalities (A.11) and (A.24) in [43]):

$$\alpha_r^2 \, \mathbb{E}_{\boldsymbol{\epsilon}} \mathbb{E}\left\langle \left(\mathcal{L}_r - \langle \mathcal{L}_r \rangle_{N,t}^{(\epsilon)} \right)^2 \right\rangle_{N,t}^{(\epsilon)} = O\left(\frac{\alpha_r}{N s_N^K} \right) \tag{3.73}$$

$$\alpha_r^2 \, \mathbb{E}_{\boldsymbol{\epsilon}} \mathbb{E}\left[\left(\langle \mathcal{L}_r \rangle_{N,t}^{(\epsilon)} - \mathbb{E}\langle \mathcal{L}_r \rangle_{N,t}^{(\epsilon)} \right)^2 \right] = O\left(\frac{1}{s_N^{4K/3} N^{1/3}} \right) . \tag{3.74}$$

The previous equalities both vanish in the thermodynamic limit with the choice $s_N \sim N^{-1/16K}$ for instance, independently on α_r. Hence the remainder of the interpolation in the proof still goes to 0 with no variation in the hypothesis.

When a form factor, say α_r, vanishes the related component of the order parameter x_r disappears from the variational pressure (3.6). Moreover, if the corresponding L_r is an intermediate layer one can see that the system decouples into two independent DBMs because the effective interaction matrix Δ becomes block diagonal and the convex ψ-term related to the mentioned layer is weighed by α_r. The global variational pressure is thus constant in x_r in the thermodynamic limit.

3.2.5   Proof of Theorem 3.4

Uniqueness of the solution of the consistency equation for positive external fields can be proven adapting the strategy used in [56], where the replica symmetric equation of a deep Boltzmann machines was proved to admit a unique solution when the couplings and the external fields are centred Gaussian random variables. In particular the layers structure permits to "decouple" the interactions as shown in the following

Remark 3.6. The consistency equation (3.13) is equivalent to the following:

$$
\begin{cases}
x_r = \mathbb{E} \tanh\left(z\sqrt{\Theta_r(a)\,x_r + h_r} + \Theta_r(a)\,x_r + h_r \right) & r = 1, \ldots, K \\
\alpha_r\, x_r\, a_r = \alpha_{r+1}\, x_{r+1} & r = 1, \ldots, K-1
\end{cases}
\tag{3.75}
$$

where we have introduced the auxiliary variables $a_1, \ldots, a_{K-1} > 0$ and the functions

$$
\Theta_r(a) \equiv
\begin{cases}
\Delta_{12}\, a_1 & \text{for } r = 1 \\[2mm]
\dfrac{\Delta_{r,r-1}}{a_{r-1}} + \Delta_{r,r+1}\, a_r & \text{for } r = 2, \ldots, K-1 \\[2mm]
\dfrac{\Delta_{K-1,K}}{a_{K-1}} & \text{for } r = K
\end{cases}
\tag{3.76}
$$

Indeed, using the definition of the matrix M, it can be easily verified that $(M\mathbf{x})_r = \Theta_r(a)\,x_r$ for $r = 1, \ldots, K$, for a satisfying the second relation in (3.75).

The proof of Theorem 3.4 relies on the following

Lemma 3.9. *Let z be a standard Gaussian random variable. For every $t, h > 0$ the equation*

$$
x = \mathbb{E} \tanh\left(z\sqrt{t\,x + h} + t\,x + h \right)
\tag{3.77}
$$

has a unique positive solution that we denote by $x = \bar{x}(t, h) > 0$. Moreover $\bar{x}$ is strictly increasing as a function of both $t > 0$ and $h > 0$.

of Theorem 3.4. Equation (3.77) rewrites as $x = F(t\,x + h)$, where $F(h) \equiv \mathbb{E}\tanh(z\sqrt{h} + h)$. By Lemma 3.6, F takes values in (0,1), is strictly increasing and concave. It follows that

equation (3.77) admits a unique solution in $(0, 1)$ and in particular we can show that the function $f(x) \equiv \frac{1}{x} F(t\,x + h)$ is strictly decreasing for $x > 0$. Indeed by Lemma 3.6 we have:

$$x^2\, f'(x) \;=\; t\,x\, F'(t\,x + h) - F(t\,x + h) \;<\; 0 \text{ in } x = 0 , \tag{3.78}$$

$$\frac{\mathrm{d}}{\mathrm{d}x}\,(x^2\, f'(x)) \;=\; t^2\, x\, F''(t\,x + h) \;<\; 0 \tag{3.79}$$

hence

$$x^2\, f'(x) \;=\; t\,x\, F'(t\,x + h) - F(t\,x + h) \;<\; 0 \quad \forall\, x > 0 . \tag{3.80}$$

Now denoting by $\bar{x}(t, h)$ the unique positive solution of equation (3.77), we can prove its monotonicity with respect to both parameters by differentiating the self-consistent equation

$$\bar{x}(t, h) = F\left(t\,\bar{x}(t, h) + h\right) , \tag{3.81}$$

which leads to

$$(1 - t\, F'(t\,\bar{x} + h))\,\frac{\mathrm{d}\bar{x}}{\mathrm{d}t} \;=\; \bar{x}\, F'(t\,\bar{x} + h) \tag{3.82}$$

$$(1 - t\, F'(t\,\bar{x} + h))\,\frac{\mathrm{d}\bar{x}}{\mathrm{d}h} \;=\; F'(t\,\bar{x} + h) . \tag{3.83}$$

Lemma 3.6 ensures that (3.82), (3.83) are positive quantities, hence to conclude it suffices to show that $1 - t\, F'(t\,\bar{x} + h) > 0$. Indeed, dividing the inequality (3.80) by x, evaluating it at $x = \bar{x}(t, h)$ and using the self-consistent equation (3.81), one finds precisely:

$$0 > t\, F'(t\,\bar{x} + h) - \frac{F(t\,\bar{x} + h)}{\bar{x}} \;=\; t\, F'(t\,\bar{x} + h) - 1 . \tag{3.84}$$

$\square$

Proof of Theorem 3.4. By Lemma 3.9, the first line of (3.75) is equivalent to:

$$x_r \;=\; \bar{x}\left(\Theta_r(a), h_r\right) \quad \forall\, r = 1, \ldots, K \tag{3.85}$$

where $\bar{x}$ is uniquely defined and strictly increasing with respect to both its arguments. On the other hand the second line of (3.75) rewrites as:

$$\alpha_1\, x_1\, a_1 \cdots a_r \;=\; \alpha_{r+1}\, x_{r+1} \quad \forall\, r = 1, \ldots, K - 1 . \tag{3.86}$$

It is convenient to set $X_1(a_1) \equiv \alpha_1\, \bar{x}\left(\Theta_1(a)\,,\, h_1\right) = \alpha_1\, \bar{x}\left(\Delta_{1,2}\, a_1\,,\, h_1\right)$ and for $r \geq 2$

$$X_r\left(\frac{1}{a_{r-1}}\,,\, a_r\right) \equiv \alpha_r\, \bar{x}\left(\Theta_r(a)\,,\, h_r\right) = \alpha_r\, \bar{x}\left(\frac{\Delta_{r,r-1}}{a_{r-1}} + \Delta_{r,r+1}\, a_r\,,\, h_r\right) . \tag{3.87}$$

Therefore equation (3.75) is equivalent to the following:

$$X_1(a_1)\, a_1 \cdots a_r \;=\; X_{r+1}\!\left(\frac{1}{a_r}, a_{r+1}\right) \quad \forall\, r = 1, \ldots, K-1\,. \tag{3.88}$$

We will show by induction on $r \geq 1$ that for any given $a_{r+1} \geq 0$ there exists a unique $a_r = \bar{a}_r(a_{r+1}) > 0$ such that

$$\begin{cases} a_{r-1} = \bar{a}_{r-1}(a_r) \\ \vdots \\ a_1 = \bar{a}_1(a_2) \\ X_1(a_1)\, a_1 \cdots a_{r-1}\, a_r = X_{r+1}\!\left(\frac{1}{a_r}, a_{r+1}\right) \end{cases} \tag{3.89}$$

and moreover $\bar{a}_r$ is a strictly increasing function with respect to a_{r+1}. The uniqueness of solution of (3.88) will follow immediately by stopping the induction at $r = K-1$ and choosing $a_K = 0$ and the Theorem will be proven thanks to Remark 3.6.

- Case $r = 1$: given $a_2 \geq 0$, let's consider the equation

$$X_1(a_1)\, a_1 \;=\; X_2\!\left(\frac{1}{a_1}, a_2\right)\,. \tag{3.90}$$

By Lemma 3.9 the left-hand side of (3.90) is a strictly increasing function of $a_1 > 0$ and takes all the values in the interval $(0, \infty)$, while the right-hand side is a decreasing function of $a_1 > 0$ and takes non-negative values. Therefore there exists a unique $a_1 = \bar{a}_1(a_2) > 0$ solution of (3.90). Now taking derivatives on both sides of (3.90) and using again Lemma 3.9, one finds:

$$\frac{d\bar{a}_1}{da_2} \;=\; \frac{\partial}{\partial a_2} X_2\!\left(\frac{1}{a_1}, a_2\right)\left[\frac{\partial}{\partial a_1}\big(X_1(a_1)\, a_1\big) - \frac{\partial}{\partial a_1} X_2\!\left(\frac{1}{a_1}, a_2\right)\right]^{-1}\Bigg|_{a_1 = \bar{a}_1(a_2)} > 0 \tag{3.91}$$

hence $\bar{a}_1$ is a strictly increasing function of a_2.

- For $r > 1$, $r - 1 \Rightarrow r$. Fix $a_{r+1} \geq 0$. By inductive hypothesis $\bar{a}_1, \ldots, \bar{a}_{r-1}$ are well-defined and strictly increasing functions. Defining the composition $A_l \equiv \bar{a}_l \circ \cdots \circ \bar{a}_{r-1}$ for every $l = 1, \ldots, r-1$, equation (3.89) rewrites as:

$$\big(X_1 \circ A_1\big)(a_r)\, A_1(a_r) \cdots A_{r-1}(a_r)\, a_r \;=\; X_{r+1}\!\left(\frac{1}{a_r}, a_{r+1}\right)\,. \tag{3.92}$$

By inductive hypothesis and Lemma 3.9, the left-hand side of (3.92) is a strictly increasing function of $a_r > 0$ and takes all the values in the interval $(0, \infty)$, while the right hand-side of

(3.92) is a decreasing function of $a_r > 0$ and takes non-negative values. Therefore for every $a_{r+1} \geq 0$ there exists a unique $a_r = \bar{a}_r(a_{r+1}) > 0$ solution of (3.92). Now taking derivatives on both sides of (3.92) one finds:

$$
\begin{aligned}
\frac{d\bar{a}_r}{da_{r+1}} &= \frac{\partial}{\partial a_{r+1}} X_{r+1}\left(\frac{1}{a_r}, a_{r+1}\right) \cdot \\
&\cdot \left[\frac{\partial}{\partial a_r}\left((X_1 \circ A_1)(a_r) \, A_1(a_r) \cdots A_{r-1}(a_r) \, a_r \right) - \frac{\partial}{\partial a_r} X_{r+1}\left(\frac{1}{a_r}, a_{r+1}\right) \right]^{-1}_{\Big|_{a_r = \bar{a}_r(a_{r+1})}}
\end{aligned}
\tag{3.93}
$$

which, using again the inductive hypothesis and Lemma 3.9, entails that $\bar{a}_r$ is a strictly increasing function of a_{r+1}. $\square$

3.3 Concluding remarks

The solution we found consists in an ordinary min-max variational principle over K real positive numbers for the thermodynamic limit of the free entropy. The properties of the optimizer show the presence of a phase transition related to the interaction strength and to the relative size of each layer defining the geometry of the system. In particular we discovered that the geometry of the system may tune the phase transition.

We mention that in the recent paper [65] the mutual information for a wide class of inference problems is rigorously computed by means of a variational principle. While it is possible to obtain our model as an instance of the one considered there, the variational principle presented has no clear correspondence to ours. We finally mention that a subsequent work [66] contains a general result that extends the one contained here. In particular the authors compute the limiting free energy with a Hamilton-Jacobi approach which proves to be effective also when dealing with lack of convexity in the interactions. On the other hand, the simplicity of our setting allows us to carry out a thorough study of the variational formula by locating the phase transition and investigating its dependency on the geometry of the system as in Theorem 3.2, Proposition 3.3 and Theorem 3.4.

Chapter 4

Mismatched rank one matrix estimation

We have seen so far that the Nishimori line is, in some sense, a privileged sub-region of the phase space of a spin glass, where replica symmetry can be shown to hold under fairly mild hypothesis [67, 39, 40]. As a consequence we were able to find rigorously finite dimensional variational principles to express the free entropy of such models. Our proofs were based on the validity of the Nishimori identities and correlation inequalities of the first and second type, that in turn derive from Bayes-optimality, *i.e.* from the fact that the Statistician knows every detail of the generative process of the data.

There are different ways to break Bayes-optimality. For instance *(i)* the Statistician may ignore the signal-to-noise ratio [68, 69]. Or, *(ii)* they use a Gaussian likelihood, whereas the noise is not Gaussian [70]. Thirdly, *(iii)* they adopt a wrong prior for the signal components. These particular settings, referred to as *mismatched*, have recently attracted interest from the Statistical Physics and Inference communities [71, 72, 73, 74].

In this chapter we work within the case *(iii)*, *i.e.* a mismatched spiked Wigner model where the Statistician has no *apriori* knowledge of the signal and tries to reconstruct it only through Ising spins. Instead of using a max-likelihood approach to estimate the signal, which would correspond to the search of the ground state of a given Hamiltonian, we choose a typical configuration of the system at finite temperature, or equivalently we adopt the receiver's posterior mean as the estimator, as a Bayes-optimal Statistician would. The emerging Statistical Mechanics model turns out to be the sum of an SK with a two-body mean-field Mattis interaction. The main result is the rigorous exact solution of this model described by two natural order parameters represented by the overlap distribution and the Mattis magnetization. We show that, while the first obeys a functional variational principle of Parisi type, the second is obtained through a classical one dimensional optimization problem. The proof relies on the crucial property of self-averaging of the Mattis magnetization. When the distribution of the ground truth signal is Gaussian the phase space is investigated and a tricritical point is identified separating paramagnetic, glassy and ferromagnetic phases.

4.1 Definitions and Main Results

Let us recall some basic definitions. Consider a system of N interacting Ising spins described by a Sherrington-Kirkpatrick Hamiltonian with external random *iid* magnetic fields and a further two body interaction of Mattis type induced by the same magnetic fields. More specifically, to each site $i = 1, \ldots, N$ we associate a spin $\sigma_i \in \{+1, -1\}$. The state of the system will be completely identified by the vector $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_N) \in \{+1, -1\}^N =: \Sigma_N$. Furthermore, we assume that the spins have a uniform prior distribution, namely $\mathbb{P}(\sigma_i = +1) = 1/2$. The Hamiltonian of the model hereby studied is

$$H_N(\boldsymbol{\sigma}; \mu, \nu, \lambda) \equiv H_N(\boldsymbol{\sigma}) = -\sum_{i,j=1}^{N} \left(z_{ij} \sqrt{\frac{\mu}{2N}} \sigma_i \sigma_j + \frac{\nu}{2N} \sigma_i \sigma_j \xi_i \xi_j \right) - \lambda \sum_{i=1}^{N} \xi_i \sigma_i \,, \qquad (4.1)$$

with $\mu, \nu \geq 0$, $\lambda \in \mathbb{R}$, $z_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$ and $\xi_i \overset{\text{iid}}{\sim} P_\xi$ independent of the z_{ij}'s, where P_ξ is any distribution such that $\mathbb{E}[\xi_1^4] < \infty$. The z_{ij}'s and ξ_i's play the role of quenched disorder in this model. The model is going to be described by the couple of order parameters

$$q_N(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \frac{1}{N} \boldsymbol{\sigma} \cdot \boldsymbol{\tau} = \frac{1}{N} \sum_{i=1}^{N} \sigma_i \tau_i \,, \quad m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i \xi_i \qquad (4.2)$$

where $\boldsymbol{\sigma}, \boldsymbol{\tau} \in \Sigma_N$ and $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_N)$. In what follows we will refer to $m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})$ as Mattis magnetization. One can now separate the three contributions in the Hamiltonian (4.1), thus obtaining

$$H_N(\boldsymbol{\sigma}) = -\sqrt{\mu} \sum_{i,j=1}^{N} \frac{z_{ij}}{\sqrt{2N}} \sigma_i \sigma_j - \frac{N\nu}{2} m_N^2(\boldsymbol{\sigma}|\boldsymbol{\xi}) - N\lambda m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) \,, \qquad (4.3)$$

where an SK-like term

$$H_N^{SK}(\boldsymbol{\sigma}) := -\sum_{i,j=1}^{N} \frac{z_{ij}}{\sqrt{2N}} \sigma_i \sigma_j \qquad (4.4)$$

at temperature $\sqrt{\mu}$ is clearly recognizable.

We will denote the random and quenched pressures of the model respectively as follows:

$$p_N(\mu, \nu, \lambda) = \frac{1}{N} \log \sum_{\boldsymbol{\sigma} \in \Sigma_N} \exp \left[-\sqrt{\mu} H_N^{SK}(\boldsymbol{\sigma}) + \frac{N\nu}{2} m_N^2(\boldsymbol{\sigma}|\boldsymbol{\xi}) + N\lambda m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) \right] \qquad (4.5)$$

$$\bar{p}_N(\mu, \nu, \lambda) = \mathbb{E} p_N(\mu, \nu, \lambda) \qquad (4.6)$$

where the expectation in the latter is taken w.r.t. all the disorder: $\mathbb{E} \equiv \mathbb{E}_{\boldsymbol{\xi}} \mathbb{E}_{\mathbf{z}}$. For the reader's convenience, we report here the definitions of the quenched pressure of an SK model with

random magnetic fields $\xi_i \overset{iid}{\sim} P_\xi$ and its limit:

$$\bar{p}_N^{SK}(\beta, h) := \frac{1}{N} \mathbb{E} \log \sum_{\boldsymbol{\sigma} \in \Sigma_N} \exp\left[-\beta H_N^{SK}(\boldsymbol{\sigma}) + h \sum_{i=1}^{N} \xi_i \sigma_i \right], \tag{4.7}$$

$$\mathcal{P}(\beta, h) := \inf_{\chi \in \mathcal{M}_{[0,1]}} \mathcal{P}(\chi; \beta, h) = \lim_{N \to \infty} \bar{p}_N^{SK}(\beta, h), \tag{4.8}$$

where $\mathcal{M}_{[0,1]}$ and $\mathcal{P}(\chi; \beta, h)$ are as in Section 1.2.3. The main result of the chapter is the variational principle for the thermodynamic limit of (4.6).

Theorem 4.1 (Variational solution). *If $\mathbb{E}[\xi_1^4] < +\infty$ then*

$$p_N(\mu, \nu, \lambda) \xrightarrow{L^2} \lim_{N \to \infty} \bar{p}_N(\mu, \nu, \lambda) =: p(\mu, \nu, \lambda) = \sup_{x \in \mathbb{R}} \varphi(x; \mu, \nu, \lambda) \tag{4.9}$$

where

$$\varphi(x; \mu, \nu, \lambda) := -\frac{\nu x^2}{2} + \mathcal{P}(\sqrt{\mu}, \nu x + \lambda). \tag{4.10}$$

From the form of the variational principle we can deduce also the differentiability properties of the limiting pressure that we have collected in the following

Corollary 4.2. *$p(\mu, \nu, \lambda)$ is λ-differentiable if and only if $\varphi(\,\cdot\,; \mu, \nu, \lambda)$ has a unique supremum point $x = \bar{x}(\mu, \nu, \lambda)$ and in that case*

$$\bar{x} = \frac{\partial}{\partial h} \mathcal{P}(\sqrt{\mu}, h) \bigg|_{h = \nu \bar{x} + \lambda} = \lim_{N \to \infty} \mathbb{E} \langle m_N(\boldsymbol{\sigma} | \boldsymbol{\xi}) \rangle_N. \tag{4.11}$$

$p(\mu, \nu, \lambda)$ is ν-differentiable if and only if $\varphi(\,\cdot\,; \mu, \nu, \lambda)$ has at most two symmetric supremum points $\{\bar{x}, -\bar{x}\}$ and it holds

$$\frac{\partial}{\partial \nu} p(\mu, \nu, \lambda) = \frac{\bar{x}^2}{2}. \tag{4.12}$$

Let $\xi \sim P_\xi$ be centered and $\nu > 0$. If $\varphi(\,\cdot\,; \mu, \nu, \lambda = 0)$ has at most two symmetric supremum points $\{\bar{x}, -\bar{x}\}$ then $p(\mu, \nu, 0)$ is μ-differentiable and it holds

$$\frac{\partial}{\partial \mu} p(\mu, \nu, 0) = \frac{1}{4} \left(1 - \int q^2 d\chi^*(\sqrt{\mu}, \nu \bar{x}; q) \right). \tag{4.13}$$

where $\chi^*(\beta, h)$ denotes the unique Parisi measure solving the Parisi variational principle in (4.8) for $\beta = \sqrt{\mu}$, $h = \nu \bar{x}$.

The proof of (4.9) relies on the adaptive interpolation. We stress that the latter has proved to be particularly successful in the optimal setting, but to our best knowledge, up to [75] it was not applied in the mismatched setting.

From a Statistical Mechanics viewpoint, a similar model was studied in [46] where the author solves a Sherrington-Kirkpatrick model with an added ferromagnetic interaction, that can be derived from (4.1) setting $P_\xi = \delta_{\sqrt{J}}$ with $J > 0$ as the interaction strength.

Notice that the variational principle in (4.9) is one dimensional, as far as x is concerned, suggesting thus the self-averaging of an order parameter to be identified with the Mattis magnetization as in (4.11). Indeed, the following concentration result holds.

Proposition 4.3. *Let $\epsilon \in [s_N, 2s_N]$ with $s_N = \frac{1}{2}N^{-\alpha}$, $\alpha \in (0, 1/2)$. Denote by $\langle \cdot \rangle_{N,y}$ the Boltzmann-Gibbs measure induced by the Hamiltonian $H_N(\boldsymbol{\sigma}; \mu, \nu, \lambda + y)$ for any $y \in \mathbb{R}$. Then*

$$\lim_{N\to\infty} \frac{1}{s_N} \int_{s_N}^{2s_N} d\epsilon \, \mathbb{E}\Big\langle \big(m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) - \mathbb{E}\langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,\epsilon}\big)^2 \Big\rangle_{N,\epsilon} = 0\,, \tag{4.14}$$

for all $\mu, \nu \geq 0$ and $\lambda \in \mathbb{R}$.

The proofs of Theorem 4.1, Corollary 4.2 and Proposition 4.3 can be found in Section 4.3.2 and require bounds on the fluctuations of $m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})$.

4.1.1   The Gaussian case

Theorem 4.1 contains a variational representation for the thermodynamic limit of the quenched pressure density $p_N(\mu, \nu, \lambda)$ under mild assumption on the distribution of the family $\boldsymbol{\xi}$. We should notice that despite the fact the variational problem is one dimensional, the potential $\varphi(x; \mu, \nu, \lambda)$ in (4.10) contains a very complicated object, namely the pressure of a SK model which is given by the Parisi formula. For these reasons it can be very hard in general to obtain analytical information on the solution of the above variational problem. For instance, an important question is when, once the potential $\varphi(x; \mu, \nu, \lambda)$ is evaluated at the optimal value for x, the Parisi term is solved by a non fluctuating order parameter, i.e. is replica symmetric. The purpose of this subsection is to obtain some detailed insights on the model by studying it on some analytically accessible case, in particular for a specific choice of the family $\boldsymbol{\xi}$ that allows a quantitative description of the phase diagram.   We choose the family $\boldsymbol{\xi}$ to be *i.i.d* centered Gaussian, $P_\xi = \mathcal{N}(0, a)$, and we set $\nu = \mu$ and $\lambda = 0$. The above choice for the parameters μ, ν, λ and its link with high dimensional inference problems is discussed in Sect. 4.2. We will show that in this setting one can use the nice result in [22] on the sharpness of the de Almeida-Thouless line for Gaussian centered external magnetic fields for the SK model, to perform an in-depth analysis of the variational problem in Theorem 4.1. With a slight abuse of notation, we denote the corresponding quenched pressure by $\bar{p}_N(\mu, a)$. We show that it is possible to identify the regions in the phase plane (μ, a) where $\mathcal{P}$ defined in (4.8) can be replaced

with its replica symmetric version, thus obtaining the following replica symmetric potential

$$\varphi_{RS}(x; \mu, a) := -\frac{\mu x^2}{2} + \frac{\mu(1 - q(x, \mu, a))^2}{4} + \mathbb{E}\log\cosh\left(z\sqrt{\mu q(x, \mu, a)} + \mu\xi x\right) \tag{4.15}$$

where $q(x, \mu, a)$ is uniquely defined, thanks to the Latala-Guerra lemma [20], by the consistency equation

$$q(x, \mu, a) = \mathbb{E}\tanh^2\left(z\sqrt{\mu q(x, \mu, a)} + \mu\xi x\right), \tag{4.16}$$

for any $x > 0$ and we extend it to $x = 0$ by continuity. The properties of φ_{RS} are hereby collected:

Proposition 4.4. *The following properties hold:*

1. *$\varphi_{RS}(-x; \mu, a) = \varphi_{RS}(x; \mu, a)$;*

2. *$\lim_{|x|\to\infty} \varphi_{RS}(x; \mu, a) = -\infty$;*

3. *there exists a unique maximum point, up to reflection, $x = \bar{x}(\mu, a) \geq 0$ which is either 0 or satisfies*

$$q(\bar{x}(\mu, a), \mu, a) = 1 - \frac{1}{\mu a}; \tag{4.17}$$

4. *the solution of (4.17) exists and is unique if and only if*

$$a \geq \frac{1}{\mu(1 - q(0, \mu, 0))}. \tag{4.18}$$

   The previous is always fulfilled if $a \geq 1/\mu$ and $a \geq 1$;

5. *under the hypothesis (4.18) the solution to (4.17) is stable:*

$$\left.\frac{d^2\varphi_{RS}(x; \mu, a)}{dx^2}\right|_{x=\pm\bar{x}(\mu,a)} < 0. \tag{4.19}$$

Finally, we give a sharp criterion to establish when the replica symmetric potential can be used to obtain the solution to the variational problem.

Proposition 4.5. *Define the function*

$$AT(\mu, a) := \mu\mathbb{E}\cosh^{-4}\left(z\sqrt{\mu q(\bar{x}(\mu, a), \mu, a)} + \mu\xi\bar{x}(\mu, a)\right). \tag{4.20}$$

Then

$$\lim_{N\to\infty} \bar{p}_N(\mu, a) = \sup_{x\in\mathbb{R}} \varphi_{RS}(x; \mu, a) \quad \textit{iff} \quad AT(\mu, a) \leq 1. \tag{4.21}$$

The proofs of Propositions 4.4 and 4.5 can be found in Section 4.3.3. The previous results for Gaussian ξ_i's and their consequences can be gathered together in the phase diagram in Figure 4.1 which will be studied in detail in the dedicated Section 4.4.

4.2   Mismatched Setting in High Dimensional Statistical Inference

The Hamiltonian (4.1) with $\nu = \mu$ (which is not restrictive, one can reabsorb ν in the ξ_i's) and $\lambda = 0$ can be derived also from the spiked Wigner model in a mismatched setting as anticipated.
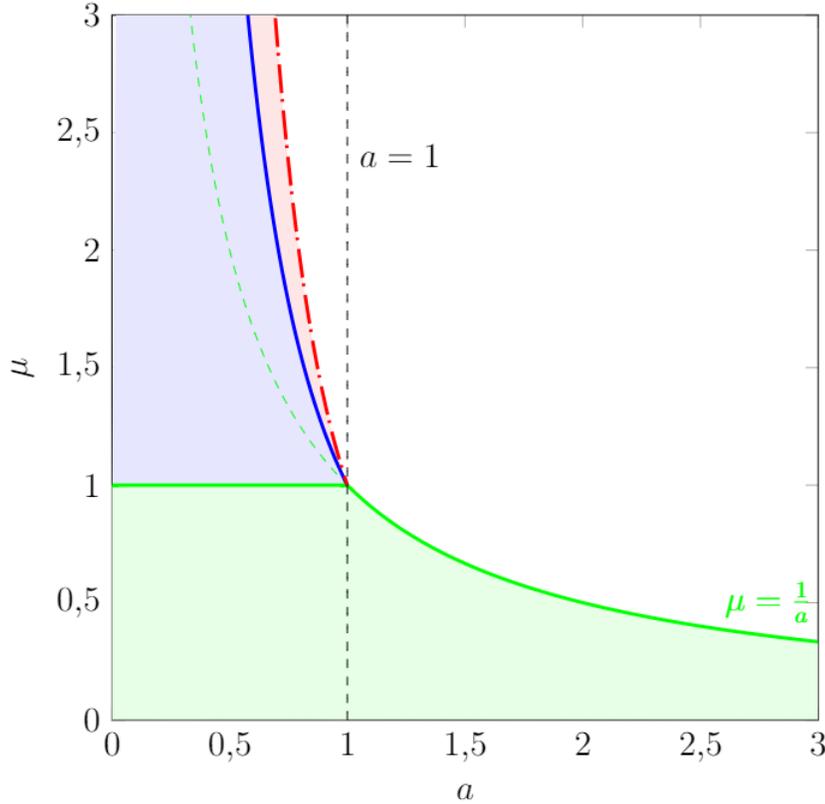

Figure 4.1:   Model phase diagram when $\xi_i \overset{\text{iid}}{\sim} \mathcal{N}(0, a)$. Green region: fully paramagnetic phase, where the Parisi overlap distribution is a Dirac delta centered at 0 as well as the Mattis magnetization. White region: ferromagnetic, replica symmetric phase. Here, the Parisi overlap distribution is still a Dirac located according to (4.16) and (4.17). The distribution of the Mattis magnetization is instead a sum of two deltas centered at $\bar{x}$ and $-\bar{x}$ with $1/2$ coefficients, namely the solutions of (4.17). The model therefore turns out to be replica symmetric in the green and white areas. The blue region is delimited by $\mu = 1$ and the blue curve drawn (here only qualitatively) by (4.18), which is above the green dashed hyperbola $\mu = 1/a$. In this region the model is in a replica symmetry breaking phase, *i.e.* the Parisi distribution is no longer concentrated at a single point and $\bar{x}$ solves the more general variational principle (4.9). With reference to Proposition 4.5, the dash-dotted red line $AT(\mu, a) = 1$, here drawn qualitatively, must contain the entire RSB phase, hence it must lie above (or at most touch) the blue curve. The analogy with the SK model (see Remark 4.6 below) would suggest the presence of a mixed phase in the red region where $\bar{x} \neq 0$ and the overlap fluctuates.

More specifically the Statistician has access to the N^2 quantities

$$y_{ij} := \sqrt{\frac{\mu}{2N}}\xi_i\xi_j + z_{ij}, \quad z_{ij} \stackrel{\text{iid}}{\sim} \mathcal{N}(0,1). \tag{4.22}$$

She also knows how the observations are generated, namely she is aware of the law (4.22) and consequently of the conditional distribution

$$p_{\mathbf{Y}|\boldsymbol{\xi}=\mathbf{x}}(\mathbf{y}) = \frac{\exp\left[-\frac{1}{2}\sum_{i,j=1}^{N}\left(y_{ij} - \sqrt{\frac{\mu}{2N}}x_ix_j\right)^2\right]}{(2\pi)^{N^2/2}}, \tag{4.23}$$

for some value $\mathbf{x}$. However, she does not know the distribution of the ξ_i's and assumes them to be binary ± 1 as the σ_i's with equal prior probability. Thus, according to Bayes' rule, the (wrong) posterior distribution used by the Statistician is

$$P_{\boldsymbol{\xi}|\mathbf{Y}=\mathbf{y}}(\boldsymbol{\sigma}) = \frac{\exp\left[-\frac{1}{2}\sum_{i,j=1}^{N}\left(y_{ij} - \sqrt{\frac{\mu}{2N}}\sigma_i\sigma_j\right)^2\right]}{2^N(2\pi)^{N^2/2}p_{\mathbf{Y}}(\mathbf{y})}, \tag{4.24}$$

where

$$p_{\mathbf{Y}}(\mathbf{y}) = \frac{1}{2^N}\sum_{\boldsymbol{\sigma}\in\Sigma_N}\frac{\exp\left[-\frac{1}{2}\sum_{i,j=1}^{N}\left(y_{ij} - \sqrt{\frac{\mu}{2N}}\sigma_i\sigma_j\right)^2\right]}{(2\pi)^{N^2/2}}. \tag{4.25}$$

A straightforward computation shows that the posterior distribution (4.24) can be rewritten as a random Boltzmann-Gibbs measure whose Hamiltonian is precisely $H_N(\boldsymbol{\sigma};\mu,\mu,0)$.

It is important to stress that nor the posterior (4.24) neither the so called *evidence* (4.25) are correct, in the sense that there is a mismatch between the receiver's prior and P_ξ. The true distribution of the y_{ij}'s is instead

$$p_{\mathbf{Y}}^*(\mathbf{y}) = \int dP_\xi(\mathbf{x})\frac{\exp\left[-\frac{1}{2}\sum_{i,j=1}^{N}\left(y_{ij} - \sqrt{\frac{\mu}{2N}}x_ix_j\right)^2\right]}{(2\pi)^{N^2/2}}, \tag{4.26}$$

with $dP_\xi(\mathbf{x}) = \prod_{i=1}^{N}dP_\xi(x_i)$. With these notations one can proceed with the computation of the cross entropy density

$$\frac{1}{N}\mathcal{H}(p_{\mathbf{Y}}^*, p_{\mathbf{Y}}) = -\frac{1}{N}\int d\mathbf{y}\, p_{\mathbf{Y}}^*(\mathbf{y})\log p_{\mathbf{Y}}(\mathbf{y}), \tag{4.27}$$

a quantity that can be evaluated only by a third party observer aware of both P_ξ and the mismatched prior. By inserting $p_{\mathbf{Y}}$ and the (4.22) in the (4.27) one obtains

$$\frac{1}{N}\mathcal{H}(p_{\mathbf{Y}}^*, p_{\mathbf{Y}}) = -\frac{1}{N}\mathbb{E}_{\mathbf{Z}}\mathbb{E}_{\boldsymbol{\xi}}\log\sum_{\boldsymbol{\sigma}\in\Sigma_N}\frac{\exp\left[-\frac{1}{2}\sum_{i,j=1}^{N}\left(z_{ij} + \sqrt{\frac{\mu}{2N}}(\xi_i\xi_j - \sigma_i\sigma_j)\right)^2\right]}{2^N(2\pi)^{N^2/2}}$$

$$= \frac{N}{2}\log 2\pi e + \frac{\mu}{4}\mathbb{E}_\xi^2[\xi_1^2] + \frac{\mu}{4} + O\left(\frac{1}{N}\right) + \log 2 - \frac{1}{N}\mathbb{E}_{\mathbf{Z}}\mathbb{E}_{\boldsymbol{\xi}}\log\sum_{\boldsymbol{\sigma}\in\Sigma_N}e^{-H_N(\boldsymbol{\sigma};\mu,\mu,0)}. \tag{4.28}$$

The first term is the Shannon entropy of the noise, whereas the last one is, up to a sign, the quenched pressure $\bar{p}_N(\mu, \mu, 0)$. In the optimal setting, namely when $P_\xi = (\delta_1 + \delta_{-1})/2$, (4.27) is just the entropy of the observations and the quantity $\frac{\mu}{4} + \frac{\mu}{4} + \log 2 - \bar{p}_N(\mu, \mu, 0)$ is the mutual information $\frac{1}{N} I(\mathbf{Y}, \boldsymbol{\xi})$ between the ground truth signal and the observations up to $O\left(\frac{1}{N}\right)$.

Using integration by parts it is straightforward to show that

$$\frac{d}{d\mu} \frac{\mathcal{H}(p_{\mathbf{Y}}^*, p_{\mathbf{Y}})}{N} = \frac{\mathbb{E}_\xi^2[\xi_1^2] - 2\mathbb{E}\langle m_N^2(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_N + \mathbb{E}\langle q_N^2(\boldsymbol{\sigma}, \boldsymbol{\tau})\rangle_N}{4} = \frac{1}{4N^2}\mathbb{E}\|\boldsymbol{\xi} \otimes \boldsymbol{\xi} - \langle \boldsymbol{\sigma} \otimes \boldsymbol{\sigma}\rangle_N\|_F^2 .$$

(4.29)

The previous equation relates the cross entropy (4.27) to the theoretical expected Mean Square Error (MSE) in Frobenius' norm that the receiver would make using the Bayesian a posteriori estimator $\langle \boldsymbol{\sigma} \otimes \boldsymbol{\sigma}\rangle = (\langle \sigma_i \sigma_j\rangle)_{i,j=1}^N$ for the ground truth diad $\boldsymbol{\xi} \otimes \boldsymbol{\xi} = (\xi_i \xi_j)_{i,j=1}^N$. As seen in Proposition 1.9, estimation performed in the matched setting produces a MSE which is the smallest possible which entails that the MSE (4.29) is sub-optimal [71]. It is worth stressing again that the MSE (4.29) can be evaluated only by the aforementioned third party observer, since it derives directly from $\mathcal{H}(p_{\mathbf{Y}}^*, p_{\mathbf{Y}})$.

The MSE in the high dimensional limit can be evaluated using Theorem 4.1 and the following

Lemma 4.6. *Let I be an open real interval, $\{g_n\}_{n\in\mathbb{N}}$ a sequence of differentiable functions defined on I converging pointwise to a differentiabile function g. Suppose there exists a differentiable function f on I such that $\{g_n + f\}_{n\in\mathbb{N}}$ is a sequence of convex differentiable functions on I. Then $\lim_{N\to\infty} g_n'(x) = g'(x)$.*

Proof. The statement follows immediately from an application of Griffith's Lemma (see for instance [64], Lemma IV.6.3) to the sequence $\tilde{g}_n = g_n + f$. □

For our cross entropy density sequence, which is not convex due to the lack of Nishimori identities, one can prove that

$$\frac{\tilde{\mathcal{H}}(p_{\mathbf{Y}}^*, p_{\mathbf{Y}})}{N} = \frac{\mathcal{H}(p_{\mathbf{Y}}^*, p_{\mathbf{Y}})}{N} - \mu \log \mu$$

(4.30)

is concave by a direct computation of its second derivative w.r.t. μ. We leave the details of the computation to the interested reader. Hence the previous Lemma, under the hypothesis for $\mu(\nu)$-differentiability of $p(\mu, \nu, \lambda = 0)$ in Corollary 4.2, implies that

$$\lim_{N\to\infty} \frac{1}{4N^2}\mathbb{E}\|\boldsymbol{\xi} \otimes \boldsymbol{\xi} - \langle \boldsymbol{\sigma} \otimes \boldsymbol{\sigma}\rangle\|_F^2 = \frac{d}{d\mu}\left[\frac{\mu}{4}\mathbb{E}_\xi^2[\xi_1^2] + \frac{\mu}{4} - p(\mu, \mu, 0)\right] =$$

$$= \frac{1}{4}\mathbb{E}_\xi^2[\xi_1^2] + \frac{1}{4} - \frac{\bar{x}^2}{2} - \frac{1}{2\sqrt{\mu}}\partial_\beta \mathcal{P}(\beta, \mu\bar{x})|_{\beta=\sqrt{\mu}} =$$

$$= \frac{1}{4}\mathbb{E}_\xi^2[\xi_1^2] + \frac{1}{4} - \frac{\bar{x}^2}{2} - \frac{1}{4}\left(1 - \int q^2 d\chi^*(\sqrt{\mu}, \mu\bar{x}; q)\right), \quad (4.31)$$

where we have replaced the derivative of the Parisi functional w.r.t. β as prescribed by Corollary 4.2. We finally come up with

$$\lim_{N\to\infty}\frac{1}{4N^2}\mathbb{E}\|\boldsymbol{\xi}\otimes\boldsymbol{\xi}-\langle\boldsymbol{\sigma}\otimes\boldsymbol{\sigma}\rangle\|_F^2=\frac{1}{4}\mathbb{E}_\xi^2[\xi_1^2]-\frac{\bar{x}^2}{2}+\frac{1}{4}\int q^2d\chi^*(\sqrt{\mu},\mu\bar{x};q)\,.\tag{4.32}$$

4.3 Proofs

4.3.1 Tools

In this section we show how to interpolate our model with a simple SK model with random *iid* magnetic fields following an adaptive path. The advantage of this approach in this mismatched setting is the possibility to confine the replica symmetry breaking phenomena in the SK part of the model which is exhaustively studied in the literature. The ultimate purpose of the interpolation hereby illustrated is thus to linearize the squared magnetization in the Hamiltonian.

The interpolating model is defined by means of the Hamiltonian

$$H_N(t;\sigma):=H_N(\sigma;\mu,(1-t)\nu,\lambda+R_\epsilon(t))=$$
$$=\sqrt{\mu}H_N^{SK}(\boldsymbol{\sigma})-(1-t)\frac{N\nu}{2}m_N^2(\boldsymbol{\sigma}|\boldsymbol{\xi})-(\lambda+R_\epsilon(t))Nm_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\tag{4.33}$$

where

$$R_\epsilon(t)=\epsilon+\nu\int_0^t ds\,r_\epsilon(s)\,,\quad\epsilon\in[s_N,2s_N]\,,\ s_N=\frac{N^{-\alpha}}{2}\tag{4.34}$$

with $\alpha\in(0,1/2)$ and where the interpolating function r_ϵ will be suitably chosen (see Remark 4.3 below for instance). The related interpolating pressure is:

$$\bar{p}_N(t):=\bar{p}_N(\mu,(1-t)\nu,\lambda+R_\epsilon(t))=\frac{1}{N}\mathbb{E}_{\boldsymbol{\xi}}\mathbb{E}_{\mathbf{Z}}\log\sum_{\boldsymbol{\sigma}\in\Sigma_N}\exp\left[-H_N(t;\sigma)\right]\,.\tag{4.35}$$

As done in Proposition 4.3, the Boltzmann-Gibbs averages relative to (4.33) will be denoted by $\langle\cdot\rangle_{N,R_\epsilon(t)}$.

Remark 4.1. The interpolation strategy that we use in this work is profoundly different from the typical one of the statistical inference literature within the Bayes optimal setting. In that case one interpolates directly at the level of the channel, namely of (4.22), to compare it with a one body channel of the type $y_i=\sqrt{R_\epsilon(t)}\xi_i+z_i$ with $z_i\overset{\text{iid}}{\sim}\mathcal{N}(0,1)$. Traveling along a trajectory that keeps an inferential interpretation ensures that the model is on the Nishimori line at any t where all the precious properties of that line, identities and correlation inequalities, provide a crucial analytical tool to obtain a finite dimensional variational principle.

In the present case instead the structural complexity of the mismatched setting implies that we cannot count in the very first place on the Nishimori line properties nor on a global absence

of fluctuations for the order parameters. The strategy to achieve the solution and overcome this difficulty is to build an interpolation scheme that, albeit not coming from a Gaussian channel of type (4.22), is able to isolate a pure SK part, described by the Parisi solution, plus a classical one dimensional variational principle.

Remark 4.2. In what follows, we will exploit the fact that the quenched pressure has bounded derivative in the external biases λ. Indeed, thanks to Cauchy-Schwartz inequality and to $\mathbb{E}[\xi_1^4] < \infty$ we have

$$\left| \frac{d}{d\lambda}\bar{p}_N(\mu,\nu,\lambda) \right| = |\mathbb{E}\langle m_N \rangle_N| \leq \frac{1}{N}\sum_{i=1}^N |\mathbb{E}\langle \sigma_i \xi_i \rangle_N| \leq \frac{1}{N}\sum_{i=1}^N \sqrt{\mathbb{E}[\xi_i^2]} = \sqrt{\mathbb{E}[\xi_1^2]} =: \sqrt{a}\,. \quad (4.36)$$

The previous bound holds for all $\mu,\nu \geq 0$ and $\lambda \in \mathbb{R}$. In particular, it holds for $\nu = 0$, namely for the SK model where this implies that $\bar{p}_N^{SK}(\cdot,h)$ is Lipschitz in h and so will be its limit $\mathcal{P}(\cdot,h)$. Furthermore, since the interpolating model (4.33) is of the type (4.1) it inherits these Lipschitz properties on its quenched pressure (4.35).

Proposition 4.7. *The following sum rule holds:*

$$\bar{p}_N(\mu,\nu,\lambda) = \bar{p}_N^{SK}(\sqrt{\mu},\lambda + R_\epsilon(1)) - \frac{\nu}{2}\int_0^1 dt[r_\epsilon^2(t) - \Delta_\epsilon(t)] + \mathcal{O}(s_N) \quad (4.37)$$

where

$$\Delta_\epsilon(t) := \mathbb{E}\Big\langle \big(m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) - r_\epsilon(t)\big)^2 \Big\rangle_{N,R_\epsilon(t)}. \quad (4.38)$$

Proof. Let us begin by computing the first derivative

$$\dot{\bar{p}}_N(t) = \mathbb{E}\Big\langle -\frac{\nu}{2}m_N^2(\boldsymbol{\sigma}|\boldsymbol{\xi}) + \nu r_\epsilon(t)m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) \Big\rangle_{N,R_\epsilon(t)} = \frac{\nu}{2}r_\epsilon^2(t) - \frac{\nu}{2}\Delta_\epsilon(t)\,. \quad (4.39)$$

Remark 4.2 implies that

$$\bar{p}_N(0) = \bar{p}_N(\mu,\nu,\lambda) + \mathcal{O}(s_N)\,; \quad (4.40)$$

$$\bar{p}_N(1) = \frac{1}{N}\mathbb{E}\log\sum_{\boldsymbol{\sigma}\in\Sigma_N}\exp\left[-\sqrt{\mu}H_N^{SK}(\boldsymbol{\sigma}) + (R_\epsilon(1)+\lambda)\sum_{i=1}^N \xi_i\sigma_i\right]$$
$$= \bar{p}_N^{SK}(\sqrt{\mu},R_\epsilon(1)+\lambda)\,. \quad (4.41)$$

An application of the fundamental theorem of calculus yields the result. $\qquad\square$

Remark 4.3. By looking at the remainder $\Delta_\epsilon(t)$ in the sum rule one may be led to choose the interpolating function as

$$r_\epsilon(t) = \mathbb{E}\langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,R_\epsilon(t)} \quad (4.42)$$

in order to apply Proposition 4.3 in some suitable form and make $\Delta_\epsilon(t)$ vanish in the thermodynamic limit. As for the multi-species models treated earlier, the choice (4.42) can be formalized by means of the ODE

$$\dot{R}_\epsilon(t) = \nu\mathbb{E}\langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,R_\epsilon(t)} =: G_N(t, R_\epsilon(t)), \quad R_\epsilon(0) = \epsilon, \tag{4.43}$$

which has always a solution by Cauchy-Lipschitz theorem because the velocity field G_N is Lipschitz for fixed N in the spatial coordinate R_ϵ

$$\frac{\partial}{\partial R_\epsilon} G_N(t, R_\epsilon(t)) = \nu N\mathbb{E}\Big\langle \big(m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) - \langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,R_\epsilon(t)}\big)^2 \Big\rangle_{N,R_\epsilon(t)} \geq 0. \tag{4.44}$$

Furthermore, by Liouville's formula (see Lemma 2.4) and the previous equation we know that the Jacobian $\partial_\epsilon R_\epsilon$ satisfies

$$\frac{\partial R_\epsilon(t)}{\partial \epsilon} = \exp\left\{\int_0^t ds\, \frac{\partial G_N(s, R_\epsilon(s))}{\partial R_\epsilon}\right\} \geq 1. \tag{4.45}$$

Equations (4.43), (4.44) and (4.45) together provide a rigorous justification to the choice (4.42).

4.3.2 Proofs of Theorem 4.1, Corollary 4.2 and Proposition 4.3

All of the above results require the L^2 convergence of the random pressure towards the limit of its expectations and a preliminary control on the fluctuations of the Mattis magnetization which are respectively contained in the two following lemmas.

Lemma 4.8 (Self-averaging of the pressure). *If $\mathbb{E}[\xi_1^4] < \infty$ then*

$$\mathbb{E}\left[(p_N(\mu, \nu, \lambda) - \bar{p}_N(\mu, \nu, \lambda))^2\right] \leq \frac{K(\mu, \nu, \lambda)}{N}, \quad K(\mu, \nu, \lambda) = C_1\mu + C_2\nu^2 + C_3\lambda^2 \tag{4.46}$$

with $C_1, C_2, C_3 > 0$.

Proof. The random pressure p_N is a function of the random variables $(\mathbf{Z}, \boldsymbol{\xi})$. For this proof we stress this dependency by writing $p_N(\mathbf{Z}, \boldsymbol{\xi})$. Define $\mathbf{Z}^{(ij)} = (z_{12}, z_{13}, \ldots, z'_{ij}, \ldots, z_{N,N-1})$ and $\boldsymbol{\xi}^{(i)} = (\xi_1, \xi_2, \ldots, \xi'_i, \ldots, \xi_N)$ where $z'_{ij} \sim \mathcal{N}(0, 1)$ and $\xi'_i \sim P_\xi$ are independent of anything else. Then, by Efron-Stein inequality

$$\mathbb{E}\left[(p_N(\mu, \nu, \lambda) - \bar{p}_N(\mu, \nu, \lambda))^2\right] \equiv \mathbb{V}[p_N(\mathbf{Z}, \boldsymbol{\xi})] \leq$$

$$\leq \frac{1}{2}\sum_{i,j=1}^N \mathbb{E}\left[\big(p_N(\mathbf{Z}^{(ij)}, \boldsymbol{\xi}) - p_N(\mathbf{Z}, \boldsymbol{\xi})\big)^2\right] + \frac{1}{2}\sum_{i=1}^N \mathbb{E}\left[\big(p_N(\mathbf{Z}, \boldsymbol{\xi}^{(i)}) - p_N(\mathbf{Z}, \boldsymbol{\xi})\big)^2\right]. \tag{4.47}$$

Let us focus on the terms in the first sum. By Lagrange's mean value theorem we have that there exists a $\tilde{z}_{ij} \in (\min(z_{ij}, z'_{ij}), \max(z_{ij}, z'_{ij}))$ such that

$$
\left(p_N(\mathbf{Z}^{(ij)}, \boldsymbol{\xi}) - p_N(\mathbf{Z}, \boldsymbol{\xi})\right)^2 = \left(\frac{\partial p_N}{\partial z_{ij}}\bigg|_{\tilde{z}_{ij}}\right)^2 (z_{ij} - z'_{ij})^2 =
$$
$$
= \left(\frac{1}{N}\sqrt{\frac{\mu}{2N}}\langle\sigma_i\sigma_j\rangle_{N,\tilde{z}_{ij}}\right)^2 (z_{ij} - z'_{ij})^2 \le \frac{\mu}{2N^3}(z_{ij} - z'_{ij})^2 \quad (4.48)
$$

where by $\langle\cdot\rangle_{N,\tilde{z}_{ij}}$ we mean the Boltzmann-Gibbs measure where z_{ij} has been replaced with $\tilde{z}_{ij}$ in the Hamiltonian (4.1). In a really similar fashion we estimate the second set of terms. Again, let $\tilde{\xi}_i \in (\min(\xi_i, \xi'_i), \max(\xi_i, \xi'_i))$

$$
\left(p_N(\mathbf{Z}, \boldsymbol{\xi}^{(i)}) - p_N(\mathbf{Z}, \boldsymbol{\xi})\right)^2 = \left(\frac{\partial p_N}{\partial \xi_i}\bigg|_{\tilde{\xi}_i}\right)^2 (\xi_i - \xi'_i)^2 =
$$
$$
= \left[\frac{\nu}{N^2}\left(\sum_{j\neq i,1}^{N} \xi_j\langle\sigma_i\sigma_j\rangle_{N,\tilde{\xi}_i} + \tilde{\xi}_i\right) + \frac{\lambda}{N}\langle\sigma_i\rangle_{N,\tilde{\xi}_i}\right]^2 (\xi_i - \xi'_i)^2 =
$$
$$
= \left[\frac{\nu}{N^2}\left(\sum_{j\neq i,1}^{N} \xi_j\langle\sigma_i\sigma_j\rangle_{N,\tilde{\xi}_i} + \tilde{\xi}_i\right) + \frac{\lambda}{N^2}\sum_{j=1}^{N}\langle\sigma_i\rangle_{N,\tilde{\xi}_i}\right]^2 (\xi_i - \xi'_i)^2. \quad (4.49)
$$

Notice that in the square bracket we have an overall sum of $2N$ terms. We can use Jensen's inequality to bring the square inside the sums. The last line of the previous is bounded by

$$
2N\left[\frac{\nu^2}{N^4}\left(\sum_{j\neq i,1}^{N} \xi_j^2\langle\sigma_i\sigma_j\rangle_{N,\tilde{\xi}_i}^2 + \tilde{\xi}_i^2\right) + \frac{\lambda^2}{N^4}\sum_{i=1}^{N}\langle\sigma_i\rangle_{N,\tilde{\xi}_i}^2\right](\xi_i - \xi'_i)^2 \quad (4.50)
$$

whence, exploiting the fact that $\tilde{\xi}_i^2 \le \max(\xi_i^2, \xi'^2_i) \le \xi_i^2 + \xi'^2_i$

$$
\left(p_N(\mathbf{Z}, \boldsymbol{\xi}^{(i)}) - p_N(\mathbf{Z}, \boldsymbol{\xi})\right)^2 \le \frac{2}{N^3}\left[\nu^2\left(\sum_{j=1}^{N}\xi_j^2 + \xi'^2_i\right) + N\lambda^2\right](\xi_i - \xi'_i)^2. \quad (4.51)
$$

From the previous equation one can clearly see that ξ appears at most at the 4th power on the r.h.s. Hence, thanks to the hypothesis, inserting the estimates (4.48) and (4.51) into (4.47) we get the claimed inequality. □

Lemma 4.9. *Let $y \in [y_1, y_2]$, $\delta \in (0, 1)$ and denote by $\langle\cdot\rangle_{N,y}$ the Boltzmann-Gibbs expectation*

associated to the Hamiltonian $H_N(\sigma; \mu, \nu, \lambda + y)$. Then

$$\mathbb{E}\Big\langle \big(m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) - \langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,y}\big)^2 \Big\rangle_{N,y} = \frac{1}{N}\frac{d^2}{dy^2}\bar{p}_N(\mu, \nu, \lambda + y) \tag{4.52}$$

$$\mathbb{E}\left[\big(\langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,y} - \mathbb{E}\langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,y}\big)^2\right] \leq \frac{12K(\mu, \nu, |\lambda| + |y| + 1)}{\delta^2 N} +$$

$$+ 8\sqrt{a}\frac{d}{dy}\left[\bar{p}_N(\mu, \nu, \lambda + y + \delta) - \bar{p}_N(\mu, \nu, \lambda + y - \delta)\right] \tag{4.53}$$

with $a := \mathbb{E}\xi_1^2$.

Proof. The concentration property (4.53) can be obtained from the self-averaging and the convexity properties of the pressure density, proved in Lemma 4.8, using a well-know argument in spin glass theory [10, 11], already employed in the optimal setting in Appendix A.1. The version of that argument applied here is analogous to the one appearing in [76]. In order to lighten the notation we neglect subscripts in the brackets for this proof. (4.52) follows from a simple computation of the second derivative on the r.h.s. Let us skip directly to (4.53). It is easy to see that both p_N and $\bar{p}_N$ are convex in the external biases λ. We first evaluate the difference

$$\left|\frac{d}{dy}\left[p_N(\mu, \nu, \lambda + y) - \bar{p}_N(\mu, \nu, \lambda + y)\right]\right| = \left|\langle m_N\rangle - \mathbb{E}\langle m_N\rangle\right|. \tag{4.54}$$

The difference between two convex differentiable functions can be bounded (see Lemma 3.2 in [18]) from above as follows

$$\left|\frac{d}{dy}\left[p_N(\mu, \nu, \lambda + y) - \bar{p}_N(\mu, \nu, \lambda + y)\right]\right| \leq \frac{1}{\delta}\sum_{u=y\pm\delta,\,y}|p_N(\mu, \nu, \lambda + u) - \bar{p}_N(\mu, \nu, \lambda + u)| +$$

$$+ \frac{d}{dy}\left(\bar{p}_N(\mu, \lambda + y + \delta) - \bar{p}_N(\mu, \lambda + y - \delta)\right) \tag{4.55}$$

for any $\delta > 0$. For our purposes, it is sufficient to restrict ourselves to $\delta \in (0, 1)$. By squaring both sides, averaging w.r.t. the disorder and using Jensen's inequality we get

$$\mathbb{E}\left[(\langle m_N\rangle - \mathbb{E}\langle m_N\rangle)^2\right] \leq \frac{4}{\delta^2}\sum_{u=y\pm\delta,\,y}\mathbb{E}\left[(p_N(\mu, \nu, \lambda + u) - \bar{p}_N(\mu, \nu, \lambda + u)^2\right] +$$

$$+ 4\left[\frac{d}{dy}\left(\bar{p}_N(\mu, \nu, \lambda + y + \delta) - \bar{p}_N(\mu, \nu, \lambda + y - \delta)\right)\right]^2. \tag{4.56}$$

By Lemma 4.8, each of the three terms in the first sum of the previous equation can be bounded by $K(\mu, \nu, |\lambda| + |y| + 1)/N$ and this explains the first term in (4.53). Concerning the second,

notice that the derivative in the square brackets is positive thanks to the convexity of $\bar{p}_N$ and bounded as seen in Remark 4.2. The previous considerations imply that

$$\left[\frac{d}{dy} \left(\bar{p}_N(\mu, \nu, \lambda + y + \delta) - \bar{p}_N(\mu, \nu, \lambda + y - \delta) \right) \right]^2 \leq$$

$$\leq 2\sqrt{a} \left[\frac{d}{dy} \left(\bar{p}_N(\mu, \nu, \lambda + y + \delta) - \bar{p}_N(\mu, \nu, \lambda + y - \delta) \right) \right] , \quad (4.57)$$

which concludes the proof.   □

We start with Proposition 4.3 that is a direct consequence the previous Lemma.

Proof of Proposition 4.3. For future convenience we introduce the notation $\mathbb{E}_\epsilon[\cdot] = \frac{1}{s_N} \int_{s_N}^{2s_N} (\cdot)$. We first decompose the quenched variance

$$\mathbb{E}\left\langle \left(m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) - \mathbb{E}\langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,\epsilon} \right)^2 \right\rangle_{N,\epsilon} = \mathbb{E}\left\langle \left(m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) - \langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,\epsilon} \right)^2 \right\rangle_{N,\epsilon} +$$

$$+ \mathbb{E}\left[\left(\langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,\epsilon} - \mathbb{E}\langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,\epsilon}^2 \right)^2 \right] . \quad (4.58)$$

The first term in the r.h.s. of the previous equation is the contribution due to the thermal fluctuations in the model, whilst the second one is due to the disorder.

Thermal fluctuations: Consider (4.52) with $y \equiv \epsilon \in [s_N, 2s_N]$ and take the expectation $\mathbb{E}_\epsilon$ of both sides:

$$\Delta_T := \mathbb{E}_\epsilon \mathbb{E}\left\langle \left(m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) - \langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,\epsilon} \right)^2 \right\rangle_{N,\epsilon} = \frac{1}{Ns_N} \int_{s_N}^{2s_N} d\epsilon \frac{d^2}{d\epsilon^2} \bar{p}_N(\mu, \nu, \lambda + \epsilon) \quad (4.59)$$

Now, recalling that the derivatives of the pressure are bounded (see (4.36)) we immediately conclude that

$$\Delta_T = \mathcal{O}\left(\frac{1}{Ns_N} \right) . \quad (4.60)$$

Disorder fluctuations: Analogously take (4.53) with $y \equiv \epsilon \in [s_N, 2s_N]$ and average w.r.t. ϵ on both sides. Considering that $\epsilon \leq 1$ we have

$$\Delta_D := \mathbb{E}_\epsilon \mathbb{E}\left[\left(\langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,\epsilon} - \mathbb{E}\langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi})\rangle_{N,\epsilon} \right)^2 \right] \leq \frac{12K(\mu, \nu, |\lambda| + 2)}{\delta^2 N} +$$

$$+ \frac{8\sqrt{a}}{s_N} \int_{s_N}^{2s_N} d\epsilon \frac{d}{d\epsilon} \left[\bar{p}_N(\mu, \nu, \lambda + \epsilon + \delta) - \bar{p}_N(\mu, \nu, \lambda + \epsilon - \delta) \right] . \quad (4.61)$$

The last integral can be explicitly computed and then bounded by $4\delta\sqrt{a}$ thanks to Lagrange's mean value theorem and (4.36). Hence

$$\Delta_D = \mathcal{O}\left(\frac{1}{\delta^2 N} + \frac{\delta}{s_N}\right) \tag{4.62}$$

which is optimized when $\delta = (s_N/N)^{1/3}$ (consistently with $\delta \in (0,1)$). This choice leads to

$$\Delta_D = \mathcal{O}\left(\frac{1}{s_N^{2/3} N^{1/3}}\right) = \mathcal{O}\left(N^{\frac{2\alpha-1}{3}}\right). \tag{4.63}$$

The latter and (4.60) both vanish in the $N \to \infty$ limit for $\alpha \in (0, 1/2)$. $\qquad\square$

We are finally ready for the proof of Theorem 4.1.

Proof of Theorem 4.1. The variational principle is proven by means of two bounds that match in the thermodynamic limit. The lower bound follows from the classical sum rule combined with the positivity of the square. The upper bound is obtained with the adaptive interpolation method. For the sake of clarity we consider each of them separately and then we prove (4.11).

Lower bound: Let us consider the sum rule (4.37) with the choice $r_\epsilon(t) = x$ constant in t. Furthermore observe that the remainder $\Delta_\epsilon(t)$ is always positive, so we discard it at the expense of an inequality:

$$\bar{p}_N(\mu, \nu, \lambda) \geq \bar{p}_N^{SK}(\sqrt{\mu}, \lambda + \epsilon + \nu x) - \frac{\nu x^2}{2} + \mathcal{O}(s_N). \tag{4.64}$$

As explained in Remark 4.2 $\bar{p}_N^{SK}$ is Lipschitz in its second entry. This allows us to reabsorb the perturbation ϵ into $\mathcal{O}(s_N)$. By sending $N \to \infty$ one obtains the bound

$$\liminf_{N\to\infty} \bar{p}_N(\mu, \nu, \lambda) \geq -\frac{\nu x^2}{2} + \mathcal{P}(\sqrt{\mu}, \nu x + \lambda) \tag{4.65}$$

which is uniform in x. We can optimize it by taking the $\sup_{x\in\mathbb{R}}$ on the r.h.s.

Upper bound: From (4.52) we see that any quenched pressure of the type (4.6) is convex in its third entry. Then, starting from the sum rule (4.37) we can use Jensen's inequality on the SK quenched pressure to obtain an upper bound

$$\bar{p}_N(\mu, \nu, \lambda) \leq \mathcal{O}(s_N) + \int_0^1 dt \left[-\frac{\nu r_\epsilon^2(t)}{2} + \bar{p}_N^{SK}(\sqrt{\mu}, \lambda + \epsilon + \nu r_\epsilon(t))\right] + \frac{\nu}{2}\int_0^1 dt\, \Delta_\epsilon(t). \tag{4.66}$$

As done in the lower bound, we throw the dependence on ϵ in $\bar{p}_N^{SK}$ into $\mathcal{O}(s_N)$ and use Guerra's uniform bound $\bar{p}_N^{SK} \leq \mathcal{P}$ [7]:

$$\bar{p}_N(\mu, \nu, \lambda) \leq \mathcal{O}(s_N) + \int_0^1 dt \left[-\frac{\nu r_\epsilon^2(t)}{2} + \mathcal{P}(\sqrt{\mu}, \lambda + \nu r_\epsilon(t)) \right] + \frac{\nu}{2} \int_0^1 dt \, \Delta_\epsilon(t) \leq$$

$$\leq \mathcal{O}(s_N) + \sup_{x \in \mathbb{R}} \varphi(x; \mu, \nu, \lambda) + \frac{\nu}{2} \int_0^1 dt \, \Delta_\epsilon(t) . \quad (4.67)$$

The only remaining dependency on the interpolation path is in $\Delta_\epsilon(t)$. To make the two bounds match we have to make sure the remainder vanishes in the limit. Hence, as suggested in Remark 4.3, we choose $r_\epsilon(\cdot)$ as in (4.42). At this point we can decompose $\Delta_\epsilon(t)$ as done in the proof of Proposition 4.3

$$\Delta_\epsilon(t) = \mathbb{E} \Big\langle \big(m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) - \langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) \rangle_{N,R_\epsilon(t)} \big)^2 \Big\rangle_{N,R_\epsilon(t)} +$$

$$+ \mathbb{E} \Big[\big(\langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) \rangle_{N,R_\epsilon(t)} - \mathbb{E} \langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) \rangle_{N,R_\epsilon(t)} \big)^2 \Big] . \quad (4.68)$$

Let us first bound the ϵ-average of the first term on the r.h.s. Using (4.52) and the inequality (4.45) on the Jacobian we get

$$\frac{1}{s_N} \int_{s_N}^{2s_N} d\epsilon \, \mathbb{E} \Big\langle \big(m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) - \langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) \rangle_{N,R_\epsilon(t)} \big)^2 \Big\rangle_{N,R_\epsilon(t)} =$$

$$= \frac{1}{N s_N} \int_{s_N}^{2s_N} d\epsilon \, \frac{d^2}{dy^2} \bar{p}_N(\mu, (1-t)\nu, \lambda + y) \Big|_{y=R_\epsilon(t)} \leq$$

$$\leq \frac{1}{N s_N} \int_{R_{s_N}(t)}^{R_{2s_N}(t)} dy \, \frac{d^2}{dy^2} \bar{p}_N(\mu, (1-t)\nu, \lambda + y) = \mathcal{O}\left(\frac{1}{N s_N} \right) \quad (4.69)$$

where the last equality follows from the bound on derivatives (4.36).

