

High-temperature expansions and message passing algorithms

Antoine Maillard, Laura Foini, Alejandro Lage Castellanos, Florent Krzakala, Marc Mezard, Lenka Zdeborová

▶ **To cite this version:**

Antoine Maillard, Laura Foini, Alejandro Lage Castellanos, Florent Krzakala, Marc Mezard, et al.. High-temperature expansions and message passing algorithms. Journal of Statistical Mechanics: Theory and Experiment, 2019, 2019 (11), pp.113301. 10.1088/1742-5468/ab4bbb . cea-02524865

HAL Id: cea-02524865

https://cea.hal.science/cea-02524865

Submitted on 30 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

arXiv:1906.08479v2 [cond-mat.dis-nn] 29 Sep 2019

High-temperature Expansions and Message Passing Algorithms

Antoine Maillard[⋆,⊗], Laura Foini[†], Alejandro Lage Castellanos[◇],
Florent Krzakala[⋆], Marc Mézard[⋆], Lenka Zdeborová[†]

Abstract

Improved mean-field technics are a central theme of statistical physics methods applied to inference and learning. We revisit here some of these methods using high-temperature expansions for disordered systems initiated by Plefka, Georges and Yedidia. We derive the Gibbs free entropy and the subsequent self-consistent equations for a generic class of statistical models with correlated matrices and show in particular that many classical approximation schemes, such as adaptive TAP, Expectation-Consistency, or the approximations behind the Vector Approximate Message Passing algorithm all rely on the same assumptions, that are also at the heart of high-temperature expansions. We focus on the case of rotationally invariant random coupling matrices in the 'high-dimensional' limit in which the number of samples and the dimension are both large, but with a fixed ratio. This encapsulates many widely studied models, such as Restricted Boltzmann Machines or Generalized Linear Models with correlated data matrices. In this general setting, we show that all the approximation schemes described before are equivalent, and we conjecture that they are exact in the thermodynamic limit in the replica symmetric phases. We achieve this conclusion by resummation of the infinite perturbation series, which generalises a seminal result of Parisi and Potters. A rigorous derivation of this conjecture is an interesting mathematical challenge. On the way to these conclusions, we uncover several diagrammatical results in connection with free probability and random matrix theory, that are interesting independently of the rest of our work.

Contents

1 Introduction **3**
1.1 Background and overview of related works . . . 3
1.2 Structure of the paper, and summary of our contributions . . . 3

2 Symmetric and bipartite spherical models with rotationally-invariant couplings **6**
2.1 Symmetric spherical model . . . 6
2.1.1 Direct free entropy computation . . . 6
2.1.2 Plefka expansion and the Georges-Yedidia formalism . . . 8
2.1.3 Stability of the paramagnetic phase . . . 11
2.2 Bipartite spherical model . . . 12
2.2.1 Direct free entropy computation . . . 13
2.2.2 Plefka expansion . . . 13

⋆ Laboratoire de Physique de l'ENS, PSL University, CNRS, Sorbonnes Universités, Paris, France.
† Institut de Physique Théorique, CNRS, CEA, Université Paris-Saclay, Saclay, France.
◇ University of Havana - Departamento de Física Teórica, Havana, Cuba.
⊗ To whom correspondence shall be sent: antoine.maillard@ens.fr

3 Plefka expansion and Expectation Consistency approximations **14**
3.1 Expectation Consistency, adaptive TAP, and Vector Approximate Message Passing approximations 14
3.1.1 Expectation Consistency approximation . . . 15
3.1.2 Adaptive TAP approximation . . . 16
3.1.3 Vector Approximate Message Passing approximation . . . 17
3.2 Plefka expansion for models of symmetric pairwise interactions . . . 18
3.2.1 A symmetric model with generic priors . . . 18
3.2.2 Connection of the Plefka expansion to EC approximations . . . 19
3.2.3 Application to the Hopfield model . . . 20
3.3 The Replica approach . . . 21
3.3.1 Expectation-Consistency approximation . . . 22
3.3.2 The adaTAP approximation . . . 22
3.3.3 The VAMP approach . . . 23
3.3.4 Plefka expansion . . . 24
3.4 Plefka expansion for models of bipartite pairwise interactions . . . 25
3.4.1 A bipartite model with generic priors . . . 25
3.4.2 Generalized Linear Models with correlated matrices . . . 27

4 Consequences for iterative algorithms **30**
4.1 Generalized Approximate Message Passing for a GLM with i.i.d. matrices . . . 31
4.2 Vector Approximate Message Passing (VAMP) in Compressed Sensing . . . 32
4.2.1 The TAP equations in Compressed sensing . . . 32
4.2.2 TAP equations and the fixed point of the VAMP algorithm . . . 33
4.3 Generalized Vector Approximate Message Passing (G-VAMP) for Generalized Linear Models . 34
4.3.1 The TAP equations from the Plefka expansion . . . 34
4.3.2 The G-VAMP algorithm for Generalized Linear Models . . . 34
4.3.3 TAP equations and fixed points of G-VAMP . . . 35

5 The diagrammatics of the Plefka expansion **36**
5.1 A weaker version of Theorem 1 . . . 37
5.2 The expectation of generic diagrams . . . 39
5.2.1 Eulerian diagrams, strongly irreducible diagrams and simple cycles . . . 40
5.2.2 Cactus diagrams . . . 41
5.3 Concentration of the diagrams: a second moment analysis . . . 41
5.4 The higher-order moments and their influence on the diagrammatics in the symmetric model 43
5.5 Extension to bipartite models . . . 43
5.5.1 Generalization of the previous results to rectangular matrices . . . 44
5.5.2 The higher order moments and their influence on the diagrammatics . . . 46
5.6 A note on i.i.d. matrices . . . 47

A The Georges-Yedidia formalism **52**

B Order 4 of the Plefka expansion for Sec. 2.1. **53**

C Some definitions and reminders of random matrix theory **55**

D Technical derivations and generalizations of the diagrammatics **56**
D.1 Hermitian matrix model . . . 56
D.2 A note on the expectation of diagrams of diverging size . . . 58

1 Introduction

1.1 Background and overview of related works

Many inference and learning tasks can be formulated as a statistical physics problem, where one needs to compute or approximate the marginal distributions of single variables in an interacting model. This is, for instance, the basis behind the popular variational mean-field approach [WJ08]. Going beyond the naive mean-field theory has been a constant goal in both physics and machine learning. One approach, for instance, has been very effective on tree-like structures: the Bethe approximation, or Belief-Propagation. Its development in the statistical physics of disordered systems can be traced back to Thouless-Anderson-Palmer (TAP) [TAP77] and has seen many developments since then [MPV87, YFW03, MM09, ZK16]. Over the last decades, in particular, there has been many works on densely connected models, leading to a myriad of different approximation schemes. In many disordered problems with i.i.d. couplings, a classical approach has been to write the TAP equations as an iterative scheme. Iterative algorithms based on this scheme are often called *Approximate Message Passing* (AMP) [DMM09, KMS+12] in this context.

AMP, or TAP, is an especially powerful approach when the coupling constants in the underlying statistical model are distributed as i.i.d. variables. This is, of course, a strong limitation and many inference schemes have been designed to improve on it: the adaptive TAP (adaTAP) method [OW01a, OW01b], approximation schemes such as *Expectation-Consistency* (EC) [Min01, OW05a] and the recent improvements of AMP such as *Vector approximate Message Passing* (VAMP) and its variants [MP17, RSF17, SRF16, OCW16, ÇOFW16]. Given all these approaches, one may wonder how different they are, and when they actually lead to asymptotically exact inference. In this paper, we wish to address this question using two main tools: high-temperature expansions and random matrix theory.

High-temperature expansions at fixed order parameters (denoted in this paper as "Plefka expansions") are an important tool of the study of disordered systems. In the context of spin glass models, they have been introduced by Plefka [Ple82] for the Sherrington-Kirkpatrick (SK) model, and have been subsequently generalized, in particular by Georges-Yedidia [GY91]. This latter paper provides a systematic way to compute high-temperature (or high-dimension) expansions of the Gibbs free entropy *for a fixed value of the order parameters* (that is Plefka expansions).

One aim of the present paper is to apply this method to a general class of inference problems with pairwise interactions, in which the coupling constants are not i.i.d., but they can have strong correlations, while keeping a rotational invariance that will be made explicit below. In particular, we generalize earlier and inspirational work by Parisi and Potters [PP95], who computed the self-consistent equations for the marginals in Ising models with orthogonal couplings via a resummation of the infinite series given by the high-temperature expansion. We shall show that a similar resummation yields the EC, adaTAP and VAMP formalisms.

1.2 Structure of the paper, and summary of our contributions

In this paper, we perform Plefka expansions for a generic class of models of pairwise interactions with correlated matrices. We provide a detailed derivation of the method, inspired by the work of Georges-Yedidia [GY91] for Ising models, and we include new results on the diagrammatics of the expansions, leveraging rigorous results of random matrix theory. This yields a general framework that encapsulates many known properties of systems sharing this pairwise structure. The main message of this work is that the three successful approximation-schemes that have been developed in the last two decades, Expectation-Consistency, adaTAP or Vector Approximate Message Passing, are equivalent and rely on the same hidden hypothesis. A careful analysis of the Plefka expansion reveals this hypothesis, as it identifies the class of high-temperature expansion diagrams that are effectively kept in these three schemes. A careful diagrammatic analysis leads us to conjecture

that all these methods are asymptotically exact for rotationally-invariant models, in the high-temperature phase. It is also worth noting that although all four methods (Expectation-Consistency, adaTAP, Vector Approximate Message Passing, Plefka expansion) lead to the same mean-field equations, the (most recent) VAMP approach presents the advantage of generating a "natural" way to iterate these equations in order to find a fixed point, which turns them into efficient algorithms. We now turn to a more precise description of the content of the paper. Throughout the paper, we will use two random matrix ensembles that we will both refer to as being *rotationally invariant*. The first one is defined as a measure over the set $\mathcal{S}_N$ of symmetric matrices:

Model S (Symmetric rotationally invariant matrix). *Let $N \geq 1$. $J \in \mathcal{S}_N$ is generated as $J = ODO^\intercal$, in which $O \in \mathcal{O}(N)$ is drawn uniformly from the (compact) orthogonal group $\mathcal{O}(N)$, and $D = \mathrm{Diag}(\{d_i\}_{i=1}^N)$ is a random diagonal matrix, such that its empirical spectral distribution $\rho_D^{(N)} \equiv \frac{1}{N}\sum_{i=1}^N \delta_{d_i}$ converges (almost surely) as $N \to \infty$ to a probability distribution ρ_D with compact support. The smallest and largest eigenvalue of D are assumed to converge almost surely to the infimum and supremum of the support of ρ_D.*

In a similar way, we define an ensemble of rectangular rotationally invariant matrices:

Model R (Rectangular rotationally invariant matrix). *Let $N \geq 1$, and $M = M(N) \geq 1$ such that $M/N \to \alpha > 0$ as $N \to \infty$. $L \in \mathbb{R}^{M\times N}$ is generated via its SVD decomposition $L = U\Sigma V^\intercal$, in which $U \in \mathcal{O}(M)$ and $V \in \mathcal{O}(N)$ are drawn uniformly from their respective orthogonal group. $D \equiv \Sigma^\intercal\Sigma = \mathrm{Diag}(\{d_i\}_{i=1}^N)$ is a diagonal matrix, such that its empirical spectral distribution $\rho_D^{(N)} \equiv \frac{1}{N}\sum_{i=1}^N \delta_{d_i}$ converges (almost surely) as $N \to \infty$ to a probability distribution ρ_D, which has compact support. The smallest and largest eigenvalue of D are assumed to converge almost surely to the infimum and supremum of the support of ρ_D.*

Examples Examples of such random matrix ensembles include matrices generated via a *potential* $V(x)$: one can generate $J \in \mathcal{S}_N$ with a probability density proportional to $e^{-\frac{N}{2}\mathrm{Tr}\, V(J)}$, and this kind of matrix satisfies the hypotheses of Model S. These ensembles also include the following well-known examples:

- The Gaussian Orthogonal Ensemble (GOE), in the case of Model S with a potential $V(x) = x^2/2$.
- The Wishart ensemble with a ratio $\psi \geq 1$. This corresponds to a random matrix $W = XX^\intercal/m$, with $X \in \mathbb{R}^{n\times m}$ an i.i.d. standard Gaussian matrix, and $n, m \to \infty$ with $m/n \to \psi$. This ensemble satisfies Model S, with a potential $V(x) = x - (\psi - 1)\log x$.
- Standard Gaussian i.i.d. rectangular matrices, for Model R. One can also think of them as generated via a potential, as the probability density of such a matrix is $\mathbb{P}(L) \propto e^{-\frac{1}{2}\mathrm{Tr}\, L^\intercal L}$.
- Generically, consider a random matrix L from Model R. Then, both $J_1 \equiv L^\intercal L$ and $J_2 \equiv LL^\intercal$ satisfy the hypotheses of Model S.

The structure of our work is as follows:

- **Spherical models with rotationally invariant couplings** In Sec. 2, we focus on spherical models and we generalize the seminal works of [MPR94a, MPR94b, PP95]. While they studied Ising models with orthogonal couplings, we consider spherical models, just assuming the coupling matrix to be rotationally invariant. We consider two types of models: "symmetric" models with an interaction of the type $\mathbf{x}^\intercal J\mathbf{x}$, in which J follows Model S, and "bipartite" models with interactions of the type $\mathbf{h}^\intercal F\mathbf{x}$, in which F follows Model R. This encapsulates orthogonal couplings, but can also be applied to other random matrix ensembles such as the Gaussian Orthogonal Ensemble (GOE), the Wishart ensemble, and many others. Using diagrammatic results that we derive with random matrix theory, we conjecture a resummation of the Plefka expansion giving the Gibbs free entropy in these models. Our results are in particular consistent with the findings of classical works for Gaussian couplings [Ple82] and orthogonal couplings [PP95].

- **Plefka expansion for statistical models with correlated couplings** Sec. 3 is devoted to the description of the Plefka expansion for different statistical models and inference problems which possess a coupling or data matrix that has rotation invariance properties. We consider models similar to the spherical models of Sec. 2, but with generic prior distributions on the underlying variables. In Sec. 3.1, we recall the Expectation-Consistency (EC), adaTAP and VAMP approximations and comment briefly on their respective history, before showing that they are equivalent. As a consequence, we will generically refer to these approximations as the *Expectation-Consistency approximations* (EC). We hope that our paper will help providing a unifying presentation of these works, generalizing them by leveraging random matrix theory. Our main conjecture for this part can be stated as the following:

  Conjecture 1. *[Informal] For statistical models of symmetric or bipartite interactions with coupling matrices that satisfy respectively Model S or Model R, the three equivalent approximations, Expectation-Consistency, adaTAP and VAMP (generically denoted EC approximations), are exact in the large size limit in the high temperature phase.*

  We believe that the validity of the above conjecture extends beyond the high temperature phase. In particular that it is correct for inference problems in the Bayes-optimal setting, and more generally anytime the system is in a replica symmetric phase as defined in [MPV87].

  The approximation behind EC approximations can be checked order by order using our high-temperature Plefka expansions technique and its resummation. We then derive Plefka expansions for these generic models, and we apply it to different situations, namely:

  - In Sec. 3.2.1 we perform a Plefka expansion for a generic symmetric rotationally invariant model with pairwise interactions. Using this method and our diagrammatic results, we show then in Sec. 3.2.2 that the EC approximations are exact for these models in the large size limit.
  - In Sec. 3.2.3 we apply our general result to the TAP free energy of the Hopfield model [Hop82], an Ising spin model with a correlated matrix of the Wishart ensemble, used as a basic model of neural network. In particular, we find back straightforwardly the results of [NT97] and [Méz17].
  - In Sec. 3.3 we extend our Plefka expansion and the corresponding diagrammatic techniques to the study of a *replicated* system, in which we constraint the overlap between different replicas. The interest for such systems comes as a consequence of the celebrated replica method of theoretical physics [MPV87].
  - Finally, we show in Sec. 3.4 how we can use these results to derive the Plefka-expanded free entropy for a very broad class of bipartite models, which includes the Generalized Linear Models (GLMs) with correlated data matrices, and the Compressed Sensing problem.

  We emphasize that we were able to derive the free entropy of all these models using very generic arguments relying only on the rotational invariance of the problem.
- **The TAP equations and message passing algorithms** Finally, we show in Sec. 4 that the TAP (or EC) equations that we derived by maximizing the Gibbs free entropy of rotationally invariant models can strikingly be understood as the fixed point equations of message passing algorithms. In the converse way, many message-passing algorithms can be seen as an iteration scheme of the TAP equations. This was known in many models in which the underlying data matrix was assumed to be i.i.d. For instance, the Generalized Approximate Message Passing (GAMP) algorithm [Ran11] was shown in [KMS+12] to be equivalent to the TAP equations, a result that we find back in Sec. 4.1, while TAP equations were already iterated for Restricted Boltzmann Machines, see [TGM+18]. In the Plefka expansion language, these results relied on the early stopping of the expansion at order 2 (in powers of the couplings) as a

consequence of the i.i.d. hypothesis. Using our resummation results to deal with the series at infinite orders, we were able to generalize these correspondences to correlated models. We argue that the stationary limit of the Vector Approximate Message Passing (VAMP) algorithm [RSF17] (that is its fixed point equations) for compressed sensing with correlated matrices gives back our TAP equations derived via Plefka expansion, see Sec. 4.2. Even more generally, the Generalized Vector Approximate Passing (G-VAMP) algorithm [SRF16], defined for the very broad class of Generalized Linear Models with correlated matrices, yields fixed point equations that are equivalent to our Plefka-expanded TAP equations, see Sec. 4.3. Combined with the results of Sec. 3, this indicates that the VAMP algorithm is an example of an approximation scheme that follows Conjecture 1.

- **Diagrammatics of the expansion and random matrix theory** Our results are largely based on a better control on the diagrammatics of the Plefka expansions for rotationally invariant random matrices, which are presented in Sec. 5. We leverage mathematically rigorous results on Harish-Chandra-Itzykson-Zuber (HCIZ) integrals [HC57, IZ80, GM05, CŚ07], involving transforms of the asymptotic spectrum of the coupling matrix, to argue that only a very specific class of diagrams contributes to the high-temperature expansion of a system with rotationally invariant couplings. These results are used throughout our study, and are detailed in Sec. 5. Some generalizations are postponed to Appendix D.

2 Symmetric and bipartite spherical models with rotationally-invariant couplings

In this section we consider two spherical models that will serve both as guidelines and building blocks for our subsequent analysis. We show in details how to perform the Plefka-Georges-Yedidia high-temperature expansion in this context, and the precise diagrammatic results that allow us to resum the Plefka series for rotationally invariant couplings. These results will be useful to clarify our subsequent derivation of the TAP equations in more involved models, and are also interesting by themselves from a random matrix theory point of view.

2.1 Symmetric spherical model

In this section $N \geq 1$, $\sigma > 0$, and we define the following pairwise interaction Hamiltonian on $\mathbb{S}^{N-1}(\sigma\sqrt{N})$, the N-th dimensional sphere of radius $\sigma\sqrt{N}$:

$$H_J(\mathbf{x}) = -\frac{1}{2}\mathbf{x}^\intercal J\mathbf{x} = -\frac{1}{2}\sum_{1\leq i,j\leq N} J_{ij}x_i x_j, \qquad \mathbf{x} \in \mathbb{S}^{N-1}(\sigma\sqrt{N}). \tag{1}$$

The coupling matrix J is a $N \times N$ symmetric random matrix drawn from Model S.

2.1.1 Direct free entropy computation

The Gibbs measure for our model at inverse temperature β is defined as:

$$P_{\beta,J}(\mathrm{d}\mathbf{x}) \equiv \frac{1}{Z_{\beta,J}} e^{\frac{\beta}{2}\sum_{i,j} J_{ij}x_i x_j}\mathrm{d}\mathbf{x}, \tag{2}$$

in which $\mathrm{d}\mathbf{x}$ is the usual surface measure on the sphere $\mathbb{S}^{N-1}(\sigma\sqrt{N})$. We write the partition function of the model introducing a Lagrange multiplier γ to enforce the condition $\|\mathbf{x}\|^2 = N\sigma^2$. We will write $A_N \simeq B_N$ to

denote that $\frac{1}{N}\log A_N = \frac{1}{N}\log B_N + \mathcal{O}_N(1)$. At leading exponential order, one has:

$$Z_{\beta,J} \equiv \int_{\mathbb{S}^{N-1}(\sigma\sqrt{N})} \mathrm{d}\mathbf{x}\, e^{\frac{\beta}{2}\sum_{i,j} J_{ij}x_i x_j}, \tag{3}$$

$$\simeq \int \mathrm{d}\gamma \prod_{i=1}^{N} \int_{\mathbb{R}} \mathrm{d}x_i\, e^{\frac{\beta}{2}\sum_{i,j} J_{ij}x_i x_j + \frac{\gamma}{2}(N\sigma^2 - \sum_i x_i^2)},$$

$$\simeq \exp\left[\sup_{\gamma}\left\{\log\left[\int \prod_{i=1}^{N} \mathrm{d}x_i\, e^{\frac{\beta}{2}\sum_{i,j} J_{ij}x_i x_j + \frac{\gamma}{2}(N\sigma^2 - \sum_i x_i^2)}\right]\right\}\right]. \tag{4}$$

Denoting $\gamma(\beta)$ the solution to the saddle-point equation in eq. (4), we have effectively defined a new Gibbs measure:

$$P_{\beta,J}(\mathbf{x}) \equiv \frac{1}{Z_{\beta,J}(\gamma)} e^{\frac{\beta}{2}\sum_{i,j} J_{ij}x_i x_j} e^{-\frac{\gamma(\beta)}{2}||\mathbf{x}||_2^2}\, \mathrm{d}\mathbf{x}, \tag{5}$$

where now $\mathrm{d}\mathbf{x}$ is the usual Euclidian measure on $\mathbb{R}^N$. Following [KTJ76] we diagonalize the Hamiltonian and we integrate over the spins in this new basis, which yields:

$$Z_{\beta,J} \simeq \exp\left[\sup_{\gamma}\left\{\frac{N}{2}\left(\log 2\pi + \gamma\sigma^2 - \frac{1}{N}\sum_{\lambda}\log(\gamma - \beta\lambda)\right)\right\}\right], \tag{6}$$

in which the sum over λ runs over the set of eigenvalues of J. Taking the $N \to \infty$ limit, the saddle point equation reads:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{\lambda} \frac{1}{\gamma - \beta\lambda} = \sigma^2, \tag{7}$$

which we can write as a function of the limiting spectral law ρ_D of the matrix J (defined in Model S):

$$\int \frac{\rho_D(\mathrm{d}\lambda)}{\gamma - \beta\lambda} = \sigma^2. \tag{8}$$

We assumed (see Model S) that the support of ρ_D is compact so that we can define its maximum $\lambda_{\max} \in \mathbb{R}$. Under these assumptions, eq. (8) has the solution:

$$\gamma = \beta\mathcal{R}_{\rho_D}(\beta\sigma^2) + \frac{1}{\sigma^2} = \beta\mathcal{S}^{-1}_{\rho_D}(-\beta\sigma^2), \tag{9}$$

as long as $-\mathcal{S}_{\rho_D}(\lambda_{\max}) \geq \beta\sigma^2$, where $\mathcal{R}_{\rho_D}$ is the $\mathcal{R}$-transform of ρ_D and $\mathcal{S}_{\rho_D}$ its Stieltjes transform (see Appendix C for their definitions). In the opposite case (if $\beta\sigma^2 > -\mathcal{S}_{\rho_D}(\lambda_{\max})$), γ 'sticks' to the solution $\gamma = \lambda_{\max}\beta$. The intensive free entropy $\Phi_J(\beta)$ is defined as:

$$\Phi_J(\beta) \equiv \lim_{N\to\infty} \frac{1}{N}\log Z_{\beta,J}. \tag{10}$$

In the end, we can compute the free entropy in the *high-temperature* phase $\beta \leq \beta_c \equiv -\sigma^{-2}\mathcal{S}_{\rho_D}(\lambda_{\max})$:

$$\Phi_J(\beta) = \frac{1}{2}\left(1 + \log 2\pi\sigma^2\right) + \frac{\beta\sigma^2}{2}\mathcal{R}_{\rho_D}(\beta\sigma^2) - \frac{1}{2}\int \rho_D(\mathrm{d}\lambda)\log\left[\beta\sigma^2\mathcal{R}_{\rho_D}(\beta\sigma^2) - \beta\sigma^2\lambda + 1\right]. \tag{11}$$

By taking the derivative of this expression with respect to β it is easy to show that this simplifies to:

$$\Phi_J(\beta) = \frac{1}{2}\left(1 + \log 2\pi\sigma^2\right) + \frac{1}{2}\int_0^{\beta\sigma^2} \mathcal{R}_{\rho_D}(x)\mathrm{d}x. \tag{12}$$

In the low temperature phase (for $\beta \geq \beta_c = -\sigma^{-2}\mathcal{S}_{\rho_D}(\lambda_{\max})$) one has

$$\Phi_J(\beta) = \frac{1}{2}\left(\log 2\pi + \lambda_{\max}\beta\sigma^2 - \log\beta - \int \rho_D(\mathrm{d}\lambda)\ \log(\lambda_{\max} - \lambda)\right). \tag{13}$$

Note that both in the high and low temperature phases the free entropy can formally be expressed as:

$$\Phi_J(\beta) = \frac{1}{2}\log 2\pi + \frac{1}{2}\sup_{\gamma}\left[\gamma\sigma^2 - \int \rho_D(\mathrm{d}\lambda)\log(\gamma - \beta\lambda)\right], \tag{14}$$

a formulation which is both more compact and easier to implement algorithmically for generic matrices J.

Remark The free entropy is usually defined as an average over the quenched disorder J, but here it is clear that the free entropy is self-averaging as a function of J, so that taking this average is trivial. Moreover, $\Phi_J(\beta)$ only depends on J via ρ_D, its asymptotic eigenvalue distribution.

Remark The derivation of the free entropy both in the high and low temperature phase has been made rigorous in [GM05], and the method of proof also essentially consists in fixing a Lagrange multiplier to enforce the condition $\sum_i s_i^2 = \sigma^2 N$.

2.1.2 Plefka expansion and the Georges-Yedidia formalism

A more generic way to compute the free entropy is to follow the formalism of [GY91] to perform a high-temperature Plefka expansion [Ple82]. The goal is to expand the free entropy at low β, in the high-temperature phase. In order to do so, we introduce the very useful U operator defined in Appendix A of [GY91]. We will compute the free entropy given the constraints on the means $\langle x_i\rangle_\beta = m_i$ and on the variances $\langle x_i^2\rangle_\beta = v_i + m_i^2$. The notation $\langle\cdot\rangle_\beta$ indicates an average over the Gibbs measure of our system at inverse temperature β, see eq. (5). A set of parameters $\{m_i, v_i\}$ will thus determine a free entropy value, and the comparison with the direct calculation of Sec. 2.1.1 will be made by maximizing the free entropy with respect to $\{m_i, v_i\}$. We can enforce the spherical constraint $\|\mathbf{x}\|_2^2 = \sigma^2 N$ by constraining our choice of parameters $\{m_i, v_i\}$ to satisfy the identity:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left[v_i + m_i^2\right]. \tag{15}$$

The Lagrange parameters introduced to fix the magnetizations are denoted $\{\lambda_i\}$, and the ones used to fix the variances are denoted $\{\gamma_i\}$. For clarity we will keep their dependency on β explicit only when needed. For a given β and a given J one defines the operator U of Georges-Yedidia:

$$U(\beta, J) \equiv H_J - \langle H_J\rangle_\beta + \sum_{i=1}^{N}\partial_\beta\lambda_i(\beta)(x_i - m_i) + \frac{1}{2}\sum_{i=1}^{N}\partial_\beta\gamma_i(\beta)\left[x_i^2 - v_i - m_i^2\right], \tag{16}$$

The derivation of U as well as its (many) useful properties are briefly recalled in Appendix A. We are now ready to compute the first orders of the expansion of the free entropy $\Phi_J(\beta)$ in terms of β. In this expansion

the Lagrange parameters $\{\lambda_i(\beta), \gamma_i(\beta)\}$ are always considered at $\beta = 0$, so we drop their β-dependency. We detail the first orders of the expansion, following Appendix A (cft. Appendix A of [GY91]).

