



**HAL**  
open science

# Rank-one matrix estimation: analysis of algorithmic and information theoretic limits by the spatial coupling method

Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, Lenka Zdeborova

## ► To cite this version:

Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, Lenka Zdeborova. Rank-one matrix estimation: analysis of algorithmic and information theoretic limits by the spatial coupling method. 2019. cea-02009706

**HAL Id: cea-02009706**

**<https://cea.hal.science/cea-02009706v1>**

Preprint submitted on 6 Feb 2019

**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.

# Rank-one matrix estimation: analysis of algorithmic and information theoretic limits by the spatial coupling method

Jean Barbier<sup>†\*</sup>, Mohamad Dia<sup>†</sup> and Nicolas Macris<sup>†</sup> FIRSTNAME.LASTNAME@EPFL.CH

<sup>†</sup> *Laboratoire de Théorie des Communications, Faculté Informatique et Communications, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.*

\* *International Center for Theoretical Physics, 11 Strada Costiera, I - 34151, Trieste, Italy.*

**Florent Krzakala**

FLORENT.KRZAKALA@ENS.FR

*Laboratoire de Physique Statistique, CNRS, PSL Universités et Ecole Normale Supérieure, Sorbonne Universités et Université Pierre & Marie Curie, 75005, Paris, France.*

**Lenka Zdeborová**

LENKA.ZDEBOROVA@GMAIL.COM

*Institut de Physique Théorique, CNRS, CEA, Université Paris-Saclay, F-91191, Gif-sur-Yvette, France.*

**Editor:**

## Abstract

Factorizing low-rank matrices is a problem with many applications in machine learning and statistics, ranging from sparse PCA to community detection and sub-matrix localization. For probabilistic models in the Bayes optimal setting, general expressions for the mutual information have been proposed using powerful heuristic statistical physics computations via the replica and cavity methods, and proven in few specific cases by a variety of methods. Here, we use the spatial coupling methodology developed in the framework of error correcting codes, to rigorously derive the mutual information for the symmetric rank-one case. We characterize the detectability phase transitions in a large set of estimation problems, where we show that there exists a gap between what currently known polynomial algorithms (in particular spectral methods and approximate message-passing) can do and what is expected information theoretically. Moreover, we show that the computational gap vanishes for the proposed spatially coupled model, a promising feature with many possible applications. Our proof technique has an interest on its own and exploits three essential ingredients: the interpolation method first introduced in statistical physics, the analysis of approximate message-passing algorithms first introduced in compressive sensing, and the theory of threshold saturation for spatially coupled systems first developed in coding theory. Our approach is very generic and can be applied to many other open problems in statistical estimation where heuristic statistical physics predictions are available.

**Keywords:** Sparse PCA, Wigner spike model, community detection, low-rank matrix estimation, spatial coupling, replica and cavity methods, interpolation method, approximate message-passing

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Setting and main results</b>	<b>5</b>

2.1	Basic underlying model . . . . .	5
2.2	AMP algorithm and state evolution . . . . .	6
2.3	Spatially coupled model . . . . .	8
2.4	Main results: basic underlying model . . . . .	9
2.5	Main results: coupled model . . . . .	12
2.6	Road map of the proof of the replica symmetric formula . . . . .	14
2.7	Connection with the planted Sherrington-Kirkpatrick spin glass model . . .	14
<b>3</b>	<b>Two Examples: spiked Wigner model and community detection</b>	<b>15</b>
3.1	Spiked Wigner model . . . . .	15
3.2	Asymmetric community detection . . . . .	17
<b>4</b>	<b>Threshold saturation</b>	<b>17</b>
4.1	State evolution of the underlying system . . . . .	18
4.2	State evolution of the coupled system . . . . .	20
4.3	Proof of Theorem 11 . . . . .	22
4.3.1	Upper bound on the potential variation under a shift . . . . .	22
4.3.2	Lower bound on the potential variation under a shift . . . . .	25
4.3.3	End of proof of threshold saturation . . . . .	26
<b>5</b>	<b>Invariance of the mutual information under spatial coupling</b>	<b>26</b>
5.1	A generic interpolation . . . . .	27
5.2	Applications . . . . .	30
<b>6</b>	<b>Proof of the replica symmetric formula (Theorem 4)</b>	<b>31</b>
6.1	Proof of Theorem 4 for $0 < \Delta \leq \Delta_{\text{opt}}$ . . . . .	32
6.2	Proof of Theorem 4 for $\Delta \geq \Delta_{\text{opt}}$ . . . . .	35
<b>7</b>	<b>Proof of corollaries 7 and 8</b>	<b>36</b>
7.1	Proof of Corollary 7 . . . . .	37
7.2	Proof of Corollary 8 . . . . .	38
<b>A</b>	<b>Upper bound on the mutual information</b>	<b>39</b>
<b>B</b>	<b>Relating the mutual information to the free energy</b>	<b>41</b>
<b>C</b>	<b>Proof of the I-MMSE relation</b>	<b>42</b>
<b>D</b>	<b>Nishimori identity</b>	<b>43</b>
<b>E</b>	<b>Proof of Lemmas 13 and 21</b>	<b>44</b>
<b>F</b>	<b>Analysis of <math>i_{\text{RS}}(E; \Delta)</math> for <math>\Delta \rightarrow 0</math></b>	<b>45</b>
<b>G</b>	<b>Proof of Proposition 30</b>	<b>45</b>

## 1. Introduction

We consider the following probabilistic rank-one matrix estimation (or rank-one matrix factorization) problem: one has access to noisy observations  $\mathbf{w} = (w_{ij})_{i,j=1}^n \in \mathbb{R}^{n,n}$  of the pair-wise product of the components of a vector  $\mathbf{s} = (s_i)_{i=1}^n \in \mathbb{R}^n$  where the components are i.i.d random variables distributed according to  $S_i \sim P_0$ ,  $i = 1, \dots, n$ . The matrix elements of  $\mathbf{w}$  are observed through a noisy element-wise (possibly non-linear) output probabilistic channel  $P_{\text{out}}(w_{ij}|s_i s_j)$ , with  $i, j = 1, \dots, n$ . The goal is to estimate the vector  $\mathbf{s}$  from  $\mathbf{w}$ , up to a global flip of sign in general, assuming that both distributions  $P_0$  and  $P_{\text{out}}$  are known. We assume the noise to be symmetric so that  $w_{ij} = w_{ji}$ . There are many important problems in statistics and machine learning that can be expressed in this way, among which:

- *Sparse PCA*: Sparse principal component analysis (PCA) is a dimensionality reduction technique where one looks for a low-rank representation of a data matrix with sparsity constraints [Zou et al. (2006)]. The following is the simplest probabilistic symmetric version where one estimates a rank-one matrix. Consider a sparse random vector  $\mathbf{S}$ , for instance drawn from a Gauss-Bernoulli distribution, and take an additive white Gaussian noise (AWGN) channel where the observations are  $W_{ij} = S_i S_j / \sqrt{n} + \Delta Z_{ij}$  with  $Z_{ij} \sim \mathcal{N}(0, 1)$ . Here<sup>1</sup>  $P_{\text{out}}(w_{ij}|s_i s_j) = \mathcal{N}(w_{ij}|s_i s_j / \sqrt{n}, \Delta)$ .
- *Spiked Wigner model*: In this model the noise is still Gaussian, but the vector  $\mathbf{S}$  is assumed to be a Bernoulli random vector with i.i.d components  $S_i \sim \text{Ber}(\rho)$ . This formulation is a particular case of the spiked covariance model in statistics introduced by [Johnstone and Lu (2004, 2012)]. It has also attracted a lot of attention in the framework of random matrix theory (see for instance [Baik et al. (2005)] and references therein).
- *Community detection*: In its simplest setting, one uses a Rademacher vector  $\mathbf{S}$  where each variable take values  $S_i \in \{-1, 1\}$  depending on the “community” it belongs to. The observation model then introduces missing information and errors such that, for instance,  $P_{\text{out}}(w_{ij}|s_i s_j) = p_1 \delta(w_{ij} - s_i s_j) + p_2 \delta(w_{ij} + s_i s_j) + (1 - p_1 - p_2) \delta(w_{ij})$ , where  $\delta(\cdot)$  is the Delta dirac function. These models have recently attracted a lot of attention both in statistics and machine learning contexts (see e.g. [Bickel and Chen (2009); Decelle et al. (2011); Karrer and Newman (2011); Saade et al. (2014); Massoulié (2014); Ricci-Tersenghi et al. (2016)]).
- *Sub-matrix localization*: This is the problem of finding a submatrix with an elevated mean in a large noisy matrix, as in [Hajek et al. (2015); Chen and Xu (2014)].
- *Matrix completion*: A last example is the matrix completion problem where a part of the information (the matrix elements) is hidden, while the rest is given with noise. For instance, a classical model is  $P_{\text{out}}(w_{ij}|s_i s_j) = p \delta(w_{ij}) + (1 - p) \mathcal{N}(w_{ij}|s_i s_j, \Delta)$ . Such problems have been extensively discussed over the last decades, in particular because of their connection to collaborative filtering (see for instance [Candès and Recht (2009); Cai et al. (2010); Keshavan et al. (2009); Saade et al. (2015)]).

---

1. In this paper  $\mathcal{N}(x|m, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp(-(x - m)^2/2\sigma^2)$

Here we shall consider the probabilistic formulation of these problems and focus on estimation in the mean square error (MSE) sense. We rigorously derive an explicit formula for the mutual information in the asymptotic limit, and for the information theoretic minimal mean square error (MMSE). Our results imply that in a large region of parameters, the posterior expectation of the underlying signal, a quantity often assumed intractable to compute, can be obtained using a polynomial-time scheme via the approximate message-passing (AMP) framework [Rangan and Fletcher (2012); Matsushita and Tanaka (2013); Deshpande and Montanari (2014); Deshpande et al. (2015); Lesieur et al. (2015b)]. We also demonstrate the existence of a region where no *known* tractable algorithm is able to find a solution correlated with the ground truth. Nevertheless, we prove explicitly that it is information theoretically possible to do so (even in this region), and discuss the implications in terms of computational complexity.

The crux of our analysis rests on an "auxiliary" spatially coupled (SC) system. The hallmark of SC models is that one can tune them so that the gap between the algorithmic and information theoretic limits is eliminated, while at the same time the mutual information is maintained unchanged for the coupled and original models. Roughly speaking, this means that it is possible to algorithmically compute the information theoretic limit of the original model because a suitable algorithm is optimal on the coupled system.

Our proof technique has an interest by its own as it combines recent rigorous results in coding theory along the study of capacity-achieving SC codes [Hassani et al. (2010); Kudekar et al. (2011); Yedla et al. (2014); Giurgiu et al. (2016); Barbier et al. (2017a); Dia (2018)] with other progress coming from developments in mathematical physics of spin glass theory [Guerra (2005)]. Moreover, our proof exploits the "threshold saturation" phenomenon of the AMP algorithm and uses spatial coupling as a proof technique. From this point of view, we believe that the theorem proven in this paper is relevant in a broader context going beyond low-rank matrix estimation and can be applied for a wide range of inference problems where message-passing algorithm and spatial coupling can be applied. Furthermore, our work provides important results on the exact formula for the MMSE and on the optimality of the AMP algorithm.