For the second term in the r.h.s. of (4.68) we use (4.53) and take its ϵ-average:

$$\frac{1}{s_N} \int_{s_N}^{2s_N} d\epsilon \, \mathbb{E} \Big[\big(\langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) \rangle_{N,R_\epsilon(t)} - \mathbb{E} \langle m_N(\boldsymbol{\sigma}|\boldsymbol{\xi}) \rangle_{N,R_\epsilon(t)} \big)^2 \Big] \leq \mathcal{O}\left(\frac{1}{N \delta^2} \right) +$$

$$+ \frac{8\sqrt{a}}{s_N} \int_{s_N}^{2s_N} d\epsilon \, \frac{d}{dy} [\bar{p}_N(\mu, (1-t)\nu, \lambda + y + \delta) - \bar{p}_N(\mu, (1-t)\nu, \lambda + y - \delta)] \Big|_{y=R_\epsilon(t)} . \quad (4.70)$$

Now, thanks again to inequality (4.45) and to the fact that the derivative of the square bracket is positive the integral in the previous equation can be bounded by

$$\int_{R_{s_N}(t)}^{R_{2s_N}(t)} dy \, \frac{d}{dy} [\bar{p}_N(\mu, (1-t)\nu, \lambda + y + \delta) - \bar{p}_N(\mu, (1-t)\nu, \lambda + y - \delta)] \leq 4\sqrt{a}\delta . \quad (4.71)$$

The last inequality follows from an application of the mean value theorem and (4.36). Equations (4.69), (4.70) and (4.71) together imply that

$$\mathbb{E}_\epsilon[\Delta_\epsilon(t)] = \mathcal{O}\left(\frac{1}{Ns_N} + \frac{1}{N\delta^2} + \frac{\delta}{s_N}\right), \tag{4.72}$$

that vanishes in the thermodynamic limit for $\delta = (s_N/N)^{1/3}$ and $s_N = 1/2N^{-\alpha}$ with $\alpha \in (0, 1/2)$ as seen for Proposition 4.3. With this information, we take the ϵ-average on both sides of (4.67) and by Fubini's Theorem and dominated convergence we have

$$\limsup_{N\to\infty} \bar{p}_N(\mu, \nu, \lambda) \leq \sup_{x\in\mathbb{R}} \varphi(x; \mu, \nu, \lambda). \tag{4.73}$$

The two bounds, together with Lemma 4.8, conclude the proof of the variational principle (4.9). □

Remark 4.4. The upper bound in the proof of (4.9) can also be obtained by adapting the elegant technique used in [46]. We opted instead for a proof that explicitly identifies the physical meaning of the vanishing distance between the upper and lower bounds in terms of the fluctuation of the order parameter. Such crucial thermodynamic property (Proposition 4.3) holds independently of the solution and it is at the origin of the (ordinary) variational principle in (4.9).

Remark 4.5. Since for $\nu > 0$ $\mathcal{P}(\beta, h)$ is h-Lipschitz we have

$$\lim_{|x|\to\infty} \varphi(x; \mu, \nu, \lambda) = -\infty, \tag{4.74}$$

therefore the supremum of $\varphi(\,\cdot\,; \mu, \nu, \lambda)$ will be attained at a finite $\bar{x} \in \mathbb{R}$. Furthermore the necessary condition for $\bar{x}$ to be a maximum point is

$$\bar{x} = \partial_h \mathcal{P}(\sqrt{\mu}, h)|_{h=\nu\bar{x}+\lambda} \tag{4.75}$$

that in turn implies $\bar{x} \in [-\sqrt{a}, \sqrt{a}]$ by (1.64). Hence one can take the supremum only over $[-\sqrt{a}, \sqrt{a}]$.

Proof of Corollary 4.2.

λ-*differentiability:* Set $\Omega(\mu, \nu, \lambda) := \text{argmax}_{[-\sqrt{a}, \sqrt{a}]} \varphi(\,\cdot\,; \mu, \nu, \lambda)$. Then, since $p(\mu, \nu, \lambda)$ is convex in λ, by Danskin's theorem (see [46] for instance) we have that the left and right derivatives satisfy respectively

$$\frac{d}{d\lambda_-}p(\mu, \nu, \lambda) = \min_{x\in\Omega(\mu,\nu,\lambda)} \frac{\partial}{\partial\lambda}\varphi(x; \mu, \nu, \lambda) = \min_{x\in\Omega(\mu,\nu,\lambda)} \frac{\partial}{\partial h}\mathcal{P}(\sqrt{\mu}, h)|_{h=\nu x+\lambda} \tag{4.76}$$

$$\frac{d}{d\lambda_+}p(\mu, \nu, \lambda) = \max_{x\in\Omega(\mu,\nu,\lambda)} \frac{\partial}{\partial\lambda}\varphi(x; \mu, \nu, \lambda) = \max_{x\in\Omega(\mu,\nu,\lambda)} \frac{\partial}{\partial h}\mathcal{P}(\sqrt{\mu}, h)|_{h=\nu x+\lambda}. \tag{4.77}$$

If $\Omega(\mu, \nu, \lambda)$ is a singleton then $p(\mu, \nu, \lambda)$ is differentiable. Conversely, suppose that $p(\mu, \nu, \lambda)$ is differentiable and that there are at least two distinct values $x_1, x_2 \in \Omega(\mu, \nu, \lambda)$, $x_1 < x_2$. Then we have

$$\frac{d}{d\lambda_-} p(\mu, \nu, \lambda) \leq \frac{\partial}{\partial h} \left. \mathcal{P}(\sqrt{\mu}, h) \right|_{h = \nu x_1 + \lambda} = x_1 < x_2 = \frac{\partial}{\partial h} \left. \mathcal{P}(\sqrt{\mu}, h) \right|_{h = \nu x_2 + \lambda} \leq \frac{d}{d\lambda_+} p(\mu, \nu, \lambda)$$
(4.78)

that is a contradiction.

Assume now that there is a unique maximum point $\bar{x}$. Thanks to the convexity of the sequence $\bar{p}_N$ in λ and Danskin's theorem we can write

$$\lim_{N \to \infty} \mathbb{E} \langle m_N(\boldsymbol{\sigma} | \boldsymbol{\xi}) \rangle_N = \lim_{N \to \infty} \frac{d}{d\lambda} \bar{p}_N(\mu, \nu, \lambda) = \frac{d}{d\lambda} p(\mu, \nu, \lambda) = \frac{\partial}{\partial \lambda} \varphi(\bar{x}; \mu, \nu, \lambda) =$$
$$= \frac{\partial}{\partial \lambda} \mathcal{P}(\sqrt{\mu}, \nu \bar{x} + \lambda) = \frac{\partial}{\partial h} \left. \mathcal{P}(\sqrt{\mu}, h) \right|_{h = \nu \bar{x} + \lambda} = \bar{x},$$
(4.79)

where it is understood that only explicit dependence on λ is taken into account when the partial derivative is taken.

ν-differentiability: The proof relies on Danskin's theorem and is a straightforward consequence of that of Proposition 2 in [46].

μ-differentiability at $\lambda = 0$: Notice that when ξ is centered then $\varphi(x; \mu, \nu, 0)$ is symmetric in x. The result then follows easily again from Danskin's Theorem (as in Theorem 2 in [46]) and from the differentiability properties w.r.t. $\beta = \sqrt{\mu}$ of the Parisi pressure in Theorem 14.11.6 of [23] and [25].

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

4.3.3   Proof of Proposition 4.4 and Proposition 4.5

Proof of Proposition 4.4. The fact that φ_{RS} is an even function of x follows directly from the symmetry of the random variable ξ. By Remark 4.5, when $|x| \to \infty$, the term $-\mu x^2/2$ is dominant in (4.15) bringing φ_{RS} to $-\infty$. As a consequence the maximum point(s) of φ_{RS} are critical point(s). The vanishing derivative condition yields

$$\frac{d\varphi_{RS}}{dx} = -\mu x + \mu \mathbb{E} \xi \tanh \left(z \sqrt{\mu q(x, \mu, a)} + \mu \xi x \right) = -\mu x + \mu^2 a x \left(1 - q(x, \mu, a) \right) = 0 \quad (4.80)$$

that is

$$x = 0 \quad \text{or} \quad q(x, \mu, a) = 1 - \frac{1}{\mu a}. \tag{4.81}$$

Since the function is $q(x, \mu, a)$ increasing for $x \geq 0$, the positive solution $\bar{x}(\mu, a)$ of (4.17) exists and is unique up to reflection if and only if

$$\lim_{x \to 0^+} q(x, \mu, a) \leq 1 - \frac{1}{\mu a} \tag{4.82}$$

which is equivalent to (4.18).

Consider now $a \geq 1$. If $1/a \leq \mu \leq 1$ we have $q(0, \mu, 0) = 0$, thus (4.18) is clearly satified. Furthermore, it turns out that

$$\mu > 1 \quad \Rightarrow \quad AT(\mu, 0) > 1. \tag{4.83}$$

In fact for $a = 0$ (4.1) reduces to an SK model with zero external magnetic field at temperature $\sqrt{\mu}$. Fix $\mu > 1$ and assume that $AT(\mu, 0) \leq 1$. Then for any $\epsilon > 0$ by the monotonicity of $q(x, \mu, a)$

$$\mu \mathbb{E} \cosh^{-4} \left(z \sqrt{\mu q(\epsilon, \mu, 1) + \epsilon^2 \mu^2} \right) < AT(\mu, 0) \leq 1. \tag{4.84}$$

[22] implies that the Parisi measure is $\chi^*(\sqrt{\mu}, \epsilon\mu) = \delta_{q(\epsilon, \mu, 1)}$ and $\chi^*(\sqrt{\mu}, \epsilon\mu) \longrightarrow \delta_{q(0, \mu, 0)}$ weakly. Since $\mathcal{P}(\beta, h)$ is continuous in h and the Parisi functional $\mathcal{P}(\chi; \beta, h)$ is weakly continuous we have that

$$\mathcal{P}(\sqrt{\mu}, 0) = \lim_{\epsilon \to 0} \mathcal{P}(\sqrt{\mu}, \epsilon\mu) = \mathcal{P}(\delta_{q(0, \mu, 0)}; \sqrt{\mu}, 0). \tag{4.85}$$

However for $\mu > 1$ we have $\mathcal{P}(\sqrt{\mu}, 0) < \mathcal{P}(\delta_{q(0, \mu, 0)}; \sqrt{\mu}, 0)$ thus the latter is a contradiction coming from the assumption $AT(\mu, 0) \leq 1$. This proves (4.83). Hence

$$1 < \mu \mathbb{E} \cosh^{-4}(z\sqrt{\mu q(0, \mu, 0)}) \leq \mu \mathbb{E} \cosh^{-2}(z\sqrt{\mu q(0, \mu, 0)}) = \mu(1 - q(0, \mu, 0)) \tag{4.86}$$

from which, when $a \geq 1$,

$$q(0, \mu, 0) < 1 - \frac{1}{\mu} \leq 1 - \frac{1}{\mu a}. \tag{4.87}$$

Finally, the solution to (4.17) is stable w.r.t. the optimization, indeed

$$\frac{d^2 \varphi_{RS}(x; \mu, a)}{dx^2}\bigg|_{x = \bar{x}(\mu, a)} = -\mu + \mu^2 a \left(1 - q(x, \mu, a)\right) - \mu^2 a \bar{x}(\mu, a) \frac{dq}{dx}(\bar{x}(\mu, a), \mu) =$$

$$= -\mu^2 a \bar{x}(\mu, a) \frac{dq}{dx}(\bar{x}(\mu, a), \mu) < 0 \tag{4.88}$$

thanks to the monotonicity of $q(x, \mu, a)$. The result for $x = -\bar{x}(\mu, a)$ follows by symmetry. $\quad\square$

Proof of Proposition 4.5. By proposition 4.4 there exists a unique (non negative) maximum point $\bar{x}(\mu, a)$ of $\varphi_{RS}(x; \mu, a)$. Given $(\mu, a) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ we introduce the following subset of the real line:

$$RS(\mu, a) = \left\{ x \in \mathbb{R} \mid \mu \mathbb{E} \cosh^{-4} \left(z\sqrt{\mu q(x, \mu, a) + \mu^2 x^2 a} \right) \leq 1 \right\}. \tag{4.89}$$

Clearly by definition $AT(\mu, a) \leq 1 \iff \bar{x}(\mu, a) \in RS(\mu, a)$. We start assuming that $\bar{x}(\mu, a) \in RS(\mu, a)$. Let us denote by $\varphi(x; \mu, a)$ the variational potential (4.10) specialized in the current setting, namely with $\nu = \mu$, $\lambda = 0$ and $\xi \sim \mathcal{N}(0, a)$. From (4.65) we already know that

$$\liminf_{N \to \infty} \bar{p}_N(\mu, a) \geq \varphi(x; \mu, a) \tag{4.90}$$

uniformly on x. Hence we can optimize (4.90) only over the region $RS(\mu, a)$ obtaining the lower bound:

$$\liminf_{N \to \infty} \bar{p}_N(\mu, a) \geq \sup_{x \in RS(\mu, a)} \varphi(x; \mu, a) \,. \tag{4.91}$$

The choice of restricting the supremum to the region $RS(\mu, a)$ allows us to replace in (4.91) the function φ with its its replica symmetric version φ_{RS}. Indeed again by [22] the AT condition is sufficient for the validity of the the replica symmetric solution of the SK model. Then from (4.91) one gets the lower bound

$$\liminf_{N \to \infty} \bar{p}_N(\mu, a) \geq \sup_{x \in RS(\mu, a)} \varphi_{RS}(x; \mu, a) \,. \tag{4.92}$$

For the upper bound we can exploit the fact that the pressure of the SK model is always bounded from above by the replica symmetric one [7]. Hence from the upper bound (4.73) we get

$$\limsup_{N \to \infty} \bar{p}_N(\mu, a) \leq \sup_x \varphi_{RS}(x; \mu, a) = \sup_{x \in RS(\mu, a)} \varphi_{RS}(x; \mu, a) \,. \tag{4.93}$$

where the last equality follows from the assumption $\bar{x}(\mu, a) \in RS(\mu, a)$. Summarising we just proved that

$$AT(\mu, a) \leq 1 \implies \lim_{N \to \infty} \bar{p}_N(\mu, a) = \sup_{x \in RS(\mu, a)} \varphi_{RS}(x; \mu, a) \,. \tag{4.94}$$

Notice that in the previous equality the supremum can be taken on the whole real line since we are assuming that $\bar{x}(\mu, a) \in RS(\mu, a)$.

Conversely, suppose that $\bar{x}(\mu, a) \in (RS(\mu, a))^c$. We are going to prove the replica symmetric solution cannot hold. By Theorem 4.1 we know that

$$\lim_{N \to \infty} \bar{p}_N(\mu, a) = \sup_{x \in \mathbb{R}} \varphi(x; \mu, a) = \varphi(\tilde{x}(\mu, a); \mu, a). \tag{4.95}$$

where $\tilde{x}(\mu, a)$ denotes a point where the supremum is attained. By Remark 4.5 one can say that $\tilde{x}(\mu, a) \in [-\sqrt{a}, \sqrt{a}]$. Let's consider two cases, first suppose that $\tilde{x}(\mu, a) \in RS(\mu, a)$, then using the result in [22] we have that

$$\varphi(\tilde{x}(\mu, a); \mu, a) = \varphi_{RS}(\tilde{x}(\mu, a); \mu, a) < \sup_{x \in \mathbb{R}} \varphi_{RS}(x; \mu, a) \tag{4.96}$$

where the last inequality follows from the assumption $\bar{x}(\mu, a) \in (RS(\mu, a))^c$. On the other hand if $\tilde{x}(\mu, a) \in (RS(\mu, a))^c$ it is known [28] that the pressure of the SK model is strictly smaller that its replica symmetric version, therefore

$$\varphi(\tilde{x}(\mu, a); \mu, a) < \varphi_{RS}(\tilde{x}(\mu, a); \mu, a) \leq \sup_{x \in \mathbb{R}} \varphi_{RS}(x; \mu, a). \tag{4.97}$$

In conclusion, we have just proved that

$$AT(\mu, a) > 1 \quad \Rightarrow \quad \lim_{N \to \infty} \bar{p}_N(\mu, a) < \sup_{x \in \mathbb{R}} \varphi_{RS}(x; \mu, a). \tag{4.98}$$

$\square$

4.4 Phase Diagram

This section collects the consequences of Propositions 4.4 and 4.5 and resumes how the phase diagram in Figure 4.1 is drawn.

When $a \leq 1$ the condition (4.18) is not trivial and identifies a curve that lies above $\mu = 1/a$. Below this curve, for $\mu > 1$, the unique stable maximizer of φ_{RS} is $\bar{x}(\mu, a) = 0$. The resulting $q(0, \mu, a) \equiv q(0, \mu, 0)$ has to be intended as the stable solution to the consistency equation for the overlap of an SK model at temperature $\sqrt{\mu}$ in absence of external magnetic field, which is known to be RSB for $\mu > 1$. Hence by (4.83)

$$AT(\mu, a) = AT(\mu, 0) = \mu \mathbb{E} \cosh^{-4}(z\sqrt{\mu q(0, \mu, 0)}) > 1. \tag{4.99}$$

This in turn implies the replica symmetry breaking in our model. The de Almeida-Thouless red line in the diagram represents the condition $AT(\mu, a) = 1$ and must lie above, or at most coincide with, the curve (4.18) since it must contain the entire RSB phase. The red region could contain a mixed phase in analogy with the SK model as explained in Remark 4.6.

From (4.99) it is also clear that in an RS phase we must have $\bar{x}(\mu, a) \neq 0$ for $\mu > 1$ otherwise $AT(\mu, a) > 1$. Similarly, for $a \geq 1$ and $1/a < \mu \leq 1$, $\bar{x}(\mu, a) = 0$ cannot be the solution to (4.17) either since

$$q(0, \mu, a) = q(0, \mu, 0) = 0 < 1 - \frac{1}{\mu a}. \tag{4.100}$$

Contrarily, in the green region, that is replica symmetric by Proposition 4.5, the unique possible maximizer is $\bar{x}(\mu, a) = 0$ because $\mu \leq 1/a$. Moreover, we have the following

Corollary 4.10 (of Proposition 4.5). *The model is always replica symmetric for any $a \geq 1$.*

Proof. Recall that for $\mu \leq 1$ one has trivially $AT(\mu, a) \leq 1$. In addition to that, thanks to Proposition 4.4 for $a \geq 1$ and $\mu \geq 1 \geq 1/a$ we can always assume (4.17). Hence

$$AT(\mu, a) \leq \mu \mathbb{E} \cosh^{-2}\left(z\sqrt{\mu q(\bar{x}(\mu, a), \mu, a) + \mu^2 \bar{x}(\mu, a)^2 a}\right) =$$

$$= \mu\left[1 - q(\bar{x}(\mu, a), \mu, a)\right] = \mu - \mu + \frac{1}{a} = \frac{1}{a} \leq 1. \quad (4.101)$$

The thesis follows from Proposition 4.5. □

Remark 4.6. Let us consider $P_\xi = (\delta_{\sqrt{a}} + \delta_{-\sqrt{a}})/2$, or equivalently $\xi_i = \sqrt{a}\tau_i$ with $\tau_i = \pm 1$. In this case, one can gauge away the signs of the variables ξ_i's in (4.1) by means of the $\mathbb{Z}_2$ gauge transformation

$$z_{ij} \mapsto z_{ij}\tau_i\tau_j, \quad \sigma_i \mapsto \sigma_i\tau_i \quad (4.102)$$

obtaining the Hamiltonian

$$\tilde{H}_N(\boldsymbol{\sigma}) = -\sum_{i,j=1}^{N}\left(z_{ij}\sqrt{\frac{\mu}{2N}}\sigma_i\sigma_j + \frac{\mu a}{2N}\sigma_i\sigma_j\right) \overset{\text{D}}{=} -\sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j, \quad J_{ij} \overset{\text{iid}}{\sim} \mathcal{N}\left(\frac{\mu a}{2N}, \frac{\mu}{2N}\right). \quad (4.103)$$

The latter describes an SK model with a peculiar parameterization. To see this it suffices to consider the parameterization [29], namely

$$\beta H_N^{SK}(\boldsymbol{\sigma}) = -\sum_{i,j=1}^{N} J_{ij}\sigma_i\sigma_j, \quad J_{ij} \overset{\text{iid}}{\sim} \mathcal{N}\left(\frac{\beta J_0}{2N}, \frac{\beta^2 J^2}{2N}\right) \quad (4.104)$$

and to identify $1/\beta J = T/J = 1/\sqrt{\mu}$ and $J_0/J = \sqrt{\mu}a$. This means that if we draw the phase diagram of the model (4.103) with $\sqrt{\mu}a$ and $1/\sqrt{\mu}$ on the x and y axes respectively we re-obtain the well known phase diagram of the SK model. In this diagram for instance the curves for fixed a are a family of hyperbolas, and among them $a = 1$ corresponds to the Nishimori line. It is then a simple exercise to show that the phase diagram of the SK model redrawn in the parameterization (4.103) is qualitatively similar to the one in Figure 4.1, meaning that the same phases are disposed in the same positions. In particular the Nishimori line is the vertical line $a = 1$.

We finally notice that the model studied in [46] can be seen as a special inference problem in a non-optimal setting where the receiver uses his own Rademacher guess to retrieve a binary signal of which he does not know the amplitude.

We conclude the analysis with the study of the behavior of the solution $\bar{x}(\mu, a)$ of the variational problem (4.21) around the critical point $(\mu, a) = (1, 1)$. By Proposition 4.4 we have that $\lim_{(\mu,a)\to(1,1)} \bar{x}(\mu, a) = 0$. Notice that the replica symmetric solution $\bar{x}(\mu, a)$ represents the limiting behaviour of the Mattis magnetization when $AT(\mu, a) \leq 1$ and it is not identically vanishing iff condition (4.18) is satisfied. By Proposition 4.4 and Corollary 4.10 the above conditions are always satisfied if $\mu a \geq 1$ and $a \geq 1$. Then it holds

Proposition 4.11. *Assuming that $\mu a \geq 1$ and $a \geq 1$ then $\bar{x}(\mu, a)$ is the unique (up to reflection) solution of*

$$\mathbb{E} \tanh^2 (Y(x, \mu, a)) = 1 - \frac{1}{\mu a}, \quad Y(x, \mu, a) = z\sqrt{\mu - \frac{1}{a}} + \mu x \xi \tag{4.105}$$

where $z \sim \mathcal{N}(0, 1)$, $\xi \sim \mathcal{N}(0, a)$ are independent Gaussian. Moreover for $(\mu, a) \to (1, 1)$ we have

$$(\bar{x}(\mu, a))^2 = \frac{(\mu - \frac{1}{a})\left[\frac{1}{\mu} - 1 + 2(\mu - \frac{1}{a})(1 + o(1))\right]}{t(\mu, a)(1 + o(\bar{x}(\mu, a)))} \tag{4.106}$$

where $t(\mu, a) = \mu^2 a \mathbb{E}\left(2 - \cosh\left(2z\sqrt{\mu - \frac{1}{a}}\right)\right) \cosh^{-4}\left(z\sqrt{\mu - \frac{1}{a}}\right)$.

Proof. Clearly (4.105) holds by Proposition 4.4. Using a Taylor expansion of $\tanh^2(b+y)$ around $y = 0$ up to order 3 one obtains

$$\mathbb{E} \tanh^2 (Y(x, \mu, a)) = \mathbb{E} \tanh^2 (Y(0, \mu, a)) + t(\mu, a)x^2 + g(x, \mu, a)$$

where $g(x, \mu, a) = \frac{(\mu x)^4}{4!}\mathbb{E}\frac{\partial^4}{\partial y^4} \tanh^2(y)\big|_{y=y(z,\xi,x,\mu,a)}\xi^4$. Since $|\frac{\partial^4}{\partial y^4} \tanh^2(y)| \leq costant$ uniformly on y, we have that $g(x, \mu, a) = o(x^3)$. Then one can write

$$\mathbb{E} \tanh^2 (Y(x, \mu, a)) = \mathbb{E} \tanh^2 \left(z\sqrt{\mu - \frac{1}{a}}\right) + t(\mu, a)x^2(1 + o(x)) \tag{4.107}$$

The term $\mathbb{E} \tanh^2 \left(z\sqrt{\mu - \frac{1}{a}}\right)$ can be represented using Taylor expansion of $\tanh^2(y)$ around $y = 0$ up to order 4 obtaining

$$\mathbb{E} \tanh^2 \left(z\sqrt{\mu - \frac{1}{a}}\right) = (\mu - \frac{1}{a}) - 2(\mu - \frac{1}{a})^2(1 + o(1)) \tag{4.108}$$

Combining (4.107) and (4.108) one obtains (4.106).

$\square$

The previous Proposition and in particular the expansion (4.106) can be used to obtain the critical behavior of $\bar{x}(\mu, a)$ as $(\mu, a) \to (1, 1)$ with the constraint $\mu a \geq 1$ and $a \geq 1$. As an example fixing $a = 1$ one gets

$$\lim_{\mu \to 1^+} \frac{\bar{x}(\mu, 1)}{\mu - 1} = 1. \tag{4.109}$$

Analogously if $\mu = 1$ and $a \to 1^+$

$$\lim_{a \to 1^+} \frac{\bar{x}(1, a)}{\sqrt{2}(1 - \frac{1}{a})} = 1. \tag{4.110}$$

More generally one can consider a family of hyperbolas

$$\mu_\alpha(a) = \frac{\alpha}{a} + 1 - \alpha, \quad \alpha \leq 1 \tag{4.111}$$

and define $x_\alpha(a) = \bar{x}(\mu_\alpha(a), a)$. Then expansion (4.106) leads to

$$\lim_{a \to 1^+} \left(\frac{x_\alpha(a)}{\frac{1}{a} - 1} \right)^2 = (\alpha - 1)(\alpha - 2). \tag{4.112}$$

The critical behavior around $(1, 1)$ of the magnetization along the above directions is therefore the same of the optimal setting.

4.5   Concluding remarks

To begin with, we point out that main result here, a nested variational principle over a distribution and a real number, can be extended beyond the Rademacher prior assumption, leading to an SK model with soft spins [77] with a Mattis interaction.

The variational principle (4.9) pinpoints the presence of the replica symmetry breaking phase in a mismatched inference problem. This is expected to have implications on the algorithms usually implemented to retrieve signal components, such as Approximate Message Passing (AMP). Indeed we have observed, with preliminary numerical tests, that in the RSB phase of the model with Gaussian signal distribution ten thousand iterations of AMP are not sufficient to reach convergence: the values of the local magnetizations keep oscillating. On the contrary less than a hundred were enough in the RS phase, thus supporting the picture in Figure 4.1. The rigorous characterization of the AMP convergence, done by E. Bolthausen [78] for the SK model, seems to be related to the Almeida-Thouless line and is left for future work.

It is interesting to notice that the model studied here is equivalent, through a Hubbard-Stratonovič transformation as done in [79, 80], to a Boltzmann Machine with one hidden analogic neuron linked to a visible layer of neurons in mean field disordered interaction, *i.e.* a non-restricted Boltzmann Machine. Our result extends also to a finite number of hidden analogic neurons and leads to a model that includes SK and Hopfield terms. In this regard we mention that the SK term can indeed be generated starting from the Hopfield model adding a form of *synaptic noise* [81, 82] (see eq. (8) in [82] in particular) that blurs the interactions, built from the patterns, precisely as in (4.22).

Chapter 5

Bayes-optimal limits in structured PCA

5.1 Introduction and related works

Thanks to their universality features spiked models, and their generalizations, find numerous applications in other central problems such as community detection [83, 84], group synchronization [85, 86], sub-matrix localization or high-dimensional clustering [87]; see [88, 89] for more. In this chapter we focus on a rank-one estimation problem, with a spike given by $\mathbf{X}^*\mathbf{X}^{*\intercal}$, $\mathbf{X}^* \in \mathbb{R}^N$, based on the data

$$\mathbf{Y} = \frac{\lambda}{N}\mathbf{X}^*\mathbf{X}^{*\intercal} + \mathbf{Z} \ \in \mathbb{R}^{N \times N} \tag{5.1}$$

with some additive noise $\mathbf{Z}$. Notice that, out of convenience, we have already re-scaled the observations $\mathbf{Y}$ so that they have $O(1)$ eigenvalues.

As described in Chapter 1, the spectral properties of finite rank perturbations of large random matrices like (5.1) were intensively investigated in random matrix theory. To the previously mentioned references we may add [90, 91, 92, 93, 94, 38, 95, 96].

Besides the spectral estimator, *i.e.* the eigenvector corresponding to the leading eigenvalue of the perturbed matrix, there exists a whole family of iterative algorithms, known as approximate message passing (AMP), that can be tailored to take further advantage of prior structural information known about the signal. AMP algorithms were first proposed for estimation in linear models [97, 98, 99, 100, 101, 102], but have since been applied to a range of statistical estimation problems, including generalized linear models [103, 104, 105, 106, 107, 108, 109] and low-rank matrix estimation [110, 111, 112, 88, 113, 114]. An attractive feature of AMP is that under suitable model assumptions, its performance in the high-dimensional limit is precisely characterized by a succinct deterministic recursion called state evolution [99, 78, 115]. Using the state evolution analysis, it has been proved that AMP achieves Bayes-optimal performance for some models [110, 116, 113, 103], and a conjecture from statistical physics posits that for a wide range of estimation problems, AMP is optimal among polynomial-time algorithms.

The references mentioned above rely on the assumption of Gaussian identically and independently distributed (i.i.d.) noise $Z_{ij} \sim \mathcal{N}(0,1)$, under which the model identified by (5.1) is the well-known spiked Wigner model [30, 117, 14]. This independence, or "absence of structure", in the noise has many advantages from the theoretical point of view due to the numerous simplifications it generates.

In order to relax this property, we can seek inspiration from the Statistical Physics literature on disordered systems. An idea that we studied in Chapter 2 and 3, that was imported also in high dimensional inference [118], is that of giving an inhomogeneous variance profile to the noise matrix elements (we mention that this idea in inference is similar to the earlier definition of "spatially coupled systems" [51, 52] in coding theory, see [47, 33] for its use in the present context). This procedure makes the (Z_{ij}) no longer identically distributed, but it leaves them independent. This is an important step towards more structure in the noise (and therefore the data). Yet, the independence assumption is a rather strong one. Actually, [118] showed that for a broad class of observation models, as long as the independence assumption holds, the model is information-theoretically equivalent to one with independent Gaussian (possibly inhomogeneous) noise.

One way to go beyond this last assumption is to consider noises that belong to the wider class of *rotationally invariant matrices*. Since the appearance of the seminal works [119, 120, 121], there has been a remarkable development in this direction, as evidenced by the rapidly growing number of papers on spin glasses [122, 123, 124, 125, 126, 127] and inference [38, 95, 128, 129, 130, 131, 132, 133, 134] that try to take into account structured disorder, including [135] on which this chapter is based. Indeed, we hereby consider a spiked model in which the noise matrix $\mathbf{Z}$ is drawn from an orthogonal matrix ensemble different from the Gaussian orthogonal ensemble (which is the only rotationally invariant ensemble such that the matrix entries are independent). Intuitively, the presence of dependencies in the noise should be exploitable by an algorithm that is sharp enough to see patterns within it and use them to retrieve the sought rank one matrix more efficiently. Going in that direction, in [132] the author proposed a version of AMP designed for rotationally invariant noises (using earlier ideas of [124, 123]) and provided also a rigorous state evolution analysis for it. Furthermore, in a recent work [70], the authors performed a rigorous analysis of a Bayes estimator and an AMP, both assuming Gaussian noise, whereas the actual noise in the data was drawn from a generic orthogonal matrix ensemble. However, besides intuition and the mentioned works, to our best knowledge there is little theoretical understanding of the true role played by noise structure in spiked matrix estimation and more generically in inference. In particular, prior to [135] there was no theoretical prediction of optimal performance to benchmark practical inference algorithms.

Notations.   Bold notations are reserved for vectors and matrices. By default a vector $\mathbf{x}$ is a column vector, and its transpose $\mathbf{x}^{\mathsf{T}}$ is therefore a row vector. Thus the usual L_2 norm $\|\mathbf{x}\|^2 = \mathbf{x}^{\mathsf{T}}\mathbf{x}$ and $\mathbf{x}\mathbf{x}^{\mathsf{T}}$ is a rank-one projector. The notation $\mathbf{x} \xrightarrow{W_2} X$ denotes convergence of the empirical distribution of the random vector $\mathbf{x}$ to the random variable X in Wasserstein-2

distance. Symbol $\propto$ means "equality up to a constant" (often, a normalization constant) and $:=$ is an equality by definition. Tr is the usual trace operator. For a vector $\mathbf{x}$, the matrix $\mathrm{diag}(\mathbf{x})$ is diagonal with $\mathbf{x}$ on its diagonal. For a diagonal matrix $\mathbf{A}$ and a function $F : \mathbb{R} \mapsto \mathbb{R}$ the matrix $F(\mathbf{A})$ is diagonal with F applied componentwise to each diagonal entry of $\mathbf{A}$. A function F applied to a real symmetric $N \times N$ matrix diagonalizable as $\mathbf{M} = \mathbf{U}\mathbf{A}\mathbf{U}^\intercal$ acts in the standard way: $F(\mathbf{M}) := \mathbf{U}F(\mathbf{A})\mathbf{U}^\intercal$. $\mathbb{E}_A$ is an expectation with respect to the random variable A; $\mathbb{E}$ is an expectation with respect to all random variables entering the ensuing expression. For a function F of one argument we denote F' its derivative. Notations like $i \leq N$ always implicitly assume that the index i starts at 1. Notation $[t] := \{1, 2, \cdots, t\} = \{i \leq t\}$. Powers for vectors apply componentwise (this is however *not* the case for matrices). We often compactly write $\mathbb{E}(\cdots)^2 = \mathbb{E}[(\cdots)^2] \geq (\mathbb{E}(\cdots))^2$ and similarly for other functions, we denote equivalently $\mathbb{E}[f(\cdots)]$ and $\mathbb{E}f(\cdots)$. Matrix I_N is the identity of size N.

5.1.1 Probabilistic model of PCA with structured noise

Consider a vector $\mathbf{X}^* = (X_i^*)_{i \leq N}$ whose components are drawn i.i.d. from a given distribution P_X with support bounded uniformly in N. Two cases will be considered: the factorized case

$$dP_X(\mathbf{X}^*) = \prod_{i \leq N} dP_X(X_i^*) = \prod_{i \leq N} P_X(X_i^*)dX_i^*,$$

and the case where dP_X is the uniform measure over the N-sphere of radius $\sqrt{N}$. If not specified the first case is assumed. We will always consider priors with unit second moment

$$\int dP_X(x)x^2 = 1.$$

This is just a convention as if one wants to consider a different normalization, it can simply be included through a proper rescaling of the SNR λ.

The inference task we are interested in is the retrieval of the rank-one spike $\mathbf{P}^* := \mathbf{X}^*\mathbf{X}^{*\intercal}$ from the following observed matrix

$$\mathbf{Y} = \frac{\lambda}{N}\mathbf{P}^* + \mathbf{Z}, \tag{5.2}$$

where $\mathbf{Z}$ is a unknown noise matrix, $\lambda \geq 0$ is the SNR. Whenever $\mathbf{Z}$ is a Wigner matrix this model corresponds to the usual Wigner spike model. But here we no longer assume that the noise is *unstructured* (namely, has independent entries). More specifically, we will assume that is drawn from a certain orthogonal rotationally invariant random matrix ensemble defined by a potential $V : \mathbb{R} \mapsto \mathbb{R}$ and a density (with normalization constant C_V)

$$dP_Z(\mathbf{Z}) = C_V d\mathbf{Z} \exp\left(-\frac{N}{2}\mathrm{Tr}V(\mathbf{Z})\right). \tag{5.3}$$

Rotational invariance means that $\mathbf{Z}$ equals in distribution $\mathbf{U}^\intercal \mathbf{Z} \mathbf{U}$ for any orthogonal matrix $\mathbf{U}$ (this follows from the trace in the exponent) [35]. More precisely, when changing variables from matrix $\mathbf{Z}$ to eigenvalues $\mathbf{D}$ and eigenbasis $\mathbf{O}$ via $\mathbf{Z} = \mathbf{O}^\intercal \mathbf{D} \mathbf{O}$ we have

$$dP_Z(\mathbf{D}, \mathbf{O}) = C_V d\mathbf{O}\, d\mathbf{D} \exp\left(-\frac{N}{2}\mathrm{Tr}V(\mathbf{D})\right) \prod_{i<j} |D_i - D_j|. \tag{5.4}$$

The measure $d\mathbf{O}$ is the Haar measure, i.e., uniform measure over the orthogonal group $\mathbf{O}(N)$, and the last term coupling all eigenvalues in a pairwise long-range fashion is the Vandermonde determinant. Note that only the special case $V(x) = x^2/(2\sigma)$ corresponding to the Gaussian orthogonal ensemble induces independent (Gaussian distributed) matrix entries (up to symmetry). Any other potential generates dependencies among matrix elements and thus *structure*. E.g., if we take $V(x) = x^4/4$,

$$dP_Z(\mathbf{Z}) = C_V d\mathbf{Z} \prod_{i,j,k,l} \exp\left(-\frac{N}{8}Z_{ij}Z_{jk}Z_{kl}Z_{li}\right) \tag{5.5}$$

which clearly is not factorizable over matrix entries.

We now introduce the Bayesian framework which we are going to analyse. Let the projector $\mathbf{P} := \mathbf{x}\mathbf{x}^\intercal$. This allows us to write the posterior measure of the inference problem:

$$dP_{X|Y}(\mathbf{x} \mid \mathbf{Y}) = \frac{C_V}{P_Y(\mathbf{Y})}dP_X(\mathbf{x})\exp\left(-\frac{N}{2}\mathrm{Tr}V\left(\mathbf{Y} - \frac{\lambda}{N}\mathbf{P}\right)\right). \tag{5.6}$$

Because both the prior $P_X(\mathbf{x})$ matches the density of the signal and the likelihood $P_{Y|X}$ matches the noise density P_Z and moreover the SNR λ is known, the posterior written above is the "correct" one and we are in the *Bayesian-optimal setting*. Studying the limits of inference in this setting draws a fundamental line between what is information-theoretically possible and what is not in terms of performance of inference. The evidence reads

$$P_Y(\mathbf{Y}) = C_V \int dP_X(\mathbf{x})\exp\left(-\frac{N}{2}\mathrm{Tr}V\left(\mathbf{Y} - \frac{\lambda}{N}\mathbf{P}\right)\right). \tag{5.7}$$

One of the main object of interest is the *free entropy* (or minus the *free energy*), which is nothing else than minus the Shannon entropy of the data, and in this chapter will be denoted as

$$F_N(\mathbf{Y}) := -H(\mathbf{Y}) = \mathbb{E}\ln P_Y(\mathbf{Y}). \tag{5.8}$$

As usual, the free entropy is related to the mutual information by an additive constant corresponding to the entropy of the noise:

$$\begin{aligned} I(\mathbf{P}^*; \mathbf{Y}) &= -F_N(\mathbf{Y}) - H(\mathbf{Y} \mid \mathbf{X}^*) \\ &= -F_N(\mathbf{Y}) - H(\mathbf{Z}) \\ &= -F_N(\mathbf{Y}) - \ln C_V + \frac{N}{2}\mathbb{E}\mathrm{Tr}V(\mathbf{Z}). \end{aligned} \tag{5.9}$$

Using the explicit form of the observation model (5.2) the free entropy reads

$$F_N(\mathbf{Y}) = \mathbb{E} \ln \int dP_X(\mathbf{x}) \exp\Big(-\frac{N}{2} \text{Tr}\Big[V\Big(\mathbf{Z} + \frac{\lambda}{N}(\mathbf{P}^* - \mathbf{P}) \Big) - V(\mathbf{Z}) \Big] \Big)$$
$$+ \ln C_V - \frac{N}{2} \mathbb{E} \text{Tr} V(\mathbf{Z}). \tag{5.10}$$

We extracted the noise entropy in the second line so that we can isolate the mutual information and to make the argument of the integrated exponential of order N. In this way the problem is naturally mapped onto a statistical mechanics model with extensive Hamiltonian given by minus the log-likelihood:

$$H_N(\mathbf{x}; \mathbf{Z}, \mathbf{X}^*) := \frac{N}{2} \text{Tr}\Big[V\Big(\mathbf{Z} + \frac{\lambda}{N}(\mathbf{P}^* - \mathbf{P}) \Big) - V(\mathbf{Z}) \Big]. \tag{5.11}$$

Indeed, our Hamiltonian can be rewritten as

$$\frac{1}{2} \text{Tr}(\mathbf{P}^* - \mathbf{P}) \int_0^\lambda dt V'\Big(\mathbf{Z} + \frac{t}{N}(\mathbf{P}^* - \mathbf{P}) \Big).$$

The difference between the two projectors has only two eigenvalues of order N and the matrix inside the potential derivative has $O(1)$ eigenvalues, hence the previous is of $O(N)$ too. The free entropy is thus directly linked to the expected log-partition function associated to this Hamiltonian:

$$\mathbb{E} \ln \mathcal{Z}(\mathbf{Y}) := \mathbb{E} \ln \int dP_X(\mathbf{x}) \exp\big(-H_N(\mathbf{x}; \mathbf{Z}, \mathbf{X}^*) \big). \tag{5.12}$$

The notation ; in $H_N(\mathbf{x}; \mathbf{Z}, \mathbf{X}^*)$ emphasizes that $\mathbf{Z}, \mathbf{X}^*$ are quenched variables while $\mathbf{x}$ fluctuates according the Gibbs-Boltzmann distribution associated to this Hamiltonian (i.e., the posterior). The same notation with same meaning for Hamiltonians will be used later on.

5.1.2 A concrete example: the quartic ensemble

Analysing this model for a generic potential V is *possible* through the novel methodology presented in this chapter. But as it will become apparent, if we take a generic polynomial potential V, the higher the order of this polynomial, the more technical and cumbersome it becomes. So for the sake of pedagogy we focus in the present contribution on a very concrete example of non trivial correction to the i.i.d. noise hypothesis. As a matter of fact, the simplest inference problem with correlated noise elements is that with the quartic matrix potential: for two positive real numbers μ and γ we restrict our analysis to the potential

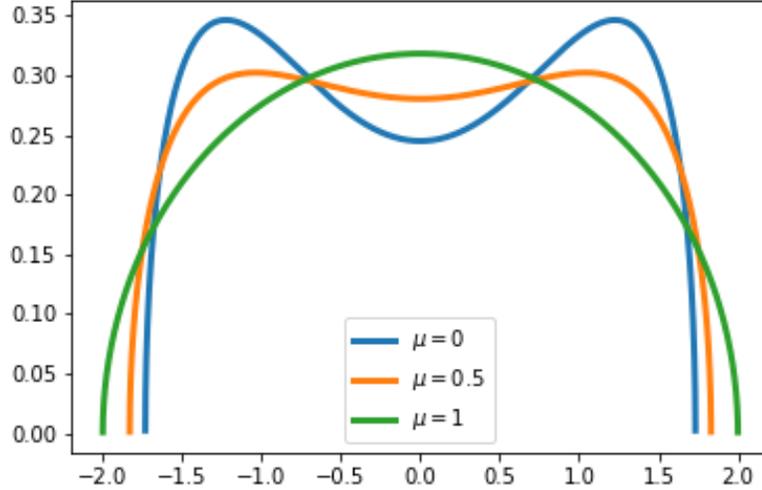$$V(x) = \frac{\mu}{2}x^2 + \frac{\gamma}{4}x^4. \tag{5.13}$$

Figure 5.1: Asymptotic spectral density (5.14) of the random noise ensemble defined by the potential (5.13) from less structured (with independent entries) at $(\mu = 1, \gamma = 0)$, corresponding to the standard semi-circle law, to the more structured $(\mu = 0, \gamma = 16/27)$ (recall relation (5.17)).

This was first introduced by Brézin et al in [136] to study the planar approximation of quantum field theories with large internal symmetry groups. We could have also considered a non-symmetric potential with a cubic term too, but for simplicity we restrict ourselves to that case as symmetry will slightly simplify the computations (but there is no barrier to applying our methods to that a more general, possibly non-even, potential).

The matrix ensemble defined by (5.13) has a known Stieltjes transform $\mathcal{S}$ and asymptotic eigenvalue density ρ, see, e.g., [35]: if $\mathbf{Z}$ is a sequence of matrices of increasing size N drawn from (5.3) with the above quartic potential and whose sequence of eigenvalues is $(D_i)_{i \leq N}$, then

$$\frac{1}{N} \sum_{i \leq N} \delta_{D_i, x} \xrightarrow{N \to \infty} \rho(x) = \frac{1}{2\pi}(\mu + 2a^2\gamma + \gamma x^2)\sqrt{4a^2 - x^2}, \tag{5.14}$$

$$\mathcal{S}(z) = \int \frac{d\rho(x)}{z - x} = \frac{1}{2}\big(\mu z + \gamma z^3 - (\mu + 2a^2\gamma + \gamma z^2)\sqrt{z^2 - 4a^2}\big), \tag{5.15}$$

for a z lying outside of the support of ρ, and where

$$a^2 := \frac{\sqrt{\mu^2 + 12\gamma} - \mu}{6\gamma}. \tag{5.16}$$

It is evident that when $\gamma \to 0^+$ one has $a^2 \to 1/\mu$ and consequently $\rho(x) \to \rho_{\mathrm{sc}}(x)$ the standard

semi-circle law, see Figure 5.1. In principle the choice of γ and μ is totally free, as long as[1] $\gamma > 0$. However, we are interested in a noise with unit variance in order to be able to make a meaningful comparison with models with unstructured noise. By enforcing this unitarity constraint one finds a relation between γ and μ:

$$\gamma = \gamma(\mu) = \frac{1}{27}\left(8 - 9\mu + \sqrt{64 - 144\mu + 108\mu^2 - 27\mu^3}\right). \tag{5.17}$$

With this choice one can check that

$$\int d\rho(x)x^2 = 1 \quad \text{for any} \quad \mu \in [0, 1].$$

When $(\mu = 1, \gamma(1) = 0)$ we recover the pure Wigner case already analyzed in great details. On the contrary $(\mu = 0, \gamma(0) = 16/27)$ corresponds to a purely quartic case with unit variance, and to the "most structured" ensemble in our restriced class of noise ensembles. Therefore, μ can be thought of as a parameter allowing to interpolate between unstructured and structured noise ensembles. Even for this simple family of potentials, as soon as $\mu < 1$, neither the Bayes-optimal nor the algorithmic limits of inference are known (except for those of a simple spectral algorithm, see [38]).

5.1.3 Main results

Our main contributions can be divided in two categories: those on the fundamental, information-theoretic, limitations of inference in structured PCA and, complementary to that, novel algorithmic ideas allowing to match these Bayes-optimal limits efficiently. Both require conceptual insights and technical advances that we emphasize. We gather here these results and state them informally; we refer to the main sections for precise statements.

Information-theoretic results

- Our analysis of the information-theoretic (Bayes-optimal) performance based on the non-rigorous replica method yields first a low-dimensional variational formulation for the free entropy (log-partition function) of the model when P_X is factorized:

Result 1 (Free entropy). The free entropy (i.e., minus Shannon entropy of the data) verifies in the limit of large size the following characterization:

$$\frac{1}{N}F_N(\mathbf{Y}) = -\frac{1}{N}H(\mathbf{Y}) \xrightarrow{N \to \infty} \operatorname{extr} f_\rho(\boldsymbol{\tau})$$

---

[1]We use implicitly the convexity of the potential, which requires $\mu, \gamma > 0$, to obtain the density of eigenvalues [35]. But we believe that this condition can be relaxed if one can get an associated well-defined asymptotic spectral density and that our analysis would still hold.

where $\boldsymbol{\tau} \in \mathbb{R}^{13}$ and for an explicit real-valued function $f_\rho : \mathbb{R}^{13} \mapsto \mathbb{R}$ depending on the noise asympotic spectral density ρ. See (5.60) for the complete statement. Here and everywhere in the chapter extr stands for the following "extremization" procedure: if $f : \mathbb{R}^k \mapsto \mathbb{R}$ then

$$\operatorname{extr} f(\boldsymbol{\tau}) := f(\boldsymbol{\tau}_*) \quad \text{where} \quad \boldsymbol{\tau}_* := \underset{\{\boldsymbol{\tau} \in \mathbb{R}^k \, : \, \nabla f(\boldsymbol{\tau}) = \mathbf{0}\}}{\operatorname{argmax}} f(\boldsymbol{\tau}).$$

We will see that, despite the apparent complication, the 13-dimensional system of equations defining $\boldsymbol{\tau}_*$ will reduce to a much simpler 3-dimensional one thanks to the Nishimori identities. The reduction of the replica saddle point equations is done in Section 5.3.3. As a consequence only three scalar quantities will remain after reduction, one denoted m and called "magnetization" quantifying the overlap between the minimum mean-square error (MMSE) estimator and the signal.

- From the solution of this variational problem we deduce our second main result, namely, an asympotically exact expression for the minimum mean-square error of inference of the hidden spike with factorized prior:

Result 2 (Minimum mean-square error)**.** The minimum mean-square error verifies

$$\lim_{N \to \infty} \frac{1}{2N^2} \mathbb{E} \| \mathbf{X}^* \mathbf{X}^{*\intercal} - \mathbb{E}[\mathbf{X}^* \mathbf{X}^{*\intercal} \mid \mathbf{Y}] \|_{\mathrm{F}}^2 = \frac{1}{2}(1 - m^2)$$

where m is one component of the solution $\boldsymbol{\tau}_*$ to the variational problem for the free entropy, studied in Section 5.3.3.

The main technical and conceptual novelties which lead to these formulas are:

- To the best of our knowledge, we provide the first adaptation of the replica method to the analysis of the fundamental limits of inference in a model with a *noise* having strongly dependent random entries (instead of a measurement operator, or matrix of covariates, in a regression setting). See Section 5.3.

- If the structure of the noise (i.e., its statistical properties) is encoded by a polynomial potential V of order $K+1$, then this induces in the posterior distribution k-wise interactions between the signal's estimator entries, for all $k \leq K + 1$. Said differently, the underlying factor graph is an hypergraph with hyperedges of degrees $K + 1, K, \ldots, 1$. However, we discovered that by exploiting the low-rank structure of the signal, all these interactions can be reduced to effective pair-wise interations. This allows to reduce the model to an Ising model more convenient for theoretical analysis (a similar reduction is useful for algorithmic approaches too, see next section). The reduction we propose is general and systematic for low-rank signals corrupted by rotational invariant noise matrices. See Section 5.3.1.

- Our analysis can be mainstreamed once we have identified a key integral that we refer to as the *inhomogeneous spherical integral*. This exactly solvable integral is a generalization of the standard low-rank spherical integral appearing in random matrix theory (as it is related to the R-transform) [35], in spin-glasses [121, 124, 123, 125, 137], the theory of large-deviations for matrix-valued stochastic processes [138, 139] and matrix models in high-energy physics [140, 141, 142]. Given the breadth of applications of this integral, we foresee that the generalization we propose and analyze in Section 5.2 may have applications well beyond the present setting, for the study of models where rotationally invariant matrices with non-independent matrices appear.

- Another important conclusion from our analysis is the fact that for signals $\mathbf{X}^*$ whose law is rotation-invariant (such as Gaussian or uniformly spherically distributed), the simple spectral PCA procedure of [38] is Bayes-optimal:

Result 3 (Optimality of spectral PCA for rotation-invariant priors). Let $\mathbf{X}^*$ be a standard Gaussian vector or uniformly sampled on the sphere of radius $\sqrt{N}$. Then its inference from $\mathbf{Y}$ can be optimally achieved from the naive spectral algorithm that constructs an estimator $C\boldsymbol{\nu}\boldsymbol{\nu}^\intercal$ of $\mathbf{P}^*$ from the eigenvector $\boldsymbol{\nu} = \boldsymbol{\nu}(\mathbf{Y})$ of $\mathbf{Y}$ with leading eigenvalue $\lambda_{\max}$ and that is then properly rescaled by a certain factor $C = C(\lambda, \rho)$, see [38].

This is verified both by the replica method and an exact computation based on Gaussian integration and a saddle point method, see Section 5.3.4. We remark that this statement is incorrect for other priors P_X.

Algorithmic results

On the algorithmic side our contributions are the following:

- We analytically show that the existing Approximate Message Passing algorithms [132, 133], whose iterates are based on the data matrix $\mathbf{Y}$, do not saturate the Bayes-optimal performance predicted by our replica theory. See Section 5.4.

- We employ in Section (5.5) the AdaTAP formalism of Opper et al [124] to analyze the model from the algorithmic perspective. What the analysis shows is that, like in the replica method, one can reduce the model with interactions of order higher than two to a pure quadratic Ising model with an effective interaction matrix $\mathbf{J}(\mathbf{Y})$ which is a non-trivial matrix polynomial of the data $\mathbf{Y}$. This explains the reason why the previously proposed AMP algorithms are sub-optimal: the data $\mathbf{Y}$ is *not* the best choice of matrix to use in the AMP iterates, despite being the most natural one. The Bayes-optimal choice is instead $\mathbf{J}(\mathbf{Y})$ obtained from our theory, which cannot be guessed a-priori. We informally state this fact as one of our main results:

Result 4 (Bayesian-optimal processing of data and optimal AMP)**.** Consider the matrix estimation model under structured noise (5.2). Given the observed matrix of data $\mathbf{Y}$, the optimal choice of matrix to use in a Bayesian inference algorithm such as AMP is *not* $\mathbf{Y}$ but instead a proper polynomial of it, i.e., $\mathbf{J}(\mathbf{Y}) = \sum_{k \leq K} c_k \mathbf{Y}^k$, with coefficients $(c_k)_{k \in [K]}$ depending on V. For example, when the potential V is given by (5.13) we show in Sections 5.5.1 and 5.5.2 that the optimal choice is

$$\mathbf{J}(\mathbf{Y}) = \mu\lambda\mathbf{Y} - \gamma\lambda^2\mathbf{Y}^2 + \gamma\lambda\mathbf{Y}^3.$$

Employing this matrix in the AMP iterates leads to a Bayesian-optimal inference algorithm whose complexity scales as the dimension N, see the result below.

- After having defined the Bayesian-optimal AMP recursion, we provide a rigorous state evolution recursion to track its asymptotic performance. We highlight that, since the data matrix $\mathbf{Y}$ is replaced by the polynomial $\mathbf{J}(\mathbf{Y})$, we cannot apply the state evolution result of [132]. More specifically, the Onsager correction terms will have a different form than the ones of [132], and their derivation requires a novel analysis.

Result 5 (State evolution of the Bayes-optimal AMP (BAMP))**.** Consider the Bayesian-optimal Approximate Message Passing (BAMP) algorithm defined by the recursion

$$\mathbf{f}^t = \mathbf{J}(\mathbf{Y})\mathbf{u}^t - \sum_{i \leq t} \mathsf{c}_{t,i}\mathbf{u}^i, \quad \mathbf{u}^{t+1} = g_{t+1}(\mathbf{f}^t), \quad t \geq 1. \tag{5.18}$$

When a proper choice of coefficients $\{\mathsf{c}_{t,j}\}_{j \in [t]}$ is considered, for a large family of functions $(g_t)_{t \geq 1}$ and ψ, the following holds almost surely:

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i \leq N} \psi(u_i^1, \ldots, u_i^{t+1}, f_i^1, \ldots, f_i^t, X_i^*) = \mathbb{E}\,\psi(U_1, \ldots, U_{t+1}, F_1, \ldots, F_t, X^*).$$

Equivalently the joint empirical distribution over the N rows of the $N \times (2t+2)$ matrix $(\mathbf{u}^1, \ldots, \mathbf{u}^{t+1}, \mathbf{f}^1, \ldots, \mathbf{f}^t, \mathbf{X}^*)$ converges in a certain sense to the $(2t+2)$-dimensional random vector $(U_1, \ldots, U_{t+1}, F_1, \ldots, F_t, X^*)$ when N increases. Here

$$U_{i+1} = g_{i+1}(F_t) \quad \text{and} \quad (F_1, \ldots, F_t) = (\mu_1, \ldots, \mu_t)X^* + (W_1, \ldots, W_t)$$

with $(W_i)_{i \leq t}$ a multivariate Gaussian vector whose covariance as well as $(\mu_i)_{i \leq t}$ can be computed via a deterministic state evolution recursion.

The precise rigorous statement can be found in Section 5.6. The idea of the argument is to construct an auxiliary AMP which tracks the quantities $(\mathbf{Y}^{j-1}\mathbf{u}^t)_{t \geq 1, j \leq K-1}$. By decomposing the iterates of this auxiliary AMP into a component aligned with previous iterates, a component in the direction of the signal and independent Gaussian noise, we obtain the form of the Onsager

correction and the state evolution. From this result we can rigorously predict the performance of the novel AMP algorithm we propose. The optimality of the pre-processed matrix $\mathbf{J}(\mathbf{Y})$ and associated AMP is then confirmed by the perfect matching of the fixed point of the state evolution recursion tracking the AMP mean-square error and our replica prediction for the MMSE.

Two important remarks are in order. First, we emphasize that the BAMP algorithm (5.18) we propose is *not* the usual AMP of [132] where the data matrix $\mathbf{Y}$ is just replaced by the pre-processed matrix $\mathbf{J}(\mathbf{Y})$. Indeed, the correct Onsager coefficients $\{c_{t,i}\}$ entering BAMP require a novel type of "multi-stage" state evolution recursion which is completely different from the one in [132], see Section 5.6. The novel acronym we introduce emphasizes that crucial distinction.

Secondly, it is true that our replica prediction for the MMSE is non-rigorous. However, our state evolution analysis of BAMP is fully rigorous (just like the analysis of the AMP in [132]). By comparing their asymptotic fixed point performance by state evolution in Section 5.7, we show that BAMP improves over the AMP in [132]. This improvement is thus a rigorous conclusion, while the conjecture is that, thanks to this improvement, BAMP saturates the Bayes-optimal performance.

Comments on the potential universality of our results

We comment the hypotheses under which our results are conjectured valid, and then extrapolate on the more general settings in which the results may still hold.

We start with a remark concerning the insensitivity of our results to the "statistical details" of the noise eigenvalues. Let us precise the hypotheses on the distribution of the noise, in particular on its eigenvalues, under which our results are conjectured valid. As seen from (5.4) the eigenvalues of the noise are strongly dependent due to the Vandermonde determinant. However, we conjecture that all our results still hold if one considers instead a simpler ensemble where the N eigenvalues are drawn i.i.d. from $\rho(x)dx$, see (5.14). The reason is that all the analysis and results rely only on the weak convergence of the empirical density of eigenvalues of the ensemble under consideration towards ρ. Hence, as long as this is the case, our results must hold, even if we do not rigorously prove it. To formally show it, from now on we consider that the diagonal matrix $\mathbf{D}$ of eigenvalues of the noise is *deterministic* with the sole constraint that the empirical density of its diagonal entries converges towards $\rho(x)dx$. This of course includes as special cases the two aforementioned settings (i.i.d. and coupled by Vandermonde determinant). We therefore work in this chapter under the following hypothesis.

Hypothesis 5.1 (Distribution of the noise). The noise $\mathbb{R}^{N \times N} \ni \mathbf{Z} = \mathbf{O}^\intercal \mathbf{D} \mathbf{O}$ in model (5.2) is a symmetric rotationally invariant matrix, namely, it is equal in law to $\mathbf{U}^\intercal \mathbf{Z} \mathbf{U}$ for any orthogonal matrix $\mathbf{U} \in \mathbb{O}(N)$ (the group of $N \times N$ orthogonal matrices). Equivalently, $\mathbf{O}$ is drawn from the Haar (uniform) measure over $\mathbb{O}(N)$. Moreover, we only require for its (possibly deterministic) eigenvalues $(D_i)_{i \leq N}$ that their empirical law $N^{-1} \sum_{i \leq N} \delta_{D_i, x}$ is tending weakly as $N \to \infty$ to a probability measure with support bounded uniformly in N and with density ρ with respect to the Lebesgue measure. As mentioned earlier, for the purpose of having a uniform measure

of SNR when tuning $(\mu, \gamma(\mu))$ we will consider cases where $\int d\rho(x)x^2 = 1$ despite this is not necessary for the analysis to hold.

A second remark concerns the rotational invariance of the noise. We believe that our results may extend beyond this hypothesis to cases where the noise eigenbasis may be invariant under more restrictive transformations (such as permutation invariant), or even "almost deterministic". This intuition comes from a very recent line of work concerning linear regression and phase retrieval with structured matrices of covariates. Indeed, the authors of [143, 144, 145] show that in this context, the class of rotationally invariant matrices leads to the same performance as a much broader class of almost deterministic matrices (with the same spectral density), also when AMP or its linearized version are used as inference algorithm. This is a different setting from the one we consider, since in our setup the structured matrix is the noise, but it nevertheless suggests that our predictions should remain true more generically. The confirmation of this universality is left for future work.

What is conjectured exact, and what is rigorous

We end this section with a remark concerning the level of rigor of our derivations. Most of our results are based on non-rigorous but well established methods from the statistical mechanics of mean-field disordered systems, in particular the replica method at the replica symmetric level, and the theory of Anderson-Thouless-Palmer equations. For a general background on these techniques we refer to [6, 146, 147, 148]. It is important to keep in mind that despite being non-rigorous, the results obtained from these techniques are conjectured to be *exact* in the present setting of *Bayesian-optimal inference* (or equivalently, statistical mechanical models living on their *Nishimori line* [29]), in the asymptotic large size limit $N \to \infty$.

This widely admitted asymptotic exactness, first proved for the Sherrington-Kirkpatrick model, spreads in numerous fields and in particular in the analysis of high-dimensional inference. In this context we have already mentioned a plethora of rigorous results confirm the validity of replica predictions, and we may add even more [117, 149]. In particular, replica symmetric formulas for the free entropy, mutual information and minimum mean-square error have been systematically proved thanks to a combination of concentration techniques specifically adapted to the context of inference together with rigorous versions of the cavity method [20, 18, 150], (adaptive) interpolation techniques or Hamilton-Jacobi approaches [151, 66, 152]. From this rapidly growing literature, we conjecture that it is only a matter of time before our replica-based predictions are proven.

Concerning our algorithmic results on the novel approximate message passing we propose (BAMP), the results are completely rigorous; full proofs are provided as appendix. They are based on the theory of message passing algorithms and associated state evolution recursions [99], in particular the most recent results for structured matrices as considered here [132, 133].

5.2 The inhomogeneous spherical integral

In this section we derive the expression of a useful general integral that will play a crucial role along the whole analysis, and that we believe may have an interest on its own. For the reader interested in the information-theoretic and algorithmic analyses directly, this section can be skipped at first reading as only its main results (5.21), (5.22) and (5.32) will be used in the rest.

Indices $\ell, \ell' \leq n$ will always indicate the "replica dimension" (with n which always remains finite), while $i, j, k \leq N$ index the "spin dimension" (where N will diverge).

5.2.1 Definition and variational characterization

Let $\mathbf{O} \sim \mathrm{Haar}(\mathbb{O}(N))$ be drawn from the Haar measure over the orthogonal group of $N \times N$ matrices. Consider a fixed matrix $\mathbf{x} \in \mathbb{R}^{N \times n}$ with rows $\mathbf{x}_i \in \mathbb{R}^n$, $i \leq N$, and columns $\mathbf{x}_\ell \in \mathbb{R}^N$, $\ell \leq n$. Assume it has the column-wise overlap structure

$$\mathbf{x}_\ell^\mathsf{T} \mathbf{x}_{\ell'} = N q_{\ell\ell'}, \qquad \ell, \ell' \leq n. \tag{5.19}$$

We let $\mathbf{q} = (q_{\ell\ell'})_{\ell,\ell' \leq n} := N^{-1}\mathbf{x}^\mathsf{T}\mathbf{x}$. Every vector is considered a column vector, so, e.g., $(\mathbf{Ox})_i$ is a n-dimensional column-vector corresponding to the transpose of the ith row of the $N \times n$ matrix $\mathbf{Ox}$, while $(\mathbf{Ox})_i^\mathsf{T}$ is a row-vector.

Let the matrices $\mathbf{C}_{\ell\ell'} = \mathrm{diag}((C_{i,\ell\ell'})_{i \leq N})$, $\mathbf{C}_i = (C_{i,\ell\ell'})_{\ell,\ell' \leq n}$, and the "external fields" $\mathbf{h}_\ell = (h_{i,\ell})_{i \leq N}$, $\mathbf{h}_i = (h_{i,\ell})_{\ell \leq n}$ all having entries bounded uniformly in N. The sequence $(\mathbf{h}_i \in \mathbb{R}^n, \mathbf{C}_i \in \mathbb{R}^{n\times n})_{i \leq N}$ is assumed to have an empirical law tending to that of the random $(\mathbf{h} \in \mathbb{R}^n, \mathbf{C} \in \mathbb{R}^{n\times n})$: for any continuous bounded function $f : \mathbb{R}^{n\times n} \times \mathbb{R}^n \mapsto \mathbb{R}^k$ with k independent of N,

$$\frac{1}{N} \sum_{i \leq N} f(\mathbf{C}_i, \mathbf{h}_i) \xrightarrow{N \to \infty} \mathbb{E} f(\mathbf{C}, \mathbf{h}).$$

We denote by $\mathbb{R} \ni I_N = I_N(\mathbf{q}, (\mathbf{C}_{\ell\ell'})_{\ell,\ell' \leq n}, (\mathbf{h}_\ell)_{\ell \leq n}) = I_N(\mathbf{q}, (\mathbf{C}_i, \mathbf{h}_i)_{i \leq N})$ the generalized low-rank spherical integral, which is defined as

$$I_N := \frac{1}{N} \ln \mathbb{E}_\mathbf{O} \exp \sum_{i \leq N} \left((\mathbf{Ox})_i^\mathsf{T} \mathbf{C}_i (\mathbf{Ox})_i + (\mathbf{Ox})_i^\mathsf{T} \mathbf{h}_i \right)$$

$$= \frac{1}{N} \ln \mathbb{E}_\mathbf{O} \exp \left(\sum_{\ell,\ell' \leq n} (\mathbf{Ox}_\ell)^\mathsf{T} \mathbf{C}_{\ell\ell'} \mathbf{Ox}_{\ell'} + \sum_{\ell \leq n} (\mathbf{Ox}_\ell)^\mathsf{T} \mathbf{h}_\ell \right)$$

$$= \frac{1}{N} \ln \mathbb{E}_\mathbf{O} \exp \left(\sum_{i,j,k \leq N} \sum_{\ell,\ell' \leq n} O_{ij} O_{ik} x_{j,\ell'} x_{k,\ell} C_{i,\ell\ell'} + \sum_{i,j \leq N} \sum_{\ell \leq n} O_{ij} x_{j,\ell} h_{i,\ell} \right). \tag{5.20}$$

Calling the columns $(\mathbf{x}_\ell)_{\ell \leq n}$ "replicas", the matrices $(\mathbf{C}_i)_{i \leq N}$, $(\mathbf{C}_{\ell\ell'})_{\ell,\ell' \leq n}$ are coupling them (after the replicas have been jointly rotated by the random $\mathbf{O}$). Therefore we call them "replica coupling matrices".

As $N \to \infty$ with n fixed this integral is given by

$$I_N \xrightarrow{N \to \infty} I_{\mathbf{C},\mathbf{h}}(\mathbf{q}), \tag{5.21}$$

with variational formula

$$I_{\mathbf{C},\mathbf{h}}(\mathbf{q}) := \frac{1}{2}\text{extr}_{\tilde{\mathbf{q}}}\big(\text{Tr}\mathbf{q}\tilde{\mathbf{q}} + \mathbb{E}\mathbf{h}^\mathsf{T}(\tilde{\mathbf{q}} - 2\mathbf{C})^{-1}\mathbf{h} - \mathbb{E}\ln\det(\tilde{\mathbf{q}} - 2\mathbf{C})\big)$$

$$- \frac{1}{2}(n + \ln\det\mathbf{q}). \tag{5.22}$$

The extremum is over symmetric matrices such $\tilde{\mathbf{q}} - 2\mathbf{C}$ is positive definite for all $\mathbf{C}$ living on its domain.