Order 0 First of all, taking $\beta = 0$ one has easily:

$$\begin{aligned}\Phi_J(\beta = 0) &= \frac{1}{2N}\sum_{i=1}^{N}\gamma_i(v_i + m_i^2) + \frac{1}{N}\sum_{i=1}^{N}\lambda_i m_i + \frac{1}{N}\log\int_{\mathbb{R}^N} e^{-\frac{1}{2}\sum_i \gamma_i x_i^2 - \sum_i \lambda_i x_i}\mathrm{d}\mathbf{x},\\ &= \frac{1}{2}\log 2\pi + \frac{1}{N}\sum_{i=1}^{N}\left[\frac{\gamma_i}{2}(v_i + m_i^2) - \frac{1}{2}\log\gamma_i + \lambda_i m_i + \frac{\lambda_i^2}{2\gamma_i}\right].\end{aligned}$$

This yields after extremization over $\{\lambda_i, \gamma_i\}$:

$$\Phi_J(\beta = 0) = \frac{1}{2}\left[1 + \log 2\pi\right] + \frac{1}{2N}\sum_{i=1}^{N}\log v_i. \tag{17}$$

Order 1 At order 1, one easily derives:

$$\left(\frac{\partial\Phi_J}{\partial\beta}\right)_{\beta=0} = -\frac{1}{N}\left\langle H_J\right\rangle_{\beta=0} = \frac{1}{2N}\sum_{i,j}J_{ij}m_i m_j + \frac{1}{2N}\sum_{i=1}^{N}J_{ii}v_i. \tag{18}$$

We can now make use of the Maxwell-type relations which are valid at any β:

$$\gamma_i(\beta) = 2N\frac{\partial\Phi_J(\beta)}{\partial v_i}, \tag{19}$$

$$m_i\gamma_i(\beta) + \lambda_i(\beta) = N\frac{\partial\Phi_J(\beta)}{\partial m_i}. \tag{20}$$

These relations plugged in eq. (18) lead to $\partial_\beta\gamma_i(\beta = 0) = J_{ii}$ and $\partial_\beta\lambda_i(\beta = 0) = \sum_{j(\neq i)}J_{ij}m_j$. We then obtain the U operator at $\beta = 0$ from eq. (16):

$$U(\beta = 0, J) = -\frac{1}{2}\sum_{i\neq j}J_{ij}(x_i - m_i)(x_j - m_j). \tag{21}$$

Order 2 Following eq. (190) in Appendix A, we have the relation:

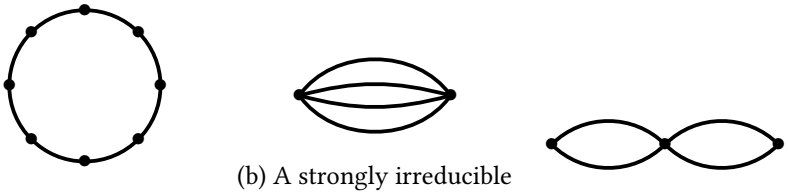
$$\frac{1}{2}\left(\frac{\partial^2\Phi_J}{\partial\beta^2}\right)_{\beta=0} = \frac{1}{2N}\left\langle U^2\right\rangle_{\beta=0} = \frac{1}{4N}\sum_{i\neq j}J_{ij}^2 v_i v_j. \tag{22}$$

Order 3 and 4 For the order 3, we obtain:

$$\frac{1}{3!}\left(\frac{\partial^3\Phi_J}{\partial\beta^3}\right)_{\beta=0} = -\frac{1}{6N}\left\langle U^3\right\rangle_{\beta=0} = \frac{1}{6N}\sum_{i,j,k}J_{ij}J_{jk}J_{ki}v_i v_j v_k + \mathcal{O}_N(1), \tag{23}$$

in which the sum is made over pairwise distinct i, j, k indices. Applying eq. (192) we reach:

$$\frac{1}{4!}\left(\frac{\partial^4\Phi_J}{\partial\beta^4}\right)_{\beta=0} = \frac{1}{8N}\sum_{i,j,k,l}J_{ij}J_{jk}J_{kl}J_{li}v_i v_j v_k v_l + \mathcal{O}_N(1), \tag{24}$$



(a) A simple cycle of order $p = 8$.

(b) A strongly irreducible negligible diagram of order $p = 4$.

(c) A cactus diagram at order $p = 4$.

Figure 1: Representation of the expansion eq. (25) with diagrams. Each vertex represents an index i_α and carries a factor v_{i_α}, while each edge is a factor J_{ij}.

where again, i, j, k, l are pairwise distinct indices. For pedagogical purposes (and since it will be useful for the following sections), we detail this calculation in Appendix B.

Larger orders By its very nature, the perturbative expansion of Georges-Yedidia [GY91] can not (somehow disappointingly) give an analytic result for an arbitrary perturbation order n. However, the results up to order 4 of eqs. (17), (18), (22), (23), (24) lead to the following natural conjecture for the free entropy *at a given realization of the disorder*:

$$\Phi_J(\beta) = \frac{1}{2}\left[1 + \log 2\pi\right] + \frac{1}{2N}\sum_{i=1}^{N} \log v_i + \frac{\beta}{2N}\sum_{i\neq j} J_{ij} m_i m_j$$
$$+ \frac{1}{N}\sum_{p=1}^{\infty} \frac{\beta^p}{2p} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distincts}}} J_{i_1 i_2} J_{i_2 i_3} \cdots J_{i_{p-1} i_p} J_{i_p i_1} \prod_{\alpha=1}^{p} v_{i_\alpha} + \mathcal{O}_N(1). \tag{25}$$

Note that in order to obtain this formula, we took the $N \to \infty$ limit at every perturbation order in β, which is part of the implicit assumptions of the Plefka expansion. The terms of this perturbative expansion can be represented diagrammatically as *simple cycles* of order p, see Fig. 1a.

In general, at any order in the expansion one can construct a diagrammatic representation of the contributing terms, and one expects that only strongly irreducible diagrams contribute to the free entropy. Strongly irreducible diagrams are those that cannot be split into two pieces by removing a vertex [GY91] (examples are given in Fig. 1a and 1b). However we retain only simple cycles as the one depicted in Fig. 1a because other diagrams as in Fig. 1b are negligible when $N \to \infty$ for rotationally invariant models, as we argue in Section 5.2. For the case of orthogonal couplings, this dominance of simple cycles was already noted in [PP95]. On the other hand, generic cactus diagrams like the one pictured in Fig. 1c are not negligible, but they cancel out and do not appear in the final form of the expansion (at order 4, this is shown in Appendix B).

We shall now prove the dominance of simple cycles, and the correctness of eq. (25), in the high-temperature phase. In this phase, the solution to the maximization of eq. (25) under $\{m_i\}$ is the paramagnetic solution $m_i = 0$. Furthermore, we expect that the $\{v_i\}$ that maximize the free entropy of eq. (25) are *homogeneous*, that is $\forall i,\ v_i = v$. The constraint of eq. (15) thus gives $v = \sigma^2$.

We can compare the result of the resummation of simple cycles, eq. (25) with the exact results of eq. (12) in the paramagnetic phase. For these two results to agree, we need the generating function for simple cycles to be related to the $\mathcal{R}$-transform of ρ_D by:

$$\mathbb{E}\left[\frac{1}{N}\sum_{p=1}^{\infty}\frac{\beta^p\sigma^{2p}}{2p}\sum_{\substack{i_1,\cdots,i_p\\ \text{pairwise distincts}}} J_{i_1i_2}J_{i_2i_3}\cdots J_{i_{p-1}i_p}J_{i_pi_1}\right] = \frac{1}{2}\int_0^{\beta\sigma^2}\mathcal{R}_{\rho_D}(x)\mathrm{d}x, \tag{26}$$

in which the outer expectation is with respect to the distribution of J. In particular, an order-by-order comparison yields that the *free cumulants* $\{c_p(\rho_D)\}_{p\in\mathbb{N}^\star}$ (see Appendix C for their definition) must satisfy:

$$\forall p\in\mathbb{N}^\star,\quad c_p(\rho_D) = \lim_{N\to\infty}\mathbb{E}\left[\frac{1}{N}\sum_{\substack{i_1,\cdots,i_p\\ \text{pairwise distincts}}} J_{i_1i_2}J_{i_2i_3}\cdots J_{i_{p-1}i_p}J_{i_pi_1}\right]. \tag{27}$$

Using rigorous results of [GM05], we were able to prove a stronger version of eq. (27), namely convergence in L^2 norm, so we state it as a theorem:

Theorem 1. *For a matrix $J\in\mathcal{S}_N$ generated by Model S, one has for every $p\in\mathbb{N}^\star$:*

$$\lim_{N\to\infty}\mathbb{E}\left|\frac{1}{N}\sum_{\substack{i_1,\cdots,i_p\\ \textit{pairwise distincts}}} J_{i_1i_2}J_{i_2i_3}\cdots J_{i_{p-1}i_p}J_{i_pi_1} - c_p(\rho_D)\right|^2 = 0.$$

We postpone the proof to Sec. 5. We assume that we can invert the summation over p and the $N\to\infty$ limit in eq. (25), so Theorem 1 implies that eq. (26) is true not only in expectation but that we can write:

$$\lim_{N\to\infty}\frac{1}{N}\sum_{p=1}^{\infty}\frac{\beta^p\sigma^{2p}}{2p}\sum_{\substack{i_1,\cdots,i_p\\ \text{pairwise distincts}}} J_{i_1i_2}J_{i_2i_3}\cdots J_{i_{p-1}i_p}J_{i_pi_1} = \frac{1}{2}\int_0^{\beta\sigma^2}\mathcal{R}_{\rho_D}(x)\mathrm{d}x, \tag{28}$$

in which the limit here means convergence in L^2 norm as $N\to\infty$. This is important, as it allows to "resum" the free entropy of eq. (25), which is valid for a given instance of J.

As a final note, we can use the results of Sec. 2.1.1 to write our result in an alternative form (dropping $\mathcal{O}_N(1)$ terms):

$$\frac{1}{N}\sum_{p=1}^{\infty}\frac{\beta^p\sigma^{2p}}{2p}\sum_{\substack{i_1,\cdots,i_p\\ \text{pairwise distincts}}} J_{i_1i_2}\cdots J_{i_pi_1} = \frac{1}{2}\sup_{\gamma}\left[\gamma\sigma^2 - \int\rho_D(\mathrm{d}\lambda)\log(\gamma-\beta\lambda)\right] - \frac{1+\log\sigma^2}{2}. \tag{29}$$

2.1.3 Stability of the paramagnetic phase

We can check whether the paramagnetic solution is stable exactly up to the temperature $\beta=\beta_c$. Recall that in this model we do not optimize the free entropy simultaneously over v and the $\{m_i\}$, because the norm $||\mathbf{x}||_2^2=\sigma^2N$ is fixed, yielding the constraint $v=\sigma^2-\frac{1}{N}\sum_i m_i^2$. Solely as a function of the $\{m_i\}$, the free

entropy therefore reads, up to $\mathcal{O}_N(1)$ terms:

$$\Phi_J(\beta) = \frac{1+\log 2\pi}{2} + \frac{1}{2}\log\left[\sigma^2 - \frac{1}{N}\sum_{i=1}^{N} m_i^2\right] + \frac{\beta}{2N}\sum_{i\neq j} J_{ij} m_i m_j + G_{\rho_D}\left(\beta\left[\sigma^2 - \frac{1}{N}\sum_{i=1}^{N} m_i^2\right]\right), \tag{30}$$

in which G_{ρ_D} is the integrated $\mathcal{R}$-transform of ρ_D, see Appendix C for its definition. The Hessian of the *extensive* free entropy $N\Phi_J$ at the paramagnetic solution $\mathbf{m} = \mathbf{0}$ is:

$$N\left(\frac{\partial^2 \Phi_J}{\partial m_i \partial m_j}\right)_{m=0} = -\frac{\delta_{ij}}{\sigma^2}\left[1 + \beta\sigma^2 \mathcal{R}_{\rho_D}(\beta\sigma^2)\right] + \beta J_{ij} + \mathcal{O}_N(1)\,. \tag{31}$$

The paramagnetic solution is stable as long as the Hessian $N\partial_m^2 \Phi_J(\beta, m=0)$ is a negative matrix. This is true as long as $\beta < \beta_c = -\sigma^{-2}\mathcal{S}_{\rho_D}(\lambda_{\max})$, because at β_c the spectrum of $N\left(\frac{\partial^2 \Phi_J}{\partial m_i \partial m_j}\right)_{m=0}$ touches zero. For $\beta > \beta_c$ the Hessian is again negative, giving the impression of stability of the paramagnetic phase, however $\mathcal{S}_{\rho_D}^{-1}(-\beta\sigma^2)$ is evaluated in the non physical solution, so the solution has to be discarded. Our Plefka expansion allows thus to compute the free entropy in the whole paramagnetic phase, coherently with the results of Sec. 2.1.1. More generically, as shown by Plefka [Ple82] in the closely-related SK model, the Hessian of the free entropy with respect to $\{m_i\}$ is related to the inverse susceptibility matrix of the system, and thus the non-inversibility of the Hessian implies a non-analyticity point of the free entropy.

Validity of the Plefka expansion and stability of the replica symmetric solution We believe it is an open question to relate the range of validity β_c of the Plefka expansion and the de Almeida-Thouless condition that characterizes the local stability of the replica symmetric solution (see [DAT78] for its original derivation, and [Kab08a, SK08] for examples of its applications in inference problems). The equivalence of these two conditions was shown in the seminal paper of Plefka [Ple82] in the Sherrington-Kirpatrick model. It is tedious but straightforward to generalize this conclusion to a model with Ising spins $x_i = \pm 1$ and a Hamiltonian given by eq. (1) with a rotationally-invariant coupling matrix J drawn from Model S. However, investigating the relation between these two conditions in a general model appears to be an open problem, and is beyond the scope of our paper.

The region of validity of the expansion and the free cumulant series In this remark, we clarify some possible confusion about the radius of convergence of the free cumulant series and the domain of validity of the Plefka expansion. We performed an expansion of $\Phi_J(\beta)$ close to $\beta = 0$, which implies that this expansion is thus valid in the region $(0, \beta_c)$, in which $\beta_c = -\sigma^{-2}\mathcal{S}(\lambda_{\max})$ is the first non-analyticity of $\Phi_J(\beta)$, see eqs. (12), (13) and the discussion above. Note that there exists spectrums for which the function $\mathcal{R}_{\rho_D}(x)$ can be analytically extended beyond $x_c \equiv -\mathcal{S}_{\rho_D}(\lambda_{\max})$, as for instance Wigner's semi-circle law, for which $\mathcal{R}_{\text{s.c.}}(x) = x$. Yet, one has to be careful that this does not imply that the free entropy $\Phi_J(\beta)$ is analytic beyond β_c, and thus even in this case our Plefka expansion is *a priori* only valid up to $\beta = \beta_c$.

2.2 Bipartite spherical model

In this section we consider $N, M \geq 1$. We let $\alpha > 0$, and we will take the limit (sometimes referred to as the *thermodynamic limit*) in which $N, M \to \infty$ with a fixed ratio $M/N \to \alpha$. We let $\sigma_x, \sigma_h > 0$. Let us consider

the following Hamiltonian, which is a function of two fields $\mathbf{h} \in \mathbb{R}^M$ and $\mathbf{x} \in \mathbb{R}^N$:

$$H_L(\mathbf{h}, \mathbf{x}) = -\mathbf{h}^\intercal L \mathbf{x} = -\sum_{\mu=1}^{M}\sum_{i=1}^{N} L_{\mu i} h_\mu x_i, \qquad \mathbf{h} \in \mathbb{S}^{M-1}(\sigma_h\sqrt{M}), \quad \mathbf{x} \in \mathbb{S}^{N-1}(\sigma_x\sqrt{N}). \tag{32}$$

The coupling matrix $L \in \mathbb{R}^{M\times N}$ is assumed to be drawn from Model R.

2.2.1 Direct free entropy computation

The calculation for this bipartite case is very similar to the calculation performed in Sec. 2.1.1, although one can not always express the result as a well-known transform of the measure ρ_D of Model S. For all values of β, the result can be expressed as:

$$\begin{aligned}\Phi_L(\beta) &\equiv \lim_{N\to\infty} \frac{1}{N} \log \int \mathrm{d}\mathbf{h} \int \mathrm{d}\mathbf{x}\, e^{\beta \mathbf{h}^\intercal L \mathbf{x}}, \\ &= \frac{1+\alpha}{2}\log 2\pi + \frac{1}{2}\sup_{\gamma_h,\gamma_x}\left[\alpha\gamma_h\sigma_h^2 + \gamma_x\sigma_x^2 - (\alpha-1)\log\gamma_h - \int \rho_D(\mathrm{d}\lambda)\log(\gamma_x\gamma_h - \beta^2\lambda)\right], \end{aligned} \tag{33}$$

where ρ_D is the asymptotic eigenvalue distribution of $L^T L$ (see the definition of Model R).

2.2.2 Plefka expansion

The Plefka expansion for this model is a straightforward generalization of Sec. 2.1.2. We will fix the averages to be $\langle h_\mu\rangle = m_\mu^h$ and $\langle x_i\rangle = m_i^x$, and the second moments $\langle h_\mu^2\rangle = v_\mu^h + (m_\mu^h)^2$ and $\langle x_i^2\rangle = v_i^x + (m_i^x)^2$, again with the constraints $\sigma_h^2 = \frac{1}{M}\sum_{\mu=1}^{M}\left[v_\mu^h + (m_\mu^h)^2\right]$ and $\sigma_x^2 = \frac{1}{N}\sum_{i=1}^{N}\left[v_i^x + (m_i^x)^2\right]$. In this problem, the U operator of [GY91] at $\beta = 0$ is given by:

$$U(\beta = 0, L) = -\sum_{\mu,i} L_{\mu i}(h_\mu - m_\mu^h)(x_i - m_i^x). \tag{34}$$

Once again, as in Sec. 2.1.2, one can study all the diagrams that appear in the Plefka expansion. We show again the L^2 concentration of the simple cycles, and the negligibility of other strongly irreducible diagrams that can be constructed from the rectangular L matrix. We state in more details these results for the bipartite case in Sec. 5.5.1. We obtain the following result, a counterpart to eq. (25) for this bipartite model:

$$\begin{aligned}\Phi_L(\beta) = \frac{1+\alpha}{2}\left[1+\log 2\pi\right] + \frac{\alpha}{2M}\sum_{\mu=1}^{M}\log v_\mu^h + \frac{1}{2N}\sum_{i=1}^{N}\log v_i^x + \frac{\beta}{N}\sum_{\mu=1}^{M}\sum_{i=1}^{N} L_{\mu i} m_\mu^h m_i^x \\ + \frac{1}{N}\sum_{p=1}^{\infty}\frac{\beta^{2p}}{2p}\sum_{\substack{\mu_1,\cdots,\mu_p \\ \text{pairwise distincts}}}\sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distincts}}} L_{\mu_1 i_1} L_{\mu_1 i_2} L_{\mu_2 i_2}\cdots L_{\mu_p i_p} L_{\mu_p i_1}\prod_{\alpha=1}^{p} v_{\mu_\alpha}^h v_{i_\alpha}^x + \mathcal{O}_N(1),\end{aligned} \tag{35}$$

in which indices $\{\mu_l\}$ run from 1 to M and indices $\{i_l\}$ run from 1 to N. We make again an assumption of uniform variances at the maximum: $v_\mu^h = v^h, v_i^x = v^x$. Comparing to eq. (33) in the paramagnetic $m_\mu^h, m_i^x = 0$ phase, we obtain the correspondence, similar to eq. (28) and valid *a priori* for any given realization of L, in the

high temperature phase:

$$\frac{\alpha}{2}\log\sigma_h^2+\frac{1}{2}\log\sigma_x^2+\frac{1}{N}\sum_{p=1}^{\infty}\frac{\beta^{2p}\sigma_h^{2p}\sigma_x^{2p}}{2p}\sum_{\substack{\mu_1,\cdots,\mu_p\\ \text{pairwise distincts}}}\sum_{\substack{i_1,\cdots,i_p\\ \text{pairwise distincts}}}L_{\mu_1 i_1}L_{\mu_1 i_2}L_{\mu_2 i_2}\cdots L_{\mu_p i_p}L_{\mu_p i_1}$$
$$=-\frac{1+\alpha}{2}+\frac{1}{2}\sup_{\gamma_h,\gamma_x}\left[\alpha\gamma_h\sigma_h^2+\gamma_x\sigma_x^2-(\alpha-1)\log\gamma_h-\int\rho_D(\mathrm{d}\lambda)\log(\gamma_h\gamma_x-\beta^2\lambda)\right]. \quad (36)$$

3 Plefka expansion and Expectation Consistency approximations

In this section we perform Plefka expansions for generic models of pairwise interactions, both symmetric and bipartite. In Sec. 3.1 we recall some known facts on the Expectation Consistency (also called Expectation Propagation [Min01]), adaTAP and VAMP approximations to compute the free entropy of such models. In Sec. 3.2 and Sec. 3.4 we generalize the results of the Plefka expansions of Sec. 2 to these models and highlight the main differences and assumptions of our method. This yields a very precise and systematic justification of the TAP equations for rotationally invariant models. We apply these results to retrieve the TAP free entropy of the Hopfield model, Compressed Sensing, as well as different variations of high-dimensional inference models called Generalized Linear Models (GLMs). Sec. 3.3 is devoted to the study of a generic replicated system using these approximations, and the Plefka expansion. We show that they can be used in the celebrated replica method [MPV87] of theoretical physics, to compute the Gibbs free entropy of a generic pairwise inference model.

3.1 Expectation Consistency, adaptive TAP, and Vector Approximate Message Passing approximations

Expectation Consistency (EC) [OW05a, OW05b], is an approximation scheme for a generic class of disordered systems that can also be applied to many inference problems. In this section we show how this scheme is derived and is closely related to the adaTAP approximation [OW01a, OW01b], and the VAMP approximation [RSF17]. Let us shortly comment on the history of these methods. The adaTAP scheme was developed and presented in 2001 in [OW01b, OW01a], and was discussed in details in the review [OS01] for systems close to the SK model. The same year, Thomas Minka's Expectation Propagation (EP) approach was presented [Min01]. Opper and Winther used an alternative view of local-consistency approximations of the EP–type which they call Expectation Consistent (EC) approximations in [OW05a, OW05b], effectively rederiving their adaTAP scheme from this new point of view. The VAMP approach is more recent [SRF16], and is again another take EP for a different problem (compressed sensing) but it has the advantage that, compared with other EP-like approaches [ÇOFW16] it leads to a practical converging algorithm, and a rigorous treatement of its time evolution. The connection between these approaches and the Parisi-Potters formulation for inference problems [JR16] was hinted several times for SK-like problems, see e.g. [OCW16, ÇO19]. We hope that our paper will help providing a unifying presentation of these works, generalizing them way beyond the SK model alone by leveraging random matrix theory. We recall briefly the main arguments of these papers which are useful for our discussion.

3.1.1 Expectation Consistency approximation

Consider a model in which the density of a vector $\mathbf{x} \in \mathbb{R}^N$ is given by a probability distribution of the form:

$$P(\mathbf{x}) = \frac{1}{Z} P_0(\mathbf{x}) P_J(\mathbf{x}). \tag{37}$$

Such distributions typically appear in Bayesian approaches to inference problems. We will use the Bayesian language and denote P_0 as a *prior* distribution on $\mathbf{x}$, which will be typically factorized (all the components of $\mathbf{x}$ are assumed to be independent under P_0) ; The distribution P_J is responsible for the interactions between the $\{x_i\}$. In this paper we are interested in pairwise interactions, which means that the log of P_J is a quadratic form in the $\{x_i\}$ variables. An example of such a model is the infinite-range Ising model of statistical physics at inverse temperature $\beta \geq 0$, with a binary prior and a quadratic interaction governed by a coupling matrix J. In this specific model, we have:

$$P_0(\mathbf{x}) = \prod_{i=1}^{N} \left[\frac{1}{2\cosh(\beta h_i)} \left(\delta(x_i - 1) e^{-\beta h_i} + \delta(x_i + 1) e^{\beta h_i} \right) \right], \tag{38}$$

$$P_J(\mathbf{x}) = \exp\left\{ \frac{\beta}{2} \sum_{i,j} J_{ij} x_i x_j \right\}, \tag{39}$$

for some $\{h_i\} \in \mathbb{R}^N$. Our goal is to compute the large N limit of the free entropy $\log Z$ in the model of eq. (37). Each of the two distributions P_0 and P_J allows for tractable computations of physical quantities (like averages), but the difficulty arises when considering their product. The idea behind EC is to simultaneously approximate P_0 and P_J by a tractable family of distributions. For the sake of the presentation we will consider the family of Gaussian probability distributions, although this can be generalized to different families, see the general framework of [OW05a]. We define the first approximation as:

$$\mu_0(\mathbf{x}) \equiv \frac{1}{Z_0(\Gamma_0, \boldsymbol{\lambda}_0)} P_0(\mathbf{x}) e^{-\frac{1}{2}\mathbf{x}^\intercal \Gamma_0 \mathbf{x} + \boldsymbol{\lambda}_0^\intercal \mathbf{x}}. \tag{40}$$

Here, the parameter Γ_0 is a symmetric positive matrix and $\boldsymbol{\lambda}_0$ is a vector. We will denote $\langle \cdot \rangle_0$ the averages with respect to μ_0. We can write the trivial identity:

$$Z = Z \times \frac{Z_0(\Gamma_0, \boldsymbol{\lambda_0})}{Z_0(\Gamma_0, \boldsymbol{\lambda_0})} = Z_0(\Gamma_0, \boldsymbol{\lambda_0}) \left\langle P_J(\mathbf{x})\, e^{\frac{1}{2}\mathbf{x}^\intercal \Gamma_0 \mathbf{x} - \boldsymbol{\lambda}_0^\intercal \mathbf{x}} \right\rangle_0 .$$

The idea of EC is to replace, when one computes the average $\left\langle P_J(\mathbf{x})\, e^{\frac{1}{2}\mathbf{x}^\intercal \Gamma_0 \mathbf{x} - \boldsymbol{\lambda}_0^\intercal \mathbf{x}} \right\rangle_0$, the distribution μ_0 by an approximate Gaussian distribution, that we can write as:

$$\mu_S(\mathbf{x}) \equiv \frac{1}{Z_S} e^{-\frac{1}{2}\mathbf{x}^\intercal (\Gamma_J + \Gamma_0)\mathbf{x} + (\boldsymbol{\lambda_0} + \boldsymbol{\lambda}_J)^\intercal \mathbf{x}}. \tag{41}$$

Performing this replacement yields the *expectation-consistency* approximation to the free entropy:

$$\begin{aligned} \log Z^{\mathrm{EC}}(\Gamma_0, \Gamma_J, \boldsymbol{\lambda}_0, \boldsymbol{\lambda}_J) = \log \left[\int \mathrm{d}\mathbf{x} P_0(\mathbf{x}) e^{-\frac{1}{2}\mathbf{x}^\intercal \Gamma_0 \mathbf{x} + \boldsymbol{\lambda}_0^\intercal \mathbf{x}} \right] + \log \left[\int \mathrm{d}\mathbf{x} P_J(\mathbf{x}) e^{-\frac{1}{2}\mathbf{x}^\intercal \Gamma_J \mathbf{x} + \boldsymbol{\lambda}_J^\intercal \mathbf{x}} \right] \\ - \log \left[\int \mathrm{d}\mathbf{x}\, e^{-\frac{1}{2}\mathbf{x}^\intercal (\Gamma_0 + \Gamma_J)\mathbf{x} + (\boldsymbol{\lambda_0} + \boldsymbol{\lambda}_J)^\intercal \mathbf{x}} \right]. \end{aligned} \tag{42}$$

Note that all three parts of this free entropy are tractable. In order to symmetrize the result we can define a third measure:

$$\mu_J(\mathbf{x}) \equiv \frac{1}{Z_J(\Gamma_J, \boldsymbol{\lambda}_J)} P_J(\mathbf{x}) e^{-\frac{1}{2}\mathbf{x}^\intercal \Gamma_J \mathbf{x} + \boldsymbol{\lambda}_J^\intercal \mathbf{x}}. \tag{43}$$