Hundreds of papers have been published in statistics, machine learning or information theory using the non-rigorous statistical physics approach. We believe that our result helps setting a rigorous foundation of a broad line of work. While we focus on rank-one symmetric matrix estimation, our proof technique is readily extendable to more generic low-rank symmetric matrix or low-rank symmetric tensor estimation. We also believe that it can be extended to other problems of interest in machine learning and signal processing. It has already been extended to linear estimation and compressed sensing [Barbier et al. (2016a, 2017b)].

We conclude this introduction by giving a few pointers to the recent literature on rigorous results. For rank-one symmetric matrix estimation problems, AMP has been introduced by [Rangan and Fletcher (2012)], who also computed the state evolution formula to analyze its performance, generalizing techniques developed by [Bayati and Montanari (2011)] and [Javanmard and Montanari (2013)]. State evolution was further studied by [Deshpande and Montanari (2014)] and [Deshpande et al. (2015)]. In [Lesieur et al. (2015a,b)], the generalization to larger rank was also considered. The mutual information was already computed in the special case when  $S_i = \pm 1 \sim \text{Ber}(1/2)$  in [Korada and Macris (2009)] where

an equivalent spin glass model was analyzed. The results of [Korada and Macris (2009)] were first generalized in [Krzakala et al. (2016)] who, notably, obtained a generic matching upper bound. The same formula was also rigorously computed following the study of AMP in [Deshpande and Montanari (2014)] for spike models (provided, however, that the signal was not *too* sparse) and in [Deshpande et al. (2015)] for strictly symmetric community detection. The general formula proposed by [Lesieur et al. (2015a)] for the conditional entropy and the MMSE on the basis of the heuristic cavity method from statistical physics was first demonstrated in full generality by the current authors in [Barbier et al. (2016b)]. This paper represents an extended version of [Barbier et al. (2016b)] that includes all the proofs and derivations along with more detailed discussions. All preexisting proofs could not reach the more interesting regime where a gap between the algorithmic and information theoretic performances appears (i.e. in the presence of “first order” phase transition), leaving a gap with the statistical physics conjectured formula. Following the work of [Barbier et al. (2016b)], the replica formula for rank-one symmetric matrix estimation has been proven again several times using totally different techniques that involve the concentration’s proof of the overlaps [Lelarge and Miolane (2017); Barbier and Macris (2018)]. Our proof strategy does not require any concentration and it uses AMP and spatial coupling as proof techniques. Hence, our result has more practical implications in terms of proving the range of optimality of the AMP algorithm for both the underlying (uncoupled) and spatially coupled models.

This paper is organized as follows: the problem statement and the main results are given in Section 2 along with a sketch of the proof, two applications for symmetric rank-one matrix estimation are presented in Section 3, the threshold saturation phenomenon and the relation between the underlying and spatially coupled models are proven in Section 4 and Section 5 respectively, the proof of the main results follows in Section 6 and Section 7.

A word about notations: in this paper, we use capital letters for random variables, and small letters for fixed realizations. Matrices and vectors are bold while scalars are not. Components of vectors or matrices are identified by the presence of lower indices.

## 2. Setting and main results

### 2.1 Basic underlying model

A standard and natural setting is to consider an additive white gaussian noise (AWGN) channel with variance  $\Delta$  assumed to be known. The model reads

$$w_{ij} = \frac{s_i s_j}{\sqrt{n}} + \sqrt{\Delta} z_{ij}, \quad (1)$$

where  $\mathbf{z} = (z_{ij})_{i,j=1}^n$  is a symmetric matrix with  $Z_{ij} \sim \mathcal{N}(0,1)$ ,  $1 \leq i \leq j \leq n$ , and  $\mathbf{s} = (s_i)_{i=1}^n$  has i.i.d components  $S_i \sim P_0$ . We set  $\mathbb{E}[S^2] = v$ . Precise hypothesis on  $P_0$  are given later.

Perhaps surprisingly, it turns out that the study of this Gaussian setting is sufficient to completely characterize all the problems discussed in the introduction, even if we are dealing with more complicated (noisy) observation models. This is made possible by a theorem of channel universality. Essentially, the theorem states that for any output channel  $P_{\text{out}}(w|y)$  such that at  $y = 0$  the function  $y \mapsto \log P_{\text{out}}(w|y)$  is three times differentiable with bounded

second and third derivatives, then the mutual information satisfies

$$I(\mathbf{S}; \mathbf{W}) = I(\mathbf{S}\mathbf{S}^\top; \mathbf{S}\mathbf{S}^\top + \sqrt{\Delta} \mathbf{Z}) + O(\sqrt{n}), \quad (2)$$

where  $\Delta$  is the inverse Fisher information (evaluated at  $y = 0$ ) of the output channel

$$\Delta^{-1} := \int dw P_{\text{out}}(w|0) \left( \frac{\partial \log P_{\text{out}}(w|y)}{\partial y} \Big|_{y=0} \right)^2. \quad (3)$$

This means that the mutual information per variable  $I(\mathbf{S}; \mathbf{W})/n$  is asymptotically equal the mutual information per variable of an AWGN channel. Informally, it implies that we only have to compute the mutual information for an “effective” Gaussian channel to take care of a wide range of problems. The statement was conjectured in [Lesieur et al. (2015a)] and can be proven by an application of the Lindeberg principle [Deshpande et al. (2015)], [Krzakala et al. (2016)].

## 2.2 AMP algorithm and state evolution

AMP has been applied for the rank-one symmetric matrix estimation problems by [Rangan and Fletcher (2012)], who also computed the state evolution formula to analyze its performance, generalizing techniques developed by [Bayati and Montanari (2011)] and [Javanmard and Montanari (2013)]. State evolution was further studied by [Deshpande and Montanari (2014)] and [Deshpande et al. (2015)]. AMP is an iterative algorithm that provides an estimate  $\hat{\mathbf{s}}^{(t)}(\mathbf{w})$ , at each iteration  $t \in \mathbb{N}$ , of the vector  $\mathbf{s}$ . It turns out that tracking the asymptotic vector and matrix MSE of the AMP algorithm is equivalent to running a simple recursion called *state evolution* (SE).

The AMP algorithm reads

$$\begin{cases} \hat{s}_j^{(t)} = \eta_t((\mathbf{w}\hat{\mathbf{s}}^{(t-1)})_j - b^{(t-1)}\hat{s}_j^{(t-2)}), \\ b^{(t)} = \frac{1}{n} \sum_{i=1}^n \eta'_t((\mathbf{w}\hat{\mathbf{s}}^{(t-1)})_i - b^{(t-1)}\hat{s}_i^{(t-2)}) \end{cases} \quad (4)$$

for  $j = 1, \dots, n$ , where  $\eta_t(y)$  is called the denoiser and  $\eta'_t(y)$  is the derivative w.r.t  $y$ . The *denoiser* is the MMSE estimate associated to an “equivalent scalar denoising problem”

$$y = s + \Sigma(E)z, \quad \Sigma(E)^{-2} := \frac{v - E}{\Delta}. \quad (5)$$

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

$$\eta(y) = \mathbb{E}[X|Y = y] = \frac{\int dx x P_0(x) e^{-\frac{(x-Y)^2}{2\Sigma(E)^2}}}{\int dx P_0(x) e^{-\frac{(x-Y)^2}{2\Sigma(E)^2}}}, \quad (6)$$

where  $E$  is updated at each time instance  $t$  according to the recursion (10).

Natural performance measures are the “vector” and “matrix” MSE’s of the AMP estimator defined below.

**Definition 1 (Vector and matrix MSE of AMP)** *The vector and matrix MSE of the AMP estimator  $\hat{\mathbf{S}}^{(t)}(\mathbf{W})$  at iteration  $t$  are defined respectively as follows*

$$\text{Vmse}_{n,\text{AMP}}^{(t)}(\Delta^{-1}) := \frac{1}{n} \mathbb{E}_{\mathbf{S}, \mathbf{W}} [\|\hat{\mathbf{S}}^{(t)} - \mathbf{S}\|_2^2], \quad (7)$$

$$\text{Mmse}_{n,\text{AMP}}^{(t)}(\Delta^{-1}) := \frac{1}{n^2} \mathbb{E}_{\mathbf{S}, \mathbf{W}} [\|\hat{\mathbf{S}}^{(t)} \hat{\mathbf{S}}^{(t)\top} - \mathbf{S}\mathbf{S}^\top\|_{\mathbb{F}}^2], \quad (8)$$

where  $\|A\|_{\mathbb{F}}^2 = \sum_{i,j} A_{ij}^2$  stands for the Frobenius norm of a matrix  $A$ .

A remarkable fact that follows from a general theorem of [Bayati and Montanari (2011)] (see [Deshpande et al. (2015)] for its use in the matrix case) is that the state evolution sequence tracks these two MSE's and thus allows to assess the performance of AMP. Consider the scalar denoising problem (5). Hence, the (scalar) mmse function associated to this problem reads

$$\text{mmse}(\Sigma(E)^{-2}) := \mathbb{E}_{S,Y} [(S - \mathbb{E}[X|Y])^2]. \quad (9)$$

The state evolution sequence  $E^{(t)}$ ,  $t \in \mathbb{N}$  is defined as

$$E^{(t+1)} = \text{mmse}(\Sigma(E^{(t)})^{-2}), \quad E^{(0)} = v. \quad (10)$$

Since the mmse function is monotone decreasing (its argument has the dimension of a signal to noise ratio) it is easy to see that that  $E^{(t)}$  is a decreasing non-negative sequence. Thus  $\lim_{t \rightarrow +\infty} E^{(t)} := E^{(\infty)}$  exists. One of the basic results of [Bayati and Montanari (2011)], [Deshpande et al. (2015)] is

$$\lim_{n \rightarrow +\infty} \text{Vmse}_{n,\text{AMP}}^{(t)}(\Delta^{-1}) = E^{(t)}, \quad \text{and} \quad \lim_{n \rightarrow +\infty} \text{Mmse}_{n,\text{AMP}}^{(t)}(\Delta^{-1}) = v^2 - (v - E^{(t)})^2. \quad (11)$$

We note that the results in [Bayati and Montanari (2011)], [Deshpande et al. (2015)] are stronger in the sense that the non-averaged algorithmic mean square errors are tracked by state evolution with probability one.

Note that when  $\mathbb{E}[S] = 0$  then  $v$  is an unstable fixed point, and as such, state evolution “does not start”, in other words we have  $E^{(t)} = v$ . While this is not really a problem when one runs AMP in practice, for analysis purposes one can circumvent this problem by slightly biasing  $P_0$  and remove the bias at the end of the analysis. For simplicity, we always assume that  $P_0$  is biased so that  $\mathbb{E}[S]$  is not zero.

**Assumption 1:** In this work we assume that  $P_0$  is discrete with bounded support. Moreover, we assume that  $P_0$  is biased such that  $\mathbb{E}[S]$  is non-zero.

A fundamental quantity computed by state evolution is the algorithmic threshold.