We remark that it may be the case that the extremum over $\tilde{\mathbf{q}}$ is actually attained on the boundary of the optimization domain, in which case the optimization requires more care than what is done in (5.34) to solve it (as (5.34) assumes the extremum to lie inside the optimization domain). This is however not expected in the settings of this chapter. When this phenomenon happens, in the standard low-rank spherical integral this leads to a "sticking phenomenon" where the solution of the optimization is dependent on the maximum eigenvalue of the full-rank random matrix entering the integral's definition, see [138].

5.2.2   Special cases

Low-rank HCIZ integral

The special case $\mathbf{h}_i = \mathbf{0}$ and replica coupling matrices $\mathbf{C}_i = \mathbf{C}D_i$ for $i \leq N$ corresponds to the standard rank-n spherical (or HCIZ) integral:

$$I_N = \frac{1}{N}\ln\mathbb{E}_{\mathbf{O}}\exp\sum_{i,j,k\leq N}\sum_{\ell,\ell'\leq n}O_{ij}O_{ik}x_{j,\ell'}x_{k,\ell}C_{\ell\ell'}D_i$$

$$= \frac{1}{N}\ln\mathbb{E}_{\mathbf{O}}\exp\sum_{\ell,\ell'\leq n}C_{\ell\ell'}(\mathbf{O}\mathbf{x}_\ell)^\mathsf{T}\mathbf{D}\mathbf{O}\mathbf{x}_{\ell'}$$

$$= \frac{1}{N}\ln\mathbb{E}_{\mathbf{O}}\exp\text{Tr}\,\mathbf{O}^\mathsf{T}\mathbf{D}\mathbf{O}(\mathbf{x}\mathbf{C}\mathbf{x}^\mathsf{T}),$$

where $\mathbf{D} = \text{diag}((D_i)_{i\leq N})$ and $\mathbf{x}\mathbf{C}\mathbf{x}^\mathsf{T}$ is an arbitrary rank-n symmetric matrix (arbitrary given that $\mathbf{x}$ and $\mathbf{C}$ are so). Its asymptotic expression can also be obtained from the results of [138] after diagonalizing $\mathbf{x}\mathbf{C}\mathbf{x}^\mathsf{T}$ and depends only on the limit of the empirical distribution of (D_i) and on the n non-zero eigenvalues of $\mathbf{x}\mathbf{C}\mathbf{x}^\mathsf{T}$.

Low-rank spherical integral with external field and diagonal replica coupling

Taking diagonal replica coupling matrices $\mathbf{C}_i = I_n D_i / 2$ gives (a generalization of) the spherical integral with external field found in [Prop. 2.7, [125]]:

$$I_N = \frac{1}{N} \ln \mathbb{E}_\mathbf{O} \exp \sum_{\ell \leq n} \left(\frac{1}{2} (\mathbf{O}\mathbf{x}_\ell)^\mathsf{T} \mathbf{D} \mathbf{O}\mathbf{x}_\ell + (\mathbf{O}\mathbf{x}_\ell)^\mathsf{T} \mathbf{h}_\ell \right). \tag{5.23}$$

Low-rank spherical integral with non-diagonal replica coupling and replica symmetric overlap

Let the $N \times N$ diagonal matrices $\mathbf{A} = \mathrm{diag}((A_i)_{i \leq N})$ and similarly for $\mathbf{B}$. The empirical law of $(A_i, B_i)_{i \leq N}$ tends to that of (A, B). Of particular interest to us corresponds to taking $\ell \in \{0, \ldots, n\}$, $\mathbf{h}_i = \mathbf{0}$ and replica coupling matrices with only non-zero entries being

$$(\mathbf{C}_i)_{\ell 0} = (\mathbf{C}_i)_{0\ell} = \frac{A_i}{2} \quad \text{for} \quad 1 \leq \ell \leq n, \qquad (\mathbf{C}_i)_{\ell\ell} = \frac{B_i}{2}(1 - \delta_{\ell,0}), \tag{5.24}$$

or equivalently,

$$\mathbf{C}_{0\ell} = \mathbf{C}_{\ell 0} = \frac{\mathbf{A}}{2} \quad \text{and} \quad \mathbf{C}_{\ell\ell} = \frac{\mathbf{B}}{2} \quad \text{for} \quad 1 \leq \ell \leq n, \qquad \mathbf{C}_{\ell\ell'} = \mathbf{0} \quad \text{else.} \tag{5.25}$$

Note that this is *not* a special case of the standard rank-n spherical integral of the first example: here $\mathbf{C}_i$ cannot be written as $\mathbf{C}$ times a function of i; instead different entries of $\mathbf{C}_i$ vary with i differently. In this case the generalized spherical integral reads (the sum over ℓ below starts at $\ell = 1$)

$$I_N = \frac{1}{N} \ln \mathbb{E}_\mathbf{O} \exp \sum_{\ell \leq n} \left((\mathbf{O}\mathbf{x}_0)^\mathsf{T} \mathbf{A} \mathbf{O}\mathbf{x}_\ell + \frac{1}{2} (\mathbf{O}\mathbf{x}_\ell)^\mathsf{T} \mathbf{B} \mathbf{O}\mathbf{x}_\ell \right). \tag{5.26}$$

So the 0th replica plays here a special role (it corresponds to the planted signal).

We consider a "replica symmetric structure" for the overlap matrix parametrized by the vector $(v_0, v, m, q) \in \mathbb{R}^4$:

$$\mathbf{q} = \begin{pmatrix} v_0 & m & m & m & \ldots & m \\ m & v & q & q & \ldots & q \\ m & q & v & q & \ldots & q \\ m & q & q & v & \ldots & q \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ m & q & q & q & \ldots & v \end{pmatrix} \in \mathbb{R}^{(n+1) \times (n+1)}, \tag{5.27}$$

and, coherently, we assume that the extremum over $\tilde{\mathbf{q}}$ is attained for a matrix having the same structure with different constants $(\tilde{v}_0, \tilde{v}, \tilde{m}, \tilde{q})$. Its determinant can be easily computed via Gauss' reduction:

$$\ln \det \mathbf{q} = \ln v_0 + n \ln(v - q) + \ln\left(1 + n\frac{v_0 q - m^2}{v_0(v - q)}\right). \tag{5.28}$$

We also need to compute

$$\mathrm{Tr} \mathbf{q}\tilde{\mathbf{q}} = v_0 \tilde{v}_0 + n(2m\tilde{m} + v\tilde{v} + (n - 1)q\tilde{q}). \tag{5.29}$$

Letting $\mathbf{C}$ be defined as (5.24) but with the random variables A, B replacing A_i, B_i, the last missing term is obtained similarly as (5.28): under the replica symmetric stucture for $\tilde{\mathbf{q}}$,

$$\mathbb{E} \ln \det(\tilde{\mathbf{q}} - 2\mathbf{C}) = \ln \tilde{v}_0 + n \ln(\tilde{v} - B - \tilde{q}) + \mathbb{E} \ln\left(1 + n\frac{\tilde{v}_0 \tilde{q} - (\tilde{m} - A)^2}{\tilde{v}_0(\tilde{v} - B - \tilde{q})}\right).$$

Combining everything in the variational formula (5.22), and taking into account that $\mathbf{q}$ here is a $(n + 1) \times (n + 1)$ matrix, we obtain the following expression for the generalized spherical integral with replica coupling (5.24), and under a replica symmetric structure for the overlap and conjugate matrices (thus the upperscript):

$$\begin{aligned}
I_N \to I_{A,B}^{\mathrm{RS}}(\mathbf{q}) :=\ & \frac{1}{2}\mathrm{extr}_{(\tilde{v}_0, \tilde{v}, \tilde{m}, \tilde{q})}\Big\{ v_0 \tilde{v}_0 - \ln \tilde{v}_0 + n(2m\tilde{m} + v\tilde{v} + (n - 1)q\tilde{q}) \\
& - n\mathbb{E} \ln(\tilde{v} - B - \tilde{q}) - \mathbb{E} \ln\left(1 + n\frac{\tilde{v}_0 \tilde{q} - (\tilde{m} - A)^2}{\tilde{v}_0(\tilde{v} - B - \tilde{q})}\right)\Big\} \\
& - \frac{1 + \ln v_0}{2} - \frac{n}{2}\big(1 + \ln(v - q)\big) - \frac{1}{2}\ln\left(1 + n\frac{v_0 q - m^2}{v_0(v - q)}\right). \tag{5.30}
\end{aligned}$$

By definition (5.26) of I_N this formula has to cancel when $n = 0$. Thus

$$\mathrm{extr}_{\tilde{v}_0}\big\{v_0 \tilde{v}_0 - \ln \tilde{v}_0\big\} - 1 - \ln v_0 = 0. \tag{5.31}$$

The saddle point equation over $\tilde{v}_0$ then yields $\tilde{v}_0 = 1/v_0$, in which case this latter formula indeed cancels. So the simplified formula reads

$$\begin{aligned}
I_{A,B}^{\mathrm{RS}}(\mathbf{q}) =\ & \frac{1}{2}\mathrm{extr}_{(\tilde{v}, \tilde{m}, \tilde{q})}\Big\{ n(2m\tilde{m} + v\tilde{v} + (n - 1)q\tilde{q}) - n\mathbb{E} \ln(\tilde{v} - B - \tilde{q}) \\
& - \mathbb{E} \ln\left(1 + n\frac{\tilde{q} - v_0(\tilde{m} - A)^2}{\tilde{v} - B - \tilde{q}}\right)\Big\} - \frac{n}{2}\big(1 + \ln(v - q)\big) \\
& - \frac{1}{2}\ln\left(1 + n\frac{v_0 q - m^2}{v_0(v - q)}\right). \tag{5.32}
\end{aligned}$$

5.2.3 Derivation of the variational formula

Let $\tilde{\mathbf{x}}_\ell := \mathbf{O}\mathbf{x}_\ell$ the columns of $\tilde{\mathbf{x}} = \mathbf{O}\mathbf{x}$. Under the law of $\mathbf{O}$ at fixed $\mathbf{x}$, these random vectors are uniform among all vectors having the overlap structure of $(\mathbf{x}_\ell)$. Thus their law conditional on $\mathbf{x}$ is just a function of the symmetric overlap $\mathbf{q} = (q_{\ell\ell'})$:

$$P(\tilde{\mathbf{x}} \mid \mathbf{x}) = P(\tilde{\mathbf{x}} \mid \mathbf{q}) = \frac{1}{\mathcal{Z}(\mathbf{q})} \prod_{\ell \geq \ell'}^{1,n} \delta(Nq_{\ell\ell'} - \tilde{\mathbf{x}}_\ell^\mathsf{T}\tilde{\mathbf{x}}_{\ell'}) = \frac{1}{\mathcal{Z}(\mathbf{q})} \delta(N\mathbf{q} - \tilde{\mathbf{x}}^\mathsf{T}\tilde{\mathbf{x}})$$

with normalization

$$\mathcal{Z}(\mathbf{q}) = \int d\tilde{\mathbf{x}}\, \delta(N\mathbf{q} - \tilde{\mathbf{x}}^\mathsf{T}\tilde{\mathbf{x}}).$$

Using the Fourier representation of the Delta function, the integral to compute reads (below $\tilde{\mathbf{q}}$ is a $n \times n$ symmetric matrix with complex entries)

$$\exp(NI_N) = \frac{1}{\mathcal{Z}(\mathbf{q})} \int d\tilde{\mathbf{x}}\, \delta(N\mathbf{q} - \tilde{\mathbf{x}}^\mathsf{T}\tilde{\mathbf{x}}) \exp \sum_{i \leq N} \left(\tilde{\mathbf{x}}_i^\mathsf{T} \mathbf{C}_i \tilde{\mathbf{x}}_i + \tilde{\mathbf{x}}_i^\mathsf{T} \mathbf{h}_i \right)$$

$$= \frac{1}{\mathcal{Z}(\mathbf{q})} \int d\tilde{\mathbf{x}} d\tilde{\mathbf{q}} \exp \left(\frac{N}{2}\mathrm{Tr}\mathbf{q}\tilde{\mathbf{q}} - \frac{1}{2}\mathrm{Tr}\tilde{\mathbf{x}}^\mathsf{T}\tilde{\mathbf{x}}\tilde{\mathbf{q}} + \sum_{i \leq N} \left(\tilde{\mathbf{x}}_i^\mathsf{T} \mathbf{C}_i \tilde{\mathbf{x}}_i + \tilde{\mathbf{x}}_i^\mathsf{T} \mathbf{h}_i \right) \right)$$

$$= \frac{1}{\mathcal{Z}(\mathbf{q})} \int d\tilde{\mathbf{q}} \exp \left(\frac{N}{2}\mathrm{Tr}\mathbf{q}\tilde{\mathbf{q}} \right) \prod_{i \leq N} \int d\tilde{\mathbf{x}}_i \exp \left(-\frac{1}{2}\tilde{\mathbf{x}}_i^\mathsf{T}(\tilde{\mathbf{q}} - 2\mathbf{C}_i)\tilde{\mathbf{x}}_i + \tilde{\mathbf{x}}_i^\mathsf{T} \mathbf{h}_i \right).$$

We will soon evaluate the $\tilde{\mathbf{q}}$-integral by saddle-point approximation. We now assume that the dominating saddle-point belongs to a set

$$D_\epsilon := \{ \tilde{\mathbf{q}} \in \mathbb{R}^{n \times n} : \tilde{\mathbf{q}} - 2\mathbf{C} \succ \epsilon I_n \text{ for all } \mathbf{C} \text{ living on its domain} \},$$

for some arbitrarily small $\epsilon > 0$ but independent of N. Thus restricting the integral to this domain yields a sub-leading correction $\exp o(N)$. For $\tilde{\mathbf{q}} \in D_\epsilon$ a Gaussian integration over $\tilde{\mathbf{x}}$ is possible: $\exp(NI_N)$ equals

$$\frac{(2\pi)^{Nn/2}e^{o(N)}}{\mathcal{Z}(\mathbf{q})} \int_{D_\epsilon} d\tilde{\mathbf{q}} \exp \frac{N}{2}\frac{1}{N} \sum_{i \leq N} \left(\mathrm{Tr}\mathbf{q}\tilde{\mathbf{q}} + \mathbf{h}_i^\mathsf{T}(\tilde{\mathbf{q}} - 2\mathbf{C}_i)^{-1}\mathbf{h}_i - \ln\det(\tilde{\mathbf{q}} - 2\mathbf{C}_i) \right)$$

$$= \frac{(2\pi)^{Nn/2}e^{o(N)}}{\mathcal{Z}(\mathbf{q})} \int_{D_\epsilon} d\tilde{\mathbf{q}} \exp \left\{ \frac{N}{2}\mathbb{E}\left(\mathrm{Tr}\mathbf{q}\tilde{\mathbf{q}} + \mathbf{h}^\mathsf{T}(\tilde{\mathbf{q}} - 2\mathbf{C})^{-1}\mathbf{h} - \ln\det(\tilde{\mathbf{q}} - 2\mathbf{C}) \right) \right\}.$$

We used the convergence of the empirical law of the sequence $(\mathbf{C}_i, \mathbf{h}_i)_i$ to turn the above empirical mean into a statistical expectation over $(\mathbf{C}, \mathbf{h})$, including the correction in the $\exp o(N)$; this is possible because over D_ϵ the summand is a bounded continuous function of $(\mathbf{C}_i, \mathbf{h}_i)$. As N

diverges at fixed n we can estimate the integral by saddle-point and reach that the generalized spherical integral is

$$I_N \to \frac{1}{2}\text{extr}_{\tilde{\mathbf{q}}}\big(\text{Tr}\mathbf{q}\tilde{\mathbf{q}} + \mathbb{E}\mathbf{h}^\intercal(\tilde{\mathbf{q}} - 2\mathbf{C})^{-1}\mathbf{h} - \mathbb{E}\ln\det(\tilde{\mathbf{q}} - 2\mathbf{C})\big)$$

$$- \frac{1}{2}\text{extr}_{\tilde{\mathbf{q}}}\big(\text{Tr}\mathbf{q}\tilde{\mathbf{q}} - \ln\det\tilde{\mathbf{q}}\big) \tag{5.33}$$

where the term $-\ln\mathcal{Z}(\mathbf{q})/N$ from the normalization has been obtained by simply setting $\mathbf{C}$ and $\mathbf{h}$ to all-zeros in the first optimization problem. The extremum is over $n \times n$ symmetric matrices $\tilde{\mathbf{q}}$ such $\tilde{\mathbf{q}} - 2\mathbf{C}$ is positive definite for all $\mathbf{C}$ on its domain.

Assuming that the extremum is attained inside the optimization domain we can perform the extremization using $\ln\det\mathbf{A} = \text{Tr}\ln\mathbf{A}$. The extremum is solution of the matrix equation

$$\mathbf{q} = \mathbb{E}\mathbf{h}^\intercal(\tilde{\mathbf{q}} - 2\mathbf{C})^{-2}\mathbf{h} + \mathbb{E}(\tilde{\mathbf{q}} - 2\mathbf{C})^{-1}. \tag{5.34}$$

The second extremization leads instead to $\tilde{\mathbf{q}} = \mathbf{q}^{-1}$. Thus the result.

5.3   Information-theoretic analysis by the replica method

Let us start with a remark. Express the noise $\mathbf{Z} = \mathbf{O}^\intercal\mathbf{D}\mathbf{O}$ in terms of its random Haar distributed basis $\mathbf{O}$ and eigenvalues $\mathbf{D}$, so that the observation model becomes

$$\mathbf{Y} = \frac{\lambda}{N}\mathbf{P}^* + \mathbf{O}^\intercal\mathbf{D}\mathbf{O}. \tag{5.35}$$

When the signal is rotationally invariant we can consider the noise diagonal right away by absorbing $\mathbf{O}$ into $\mathbf{x}, \mathbf{X}^*$. If the law P_X is uniform on the sphere, then this joint rotation does not change the distribution of $\mathbf{x}, \mathbf{X}^*$ which greatly simplifies the analysis. In this simpler case, the replica method is not needed as the computation of the free entropy can be carried out simply using a saddle point method. We provide this analysis in Section 5.3.4. The rotational invariance of the Gaussian law implies that also that case could be treated similarly by direct computation. On the contrary, for other priors than spherical or Gaussian this is no longer possible and the replica method is needed.

In order to deal with such non-rotational invariant priors we are going to adapt an approach developed by Kabashima in [153, 131] to study certain inference models where rotational invariant random matrices appear as quenched disorder. The main difference compared to the works is the fact that because they consider (generalized) linear regression, the structured matrix plays the role of covariates/data and therefore does not influence the form of the likelihood when writing the posterior. A novelty of the present setting is the fact that because the structured matrix is now the noise itself, the likelihood is a function of its statistics which in turn complicates the analysis.

The goal here is to compute the log-partition function (5.12) using the replica trick

$$\lim_{N \to \infty} \frac{1}{N} \mathbb{E} \ln \mathcal{Z} = \lim_{N \to \infty} \frac{1}{N} \lim_{n \to 0} \partial_n \ln \mathbb{E} \mathcal{Z}^n = \lim_{n \to 0} \partial_n \lim_{N \to \infty} \frac{1}{N} \ln \mathbb{E} \mathcal{Z}^n. \tag{5.36}$$

The expectation is with respect to $\mathbf{Y}$ or equivalently the independent $\mathbf{O}, \mathbf{x}_0$ (recall $\mathbf{D}$ is deterministic). The last equality *assumes* the commutation of the two limits. Another key assumption of the method is that we are going to make the computation considering $n \in \mathbb{N}$ and then assume an analytic continuation to n in a small neighborhood of 0. Before doing all that we are going to first re-express our model in a form more convenient for analysis.

5.3.1 An equivalent quadratic model

The Hamiltonian (5.11) of the model can be written in a more convenient way by introducing the following shorthand notations for order parameters. Despite at the moment only vector $\mathbf{x}$ has been introduced, soon a family of vectors $(\mathbf{x}_\ell)$ will be introduced when "replicating" the system. So we directly introduce the order parameters for these:

$$v_\ell = v(\mathbf{x}_\ell) := \frac{1}{N} \|\mathbf{x}_\ell\|^2, \tag{5.37}$$

$$M_{(k)\ell} = M_{(k)}(\mathbf{x}_\ell, \mathbf{Z}) := \frac{1}{N} \mathbf{x}_\ell^\mathsf{T} \mathbf{Z}^k \mathbf{x}_\ell, \tag{5.38}$$

$$\kappa_\ell = \kappa(\mathbf{x}_\ell, \mathbf{x}_0, \mathbf{Z}) := \frac{1}{N} \mathbf{x}_\ell^\mathsf{T} \mathbf{Z} \mathbf{x}_0, \tag{5.39}$$

$$q_{\ell\ell'} = q(\mathbf{x}_\ell, \mathbf{x}_{\ell'}) := \frac{1}{N} \mathbf{x}_\ell^\mathsf{T} \mathbf{x}_{\ell'}, \tag{5.40}$$

where the replica indices $0 \le \ell, \ell' \le n$ with the identification $\mathbf{x}_0 := \mathbf{X}^*$.

We now treat the quadratic and quartic part of the matrix potential separately. Let us denote

$$\Delta := \frac{1}{N}(\mathbf{P}^* - \mathbf{P}), \qquad M_\ell := M_{(1)\ell}.$$

The quadratic part yields a contribution:

$$\frac{N}{4} \mathrm{Tr}[(\mathbf{Z} + \lambda\Delta)^2 - \mathbf{Z}^2] = \frac{1}{2} \left[\lambda(\mathbf{x}_0^\mathsf{T} \mathbf{Z} \mathbf{x}_0 - \mathbf{x}^\mathsf{T} \mathbf{Z} \mathbf{x}) + N\lambda^2 \left(\frac{1}{2N^2}(\|\mathbf{x}_0\|^4 + \|\mathbf{x}\|^4) - q_{01}^2 \right) \right]$$

$$= -\frac{N\lambda}{2} M_1 + \frac{N\lambda^2}{2} \left(\frac{1}{2}(v_0^2 + v_1^2) - q_{01}^2 \right) + o(N). \tag{5.41}$$

The subscript 1 indicates that only one replica $\mathbf{x}_1 := \mathbf{x}$ is involved yet, and by convention it is replica number one. We used that by the law of large numbers, and thanks to the symmetry of the chosen matrix potential, we can assert that

$$M_{(2k+1)0} = o_N(1), \qquad M_{(2)0} = 1 + o_N(1)$$

due to our choice of normalization, so in particular $M_0 = M_{(1)0} = o_N(1)$. Again by the law of large numbers we have

$$v_0 = \mathbb{E}(X_1^*)^2 + o_N(1) = 1 + o_N(1).$$

The quartic contribution is more complicated due to the non-commutativity of matrices:

$$
\frac{N}{8}\mathrm{Tr}[(\mathbf{Z} + \lambda\Delta)^4 - \mathbf{Z}^4]
$$

$$
= \frac{N}{8}\mathrm{Tr}[\lambda^4\Delta^4 + 4\lambda^3\mathbf{Z}\Delta^3 + 4\lambda^2\mathbf{Z}^2\Delta^2 + 4\lambda\mathbf{Z}^3\Delta + 2\lambda^2\mathbf{Z}\Delta\mathbf{Z}\Delta]
$$

$$
= \frac{N}{8}\Big[\lambda^4(2q_{01}^4 + v_1^4 + 1 - 4q_{01}^2(v_1^2 + 1 - v_1))
$$

$$
+ 4\lambda^3(M_0(1 - q_{01}^2) - M_1(v_1^2 - q_{01}^2) + 2q_{01}(v_1 - 1)\kappa_1)
$$

$$
+ 4\lambda^2\Big(M_{(2)0} + v_1\frac{1}{N}\mathbf{x}^\mathsf{T}\mathbf{Z}^2\mathbf{x} - 2q_{01}\frac{1}{N}\mathbf{x}^\mathsf{T}\mathbf{Z}^2\mathbf{x}_0\Big) + 4\lambda\Big(M_{(3)0} - \frac{1}{N}\mathbf{x}^\mathsf{T}\mathbf{Z}^3\mathbf{x}\Big)
$$

$$
+ 2\lambda^2(M_0^2 + M_1^2 - 2\kappa_1^2)\Big]. \tag{5.42}
$$

Note that the only three terms which we did not write in a compact form using order parameters are linear and quadratic forms in $\mathbf{x}$ that do *not* appear elsewhere to a power greater than 1. This is because introducing order parameters for these would add useless redundancy in the final equations (but it is necessary for the other order parameters due to powers of them appearing in the Hamiltonian). Let

$$
f_\ell = f(q_{0\ell}, v_\ell, M_\ell, \kappa_\ell) := \gamma\frac{\lambda^4}{8}\Big(2q_{0\ell}^4 + v_\ell^4 - 4q_{0\ell}^2(v_\ell^2 + 1 - v_\ell)\Big) - \gamma\frac{\lambda^3}{2}M_\ell(v_\ell^2 - q_{0\ell}^2)
$$

$$
+ \gamma\lambda^3 q_{0\ell}(v_\ell - 1)\kappa_\ell + \gamma\frac{\lambda^2}{4}M_\ell^2 - \gamma\frac{\lambda^2}{2}\kappa_\ell^2 + \mu\frac{\lambda^2}{2}\Big(\frac{1}{2}v_\ell^2 - q_{0\ell}^2\Big) - \mu\frac{\lambda}{2}M_\ell. \tag{5.43}
$$

Plugging the contributions we computed into (5.11) shows that the Hamiltonian is equivalently written as

$$
H_N(\mathbf{x}; \mathbf{Z}, \mathbf{x}_0) = Nf_1 + \gamma\frac{\lambda}{2}\mathbf{x}^\mathsf{T}(\lambda v_1\mathbf{Z}^2 - \mathbf{Z}^3)\mathbf{x} - \gamma q_{01}\lambda^2\mathbf{x}^\mathsf{T}\mathbf{Z}^2\mathbf{x}_0 + C + o(N), \tag{5.44}
$$

where we have put all irrelevant constants inside C. We will neglect the $o(N)$ contribution in the following as it yields a subleading correction to the free entropy. Also the constant C is irrelevant, so we simply forget about it. Keep in mind that at the moment f_1 is still a function of $\mathbf{x}$. This model is thus *not* (yet) quadratic in $\mathbf{x}$ due to terms such as $M_1(\mathbf{x}, \mathbf{Z})^2$ appearing in f_1.

We now use delta functions to fix various order parameters. We are going to use repeatedly the Fourier representation of the delta function, namely

$$
\delta(x) = \frac{1}{2\pi}\int d\hat{x}\exp(i\hat{x}x). \tag{5.45}
$$

Because the integrals we will end-up with will always be at some point evaluated by saddle point, implying a deformation of the integration contour in the complex plane, tracking the imaginary unit i in the delta functions will be irrelevant. Similarly, the normalization $1/(2\pi)$ will always contribute to sub-exponential corrections in the integrals at hand. Therefore, we will allow ourselves to formally write

$$\delta(x) = \int d\hat{x} \exp(r\hat{x}x) \tag{5.46}$$

for a convenient constant r, keeping in mind these considerations (again, as we evaluate the final integrals by saddle point, the choice of r ends-up being irrelevant).

We denote jointly $\boldsymbol{\tau} := (v_1, M_1, \kappa_1, q_{01})$ and $\hat{\boldsymbol{\tau}}$ for their Fourier conjugates. Coming back to the the partition function for this equivalent model (5.44), it can be re-expressed using delta functions as

$$\int dP_X(\mathbf{x}) d\boldsymbol{\tau} \exp\left(-H_N(\mathbf{x}; \mathbf{Z}, \mathbf{x}_0)\right)$$
$$\times \delta(Nq_{01} - \mathbf{x}^\mathsf{T}\mathbf{x}_0)\delta(Nv_1 - \|\mathbf{x}\|^2)\delta(NM_1 - \mathbf{x}^\mathsf{T}\mathbf{Z}\mathbf{x})\delta(N\kappa_1 - \mathbf{x}^\mathsf{T}\mathbf{Z}\mathbf{x}_0)$$
$$= \int dP_X(\mathbf{x}) d\boldsymbol{\tau} d\hat{\boldsymbol{\tau}} \exp\left(-H_N(\boldsymbol{\tau}, \hat{\boldsymbol{\tau}}, \mathbf{x}; \mathbf{x}_0, \mathbf{Z})\right), \tag{5.47}$$

where

$$H_N(\boldsymbol{\tau}, \hat{\boldsymbol{\tau}}, \mathbf{x}; \mathbf{x}_0, \mathbf{Z}) := Nh(\boldsymbol{\tau}, \hat{\boldsymbol{\tau}}) + \mathbf{x}^\mathsf{T}\mathbf{J}_1(\boldsymbol{\tau}, \hat{\boldsymbol{\tau}}, \mathbf{Z})\mathbf{x} + \mathbf{x}^\mathsf{T}\mathbf{J}_0(\boldsymbol{\tau}, \hat{\boldsymbol{\tau}}, \mathbf{Z})\mathbf{x}_0 \tag{5.48}$$

and

$$h(\boldsymbol{\tau}, \hat{\boldsymbol{\tau}}) := f_1 - \hat{q}_{01}q_{01} - \frac{\hat{v}_1 v_1}{2} - \frac{\hat{M}_1 M_1}{2} - \hat{\kappa}_1 \kappa_1, \tag{5.49}$$

$$\mathbf{J}_1(\boldsymbol{\tau}, \hat{\boldsymbol{\tau}}, \mathbf{Z}) := \frac{\hat{v}_1}{2} I_N + \frac{\hat{M}_1}{2}\mathbf{Z} + \gamma\frac{\lambda^2}{2}v_1\mathbf{Z}^2 - \gamma\frac{\lambda}{2}\mathbf{Z}^3, \tag{5.50}$$

$$\mathbf{J}_0(\boldsymbol{\tau}, \hat{\boldsymbol{\tau}}, \mathbf{Z}) := \hat{q}_{01} I_N + \hat{\kappa}_1 \mathbf{Z} - \gamma q_{01}\lambda^2\mathbf{Z}^2. \tag{5.51}$$

So what this shows is that by introducing new variables (order parameters and conjugate Fourier parameters), the original model turns out being equivalent to an extended system with Hamiltonian (5.48). The key point of all this analysis is that by introducing the new variables $\boldsymbol{\tau}, \hat{\boldsymbol{\tau}}$ we have turned the interactions between the $(x_i)_{i \le N}$ into purely quadratic ones. This form is now more approriate to be solved using (generalizations of) known techniques. We emphasize that despite the algebraic manipulations leading from (5.11) to (5.48) are cumbersome, given a more complicated polynomial potential V the very same strategy could be applied but would require the introduction of more order parameters. Yet, the equivalent model would still collapse into a quadratic one of the above form but with a more complicated function h and matrices $\mathbf{J}_1, \mathbf{J}_0$ (still being polynomials of the noise $\mathbf{Z}$ of order one less than the order of V). The reason is that the key mechanisms behind these simplifications when expanding the original Hamiltonian (5.11) are stemming from the low-rank structure of the spike.

5.3.2　Replica symmetric free entropy using the inhomogeneous spherical integral

Having reduced the model to a quadratic one, we are now ready to replicate the system to compute the free entropy. The partition function $\mathcal{Z}$ is now computed using the equivalent model (5.48). The expected replicated partition function is

$$\mathbb{E}\mathcal{Z}(\mathbf{x}_0, \mathbf{Z})^n = \int \prod_{\ell=0}^n dP_X(\mathbf{x}_\ell) \prod_{\ell \leq n} d\boldsymbol{\tau}_\ell d\hat{\boldsymbol{\tau}}_\ell \, \mathbb{E}_{\mathbf{Z}} \exp\Big(-\sum_{\ell \leq n} H_N(\boldsymbol{\tau}_\ell, \hat{\boldsymbol{\tau}}_\ell, \mathbf{x}_\ell; \mathbf{x}_0, \mathbf{Z}) \Big), \qquad (5.52)$$

with replicas $(\mathbf{x}_\ell, \boldsymbol{\tau}_\ell, \hat{\boldsymbol{\tau}}_\ell)_{\ell \leq n}$ and shared quenched disorder $\mathbf{x}_0, \mathbf{Z}$. What we do next is to replace $\mathbf{Z}$ by $\mathbf{O}^{\mathsf{T}}\mathbf{D}\mathbf{O}$ and fix the overlap structure between replicas

$$\mathbf{x}_\ell^{\mathsf{T}}\mathbf{x}_{\ell'} = Nq_{\ell\ell'}, \qquad \ell, \ell' \leq n \qquad (5.53)$$

by introducing further variables and their Fourier conjugates (this is already taken care of for the overlaps $\mathbf{x}_\ell^{\mathsf{T}}\mathbf{x}_0$ with the planted signal). The purpose will become clear soon. Redefining $\boldsymbol{\tau}_\ell := (v_\ell, M_\ell, \kappa_\ell)$ and similarly for $\hat{\boldsymbol{\tau}}_\ell$, and defining the overlaps $\mathbf{q} = (q_{\ell\ell'})_{0 \leq \ell < \ell' \leq n}$ and similarly for $\hat{\mathbf{q}}$, the log-partition function can be recast as

$$\mathbb{E}\mathcal{Z}^n = \int d\mathbf{q}\, d\hat{\mathbf{q}} \prod_{\ell \leq n} d\boldsymbol{\tau}_\ell d\hat{\boldsymbol{\tau}}_\ell \exp N\Big(\sum_{\ell \leq n} \Big(\frac{\hat{v}_\ell v_\ell}{2} + \frac{\hat{M}_\ell M_\ell}{2} + \hat{\kappa}_\ell \kappa_\ell - f_\ell \Big) + \sum_{0 \leq \ell < \ell' \leq n} \hat{q}_{\ell\ell'} q_{\ell\ell'} \Big)$$
$$\times \int \prod_{\ell=0}^n dP_X(\mathbf{x}_\ell) \exp\Big(-\sum_{0 \leq \ell < \ell' \leq n} \hat{q}_{\ell\ell'} \mathbf{x}_\ell^{\mathsf{T}}\mathbf{x}_{\ell'} - \frac{1}{2}\sum_{\ell \leq n} \hat{v}_\ell \|\mathbf{x}_\ell\|^2 \Big)$$
$$\times \mathbb{E}_{\mathbf{O}} \exp \sum_{\ell \leq n} \Big((\mathbf{O}\mathbf{x}_0)^{\mathsf{T}}\mathbf{A}_\ell \mathbf{O}\mathbf{x}_\ell + \frac{1}{2}(\mathbf{O}\mathbf{x}_\ell)^{\mathsf{T}}\mathbf{B}_\ell \mathbf{O}\mathbf{x}_\ell \Big) \qquad (5.54)$$

where the $N \times N$ "replica coupling matrices" are

$$\mathbf{A}_\ell := -\hat{\kappa}_\ell \mathbf{D} + \gamma q_{0\ell} \lambda^2 \mathbf{D}^2, \qquad (5.55)$$
$$\mathbf{B}_\ell := -\hat{M}_\ell \mathbf{D} - \gamma \lambda^2 v_\ell \mathbf{D}^2 + \gamma \lambda \mathbf{D}^3. \qquad (5.56)$$

We now assume a replica-symmetric ansatz which should lead to the correct solution due to the strong concentration-of-measure effects taking place in the Bayes-optimal setting as well as the Nishimori identities [29, 40]. It means that we assume that the saddle point over the order parameters dominating the partition function as $N \to \infty$, which are finitely many, lies in the subset verifying the following (note the minus sign introduced for $-\hat{q}$ and $-\hat{m}$ for convenience):

for all $\ell \neq \ell' = 1, \ldots, n$

$$\text{Replica Symmetry Ansatz:} \quad \begin{cases} M_\ell = M, & \hat{M}_\ell = \hat{M}, \\ \kappa_\ell = \kappa, & \hat{\kappa}_\ell = \hat{\kappa}, \\ v_\ell = v, & \hat{v}_\ell = \hat{v}, \\ q_{\ell\ell'} = q, & \hat{q}_{\ell\ell'} = -\hat{q}, \\ q_{0\ell} = m, & \hat{q}_{0\ell} = -\hat{m}. \end{cases} \tag{5.57}$$

Using this ansatz, the matrices $(\mathbf{A}_\ell, \mathbf{B}_\ell)_{\ell \leq n}$ become independent of ℓ. We thus call their common value $\mathbf{A}, \mathbf{B}$. As a consequence the term $\mathbb{E}_{\mathbf{O}}(\cdot)$ at the third line in (5.54) is recognized to be what we call an *inhomogeneous spherical integral* defined and analyzed in a devoted Section 5.2.2. From Section 5.2 we know that the result of such integral depends only on the overlap structure; this is the reason why we fixed it earlier. We will thus replace it by $\exp N I_{A,B}^{\text{RS}}(n, v, m, q)$ whose formula is (5.32) and which is parametrized by the random variables (below $D \sim \rho$)

$$A = -\hat{\kappa}D + \gamma m\lambda^2 D^2, \tag{5.58}$$

$$B = -\hat{M}D - \gamma\lambda^2 v D^2 + \gamma\lambda D^3. \tag{5.59}$$

Notice that at this point the only $\mathbf{x}$-integrals remaining (second line of (5.54)) are completely factorized over the spin indices i. Hence after taking the saddle point the log-replicated free entropy becomes in the limit $N \to \infty$

$$\frac{1}{N} \ln \mathbb{E}\mathcal{Z}^n \to \text{extr}\left\{ n\left(\frac{\hat{v}v}{2} + \frac{\hat{M}M}{2} + \hat{\kappa}\kappa - \hat{m}m + \frac{1-n}{2}\hat{q}q - f(m, v, M, \kappa)\right) \right.$$
$$\left. + I_{A,B}^{\text{RS}}(n, v, m, q) + \ln \int \prod_{\ell=0}^{n} dP_X(x_\ell) e^{\hat{q}\sum_{\ell < \ell' \leq n} x_\ell x_{\ell'} + \hat{m}\sum_{\ell \leq n} x_0 x_\ell - \frac{\hat{v}}{2}\sum_{\ell \leq n} x_\ell^2}\right\}$$

where the extremum is over all scalars in (5.57). The last line can be treated by a Hubbard-Stratonovič transform (i.e., Gaussian integral formula) to decouple the integral over the replica indices. Doing so it becomes

$$\mathbb{E}\left(\int dP_X(x) \exp\left(\sqrt{\hat{q}}Zx - \frac{\hat{q}+\hat{v}}{2}x^2 + \hat{m}X_0 x\right)\right)^n,$$

with $Z \sim \mathcal{N}(0, 1), X_0 \sim P_X$.

We now consider the limit of number of replicas going to 0 assuming the analytic continuation of our formulas from integer n to real. To expand the latter term we use $\ln \mathbb{E}X^n = n\mathbb{E}\ln X + O(n^2)$. The inhomogeneous spherical integral given by (5.32) (with $v_0 = 1$) also has to be expanded in n. We get

$$I_{A,B}^{\text{RS}}(n, v, m, q) = \frac{n}{2}\text{extr}_{(\tilde{v}, \tilde{m}, \tilde{q})}\left\{2m\tilde{m} + v\tilde{v} - q\tilde{q} - \mathbb{E}\ln(\tilde{v} - B - \tilde{q})\right.$$
$$\left. - \mathbb{E}\frac{\tilde{q} - (\tilde{m} - A)^2}{\tilde{v} - B - \tilde{q}}\right\} - \frac{n}{2}\big(1 + \ln(v - q)\big) - \frac{n}{2}\frac{q - m^2}{v - q} + O(n^2)$$

with an expectation over $D \sim \rho$ entering A, B. Now we plug the previous expressions in the log-replicated partition function and expand up to $O(n)$ the resulting expression:

$$
\frac{1}{N} \ln \mathbb{E} \mathcal{Z}^n \to \operatorname{extr}\Big\{ n\Big(\frac{\hat{v}v}{2} + \frac{\hat{M}M}{2} + \hat{\kappa}\kappa - \hat{m}m + \frac{1-n}{2}\hat{q}q - f(m,v,M,\kappa)\Big)
$$
$$
+ I^{\mathrm{RS}}_{A,B}(n,v,m,q) + n \mathbb{E} \ln \int dP_X(x) \exp\Big(\sqrt{\hat{q}}Zx - \frac{\hat{q}+\hat{v}}{2}x^2 + \hat{m}X_0 x\Big) \Big\} + O(n^2).
$$

One can check that as it should $\lim_{N \to \infty} N^{-1} \ln \mathbb{E} Z^n$ vanishes when $n \to 0$. Taking the n-derivative (recall (5.36)) and then sending $n \to 0$ the final formula for the free entropy is obtained (and recalling that we dropped irrelevant constants along the computation):

$$
\frac{1}{N}\mathbb{E}\ln\mathcal{Z} \to \operatorname{extr}\Big\{ \frac{\hat{v}v}{2} + \frac{\hat{M}M}{2} + \hat{\kappa}\kappa - \hat{m}m + \frac{\hat{q}q}{2} + m\tilde{m} + \frac{v\tilde{v}}{2} - \frac{q\tilde{q}}{2}
$$
$$
- \gamma\frac{\lambda^4}{8}\Big(2m^4 + v^4 - 4m^2(v^2 + 1 - v)\Big) + \gamma\frac{\lambda^3}{2}M(v^2 - m^2)
$$
$$
- \gamma\lambda^3 m(v-1)\kappa - \gamma\frac{\lambda^2}{4}M^2 + \gamma\frac{\lambda^2}{2}\kappa^2 - \mu\frac{\lambda^2}{2}\Big(\frac{1}{2}v^2 - m^2\Big) + \mu\frac{\lambda}{2}M
$$
$$
+ \mathbb{E}\ln\int dP_X(x)\exp\Big(\sqrt{\hat{q}}Zx - \frac{\hat{q}+\hat{v}}{2}x^2 + \hat{m}X_0 x\Big)
$$
$$
- \frac{1}{2}\mathbb{E}\ln(\tilde{v} - \tilde{q} + \hat{M}D + \gamma\lambda^2 vD^2 - \gamma\lambda D^3) - \frac{1}{2}\ln(v - q) - \frac{q - m^2}{2(v-q)}
$$
$$
+ \frac{1}{2}\mathbb{E}\frac{(\tilde{m} + \hat{\kappa}D - \gamma m\lambda^2 D^2)^2 - \tilde{q}}{\tilde{v} - \tilde{q} + \hat{M}D + \gamma\lambda^2 vD^2 - \gamma\lambda D^3}\Big\} + \text{constant}. \tag{5.60}
$$

The extremization is intended over the set of 13 variational parameters v, $\hat{v}$, $\tilde{v}$, m, $\hat{m}$, $\tilde{m}$, q , $\hat{q}$, $\tilde{q}$, M, $\hat{M}$, κ, $\hat{\kappa}$. However, as we shall see later the saddle point equations will reduce only to two, because thanks to the Nishimori identities the saddle point values of many order parameters can be found right away. This is a specific and rather convenient feature of the Bayesian-optimal setting.

5.3.3   Replica saddle point equations

Define the following random local measure

$$
\langle \,\cdot\, \rangle_{\hat{m},\hat{q},\hat{v}} = \frac{\int dP_X(x)e^{\sqrt{\hat{q}}Zx + \hat{m}xX_0 - \frac{\hat{q}+\hat{v}}{2}x^2}(\,\cdot\,)}{\int dP_X(x)e^{\sqrt{\hat{q}}Zx + \hat{m}xX_0 - \frac{\hat{q}+\hat{v}}{2}x^2}}, \tag{5.61}
$$

the randomness being $Z \sim \mathcal{N}(0,1)$ and $X_0 \sim P_X$, and the random functions (random in $D \sim \rho$)

$$
H = (\tilde{v} - \tilde{q} + \hat{M}D + \gamma\lambda^2 vD^2 - \gamma\lambda D^3)^{-1}, \tag{5.62}
$$
$$
Q = \gamma m\lambda^2 D^2 - \hat{\kappa}D - \tilde{m}. \tag{5.63}
$$

Below follow the saddle point equations obtained by equating to 0 the gradient w.r.t. the variational parameters of the variational free entropy in (5.60). The parameter associated to each equation are reported in the round parenthesis:

$$(m) \quad \mu\lambda^2 m + \gamma\lambda^4 m(v^2 + 1 - v - m^2) - \gamma\lambda^3 Mm - \hat{m} + \tilde{m} + \frac{m}{v-q}$$
$$- \gamma\lambda^3(v-1)\kappa + \gamma\lambda^2 \mathbb{E}QHD^2 = 0$$

$$(\hat{m}) \quad m = \mathbb{E}X_0\langle X\rangle_{\hat{m},\hat{q},\hat{v}}$$

$$(\tilde{m}) \quad m = \mathbb{E}QH$$

$$(q) \quad \hat{q} - \tilde{q} = \frac{q - m^2}{(v-q)^2}$$

$$(\hat{q}) \quad q = \mathbb{E}\langle X\rangle^2_{\hat{m},\hat{q},\hat{v}}$$

$$(\tilde{q}) \quad q = \mathbb{E}(Q^2 - \tilde{q})H^2$$

$$(v) \quad -\mu\lambda^2 v - \gamma\lambda^4(v^3 - m^2(2v-1)) + 2\gamma\lambda^3 Mv + \hat{v} + \tilde{v} - \frac{1}{v-q} - \frac{m^2 - q}{(v-q)^2}$$
$$- \gamma\lambda^2 \mathbb{E}HD^2 - 2\gamma m\lambda^3\kappa + \gamma\lambda^2\mathbb{E}D^2(\tilde{q} - Q^2)H^2 = 0$$

$$(\hat{v}) \quad v = \mathbb{E}\langle X^2\rangle_{\hat{m},\hat{q},\hat{v}}$$

$$(\tilde{v}) \quad v = \mathbb{E}[H + H^2(Q^2 - \tilde{q})]$$

$$(M) \quad \mu\lambda + \gamma\lambda^3(v^2 - m^2) - \gamma\lambda^2 M + \hat{M} = 0$$

$$(\hat{M}) \quad M = \mathbb{E}D[H + H^2(Q^2 - \tilde{q})]$$

$$(\kappa) \quad \hat{\kappa} = \gamma\lambda^3 m(v-1) - \gamma\lambda^2\kappa$$

$$(\hat{\kappa}) \quad \kappa = \mathbb{E}DQH$$

As in any replica symmetric mean-field theory, the physical meaning of some order parameters makes it possible to fix their values to their expectation, obtainable using the Nishimori identities and, as a consequence, to drastically reduce this 13-dimensional system. To begin with, recall that we fixed v to be the squared norm of a sample from the posterior re-scaled by the number of components. Assuming concentration effects take place as they should in this optimal setting, and denoting the posterior mean by $\langle\cdot\rangle$, using the Nishimori identity we have that

$$v = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\langle\|\mathbf{x}\|^2\rangle = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\|\mathbf{X}^*\|^2 = 1. \tag{5.64}$$

We have $\hat{v} = 0$ because the constraint $v = 1$ is enforced by the prior without the need of a delta constraint. The (κ)-equation can then be used to directly eliminate $\hat{\kappa}$ by inserting $\hat{\kappa} = -\gamma\lambda^2\kappa$ into Q. The Nishimori identity also imposes

$$m = \mathbb{E}X_0\langle X\rangle_{\hat{m},\hat{q},0} = q = \mathbb{E}\langle X\rangle^2_{\hat{m},\hat{q},0}. \tag{5.65}$$

It is not difficult to realize that for this to be true one also needs necessarily $\hat{m} = \hat{q}$. So we have 8 variables left. The most tricky parameter is M, that we introduced to decouple the four body interactions in the Hamiltonian. Notice first that (recall definitions (5.38) and (5.40))

$$
\begin{aligned}
\frac{1}{N}\mathbb{E}\langle \mathbf{x}^\mathsf{T}\mathbf{Z}\mathbf{x}\rangle &= \frac{1}{N}\mathbb{E}\Big\langle \mathbf{x}^\mathsf{T}\Big(\mathbf{Y} - \frac{\lambda}{N}\mathbf{P}^*\Big)\mathbf{x}\Big\rangle \\
&= \frac{1}{N}\mathbb{E}\mathbf{X}^{*\mathsf{T}}\mathbf{Y}\mathbf{X}^* - \lambda\mathbb{E}\Big\langle \Big(\frac{1}{N}\mathbf{x}^\mathsf{T}\mathbf{X}^*\Big)^2\Big\rangle \\
&= \lambda\Big(1 - \mathbb{E}\Big\langle \Big(\frac{1}{N}\mathbf{x}^\mathsf{T}\mathbf{X}^*\Big)^2\Big\rangle\Big) + O\Big(\frac{1}{N}\Big).
\end{aligned}
$$

We used that by the Nishimori identity

$$
\frac{\mathbb{E}\langle \mathbf{x}^\mathsf{T}\mathbf{Y}\mathbf{x}\rangle}{N} = \frac{\mathbb{E}\mathbf{X}^{*\mathsf{T}}\mathbf{Y}\mathbf{X}^*}{N} = \frac{1}{N}\mathbb{E}\mathbf{X}^{*\mathsf{T}}\Big(\frac{\lambda}{N}\mathbf{X}^*\mathbf{X}^{*\mathsf{T}} + \mathbf{Z}\Big)\mathbf{X}^* = (\mathbb{E}(X_1^*)^2)^2\lambda = \lambda. \tag{5.66}
$$

Indeed, by diagonalizing the noise,

$$
\mathbb{E}\mathbf{X}^{*\mathsf{T}}\mathbf{Z}\mathbf{X}^* = \mathbb{E}\sum_{i\leq N} s_i^2 D_i = \mathbb{E}\|\mathbf{X}^*\|^2\mathbb{E}D_1 = 0,
$$

where $\mathbf{s}$ is a uniform spherical vector of same norm as $\mathbf{X}^*$, and $\mathbb{E}D_1 = 0$ by symmetry. By concentration happening on the Nishimori line [40] we have

$$
\mathbb{E}\Big\langle \Big(\frac{1}{N}\mathbf{x}^\mathsf{T}\mathbf{X}^*\Big)^2\Big\rangle = \Big(\mathbb{E}\Big\langle \frac{1}{N}\mathbf{x}^\mathsf{T}\mathbf{X}^*\Big\rangle\Big)^2 + o_N(1) = m^2 + o_N(1).
$$

Hence

$$
M = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\langle \mathbf{x}^\mathsf{T}\mathbf{Z}\mathbf{x}\rangle = \lambda(1 - m^2). \tag{5.67}
$$

The (M)-equation together with the other identities implies $\hat{M} = -\mu\lambda$. To summarize the Nishimori identities and concentration properties enforce five constraints:

$$
v = 1, \quad \hat{v} = 0, \quad m = q, \quad \hat{m} = \hat{q}, \quad M = \lambda(1 - m^2) \tag{5.68}
$$

and we have 6 variables left. Our updated definitions of Q and H are

$$
Q = \gamma m\lambda^2 D^2 + \gamma\lambda^2\kappa D - \tilde{m}, \tag{5.69}
$$
$$
H = (\tilde{v} - \tilde{q} - \mu\lambda D + \gamma\lambda^2 D^2 - \gamma\lambda D^3)^{-1}. \tag{5.70}
$$

Using the Nishimori identities we see from the $(\tilde{v})$ and $(\tilde{q})$-equations that

$$
m = \mathbb{E}H^2(Q^2 - \tilde{q}) \quad \Rightarrow \quad \mathbb{E}H = 1 - m. \tag{5.71}
$$

The latter has to be interpreted as an equation for the quantity $\tilde{V} := \tilde{v} - \tilde{q}$ as a function of m. Furthermore, one can now express $\tilde{m}$ as a function of κ and m. In fact from equation $(\tilde{m})$, unfolding Q and then solving for $\tilde{m}$, one gets

$$\tilde{m} = \frac{\gamma\lambda^2}{1-m}\mathbb{E}D(mD+\kappa)H - \frac{m}{1-m}. \tag{5.72}$$

Plugging this back into the (m)-equation we get $\hat{m}$, equation (5.78). We stress that inside H there is still an m dependency through $\tilde{V}$.

With all these simplifications we can close the equations on (m, κ) only:

$$(\hat{m}) \quad m = \mathbb{E}X_0\langle X\rangle_{\hat{m},\hat{m},0} \tag{5.73}$$

$$(\tilde{\kappa}) \quad \kappa = \mathbb{E}DQH, \tag{5.74}$$

where the random variables $Q = Q(m, \kappa, D)$ and $H = H(m, D)$ are

$$Q = \gamma m\lambda^2 D^2 + \gamma\lambda^2\kappa D - \frac{\gamma\lambda^2}{1-m}\mathbb{E}D(mD+\kappa)H + \frac{m}{1-m}, \tag{5.75}$$

$$H = (\tilde{V} - \mu\lambda D + \gamma\lambda^2 D^2 - \gamma\lambda D^3)^{-1}, \tag{5.76}$$

with $\tilde{V} = \tilde{V}(m)$ and $\hat{m} = \hat{m}(m, \kappa)$ being determined respectively by

$$\mathbb{E}H = 1 - m, \tag{5.77}$$

$$\hat{m} = \gamma\lambda^2\mathbb{E}HD\Big(\frac{mD+\kappa}{1-m} + DQ\Big) + \mu\lambda^2 m. \tag{5.78}$$

Then the replica prediction for the MMSE is

$$\lim_{N\to\infty}\frac{1}{2N^2}\mathbb{E}\|\mathbf{X}^*\mathbf{X}^{*\intercal} - \mathbb{E}[\mathbf{X}^*\mathbf{X}^{*\intercal} \mid \mathbf{Y}]\|_{\mathrm{F}}^2 = \frac{1}{2}(1 - m^2). \tag{5.79}$$

From (5.78) it is evident that when $\gamma = 0$ and $\mu = 1$ (to preserve unit variance of the noise), κ and $\hat{m}$ decouple, $\hat{m} = \lambda^2 m$, and the equation (5.73) reduces to the standard replica saddle point equation for the Wigner spike model.

There would be also an equation for $\tilde{q}$, that is decoupled though, meaning that $\tilde{q}$ is a simple function of m and κ in the end:

$$(q) \quad \tilde{q} = \hat{m}(m, \kappa) - \frac{m}{1-m}. \tag{5.80}$$

5.3.4 Spectral PCA is optimal for rotation-invariant signals

In this section we show that spectral PCA [38] is optimal for inferring $\mathbf{X}^*$ such that $\mathbf{X}^*$ equals in law $\mathbf{OX}^*$ for any orthogonal matrix $\mathbf{O}$. This is the case for Gaussian and spherically uniformly distributed $\mathbf{X}^*$.

To do so, we first show that the previous computations can be straightforwardly modified to accommodate the case of spherical prior. Let us assume that the signal $\mathbf{X}^*$ is uniformly distributed on a sphere of radius $\sqrt{N}$. We denote the uniform measure on this sphere by ω. Thanks to the invariance property of the measure on the sphere under rotations we know that $\mathbf{x}$ equals in law $\mathbf{Ox}$ for $\mathbf{x} \sim \omega$ and any orthogonal matrix $\mathbf{O}$. Therefore, we can directly diagonalize the noise without loss of generality and work with the equivalent model

$$\mathbf{Y} = \frac{\lambda}{N}\mathbf{P}^* + \mathbf{D}. \tag{5.81}$$

In this way we can get rid of $\mathbf{O}$ and as a consequence replicating the system and the inhomogeneous spherical integral becomes useless. Only Gaussian integrations and a saddle point estimation are needed.

The partition function is (5.48)–(5.51) but with the diagonal matrix $\mathbf{D}$ replacing $\mathbf{Z}$ (the constraint $\|\mathbf{x}\|^2 = N$ is taken care of by the Hamiltonian):

$$\int d\mathbf{x}\, d\boldsymbol{\tau}\, d\hat{\boldsymbol{\tau}}\, \exp\big(-Nh(\boldsymbol{\tau},\hat{\boldsymbol{\tau}}) - \mathbf{x}^\mathsf{T}\mathbf{J}_1(\boldsymbol{\tau},\hat{\boldsymbol{\tau}},\mathbf{D})\mathbf{x} - \mathbf{x}^\mathsf{T}\mathbf{J}_0(\boldsymbol{\tau},\hat{\boldsymbol{\tau}},\mathbf{D})\mathbf{x}_0 \big). \tag{5.82}$$

Because now $\mathbf{J}_1$ and $\mathbf{J}_0$ are diagonal matrices, the $\mathbf{x}$-integral in the partition function is just a Gaussian integral: it is (up to an irrelevant multiplicative constant)

$$\int d\boldsymbol{\tau}\, d\hat{\boldsymbol{\tau}}\, \exp N\Big(-h(\boldsymbol{\tau},\hat{\boldsymbol{\tau}}) - \frac{1}{2N}\sum_{i\leq N}\ln J_{1,i} + \frac{1}{4N}\sum_{i\leq N}x_{0,i}^2\frac{J_{0,i}^2}{J_{1,i}} \Big) \tag{5.83}$$

with $v_1 = 1$ (appearing in h). Because $\mathbf{x}_0$ is a uniform spherical vector combined with the convergence of the empirical law of (D_i) we have

$$-\frac{1}{2N}\sum_{i\leq N}\ln J_{1,i} + \frac{1}{4N}\sum_{i\leq N}x_{0,i}^2\frac{J_{0,i}^2}{J_{1,i}} = -\frac{1}{2}\mathbb{E}\ln J_{1,1} + \frac{1}{4}\mathbb{E}\frac{J_{0,1}^2}{J_{1,1}} + o_N(1).$$

Thus saddle point estimation of (5.82) yields

$$\frac{1}{N}\ln \mathcal{Z} \to \mathrm{const} + \mathrm{extr}\Big\{ -f(m,1,M,\kappa) + \hat{m}m + \frac{\hat{v}v}{2} + \frac{\hat{M}M}{2} + \hat{\kappa}\kappa$$
$$-\frac{1}{2}\mathbb{E}\ln\big(\hat{v} + \hat{M}D + \gamma\lambda^2D^2 - \gamma\lambda D^3\big) + \frac{1}{2}\mathbb{E}\frac{(\hat{m} + \hat{\kappa}D - \gamma\lambda^2mD^2)^2}{\hat{v} + \hat{M}D + \gamma\lambda^2D^2 - \gamma\lambda D^3}\Big\}, \tag{5.84}$$

where recall that f is defined by (5.43). Note that this strategy does not require the replica method, and it could also be applied in the case of Gaussian prior $P_X = \mathcal{N}(0,1)$, due to its rotational invariance.

At this point, the saddle point equations can be written and simplified similarly as in the previous section. After doing so and from the numerical solution of the saddle point equations,

one can deduce that: (*i*) in the case of spherical and Gaussian priors the MMSE is the same; and (*ii*) this MMSE matches the performance of the spectral PCA algorithm studied in [38]. Additionally, (*iii*) the MMSE obtained from this exact approach matches the replica prediction of the previous section in the case of Gaussian prior (a special case of factorized P_X tackled by our replica theory). This further confirms the validity and consistency of our methodology. Therefore we conclude that spectral PCA is Bayes-optimal in the special case of rotationally invariant priors and noise.

Let us provide a further argument in support of Bayes-optimality of PCA in the present setting. In this argument we consider the noise eigenvalues as quenched random variables, and we are going to average over them. We first notice that the MMSE estimator is diagonal in the basis of the matrix of data $\mathbf{Y}$. Indeed, letting $\mathbf{Y}$ be diagonalized as $\mathbf{Y} = \mathbf{U}^{\mathsf{T}}\mathbf{S}\mathbf{U}$ then using the posterior (5.6),

$$
\begin{aligned}
\mathbb{E}[\mathbf{X}^*\mathbf{X}^{*\mathsf{T}} \mid \mathbf{Y}] &= \frac{C_V}{P_Y(\mathbf{Y})} \int dP_X(\mathbf{x}) \exp\Big(-\frac{N}{2}\mathrm{Tr}V\Big(\mathbf{S} - \frac{\lambda}{N}(\mathbf{U}\mathbf{x})(\mathbf{U}\mathbf{x})^{\mathsf{T}}\Big)\Big)\mathbf{x}\mathbf{x}^{\mathsf{T}} \\
&= \frac{C_V}{P_Y(\mathbf{Y})}\mathbf{U}^{\mathsf{T}}\Big(\int dP_X(\mathbf{x})\exp\Big(-\frac{N}{2}\mathrm{Tr}V\Big(\mathbf{S} - \frac{\lambda}{N}\mathbf{x}\mathbf{x}^{\mathsf{T}}\Big)\Big)\mathbf{x}\mathbf{x}^{\mathsf{T}}\Big)\mathbf{U} \quad (5.85)
\end{aligned}
$$

where we changed $\mathbf{U}\mathbf{x}$ to $\mathbf{x}$, which leaves the prior invariant by rotational invariance. We would then like to see that the matrix

$$
\mathbf{L} = \frac{C_V}{P_Y(\mathbf{Y})} \int dP_X(\mathbf{x}) \exp\Big(-\frac{N}{2}\mathrm{Tr}V\Big(\mathbf{S} - \frac{\lambda}{N}\mathbf{x}\mathbf{x}^{\mathsf{T}}\Big)\Big)\mathbf{x}\mathbf{x}^{\mathsf{T}}
$$

is a diagonal. Indeed, because $\mathbf{S} = \mathrm{diag}(s_1, \ldots, s_N)$ is diagonal, $\mathrm{Tr}V(\mathbf{S} - (\lambda/N)\mathbf{x}\mathbf{x}^{\mathsf{T}})$ can be easily seen (see, e.g., the steps leading to (5.151)) to be a polynomial of degree k of the k variables

$$
\Big(\sum_{i \leq N} x_i^2, \sum_{i \leq N} s_i x_i^2, \ldots, \sum_{i \leq N} s_i^{k-1} x_i^2\Big).
$$

Then, for every $1 \leq j \leq N$, the integrand that defines $\mathbf{L}$ takes the same value for $\mathbf{x}$ and the point $\mathbf{x}'$ which results from changing the sign of the j-th coordinate of $\mathbf{x}$. We thus have that $\mathbf{L}$ is a diagonal matrix.

For $1 \leq k \leq N$, let $\mathbf{u}_k$ be the eigenvector of the k-largest eigenvalue of $\mathbf{Y}$. Then we can express $\mathbf{L}(\mathbf{Y})$ as $\mathrm{diag}(\gamma_1(\mathbf{Y}), \ldots, \gamma_N(\mathbf{Y}))$, where by definition we have that

$$
\mathbb{E}[\mathbf{X}^*\mathbf{X}^{*\mathsf{T}} \mid \mathbf{Y}] = \sum_{k \leq N} \gamma_k \mathbf{u}_k \mathbf{u}_K^{\mathsf{T}}, \quad (5.86)
$$

i.e., $\gamma_k = \mathbf{u}_k^{\mathsf{T}}\mathbb{E}[\mathbf{X}^*\mathbf{X}^{*\mathsf{T}} \mid \mathbf{Y}]\mathbf{u}_k$ with the ordering $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_N$. This therefore means that the "matrix magnetization" may be written according to

$$
\frac{1}{N^2}\mathbb{E}\mathrm{Tr}(\mathbb{E}[\mathbf{X}^*\mathbf{X}^{*\mathsf{T}} \mid \mathbf{Y}]\mathbf{X}^*\mathbf{X}^{*\mathsf{T}}) = \frac{1}{N^2}\sum_{k \leq N}\mathbb{E}[(\mathbf{u}_k^{\mathsf{T}}\mathbf{X}^*)^2\gamma_k].
$$

We would like now to compute the asymptotic magnetization of the Bayes estimator. For this we will use Nishimori identities and a bound over the projections of $\mathbf{X}^*$ onto the eigenvectors of $\mathbf{Y}$ that we verify numerically. More specifically, we will assume that there is some constant $K > 0$ such that for all $k \geq 2$ it holds that

$$(\mathbf{u}_k^\mathsf{T}\mathbf{X}^*)^2 \leq K. \tag{5.87}$$

As mentioned before, inequality (5.87), which is an explicit rate of convergence for the limit in [38, Theorem 2], has been verified through many numerical experiments for different noise potentials and SNRs. In every case, a bound of this type is observed, although for experiments close to the corresponding phase transition, the constant K takes larger values and the quantity bounded exhibits a larger variance (this type of behavior is expected to hold very close to the transition point).

Now, notice that by Nishimori identities the following holds

$$\mathbb{E}\gamma_k = \mathbb{E}(\mathbf{u}_k^\mathsf{T}\mathbf{X}^*)^2. \tag{5.88}$$

Also, by [38, Theorem 2] we have that (below R is the R-transform associated with the noise spectral density ρ)

$$\frac{1}{N^2}\mathbb{E}[\gamma_1(\mathbf{u}_1^\mathsf{T}\mathbf{X}^*)^2] = \frac{1}{N}\Big(1 - \frac{R'(1/\lambda)}{\lambda^2}\Big)\mathbb{E}\gamma_1 + \frac{1}{N}\mathbb{E}\Big[\gamma_1\Big(\frac{(\mathbf{u}_1^\mathsf{T}\mathbf{X}^*)^2}{N} - 1 + \frac{R'(1/\lambda)}{\lambda^2}\Big)\Big],$$

where the second term on the r.h.s. is a vanishing function of N. If we use (5.88) and [38, Theorem 2] a second time, we get that

$$\lim_{N\to\infty}\frac{1}{N^2}\mathbb{E}[\gamma_1(\mathbf{u}_1^\mathsf{T}\mathbf{X}^*)^2] = \Big(1 - \frac{R'(1/\lambda)}{\lambda^2}\Big)^2.$$

On the other hand, by inequality (5.87) and the Nishimori identities (5.88) we get

$$\frac{1}{N^2}\sum_{2\leq k\leq N}\mathbb{E}[\gamma_k(\mathbf{u}_k^\mathsf{T}\mathbf{X}^*)^2] \leq \frac{K}{N^2}\sum_{2\leq k\leq N}\mathbb{E}\gamma_k = \frac{K}{N^2}\sum_{2\leq k\leq N}\mathbb{E}(\mathbf{u}_k^\mathsf{T}\mathbf{X}^*)^2.$$

that by [38, Theorem 2], vanishes in the limit. We then conclude that

$$\frac{1}{N^2}\mathbb{E}\mathrm{Tr}(\mathbb{E}[\mathbf{X}^*\mathbf{X}^{*\mathsf{T}} \mid \mathbf{Y}]\mathbf{X}^*\mathbf{X}^{*\mathsf{T}}) = \Big(1 - \frac{R'(1/\lambda)}{\lambda^2}\Big)^2 + o_N(1).$$

This in turn implies that

$$\lim_{N\to\infty}\frac{1}{2N^2}\mathbb{E}\|\mathbb{E}[\mathbf{X}^*\mathbf{X}^{*\mathsf{T}} \mid \mathbf{Y}] - \mathbf{X}^*\mathbf{X}^{*\mathsf{T}}\|^2 = 1 - \Big(1 - \frac{R'(1/\lambda)}{\lambda^2}\Big)^2,$$

which is the MSE of the optimally scaled PCA estimator [38].