The final free entropy should not depend on the values of the parameters, so we expect that the best values for $\Gamma_0, \Gamma_J, \boldsymbol{\lambda}_0, \boldsymbol{\lambda}_J$ make Z^{EC} stationary. This is a strong hypothesis, and the reader can refer to [OW05a] for more details and justifications. This yields the *Expectation Consistency* conditions, giving their name to the procedure:

$$\begin{cases} \langle x_i \rangle_0 & = \langle x_i \rangle_J = \langle x_i \rangle_S \,, \\ \langle x_i x_j \rangle_0 & = \langle x_i x_j \rangle_J = \langle x_i x_j \rangle_S \,. \end{cases} \tag{44}$$

3.1.2 Adaptive TAP approximation

The adaptive TAP approximation (or adaTAP) [OW01a, OW01b] provides an equivalent way to derive the free entropy of eq. (42) for models with pairwise interactions. Let us briefly sketch its derivation and the main arguments behind it. We follow the formulation of [HK13] and we consider again the infinite-range Ising model of eq. (38). The extensive Gibbs free entropy $N\Phi = \log Z$ at fixed values of the magnetizations $m_i = \langle x_i \rangle$ and $v_{ij} = \langle x_i x_j \rangle_c$ can be written using Lagrange parameters: a vector $\boldsymbol{\lambda}$ and a symmetric matrix Γ.

$$\Phi(\beta, \boldsymbol{m}, \mathbf{v}) = \mathrm{extr}_{\boldsymbol{\lambda}, \Gamma} \left[-\boldsymbol{\lambda}^\intercal \boldsymbol{m} + \frac{1}{2} \sum_{i,j} \Gamma_{ij} (v_{ij} + m_i m_j) + \log \int \mathrm{d}\mathbf{x}\, P_0(\mathbf{x})\, e^{\frac{\beta}{2}\mathbf{x}^\intercal J \mathbf{x} - \frac{1}{2}\mathbf{x}^\intercal \Gamma \mathbf{x} + \boldsymbol{\lambda}^\intercal \mathbf{x}} \right]. \tag{45}$$

The adaTAP approximation consists in writing:

$$\begin{aligned} N\Phi(\beta, \boldsymbol{m}, \mathbf{v}) &= \Phi(0, \boldsymbol{m}, \mathbf{v}) + \int_0^\beta \mathrm{d}l\, \frac{\partial \Phi(l, \boldsymbol{m}, \mathbf{v})}{\partial l}, \\ &\simeq \Phi(0, \boldsymbol{m}, \mathbf{v}) + \Phi_g(\beta, \boldsymbol{m}, \mathbf{v}) - \Phi_g(0, \boldsymbol{m}, \mathbf{v}). \end{aligned} \tag{46}$$

In this expression, $\Phi_g(\beta, \boldsymbol{m}, \mathbf{v})$ denotes the free entropy of the same system, but where the spins have a Gaussian statistics. The idea behind the adaTAP approximation is as follows. The derivative $\partial_l \Phi(l, \boldsymbol{m}, \mathbf{v}) = \frac{1}{2N} \sum_{ij} J_{ij} \langle x_i x_j \rangle$ is an expectation of a sum over a large number of terms; therefore it is reasonable to assume that this expectation is the same as if the underlying variables were Gaussian. This assumption of adaTAP, although reasonable, is *a priori* hard to justify more rigorously and systematically. It is important to notice that the free entropy (46) of adaTAP is equivalent to the one derived using Expectation Consistency in eq. (42). Indeed, using Lagrange parameters we can write the three terms of eq. (46) as:

$$\begin{aligned} N\Phi^{\mathrm{adaTAP}}(\beta, \boldsymbol{m}, \mathbf{v}) = &\underset{\boldsymbol{\lambda_0}, \Gamma_0}{\mathrm{extr}} \left[\log \left\{ \int \mathrm{d}\mathbf{x} P_0(\mathbf{x})\, e^{-\frac{1}{2}\mathbf{x}^\intercal \Gamma_0 \mathbf{x} + \boldsymbol{\lambda}_0^\intercal \mathbf{x}} \right\} - \boldsymbol{\lambda_0}^\intercal \boldsymbol{m} + \frac{1}{2} \sum_{i,j} (\Gamma_0)_{ij} (v_{ij} + m_i m_j) \right] \\ &+ \underset{\boldsymbol{\lambda_J}, \Gamma_J}{\mathrm{extr}} \left[\log \left\{ \int \mathrm{d}\mathbf{x} P_J(\mathbf{x})\, e^{-\frac{1}{2}\mathbf{x}^\intercal \Gamma_J \mathbf{x} + \boldsymbol{\lambda}_J^\intercal \mathbf{x}} \right\} - \boldsymbol{\lambda_J}^\intercal \boldsymbol{m} + \frac{1}{2} \sum_{i,j} (\Gamma_J)_{ij} (v_{ij} + m_i m_j) \right] \\ &- \underset{\boldsymbol{\lambda_S}, \Gamma_S}{\mathrm{extr}} \left[\log \left\{ \int \mathrm{d}\mathbf{x}\, e^{-\frac{1}{2}\mathbf{x}^\intercal \Gamma_S \mathbf{x} + \boldsymbol{\lambda}_S^\intercal \mathbf{x}} \right\} - \boldsymbol{\lambda_S}^\intercal \boldsymbol{m} + \frac{1}{2} \sum_{i,j} (\Gamma_S)_{ij} (v_{ij} + m_i m_j) \right]. \end{aligned} \tag{47}$$

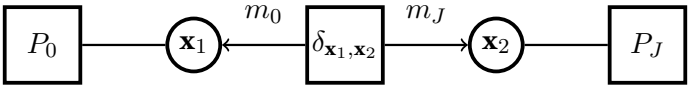


Figure 2: Duplicated factor graph for the VAMP approximation. Circles represent vector nodes and squares factor nodes. We represent the two messages m_0 and m_J in terms of which we can write the full BP equations.

Once written in this form, the extremization over $\boldsymbol{m}$ and $\mathbf{v}$ of the free entropy implies that $\Gamma_S = \Gamma_0 + \Gamma_J$ and $\boldsymbol{\lambda}_S = \boldsymbol{\lambda}_0 + \boldsymbol{\lambda}_J$. It is then clear that we found back $\log Z^{\rm EC}$ of eq. (42).

3.1.3 Vector Approximate Message Passing approximation

The Vector Approximate Message Passing (VAMP) algorithm [RSF17] extends previous message-passing approaches like the GAMP algorithm [Ran11] (that we will describe in more details in Sec. 4.1) to a class of correlated interaction matrices, namely matrices that satisfy a right-rotation invariance property, similarly to Model S and Model R. The algorithm itself can be derived in several ways (see [RSF17]). Here we briefly recall the use of belief-propagation equations on a "duplicated" factor graph and their Gaussian projection. As we shall see, the Bethe free entropy, given as a function of the BP messages, is then equivalent to the expectation-consistency free entropy. For simplicity, we consider again the problem of eq. (38) with a pairwise interaction involving a matrix J following Model S.

The idea behind VAMP is to consider two vector spin variables $\mathbf{x_1}$, with measure P_0 and $\mathbf{x_2}$ with measure P_J, and to impose that they are equal. The partition function can be written using a trivial decomposition:

$$Z \equiv e^{N\Phi} = \int_{\mathbb{R}^N} \mathrm{d}\mathbf{x}_1 \, \mathrm{d}\mathbf{x}_2 \; P_0(\mathbf{x}_1) P_J(\mathbf{x}_2) \; \delta(\mathbf{x}_1 - \mathbf{x}_2). \tag{48}$$

This partition function can be represented as a "duplicated" factor graph involving two vector nodes, see Fig. 2. One then writes the BP equations for this problem using this factor graph representation (the reader who is not familiar with BP equations and factor graph representations can consult [MM09]). The BP equations can be written in terms of two "messages" $m_0(\mathbf{x_2})$ and $m_J(\mathbf{x_1})$. They can be obtained by looking at the stationarity conditions of the following "Bethe free entropy":

$$\Phi_{\rm Bethe} \equiv \log\left(\int \mathrm{d}\mathbf{x} \; P_0(\mathbf{x}) m_J(\mathbf{x})\right) + \log\left(\int \mathrm{d}\mathbf{x} \; P_J(\mathbf{x}) m_0(\mathbf{x})\right) - \log\left(\int \mathrm{d}\mathbf{x} \; m_0(\mathbf{x}) m_J(\mathbf{x})\right). \tag{49}$$

As the factor graph has a tree structure, these BP equations are an exact representation of the original problem, but they are in general intractable. In order to make the computation tractable one can make a Gaussian approximation, which is at the core of the VAMP algorithm: the messages m_0 and m_J on the factor graph of Fig. 2 are assumed to be Gaussian, and thus are only characterized by their mean and their covariance. We can thus write:

$$m_0(\mathbf{x}) \propto e^{-\frac{1}{2}\mathbf{x}^\intercal \Gamma_0 \mathbf{x} + \boldsymbol{\lambda}_0^\intercal \mathbf{x}}, \qquad m_J(\mathbf{x}) \propto e^{-\frac{1}{2}\mathbf{x}^\intercal \Gamma_J \mathbf{x} + \boldsymbol{\lambda}_J^\intercal \mathbf{x}}. \tag{50}$$

Writing the BP update rule with this assumption yields the VAMP algorithm of [RSF17]. In the present case it reads:

$$\begin{cases} \Gamma_0^t = \langle \mathbf{x}\mathbf{x}^T \rangle_{\mu_0,c}^{t-1} - \Gamma_J^{t-1}, & \text{(51a)} \\ \boldsymbol{\lambda}_0^t = (\Gamma_0^t + \Gamma_J^{t-1}) \langle \mathbf{x} \rangle_{\mu_0}^{t-1} - \boldsymbol{\lambda}_J^{t-1}, & \text{(51b)} \\ \Gamma_J^t = \langle \mathbf{x}\mathbf{x}^T \rangle_{\mu_J,c}^{t} - \Gamma_0^t, & \text{(51c)} \\ \boldsymbol{\lambda}_J^t = (\Gamma_0^t + \Gamma_J^t) \langle \mathbf{x} \rangle_{\mu_J}^{t} - \boldsymbol{\lambda}_0^t, & \text{(51d)} \end{cases}$$

where the measures μ_0 and μ_J are respectively

$$\mu_0(\mathbf{x}) \propto P_0(\mathbf{x}) m_J(\mathbf{x}), \qquad \mu_J(\mathbf{x}) \propto P_J(\mathbf{x}) m_0(\mathbf{x}). \tag{52}$$

Plugging the ansatz of eq. (50) into eq. (49) immediately gives back the EC free entropy of eq. (42). And the BP equations of eq. (51) are identical to the Expectation Consistency conditions of eq. (44).

In the end, the three approximation schemes, Expectation Consistency, adaptive TAP and VAMP give the same expression for the free entropy. However, an important advantage of the VAMP approach is that it "naturally" gives an iterative scheme to solve the fixed point equations because it was derived via the belief propagation equations. This iterative scheme turns the fixed point equations into an efficient algorithm, as noticed by [RSF17] and as we show later in Sec. 4.2.

3.2 Plefka expansion for models of symmetric pairwise interactions

3.2.1 A symmetric model with generic priors

In this subsection we consider a generic model of N "spin" variables $\mathbf{x} = \{x_1, \cdots, x_N\} \in \mathbb{R}^N$. They interact via a pairwise interaction and are subject to a possible external field, which is modeled by the following Hamiltonian:

$$H_J(\mathbf{x}) = -\frac{1}{2} \sum_{i,j} J_{ij} x_i x_j + \sum_i h_i x_i. \tag{53}$$

We will consider random symmetric coupling matrices $\{J_{ij}\}$, generated from a rotationally invariant ensemble satisfying Model S. We assume that the variables $\{x_i\}$ have a prior distribution under which they are *independent* variables, each with a distribution P_i. For instance, this includes Ising (binary) spins by choosing $P_i = \frac{1}{2}(\delta_1 + \delta_{-1})$. At a given inverse temperature $\beta \geq 0$ and a fixed realization of the coupling matrix J we define the Gibbs-Boltzmann distribution of the spins, and the partition function, as:

$$P_{\beta,J}(\mathrm{d}\mathbf{x}) \equiv \frac{1}{Z_J(\beta)} \prod_i P_i(\mathrm{d}x_i) \, \exp\left\{ \frac{\beta}{2} \sum_{i,j} J_{ij} x_i x_j - \beta \sum_i h_i x_i \right\}, \tag{54}$$

$$Z_J(\beta) = \int \prod_i P_i(\mathrm{d}x_i) \, \exp\left\{ \frac{\beta}{2} \sum_{i,j} J_{ij} x_i x_j - \beta \sum_i h_i x_i \right\}. \tag{55}$$

We will compute the large N limit of the free entropy $\Phi_J(\beta) \equiv \frac{1}{N} \log Z_J(\beta)$ at fixed values of the magnetizations $m_i = \langle x_i \rangle$ and variances $v_i = \langle (x_i - m_i)^2 \rangle$ using the Plefka expansion. We will fix these variables using Lagrange multipliers $\{\lambda_i\}$ for the $\{m_i\}$ variables, and $\{\gamma_i\}$ for the $\{v_i\}$. The goal of this section is to show how, and under which assumptions, the calculation of Sec. 2.1.2 can be generalized in this context. Clearly the zeroth order term is different from the spherical case and it is given by:

$$\Phi_J(\beta = 0) = \frac{1}{N} \sum_i \lambda_i m_i + \frac{1}{2N} \sum_i \gamma_i (v_i + m_i^2) + \frac{1}{N} \log \int \prod_i P_i(\mathrm{d}x_i) \, e^{-\frac{1}{2}\sum_i \gamma_i x_i^2 - \sum_i \lambda_i x_i}. \tag{56}$$

As we underlined in Sec. 2.1.2, in the Georges-Yedidia method the Lagrange parameters are always considered at $\beta = 0$. At order 1 in β we obtain at leading order:

$$\left(\frac{\partial \Phi_J}{\partial \beta} \right)_{\beta=0} = -\frac{1}{N} \langle H_J \rangle_0 = \frac{1}{2N} \sum_{i,j} J_{ij} m_i m_j + \frac{1}{2N} \sum_i J_{ii} v_i - \frac{1}{N} \sum_i h_i m_i. \tag{57}$$

The operator U of eq. (16), defined in [GY91] and taken at $\beta = 0$, is thus exactly the same as the one of eq. (21). Using this remark we can see that many of the results obtained in Sec. 2.1.2 will apply to the present case. For instance the second order term is identical and given in eq. (22). We then conjecture, backed by our diagrammatic results in Sec. 5, that the higher order terms are different from the spherical model only in terms which are sub-leading in N. For instance at third order we obtain:

$$\frac{1}{3!}\left(\frac{\partial^3 \Phi_J}{\partial \beta^3}\right)_{\beta=0} = -\frac{1}{6N}\langle U^3\rangle_0 = \frac{1}{6N}\sum_{\substack{i_1,i_2,i_3 \\ \text{pairwise distincts}}} J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_1} v_{i_1} v_{i_2} v_{i_3} + \frac{1}{6N}\sum_{i\neq j} J_{ij}^3 \kappa_i^{(3)} \kappa_j^{(3)}. \tag{58}$$

In this equation, we denoted $\kappa_i^{(p)}$ the *cumulant* of order p of the distribution of x_i at $\beta = 0$. Note that the rotation invariance of Model S implies that if $i \neq j$, typically $J_{ij} \sim \frac{1}{\sqrt{N}}$. Therefore a term like $\sum_{i\neq j} J_{ij}^3$ gives a negligible contribution to the free entropy. We shall therefore assume that the second part of the RHS of eq. (58) is negligible as $N \to \infty$. This is correct provided that the possible correlations of the third order cumulants $\kappa_i^{(3)}$ with the matrix J do not change the scaling of this term sufficiently to make it thermodynamically relevant (see Sec. 5.4 for more details on this particular point). The first term corresponds to a simple cycle of order 3 and is the same term that appeared in Sec. 2.1.2.

Higher orders We can carry on the computation of the derivatives $\frac{\partial^p}{\partial \beta^p}\Phi_J(\beta = 0)$. We explain in more details in Sec. 5.4 under which precise results and assumptions we can leverage the results of Sec. 2.1.2, summarized in eq. (25), to conjecture the following value of the free entropy at all orders of perturbation and at leading order in N:

$$\Phi_J(\beta) = \Phi_J(0) + \frac{\beta}{2N}\sum_{i,j} J_{ij} m_i m_j - \frac{\beta}{N}\sum_i h_i m_i + \frac{1}{N}\sum_{p=1}^{\infty}\frac{\beta^p}{2p}\sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distincts}}} J_{i_1 i_2}\cdots J_{i_p i_1}\prod_{\alpha=1}^{p} v_{i_\alpha}. \tag{59}$$

Homogeneous variances For the remainder of this section we assume that the maximum of the free entropy of eq. (59) is attained for variables $\{v_i\}$ such that $v_i = v$. This hypothesis can be argued as reasonable for many models, but we postpone this argumentation to the applications of eq. (59) to specific models. We obtain a resummation of the Plefka free entropy using the correspondence of eq. (28):

$$\Phi_J(\beta) = \Phi_J(0) + \frac{\beta}{2N}\sum_{i,j} J_{ij} m_i m_j - \frac{\beta}{N}\sum_i h_i m_i + \frac{1}{2}\int_0^{\beta v} \mathcal{R}_{\rho_D}(u)\,\mathrm{d}u. \tag{60}$$

Recall finally that $\Phi_J(0)$ is given by eq. (56). We were able to perform this expansion and its resummation almost only by applying our results on the spherical models of Sec. 2. The study of the large-N behavior of diagrams made out of matrix elements of J, performed in Sec. 5, proves to be of crucial importance both in the expansion and its resummation. As discussed in Sec. 2.1.2, we expect this Plefka expansion of the free entropy to hold for $\beta < \beta_c$, in which $\beta_c \equiv -v^{-1}\,\mathcal{S}_{\rho_D}(\lambda_{\max})$.

3.2.2 Connection of the Plefka expansion to EC approximations

Although not obvious at first, the result of the Plefka expansion in eq. (59) provides a systematic and precise analysis of asymptotic exactness for rotationally invariant models of the Expectation Consistency approximations. The more straightforward way to see how the Plefka expansion relates to these approximations is to start from the adaTAP approximation, see eq. (46). In the language of the Plefka expansion, adaTAP amounts to assuming that at every order $p \geq 1$ of perturbation in β, one can perform the calculation *as if* the statistics of the variables were Gaussian. An equivalent formulation is that all the terms of order $p \geq 1$ in the low-β

expansion of the free entropy should be the same for the model with a generic prior of Sec. 3.2.1 and for the spherical model of Sec. 2.1. This statement, which generalizes the Parisi-Potters result of [PP95], is exactly what we argued in the calculation of Sec. 3.2.1, using the diagrammatic analysis of Sec. 5. Therefore, the diagrammatic analysis 'à la Plefka' provides a clear meaning to the EC approximations: the class of diagrams that are neglected in these approximations are explicited in Sec. 5. We believe that these diagrams are actually negligible in the large N limit, so that the EC approximations are actually exact asymptotically for rotationally invariant models, in the high temperature phase as captured by the resummation of the Plefka expansion, which we summarized in Conjecture 1. We believe that this asymptotic exactness extends beyond the high temperature phase to any model in the replica symmetric phase. The diagrammatic analysis provides a route to proving this statement rigorously.

3.2.3 Application to the Hopfield model

In the Hopfield model [Hop82] we consider binary spins $\mathbf{x} \in \{\pm 1\}^N$ and the coupling matrix J is constructed out of P *patterns*, which are spin configurations $\xi^l \in \{\pm 1\}^N$, for $l \in \{1, \cdots, P\}$. The coupling constants are defined as:

$$\begin{cases} J_{ij} &= \frac{1}{N} \sum_{l=1}^{P} \xi_i^l \xi_j^l \qquad (i \neq j), \\ J_{ii} &= 0, \end{cases} \tag{61}$$

and we assume that the $\{\xi_i^l\}$ are i.i.d. variables with equal probability in $\{\pm 1\}$, so that $\mathbb{E} J_{ij} = 0$ and $\mathbb{E} J_{ij}^2 = P/N^2$. We study this system in the limit in which both $P, N \to \infty$ with a fixed ratio $P/N \to \alpha$. The derivation of the TAP free energy for these models has been performed in [NT97, Méz17] via the Plefka expansion, and via the cavity method in [MPV87]. If the random matrix ensemble of eq. (61) is *a priori* not rotationally invariant, one can show that since the variables $\{\xi_i^l\}$ are i.i.d., only the first and second moment of their distributions will contribute to the thermodynamic limit of the free entropy, so that we can assume that they are actually standard centered Gaussian variables without changing the free entropy. This is strengthened by the classical results of [MP67], who only need to consider i.i.d. variables $\{\xi_\mu^l\}$ to obtain that the spectral law of the covariance matrix written in eq. (61) converges weakly to the celebrated Marchenko-Pastur distribution.

The ensemble of eq. (61) is thus for our purposes essentially a Wishart matrix model in which the diagonal has been removed. It is then a known result of random matrix theory (see for instance [TV04]) that its asymptotic $\mathcal{R}$-transform reads:

$$\mathcal{R}_J(x) = \frac{\alpha}{1-x} - \alpha = \alpha \frac{x}{1-x}. \tag{62}$$

The term $-\alpha$ in the first equality accounts for the "removal" of the diagonal of the Wishart matrix. Let us apply eq. (59) and eq. (60) for this model. At $\beta = 0$, the free entropy is given by eq. (56). For an Ising prior on the spins, it reads:

$$\Phi_J(\beta = 0) = -\frac{1}{N} \sum_i \left[\frac{1+m_i}{2} \ln \frac{1+m_i}{2} + \frac{1-m_i}{2} \ln \frac{1-m_i}{2} \right]. \tag{63}$$

Note that because $x_i^2 = 1$ the variances v_i of the variables x_i are fixed by the magnetizations m_i by the relation $v_i = 1 - m_i^2$. Eq. (59) becomes:

$$\Phi_J(\beta) = \Phi_J(0) + \frac{\beta}{2} \sum_{i,j} J_{ij} m_i m_j + \frac{1}{N} \sum_{p=1}^{\infty} \frac{\beta^p}{2p} \sum_{\substack{i_1, \cdots, i_p \\ \text{pairwise distincts}}} J_{i_1 i_2} \cdots J_{i_p i_1} \prod_{\alpha=1}^{p} (1 - m_{i_\alpha}^2) + \mathcal{O}_N(1). \tag{64}$$

Substituting the squared means m_i^2 by the spin glass order parameter $q \equiv (1/N)\sum_i m_i^2$ in eq. (64) and using the resummed form of eq. (60), we obtain:

$$\Phi_J(\beta) = \Phi_J(0) + \frac{\beta}{2}\sum_{i,j} J_{ij} m_i m_j + \frac{1}{2}\int_0^{\beta(1-q)} \mathcal{R}_{\rho_D}(u)\,\mathrm{d}u + \mathcal{O}_N(1). \tag{65}$$

Starting from eq. (65) and eq. (62), we reach the final form for the free entropy:

$$\begin{aligned}\Phi_J(\beta) = -\frac{1}{N}\sum_i \left[\frac{1+m_i}{2}\ln\frac{1+m_i}{2} + \frac{1-m_i}{2}\ln\frac{1-m_i}{2}\right] + \frac{\beta}{2}\sum_{i,j} J_{ij} m_i m_j \\ - \frac{\alpha\beta(1-q)}{2} - \frac{\alpha}{2}\ln\left[1-\beta(1-q)\right] + \mathcal{O}_N(1).\end{aligned} \tag{66}$$

Maximizing it over the continuous set of magnetizations $\{m_i\}$ yields the TAP equations:

$$\beta^{-1}\tanh^{-1}(m_i) = \sum_{j(\neq i)} J_{ij} m_j - \alpha\beta\frac{1-q}{1-\beta(1-q)} m_i. \tag{67}$$

This is in agreement with the findings of [MPV87, NT97, Méz17]. However, our framework and results allowed us to treat this kind of model in a very generic way.

3.3 The Replica approach

Interestingly, the same approaches can be used to study the partition function and derive the free entropy of rotationally invariant models using the replica approach. The essence of the replica method is well known [MPV87]: one studies the n-th moment of the partition function Z by introducing n 'replicas' of each original variable, and one then studies the average of Z^n in the limit $n \to 0$, which allows to reconstruct the average of $\log Z$, and therefore the free entropy. We shall illustrate in this section how the three approximation schemes (EC, adaTAP and VAMP) can be used to derive the replica free entropy, and how the high-temperature Plefka expansion justifies these approximations.

Let us consider again the generic model of eq. (53), with $h_i = 0$ for simplicity. The n-th moment of the partition function reads (a, b indices always denote replica indices running from 1 to n):

$$Z_J^n(\beta) = \left[\int \prod_a \prod_i P_i(\mathrm{d}x_i^a)\right] \exp\left\{\frac{\beta}{2}\sum_a \sum_{i,j} J_{ij} x_i^a x_j^a\right\}. \tag{68}$$

This "replicated partition function" should then be averaged over the disorder. Here, we deal with a J matrix generated from Model S, and the disorder average means an average over the orthogonal matrix O in the decomposition $J = ODO^T$ (keeping the eigenvalues D fixed). Let us see how it can be analyzed using the four (equivalent) approximation schemes. Interestingly the average over O is in fact not needed in these approaches, as they show that the repliacted partition function is "self-averaging", meaning that it gives the same free entropy density for almost all realizations of J. The order parameter we will fixe here is the $n \times n$ symmetric matrix Q with elements $Q_{ab} = \frac{1}{N}\sum_i \langle x_i^a x_i^b \rangle_c$, called the *overlap* matrix in the statistical physics language.

3.3.1 Expectation-Consistency approximation

As we have seen in Sec. 3.1.1, in the Expectation-Consistency (EC) approximation, we decompose the replicated free entropy $\Phi = \frac{1}{N}\log Z_J^n(\beta)$ as a function of three auxiliary free entropies:

$$\Phi = \underset{\Gamma_0,\Gamma_J}{\text{extr}} \left\{ \frac{1}{N}\log\left[\int \prod_{a,i} P_i(\mathrm{d}x_i^a) e^{-\frac{1}{2}\sum_i \mathbf{x}_i^\intercal \Gamma_0 \mathbf{x}_i}\right] + \frac{1}{N}\log\left[\int \prod_{a,i} \mathrm{d}x_i^a \, e^{\frac{\beta}{2}\sum_{i,j,a} J_{ij} x_i^a x_j^a - \frac{1}{2}\sum_i \mathbf{x}_i^\intercal \Gamma_J \mathbf{x}_i}\right] \right.$$
$$\left. - \frac{1}{N}\log\left[\int \prod_{a,i} \mathrm{d}x_i^a e^{-\frac{1}{2}\sum_i \mathbf{x}_i^\intercal (\Gamma_0+\Gamma_J)\mathbf{x}_i}\right] \right\}. \tag{69}$$

The Γ_0, Γ_J matrices are symmetric $n \times n$ matrices, and $\mathbf{x}_i = (x_i^a)_{a=1}^n$. Indeed, as we will only fix the Q_{ab}, we expect the first moments to vanish and the second moments to be uniform in space (but not in replica space). The extremization over Γ_0, Γ_J yields the Expectation-Consistency equations, as described in eq. (44). In particular, the average $\frac{1}{N}\sum_i \langle x_i^a x_i^b \rangle$ is constrained to be the same under the three different measures appearing in eq. (69). As we want to impose this average to be Q_{ab}, we can introduce a Lagrange parameter Γ_S to fix this average for instance in the third part of the r.h.s. of eq. (69). This third term becomes

$$-\underset{\Gamma_S}{\text{extr}} \left\{ \frac{1}{2}\mathrm{Tr}\,(\Gamma_S Q) + \frac{1}{N}\log\left[\int \prod_{a,i} \mathrm{d}x_i^a e^{-\frac{1}{2}\sum_i \mathbf{x}_i^\intercal (\Gamma_0+\Gamma_J+\Gamma_S)\mathbf{x}_i}\right] \right\}.$$

Changing $\Gamma_S \to \Gamma_S + \Gamma_0 + \Gamma_J$ gives:

$$\frac{1}{2}\mathrm{Tr}\,[Q(\Gamma_0+\Gamma_J)] - \underset{\Gamma_S}{\text{extr}} \left\{ \frac{1}{2}\mathrm{Tr}\,[\Gamma_S Q] + \frac{1}{N}\log\left[\int \prod_{a,i} \mathrm{d}x_i^a e^{-\frac{1}{2}\sum_i \mathbf{x}_i^\intercal \Gamma_S \mathbf{x}_i}\right] \right\}$$
$$= \frac{1}{2}\mathrm{Tr}\,[Q(\Gamma_0+\Gamma_J)] - \frac{n}{2}(1+\log 2\pi) - \frac{1}{2}\mathrm{Tr}\log Q.$$

In the end, performing explicitely the Gaussian integration in the basis of eigenvectors of J in eq. (69), we obtain the free entropy at fixed overlap Q (recall that ρ_D is the asymptotic eigenvalue density of J):

$$\Phi(Q) = \underset{\Gamma_0}{\text{extr}} \left\{ \frac{1}{2}\mathrm{Tr}(\Gamma_0 Q) + \frac{1}{N}\sum_i \log\left[\int \prod_a P_i(\mathrm{d}x^a) e^{-\frac{1}{2}\mathbf{x}^\intercal \Gamma_0 \mathbf{x}}\right] \right\}$$
$$+ \underset{\Gamma_J}{\text{extr}} \left\{ \frac{1}{2}\mathrm{Tr}(\Gamma_J Q) - \frac{1}{2}\int \rho_D(\mathrm{d}\lambda)\mathrm{Tr}\log(\Gamma_J - \beta\lambda \mathrm{I}_n) \right\} - \frac{1}{2}\mathrm{Tr}\log Q - \frac{n}{2}. \tag{70}$$

The total free entropy is obtained simply as $\underset{Q}{\text{extr}}\, \Phi(Q)$. Assuming an ultrametric structure in replica space [MPV87] (for instance in the replica symmetric case the matrix elements of Q take only two values, one on the diagonal and one out of the diagonal, and the same holds for Λ and R), one can perform an explicit computation. For random orthogonal models, this expression was first derived using Plefka expansion in [MPR94a, MPR94b], and written in exactly the same form as ours in [CDL03].