**Definition 2 (AMP threshold)** *For  $\Delta > 0$  small enough, the fixed point equation corresponding to (10) has a unique solution for all noise values in  $]0, \Delta[$ . We define  $\Delta_{\text{AMP}}$  as the supremum of all such  $\Delta$ .*



### 2.3 Spatially coupled model

The present spatially coupled construction is similar to the one used for the coupled Curie-Weiss model [Hassani et al. (2010)] and is also similar to mean field spin glass systems introduced in [Franz and Toninelli (2004); Caltagirone et al. (2014)]. We consider a chain (or a ring) of underlying systems positioned at  $\mu \in \{0, \dots, L\}$  and coupled to neighboring blocks  $\{\mu-w, \dots, \mu+w\}$ . Positions  $\mu$  are taken modulo  $L+1$  and the integer  $w \in \{0, \dots, L/2\}$  equals the size of the *coupling window*. The coupled model is

$$w_{i_\mu j_\nu} = s_{i_\mu} s_{j_\nu} \sqrt{\frac{\Lambda_{\mu\nu}}{n}} + z_{i_\mu j_\nu} \sqrt{\Delta}, \quad (12)$$

where the index  $i_\mu \in \{1, \dots, n\}$  (resp.  $j_\nu$ ) belongs to the block  $\mu$  (resp.  $\nu$ ) along the ring,  $\Lambda$  is an  $(L+1) \times (L+1)$  matrix which describes the strength of the coupling between blocks, and  $Z_{i_\mu j_\nu} \sim \mathcal{N}(0, 1)$  are i.i.d. For the analysis to work, the matrix elements have to be chosen appropriately. We assume that:

- i)  $\Lambda$  is a doubly stochastic matrix;
- ii)  $\Lambda_{\mu\nu}$  depends on  $|\mu - \nu|$ ;
- iii)  $\Lambda_{\mu\nu}$  is not vanishing for  $|\mu - \nu| \leq w$  and vanishes for  $|\mu - \nu| > w$ ;
- iv)  $\Lambda$  is *smooth* in the sense  $|\Lambda_{\mu\nu} - \Lambda_{\mu+1\nu}| = \mathcal{O}(w^{-2})$  and  $\Lambda^* := \sup_{\mu, \nu} \Lambda_{\mu\nu} = \mathcal{O}(w^{-1})$ ;
- v)  $\Lambda$  has a non-negative Fourier transform.

All these conditions can easily be met, the simplest example being a triangle of base  $2w+1$  and height  $1/(w+1)$ , more precisely:

$$\Lambda_{\mu\nu} = \begin{cases} \frac{1}{w+1} \left(1 - \frac{|\mu-\nu|}{w+1}\right), & |\mu - \nu| \leq w \\ 0, & |\mu - \nu| > w \end{cases} \quad (13)$$

We will always denote by  $\mathcal{S}_\mu := \{\nu \mid \Lambda_{\mu\nu} \neq 0\}$  the set of  $2w+1$  blocks coupled to block  $\mu$ .

The construction of the coupled system is completed by introducing a *seed* in the ring: we assume perfect knowledge of the signal components  $\{s_{i_\mu}\}$  for  $\mu \in \mathcal{B} := \{-w-1, \dots, w-1\} \bmod L+1$ . This seed is what allows to close the gap between the algorithmic and information theoretic limits and therefore plays a crucial role. We sometimes refer to the seed as the *pinning construction*. Note that the seed can also be viewed as an “opening” of the chain with fixed boundary conditions.

AMP has been applied for the rank-one symmetric matrix estimation problems by [Rangan and Fletcher (2012)], who also computed the state evolution formula to analyze its performance, generalizing techniques developed by [Bayati and Montanari (2011)] and [Javanmard and Montanari (2013)]. State evolution was further studied by [Deshpande and Montanari (2014)] and [Deshpande et al. (2015)].

The AMP algorithm and the state evolution recursion [Deshpande and Montanari (2014); Deshpande et al. (2015)] can be easily adapted to the spatially coupled model as done in Section 4. The proof that the state evolution for the symmetric rank-one matrix estimation

problem tracks the AMP on a spatially coupled model is an extension of the analysis done in [Deshpande and Montanari (2014); Deshpande et al. (2015)] for the uncoupled model. The full re-derivation of such result would be lengthy and beyond the scope of our analysis. We thus assume that state evolution tracks the AMP performance for our coupled problem. However, we believe that the proof will be similar to the one done for the spatially coupled compressed sensing problem [Javanmard and Montanari (2013)]. This assumption is vindicated numerically.

**Assumption 2:** We consider the spatially coupled model (12) with  $P_0$  satisfying Assumption 1. We assume that state evolution tracks the AMP algorithm for this model.

## 2.4 Main results: basic underlying model

One of our central results is a proof of the expression for the asymptotic mutual information per variable via the so-called *replica symmetric (RS) potential*. This is the function  $E \in [0, v] \mapsto i_{\text{RS}}(E; \Delta) \in \mathbb{R}$  defined as

$$i_{\text{RS}}(E; \Delta) := \frac{(v - E)^2 + v^2}{4\Delta} - \mathbb{E}_{S, Z} \left[ \ln \left( \int dx P_0(x) e^{-\frac{x^2}{2\Sigma(E)^2} + x \left( \frac{S}{\Sigma(E)^2} + \frac{Z}{\Sigma(E)} \right)} \right) \right], \quad (14)$$

with  $Z \sim \mathcal{N}(0, 1)$ ,  $S \sim P_0$ . Most of our results will assume that  $P_0$  is a discrete distribution over a finite bounded real alphabet  $P_0(s) = \sum_{\alpha=1}^{\nu} p_{\alpha} \delta(s - a_{\alpha})$  (see Assumption 1). Thus the only continuous integral in (14) is the Gaussian over  $Z$ . The extension to mixtures of continuous and discrete signals can be obtained by approximation methods not discussed in this paper (see e.g. the methods in [Lelarge and Miolane (2017)]).

It turns out that both the information theoretic and algorithmic AMP thresholds are determined by the set of stationary points of (14) (w.r.t  $E$ ). It is possible to show that for all  $\Delta > 0$  there always exist at least one stationary minimum.<sup>2</sup> In this contribution we suppose that at most three stationary points exist, corresponding to situations with at most one phase transition as depicted in Fig. 1 (see Assumption 3 below). Situations with multiple transitions could also be covered by our techniques.

**Assumption 3:** We assume that  $P_0$  is such that there exist at most three stationary points for the potential (14).

To summarize, our main assumptions in this paper are:

- (A1) The prior  $P_0$  is discrete with bounded support. Moreover, we assume that  $P_0$  is biased such that  $\mathbb{E}[S]$  is non-zero.
- (A2) We consider the spatially coupled model (12) with  $P_0$  satisfying Assumption (A1). We assume that state evolution tracks the AMP algorithm for this model.
- (A3) We assume that  $P_0$  is such that there exist at most three stationary points for the potential (14).

**Remark 3** *An important property of the replica symmetric potential is that the stationary points satisfy the state evolution fixed point equation. In other words  $\partial i_{\text{RS}} / \partial E = 0$  implies  $E = \text{mmse}(\Sigma(E)^{-2})$  and conversely. Moreover it is not difficult to see that the  $\Delta_{\text{AMP}}$*

---

2. Note  $E=0$  is never a stationary point (except for the trivial case of  $P_0$  a single Dirac mass which we exclude from the discussion) and  $E=v$  is stationary only if  $\mathbb{E}[S] = 0$ .

is given by the smallest solution of  $\partial i_{\text{RS}}/\partial E = \partial^2 i_{\text{RS}}/\partial E^2 = 0$ . In other words the AMP threshold is the “first” horizontal inflexion point appearing in  $i_{\text{RS}}(E; \Delta)$  when  $\Delta$  increases from 0 to  $+\infty$ .

One of the main results of this paper is formulated in the following theorem which provides a proof of the conjectured single-letter formula for the asymptotic mutual information per variable.

**Theorem 4 (RS formula for the mutual information)** *Fix  $\Delta > 0$  and let  $P_0$  satisfy Assumptions (A1)-(A3). Then*

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) = \min_{E \in [0, v]} i_{\text{RS}}(E; \Delta). \quad (15)$$

**Proof** See Section 6. ■

The proof of the *existence of the limit* does not require the above hypothesis on  $P_0$ . Also, it was first shown in [Krzakala et al. (2016)] that

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) \leq \min_{E \in [0, v]} i_{\text{RS}}(E; \Delta), \quad (16)$$

an inequality that *we will use* in the proof section. Note that, interestingly, and perhaps surprisingly, the analysis of [Krzakala et al. (2016)] leads to a sharp upper bound on the “free energy” for all finite  $n$ . We will make extensive use of this inequality and for sake of completeness, we summarize its proof in Appendix A.

Theorem 4 allows to compute the information theoretic phase transition threshold which we define in the following way.

**Definition 5 (Information theoretic or optimal threshold)** *Define  $\Delta_{\text{opt}}$  as the first non-analyticity point of the mutual information as  $\Delta$  increases. Formally*

$$\Delta_{\text{opt}} := \sup\{\Delta \mid \lim_{n \rightarrow \infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) \text{ is analytic in } ]0, \Delta[\}. \quad (17)$$

The information theoretic threshold is also called “optimal threshold” because we expect  $\Delta_{\text{AMP}} \leq \Delta_{\text{opt}}$ . This is indeed proven in Lemma 37.

When  $P_0$  is s.t (14) has at most three stationary points, then  $\min_{E \in [0, v]} i_{\text{RS}}(E; \Delta)$  has at most one *non-analyticity* point denoted  $\Delta_{\text{RS}}$  (see Fig. 1). In case of analyticity over all  $\mathbb{R}_+$ , we set  $\Delta_{\text{RS}} = \infty$ . We call  $\Delta_{\text{RS}}$  the RS or potential threshold. Theorem 4 gives us a mean to concretely *compute* the information theoretic threshold:  $\Delta_{\text{opt}} = \Delta_{\text{RS}}$ .

From Theorem 4 we will also deduce the expressions for the *vector* MMSE and the *matrix* MMSE defined below.