5.4 Sub-optimality of the previously proposed AMP

Consider the following AMP iteration for $t \geq 1$:

$$\mathbf{f}^t = \mathbf{Y}\mathbf{u}^t - \sum_{i=1}^{t} \mathsf{b}_{t,i}\mathbf{u}^i, \quad \mathbf{u}^{t+1} = h_{t+1}(\mathbf{f}^t). \tag{5.89}$$

Here, $\mathbf{f}^t = (f_1^t, \ldots, f_N^t), \mathbf{u}^{t+1} = (u_1^{t+1}, \ldots, u_N^{t+1}) \in \mathbb{R}^N$ and the *denoiser* function $h_{t+1} : \mathbb{R} \to \mathbb{R}$ is continuously differentiable, Lipschitz and applied component-wise, namely $h_{t+1}(\mathbf{f}^t) = (h_{t+1}(f_1^t), \ldots, h_{t+1}(f_N^t))$. The time-dependent AMP estimate of the spike $\mathbf{P}^*$ is $(\mathbf{u}^t)^\mathsf{T}\mathbf{u}^t$.

The Onsager coefficients $\{\mathsf{b}_{t,i}\}_{i\in[t],t\geq 1}$ are carefully chosen so that, conditioned on the signal, the empirical distribution of the components of iterate $\mathbf{f}^t$ is Gaussian. The form of these Onsager coefficients was derived by [123] using non-rigorous dynamic functional theory techniques, and a rigorous state evolution result was recently proved in [132]. More formally, assume that $\mathbf{X}^* \xrightarrow{W_2} X^*$. Then, the state evolution result of [132] gives that

$$(\mathbf{f}^1, \ldots, \mathbf{f}^t) \xrightarrow{W_2} (F_1, \ldots, F_t) := \boldsymbol{\mu}_t X^* + \mathbf{W}_t, \tag{5.90}$$

where $\boldsymbol{\mu}_t = (\mu_1, \ldots, \mu_t)$ and $\mathbf{W}_t = (W_1, \ldots, W_t)$ is a multivariate Gaussian with zero mean and covariance $\boldsymbol{\Sigma}_t = (\sigma_{ij})_{i,j\leq t}$ independent of X^*. Furthermore, the mean vectors $\{\boldsymbol{\mu}_t\}_{t\geq 1}$ and the covariance matrices $\{\boldsymbol{\Sigma}_t\}_{t\geq 1}$ are tracked by a deterministic state evolution recursion. We refer to [132] for more details on this AMP and associated state evolution. Such details won't be crucial for our argument, as we are going to focus directly on the fixed point performance, and not on the dynamics.

For this section, we restrict the analysis to (*i*) Rademacher prior $P_X = \frac{1}{2}(\delta_1 + \delta_{-1})$, and (*ii*) a "large enough" signal-to-noise ratio. We remark that our methodology extends to more generic factorized priors. However, since our goal is to prove sub-optimality of AMP, this setting suffices. Moreover, we will further restrict our proof of sub-optimality to (*iii*) the "one-step memory" version of the AMP in [132]. This means that the denoiser h_{t+1} in (5.89) is allowed to depend only on the past iterate $\mathbf{f}^t$. A more general "multi-step memory AMP" was proposed in [133], where the denoiser h_{t+1} can depend on *all* the past iterates $\mathbf{f}^1, \ldots, \mathbf{f}^t$. We remark that the analysis of [123] suggests that the fixed points of both these versions are the same; the longer memory of the latter AMP being only useful to improve its convergence properties. Therefore, despite our analysis below holds under hypotheses (*i*)–(*iii*), we conclude more generically that *the existing AMP algorithms for structured PCA in [132, 133] are sub-optimal, and this is the case for most SNR values and prior/signal's distributions that are not rotationally invariant*[2]. From the findings in the following sections, the reason for the sub-optimality of these AMPs

---

[2]We do not discard the possibility that for very peculiar choices of SNR regimes and/or priors these generically sub-optimal AMPs end-up being optimal, but that would be for highly specific setting-dependent reasons. One case where the AMPs of [132], and also the spectral PCA algorithm [38], are actually optimal is when the prior is rotationally invariant (spherical or Gaussian prior), see Section 5.3.4.

will become clear. Essentially, the data $\mathbf{Y}$ is *not* the best choice of matrix to use in the AMP iterates, despite being the most natural one.

5.4.1   Analysis of the one-step AMP fixed point performance

In this section we analyse the AMP algorithm (5.89) for structured PCA proposed in [132], with a posterior mean denoiser with a single-step memory term:

$$h_{t+1}(f_i^t) = \mathbb{E}[X \mid f_i^t]. \tag{5.91}$$

In [132, Section 3] it is shown that the fixed point of this AMP algorithm is, for λ sufficiently large, described by the following system:

$$1 - \Delta_* = \text{mmse}\Big(\frac{\lambda^2 \Delta_*^2}{\Sigma_*}\Big), \qquad \Sigma_* = \Delta_* R'\Big(\frac{\lambda \Delta_*(1 - \Delta_*)}{\Sigma_*}\Big). \tag{5.92}$$

Here, $R'(\cdot)$ denotes the derivative of the R transform of the (limiting) distribution of the noise eigenvalues D. For details about the R-transform, the interested reader is referred to [154]. The above is related to the asymptotic overlap of the AMP estimator through

$$\lim_{t \to \infty} \lim_{N \to \infty} \Big|\frac{1}{N}\mathbf{X}^{*\mathsf{T}}\hat{\mathbf{x}}^t\Big| = \lim_{t \to \infty} \lim_{N \to \infty} \frac{1}{N}\|\hat{\mathbf{x}}^t\|^2 = \Delta_* \tag{5.93}$$

and thus the AMP mean-square error is

$$\lim_{t \to \infty} \lim_{N \to \infty} \frac{1}{2N^2}\mathbb{E}\|\hat{\mathbf{x}}^t(\hat{\mathbf{x}}^t)^\mathsf{T} - \mathbf{X}^*\mathbf{X}^{*\mathsf{T}}\|^2 = \frac{1}{2}(1 - \Delta_*^2). \tag{5.94}$$

In the case of Rademacher prior the explicit form of the posterior-mean denoiser is

$$h_{t+1}(f^t) = \tanh\Big(\frac{f^t \mu_t}{\sigma_{tt}^2}\Big) \tag{5.95}$$

where (μ_t, σ_{tt}) are the mean and variance of the (empirically) "Gaussian observation" f^t computed from the state evolution of [132]. The associated mmse function is (below $Z \sim \mathcal{N}(0,1)$ is a standard Gaussian random variable and $X^* \sim \frac{1}{2}(\delta_{-1} + \delta_1)$)

$$\text{mmse}(x) = 1 - \mathbb{E}\Big[X^* \frac{\int dP_X(x)\, x\, e^{Zx\sqrt{\hat{m}} + \hat{m}xX^* - \frac{\hat{m}}{2}x^2}}{\int dP_X(x) e^{Zx\sqrt{\hat{m}} + \hat{m}xX^* - \frac{\hat{m}}{2}x^2}}\Big] \tag{5.96}$$

$$= 1 - \mathbb{E}\tanh(x + \sqrt{x}Z). \tag{5.97}$$

We now consider the limit $(\lambda, \Delta_*, \Sigma_*) \to (\infty, 1, 1)$ which indeed is a fixed point of (5.92) as we verify at the end of this section. Moreover it is unique, see [132, Theorem 3.1]. It implies

$x := \lambda^2 \Delta_*^2 / \Sigma_* \to \lambda^2 \to \infty$. We have in this limit

$$
\begin{aligned}
\mathrm{mmse}(x) &= 1 - \int dt \frac{e^{-\frac{1}{2x}(t-x)^2}}{\sqrt{2\pi x}} \tanh(t) \\
&= \sqrt{\frac{\pi}{2}} \frac{e^{-\frac{x}{2}}}{\sqrt{x}} (1 + O(1/x)) \\
&= \exp\left(-\frac{x}{2}(1 + o_x(1))\right).
\end{aligned}
\tag{5.98}
$$

We plug this in the first equation of (5.92) which gives at leading order

$$
\Delta_* = 1 - \exp\left(-\frac{\lambda^2}{2}(1 + o_\lambda(1))\right).
\tag{5.99}
$$

It just remains to check that $(\lambda, \Delta_*, \Sigma_*) = (\infty, 1, 1)$ is indeed the unique fixed point of (5.92) in the large SNR regime. From our analysis we already know that this fixed point is consistent with the first equation of (5.92). So we simply need to verify the second one, namely,

$$
R'\big(\lambda(1 - \Delta_*)(1 + o_\lambda(1))\big) \to 1
\tag{5.100}
$$

as $\lambda \to \infty$. From (5.99) we have in this limit $\lambda(1 - \Delta_*) \to 0$ exponentially fast in λ, and it can be readily verified that $R'(0) = 1$, as the noise distribution D has unit second moment. This ends the argument.

5.4.2   Analysis of the replica Bayes-optimal fixed point

We now analyse in the same large SNR regime the replica fixed point equations that we recall below for convenience: let us rename $\tilde{V} := \tilde{v} - \tilde{q}$ as they always appear together. We consider that all quantities below are at their saddle point values maximizing the replica free entropy (5.60).

Let us recall the outcome of the Section 5.3.3 on the saddle point equations. Consider the random variables (random through their dependence in D)

$$
Q = \gamma m \lambda^2 D^2 + \gamma \lambda^2 \kappa D - \frac{\gamma \lambda^2}{1-m} \mathbb{E}D(mD + \kappa)H + \frac{m}{1-m},
\tag{5.101}
$$

$$
H = (\tilde{V} - \mu\lambda D + \gamma\lambda^2 D^2 - \gamma\lambda D^3)^{-1}.
\tag{5.102}
$$

For a given value of the parameter m, the saddle point equations require $\tilde{V} = \tilde{V}(m)$ to be the solution of the implicit equation

$$
\mathbb{E}H = 1 - m.
\tag{5.103}
$$

Using this implicit solution, H is a function $H(m)$ and $Q = Q(m, \kappa)$. Let $Z \sim \mathcal{N}(0, 1)$ and $X^* \sim P_X$. The saddle point equations over the order parameters (m, κ) read

$$m = 1 - \text{mmse}(\hat{m}), \tag{5.104}$$

$$\kappa = \mathbb{E}DQH, \tag{5.105}$$

where $\text{mmse}(\hat{m})$ is the same function (5.97) as before and

$$\hat{m} = \hat{m}(m, \kappa) = \gamma\lambda^2 \mathbb{E}H\Big(\frac{mD^2 + \kappa D}{1 - m} + D^2 Q\Big) + \mu\lambda^2 m. \tag{5.106}$$

Recall that the replica prediction for the MMSE is (5.79). In the regime $\lambda \to \infty$ we thus necessarily have $m \to 1^-$. Since the solution (m, κ) of the replica saddle point equations yields the MMSE (5.79) which must be at least as good as the AMP MSE (5.94) then $m \geq \Delta_*$. Thus from (5.99) we deduce

$$1 - m = O\Big(\exp\Big(-\frac{\lambda^2}{2}(1 + o_\lambda(1))\Big)\Big). \tag{5.107}$$

The support of the density of D is bounded, therefore from (5.102) it is then clear that for (5.103) to be verified under the scaling (5.107) in the large λ limit, the solution $\tilde{V}$ of (5.103) must verify

$$\frac{\lambda^2}{\tilde{V}} = o_\lambda(1). \tag{5.108}$$

Thus from (5.103) we obtain

$$(1 - m)\tilde{V} = \mathbb{E}\Big(1 + \frac{\gamma\lambda D^2(\lambda - D) - \mu\lambda D}{\tilde{V}}\Big)^{-1} = 1 + o_\lambda(1) \tag{5.109}$$

from which we deduce using (5.107) that

$$\tilde{V} = \Theta\Big(\frac{1}{1 - m}\Big) = \Omega\Big(\exp\Big(\frac{\lambda^2}{2}(1 + o_\lambda(1))\Big)\Big). \tag{5.110}$$

This also implies that in the limit of large SNR, H becomes deterministic:

$$H = \tilde{V}^{-1} + O\Big(\frac{\lambda^2}{\tilde{V}^2}\Big). \tag{5.111}$$

This equality means that H can be written as $\tilde{V}^{-1}$ plus a possibly random term dependent of D, which can be bounded by a non-random constant of order $O(\lambda^2/\tilde{V}^2)$. Similarly for Q: using

that κ is bounded (recall that it is the limit of the expectation of (5.39)), (5.111) and (5.109), we get the following deterministic scaling in the large SNR regime:

$$Q = \frac{m}{1 - m} + O(\lambda^2). \tag{5.112}$$

Using all these scalings together with the fact that $\mathbb{E}D = 0$ and κ is bounded (actually it can now be seen from the $(\hat{\kappa})$-equation of Section 5.3.3 that $\kappa = o_\lambda(1)$) we reach, using $\mathbb{E}D^2 = 1$ and (5.108), (5.109),

$$
\begin{aligned}
\hat{m} &= \gamma \lambda^2 \mathbb{E}H\Big(\frac{mD^2 + \kappa D}{1 - m} + D^2 Q\Big) + \mu \lambda^2 m \\
&= \gamma \lambda^2 \Big(\tilde{V}^{-1} + O\Big(\frac{\lambda^2}{\tilde{V}^2}\Big)\Big)\Big(\frac{2m}{1 - m} + O(\lambda^2)\Big) + \mu \lambda^2 m \\
&= \gamma \lambda^2 \Big(\frac{2m}{\tilde{V}(1 - m)} + O\Big(\frac{\lambda^2}{\tilde{V}}\Big) + O\Big(\frac{\lambda^2}{\tilde{V}} \times \frac{1}{\tilde{V}(1 - m)}\Big) + O\Big(\frac{\lambda^4}{\tilde{V}^2}\Big)\Big) + \mu \lambda^2 m \\
&= \gamma \lambda^2 (2m + o_\lambda(1)) + \mu \lambda^2 m \\
&= \lambda^2 (2\gamma + \mu)(1 + o_\lambda(1))
\end{aligned}
\tag{5.113}
$$

where also used $m = 1 + o_\lambda(1)$, see (5.107). Recall $m = 1 - \mathrm{mmse}(\hat{m})$ as well as the scaling (5.98). So we have

$$m = 1 - \exp\Big(-\frac{\lambda^2}{2}(2\gamma + \mu)(1 + o_\lambda(1))\Big). \tag{5.114}$$

By comparing with (5.99) we see that $m \neq \Delta_*$. Moreover, since m is the Bayes-optimal overlap, it has to be the case that $m \geq \Delta_*$, namely, $2\gamma + \mu \geq 1$. From (5.17) it can be verified that $2\gamma + \mu > 1$ strictly for $\mu < 1$. Equality holds for the pure Wigner case ($\mu = 1, \gamma = 0$), as expected. This ends the proof that the MMSE (5.79) is asympotically in λ strictly exponentially smaller than the MSE of AMP with one-term memory (5.94) whenever $\mu < 1, \gamma > 0$.

5.4.3 What is actually doing this sub-optimal AMP? Mismatched estimation with Gaussian likelihood

In the same spirit as [70], we study here a mismatched estimation where the statistician assumes the noise to be Gaussian, thus a wrong likelihood, whereas the noise is drawn from the quartic ensemble with potential (5.13). In the same way as we did for the quartic potential, the mismatched posterior associated to (5.2) is written as

$$d\bar{P}_{X|Y}(\mathbf{x}|\mathbf{Y}) = \frac{1}{\bar{\mathcal{Z}}(\mathbf{Y})} dP_X(\mathbf{x}) \exp\Big(\frac{\lambda}{2} \mathrm{Tr}\mathbf{Y}\mathbf{x}\mathbf{x}^\intercal - \frac{\lambda^2}{4N} \|\mathbf{x}\|^4\Big) \tag{5.115}$$

where we have re-absorbed $\mathbf{x}$-independent terms in the normalization. The corresponding log-partition function is

$$\mathbb{E}\ln\bar{\mathcal{Z}}(\mathbf{Y}). \tag{5.116}$$

Notice that we have barred some quantities to distinguish them from their Bayes-optimal analogues. We further stress that, with Gaussian likelihood, the spin-glass model that arises already contains only two body interactions.

We aim at *approximating* (5.116). Indeed, we are going to perform a replica symmetric computation, which has no a-priori reasons to be exact as we are not anymore in the Bayesian-optimal setting [40] (nor the mismatched posterior is log-concave [72]; Chapter 4 serves as a counter-example). We denote jointly $\boldsymbol{\tau} = (v_1, q_{01})$ and $\hat{\boldsymbol{\tau}}$ their Fourier conjugates. The partition function can then be expressed using deltas to fix the $\boldsymbol{\tau}$ parameters and expanding $\mathbf{Y}$ as in (5.2). Up to irrelevant constants it reads

$$\bar{\mathcal{Z}}(\mathbf{Y}) = \int dP_X(\mathbf{x})d\boldsymbol{\tau}d\hat{\boldsymbol{\tau}}\exp\big(-\bar{H}_N(\boldsymbol{\tau},\hat{\boldsymbol{\tau}},\mathbf{x};\mathbf{x}_0,\mathbf{Z})\big) \tag{5.117}$$

where

$$\bar{H}_N(\boldsymbol{\tau},\hat{\boldsymbol{\tau}},\mathbf{x};\mathbf{x}_0,\mathbf{Z}) := N\bar{h}(\boldsymbol{\tau},\hat{\boldsymbol{\tau}}) + \mathbf{x}^\mathsf{T}\bar{\mathbf{J}}(\boldsymbol{\tau},\hat{\boldsymbol{\tau}},\mathbf{Z})\mathbf{x} + \hat{q}_{01}\mathbf{x}^\mathsf{T}\mathbf{x}_0 \tag{5.118}$$

and

$$\bar{h}(\boldsymbol{\tau},\hat{\boldsymbol{\tau}}) := \frac{\lambda^2}{4}v_1^2 - \frac{\lambda^2}{2}q_{01}^2 - q_{01}\hat{q}_{01} - \frac{v_1\hat{v}_1}{2}, \tag{5.119}$$

$$\bar{\mathbf{J}}(\boldsymbol{\tau},\hat{\boldsymbol{\tau}},\mathbf{Z}) := \frac{\hat{v}_1}{2}I_N - \frac{\lambda}{2}\mathbf{Z}. \tag{5.120}$$

While replicating we will need as before to fix the entire overlap structure (and not only q_{01}), i.e., $(N\mathbf{q})_{\ell\ell'} = Nq_{\ell\ell'} = \mathbf{x}_\ell^\mathsf{T}\mathbf{x}_{\ell'}$, the diagonal elements being denoted as v_ℓ. As usual, we also introduce the corresponding Fourier conjugates $\hat{\mathbf{q}}$. The expected replicated partition function then reads as

$$\mathbb{E}\bar{\mathcal{Z}}^n = \int d\mathbf{q}d\hat{\mathbf{q}}\exp N\Big(\sum_{\ell\leq n}\Big(\frac{\lambda^2}{2}q_{0\ell}^2 - \frac{\lambda^2}{4}v_\ell^2 + \frac{v_\ell\hat{v}_\ell}{2}\Big) + \sum_{0\leq\ell<\ell'\leq n}\hat{q}_{\ell\ell'}q_{\ell\ell'}\Big)$$

$$\times\int\prod_{\ell=0}^{n}dP_X(\mathbf{x}_\ell)\exp\Big(-\sum_{0\leq\ell<\ell'\leq n}\hat{q}_{\ell\ell'}\mathbf{x}_\ell^\mathsf{T}\mathbf{x}_{\ell'} - \frac{1}{2}\sum_{\ell\leq n}\hat{v}_\ell\|\mathbf{x}_\ell\|^2\Big)$$

$$\times\mathbb{E}_{\mathbf{O}}\exp\Big(\frac{\lambda}{2}\mathrm{Tr}\mathbf{O}\mathbf{D}\mathbf{O}^\mathsf{T}\sum_{\ell\leq n}\mathbf{x}_\ell\mathbf{x}_\ell^\mathsf{T}\Big). \tag{5.121}$$

In the last line we recognize a rank-n (standard) spherical integral, see Section 5.2.2 and [138]. Recall the spectrum is deterministic with empirical law tending weakly to ρ. Hence we can use

the results from Section 5.2.2, with the difference that $\mathbf{C} = I_n \frac{\lambda}{2} D$ is virtually a scalar random variable, and thus w.l.o.g. we can also assume $\mathbf{q}$, and thus $\tilde{\mathbf{q}}$ to be diagonal in (5.22). If we aim for a replica symmetric ansatz

$$\text{Replica Symmetric Ansatz:} \quad \begin{cases} v_\ell = v, & \hat{v}_\ell = \hat{v} \\ q_{0\ell} = m, & \hat{q}_{0\ell} = -\hat{m} \\ q_{\ell\ell'} = q, & \hat{q}_{\ell\ell'} = -\hat{q} \ (\ell \neq \ell') \end{cases} \tag{5.122}$$

then $\mathbf{q}$ has a non degenerate eigenvalue $v + (n-1)q$ and $n-1$ degenerate eigenvalues $v - q$. Within this ansatz we can thus replace the mentioned spherical integral with

$$\mathbb{E}_{\mathbf{O}} \exp\left(\frac{\lambda}{2} \text{Tr} \mathbf{O} D \mathbf{O}^{\mathsf{T}} \sum_{\ell \leq n} \mathbf{x}_\ell \mathbf{x}_\ell^{\mathsf{T}}\right) = \exp N\Big((n-1) I_D(v-q) + I_D(v-q+nq)\Big)$$

$$= \exp Nn\Big(I_D(v-q) + I_D'(v-q)q + O(n)\Big) \tag{5.123}$$

as done in [124], where $I_D(\cdot)$ are rank-one spherical integrals. The rest can be treated exactly as in Section 5.3.2, yielding

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E} \ln \bar{\mathcal{Z}}(\mathbf{Y}) = \text{extr}\Big\{ \frac{\lambda^2}{2} m^2 - \frac{\lambda^2}{4} v^2 + \frac{\hat{v}v}{2} - \hat{m}m + \frac{\hat{q}q}{2} + I_D(v-q)$$

$$+ q I_D'(v-q) + \mathbb{E} \ln \int dP_X(x) \exp\Big(\sqrt{\hat{q}} Z x - \frac{\hat{q}+\hat{v}}{2} x^2 + \hat{m} X_0 x\Big)\Big\} \tag{5.124}$$

where extremization is intended over $m, \hat{m}, q, \hat{q}, v, \hat{v}$. With the same notation for the local measure (5.61), the fixed point equations read

$$(m) \qquad \hat{m} = \lambda^2 m \tag{5.125}$$

$$(\hat{m}) \qquad m = \mathbb{E} X_0 \langle X \rangle_{\hat{m},\hat{q},\hat{v}} \tag{5.126}$$

$$(q) \qquad \hat{q} = 2q I_D''(v-q) \tag{5.127}$$

$$(\hat{q}) \qquad q = \mathbb{E} \langle X \rangle_{\hat{m},\hat{q},\hat{v}}^2 \tag{5.128}$$

$$(v) \qquad \hat{v} = \lambda^2 v - 2 I_D'(v-q) - 2q I_D''(v-q) \tag{5.129}$$

$$(\hat{v}) \qquad v = \mathbb{E} \langle X^2 \rangle_{\hat{m},\hat{q},\hat{v}}. \tag{5.130}$$

The computation above follows the same lines as that in [124], with the only difference being the presence of a planted signal. In case of Gaussian likelihood, the term arising from the spike though is easily tractable, as well as the term containing the fourth norm of the estimator (see (5.115)). This suggests that the AMP algorithm designed in [132], whose aim was to make the results in [124, 123] rigorous, has to match the performance predicted by our replica computation, measured by the MSE

$$\lim_{N\to\infty} \frac{1}{2N^2} \mathbb{E} \|\mathbf{X}^* \mathbf{X}^{*\mathsf{T}} - \bar{\mathbb{E}}[\mathbf{X}^* \mathbf{X}^{*\mathsf{T}} \mid \mathbf{Y}]\|_{\text{F}}^2 = \frac{1}{2}(1 - 2m^2 + q^2). \tag{5.131}$$

in the large N limit, where the $\bar{\mathbb{E}}$ denotes the expectation w.r.t. (5.115), and m and q solve (5.125)–(5.130).

An alternative to (5.125)–(5.130), which turns out to be more practical from the numerical point of view, can be obtained by keeping $\mathbf{q}$ as it is, without diagonalizing it. In this case one needs the entire formula (5.22), with $\tilde{\mathbf{q}}$ having the same RS structure as $\mathbf{q}$, in a similar fashion as that of Section 5.2.2. The spherical integral then takes the form (up to constants)

$$
\mathbb{E}_{\mathbf{O}} \exp\left(\frac{\lambda}{2}\mathrm{Tr}\mathbf{O}\mathbf{D}\mathbf{O}^{\mathsf{T}}\sum_{\ell\leq n}\mathbf{x}_\ell\mathbf{x}_\ell^{\mathsf{T}}\right) \propto \exp\left(Nn\,\mathrm{extr}\left\{\frac{v\tilde{v} - q\tilde{q}}{2} - \frac{1}{2}\mathbb{E}\ln(\tilde{v} - \tilde{q} - \lambda D)\right.\right.
$$
$$
\left.\left. - \frac{\tilde{q}}{2}\mathbb{E}(\tilde{v} - \tilde{q} - \lambda D)^{-1} - \frac{1}{2}\ln(v - q) - \frac{q}{2(v - q)} + O(n)\right\}\right) \tag{5.132}
$$

where extremization is w.r.t. the tilded variables only, for now. Consequently, the free entropy rewrites as follows

$$
\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\ln\bar{\mathcal{Z}}(\mathbf{Y}) = \mathrm{extr}\left\{\frac{\lambda^2}{2}m^2 - \frac{\lambda^2}{4}v^2 + \frac{(\hat{v} + \tilde{v})v}{2} - \hat{m}m + \frac{(\hat{q} - \tilde{q})q}{2}\right.
$$
$$
- \frac{1}{2}\mathbb{E}\ln(\tilde{v} - \tilde{q} - \lambda D) - \frac{\tilde{q}}{2}\mathbb{E}(\tilde{v} - \tilde{q} - \lambda D)^{-1} - \frac{1}{2}\ln(v - q) - \frac{q}{2(v - q)}
$$
$$
\left. + \mathbb{E}\ln\int dP_X(x)\exp\left(\sqrt{\hat{q}}Zx - \frac{\hat{q} + \hat{v}}{2}x^2 + \hat{m}X_0 x\right)\right\}. \tag{5.133}
$$

Here instead, extremization is intended over the tilded and hatted variables, together with m, q, v.

The fixed point equations are

$$(m) \qquad \hat{m} = \lambda^2 m \tag{5.134}$$
$$(\hat{m}) \qquad m = \mathbb{E}X_0\langle X\rangle_{\hat{m},\hat{q},\hat{v}} \tag{5.135}$$
$$(q) \qquad \hat{q} - \tilde{q} = \frac{q}{(v - q)^2} \tag{5.136}$$
$$(\hat{q}) \qquad q = \mathbb{E}\langle X\rangle^2_{\hat{m},\hat{q},\hat{v}} \tag{5.137}$$
$$(\tilde{q}) \qquad q = -\tilde{q}\mathbb{E}(\tilde{v} - \tilde{q} - \lambda D)^{-2} \tag{5.138}$$
$$(v) \qquad \hat{v} + \tilde{v} - \lambda^2 v - \frac{1}{v - q} + \frac{q}{(v - q)^2} = 0 \tag{5.139}$$
$$(\hat{v}) \qquad v = \mathbb{E}\langle X^2\rangle_{\hat{m},\hat{q},\hat{v}} \tag{5.140}$$
$$(\tilde{v}) \qquad v - \mathbb{E}(\tilde{v} - \tilde{q} - \lambda D)^{-1} + \tilde{q}\mathbb{E}(\tilde{v} - \tilde{q} - \lambda D)^{-2} = 0. \tag{5.141}$$

Plugging $(\tilde{q})$ into $(\tilde{v})$ we readily see that

$$
v - q = \mathbb{E}(\tilde{V} - \lambda D)^{-1} \tag{5.142}
$$

that works as an equation for $\tilde{V} := \tilde{v} - \tilde{q}$ as a function of v, q. Analogously, we can plug (q) into (v) obtaining

$$\hat{v} + \hat{q} = \lambda^2 v + \frac{1}{v - q} - \tilde{V} \tag{5.143}$$

that determines $\hat{v} + \hat{q}$ as a function of v and q, thanks to the above equation for $\tilde{V}$. Finally, from $(\tilde{q})$ and (q) we have respectively

$$\tilde{q} = -\frac{q}{\mathbb{E}(\tilde{V} - \lambda D)^{-2}} \tag{5.144}$$

$$\hat{q} = \frac{q}{(v - q)^2} + \tilde{q}. \tag{5.145}$$

Notice that, being in a mismatched setting, there cannot be any simplifications due to the Nishimori identities.

It is not difficult to verify a posteriori that the systems (5.125)–(5.130) and (5.134)–(5.141) are equivalent. The extremization over the tilded variables has indeed the purpose of reproducing I_D and its derivatives. From (5.142) one can infer

$$\tilde{V} = R_{\lambda \mathbf{D}}(v - q) + \frac{1}{v - q} \tag{5.146}$$

where $R_{\lambda \mathbf{D}}$ denotes the R-transform of $\lambda \mathbf{D}$, and deriving both sides w.r.t. v one also has

$$\tilde{V}' = -\frac{1}{\mathbb{E}(\tilde{V} - \lambda D)^{-2}} = R'_{\lambda \mathbf{D}}(v - q) - \frac{1}{(v - q)^2}. \tag{5.147}$$

Therefore, from (5.144)

$$\tilde{q} = q R'_{\lambda \mathbf{D}}(v - q) - \frac{q}{(v - q)^2} \quad \Rightarrow \quad \hat{q} = q R'_{\lambda \mathbf{D}}(v - q), \tag{5.148}$$

and from (5.143)

$$\hat{v} + \hat{q} = \lambda^2 v - R_{\lambda \mathbf{D}}(v - q), \tag{5.149}$$

both in perfect agreement with (5.127) and (5.129), as long as $R_{\lambda \mathbf{D}} = 2I'_D$ [138].

The system of fixed point equations (5.134)–(5.141) can be solved numerically as follows: *(i)* initialize $m = m_0$, $q = q_0$, $v = v_0$ (the latter being identically 1 if we use a Rademacher prior); *(ii)* solve (5.142) for $\tilde{V}$; *(iii)* compute $\hat{q}$, $\tilde{q}$, $\hat{m}$ and $\hat{v} + \hat{q}$ from (5.145), (5.144), (5.134) and (5.143) respectively; *(iv)* update the values of m, q, v through $(\hat{m})$, $(\hat{q})$ and $(\hat{v})$ obtaining m^1, q^1 and v^1; *(v)* repeat the steps *(i)–(iv)* starting from $m = m^1$, $q = q^1$ and $v = v^1$, thus obtaining m^2, q^2 and v^2, and so forth.
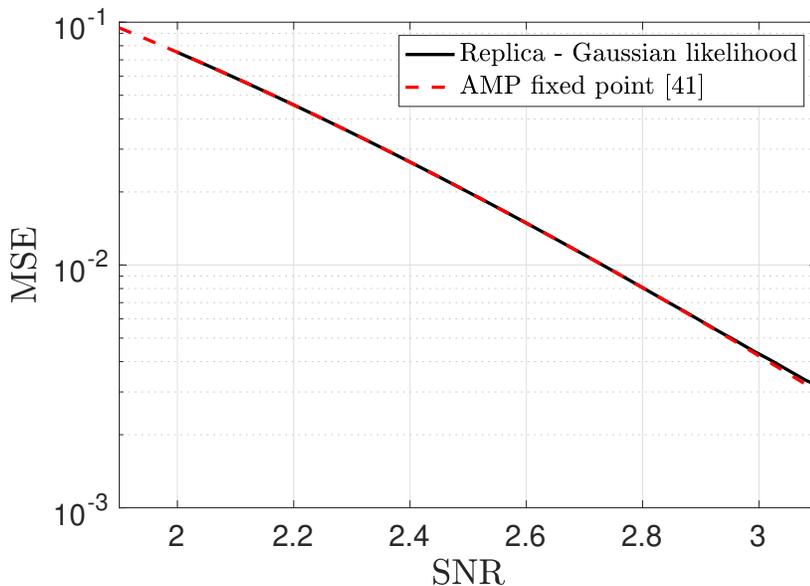
Figure 5.2: Comparison between the fixed point of the AMP algorithm in [132] and that obtained via the replica computation (cf. (5.134)–(5.141)), for i.i.d. Rademacher distributed $(X_i^*)_i$. The agreement between these two fixed points is excellent when the SNR is between 2 and 3.

The numerics arising from this procedure though turns out to be delicate for extreme values of the overlap, namely when $v-q$ is really small, which in turn happens when λ is large (typically > 3 for Rademacher prior). The equation that seems to generate numerical instability is (5.145), and in particular the two contributions there appearing. With reference to the Rademacher prior, and the related Figure 5.2, when $\lambda > 3$ the overlap gets close to ~ 0.999. At this value $1/(v-q)^2 \sim 10^6$. $\tilde{q}$, that is also contributing to (5.145), on the contrary becomes really negative, and is such that $\hat{q}$ is typically ~ 10 near $\lambda \sim 3$. The subtraction of these two big numbers apparently dooms the iterations for larger SNRs. This was not the case in the Bayes-optimal setting, thanks to the simplifications introduced by the Nishimori identities. Indeed, from (5.75), (5.76), (5.77) and (5.78) we see that $1 - m$ appears at most at the first power in denominators. The only issue there was that $\tilde{V}$ can grow exponentially fast, and this can be solved by allowing for a wide range of search of the solution of (5.77).

The fixed point of the MSE arising from (5.134)–(5.141) is compared with the fixed point (5.92), which corresponds to the MSE of the AMP proposed in [132]. The match between these two computations is excellent, as long as the SNR is not too large, because of the aforementioned numerical issues in iterating (5.134)–(5.141). The plot of Figure 5.2 is a compelling numerical confirmation of the arguments put forward in this section. The conclusion is the following: the AMP algorithm of [132] is solving a *replica symmetric approximation* to the TAP equations

associated with the mismatched posterior distribution (5.115).

5.5 Towards an optimal AMP: AdaTAP formalism

We have previously shown that the AMP found in the literature for structured PCA [132] is sub-optimal. In this section we understand the fundamental reason behind this issue by generalizing the Adaptive Thouless-Anderson-Palmer (AdaTAP) formalism of [124, 123]. Using our new insights we will then be able in the next section to cure the issue and derive a Bayes-optimal AMP. Like in the replica method and in particular Section 5.3.1, a key ingredient will be to reduce the model to a quadratic one of the Ising type.

5.5.1 The AdaTAP single-instance free entropy

Recall that the posterior distribution is given by (5.6). Denoting $\mathbf{p} := \mathbf{x}\mathbf{x}^\mathsf{T}/N$ and $v := \|\mathbf{x}\|^2/N$ the trace of the matrix potential (5.13) can be expanded as follows:

$$\mathrm{Tr}V(\mathbf{Y} - \lambda\mathbf{p}) = C + \frac{\mu}{2}\mathrm{Tr}\{\lambda^2 v^2 - 2\lambda\mathbf{Y}\mathbf{p}\}$$
$$+ \frac{\gamma}{4}\mathrm{Tr}\{\lambda^4 v^4 - 4\lambda^3 v^2\mathbf{Y}\mathbf{p} + 4\lambda^2 v\mathbf{Y}^2\mathbf{p} - 4\lambda\mathbf{Y}^3\mathbf{p} + 2\lambda^2\mathbf{Y}\mathbf{p}\mathbf{Y}\mathbf{p}\}$$

where C is independent of $\mathbf{x}$. Define the matrix polynomial:

$$\mathbf{R}(v, \mathbf{Y}) := -(\mu\lambda + \gamma\lambda^3 v^2)\mathbf{Y} + \gamma\lambda^2 v\mathbf{Y}^2 - \gamma\lambda\mathbf{Y}^3. \tag{5.150}$$

Then

$$-\frac{N}{2}\mathrm{Tr}V(\mathbf{Y} - \lambda\mathbf{p}) \propto -\frac{N}{4}\lambda^2 v^2\left(\mu + \frac{\gamma\lambda^2 v^2}{2}\right) - \frac{\mathbf{x}^\mathsf{T}\mathbf{R}(v, \mathbf{Y})\mathbf{x}}{2} - \frac{N}{4}\gamma\lambda^2\left(\frac{\mathbf{x}^\mathsf{T}\mathbf{Y}\mathbf{x}}{N}\right)^2. \tag{5.151}$$

The partition function of the model defined by (5.12) can then be written in the form

$$\mathcal{Z} \propto \int dP_X(\mathbf{x})dvdf\delta(Nv - \|\mathbf{x}\|^2)\delta(Nf - \mathbf{x}^\mathsf{T}\mathbf{Y}\mathbf{x})$$
$$\times \exp\left(-\frac{N}{4}\lambda^2 v^2\left(\mu + \frac{\gamma\lambda^2 v^2}{2}\right) - \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{R}(v, \mathbf{Y})\mathbf{x} - \frac{N}{4}\gamma\lambda^2 f^2\right)$$
$$= \int dvd\hat{v}dfd\hat{f}\exp\left(N\hat{v}v + N\hat{f}f - \frac{N}{4}\lambda^2 v^2\left(\mu + \frac{\gamma\lambda^2 v^2}{2}\right) - \frac{N}{4}\gamma\lambda^2 f^2\right)$$
$$\times \int dP_X(\mathbf{x})\exp\left(-\hat{v}\|\mathbf{x}\|^2 - \hat{f}\mathbf{x}^\mathsf{T}\mathbf{Y}\mathbf{x} - \frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{R}(v, \mathbf{Y})\mathbf{x}\right)$$
$$= \int dvd\hat{v}dfd\hat{f}\exp\left(N\hat{v}v + N\hat{f}f - \frac{N}{4}\lambda^2 v^2\left(\mu + \frac{\gamma\lambda^2 v^2}{2}\right) - \frac{N}{4}\gamma\lambda^2 f^2\right)$$
$$\times \int dP_X(\mathbf{x})\exp\left(\frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{J}(v, \hat{v}, \hat{f}, \mathbf{Y})\mathbf{x}\right), \tag{5.152}$$

where the overall symmetric interaction matrix of this "Ising model" is

$$\mathbf{J}(v, \hat{v}, \hat{f}, \mathbf{Y}) := -\mathbf{R}(v, \mathbf{Y}) - 2\hat{v}\boldsymbol{I}_N - 2\hat{f}\mathbf{Y}. \tag{5.153}$$

Now, defining the free entropy at fixed $(v, \hat{v}, \hat{f})$

$$\Phi_N(v, \hat{v}, \hat{f}, \mathbf{Y}) := \ln \int dP_X(\mathbf{x}) \exp\left(\frac{1}{2}\mathbf{x}^{\mathsf{T}}\mathbf{J}(v, \hat{v}, \hat{f}, \mathbf{Y})\mathbf{x}\right), \tag{5.154}$$

because the prior is factorized and we have an Ising-type of model, we can directly use the AdaTAP result [124]: it tells us that

$$\Phi_N(v, \hat{v}, \hat{f}, \mathbf{Y}) = -\mathrm{extr}_{\mathbf{m}, \boldsymbol{\tau}, \mathbf{V}}\Big\{\frac{1}{2}\mathbf{m}^{\mathsf{T}}\mathbf{J}(v, \hat{v}, \hat{f}, \mathbf{Y})\mathbf{m}$$
$$+ \frac{1}{2}\ln\det\left(\boldsymbol{\Omega} - \mathbf{J}(v, \hat{v}, \hat{f}, \mathbf{Y})\right) - \frac{1}{2}\mathbf{V}^{\mathsf{T}}\mathbf{m}^2 + \frac{1}{2}\sum_{i \leq N}\ln(\tau_i - m_i^2)$$
$$- \sum_{i \leq N}\ln\int dP_X(x)\exp\left(\frac{1}{2}V_i x^2 + \left((\mathbf{J}(v, \hat{v}, \hat{f}, \mathbf{Y})\mathbf{m})_i - V_i m_i\right)x\right)\Big\} + o_N(1). \tag{5.155}$$

The extremization is over $(\mathbf{m}, \boldsymbol{\tau}, \mathbf{V}) \in \mathbb{R}^N \times (\mathbb{R}_{\geq 0}^N)^2$, $\mathbf{m}^2 = (m_i^2)_{i \leq N}$, and the diagonal matrix

$$\boldsymbol{\Omega} := \mathrm{diag}(\mathbf{V} + (\boldsymbol{\tau} - \mathbf{m}^2)^{-1}). \tag{5.156}$$

Let the bracket notation $\langle \cdot \rangle$ be used as expectation with respect to the posterior (5.6), while $\langle \cdot \rangle_{\backslash i}$ is the mean with respect to the Gibbs measure of the "cavity graph" where $(J_{ij})_j$ are set to 0. Define also the cavity fields

$$h_i := (\mathbf{Jx})_i.$$

The various variables at their extremum values are (asymptotically exact approximations to) the marginals means, second moments and variances of the cavity fields

$$m_i = \langle x_i \rangle, \quad \tau_i = \langle x_i^2 \rangle, \quad V_i = \langle h_i^2 \rangle_{\backslash i} - \langle h_i \rangle_{\backslash i}^2.$$

From the AdaTAP free entropy at fixed $(v, \hat{v}, \hat{f})$ we can compute the total log-partition function by saddle-point and get

$$\frac{1}{N}\ln\mathcal{Z}(\mathbf{Y}) \propto o_N(1)$$
$$+ \mathrm{extr}\Big\{\hat{v}v + \hat{f}f - \frac{1}{4}\lambda^2 v^2\Big(\mu + \frac{\gamma\lambda^2 v^2}{2}\Big) - \frac{1}{4}\gamma\lambda^2 f^2 + \Phi_N(v, \hat{v}, \hat{f}, \mathbf{Y})\Big\} \tag{5.157}$$

where the extremization is over $(v, \hat{v}, f, \hat{f})$.

5.5.2 Saddle point: reduction to an Ising model, AdaTAP equations and optimal pre-processing of the data

By extremization of the AdaTAP single-instance free entropy (5.157) we derive the AdaTAP equations. We start with the intensive parameters. The extremization with respect to f is trivial and gives

$$\hat{f} = \frac{1}{2}\gamma\lambda^2 f.$$

So the leading order of the AdaTAP free entropy simplifies to

$$\text{extr}_{v,\hat{v},f}\left\{\hat{v}v + \frac{1}{4}\gamma\lambda^2 f^2 - \frac{1}{4}\lambda^2 v^2\left(\mu + \frac{\gamma\lambda^2 v^2}{2}\right) + \Phi_N\left(v, \hat{v}, \frac{1}{2}\gamma\lambda^2 f, \mathbf{Y}\right)\right\}. \tag{5.158}$$

The remaining saddle point equations can simply be written down. But this is not necessary as the solution of the three remaining intensive order parameters at the saddle point is simply deduced from their physical meaning, concentration properties, and the Nishimori identity: in the large size limit,

$$v \to \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\langle\|\mathbf{x}\|^2\rangle = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\|\mathbf{X}^*\|^2 = 1,$$

as well as (recall (5.66))

$$f \to \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\langle\mathbf{x}^\mathsf{T}\mathbf{Y}\mathbf{x}\rangle = \lambda.$$

Moreover we know that

$$\hat{v} \to 0$$

because the prior is already enforcing the constraint that $v = \|\mathbf{x}\|^2/N \to 1$ in (5.152) without the need to introducing a further, redundant, delta constraint; note that for Rademacher or spherical prior this is simply true as no delta function is needed. Therefore the AdaTAP free entropy becomes

$$\frac{1}{8}\gamma\lambda^4 - \frac{1}{4}\mu\lambda^2 + \Phi_N\left(1, 0, \frac{1}{2}\gamma\lambda^3, \mathbf{Y}\right) + o_N(1). \tag{5.159}$$

From this AdaTAP free entropy we see that the values of the marginal means and variances correspond to the solution of the variational problem (5.155) with interaction matrix

$$\mathbf{J}\left(1, 0, \frac{1}{2}\gamma\lambda^3, \mathbf{Y}\right) = \mu\lambda\mathbf{Y} - \gamma\lambda^2\mathbf{Y}^2 + \gamma\lambda\mathbf{Y}^3 =: \mathbf{J}(\mathbf{Y}). \tag{5.160}$$

So we end-up with the following effective partition function of an Ising-like model:

$$\int dP_X(\mathbf{x})\exp\left(\frac{1}{2}\mathbf{x}^\mathsf{T}\mathbf{J}(\mathbf{Y})\mathbf{x}\right). \tag{5.161}$$

This shows that the original model is equivalent to an Ising model with interaction matrix $\mathbf{J}(\mathbf{Y})$, which can thus be interpreted as a *Bayes-optimal pre-processing of the data*. This will be verified

in Section 5.6, as the use of $\mathbf{J}(\mathbf{Y})$ instead of $\mathbf{Y}$ will turn AMP into an optimal algorithm. Ising models like this are precisely studied in [124] and we can therefore again exploit directly the AdaTAP formalism. Let

$$\eta_i(\mathbf{J}, \mathbf{m}, V_i) := \frac{\int dP_X(x)\, x\, e^{\frac{1}{2}V_i x^2 + ((\mathbf{Jm})_i - V_i m_i)x}}{\int dP_X(x) e^{\frac{1}{2}V_i x^2 + ((\mathbf{Jm})_i - V_i m_i)x}}, \tag{5.162}$$

$$g_i(\mathbf{J}, \mathbf{m}, V_i) := \frac{\int dP_X(x)\, x^2\, e^{\frac{1}{2}V_i x^2 + ((\mathbf{Jm})_i - V_i m_i)x}}{\int dP_X(x) e^{\frac{1}{2}V_i x^2 + ((\mathbf{Jm})_i - V_i m_i)x}}. \tag{5.163}$$

The associated AdaTAP equations over $(\mathbf{m}, \boldsymbol{\tau}, \mathbf{V})$, namely the saddle point equations associated with the AdaTAP free entropy (5.155) with $\mathbf{J}(v, \hat{v}, \hat{f}, \mathbf{Y})$ replaced by $\mathbf{J} = \mathbf{J}(\mathbf{Y})$, read

$$m_i = \eta_i(\mathbf{J}, \mathbf{m}, V_i), \tag{5.164}$$

$$\tau_i = g_i(\mathbf{J}, \mathbf{m}, V_i), \tag{5.165}$$

$$\tau_i - m_i^2 = \left([\operatorname{diag}(\mathbf{V} + (\boldsymbol{\tau} - \mathbf{m}^2)^{-1}) - \mathbf{J}]^{-1}\right)_{ii}, \tag{5.166}$$

where the last equation is understood as an implicit equation for $\mathbf{V}$.

5.5.3 Simplifying the AdaTAP equations by self-averaging of the Onsager reaction term

The variances V_i are expected to be self-averaging with respect to the interaction matrix, i.e., in the large size limit $V_i = \bar{V} := \lim_{N \to \infty} \mathbb{E}_{\mathbf{J}} V_i$. The computation we are going to carry out now could be performed in various ways leading to different but equivalent expressions. For pedagogical reasons we take a path that remains as close as possible to the approach of [124]. Following this reference we compute the expectation of the AdaTAP equation for $\mathbf{V}$. In this section, all quantities $\mathbf{V}$, $\mathbf{m}$ and $\boldsymbol{\tau}$ are fixed to a solution of the AdaTAP equations (5.164)–(5.166).

We start from the convenient identity

$$\left([\boldsymbol{\Omega} - \mathbf{J}]^{-1}\right)_{ii} = \partial_{\Omega_{ii}} \ln \det(\boldsymbol{\Omega} - \mathbf{J}). \tag{5.167}$$

We are going to average the right-hand side. As for a Gaussian model there is no spin glass phase and strong concentrations take place, the quenched and annealed averages match [124]: we can thus simply compute the logarithm of the average of the determinant. A Gaussian identity then gives

$$\mathbb{E} \det(\boldsymbol{\Omega} - \mathbf{J})^{-1/2} = \int \frac{d\mathbf{z}}{(2\pi)^{N/2}} \exp\left(-\frac{1}{2}\mathbf{z}^\mathsf{T}\boldsymbol{\Omega}\mathbf{z}\right) \mathbb{E} \exp\left(\frac{1}{2}\mathbf{z}^\mathsf{T}\mathbf{J}\mathbf{z}\right). \tag{5.168}$$

We denote $\mathbf{J} = \sum_{k \le 3} c_k \mathbf{Y}^k$ where $\mathbf{c} = (\mu\lambda, -\gamma\lambda^2, \gamma\lambda)$. The term we need to compute therefore reads

$$\mathbb{E} \exp\left(\frac{1}{2}\mathbf{z}^\mathsf{T}\mathbf{J}\mathbf{z}\right) = \mathbb{E} \exp\frac{1}{2}(\mathbf{z}^\mathsf{T}(c_1 \mathbf{Y} + c_2 \mathbf{Y}^2 + c_3 \mathbf{Y}^3)\mathbf{z}) \tag{5.169}$$

Define the order parameters

$$p := \frac{1}{N}\mathbf{z}^\mathsf{T}\mathbf{X}^*, \quad v := \frac{1}{N}\|\mathbf{z}\|^2, \quad p_D := \frac{1}{N}(\mathbf{Oz})^\mathsf{T}\mathbf{DOX}^*. \tag{5.170}$$

We also have $\|\mathbf{X}^*\|^2/N = 1 + o_N(1)$. Our goal is to identify the generalized spherical integral (5.20). Replacing $\mathbf{Y}$ by $\lambda\mathbf{p}^* + \mathbf{O}^\mathsf{T}\mathbf{DO}$ (with $\mathbf{p}^* := \mathbf{X}^*\mathbf{X}^{*\mathsf{T}}/N$) we expand the various terms. The first term is then simply

$$c_1\mathbf{z}^\mathsf{T}\big(\lambda\mathbf{p}^* + \mathbf{O}^\mathsf{T}\mathbf{DO}\big)\mathbf{z} = c_1\big(\lambda Np^2 + (\mathbf{Oz})^\mathsf{T}\mathbf{DOz}\big). \tag{5.171}$$

The second term is

$$\begin{aligned}
c_2\mathbf{z}^\mathsf{T}\big(\lambda^2(\|\mathbf{X}^*\|^2/N)\mathbf{p}^* + \lambda\mathbf{p}^*\mathbf{O}^\mathsf{T}\mathbf{DO} + \lambda\mathbf{O}^\mathsf{T}\mathbf{DOp}^* + \mathbf{O}^\mathsf{T}\mathbf{D}^2\mathbf{O}\big)\mathbf{z} \\
= c_2\big(N\lambda^2p^2 + 2N\lambda pp_D + (\mathbf{Oz})^\mathsf{T}\mathbf{D}^2\mathbf{Oz}\big) + o(N).
\end{aligned} \tag{5.172}$$

Finally the last term is a bit more cumbersome:

$$\begin{aligned}
c_3\mathbf{z}^\mathsf{T}\big(\lambda^3(\|\mathbf{X}^*\|^4/N^2)\mathbf{p}^* + \lambda^2\mathbf{p}^*\mathbf{O}^\mathsf{T}\mathbf{DOp}^* + \lambda^2(\|\mathbf{X}^*\|^2/N)\mathbf{O}^\mathsf{T}\mathbf{DOp}^* + \lambda\mathbf{O}^\mathsf{T}\mathbf{D}^2\mathbf{Op}^* \\
+ \lambda^2(\|\mathbf{X}^*\|^2/N)\mathbf{p}^*\mathbf{O}^\mathsf{T}\mathbf{DO} + \lambda\mathbf{p}^*\mathbf{O}^\mathsf{T}\mathbf{D}^2\mathbf{O} + \lambda\mathbf{O}^\mathsf{T}\mathbf{DOp}^*\mathbf{O}^\mathsf{T}\mathbf{DO} + \mathbf{O}^\mathsf{T}\mathbf{D}^3\mathbf{O}\big)\mathbf{z} \\
= c_3\big(N\lambda^3p^2 + \lambda^2p^2(\mathbf{OX}^*)^\mathsf{T}\mathbf{DOX}^* + 2\lambda p(\mathbf{Oz})^\mathsf{T}\mathbf{D}^2\mathbf{OX}^* \\
+ 2N\lambda^2pp_D + \lambda Np_D^2 + (\mathbf{Oz})^\mathsf{T}\mathbf{D}^3\mathbf{Oz}\big) + o(N).
\end{aligned} \tag{5.173}$$

Combining all we reach

$$\begin{aligned}
\mathbb{E}\exp\Big(\frac{1}{2}\mathbf{z}^\mathsf{T}\mathbf{Jz}\Big) = \int d\boldsymbol{\tau}d\hat{\boldsymbol{\tau}}\,\exp\Big(NK + \frac{1}{2}\hat{v}\|\mathbf{z}\|^2 - \frac{N}{2}\hat{v}v + o(N)\Big) \\
\times \mathbb{E}_\mathbf{O}\exp\Big((\mathbf{Oz})^\mathsf{T}\mathbf{C}_{\mathbf{z},\mathbf{z}}\mathbf{Oz} + (\mathbf{OX}^*)^\mathsf{T}\mathbf{C}_{*,*}\mathbf{OX}^* + (\mathbf{OX}^*)^\mathsf{T}\mathbf{C}_{\mathbf{z},*}\mathbf{Oz}\Big)
\end{aligned} \tag{5.174}$$

with $d\boldsymbol{\tau} := (dp, dv, dp_D)$ and $d\hat{\boldsymbol{\tau}} := (d\hat{p}, d\hat{v}, d\hat{p}_D)$, and (all coupling matrices below are $N \times N$ and symmetric)

$$K := \frac{1}{2}\big(\mu\lambda^2p^2 - \gamma\lambda^2(\lambda^2p^2 + 2\lambda pp_D) + \gamma\lambda(\lambda^3p^2 + 2\lambda^2pp_D + \lambda p_D^2) + \hat{p}p + \hat{p}_Dp_D\big),$$

$$\mathbf{C}_{*,*} := \frac{1}{2}\gamma\lambda^3p^2\mathbf{D},$$

$$\mathbf{C}_{\mathbf{z},\mathbf{z}} := \frac{1}{2}\big(\mu\lambda\mathbf{D} - \gamma\lambda^2\mathbf{D}^2 + \gamma\lambda\mathbf{D}^3\big),$$

$$\mathbf{C}_{\mathbf{z},*} := \frac{1}{2}\big(-\hat{p}I_N - \hat{p}_D\mathbf{D} + 2\gamma\lambda^2p\mathbf{D}^2\big).$$

Note the asymmetry for the variable $\hat{v}$ compared to the other hat-variables, which has not been injected in the definition of the coupling matrices as the others, but instead leads to a term

appearing explicitly in (5.174) (both choices are equivalently valid ones). The term averaged over $\mathbf{O}$ is an inhomogeneous spherical integral as studied in Section 5.2. In particular, we are in the case of Section 5.2.2 with $\ell \in \{0, 1\}$ with the exception that $\mathbf{X}^*$ also (playing the role of the 0th replica) has a non-zero self-coupling. So this trivial modification of the computation of Section 5.2.2 yields

$$\mathbb{E}e^{\frac{1}{2}\mathbf{z}^{\mathsf{T}}\mathbf{J}\mathbf{z}} = \int d\boldsymbol{\tau}d\hat{\boldsymbol{\tau}} \, \exp\left(NK + \frac{1}{2}\hat{v}\|\mathbf{z}\|^2 - \frac{N}{2}\hat{v}v + NI_{\mathbf{C}}(p, v, \hat{p}, \hat{p}_D) + o(N)\right)$$

where the 2×2 random coupling matrix $\mathbf{C}$ has entries

$$2C_{00} = \gamma\lambda^3 p^2 D, \tag{5.175}$$
$$2C_{11} = \mu\lambda D - \gamma\lambda^2 D^2 + \gamma\lambda D^3, \tag{5.176}$$
$$2C_{01} = 2C_{10} = \frac{1}{2}(-\hat{p} - \hat{p}_D D + 2\gamma\lambda^2 pD^2), \tag{5.177}$$

with $D \sim \rho$ drawn from the noise asymptotic spectral density, and

$$I_{\mathbf{C}}(p, v, \hat{p}, \hat{p}_D) = \frac{1}{2}\mathrm{extr}_{(\tilde{v}_0, \tilde{v}, \tilde{p})}\Big\{\tilde{v}_0 + 2\tilde{p}p + \tilde{v}v$$
$$- \mathbb{E}\ln\left((\tilde{v}_0 - 2C_{00})(\tilde{v} - 2C_{11}) - (\tilde{p} - 2C_{01})^2\right)\Big\} - \frac{1}{2}\ln(v - p^2) - 1. \tag{5.178}$$

One can check that $I_{\mathbf{C}}$ is null when $C_{00} = C_{11} = C_{01}$ as it should. Therefore equation (5.168) becomes at leading exponential order

$$\ln\mathbb{E}\det(\boldsymbol{\Omega} - \mathbf{J})^{-1/2}$$
$$= \ln\int\frac{d\mathbf{z}}{(2\pi)^{N/2}}d\boldsymbol{\tau}d\hat{\boldsymbol{\tau}}\exp\left(-\frac{1}{2}\mathbf{z}^{\mathsf{T}}(\boldsymbol{\Omega} - \hat{v}I_N)\mathbf{z} + NK - \frac{N}{2}\hat{v}v + NI_{\mathbf{C}} + o(N)\right)$$
$$= \ln\int d\boldsymbol{\tau}d\hat{\boldsymbol{\tau}}\exp\left(NK - \frac{N}{2}\hat{v}v + NI_{\mathbf{C}} - \frac{1}{2}\ln\det(\boldsymbol{\Omega} - \hat{v}I_N) + o(N)\right)$$
$$= \mathrm{extr}\left\{NK - \frac{N}{2}\hat{v}v + NI_{\mathbf{C}} - \frac{1}{2}\ln\det(\boldsymbol{\Omega} - \hat{v}I_N)\right\} + o(N),$$

where we used Gaussian integration followed by a saddle point estimation. By the aforementioned strong concentration properties of the Gaussian model, this is also equal to $-\frac{1}{2}\ln\mathbb{E}\det(\boldsymbol{\Omega} - \mathbf{J}) \approx -\frac{1}{2}\mathbb{E}\ln\det(\boldsymbol{\Omega} - \mathbf{J})$ so we reach at leading order

$$\mathbb{E}\ln\det(\boldsymbol{\Omega} - \mathbf{J}) \approx \mathrm{extr}\left\{-2NK + N\hat{v}v - 2NI_{\mathbf{C}} + \ln\det(\boldsymbol{\Omega} - \hat{v}I_N)\right\}$$
$$= \mathrm{extr}_{(\hat{v}, v)}\left\{N\hat{v}v + \sum_{i \leq N}\ln(\Omega_{ii} - \hat{v}) - 2N\tilde{G}(v)\right\} \tag{5.179}$$

where the extremization is over all variables and

$$\tilde{G}(v) := \mathrm{extr}_{(p, p_D, \hat{p}, \hat{p}_D)}\left\{I_{\mathbf{C}}(p, v, \hat{p}, \hat{p}_D) + K(p, p_D, \hat{p}, \hat{p}_D)\right\}. \tag{5.180}$$

This is the analogue of the G-function appearing, e.g., in [124]. The extremization over $\hat{v}$ in (5.179) yields that at the saddle point,

$$v = \frac{1}{N} \sum_{i \leq N} \frac{1}{\Omega_{ii} - \hat{v}}.$$

Moreover, combining the TAP equation (5.166) with (5.167) and (5.179) we have

$$\mathbb{E}(\tau_i - m_i^2) = \partial_{\Omega_{ii}} \mathbb{E} \ln \det(\boldsymbol{\Omega} - \mathbf{J}) = \frac{1}{\Omega_{ii} - \hat{v}} \tag{5.181}$$

where $\hat{v}$ is evaluated at its saddle point value. Therefore, summing over i the last identity and recalling the definition of Ω_{ii} we reach

$$\bar{\chi} := \frac{1}{N} \mathbb{E} \sum_{i \leq N} (\tau_i - m_i^2) = v = \frac{1}{N} \mathbb{E} \sum_{i \leq N} \frac{1}{V_i + (\tau_i - m_i^2)^{-1} - \hat{v}}. \tag{5.182}$$

Under the concentration assumption $V_i = \bar{V}$ for all $i \leq N$, this identity implies

$$V_i = \hat{v}. \tag{5.183}$$

Additionally the saddle point equation for v extracted from (5.179) yields

$$\hat{v} = 2\partial_v \tilde{G}(v)|_{v=\bar{\chi}} \quad \Rightarrow \quad V_i = \bar{V} := 2\partial_v \tilde{G}(v)|_{v=\bar{\chi}}. \tag{5.184}$$

The variable $\bar{\chi}$ is instance-independent and can be deduced from our replica theory: it is equal to twice the MMSE (5.79), namely,

$$\bar{\chi} = 1 - m^2 \tag{5.185}$$

where m is solution to the replica fixed point equations (5.73)–(5.78). Computing $\bar{V}$ from (5.184) is then easy, as taking a derivative w.r.t. v of $\tilde{G}(v)$ is straightforward: all the quantities appearing on the right-hand side of (5.180) are at the saddle point, so it simply amounts to a partial derivative of (5.178). It gives

$$\bar{V} = \tilde{v} - \frac{1}{1 - m^2 - p^2} \tag{5.186}$$

where $\tilde{v} = \tilde{v}(p, v)$ takes its saddle point value from (5.178) while $p = p(v)$ from (5.180) with $v = \bar{\chi}$ fixed.

Thanks to these simplifications the AdaTAP equation reads in the large size limit

$$m_i = \eta_i(\mathbf{J}, \mathbf{m}, \bar{V}). \tag{5.187}$$

Or, when written in a fashion closer to the form of AMP algorithms, the AdaTAP equations read

$$\mathbf{f} = \mathbf{J}\mathbf{m} - \bar{V}\mathbf{m}, \qquad \mathbf{m} = \eta_{\bar{V}}(\mathbf{f}), \tag{5.188}$$

where the "denoiser", which is applied component-wise above, is

$$\eta_{\bar{V}}(f) := \frac{\int dP_X(x)\, x\, e^{\frac{1}{2}\bar{V}x^2 + fx}}{\int dP_X(x) e^{\frac{1}{2}\bar{V}x^2 + fx}}. \tag{5.189}$$

5.6   Approximate message passing, optimally

We will now describe an AMP algorithm that matches the replica prediction for the minimum mean-square error. We therefore conjecture it to be Bayes-optimal. The main difference between this new AMP and the previously proposed one for structured PCA is that it is constructed from iterates based on the pre-processed matrix $\mathbf{J}(\mathbf{Y})$ rather than $\mathbf{Y}$ as in [132]. Consequently, the Onsager reaction terms will have to be adapted.

5.6.1   BAMP: Bayes-optimal AMP

The AdaTAP approach described in Section 5.5 suggests that, in order to achieve Bayes-optimal performance, one should consider the BAMP iteration which is of the form

$$\mathbf{f}^t = \mathbf{J}(\mathbf{Y})\mathbf{u}^t - \sum_{i=1}^{t} \mathsf{c}_{t,i}\mathbf{u}^i, \quad \mathbf{u}^{t+1} = g_{t+1}(\mathbf{f}^t), \quad t \geq 1. \tag{5.190}$$

As in the AMP iteration (5.89), the *denoiser* function $g_{t+1} : \mathbb{R} \to \mathbb{R}$ is continuously differentiable, Lipschitz and applied component-wise. Crucially, the Onsager coefficients $\{\mathsf{c}_{t,i}\}_{i\in[t],t\geq 1}$ need to ensure that, conditioned on the signal, the empirical distribution of the iterate $\mathbf{f}^t$ is Gaussian, namely, the convergence result in (5.90) holds for some mean vector $\boldsymbol{\mu}_t$ and covariance matrix $\boldsymbol{\Sigma}_t$.

We highlight that the matrix $\mathbf{Y}$ in (5.89) is replaced by the matrix $\mathbf{J}(\mathbf{Y})$ in (5.190). This means that the state evolution result of [132] cannot be applied and the Onsager coefficients $\{\mathsf{c}_{t,i}\}_{i\in[t],t\geq 1}$ will have a different form with respect to $\{\mathsf{b}_{t,i}\}_{i\in[t],t\geq 1}$.

In what follows, we will consider the general case in which $\mathbf{J}(\mathbf{Y})$ is an arbitrary polynomial of degree K in $\mathbf{Y}$, namely,

$$\mathbf{J}(\mathbf{Y}) = \sum_{i \leq K} c_i \mathbf{Y}^i.$$

To compute $\{\mathsf{c}_{t,i}\}_{i\in[t],t\geq 1}$ and obtain a state evolution result for the iteration (5.190), the key idea is to map the first T iterations of (5.190) to the first $K \times T$ iterations of an *auxiliary* AMP with iterates $(\tilde{\mathbf{z}}^t, \tilde{\mathbf{u}}^t)_{t\in[KT]}$ and denoisers $\{\tilde{h}_{t+1}\}_{t\in[KT]}$, whose state evolution can be deduced

from [132]. The denoisers $\{\tilde{h}_{t+1}\}_{t \in [KT]}$ of this auxiliary AMP are chosen so that, for $t \in [T]$ and $\ell \in [K]$,

$$\lim_{N \to \infty} \frac{1}{N} \|\tilde{\mathbf{u}}^{K(t-1)+\ell} - \mathbf{Y}^{\ell-1} \mathbf{u}^t\|_2^2 = 0. \tag{5.191}$$

More specifically, for $t \in [T]$ and $\ell \in \{2, \dots, K\}$, the denoiser $\tilde{h}_{K(t-1)+\ell}$ giving $\tilde{\mathbf{u}}^{K(t-1)+\ell}$ is a linear combinations of the past iterates $\tilde{\mathbf{u}}^1, \dots, \tilde{\mathbf{u}}^{K(t-1)+\ell-1}$ and of $\tilde{\mathbf{z}}^{K(t-1)+\ell-1}$; furthermore, the coefficients of these linear combinations are chosen to ensure that $\tilde{\mathbf{u}}^{K(t-1)+\ell} \approx \mathbf{Y}^{\ell-1} \mathbf{u}^t$. Hence, from $\tilde{\mathbf{z}}^{Kt}$ and $(\tilde{\mathbf{u}}^{K(t-1)+\ell})_{\ell \in \{2,\dots,K\}}$, one obtains $(\mathbf{Y}^\ell \mathbf{u}^t)_{\ell \in [K]}$ (up to an $o_N(1)$ error). As a result, $\mathbf{J}(\mathbf{Y})\mathbf{u}^t$ can be expressed as a linear combination of $(\tilde{\mathbf{u}}^1, \dots, \tilde{\mathbf{u}}^{Kt}, \tilde{\mathbf{z}}^{Kt})$, which in turn is a linear combination of *(i)* the past iterates $\{\mathbf{u}^i\}_{i \in [t]}$, *(ii)* the signal $\mathbf{X}^*$, plus *(iii)* independent Gaussian noise. By inspecting the coefficients of this linear combination, one deduces *(i)* the values of the Onsager coefficients $\{\mathsf{c}_{t,i}\}_{i \in [t], t \geq 1}$ (as the coefficients multiplying the past iterates $\{\mathbf{u}^i\}_{i \in [t]}$), *(ii)* the mean μ_t (as the coefficient multiplying the signal $\mathbf{X}^*$), and *(iii)* the covariance matrix $\mathbf{\Sigma}_t$ (as the covariance matrix of the remaining noise terms). Finally, by making $\tilde{h}_{Kt+1}$ depend on g_{t+1}, we enforce that $\tilde{\mathbf{u}}^{Kt+1} \approx \mathbf{u}^{t+1}$. We highlight that the auxiliary AMP is employed purely as a proof technique. Its formal description is deferred to Appendix B.2.1, and its state evolution follows in Appendix B.2.2.