3.3.2 The adaTAP approximation

Let us now describe the adaTAP approach to this problem. We introduce Lagrange parameters Λ_{ab} which act as external fields, fixing the values of the order parameter Q_{ab}. The matrix Λ is a $n \times n$ symmetric matrix.

The free entropy at fixed overlap matrix $\Phi(Q)$ is expressed as:

$$\Phi(Q) = \underset{\Lambda}{\text{extr}} \left[\frac{1}{2} \sum_{a,b} \Lambda_{ab} Q_{ab} + \frac{1}{N} \log \left[\int \prod_{a,i} P_i(\mathrm{d}x_i^a) \right] \exp \left\{ \frac{\beta}{2} \sum_{a,i,j} J_{ij} x_i^a x_j^a - \frac{1}{2} \sum_{a,b,i} \Lambda_{ab} x_i^a x_i^b \right\} \right]. \quad (71)$$

The total free entropy will thus be obtained as $\underset{Q}{\text{extr}}\, \Phi(Q)$. We use the adaTAP approximation, which gives:

$$\Phi(Q) = \Phi(Q, \beta = 0) + \Phi_G(Q, \beta) - \Phi_G(Q, \beta = 0), \quad (72)$$

where Φ_G is the Gibbs free energy for a model with Gaussian spins, at fixed overlap between spins. The three pieces of Φ can be written as:

$$\Phi(Q, \beta = 0) = \underset{\Lambda}{\text{extr}} \left[\frac{1}{2} \mathrm{Tr}\,(\Lambda Q) + \frac{1}{N} \sum_i \log \left(\int \prod_a P_i(dx^a) e^{-\frac{1}{2} \sum_{a,b} \Lambda_{ab} x^a x^b} \right) \right], \quad (73)$$

$$\Phi_G(Q, \beta) = \underset{\Lambda}{\text{extr}} \left[\frac{1}{2} \mathrm{Tr}\,(\Lambda Q) + \frac{1}{N} \sum_{\lambda \in \mathrm{Sp}(J)} \log \int \prod_a \mathrm{d}x^a e^{-\frac{1}{2} \sum_{a,b} \Lambda_{ab} x^a x^b + \frac{\beta}{2} \lambda \sum_a (x^a)^2} \right], \quad (74)$$

$$\Phi_G(Q, \beta = 0) = \underset{\Lambda}{\text{extr}} \left[\frac{1}{2} \mathrm{Tr}\,(\Lambda Q) + \log \int \prod_a \mathrm{d}x^a e^{-\frac{1}{2} \sum_{a,b} \Lambda_{ab} x^a x^b} \right]. \quad (75)$$

In $\Phi_G(Q, \beta)$ we carry the integral over s_i^a in an (orthonormal) basis of eigenvectors of $\{J_{ij}\}$. $\Phi_G(Q, \beta = 0)$ is easily evaluated: the extremization gives $\Lambda = Q^{-1}$, and therefore

$$\Phi_G(Q, \beta = 0) = n \frac{1 + \log 2\pi}{2} + \frac{1}{2} \mathrm{Tr} \log Q. \quad (76)$$

We change notation and denote by R the matrix Λ that appears in $\Phi_G(Q, \beta)$. This gives finally:

$$\Phi(Q) = \underset{\Lambda, R}{\text{extr}}\, \Phi(Q, \Lambda, R), \quad (77)$$

with

$$\begin{aligned} \Phi(Q, \Lambda, R) = {} & \frac{1}{2} \mathrm{Tr}(\Lambda Q) + \frac{1}{N} \sum_i \log \left(\int \prod_a P_i(dx^a) e^{-\frac{1}{2} \sum_{a,b} \Lambda_{ab} x^a x^b} \right) \\ & + \frac{1}{2} \mathrm{Tr}(RQ) - \frac{1}{2} \int \rho_D(\mathrm{d}\lambda) \mathrm{Tr} \log(R - \beta \lambda \mathrm{I}_n) - \frac{1}{2} \mathrm{Tr} \log Q - \frac{n}{2}. \end{aligned} \quad (78)$$

We found back exactly the EC free entropy of eq. (70).

3.3.3 The VAMP approach

Following eq. (48), the replicated partition function can be written as a "duplicated" integral:

$$Z_J^n(\beta) = \int_{\mathbb{R}^{nN} \times \mathbb{R}^{nN}} P_0(\mathrm{d}\mathbf{x}) \mathrm{P}_\mathrm{J}(\mathrm{d}\mathbf{y}) \delta(\mathbf{x} - \mathbf{y}), \quad (79)$$

where

$$P_0(\mathrm{d}\mathbf{x}) = \prod_i P_i(dx_i^a), \tag{80}$$

$$P_J(\mathrm{d}\mathbf{y}) = \exp\left\{\frac{\beta}{2}\sum_a\sum_{i,j} J_{ij}y_i^a y_j^a\right\}\mathrm{d}\mathbf{y}. \tag{81}$$

We now write the Bethe free entropy Φ_{Bethe} as in eq. (49), and project it onto the space of Gaussian messages which are assumed to be proportional to identity in space, but with an arbitrary replica structure:

$$m_0(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\sum_{a,b} A^0_{ab}\sum_i x_i^a x_i^b\right), \tag{82}$$

$$m_J(\mathbf{x}) \propto \exp\left(-\frac{1}{2}\sum_{a,b} A^J_{ab}\sum_i x_i^a x_i^b\right). \tag{83}$$

We did not include first moments because we expect all the first moments to vanish in the replicated system. We can now write the stationarity equations of Φ_{Bethe} with respect to A^0 and A^J:

$$\left([A^0+A^J]^{-1}\right)_{ab} = -2\frac{\partial}{\partial A^J_{ab}}\frac{1}{N}\sum_i \log\left(\int\prod_a P_i(dx^a)\exp\left(-\frac{1}{2}\sum_{a,b}A^J_{ab}x^a x^b\right)\right), \tag{84}$$

$$\left([A^0+A^J]^{-1}\right)_{ab} = \int \rho_D(\mathrm{d}\lambda)\left([A^0-\beta\lambda]^{-1}\right)_{ab}. \tag{85}$$

It is easy to derive the replicated free entropy Φ. One finds

$$\Phi = \underset{Q,A^0,A^J}{\mathrm{extr}}\,[\Phi(Q,A^0,A^J)], \tag{86}$$

where

$$\Phi(Q,A^0,A^J) = \frac{1}{N}\sum_i \log\left[\int\prod_a P_i(dx^a)\exp\left(-\frac{1}{2}\sum_{a,b}A^J_{ab}x^a x^b\right)\right] - \frac{1}{2}\int\rho_D(\mathrm{d}\lambda)\mathrm{Tr}\log(A^0-\beta\lambda\mathrm{I}_n)$$
$$-\frac{1}{2}\mathrm{Tr}\log(Q) + \frac{1}{2}\mathrm{Tr}(A^J Q) + \frac{1}{2}\mathrm{Tr}(A^0 Q) - \frac{n}{2}. \tag{87}$$

This is again exactly the same result as the replicated free entropy found with EC or adaTAP in the previous sections.

3.3.4 Plefka expansion

As before, we fix only the overlap Q_{ab}, via Lagrange parameters γ^{ab}. Let us denote $\Phi(Q,\beta)$ the corresponding free entropy. At $\beta = 0$ the replicas are not coupled so that we obtain

$$\Phi(Q,\beta=0) = \frac{1}{2}\sum_{a,b}\gamma^{ab}Q_{ab} + \frac{1}{N}\sum_i \log\int\prod_{a=1}^n P_i(\mathrm{d}x^a)\, e^{-\frac{1}{2}\sum_{a,b}\gamma^{ab}x^a x^b}. \tag{88}$$

One can then compute the order 1 perturbation and the U operator of Georges-Yedidia:

$$\left(\frac{\partial \Phi}{\partial \beta}\right)_{\beta=0} = \frac{1}{2N}\sum_{a,i} J_{ii} q_{aa}, \tag{89}$$

$$U(\beta=0) = -\frac{1}{2}\sum_{a}\sum_{i\neq j} J_{ij} x_i^a x_j^a. \tag{90}$$

Note that at $\beta = 0$ we have $\langle x_i^a \rangle_0 = 0$. We obtain the order 2 correction in the same way as before:

$$\frac{1}{2}\left(\frac{\partial^2 \Phi}{\partial \beta^2}\right)_{\beta=0} = \frac{1}{8N}\sum_{i\neq j} J_{ij}^2 \mathrm{Tr}\,[Q^2]. \tag{91}$$

Here the trace is taken in the replica space. One can continue the Plefka expansion at any order, and one obtains very similar results to the non-replicated free entropy, simply the product of variances is replaced by traces of the overlap matrix Q. Indeed, the diagrams constructed from the matrix indices $\{J_{ij}\}$ that appear in the replicated free entropy are exactly the same as in the non-replicated case, so that all the results of Sec. 5 stay valid for this replicated calculation. In the end, the resummation yields the *single-graph* replicated free entropy as a function of the overlap matrix Q:

$$\Phi(Q,\beta) = \Phi(Q,\beta=0) + \sum_{p=1}^{\infty} \frac{\beta^p}{2p} c_p(\rho_D) \mathrm{Tr}\,[Q^p]\ . \tag{92}$$

The series in the second part of this equation is nothing but $\mathrm{Tr}\,[G_{\rho_D}(Q)]$ (see Appendix C for the definition of this function). Recalling the expression (88) for the first, $\beta = 0$, piece, we get finally:

$$\Phi(Q,\beta) = \underset{\gamma}{\mathrm{extr}} \left\{ \frac{1}{2}\mathrm{Tr}(\gamma Q) + \frac{1}{N}\sum_i \log \int \prod_{a=1}^{n} P_i(\mathrm{d}x^a)\, e^{-\frac{1}{2}\sum_{a,b}\gamma^{ab}x^a x^b} + \mathrm{Tr}\,[G_{\rho_D}(Q)] \right\}, \tag{93}$$

and recall that the $\beta = 0$ term is given by eq. (88) and that the G_{ρ_D} function is the integrated $\mathcal{R}$-transform defined in Appendix C, which verifies:

$$G_{\rho_D}(z) = \underset{x}{\mathrm{extr}} \left[\frac{zx}{2} - \frac{1}{2}\int \rho_D(\mathrm{d}\lambda) \log(\lambda - x) \right]. \tag{94}$$

Thus we recognize once again in eq. (93) the expression obtained via EC and adaTAP, see eq. (70).

3.4 Plefka expansion for models of bipartite pairwise interactions

3.4.1 A bipartite model with generic priors

We consider here another generic model, which is an extension of the bipartite model studied in Sec. 2.2, the difference being that we assume a generic prior on the variables rather than a spherical constraint. As we will see, this model is closely related to the symmetric model of Sec. 3.2.1. Let $M, N \geq 1$. We consider two types of variables: a vector $\mathbf{x} \in \mathbb{R}^N$ and a vector $\mathbf{h} \in \mathbb{R}^M$. These two fields are assumed to follow prior distributions under which they are independent and the distribution of all their components decouples. For instance, the prior P_X on $\mathbf{x}$ can be written as $P_X(\mathrm{d}\mathbf{x}) = \prod_i P_i(\mathrm{d}x_i)$. The two fields $\mathbf{h}, \mathbf{x}$ interact via the following Hamiltonian:

$$H_F(\mathbf{h},\mathbf{x}) = -\sum_{\mu,i} F_{\mu i} h_\mu x_i. \tag{95}$$

As in Sec. 2.2, Greek indices μ, ν will always run from 1 to M while Latin indices i, j run from 1 to N. We assume that the coupling matrix $F \in \mathbb{R}^{M\times N}$ satisfies the rotation invariance property described in Model R. For a fixed $\beta \geq 0$ and a realization of F we define the Gibbs-Boltzmann distribution and the partition function:

$$P_{\beta,F}(\mathbf{h},\mathbf{x}) \equiv \frac{1}{Z_{\beta,F}} \prod_{\mu} P_\mu(\mathrm{d}h_\mu) \prod_i P_i(\mathrm{d}x_i) \exp\left\{\beta \sum_{\mu,i} F_{\mu i} h_\mu x_i\right\}, \tag{96}$$

$$Z_{\beta,F} = \int \prod_{\mu} P_\mu(\mathrm{d}h_\mu) \prod_i P_i(\mathrm{d}x_i) \exp\left\{\beta \sum_{\mu,i} F_{\mu i} h_\mu x_i\right\}. \tag{97}$$

As in the previous section we will compute the large N limit of the free entropy $\Phi_F(\beta) \equiv \frac{1}{N}\log Z_{\beta,F}$. We look at this problem in the thermodynamic limit, with $N, M \to \infty$ and a fixed ratio $M/N \to \alpha > 0$. We constraint the first and second moments of $\{x_i\}$ and $\{h_\mu\}$ under the Gibbs measure to be $\langle x_i\rangle = m_i^x$, $\langle h_\mu\rangle = m_\mu^h$, $\langle (x_i - m_i^x)^2\rangle = v_i^x$, $\langle (h_\mu - m_\mu^h)^2\rangle = v_\mu^h$. The Lagrange multipliers introduced to enforce these conditions will be denoted λ_i^x, λ_μ^h for the first moments and γ_μ^x, γ_μ^h for the second moments. At order 0, we obtain:

$$\begin{aligned}\Phi_F(\beta=0) = &\frac{1}{N}\sum_i \left[\lambda_i^x m_i^x + \frac{1}{2}\gamma_i^x (v_i^x + (m_i^x)^2)\right] + \frac{1}{N}\sum_\mu \left[\lambda_\mu^h m_\mu^h + \frac{1}{2}\gamma_\mu^h (v_\mu^h + (m_\mu^h)^2)\right] \\ &+ \frac{1}{N}\log\left[\int \prod_\mu P_\mu(\mathrm{d}h_\mu) \prod_i P_i(\mathrm{d}x_i)\, e^{-\sum_i \left[\lambda_i^x x_i + \frac{1}{2}\gamma_i^x x_i^2\right] - \sum_\mu \left[\lambda_\mu^h h_\mu + \frac{1}{2}\gamma_\mu^h h_\mu^2\right]}\right].\end{aligned} \tag{98}$$

The calculations at order 1 and 2 are very similar to the ones of the spherical model of Sec. 2.2.2. One can also refer to the symmetric case of Sec. 3.2.1. We obtain the first four orders:

$$\left(\frac{\partial \Phi_F}{\partial \beta}\right)_{\beta=0} = \frac{1}{N}\sum_{\mu,i} F_{\mu i} m_\mu^h m_i^x, \tag{99}$$

$$\frac{1}{2}\left(\frac{\partial^2 \Phi_F}{\partial \beta^2}\right)_{\beta=0} = \frac{1}{2N}\sum_{\mu,i} F_{\mu i}^2 v_\mu^h v_i^x, \tag{100}$$

$$\frac{1}{3!}\left(\frac{\partial^3 \Phi_F}{\partial \beta^3}\right)_{\beta=0} = \frac{1}{6N}\sum_{\mu,i} F_{\mu i}^3 \kappa_\mu^{(3,h)} \kappa_i^{(3,x)}, \tag{101}$$

$$\frac{1}{4!}\left(\frac{\partial^4 \Phi_F}{\partial \beta^4}\right)_{\beta=0} = \frac{1}{8N}\sum_{\mu_1\neq\mu_2}\sum_{i_1\neq i_2} F_{\mu_1 i_1} F_{\mu_1 i_2} F_{\mu_2 i_2} F_{\mu_2 i_1} v_{\mu_1}^h v_{\mu_2}^h v_{i_1}^x v_{i_2}^x + \mathcal{O}_N(1). \tag{102}$$

Recall that $\kappa_i^{(p,x)}$ is the p-th order cumulant of x_i at $\beta = 0$, and we define $\kappa_\mu^{(p,h)}$ in the same way for h_μ.

On higher orders The discussion on the higher orders in perturbation is very similar to the one we made in Sec. 3.2.1. In the end, we only retain the simple cycles made of matrix elements $\{F_{\mu i}\}$ in our Plefka expansion. As we explain in more details in Sec. 5, and in particular in Sec. 5.5 for the bipartite case, we make two crucial statements to obtain this result:

- Let us forget for the moment about the factors of variances or higher-order moments of the distributions of x_i, h_μ at $\beta = 0$. We can then study all the possible terms appearing in the Plefka expansion at every order as *diagrams* made of matrix elements $\{F_{\mu i}\}$, and show that the only non-vanishing diagrams are the *simple cycles*. This is shown in more details in Sec. 5.5.
- We assume that the factors arising from the variances v_i^x, v_μ^h or higher-order cumulants of the variables x_i and h^μ do not change significantly the scaling of the diagrams, that is we can still only retain the

simple cycles in the thermodynamic limit. More details are given in Sec. 5.5.2.

For instance, at order 3 these statements yield $\partial_\beta^3 \Phi_{N,F} = \mathcal{O}_N(1)$ since one can not construct a simple cycle for bipartite models at order 3. This also explains the $\mathcal{O}_N(1)$ term in the order 4, see eq. (102). In the end, we obtain the following value of the free entropy at leading order in N:

$$\Phi_F(\beta) = \Phi_F(0) + \frac{\beta}{N}\sum_{\mu,i} F_{\mu i} m_\mu^h m_i^x + \frac{1}{N}\sum_{p=1}^{\infty} \frac{\beta^{2p}}{2p} \sum_{\substack{\mu_1,\cdots,\mu_p \\ i_1,\cdots,i_p}} F_{\mu_1 i_1} F_{\mu_1 i_2} \cdots F_{\mu_p i_p} F_{\mu_p i_1} \prod_{\alpha=1}^{p} v_{\mu_\alpha}^h v_{i_\alpha}^x . \tag{103}$$

In the summation all indices $\mu_1, \cdots, \mu_p$ are pairwise distinct, and so are $i_1, \cdots, i_p$.

Homogeneous variances Let us assume that the maximum of the free entropy of eq. (103) is attained for variables $\{v_\mu^h, v_i^x\}$ such that $v_\mu^h = v^h$ and $v_i^x = v^x$. As in the symmetric case this hypothesis can be justified for many models that we will analyze later on. Using eq. (36) the free entropy of eq. (103) can be resummed:

$$\begin{aligned}\Phi_F(\beta) =& \Phi_F(0) + \frac{\beta}{N}\sum_{\mu,i} F_{\mu i} m_\mu^h m_i^x - \frac{1+\log v^x}{2} - \alpha\frac{1+\log v^h}{2} \\ &+ \frac{1}{2}\sup_{\zeta^x,\zeta^h}\left[\alpha\zeta^h v^h + \zeta^x v^x - (\alpha-1)\log\zeta^h - \int \rho_D(\mathrm{d}\lambda)\log(\zeta^x\zeta^h - \beta^2\lambda)\right] ,\end{aligned} \tag{104}$$

where ρ_D is the spectral distribution of $F^T F$. At $\beta = 0$, $\Phi_F(0)$ is given by eq. (98). As in the symmetric model of Sec. 3.2 we were able to perform this expansion and its resummation by applying our results on the spherical models of Sec. 2. The diagrammatic study performed in Sec. 5.5, which is a generalization of the diagrammatic for symmetric matrices, plays a decisive role in this analysis.

A remark on Restricted Boltzmann machines In the case of an i.i.d. matrix F (see Sec. 5.6 for a remark on how to apply our results to this class of matrices), we recognize in particular in eq. (103) the result obtained in eq. (36) of [TGM+18] for Restricted Boltzmann Machines (RBMs). For generic rotationally invariant F, the fixed point equations corresponding to the extremization of Φ in eq. (104) also correspond to the fixed point of Algorithm 3 of [TGM+18] (which was described there as "adaTAP inference algorithm"). This generic property of the fixed points of the Plefka-expanded free entropy will be investigated in Sec. 4.

3.4.2 Generalized Linear Models with correlated matrices

Generalized Linear Models (GLMs) [NW72, McC18] arise as a generalization of the Compressed Sensing problem (see below). They are of primary importance in a very wide variety of scientific and engineering fields, such as phase retrieval in optics [Fie82] and classification problems in statistics. GLMs can also be thought of as the building blocks of fully connected neural networks [LBH15]. Let us now define more precisely the model we will study. Consider $M, N \geq 1$ both going to infinity with a fixed ratio $M/N \to \alpha > 0$. We are given a (random) measurement matrix $F \in \mathbb{R}^{M\times N}$ which comes from the ensemble of Model R. Given F, data samples $\{Y_\mu\}$ are generated as:

$$\forall \mu \in \{1, \cdots, M\}, \qquad Y_\mu \sim P_{\text{out}}\left(\cdot \,\middle|\, (F\mathbf{X})_\mu\right), \tag{105}$$

in which $\mathbf{X} \in \mathbb{R}^N$ is the vector we try to recover from the observation of $\{Y_\mu\}_{\mu=1}^M$, and P_{out} is a fixed probabilistic channel. Recall that Greek indices μ, ν will always run between 1 and M and Latin indices i, j between 1 and N. The vector $\mathbf{X} \in \mathbb{R}^N$ is assumed to be drawn with i.i.d. coordinates $\{X_i\}_{i=1}^N$ according to a prior P_X with zero mean and variance $\rho > 0$.

Compressed Sensing, the Gaussian channel case Compressed Sensing [Don06] arises as a particular case of channel distribution in eq. (105), in which the channel is taken to be a Gaussian distribution with zero mean and variance Δ. Equivalently, it can be formulated as:

$$Y_\mu = \sum_i F_{\mu i} X_i + \sqrt{\Delta} z_\mu. \tag{106}$$

Our aim is to infer the vector $\mathbf{X}$ from these observations. In this equation we modeled the noise by a standard Gaussian variable z_μ, the strength of the noise being $\Delta > 0$. In [KMS+12] the authors have considered a subclass of matrices F, namely i.i.d. matrices. We follow here the same probabilistic inference approach, as we aim to study this problem by sampling from the following distribution:

$$P(\mathbf{x}|\mathbf{Y}) = \frac{1}{Z_{\mathbf{Y},F}} P_X(\mathbf{x}) \exp\left\{ -\frac{1}{2\Delta} \sum_\mu \left(Y_\mu - \sum_i F_{\mu i}\, x_i \right)^2 \right\}. \tag{107}$$

As the parameters (P_X, Δ) of the signal are known, we could use them in our model, a setting which is known as the *Bayes-optimal* setting. Using the matrix $J = -F^\intercal F$ (which follows Model S) can rewrite the posterior distribution, up to a normalization, as:

$$P(\mathbf{x}|\mathbf{Y}) = \frac{1}{Z_{\mathbf{Y},F}} P_X(\mathbf{x}) \exp\left\{ \frac{1}{2\Delta} \sum_{i,j} J_{ij}\, x_i x_j + \frac{1}{\Delta} \sum_{\mu,i} F_{\mu i} Y_\mu x_i \right\}. \tag{108}$$

Defining $\beta \equiv \Delta^{-1}$ it becomes clear that this can be written as a Gibbs-Boltzmann distribution of the model we studied in Sec. 3.2.1, with $J = -F^\intercal F$ and $h_i = -\sum_\mu F_{\mu i} Y_\mu$. Assuming that the variance variables at the maximum of the free entropy are homogeneous ($v_i = v$) we can use eq. (60) to directly obtain the free entropy of this problem as a function of $\mathcal{R}_J$, the $\mathcal{R}$-transform of the asymptotic spectrum of the J matrix:

$$\Phi_{\mathbf{Y},F}(\beta) = \Phi_{\mathbf{Y},F}(0) - \frac{\beta}{2N} \sum_{i,j} (F^\intercal F)_{ij} m_i m_j + \frac{\beta}{N} \sum_{\mu,i} F_{\mu i} Y_\mu m_i + \frac{1}{2} \int_0^{\beta v} \mathcal{R}_J(u) \mathrm{d}u. \tag{109}$$

We postpone the analysis of the corresponding fixed point equations to our algorithmic discussion in Sec. 4.1 and Sec. 4.2.