**Definition 6 (Vector and matrix MMSE)** *The vector and matrix MMSE are defined respectively as follows*

$$\text{Vmmse}_n(\Delta^{-1}) := \frac{1}{n} \mathbb{E}_{\mathbf{S}, \mathbf{W}} \left[ \|\mathbf{S} - \mathbb{E}[\mathbf{X} | \mathbf{W}]\|_2^2 \right]. \quad (18)$$

$$\text{Mmmse}_n(\Delta^{-1}) := \frac{1}{n^2} \mathbb{E}_{\mathbf{S}, \mathbf{W}} \left[ \|\mathbf{S}\mathbf{S}^\top - \mathbb{E}[\mathbf{X}\mathbf{X}^\top | \mathbf{W}]\|_{\text{F}}^2 \right]. \quad (19)$$

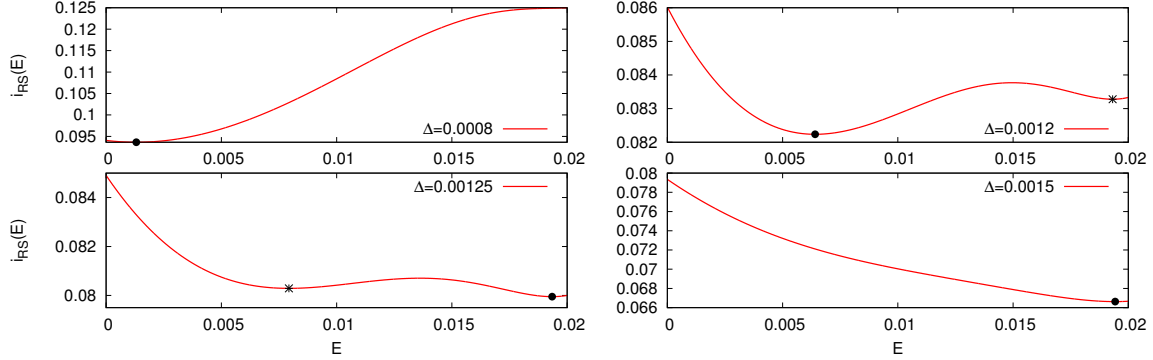


Figure 1: The replica symmetric potential  $i_{\text{RS}}(E)$  for four values of  $\Delta$  in the spiked Wigner model with  $S_i \sim \text{Ber}(\rho)$ . The normalized mutual information is  $\min i_{\text{RS}}(E)$  (the black dot, while the black cross corresponds to the local minimum) and the asymptotic MMSE is  $\text{argmin} i_{\text{RS}}(E)$ , where  $v = \rho$  in this case with  $\rho = 0.02$ . From top left to bottom right: *i*) For low noise values, here  $\Delta = 0.0008 < \Delta_{\text{AMP}}$ , there exists a unique “good” minimum corresponding to the MMSE and AMP is Bayes optimal. *ii*) As the noise increases, a second local “bad” minimum appears: this is the situation at  $\Delta_{\text{AMP}} < \Delta = 0.0012 < \Delta_{\text{RS}}$ . *iii*) For  $\Delta = 0.00125 > \Delta_{\text{RS}}$ , the “bad” minimum becomes the global one and the MMSE suddenly deteriorates. *iv*) For larger values of  $\Delta$ , only the “bad” minimum exists. AMP can be seen as a naive minimizer of this curve starting from  $E = v = 0.02$ . AMP can reach the global minimum in situations *i*), *iii*) and *iv*). However, in situation *ii*), when  $\Delta_{\text{AMP}} < \Delta < \Delta_{\text{RS}}$ , AMP is trapped by the local minimum with large MSE instead of reaching the global one corresponding to the MMSE.

The conditional expectation  $\mathbb{E}[\cdot | \mathbf{W}]$  in Definition 6 is w.r.t the posterior distribution

$$P(\mathbf{x} | \mathbf{w}) = \frac{1}{\tilde{\mathcal{Z}}(\mathbf{w})} e^{-\frac{1}{2\Delta} \sum_{i \leq j} \left( \frac{x_i x_j}{\sqrt{n}} - w_{ij} \right)^2} \prod_{i=1}^n P_0(x_i), \quad (20)$$

with the normalizing factor depending on the observation given by

$$\tilde{\mathcal{Z}}(\mathbf{w}) = \int \left\{ \prod_{i=1}^n dx_i P_0(x_i) \right\} e^{-\frac{1}{2\Delta} \sum_{i \leq j} \left( \frac{x_i x_j}{\sqrt{n}} - w_{ij} \right)^2} \quad (21)$$

The expectation  $\mathbb{E}_{\mathbf{s}, \mathbf{w}}[\cdot]$  is the one w.r.t  $P(\mathbf{w})P(\mathbf{s}) = \tilde{\mathcal{Z}}(\mathbf{w}) \prod_{i=1}^n P_0(s_i)$ . The expressions for the MMSE’s in terms of (14) are given in the following corollary.

**Corollary 7 (Exact formula for the MMSE)** *For all  $\Delta \neq \Delta_{\text{RS}}$ , the matrix MMSE is asymptotically*

$$\lim_{n \rightarrow \infty} \text{Mmmse}_n(\Delta^{-1}) = v^2 - (v - \text{argmin}_{E \in [0, v]} i_{\text{RS}}(E; \Delta))^2. \quad (22)$$

Moreover, if  $\Delta < \Delta_{\text{AMP}}$  or  $\Delta > \Delta_{\text{RS}}$ , then the usual vector MMSE satisfies

$$\lim_{n \rightarrow \infty} \text{Vmmse}_n(\Delta^{-1}) = \text{argmin}_{E \in [0, v]} i_{\text{RS}}(E; \Delta). \quad (23)$$

**Proof** See Section 7. ■

It is natural to conjecture that the vector MMSE is given by  $\operatorname{argmin}_{E \in [0, v]} i_{\text{RS}}(E; \Delta)$  in the whole range  $\Delta \neq \Delta_{\text{RS}}$ , but our proof does not quite yield the full statement.

Another fundamental consequence of Theorem 4 concerns the optimality of the performance of AMP.

**Corollary 8 (Optimality of AMP)** *For  $\Delta < \Delta_{\text{AMP}}$  or  $\Delta > \Delta_{\text{RS}}$ , the AMP is asymptotically optimal in the sense that it yields upon convergence the asymptotic vector-MMSE and matrix-MMSE of Corollary 7. Namely,*

$$\lim_{t \rightarrow +\infty} \lim_{n \rightarrow +\infty} \text{Mmse}_{n, \text{AMP}}^{(t)}(\Delta^{-1}) = \lim_{n \rightarrow \infty} \text{Mmmse}_n(\Delta^{-1}). \quad (24)$$

$$\lim_{t \rightarrow +\infty} \lim_{n \rightarrow +\infty} \text{Vmse}_{n, \text{AMP}}^{(t)}(\Delta^{-1}) = \lim_{n \rightarrow \infty} \text{Vmmse}_n(\Delta^{-1}). \quad (25)$$

On the other hand, for  $\Delta_{\text{AMP}} < \Delta < \Delta_{\text{RS}}$  the AMP algorithm is strictly suboptimal, namely

$$\lim_{t \rightarrow +\infty} \lim_{n \rightarrow +\infty} \text{Mmse}_{n, \text{AMP}}^{(t)}(\Delta^{-1}) > \lim_{n \rightarrow \infty} \text{Mmmse}_n(\Delta^{-1}). \quad (26)$$

$$\lim_{t \rightarrow +\infty} \lim_{n \rightarrow +\infty} \text{Vmse}_{n, \text{AMP}}^{(t)}(\Delta^{-1}) > \lim_{n \rightarrow \infty} \text{Vmmse}_n(\Delta^{-1}). \quad (27)$$

**Proof** See Section 7. ■

This leaves the region  $\Delta_{\text{AMP}} < \Delta < \Delta_{\text{RS}}$  algorithmically open for efficient polynomial time algorithms. A natural conjecture, backed up by many results in spin glass theory, coding theory, planted models and the planted clique problems, is:

**Conjecture 9** *For  $\Delta_{\text{AMP}} < \Delta < \Delta_{\text{RS}}$ , no polynomial time efficient algorithm that outperforms AMP exists.*

## 2.5 Main results: coupled model

In this work, the spatially coupled construction is used for the purposes of the proof. However, one can also imagine interesting applications of the spatially coupled estimation problem, specially in view of the fact that AMP turns out to be optimal for the spatially coupled system. In coding theory for example, the use of spatially coupled systems as a proof device historically followed their initial construction which was for engineering purposes and led to the construction of capacity achieving codes.

Our first crucial result states that the mutual information of the coupled and original systems are the same in a suitable limit. The mutual information of the coupled system of length  $L$  and with coupling window  $w$  is denoted  $I_{w, L}(\mathbf{S}; \mathbf{W})$ .

**Theorem 10 (Equality of mutual informations)** *For any fixed  $w$  s.t.  $P_0$  satisfies Assumption (A1), the following limits exist and are equal*

$$\lim_{L \rightarrow \infty} \lim_{n \rightarrow \infty} \frac{1}{n(L+1)} I_{w, L}(\mathbf{S}; \mathbf{W}) = \lim_{n \rightarrow \infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}). \quad (28)$$

**Proof** See Section 5. ■

An immediate corollary is that the non-analyticity points (w.r.t  $\Delta$ ) of the mutual informations are the same in the coupled and underlying models. In particular, defining  $\Delta_{\text{opt}}^c := \sup\{\Delta \mid \lim_{L \rightarrow \infty} \lim_{n \rightarrow \infty} I_{w,L}(\mathbf{S}; \mathbf{W})/(n(L+1))$  is analytic in  $]0, \Delta[$ , we have  $\Delta_{\text{opt}}^c = \Delta_{\text{opt}}$ .

The second crucial result states that the AMP threshold of the spatially coupled system is at least as good as  $\Delta_{\text{RS}}$  (threshold saturation result of Theorem 11). The analysis of AMP applies to the coupled system as well [Bayati and Montanari (2011); Javanmard and Montanari (2013)] and it can be shown that the performance of AMP is assessed by SE. Let

$$E_{\mu}^{(t)} := \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\|\mathbf{S}_{\mu} - \hat{\mathbf{S}}_{\mu}^{(t)}\|_2^2] \quad (29)$$

be the asymptotic average vector-MSE of the AMP estimate  $\hat{\mathbf{S}}_{\mu}^{(t)}$  at time  $t$  for the  $\mu$ -th “block” of  $\mathbf{S}$ . We associate to each position  $\mu \in \{0, \dots, L\}$  an *independent* scalar system with AWGN of the form  $y = s + \Sigma_{\mu}(\mathbf{E}; \Delta)z$ , with

$$\Sigma_{\mu}(\mathbf{E})^{-2} := \frac{v - \sum_{\nu=0}^L \Lambda_{\mu\nu} E_{\nu}}{\Delta} \quad (30)$$

and  $S \sim P_0$ ,  $Z \sim \mathcal{N}(0, 1)$ . Taking into account knowledge of the signal components in the seed  $\mathcal{B}$ , SE reads

$$\begin{cases} E_{\mu}^{(t+1)} = \text{mmse}(\Sigma_{\mu}(\mathbf{E}^{(t)}; \Delta)^{-2}), & E_{\mu}^{(0)} = v \text{ for } \mu \in \{0, \dots, L\} \setminus \mathcal{B}, \\ E_{\mu}^{(t)} = 0 \text{ for } \mu \in \mathcal{B}, t \geq 0 \end{cases} \quad (31)$$

where the mmse function is defined as in (9).

From the monotonicity of the mmse function we have  $E_{\mu}^{(t+1)} \leq E_{\mu}^{(t)}$  for all  $\mu \in \{0, \dots, L\}$ , a partial order which implies that  $\lim_{t \rightarrow \infty} \mathbf{E}^{(t)} = \mathbf{E}^{(\infty)}$  exists. This allows to define an algorithmic threshold for the coupled system on a finite chain:

$$\Delta_{\text{AMP}, w, L} := \sup\{\Delta \mid E_{\mu}^{(\infty)} \leq E_{\text{good}}(\Delta) \forall \mu\}$$

where  $E_{\text{good}}(\Delta)$  is the trivial fixed point solution of the SE starting with the initial condition  $E^{(0)} = 0$ . A more formal but equivalent definition of  $\Delta_{\text{AMP}, w, L}$  is given in Section 4.