For simplicity, we assume to have access to an initialization $\mathbf{u}^1 \in \mathbb{R}^N$, which is independent of the noise $\mathbf{Z}$ and has a strictly positive correlation with $\mathbf{X}^*$, i.e.,

$$(\mathbf{X}^*, \mathbf{u}^1) \xrightarrow{W_2} (X^*, U_1), \quad \mathbb{E}[X^* U_1] := \epsilon > 0, \quad \mathbb{E}[U_1^2] = 1. \tag{5.192}$$

The requirement (5.192) is rather standard in the analysis of AMP algorithms. However, as having access to such an initialization is often impractical, a recent line of work has designed AMP iterations which are initialized with the eigenvector of the data matrix $\mathbf{Y}$ associated to the largest eigenvalue, see [113, 155, 133]. By following the approach detailed in [155], one can design a Bayes-optimal AMP with spectral initialization. As this would be out of the scope of the current contribution – whose goal is to obtain an algorithm with a Bayes-optimal *fixed point* – we will not pursue this extension here.

5.6.2 Onsager coefficients and state evolution recursion

We now detail the calculation of the Onsager coefficients $\{\mathsf{c}_{t,i}\}_{i \in [t], t \geq 1}$ and of the state evolution parameters $\boldsymbol{\mu}_t, \mathbf{\Sigma}_t$ associated to the AMP algorithm (5.190). We obtain these quantities from the state evolution recursion of the auxiliary AMP which, up to a $o_N(1)$ error, tracks $(\mathbf{Y}^{\ell-1} \mathbf{u}^t)_{\ell \in [K]}$ and, as such, has a number of iterations K times larger. To express the latter, we define a number of auxiliary quantities: the vector $\tilde{\boldsymbol{\mu}}_{Kt} \in \mathbb{R}^{Kt}$, the matrices $\tilde{\boldsymbol{\Delta}}_{Kt}, \tilde{\boldsymbol{\Phi}}_{Kt}, \tilde{\boldsymbol{\Sigma}}_{Kt}, \tilde{\mathbf{B}}_{Kt} \in \mathbb{R}^{Kt \times Kt}$, and the coefficients $\{\alpha_{i,j}\}_{j \in [i], i \in [Kt]}$, $\{\beta_{i,j}\}_{j \in [\lfloor (i-1)/K \rfloor + 1], i \in [Kt]}$, $\{\gamma_i\}_{i \in [Kt]}$, $\{\theta_{i,j}\}_{i \in [t], j \in [Kt]}$. The quantities $\tilde{\boldsymbol{\mu}}_{Kt}, \tilde{\boldsymbol{\Delta}}_{Kt}, \tilde{\boldsymbol{\Phi}}_{Kt}, \tilde{\boldsymbol{\Sigma}}_{Kt}, \tilde{\mathbf{B}}_{Kt}$ are directly connected to the state evolution of the auxiliary AMP (see the remark at the end of Appendix B.2.2). Furthermore, the coefficients $\{\alpha_{i,j}\}_{j \in [i], i \in [Kt]}$,

$\{\beta_{i,j}\}_{j \in [\lfloor (i-1)/K \rfloor +1], i \in [Kt]}$, $\{\gamma_i\}_{i \in [Kt]}$, $\{\theta_{i,j}\}_{i \in [t], j \in [Kt]}$ allow for a useful (approximate) decomposition of the vectors $(\mathbf{Y}^\ell \mathbf{u}^t)_{\ell \in [K-1]}$, see the remark at the end of this section.

We start with the initialization

$$\tilde{U}_1 := U_1, \tag{5.193}$$

where U_1 satisfies (5.192), and we set

$$\tilde{\mu}_1 := \lambda \epsilon, \quad (\tilde{\boldsymbol{\Delta}}_1)_{1,1} := 1, \quad (\tilde{\boldsymbol{\Phi}}_1)_{1,1} := 0, \quad (\tilde{\mathbf{B}}_1)_{1,1} := \bar{\kappa}_1, \quad (\tilde{\boldsymbol{\Sigma}}_1)_{1,1} := \bar{\kappa}_2,$$
$$\alpha_{1,1} := 0, \quad \beta_{1,1} := 1, \quad \gamma_1 := 0. \tag{5.194}$$

Here and in what follows, we denote by $\{\bar{\kappa}_k\}_{k \geq 1}$ the sequence of free cumulants associated to D. The free cumulants can be recursively computed from the moments, see e.g. [154, Section 2.5].

For $t \geq 1$, let us define

$$\tilde{U}_{K(t-1)+1+\ell} := \tilde{Z}_{K(t-1)+\ell} + \tilde{\mu}_{K(t-1)+\ell} X^* + \sum_{j=1}^{K(t-1)+\ell} (\tilde{\mathbf{B}}_{K(t-1)+\ell})_{K(t-1)+\ell,j} \tilde{U}_j, \; \ell \in [K-1], \tag{5.195}$$

$$\tilde{U}_{Kt+1} := g_{t+1}\left(\mu_t X^* + \sum_{j=1}^{Kt} \theta_{t,j} \tilde{Z}_j\right), \tag{5.196}$$

$$(\tilde{Z}_1, \ldots, \tilde{Z}_{Kt}) \sim \mathcal{N}(0, \tilde{\boldsymbol{\Sigma}}_{Kt}) \text{ and independent of } X^*, U_1. \tag{5.197}$$

We note that the function g_{t+1} in (5.196) is the AMP denoiser in (5.190). Let us also define

$$\tilde{\mu}_{K(t-1)+1+\ell} = \lambda \mathbb{E}[\tilde{U}_{K(t-1)+1+\ell} X^*], \tag{5.198}$$

$$(\tilde{\boldsymbol{\Delta}}_{K(t-1)+1+\ell})_{K(t-1)+1+\ell,j} = (\tilde{\boldsymbol{\Delta}}_{K(t-1)+1+\ell})_{j,K(t-1)+1+\ell} = \mathbb{E}[\tilde{U}_{K(t-1)+1+\ell} \tilde{U}_j], \tag{5.199}$$
$$j \in [K(t-1)+1+\ell],$$

$$(\tilde{\boldsymbol{\Phi}}_{K(t-1)+1+\ell})_{K(t-1)+1+\ell,j} = \mathbb{E}[\partial_{\tilde{Z}_j} \tilde{U}_{K(t-1)+1+\ell}], \quad j \in [K(t-1)+\ell], \tag{5.200}$$

$$\tilde{\mathbf{B}}_{K(t-1)+1+\ell} = \sum_{j=0}^{K(t-1)+\ell} \bar{\kappa}_{j+1} \tilde{\boldsymbol{\Phi}}_{K(t-1)+1+\ell}^j, \tag{5.201}$$

$$\tilde{\boldsymbol{\Sigma}}_{K(t-1)+1+\ell} = \sum_{j=0}^{2(K(t-1)+\ell)} \bar{\kappa}_{j+2} \sum_{k=0}^{j} (\tilde{\boldsymbol{\Phi}}_{K(t-1)+1+\ell})^k \tilde{\boldsymbol{\Delta}}_{K(t-1)+1+\ell} (\tilde{\boldsymbol{\Phi}}_{K(t-1)+1+\ell}^\mathsf{T})^{j-k}. \tag{5.202}$$

Now, we obtain $\tilde{\boldsymbol{\mu}}_{K(t-1)+1}, \tilde{\boldsymbol{\Delta}}_{K(t-1)+1}, \tilde{\boldsymbol{\Phi}}_{K(t-1)+1}, \tilde{\mathbf{B}}_{K(t-1)+1}, \tilde{\boldsymbol{\Sigma}}_{K(t-1)+1}$ by setting $\ell = 0$ in (5.198)–(5.202) (and by using the initialization (5.194) for $t = 1$). This allows us to define $\tilde{U}_{K(t-1)+2}$ by setting $\ell = 1$ in (5.195). Next, we obtain $\tilde{\boldsymbol{\mu}}_{K(t-1)+2}, \tilde{\boldsymbol{\Delta}}_{K(t-1)+2}, \tilde{\boldsymbol{\Phi}}_{K(t-1)+2}, \tilde{\mathbf{B}}_{K(t-1)+2}, \tilde{\boldsymbol{\Sigma}}_{K(t-1)+2}$ by setting $\ell = 1$ in (5.198)–(5.202). This allows us to define $\tilde{U}_{K(t-1)+2}$ by setting $\ell = 2$ in (5.195). We iterate this procedure until we have obtained $(\tilde{\boldsymbol{\mu}}_{K(t-1)+\ell}, \tilde{\boldsymbol{\Delta}}_{K(t-1)+\ell}, \tilde{\boldsymbol{\Phi}}_{K(t-1)+\ell}, \tilde{\mathbf{B}}_{K(t-1)+\ell},$

$\tilde{\boldsymbol{\Sigma}}_{K(t-1)+\ell})_{\ell\in[K]}$ and $(\tilde{U}_{K(t-1)+1+\ell})_{\ell\in[K-1]}$. We note that, for any $i \geq 1$, $\tilde{\mathbf{B}}_i$ and $\tilde{\boldsymbol{\Sigma}}_i$ are the top left sub-matrices of $\tilde{\mathbf{B}}_{i+1}$ and $\tilde{\boldsymbol{\Sigma}}_{i+1}$, respectively.

At this point, for $\ell \in [K-1]$, we compute the quantities $\{\alpha_{K(t-1)+1+\ell,j}\}_{j\in[K(t-1)+\ell]}$, $\{\beta_{K(t-1)+1+\ell,j}\}_{j\in[t]}$, $\gamma_{K(t-1)+1+\ell}$ as

$$\alpha_{K(t-1)+1+\ell,j} = \delta_{K(t-1)+\ell,j} + \sum_{\substack{i=1 \\ i\not\equiv 1(\mathrm{mod}\,K)}}^{K(t-1)+\ell} \alpha_{i,j}\,(\tilde{\mathbf{B}}_{K(t-1)+\ell})_{K(t-1)+\ell,i}, \quad j \in [K(t-1)+\ell], \quad (5.203)$$

$$\beta_{K(t-1)+1+\ell,j} = (\tilde{\mathbf{B}}_{K(t-1)+\ell})_{K(t-1)+\ell,K(j-1)+1} + \sum_{\substack{i=1 \\ i\not\equiv 1(\mathrm{mod}\,K)}}^{K(t-1)+\ell} \beta_{i,j}\,(\tilde{\mathbf{B}}_{K(t-1)+\ell})_{K(t-1)+\ell,i}, \; j \in [t], \quad (5.204)$$

$$\gamma_{K(t-1)+1+\ell} = \tilde{\mu}_{K(t-1)+\ell} + \sum_{\substack{i=1 \\ i\not\equiv 1(\mathrm{mod}\,K)}}^{K(t-1)+\ell} (\tilde{\mathbf{B}}_{K(t-1)+\ell})_{K(t-1)+\ell,i}\gamma_i. \quad (5.205)$$

In (5.203), $\delta_{i,j}$ denotes the Kronecker symbol ($\delta_{i,j} = 1$ if $i = j$ and 0 otherwise), and $\alpha_{i,j}$ is assumed to be 0 if $j \geq i$; in (5.204), $\beta_{i,j}$ is assumed to be 0 if $j > \lceil (i-1)/K \rceil$.

Recall that $\{c_i\}_{i=1}^K$ are the coefficients of the polynomial $\mathbf{J}(\mathbf{Y})$ (in $\mathbf{Y}$), i.e., $\mathbf{J}(\mathbf{Y}) = \sum_{i=1}^K c_i\mathbf{Y}^i$. Finally, we are ready to express μ_t, $\{\theta_{t,j}\}_{j\in[Kt]}$:

$$\mu_t = \sum_{i=1}^K c_i\Big(\tilde{\mu}_{K(t-1)+i} + \sum_{k=1}^{K(t-1)+i} \gamma_k\,(\tilde{\mathbf{B}}_{K(t-1)+i})_{K(t-1)+i,k}\Big), \quad (5.206)$$

$$\theta_{t,j} = \sum_{i=1}^K c_i\Big(\delta_{K(t-1)+i,j} + \sum_{k=1}^{K(t-1)+i} \alpha_{k,j}\,(\tilde{\mathbf{B}}_{K(t-1)+i})_{K(t-1)+i,k}\Big), \quad j \in [Kt]. \quad (5.207)$$

As before, $\alpha_{i,j}$ is assumed to be 0 if $j \geq i$. This allows us to define $\tilde{U}_{Kt+1}$ via (5.196) and, after setting $\beta_{Kt+1,t} = 1$, $\beta_{Kt+1,j} = 0$ for all $j \in [t-1]$, $\alpha_{Kt+1,j} = 0$ for all $j \in [Kt+1]$ and $\gamma_{Kt+1} = 0$, the definition of the state evolution recursion is complete.

From the state evolution recursion defined above, we can derive the Onsager coefficients $\{\mathsf{c}_{t,j}\}_{j\in[t]}$ as

$$\mathsf{c}_{t,j} = \sum_{i=1}^K c_i \sum_{k=1}^{K(t-1)+i} \beta_{k,j}\,(\tilde{\mathbf{B}}_{K(t-1)+i})_{K(t-1)+i,k}, \quad j \in [t]. \quad (5.208)$$

At this point, we are ready to present our result concerning the characterization of the iterates of the AMP algorithm (5.190), with Onsager coefficients given by (5.208), in the high-dimensional limit $N \to \infty$: we prove that the convergence (5.90) holds, where μ_t is given by (5.206) and $W_t = \sum_{j=1}^{Kt} \theta_{t,j}\tilde{Z}_j$, with $\{\theta_{t,j}, \tilde{Z}_j\}_{j\in[Kt]}$ described by the recursion above. Equivalently [156, Corollary 7.21], the convergence can be expressed in terms of pseudo-Lipschitz test

functions. A function $\psi \colon \mathbb{R}^m \to \mathbb{R}$ is *pseudo-Lipschitz of order* 2, denoted by $\psi \in \mathrm{PL}(2)$, if there exists a constant $C > 0$ such that

$$\|\psi(\mathbf{x}) - \psi(\mathbf{y})\|_2 \le C\Big(1 + \|\mathbf{x}\|_2 + \|\mathbf{y}\|_2\Big)\|\mathbf{x} - \mathbf{y}\|_2,$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$.

Theorem 5.1 (State evolution of the BAMP). *Let* $\mathbf{Y}$ *be given by* (5.2) *and which verifies Hypothesis 5.1, and let* $\mathbf{J}(\mathbf{Y}) = \sum_{i=1}^K c_i \mathbf{Y}^i$. *Consider the AMP algorithm* (5.190), *with initialization* (5.192), *Onsager coefficients* $\{\mathsf{c}_{t,j}\}_{j\in[t]}$ *given by* (5.208) *and where, for* $t \ge 1$, g_{t+1} *is continuously differentiable and Lipschitz. Then, the following limit holds almost surely for any* $\mathrm{PL}(2)$ *function* $\psi : \mathbb{R}^{2t+2} \to \mathbb{R}$, *for* $t \ge 1$ *as* $N \to \infty$:

$$\frac{1}{N}\sum_{i\le N}\psi(u_i^1,\ldots,u_i^{t+1},f_i^1,\ldots,f_i^t,X_i^*) \to \mathbb{E}\,\psi(U_1,\ldots,U_{t+1},F_1,\ldots,F_t,X^*). \tag{5.209}$$

Equivalently, as $N \to \infty$, *the joint empirical distribution of* $(\mathbf{u}^1,\ldots,\mathbf{u}^{t+1},\mathbf{f}^1,\ldots,\mathbf{f}^t,\mathbf{X}^*)$ *converges almost surely in Wasserstein-2 distance to* $(U_1,\ldots,U_{t+1},F_1,\ldots,F_t,X^*)$. *Here, for* $i \in [t]$, $U_{i+1} = g_{i+1}(F_t)$ *and* $(F_1,\ldots,F_t) = \boldsymbol{\mu}_t X^* + (W_1,\ldots,W_t)$, *with* $W_t = \sum_{j=1}^{Kt} \theta_{t,j}\tilde{Z}_j$ *and where* $\boldsymbol{\mu}_t$ *can be computed via* (5.206), $\{\theta_{t,j}\}_{j\in[Kt]}$ *via* (5.207) *and* $\{Z_j\}_{j\in[Kt]}$ *is given by* (5.197).

The proof of Theorem 5.1 is deferred to Appendix B.2.3. A few remarks are now in order. First, we highlight that (5.209) directly implies a high-dimensional characterization of the performance of the AMP (5.190). In fact, by taking the pseudo-Lipschitz functions $\psi(U_{t+1}, X^*) = (U_{t+1} - X^*)^2$, $\psi(U_{t+1}, X^*) = U_{t+1} \cdot X^*$ and $\psi(U_{t+1}, X^*) = (U_{t+1})^2$, we obtain the limit mean-square error and overlap of the AMP iterates as

$$\lim_{N\to\infty}\frac{1}{2N^2}\mathbb{E}\big\|\mathbf{X}^*(\mathbf{X}^*)^\mathsf{T} - \mathbf{u}^t(\mathbf{u}^t)^\mathsf{T}\big\|_\mathrm{F}^2 = \frac{1}{2}\Big(1 - 2\Big(\mathbb{E}[U_t \cdot X^*]\Big)^2 + (\mathbb{E}[(U_t)^2])^2\Big),$$

$$\lim_{N\to\infty}\frac{|\langle \mathbf{X}^*, \mathbf{u}^t\rangle|}{\|\mathbf{u}^t\|\cdot\|\mathbf{X}^*\|} = \frac{|\mathbb{E}[U_t \cdot X^*]|}{\sqrt{\mathbb{E}[(U_t)^2]}}. \tag{5.210}$$

Next, note that Theorem 5.1 holds for any family of denoisers $\{g_{t+1}\}_{t\ge 1}$, subject to some mild regularity requirement. A natural choice is to pick the posterior mean

$$g_{t+1}(f) = \mathbb{E}[U_* \mid F_t = f]. \tag{5.211}$$

Such a choice requires estimating the state evolution parameters μ_t, $\{\theta_{t,j}\}_{j\in[Kt]}$ and $\tilde{\boldsymbol{\Sigma}}_{Kt}$. These parameters, as well as the Onsager coefficients (5.208), can be estimated consistently from the data. To do so, first we obtain $\tilde{\boldsymbol{\Delta}}_{Kt}$ and $\tilde{\boldsymbol{\Phi}}_{Kt}$ by replacing expectations with empirical averages in (5.199) and (5.200), respectively. Next, we compute $\tilde{\mathbf{B}}_{Kt}$ and $\tilde{\boldsymbol{\Sigma}}_{Kt}$ by plugging in such estimates in (5.201) and (5.202), respectively. Having done that, we obtain $\{\alpha_{K(t-1)+1+\ell,j}\}_{j\in[K(t-1)+\ell],\ell\in[K-1]}$, $\{\beta_{K(t-1)+1+\ell,j}\}_{j\in[t],\ell\in[K-1]}$, $\{\gamma_{K(t-1)+1+\ell}\}_{\ell\in[K-1]}$ via (5.203)-(5.205).

Finally, μ_t, $\{\theta_{t,j}\}_{j\in[Kt]}$ and $\{c_{t,j}\}_{j\in[t]}$ can be computed from (5.206), (5.207) and (5.208), respectively.

As a final remark, we provide an interpretation of the coefficients $\{\alpha_{i,j}\}$, $\{\beta_{i,j}\}$, $\{\gamma_i\}$. As a by-product of the argument proving Theorem 5.1, we will show that, for $\ell \in [K-1]$, (cf. (B.20)–(B.21))

$$\lim_{N\to\infty} \frac{\|\mathbf{Y}^\ell \mathbf{u}^t - \sum_{j=1}^{K(t-1)+\ell} \alpha_{K(t-1)+1+\ell,j} \tilde{\mathbf{z}}^j - \sum_{j=1}^t \beta_{K(t-1)+1+\ell,j} \mathbf{u}^j - \gamma_{K(t-1)+1+\ell} \mathbf{X}^*\|^2}{N} = 0. \quad (5.212)$$

This formalizes the fact that $\mathbf{Y}^\ell \mathbf{u}^t$ can be approximately expressed as a linear combination of *(i)* the past iterates $\{\mathbf{u}^j\}_{j\in[t]}$, *(ii)* the signal $\mathbf{X}^*$, plus *(iii)* independent Gaussian noise (represented by the $\tilde{\mathbf{z}}^j$'s). The quantities $\{\alpha_{i,j}\}$, $\{\beta_{i,j}\}$, $\{\gamma_i\}$ represent the coefficients of this linear combination. The characterization (5.212) allows to subtract from $\mathbf{J}(\mathbf{Y})\mathbf{u}^k$ just the right Onsager terms, so that this difference equals a component in the direction of the signal (whose size is captured by μ_t) plus independent Gaussian noise (given by the linear combination of the $\tilde{\mathbf{z}}^j$'s via the coefficients $\{\theta_{i,j}\}$).

5.7 Numerics

For all experiments in this section, random instances of $\mathbf{Y}$ are generated according to the model (5.2). The signal has Rademacher prior, i.e., i.i.d. entries $X_i^* \sim \frac{1}{2}(\delta_1 + \delta_{-1})$. The noise matrices $\mathbf{Z} = \mathbf{O}^\intercal \mathbf{D} \mathbf{O}$ are generated by first drawing N i.i.d. eigenvalues $(D_i)_{i\leq N}$ according to the density (5.14), and then multiplying from left and right the diagonal matrix of eigenvalues $\mathbf{D}$ by a random Haar distributed orthogonal matrix $\mathbf{O}$ sampled independently for each realization. As mentioned at the end of Section 5.1.3, the results are expected to be the same if we were to draw $\mathbf{Z}$ according to the harder to sample[3] measure (5.3).

5.7.1 Spectral properties of the pre-processed matrix $\mathbf{J}(\mathbf{Y})$

Let us discuss the effect on the spectrum of $\mathbf{Y}$ that has the application of the optimal pre-processing function $J(\cdot)$; clearly, this function does not influence the eigenvectors of $\mathbf{Y}$ which therefore has the same basis as $\mathbf{J}(\mathbf{Y})$. From Figures 5.3 and 5.4, the effect is clear: the function J (Figure 5.4, middle plots (b)) "cleans" the eigenvalues of the data $\mathbf{Y}$ (Figure 5.4, upper plots (a)) by shifting the non-informative bulk eigenvalues of $\mathbf{Y}$ to negative values, while the largest, informative, eigenvalue is further separated from the bulk. This results in the histograms (Figure 5.4, lower plots (c)) for the processed data $\mathbf{J}(\mathbf{Y})$. It thus becomes much easier to distinguish the informative eigenvalue, which may be of interest for smaller instances where the finite-size effects are stronger.

---

[3]This can be done using the Dyson Brownian motion, see [35].
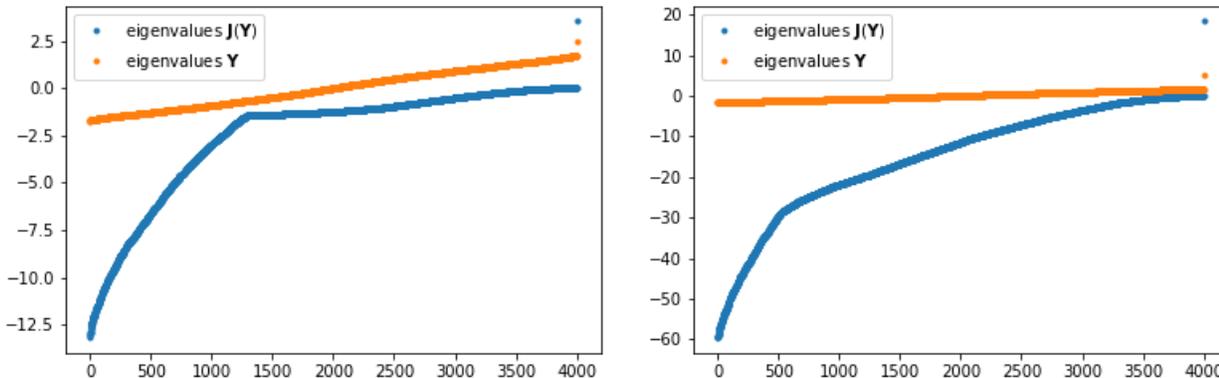
Figure 5.3: Ranked eigenvalues of the data matrix $\mathbf{Y}$ (orange) and the optimally pre-processed matrix $\mathbf{J}(\mathbf{Y})$ (blue) for $N = 4000$ for (left) $\lambda = 2$ and (right) $\lambda = 5$. The gap between the largest detached eigenvalue on the extreme right and the second highest one is much bigger for the pre-processed matrix. Moreover, all the eigenvalues of $\mathbf{J}(\mathbf{Y})$ in its non-informative bulk are negative.

Since the eigenbasis of $\mathbf{Y}$ remains untouched by the pre-processing, an algorithm based on spectral analysis only, like PCA, applied to $\mathbf{J}(\mathbf{Y})$ cannot hope to produce an improvement in performance. However, algorithms that exploit prior structural information about the signal, such as AMP, seem to be sensible this extra "detaching speed" of the leading eigenvalue from the bulk resulting from the pre-processing, thus getting better results as hereby reported.

5.7.2   BAMP improves over the existing AMP and matches the replica prediction for the MMSE

The plots of Figure 5.5 consider the quartic ensemble discussed in Section 5.1.2 for three values of the parameter μ, namely, $\mu \in \{0, 0.5, 1\}$ (recall $\gamma = \gamma(\mu)$ is fixed by relation (5.17)). Again, in all cases the signal $\mathbf{X}^*$ is assumed to have a Rademacher prior. The estimators of the spike $\mathbf{X}^*\mathbf{X}^{*\intercal}$ are compared in terms of the MSE (y-axis) achieved at the fixed point, as a function of the SNR λ (x-axis). All algorithms are run for $N = 8000$ and the results are averaged over $n_{\text{trials}} = 50$ independent trials; the state evolution recursions (and the replica prediction as well) correspond to $N \to \infty$. We compare the following inference procedures:

- In black, we plot the replica prediction (5.79), obtained as the fixed point of (5.73)–(5.78).

- In red, we plot the performance of the BAMP algorithm described in Section 5.6, where g_{t+1} is the posterior mean denoiser (5.211). More specifically, the red line corresponds
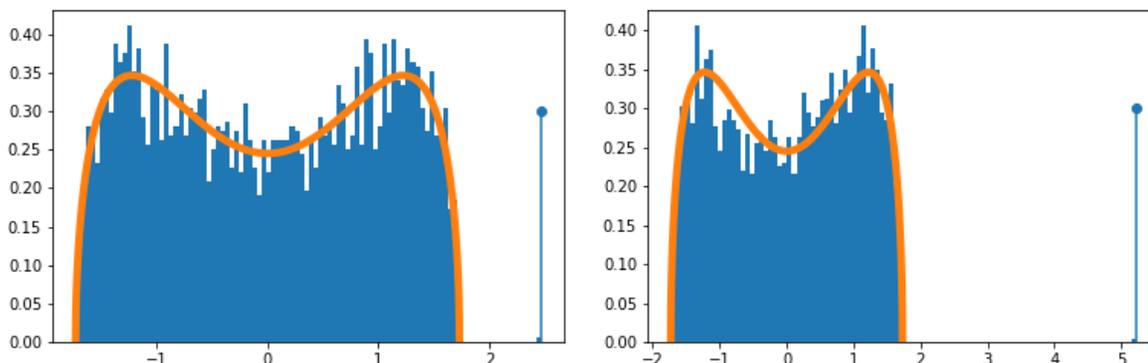
(a) Empirical spectral density of $\mathbf{Y}$. The largest, informative, eigenvalue is emphasized.
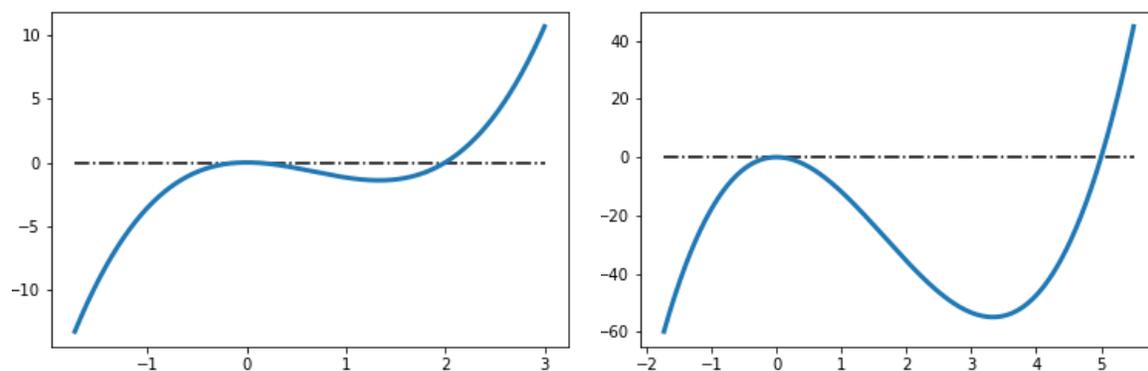
(b) The function $J(x) = \mu\lambda x - \gamma\lambda^2 x^2 + \gamma\lambda x^3$ with $(\mu = 0, \gamma(0) = 16/27)$ is used to optimally pre-process the (eigenvalues of the) data $\mathbf{Y}$ and obtain $\mathbf{J}(\mathbf{Y})$. The dashed curve indicates 0. By comparison with the plots (a) above, we understand that the noise bulk will be pushed to negative values, while the spike towards the right, which results in a "cleaning" effect.
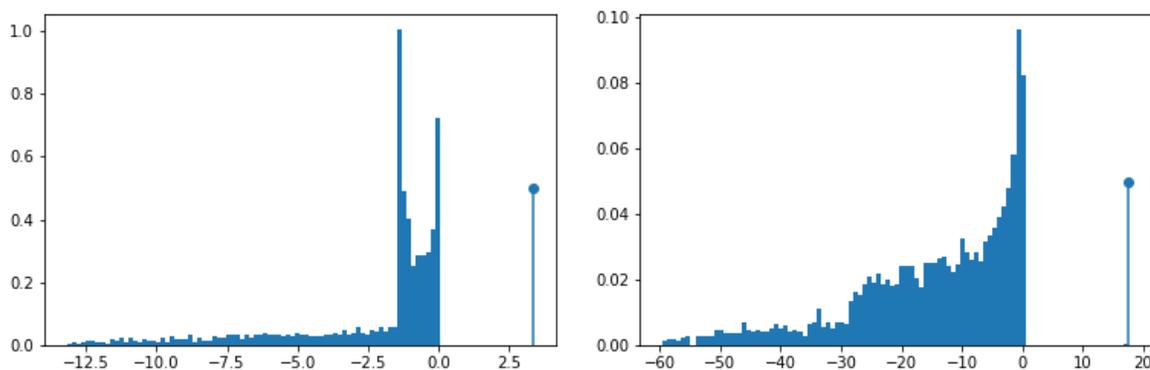
(c) Empirical spectral density of the pre-processed matrix $\mathbf{J}(\mathbf{Y})$. The largest eigenvalue is emphasized and well separated from the negative bulk by the application of $J(x)$.

Figure 5.4: Effect of the optimal pre-processing $J(x)$ on the eigenvalues of $\mathbf{Y}$. All experiments are for the most structured noise ensemble $(\mu = 0, \gamma(0) = 16/27)$ and $N = 4000$. The left column corresponds to $\lambda = 2$, while the right column to $\lambda = 5$.

to the fixed point of the MSE given by the state evolution recursion discussed in Section 5.6.2 (cf. (5.210)), and the red stars denote the MSE obtained by running the BAMP algorithm (5.190).

- In blue, we plot the performance of the AMP proposed in [132]. More specifically, the blue line corresponds to the fixed point of the MSE (5.92) obtained by choosing the posterior mean denoiser with a single-step memory term (5.91). The blue diamonds denote the MSE obtained by running the AMP (5.89) with this single-step denoiser.

- Finally, the green squares denote the MSE obtained by running the AMP in [133], which consists in employing the following posterior mean denoiser with a multi-step memory term in the iteration (5.89):

$$h_{t+1}(f_1, \ldots, f_t) = \mathbb{E}[X^* \mid (F_1, \ldots, F_t) = (f_1, \ldots, f_t)]. \tag{5.213}$$

We note that all algorithms converge rapidly: 10 iterations are sufficient to reach the corresponding fixed points. A few remarks concerning the numerical results displayed in Figure 5.5 are now in order:

- In all settings, the fixed point of the BAMP state evolution (in red) matches the replica prediction (in black). This is a strong numerical evidence supporting our conjecture that the proposed BAMP algorithm is Bayes-optimal. These theoretical curves for $N \to \infty$ are also remarkably close to the MSE achieved by the BAMP algorithm (5.190) at $N = 8000$.

- When $\mu = 0$, i.e., the noise is sufficiently far from being independent Gaussian, there is a clear performance gap between our proposed BAMP (in red) and the existing AMP algorithms [132, 133] (single-step denoiser in blue, and multi-step in green). As predicted by our theory, this gap is reduced for $\mu = 0.5$, and all curves collapse for $\mu = 1$.

- Finally, we note that the BAMP algorithm exhibits a numerical instability for low SNR. More specifically, when $\mu = 0$ and $\lambda = 2.3$, 5 out of the 50 trials of the iteration (5.190) do not reach the fixed point of state evolution (and are therefore discarded). Furthermore, by inspecting Figure 5.5c, one notices that the curve representing the BAMP state evolution detaches from the replica prediction as the SNR get smaller than 2.3. As expected, considering an initialization closer to the fixed point mitigates the issue. This numerical instability is likely due to the state evolution of BAMP corresponding to the recursion of an auxiliary AMP that *triples* the number of iterations. This fact leads to an amplification of the numerical errors.

Let us re-emphasize that all these results hold in the Bayesian-optimal setting where all hyper-parameters of the model are known and optimally used. In practical situations this may not be the case. In particular the statistical properties of the correlated noise $\mathbf{Z}$ may be
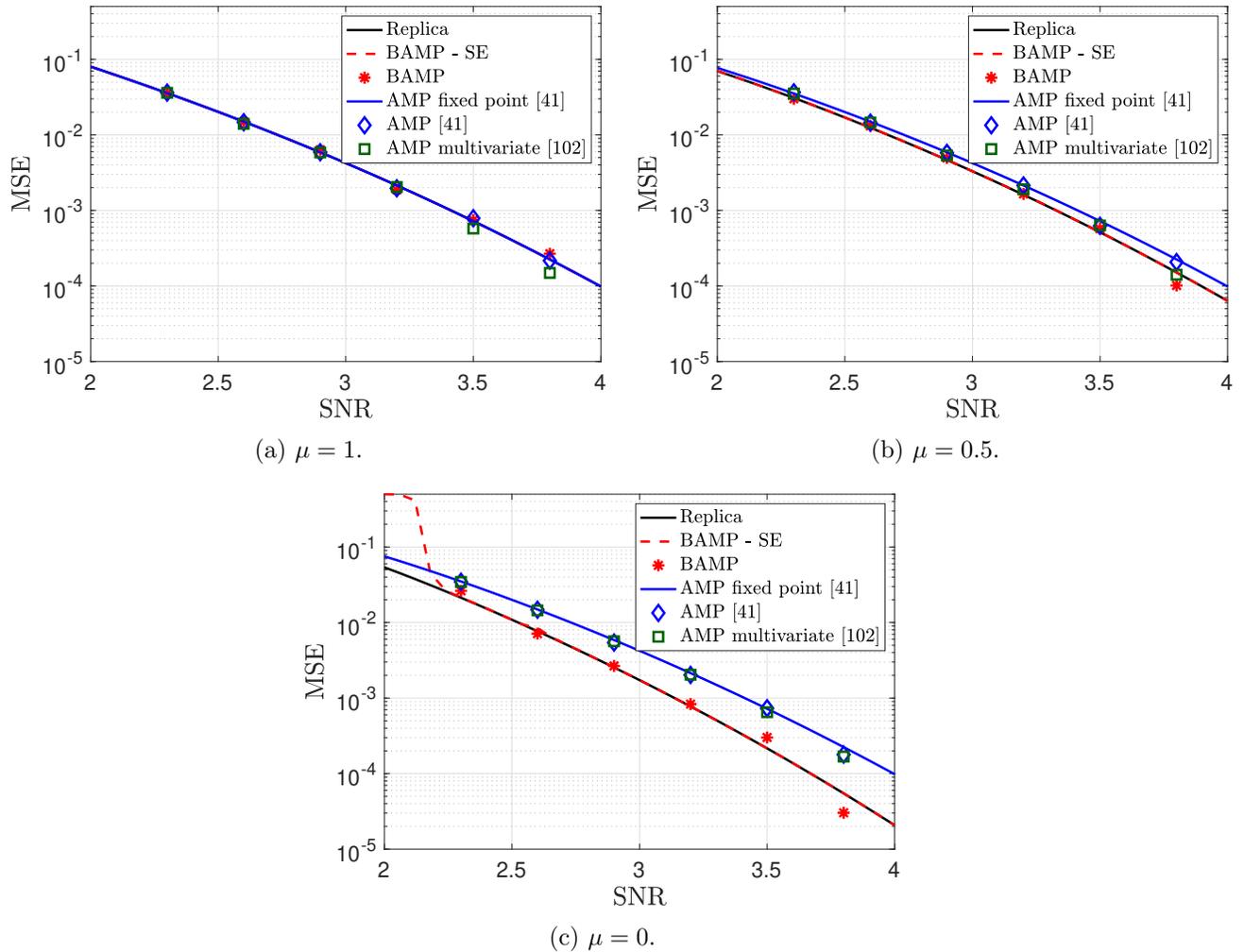
(a) $\mu = 1$.

(b) $\mu = 0.5$.

(c) $\mu = 0$.

Figure 5.5: Performance comparison between the replica prediction for the MMSE (in black), the proposed BAMP (in red), and the existing AMP [132, 133] (in blue and green). BAMP matches the Bayes-optimal MSE predicted via the replica method, and it outperforms the existing AMP when the noise is not Gaussian. This improvement is more evident as the noise distribution gets further from a Wigner distribution. Taken all together, these numerical results provide an empirical confirmation of the (Bayes-)optimality of the proposed BAMP algorithm.

only partially known, preventing one to obtain the coefficients (c_k) defining the optimal pre-processing of the data $\mathbf{J}(\mathbf{Y}) = \sum_{k \leq K} c_k \mathbf{Y}^k$ as done in Section 5.5. In Appendix B.1 we provide a learning procedure based on expectation maximization to overcome this issue and which can be of help to practitioners aiming at using BAMP in more realistic situations. Its testing is left for future work.

Chapter 6

Matrix factorization

Consider the problem of reconstructing the two factors $\mathbf{D} \in \mathbb{R}^{N \times P}$ and $\mathbf{A} \in \mathbb{R}^{P \times M}$ composing a matrix $\mathbf{Y}^* = \mathbf{DA} \in \mathbb{R}^{N \times M}$ from noisy observations of it, $\mathbf{Y}$. Assuming that $\mathbf{Y}^*$'s elements are blurred by Gaussian additive independent noise of variance σ^2, the Bayes posterior measure of matrix factorization reads as

$$dP_{\mathbf{D},\mathbf{A}|\mathbf{Y}}(\mathbf{d},\mathbf{a}) \propto \prod_{i=1}^{N} \prod_{j=1}^{M} e^{-\frac{1}{2\sigma^2}(Y_{ij}-(\mathbf{da})_{ij})^2} dP_{\mathbf{D}}(\mathbf{d}) dP_{\mathbf{A}}(\mathbf{a}) \tag{6.1}$$

where possible constraints, such as sparsity, may be enforced via the priors $P_{\mathbf{D}}$ and $P_{\mathbf{A}}$.

The need to factorize a given matrix manifests itself in several concrete tasks. In fact, a possible strategy to denoise large matrices, *i.e.* to estimate simply $\mathbf{Y}^*$ from $\mathbf{Y}$, is to find a more convenient representation of the data at disposal [157]. In this context, we mention *dictionary learning* [158, 159, 160], that aims to find two factors $\mathbf{D}$ and $\mathbf{A}$ of $\mathbf{Y}^*$ such that the columns of $\mathbf{D}$ form an over-complete basis of $\mathbb{R}^N$ ($P > N$) and $\mathbf{A}$ is sparse. Furthermore, also Boltzmann machines are an instance of the implementation of the same strategy [161, 162]: given a set of observations on their visible units, while training they create an internal representation on the hidden units. The core idea of representation learning, and presumably the reason of its effectiveness [163, 164, 165], is the extraction of characteristic features, *e.g.* the over-complete basis in $\mathbf{D}$, that if properly recombined can reconstruct the data, or even some of their missing parts, as in recommender systems [166]. Another motivation for studying matrix factorization is to find new ways of training deep networks. Given a desired output, the process of learning can be decomposed layer by layer, in which the task in each layer is to find a set of synaptic weights together with the internal representation of the data in the previous layer. This is a matrix factorization task (complicated by the non-linearity), that one can hope to turn into a self-consistent solution of deep network training, following what was done for multi-layer generalized linear estimation [167].

The analysis that follows can be carried out for the generic $\mathbf{D}$-$\mathbf{A}$ problem in (6.1). However,

for the sake of clarity we present it here in its symmetric version, that is when

$$\mathbf{Y} = \frac{\boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}}{\sqrt{N}} + \sqrt{\Delta}\mathbf{Z} \ \in \mathbb{R}^{N \times N} \tag{6.2}$$

with $\Delta \geq 0$, $\boldsymbol{\xi} = (\xi_i^\mu)_{i \leq N}^{\mu \leq P}$, $\xi_i^\mu \overset{\text{iid}}{\sim} P_\xi$ and $\mathbf{Z} = (Z_{ij})_{i,j \leq N}$, $Z_{ij} \sim \mathcal{N}(0, 1 + \delta_{ij})$ all independent. We shall restrict to the family of priors $P_\xi = (1 - \rho)\delta_0 + \frac{\rho}{2}[\delta_{1/\sqrt{\rho}} + \delta_{-1/\sqrt{\rho}}]$, $\rho \in (0, 1]$. If we imagine μ as a "time" index, the symmetric matrix $\boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}/N$ in (6.2) can help us capture the correlation between the time series $(\xi_i^t)^{t \leq P}$ and $(\xi_j^t)^{t \leq P}$ of two quantities, that might be stock prices of two different assets for instance [168, 169]. Therefore, studying the performance of denoising procedures for sample covariance matrices as $\boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}/N$ is of paramount importance to establish their reliability and hence to characterize the main dynamical modes of a stochastic process.

We have already studied the finite rank (P) case, *i.e.* the spiked Wigner model, and we now know that in the Bayes-optimal setting the average means square error made by a Statistician is the least possible, and AMP algorithms saturate such information theoretical bounds. The most challenging regime so far though is that of extensive rank, namely when $P/N \to \alpha > 0$ as N grows. A possible route to tackle the denoising problem of $\boldsymbol{\xi}\boldsymbol{\xi}^\mathsf{T}$ in this regime was introduced in [170] and it amounts to restrict the class of possible estimators for the hidden matrix to the rotational invariant ones. Such estimators are indeed called *Rotationally Invariant Estimators* (RIE). RIEs have the peculiarity of being diagonal on the same eigenbasis $\mathbf{O_Y}$ of $\mathbf{Y}$. What remains to do then is to find a good cleaning procedure to produce estimates of the eigenvalues $\hat{\boldsymbol{\lambda}}_\mathbf{S}$ of the hidden matrix and finally recompose the estimator: $\hat{\mathbf{S}} = \mathbf{O_Y}\hat{\boldsymbol{\lambda}}_\mathbf{S}\mathbf{O_Y}^\mathsf{T}$. Notice however, that RIEs do not answer completely to the matrix factorization problem, their only goal is denoising, and no representation of the data is searched for. Furthermore, in general they are not Bayes-optimal if there is no rotational invariance encoded in the prior P_ξ.

Bayes-optimal limits for (symmetric) matrix factorization have been investigated through perturbative [112, 171] and non-perturbative approaches [172], but the high rank regime seems to be an insurmountable obstacle to the production of closed formulae for the information theoretically optimal reconstruction performances. A substantial difficulty is indeed introduced by the use of a prior over the entire matrix $\boldsymbol{\xi}$. Therefore, as opposed to that we propose here an alternative strategy: being $\boldsymbol{\xi}$ composed by P vectors $\boldsymbol{\xi}^\mu$, $\mu = 1, 2, \ldots, P$, we aim for one of them at a time. Assuming that we are able to get efficiently an estimate of the first one, say $\boldsymbol{\xi}^P$, denoted by $\boldsymbol{\eta}^P$, we can build a rank one contribution and subtract it from $\mathbf{Y} \equiv \mathbf{Y}(0)$, obtaining $\mathbf{Y}(1/P) = \mathbf{Y}(0) - \boldsymbol{\eta}^P\boldsymbol{\eta}^{P\mathsf{T}}/\sqrt{N}$. Then we iterate the process until, if possible, we have an estimate $\boldsymbol{\eta}^\mu$ for each $\boldsymbol{\xi}^\mu$. We will refer to this iterative scheme as *decimation*.

Using decimation we show that probabilistic symmetric matrix factorization is possible, that it has a better performance in denoising w.r.t. the RIE in some range of the model parameters, and that a strong sparsity, tuned by ρ, can make the system more robust against noise.

6.1 Definition of the problem

In this chapter we deal with the Gaussian observation channel

$$\mathbf{Y} = \mathbf{Y}^* + \sqrt{\Delta}\mathbf{Z} = \frac{\boldsymbol{\xi}\boldsymbol{\xi}^T}{\sqrt{N}} + \sqrt{\Delta}\mathbf{Z} \ \in \mathbb{R}^{N \times N} \tag{6.3}$$

where $\Delta \geq 0$ will be referred to as *noise-to-signal ratio* and $Z_{ij} \sim \mathcal{N}(0, 1 + \delta_{ij})$ are all independent. Let us assume that the Statistician was able to estimate, in a way we shall clarify soon, the last R $(= 0, 1, \ldots, P - 1)$ patterns without loss of generality, denote the estimates by $(\boldsymbol{\eta}^\mu)_{\mu > P-R}^P$ and the fraction of retrieved patterns as $t = R/P$. We define the modified observations according to our procedure now appear as

$$\mathbf{Y}(t) = \sum_{\mu=1}^{P} \frac{\boldsymbol{\xi}^\mu \boldsymbol{\xi}^{\mu\,T}}{\sqrt{N}} - \sum_{\mu=P(1-t)+1}^{P} \frac{\boldsymbol{\eta}^\mu \boldsymbol{\eta}^{\mu\,T}}{\sqrt{N}} + \sqrt{\Delta}\mathbf{Z} \,. \tag{6.4}$$

For the moment we assume that in order to obtain the estimates $\boldsymbol{\eta}^\mu$ the Statistician samples from the Boltzmann-Gibbs distribution

$$d\mu(\mathbf{x}) = \frac{1}{\mathcal{Z}_N} dP_\xi(\mathbf{x}) \exp\left[\frac{\beta}{2\sqrt{N}} \mathrm{Tr}\,\mathbf{Y}(t)\mathbf{x}\mathbf{x}^T - \beta \frac{\|\mathbf{x}\|^4}{4N} + \frac{\beta\hat{v}}{2}(N - \|\mathbf{x}\|^2) \right], \quad dP_\xi(\mathbf{x}) := \prod_{i=1}^{N} dP_\xi(x_i) \tag{6.5}$$

that corresponds to the posterior distribution of a strongly mismatched inference problem, where the Statistician tries to reconstruct an extensive rank matrix with only a rank one matrix. Notice that we have also artificially introduced a Lagrange multiplier $\hat{v}$ w.r.t. which we require stationarity. This allows us to enforce the constraint

$$\mathbb{E}\left\langle \frac{\|\mathbf{x}\|^2}{N} \right\rangle = 1 \tag{6.6}$$

with $\langle \cdot \rangle$ the expectation w.r.t. (6.5) and $\mathbb{E}$ averages over all the remaining quenched noise. The Lagrange multiplier $\hat{v}$ is needed because, as mentioned above, this kind of estimation is mismatched, thus the Nishimori identities break down and (6.6) would not be true in general in absence of $\hat{v}$. Another good reason to introduce it is that, although the Statistician is performing a mismatched estimation, she is Bayes optimal and she knows her estimates should be properly normalized as in (6.6). Furthermore, as we shall see in the following, the introduction of $\hat{v}$ simplifies a lot the fixed point equations, and extends the range of parameters α, Δ in which retrieval is possible when $\beta = 1/\Delta$, that is a somehow natural choice of parameters.

Let us further expand the Hamiltonian at the exponent in (6.5):

$$-H_N(\mathbf{x}) = \frac{\sqrt{\Delta}}{2\sqrt{N}} \sum_{i,j=1}^{N} Z_{ij} x_i x_j + \frac{1}{2N} \sum_{\mu=1}^{P} \left(\sum_{i=1}^{N} \xi_i^\mu x_i \right)^2 - \frac{1}{2N} \sum_{\mu=P(1-t)+1}^{P} \left(\sum_{i=1}^{N} \eta_i^\mu x_i \right)^2$$

$$- \frac{\|\mathbf{x}\|^4}{4N} + \frac{\hat{v}}{2}(N - \|\mathbf{x}\|^2). \quad (6.7)$$

The previous Hamiltonian is really similar to that of the classical Hopfield model [173], except for the presence of the first noise terms and the η-terms at the end of the first line. For this reason we shall refer to the $\boldsymbol{\xi}^\mu$'s also as patterns. If the Statistician has "good enough" estimates $\boldsymbol{\eta}^\mu$ of the R patterns, namely with a good overlap with the corresponding $\boldsymbol{\xi}^\mu$, then $-\frac{\beta}{2N} \sum_{\mu=P(1-t)+1}^{P} \left(\sum_{i=1}^{N} \eta_i^\mu x_i \right)^2$ acts as a repulsion from those very same patterns, penalizing them in probability. Therefore, we expect the estimates to condense onto the other $P - R$ non retrieved patterns, and in particular, in analogy with the Hopfield model, onto a finite number of them.

It is convenient at this point to introduce the Mattis magnetizations

$$m^\mu(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu x_i, \quad \mu = 1, \ldots, P \quad (6.8)$$

$$p^\mu(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} \eta_i^\mu x_i, \quad \mu = P(1-t) + 1, \ldots, P. \quad (6.9)$$

The Hamiltonian then takes the useful form

$$-H_N(\mathbf{x}) = \frac{\sqrt{\Delta}}{2\sqrt{N}} \sum_{i,j=1}^{N} Z_{ij} x_i x_j + \frac{N}{2} \sum_{\mu=1}^{P} (m^\mu(\mathbf{x}))^2 - \frac{N}{2} \sum_{\mu=P(1-t)+1}^{P} (p^\mu(\mathbf{x}))^2$$

$$- \frac{\|\mathbf{x}\|^4}{4N} + \frac{\hat{v}}{2}(N - \|\mathbf{x}\|^2). \quad (6.10)$$

Notice that the $\boldsymbol{\eta}^\mu$'s depend in general on the specific instances of the $\boldsymbol{\xi}^\mu$'s, so they must be treated as quenched noise with a dependency on the $\boldsymbol{\xi}^\mu$'s that we need to model. First off, we argue that conditionally on the patterns $(\boldsymbol{\xi}^\mu)_\mu$ the randomness in η_i^μ is only through ξ_i^μ. Secondly, define the first two conditional moments

$$\mathbb{E}_{\boldsymbol{\eta}|\boldsymbol{\xi}}[\eta_i^\mu] = m_i^\mu, \quad \mathbb{E}_{\boldsymbol{\eta}|\boldsymbol{\xi}}[(\eta_i^\mu)^2] = v_i^\mu. \quad (6.11)$$

$\boldsymbol{\eta}^\mu$ tends to align to $\boldsymbol{\xi}^\mu$ whose components are *i.i.d.* from a centered prior and therefore each of them has vanishing expectation. A further justification to this assumption will be given after the computation of the free entropy of the model (see Remark 6.2). So, to sum up

$$\mathbb{E}_{\boldsymbol{\xi}}[\eta_i^\mu] = \mathbb{E}_{\boldsymbol{\xi}} \mathbb{E}_{\boldsymbol{\eta}|\boldsymbol{\xi}}[\eta_i^\mu] = 0. \quad (6.12)$$

Since the Statistician enforces the condition (6.6) at each step of decimation, it holds also that

$$\frac{1}{N}\sum_{i=1}^{N}(\eta_i^\mu)^2 \simeq 1 \tag{6.13}$$

with high probability, thus can conjecture the following consistency constraint:

$$\mathbb{E}[v_i^\mu] = 1. \tag{6.14}$$

At the $R = tP$ step of decimation, the $\boldsymbol{\eta}^{P(1-t)+1}$ estimate for the $\boldsymbol{\xi}^{P(1-t)+1}$ pattern is found. We define thus the error

$$\epsilon(t) = \frac{1}{2N}\sum_{i=1}^{N}\left(\eta_i^{P(1-t)+1} - \xi_i^{P(1-t)+1}\right)^2 =$$

$$= \frac{1}{2N}\sum_{i=1}^{N}[(\eta_i^{P(1-t)+1})^2 + (\xi_i^{P(1-t)+1})^2] - \frac{1}{N}\boldsymbol{\eta}^{P(1-t)+1}\cdot\boldsymbol{\xi}^{P(1-t)+1}. \tag{6.15}$$

According to the previous considerations, the first normalized sum equals 2 with high probability, whereas the normalized scalar product is expected to converge to the equilibrium Mattis magnetization of the R-th step of decimation. Hence

$$\bar{\epsilon}(t) := \mathbb{E}\epsilon(t) = 1 - \mathbb{E}[m_i^{P(1-t)+1}\xi_i^{P(1-t)+1}]. \tag{6.16}$$

Analogous relations hold for other values $\tau \in [0, t)$ corresponding to some decimation step.
  Within this framework the decimation recursion goes as follows. From 0-th to first step:

- start at $t = 0$, which means $R = 0$ patterns recovered yet;

- sample the first estimate $\boldsymbol{\eta}^P$ from (6.5) with $\mathbf{Y}(0) = \mathbf{Y}$;

- obtain $\mathbf{Y}(1/P) = \mathbf{Y}(0) - \frac{1}{\sqrt{N}}\boldsymbol{\eta}^P\boldsymbol{\eta}^{P\intercal}$.

From R-th to $R + 1$-th step:

- sample $\boldsymbol{\eta}^{P-R}$ from (6.5) with $\mathbf{Y}(t) = \mathbf{Y}(R/P) = \mathbf{Y}((R-1)/P) - \frac{1}{\sqrt{N}}\boldsymbol{\eta}^{P-R+1}\boldsymbol{\eta}^{P-R+1\intercal}$, the last contribution coming from the R-th step;

- obtain $\mathbf{Y}((R+1)/P) = \mathbf{Y}(R/P) - \frac{1}{\sqrt{N}}\boldsymbol{\eta}^{P-R}\boldsymbol{\eta}^{P-R\intercal}$.

Remark 6.1. There are three noise sources in the decimation procedure:

- the original noise $\mathbf{Z}$, which is there by definition of the problem;

- interference among the patterns: this phenomenon appears because of the high rank nature of the Hopfield-like interaction [174, 175]. In the (large) N dimensional space of configurations of $\mathbf{x}$ the directions pointed by the $\boldsymbol{\xi}^{\mu}$'s act as attractors, meaning that they gather a lot of probability according to (6.5). However, when these directions are a finite ratio of all the N independent ones, intereference phenomena may occur. One way to understand it is from a geometrical point of view: an estimate $\boldsymbol{\eta}^{\mu}$ cannot have an extensive $O(N)$ projection on every direction $(\boldsymbol{\xi}^{\mu})_{\mu=1}^{P}$, otherwise it would have a super-extensive norm. The other projections must be at most of order $O(\sqrt{N})$, but they are very numerous, so when summed together they give back another $O(N)$ contribution that acts as a noise, whose intensity depends on α;

- the decimation procedure itself: each time we subtract a rank one contribution of the type $\frac{1}{\sqrt{N}}\boldsymbol{\eta}^{\mu}\boldsymbol{\eta}^{\mu\mathsf{T}}$ we are decreasing the *effective rank* of the hidden high rank matrix we want to estimate. This goes in the right direction, since from the previous point we also expect the patterns interference to decrease. However, the estimates $\boldsymbol{\eta}^{\mu}$ are noisy versions of $\boldsymbol{\eta}^{\mu}$ themselves, hence each subtraction adds inevitably some more noise.

From the previous considerations we understand that the viability of decimation is strongly affected by the interplay between the last two noise sources listed above. In fact, the whole point of the replica computation that follows is to compare the contributions of the two mentioned noise sources.