Generic channel distributions We now turn to generic P_{out} distributions. We assume that both P_{out} and P_X are *known* (this is the Bayes-optimal setting, known in statistical physics as the *Nishimori line*), so that we can use them in the posterior distribution:

$$P(\mathbf{x}|\mathbf{Y}) = \frac{1}{Z(\mathbf{Y}, F)} \prod_i P_X(x_i) \prod_\mu P_{\text{out}}\left[Y_\mu | (F\mathbf{x})_\mu \right], \tag{110}$$

from which we will sample to obtain an estimate of $\mathbf{X}$. While in the compressed sensing setting $\beta = \Delta^{-1}$ played naturally the role of an inverse temperature, in the general setting of eq. (105) there is *a priori* no way to perform a Plefka expansion. As it turns out, there is a way to introduce an auxiliary parameter in terms of which we will perform the expansion, similarly to what is done in [AFP16, Alt18]. Introducing the usual Lagrange parameters to fix the mean and variance of $\{x_i\}$, we obtain the free entropy:

$$\Phi_{\mathbf{Y},F} \equiv \frac{1}{N} \sum_i \lambda_i m_i + \frac{1}{2N} \sum_i \gamma_i \left(v_i + m_i^2 \right) + \frac{1}{N} \log \left[\int_{\mathbb{R}^N} \mathrm{d}\mathbf{x}\, e^{-S[\mathbf{x}]} \right], \tag{111}$$

in which we introduced an *action* $S[\mathbf{x}]$:

$$S[\mathbf{x}] \equiv \sum_i \lambda_i x_i + \frac{1}{2}\sum_i \gamma_i x_i^2 - \sum_\mu \log P_{\text{out}}\left(Y_\mu \Big| \sum_i F_{\mu i} x_i\right) - \sum_i \log P_X(x_i). \tag{112}$$

As before $\{\lambda_i, \gamma_i\}$ are Lagrange parameters used to enforce the condition $\langle x_i \rangle = m_i$ and $\langle x_i \rangle^2 = v_i + m_i^2$. Introducing an auxiliary field $\mathbf{h} \equiv \mathbf{F}\mathbf{x} \in \mathbb{R}^M$, and using the Fourier representation of the Dirac distribution, we reach:

$$\Phi_{\mathbf{Y},F} = -\alpha \log 2\pi + \frac{1}{N}\sum_i \lambda_i m_i + \frac{1}{2N}\sum_i \gamma_i \left(v_i + m_i^2\right) + \frac{1}{N}\log\left[\int \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{h}\, \mathrm{d}\tilde{\mathbf{h}}\, e^{-S_{\text{eff}}\left[\mathbf{x},\mathbf{h},\tilde{\mathbf{h}}\right]}\right], \tag{113}$$

with a new effective action S_{eff}:

$$\begin{aligned} S_{\text{eff}}\left[\mathbf{x},\mathbf{h},\tilde{\mathbf{h}}\right] \equiv \sum_i \left[\lambda_i x_i + \frac{1}{2}\gamma_i x_i^2\right] - \sum_i \log P_X(x_i) - \sum_\mu \left[\log P_{\text{out}}(Y_\mu|h_\mu) + h_\mu(i\tilde{h}_\mu)\right] \\ + \sum_{\mu,i} (i\tilde{h}_\mu) F_{\mu i} x_i. \end{aligned} \tag{114}$$

The key idea is to treat $\mathbf{x}$ and $i\tilde{\mathbf{h}}$ as two independent non-Gaussian fields that interact via the last (quadratic) term of eq. (114) and to perform a Plefka expansion in terms of this *effective* Hamiltonian, which is exactly the bipartite Hamiltonian of the general model of Sec. 3.4.1. This mapping of a generalized linear model to a bipartite model using Fourier transformation has already been successfully applied in the context of the replica method, see [Kab08a, Kab08b]. We will call η the "inverse temperature", that is in eq. (114) we substitute:

$$\sum_{\mu,i} F_{\mu i}\, x_i\, (i\tilde{h}_\mu) \to \eta \sum_{\mu,i} F_{\mu i}\, x_i\, (i\tilde{h}_\mu), \tag{115}$$

and at the end of the expansion we will set $\eta = 1$. Similarly as for the field $\mathbf{x}$ we will fix the first and second moments of the field $i\tilde{\mathbf{h}}$ as $\langle i\tilde{h}_\mu \rangle_\eta = f_\mu$ and $\langle (i\tilde{h}_\mu)^2 \rangle_\eta = -r_\mu + f_\mu^2$, conditions that will be enforced by new Lagrange parameters $\{\omega_\mu, b_\mu\}$. Although a bit tedious, this is straightforward, and we obtain a free entropy in which we will perform a low-η expansion:

$$\begin{aligned} \Phi_{\mathbf{Y},F}(\eta) = -\alpha \log 2\pi + \frac{1}{N}\sum_i \lambda_i m_i + \frac{1}{2N}\sum_i \gamma_i \left(v_i + m_i^2\right) + \frac{1}{N}\sum_\mu \omega_\mu f_\mu - \frac{1}{2N}\sum_\mu b_\mu \left(-r_\mu + f_\mu^2\right) \\ + \frac{1}{N}\log\left[\int \mathrm{d}\mathbf{x}\, \mathrm{d}\mathbf{h}\, \mathrm{d}\tilde{\mathbf{h}}\, e^{-S_{\text{eff}}\left[\mathbf{x},\mathbf{h},\tilde{\mathbf{h}}\right]}\right]. \end{aligned} \tag{116}$$

The effective action and Hamiltonian are expressed as follows:

$$\begin{aligned} S_{\text{eff}}\left[\mathbf{x},\mathbf{h},\tilde{\mathbf{h}}\right] \equiv \sum_i \lambda_i x_i + \frac{1}{2}\sum_i \gamma_i x_i^2 + \sum_\mu \omega_\mu (i\tilde{h}_\mu) - \frac{1}{2}\sum_\mu b_\mu (i\tilde{h}_\mu)^2 \\ - \sum_i \log P_X(x_i) - \sum_\mu \log P_{\text{out}}\left(Y_\mu|h_\mu\right) - \sum_\mu h_\mu(i\tilde{h}_\mu) + \eta\, H_{\text{eff}}\left[\mathbf{x},\tilde{\mathbf{h}}\right], \end{aligned} \tag{117}$$

$$H_{\text{eff}}[\mathbf{x},\tilde{\mathbf{h}}] \equiv \sum_{\mu,i} F_{\mu i}\, x_i\, (i\tilde{h}_\mu). \tag{118}$$

From these equations it is clear that:

(i) The priors on the variables $\{x_i\}$ and $\{(i\tilde{h}_\mu)\}$ decouple. The prior on x_i is $P_X(x_i)$, while the prior

distribution on $(i\tilde{h})_\mu$ is related to the Fourier transform of the channel distribution:

$$P_{\tilde{H}}(i\tilde{h}_\mu) = \int \frac{\mathrm{d}h}{2\pi} e^{ih\tilde{h}_\mu} P_{\text{out}}(Y_\mu|h). \tag{119}$$

(ii) The interaction Hamiltonian of eq. (118) is a bipartite Hamiltonian of the type of the model we studied in Sec. 3.4.1, in terms of the variables x_i and $(i\tilde{h}_\mu)$.

Points (i) and (ii) allow us to use the general results of Sec. 3.4.1. We can therefore directly conjecture the final form of the free entropy we were seeking:

$$\begin{aligned}\Phi_{\mathbf{Y},F}(\eta=1) = \Phi_{\mathbf{Y},F}(\eta=0) - \frac{1}{N}\sum_{\mu i} F_{\mu i} f_\mu m_i \\ + \frac{1}{N}\sum_{p=1}^{\infty} \frac{(-1)^p}{2p} \sum_{\substack{\mu_1,\cdots,\mu_p \\ \text{pairwise distincts}}} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distincts}}} F_{\mu_1 i_1} F_{\mu_2 i_1} \cdots F_{\mu_p i_p} F_{\mu_1 i_p} \prod_{\alpha=1}^{p} r_{\mu_\alpha} v_{i_\alpha} + \mathcal{O}_N(1).\end{aligned} \tag{120}$$

Homogeneous variances Once again we can assume that the maximum of the free entropy written in eq. (120) will be attained for variance variables $\{v_i, r_\mu\}$ that are *homogeneous*: they satisfy $v_i = v$ and $r_\mu = r$. Using the resummation of eq. (104) this leads to a simplified expression for eq. (120):

$$\begin{aligned}\Phi_{\mathbf{Y},F}(\eta=1) = \Phi_{\mathbf{Y},F}(\eta=0) - \frac{1}{N}\sum_{\mu i} F_{\mu i} f_\mu m_i - \alpha\,\frac{1+\log r}{2} - \frac{1+\log v}{2} \\ + \frac{1}{2}\sup_{\zeta,\zeta'}\left[\alpha\zeta r + \zeta' v - (\alpha-1)\log\zeta - \frac{1}{N}\log\det\left[\zeta\zeta' \mathrm{I}_N + F^\intercal F\right]\right] + \mathcal{O}_N(1).\end{aligned} \tag{121}$$

We will study the fixed point equations corresponding to the free entropies of eq. (120) and eq. (121) in Sec. 4.1 and Sec. 4.3.

4 Consequences for iterative algorithms

In the previous section we showed how to use the Plefka expansion to derive the single-graph free entropy of a large class of systems with pairwise interactions. One then needs to maximize this free entropy, which yields fixed point equations. Iterating these fixed point equations is in itself a challenge since different choices for the iteration scheme can lead to drastically different convergence properties. In the context of the Plefka expansion, or equivalently in the adaTAP or EC approximation, several iterations schemes for the TAP equations have been studied, see for instance [OCW16, ÇOFW16].

On a parallel point of view, message-passing algorithms have been extensively studied in the statistical physics literature. In particular the belief-propagation equations [MM09] can be shown in the large N limit to reduce to a simpler algorithm called Approximate Message Passing (AMP) [DMM09], and derived initially for i.i.d. coupling matrices. It is well understood that for these matrices the stationary limit of the AMP equations is directly related to the fixed point equations of the Plefka free entropy (stopping here at order 2 in the couplings). Extending these algorithms to correlated matrices has been the subject of extensive studies [CWF14, MP17]. The generic (and appealing by its simplicity) iteration scheme called Vector Approximate Message Passing (VAMP) [RSF17] has proven to be very successful both numerically and in its theoretical justification in the case of rotationally-invariant sensing matrices. It has then been generalized to the broader class of generalized linear models with generic output channels [SRF16].

In Sec. 4.1 we describe the connection between AMP equations and the Plefka expansion in the context of generalized linear models with i.i.d. matrices, retrieving the GAMP algorithm [Ran11] and the analysis of

[KMS+12]. In Sec. 4.2 we relate the VAMP algorithm to the Expectation-Consistency equations by showing that the stationary limit of the algorithm yields the TAP equations. We note that in the naïve Plefka expansion performed in terms of Δ^{-1} for the compressed sensing problem there is no clear way to iterate the TAP equations to find back an asymptotically exact algorithm. Finally, in Sec. 4.3 we extend this analysis to Generalized Linear Models with correlated matrices and the G-VAMP algorithm. In contrast with the naïve Δ^{-1}-expansion in compressed sensing, the mapping to a bipartite model that we performed in Sec. 3.4.2 allows to retrieve the stationary limit of the more general G-VAMP algorithm as our TAP equations.

4.1 Generalized Approximate Message Passing for a GLM with i.i.d. matrices

We consider in this subsection the Generalized Linear Model (GLM) of Sec. 3.4.2. We will concentrate on a particular subclass of sensing matrices, namely matrices $F \in \mathbb{R}^{M\times N}$ that are generated with i.i.d. centered standard Gaussian matrix elements $\{F_{\mu i}\}$. Note that the remarks of Sec. 5.6 (coherently with the analysis of [Ran11, BKM+19]) show that the particular distribution of the elements does not need to be Gaussian, as long as the $\{F_{\mu i}\}$ are distributed independently and identically. Generalized linear estimation with such i.i.d. matrices F and generic prior and channel distributions has received a lot of attention recently. In particular Generalized Approximate Message Passing (GAMP), an algorithm first developed in [Ran11], has been shown to be optimal among all polynomial-time algorithms for this problem, see [BKM+19]. A description of the algorithm in our case can be found in eqs. (171)-(177) of [ZK16]. We state here the iterative GAMP equations with our notations:

$$\begin{cases} b^t &= \frac{1}{N}\sum_i v_i^{t-1}, \\ \omega_\mu^t &= \sum_i F_{\mu i} m_i^{t-1} + f_\mu^{t-1} b^t, \\ f_\mu^t &= -g_{\text{out}}(Y_\mu, \omega_\mu^t, b^t), \\ \gamma^t &= -\frac{1}{N}\sum_\mu \partial_\omega g_{\text{out}}(Y_\mu, \omega_\mu^t, b^t), \\ \lambda_i^t &= -\gamma^t m_i^{t-1} + \sum_\mu F_{\mu i} f_\mu^t, \\ m_i^t &= \mathbb{E}_{P_X(\lambda_i^t,\gamma^t)}\left[x\right], \\ v_i^t &= \mathbb{E}_{P_X(\lambda_i^t,\gamma^t)}\left[(x-m_i^t)^2\right]. \end{cases} \tag{122}$$

In these equations $P_X(\lambda_i, \gamma_i)$ is the probability measure with density:

$$P_X(\lambda_i,\gamma_i)(x) \propto P_X(x) e^{-\frac{1}{2}\gamma_i x^2 - \lambda_i x}, \tag{123}$$

and g_{out} is defined from the channel distribution P_{out} as:

$$g_{\text{out}}(y,\omega,b) \equiv \frac{1}{b}\frac{\int \mathrm{d}z\, P_{\text{out}}(y|z)\,(z-\omega)\, e^{-\frac{(z-\omega)^2}{2b}}}{\int \mathrm{d}z\, P_{\text{out}}(y|z)\, e^{-\frac{(z-\omega)^2}{2b}}}. \tag{124}$$

We now turn to the TAP equations that we can derive from the extremization of the Plefka-expanded free entropy of eq. (120). Note that as F is an i.i.d. matrix, all the terms with $p \geq 2$ in eq. (120) will be negligible. More discussion on this can be found in Sec. 5.6. We obtain:

$$\Phi_F(\eta=1) = \Phi_F(\eta=0) - \frac{1}{N}\sum_{\mu i} F_{\mu i} f_\mu m_i - \frac{1}{2N}\sum_{\mu,i} F_{\mu i}^2 r_\mu v_i + \mathcal{O}_N(1). \tag{125}$$

Recall that $\Phi_F(\eta = 0)$ is given by eq. (98). Extremizing $\Phi_F(\eta = 0)$ over the Lagrange parameters yields the moments conditions that we wish to enforce:

$$\begin{cases} m_i &= \mathbb{E}_{P_X(\lambda_i,\gamma_i)}\left[x\right], \\ v_i &= \mathbb{E}_{P_X(\lambda_i,\gamma_i)}\left[(x-m_i)^2\right], \\ f_\mu &= -g_{\text{out}}(y_\mu, \omega_\mu, b_\mu), \\ r_\mu &= -\partial_\omega g_{\text{out}}(y_\mu, \omega_\mu, b_\mu). \end{cases} \tag{126}$$

On the other hand, the maximization of eq. (125) with respect to the physical parameters $\{m_i, v_i, f_\mu, r_\mu\}$ leads to four additional equations:

$$\begin{cases} \lambda_i &= -\gamma_i m_i + \sum_\mu F_{\mu i} f_\mu, \\ \omega_\mu &= f_\mu b_\mu + \sum_i F_{\mu i} m_i, \\ \gamma_i &= \frac{1}{N}\sum_\mu r_\mu, \\ b_\mu &= \frac{1}{N}\sum_i v_i. \end{cases} \tag{127}$$

Combining eq. (126) and eq. (127) we see easily that the equations that one has to solve are exactly the ones of the GAMP algorithm of eq. (122) (without time indices).

The Gaussian channel case In the case of an additive gaussian channel with variance Δ the problem reduces to the compressed sensing studied in Sec. 3.4.2 and Sec. 4.2, with a Gaussian i.i.d. sensing matrix. In this case, the function g_{out} of eq. (124) is computable explicitly, and leads to $f_\mu = (\omega_\mu - y_\mu)/(\Delta + b_\mu)$ and $r_\mu = (\Delta + b_\mu)^{-1}$. In this particular case one recovers the AMP algorithm of [KMS+12], which is also compatible with the VAMP algorithm of Sec. 4.2.

4.2 Vector Approximate Message Passing (VAMP) in Compressed Sensing

4.2.1 The TAP equations in Compressed sensing

We analyze here the fixed point equations for the Compressed Sensing (CS) problem, see Sec. 3.4.2 for the corresponding free entropy derivation. Our starting point is the Plefka-expanded free entropy written in eq. (109). The extremization over the Lagrange parameters $\{\lambda_i, \gamma_i\}$ (which are considered at $\beta = 0$) yields:

$$\begin{cases} m_i = \dfrac{\int \mathrm{d}x P_X(x)\, x\, e^{-\frac{1}{2}\gamma_i x^2 - \lambda_i x}}{\int \mathrm{d}x\, P_X(x)\, e^{-\frac{1}{2}\gamma_i x^2 - \lambda_i x}} \overset{(a)}{\equiv} F_m(\lambda_i, \gamma_i), & \text{(128a)} \\ v_i = \dfrac{\int \mathrm{d}x\, P_X(x)\, (x-m_i)^2\, e^{-\frac{1}{2}\gamma_i x^2 - \lambda_i x}}{\int \mathrm{d}x\, P_X(x)\, e^{-\frac{1}{2}\gamma_i x^2 - \lambda_i x}} \overset{(b)}{\equiv} F_v(\lambda_i, \gamma_i), & \text{(128b)} \end{cases}$$

where (a) and (b) respectively define the functions F_m and F_v. Maximizing the free entropy with respect to the physical parameters $\{m_i, v\}$ results in the following equations (recall that $\beta = \Delta^{-1}$):

$$\begin{cases} \gamma = -\dfrac{1}{\Delta}\mathcal{R}_{-F^\intercal F}\left(\dfrac{v}{\Delta}\right) = \mathcal{R}_{F^\intercal F/\Delta}\left(-v\right), & \text{(129a)} \\ \lambda_i = -\gamma m_i + \dfrac{1}{\Delta}\displaystyle\sum_j (F^\intercal F)_{ij} m_j - \frac{1}{\Delta}\sum_\mu F_{\mu i} Y_\mu. & \text{(129b)} \end{cases}$$

Eq. (128) and eq. (129) define a set of fixed point equations that one has to solve in order to retrieve the maximum of the free entropy of eq. (109).

4.2.2 TAP equations and the fixed point of the VAMP algorithm

A remark on i.i.d. matrices We start with a remark on the case of an i.i.d. matrix F. Remarkably, eqs. (128) and (129) are compatible with the fixed points of AMP, see eqs. (22) and (23) in [KMTZ14] with $R_i = -\lambda_i/\gamma$ and $\Sigma^{-2} = \gamma$, since in this case $\mathcal{R}_{F^\intercal F/\Delta}(-v) = \alpha/(\Delta + v)$, see for instance [TV04]. We now turn to the VAMP algorithm for a general rotationally invariant matrix F. Applying the VAMP derivation of Sec. 3.1.3 to the Compressed Sensing problem of Sec. 3.4.2, the VAMP algorithm reads[1]:

$$(\mathbf{m}_1^t)_i = F_m((\boldsymbol{\lambda}_J^t)_i, a_J^t), \qquad v_1^t = \frac{1}{N}\sum_i F_v((\boldsymbol{\lambda}_J^t)_i, \gamma_J^t), \tag{130}$$

$$\boldsymbol{\lambda}_0^t = -\frac{\mathbf{m}_1^t}{v^t} - \boldsymbol{\lambda}_J^t, \qquad \gamma_0^t = \frac{1}{v^t} - \gamma_J^t, \tag{131}$$

$$\mathbf{m}_2^t = -(\gamma_0^t + F^\intercal F/\Delta)^{-1}(\boldsymbol{\lambda}_0^t - \frac{1}{\Delta}F^\intercal \mathbf{Y}), \qquad v_2^t = \frac{1}{N}\mathrm{Tr}\frac{1}{\gamma_0^t + F^\intercal F/\Delta}, \tag{132}$$

$$\boldsymbol{\lambda}_J^{t+1} = -\frac{\mathbf{m}_2^t}{v_2^t} - \boldsymbol{\lambda}_0^t, \qquad \gamma_J^{t+1} = \frac{1}{v_2^t} - \gamma_0^t, \tag{133}$$

where F_m and F_v were defined in eq. (128). Note that in Compressed Sensing the matrix $\gamma_0^t + F^\intercal F/\Delta$ has only strictly positive eigenvalues since $\gamma_0^t \geq 0$, so the previous iterative equations are always well defined. At the fixed point, we expect $\mathbf{m}_1 = \mathbf{m}_2 = \mathbf{m}$ and $v_1 = v_2 = v$. In the stationary limit eq. (132) yields:

$$v = \mathcal{S}_{F^\intercal F/\Delta}(-\gamma_0) \qquad \Rightarrow \qquad \gamma_0 = -\mathcal{S}^{-1}_{F^\intercal F/\Delta}(v). \tag{134}$$

From eq. (133), one has

$$\gamma_J = \frac{1}{v} + S^{-1}_{F^\intercal F/\Delta}(v) = R_{F^\intercal F/\Delta}(-v). \tag{135}$$

And from eq. (132), we obtain

$$(\gamma_0 + F^\intercal F/\Delta)\mathbf{m} = -\boldsymbol{\lambda}_0 + \frac{1}{\Delta}F^\intercal \mathbf{Y}, \tag{136}$$

and

$$\Big(\frac{1}{v} - \gamma_J + \frac{F^\intercal F}{\Delta}\Big)\mathbf{m} = \frac{\mathbf{m}}{v} + \boldsymbol{\lambda}_J + \frac{1}{\Delta}F^\intercal \mathbf{Y}\,, \tag{137}$$

which gives

$$\boldsymbol{\lambda}_J = \Big(-\gamma_J + \frac{F^\intercal F}{\Delta}\Big)\mathbf{m} - \frac{1}{\Delta}F^\intercal \mathbf{Y} = -\gamma_J \mathbf{m} - \frac{1}{\Delta}F^\intercal\Big(\mathbf{Y} - F\mathbf{m}\Big)\,. \tag{138}$$

One now recognizes easily the fixed points obtained with the Plefka expansion in Sec. 4.2.1, namely eq. (128) and eq. (129), with $\boldsymbol{\lambda}_J = \boldsymbol{\lambda}$ and $\gamma_J = \gamma$.

A remark on iterating the TAP equations in the i.i.d. case Note that in the i.i.d. case (Sec. 4.1), doing the Plefka expansion in terms of the η parameter after having mapped the GLM to a bipartite problem allows us not only to retrieve the fixed point of the GAMP algorithm (and even the G-VAMP for non-i.i.d. matrices, as we will see in Sec. 4.3), but there is a simple iterating scheme of the TAP equations that exactly yields the GAMP algorithm. We insist that this is not true when making the correspondence of the VAMP algorithm with the Plefka expansion in Δ^{-1} for compressed sensing with an i.i.d. matrix. This underlines one of the possible limitations of the EC, adaTAP and Plefka methods for these problems, as iterating the TAP equations with an algorithmic scheme that guarantees convergence is a very involved task, while the VAMP derivation provides an iteration scheme of the equations.

[1] Note that instead of fixing all the correlations $\langle x_i x_j \rangle$, we only fix the 'diagonal' second moments $\langle x_i^2 \rangle$.

4.3 Generalized Vector Approximate Message Passing (G-VAMP) for Generalized Linear Models

We focus in this section on Generalized Linear Models with a correlated matrix F that satisfies rotation invariance (Model R). We first derive the TAP equations from the Plefka expansion we performed in Sec. 3.4.2, before stating the G-VAMP algorithm for this problem following [SRF16]. We then analyze how the stationary limit of G-VAMP is equivalent to these TAP equations.

4.3.1 The TAP equations from the Plefka expansion

Recall that the Plefka-expanded free entropy was computed in Sec. 3.4.2. Following the assumptions of the VAMP and G-VAMP algorithms [SRF16, RSF17] we assume that the variances $\{v_i, r_\mu\}$ are *homogeneous*, that is $r_\mu = r$ and $v_i = v$. We can then use the resummed expression of the Plefka free entropy expressed in eq. (121). We first extremize this expression with respect to the Lagrange parameters $\{\lambda_i, \gamma_i, \omega_\mu, b_\mu\}$ and we obtain an equivalent expression to eq. (128). We reach more precisely:

$$\begin{cases} m_i &= \mathbb{E}_{P_X(\lambda_i,\gamma)}\left[x\right], \\ v_i &= \mathbb{E}_{P_X(\lambda_i,\gamma)}\left[(x-m_i)^2\right], \\ f_\mu &= -g_{\text{out}}(y_\mu, \omega_\mu, b), \\ r &= -\frac{1}{M}\sum_\mu \partial_\omega g_{\text{out}}(y_\mu, \omega_\mu, b). \end{cases} \tag{139}$$

Recall the definitions of $P_X(\lambda, \gamma)$ and $g_{\text{out}}(y, \omega, b)$ from eq. (123) and eq. (124). The remaining equations are obtained by maximizing eq. (121) with respect to the physical parameters. We make use of the Jacobi formula for a symmetric positive definite matrix $J \in \mathcal{S}_N^{++}$: $\frac{\partial}{\partial J_{ij}} \log\det J = (J^{-1})_{ij}$. We reach:

$$\begin{cases} \lambda_i = -\gamma m_i + \sum_\mu F_{\mu i} f_\mu, & \text{(140a)} \\ \omega_\mu = f_\mu b + \sum_i F_{\mu i} m_i, & \text{(140b)} \\ \zeta \mathcal{S}_{F^\intercal F}\left(-\zeta\zeta'\right) = v, & \text{(140c)} \\ \zeta' \mathcal{S}_{F^\intercal F}\left(-\zeta\zeta'\right) = \alpha r - \dfrac{\alpha - 1}{\zeta}, & \text{(140d)} \\ \gamma = \dfrac{1}{v} - \zeta', & \text{(140e)} \\ b = \dfrac{1}{r} - \zeta. & \text{(140f)} \end{cases}$$

Remark: Additive gaussian channel In the case of an additive Gaussian channel with variance Δ we find $r = (\Delta + b)^{-1}$, which gives $\zeta = \Delta$ and $\gamma = \mathcal{R}_{F^\intercal F/\Delta}(-v)$. We thus coherently recover the TAP equations for the compressed sensing problem (see Sec. 4.2.1) even though these equations were derived with a "naïve" Plefka expansion in powers of $\beta \equiv \Delta^{-1}$.

4.3.2 The G-VAMP algorithm for Generalized Linear Models

With a similar reasoning that we used to derive the VAMP algorithm for a symmetric pairwise models, we can write a VAMP algorithm for a bipartite model. We do not describe its full derivation here, and we simply

report the G-VAMP algorithm for the GLM as stated in [SRF16]. We define a set of functions:

$$\begin{cases} \tilde{F}_m(r,\gamma) \equiv \dfrac{\int \mathrm{d}x P_X(x)\, x\, e^{-\frac{1}{2}\gamma(x-r)^2}}{\int \mathrm{d}x\, P_X(x)\, e^{-\frac{1}{2}\gamma(x-r)^2}} = F_m(-\gamma r,\gamma), & \text{(141a)} \\ \tilde{F}_v(r,\gamma) \equiv \dfrac{\int \mathrm{d}x P_X(x)\, x^2\, e^{-\frac{1}{2}\gamma(x-r)^2}}{\int \mathrm{d}x\, P_X(x)\, e^{-\frac{1}{2}\gamma(x-r)^2}} - (\tilde{F}_m(r,\gamma))^2 = F_v(-\gamma r,\gamma), & \text{(141b)} \\ \tilde{F}_z(\omega,\tau) \equiv \dfrac{\int \mathrm{d}z P_{\rm out}(y|z)\, z\, e^{-\frac{1}{2}\tau(z-\omega)^2}}{\int \mathrm{d}z P_{\rm out}(y|z)\, e^{-\frac{1}{2}\tau(z-\omega)^2}} = g_{\rm out}(y,\omega,\tau^{-1})\tau^{-1} + \omega, & \text{(141c)} \\ \tilde{F}_\kappa(\omega,\tau) \equiv \dfrac{\int \mathrm{d}z P_{\rm out}(y|z)\, z^2\, e^{-\frac{1}{2}\tau(z-\omega)^2}}{\int \mathrm{d}z P_{\rm out}(y|z)\, e^{-\frac{1}{2}\tau(z-\omega)^2}} - (\tilde{F}_z(\omega,\tau))^2 = \partial_\omega g_{\rm out}(y,\omega,\tau^{-1})\tau^{-2} + \tau^{-1}. & \text{(141d)} \end{cases}$$

The full algorithm then amounts to iterate the following equations:

$$\mathbf{m}_{1i}^t = \tilde{F}_m((\mathbf{r}_J^t)_i, \gamma_J^t), \qquad v_1^t = \frac{1}{N}\sum_i \tilde{F}_v((\mathbf{r}_J^t)_i, \gamma_J^t), \tag{142a}$$

$$\mathbf{r}_0^t = \frac{(\mathbf{m}^t - \gamma_J^t v^t \mathbf{r}_J^t)}{(1-\gamma_J^t v^t)}, \qquad \gamma_0^t = \frac{1}{v^t} - \gamma_J^t, \tag{142b}$$

$$\mathbf{z}_{1\mu}^t = \tilde{F}_z((\boldsymbol{\omega}_J^t)_\mu, \tau_J^t), \qquad \kappa_1^t = \frac{1}{M}\sum_\mu \tilde{F}_\kappa((\boldsymbol{\omega}_J^t)_\mu, \tau_J^t), \tag{142c}$$

$$\boldsymbol{\omega}_0^t = \frac{(z^t - \tau_J^t \kappa^t \boldsymbol{\omega}_J^t)}{(1-\tau_J^t\kappa^t)}, \qquad \tau_0^{t+1} = \frac{1}{\kappa^t{}_1} - \tau_J^t, \tag{142d}$$

$$\mathbf{m}_2^t = \frac{1}{\tau_0^{t+1}F^TF + \gamma_0^t}\left(\gamma_0^t \mathbf{r}_0^t + F^T\tau_0^t\boldsymbol{\omega}_0^t\right), \qquad v_2^t = \frac{1}{N}\mathrm{Tr}\frac{1}{\tau_0^t F^TF + \gamma_0^t}, \tag{142e}$$

$$\mathbf{r}_J^{t+1} = \frac{(\mathbf{m}_2^t - \gamma_0^t v_2^t \mathbf{r}_0^t)}{(1-\gamma_0^t v_2^t)}, \qquad \gamma_J^{t+1} = \frac{1}{v_2^t} - \gamma_0^t, \tag{142f}$$

$$\mathbf{z}_2^t = F\frac{1}{\tau_0^t F^\intercal F + \gamma_0^t}\left(\gamma_0^t \mathbf{r}_0^t + F^\intercal\tau_0^t\boldsymbol{\omega}_0^t\right), \qquad \kappa_2^t = \frac{1}{M}\mathrm{Tr}F^\intercal F\frac{1}{\tau_0^t F^\intercal F + \gamma_0^t}, \tag{142g}$$

$$\boldsymbol{\omega}_J^{t+1} = \frac{(\mathbf{z}_2^t - \tau_0^t\kappa_2^t\boldsymbol{\omega}_0^t)}{(1-\tau_0^t\kappa_2^t)}, \qquad \tau_J^{t+1} = \frac{1}{\kappa_2^t} - \tau_0^t. \tag{142h}$$

4.3.3 TAP equations and fixed points of G-VAMP

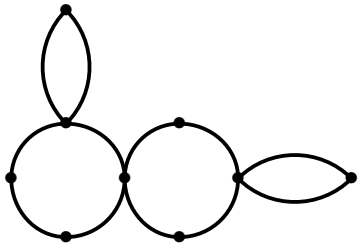
We want to see if the stationary limit of G-VAMP, that is the G-VAMP equations without time indices, is related to the TAP equations. At the fixed points of the G-VAMP algorithm written in eq. (142), one expects the following equalities to take place: $\mathbf{m}_1 = \mathbf{m}_2 = \mathbf{m}$, $\mathbf{z}_1 = \mathbf{z}_2 = \mathbf{z}$, $v_1 = v_2 = v$ and $\kappa_1 = \kappa_2 = \kappa$. We start from the TAP equations, eq. (139) and eq. (140), and we will try to recover every equation in eq. (142).