**Theorem 11 (Threshold saturation)** *Let  $\Delta_{\text{AMP}}^c$  be the algorithmic threshold on an infinite chain,  $\Delta_{\text{AMP}}^c := \liminf_{w \rightarrow \infty} \liminf_{L \rightarrow \infty} \Delta_{\text{AMP}, w, L}$ , s.t.  $P_0$  satisfies Assumptions (A1) and (A2). We have  $\Delta_{\text{AMP}}^c \geq \Delta_{\text{RS}}$ .*

**Proof** See Section 4. ■

Our techniques also allow to prove the equality  $\Delta_{\text{AMP}}^c = \Delta_{\text{RS}}$ , but this is not directly needed.

## 2.6 Road map of the proof of the replica symmetric formula

Here we give a road map of the proof of Theorem 4 that will occupy Sections 4–6. A fruitful idea is to concentrate on the question whether  $\Delta_{\text{opt}} = \Delta_{\text{RS}}$ . The proof of this equality automatically generates the proof of Theorem 4.

We first prove in Section 6.1 that  $\Delta_{\text{opt}} \leq \Delta_{\text{RS}}$ . This proof is based on a joint use of the I-MMSE relation (Lemma 33), the replica bound (16) and the suboptimality of the AMP algorithm. In the process of proving  $\Delta_{\text{opt}} \leq \Delta_{\text{RS}}$ , we in fact get as a direct bonus the proof of Theorem 4 for  $\Delta < \Delta_{\text{opt}}$ .

The proof of  $\Delta_{\text{opt}} \geq \Delta_{\text{RS}}$  requires the use of spatial coupling. The main strategy is to show

$$\Delta_{\text{RS}} \leq \Delta_{\text{AMP}}^c \leq \Delta_{\text{opt}}^c = \Delta_{\text{opt}}. \quad (32)$$

The first inequality in (32) is proven in Section 4 using methods first invented in coding theory: The algorithmic AMP threshold of the spatially coupled system  $\Delta_{\text{AMP}}^c$  saturates (tends in a suitable limit) towards  $\Delta_{\text{RS}}$ , i.e.  $\Delta_{\text{RS}} \leq \Delta_{\text{AMP}}^c$  (Theorem 11). To prove the (last) equality we show in Section 5 that the free energies, and hence the mutual informations, of the underlying and spatially coupled systems are equal in a suitable asymptotic limit (Theorem 10). This implies that their non-analyticities occur at the same point and hence  $\Delta_{\text{opt}}^c = \Delta_{\text{opt}}$ . This is done through an interpolation which, although similar in spirit, is different than the one used to prove replica bounds (e.g. (16)). In the process of showing  $\Delta_{\text{opt}}^c = \Delta_{\text{opt}}$ , we will also derive the existence of the limit for  $I(\mathbf{S}; \mathbf{W})/n$ . Finally, the second inequality is due the suboptimality of the AMP algorithm. This follows by a direct extension of the SE analysis of [Deshpande and Montanari (2014); Deshpande et al. (2015)] to the spatially coupled case as done in [Javanmard and Montanari (2013)].

Once  $\Delta_{\text{opt}} = \Delta_{\text{RS}}$  is established it is easy to put everything together and conclude the proof of Theorem 4. In fact all that remains is to prove Theorem 4 for  $\Delta > \Delta_{\text{opt}}$ . This follows by an easy argument in section 6.2 which combines  $\Delta_{\text{opt}} = \Delta_{\text{RS}}$ , the replica bound (16) and the suboptimality of the AMP algorithm. Note that in the proof sections that follow, we assume that Assumptions (A1)-(A3) hold.

## 2.7 Connection with the planted Sherrington-Kirkpatrick spin glass model

Let us briefly discuss the connection of the matrix factorization problem (1) with a statistical mechanical spin glass model which is a variant of the classic Sherrington-Kirkpatrick (SK) model. This is also the occasion to express the mutual information as a “free energy” through a simple relation that will be used in various guises later on.

Replacing  $w_{ij} = n^{-1/2}s_i s_j + \sqrt{\Delta}z_{ij}$  in (20) and simplifying the fraction after expanding the squares, the posterior distribution can be expressed in terms of  $\mathbf{s}, \mathbf{z}$  as follows

$$P(\mathbf{x}|\mathbf{s}, \mathbf{z}) = \frac{1}{\mathcal{Z}} e^{-\mathcal{H}(\mathbf{x}|\mathbf{s}, \mathbf{z})} \prod_{i=1}^n P_0(x_i), \quad (33)$$

where

$$\mathcal{H}(\mathbf{x}|\mathbf{s}, \mathbf{z}) = \sum_{i \leq j=1}^n \left( \frac{x_i^2 x_j^2}{2n\Delta} - \frac{s_i s_j x_i x_j}{n\Delta} - \frac{z_{ij} x_i x_j}{\sqrt{n\Delta}} \right) \quad (34)$$

and

$$\mathcal{Z} = \int \left\{ \prod_{i=1}^n dx_i P_0(x_i) \right\} e^{-\mathcal{H}(\mathbf{x}|\mathbf{s}, \mathbf{z})}. \quad (35)$$

In the language of statistical mechanics, (34) is the ‘‘Hamiltonian’’, (35) the ‘‘partition function’’, and (33) is the Gibbs distribution. This distribution is random since it depends on the realizations of  $\mathbf{S}$ ,  $\mathbf{Z}$ . Conditional expectations with respect to (33) are denoted by the Gibbs ‘‘bracket’’  $\langle - \rangle$ . More precisely

$$\mathbb{E}_{\mathbf{X}|\mathbf{S}, \mathbf{Z}}[A(\mathbf{X})|\mathbf{S} = \mathbf{s}, \mathbf{Z} = \mathbf{z}] = \langle A(\mathbf{X}) \rangle. \quad (36)$$

The free energy is defined as

$$f_n = -\frac{1}{n} \mathbb{E}_{\mathbf{S}, \mathbf{Z}}[\ln \mathcal{Z}]. \quad (37)$$

Notice the difference between  $\tilde{\mathcal{Z}}$  in (20) and  $\mathcal{Z}$  in (33). The former is the partition function with a complete square, whereas the latter is the partition function that we obtain after expanding the square and simplifying the posterior distribution.

In Appendix B, we show that mutual information and free energy are essentially the same object up to a trivial term. For the present model

$$\frac{1}{n} I(\mathbf{S}; \mathbf{W}) = -\frac{1}{n} \mathbb{E}_{\mathbf{S}, \mathbf{Z}}[\ln \mathcal{Z}] + \frac{v^2}{4\Delta} + \frac{1}{4\Delta n} (2\mathbb{E}[S^4] - v^2), \quad (38)$$

where recall  $v = \mathbb{E}[S^2]$ . This relationship turns out to be very practical and will be used several times.

For binary signals we have  $s_i$  and  $x_i \in \{-1, +1\}$ , so the model is a binary spin glass model. The first term in the Hamiltonian is a trivial constant, the last term corresponds exactly to the SK model with random Gaussian interactions, and the second term can be interpreted as an external random field that biases the spins. This is sometimes called a ‘‘planted’’ SK model.

The rest of the paper is organized as follows. In Section 3 we provide two examples of the symmetric rank-one matrix estimation problem. Threshold saturation and the invariance of the mutual information due the spatial coupling are shown in Section 4 and 5 respectively. The proof of Theorem 4 follows in Section 6. Section 7 is dedicated to the proof of Corollary 7 and Corollary 8.

### 3. Two Examples: spiked Wigner model and community detection

In order to illustrate our results, we shall present them here in the context of two examples: the spiked Wigner model, where we close a conjecture left open by [Deshpande and Montanari (2014)], and the case of asymmetric community detection.

#### 3.1 Spiked Wigner model

The first model is defined as follows: we are given data distributed according to the spiked Wigner model where the vector  $\mathbf{s}$  is assumed to be a Bernoulli 0/1 random variable with



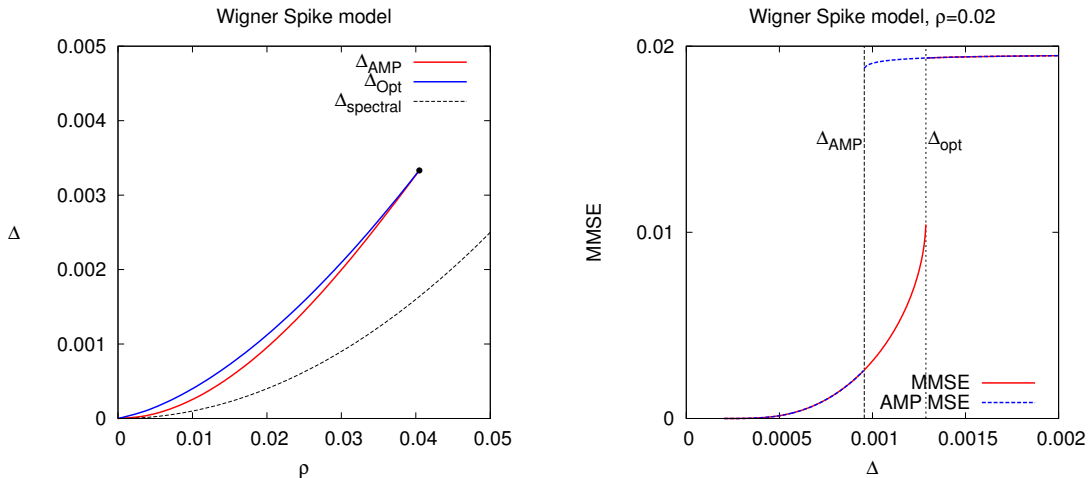


Figure 2: Left: close up of the phase diagram in the density  $\rho$  (where  $P_0(s) = \rho\delta(s-1) + (1-\rho)\delta(x)$ ) versus noise  $\Delta$  plane for the rank-one spiked Wigner model (data from [Lesieur et al. (2015b)]). In [Deshpande and Montanari (2014)], the authors proved that AMP achieves the information theoretic MMSE for all noise as long as  $\rho > 0.041(1)$ . We show that AMP is actually achieving the optimal reconstruction in the whole phase diagram except in the small region between the blue and red line. Notice the large gap with spectral methods (dotted black line). Right: MMSE (red) at  $\rho = 0.02$  as a function of the noise variance  $\Delta$ . AMP (dashed blue) provably achieve the MMSE except in the region  $\Delta_{\text{AMP}} < \Delta < \Delta_{\text{opt}}$ . We conjecture that no polynomial-time algorithm will do better than AMP in this region.

probability  $\rho$ . Data then consists of a sparse, rank-one matrix observed through a Gaussian noise. In [Deshpande and Montanari (2014)], the authors proved that, for  $\rho > 0.041$ , AMP is a computationally efficient algorithm that asymptotically achieves the information theoretically optimal mean-square error for *any* value of the noise  $\Delta$ .