6.2   Replica symmetric free entropy

The goal of this section is the computation of the large N limit of free entropy

$$\bar{p}_N = \frac{1}{N}\mathbb{E}\log \int dP_\xi(\mathbf{x})\exp\left[-\beta H_N(\mathbf{x})\right], \tag{6.17}$$

where $\mathbb{E}$ is taken w.r.t. all the disorder: $\mathbf{Z}, \boldsymbol{\xi}, \boldsymbol{\eta}$. In order to do it, we employ the *replica method*. We recall it is based on the following identity:

$$\bar{p}_{N,n} := \frac{1}{nN}\log\mathbb{E}\mathcal{Z}_N^n \xrightarrow{n\to 0} \frac{1}{N}\mathbb{E}\log\mathcal{Z}_N = \bar{p}_N. \tag{6.18}$$

The limit would require n to be a real (or at least rational) parameter, but we will carry out the computations as if it were an integer. We will also assume to be able to exchange the $N \to \infty$ limit with the $n \to 0$ one. If we accept these premises, we have the great advantage of having an annealed expectation in front of the partition function, at the cost of analysing a "replicated" system. Hence, from now on we will focus on the quantity

$$\mathbb{E}\mathcal{Z}_N^n := \mathbb{E}_{\mathbf{Z}}\mathbb{E}_{\boldsymbol{\xi},\boldsymbol{\eta}}\int \prod_{a=1}^{n} dP_\xi(\mathbf{x}_a)\exp\left[-\beta\sum_{a=1}^{n}H_N(\mathbf{x}_a)\right]. \tag{6.19}$$

The strategy now is to carry out the averages w.r.t. the quenched disorders first. To avoid proliferation of many terms in the exponential, we need to treat them separately. Let us begin with the first noise terms in (6.10), and the related $\mathbb{E}_{\mathbf{Z}}$ average

$$
\mathbb{E}_{\mathbf{Z}} \exp\left(\frac{\beta\sqrt{\Delta}}{2\sqrt{N}} \sum_{i,j=1}^{N} Z_{ij} \sum_{a=1}^{n} x_{a,i}x_{a,j}\right) = \exp\left(\frac{\beta^2\Delta}{4N} \sum_{i,j=1}^{N} \sum_{a,b=1}^{n} x_{a,i}x_{a,j}x_{b,i}x_{b,j}\right)
$$

$$
= \exp\left(\frac{N\beta^2\Delta}{4} \sum_{a\neq b}^{n} Q^2(\mathbf{x}_a, \mathbf{x}_b) + \beta^2\Delta \sum_{a=1}^{n} \frac{\|\mathbf{x}_a\|^4}{4N}\right) . \tag{6.20}
$$

Now we take care of the penalizing p-terms in (6.10). After replicating, their contribution to the partition function is

$$
A := \prod_{\mu=P(1-t)+1}^{P} \prod_{a=1}^{n} e^{-\frac{N\beta}{2}(p^{\mu}(\mathbf{x}_a))^2} = \prod_{\mu=P(1-t)+1}^{P} \prod_{a=1}^{n} \int \frac{ds_a^{\mu}}{\sqrt{2\pi}} e^{-\frac{(s_a^{\mu})^2}{2} + i\sqrt{\frac{\beta}{N}} s_a^{\mu} \sum_{j=1}^{N} \eta_j^{\mu} x_{a,j}} . \tag{6.21}
$$

Notice that, thanks to the introduction of the auxiliary Gaussian variables $(s_a^{\mu})_{a\leq n, P(1-t)<\mu\leq P}$, the exponential is now decoupled over the particle indices j. We now take the expectation of A w.r.t. $\boldsymbol{\eta}$, conditioning on $\boldsymbol{\xi}$, keeping in mind our assumptions on the $\boldsymbol{\eta}^{\mu}$'s:

$$
\mathbb{E}_{\boldsymbol{\eta}|\boldsymbol{\xi}}[A] = \prod_{\mu=P(1-t)+1}^{P} \prod_{a=1}^{n} \int \frac{ds_a^{\mu}}{\sqrt{2\pi}} e^{-\frac{(s_a^{\mu})^2}{2}} \prod_{i=1}^{N} \mathbb{E}_{\eta_i^{\mu}|\xi_i^{\mu}} \exp\left(i\sqrt{\frac{\beta}{N}} \eta_i^{\mu} \sum_{a=1}^{n} s_a^{\mu} x_{a,i}\right)
$$

$$
= \prod_{\mu=P(1-t)+1}^{P} \prod_{a=1}^{n} \int \frac{ds_a^{\mu}}{\sqrt{2\pi}} \exp\left(-\frac{(s_a^{\mu})^2}{2} + \sum_{i=1}^{N} \log \mathbb{E}_{\eta_i^{\mu}|\xi_i^{\mu}} e^{i\sqrt{\frac{\beta}{N}} \eta_i^{\mu} \sum_{a=1}^{n} s_a^{\mu} x_{a,i}}\right) . \tag{6.22}
$$

Now we can expand the exponential inside the log up to second order, the remaining terms will be of sub-leading order and thus neglected in the following:

$$
\mathbb{E}_{\boldsymbol{\eta}|\boldsymbol{\xi}}[A] = \prod_{\mu=P(1-t)+1}^{P} \prod_{a=1}^{n} \int \frac{ds_a^{\mu}}{\sqrt{2\pi}} \exp\left(-\frac{(s_a^{\mu})^2}{2} + \sum_{a=1}^{n} is_a^{\mu}\sqrt{\frac{\beta}{N}} \sum_{i=1}^{N} m_i^{\mu} x_{a,i}\right.
$$

$$
\left.-\frac{\beta}{2} \sum_{a,b=1}^{n} s_a^{\mu}s_b^{\mu} \sum_{i=1}^{N} \frac{(v_i^{\mu} - (m_i^{\mu})^2)}{N} x_{a,i}x_{b,i}\right)
$$

$$
= \prod_{\mu=P(1-t)+1}^{P} \prod_{a=1}^{n} \int \frac{ds_a^{\mu}}{\sqrt{2\pi}} \exp\left[-\frac{1}{2} \sum_{a,b=1}^{n} s_a^{\mu}s_b^{\mu} \left(\delta_{ab} + \beta \sum_{i=1}^{N} \frac{(v_i^{\mu} - (m_i^{\mu})^2)}{N} x_{a,i}x_{b,i}\right)\right.
$$

$$
\left.+ \sum_{a=1}^{n} is_a^{\mu}\sqrt{\frac{\beta}{N}} \sum_{i=1}^{N} m_i^{\mu} x_{a,i}\right] . \tag{6.23}
$$

To continue, we need to assume condensation on a finite number of patterns that are not the condensed ones, say the first k without loss of generality. We focus now on the remaining ones, those with $\mu > k$:

$$B := \exp\left[\frac{\beta N}{2} \sum_{a=1}^{n} \sum_{\mu=k+1}^{P} (m^\mu(\mathbf{x}_a))^2\right] =$$

$$= \int \prod_{\mu=k+1}^{P} \prod_{a=1}^{n} \frac{dz_a^\mu}{\sqrt{2\pi}} \exp\left[-\sum_{a=1}^{n}\sum_{\mu=k+1}^{P} \frac{(z_a^\mu)^2}{2} + \sqrt{\frac{\beta}{N}} \sum_{a=1}^{n}\sum_{\mu=k+1}^{P} z_a^\mu \sum_{i=1}^{N} x_{a,i}\xi_i^\mu\right]. \quad (6.24)$$

Putting A and B together, their overall average over $(\boldsymbol{\xi}^\mu)_{\mu>k}$ takes the form

$$\mathbb{E}_{(\boldsymbol{\xi}^\mu)_{\mu>k}}[AB] = \int \prod_{\mu=P(1-t)+1}^{P}\prod_{a=1}^{n}\frac{ds_a^\mu}{\sqrt{2\pi}} \int \prod_{\mu=k+1}^{P}\prod_{a=1}^{n}\frac{dz_a^\mu}{\sqrt{2\pi}} e^{-\frac{1}{2}\sum_{a=1}^{n}\left(\sum_{\mu=P(1-t)+1}^{P}\frac{(s_a^\mu)^2}{2}+\sum_{\mu=k+1}^{P}\frac{(z_a^\mu)^2}{2}\right)}$$

$$\exp\left[\sum_{i=1}^{N}\sum_{\mu=k+1}^{P} \log \mathbb{E}_{\xi_i^\mu} e^{\sqrt{\frac{\beta}{N}}\sum_{a=1}^{n} x_{a,i}(\xi_i^\mu z_a^\mu + i\chi(\mu>P-R)m_i^\mu s_a^\mu) - \chi(\mu>P-R)\sum_{a,b=1}^{n} s_a^\mu s_b^\mu \sum_{i=1}^{N}\frac{\beta(v_i^\mu - (m_i^\mu)^2)x_{a,i}x_{b,i}}{2N}}\right]. \quad (6.25)$$

If we call $\mathbb{E}_{\boldsymbol{\xi}} m_i^{\mu\,2} =: \bar{M}^{\mu\,2}$, a further expansion of the exponential yields:

$$\mathbb{E}_{(\boldsymbol{\xi}^\mu)_{\mu>k}}[AB] = \int \prod_{\mu=P(1-t)+1}^{P}\prod_{a=1}^{n}\frac{ds_a^\rho}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\sum_{\mu=P(1-t)+1}^{P} \mathbf{s}^\mu \cdot \left(\mathbb{1} + \beta(1 - \bar{M}^{\mu\,2})Q\right) \mathbf{s}^\mu\right]$$

$$\int \prod_{\mu=k+1}^{P}\prod_{a=1}^{n}\frac{dz_a^\mu}{\sqrt{2\pi}} \exp\left\{-\sum_{\mu=k+1}^{P}\sum_{a=1}^{n}\frac{(z_a^\mu)^2}{2} + \frac{\beta}{2}\sum_{\mu=k+1}^{P}\sum_{a,b=1}^{n} z_a^\mu z_b^\mu Q(\mathbf{x}_a,\mathbf{x}_b) + \right. \quad (6.26)$$

$$\left. + i\beta \sum_{\mu=P(1-t)+1}^{P} \mathbb{E}_{\boldsymbol{\xi}}[\xi_1^\mu m_1^\mu] \sum_{a,b=1}^{n} z_a^\mu s_b^\mu Q(\mathbf{x}_a,\mathbf{x}_b) - \frac{\beta}{\Delta}\sum_{\mu=P(1-t)+1}^{P}\sum_{a,b=1}^{n}(\bar{M}^\mu)^2 s_a^\mu s_b^\mu Q(\mathbf{x}_a,\mathbf{x}_b)\right\}$$

Notice that $\bar{M}^\mu$ disappears from the computation. We can now perform a Gaussian integration over the variables $\mathbf{z}^\mu = (z_a^\mu)_{a\leq n}$:

$$\mathbb{E}_{(\boldsymbol{\xi}^\mu)_{\mu>k}}[AB] = \int \prod_{\mu=P(1-t)+1}^{P}\prod_{a=1}^{n}\frac{ds_a^\rho}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\sum_{\mu=P(1-t)+1}^{P} \mathbf{s}^\mu \cdot \left(\mathbb{1} + \beta Q + \beta^2 Q\frac{\mathbb{E}_{\boldsymbol{\xi}}^2[\xi_1^\mu m_1^\mu]}{\mathbb{1} - \beta Q}Q\right) \mathbf{s}^\mu\right]$$

$$\times \exp\left[-\frac{\alpha N}{2}\log\det\left(\mathbb{1} - \beta Q\right)\right]. \quad (6.27)$$

Finally, after an integration over the remaining Gaussian variables $\mathbf{s}^\mu$ we get

$$
\mathbb{E}_{(\boldsymbol{\xi}^\mu)_{\mu>k}}[AB] = \exp\left[-\frac{\alpha(1-t)N}{2}\log\det\left(\mathbb{1} - \beta Q\right)\right.
$$
$$
\left. -\frac{1}{2}\sum_{\mu=P(1-t)+1}^{P} \log\det\left(\mathbb{1} - (1 - \left(1 - \bar\epsilon\left(1 - \frac{(\mu-1)}{P}\right)\right)^2)\beta^2 Q^2\right)\right]. \quad (6.28)
$$

It remains to analyze the contribution given by $(\boldsymbol{\xi})_{\mu\leq k}$:

$$
C := \exp\left[\frac{\beta N}{2}\sum_{a=1}^{n}\sum_{\mu=1}^{k}(m^\mu(\mathbf{x}_a))^2\right] =
$$
$$
= \int \prod_{a=1}^{n}\prod_{\mu=1}^{k} dm_a^\mu\sqrt{\frac{\beta N}{2\pi}}\exp\left[\sum_{a=1}^{n}\sum_{\mu=1}^{k}\left(-N\beta\frac{(m_a^\mu)^2}{2} + \beta m_a^\mu\sum_{i=1}^{N}\xi_i^\mu x_{a,i}\right)\right]. \quad (6.29)
$$

Before plugging the contributions coming from A, B and C into $\mathbb{E}\mathcal{Z}_N^n$ we need to introduce a collection of Dirac deltas to fix the desired order parameters, that are organized in the overlap matrix $(Q(\mathbf{x}_a, \mathbf{x}_b))_{a,b=1}^n$:

$$
1 = \int \prod_{a\leq b\leq n} dq_{ab}\delta(Q(\mathbf{x}_a, \mathbf{x}_b) - q_{ab}) = \int \prod_{a\leq b\leq n} \frac{N dr_{ab}dq_{ab}}{4\pi i}\exp\left[-\frac{1}{2}\sum_{a,b=1}^{n}r_{ab}\left(Nq_{ab} - \sum_i x_{a,i}x_{b,i}\right)\right].
$$
$$
(6.30)
$$

Hence, the averaged replicated partition function, at leading exponential order in N, takes the form

$$
\mathbb{E}\mathcal{Z}_N^n = \int \prod_{a\leq b\leq n}\frac{N dr_{ab}dq_{ab}}{4\pi i}\int\prod_{a=1}^{n}\prod_{\mu=1}^{k} dm_a^\mu\sqrt{\frac{N\beta}{2\pi}}\exp\left[-\frac{N}{2}\sum_{a,b}r_{ab}q_{ab} - \frac{\beta N}{2}\sum_{a=1}^{n}\sum_{\mu=1}^{k}(m_a^\mu)^2\right]
$$
$$
\times \exp\left[-\frac{1}{2}\sum_{\mu=P(1-t)+1}^{P}\log\det\left(\mathbb{1} - (1 - \left(1 - \bar\epsilon\left(1 - \frac{(\mu-1)}{P}\right)\right)^2)\beta^2 Q^2\right)\right]
$$
$$
\times \exp\left[-\frac{\alpha(1-t)N}{2}\log\det\left(\mathbb{1} - \beta Q\right) + N\beta^2\Delta\sum_{a\neq b,1}^{n}\frac{q_{ab}^2}{4} + N\beta\sum_{a=1}^{n}\left(\frac{\beta\hat v}{2}(1 - q_{aa}) + \frac{\beta\Delta - 1}{4}q_{aa}^2\right)\right]
$$
$$
\times \prod_{i=1}^{N}\int\prod_{\mu=1}^{k} dP_\xi(\xi_i^\mu)\prod_{a=1}^{n} dP_\xi(x_{a,i})\exp\left[\frac{1}{2}\sum_{a,b=1}^{n}r_{ab}x_{a,i}x_{b,i} + \beta\sum_{\mu=1}^{k}\sum_{a=1}^{n}m_a^\mu\xi_i^\mu x_{a,i}\right].
$$
$$
(6.31)
$$

Notice that in the last line there is the product of N identical integrals (indexed by i). With a little abuse of notation, since the overlaps have been fixed by the deltas, we denote $Q = (q_{ab})_{a,b=1}^n$. We can finally express the replicated free entropy with a variational principle coming from a saddle point argument applied to the formula above:

$$
\bar{p}_n := \lim_{N\to\infty} \bar{p}_{N,n} = \frac{1}{n}\mathrm{Extr}\Big\{ -\frac{1}{2}\sum_{a,b} r_{ab}q_{ab} - \frac{\beta}{2}\sum_{a=1}^n\sum_{\mu=1}^k (m_a^\mu)^2 - \frac{\alpha(1-t)N}{2}\log\det(\mathbb{1}-\beta Q)
$$

$$
+\beta\sum_{a=1}^n\Big(\frac{\hat{v}(1-q_{aa})}{2} + \frac{\beta\Delta-1}{4}q_{aa}^2\Big) - \frac{\alpha t}{2R}\sum_{\mu=P(1-t)+1}^P \log\det\Big(\mathbb{1}-(1-\Big(1-\bar{\epsilon}\Big(1-\frac{(\mu-1)}{P}\Big)\Big))^2)\beta^2 Q\Big)
$$

$$
+\beta^2\Delta\sum_{a\neq b,1}^n \frac{q_{ab}^2}{4} + \log\int\prod_{\mu=1}^k\mathbb{E}_{\xi^\mu}\int\prod_{a=1}^n dP_\xi(x_a)\exp\Big[\frac{1}{2}\sum_{a,b=1}^n r_{ab}x_a x_b + \beta\sum_{\mu=1}^k\sum_{a=1}^n m_a^\mu\xi^\mu x_a\Big]\Big\}.
$$

$$(6.32)$$

The normalized sum over $\mu = P(1-t)+1,\ldots,P$ on the second line can be turned into an integral in the large N limit as follows:

$$
\bar{p}_n := \lim_{N\to\infty} \bar{p}_{N,n} = \frac{1}{n}\mathrm{Extr}\Big\{ -\frac{1}{2}\sum_{a,b} r_{ab}q_{ab} - \frac{\beta}{2}\sum_{a=1}^n\sum_{\mu=1}^k (m_a^\mu)^2 - \frac{\alpha(1-t)N}{2}\log\det(\mathbb{1}-\beta Q)
$$

$$
+\beta\sum_{a=1}^n\Big(\frac{\hat{v}(1-q_{aa})}{2} + \frac{\beta\Delta-1}{4}q_{aa}^2\Big) - \frac{\alpha t}{2}\int_0^t d\tau\log\det\big(\mathbb{1}-(1-(1-\bar{\epsilon}(\tau))^2)\beta^2 Q^2\big)
$$

$$
+\beta^2\Delta\sum_{a\neq b,1}^n \frac{q_{ab}^2}{4} + \log\int\prod_{\mu=1}^k\mathbb{E}_{\xi^\mu}\int\prod_{a=1}^n dP_\xi(x_a)\exp\Big[\frac{1}{2}\sum_{a,b=1}^n r_{ab}x_a x_b + \beta\sum_{\mu=1}^k\sum_{a=1}^n m_a^\mu\xi^\mu x_a\Big]\Big\}.
$$

$$(6.33)$$

The extremization is taken w.r.t. the collection of parameters $(r_{ab}, q_{ab})_{a,b=1}^n$, $(m_a^\mu)_{a=1,\mu=1}^{n,k}$. Within the replica symmetric ansatz

$$
\begin{cases} r_{ab}=r, & a\neq b \\ r_{aa}=-u \end{cases}\qquad \begin{cases} q_{ab}=q, & a\neq b \\ q_{aa}=v \end{cases}\qquad m_a^\mu = m^\mu, \tag{6.34}
$$

the matrix Q assumes the form

$$
Q = \begin{pmatrix} v & q & q & \cdots & q \\ q & v & q & \cdots & q \\ q & q & v & \cdots & q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ q & q & q & \cdots & v \end{pmatrix} \in \mathbb{R}^{n\times n}, \tag{6.35}
$$

whose eigenvalues can be easily found by Gauss' reduction to triangular form:

$$\text{Eig}Q = \text{diag}\left(\underbrace{v - q, v - q, \ldots, v - q}_{n-1}, v + (n-1)q\right).$$ (6.36)

from which it follows that

$$\det\left(\mathbb{1} - \beta Q\right) = (1 - \beta(v - q))^n \left[1 - n\frac{\beta q}{1 - \beta(v - q)}\right]$$ (6.37)

$$\det\left(\mathbb{1} - (1 - (1 - \bar{\epsilon}(\tau))^2)\beta^2 Q^2\right) = \left[1 - (1 - (1 - \bar{\epsilon}(\tau))^2)\beta^2(v - q)^2\right]^{n-1}$$
$$\times \left[1 - (1 - (1 - \bar{\epsilon}(\tau))^2)\beta^2\left(v + (n-1)q\right)^2\right].$$ (6.38)

Further simplifications occur for the other terms in the replicated free entropy. In particular the remaining log integral has to be treated as follows:

$$\int \prod_{\mu=1}^k \mathbb{E}_{\xi^\mu} \int \prod_{a=1}^n dP_\xi(x_a) \exp\left[\frac{r}{2}\sum_{a\neq b,1}^n x_a x_b - \frac{u}{2}\sum_{a=1}^n x_a^2 + \beta\sum_{\mu=1}^k m^\mu \xi^\mu \sum_{a=1}^n x_a\right] =$$

$$= \int \prod_{\mu=1}^k \mathbb{E}_{\xi^\mu} \int \prod_{a=1}^n dP_\xi(x_a) \exp\left[\frac{r}{2}\left(\sum_{a=1}^n x_a\right)^2 - \frac{u+r}{2}\sum_{a=1}^n x_a^2 + \beta\sum_{\mu=1}^k m^\mu \xi^\mu \sum_{a=1}^n x_a\right] =$$

$$= \mathbb{E}_Z \int \prod_{\mu=1}^k \mathbb{E}_{\xi^\mu} \prod_{a=1}^n \int dP_\xi(x_a) \exp\left[\sqrt{r}Z x_a - \frac{u+r}{2}x_a^2 + \beta\sum_{\mu=1}^k m^\mu \xi^\mu x_a\right] =$$

$$= \mathbb{E}_Z \mathbb{E}_{\boldsymbol{\xi}}\left[\int dP_\xi(x) \exp\left((Z\sqrt{r} + \beta \mathbf{m} \cdot \boldsymbol{\xi})x - \frac{u+r}{2}x^2\right)\right]^n$$ (6.39)

where $Z \sim \mathcal{N}(0,1)$, $\boldsymbol{\xi} = (\xi^1, \ldots, \xi^k)$, $\mathbf{m} = (m^1, \ldots, m^k)$. Finally, expanding at first order in n one has:

$$\bar{p}_n := \text{Extr}\left\{\frac{rq + uv}{2} - \beta\sum_{\mu=1}^k \frac{(m^\mu)^2}{2} - \frac{\beta^2\Delta q^2}{4} - \frac{\alpha(1-t)}{2}\left[\log\left(1 - \beta(v - q)\right) - \frac{\beta q}{1 - \beta(v - q)}\right]\right.$$

$$-\frac{\alpha t}{2}\int_0^t d\tau\left[\log\left(1 - (1 - (1 - \bar{\epsilon}(\tau))^2)\beta^2(v - q)^2\right) - \frac{2\beta^2 q(v - q)(1 - (1 - \bar{\epsilon}(\tau))^2)}{1 - \beta^2(1 - (1 - \bar{\epsilon}(\tau))^2)(v - q)^2}\right]$$

$$\left. + \beta\left(\frac{\hat{v}(1 - v)}{2} + \frac{\beta\Delta - 1}{4}v^2\right) + \mathbb{E}_{Z,\boldsymbol{\xi}}\log\int dP_\xi(x)\exp\left((Z\sqrt{r} + \beta\mathbf{m}\cdot\boldsymbol{\xi})x - \frac{u+r}{2}x^2\right)\right\} + O(n).$$ (6.40)

Hence, finally

$$\bar{p} := \lim_{n \to 0} \bar{p}_n = \text{Extr}\Big\{ \frac{rq + uv}{2} - \beta \sum_{\mu=1}^{k} \frac{(m^\mu)^2}{2} - \frac{\alpha(1-t)}{2} \Big[\log\left(1 - \beta(v-q)\right) - \frac{\beta q}{1 - \beta(v-q)} \Big]$$

$$- \frac{\beta^2 \Delta q^2}{4} - \frac{\alpha t}{2} \int_0^t d\tau \Big[\log\left(1 - (1 - (1 - \bar{\epsilon}(\tau))^2)\beta^2 (v-q)^2\right) - \frac{2\beta^2 q(v-q)(1 - (1 - \bar{\epsilon}(\tau))^2)}{1 - \beta^2(1 - (1 - \bar{\epsilon}(\tau))^2)(v-q)^2} \Big]$$

$$+ \beta\Big(\frac{\hat{v}(1-v)}{2} + \frac{\beta \Delta - 1}{4} v^2 \Big) + \mathbb{E}_{Z,\boldsymbol{\xi}} \log \int dP_\xi(x) \exp\Big((Z\sqrt{r} + \beta \mathbf{m} \cdot \boldsymbol{\xi}) x - \frac{u+r}{2} x^2 \Big) \Big\},$$

$$(6.41)$$

where extremization is intended over r, q, u, v, $\hat{v}$ and $(m^\mu)^{\mu \le k}$.

6.2.1 Fixed point equations

In order to have the fixed point equations in compact form define the following local random measure

$$\langle \cdot \rangle_{\boldsymbol{\xi},Z} = \frac{\int dP_\xi(x) e^{(Z\sqrt{r} + \beta \mathbf{m} \cdot \boldsymbol{\xi})x - \frac{r+u}{2}x^2}(\cdot)}{\int dP_\xi(x) e^{(Z\sqrt{r} + \beta \mathbf{m} \cdot \boldsymbol{\xi})x - \frac{r+u}{2}x^2}} \tag{6.42}$$

Then the stationarity conditions coming from (6.41) are

$$v = 1 \tag{6.43}$$
$$v = \mathbb{E}\langle X^2 \rangle_{\boldsymbol{\xi},Z} \tag{6.44}$$
$$m^\mu = \mathbb{E}\xi^\mu \langle X \rangle_{\boldsymbol{\xi},Z} \tag{6.45}$$
$$q = \mathbb{E}\langle X \rangle_{\boldsymbol{\xi},Z}^2 \tag{6.46}$$

$$r = \frac{\alpha(1-t)\beta^2 q}{(1 - \beta(v-q))^2} + \beta^2 \Delta q$$

$$+ \alpha t \int_0^t d\tau \frac{2q\beta^2(1 - (1 - \bar{\epsilon}(\tau))^2)}{1 - \beta^2(1 - (1 - \bar{\epsilon}(\tau))^2)(v-q)^2} \Big[1 + \frac{2\beta^2(v-q)^2(1 - (1 - \bar{\epsilon}(\tau))^2)}{1 - \beta^2(1 - (1 - \bar{\epsilon}(\tau))^2)(v-q)^2} \Big] \tag{6.47}$$

$$u = \beta\hat{v} + \beta(1 - \beta\Delta)v - \alpha(1-t)\beta \frac{1 - \beta(v - 2q)}{(1 - \beta^2(v-q))^2}$$

$$- \alpha t \int_0^t d\tau \Big[\frac{2v\beta^2(1 - (1 - \bar{\epsilon}(\tau))^2)}{1 - \beta^2(1 - (1 - \bar{\epsilon}(\tau))^2)(v-q)^2} + q \frac{4\beta^4(v-q)^2(1 - (1 - \bar{\epsilon}(\tau))^2)^2}{(1 - \beta^2(1 - (1 - \bar{\epsilon}(\tau))^2)(v-q)^2)^2} \Big]. \tag{6.48}$$

The first one in particular can be directly eliminated. Notice that the effect of decimation is visible only in the variables u and r that affect the local measure (6.42).

The value of the Mattis magnetization m^μ depends on the errors $\bar\epsilon$ made on the pattern reconstruction, namely on the magnetizations, corresponding to the previous steps of decimation. The representation of their contributions in integral form holds only in the thermodynamic limit, for it is a limit of a Riemann sum. For all practical purposes, we will make finite size simulations and use the discretized form present in (6.32), starting from step 0, when no pattern has been retrieved yet.

Remark 6.2 (Assumptions on decimation). If the Statistician builds her estimates starting from the posterior Boltzmann-Gibbs distribution (6.5), then the assumptions we have made on the $\boldsymbol{\eta}^\mu$'s mean that each site i is asymptotically decoupled from the others and interacts only with a random *mean field*, composed by a Gaussian contribution, whose variance is tuned by r, and a magnetic field produced by the other (possibly) magnetized sites. Notice that r comprises all the three noise sources as discussed in Remark 6.1 and treats them as independent (the global variance is the sum of the three contributions). Hence, asymptotically it is reasonable that the sampling from (6.5) is equivalent to the one from a local measure like (6.42), from which our assumptions on the $\boldsymbol{\eta}^\mu$'s easily follow. Furthermore, in the case of one pattern condensation $k = 1$, we are now able to relate $\bar\epsilon$ to the Mattis magnetization m found at the $R + 1$-th step:

$$\bar\epsilon((R+1)/P) = 1 - m. \tag{6.49}$$

Therefore the magnetization and $\bar\epsilon(t)$ depend on the history of the decimation process. Once a magnetization is found, it can be used to perform the step $R + 2$, and so on, till the patterns are all retrieved with a certain accuracy. Hence, starting from the 0-th step, we can study a flow in t (that becomes ideally continuous when $N \to \infty$) of the magnetization values.

6.2.2 Sanity checks

First off, we need to prove that in case there is only one pattern $P = 1$, and therefore $\alpha = 0$, our computation coincides with the Bayes-optimal one ($\beta = 1/\Delta$). Let us set ourselves on the 0-th step. In that case one has

$$r = \frac{q}{\Delta}, \quad u = \hat v, \tag{6.50}$$

and the free entropy can be simplified to

$$\bar p = \mathrm{Extr}\left\{ \frac{q^2}{4\Delta} - \frac{m^2}{2\Delta} + \frac{\hat v}{2\Delta} + \mathbb{E}_{Z,\xi} \log \int dP_\xi(x) \exp\left(\left(Z\sqrt{\frac{q}{\Delta}} + \frac{m}{\Delta}\xi \right) x - \left(\frac{q + \hat v}{\Delta} \right) \frac{x^2}{2} \right) \right\}. \tag{6.51}$$

The local measure becomes

$$\langle \cdot \rangle_{Z,\xi} = \frac{\int dP_\xi(x) \exp\left(\left(Z\sqrt{\frac{q}{\Delta}} + \frac{m}{\Delta}\xi \right) x - \left(\frac{q+\hat v}{\Delta} \right) \frac{x^2}{2} \right) (\cdot)}{\int dP_\xi(x) \exp\left(\left(Z\sqrt{\frac{q}{\Delta}} + \frac{m}{\Delta}\xi \right) x - \left(\frac{q+\hat v}{\Delta} \right) \frac{x^2}{2} \right)}. \tag{6.52}$$

It is easy to realize now that there is a consistent solution to the simplified system of fixed point equations with the previous local measure:

$$q = m = \mathbb{E}\xi\langle X\rangle_{Z,\xi}\,, \quad \hat{v} = 0\,, \tag{6.53}$$

that corresponds to enforce the Nishimori identities. Notice that this would hold also for any finite rank case $P = K$ with one pattern condensation. We stress that the effect of decimation is sensible only in the extensive rank case as it is clear from (6.41) and the consequent fixed point equations. The reason is that, if we subtract a finite number of p-like terms in (6.10), we are penalizing a finite number of directions in a space of dimension growing to infinity. This is not enough to create an additional noise contribution, since the system can easily find other favoured directions to termalize in. In other words, the $p^\mu(\mathbf{x})$'s give a sub-extensive contribution that can be neglected in the finite rank case.

Another really important thing to check, in order to have a well defined zero temperature limit $\beta \to +\infty$, is that $u + r = O(\beta)$, which is not trivial given the $O(\beta^2)$ contributions in (6.47)-(6.48). Summing them together we see that

$$r + u = \beta\hat{v} - \beta^2\Delta(1 - q) + \beta - \frac{\alpha(1 - t)\beta}{1 - \beta(1 - q)} - \alpha t\int_0^t d\tau\left[\frac{2\beta^2(1 - q)(1 - (1 - \bar{\epsilon}(\tau))^2)}{1 - \beta^2(1 - (1 - \bar{\epsilon}(\tau))^2)(1 - q)^2}\right] \tag{6.54}$$

where we imposed $v = 1$. From the expression of the free entropy (6.41) we see that we can only accept solutions for q such that $\beta(1 - q) < 1$, otherwise the logarithm gives rise to a singularity. Hence $C := \beta(1 - q) = O(1)$. Using this definition the previous formula becomes

$$r + u = \beta\hat{v} + \beta(1 - \Delta C) - \frac{\alpha(1 - t)\beta}{1 - C} - \alpha t\int_0^t d\tau\left[\frac{2\beta C(1 - (1 - \bar{\epsilon}(\tau))^2)}{1 - (1 - (1 - \bar{\epsilon}(\tau))^2)C^2}\right]\,. \tag{6.55}$$

From the fixed point equation (6.44) with $v = 1$, $\hat{v}$ must be at most of order $O(1)$, since it only has to compete with the other two terms in the previous equation.

Finally, as in the standard Hopfield model, when $P_\xi = \mathcal{N}(0, 1)$ the only magnetization value we can accept is zero. In fact, in that case we can integrate by parts in (6.45) obtaining:

$$m^\mu = \beta m^\mu\left(\mathbb{E}\langle X^2\rangle_{\boldsymbol{\xi},Z} - \mathbb{E}\langle X\rangle_{\boldsymbol{\xi},Z}^2\right) = m^\mu\beta(v - q)\,. \tag{6.56}$$

The latter implies wither $m = 0$ or

$$q = v - \frac{1}{\beta}\,, \tag{6.57}$$

which is not possible, since it would make the log term in the first line of (6.41) explode. The only case when $m \neq 0$ is acceptable is $\alpha = 0$, *i.e.* for finit rank. Indeed, using (6.44) and choosing $\beta = 1/\Delta$, we recover the PCA overlap $q = 1 - \Delta$ [37, 38].

6.2.3 The zero temperature limit

In this section we study the $\beta \to \infty$ limit of the free entropy with a prior in the form

$$P_\xi = (1 - \rho)\delta_0 + \frac{\rho}{2} \left[\delta_{-1/\sqrt{\rho}} + \delta_{1/\sqrt{\rho}} \right], \quad \rho \in (0, 1]. \tag{6.58}$$

It can be easily verified that P_ξ is symmetric, and thus centered, and has unit variance as required by our analysis. ρ is a parameter measuring sparsity in the ξ's. If $\rho = 1$ then we recover the Ising case, sparsity being totally absent.

For future convenience we introduce the notations

$$C := \beta(1 - q) \in [0, 1), \tag{6.59}$$

$$\bar{r} := r/\beta^2 \tag{6.60}$$

$$U := \frac{u + r}{\beta} \tag{6.61}$$

where q is intended as the stationary value of the overlap solving the fixed point equations. Denote $\mathbf{m} = (m^\mu)_{\mu=1}^k$, where k is the maximum number of condensed patterns. In the low temperature limit the free entropy, re-scaled by β, and evaluated at the stationary values of the parameters involved has the form

$$\frac{1}{\beta}\bar{p} = -\frac{\bar{r}C}{2} + \frac{U}{2} + \frac{\alpha(1 - t)}{2(1 - C)} - \frac{1}{4} - \frac{\mathbf{m}^2}{2} + \frac{\Delta C}{2} + \Gamma + \alpha t \int_0^t d\tau \frac{C(1 - (1 - \bar{\epsilon}(\tau))^2)}{1 - (1 - (1 - \bar{\epsilon}(\tau))^2)C^2} \tag{6.62}$$

where

$$\Gamma = \frac{1}{\beta}\mathbb{E}_{\boldsymbol{\xi},Z} \log \int dP_\xi(x) \exp \left[\beta \left(Z\sqrt{\bar{r}} + \mathbf{m} \cdot \boldsymbol{\xi} \right) x - \frac{\beta U}{2}x^2 \right] =$$

$$= \frac{1}{\beta}\mathbb{E}_{\boldsymbol{\xi},Z} \log \left[1 - \rho + \rho \cosh \frac{\beta}{\sqrt{\rho}} \left(Z\sqrt{\bar{r}} + \mathbf{m} \cdot \boldsymbol{\xi} \right) \exp \left(-\frac{\beta U}{2\rho} \right) \right]. \tag{6.63}$$

When $\beta \to \infty$ we have to distinguish two cases in the Z average:

$$\Gamma = O(\beta^{-1}) + \frac{1}{\beta}\mathbb{E}_{\boldsymbol{\xi}} \left(\int_{-\mathbf{m}\cdot\boldsymbol{\xi}/\sqrt{\bar{r}}+U/2\sqrt{\bar{r}\rho}}^{\infty} + \int_{-\infty}^{-\mathbf{m}\cdot\boldsymbol{\xi}/\sqrt{\bar{r}}-U/2\sqrt{\bar{r}\rho}} \right) \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \times$$

$$\times \log \left[1 - \rho + \rho \cosh \frac{\beta}{\sqrt{\rho}} \left(z\sqrt{\bar{r}} + \mathbf{m} \cdot \boldsymbol{\xi} \right) e^{-\frac{\beta U}{2\rho}} \right] \tag{6.64}$$

The $O(\beta^{-1})$ instead comes from the contribution:

$$\frac{1}{\beta} \int_{-\mathbf{m}\cdot\boldsymbol{\xi}/\sqrt{\bar{r}}-U/2\sqrt{\bar{r}\rho}}^{-\mathbf{m}\cdot\boldsymbol{\xi}/\sqrt{\bar{r}}+U/2\sqrt{\bar{r}\rho}} \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \log \left[1 - \rho + \rho \cosh \frac{\beta}{\sqrt{\rho}} \left(z\sqrt{\bar{r}} + \mathbf{m} \cdot \boldsymbol{\xi} \right) e^{-\frac{\beta U}{2\rho}} \right] \tag{6.65}$$

that can be bounded from below by

$$\frac{1}{\beta} \int_{-\mathbf{m}\cdot\boldsymbol{\xi}/\sqrt{\bar{r}}-U/2\sqrt{\bar{r}\rho}}^{-\mathbf{m}\cdot\boldsymbol{\xi}/\sqrt{\bar{r}}+U/2\sqrt{\bar{r}\rho}} \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \log\left[1-\rho\right] \tag{6.66}$$

and from above by 0. Hence, it vanishes linearly in β^{-1} as $\beta \to \infty$.

Let us now focus on the first integral in (6.64). The hyperbolic cosine and the exponential in U dominate on the other terms in the log. Taking into account the exponential growth in the selected range of z-values the first integral can be approximated with:

$$\mathbb{E}_{\boldsymbol{\xi}} \int_{-\mathbf{m}\cdot\boldsymbol{\xi}/\sqrt{\bar{r}}+U/2\sqrt{\bar{r}\rho}}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \left(\frac{Z\sqrt{\bar{r}}+\mathbf{m}\cdot\boldsymbol{\xi}}{\sqrt{\rho}} - \frac{U}{2\rho}\right) = \sqrt{\frac{\bar{r}}{2\pi\rho}} \mathbb{E}_{\boldsymbol{\xi}} e^{-\frac{1}{2\bar{r}}\left(\frac{U}{2\sqrt{\rho}}-\mathbf{m}\cdot\boldsymbol{\xi}\right)^2} +$$

$$+ \mathbb{E}_{\boldsymbol{\xi}} \left(\frac{\mathbf{m}\cdot\boldsymbol{\xi}}{\sqrt{\rho}} - \frac{U}{2\rho}\right) \int_{-\mathbf{m}\cdot\boldsymbol{\xi}/\sqrt{\bar{r}}+U/2\sqrt{\bar{r}\rho}}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} . \tag{6.67}$$

The second integral in (6.64) can be treated similarly. Putting all the terms together one gets

$$\frac{1}{\beta}\bar{p} = -\frac{\bar{r}C}{2} + \frac{\Delta C}{2} + \frac{U}{2} + \frac{\alpha(1-t)}{2(1-C)} - \frac{1}{4} - \frac{\mathbf{m}^2}{2} + \sqrt{\frac{\bar{r}}{2\pi\rho}}\mathbb{E}_{\boldsymbol{\xi}}\left[e^{-\frac{1}{2\bar{r}}\left(\frac{U}{2\sqrt{\rho}}-\mathbf{m}\cdot\boldsymbol{\xi}\right)^2} + e^{-\frac{1}{2\bar{r}}\left(\frac{U}{2\sqrt{\rho}}+\mathbf{m}\cdot\boldsymbol{\xi}\right)^2}\right] +$$

$$+ \mathbb{E}_{\boldsymbol{\xi}}\frac{\mathbf{m}\cdot\boldsymbol{\xi}}{2\sqrt{\rho}}\left[\mathrm{erf}\left(\frac{\mathbf{m}\cdot\boldsymbol{\xi}+\frac{U}{2\sqrt{\rho}}}{\sqrt{2\bar{r}}}\right) + \mathrm{erf}\left(\frac{\mathbf{m}\cdot\boldsymbol{\xi}-\frac{U}{2\sqrt{\rho}}}{\sqrt{2\bar{r}}}\right)\right]$$

$$- \frac{U}{4\rho}\mathbb{E}_{\boldsymbol{\xi}}\left[2 + \mathrm{erf}\left(\frac{\mathbf{m}\cdot\boldsymbol{\xi}-\frac{U}{2\sqrt{\rho}}}{\sqrt{2\bar{r}}}\right) - \mathrm{erf}\left(\frac{\mathbf{m}\cdot\boldsymbol{\xi}+\frac{U}{2\sqrt{\rho}}}{\sqrt{2\bar{r}}}\right)\right] + \alpha t \int_0^t d\tau \frac{C(1-(1-\bar{\epsilon}(\tau))^2)}{1-(1-(1-\bar{\epsilon}(\tau))^2)C^2} . \tag{6.68}$$

Considering the all the parameters are evaluated at their stationary values, the previous formula can be further simplified by looking at the limiting version of the fixed point equations. Starting from (6.46), (6.59) and (6.60) we see that

$$C = \beta(1-q) = \beta(\mathbb{E}\langle X^2 \rangle_{Z,\boldsymbol{\xi}} - \mathbb{E}\langle X \rangle_{Z,\boldsymbol{\xi}}^2) = \mathbb{E}\frac{\beta Z}{\sqrt{r}}\langle X \rangle_{Z,\boldsymbol{\xi}} = \mathbb{E}\frac{Z}{\sqrt{\bar{r}}}\langle X \rangle_{Z,\boldsymbol{\xi}} = 2\frac{\partial\Gamma}{\partial\bar{r}} \tag{6.69}$$

where we used $v=1$ and reversed Gaussian integration by parts. Hence:

$$C = \frac{1}{\sqrt{2\pi\rho\bar{r}}}\mathbb{E}_{\boldsymbol{\xi}}\left[\exp\left(-\left(\frac{U/2\sqrt{\rho}-\mathbf{m}\cdot\boldsymbol{\xi}}{\sqrt{2\bar{r}}}\right)^2\right) + \exp\left(-\left(\frac{U/2\sqrt{\rho}+\mathbf{m}\cdot\boldsymbol{\xi}}{\sqrt{2\bar{r}}}\right)^2\right)\right] . \tag{6.70}$$

The value of $\bar{r}$ can be found directly from (6.47) by multiplying it by β^{-2}:

$$\bar{r} = \frac{\alpha(1-t)}{(1-C)^2} + \Delta + \alpha t \int_0^t d\tau \frac{2(1-(1-\bar{\epsilon}(\tau))^2)}{1-(1-(1-\bar{\epsilon}(\tau))^2)C^2}\left[1 + \frac{2C^2(1-(1-\bar{\epsilon}(\tau))^2)}{1-(1-(1-\bar{\epsilon}(\tau))^2)C^2}\right] . \tag{6.71}$$

There are at least two ways to determine the value of U. Here we proceed from (6.44) with $v = 1$. In particular, from (6.63) it follows that

$$1 = \mathbb{E}\langle X^2 \rangle_{Z,\boldsymbol{\xi}} = -2\frac{\partial \Gamma}{\partial U} = \frac{1}{2\rho}\left[2 + \operatorname{erf}\left(\frac{\mathbf{m} \cdot \boldsymbol{\xi} - \frac{U}{2\sqrt{\rho}}}{\sqrt{2\bar{r}}}\right) - \operatorname{erf}\left(\frac{\mathbf{m} \cdot \boldsymbol{\xi} + \frac{U}{2\sqrt{\rho}}}{\sqrt{2\bar{r}}}\right)\right] \tag{6.72}$$

that is

$$\mathbb{E}_{\boldsymbol{\xi}}\left[\operatorname{erf}\left(\frac{U/2\sqrt{\rho} - \mathbf{m} \cdot \boldsymbol{\xi}}{\sqrt{2\bar{r}}}\right) + \operatorname{erf}\left(\frac{U/2\sqrt{\rho} + \mathbf{m} \cdot \boldsymbol{\xi}}{\sqrt{2\bar{r}}}\right)\right] = 2(1 - \rho). \tag{6.73}$$

We stress that the l.h.s. of the previous equation is monotonic in U, and thus (6.48) has a unique solution for fixed sparsity parameter ρ. Finally, from (6.45) and (6.63)

$$\mathbf{m} = \mathbb{E}\boldsymbol{\xi}\langle X \rangle_{Z,\boldsymbol{\xi}} = \frac{\partial \Gamma}{\partial \mathbf{m}} = \mathbb{E}_{\boldsymbol{\xi}}\frac{\boldsymbol{\xi}}{2\sqrt{\rho}}\left[\operatorname{erf}\left(\frac{\mathbf{m} \cdot \boldsymbol{\xi} - U/2\sqrt{\rho}}{\sqrt{2\bar{r}}}\right) + \operatorname{erf}\left(\frac{U/2\sqrt{\rho} + \mathbf{m} \cdot \boldsymbol{\xi}}{\sqrt{2\bar{r}}}\right)\right]. \tag{6.74}$$

If we insert these conditions in (6.68) we get

$$\frac{\bar{p}}{\beta} = \frac{\alpha(1 - t)}{2(1 - C)^2} + \Delta C - \frac{1}{4} + \frac{\mathbf{m}^2}{2} + 2\alpha t \int_0^t d\tau \frac{C(1 - (1 - \bar{\epsilon}(\tau))^2)}{(1 - (1 - (1 - \bar{\epsilon}(\tau))^2)C^2)^2}. \tag{6.75}$$

Notice that sparsity does not appear explicitly in $\bar{p}/\beta$, it affects its value only through $\mathbf{m}$ and C. If we wish to analyze the model in a Bayes-optimal fashion, meaning trying to recover finite rank results, it is sufficient to set $\Delta = 1/\beta \to 0$ in the low temperature limit, and the form of the previous expression reduces to the well known free entropy in the low temperature limit of the Hopfield model for $t = 0$, where again no sparsity appears.

6.3   0-th step phase diagrams

In order to be viable, decimation needs to start in a region of the α-Δ(-β) phase diagram of the 0-step model where pattern retrieval is possible, otherwise our estimates will be orthogonal to the ground truth we seek, and the subtraction of the corresponding rank one contribution only introduces noise. For this reason the 0-th step acquires a great importance in the analysis. For $t = 0$ the RS free entropy becomes

$$\bar{p} = \operatorname{Extr}\left\{\frac{rq + uv}{2} - \beta\sum_{\mu=1}^{k}\frac{(m^\mu)^2}{2} + \frac{\beta^2\Delta(v^2 - q^2)}{4} - \frac{\alpha}{2}\left[\log\left(1 - \beta(v - q)\right) - \frac{\beta q}{1 - \beta(v - q)}\right]\right.$$

$$\left. + \beta\left(\frac{\hat{v}(1 - v)}{2} - \frac{v^2}{4}\right) + \mathbb{E}_{Z,\boldsymbol{\xi}}\log\int dP_\xi(x)\exp\left(\left(Z\sqrt{r} + \beta\mathbf{m} \cdot \boldsymbol{\xi}\right)x - \frac{u + r}{2}x^2\right)\right\}. \tag{6.76}$$

and the fixed point equations for u and r simplify to

$$r = \frac{\alpha\beta^2 q}{(1 - \beta(1 - q))^2} + \beta^2 \Delta q \tag{6.77}$$

$$u = \beta\hat{v} + \beta(1 - \beta\Delta)v - \alpha\beta\frac{1 - \beta(1 - 2q)}{(1 - \beta(1 - q))^2} \,, \tag{6.78}$$

where we have already enforced the constraint $v = 1$.

Another important feature of these models to be understood is their robustness against the interference among patterns (and decimation noise after the 0-th step). This can be evaluated in the low temperature limit, that will tell us what is the maximum value of α, that measures indeed the interference, that we can bear still having a positive magnetization, and a corresponding reconstruction error smaller than one.

6.3.1   The pure Ising case

In this subsection we analyze the case when $\rho = 1$, that corresponds to the Rademacher prior $P_\xi = \frac{\delta_{-1} + \delta_1}{2}$. The local measure (6.42) simplifies to

$$\langle\cdot\rangle_{\boldsymbol{\xi},Z} = \sum_{x=\pm 1} \frac{e^{\left(Z\sqrt{r} + \beta\mathbf{m}\cdot\boldsymbol{\xi}\right)x}}{\cosh\left(Z\sqrt{r} + \beta\mathbf{m}\cdot\boldsymbol{\xi}\right)}(\cdot) \tag{6.79}$$

and the value of u becomes irrelevant. Furthermore $v = 1$ is verified by default, which means $\hat{v} = 0$. Hence we just have to analyze the three equations

$$m^\mu = \mathbb{E}\xi^\mu \tanh\left(Z\sqrt{r} + \beta\mathbf{m}\cdot\boldsymbol{\xi}\right) \tag{6.80}$$

$$q = \mathbb{E}\tanh^2\left(Z\sqrt{r} + \beta\mathbf{m}\cdot\boldsymbol{\xi}\right) \tag{6.81}$$

$$r = \beta^2 \Delta q + \frac{\alpha\beta^2 q}{(1 - \beta(1 - q))^2} \,. \tag{6.82}$$

The previous system of fixed point equations can be solved iterating (6.80) and solving (6.81) by dichotomy at each step. Notice that, a part from the noise contribution whose amplitude is governed by $\beta^2 \Delta q$ in (6.82), the equations have the same form of those of the Hopfield model. Then, as intuition would suggest, the phase diagram should be characterized by the same regions.

To start with, we choose $\beta = 1/\Delta$ and the related phase diagram is depicted in Figure 6.1. The blue curve represents the transition line from the spin glass phase above it, where the expected overlap $q > 0$ and $m = 0$, and the retrieval phase below it. Between the blue and the red curves there exists a metastable state with a non vanishing magnetization, although it is not dominant in probability, meaning that there is a solution to the fixed point equations with zero magnetization that has a bigger associated free entropy. Finally, below the red curve there
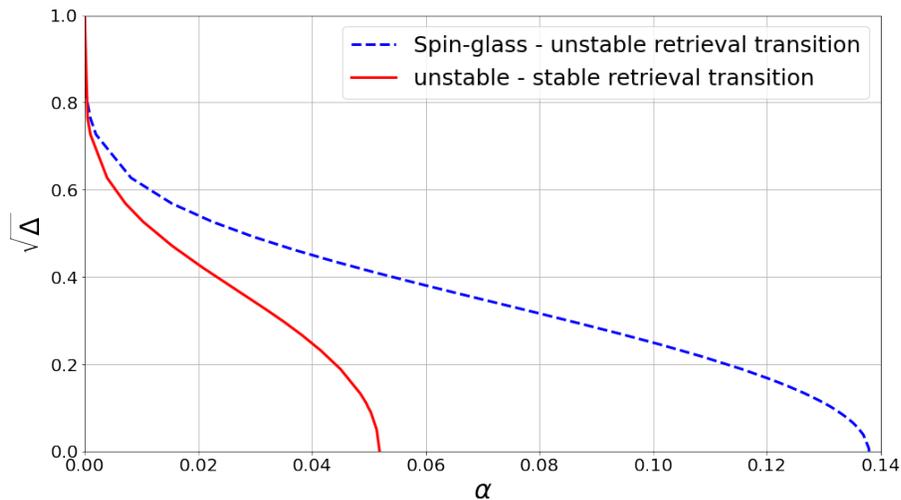
Figure 6.1: Phase diagram at $t = 0$, with $\rho = 1$ and $\beta = 1/\Delta$.

is a stable state with non vanishing magnetization. Strictly speaking, from a thermodynamic point of view, the blue line is not a true transition line, because the spin glass state remains stable. However, it is still possible that a sharp enough algorithm can find a configuration with non zero magnetization even right below the blue curve.

Concerning the low temperature limit, the r.h.s. of equation (6.73) vanishes when $\rho = 1$, and thanks to the monotonicity of the l.h.s. in U we know that the only solution is $U = 0$. Notice that if we select $\Delta = 1/\beta \to 0$ then the fixed point equations turn into those of the well-known low temperature limit of the Hopfield model:

$$C = \sqrt{\frac{2}{\pi \alpha}} (1 - C) \mathbb{E}_{\boldsymbol{\xi}} \exp \left[-\frac{(1 - C)^2}{2\alpha} (\mathbf{m} \cdot \boldsymbol{\xi})^2 \right] , \tag{6.83}$$

$$\mathbf{m} = \mathbb{E}_{\boldsymbol{\xi}} \boldsymbol{\xi} \operatorname{erf} \left(\frac{\mathbf{m} \cdot \boldsymbol{\xi}}{\sqrt{2\alpha}} (1 - C) \right) , \tag{6.84}$$

Hence, the model behaves exactly as the Hopfield model and shares its very same storage capacity $\alpha_c \simeq 0.138$, and value below which the patterns become pure states $\alpha_F \simeq 0.051$.

Another possibility is to study the model with Δ fixed and $\beta \to \infty$, namely to study the performance of the Ground State (GS) search. With reference to Figure 6.2, the dashed blue curve represents as before the "transition" from the spin glass phase to the retrieval one, and below the red curve the magnetized state dominates the probability. The curves touch the x-axis at the values α_c and α_F respectively, as expected. See the following section for an explanation of the other curves. We find another curious transition that occurs at $\alpha = 0$, $\Delta = \bar{\Delta}$, that is related to the performance of GS search algorithms. To determine $\bar{\Delta}$ we specialize our formulae
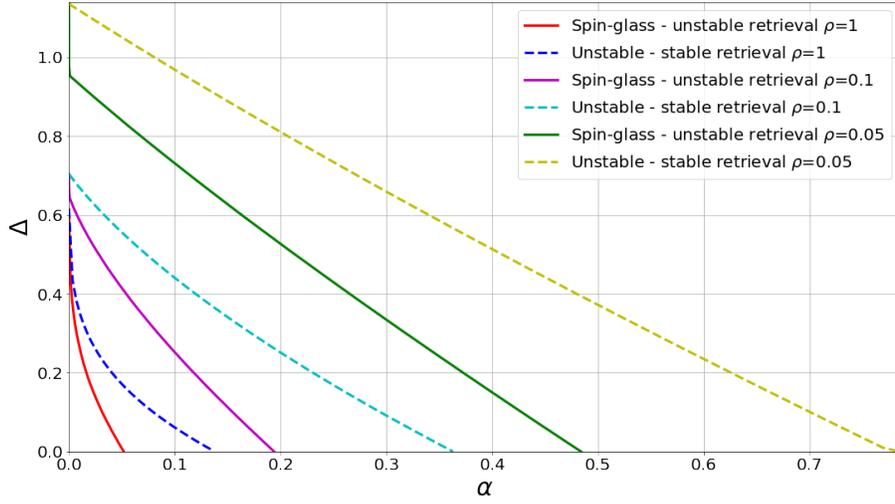
Figure 6.2: α-Δ phase diagram at zero temperature.

to the case of condensation onto a single pattern, $\rho = 1$ and $\alpha = 0$:

$$\bar{r} = \Delta \tag{6.85}$$

$$C = \sqrt{\frac{2}{\Delta\pi}} \exp\left(-\frac{m^2}{2\Delta}\right) \tag{6.86}$$

$$m = \mathrm{erf}\left(\frac{m}{\sqrt{2\Delta}}\right), \tag{6.87}$$

from which we see that m has decoupled from C. The equation for m admits a solution only when the r.h.s. has a derivative greater than 1 w.r.t. m at $m = 0$, namely

$$\frac{2}{\pi\Delta} \geq 1 \quad \Rightarrow \quad \Delta \leq \bar{\Delta} = \frac{2}{\pi}. \tag{6.88}$$

This transition happens below the Bayes-optimal threshold $\Delta_{BO} = 1$, that we have for $\alpha = 0$ in Figure 6.1. As we shall see later, even if it is clear that the presence of additional noise Δ here compromises the stability of the magnetized state, the GS search has a better performance than the choice $\beta = 1/\Delta$.

6.3.2 Sparse prior

When $\rho < 1$, the equations modify non trivially, because the constraint $v = 1$ is no longer automatically satisfied, and it is enforced by a non zero value of $\hat{v}$. The first and second

moment of the local measure become

$$\langle X \rangle_{\boldsymbol{\xi},Z} = \frac{\sqrt{\rho} \sinh\left(Z\sqrt{\frac{r}{\rho}} + \frac{\mathbf{m}\cdot\boldsymbol{\xi}}{\Delta\sqrt{\rho}}\right) e^{-\frac{r+u}{2\rho}}}{1 - \rho + \rho\cosh\left(Z\sqrt{\frac{r}{\rho}} + \frac{\mathbf{m}\cdot\boldsymbol{\xi}}{\Delta\sqrt{\rho}}\right) e^{-\frac{r+u}{2\rho}}} \tag{6.89}$$

$$\langle X^2 \rangle_{\boldsymbol{\xi},Z} = \frac{\cosh\left(Z\sqrt{\frac{r}{\rho}} + \frac{\mathbf{m}\cdot\boldsymbol{\xi}}{\Delta\sqrt{\rho}}\right) e^{-\frac{r+u}{2\rho}}}{1 - \rho + \rho\cosh\left(Z\sqrt{\frac{r}{\rho}} + \frac{\mathbf{m}\cdot\boldsymbol{\xi}}{\Delta\sqrt{\rho}}\right) e^{-\frac{r+u}{2\rho}}} . \tag{6.90}$$

Now all the equations (6.44)-(6.48) come into play. The entire system of fixed point equations can be solved iterating (6.45) and solving exactly and simultaneously (6.44), (6.46) at each iteration step. By doing so, one would obtain a picture similar to that in Figure 6.1, with α_c and α_F depending on ρ. Using the fixed point equations from the previous subsection we draw a phase diagram in the α-Δ plane with $\beta \to \infty$ as in Figure 6.2. The cyan dashed line for instance is the transition from the spin glass phase to the retrieval phase when $\rho = 0.1$, whereas the solid magenta line stands for the transition to the true ferromagnetic phase. The values of α_c and α_F have shifted sensibly in the positive direction, signalling that sparsity increases robustness against patterns interference. In Figure 6.2 we observe, as for the Ising case, transitions in Δ when $\alpha = 0$ that are below the optimal threshold that up to now we know only numerically. Notice that, contrary to what happens at $\rho = 1$, the dashed lines and the corresponding solid lines do not intersect the y-axis in the same point. Though it might look unusual at first sight, we recall that sparsity induces similar gaps even in the low rank case [32, 47, 88, 176], where a solution with non-vanishing magnetization can appear even before the information theoretic threshold. Here the theoretic thresholds of the decimation procedure are represented by the solid lines.
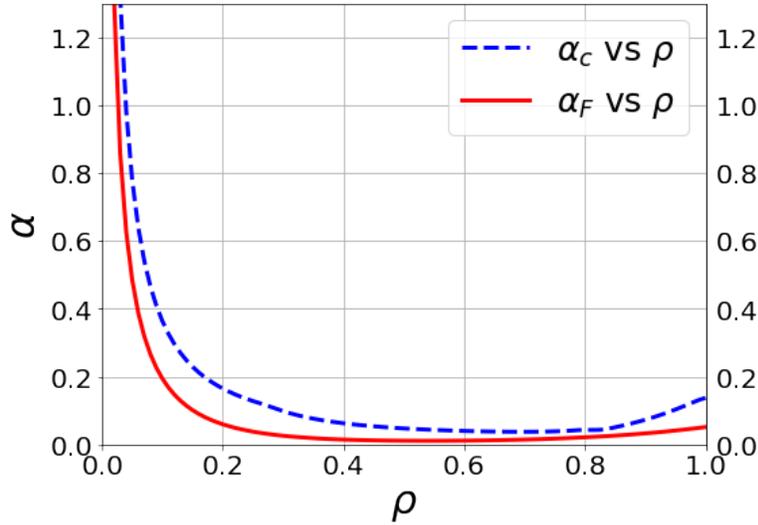
Figure 6.3 describes α_c and α_F as functions of ρ, that are the intersections of the dashed and solid lines in Figure 6.2 with the x-axis ($\Delta = 0$). Looking at Figure 6.3 we realize that α_c and α_F can both become large, exceeding the standard limits of the Hopfield model, when ρ approaches zero, confirming that a strong sparsity can increase the storage capacity.

Let us focus on the limit $\rho \to 0$ in the hypothesis of condensation only onto one pattern. Fix $\alpha > 0$. First, we need to expand the fixed point equations using the sparse Ising prior on ξ. Equation (6.73) becomes for instance

$$2(1 - \rho)\text{erf}\left(\frac{U}{2\sqrt{2\bar{r}\rho}}\right) + \rho\left[\text{erf}\left(\frac{U/2 - m}{\sqrt{2\bar{r}\rho}}\right) + \text{erf}\left(\frac{U/2 + m}{\sqrt{2\bar{r}\rho}}\right)\right] = 2(1 - \rho). \tag{6.91}$$

The equation for C is

$$C = \frac{2(1 - \rho)}{\sqrt{2\pi\bar{r}\rho}} e^{-\frac{U^2}{8\bar{r}\rho}} + \frac{\rho}{\sqrt{2\pi\bar{r}\rho}}\left[e^{-\left(\frac{U/2+m}{\sqrt{2\bar{r}\rho}}\right)^2} + e^{-\left(\frac{U/2-m}{\sqrt{2\bar{r}\rho}}\right)^2}\right] \xrightarrow{\rho\to 0} 0, \tag{6.92}$$

Figure 6.3: α_c and α_F as functions of ρ.

since U has to be non zero when $\rho \to 0$ by (6.91) and $C \in [0, 1)$. This implies also that $\bar{r} \xrightarrow{\rho \to 0} \alpha$ by (6.71), recalling that $t = \Delta = 0$. Notice that, when $\rho \to 0$ the first erf term in (6.91) is exponentially close to $2(1 - \rho)$. Using the asymptotics for erf

$$\mathrm{erf}(x) = 1 - \frac{e^{-x^2}}{\sqrt{\pi}x} + \dots \tag{6.93}$$

which entails that

$$\rho \left[\mathrm{erf}\left(\frac{U/2 - m}{\sqrt{2\bar{r}\rho}} \right) + \mathrm{erf}\left(\frac{U/2 + m}{\sqrt{2\bar{r}\rho}} \right) \right] = O\left(e^{-K/(2\bar{r}\rho)} \right) \tag{6.94}$$

with some constant $K > 0$. The only way for the previous to hold is that one of the two erf in the square brackets go exponentially fast to -1, or in other words that

$$\frac{U}{2} < m . \tag{6.95}$$

This is already sufficient to prove that the magnetization goes to 1. Indeed, one has

$$m = \frac{1}{2} \left[\underbrace{\mathrm{erf}\left(\frac{m - U/2}{\sqrt{2\bar{r}\rho}} \right)}_{\to \infty} + \mathrm{erf}\left(\frac{m + U/2}{\sqrt{2\bar{r}\rho}} \right) \right] \xrightarrow{\rho \to 0} 1 . \tag{6.96}$$

We have just shown that for any fixed α, there is a sufficiently small ρ that induces a non vanishing magnetization. In other words, α_c as a function of ρ has to approach ∞ when $\rho \to 0$.

Remark 6.3. Here the introduction of the prior information on the second moment through $\hat{v}$ is crucial to enlarge the retrieval region in the phase diagram α-Δ. As an example for $\rho = 0.05$ fixed, at $\alpha = 0.3$ and $\Delta = 0.4$, in presence of the $v = 1$ constraint, the fixed point equations under the hypothesis of one pattern condensation produce a magnetization $m \simeq 0.970$. If we remove the constraint instead, by setting $\hat{v} = 0$, the equilibrium magnetization vanishes.

6.4 Decimation performance

In this section we discuss the viability of the decimation procedure and list the numerical results obtained from the fixed point equations (6.44)-(6.48). Analogously to what discussed in Section 6.3, (6.44)-(6.48) can be solved by iterating (6.45) and solving exactly (6.44), (6.46) at each iteration step. In order to implement a truly recursive decimation we need to discretize the integrals in (6.47) and (6.48) as follows

$$r = \frac{\alpha(1 - R/N)\beta^2 q}{(1 - \beta(1 - q))^2} + \beta^2 \Delta q$$

$$+ \frac{1}{N} \sum_{\mu=P-R+1}^{P} \frac{2q\beta^2(1 - (1 - \bar{\epsilon}(\tau^\mu))^2)}{1 - \beta^2(1 - (1 - \bar{\epsilon}(\tau^\mu))^2)(1 - q)^2} \left[1 + \frac{2\beta^2(1 - q)^2(1 - (1 - \bar{\epsilon}(\tau^\mu))^2)}{1 - \beta^2(1 - (1 - \bar{\epsilon}(\tau^\mu))^2)(1 - q)^2} \right]$$

(6.97)

$$u = \beta\hat{v} + \beta(1 - \beta\Delta)v - \alpha(1 - R/N)\beta \frac{1 - \beta(1 - 2q)}{(1 - \beta(1 - q))^2}$$

$$- \frac{1}{N} \sum_{\mu=P-R+1}^{P} \left[\frac{2v\beta^2(1 - (1 - \bar{\epsilon}(\tau^\mu))^2)}{1 - \beta^2(1 - (1 - \bar{\epsilon}(\tau^\mu))^2)(1 - q)^2} + q\frac{4\beta^4(1 - q)^2(1 - (1 - \bar{\epsilon}(\tau^\mu))^2)^2}{(1 - \beta^2(1 - (1 - \bar{\epsilon}(\tau^\mu))^2)(1 - q)^2)^2} \right] ,$$

(6.98)

where $\tau^\mu = 1 - \frac{\mu-1}{P}$. When $R = 0$ by convention we set the sums to 0, and the analysis from the previous section holds. In what follows we consider only one pattern condensation. Furthermore, we stress that despite the magnetized state can be not truly stable in the thermodynamic limit, it is still a critical point for the free entropy. At the end of the procedure we have P values of magnetizations, and reconstruction errors on each pattern $(\bar{\epsilon}(\tau^\mu))_{\mu=1}^P$. We can use them to evaluate the theoretical Mean Square Error (MSE) associated with decimation. We define the matrix MSE as follows

$$\text{MSE}(\boldsymbol{\eta}) = \frac{1}{2N} \left\| \frac{\boldsymbol{\eta}\boldsymbol{\eta}^\intercal}{N} - \frac{\boldsymbol{\xi}\boldsymbol{\xi}^\intercal}{N} \right\|_F^2 , \tag{6.99}$$

where we adopted $\hat{\mathbf{S}} = \frac{\boldsymbol{\eta}\boldsymbol{\eta}^{\mathsf{T}}}{\sqrt{N}}$ as an estimator for $\frac{\boldsymbol{\xi}\boldsymbol{\xi}^{\mathsf{T}}}{\sqrt{N}}$. The subscript F denotes the Frobenius' norm. Expanding the average of the previous formula we can evaluate the expected MSE

$$\mathbb{EMSE}(\boldsymbol{\eta}) = \frac{1}{2N^3}\sum_{i,j=1}^{N}\sum_{\mu,\nu=1}^{P}\mathbb{E}[\eta_i^\mu\eta_i^\nu\eta_j^\mu\eta_j^\nu + \xi_i^\mu\xi_i^\nu\xi_j^\mu\xi_j^\nu] - \frac{1}{N^3}\sum_{i,j=1}^{N}\sum_{\mu,\nu=1}^{P}\mathbb{E}\xi_i^\mu\eta_i^\nu\xi_j^\mu\eta_j^\nu. \tag{6.100}$$

Concerning the four ξ's contribution, the only chance to have non zero expectation is that $\mu = \nu$ or $i = j$, because ξ_i^μ are *i.i.d.* from a centered distribution. Hence, neglecting the diagonal contributions $\mu = \nu$ and $i = j$ that are sub-leading, a simple computation yields

$$\frac{1}{2N^3}\sum_{i,j=1}^{N}\sum_{\mu,\nu=1}^{P}\mathbb{E}\xi_i^\mu\xi_i^\nu\xi_j^\mu\xi_j^\nu = \frac{\alpha + \alpha^2}{2} + O\left(\frac{1}{N}\right). \tag{6.101}$$

By our assumptions on the estimates $\boldsymbol{\eta}$ (see Remark 6.2) we analogously get

$$\frac{1}{2N^3}\sum_{i,j=1}^{N}\sum_{\mu,\nu=1}^{P}\mathbb{E}\xi_i^\mu\xi_i^\nu\xi_j^\mu\xi_j^\nu = \frac{\alpha + \alpha^2}{2} + O\left(\frac{1}{N}\right), \tag{6.102}$$

$$\frac{1}{N^3}\sum_{i,j=1}^{N}\sum_{\mu,\nu=1}^{P}\mathbb{E}\xi_i^\mu\eta_i^\nu\xi_j^\mu\eta_j^\nu = \alpha^2 + \frac{1}{N}\sum_{\mu=1}^{P}(1 - \bar{\epsilon}(\tau^\mu))^2 + O\left(\frac{1}{N}\right). \tag{6.103}$$

Finally, the average matrix MSE takes the form

$$\mathbb{EMSE}(\boldsymbol{\eta}) = \alpha\left(1 - \frac{1}{P}\sum_{\mu=1}^{P}(1 - \bar{\epsilon}(\tau^\mu))^2\right) + O(N^{-1}) \tag{6.104}$$

or in integral form in the thermodynamic limit

$$\mathbb{EMSE}(\boldsymbol{\eta}) = \alpha - \alpha\int_0^1 d\tau(1 - \bar{\epsilon}(\tau))^2. \tag{6.105}$$

6.4.1   Flow of magnetization

Each decimation step is associated with an equilibrium value of the Mattis magnetization, and a corresponding reconstruction error. To begin with consider a Rademacher prior $P_\xi = \frac{\delta_{-1}+\delta_1}{2}$ and $\beta = 1/\Delta$. As it is clear from Figure 6.4, the magnetization increases along the decimation procedure, meaning that the amount of noise we add while decimating is less compared the one we remove by diminishing the effective rank of the hidden matrix $\frac{\boldsymbol{\xi}\boldsymbol{\xi}^{\mathsf{T}}}{\sqrt{N}}$. This confirms, at least for a simple binary prior, that decimation is a viable option for symmetric matrix factorization.

The points and error bars are obtained by sampling the decimation Gibbs measure (6.5) with an AMP algorithm with an informative initialization, pointing in the direction of one of
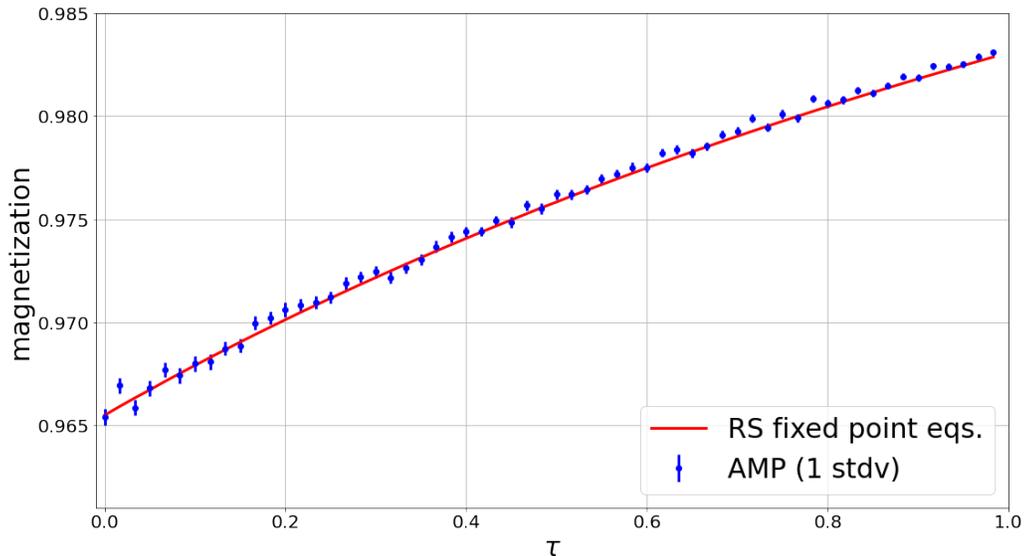
Figure 6.4: Magnetization vs decimation step, condensation on a single pattern. $N = 2000$, $P = 60$, $\Delta = 1/\beta = 0.15$, $\rho = 1$.

the patterns. Obviously, such initialization is not accessible to the Statistician that aims for the patterns, but it is still a good way to sample (6.5) and to verify numerically that we correctly evaluated the decimation noise contribution. To obtain error bars we run AMP 400 times, generating each time a new spike independently, for the entire decimation, thus obtaining 400 different values of magnetization for each decimation step. Error bars correspond to 1 standard deviation of such samples.

There are at least two possible objections to this test:

- the size of the system considered is $N = 2000$, hence with 0.03 we only have rank $P = 60$, which is not truly a high rank regime;

- the range of Δ under scrutiny, the one where the starting magnetization is non-vanishing, is not really satisfactory.

We can answer to both by introducing a strong sparsity in the ground truth, that increases robustness to noise in general, allowing for larger α's and a wider Δ range.

6.4.2 Decimation vs Rotationally Invariant Estimators (RIE)

A problem that is strongly related to dictionary learning is matrix denoising. Matrix denoising requires to clean a matrix $\mathbf{S} \in \mathbb{R}^{N \times N}$, that we suppose to have $O(1)$ eigenvalues, from the noisy observation

$$\mathbf{Y}_D = \mathbf{S} + \sqrt{\Delta}\mathbf{Z} \tag{6.106}$$

with $\mathbf{Z}$ a Wigner matrix with $O(1)$ eigenvalues (one could actually take any rotationally invariant matrix). Notice that the task is somehow simpler than factorization, which instead requires in addition to find the rank factorization of $\mathbf{S}$. However, it is clear that decimation can carry out both tasks.

In absence of additional information on the matrix $\mathbf{S}$ than the observations $\mathbf{Y}_D$, a possible strategy is to look for an estimator $\hat{\mathbf{S}}$ that diagonalizes on the same eigenbasis of $\mathbf{Y}_D$:

$$\mathbf{Y}_D = \mathbf{O}\boldsymbol{\lambda_Y}\mathbf{O}^{\mathsf{T}} \quad \Rightarrow \quad \hat{\mathbf{S}} = \mathbf{O}\hat{\boldsymbol{\lambda}}_{\mathbf{S}}\mathbf{O}^{\mathsf{T}}. \tag{6.107}$$

It remains then to estimate the eigenvalues $\hat{\boldsymbol{\lambda}}_{\mathbf{S}}$. In [170] the authors propose the cleaning procedure

$$\hat{\boldsymbol{\lambda}}_{\mathbf{S}} = \boldsymbol{\lambda_Y} - 2\Delta\mathcal{H}[\rho_{\mathbf{Y}}](\boldsymbol{\lambda_Y}) \tag{6.108}$$

where

$$\mathcal{H}[\rho_{\mathbf{Y}}](x) := \mathrm{P.V.} \int \frac{dy\rho_{\mathbf{Y}}(y)}{x - y} \tag{6.109}$$

is (proportional to) the Hilbert transform of $\rho_{\mathbf{Y}}$, the density of eigenvalues of $\mathbf{Y}_D$, and it is applied component-wise to the diagonal matrix elements in (6.108). $\hat{\mathbf{S}}$ is thus called Rotationally Invariant Estimator (RIE).

Now, if we take $\mathbf{S} = \frac{\boldsymbol{\xi}\boldsymbol{\xi}^{\mathsf{T}}}{N}$ (that has indeed $O(1)$ eigenvalues) we can compare numerically the performances of the RIE with that of decimation. The MSE is a random quantity, hence for each value of Δ that we tested (see Figure 6.5a) we repeated the same numerical experiment 30 times starting from independently generated $\frac{\boldsymbol{\xi}\boldsymbol{\xi}^{\mathsf{T}}}{N}$, in order to be able to estimate its fluctuations and exclude possible compatibility between the decimation MSE and the RIE's.