- From eq. (140f) and eq. (141d) we can write

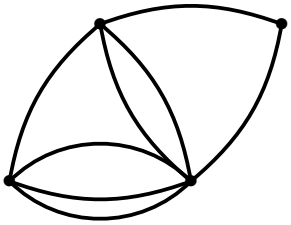
$$\frac{1}{b} = \frac{1}{\kappa} - \frac{1}{\zeta}, \tag{143}$$

  which can be identified with eq. (142h), with $b = \tau_J^{-1}$ and $\zeta = \tau_0^{-1}$.
- Using eq. (141d) we write eq. (140d) as

$$\frac{r}{\zeta'} = -\frac{\kappa}{\zeta' b^2} + \frac{1}{\zeta' b} = \frac{1}{M}\mathrm{Tr}\left[\frac{1}{\zeta\zeta' + F^\intercal F}\right] + \frac{1-\alpha^{-1}}{\zeta'\zeta}. \tag{144}$$



(a) A possible diagram at perturbation order $p = 12$.



(b) Another diagram at order $p = 8$

Figure 3: Cactus and non-cactus diagrams. Each vertex represents an index i over which we sum, and each edge is a factor J_{ij}. Each connected component of the diagrams carries a global $\frac{1}{N}$ factor.

Finally from eq. (143) we obtain

$$\frac{\kappa}{\zeta} = \frac{1}{\alpha} - \frac{1}{M}\mathrm{Tr}\frac{\zeta\zeta'}{\zeta\zeta' + F^\intercal F}, \tag{145}$$

which is compatible with the second part of eq. (142g), with $\zeta = \tau_0^{-1}$ and $\zeta' = \gamma_0$.

- Eq. (140c) and eq. (140e) are equivalent to the second parts of eq. (142e) and eq. (142f), with $\zeta' = \gamma_0$, $\zeta = \tau_0^{-1}$ and $\gamma = \gamma_J$.
- We write eq. (142e) as

$$(\tau_0 F^\intercal F + \gamma_0)\mathbf{m} = (\gamma_0 \mathbf{r}_0 + F^\intercal \tau_0 \boldsymbol{\omega}_0)\,, \tag{146}$$

and using that $F\mathbf{m} = \mathbf{z}$, as well as eq. (142b) and eq. (142d), we arrive at

$$\gamma_J \mathbf{r}_J = \gamma_J \mathbf{m} + \tau_J F^T(\mathbf{z} - \boldsymbol{\omega}_J), \tag{147}$$

which is exactly eq. (140a) with $\omega = \omega_J$, $\boldsymbol{\lambda} = -\mathbf{r}_J\gamma_J$ and $\tau_J = b^{-1}$.

- Finally we note that eq. (140b) at the fixed point is nothing but $\mathbf{z} = F\mathbf{m}$, which gives eq. (142g).

All these relations show the equivalence between the stationary limit of the G-VAMP algorithm of [SRF16] and the (TAP) maximization equations of the free entropy that we derived with our Plefka expansion in Sec. 3.4.2.

5 The diagrammatics of the Plefka expansion

The goal of this section is to precise how the different diagrams arising in our Plefka expansions in Sec. 3 can be computed. Recall that for symmetric random matrices J we construct diagrams as described in Fig. 3. For instance the diagram depicted in Fig. 3a is equal to:

$$\frac{1}{N} \sum_{\substack{i_1,\cdots,i_9 \\ \text{pairwise distincts}}} \left(J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_4} J_{i_4 i_1}\right) \left(J_{i_3 i_5} J_{i_5 i_6} J_{i_6 i_7} J_{i_7 i_3}\right) J^2_{i_6 i_8} J^2_{i_2 i_9}. \tag{148}$$

The perturbation order of any diagram is equal to its number of edges, since each of them represents a factor J_{ij}. In this whole section we will only consider *connected* diagrams (unless stated otherwise). The structure of the section is the following:

- In Sec. 5.1 we prove a first rigorous result on the 'simple cycles' arising in the Plefka expansion of Sec. 2, namely we study these diagrams *in expectation over* J and show a weaker version of Theorem 1.
- In Sec. 5.2 we extend this study to all possible diagrams, in expectation over J.

- In Sec. 5.3 we show how the results of Sec. 5.1 and Sec. 5.2 can be extended to study the second moments of these diagrams, and use it to show concentration results. This will in particular imply the full statement of Theorem 1.
- In Sec. 5.4 we explain how to handle the higher-order moments that can appear as additional factors in these diagrams for the statistical models studied in Sec. 3.
- In Sec. 5.5 we explain how to generalize all these techniques and results to diagrams made of rectangular matrices, that arise in the Plefka expansion for bipartite models.
- Finally, in Sec. 5.6 we show that if one considers an i.i.d. coupling matrix, all the diagrams of order greater than 3 will not contribute in the thermodynamic limit and that one can effectively consider the distribution of the matrix elements to be Gaussian.

Some technicalities, as well as side results and generalizations of these diagrammatics for Hermitian matrices and diverging-size diagrams, which are not directly useful for our expansions, are detailed in Appendix D. We finally note that some of our results are similar to the recent independent work of [BBJ19] that was recently brought to our attention.

5.1 A weaker version of Theorem 1

We will consider the random matrix ensemble defined by Model S. In the following, $J \in \mathcal{S}_N$ is a random matrix from this ensemble. Recall as well the random matrix tools defined in Appendix C, in particular the free cumulants $\{c_p(\rho_D)\}$. We first show a weaker version of Theorem 1:

Theorem 2 (Expectation of simple cycles and free cumulants)**.** *For J following Model S, for any $p \geq 1$, and any set of pairwise distinct indices $i_1, \cdots, i_r \in \mathbb{N}^p$, one has:*

$$\lim_{N\to\infty} \mathbb{E}\left[N^{p-1} J_{i_1 i_2} J_{i_2 i_3} \cdots J_{i_{p-1} i_p} J_{i_p i_1}\right] = c_p(\rho_D). \tag{149}$$

A stronger result actually takes place, that is we only need to average over O to obtain the result:

$$\lim_{N\to\infty} N^{p-1} \int_{\mathcal{O}(N)} \mathcal{D}O \left[(ODO^\intercal)_{i_1 i_2} (ODO^\intercal)_{i_2 i_3} \cdots (ODO^\intercal)_{i_p i_1}\right] = c_p(\rho_D). \tag{150}$$

This last equality is true a.s. with respect to the law of D.

Note that in the Plefka expansions we perform in Sec. 2.1.2 and Sec. 3.2 we consider sums over all distinct pairwise indices of eq. (149). The expectation of these sums over O is an immediate consequence of Theorem 2:

$$\forall p \in \mathbb{N}^\star, \quad \lim_{N\to\infty} \mathbb{E}_O \left[\frac{1}{N} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distincts}}} J_{i_1 i_2} J_{i_2 i_3} \cdots J_{i_{p-1} i_p} J_{i_p i_1}\right] = c_p(\rho_D).$$

We now turn to the proof of Theorem 2.

Proof of Theorem 2. A first pedestrian way to show eq. (150) for small values of p is to use explicit integration of polynomials over the Haar measure of the orthogonal or unitary group, see for instance [CŚ06]. This can be used to check eq. (150) for the first values of p. Since we aim at a generic proof we will choose a different path, leveraging from HCIZ-type integrals [HC57] [IZ80], in the particular case in which one matrix has finite rank. In our setting, the computation of these integrals has been made rigorous in [GM05]. Let us denote:

$$L_p^{(N)} \equiv N^{p-1} \int_{\mathcal{O}(N)} \mathcal{D}O \left[(ODO^\intercal)_{i_1 i_2} (ODO^\intercal)_{i_2 i_3} \cdots (ODO^\intercal)_{i_p i_1}\right]. \tag{151}$$

In order to simplify the following calculation, we assume that $(i_1, \cdots, i_p) = (1, \cdots, p)$. Since the sought result does not depend on the particular choice of indices (as is clear by rotational invariance), this does not remove any generality. We first note that the case $p = 1$ and $p = 2$ are trivial to show by an explicit computation, so we will assume $p \geq 3$ in the following. One can rewrite eq. (151) as:

$$L_p^{(N)} = \frac{1}{N} \prod_{l=1}^{p} \frac{\partial}{\partial b_l} \left[\int_{\mathcal{O}(N)} \mathcal{D}O \, e^{\frac{N}{2} \mathrm{Tr}\,[M(\mathbf{b}) O D O^\intercal]} \right]_{\mathbf{b}=0}, \tag{152}$$

in which we denoted $\mathbf{b} \equiv (b_1, \cdots, b_p)$ and $M(\mathbf{b}) \in \mathcal{S}_N$ the following symmetric block matrix of rank p:

$$M(\mathbf{b}) \equiv \begin{pmatrix} M_1(\mathbf{b}) & (0) \\ (0) & (0) \end{pmatrix}, \tag{153}$$

in which $M_1(\mathbf{b}) \in \mathcal{S}_p$ with:

$$M_1(\mathbf{b}) \equiv \begin{pmatrix} 0 & b_1 & 0 & \cdots & 0 & b_p \\ b_1 & 0 & b_2 & \cdots & 0 & 0 \\ 0 & b_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & b_{p-1} \\ b_p & 0 & 0 & \cdots & b_{p-1} & 0 \end{pmatrix}. \tag{154}$$

Now we can apply Theorem 2 of [GM05]. Recall that G_{ρ_D} is (up to a factor) the integrated $\mathcal{R}$-transform of ρ_D. We obtain:

$$\begin{aligned} \lim_{N\to\infty} L_p^{(N)} &= \lim_{N\to\infty} \frac{1}{N} \left[\prod_{l=1}^{p} \frac{\partial}{\partial b_l} \right] \left[\exp\left\{ N \mathrm{Tr}\, G_{\rho_D}\left[M(\mathbf{b}) \right] \right\} \right]_{\mathbf{b}=0}, \\ &= \lim_{N\to\infty} \frac{1}{N} \left[\prod_{l=1}^{p} \frac{\partial}{\partial b_l} \right] \left[\exp\left\{ \frac{N}{2} \sum_{n=1}^{\infty} \frac{c_n(\rho_D)}{n} \mathrm{Tr}\left[M(\mathbf{b})^n \right] \right\} \right]_{\mathbf{b}=0}. \end{aligned} \tag{155}$$

Let us denote $Z(\mathbf{b}) \equiv \exp\left\{ \frac{N}{2} \sum_{n=1}^{\infty} \frac{c_n(\rho_D)}{n} \mathrm{Tr}\left[M(\mathbf{b})^n \right] \right\}$. Note that differentiating $Z(\mathbf{b})$ with respect to b_1 yields (by cyclicity of the trace):

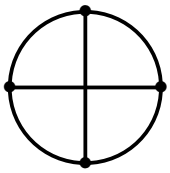
$$\frac{1}{Z(\mathbf{b})} \frac{\partial}{\partial b_1} Z(\mathbf{b}) = \frac{N}{2} \sum_{n=1}^{\infty} c_n(\rho_D) \mathrm{Tr}\, \left[\left\{ \frac{\partial}{\partial b_1} M(\mathbf{b}) \right\} M(\mathbf{b})^{n-1} \right], \tag{156}$$

$$= \frac{N}{2} \sum_{n=1}^{\infty} c_n(\rho_D) \mathrm{Tr}\, \left[E_{12} M(\mathbf{b})^{n-1} \right], \tag{157}$$

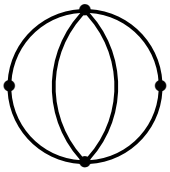
with elementary symmetric matrices $(E_{ab})_{ll'} \equiv \delta_{l,a}\delta_{l',b} + \delta_{l',a}\delta_{l,b}$. These matrices are such that for each $a < b$ and $c < d$, $E_{ab}E_{cd} = 0$ if $\{c, d\} \cap \{a, b\} = \emptyset$. The only way to obtain a matrix of non-zero trace with a product of matrices $\{E_{ab}\}$ is to have a cycle structure in the indices of the matrices. Recall that the indices are symmetric, that is $E_{ba} = E_{ab}$. For instance:

$$\begin{aligned} &\mathrm{Tr}\, \left[E_{12}^2 E_{13} E_{23} E_{12} \right] = \mathrm{Tr}\, \left[E_{12} E_{21} E_{13} E_{32} E_{21} \right] \neq 0, \\ &\mathrm{Tr}\, \left[E_{12}^2 E_{24} E_{23} E_{12} \right] = 0. \end{aligned}$$

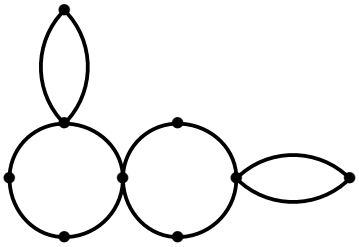
Because of this and the fact that $M(\mathbf{b} = 0) = 0$, it is easy to see that the only term that will survive after



(a) A non-Eulerian diagram.



(b) An Eulerian strongly irreducible diagram.



(c) A cactus diagram.

Figure 4: Cactus and non-cactus diagrams. Each vertex represents an index i over which we sum, and each edge is a factor J_{ij}. Each diagram carries a global $\frac{1}{N}$ factor.

taking all the successive derivatives and taking $\mathbf{b} = 0$ will be the derivatives of the right-hand-side of eq. (157), and not other derivatives of $Z(\mathbf{b})$. Let us analyze what differentiating this term yields. As we saw, taking derivative with respect to b_1 yields a matrix E_{12}. When differentiating with respect to b_2 this yields a matrix E_{23}. Note that *a priori*, one would have:

$$\frac{\partial}{\partial b_2}\mathrm{Tr}\left[E_{12}M(\mathbf{b})^{n-1}\right] = \mathrm{Tr}\left[E_{12}\sum_{k=0}^{n-2} M(\mathbf{b})^k E_{23} M(\mathbf{b})^{n-2-k}\right]. \tag{158}$$

However, the following differentiations with respect to $b_3, \cdots, b_p$ will never yield a matrix E_{ab} with one of the indices being equal to 2. So in eq. (158) it is clear that only two terms of the sum, the term $k = 0$ and $k = n-2$, will yield a non-zero contribution. In the end, after taking all the p successive derivatives, only two terms will remain, which correspond to the two possible orientations of the simple cycle:

$$\begin{aligned}\lim_{N\to\infty} L_p^{(N)} &= \frac{1}{2}\sum_{n=p}^{\infty} c_n(\rho_D)\mathrm{Tr}\left[\left(E_{12}E_{23}\cdots E_{p1} + E_{1p}E_{pp-1}\cdots E_{32}E_{21}\right) M(0)^{n-p}\right], \\ &= c_p(\rho_D),\end{aligned} \tag{159}$$

using that $M(0) = 0$, which finishes the proof. □

5.2 The expectation of generic diagrams

Following the remarks of [GY91] and [PP95], we can separate some of the diagrams constructed as in Fig. 3 in three disjoint categories or types:

- **T.1** *Non-Eulerian diagrams.* By definition, a diagram is Eulerian if one can construct a cyclic path in the graph that goes through each edge exactly once. Note that this is a classic result of graph theory (the Euler–Hierholzer theorem) that these graphs are exactly the connected graphs with even degree in each vertex. For instance, the graph depicted in eq. (4a) is not Eulerian, whereas the one of Fig. 4b is Eulerian.
- **T.2** *Eulerian diagrams that are strongly irreducible but not simple cycles.* By strongly irreducible, we mean [GY91] that one can not make it disconnected by removing any single vertex. For instance, the diagram of Fig. 4b is strongly irreducible, whereas the diagram of Fig. 4c is not.
- **T.3** *Cactus diagrams.* These diagrams, like the one of Fig. 4c, are trees made of simple cycles joining at their vertices. Among them are of course the *simple cycles*.

We are not interested in Eulerian diagrams that are not strongly irreducible. Indeed, as argued in [GY91], only strongly irreducible diagrams will appear in the Plefka expansions. This is an important hypothesis of the Plefka expansion, somehow a bit hidden by the formalism. We give precise descriptions of the large N limit

of the *expectation* of all these diagrams in the following. When we write "expectation" we will always mean expectation over the orthogonal matrix of Model S. More precisely, we will show:

- (i) All non-Eulerian graphs of type T.1 have a vanishing expectation in the $N \to \infty$ limit.
- (ii) All strongly irreducible diagrams of type T.2 also have a vanishing expectation in the $N \to \infty$ limit.
- (iii) We already showed that the expectation of a simple cycle of size p converges to the p-th free cumulant of ρ_D in Sec. 5.1. We show that the expectation of a cactus diagram converges to the product of the expectations of all its constituent simple cycles. For instance, for the diagram $\mathcal{C}$ of Fig. 4c we obtain that its expectation converges to:

$$\lim_{N\to\infty} \mathbb{E}\,\mathcal{C} = c_2(\rho_D)^2 c_4(\rho_D)^2. \tag{160}$$

Results (i) and (ii) are justified in Sec. 5.2.1, and are directly useful for our diagrammatic expansions. Result (iii) on the other hand is a side result that is not used in our expansions, as we argued that only strongly irreducible diagrams come up in our expansions [GY91]. It is justified in Sec. 5.2.2.

5.2.1 Eulerian diagrams, strongly irreducible diagrams and simple cycles

Let us consider a connected diagram G with V vertices and E edges. We will show that:

- If G is not Eulerian, its expectation goes to 0 as $N \to \infty$.
- If G is Eulerian and strongly irreducible, but is not a simple cycle, its expectation also goes to 0 as $N \to \infty$.

Once averaged over the orthogonal matrices, the permutation invariance of the indices allows us to write

$$\mathbb{E}\,G = N^{V-1} \int_{\mathcal{O}(N)} \mathcal{D}O \prod_{1\le l<l'\le V} (ODO^\intercal)_{ll'}^{\epsilon_{ll'}}\,, \tag{161}$$

in which the $\epsilon_{ll'}$ are positive integers such that $\sum_{l<l'} \epsilon_{ll'} = E$. We can now use the results of [GM05], as we did in Sec. 5.1, to write this diagram as (in the $N \to \infty$ limit):

$$\mathbb{E}\,G = N^{V-E-1} \left[\prod_{l<l'} \frac{\partial^{\epsilon_{ll'}}}{\partial b_{ll'}^{\epsilon_{ll'}}}\right] \left[\exp\left\{\frac{N}{2}\sum_{n=1}^{\infty} \frac{c_n(\rho_D)}{n} \mathrm{Tr}\,[M(\mathbf{b})]^n\right\}\right]_{\mathbf{b}=0}. \tag{162}$$

In this expression, $M(\mathbf{b})_{ll'} \equiv b_{ll'} = M(\mathbf{b})_{l'l}$ for $l < l'$,and the diagonal is zero: $M(\mathbf{b})_{ll} = 0$. Exactly as in Sec. 5.1, the elementary matrices $\{E_{ll'}\}$ will appear in eq. (162) by successive derivatives of the exponential, using the fact that $\frac{\partial}{\partial b_{ll'}} M(\mathbf{b}) = E_{ll'}$ and then using $M(\mathbf{b}=0) = 0$. As we explained in Sec. 5.1, a trace of the products of the $\{E_{ll'}\}$ matrices will only be non-zero if and only if the indices in the products form a cycle. Moreover, as is clear in eq. (162), the terms corresponding to the decomposition of $\mathbb{E}\,G$ into the maximum number of such cycles will dominate in the large N limit, as each derivation of the exponential term adds a multiplicative factor N [2]. These two facts together imply that:

- If G is not Eulerian, as in Fig. 4a, its expectation will be 0 in the limit $N \to \infty$ since it is not possible to decompose it into disjoint cycles by definition.
- If G is Eulerian, strongly irreducible, but not a simple cycle, the dominant contribution to $\mathbb{E}\,G$ in eq. (162) will arise from decomposing the graph G into simple cycles, as this decomposition maximizes the number of cycles, and we already showed that each simple cycle has a $\mathcal{O}_N(1)$ contribution. For the graph of Fig. 4b, we show two such possible decompositions in Fig. 5.

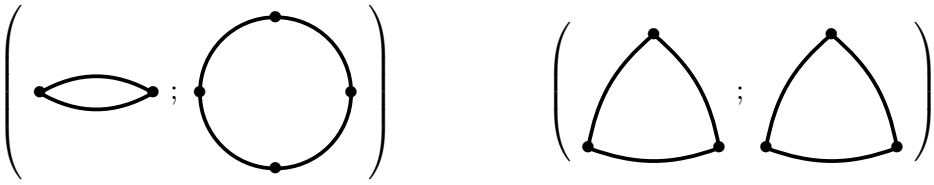


Figure 5: Two possible decompositions of the diagram of Fig. 4b into simple cycles.

Given the remarks above we assume now that G is Eulerian and strongly irreducible. Let us denote P the maximal number of simple cycles in such a decomposition of the graph G. Then one can see that the scaling of eq. (162) will be:

$$\mathbb{E}\, G \sim N^{V+P-E-1}.$$

One can easily be convinced that for a strongly irreducible diagram G we have $V + P - E - 1 \leq 0$, and we have equality only if G is a simple cycle. This implies that all the strongly irreducible diagrams that are not simple cycles and that appear in our Plefka expansions in Sec. 2 and Sec. 3 will not contribute in the $N \to \infty$ limit.

5.2.2 Cactus diagrams

As a side result, although it's not directly useful for our Plefka expansions, we show that we can compute the large N limit of any "cactus" [PP95] diagram (like the one of Fig. 4c) as a function of the free cumulants of ρ_D. The argument is straightforward and uses the same technique as in Sec. 5.2.1. Consider a cactus diagram G with V vertices and E edges. One can write the same equation as eq. (162):

$$\mathbb{E}\, G = N^{V-E-1} \left[\prod_{l<l'} \frac{\partial^{\epsilon_{ll'}}}{\partial b_{ll'}^{\epsilon_{ll'}}} \right] \left[\exp\left\{ \frac{N}{2} \sum_{n=1}^{\infty} \frac{c_n(\rho_D)}{n} \mathrm{Tr}\left[M(\mathbf{b})\right]^n \right\} \right]_{\mathbf{b}=0}. \tag{163}$$

Again, the dominant contribution is obtained by decomposing G in as many simple cycles as possible. For a cactus diagram it is easy to see that there is only *one* such decomposition, which corresponds to its natural decomposition into its constituent simple cycles, and that the number of such cycles is $P = E + V - 1$. Let us denote $\{r_1, \cdots, r_P\}$ the number of vertices in each of these P simple cycles. The dominant contribution corresponds to differentiating P times inside the exponential of eq. (163). Using exactly the argument of Sec. 5.1 for each of the P simple cycles we finally obtain:

$$\begin{aligned} \mathbb{E}\, G &= N^{P+V-E-1} \prod_{\alpha=1}^{P} c_{r_\alpha}(\rho_D) + \mathcal{O}_N(1), \\ \mathbb{E}\, G &= \prod_{\alpha=1}^{P} c_{r_\alpha}(\rho_D) + \mathcal{O}_N(1). \end{aligned} \tag{164}$$

This justifies the point (iii) that we gave in the introductory part of the section: the expectation of the cactus diagrams decouple into the products of their simple cycles constituents.

5.3 Concentration of the diagrams: a second moment analysis

Using our first moment results of Sec. 5.1 and Sec. 5.2, we will show the following results:

[2] There might be a confusion, so we emphasize that this "decomposition" of $\mathbb{E}\, G$ is a decomposition of the *graph* representing $\mathbb{E}\, G$.

(i) If $\mathcal{C}_p$ is the simple cycle of order p, then we have that $\lim_{N\to\infty} \mathcal{C}_p \overset{L^2}{=} c_p(\rho_D)$, which implies directly Theorem 1 and thus ends its proof. Moreover, if G is a cactus diagram then it converges in L^2 to the products of the free cumulants corresponding to its constituent simple cycles.

(ii) If G is of the type T.1 or T.2, we have:

$$\lim_{N\to\infty} \mathbb{E}\, G^2 = 0. \tag{165}$$

This implies that the diagram G will be negligible in the $N \to \infty$ limit.

Note that following the arguments of [GY91], one can convince oneself that *only* strongly irreducible diagrams will contribute in general to the expansion in our models. Together with point (ii) this shows in more detail why only the simple cycles contribute in our Plefka expansions, like in eq. (25) for the spherical model of Sec. 2.1. In order to show (i) and (ii) we will establish the following fact. Consider a diagram G with V vertices and E edges, of any of the types T.1, T.2, or T.3. Then one has:

$$\mathbb{E}\, G^2 = (\mathbb{E} G)^2 + \frac{1}{N} \sum_{\alpha} \mathbb{E}\, \mathcal{C}_\alpha + \mathcal{O}_N(1). \tag{166}$$

In this formula, the sum $\sum_\alpha \mathcal{C}_\alpha$ represents *all the possible diagrams* that one can obtain by 'gluing' together two replicas of the diagram G. Indeed, one can write the generic form of a diagram G as:

$$G = \frac{1}{N} \sum_{\substack{i_1,\cdots,i_V \\ \text{pairwise distincts}}} \prod_{1\leq l<l'\leq V} J_{i_l i_{l'}}^{\epsilon_{ll'}},$$

in which the integers $\epsilon_{ll'}$ verify $\sum_{l<l'} \epsilon_{ll'} = E$. Thus one has:

$$\mathbb{E}\, G^2 = \mathbb{E}\left[\frac{1}{N^2} \sum_{\substack{i_1,\cdots,i_V \\ \text{pairwise distincts}}} \sum_{\substack{j_1,\cdots,j_V \\ \text{pairwise distincts}}} \prod_{1\leq l<l'\leq V} J_{i_l i_{l'}}^{\epsilon_{ll'}} J_{j_l j_{l'}}^{\epsilon_{ll'}} \right].$$

In this expression, one can see that two types of terms have to be taken into account:

- A term for which *all indices* $\{i_1, \cdots, i_V, j_1, \cdots, j_V\}$ are pairwise distinct. Diagrammatically, this corresponds to a graph with two disconnected components that are identical and equal to G. Therefore, one can repeat the arguments of Sec. 5.1 and Sec. 5.2 straightforwardly. Indeed, as all the indices are distincts, the decomposition of this diagram into the maximum number of simple cycles will be two copies of the maximal decomposition of G. This yields that this term is equal in the $N \to \infty$ limit to $(\mathbb{E}\, G)^2$.
- Terms for which there is at least one equality of the type $i_l = j_{l'}$ for $1 \leq l, l' \leq V$. Such a term thus corresponds to a diagram with a *single* connected component and constructed by 'gluing' some of the vertices of two identical copies of G. Since these diagrams have a single connected component, they carry a single $\frac{1}{N}$ factor, which explains the term $\frac{1}{N}\sum_\alpha \mathbb{E}\, \mathcal{C}_\alpha$ in eq. (166), if we denote $\mathcal{C}_\alpha$ each of these possible terms.