For very small densities (i.e. when  $\rho$  is  $o(1)$ ), there is a well known large gap between what is information theoretically possible and what is tractable with current algorithms in support recovery [Amini and Wainwright (2008)]. This gap is actually related to the planted clique problem [d’Aspremont et al. (2007); Barak et al. (2016)], where it is believed that no polynomial algorithm is able to achieve information theoretic performances. It is thus perhaps not surprising that the situation for  $\rho < 0.041$  becomes a bit more complicated. This is summarized in Fig. 2 and discussed in [Lesieur et al. (2015b)] on the basis of statistical physics consideration which we now prove.

For such values of  $\rho$ , as  $\Delta$  changes there is a region when two local minima appears in  $i_{\text{RS}}(E; \Delta)$  (see the RS formula (14)). In particular for  $\Delta_{\text{AMP}} < \Delta < \Delta_{\text{opt}}$ , the global minimum differs from the AMP one and a computational gap appears (see right panel in Fig. 2). Interestingly, in this problem, the region where AMP is Bayes optimal is still quite large.

The region where AMP is not Bayes optimal is perhaps the most interesting one. While this is by no means evident, statistical physics analogies with actual phase transition in nature suggest that this region will be hard for a very large class of algorithms. A fact that add credibility to this prediction is the following: when looking to small  $\rho$  regime, we find that both the information theoretic threshold *and* the AMP one corresponds to what has been predicted in sparse PCA for sub-extensive values of  $\rho$  [Amini and Wainwright (2008)].

Finally, another interesting line of work for such probabilistic models has appeared in the context of random matrix theory (see for instance [Baik et al. (2005)] and references therein). The focus is to analyze the limiting distribution of the eigenvalues of the observed matrix. The typical picture that emerges from this line of work is that a sharp phase transition occurs at a well-defined critical value of the noise. Above the threshold an outlier eigenvalue (and the principal eigenvector corresponding to it) has a positive correlation with the hidden signal. Below the threshold, however, the spectral distribution of the observation is indistinguishable from that of the pure random noise. In this model, this happens at  $\Delta_{\text{spectral}} = \rho^2$ . Note that for  $\Delta > \Delta_{\text{spectral}}$  spectral methods are not able to distinguish data coming from the model from random ones, while AMP is able to sort (partly) data from noise for any values of  $\Delta$  and  $\rho$ .

### 3.2 Asymmetric community detection

The second model is a problem of detecting two communities (groups) with different sizes  $\rho n$  and  $(1-\rho)n$ , that generalizes the one considered in [Deshpande et al. (2015)]. One is given a graph where the probability to have a link between nodes in the first group is  $p+\mu(1-\rho)/(\rho\sqrt{n})$ , between those in the second group is  $p+\mu\rho/(\sqrt{n}(1-\rho))$ , while interconnections appear with probability  $p-\mu/\sqrt{n}$ . With this peculiar “balanced” setting, the nodes in each group have the same degree distribution with mean  $\rho n$ , making them harder to distinguish.

According to the universality property described in Section 2, this is equivalent to the AWGN model (1) with variance  $\Delta = p(1-p)/\mu^2$  where each variable  $s_i$  is chosen according to

$$P_0(s) = \rho\delta(s - \sqrt{(1-\rho)/\rho}) + (1-\rho)\delta(s + \sqrt{\rho/(1-\rho)}). \quad (39)$$

Our results for this problem<sup>3</sup> are summarized in Fig. 3. For  $\rho > \rho_c = 1/2 - \sqrt{1/12}$  (black point), it is asymptotically information theoretically *possible* to get an estimation better than chance if and only if  $\Delta < 1$ . When  $\rho < \rho_c$ , however, it becomes possible for much larger values of the noise. Interestingly, AMP and spectral methods have the same transition and can find a positive correlation with the hidden communities for  $\Delta < 1$ , regardless of the value of  $\rho$ .

## 4. Threshold saturation

The main goal of this section is to prove that for a proper spatially coupled (SC) system, threshold saturation occurs (Theorem 11), that is  $\Delta_{\text{RS}} \leq \Delta_{\text{AMP}}^c$ . We begin with some

---

3. Note that here since  $E = v = 1$  is an extremum of  $i_{\text{RS}}(E; \Delta)$ , one must introduce a small bias in  $P_0$  and let it then tend to zero at the end of the proofs.

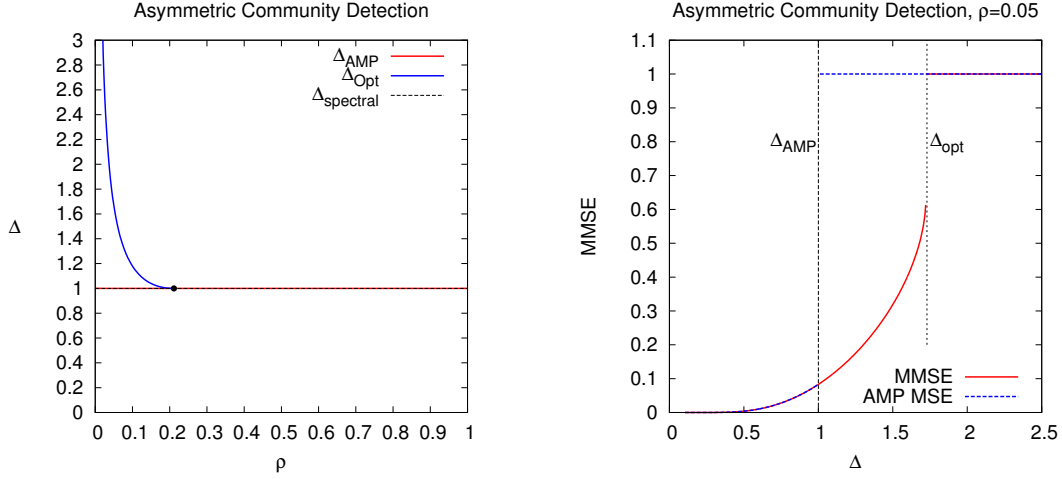


Figure 3: Left: Phase diagram in the density  $\rho$  (where  $P_0(s) = \rho\delta(s - \sqrt{1 - \rho/\rho}) + (1 - \rho)\delta(s + \sqrt{\rho/1 - \rho})$ ) corresponding to an asymmetric community detection problem with two communities. For  $\rho > 1 - \sqrt{1/12}$  (black point), it is information theoretically impossible when  $\Delta > 1$  to find any overlap with the true communities and the optimal MMSE is equal 1, while it is possible for  $\Delta < 1$ . AMP is again always achieving the MMSE, and spectral method can find a non-trivial overlap with the truth starting from  $\Delta < 1$ . For  $\rho < 1 - \sqrt{1/12}$ , however, it is information theoretically possible to find information on the hidden assignment (below the blue line), but both AMP and spectral methods miss this information: Right: MMSE (red) at  $\rho = 0.05$  as a function of the noise variance  $\Delta$ . AMP (dashed blue) provably achieve the MMSE except in the region  $\Delta_{\text{spectral}} = \Delta_{\text{AMP}} < \Delta < \Delta_{\text{opt}}$ .

preliminary formalism in Sec. 4.1, 4.2 on state evolution for the underlying and coupled systems.

#### 4.1 State evolution of the underlying system

First, define the following posterior average

$$\langle A \rangle := \frac{\int dx A(x) P_0(x) e^{-\frac{x^2}{2\Sigma(E,\Delta)^2} + x\left(\frac{s}{\Sigma(E,\Delta)^2} + \frac{z}{\Sigma(E,\Delta)}\right)}}{\int dx P_0(x) e^{-\frac{x^2}{2\Sigma(E,\Delta)^2} + x\left(\frac{s}{\Sigma(E,\Delta)^2} + \frac{z}{\Sigma(E,\Delta)}\right)}}, \quad (40)$$

where  $S \sim P_0$ ,  $Z \sim \mathcal{N}(0, 1)$ . The dependence on these variables, as well as on  $\Delta$  and  $E$  is implicit and dropped from the notation of  $\langle A \rangle$ . Let us define the following operator.

**Definition 12 (SE operator)** *The state evolution operator associated with the underlying system is*

$$T_u(E) := \text{mmse}(\Sigma(E)^{-2}) = \mathbb{E}_{S,Z}[(S - \langle X \rangle)^2], \quad (41)$$

where  $S \sim P_0$ ,  $Z \sim \mathcal{N}(0, 1)$ .

The fixed points of this operator play an important role. They can be viewed as the stationary points of the replica symmetric potential function  $E \in [0, v] \mapsto i_{\text{RS}}(E; \Delta) \in \mathbb{R}$  or equivalently of  $f_{\text{RS}}^{\text{u}} : E \in [0, v] \mapsto f_{\text{RS}}^{\text{u}}(E) \in \mathbb{R}$  where

$$f_{\text{RS}}^{\text{u}}(E) := i_{\text{RS}}(E; \Delta) - \frac{v^2}{4\Delta}. \quad (42)$$

It turns out to be more convenient to work with  $f_{\text{RS}}^{\text{u}}$  instead of  $i_{\text{RS}}$ . We have

**Lemma 13** *Any fixed point of the SE corresponds to a stationary point of  $f_{\text{RS}}^{\text{u}}$ :*

$$E = T_{\text{u}}(E) \Leftrightarrow \left. \frac{\partial f_{\text{RS}}^{\text{u}}(E; \Delta)}{\partial E} \right|_E = 0. \quad (43)$$

**Proof** See Appendix D. ■

The asymptotic performance of the AMP algorithm can be tracked by iterating the SE recursion as follows (this is the same as equation (10) expressed here with the help of  $T_{\text{u}}$ )

$$E^{(t+1)} = T_{\text{u}}(E^{(t)}), \quad t \geq 0, \quad E^{(0)} = v, \quad (44)$$

where the iteration is initialized without any knowledge about the signal other than its prior distribution (in fact, both the asymptotic vector and matrix MSE of the AMP are tracked by the previous recursion as reviewed in Section 2.2). Let  $E_{\text{good}}(\Delta) = T_{\text{u}}^{(\infty)}(0)$ , the fixed point reached by initializing iterations at  $E = 0$ . With our hypothesis on  $P_0$  it is not difficult to see that definition 2 is equivalent to

$$\Delta_{\text{AMP}} := \sup \{ \Delta > 0 \mid T_{\text{u}}^{(\infty)}(v) = E_{\text{good}}(\Delta) \}. \quad (45)$$

The following definition is handy

**Definition 14 (Basin of attraction)** *The basin of attraction of the good solution  $E_{\text{good}}(\Delta)$  is  $\mathcal{V}_{\text{good}} := \{E \mid T_{\text{u}}^{(\infty)}(E) = E_{\text{good}}(\Delta)\}$ .*

Finally, we introduce the notion of *potential gap*. This is a function  $\delta f_{\text{RS}}^{\text{u}} : \Delta \in \mathbb{R}_+ \mapsto \delta f_{\text{RS}}^{\text{u}}(\Delta) \in \mathbb{R}$  defined as follows:

**Definition 15 (Potential gap)** *Define*

$$\delta f_{\text{RS}}^{\text{u}}(\Delta) := \inf_{E \notin \mathcal{V}_{\text{good}}} (f_{\text{RS}}^{\text{u}}(E) - f_{\text{RS}}^{\text{u}}(E_{\text{good}})) \quad (46)$$

*as the potential gap, with the convention that the infimum over the empty set is  $\infty$  (this happens for  $\Delta < \Delta_{\text{AMP}}$  where the complement of  $\mathcal{V}_{\text{good}}$  is the empty set).*

Our hypothesis on  $P_0$  imply that

$$\Delta_{\text{RS}} := \sup \{ \Delta > 0 \mid \delta f_{\text{RS}}^{\text{u}}(\Delta) > 0 \} \quad (47)$$

## 4.2 State evolution of the coupled system

For the SC system, the performance of the AMP decoder is tracked by an *MSE profile* (or just *profile*)  $\mathbf{E}^{(t)}$ , defined componentwise by

$$E_\mu^{(t)} = \lim_{n \rightarrow +\infty} \frac{1}{n} \mathbb{E}_{\mathbf{S}, \mathbf{W}} \|\mathbf{S}_\mu - \hat{\mathbf{S}}_\mu^{(t)}(\mathbf{W})\|_2^2. \quad (48)$$

It has  $L + 1$  components and describes the scalar MSE in each block  $\mu$ . Let us introduce the SE associated with the AMP algorithm for the inference over this SC system. First, denote the following posterior average at fixed  $s, z$  and  $\Delta$ .

$$\langle A \rangle_\mu := \frac{\int dx A(x) P_0(x) e^{-\frac{x^2}{2\Sigma_\mu(\mathbf{E}, \Delta)^2} + x \left( \frac{s}{\Sigma_\mu(\mathbf{E}, \Delta)^2} + \frac{z}{\Sigma_\mu(\mathbf{E}, \Delta)} \right)}}{\int dx P_0(x) e^{-\frac{x^2}{2\Sigma_\mu(\mathbf{E}, \Delta)^2} + x \left( \frac{s}{\Sigma_\mu(\mathbf{E}, \Delta)^2} + \frac{z}{\Sigma_\mu(\mathbf{E}, \Delta)} \right)}}. \quad (49)$$

where the *effective noise variance* of the SC system is defined as

$$\Sigma_\mu(\mathbf{E})^{-2} := \frac{v - \sum_{\nu \in \mathcal{S}_\mu} \Lambda_{\mu\nu} E_\nu}{\Delta}, \quad (50)$$

where we recall  $\mathcal{S}_\mu := \{\nu \mid \Lambda_{\mu\nu} \neq 0\}$  is the set of  $2w + 1$  blocks coupled to block  $\mu$ .

**Definition 16 (SE operator of the coupled system)** *The state evolution operator associated with the coupled system (12) is defined component-wise as*

$$[T_c(\mathbf{E})]_\mu := \mathbb{E}_{S, Z} [(S - \langle X \rangle_\mu)^2]. \quad (51)$$

$T_c(\mathbf{E})$  is vector valued and here we have written its  $\mu$ -th component.

We assume perfect knowledge of the variables  $\{s_{i_\mu}\}$  inside the blocks  $\mu \in \mathcal{B} := \{0 : w - 1\} \cup \{L - w : L\}$  as mentioned in Section 2.3, that is  $x_{i_\mu} = s_{i_\mu}$  for all  $i_\mu$  such that  $\mu \in \mathcal{B}$ . This implies  $E_\mu = 0 \forall \mu \in \mathcal{B}$ . We refer to this as the *pinning condition*. The SE iteration tracking the scalar MSE profile of the SC system reads for  $\mu \notin \mathcal{B}$

$$E_\mu^{(t+1)} = [T_c(\mathbf{E}^{(t)})]_\mu \quad \forall t \geq 0, \quad (52)$$

with the initialization  $E_\mu^{(0)} = v$ . For  $\mu \in \mathcal{B}$ , the pinning condition forces  $E_\mu^{(t)} = 0 \forall t$ . This equation is the same as (31) but is expressed here in terms of the operator  $T_c$ .

Let us introduce a suitable notion of degradation that will be very useful for the analysis.

**Definition 17 (Degradation)** *A profile  $\mathbf{E}$  is degraded (resp. strictly degraded) w.r.t another one  $\mathbf{G}$ , denoted as  $\mathbf{E} \succeq \mathbf{G}$  (resp.  $\mathbf{E} \succ \mathbf{G}$ ), if  $E_\mu \geq G_\mu \forall \mu$  (resp. if  $\mathbf{E} \succeq \mathbf{G}$  and there exists some  $\mu$  such that  $E_\mu > G_\mu$ ).*

Define an error profile  $\mathbf{E}_{\text{good}}(\Delta)$  as the vector with all  $L + 1$  components equal to  $E_{\text{good}}(\Delta)$ .

**Definition 18 (AMP threshold of coupled ensemble)** *The AMP threshold of the coupled system is defined as*

$$\Delta_{\text{AMP}}^c := \liminf_{w, L \rightarrow \infty} \sup \{ \Delta > 0 \mid T_c^{(\infty)}(\mathbf{v}) \prec \mathbf{E}_{\text{good}}(\Delta) \} \quad (53)$$

where  $\mathbf{v}$  is the all  $v$  vector. The  $\liminf_{w, L \rightarrow \infty}$  is taken along sequences where first  $L \rightarrow \infty$  and then  $w \rightarrow \infty$ . We also set for a finite system  $\Delta_{\text{AMP}, w, L} := \sup \{ \Delta > 0 \mid T_c^{(\infty)}(\mathbf{v}) \prec \mathbf{E}_{\text{good}}(\Delta) \}$ .

The proof presented in the next subsection uses extensively the following monotonicity properties of the SE operators.

**Lemma 19** *The SE operator of the SC system maintains degradation in space, i.e.  $\mathbf{E} \succeq \mathbf{G} \Rightarrow T_c(\mathbf{E}) \succeq T_c(\mathbf{G})$ . This property is verified by  $T_u(E)$  for a scalar error as well.*

**Proof** From (50) one can immediately see that  $\mathbf{E} \succeq \mathbf{G} \Rightarrow \Sigma_\mu(\mathbf{E}) \geq \Sigma_\mu(\mathbf{G}) \forall \mu$ . Now, the SE operator (51) can be interpreted as the mmse function associated to the Gaussian channel  $y = s + \Sigma_\mu(\mathbf{E}, \Delta)z$ . This is an increasing function of the noise intensity  $\Sigma_\mu^2$ : this is intuitively clear but we provide an explicit formula for the derivative below. Thus  $[T_c(\mathbf{E})]_\mu \geq [T_c(\mathbf{G})]_\mu \forall \mu$ , which means  $T_c(\mathbf{E}) \succeq T_c(\mathbf{G})$ .

The derivative of the mmse function of the Gaussian channel can be computed as

$$\frac{d \text{mmse}(\Sigma^{-2})}{d \Sigma^{-2}} = -2 \mathbb{E}_{X,Y} \left[ \|X - \mathbb{E}[X|Y]\|_2^2 \text{Var}[X|Y] \right]. \quad (54)$$

This formula explicitly confirms that  $T_u(E)$  (resp.  $[T_c(\mathbf{E})]_\mu$ ) is an increasing function of  $\Sigma^2$  (resp.  $\Sigma_\mu^2$ ). ■

**Corollary 20** *The SE operator of the coupled system maintains degradation in time, i.e.,  $T_c(\mathbf{E}^{(t)}) \preceq \mathbf{E}^{(t)} \Rightarrow T_c(\mathbf{E}^{(t+1)}) \preceq \mathbf{E}^{(t+1)}$ . Similarly  $T_c(\mathbf{E}^{(t)}) \succeq \mathbf{E}^{(t)} \Rightarrow T_c(\mathbf{E}^{(t+1)}) \succeq \mathbf{E}^{(t+1)}$ . Furthermore, the limiting error profile  $\mathbf{E}^{(\infty)} := T_c^{(\infty)}(\mathbf{E}^{(0)})$  exists. These properties are verified by  $T_u(E)$  as well.*

**Proof** The degradation statements are a consequence of Lemma 19. The existence of the limits follows from the monotonicity of the operator and boundedness of the scalar MSE. ■

Finally we will also need the following generalization of the (replica symmetric) potential function to a spatially coupled system:

$$\begin{aligned} f_{\text{RS}}^c(\mathbf{E}) &= \sum_{\mu=0}^L \sum_{\nu \in \mathcal{S}_\mu} \frac{\Lambda_{\mu\nu}}{4\Delta} (v - E_\mu)(v - E_\nu) \\ &\quad - \sum_{\mu=0}^L \mathbb{E}_{S,Z} \left[ \ln \left( \int dx P_0(x) e^{-\frac{1}{2\Sigma_\mu(\mathbf{E})^2} (x^2 - 2xS + xZ\Sigma_\mu(\mathbf{E}, \Delta))} \right) \right], \end{aligned} \quad (55)$$

where  $Z \sim \mathcal{N}(z|0, 1)$  and  $S \sim P_0(s)$ . As for the underlying system, the following Lemma links the SE and RS formulations.

**Lemma 21** *If  $\mathbf{E}$  is a fixed point of (52), i.e.  $E_\mu = [T_c(\mathbf{E})]_\mu \Rightarrow \frac{\partial f_{\text{RS}}^c(\mathbf{E})}{\partial E_\mu} \Big|_{\mathbf{E}} = 0 \forall \mu \in \mathcal{B}^c = \{w : L - w - 1\}$ .*

**Proof** The proof is similar to the proof of Lemma 13 in Appendix D. We skip the details for brevity. ■

Now that we have settled the required definitions and properties, we can prove threshold saturation.

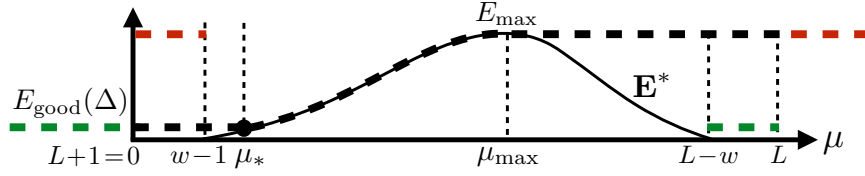


Figure 4: A fixed point profile  $\mathbf{E}^*$  of the coupled SE iteration (solid line) is necessarily null  $\forall \mu \in \{0 : w - 1\}$  because of the pinning condition, and then it increases to reach  $E_{\max} \in [0, v]$  at some  $\mu_{\max} \in \{w : L - w - 1\}$  (for a symmetric coupling matrix  $\mu_{\max} = L/2$ ). Then, it starts decreasing and it is again null  $\forall \mu \in \{L - w : L\}$ . By definition, the associated *saturated* profile  $\mathbf{E}$  (dashed line) starts at  $E_{\text{good}}(\Delta) \forall \mu \leq \mu_*$ , where  $\mu_*$  is such that:  $E_{\mu}^* \leq E_{\text{good}}(\Delta) \forall \mu \in \{0 : \mu_*\}$  and  $E_{\mu'}^* > E_{\text{good}}(\Delta) \forall \mu' \in \{\mu_* + 1 : L\}$ . Then,  $\mathbf{E}$  matches  $\mathbf{E}^* \forall \mu \in \{\mu_* : \mu_{\max}\}$  and saturates at  $E_{\max} \forall \mu \geq \mu_{\max}$ . The saturated profile is extended for  $\mu < 0$  and  $\mu > L$  indices. The green (resp. red) branch shows that when the block indices of  $\mathbf{E}$  are  $\mu < 0$  (resp.  $\mu > L$ ), then  $E_{\mu}$  equals to  $E_{\text{good}}(\Delta)$  (resp.  $E_{\max}$ ). By construction,  $\mathbf{E}$  is non decreasing in  $\mu$  and is degraded w.r.t the fixed point profile  $\mathbf{E} \succeq \mathbf{E}^*$ .