With reference to Figure 6.5a, the red curve represents the MSE associated with the RIE when P_{ξ} is Rademacher. Error bars are too small to be appreciated in the plot. The blue and the green curves are instead the MSEs associated with the decimation procedures at $\beta = 1/\Delta$ and $\beta \to +\infty$ respectively. The low temperature one turns out to have the best predicted performance. In a Bayes-optimal setting this would obviously be not the case. However, since at each decimation step we have an associative memory model, it is somehow reasonable that the best results are attained at really low temperature, when it is easier to retrieve patterns. The cyan and yellow data points are obtained through an AMP that samples at $\beta = 10$, whose error bars are still invisible, and $\beta = 20$ respectively. The latter in particular is in good agreement with our theoretical prediction within the confidence of 1 standard deviation. In Figure 6.5a both MSEs are really small and the range of Δ which is accessible (that grants non-vanishing magnetization) is rather narrow. Furthermore, as anticipated, it can be objected that $\alpha = 0.03$, with finite size experiments, is not really an extensive rank regime.

For this reason we repeated the same experiment with a strongly sparse prior, $\rho = 0.05$, and $\alpha = 0.4$. In this way when $N = 1000$ the rank is 400, as in Figure 6.5b. The latter illustrates

(a) $N = 2000$, $P = 60$, Rademacher prior.

(b) $N = 1000$, $P = 400$, strongly sparse Rademacher prior: $\rho = 0.05$.
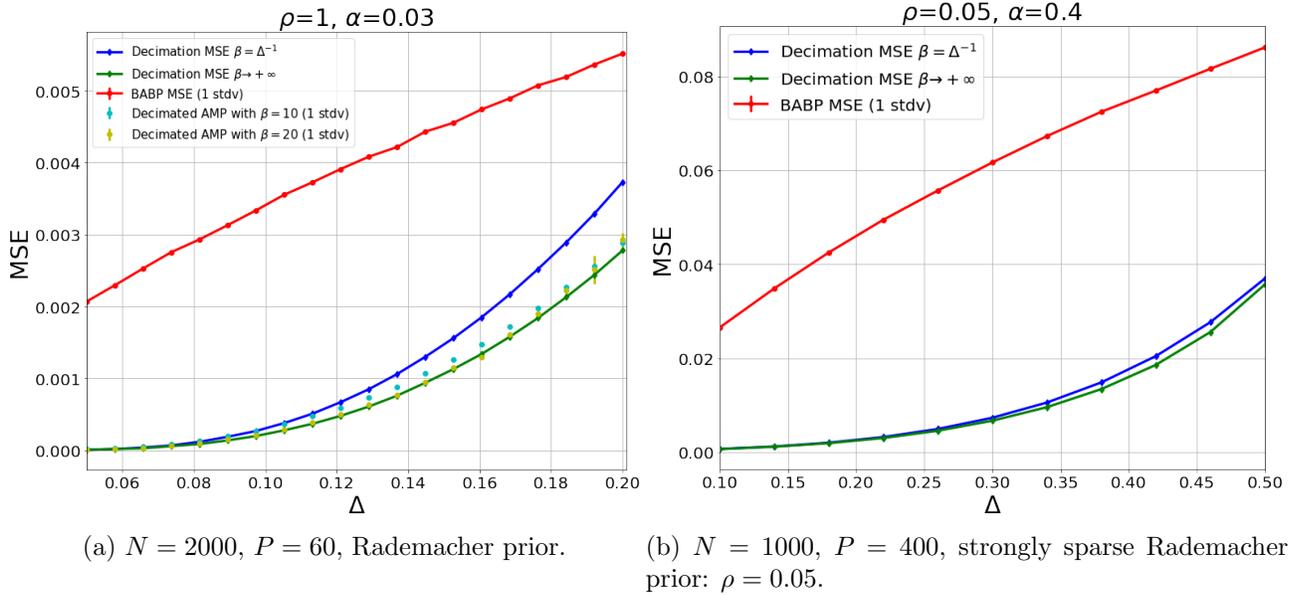
Figure 6.5: Comparison between RIE and decimation performances for matrix denoising.

again the comparison between RIE's and the predicted decimation performances. The colors are associated as before. We see that the ordering of the performances does not change, but the MSEs are bigger and the difference between those of the finite temperature sampling $\beta = 1/\Delta$ and the low temperature sampling is no longer that large if compared to the MSEs themselves.

Recall that decimation is possible only if we start from an (α, Δ), be it for $\beta = 1/\Delta$ or $\beta \to \infty$, such that the first magnetization is non-zero, whereas it is always possible to build a RIE starting from $\mathbf{Y}_D$. In fact, for larger values of Δ the first magnetization may vanish, and so do all the following ones yielding a trivial estimator. Nevertheless, when viable, decimation not only has a better performance than the RIE, but it retrieves also the factors composing the hidden matrix.

6.5 Concluding remarks

In this chapter we have proposed a novel iterative procedure to carry out (symmetric) high rank matrix factorization. The main idea is that of abandoning the Bayes-optimal setting, to reduce the problem to a Hopfield-like model in which the high rank effects induced by the coupling matrix are known. However, as the reader will have noticed at this point, the Hopfield model works as an associative memory model, *i.e.* it is able to retrieve the patterns in some regions if a hint on their directions is given. It is clear that no hint is accessible to a Statistician the wants to infer them. With an uninformative start, algorithms such as AMP typically remain stuck in some *mixture states*, in which the **x**-configuration has a finite overlap with an odd (> 1) number

of patterns, that is undesirable for inference. Nevertheless, our theoretical analysis clearly shows that those maxima of the free entropy corresponding to magnetized states are there in some range of parameters, and hence it is indeed theoretically possible to retrieve the hidden matrix.

In a continuation [177] (or completion) of the present work, we designed an algorithm, a "ground state oracle", able to find the ground states of the neural network models at each decimation step without any prior hint. The algorithm works for binary spins and is based on simulated annealing. In order to prevent the oracle to get stuck in the aforementioned mixture states, we inserted an energy-based restarting criterion: once a candidate pattern is found we compute its energy and compare it with the one of a ground state, which is theoretically known. If it is too high we discard the result and start over. We defer the reader to the preprint [177] for further details. One drawback of the ground state oracle is that it needs a number of restarts growing exponentially in the number of signal components. We were able to run the algorithm with N up to ~ 2000, and the execution time depends also on the noise level Δ.

Appendix A

Multi-species on the Nishimori line

A.1 Proof of the Concentration Lemma

Proof of Lemma 2.5. Let us split the proof into three steps for the sake of clarity. As anticipated, it is convenient to split the total fluctuation of $\mathcal{L}_r$ into two parts, thus proving that:

$$\mathbb{E}_{\boldsymbol{\epsilon}}\mathbb{E}\Big\langle \Big(\mathcal{L}_r - \langle\mathcal{L}_r\rangle_{N,t}^{(\epsilon)} \Big)^2 \Big\rangle_{N,t}^{(\epsilon)} \longrightarrow 0 \quad \text{as } N \to \infty \tag{A.1}$$

$$\mathbb{E}_{\boldsymbol{\epsilon}}\mathbb{E}\Big(\langle\mathcal{L}_r\rangle_{N,t}^{(\epsilon)} - \mathbb{E}\langle\mathcal{L}_r\rangle_{N,t}^{(\epsilon)} \Big)^2 \longrightarrow 0 \quad \text{as } N \to \infty . \tag{A.2}$$

From this moment on, we neglect sub and superscripts in the Gibbs brackets as well as t-dependencies. We start by proving the inequality (2.45).

Proof of inequality (2.45)**:** To begin with, we compute:

$$\mathbb{E}[\langle\mathcal{L}_r\rangle] = \frac{1}{N_r}\sum_{i\in\Lambda_r}\mathbb{E}\Big\langle\sigma_i + \frac{J_i^r\sigma_i}{2\sqrt{Q_{\epsilon,r}}}\Big\rangle = \mathbb{E}\langle m_r\rangle + \frac{1}{2}\mathbb{E}[1 - \langle m_r\rangle] = \frac{1}{2}\mathbb{E}[1 + \langle m_r\rangle] \tag{A.3}$$

where integration by parts has been used.

Then, we proceed with:

$$\mathbb{E}\langle\mathcal{L}_r^2\rangle = \frac{1}{N_r^2}\sum_{i,j\in\Lambda_r}\mathbb{E}\Big\langle\sigma_i\sigma_j + \frac{J_i^r\sigma_i\sigma_j}{\sqrt{Q_{\epsilon,r}}} + \frac{J_i^r J_j^r\sigma_i\sigma_j}{4Q_{\epsilon,r}}\Big\rangle = \underbrace{\mathbb{E}\langle m_r^2\rangle}_{R_1} + \underbrace{\frac{1}{N_r^2}\sum_{i,j\in\Lambda_r}\mathbb{E}\Big\langle\frac{J_i^r\sigma_i\sigma_j}{\sqrt{Q_{\epsilon,r}}}\Big\rangle}_{R_2} +$$

$$+ \underbrace{\frac{1}{N_r^2}\sum_{i,j\in\Lambda_r}\mathbb{E}\Big\langle\frac{J_i^r J_j^r\sigma_i\sigma_j}{4Q_{\epsilon,r}}\Big\rangle}_{R_3} . \tag{A.4}$$

We treat the three terms R_1, R_2 and R_3 separately with repeated integrations by parts.

$$R_2 = \frac{1}{N_r^2} \sum_{i,j \in \Lambda_r} \mathbb{E}\left[\langle\sigma_j\rangle - \langle\sigma_i\sigma_j\rangle\langle\sigma_i\rangle\right] = \mathbb{E}\langle m_r\rangle - \mathbb{E}\langle m_r\rangle^2 \tag{A.5}$$

where we have used the Nishimori identity (2.16).

$$R_3 = \frac{1}{4N_r^2} \sum_{i,j \in \Lambda_r} \mathbb{E}\left[\frac{\delta_{ij}\overbrace{\langle\sigma_i\sigma_j\rangle}^{=1}}{Q_{\epsilon,r}} + J_j^r\frac{(\langle\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_i\sigma_j\rangle)}{\sqrt{Q_{\epsilon,r}}}\right] = \frac{1}{4N_rQ_{\epsilon,r}} +$$

$$+ \frac{1}{4N_r^2} \sum_{i,j \in \Lambda_r} \mathbb{E}\left[1 - \langle\sigma_j\rangle^2 - \langle\sigma_i\rangle(\langle\sigma_i\rangle - \langle\sigma_j\rangle\langle\sigma_i\sigma_j\rangle) - \langle\sigma_i\sigma_j\rangle(\langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle)\right] =$$

$$= \frac{1}{4N_rQ_{\epsilon,r}} + \frac{1}{4} - \frac{1}{2}\mathbb{E}\langle m_r\rangle + \frac{1}{2}\mathbb{E}\langle m_r\rangle^2 - \frac{1}{4}\mathbb{E}\langle m_r^2\rangle . \tag{A.6}$$

Hence:

$$R_1 + R_2 + R_3 = \frac{1}{4} + \frac{1}{4N_rQ_{\epsilon,r}} + \frac{3}{4}\mathbb{E}\langle m_r^2\rangle + \frac{1}{2}\mathbb{E}\langle m_r\rangle - \frac{1}{2}\mathbb{E}\langle m_r\rangle^2 . \tag{A.7}$$

Summing up all the contributions:

$$\mathbb{E}\langle\mathcal{L}_r^2\rangle - (\mathbb{E}\langle\mathcal{L}_r\rangle)^2 = \frac{1}{4N_rQ_{\epsilon,r}} + \frac{3}{4}\mathbb{E}\langle m_r^2\rangle - \frac{1}{2}\mathbb{E}\langle m_r\rangle^2 - \frac{1}{4}(\mathbb{E}\langle m_r\rangle)^2 =$$

$$= \frac{1}{4N_rQ_{\epsilon,r}} + \frac{1}{4}(\mathbb{E}\langle m_r^2\rangle - (\mathbb{E}\langle m_r\rangle)^2) + \frac{1}{2}(\mathbb{E}\langle m_r^2\rangle - \mathbb{E}\langle m_r\rangle^2) \geq \frac{1}{4}\mathbb{E}\left\langle(m_r - \mathbb{E}\langle m_r\rangle)^2\right\rangle . \tag{A.8}$$

Proof of (A.1): Notice that:

$$\frac{\partial\bar{p}_{N,\epsilon}}{\partial Q_{\epsilon,r}} = \frac{1}{N}\mathbb{E}\left\langle\sum_{i \in \Lambda_r}\left(\sigma_i + \frac{J_i^r\sigma_i}{2\sqrt{Q_{\epsilon,r}}}\right)\right\rangle = \alpha_r\mathbb{E}\langle\mathcal{L}_r\rangle = \frac{\alpha_r}{2}\mathbb{E}[1 + \langle m_r\rangle], \quad J_i^r \overset{\text{iid}}{\sim} \mathcal{N}(0,1) \tag{A.9}$$

$$\frac{\partial^2\bar{p}_{N,\epsilon}}{\partial Q_{\epsilon,r}^2} = \alpha_rN_r\mathbb{E}\langle(\mathcal{L}_r - \langle\mathcal{L}_r\rangle)^2\rangle - \frac{1}{4NQ_{\epsilon,r}^{3/2}}\sum_{i \in \Lambda_r}\mathbb{E}\langle J_i^r\sigma_i\rangle . \tag{A.10}$$

From the last one, after an integration by parts and using the regularity of the map $\epsilon \longmapsto \mathbf{Q}_\epsilon(\cdot)$ and Lemma 2.4 we get:

$$\mathbb{E}_\epsilon \mathbb{E}\langle (\mathcal{L}_r - \langle \mathcal{L}_r \rangle)^2 \rangle \leq \frac{1}{N_r \alpha_r s_N^K} \prod_{s=1}^K \int_{Q_{s_N,s}}^{Q_{2s_N,s}} dQ_{\epsilon,s} \frac{\partial^2 \bar{p}_{N,\epsilon}}{\partial Q_{\epsilon,r}^2} + \mathbb{E}_\epsilon \frac{1}{4 N_r \epsilon_r} \mathbb{E}[1 - \langle m_r \rangle] \leq$$

$$\leq \frac{1}{N_r \alpha_r s_N^K} \prod_{s \neq r,1}^K \int_{Q_{s_N,s}}^{Q_{2s_N,s}} dQ_{\epsilon,s} \left[\left. \frac{\partial \bar{p}_{N,\epsilon}}{\partial Q_{\epsilon,r}} \right|_{Q_{2s_N,r}} - \left. \frac{\partial \bar{p}_{N,\epsilon}}{\partial Q_{\epsilon,r}} \right|_{Q_{s_N,r}} \right] + \frac{\log 2}{4 N_r s_N} \leq$$

$$\leq \frac{2 K_r(\Delta)}{N_r s_N^K} + \frac{\log 2}{4 N_r s_N} = \mathcal{O}\left(\frac{1}{N_r s_N^K} \right) \longrightarrow 0 \quad \text{(A.11)}$$

where:

$$\prod_{s \neq r,1}^K (Q_{2s_N,s} - Q_{s_N,s}) \leq K_r(\Delta). \quad \text{(A.12)}$$

Proof of (A.2): Let $p_{N,\epsilon}$ be the random interpolating pressure, such that $\mathbb{E} p_{N,\epsilon} = \bar{p}_{N,\epsilon}$. Define:

$$\hat{p}_{N,\epsilon} = p_{N,\epsilon} - \alpha_r \sqrt{Q_{\epsilon,r}} \sum_{i \in \Lambda_r} \frac{|J_i^r|}{N_r}, \qquad \hat{\bar{p}}_{N,\epsilon} = \mathbb{E} \hat{p}_{N,\epsilon} \quad \text{(A.13)}$$

$$\frac{\partial^2 \hat{p}_{N,\epsilon}}{\partial Q_{\epsilon,r}^2} = \alpha_r N_r \langle (\mathcal{L}_r - \langle \mathcal{L}_r \rangle)^2 \rangle + \frac{\alpha_r}{4 Q_{\epsilon,r}^{3/2}} \sum_{i \in \Lambda_r} \frac{|J_i^r| - J_i^r \langle \sigma_i \rangle}{N_r} \geq 0. \quad \text{(A.14)}$$

Let us evaluate:

$$\left| \frac{\partial \hat{p}_{N,\epsilon}}{\partial Q_{\epsilon,r}} - \frac{\partial \hat{\bar{p}}_{N,\epsilon}}{\partial Q_{\epsilon,r}} \right| \geq \alpha_r \left| \langle \mathcal{L}_r \rangle - \mathbb{E}\langle \mathcal{L}_r \rangle \right| - \frac{\alpha_r |A_r|}{2 \sqrt{Q_{\epsilon,r}}} \quad \text{(A.15)}$$

where:

$$A_r := \frac{1}{N_r} \sum_{i \in \Lambda_r} [|J_i^r| - \mathbb{E}|J_i^r|]. \quad \text{(A.16)}$$

Thanks to the independence of the J_i^r it is immediate to verify that $\exists a \geq 0$ s.t.:

$$\mathbb{E}[A_r^2] \leq \frac{a}{N_r}. \quad \text{(A.17)}$$

Using Lemma 3.2 in [18], with notations used in [13]:

$$\left| \frac{\partial \hat{p}_{N,\epsilon}}{\partial Q_{\epsilon,r}} - \frac{\partial \hat{\bar{p}}_{N,\epsilon}}{\partial Q_{\epsilon,r}} \right| \leq \frac{1}{\delta} \sum_{u \in \{Q_{\epsilon,r}+\delta, Q_{\epsilon,r}, Q_{\epsilon,r}-\delta\}} [|\hat{p}_{N,\epsilon} - \hat{\bar{p}}_{N,\epsilon}| + \alpha_r \sqrt{u} |A_r|] +$$

$$+ C_\delta^+(Q_{\epsilon,r}) + C_\delta^-(Q_{\epsilon,r}) \quad \text{(A.18)}$$

with:

$$C_\delta^\pm(Q_{\epsilon,r}) = |\hat{\tilde{p}}'_{N,\epsilon}(Q_{\epsilon,r} \pm \delta) - \hat{\tilde{p}}'_{N,\epsilon}(Q_{\epsilon,r})| \tag{A.19}$$

$$|\hat{\tilde{p}}'_{N,\epsilon}| = \left| \frac{\alpha_r}{2}\mathbb{E}[1 + \langle m_r \rangle] - \frac{\alpha_r \mathbb{E}|J_1^r|}{2\sqrt{Q_{\epsilon,r}}} \right| \leq \alpha_r \left(1 + \frac{C}{2\sqrt{s_N}} \right) \tag{A.20}$$

$$C_\delta^\pm(Q_{\epsilon,r}) \leq \alpha_r \left(2 + \frac{C}{\sqrt{s_N}} \right) \tag{A.21}$$

where for simplicity we have kept the dependence on $Q_{\epsilon,r}$ only, $\hat{\tilde{p}}'_{N,\epsilon}$ is the derivative w.r.t. it and $\delta > 0$. Notice that δ will be chosen strictly smaller than s_N, so that $Q_{\epsilon,r} - \delta \geq \epsilon - \delta \geq s_N - \delta > 0$.

Then, using (A.15), (A.16) and (A.18), and thanks to the fact that $(\sum_{i=1}^p \nu_i)^2 \leq p \sum_{i=1}^p \nu_i^2$, we get:

$$\frac{\alpha_r^2}{9} |\langle \mathcal{L}_r \rangle - \mathbb{E}\langle \mathcal{L}_r \rangle|^2 \leq \frac{1}{\delta^2} \sum_{u \in \{Q_{\epsilon,r}+\delta, Q_{\epsilon,r}, Q_{\epsilon,r}-\delta\}} [|\hat{p}_{N,\epsilon} - \hat{\tilde{p}}_{N,\epsilon}|^2 + \alpha_r^2 u |A_r|^2] +$$

$$+ C_\delta^+(Q_{\epsilon,r})^2 + C_\delta^-(Q_{\epsilon,r})^2 + \frac{\alpha_r^2 A_r^2}{4\epsilon_r}. \tag{A.22}$$

We first evaluate the two terms containing $C_\delta^\pm$:

$$\mathbb{E}_\epsilon[C_\delta^+(Q_{\epsilon,r})^2 + C_\delta^-(Q_{\epsilon,r})^2] \leq 2\alpha_r \left(2 + \frac{C}{\sqrt{s_N}} \right) \mathbb{E}_\epsilon[C_\delta^+(Q_{\epsilon,r}) + C_\delta^-(Q_{\epsilon,r})] \leq$$

$$\leq \frac{2\alpha_r}{s_N^K} \left(2 + \frac{C}{\sqrt{s_N}} \right) \prod_{s=1}^K \int_{Q_{s_N,s}}^{Q_{2s_N,s}} dQ_{\epsilon,s} [\hat{\tilde{p}}'_{N,\epsilon}(Q_{\epsilon,r} + \delta) - \hat{\tilde{p}}'_{N,\epsilon}(Q_{\epsilon,r} - \delta)] =$$

$$= \frac{2\alpha_r}{s_N^K} \left(2 + \frac{C}{\sqrt{s_N}} \right) \prod_{s \neq r, 1}^K \int_{Q_{s_N,s}}^{Q_{2s_N,s}} dQ_{\epsilon,s} [\hat{\tilde{p}}_{N,\epsilon}(Q_{2s_N,r} + \delta) - \hat{\tilde{p}}_{N,\epsilon}(Q_{2s_N,r} - \delta) +$$

$$- \hat{\tilde{p}}_{N,\epsilon}(Q_{s_N,r} + \delta) + \hat{\tilde{p}}_{N,\epsilon}(Q_{s_N,r} - \delta)] \leq \frac{8\alpha_r^2 K_r(\Delta)}{s_N^K} \delta \left(2 + \frac{C}{\sqrt{s_N}} \right)^2. \tag{A.23}$$

Taking the expectation $\mathbb{E}_\epsilon \mathbb{E}$ in (A.22), and defining W_r s.t. $Q_{\epsilon,r} \leq W_r$, we get:

$$\frac{\alpha_r^2}{9} \mathbb{E}_\epsilon \mathbb{E} |\langle \mathcal{L}_r \rangle - \mathbb{E}\langle \mathcal{L}_r \rangle|^2 \leq \frac{3}{\delta^2} \left[\frac{S}{N} + \frac{\alpha_r^2 W_r a}{N_r} \right] +$$

$$+ \frac{8\alpha_r^2 K_r(\Delta)}{s_N^K} \delta \left(2 + \frac{C}{\sqrt{s_N}} \right)^2 + \frac{\alpha_r^2 a \log 2}{4N_r s_N}. \tag{A.24}$$

We can make the r.h.s. vanish by choosing for example: $\delta = s_N^{2K/3} N^{-1/3}$. The choice $s_N \propto N^{-1/16K}$ makes the r.h.s. (A.24) behave like $\mathcal{O}(N^{-1/4})$. $\qquad \square$

A.2 The monospecies case

In the case $K = 1$ the equation (2.46) reduces to:

$$\lim_{N \to \infty} \bar{p}_N(\mu, \mathbf{h}) = \sup_{x \in \mathbb{R}_{\geq 0}} \left\{ \mu \frac{(1-x)^2}{4} - \frac{\mu x^2}{2} + \psi(\mu x + h) \right\} \tag{A.25}$$

while (2.48) simply becomes:

$$x = \mathbb{E}_z \tanh \left(z \sqrt{\mu x + h} + \mu x + h \right) := T(x; \mu, h) . \tag{A.26}$$

We collect the main results on this model in the following proposition.

Proposition A.1. *Define:*

$$\bar{p}_{var}(x; \mu, h) = \mu \frac{(1-x)^2}{4} - \frac{\mu x^2}{2} + \psi(\mu x + h) . \tag{A.27}$$

The following hold:

1. *if $\mu < 1$ then $\bar{p}_{var}$ is concave in x. Equivalently if $\mu < 1$ then $T(x; \mu, h)$ is a contraction, and if further $h = 0$ then $x = 0$ is its fixed point;*

2. *the stable solution of the consistency equation (A.26) is continuous at $(\mu, h) = (1, 0)$:*

$$\lim_{(\mu,h) \to (1,0)} \bar{x}(\mu, h) = 0 = \bar{x}(1, 0) ; \tag{A.28}$$

3. *for fixed $h = 0$, the magnetization goes to 0 linearly with $\mu - 1$ as $\mu \to 1_+$, more precisely:*

$$\bar{x} = (1 + o(1)) \frac{\mu - 1}{\mu^2} \tag{A.29}$$

   where $o(1)$ goes to 0 when $\mu \to 1_+$. Therefore the critical exponent β (in the Landau classification) is 1, which means that the derivative of the magnetization w.r.t. μ does not diverge at the critical point, it only jumps from 0 to 1 and then decreases;

4. *Along the line $(\mu, \lambda(\mu - 1))$, $\lambda > 0$ in the plane (μ, h) the magnetization goes to 0 as follows:*

$$\bar{x} = \sqrt{\frac{\lambda(\mu - 1)}{\mu^2}} (1 + o(1)) \tag{A.30}$$

   when $\mu \to 1_+$, therefore with a critical exponent $1/2$;

5. *For fixed $\mu = 1$ and $h \to 0_+$ the magnetization behaves as:*

$$\bar{x}^2 = h(1 + o(1)) \tag{A.31}$$

where $o(1) \to 0$ when $h \to 0_+$. Therefore we have a critical exponent $\delta = 2$ (according to Landau's classification).

Proof. 1. The fist assertion follows immediately from (2.56), since $\hat{\alpha} \equiv 1$ and $\Delta \equiv \mu$. Then, by (2.51):

$$\frac{dT}{dx}(x; \mu, h) = \mu \mathbb{E}_z \left[\left(1 - \tanh^2 \left(z\sqrt{\mu x + h} + \mu x + h \right) \right)^2 \right] \leq \mu < 1$$

that implies T is a contraction. It is easy to see that if $h = 0$ then $x = 0$ is a solution of the fixed point equation which must be unique by Banach's fixed point theorem.

2. Using continuity and monotonicity of T (see (2.51)):

$$\limsup_{(\mu,h)\to(1,0)} \bar{x}(\mu, h) = T(\limsup_{(\mu,h)\to(1,0)} \bar{x}(\mu, h); 1, 0)$$

$$\liminf_{(\mu,h)\to(1,0)} \bar{x}(\mu, h) = T(\liminf_{(\mu,h)\to(1,0)} \bar{x}(\mu, h); 1, 0)$$

hence both $\limsup_{(\mu,h)\to(1,0)} \bar{x}(\mu, h)$ and $\liminf_{(\mu,h)\to(1,0)} \bar{x}(\mu, h)$ satisfy the consistency equation:

$$m = \mathbb{E}_z \tanh(z\sqrt{m} + m)$$

whose solution $m = 0$ is unique, since the derivative of $T(m; 1, 0)$ is ≤ 1 and equality holds only at $m = 0$. We conclude that there exists

$$\lim_{(\mu,h)\to(1,0)} \bar{x}(\mu, h) = 0 = \bar{x}(1, 0) . \tag{A.32}$$

3. We first notice that

$$\mathbb{E} \tanh^2 \left(z\sqrt{Q} + Q \right) = \mathbb{E} \tanh \left(z\sqrt{Q} + Q \right) , \quad Q \geq 0, z \sim \mathcal{N}(0, 1) \tag{A.33}$$

which simply follows from the third relation in (2.51) and the identity (2.14). Indeed, the quantity in (A.33) is nothing but the quenched average magnetization of a free system. Now, by computing the first and second derivatives of the map $T(x; \mu, 0)$ and using (A.33) we get:

$$T'(0; \mu, 0) = \mu , \quad T''(0; \mu, 0) = -2\mu^2$$

$$\bar{x} = \mu\bar{x} - \mu^2\bar{x}^2(1 + o(1)) \quad \Rightarrow \quad \bar{x} = (1 + o(1)) \left(\frac{\mu - 1}{\mu^2} \right) ,$$

which implies that, in proximity of $\mu = 1$, the magnetization goes to 0 with a critical exponent $\beta = 1$ (not to be confused with inverse absolute temperature) and with slope 1.

4. An analogous expansion of T yields:

$$\bar{x} = T(\bar{x}; \mu, \lambda(\mu - 1)) = \mu\bar{x} + \lambda(\mu - 1) - (\mu^2\bar{x}^2 + o(\mu - 1))(1 + o(1))$$

which in turn entails:

$$\bar{x}^2 = \frac{\lambda(\mu - 1)}{\mu^2}(1 + o(1)) \ .$$

5. Here by $o(1)$ we mean a quantity that approaches 0 as $h \to 0_+$. As in the previous steps:

$$\bar{x} = T(\bar{x}; 1, h) = \bar{x} + h - (\bar{x}^2 + o(h))(1 + o(1)) \ ,$$

then we get:

$$\bar{x}^2 = h(1 + o(1)) \quad \Rightarrow \quad \delta = 2 \ .$$

$\square$

Appendix B

Structured PCA

B.1 Learning the optimal pre-processing $\mathbf{J}(\mathbf{Y})$ by expectation maximization

Until now we have assumed that we are in the Bayesian-optimal setting where, in particular, the polynomial potential V defining the noise statistics is completely known and correctly exploited. As seen from section 5.5.2, given a potential V we could deduce from the AdaTAP formalism an optimal polynomial

$$\mathbf{J} = \mathbf{J}(\mathbf{Y}) = \sum_{k \leq K} c_k \mathbf{Y}^k$$

to pre-process the data $\mathbf{Y}$ before using it in AMP. The Bayes-optimal case corresponds to matrix (5.160), i.e., $\mathbf{J} = c_1 \mathbf{Y} + c_2 \mathbf{Y}^2 + c_3 \mathbf{Y}^3$ with $\mathbf{c} = (\mu\lambda, -\gamma\lambda^2, \gamma\lambda)$.

We here consider an extension of the previously derived AMP to a case where V is not known and therefore the optimal $\mathbf{J}$ cannot be deduced by the AdaTAP approach as we did in Section 5.5.2. What is known instead is an upper bound on the order of V. In the base-case model studied in details in the present paper the order is four. The procedure we propose below will not be tested numerically yet, but we believe it may be of interest to practitioners eager to improve the Bayes-optimal AMP for more practical settings than the specific ones studied here.

To directly learn the coefficients $(c_k)_{k \leq K}$ from the data, we propose to use an expectation maximization (EM) approach, with a routine inside AMP performing the parameter estimation by maximizing the current estimate of the free entropy, i.e., of the log-likelihood of the observed data $\ln P(\mathbf{Y} \mid \mathbf{c})$.

Assume that, at the AMP iterate t, the current estimate of the unknown coefficients $\mathbf{c} = (c_k)_{k \leq K}$ is $\mathbf{c}(t) = (c_k(t))_{k \leq K}$, the AMP estimate of the marginal means is $\mathbf{m}(t)$, and of the Onsager reaction term is $\bar{V}(t)$ (which is related to the set of Onsager coefficients, see Section 5.6.2). Let also the data matrix polynomial currently used by AMP be

$$\mathbf{J}(t) := \sum_{k \leq K} c_k(t)\mathbf{Y}^k.$$

From the analysis of Section 5.5.3 we know that at the saddle point we can safely replace the Onsager reaction term V_i by $\bar{V}$ in the AdaTAP equations. When this is plugged back into (5.181), this identity implies that also the following concentration is consistently valid: $\mathbb{E}(\tau_i - m_i^2) = \tau_i - m_i^2$, which is also equal by exchangeability to $N^{-1} \sum_{i \leq N} \mathbb{E}(\tau_i - m_i^2)$. Let us call $\bar{\chi}(t)$ the AMP estimate of the variance $\mathbb{E}(\tau_i - m_i^2)$. Applying these simplifications to the AMP iterates we get that the matrix $\mathbf{\Omega}(t) := \mathrm{diag}(\mathbf{V}(t) + (\boldsymbol{\tau}(t) - \mathbf{m}(t)^2)^{-1})$ can be simplified as

$$\mathbf{\Omega}(t) = (\bar{V}(t) + \bar{\chi}(t)^{-1})I_N.$$

From section 5.5.1 the AdaTAP approximation to the free entropy at iterate t then reads, using these simplifications, as

$$\Phi_N(t, \mathbf{c}(t)) = \frac{1}{2}\mathbf{m}(t)^\intercal \mathbf{J}(t)\mathbf{m}(t) + \frac{1}{2}\ln\det\left(\mathbf{\Omega}(t) - \mathbf{J}(t)\right) - \frac{1}{2}\bar{V}(t)\sum_{i \leq N} m_i(t)^2 + \frac{1}{2}\bar{\chi}(t)$$

$$- \sum_{i \leq N} \ln \int dP_X(x) \exp\left(\frac{1}{2}\bar{V}(t)x^2 + \left((\mathbf{J}(t)\mathbf{m}(t))_i - \bar{V}(t)m_i(t)\right)x\right). \tag{B.1}$$

The free entropy $\Phi_N(t, \mathbf{c}(t))$ is the current best approximation to the marginal log-likelihood of the data $\ln P(\mathbf{Y} \mid \mathbf{c})$, which we thus aim at maximizing with respect to the unknown parameters, all other quantities being fixed at their current values:

$$\partial_{c_k}\Phi_N|_{t,\mathbf{c}(t)} = \mathbf{m}(t)^\intercal \mathbf{Y}^k\left(\frac{1}{2}\mathbf{m}(t) - \eta(\mathbf{J}(t), \mathbf{m}(t), \bar{V}(t))\right)$$

$$- \frac{1}{2}\mathrm{Tr}\left(\mathbf{Y}^k(\mathbf{\Omega}(t) - \mathbf{J}(t))^{-1}\right), \tag{B.2}$$

where we used (5.162) and the notation $\eta(\mathbf{J}(t), \mathbf{m}(t), \bar{V}(t)) = (\eta_i(\mathbf{J}(t), \mathbf{m}(t), \bar{V}(t))_{i \leq N}$. Because $\mathbf{J}$ is diagonalizable in the same basis as the data $\mathbf{Y}$, the eigenvalues of which are denoted $\sigma_i = \sigma_i(\mathbf{Y})$, we have

$$\mathrm{Tr}\left(\mathbf{Y}^k(\mathbf{\Omega}(t) - \mathbf{J}(t))^{-1}\right) = \sum_{i \leq N} \frac{\sigma_i^k}{\bar{V}(t) + \bar{\chi}(t)^{-1} - \sum_{\ell \leq K} c_\ell(t)\sigma_i^\ell}. \tag{B.3}$$

Then

$$\partial_{c_k}\Phi_N|_{t,\mathbf{c}(t)} = \mathbf{m}(t)^\intercal \mathbf{Y}^k\left(\frac{1}{2}\mathbf{m}(t) - \eta(\mathbf{J}(t), \mathbf{m}(t), \bar{V}(t))\right)$$

$$- \frac{1}{2}\sum_{i \leq N} \frac{\sigma_i^k}{\bar{V}(t) + \bar{\chi}(t)^{-1} - \sum_{\ell \leq K} c_\ell(t)\sigma_i^\ell}. \tag{B.4}$$

We aim at maximizing the free entropy so given a learning rate $\zeta > 0$ the learning rule finally reads

$$c_k(t+1) = c_k(t) + \zeta\partial_{c_k}\Phi_N|_{t,\mathbf{c}(t)}. \tag{B.5}$$

B.2 Proofs for BAMP

B.2.1 Auxiliary AMP

The iterates of the auxiliary AMP are denoted by $\tilde{\mathbf{z}}^t, \tilde{\mathbf{u}}^t \in \mathbb{R}^N$, and they are computed as follows, for $t \geq 1$:

$$\tilde{\mathbf{z}}^t = \mathbf{Z}\tilde{\mathbf{u}}^t - \sum_{i=1}^{t} \bar{\mathsf{b}}_{t,i}\tilde{\mathbf{u}}^i, \quad \tilde{\mathbf{u}}^{t+1} = \tilde{h}_{t+1}(\tilde{\mathbf{z}}^1, \dots, \tilde{\mathbf{z}}^t, \mathbf{u}^1, \mathbf{X}^*). \tag{B.6}$$

The iteration (B.6) is initialized with $\tilde{\mathbf{u}}^1 = \mathbf{u}^1$, where $\mathbf{u}^1$ satisfies (5.192). For $t \geq 1$, the functions $\tilde{h}_{t+1} : \mathbb{R}^{t+2} \to \mathbb{R}$ are applied component-wise, and they are recursively defined as

$$\tilde{h}_{K(t-1)+1+\ell}(z_1, \dots, z_{K(t-1)+\ell}, u_1, x^*) = z_{K(t-1)+\ell} + (\tilde{\mathbf{B}}_{K(t-1)+\ell})_{K(t-1)+\ell,1}\, u_1$$
$$+ \sum_{i=2}^{K(t-1)+\ell} (\tilde{\mathbf{B}}_{K(t-1)+\ell})_{K(t-1)+\ell,i}\, \tilde{h}_i\Big(z_1, \dots, z_{i-1}, u_1, x^*\Big) + \tilde{\mu}_{K(t-1)+\ell} x^*, \quad \ell \in [K-1],$$

$$\tilde{h}_{Kt+1}(z_1, \dots, z_{Kt}, u_1, x^*) = g_{t+1}\Big(\mu_t x^* + \sum_{i=1}^{Kt} \theta_{t,i} z_i\Big). \tag{B.7}$$

The idea is that the choice (B.7) for the denoisers $\{\tilde{h}_{t+1}\}_{t \geq 1}$ ensures that $\tilde{\mathbf{u}}^{K(t-1)+\ell}$ tracks the quantity $\mathbf{Y}^{\ell-1}\mathbf{u}^t$ for $\ell \in [K]$ and $t \geq 1$, where $\{\mathbf{u}^t\}$ are the iterates of the AMP iteration (5.190) we are interested in analyzing.

In (B.7), g_{t+1} is the denoiser of the AMP (5.190). The parameters $(\tilde{\mathbf{B}}_{K(t-1)+\ell}, \tilde{\mu}_{K(t-1)+\ell}, \mu_t, \theta_{t,i})$ come from the state evolution recursion detailed in Section 5.6.2: $\tilde{\mathbf{B}}_{K(t-1)+\ell}$ is given by (5.201), $\tilde{\mu}_{K(t-1)+\ell}$ by (5.198), μ_t by (5.206) and $\theta_{t,i}$ by (5.207). We now discuss how to obtain the coefficients $\{\bar{\mathsf{b}}_{t,i}\}_{i=1}^{t}$ needed in (B.6). Let us define the matrix $\bar{\mathbf{\Phi}}_t \in \mathbb{R}^{t \times t}$ as

$$(\bar{\mathbf{\Phi}}_t)_{i,j} = 0, \quad \text{for } i \leq j, \qquad (\bar{\mathbf{\Phi}}_t)_{i,j} = \langle \partial_j \tilde{\mathbf{u}}^i \rangle, \quad \text{for } i > j, \tag{B.8}$$

where, for $j < i$, the vector $\langle \partial_j \tilde{\mathbf{u}}^i \rangle \in \mathbb{R}^N$ denotes the partial derivative of $\tilde{h}_i : \mathbb{R}^{i+1} \to \mathbb{R}$ with respect to the j-th input (applied component-wise). Then, the vector $(\bar{\mathsf{b}}_{t,1}, \dots, \bar{\mathsf{b}}_{t,t})$ is given by the last row of the matrix $\bar{\mathbf{B}}_t \in \mathbb{R}^{t \times t}$ defined as

$$\bar{\mathbf{B}}_t = \sum_{j=0}^{t-1} \kappa_{j+1} \bar{\mathbf{\Phi}}_t^j. \tag{B.9}$$

where $\{\kappa_k\}_{k \geq 1}$ denotes the sequence of free cumulants associated to the matrix $\mathbf{Z}$.

B.2.2 State evolution of auxiliary AMP

Using Theorem 2.3 in [133], we provide a state evolution result for the auxiliary AMP (B.6). In particular, we show in Proposition B.1 that the joint empirical distribution of $(\tilde{\mathbf{z}}^1, \dots, \tilde{\mathbf{z}}^t)$ converges to a t-dimensional Gaussian $\mathcal{N}(\mathbf{0}, \hat{\mathbf{\Sigma}}_t)$.

The covariance matrices $\{\hat{\boldsymbol{\Sigma}}_t\}_{t \geq 1}$ are defined recursively, starting with $\hat{\boldsymbol{\Sigma}}_1 = \bar{\kappa}_2 \mathbb{E}[U_1^2]$, where U_1 is defined in (5.192). Given $\hat{\boldsymbol{\Sigma}}_t$, let

$$
\begin{aligned}
(\hat{Z}_1, \ldots, \hat{Z}_t) &\sim \mathcal{N}(\mathbf{0}, \hat{\boldsymbol{\Sigma}}_t) \text{ and independent of } (X^*, U_1), \\
\hat{U}_s &= \tilde{h}_s\Big(\hat{Z}_1, \ldots, \hat{Z}_{s-1}, U_1, X^*\Big), \quad \text{for } s \in \{2, \ldots, t+1\},
\end{aligned}
\tag{B.10}
$$

where $\tilde{h}_s$ is defined via (B.7) and we set $\hat{U}_1 = U_1$. Let $\hat{\boldsymbol{\Phi}}_{t+1}, \hat{\boldsymbol{\Delta}}_{t+1} \in \mathbb{R}^{(t+1) \times (t+1)}$ be matrices with entries given by

$$
\begin{aligned}
(\hat{\boldsymbol{\Phi}}_{t+1})_{i,j} &= 0, \quad \text{for } i \leq j, \qquad (\hat{\boldsymbol{\Phi}}_{t+1})_{i,j} = \mathbb{E}[\partial_j \hat{U}_i], \quad \text{for } i > j, \\
(\tilde{\boldsymbol{\Delta}}_{t+1})_{i,j} &= \mathbb{E}[\hat{U}_i \hat{U}_j], \quad 1 \leq i, j \leq t+1,
\end{aligned}
\tag{B.11}
$$

where $\partial_j \hat{U}_i$ denotes the partial derivative $\partial_{\hat{Z}_j} \tilde{h}_i(\hat{Z}_1, \ldots, \hat{Z}_{i-1}, U_1, X)$. Then, we compute the covariance matrix $\hat{\boldsymbol{\Sigma}}_{t+1}$ as

$$
\hat{\boldsymbol{\Sigma}}_{t+1} = \sum_{j=0}^{2t} \bar{\kappa}_{j+2} \sum_{i=0}^{j} (\hat{\boldsymbol{\Phi}}_{t+1})^i \hat{\boldsymbol{\Delta}}_{t+1} (\hat{\boldsymbol{\Phi}}_{t+1}^{\intercal})^{j-i}.
\tag{B.12}
$$

It can be verified that the $t \times t$ top left sub-matrix of $\hat{\boldsymbol{\Sigma}}_{t+1}$ is given by $\hat{\boldsymbol{\Sigma}}_t$.

Proposition B.1 (State evolution for auxiliary AMP). *Consider the auxiliary AMP in* (B.6) *and the state evolution random variables defined in* (B.10). *Let $\tilde{\psi} : \mathbb{R}^{2t+2} \to \mathbb{R}$ be a PL(2) function. Then, for each $t \geq 1$, we almost surely have*

$$
\begin{aligned}
\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} \tilde{\psi}(\tilde{z}_i^1, \ldots, \tilde{z}_i^t, \tilde{u}_i^1, \ldots, \tilde{u}_i^{t+1}, X_i^*) & \\
= \mathbb{E}[\tilde{\psi}(\hat{Z}_1, \ldots, \hat{Z}_t, \hat{U}_1, \ldots, \hat{U}_{t+1}, X^*)]. &
\end{aligned}
\tag{B.13}
$$

Equivalently, as $N \to \infty$, almost surely:

$$
(\tilde{\mathbf{z}}^1, \ldots, \tilde{\mathbf{z}}^t, \tilde{\mathbf{u}}^1, \ldots, \tilde{\mathbf{u}}^{t+1}, \mathbf{X}^*) \xrightarrow{W_2} (\hat{Z}_1, \ldots, \hat{Z}_t, \hat{U}_1, \ldots, \hat{U}_{t+1}, X^*).
\tag{B.14}
$$

Furthermore,

$$
(\hat{Z}_1, \ldots, \hat{Z}_t, \hat{U}_1, \ldots, \hat{U}_{t+1}, X^*) \stackrel{\mathrm{d}}{=} (\tilde{Z}_1, \ldots, \tilde{Z}_t, \tilde{U}_1, \ldots, \tilde{U}_{t+1}, X^*),
\tag{B.15}
$$

where $(\tilde{Z}_1, \ldots, \tilde{Z}_t, \tilde{U}_1, \ldots, \tilde{U}_{t+1}, X^*)$ are obtained via (5.195)–(5.197).

Proof. The result follows from Theorem 2.3 in [133]. In fact, Assumption 2.1 of [133] holds because of the model assumptions on $\mathbf{Z}$, Assumption 2.2(a) holds because $(\mathbf{X}^*, \tilde{\mathbf{u}}^1) = (\mathbf{X}^*, \mathbf{u}^1) \xrightarrow{W_2}$

(X^*, U_1) from (5.192), and Assumption 2.2(b) follows from the definition of $\tilde{h}_{t+1}$ in (B.7) and the fact that g_{t+1} is continuously differentiable and Lipschitz. As the auxiliary AMP in (B.6) is of the standard form for which the state evolution result of Theorem 2.3 in [133] holds, we readily obtain (B.14). The equivalence between (B.14) and (B.13) follows from [156, Corollary 7.21]. Finally, by inspecting the state evolution recursions (5.195)–(5.197) and (B.10) giving $(\tilde{Z}_1, \ldots, \tilde{Z}_t, \tilde{U}_1, \ldots, \tilde{U}_{t+1}, X^*)$ and $(\hat{Z}_1, \ldots, \hat{Z}_t, \hat{U}_1, \ldots, \hat{U}_{t+1}, X^*)$ respectively, (B.15) is readily obtained. $\qquad\square$

Proposition B.1 gives that the state evolution recursion discussed in Section 5.6.2 (cf. (5.195)–(5.197)) coincides with the state evolution tracking the iterates of the auxiliary AMP algorithm (B.6). In particular, $\tilde{\boldsymbol{\Delta}}_{3t} = \hat{\boldsymbol{\Delta}}_{3t}$, $\tilde{\boldsymbol{\Phi}}_{3t} = \hat{\boldsymbol{\Phi}}_{3t}$, and $\tilde{\boldsymbol{\Sigma}}_{3t} = \bar{\boldsymbol{\Sigma}}_{3t}$. Furthermore, in the proof of Theorem 5.1 contained in Appendix B.2.3, we will show that $\bar{\mathbf{B}}_{3t} \to \tilde{\mathbf{B}}_{3t}$ as $N \to \infty$.

B.2.3 Proof of Theorem 5.1

We start by presenting a useful technical lemma.

Lemma B.2. *Let $F : \mathbb{R}^t \to \mathbb{R}$ be a Lipschitz function, and let $\partial_k F$ denote its derivative with respect to the k-th argument, for $1 \le k \le t$. Assume that $\partial_k F$ is continuous almost everywhere in the k-th argument, for each k. Let $(V_1^{(m)}, \ldots, V_t^{(m)})$ be a sequence of random vectors in $\mathbb{R}^t$ converging in distribution to the random vector $(V_1, \ldots, V_t)$ as $m \to \infty$. Furthermore, assume that the distribution of $(V_1, \ldots, V_t)$ is absolutely continuous with respect to the Lebesgue measure. Then,*

$$\lim_{m \to \infty} \mathbb{E}[\partial_k F(V_1^{(m)}, \ldots, V_t^{(m)})] = \mathbb{E}[\partial_k F(V_1, \ldots, V_t)], \qquad 1 \le k \le t. \tag{B.16}$$

The result was proved for $t = 2$ in [99, Lemma 6]. The proof for $t > 2$ is basically the same, see also [156, Lemma 7.14]. At this point, we are ready to give the proof of Theorem 5.1.

Proof of Theorem 5.1. We show that, for any PL(2) function $\psi : \mathbb{R}^{2t+2} \to \mathbb{R}$, the following limit holds almost surely for $t \ge 1$:

$$\lim_{N \to \infty} \left| \frac{1}{N} \sum_{i=1}^N \psi\left(u_i^1, u_i^2, \ldots, u_i^{t+1}, f_i^1, f_i^2, \ldots, f_i^t, X_i^*\right) \right.$$
$$\left. - \frac{1}{N} \sum_{i=1}^N \psi\left(\tilde{u}_i^1, \tilde{u}_i^{K+1}, \ldots, \tilde{u}_i^{Kt+1}, \tilde{f}_i^1, \tilde{f}_i^2, \ldots, \tilde{f}_i^t, X_i^*\right) \right| = 0, \tag{B.17}$$

where we have defined for $s \in \{1, \ldots, t\}$,

$$\tilde{\mathbf{f}}^s = \mu_s \mathbf{X}^* + \sum_{i=1}^{Ks} \theta_{s,i} \tilde{\mathbf{z}}^i. \tag{B.18}$$

From here till the end of the argument, all the limits hold almost surely, and we use C to denote a generic positive constant, which can change from line to line and is independent of N. By using that ψ is pseudo-Lipschitz, we have that

$$
\begin{aligned}
\Big| \frac{1}{N} \sum_{i=1}^{N} & \psi\big(u_i^1, u_i^2, \ldots, u_i^{t+1}, f_i^1, f_i^2, \ldots, f_i^t, X_i^*\big) \\
& - \frac{1}{N} \sum_{i=1}^{N} \psi\big(\tilde{u}_i^1, \tilde{u}_i^{K+1}, \ldots, \tilde{u}_i^{Kt+1}, \tilde{f}_i^1, \tilde{f}_i^2, \ldots, \tilde{f}_i^t, X_i^*\big) \Big| \\
\le \frac{C}{N} \sum_{i=1}^{N} & \Big(1 + |X_i^*| + 2|u_i^1| + \sum_{k=1}^{t} \big(|f_i^k| + |\tilde{f}_i^k| + |u_i^{k+1}| + |\tilde{u}_i^{Kk+1}| \big) \Big) \\
& \qquad\qquad \cdot \Big(\sum_{k=1}^{t} \big(|f_i^k - \tilde{f}_i^k|^2 + |u_i^{k+1} - \tilde{u}_i^{Kk+1}|^2 \big) \Big)^{1/2} \\
\le C(4t+3) & \Big[1 + \frac{\|\mathbf{X}^*\|^2}{N} + \sum_{k=1}^{t} \Big(\frac{\|\mathbf{f}^k\|^2}{N} + \frac{\|\tilde{\mathbf{f}}^k\|^2}{N} + \frac{\|\mathbf{u}^{k+1}\|^2}{N} + \frac{\|\tilde{\mathbf{u}}^{Kk+1}\|^2}{N} \Big) \Big]^{1/2} \\
& \qquad\qquad \cdot \Big(\sum_{k=1}^{t} \Big(\frac{\|\mathbf{f}^k - \tilde{\mathbf{f}}^k\|^2}{N} + \frac{\|\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{Kk+1}\|^2}{N} \Big) \Big)^{1/2},
\end{aligned}
\tag{B.19}
$$

where the last step uses twice Cauchy-Schwarz inequality. We now inductively show that as $N \to \infty$: *(i)* each of the terms in the last line of (B.19) converges to zero, and *(ii)* the terms within the square brackets in (B.19) all converge to finite, deterministic limits. To achieve this goal, we will also show that, for $k \in [t]$ and $\ell \in [K-1]$,

$$
\lim_{N\to\infty} \frac{\|\mathbf{Y}^\ell \mathbf{u}^k - \tilde{\mathbf{u}}^{K(k-1)+1+\ell}\|^2}{N} = 0,
\tag{B.20}
$$

$$
\lim_{N\to\infty} \frac{\|\tilde{\mathbf{u}}^{K(k-1)+1+\ell} - \sum_{j=1}^{K(k-1)+\ell} \alpha_{K(k-1)+1+\ell,j} \tilde{\mathbf{z}}^j - \sum_{j=1}^{k} \beta_{K(k-1)+1+\ell,j} \mathbf{u}^j - \gamma_{K(k-1)+1+\ell} \mathbf{X}^*\|^2}{N}
$$
$$
= 0. \quad \text{(B.21)}
$$

The limit (B.20) formalizes the idea discussed in Section 5.6.1 (see (5.191)) that the iterate $\tilde{\mathbf{u}}^{K(k-1)+1+\ell}$ of the auxiliary AMP tracks the quantity $\mathbf{Y}^\ell \mathbf{u}^k$, where $\mathbf{u}^k$ is the iterate of the AMP we wish to analyze, up to an $o_N(1)$ error. The limit (B.21) formalizes the interpretation of the coefficients $\{\alpha_{i,j}\}$, $\{\beta_{i,j}\}$, $\{\gamma_i\}$ provided at the end of Section 5.6.2 (see (5.212)).

<u>Base case ($t = 1$).</u> We have that

$$
\begin{aligned}
\mathbf{Y}\mathbf{u}^1 - \tilde{\mathbf{u}}^2 &= \mathbf{Z}\mathbf{u}^1 + \lambda \frac{\langle \mathbf{X}^*, \mathbf{u}^1 \rangle}{N} \mathbf{X}^* - \tilde{\mathbf{z}}^1 - (\tilde{\mathbf{B}}_1)_{1,1} \mathbf{u}^1 - \tilde{\mu}_1 \mathbf{X}^* \\
&= \Big(\lambda \frac{\langle \mathbf{X}^*, \mathbf{u}^1 \rangle}{N} - \tilde{\mu}_1 \Big) \mathbf{X}^* + \big(\bar{\mathsf{b}}_{1,1} - (\tilde{\mathbf{B}}_1)_{1,1} \big) \mathbf{u}^1,
\end{aligned}
\tag{B.22}
$$

where the first equality uses the definition of $\mathbf{Y}$ and of $\tilde{h}_2$ (see (B.7)), and the second equality uses (B.6) and that $\tilde{\mathbf{u}}^1 = \mathbf{u}^1$. Hence, by triangle inequality,

$$
\begin{aligned}
\frac{\|\mathbf{Y}\mathbf{u}^1 - \tilde{\mathbf{u}}^2\|^2}{N} &\le 2\Big(\lambda\frac{\langle\mathbf{X}^*, \mathbf{u}^1\rangle}{N} - \tilde{\mu}_1\Big)^2\frac{\|\mathbf{X}^*\|^2}{N} + 2\big(\bar{\mathsf{b}}_{1,1} - (\tilde{\mathbf{B}}_1)_{1,1}\big)^2\frac{\|\mathbf{u}^1\|^2}{N} \\
&\le C\Big(\Big(\lambda\frac{\langle\mathbf{X}^*, \mathbf{u}^1\rangle}{N} - \tilde{\mu}_1\Big)^2 + \big(\bar{\mathsf{b}}_{1,1} - (\tilde{\mathbf{B}}_1)_{1,1}\big)^2\Big),
\end{aligned}
\tag{B.23}
$$

where the last inequality uses that $(\mathbf{X}^*, \mathbf{u}^1)$ converges in W_2 to a pair of random variables with finite second moments. As $\tilde{\mu}_1 = \lambda\epsilon$ (cf. (5.194)), we have

$$
\lim_{N\to\infty} \lambda\frac{\langle\mathbf{X}^*, \mathbf{u}^1\rangle}{N} = \lambda\mathbb{E}[U_1 X^*] = \lambda\epsilon = \tilde{\mu}_1.
\tag{B.24}
$$

Furthermore, note that $(\tilde{\mathbf{B}}_1)_{1,1} = \bar{\kappa}_1$ (cf. (5.194)) and $\bar{\mathsf{b}}_{1,1} = \kappa_1$ (cf. (B.9)). Hence, by the model assumptions, as $N \to \infty$, $\kappa_1 \to \bar{\kappa}_1$ and, therefore, $\bar{\mathsf{b}}_{1,1} \to (\tilde{\mathbf{B}}_1)_{1,1}$. By combining this observation with (B.23) and (B.24), we obtain that (B.20) holds for $k = 1$ and $\ell = 1$.

By using (5.203)–(5.205), we readily obtain that $\alpha_{2,1} = 1$, $\beta_{2,1} = (\tilde{\mathbf{B}}_1)_{1,1}$ and $\gamma_2 = \tilde{\mu}_1$. Hence, by using the definition (B.7) of $\tilde{h}_2$, we obtain that (B.21) holds for $k = 1$ and $\ell = 1$.

Next, by using the definitions of $\mathbf{Y}$, of the auxiliary AMP (B.6) and of $\tilde{h}_3$ (cf. (B.7)), we have

$$
\begin{aligned}
\mathbf{Y}^2\mathbf{u}^1 - \tilde{\mathbf{u}}^3 &= \mathbf{Y}(\mathbf{Y}\mathbf{u}^1 - \tilde{\mathbf{u}}^2) + \mathbf{Y}\tilde{\mathbf{u}}^2 - \tilde{\mathbf{z}}^2 - (\tilde{\mathbf{B}}_2)_{2,1}\mathbf{u}^1 - (\tilde{\mathbf{B}}_2)_{2,2}\tilde{\mathbf{u}}^2 - \tilde{\mu}_2\mathbf{X}^* \\
&= \mathbf{Y}(\mathbf{Y}\mathbf{u}^1 - \tilde{\mathbf{u}}^2) + \mathbf{Z}\tilde{\mathbf{u}}^2 - \tilde{\mathbf{z}}^2 - (\tilde{\mathbf{B}}_2)_{2,1}\mathbf{u}^1 - (\tilde{\mathbf{B}}_2)_{2,2}\tilde{\mathbf{u}}^2 + \Big(\lambda\frac{\langle\mathbf{X}^*, \tilde{\mathbf{u}}^2\rangle}{N} - \tilde{\mu}_2\Big)\mathbf{X}^* \\
&= \mathbf{Y}(\mathbf{Y}\mathbf{u}^1 - \tilde{\mathbf{u}}^2) + \big(\bar{\mathsf{b}}_{2,1} - (\tilde{\mathbf{B}}_2)_{2,1}\big)\mathbf{u}^1 + \big(\bar{\mathsf{b}}_{2,2} - (\tilde{\mathbf{B}}_2)_{2,2}\big)\tilde{\mathbf{u}}^2 + \Big(\lambda\frac{\langle\mathbf{X}^*, \tilde{\mathbf{u}}^2\rangle}{N} - \tilde{\mu}_2\Big)\mathbf{X}^*.
\end{aligned}
\tag{B.25}
$$

Hence, by triangle inequality,

$$
\begin{aligned}
\frac{\|\mathbf{Y}^2\mathbf{u}^1 - \tilde{\mathbf{u}}^3\|^2}{N} &\le C\Big(\frac{\|\mathbf{Y}(\mathbf{Y}\mathbf{u}^1 - \tilde{\mathbf{u}}^2)\|^2}{N} + \big(\bar{\mathsf{b}}_{2,1} - (\tilde{\mathbf{B}}_2)_{2,1}\big)^2\frac{\|\mathbf{u}^1\|^2}{N} \\
&\quad + \big(\bar{\mathsf{b}}_{2,2} - (\tilde{\mathbf{B}}_2)_{2,2}\big)^2\frac{\|\tilde{\mathbf{u}}^2\|^2}{N} + \Big(\lambda\frac{\langle\mathbf{X}^*, \tilde{\mathbf{u}}^2\rangle}{N} - \tilde{\mu}_2\Big)^2\frac{\|\mathbf{X}^*\|^2}{N}\Big) \\
&:= C(T_1 + T_2 + T_3 + T_4).
\end{aligned}
\tag{B.26}
$$

Consider the first term. As $\mathbf{Y}$ has bounded operator norm and (B.20) holds for $k = 1$ and $\ell = 1$, we have that $T_1 \to 0$ as $N \to \infty$.

Consider the second and third terms. The following chain of equalities holds

$$
\lim_{N\to\infty} (\bar{\boldsymbol{\Phi}}_2)_{2,1} = \lim_{N\to\infty} \langle\partial_1\tilde{\mathbf{u}}^2\rangle = \mathbb{E}[\partial_1\hat{U}_2] = \mathbb{E}[\partial_1\tilde{U}_2] = (\tilde{\boldsymbol{\Phi}}_2)_{2,1}.
\tag{B.27}
$$

Here, the first equality uses the definition (B.8); the second equality follows from Lemma B.2, as $\tilde{\mathbf{u}}^2$ converges in W_2 (and therefore in distribution) to $\tilde{U}_2$ and $\partial_1\tilde{U}_2$ is continuous; the third

equality uses (B.15); and the fourth equality uses the definition of $(\tilde{\boldsymbol{\Phi}}_2)_{2,1}$ in (5.200). By the model assumptions, as $N \to \infty$, $\kappa_j \to \bar{\kappa}_j$ for all j. Thus, by combining (B.27) with the definitions of $\bar{\mathbf{B}}_2$ and $\tilde{\mathbf{B}}_2$ in (B.9) and (5.201), respectively, we conclude that, as $N \to \infty$, $\bar{\mathsf{b}}_{2,i} \to (\tilde{\mathbf{B}}_2)_{2,i}$ for $i \in \{1, 2\}$. By Proposition B.1, $\|\tilde{\mathbf{u}}^2\|^2/N$ converges to a finite limit, hence we conclude that $T_2, T_3 \to 0$ as $N \to \infty$.

Consider the fourth term. Then,

$$\lim_{N \to \infty} \lambda \frac{\langle \mathbf{X}^*, \tilde{\mathbf{u}}^2 \rangle}{N} = \lambda \mathbb{E}[X^* \tilde{U}_2] = \tilde{\mu}_2.$$

Here, the first equality uses Proposition B.1 and the second equality uses the definition of $\tilde{\mu}_2$ in (5.198). As $\|\mathbf{X}^*\|^2/N = 1$, we conclude that $T_4 \to 0$ as $N \to \infty$. This proves that the RHS of (B.26) vanishes and gives that (B.20) holds for $k = 1$ and $\ell = 2$.

By using (5.203)–(5.205), we readily obtain that $\alpha_{3,1} = (\tilde{\mathbf{B}}_2)_{2,2}$, $\alpha_{3,2} = 1$, $\beta_{3,1} = (\tilde{\mathbf{B}}_2)_{2,1} + (\tilde{\mathbf{B}}_2)_{2,2}(\tilde{\mathbf{B}}_2)_{1,1}$ and $\gamma_3 = \tilde{\mu}_2 + \tilde{\mu}_1 (\tilde{\mathbf{B}}_2)_{2,2}$. Hence, by using the definition (B.7) of $\tilde{h}_3$, we obtain that (B.21) holds for $k = 1$ and $\ell = 2$.

The proof of (B.20)–(B.21) for $k = 1$ and $\ell \in \{3, \ldots, K-1\}$ follows from similar arguments. In particular, we write

$$\mathbf{Y}^\ell \mathbf{u}^1 - \tilde{\mathbf{u}}^{1+\ell} = \mathbf{Y}(\mathbf{Y}^{\ell-1}\mathbf{u}^1 - \tilde{\mathbf{u}}^\ell) + \mathbf{Y}\tilde{\mathbf{u}}^\ell - \tilde{\mathbf{z}}^\ell - (\tilde{\mathbf{B}}_\ell)_{\ell,1}\mathbf{u}^1 - \sum_{j=2}^\ell (\tilde{\mathbf{B}}_\ell)_{\ell,j}\tilde{\mathbf{u}}^j - \tilde{\mu}_\ell \mathbf{X}^*$$

$$= \mathbf{Y}(\mathbf{Y}^{\ell-1}\mathbf{u}^1 - \tilde{\mathbf{u}}^\ell) + \mathbf{Z}\tilde{\mathbf{u}}^\ell - \tilde{\mathbf{z}}^\ell - (\tilde{\mathbf{B}}_\ell)_{\ell,1}\mathbf{u}^1 - \sum_{j=2}^\ell (\tilde{\mathbf{B}}_\ell)_{\ell,j}\tilde{\mathbf{u}}^j + \left(\lambda \frac{\langle \mathbf{X}^*, \tilde{\mathbf{u}}^\ell \rangle}{N} - \tilde{\mu}_\ell\right)\mathbf{X}^* \qquad \text{(B.28)}$$

$$= \mathbf{Y}(\mathbf{Y}^{\ell-1}\mathbf{u}^1 - \tilde{\mathbf{u}}^\ell) + \left(\bar{\mathsf{b}}_{\ell,1} - (\tilde{\mathbf{B}}_\ell)_{\ell,1}\right)\mathbf{u}^1 + \sum_{j=2}^\ell \left(\bar{\mathsf{b}}_{\ell,j} - (\tilde{\mathbf{B}}_\ell)_{\ell,j}\right)\tilde{\mathbf{u}}^j$$

$$+ \left(\lambda \frac{\langle \mathbf{X}^*, \tilde{\mathbf{u}}^\ell \rangle}{N} - \tilde{\mu}_\ell\right)\mathbf{X}^*,$$

which by triangle inequality gives

$$\frac{\|\mathbf{Y}^\ell \mathbf{u}^1 - \tilde{\mathbf{u}}^{1+\ell}\|^2}{N} \le C\Big(\frac{\|\mathbf{Y}(\mathbf{Y}^{\ell-1}\mathbf{u}^1 - \tilde{\mathbf{u}}^\ell)\|^2}{N} + \left(\bar{\mathsf{b}}_{\ell,1} - (\tilde{\mathbf{B}}_\ell)_{\ell,1}\right)^2 \frac{\|\mathbf{u}^1\|^2}{N}$$

$$+ \sum_{j=2}^\ell \left(\bar{\mathsf{b}}_{\ell,j} - (\tilde{\mathbf{B}}_\ell)_{\ell,j}\right)^2 \frac{\|\tilde{\mathbf{u}}^j\|^2}{N} + \left(\lambda \frac{\langle \mathbf{X}^*, \tilde{\mathbf{u}}^\ell \rangle}{N} - \tilde{\mu}_\ell\right)^2 \frac{\|\mathbf{X}^*\|^2}{N}\Big). \qquad \text{(B.29)}$$

As $\mathbf{Y}$ has bounded operator norm and $\|\mathbf{Y}^{\ell-1}\mathbf{u}^1 - \tilde{\mathbf{u}}^\ell\|^2/N \to 0$ (by the previous step), we have that

$$\lim_{N \to \infty} \frac{\|\mathbf{Y}(\mathbf{Y}^{\ell-1}\mathbf{u}^1 - \tilde{\mathbf{u}}^\ell)\|^2}{N} = 0.$$

Next, by following passages analogous to those in (B.27), we have that $\lim_{N\to\infty} \bar{\Phi}_\ell = \tilde{\Phi}_\ell$. As $\kappa_j \to \bar{\kappa}_j$ for all j, this implies that $\lim_{N\to\infty} \bar{\mathbf{B}}_\ell = \tilde{\mathbf{B}}_\ell$. Hence, for all $j \in [\ell]$, as $\|\tilde{\mathbf{u}}^j\|/N$ is bounded, we have that

$$\lim_{N\to\infty} \left(\left(\bar{\mathsf{b}}_{\ell,1} - (\tilde{\mathbf{B}}_\ell)_{\ell,1}\right)^2 \frac{\|\mathbf{u}^1\|^2}{N} + \sum_{j=2}^{\ell} \left(\bar{\mathsf{b}}_{\ell,j} - (\tilde{\mathbf{B}}_\ell)_{\ell,j}\right)^2 \frac{\|\tilde{\mathbf{u}}^j\|^2}{N} \right) = 0.$$

Finally, as

$$\lim_{N\to\infty} \lambda \frac{\langle \mathbf{X}^*, \tilde{\mathbf{u}}^\ell \rangle}{N} = \lambda \mathbb{E}[X^* \tilde{U}_\ell] = \tilde{\mu}_\ell,$$

we conclude that the last term in the RHS of (B.29) vanishes as well, which proves that (B.20) holds for $k = 1$ and a generic $\ell \in \{3, \dots, K-1\}$. Furthermore, by using (5.203)–(5.205) and the definition (B.7) of $\tilde{h}_{\ell+1}$, one can readily verify that (B.21) holds for $k = 1$ and a generic $\ell \in \{3, \dots, K-1\}$.