We give a schematic representation of eq. (166) for a simple cycle in Fig. 6. It is now possible to see why it implies our results (i) and (ii). Indeed, all the diagrams $\mathcal{C}_\alpha$ have an expectation that is $\mathcal{O}_N(1)$ by the first moment analysis we performed in Sec. 5.1 and Sec. 5.2. So very generically, for every kind of diagram we described we have:

$$\mathbb{E}\, G^2 = (\mathbb{E}\, G)^2 + \mathcal{O}_N(1). \tag{167}$$

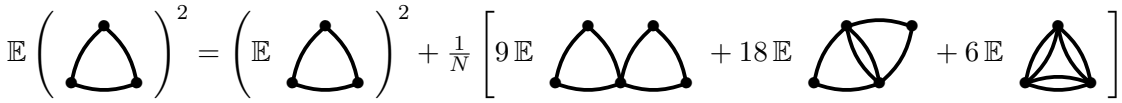


Figure 6: Second moment decomposition of the simple cycle of order 3. We detail the combinatorial factors.

Given our previous computations of the first moments this implies results (i) and (ii).

5.4 The higher-order moments and their influence on the diagrammatics in the symmetric model

All the results of Sec. 5.1, Sec. 5.2 and Sec. 5.3 that we derived for the diagrammatics of the Plefka expansion in this context were valid for diagrams solely made out of the matrix elements $\{J_{ij}\}$, without any additional factors. However in the Plefka expansions there generically are possible factors that are the cumulants (or the moments) of the variables x_i at $\beta = 0$, see Sec. 3.2. Recall that we denote $\kappa_i^{(p)}$ the cumulant of order p of x_i at $\beta = 0$. As an example, consider the diagram of Fig. 4b. Two possible contributions to the free entropy in our Plefka expansion at order 6 would be:

$$\frac{1}{N} \sum_{\substack{i_1,i_2,i_3 \\ \text{pairwise distincts}}} J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_4} J_{i_4 i_1} J_{i_1 i_3}^2 \, v_{i_1}^2 v_{i_2} v_{i_3}^2 v_{i_4}, \tag{168}$$

$$\frac{1}{N} \sum_{\substack{i_1,i_2,i_3 \\ \text{pairwise distincts}}} J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_4} J_{i_4 i_1} J_{i_1 i_3}^2 \, \kappa_{i_1}^{(4)} v_{i_2} \kappa_{i_3}^{(4)} v_{i_4}. \tag{169}$$

Note that both these contributions are represented by the diagram of Fig. 4b. One can now clearly see that in order to apply the diagrammatic results of Sec. 5.1, Sec. 5.2 and Sec. 5.3 to our Plefka expansion, and justify eq. (59), we need to make some additional assumptions that we detail here:

A.1 From the construction of the diagrams, odd cumulants of order greater or equal to 3 only appear in *non-Eulerian* graphs. By the results of Sec. 5.2 and Sec. 5.3 we know that these diagrams, without the moments or cumulants as factors, are negligible. We assume that the possible correlations of the higher order moments of x_i with the matrix elements $\{F_{\mu i}\}$ are not strong enough to yield thermodynamically relevant corrections to the free entropy.

A.2 Eulerian strongly irreducible diagrams that are not simple cycles are negligible by our previous result. We assume that the higher order (even) moments that appear as additional factors do not change their scaling, so that they remain negligible in the thermodynamic limit.

For instance, A.2 implies that the contributions of both eq. (168) and eq. (169) are negligible in the $N \to \infty$ limit, as the diagram of Fig. 4b is strongly irreducible but is not a simple cycle. Concerning the simple cycles, we already know that they are not thermodynamically negligible. So we do not need to assume anything additional regarding them. Note however that in order to "resum" the free entropy of the Plefka expansion, as we did in Sec. 3.2, we will need to assume that all the variance factors appearing in these simple cycles will be the same, that is $v_i = v$ (at the maximum of the free entropy).

5.5 Extension to bipartite models

We detail here how we can treat the diagrams that arise in the Plefka expansion of bipartite models with pairwise interactions (as the generalized linear models) that we perform in Sec. 3.4. The structure of this section is the following:

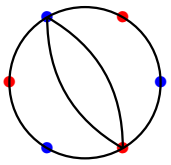


Figure 7: A diagram constructed from the non square matrix F. Each blue vertex is an index μ, each red vertex an index i. Each edge is a factor $F_{\mu i}$, and we sum on each vertex the resulting quantity. Each connected component of the diagram carries a global factor $\frac{1}{N}$. Note that there can only be edges between red and blue vertices.

- We show in Sec. 5.5.1 how we can generalize all the techniques and results already seen in the rest of Sec. 5 to diagrams constructed from a random rectangular matrix L drawn from the rotationally invariant ensemble given by Model R.
- In Sec. 5.5.2 we transpose the assumptions of Sec. 5.4 to this bipartite case, to deal with the higher-order moments of the fields that can arise in the high-temperature Plefka expansions.

5.5.1 Generalization of the previous results to rectangular matrices

Consider a random matrix $F \in \mathbb{R}^{M\times N}$ drawn from a rotation invariant ensemble satisfying Model R. We are interested in the limit $M, N \to \infty$ with a finite ratio $M/N \to \alpha > 0$. In the Plefka expansions performed for bipartite models in Sec. 2.2.2 and Sec. 3.4 they appear some quantities that we can represent as *diagrams*. In this subsection, we construct diagrams as explained Fig. 7. For instance, the diagram depicted in this figure represents the quantity:

$$\frac{1}{N} \sum_{\substack{\mu_1,\mu_2,\mu_3 \\ \text{pairwise distincts}}} \sum_{\substack{i_1,i_2,i_3 \\ \text{pairwise distincts}}} F_{\mu_1 i_1} F_{\mu_1 i_2} F_{\mu_2 i_2} F_{\mu_2 i_3} F_{\mu_3 i_3} F_{\mu_3 i_1} F^2_{\mu_1 i_3}. \tag{170}$$

The analogous to Theorem 2 of [GM05] for this setting can be stated. Let $\Sigma \in \mathbb{R}^{M\times N}$ be a matrix such that the empirical spectral distribution of $D \equiv \Sigma^\intercal \Sigma$ converges (almost surely) as $N \to \infty$ to a probability measure ρ_D. Denote $\mathcal{G}_{\alpha,\rho_D}$ the following function:

$$\mathcal{G}_{\alpha,\rho_D}(x) \equiv \frac{1}{2} \sup_{\gamma_1,\gamma_2} \left[\alpha\gamma_1 + \gamma_2 - (\alpha-1)\log\gamma_1 - \int \rho_D(\mathrm{d}\lambda) \log(\gamma_1\gamma_2 - x^2\lambda) \right] - \frac{1+\alpha}{2}. \tag{171}$$

Note that this is obviously an even function of x, and that $\mathcal{G}_{\alpha,\rho_D}(0) = 0$. The function $\mathcal{G}_{\alpha,\rho_D}$ stands as an analog to the integrated $\mathcal{R}$-transform G_{ρ_D} for this problem. Since one could expand the function G_{ρ_D} using the free cumulants of ρ_D, we analogously expand formally $\mathcal{G}_{\alpha,\rho_D}(x)$ around $x = 0$, and define the coefficients $\Gamma_p(\alpha, \rho_D)$ by:

$$\mathcal{G}_{\alpha,\rho_D}(x) \equiv \sum_{p=1}^{\infty} \frac{1}{2p} \Gamma_p(\alpha,\rho_D) x^{2p}. \tag{172}$$

Recall that for any function $f(x)$, and any symmetric matrix $J = ODO^\intercal \in \mathcal{S}_N$, one can define $f(J) \equiv Of(D)O^\intercal$, with $f(D) = \mathrm{Diag}\left(\{f(d_i)\}_{1\le i\le N}\right)$. If one can expand $f(x) = \sum_{k\ge 0} c_k x^k$, this definition is coherent with $f(J) = \sum_{k\ge 0} c_k J^k$. Our generalization of Theorem 2 of [GM05] is the following: consider a rectangular matrix $\Lambda \in \mathbb{R}^{M\times N}$ of finite rank p. In other terms, one can write its SVD decomposition as:

$$\Lambda = U_0 \Delta V_0^\intercal, \qquad \Delta = \begin{pmatrix} \Lambda_p & 0 \\ 0 & 0 \end{pmatrix} \in \mathbb{R}^{M\times N},$$

with $\Lambda_p \in \mathbb{R}^{p\times p}$ a square diagonal matrix, and U_0, V_0 orthogonal matrices. We can now state:

$$\lim_{N\to\infty} \frac{1}{N} \log \int_{\mathcal{O}(M)} \mathcal{D}U \int_{\mathcal{O}(N)} \mathcal{D}V e^{\sqrt{\alpha} N \mathrm{Tr}\,[\Lambda^\intercal U \Sigma V^\intercal]} = \mathrm{Tr}\,\left[\mathcal{G}_{\alpha,\rho_D}(\Lambda_p)\right], \tag{173}$$

Note first that the right hand side of this equation can also be written as:

$$\mathrm{Tr}\,\left[\mathcal{G}_{\alpha,\rho_D}(\Lambda_p)\right] = \mathrm{Tr}\,\left[\mathcal{G}_{\alpha,\rho_D}(\sqrt{\Lambda^\intercal\Lambda})\right] \overset{(a)}{=} \mathrm{Tr}\,\left[\mathcal{G}_{\alpha,\rho_D}(\sqrt{\Lambda\Lambda^\intercal})\right], \tag{174}$$

since Λ is of finite rank p. Equality (a) is true since $\mathcal{G}_{\alpha,\rho_D}(0) = 0$, and the spectrum of $\Lambda^\intercal\Lambda$ and $\Lambda\Lambda^\intercal$ only differ by eigenvalues which are all equal to 0. Note also that we already derived and used this relation, for $p = 1$, when computing the free entropy of the model of Sec. 2.2, as stated in eq. (33). Equipped with the definitions of $\mathcal{G}_{\alpha,\rho_D}$, $\Gamma_p(\alpha,\rho_D)$, and eq. (173), we can state the counterpart of all our previous results in this rectangular setting:

R.1 Consider a simple cycle of size $2p$. Then it converges (in L^2) to $\Gamma_p(\alpha,\rho_D)$ as $N\to\infty$. More precisely we have:

$$\lim_{N\to\infty} \mathbb{E}\left| \frac{1}{N} \sum_{\substack{\mu_1,\cdots,\mu_p \\ \text{pairwise distincts}}} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distincts}}} F_{\mu_1 i_1} F_{\mu_1 i_2} F_{\mu_2 i_2} \cdots F_{\mu_p i_p} F_{\mu_p i_1} - \Gamma_p(\alpha,\rho_D) \right|^2 = 0. \tag{175}$$

R.2 Any diagram G that is not *Eulerian* will have a vanishing first and second moment as $N\to\infty$:

$$\lim_{N\to\infty} \mathbb{E}\, G^2 = 0. \tag{176}$$

R.3 Any diagram G that is *strongly irreducible* (that is it can not be disconnected by removing a single vertex) but not a simple cycle, as in Fig. 7, will also have a vanishing first and second moment.

R.4 If G is a *cactus* (a tree made of simple cycles joining at vertices [PP95]) made of r simple cycles of size $(2p_1,\cdots,2p_r)$, we have:

$$\lim_{N\to\infty} \mathbb{E}\left| G - \prod_{l=1}^{r} \Gamma_{p_l}(\alpha,\rho_D) \right|^2 = 0. \tag{177}$$

Since every argument to show points R.1 to R.4 is straightforwardly given by slightly modifying what we already did in Sec. 5, the rest of Sec. 5.5.1 will be devoted to show point R.1, and we leave the remaining points for the reader.

Justifying R.1 In order to show eq. (175), we proceed as in Sec. 5.1 and begin by showing:

$$\lim_{N\to\infty} \mathbb{E}\left[\frac{1}{N} \sum_{\substack{\mu_1,\cdots,\mu_p \\ \text{pairwise distincts}}} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distincts}}} F_{\mu_1 i_1} F_{\mu_1 i_2} F_{\mu_2 i_2} \cdots F_{\mu_p i_p} F_{\mu_p i_1} \right] = \Gamma_p(\alpha,\rho_D) + \mathcal{O}_N(1). \tag{178}$$

By rotation invariance of the indices we can replace the left-hand side of eq. (178) by a term without summation on the indices, and as in Sec. 5.1 we obtain at leading order in N:

$$\alpha^p N^{2p-1} \mathbb{E}\left[F_{11}F_{12}F_{22}\cdots F_{pp}F_{p1}\right] = \frac{1}{N} \frac{\partial^{2p}}{\partial b_1 \cdots \partial b_p \partial c_1 \cdots \partial c_p} \left[\int \mathcal{D}U\mathcal{D}V e^{\sqrt{\alpha} N \mathrm{Tr}\,[M(\mathbf{b},\mathbf{c})^\intercal U \Sigma V^\intercal]} \right]_{\mathbf{b},\mathbf{c}=0},$$

with $M(\mathbf{b},\mathbf{c})$ a block matrix of rank p defined as:

$$M(\mathbf{b},\mathbf{c}) = \begin{pmatrix} M_1(\mathbf{b},\mathbf{c}) & 0 \\ 0 & 0 \end{pmatrix},$$

$$M_1(\mathbf{b},\mathbf{c}) \equiv \begin{pmatrix} b_1 & c_1 & 0 & \cdots & 0 & 0 \\ 0 & b_2 & c_2 & \cdots & 0 & 0 \\ 0 & 0 & b_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & b_{p-1} & c_{p-1} \\ c_p & 0 & 0 & \cdots & 0 & b_p \end{pmatrix}.$$

Using eq. (173), we obtain:

$$\alpha^p N^{2p-1}\mathbb{E}\left[F_{11}F_{12}F_{22}\cdots F_{pp}F_{p1}\right] = \frac{1}{N}\frac{\partial^{2p}}{\partial b_1\cdots\partial b_p\partial c_1\cdots\partial c_p}\left[\exp\left\{N\sum_{n=1}^{\infty}\frac{\Gamma_n(\alpha,\rho_D)}{2n}\mathrm{Tr}\left[(M(\mathbf{b},\mathbf{c})^\intercal M(\mathbf{b},\mathbf{c}))^n\right]\right\}\right]_{\mathbf{b},\mathbf{c}=0} + \mathcal{O}_N(1), \tag{179}$$

We define the elementary matrices $(T_{ab})_{ll'} = \delta_{al}\delta_{bl'}$ and the symmetric elementary matrices $E_{ab} = T_{ab} + T_{ba}$. One easily derives that $\frac{\partial^2}{\partial b_1\partial c_1}M(\mathbf{b},\mathbf{c})^\intercal M(\mathbf{b},\mathbf{c}) = E_{12}$. In a very similar way to what was done in Sec. 5.1, the dominant terms in eq. (179) will be given by the maximum number of differentiations of the exponential term. However, one can see that the exponential can only be differentiated once: since $M(0,0) = 0$, one would need to create cycles with the matrices E_{ab}, and such a cycle can only appear if one derives a single time the exponential term. As in Sec. 5.1, there are two cycles that are created by the successive derivatives: $E_{12}E_{23}\cdots E_{p1}$ and $E_{21}E_{1p}\cdots E_{32}$. These two cycles yield the dominant contribution:

$$\begin{aligned}
&\alpha^p N^{2p-1}\mathbb{E}\left[F_{11}F_{12}F_{22}\cdots F_{pp}F_{p1}\right] \\
&\quad= \sum_{n=p}^{\infty}\frac{\Gamma_n(\alpha,\rho_D)}{2}\mathrm{Tr}\left[\left(E_{12}E_{23}\cdots E_{p1} + E_{21}E_{1p}\cdots E_{32}\right)\left(M(0,0)^\intercal M(0,0)\right)^{n-p}\right] + \mathcal{O}_N(1), \\
&\quad= \frac{1}{2}\Gamma_p(\alpha,\rho_D)\mathrm{Tr}\left(E_{12}E_{23}\cdots E_{p1} + E_{21}E_{1p}\cdots E_{32}\right) + \mathcal{O}_N(1), \\
&\quad= \Gamma_p(\alpha,\rho_D) + \mathcal{O}_N(1).
\end{aligned}$$

This shows eq. (178). The exact same arguments as the ones used in Sec. 5.3 show that we have L^2 concentration, which means:

$$\lim_{N\to\infty}\mathbb{E}\left|\frac{1}{N}\sum_{\substack{\mu_1,\cdots,\mu_p \\ \text{pairwise distincts}}}\sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distincts}}} F_{\mu_1 i_1}F_{\mu_1 i_2}F_{\mu_2 i_2}\cdots F_{\mu_p i_p}F_{\mu_p i_1} - \Gamma_p(\alpha,\rho_D)\right|^2 = 0,$$

which is the point R.1 we wanted to show.

5.5.2 The higher order moments and their influence on the diagrammatics

In Sec. 3.4, we deal with diagrams which have additional factors coming from the higher order moments of the fields x_i and h_μ at $\beta = 0$, while all the results R.1 to R.4 that we derived for the diagrammatics of the Plefka expansion in this context were made solely out of the matrix elements $\{F_{\mu i}\}$, without any additional factors. We adopt the notation of Sec. 5.4 for the higher order cumulants. Exactly as in Sec. 5.4, when considering the

diagram of Fig. 7, two possible contributions to the free entropy at order 8 would be:

$$\frac{1}{N}\sum_{\substack{\mu_1,\mu_2,\mu_3\\ \text{pairwise distincts}}}\sum_{\substack{i_1,i_2,i_3\\ \text{pairwise distincts}}} F_{\mu_1 i_1}F_{\mu_1 i_2}F_{\mu_2 i_2}F_{\mu_2 i_3}F_{\mu_3 i_3}F_{\mu_3 i_1}F^2_{\mu_1 i_3}\,(v^h_{\mu_1})^2 v^h_{\mu_2}v^h_{\mu_3}v^x_{i_1}v^x_{i_2}(v^x_{i_3})^2, \tag{180}$$

$$\frac{1}{N}\sum_{\substack{\mu_1,\mu_2,\mu_3\\ \text{pairwise distincts}}}\sum_{\substack{i_1,i_2,i_3\\ \text{pairwise distincts}}} F_{\mu_1 i_1}F_{\mu_1 i_2}F_{\mu_2 i_2}F_{\mu_2 i_3}F_{\mu_3 i_3}F_{\mu_3 i_1}F^2_{\mu_1 i_3}\,\kappa^{(4,h)}_{\mu_1} v^h_{\mu_2}v^h_{\mu_3}v^x_{i_1}v^x_{i_2}\kappa^{(4,x)}_{i_3}. \tag{181}$$

The assumptions we need to make in order to deal with these diagrams are very similar to A.1 and A.2, and we state them here for completeness:

B.1 From the construction of the diagrams, odd moments of order greater or equal to 3, like $\kappa^{(3,x)}$, only appear in *non-Eulerian* graphs. By R.2 we know that these diagrams (without the moments as factors) are negligible. We assume that the possible correlations of the higher order moments with the matrix elements $\{F_{\mu i}\}$ are not strong enough to yield thermodynamically relevant corrections to the free entropy.

B.2 Eulerian strongly irreducible diagrams that are not simple cycles are negligible by R.3. We assume that the higher order (even) moments that appear as additional factors do not change their scaling, so that they remain negligible in the thermodynamic limit.

5.6 A note on i.i.d. matrices

We make here a side comment on i.i.d. rectangular matrices. We consider a random matrix $F \in \mathbb{R}^{M\times N}$ whose elements $\{F_{\mu i}\}$ are taken i.i.d., such that $\sqrt{N}F_{\mu i}$ is drawn from a given probability measure ρ. We assume that ρ has zero mean and finite moments of all orders. These matrices appear in our study of the GAMP algorithm in Sec. 4.1. Except if ρ is a Gaussian probability measure, the matrix F is not rotationally invariant, in the sense that it does not satisfy Model R. However, one can still derive strong results on the diagrammatics of F. We still assume B.1 and B.2, that is we assume that the additional factors in the diagrams do not change the scaling of a negligible diagram enough to make it thermodynamically relevant. It is then easy to see that because the $\{F_{\mu i}\}$ are uncorrelated, *all diagrams with order $p \geq 3$ are negligible in the $N \to \infty$ limit.* The only diagram that remains in the $N \to \infty$ limit is:



$$= \frac{1}{N}\sum_{\mu,i} F^2_{\mu i} v^h_\mu v^x_i. \tag{182}$$

In particular, we can only retain this diagram and apply the results of our Plefka expansions in this case as well, despite the fact that F is not rotationally invariant.

Acknowledgments

The authors would like to thank Yoshiyuki Kabashima, Marylou Gabrié, Bertrand Eynard and Jorge Kurchan for many insightful discussions. This work is supported by “Investissements d’Avenir" LabEx PALM (ANR-10-LABX-0039-PALM) (EquiDystant project, L. Foini), as well as by the French Agence Nationale de la Recherche under grant ANR-17-CE23-0023-01 PAIL, the European Union’s Horizon 2020 Research and Innovation Program 714608-SMiLe, and the ERC 307087 SPARCS. Additional funding is acknowledged by AM from ‘Chaire de recherche sur les modèles et sciences des données’, Fondation CFM pour la Recherche-ENS.

References

[AFP16] Ada Altieri, Silvio Franz, and Giorgio Parisi. The jamming transition in high dimension: an analytical study of the TAP equations and the effective thermodynamic potential. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(9):093301, 2016.

[AGZ10] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118. Cambridge university press, 2010.

[Alt18] Ada Altieri. Higher-order corrections to the effective potential close to the jamming transition in the perceptron model. *Physical Review E*, 97(1):012103, 2018.

[BBJ19] M. Bauer, D. Bernard, and T. Jin. Equilibrium Fluctuations in Maximally Noisy Extended Quantum Systems. *SciPost Phys.*, 6:45, 2019.

[BKM$^+$19] Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová. Optimal errors and phase transitions in high-dimensional generalized linear models. *Proceedings of the National Academy of Sciences*, page 201802705, 2019.

[CDL03] Raphaël Cherrier, David S Dean, and Alexandre Lefèvre. Role of the interaction matrix in mean-field spin glass models. *Physical Review E*, 67(4):046112, 2003.

[ÇO19] Burak Çakmak and Manfred Opper. Convergent dynamics for solving the tap equations of ising models with arbitrary rotation invariant coupling matrices. *arXiv preprint arXiv:1901.08583*, 2019.

[ÇOFW16] Burak Çakmak, Manfred Opper, Bernard H Fleury, and Ole Winther. Self-averaging expectation propagation. *arXiv preprint arXiv:1608.06602*, 2016.

[CŚ06] Benoît Collins and Piotr Śniady. Integration with respect to the Haar measure on unitary, orthogonal and symplectic group. *Communications in Mathematical Physics*, 264(3):773–795, 2006.

[CŚ07] Benoît Collins and Piotr Śniady. New scaling of Itzykson-Zuber integrals. In *Annales de l'IHP Probabilités et statistiques*, volume 43, pages 139–146, 2007.

[CWF14] Burak Cakmak, Ole Winther, and Bernard H Fleury. S-amp: Approximate message passing for general matrix ensembles. In *2014 IEEE Information Theory Workshop (ITW 2014)*, pages 192–196. IEEE, 2014.

[DAT78] JRL De Almeida and David J Thouless. Stability of the sherrington-kirkpatrick solution of a spin glass model. *Journal of Physics A: Mathematical and General*, 11(5):983, 1978.

[DMM09] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[Don06] David L Donoho. Compressed sensing. *IEEE Transactions on information theory*, 52(4):1289–1306, 2006.

[Fie82] James R Fienup. Phase retrieval algorithms: a comparison. *Applied optics*, 21(15):2758–2769, 1982.

[GM05] Alice Guionnet and M Maı. A Fourier view on the R-transform and related asymptotics of spherical integrals. *Journal of functional analysis*, 222(2):435–490, 2005.

[GY91] Antoine Georges and Jonathan S Yedidia. How to expand around mean-field theory using high-temperature expansions. *Journal of Physics A: Mathematical and General*, 24(9):2173, 1991.

[HC57] Harish-Chandra. Differential operators on a semisimple lie algebra. *American Journal of Mathematics*, pages 87–120, 1957.

[HK13] Haiping Huang and Yoshiyuki Kabashima. Adaptive Thouless-Anderson-Palmer approach to inverse Ising problems with quenched random fields. *Physical Review E*, 87(6):062129, 2013.

[Hop82] John J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, 79(8):2554–2558, 1982.

[IZ80] Claude Itzykson and J-B Zuber. The planar approximation. II. *Journal of Mathematical Physics*, 21(3):411–421, 1980.

[JR16] Hugo Jacquin and A Rançon. Resummed mean-field inference for strongly coupled data. *Physical Review E*, 94(4):042118, 2016.

[Kab08a] Yoshiyuki Kabashima. Inference from correlated patterns: a unified theory for perceptron learning and linear vector channels. In *Journal of Physics: Conference Series*, volume 95, page 012001. IOP Publishing, 2008.

[Kab08b] Yoshiyuki Kabashima. An integral formula for large random rectangular matrices and its application to analysis of linear vector channels. In *2008 6th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks and Workshops*, pages 620–624. IEEE, 2008.

[KMS+12] Florent Krzakala, Marc Mézard, Francois Sausset, Yifan Sun, and Lenka Zdeborová. Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(08):P08009, 2012.

[KMTZ14] Florent Krzakala, Andre Manoel, Eric W Tramel, and Lenka Zdeborová. Variational free energies for compressed sensing. In *2014 IEEE International Symposium on Information Theory*, pages 1499–1503. IEEE, 2014.

[KTJ76] JM Kosterlitz, DJ Thouless, and RC Jones. Spherical model of a spin-glass. *Physical Review Letters*, 36(20):1217, 1976.

[LBH15] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

[McC18] Peter McCullagh. *Generalized linear models*. Routledge, 2018.

[Meh04] Madan Lal Mehta. *Random matrices*, volume 142. Elsevier, 2004.

[Méz17] Marc Mézard. Mean-field message-passing equations in the Hopfield model and its generalizations. *Physical Review E*, 95(2):022117, 2017.

[Min01] Thomas P Minka. Expectation propagation for approximate Bayesian inference. In *Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence*, pages 362–369. Morgan Kaufmann Publishers Inc., 2001.

[MM09] Marc Mézard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[MP67] Vladimir Alexandrovich Marchenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Matematicheskii Sbornik*, 114(4):507–536, 1967.

[MP17] Junjie Ma and Li Ping. Orthogonal amp. *IEEE Access*, 5:2020–2033, 2017.

[MPR94a] Enzo Marinari, Giorgio Parisi, and Felix Ritort. Replica field theory for deterministic models: I. binary sequences with low autocorrelation. *Journal of Physics A: Mathematical and General*, 27(23):7615, 1994.

[MPR94b] Enzo Marinari, Giorgio Parisi, and Felix Ritort. Replica field theory for deterministic models. II. a non-random spin glass with glassy behaviour. *Journal of Physics A: Mathematical and General*, 27(23):7647, 1994.

[MPV87] Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

[NT97] Kazuo Nakanishi and Hajime Takayama. Mean-field theory for a spin-glass model of neural networks: TAP free energy and the paramagnetic to spin-glass transition. *Journal of Physics A: Mathematical and General*, 30(23):8085, 1997.

[NW72] John Ashworth Nelder and Robert WM Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, 135(3):370–384, 1972.

[OCW16] Manfred Opper, Burak Cakmak, and Ole Winther. A theory of solving TAP equations for Ising models with general invariant random matrices. *Journal of Physics A: Mathematical and Theoretical*, 49(11):114002, 2016.

[OS01] Manfred Opper and David Saad. *Advanced mean field methods: Theory and practice*. MIT press, 2001.

[OW01a] Manfred Opper and Ole Winther. Adaptive and self-averaging Thouless-Anderson-Palmer mean-field theory for probabilistic modeling. *Physical Review E*, 64(5):056131, 2001.

[OW01b] Manfred Opper and Ole Winther. Tractable approximations for probabilistic models: The adaptive Thouless-Anderson-Palmer mean field approach. *Physical Review Letters*, 86(17):3695, 2001.