By using (B.6) and the definition of $\mathbf{Y}$, we have that

$$\mathbf{Y}^K \mathbf{u}^1 - \tilde{\mathbf{z}}^K - \sum_{i=1}^{K} \bar{\mathsf{b}}_{K,i} \tilde{\mathbf{u}}^i - \tilde{\mu}_K \mathbf{X}^* = \mathbf{Z}\left(\mathbf{Y}^{K-1}\mathbf{u}^1 - \tilde{\mathbf{u}}^K\right) + \left(\lambda \frac{\langle \mathbf{X}^*, \mathbf{Y}^{K-1}\mathbf{u}^1 \rangle}{N} - \tilde{\mu}_K\right)\mathbf{X}^*. \quad (B.30)$$

Hence, by using the definition of $\tilde{\mu}_K$ in (5.198) and (B.20) with $k = 1$, $\ell = K-1$, we obtain

$$\lim_{N\to\infty} \frac{\|\mathbf{Y}^K \mathbf{u}^1 - \tilde{\mathbf{z}}^K - \sum_{i=1}^{K} \bar{\mathsf{b}}_{K,i} \tilde{\mathbf{u}}^i - \tilde{\mu}_K \mathbf{X}^*\|^2}{N} = 0. \quad (B.31)$$

Recall that $\mathbf{J}(\mathbf{Y}) = \sum_{j=1}^{K} c_j \mathbf{Y}^j$. Then, by combining (B.20) with $k = 1$ and (B.31), we have

$$\lim_{N\to\infty} \frac{\|\mathbf{J}(\mathbf{Y})\mathbf{u}^1 - \sum_{j=1}^{K} c_j\left(\tilde{\mathbf{z}}^j + \sum_{i=1}^{j} \bar{\mathsf{b}}_{j,i} \tilde{\mathbf{u}}^i + \tilde{\mu}_j \mathbf{X}^*\right)\|^2}{N} = 0. \quad (B.32)$$

By following the same argument as in (B.27), we have that $\lim_{N\to\infty} \bar{\Phi}_K = \tilde{\Phi}_K$. As $\kappa_j \to \bar{\kappa}_j$ for all j, this implies that $\lim_{N\to\infty} \bar{\mathbf{B}}_K = \tilde{\mathbf{B}}_K$. Therefore,

$$\lim_{N\to\infty} \frac{\left\| \sum_{j=1}^{K} c_j\left(\tilde{\mathbf{z}}^j + \sum_{i=1}^{j} \bar{\mathsf{b}}_{j,i} \tilde{\mathbf{u}}^i + \tilde{\mu}_j \mathbf{X}^*\right) - \sum_{j=1}^{K} c_j\left(\tilde{\mathbf{z}}^j + \sum_{i=1}^{j} (\tilde{\mathbf{B}}_j)_{j,i} \tilde{\mathbf{u}}^i + \tilde{\mu}_j \mathbf{X}^*\right) \right\|^2}{N} = 0. \quad (B.33)$$

Recall that $\tilde{\mathbf{u}}^1 = \mathbf{u}^1$ and (B.21) holds for $k = 1$. Hence, by plugging in the formulas for $\mathsf{c}_{1,1}$, μ_1 and $\{\theta_{1,i}\}_{i\in[K]}$ (cf. (5.208), (5.206) and (5.207)), we have

$$\lim_{N\to\infty} \frac{\left\| \sum_{j=1}^{K} c_j\left(\tilde{\mathbf{z}}^j + \sum_{i=1}^{j} (\tilde{\mathbf{B}}_j)_{j,i} \tilde{\mathbf{u}}^i + \tilde{\mu}_j \mathbf{X}^*\right) - \mathsf{c}_{1,1}\mathbf{u}^1 - \mu_1 \mathbf{X}^* - \sum_{i=1}^{K} \theta_{1,i} \tilde{\mathbf{z}}^i \right\|^2}{N} = 0. \quad (B.34)$$

By combining (B.32)–(B.34) with the definitions of $\mathbf{f}^1$ and $\tilde{\mathbf{f}}^1$(cf. (5.190) and (B.18)), we conclude that

$$\lim_{N\to\infty} \frac{\|\mathbf{f}^1 - \tilde{\mathbf{f}}^1\|^2}{N} = 0. \tag{B.35}$$

As g_2 is Lipschitz, (B.35) immediately implies that

$$\lim_{N\to\infty} \frac{\|\mathbf{u}^2 - \tilde{\mathbf{u}}^{K+1}\|^2}{N} = 0. \tag{B.36}$$

An application of the triangle inequality gives that, for any $i \geq 1$,

$$\|\tilde{\mathbf{f}}^i\| - \|\mathbf{f}^i - \tilde{\mathbf{f}}^i\| \leq \|\mathbf{f}^i\| \leq \|\tilde{\mathbf{f}}^i\| + \|\mathbf{f}^i - \tilde{\mathbf{f}}^i\|,$$
$$\|\tilde{\mathbf{u}}^{Ki+1}\| - \|\mathbf{u}^{i+1} - \tilde{\mathbf{u}}^{Ki+1}\| \leq \|\mathbf{u}^{i+1}\| \leq \|\tilde{\mathbf{u}}^{Ki+1}\| + \|\mathbf{u}^{i+1} - \tilde{\mathbf{u}}^{Ki+1}\|. \tag{B.37}$$

Thus, by using (B.37) with $i = 1$ and Proposition B.1, we obtain that

$$\lim_{N\to\infty} \frac{\|\mathbf{f}^1\|^2}{N} = \lim_{N\to\infty} \frac{\|\tilde{\mathbf{f}}^1\|^2}{N} = \mathbb{E}\Big[\Big(\mu_1 X^* + \sum_{i=1}^{K} \theta_{1,i}\tilde{Z}_i\Big)^2\Big],$$
$$\lim_{N\to\infty} \frac{\|\mathbf{u}^2\|^2}{N} = \lim_{N\to\infty} \frac{\|\tilde{\mathbf{u}}^{K+1}\|^2}{N} = \mathbb{E}[(\tilde{U}_{K+1})^2], \tag{B.38}$$

which concludes the base step.

Induction step. Assume towards induction that (B.20)–(B.21) hold for $k \in [t]$, $\ell \in [K-1]$ and that, for $k \in [t]$,

$$\lim_{N\to\infty} \frac{\|\mathbf{f}^k - \tilde{\mathbf{f}}^k\|^2}{N} = 0, \tag{B.39}$$

$$\lim_{N\to\infty} \frac{\|\mathbf{u}^{k+1} - \tilde{\mathbf{u}}^{Kk+1}\|^2}{N} = 0, \tag{B.40}$$

$$\lim_{N\to\infty} \frac{\|\mathbf{f}^k\|^2}{N} = \lim_{N\to\infty} \frac{\|\tilde{\mathbf{f}}^k\|^2}{N} = \mathbb{E}\Big[\Big(\mu_k X^* + \sum_{i=1}^{Kk} \theta_{k,i}\tilde{Z}_i\Big)^2\Big], \tag{B.41}$$

$$\lim_{N\to\infty} \frac{\|\mathbf{u}^{k+1}\|^2}{N} = \lim_{N\to\infty} \frac{\|\tilde{\mathbf{u}}^{Kk+1}\|^2}{N} = \mathbb{E}[\tilde{U}_{Kk+1}^2]. \tag{B.42}$$

We now show that (B.39)–(B.42) hold for $k = t+1$, and that (B.20)–(B.21) hold for $k = t+1$, $\ell \in [K-1]$. By doing so, we will have proved also the induction step and consequently that (B.17) holds.

Using similar passages as in (B.22), we obtain

$$\mathbf{Y}\mathbf{u}^{t+1} - \tilde{\mathbf{u}}^{Kt+2} = \mathbf{Z}\mathbf{u}^{t+1} + \lambda\frac{\langle \mathbf{X}^*, \mathbf{u}^{t+1}\rangle}{N}\mathbf{X}^* - \tilde{\mathbf{z}}^{Kt+1} - \sum_{i=1}^{Kt+1}(\tilde{\mathbf{B}}_{Kt+1})_{Kt+1,i}\tilde{\mathbf{u}}^i - \tilde{\mu}_{Kt+1}\mathbf{X}^*$$
$$= \mathbf{Z}(\mathbf{u}^{t+1} - \tilde{\mathbf{u}}^{Kt+1}) + \Big(\lambda\frac{\langle \mathbf{X}^*, \mathbf{u}^{t+1}\rangle}{N} - \tilde{\mu}_{Kt+1}\Big)\mathbf{X}^* + \sum_{i=1}^{Kt+1}\big(\bar{\mathsf{b}}_{Kt+1,i} - (\tilde{\mathbf{B}}_{Kt+1})_{Kt+1,i}\big)\tilde{\mathbf{u}}^i. \tag{B.43}$$

Hence, by triangle inequality,

$$
\frac{\|\mathbf{Y}\mathbf{u}^{t+1} - \tilde{\mathbf{u}}^{Kt+2}\|^2}{N} \le C\Big(\frac{\|\mathbf{Z}(\mathbf{u}^{t+1} - \tilde{\mathbf{u}}^{Kt+1})\|^2}{N}
$$
$$
+ \Big(\lambda\frac{\langle\mathbf{X}^*, \mathbf{u}^{t+1}\rangle}{N} - \tilde{\mu}_{Kt+1}\Big)^2\frac{\|\mathbf{X}^*\|^2}{N} + \sum_{i=1}^{Kt+1}\big(\bar{\mathbf{b}}_{Kt+1,i} - (\tilde{\mathbf{B}}_{Kt+1})_{Kt+1,i}\big)^2\frac{\|\tilde{\mathbf{u}}^i\|^2}{N}\Big) \tag{B.44}
$$
$$
:= C(\bar{T}_1 + \bar{T}_2 + \bar{T}_3).
$$

Consider the first term. Since $\|\mathbf{Z}\|_{\mathrm{op}} \le C$, the induction hypothesis (B.40) implies that $\bar{T}_1 \to 0$ as $N \to \infty$.

Consider the second term. The following chain of equalities holds:

$$
\lim_{N\to\infty}\lambda\frac{\langle\mathbf{X}^*, \mathbf{u}^{t+1}\rangle}{N} = \lim_{N\to\infty}\lambda\frac{\langle\mathbf{X}^*, \tilde{\mathbf{u}}^{Kt+1}\rangle}{N} = \lambda\mathbb{E}[X\,\tilde{U}_{Kt+1}] = \tilde{\mu}_{Kt+1}. \tag{B.45}
$$

Here, the first equality uses (B.40) together with the fact that $\|\mathbf{X}^*\|^2/N = 1$; the second equality follows from Proposition B.1; and the third equality uses the definition of $\tilde{\mu}_{Kt+1}$ in (5.198). Finally, using (B.45) and again that $\|\mathbf{X}^*\|^2/N = 1$ gives that $\bar{T}_2 \to 0$ as $N \to \infty$.

Consider the third term. By following the same argument as in (B.27), we have that $\lim_{N\to\infty}\bar{\boldsymbol{\Phi}}_{Kt+1} = \tilde{\boldsymbol{\Phi}}_{Kt+1}$. As $\kappa_j \to \bar{\kappa}_j$ for all j, this implies that $\lim_{N\to\infty}\bar{\mathbf{B}}_{Kt+1} = \tilde{\mathbf{B}}_{Kt+1}$. By using the induction hypothesis (B.42), which shows that $\|\tilde{\mathbf{u}}^i\|^2/N$ converges to a finite limit, we conclude that $\bar{T}_3 \to 0$ as $N \to \infty$. This proves that the RHS of (B.44) vanishes and gives that (B.20) holds for $k = t+1$ and $\ell = 1$.

For $\ell \in \{2, \dots, K-1\}$, by following passages similar to (B.28), we have

$$
\mathbf{Y}^\ell\mathbf{u}^{t+1} - \tilde{\mathbf{u}}^{Kt+\ell+1} = \mathbf{Y}(\mathbf{Y}^{\ell-1}\mathbf{u}^{t+1} - \tilde{\mathbf{u}}^{Kt+\ell}) + \sum_{i=1}^{Kt+\ell}\big(\bar{\mathbf{b}}_{Kt+\ell,i} - (\tilde{\mathbf{B}}_{Kt+\ell})_{Kt+\ell,i}\big)\tilde{\mathbf{u}}^i
$$
$$
+ \Big(\lambda\frac{\langle\mathbf{X}^*, \tilde{\mathbf{u}}^{Kt+\ell}\rangle}{N} - \tilde{\mu}_{Kt+\ell}\Big)\mathbf{X}^*,
$$

which by triangle inequality gives

$$
\frac{\|\mathbf{Y}^\ell\mathbf{u}^{t+1} - \tilde{\mathbf{u}}^{Kt+\ell+1}\|^2}{N} \le C\Big(\frac{\|\mathbf{Y}(\mathbf{Y}^{\ell-1}\mathbf{u}^{t+1} - \tilde{\mathbf{u}}^{Kt+\ell})\|^2}{N}
$$
$$
+ \sum_{i=1}^{Kt+\ell}\big(\bar{\mathbf{b}}_{Kt+\ell,i} - (\tilde{\mathbf{B}}_{Kt+\ell})_{Kt+\ell,i}\big)^2\frac{\|\tilde{\mathbf{u}}^i\|^2}{N} + \Big(\lambda\frac{\langle\mathbf{X}^*, \tilde{\mathbf{u}}^{Kt+\ell}\rangle}{N} - \tilde{\mu}_{Kt+\ell}\Big)^2\frac{\|\mathbf{X}^*\|^2}{N}\Big). \tag{B.46}
$$

The first term on the RHS of (B.46) vanishes as $\mathbf{Y}$ has bounded operator norm and we have just proved in the previous step that $\|\mathbf{Y}^{\ell-1}\mathbf{u}^{t+1} - \tilde{\mathbf{u}}^{Kt+\ell}\|^2/N \to 0$. To bound the second term, note that, by following the same argument as in (B.27), we have that $\lim_{N\to\infty}\bar{\boldsymbol{\Phi}}_{Kt+\ell} = \tilde{\boldsymbol{\Phi}}_{Kt+\ell}$. As $\kappa_j \to \bar{\kappa}_j$ for all j, this implies that $\lim_{N\to\infty}\bar{\mathbf{B}}_{Kt+\ell} = \tilde{\mathbf{B}}_{Kt+\ell}$. By using the induction hypothesis

(B.42), we have that $\|\tilde{\mathbf{u}}^i\|^2/N$ converges to a finite limit for $i \in [Kt + \ell - 1]$. Furthermore, as $\|\mathbf{Y}^{\ell-1}\mathbf{u}^{t+1} - \tilde{\mathbf{u}}^{Kt+\ell}\|^2/N \to 0$, we also have that $\|\tilde{\mathbf{u}}^{Kt+\ell}\|^2/N$ converges to a finite limit. As a result, the second term on the RHS of (B.46) vanishes. Finally, we can write a chain of equalities analogous to (B.45) with $Kt + \ell$ in place of $Kt + 1$, from which we deduce that the third term vanishes. This concludes the proof that (B.20) holds for $k = t + 1$ and $\ell \in [K - 1]$.

For $\ell \in [K - 1]$, by definition (B.7) of $h_{Kt+1+\ell}$, we have

$$\tilde{\mathbf{u}}^{Kt+1+\ell} = \tilde{\mathbf{z}}^{Kt+\ell} + \tilde{\mu}_{Kt+\ell}\mathbf{X}^* + \sum_{i=1}^{Kt+\ell} (\tilde{\mathbf{B}}_{Kt+\ell})_{Kt+\ell,i}\tilde{\mathbf{u}}^i. \tag{B.47}$$

Let us define:

$$
\begin{aligned}
\hat{\mathbf{u}}^{Kt+1+\ell} :=\ & \tilde{\mathbf{z}}^{Kt+\ell} + \tilde{\mu}_{Kt+\ell}\mathbf{X}^* + \sum_{i=1}^{t+1} (\tilde{\mathbf{B}}_{Kt+\ell})_{Kt+\ell,K(i-1)+1}\mathbf{u}^i \\
& + \sum_{\substack{i=1 \\ i \not\equiv 1(\mathrm{mod}K)}}^{Kt+\ell} (\tilde{\mathbf{B}}_{Kt+\ell})_{Kt+\ell,i}\Big(\sum_{j=1}^{i-1} \alpha_{i,j}\tilde{\mathbf{z}}^j + \sum_{j=1}^{\lceil (i-1)/K \rceil} \beta_{i,j}\mathbf{u}^j + \gamma_i\mathbf{X}^* \Big).
\end{aligned}
\tag{B.48}
$$

Then, by using the recursive definitions (5.203)–(5.205), we readily have that the RHS of (B.48) is equal to

$$\sum_{j=1}^{Kt+\ell} \alpha_{Kt+1+\ell,j}\tilde{\mathbf{z}}^j + \sum_{j=1}^{t+1} \beta_{Kt+1+\ell,j}\mathbf{u}^j + \gamma_{Kt+1+\ell}\mathbf{X}^*. \tag{B.49}$$

Recall that, by induction hypothesis, (B.40) holds for $k \in [t]$, and (B.21) holds for $k \in [t]$ and $\ell \in [K - 1]$. Thus, by using the expressions in (B.47) and (B.48) for $\ell = 1$, one readily obtains that

$$\lim_{N\to\infty} \frac{\|\tilde{\mathbf{u}}^{Kt+2} - \hat{\mathbf{u}}^{Kt+2}\|^2}{N} = 0. \tag{B.50}$$

Since the RHS of (B.48) is equal to the expression in (B.49) for $\ell = 1$, we conclude that (B.21) holds for $k = t + 1$ and $\ell = 1$. At this point, we have that (B.21) holds for $k \in [t]$, $\ell \in [K - 1]$ and also for $k = t + 1, \ell = 1$. Hence, by using the expressions in (B.47) and (B.48) for $\ell = 2$, we obtain

$$\lim_{N\to\infty} \frac{\|\tilde{\mathbf{u}}^{Kt+3} - \hat{\mathbf{u}}^{Kt+3}\|^2}{N} = 0. \tag{B.51}$$

Since the RHS of (B.48) is equal to the expression in (B.49) for $\ell = 2$, we conclude that (B.21) holds for $k = t + 1$, $\ell = 2$. By iterating this procedure for $\ell \in \{3, \ldots, K - 1\}$, we obtain that (B.21) holds for $k = t + 1$, $\ell \in [K - 1]$.

By using (B.6) and the definition of $\mathbf{Y}$, we have that

$$
\mathbf{Y}^K \mathbf{u}^{t+1} - \tilde{\mathbf{z}}^{K(t+1)} - \sum_{i=1}^{K(t+1)} \bar{\mathsf{b}}_{K(t+1),i} \tilde{\mathbf{u}}^i - \tilde{\mu}_{K(t+1)} \mathbf{X}^*
$$

$$
= \mathbf{Z} \big(\mathbf{Y}^{K-1} \mathbf{u}^{t+1} - \tilde{\mathbf{u}}^{K(t+1)} \big) + \Big(\lambda \frac{\langle \mathbf{X}^*, \mathbf{Y}^{K-1} \mathbf{u}^{t+1} \rangle}{N} - \tilde{\mu}_{K(t+1)} \Big) \mathbf{X}^*. \tag{B.52}
$$

Hence, by using (B.20) with $k = t+1$, $\ell = K-1$ and the definition of $\tilde{\mu}_{K(t+1)}$ in (5.198), we obtain

$$
\lim_{N \to \infty} \frac{\| \mathbf{Y}^K \mathbf{u}^{t+1} - \tilde{\mathbf{z}}^{K(t+1)} - \sum_{i=1}^{K(t+1)} \bar{\mathsf{b}}_{K(t+1),i} \tilde{\mathbf{u}}^i - \tilde{\mu}_{K(t+1)} \mathbf{X}^* \|^2}{N} = 0. \tag{B.53}
$$

As $\mathbf{J}(\mathbf{Y}) = \sum_{j=1}^K c_j \mathbf{Y}^j$, by combining (B.53) with (B.20) with $k = t+1$, $\ell \in [K-1]$, we have

$$
\lim_{N \to \infty} \frac{\| \mathbf{J}(\mathbf{Y}) \mathbf{u}^{t+1} - \sum_{j=1}^K c_j \big(\tilde{\mathbf{z}}^{Kt+j} + \sum_{i=1}^{Kt+j} \bar{\mathsf{b}}_{Kt+j,i} \tilde{\mathbf{u}}^i + \tilde{\mu}_{Kt+j} \mathbf{X}^* \big) \|^2}{N} = 0. \tag{B.54}
$$

By following the same argument as in (B.27), we have that $\lim_{N \to \infty} \bar{\boldsymbol{\Phi}}_{Kt+j} = \tilde{\boldsymbol{\Phi}}_{Kt+j}$ for all $j \in [K]$. As $\kappa_j \to \bar{\kappa}_j$ for all j, this implies that $\lim_{N \to \infty} \bar{\mathbf{B}}_{Kt+j} = \tilde{\mathbf{B}}_{Kt+j}$ for all $j \in [K]$. Therefore, (B.54) implies that

$$
\lim_{N \to \infty} \frac{\| \mathbf{J}(\mathbf{Y}) \mathbf{u}^{t+1} - \sum_{j=1}^K c_j \big(\tilde{\mathbf{z}}^{Kt+j} + \sum_{i=1}^{Kt+j} (\tilde{\mathbf{B}}_{Kt+j})_{Kt+j,i} \tilde{\mathbf{u}}^i + \tilde{\mu}_{Kt+j} \mathbf{X}^* \big) \|^2}{N} = 0. \tag{B.55}
$$

Recall that (B.40) holds for $k \in [t]$ by the induction hypothesis and (B.21) holds for $k \in [t+1]$, $\ell \in [K-1]$ (thanks to the induction hypothesis and the argument above). Hence, by plugging in the formulas for $\{\mathsf{c}_{t+1,i}\}_{i \in [t+1]}$, μ_{t+1} and $\{\theta_{t+1,i}\}_{i \in [K(t+1)]}$ (cf. (5.208), (5.206) and (5.207)), we have

$$
\lim_{N \to \infty} \frac{\| \mathbf{J}(\mathbf{Y}) \mathbf{u}^{t+1} - \sum_{i=1}^{t+1} \mathsf{c}_{t+1,i} \mathbf{u}^i - \mu_t \mathbf{X}^* - \sum_{i=1}^{K(t+1)} \theta_{t+1,i} \tilde{\mathbf{z}}^i \|^2}{N} = 0. \tag{B.56}
$$

By recalling the definitions of $\mathbf{f}^{t+1}$ and $\tilde{\mathbf{f}}^{t+1}$ (cf. (5.190) and (B.18)), (B.56) implies that

$$
\lim_{N \to \infty} \frac{\| \mathbf{f}^{t+1} - \tilde{\mathbf{f}}^{t+1} \|^2}{N} = 0. \tag{B.57}
$$

As g_{t+2} is Lipschitz, (B.57) also gives that

$$
\lim_{N \to \infty} \frac{\| \mathbf{u}^{t+2} - \tilde{\mathbf{u}}^{K(t+1)+1} \|^2}{N} = 0. \tag{B.58}
$$

Then, by using (B.37) with $i = t+1$ and Proposition B.1, we obtain that (B.41) and (B.42) hold for $k = t+1$, thus concluding the inductive proof. The result we have just proved by induction, combined with (B.19), gives that (B.17) holds.

Another application of Proposition B.1, together with (B.17), gives that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \psi\big(\tilde{u}_i^1, \tilde{u}_i^{K+1}, \ldots, \tilde{u}_i^{Kt+1}, \tilde{f}_i^1, \tilde{f}_i^2, \ldots, \tilde{f}_i^t, X_i^*\big)$$
$$= \mathbb{E}[\psi(\tilde{U}_1, \tilde{U}_{K+1}, \ldots, \tilde{U}_{Kt+1}, F_1, \ldots, F_t, X^*)], \tag{B.59}$$

where we recall that, by the definition in the theorem statement, for $s \in \{1, \ldots, t\}$,

$$F_s = \mu_s X^* + \sum_{i=1}^{Ks} \theta_{s,i} \tilde{Z}_i. \tag{B.60}$$

As $U_{s+1} = g_{s+1}(F_s)$, we have $\tilde{U}_{Ks+1} = U_{s+1}$ for all $s \in [t]$, and the proof is complete. $\square$

Acknowledgements

The thesis is built upon the papers, preprints and works in progress I have published, or I have been working on in the last three years. I deem it appropriate to thank and credit my collaborators in the various projects.

References

[1] David Sherrington and Scott Kirkpatrick. "Solvable Model of a Spin-Glass". In: *Phys. Rev. Lett.* 35 (26 1975), pp. 1792–1796. DOI: 10.1103/PhysRevLett.35.1792. URL: https://link.aps.org/doi/10.1103/PhysRevLett.35.1792.

[2] S F Edwards and P W Anderson. "Theory of spin glasses". In: *Journal of Physics F: Metal Physics* 5.5 (1975), pp. 965–974. DOI: 10.1088/0305-4608/5/5/017. URL: https://doi.org/10.1088/0305-4608/5/5/017.

[3] V. Cannella and J. A. Mydosh. "Magnetic Ordering in Gold-Iron Alloys". In: *Phys. Rev. B* 6 (11 1972), pp. 4220–4237. DOI: 10.1103/PhysRevB.6.4220. URL: https://link.aps.org/doi/10.1103/PhysRevB.6.4220.

[4] G. Parisi. "An Infinite Number of Order Parameters for Spin Glasses". In: *Phys. Rev. Lett.* 43 (1979), pp. 1754–1756. DOI: 10.1103/PhysRevLett.43.1754.

[5] G. Parisi. "A Sequence of Approximated Solutions to the S-K Model for Spin Glasses". In: *J. Phys. A* 13 (1980), p. L115. DOI: 10.1088/0305-4470/13/4/009.

[6] M Mezard, G Parisi, and M Virasoro. *Spin Glass Theory and Beyond*. WORLD SCIENTIFIC, 1986. DOI: 10.1142/0271.

[7] Francesco Guerra. "Broken Replica Symmetry Bounds in the Mean Field Spin Glass Model". In: *Communications in Mathematical Physics* 233 (2003).

[8] Michel Talagrand. "The Parisi Formula". In: *Annals of Mathematics* 163.1 (2006), pp. 221–263. ISSN: 0003486X. DOI: 10.4007/annals.2006.163.221.

[9] Stefano Ghirlanda and Francesco Guerra. "General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity". In: *Journal of Physics A: Mathematical and General* 31.46 (1998), pp. 9149–9155. DOI: 10.1088/0305-4470/31/46/006. URL: https://doi.org/10.1088/0305-4470/31/46/006.

[10] Francesco Guerra. "About The Overlap Distribution In Mean Field Spin Glass Models". In: *Int. J. Phys. B* 10 (1997), pp. 1675–1684.

[11] Pierluigi Contucci and Cristian Giardina. "The Ghirlanda-Guerra Identities". In: *Journal of Statistical Physics* 126 (2005).

[12]   Dmitry Panchenko. "A connection between the Ghirlanda–Guerra identities and ultra-metricity". In: *The Annals of Probability* 38.1 (2010), pp. 327–347. DOI: 10.1214/09-AOP484.

[13]   Jean Barbier and Nicolas Macris. "The adaptive interpolation method: a simple scheme to prove replica formulas in Bayesian inference". In: *Probability Theory and Related Fields* 174 (2019).

[14]   Jean Barbier and Nicolas Macris. "The adaptive interpolation method for proving replica formulas. Applications to the Curie–Weiss and Wigner spike models". In: *Journal of Physics A: Mathematical and Theoretical* 52.29 (2019), p. 294002. DOI: 10.1088/1751-8121/ab2735. URL: https://doi.org/10.1088/1751-8121/ab2735.

[15]   Jean Barbier and Galen Reeves. "Information-theoretic limits of a multiview low-rank symmetric spiked matrix model". In: *2020 IEEE International Symposium on Information Theory (ISIT)*. 2020, pp. 2771–2776. DOI: 10.1109/ISIT44484.2020.9173970.

[16]   Clément Luneau, N. Macris, and Jean Barbier. "High-dimensional rank-one nonsymmetric matrix decomposition: the spherical case". In: *2020 IEEE International Symposium on Information Theory (ISIT)* (2020), pp. 2646–2651.

[17]   Jean Barbier, Nicolas Macris, and Léo Miolane. "The layered structure of tensor estimation and its mutual information". In: *55th Annual Allerton Conference on Communication, Control, and Computing*. 2017.

[18]   Dmitry Panchenko. *The Sherrington-Kirkpatrick Model*. Springer, 2015.

[19]   Pierluigi Contucci and Cristian Giardinà. *Perspectives on Spin Glasses*. Cambridge University Press, 2012. DOI: 10.1017/CBO9781139049306.

[20]   Michel Talagrand. *Mean Field Models for Spin Glasses: Volume I: Basic Examples*. Springer, 2010.

[21]   Francesco Guerra and Fabio Lucio Toninelli. "The Thermodynamic Limit in Mean Field Spin Glass Models". In: *Communications in Mathematical Physics* 230 (2002).

[22]   Wei-Kuo Chen. "On the Almeida-Thouless transition line in the SK model with centered Gaussian external field". In: *arXiv e-prints* (2021). arXiv: 2103.04802 [cond-mat.dis-nn].

[23]   Michel Talagrand. *Mean Field Models for Spin Glasses: Volume II: Advanced Replica-Symmetry and Low Temperature*. Vol. 55. Springer, Jan. 2011. ISBN: 978-3-642-22252-8. DOI: 10.1007/978-3-642-22253-5.

[24]   A. Auffinger and Wei-Kuo Chen. "The Parisi Formula has a Unique Minimizer". In: *Communications in Mathematical Physics* 335 (2015), pp. 1429–1444.

[25]   Dmitry Panchenko. "On differentiability of the Parisi formula". In: *Electronic Communications in Probability* 13 (Oct. 2007). DOI: 10.1214/ECP.v13-1365.

[26]   Antonio Auffinger and Wei-Kuo Chen. "Free Energy and Complexity of Spherical Bipartite Models". In: *Journal of Statistical Physics* 157 (2014).

[27]   J R L de Almeida and D J Thouless. "Stability of the Sherrington-Kirkpatrick solution of a spin glass model". In: *Journal of Physics A: Mathematical and General* 11.5 (1978), pp. 983–990. DOI: `10.1088/0305-4470/11/5/028`.

[28]   F. Toninelli. "About the Almeida-Thouless transition line in the Sherrington-Kirkpatrick mean-field spin glass model". In: *EPL* 60 (2002), pp. 764–767.

[29]   Hidetoshi Nishimori. *Statistical Physics of Spin Glasses and Information Processing: an Introduction*. Oxford; New York: Oxford University Press, 2001.

[30]   Iain M. Johnstone. "On the distribution of the largest eigenvalue in principal components analysis". In: *The Annals of Statistics* 29.2 (2001), pp. 295–327. DOI: `10.1214/aos/1009210544`.

[31]   Florent Krzakala, Jiaming Xu, and Lenka Zdeborová. "Mutual information in rank-one matrix estimation". In: *2016 IEEE Information Theory Workshop (ITW)*. 2016, pp. 71–75. DOI: `10.1109/ITW.2016.7606798`.

[32]   M. Lelarge and Léo Miolane. "Fundamental limits of symmetric low-rank matrix estimation". In: *Probability Theory and Related Fields* 173 (2017), pp. 859–929.

[33]   Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, and Lenka Zdeborová. "Rank-one matrix estimation: analysis of algorithmic and information theoretic limits by the spatial coupling method". In: *arXiv e-prints*, arXiv:1812.02537 (2018).

[34]   Ahmed El Alaoui, Florent Krzakala, and Michael Jordan. "Fundamental limits of detection in the spiked Wigner model". In: *The Annals of Statistics* 48.2 (2020), pp. 863–885. DOI: `10.1214/19-AOS1826`. URL: `https://doi.org/10.1214/19-AOS1826`.

[35]   Marc Potters and Jean-Philippe Bouchaud. *A First Course in Random Matrix Theory: for Physicists, Engineers and Data Scientists*. Cambridge University Press, 2020. DOI: `10.1017/9781108768900`.

[36]   Giacomo Livan, Marcel Novaes, and Pierpaolo Vivo. *Introduction to Random Matrices - Theory and Practice*. Vol. 26. Dec. 2017. ISBN: 978-3-319-70883-6. DOI: `10.1007/978-3-319-70885-0`.

[37]   Jinho Baik, Gérard Ben Arous, and Sandrine Péché. "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices". In: *The Annals of Probability* 33.5 (2005), pp. 1643–1697. DOI: `10.1214/009117905000000233`.

[38]   Florent Benaych-Georges and Raj Rao Nadakuditi. "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices". In: *Advances in Mathematics* 227.1 (2011), pp. 494–521. ISSN: 0001-8708. DOI: `https://doi.org/10.1016/j.aim.2011.02.007`.

[39]   Jean Barbier. "Overlap matrix concentration in optimal Bayesian inference". In: *Information and Inference: A Journal of the IMA* 10.2 (May 2020), pp. 597–623. ISSN: 2049-8772. DOI: `10.1093/imaiai/iaaa008`.

[40]   Jean Barbier and Dmitry Panchenko. "Strong Replica Symmetry in High-Dimensional Optimal Bayesian Inference". In: *Communications in Mathematical Physics* 393 (Aug. 2022), pp. 1–41. DOI: `10.1007/s00220-022-04387-w`.

[41]   Michael Aizenman and Pierluigi Contucci. "On the Stability of the Quenched State in Mean Field Spin Glass Models". In: *Journal of Statistical Physics* 92 (Jan. 1998). DOI: `10.1023/A:1023080223894`.

[42]   Satoshi Morita, Hidetoshi Nishimori, and Pierluigi Contucci. "Griffiths inequalities for the Gaussian spin glass". In: *Journal of Physics A Mathematical General* 37 (2004).

[43]   Diego Alberici, Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione. "The multi-species mean-field spin-glass on the Nishimori line". In: *Journal of Statistical Physics* 182.1 (2021), pp. 1–20.

[44]   Adriano Barra, Pierluigi Contucci, Emanuele Mingione, and Daniele Tantari. "Multi-Species Mean Field Spin Glasses. Rigorous Results". In: *Annales Institut Henri Poincaré* 16 (2013).

[45]   Dmitry Panchenko. "The free energy in a multi-species Sherrington–Kirkpatrick model". In: *Annals of Probability* 43 (2015).

[46]   Wei Kuo Chen. "On the mixed even-spin Sherrington-Kirkpatrick model with ferromagnetic interaction". In: *Annales Institut Henri Poincare* 50 (2014).

[47]   Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, and Lenka Zdeborová. "Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula". In: *Advances in Neural Information Processing Systems* 29 (2016).

[48]   Elena Agliari, Linda Albanese, Adriano Barra, and Gabriele Ottaviani. "Replica symmetry breaking in neural networks: a few steps toward rigorous results". In: *Journal of Physics A: Mathematical and Theoretical* 53 (2020).

[49]   Elena Agliari, Adriano Barra, Peter Sollich, and Lenka Zdeborova. "Machine Learning and Statistical Physics: Theory, Inspiration, Application". In: *Journal of Physics A: Mathematical and Theoretical* 53 (2020).

[50]   Satoshi Morita, Hidetoshi Nishimori, and Pierluigi Contucci. "Griffiths Inequalities in the Nishimori Line". In: *Progress of Theoretical Physics Supplement* 157 (2005).

[51]   Alberto Jimenez Felstrom and Kamil Sh Zigangirov. "Time-varying periodic convolutional codes with low-density parity-check matrix". In: *IEEE Transactions on Information Theory* 45.6 (1999), pp. 2181–2191.

[52] Shrinivas Kudekar, Thomas J Richardson, and Rüdiger L Urbanke. "Threshold saturation via spatial coupling: Why convolutional LDPC ensembles perform so well over the BEC". In: *IEEE Transactions on Information Theory* 57.2 (2011), pp. 803–834.

[53] Hidetoshi Nishimori. "Internal Energy, Specific Heat and Correlation Function of the Bond-Random Ising Model". In: *Progress of Theoretical Physics* 66 (1981).

[54] Pierluigi Contucci, Satoshi Morita, and Hidetoshi Nishimori. "Surface Terms on the Nishimori Line of the Gaussian Edwards-Anderson Model". In: *Journal of Statistical Physics* 122 (2005).

[55] Diego Alberici, Adriano Barra, Pierluigi Contucci, and Emanuele Mingione. "Annealing and Replica-Symmetry in Deep Boltzmann Machines". In: *Journal of Statistical Physics* 180 (2020).

[56] Diego Alberici, Pierlugi Contucci, and Emanuele Mingione. "Deep Boltzmann Machines: rigorous results at arbitrary depth". In: *Annales Institut Henri Poincaré (to appear)* (2021).

[57] Adriano Barra, Giuseppe Genovese, and Francesco Guerra. "Equilibrium statistical mechanics of bipartite spin systems". In: *Journal of Physics A: Mathematical and Theoretical* 44 (2011).

[58] Giuseppe Genovese. "Minimax formula for the replica symmetric free energy of deep restricted Boltzmann machines". In: *arXiv preprint arXiv:2005.09424* (2020).

[59] Diego Alberici, Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione. "The solution of the deep Boltzmann machine on the Nishimori line". In: *Communications in Mathematical Physics (to appear)* (July 2021). DOI: 10.1007/s00220-021-04165-0.

[60] Erik Bates, Leila Sloman, and Youngtak Sohn. "Replica symmetry breaking in multi-species Sherrington–Kirkpatrick model". In: *Journal of Statistical Physics* 174 (2018).

[61] Jinho Baik and Ji Oon Lee. "Free energy of bipartite spherical Sherrington–Kirkpatrick model". In: *Annales Institut Henri Poincaré* 56 (2020).

[62] Jean-Christophe Mourrat. "Free energy upper bound for mean-field vector spin glasses". In: *arXiv e-prints*, arXiv:2010.0911 (2020).

[63] Jean-Christophe Mourrat. "Nonconvex interactions in mean-field spin glasses". In: *arXiv e-prints*, arXiv:2004.01679 (2020).

[64] Richard Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Springer, 2006.

[65] Galen Reeves. "Information-Theoretic Limits for the Matrix Tensor Product". In: *IEEE Journal on Selected Areas in Information Theory* 1 (2020).

[66] Hong-Bin Chen, Jean-Christophe Mourrat, and Jiaming Xia. "Statistical inference of finite-rank tensors". In: *arXiv e-prints* (2021). arXiv: 2104.05360.

[67]   Jean Barbier. "Concentration of the Matrix-Valued Minimum Mean-Square Error in Optimal Bayesian Inference". In: *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*. 2019, pp. 644–648. DOI: `10.1109/CAMSAP45676.2019.9022463`.

[68]   Farzad Pourkamali and Nicolas Macris. *Mismatched Estimation of rank-one symmetric matrices under Gaussian noise*. 2021. arXiv: `2107.08927 [cs.IT]`.

[69]   Farzad Pourkamali and Nicolas Macris. "Mismatched Estimation of Non-Symmetric Rank-One Matrices Under Gaussian Noise". In: *2022 IEEE International Symposium on Information Theory (ISIT)*. 2022, pp. 1288–1293. DOI: `10.1109/ISIT50566.2022.9834858`.

[70]   Jean Barbier, TianQi Hou, Marco Mondelli, and Manuel Sáenz. "The price of ignorance: how much does it cost to forget noise structure in low-rank matrix estimation?" In: *Advances in Neural Information Processing Systems*. 2022.

[71]   Sergio Verdú. "Mismatched Estimation and Relative Entropy". In: *IEEE Transactions on Information Theory* 56.8 (2010), pp. 3712–3720. DOI: `10.1109/TIT.2010.2050800`.

[72]   Jean Barbier, Dmitry Panchenko, and Manuel Sáenz. *Strong replica symmetry for high-dimensional disordered log-concave Gibbs measures*. Dec. 2021. DOI: `10.1093/imaiai/iaab027`.

[73]   Jean Barbier, Wei-Kuo Chen, Dmitry Panchenko, and Manuel Sáenz. *Performance of Bayesian linear regression in a model with mismatch*. 2021. arXiv: `2107.06936 [math.PR]`.

[74]   Sumit Mukherjee and Subhabrata Sen. *Variational Inference in high-dimensional linear regression*. 2021. arXiv: `2104.12232 [math.ST]`.

[75]   Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione. "An inference problem in a mismatched setting: a spin-glass model with Mattis interaction". In: *SciPost Phys.* 12 (4 2022), p. 125. DOI: `10.21468/SciPostPhys.12.4.125`.

[76]   Satish Babu Korada and Nicolas Macris. "Tight Bounds on the Capacity of Binary Input Random CDMA Systems". In: *IEEE Transactions on Information Theory* 56.11 (2010), pp. 5590–5613. DOI: `10.1109/TIT.2010.2070131`.

[77]   Dmitry Panchenko. "Free energy in the mixed p-spin models with vector spins". In: *The Annals of Probability* 46.2 (2018), pp. 865–896. DOI: `10.1214/17-AOP1194`.

[78]   Erwin Bolthausen. "An Iterative Construction of Solutions of the TAP Equations for the Sherrington-Kirkpatrick Model". In: *Communications in Mathematical Physics* 325 (Jan. 2012). DOI: `10.1007/s00220-013-1862-3`.

[79]   Adriano Barra, Alberto Bernacchia, Enrica Santucci, and Pierluigi Contucci. "On the equivalence of Hopfield Networks and Boltzmann Machines". In: *Neural networks : the official journal of the International Neural Network Society* 34 (June 2012), pp. 1–9. DOI: `10.1016/j.neunet.2012.06.003`.

[80]   Marc Mézard. "Mean-field message-passing equations in the Hopfield model and its generalizations". In: *Physical Review E* 95 (Aug. 2016). DOI: 10.1103/PhysRevE.95.022117.

[81]   E. Agliari, F. Alemanno, A. Barra, M. Centonze, and A. Fachechi. "Neural networks with a redundant representation: detecting the undetectable". In: *Physical Review Letters* 124.2 (2020), p. 028301.

[82]   E. Agliari and G. De Marzo. "Tolerance versus synaptic noise in dense associative memories". In: *The European Physics Journal Plus* 135 (2020), p. 883.

[83]   Emmanuel Abbe. "Community Detection and Stochastic Block Models: Recent Developments". In: *Journal of Machine Learning Research* 18.177 (2018), pp. 1–86. URL: http://jmlr.org/papers/v18/16-480.html.

[84]   Chiheon Kim, Afonso S Bandeira, and Michel X Goemans. "Community detection in hypergraphs, spiked tensor models, and sum-of-squares". In: *2017 International Conference on Sampling Theory and Applications (SampTA)*. IEEE. 2017, pp. 124–128.

[85]   Amelia Perry, Alexander S. Wein, Afonso S. Bandeira, and Ankur Moitra. "Message-Passing Algorithms for Synchronization Problems over Compact Groups". In: *Communications on Pure and Applied Mathematics* 71.11 (2018), pp. 2275–2322.

[86]   Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. "Optimality and sub-optimality of PCA for spiked random matrices and synchronization". In: *arXiv preprint arXiv:1609.05573* (2016).

[87]   Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. "MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel". In: *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 2015, pp. 680–687.

[88]   Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. "Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications". In: *Journal of Statistical Mechanics: Theory and Experiment* 2017.7 (2017), p. 073403.

[89]   Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. "Optimality and sub-optimality of PCA I: Spiked random matrix models". In: *The Annals of Statistics* 46.5 (2018), pp. 2416–2451.

[90]   Jinho Baik and Jack W. Silverstein. "Eigenvalues of large sample covariance matrices of spiked population models". In: *Journal of multivariate analysis* 97.6 (2006), pp. 1382–1408.

[91]   Sandrine Péché. "The largest eigenvalue of small rank perturbations of Hermitian random matrices. Probab. Theory Relat. Fields 134, 127-173". In: *Probability Theory and Related Fields* 134 (Jan. 2006), pp. 127–173. DOI: 10.1007/s00440-005-0466-z.

[92]   Delphine Féral and Sandrine Péché. "The largest eigenvalue of rank one deformation of large Wigner matrices". In: *Communications in mathematical physics* 272.1 (2007), pp. 185–228.

[93]   Mireille Capitaine, Catherine Donati-Martin, and Delphine Féral. "The largest eigenvalues of finite rank deformation of large Wigner matrices: convergence and nonuniversality of the fluctuations". In: *The Annals of Probability* 37.1 (2009), pp. 1–47.

[94]   Raj Rao Nadakuditi and Jack W. Silverstein. "Fundamental Limit of Sample Generalized Eigenvalue Based Detection of Signals in Noise Using Relatively Few Signal-Bearing and Noise-Only Samples". In: *IEEE Journal of Selected Topics in Signal Processing* 4.3 (2010), pp. 468–480. DOI: 10.1109/JSTSP.2009.2038310.

[95]   Florent Benaych-Georges and Raj Rao Nadakuditi. "The singular values and vectors of low rank perturbations of large rectangular random matrices". In: *Journal of Multivariate Analysis* 111 (2012), pp. 120–135.

[96]   Zhidong Bai and Jianfeng Yao. "On sample eigenvalues in a generalized spiked population model". In: *Journal of Multivariate Analysis* 106 (2012), pp. 167–177.

[97]   Yoshiyuki Kabashima. "A CDMA multiuser detection algorithm on the basis of belief propagation". In: *Journal of Physics A: Mathematical and General* (2003).

[98]   Mohsen Bayati and Andrea Montanari. "The LASSO Risk for Gaussian Matrices". In: *IEEE Transactions on Information Theory* 58.4 (2012), pp. 1997–2017. DOI: 10.1109/TIT.2011.2174612.

[99]   Mohsen Bayati and Andrea Montanari. "The Dynamics of Message Passing on Dense Graphs, with Applications to Compressed Sensing". In: *IEEE Transactions on Information Theory* 57.2 (2011), pp. 764–785. DOI: 10.1109/TIT.2010.2094817.

[100]   David Donoho, Arian Maleki, and Andrea Montanari. "Message Passing Algorithms for Compressed Sensing". In: *Proceedings of the National Academy of Sciences of the United States of America* 106 (Nov. 2009), pp. 18914–9. DOI: 10.1073/pnas.0909892106.

[101]   Florent Krzakala, Marc Mézard, Francois Sausset, Yifan Sun, and Lenka Zdeborová. "Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices". In: *Journal of Statistical Mechanics: Theory and Experiment* (2012).

[102]   Arian Maleki, Laura Anitori, Zai Yang, and Richard G Baraniuk. "Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP)". In: *IEEE Transactions on Information Theory* 59.7 (2013), pp. 4290–4308.

[103]   Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. "Optimal errors and phase transitions in high-dimensional generalized linear models". In: *Proceedings of the National Academy of Sciences* 116.12 (2019), pp. 5451–5460.

[104] Junjie Ma, Ji Xu, and Arian Maleki. "Optimization-Based AMP for Phase Retrieval: The Impact of Initialization and ℓ_2 Regularization". In: *IEEE Transactions on Information Theory* 65.6 (2019), pp. 3600–3629.

[105] Antoine Maillard, Bruno Loureiro, Florent Krzakala, and Lenka Zdeborová. "Phase retrieval in high dimensions: Statistical and computational phase transitions". In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020.

[106] Marco Mondelli and Ramji Venkataramanan. "Approximate Message Passing with Spectral Initialization for Generalized Linear Models". In: *24th International Conference on Artificial Intelligence and Statistics*. Vol. 130. 2021, pp. 397–405.

[107] Sundeep Rangan. "Generalized approximate message passing for estimation with random linear mixing". In: *International Symposium on Information Theory*. 2011, pp. 2168–2172.

[108] Philip Schniter and Sundeep Rangan. "Compressive phase retrieval via generalized approximate message passing". In: *IEEE Transactions on Signal Processing* 63.4 (2014), pp. 1043–1055.

[109] Pragya Sur and Emmanuel J. Candès. "A modern maximum-likelihood theory for high-dimensional logistic regression". In: *Proceedings of the National Academy of Sciences* 116.29 (2019), pp. 14516–14525.

[110] Yash Deshpande and Andrea Montanari. "Information-theoretically optimal sparse PCA". In: *IEEE International Symposium on Information Theory* (2014), pp. 2197–2201. DOI: 10.1109/ISIT.2014.6875223.

[111] Alyson Fletcher and Sundeep Rangan. "Iterative reconstruction of rank-one matrices in noise". In: *Information and Inference: A Journal of the IMA* 7 (Sept. 2018), pp. 531–562. DOI: 10.1093/imaiai/iax014.

[112] Yoshiyuki Kabashima, Florent Krzakala, Marc Mézard, Ayaka Sakata, and Lenka Zdeborová. "Phase Transitions and Sample Complexity in Bayes-Optimal Matrix Factorization". In: *IEEE Transactions on Information Theory* 62.7 (2016), pp. 4228–4265. DOI: 10.1109/TIT.2016.2556702.

[113] Andrea Montanari and Ramji Venkataramanan. "Estimation of Low-Rank Matrices via Approximate Message Passing". In: *Annals of Statistics* 45.1 (2021), pp. 321–345.

[114] Jean Barbier, Nicolas Macris, and Cynthia Rush. "All-or-nothing statistical and computational phase transitions in sparse spiked matrix estimation". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 14915–14926.

[115] Adel Javanmard and Andrea Montanari. "State evolution for general approximate message passing algorithms, with applications to spatial coupling". In: *Information and Inference: A Journal of the IMA* 2.2 (2013), pp. 115–144.

[116]   David L. Donoho, Adel Javanmard, and Andrea Montanari. "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing". In: *IEEE Transactions on Information Theory* 59.11 (2013), pp. 7434–7464.

[117]   Ahmed El Alaoui and Florent Krzakala. "Estimation in the spiked Wigner model: a short proof of the replica formula". In: *IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2018, pp. 1874–1878.

[118]   Alice Guionnet, Justin Ko, Florent Krzakala, and Lenka Zdeborová. "Low-rank Matrix Estimation with Inhomogeneous Noise". In: *arXiv preprint arXiv:2208.05918* (2022).

[119]   Enzo Marinari, Giorgio Parisi, and Felix Ritort. "Replica field theory for deterministic models: I. Binary sequences with low autocorrelation". In: *Journal of Physics A* 27 (1994), pp. 7615–7645.

[120]   Enzo Marinari, Giorgio Parisi, and Felix Ritort. "Replica field theory for deterministic models: II. A non-random spin glass with glassy behaviour". In: *Journal of Physics A* 27 (1994), pp. 7647–7668.

[121]   Giorgio Parisi and Marc Potters. "Mean-Field Equations for Spin Models with Orthogonal Interaction Matrices". In: *Journal of Physics A: Mathematical and General* 28 (Jan. 1999), p. 5267. DOI: 10.1088/0305-4470/28/18/016.

[122]   Bhaswar B. Bhattacharya and Subhabrata Sen. "High Temperature Asymptotics of Orthogonal Mean-Field Spin Glasses". In: *Journal of Statistical Physics* 162.1 (2016), pp. 63–80.

[123]   Manfred Opper, Burak Cakmak, and Ole Winther. "A theory of solving TAP equations for Ising models with general invariant random matrices". In: *Journal of Physics A: Mathematical and Theoretical* 49.11 (2016), p. 114002.

[124]   Manfred Opper and Ole Winther. "Adaptive and Self-Averaging Thouless-Anderson-Palmer Mean-Field Theory for Probabilistic Modeling". In: *Physical Review E* 64 (2011).

[125]   Zhou Fan and Yihong Wu. "The replica-symmetric free energy for Ising spin glasses with orthogonally invariant couplings". In: *arXiv preprint arXiv:2105.02797* (2021).

[126]   Antoine Maillard, Laura Foini, Alejandro Lage Castellanos, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. "High-temperature expansions and message passing algorithms". In: *Journal of Statistical Mechanics: Theory and Experiment* 2019.11 (2019), p. 113301.

[127]   Laura Foini and Jorge Kurchan. "Annealed averages in spin and matrix models". In: *SciPost Phys.* 12 (2022), p. 080. DOI: 10.21468/SciPostPhys.12.3.080.

[128] Marylou Gabrié, Andre Manoel, Clément Luneau, Jean Barbier, Nicolas Macris, Florent Krzakala, and Lenka Zdeborová. "Entropy and mutual information in models of deep neural networks". In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Vol. 31. Curran Associates, Inc., 2018, pp. 1821–1831.

[129] Cedric Gerbelot, Alia Abbara, and Florent Krzakala. "Asymptotic errors for teacher-Student Convex Generalized Linear Models (or: how to Prove Kabashima's Replica Formula)". In: *arXiv preprint, arXiv:2006.06581* (2020).

[130] Junjie Ma, Ji Xu, and Arian Maleki. "Analysis of sensing spectral for signal recovery under a generalized linear model". In: *Advances in Neural Information Processing Systems*. Vol. 34. 2021, pp. 22601–22613.

[131] Takashi Takahashi and Yoshiyuki Kabashima. "Macroscopic analysis of vector approximate message passing in a model mismatch setting". In: *International Symposium on Information Theory*. 2020, pp. 1403–1408.

[132] Zhou Fan. "Approximate message passing algorithms for rotationally invariant matrices". In: *The Annals of Statistics* 50.1 (2022), pp. 197–224.

[133] Xinyi Zhong, Tianhao Wang, and Zhou Fan. "Approximate Message Passing for orthogonally invariant ensembles: Multivariate non-linearities and spectral initialization". In: *arXiv preprint, arXiv:2110.02318* (2021).

[134] Ramji Venkataramanan, Kevin Kögler, and Marco Mondelli. "Estimation in rotationally invariant generalized linear models via approximate message passing". In: *International Conference on Machine Learning*. 2022, pp. 22120–22144.

[135] Jean Barbier, Francesco Camilli, Marco Mondelli, and Manuel Saenz. "Bayes-optimal limits in structured PCA, and how to reach them". In: (2022). DOI: `10.48550/ARXIV.2210.01237`.

[136] Edouard Brézin, Claude Itzykson, Giorgio Parisi, and Jean-Bernard Zuber. "Planar diagrams". In: *Communications in Mathematical Physics* 59.1 (1978), pp. 35–51. DOI: `cmp/1103901558`.

[137] Jean Barbier and Manuel Sáenz. "Marginals of a spherical spin glass model with correlated disorder". In: *arXiv preprint arXiv:2112.02066* (2021).

[138] Alice Guionnet and Mylène Maida. "A Fourier view on the R-transform and related asymptotics of spherical integrals". In: *Journal of Functional Analysis* 222.2 (2005), pp. 435–490.

[139] Alice Guionnet and Ofer Zeitouni. "Large deviations asymptotics for spherical integrals". In: *Journal of functional analysis* 188.2 (2002), pp. 461–515. URL: `https://www.sciencedirect.com/science/article/pii/S0022123601938339`.

[140]   Claude Itzykson and Jean-Bernard Zuber. "The planar approximation. II". In: *Journal of Mathematical Physics* 21.3 (1980), pp. 411–421.

[141]   Vladimir Kazakov. "Solvable matrix models". In: *Random matrix models and their applications* 40 (2001), pp. 271–283.

[142]   Alice Guionnet. "First order asymptotics of matrix integrals: a rigorous approach towards the understanding of matrix models". In: *Communications in mathematical physics* 244.3 (2004), pp. 527–569.

[143]   Rishabh Dudeja and Milad Bakhshizadeh. "Universality of linearized message passing for phase retrieval with structured sensing matrices". In: *IEEE Transactions on Information Theory* (2022).

[144]   Rishabh Dudeja, Yue M Lu, and Subhabrata Sen. "Universality of Approximate Message Passing with Semi-Random Matrices". In: *arXiv preprint arXiv:2204.04281* (2022).

[145]   Rishabh Dudeja, Subhabrata Sen, and Yue M Lu. "Spectral Universality of Regularized Linear Regression with Nearly Deterministic Sensing Matrices". In: *arXiv preprint arXiv:2208.02753* (2022).

[146]   Marc Mézard and Andrea Montanari. *Information, physics, and computation.* Oxford University Press, 2009.

[147]   Manfred Opper and David Saad. *Advanced mean field methods: Theory and practice.* MIT press, 2001.

[148]   Andrea Montanari and Subhabrata Sen. "A Short Tutorial on Mean-Field Spin Glass Techniques for Non-Physicists". In: *arXiv preprint arXiv:2204.02909* (2022).

[149]   Jean Barbier, Mohamad Dia, and Nicolas Macris. "Proof of threshold saturation for spatially coupled sparse superposition codes". In: *IEEE International Symposium on Information Theory (ISIT).* 2016, pp. 1173–1177.

[150]   Amin Coja-Oghlan, Florent Krzakala, Will Perkins, and Lenka Zdeborová. "Information-theoretic thresholds from the cavity method". In: *Advances in Mathematics* 333 (2018), pp. 694–795.

[151]   J.-C. Mourrat. "Hamilton–Jacobi equations for finite-rank matrix inference". In: *The Annals of Applied Probability* 30.5 (2020), pp. 2234–2260. DOI: `10.1214/19-AAP1556`.

[152]   Hong-Bin Chen and Jiaming Xia. "Limiting free energy of multi-layer generalized linear models". In: *arXiv preprint arXiv:2108.12615* (2021).

[153]   Yoshiyuki Kabashima. "Inference from correlated patterns: a unified theory for perceptron learning and linear vector channels". In: *Journal of Physics: Conference Series.* Vol. 95. 1. IOP Publishing. 2008, p. 012001.

[154]   Jonathan Novak. "Three lectures on free probability". In: *Random matrix theory, interacting particle systems, and integrable systems* 65.309-383 (2014), p. 13.

[155] Marco Mondelli and Ramji Venkataramanan. "PCA Initialization for Approximate Message Passing in Rotationally Invariant Models". In: *Advances in Neural Information Processing Systems*. Vol. 34. 2021, pp. 29616–29629. URL: https://proceedings.neurips.cc/paper/2021/file/f7ac67a9aa8d255282de7d11391e1b69-Paper.pdf.

[156] Oliver Y Feng, Ramji Venkataramanan, Cynthia Rush, Richard J Samworth, et al. "A unifying tutorial on Approximate Message Passing". In: *Foundations and Trends® in Machine Learning* 15.4 (2022), pp. 335–536.

[157] Yoshua Bengio, Aaron Courville, and Pascal Vincent. "Representation Learning: A Review and New Perspectives". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.8 (2013), pp. 1798–1828. DOI: 10.1109/TPAMI.2013.50.

[158] Bruno A Olshausen and David J Field. "Emergence of simple-cell receptive field properties by learning a sparse code for natural images". In: *Nature* 381.6583 (1996), pp. 607–609.

[159] Bruno A. Olshausen and David J. Field. "Sparse coding with an overcomplete basis set: A strategy employed by V1?" In: *Vision Research* 37.23 (1997), pp. 3311–3325. ISSN: 0042-6989. DOI: https://doi.org/10.1016/S0042-6989(97)00169-7.

[160] Kenneth Kreutz-Delgado, Joseph F. Murray, Bhaskar D. Rao, Kjersti Engan, Te-Won Lee, and Terrence J. Sejnowski. "Dictionary Learning Algorithms for Sparse Representation". In: *Neural Computation* 15.2 (Feb. 2003), pp. 349–396. ISSN: 0899-7667. DOI: 10.1162/089976603762552951. URL: https://doi.org/10.1162/089976603762552951.

[161] David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. "A Learning Algorithm for Boltzmann Machines*". In: *Cognitive Science* 9.1 (1985), pp. 147–169. DOI: https://doi.org/10.1207/s15516709cog0901{_}7.

[162] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. "Restricted Boltzmann Machines for Collaborative Filtering". In: *Proceedings of the 24th International Conference on Machine Learning*. ICML '07. Corvalis, Oregon, USA: Association for Computing Machinery, 2007, pp. 791–798. ISBN: 9781595937933. DOI: 10.1145/1273496.1273596. URL: https://doi.org/10.1145/1273496.1273596.

[163] Julien Mairal, Michael Elad, and Guillermo Sapiro. "Sparse Representation for Color Image Restoration". In: *IEEE Transactions on Image Processing* 17.1 (2008), pp. 53–69. DOI: 10.1109/TIP.2007.911828.

[164] John Wright, Yi Ma, Julien Mairal, Guillermo Sapiro, Thomas S. Huang, and Shuicheng Yan. "Sparse Representation for Computer Vision and Pattern Recognition". In: *Proceedings of the IEEE* 98.6 (2010), pp. 1031–1044. DOI: 10.1109/JPROC.2010.2044470.

[165] M. Elad and M. Aharon. "Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries". In: *Trans. Img. Proc.* 15.12 (2006), pp. 3736–3745. ISSN: 1057-7149. DOI: 10.1109/TIP.2006.881969. URL: https://doi.org/10.1109/TIP.2006.881969.

[166] Andriy Mnih and Russ R Salakhutdinov. "Probabilistic Matrix Factorization". In: *Advances in Neural Information Processing Systems*. Ed. by J. Platt, D. Koller, Y. Singer, and S. Roweis. Vol. 20. Curran Associates, Inc., 2007. URL: https://proceedings.neurips.cc/paper/2007/file/d7322ed717dedf1eb4e6e52a37ea7bcd-Paper.pdf.

[167] Andre Manoel, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. "Multi-Layer Generalized Linear Estimation". In: *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE Press, 2017, pp. 2098–2102. DOI: 10.1109/ISIT.2017.8006899.

[168] Joël Bun, Jean-Philippe Bouchaud, and Marc Potters. "Cleaning large correlation matrices: Tools from Random Matrix Theory". In: *Physics Reports* 666 (2017). Cleaning large correlation matrices: tools from random matrix theory, pp. 1–109. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2016.10.005. URL: https://www.sciencedirect.com/science/article/pii/S0370157316303337.

[169] Harry Markowitz. "Portfolio Selection". In: *The Journal of Finance* 7.1 (1952), pp. 77–91. ISSN: 00221082, 15406261. URL: http://www.jstor.org/stable/2975974 (visited on 10/04/2022).

[170] Joel Bun, Romain Allez, Jean-Philippe Bouchaud, and Marc Potters. "Rotational Invariant Estimator for General Noisy Matrices". In: *IEEE Transactions on Information Theory* PP (Feb. 2015). DOI: 10.1109/TIT.2016.2616132.

[171] Antoine Maillard, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. "Perturbative construction of mean-field equations in extensive-rank matrix factorization and denoising". In: *Journal of Statistical Mechanics: Theory and Experiment* 2022.8 (2022), p. 083301. DOI: 10.1088/1742-5468/ac7e4c. URL: https://doi.org/10.1088/1742-5468/ac7e4c.

[172] Jean Barbier and Nicolas Macris. "Statistical limits of dictionary learning: Random matrix theory and the spectral replica method". In: *Phys. Rev. E* 106 (2 2022), p. 024136. DOI: 10.1103/PhysRevE.106.024136.

[173] J J Hopfield. "Neural networks and physical systems with emergent collective computational abilities." In: *Proceedings of the National Academy of Sciences* 79.8 (1982), pp. 2554–2558. DOI: 10.1073/pnas.79.8.2554. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.79.8.2554. URL: https://www.pnas.org/doi/abs/10.1073/pnas.79.8.2554.

[174] Daniel J. Amit, Hanoch Gutfreund, and H. Sompolinsky. "Spin-glass models of neural networks". In: *Phys. Rev. A* 32 (2 Aug. 1985), pp. 1007–1018. DOI: 10.1103/PhysRevA.32.1007. URL: https://link.aps.org/doi/10.1103/PhysRevA.32.1007.

[175] Daniel J. Amit, Hanoch Gutfreund, and H. Sompolinsky. "Storing Infinite Numbers of Patterns in a Spin-Glass Model of Neural Networks". In: *Phys. Rev. Lett.* 55 (14 1985), pp. 1530–1533. DOI: 10.1103/PhysRevLett.55.1530.

[176] Jess Banks, Cristopher Moore, Roman Vershynin, Nicolas Verzelen, and Jiaming Xu. "Information-theoretic bounds and phase transitions in clustering, sparse PCA, and sub-matrix localization". In: *2017 IEEE International Symposium on Information Theory (ISIT)*. 2017, pp. 1137–1141. DOI: 10.1109/ISIT.2017.8006706.

[177] Francesco Camilli and Marc Mézard. "Matrix factorization with neural networks". In: *arXiv e-prints* (Dec. 2022). arXiv: 2212.02105.

Ringraziamenti

Desidero ringraziare anzitutto il professor Contucci per essere stato, e per essere tuttora, un costante stimolo nel mio percorso, nonchè per aver condiviso con me le sue conoscenze scientifiche e personali. Ci tengo a sottolineare che il progetto di cotutela è stato realizzato in buona sostanza grazie al suo contributo e al suo tempo. Il ringraziamento si estende anche al dottor Mingione, il dottor Alberici ed il professor Tantari, per le vivaci chiacchierate scientifiche, e non solo.

Un grande grazie anche al professor Mézard per avermi accettato come suo studente di dottorato e coinvolto in un progetto così entusiasmante, ma in particolare per l'iniezione di ottimismo. Dalla nostra interazione ho non solo appreso nozioni fondamentali per la mia maturità scientifica, ma anche imparato a guardare le cose da un'altra prospettiva.

Dedico un ringraziamento speciale al professor Barbier, e per estensione anche al dottor Saenz e il professor Mondelli, per aver condiviso con me il progetto da cui è nato il Capitolo 5, una vera e propria avventura, parzialmente vissuta nella bellissima Trieste dove posso dire di aver lasciato una parte di me.

Grazie Arianna, basta un'ora con te per farmi dimenticare ogni preoccupazione. Il problema è che solitamente passa troppo in fretta. Sei veramente la mia fonte di tranquillità.

Alle mie rocce, mia madre e mio padre, va tutta la mia gratitudine per avermi donato il loro inesauribile e incondizionato sostegno. Avere due genitori come voi è un'immensa fortuna.

Per finire, ringrazio tutti coloro che mi hanno sopportato nei miei periodi di irrequietudine, tra cui i miei coinquilini, i miei colleghi, e i miei amici. Vorrei menzionare anche Armand e Fabien, con cui è stato difficile incontrarsi questi anni, ma su di loro si può sempre contare.