[OW05a] Manfred Opper and Ole Winther. Expectation consistent approximate inference. *Journal of Machine Learning Research*, 6(Dec):2177–2204, 2005.

[OW05b] Manfred Opper and Ole Winther. Expectation consistent free energies for approximate inference. In *Advances in Neural Information Processing Systems*, pages 1001–1008, 2005.

[Ple82] Timm Plefka. Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. *Journal of Physics A: Mathematical and general*, 15(6):1971, 1982.

[PP95] Giorgio Parisi and Marc Potters. Mean-field equations for spin models with orthogonal interaction matrices. *Journal of Physics A: Mathematical and General*, 28(18):5267, 1995.

[Ran11] Sundeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2168–2172. IEEE, 2011.

[RSF17] Sundeep Rangan, Philip Schniter, and Alyson K Fletcher. Vector approximate message passing. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 1588–1592. IEEE, 2017.

[SK08] Takashi Shinzato and Yoshiyuki Kabashima. Perceptron capacity revisited: classification ability for correlated patterns. *Journal of Physics A: Mathematical and Theoretical*, 41(32):324013, 2008.

[SRF16] Philip Schniter, Sundeep Rangan, and Alyson K Fletcher. Vector approximate message passing for the generalized linear model. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 1525–1529. IEEE, 2016.

[TAP77] David J Thouless, Philip W Anderson, and Robert G Palmer. Solution of'solvable model of a spin glass'. *Philosophical Magazine*, 35(3):593–601, 1977.

[TGM+18] Eric W Tramel, Marylou Gabrié, Andre Manoel, Francesco Caltagirone, and Florent Krzakala. Deterministic and generalized framework for unsupervised learning with restricted boltzmann machines. *Physical Review X*, 8(4):041006, 2018.

[TV04] Antonia M Tulino and Sergio Verdú. Random matrix theory and wireless communications. *Foundations and Trends® in Communications and Information Theory*, 1(1):1–182, 2004.

[WJ08] Martin J. Wainwright and Michael I. Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008.

[YFW03] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Understanding belief propagation and its generalizations. *Exploring artificial intelligence in the new millennium*, 8:236–239, 2003.

[ZK16] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.

A The Georges-Yedidia formalism

In this section we recall the formalism of [GY91] which allows to systematically expand the free entropy around $\beta = 0$, at fixed values of the first and second moments of the variables. We consider here a generic Hamiltonian $H_J(\mathbf{x})$, with variables $(x_1, \cdots, x_N)$. We fix the first and second moments $\langle x_i \rangle_\beta = m_i$ and $\langle (x_i - m_i)^2 \rangle_\beta = v_i$, using Lagrange parameters respectivelly denoted $\lambda_i(\beta)$ and $\gamma_i(\beta)$. Recall that $\langle \cdot \rangle_\beta$ stands for the expectation over the Gibbs measure at inverse temperature β, constrained by the Lagrange multipliers λ_i and γ_i. From now on, we will drop the β supbscript.

We introduce the operator U from Appendix A of [GY91]:

$$U(\beta, J) = H_J - \langle H_J \rangle + \sum_{i=1}^{N} \partial_\beta \lambda_i(\beta)(x_i - m_i) + \frac{1}{2} \sum_{i=1}^{N} \partial_\beta \gamma_i(\beta) \left[x_i^2 - v_i - m_i^2 \right]. \tag{183}$$

Then the derivative of the thermal average of any observable O is given by

$$\frac{\partial \langle O \rangle}{\partial \beta} = \left\langle \frac{\partial O}{\partial \beta} \right\rangle - \langle OU \rangle. \tag{184}$$

As the Lagrange multipliers λ_i and γ_i have been introduced to fix the average of x_i and its variance one has the following easy identity, valid at any β:

$$\langle U \rangle = 0. \tag{185}$$

Moreover, given that the magnetizations $\{m_i\}$ and the variances $\{v_i\}$ do not depend on β one has:

$$0 = \frac{\partial \langle x_i \rangle}{\partial \beta} = -\langle x_i U \rangle = -\langle (x_i - m_i) U \rangle. \tag{186}$$

$$0 = \frac{\partial \langle x_i^2 \rangle}{\partial \beta} = -\langle x_i^2 U \rangle = -\langle (x_i^2 - v_i - m_i^2) U \rangle. \tag{187}$$

Considering the previous results one can compute the derivative of U:

$$\begin{aligned} \frac{\partial U}{\partial \beta} &= \langle H_J U \rangle + \sum_{i=1}^{N} \partial_\beta^2 \lambda_i(\beta)(x_i - m_i) + \frac{1}{2} \sum_{i=1}^{N} \partial_\beta^2 \gamma_i(\beta) \left[x_i^2 - v_i - m_i^2 \right], \\ &= \langle U^2 \rangle + \sum_{i=1}^{N} \partial_\beta^2 \lambda_i(\beta)(x_i - m_i) + \frac{1}{2} \sum_{i=1}^{N} \partial_\beta^2 \gamma_i(\beta) \left[x_i^2 - v_i - m_i^2 \right]. \end{aligned} \tag{188}$$

Equipped with these relations one can compute the derivatives of the free entropy up to fourth order. Recall

that Φ_J is the *intensive* free entropy of the system. We obtain its derivatives:

$$\frac{\partial N\Phi_J}{\partial\beta} = -\langle H_J\rangle + \sum_{i=1}^{N}\partial_\beta\lambda_i(\beta)\langle(x_i-m_i)\rangle + \frac{1}{2}\sum_{i=1}^{N}\partial_\beta\gamma_i(\beta)\langle\left[x_i^2-v_i-m_i^2\right]\rangle = -\langle H_J\rangle, \tag{189}$$

$$\frac{\partial^2 N\Phi_J}{\partial\beta^2} = \langle H_J U\rangle = \langle U^2\rangle\,, \tag{190}$$

$$\frac{\partial^3 N\Phi_J}{\partial\beta^3} = -\langle U^3\rangle + 2\,\langle U\frac{\partial U}{\partial\beta}\rangle = -\langle U^3\rangle\,, \tag{191}$$

$$\begin{aligned}\frac{\partial^4 N\Phi_J}{\partial\beta^4} &= \langle U^4\rangle - 3\,\langle U^2\frac{\partial U}{\partial\beta}\rangle = \langle U^4\rangle - 3\,\langle U^2\rangle^2 - 3\sum_{i=1}^{N}\partial_\beta^2\lambda_i(\beta)\,\langle U^2(x_i-m_i)\rangle \\ &\quad - \frac{3}{2}\sum_{i=1}^{N}\partial_\beta^2\gamma_i(\beta)\,\langle U^2\left[x_i^2-v_i-m_i^2\right]\rangle\,.\end{aligned} \tag{192}$$

These relations are valid at any inverse temperature β ! In the main sections we derive the explicit expression of the operator U for our particular choice of Hamiltonian, and we will use these relations (and show how to conjecture their higher order counterparts) to compute the expansion of the free entropy around $\beta = 0$.

B Order 4 of the Plefka expansion for Sec. 2.1.

We start from eq. (192) in Appendix A, that we consider at $\beta = 0$:

$$N\frac{\partial^4\Phi_J(\beta)}{\partial\beta^4} = \langle U^4\rangle_0 - 3\,\langle U^2\rangle_0^2 - 3\sum_{i=1}^{N}\partial_\beta^2\lambda_i\,\langle U^2(x_i-m_i)\rangle_0 - \frac{3}{2}\sum_{i=1}^{N}\partial_\beta^2\gamma_i\,\langle U^2(x_i^2-m_i^2-v_i)\rangle_0\,. \tag{193}$$

For simplicity we will denote $\tilde{x}_i \equiv (x_i - m_i)$, so that at $\beta = 0$ the $\{\tilde{x}_i\}$ variables are Gaussian variables with mean $\langle\tilde{x}_i\rangle = 0$ and covariance $\langle\tilde{x}_i\tilde{x}_j\rangle = \delta_{ij}v_i$. In particular eq. (21) becomes:

$$U = -\frac{1}{2}\sum_{i\neq j}J_{ij}\tilde{x}_i\tilde{x}_j. \tag{194}$$

From the calculation at order 2 we obtain the following relation that we can represent diagrammatically:

$$-3\,\langle U^2\rangle_0^2 = -\frac{3}{4}\left[\sum_{i\neq j}J_{ij}^2 v_i v_j\right]^2 = -\frac{3N}{4}\Big[\ \Big]^2. \tag{195}$$

We now turn to the next term:

$$
\begin{aligned}
&-\frac{3}{2}\sum_{i=1}^{N}\partial_\beta^2\gamma_i\left\langle U^2(x_i^2-m_i^2-v_i)\right\rangle_0-3\sum_{i=1}^{N}\partial_\beta^2\lambda_i\left\langle U^2(x_i-m_i)\right\rangle_0\\
&=-\frac{3}{2}\sum_{i=1}^{N}\partial_\beta^2\gamma_i\left\langle U^2(\tilde{x}_i^2-v_i)\right\rangle_0-3\sum_{i=1}^{N}\left\langle U^2\tilde{x}_i\left(\partial_\beta^2\lambda_i+m_i\partial_\beta^2\gamma_i\right)\right\rangle_0,\\
&\overset{(a)}{=}-\frac{3}{2}\sum_{i=1}^{N}\partial_\beta^2\gamma_i\left\langle U^2(\tilde{x}_i^2-v_i)\right\rangle_0-3N\sum_{i=1}^{N}\left\langle U^2\tilde{x}_i\frac{\partial}{\partial m_i}\left(\partial_\beta^2\Phi_J\right)\right\rangle_0,\\
&\overset{(b)}{=}-\frac{3}{2}\sum_{i=1}^{N}\partial_\beta^2\gamma_i\left\langle U^2(\tilde{x}_i^2-v_i)\right\rangle_0+\mathcal{O}_N(1).
\end{aligned}
\tag{196}
$$

In (a) we used the Maxwell equation eq. (20), while in (b) we made use of the fact that the order 2 of the free entropy does not depend on the m_i variables. We obtain:

$$
-\frac{3}{2}\sum_{i=1}^{N}\partial_\beta^2\gamma_i\left\langle U^2(\tilde{x}_i^2-v_i)\right\rangle_0=\frac{3}{2}\left[\sum_{i\neq j}J_{ij}^2v_iv_j\right]^2-3\sum_{i\neq j}J_{ij}^2v_j\left\langle U^2\tilde{x}_i^2\right\rangle_0,
$$

in which we used the Maxwell relation eq. (19) to compute $\partial_\beta^2\gamma_i$. To compute $\langle U^2\tilde{x}_i^2\rangle_0$, we expand:

$$
\left\langle U^2\tilde{x}_i^2\right\rangle_0=\frac{1}{4}\sum_{i_1\neq j_1}\sum_{i_2\neq j_2}J_{i_1j_1}J_{i_2j_2}\left\langle\tilde{x}_i^2\tilde{x}_{i_1}\tilde{x}_{j_1}\tilde{x}_{i_2}\tilde{x}_{j_2}\right\rangle_0.
$$

We can then use Wick's theorem to simplify the average. There are two types of contractions (or pairings):

- Contractions that do not mix indices i_1, j_1, i_2, j_2 with i. There are 2 such possible pairings and in $\frac{\partial^4\Phi_J}{\partial\beta^4}$ they give rise to the diagram $\Big[N$  $\Big]^2$.
- Contractions that mix these indices with i. There are all equivalent and there are 8 of them, which gives rise to the diagram: N  .

In the end, we reach:

$$
\left\langle U^2\tilde{x}_i^2\right\rangle_0=\frac{v_i}{2}\sum_{k\neq l}J_{kl}^2v_kv_l+2\sum_{k(\neq i)}J_{ik}^2v_i^2v_k.
$$

We can finally compute the term we were seeking:

$$
-\frac{3}{2}\sum_{i=1}^{N}\partial_\beta^2\gamma_i\left\langle U^2(\tilde{x}_i^2-v_i)\right\rangle_0=-6N\sum_{\substack{i,j,k\\ \text{pairwise distincts}}}J_{ij}^2J_{ik}^2v_i^2v_jv_k=-6
$$  .
(197)

Note that in this last equation we could add the hypothesis that $j\neq k$. Indeed the term $j=k$ would give rise to the diagram N  , which is negligible since for every $i\neq j$ one has $J_{ij}=\mathcal{O}(\frac{1}{\sqrt{N}})$ as a consequence of rotational invariance (Model S). We finally turn to the computation of $\langle U^4\rangle_0$:

$$
\left\langle U^4\right\rangle_0=\frac{1}{16}\prod_{\alpha=0}^{3}\left[\sum_{i_\alpha\neq j_\alpha}J_{i_\alpha j_\alpha}\right]\left\langle\prod_{\alpha=0}^{3}\tilde{x}_{i_\alpha}\tilde{x}_{j_\alpha}\right\rangle_0.
$$



(a) (b)

Figure 8: Different types of diagrams of indices appearing in $\langle U^4 \rangle_0$

The possible contractions arising from Wick's theorem yield several contributions, that we can represent by diagrams. Note that these diagrams are very different from the diagrams that we described for instance in Fig. 1, and are merely a way to visualize the contractions in Wick's theorem. The first column contains the i_α indices and the second contains the j_α. Note that we always have $i_\alpha \neq j_\alpha$. The two different types of contractions are represented as Fig. 8a and Fig. 8b. They are 12 possible contractions of the type of Fig. 8a and 48 of Fig. 8b. We also take into account that in the pairings of Fig. 8b indices are not all necessarily pairwise distinct. Discarding terms that are $\mathcal{O}_N(N)$, we finally reach:

$$\langle U^4 \rangle_0 = \frac{3}{4}\left[\sum_{i\neq j} J_{ij}^2 v_i v_j\right]^2 + 6 \sum_{\substack{i,j,k \\ \text{pairwise distincts}}} J_{ij}^2 J_{ik}^2 v_i^2 v_j v_k + 3 \sum_{\substack{i_0,i_1,i_2,i_3 \\ \text{pairwise distincts}}} J_{i_0 i_1} J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_0} v_{i_0} v_{i_1} v_{i_2} v_{i_3},$$

$$= \frac{3}{4}\Big[N \quad \Big]^2 + 6N \quad . + 3N \quad . \tag{198}$$



Finally, combining eq. (195), eq. (197), and eq. (198) to plug them into eq. (193), we reach:

$$\frac{1}{4!}\frac{\partial^4 \Phi_J}{\partial \beta^4} = \frac{1}{8N} \sum_{\substack{i_0,i_1,i_2,i_3 \\ \text{pairwise distincts}}} J_{i_0 i_1} J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_0} v_{i_0} v_{i_1} v_{i_2} v_{i_3} + \mathcal{O}_N(1), \tag{199}$$

$$= \frac{1}{8} \quad , \tag{200}$$



which is what we wanted to show !

C Some definitions and reminders of random matrix theory

For a complete mathematical introduction to random matrix theory, the reader can refer to [Meh04, AGZ10], while a more practical approach is carried out in [TV04]. Let us consider a compactly supported probability measure μ on $\mathbb{R}$. We denote $\lambda_{\max} \equiv \max \operatorname{supp}(\mu)$ and $\lambda_{\min} \equiv \min \operatorname{supp}(\mu)$. One can introduce the *Stieltjes* transform of μ as:

$$\mathcal{S}_\mu(z) \equiv \mathbb{E}\left[\frac{1}{X - z}\right] = \int_{\mathbb{R}} \mu(\mathrm{d}\lambda) \frac{1}{\lambda - z}. \tag{201}$$

On $(\lambda_{\max}, +\infty)$, $\mathcal{S}_\mu$ induces a strictly increasing $\mathcal{C}^\infty$ diffeomorphism $\mathcal{S}_\mu : (\lambda_{\max}, \infty) \hookrightarrow (-\infty, 0)$, and we denote its inverse $\mathcal{S}_\mu^{-1}$. One can then introduce the $\mathcal{R}$-*transform* of μ as:

$$\forall z > 0, \quad \mathcal{R}_\mu(z) \equiv \mathcal{S}_\mu^{-1}(-z) - \frac{1}{z}. \tag{202}$$

$\mathcal{R}_\mu(z)$ is *a priori* defined for $-z \in \mathcal{S}_\mu\left[(\lambda_{\min}, \lambda_{\max})^c\right]$ and admits an analytical expansion around $z = 0$. We can write this expansion as:

$$\mathcal{R}_\mu(z) = \sum_{k=0}^{\infty} c_{k+1}(\mu)\, z^k. \tag{203}$$

The elements of the sequence $\{c_k(\mu)\}_{k\in\mathbb{N}^\star}$ are called the *free cumulants* of μ. In particular, one can show that $c_1(\mu) = \mathbb{E}_\mu(X)$ and $c_2(\mu) = \mathbb{E}_\mu(X^2) - (\mathbb{E}_\mu X)^2$. The free cumulants can be recursively computed from the moments of the measure using the so-called *free cumulant equation*:

$$\forall k \in \mathbb{N}^*, \quad \mathbb{E}_\mu X^k = \sum_{m=1}^{k} c_m(\mu) \sum_{\substack{\{k_i\}_{i\in[|1,m|]} \\ \text{s.t } \sum_i k_i = k}} \prod_{i=1}^{m} \mathbb{E}_\mu X^{k_i - 1}. \tag{204}$$

For practical purposes, for all $x \in (-\mathcal{S}_\mu(\lambda_{\min}), -\mathcal{S}_\mu(\lambda_{\max}))$ we can define:

$$G_\mu(x) \equiv \frac{1}{2} \int_0^x \mathrm{d}u\, \mathcal{R}_\mu(u). \tag{205}$$

D Technical derivations and generalizations of the diagrammatics

We detail here some extensions of the results of Sec. 5. In Sec. D.1, we explain how to transpose these results to Hermitian matrix models, and in Sec. D.2 we show how to extend some of them to diagrams of diverging size (as $N \to \infty$).

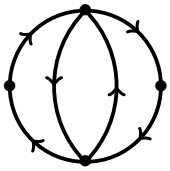
D.1 Hermitian matrix model

One can generalize the results of Sec. 5 to the following Hermitian matrix model, similar to Model S:

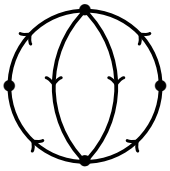
Model 3. *Let $N \geq 1$ and $\mathcal{U}(N)$ be the unitary group. Let $J \in \mathbb{C}^{N\times N}$ be a random matrix generated as $J = UDU^\dagger$ with $U \in \mathcal{U}(N)$ drawn uniformly and independently from D. D is a real diagonal matrix such that its empirical spectral distribution $\rho_D^{(N)} \equiv \frac{1}{N}\sum_{i=1}^N \delta_{d_i}$ converges (almost surely) as $N \to \infty$ a.s. to a probability distribution ρ_D with compact support. The smallest and largest eigenvalue of D are assumed to converge almost surely to the infimum and supremum of the support of ρ_D.*

Note that the diagrams are now directed, as $J_{ij} = \overline{J_{ji}}$. We describe such diagrams in Fig. 9. For instance, the diagram of Fig. 9a is equal to:

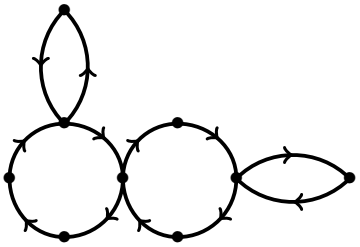
$$\frac{1}{N} \sum_{\substack{i_1,\cdots,i_4 \\ \text{pairwise distincts}}} J_{i_1 i_2} J_{i_2 i_3} J_{i_3 i_4} J_{i_4 i_1} |J_{i_2 i_4}|^2, \tag{206}$$



(a) A directed Eulerian diagram for Hermitian matrices.



(b) A different Eulerian diagram, but with the same undirected edges.



(c) An Eulerian directed cactus diagram. A cactus diagram is Eulerian if and only if all its simple cycles are Eulerian.

Figure 9: Diagrams similar to the ones of Fig. 4, but for Hermitian matrices. Note that the diagrams of Fig. 9a and Fig. 9b are different because of the different directions of the edges, but that both are Eulerian.

while the diagram of Fig. 9b represents the quantity:

$$\frac{1}{N} \sum_{\substack{i_1,\cdots,i_4 \\ \text{pairwise distincts}}} J_{i_1 i_2} \overline{J_{i_2 i_3} J_{i_3 i_4}} J_{i_4 i_1} J_{i_2 i_4}^2. \tag{207}$$

In the complex case, an *Eulerian* graph is similarly defined as a graph in which one can construct a cyclic path (following the directions of the edges) that visits each edge exactly once. Note that a *simple cycle* is defined such that the arrows on its edges themselves form a cycle, like the constituent cycles of Fig. 9c. We describe the main results we get, using the same kind of techniques as used in Sec. 5.1:

- (i) Only Eulerian diagrams contribute in the $N \to \infty$ limit.
- (ii) Consider a simple cycle $\mathcal{C}_p$ with p vertices. Then this diagram converges in the $N \to \infty$ limit to the free cumulant $c_p(\rho_D)$ in L^2 norm, as in the real case. More precisely:

$$\lim_{N\to\infty} \mathbb{E} \left| \frac{1}{N} \sum_{\substack{i_1,\cdots,i_p \\ \text{pairwise distincts}}} \left(UDU^\dagger\right)_{i_1 i_2} \left(UDU^\dagger\right)_{i_2 i_3} \cdots \left(UDU^\dagger\right)_{i_p i_1} - c_p(\rho_D) \right|^2 = 0. \tag{208}$$

- (iii) Any Eulerian strongly irreducible diagram that is not a simple cycle will be negligible in the $N \to \infty$ limit (in L^2 norm).
- (iv) Any Eulerian cactus diagram (like in Fig. 9c) will converge in L^2 to the products of the free cumulants of ρ_D corresponding to each one of its constituent simple cycles.

These results are straightforward generalizations of the ones obtained for real matrices in Sec. 5. For completeness, we describe how to show a weaker version of (ii), and leave other statements as easy generalizations of Sec. 5. Let us now show that the limit of the expectation of the term in (ii) is the free cumulant, as in Sec. 5.1. As before, by unitary invariance we can assume that $(i_1, \cdots, i_p) = (1, , \cdots, p)$, and we can apply the results of [GM05] to obtain a similar equation to eq. (155):

$$\begin{aligned} L_p &\equiv \lim_{N\to\infty} N^{p-1} \mathbb{E}\left[(UDU^\dagger)_{12} \cdots (UDU^\dagger)_{p1}\right], \\ &= \lim_{N\to\infty} \frac{1}{N} \prod_{l=1}^{p} \left[\frac{\partial}{\partial b_i} + i\frac{\partial}{\partial c_i}\right] \left[\exp\left\{N \sum_{n=1}^{\infty} \frac{c_n(\rho_D)}{n} \mathrm{Tr}\left[M(\mathbf{b},\mathbf{c})^n\right]\right\}\right]_{\mathbf{b},\mathbf{c}=0}, \end{aligned} \tag{209}$$

with now the matrix $M(\mathbf{b},\mathbf{c})$ defined as:

$$M(\mathbf{b},\mathbf{c}) \equiv \frac{1}{2}\begin{pmatrix} 0 & b_1+ic_1 & 0 & \cdots & 0 & b_p-ic_p \\ b_1-ic_1 & 0 & b_2+ic_2 & \cdots & 0 & 0 \\ 0 & b_2-ic_2 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & b_{p-1}+ic_{p-1} \\ b_p+ic_p & 0 & 0 & \cdots & b_{p-1}-ic_{p-1} & 0 \end{pmatrix}. \tag{210}$$

Now, we have $\left[\frac{\partial}{\partial b_i}+i\frac{\partial}{\partial c_i}\right] M(\mathbf{b},\mathbf{c}) = F_{i+1,i}$, in which $(F_{a,b})_{ll'} \equiv \delta_{al}\delta_{bl'}$ are elementary non-symmetric matrices. In the exact same way as in Sec. 5.1, the dominant contribution in eq. (209) will be given by differentiating a single time the exponential term, and creating a cycle with the matrices $F_{i+1,i}$. Note that contrary to the symmetric case of Sec. 5.1, here *only the directed cycle will contribute*, whereas both possible directions of the cycle contributed in eq. (159). Indeed, the cycles in terms of the matrices $\{F_{a,b}\}$ have to be directed in order to yield a non-zero contribution:

$$\begin{aligned} &\mathrm{Tr}\ [F_{1,2}F_{2,1}F_{1,3}F_{3,2}F_{2,1}] \neq 0, \\ &\mathrm{Tr}\ [F_{1,2}F_{2,1}F_{2,3}F_{3,2}F_{2,1}] = 0. \end{aligned}$$

Thus we have:

$$\begin{aligned} L_p &= \sum_{n=p}^{\infty} c_n(\rho_D)\mathrm{Tr}\ \left[(F_{1,p}F_{p,p-1}\cdots F_{2,1})\ M(0,0)^{n-p}\right], \\ &= c_p(\rho_D). \end{aligned}$$

In order to get L^2 concentration of the simple cycle on the free cumulant, one can exactly repeat the arguments of Sec. 5.3.

D.2 A note on the expectation of diagrams of diverging size

Although it is not directly useful in our Plefka expansions, another side question one can ask on the behavior of these diagrams is: how do diagrams that have a number of edges that diverge with N behave in the $N \to \infty$ limit ? In all of Sec. 5 we only considered diagrams of finite size. The behavior of the HCIZ-type integrals with a matrix with diverging rank (as opposed to the finite-rank case) has been rigorously treated in [GM05] and then generalized in [CŚ07] as soon as the rank of the matrix diverges sub linearly in N. We recall the main result of [CŚ07]:

Theorem 3 (Collins-Śniadyc). *Let A_N, B_N be diagonal real matrices of size N. Assume that the rank $M(N)$ of A_N is such that $M(N) = o(N)$, and denote $a_{1,N} \geq \cdots \geq a_{M,N}$ the eigenvalues of A_N. Assume that the spectral measure of B_N converges a.s. and in the weak sense to a probability measure ρ_B, and that all elements of A_N are bounded by a constant independent of N. Then one has:*

$$\frac{1}{NM(N)}\log\int_{\mathcal{U}(N)} \mathcal{D}U\, e^{N\mathrm{Tr}\left[A_N U B_N U^\dagger\right]} = \frac{2}{M(N)}\mathrm{Tr}\left[G_{\rho_B}\left(A_N\right)\right] + o_N(1). \tag{211}$$

A similar result holds for real orthogonal matrices:

$$\frac{1}{NM(N)}\log\int_{\mathcal{O}(N)} \mathcal{D}O\, e^{\frac{N}{2}\mathrm{Tr}\left[A_N O B_N O^\intercal\right]} = \frac{1}{M(N)}\mathrm{Tr}\left[G_{\rho_B}\left(A_N\right)\right] + o_N(1). \tag{212}$$

The techniques of Sec. 5 thus generalize to this case. We consider real symmetric matrices under Model S (in the Hermitian case, the results also generalize following the line of Appendix D.1). We say that a sequence $\{p(N)\}$ satisfies the *bounded free cumulant property* if it satisfies the following:

Property 1. *There exists $C > 0$ such that for all $N \in \mathbb{N}$, $|c_{p(N)}(\rho_D)| < C$.*

We state two of the results of Sec. 5 that can be easily generalized to the diverging size case without changing any of the arguments:

(a) Consider a sequence $p(N) = \mathcal{O}_N(N)$ that satisfies the bounded free cumulant property. Then one obtains the generalization of eq. (150):

$$N^{p(N)-1} \int_{\mathcal{O}(N)} \mathcal{D}O \left[(ODO^\intercal)_{12} (ODO^\intercal)_{23} \cdots (ODO^\intercal)_{p(N)1} \right] = c_{p(N)}(\rho_D) + \mathcal{O}_N(1) \,. \tag{213}$$

(c) Consider a cactus diagram G composed of $P(N)$ simple cycles of size $(r_1(N), \cdots, r_P(N))$, joining at vertices. Assume that $\sum_{i=1}^{P(N)} r_i(N) = \mathcal{O}_N(N)$ and that all the sequences $r_i(N)$ satisfy the bounded free cumulant property. Then one has:

$$\mathbb{E}\, G = \left[\prod_{i=1}^{P(N)} c_{r_i(N)}(\rho_D) \right] (1 + \mathcal{O}_N(1)) \,. \tag{214}$$

Other results obtained in Sec. 5 for finite-size diagrams might also be applicable to the diverging size case, and we leave them for future work.