### 4.3 Proof of Theorem 11

The proof will proceed by contradiction. Let  $\mathbf{E}^*$  a fixed point profile of the SE iteration (52). We suppose that  $\mathbf{E}^*$  *does not* satisfy  $\mathbf{E}^* \prec \mathbf{E}_{\text{good}}(\Delta)$ , and exhibit a contradiction for  $\Delta < \Delta_{\text{RS}}$  and  $w$  large enough (but independent of  $L$ ). Thus we must have  $\mathbf{E}^* \prec \mathbf{E}_{\text{good}}(\Delta)$ . This is the statement of Theorem 28 in Sec. 4.3.3 and directly implies Theorem 11.

The pinning condition together with the monotonicity properties of the coupled SE operator (Lemma 19 and Corollary 20) ensure that any fixed point profile  $\mathbf{E}^*$  which does not satisfy  $\mathbf{E}^* \prec \mathbf{E}_{\text{good}}(\Delta)$  necessarily has a shape as described in Fig. 4. We construct an associated *saturated profile*  $\mathbf{E}$  as described in Fig. 4. From now on we work with a *saturated profile*  $\mathbf{E}$  which verifies  $\mathbf{E} \succeq \mathbf{E}^*$  and  $\mathbf{E} \succeq \mathbf{E}_{\text{good}}(\Delta)$ . In the following we will need the following operator.

**Definition 22 (Shift operator)** *The shift operator  $S$  is defined componentwise as  $[S(\mathbf{E})]_{\mu} := E_{\mu-1}$ .*

#### 4.3.1 UPPER BOUND ON THE POTENTIAL VARIATION UNDER A SHIFT

The first step in the proof of threshold saturation is based on the Taylor expansion of the RS free energy of the SC system.

**Lemma 23** *Let  $\mathbf{E}$  be a saturated profile. Set  $\mathbf{E}_{\lambda} := (1 - \lambda)\mathbf{E} + \lambda S(\mathbf{E})$  for  $\lambda \in [0, 1]$  and  $\delta E_{\mu} := E_{\mu} - E_{\mu-1}$ . There exists some  $\lambda \in [0, 1]$  such that*

$$f_{\text{RS}}^c(S(\mathbf{E})) - f_{\text{RS}}^c(\mathbf{E}) = \frac{1}{2} \sum_{\mu, \mu'=0}^L \delta E_{\mu} \delta E_{\mu'} \left. \frac{\partial^2 f_{\text{RS}}^c(\mathbf{E})}{\partial E_{\mu} \partial E_{\mu'}} \right|_{\mathbf{E}_{\lambda}}. \quad (56)$$

**Proof** Using the remainder Theorem, the free energy difference can be expressed as

$$f_{\text{RS}}^c(\mathbf{S}(\mathbf{E})) - f_{\text{RS}}^c(\mathbf{E}) = - \sum_{\mu=0}^L \delta E_{\mu} \frac{\partial f_{\text{RS}}^c(\mathbf{E})}{\partial E_{\mu}} \Big|_{\mathbf{E}} + \frac{1}{2} \sum_{\mu, \mu'=0}^L \delta E_{\mu} \delta E_{\mu'} \frac{\partial^2 f_{\text{RS}}^c(\mathbf{E})}{\partial E_{\mu} \partial E_{\mu'}} \Big|_{\mathbf{E}_{\lambda}}. \quad (57)$$

for some  $\lambda \in [0, 1]$ . By definition of the saturated profile  $\mathbf{E}$ , we have  $\delta E_{\mu} = 0 \forall \mu \in \mathcal{A} := \{0 : \mu_*\} \cup \{\mu_{\max} + 1 : L\}$  and  $E_{\mu} = [T_c(\mathbf{E})]_{\mu}$  for  $r \notin \mathcal{A}$ . Recalling Lemma 21 we see that the derivative in the first sum cancels for  $r \notin \mathcal{A}$ . Hence the first sum in (57) vanishes.  $\blacksquare$

**Lemma 24** *The saturated profile  $\mathbf{E}$  is smooth, i.e.  $\delta E^* := \max_{\mu} |\delta E_{\mu}| = \mathcal{O}(1/w)$  uniformly in  $L$ .*

**Proof** By definition of the saturated profile  $\mathbf{E}$  we have  $\delta E_{\mu} = 0 \forall \mu \in \mathcal{A} := \{0 : \mu_*\} \cup \{\mu_{\max} + 1 : L\}$ . For  $\mu \notin \mathcal{A}$ , we can replace the fixed point profile component  $E_{\mu}$  by  $[T_c(\mathbf{E})]_{\mu}$  so that  $\delta E_{\mu} = [T_c(\mathbf{E})]_{\mu} - [T_c(\mathbf{E})]_{\mu-1}$ . We will Taylor expand the SE operator. To this end, we define  $\delta \Sigma_{\mu}^{-2} := \Sigma_{\mu}(\mathbf{E})^{-2} - \Sigma_{\mu-1}(\mathbf{E})^{-2}$  for  $\mu \in \{\mu_* + 1 : \mu_{\max}\}$ . Recall that  $\Lambda_{\mu-1, \nu-1} = \Lambda_{\mu\nu}$ ,  $\Lambda_{\mu\nu} \geq 0$  and  $\Lambda^* := \sup_{\mu, \nu} \Lambda_{\mu\nu} = \mathcal{O}(1/w)$ . Thus from (50) we get

$$\begin{aligned} |\delta \Sigma_{\mu}^{-2}| &= \frac{1}{\Delta} \left| \sum_{\nu \in \mathcal{S}_{\mu}} \Lambda_{\mu\nu} (E_{\nu} - E_{\nu-1}) \right| \\ &\leq \frac{\Lambda^*}{\Delta} \sum_{\nu \in \mathcal{S}_{\mu}} (E_{\nu} - E_{\nu-1}) \\ &\leq \frac{2v\Lambda^*}{\Delta} = \mathcal{O}\left(\frac{1}{w}\right) \end{aligned} \quad (58)$$

where we have used  $E_{\nu} - E_{\nu-1} \geq 0$  to get rid of the absolute value. Note that the first and second derivatives of the SE operator (51) w.r.t  $\Sigma_{\mu}^{-2}$  are bounded as long as the five first moments of the posterior (49) exist and are bounded (which is true under our assumptions). Then by Taylor expansion at first order in  $\delta \Sigma_{\mu}^{-2}$  and using the remainder theorem, we obtain

$$|\delta E_{\mu}| = |[T_c(\mathbf{E})]_{\mu} - [T_c(\mathbf{E})]_{\mu-1}| \leq |\delta \Sigma_{\mu}^{-2}| \left| \frac{\partial [T_c(\mathbf{E})]_{\mu}}{\partial \Sigma_{\mu}^{-2}} \right| + \mathcal{O}(\delta \Sigma_{\mu}^{-4}) \leq \mathcal{O}\left(\frac{1}{w}\right), \quad (59)$$

where the last inequality follows from (58).  $\blacksquare$

**Proposition 25** *Let  $\mathbf{E}$  be a saturated profile. Then for all  $\Delta > 0$  there exists a constant  $0 < C(\Delta) < +\infty$  independent of  $L$  such that*

$$|f_{\text{RS}}^c(\mathbf{S}(\mathbf{E})) - f_{\text{RS}}^c(\mathbf{E})| \leq \frac{C(\Delta)}{w}. \quad (60)$$

**Proof** From Lemma 23, in order to compute the free energy difference between the shifted and non-shifted profiles, we need to compute the Hessian associated with this free energy.



We have

$$\begin{aligned} \frac{\partial f_{\text{RS}}^c(\mathbf{E})}{\partial E_\mu} &= \sum_{\nu \in \mathcal{S}_\mu} \frac{\Lambda_{\mu,\nu} E_\nu}{2\Delta} - \frac{1}{2} \sum_{\nu} \frac{\partial \Sigma_\nu(\mathbf{E})^{-2}}{\partial E_\mu} [T_c(\mathbf{E})]_\nu - \frac{v}{2\Delta} \\ &= \frac{1}{2\Delta} \left( \sum_{\nu \in \mathcal{S}_\mu} \Lambda_{\mu,\nu} E_\nu + \sum_{\nu \in \mathcal{S}_\mu} \Lambda_{\mu,\nu} [T_c(\mathbf{E})]_\nu - v \right) \end{aligned} \quad (61)$$

and

$$\frac{\partial^2 f_{\text{RS}}^c(\mathbf{E})}{\partial E_\mu \partial E_{\mu'}} = \frac{1}{2\Delta} \left( \Lambda_{\mu,\mu'} \mathbb{I}(\mu' \in \mathcal{S}_\mu) - \frac{1}{\Delta} \sum_{\nu \in \mathcal{S}_\mu \cap \mathcal{S}_{\mu'}} \Lambda_{\mu,\nu} \Lambda_{\mu',\nu} \frac{\partial [T_c(\mathbf{E})]_\nu}{\partial \Sigma_\nu^{-2}} \right). \quad (62)$$

We can now estimate the sum in the Lemma 23. The contribution of the first term on the r.h.s of (62) can be bounded as

$$\frac{1}{2\Delta} \left| \sum_{\mu=0}^L \sum_{\mu' \in \mathcal{S}_\mu} \Lambda_{\mu,\mu'} \delta E_\mu \delta E_{\mu'} \right| \leq \frac{\delta E^* \Lambda^* (2w+1)}{2\Delta} \left| \sum_{\mu=0}^L \delta E_\mu \right| \leq \mathcal{O}\left(\frac{1}{w}\right), \quad (63)$$

where we used the facts:  $\delta E_\mu \geq 0$ , the sum over  $\mu = 0, \dots, L$  is telescopic,  $E_\mu \in [0, v]$ ,  $\Lambda^* = \mathcal{O}(1/w)$  and  $\delta E^* = (w^{-1})$  (Lemma 24). We now bound the contribution of the second term on the r.h.s of (62). Recall the first derivative w.r.t  $\Sigma_\nu^{-2}$  of the SE operator is bounded uniformly in  $L$ . Call this bound  $K = \mathcal{O}(1)$ . We obtain

$$\begin{aligned} & \frac{1}{2\Delta^2} \left| \sum_{\mu,\mu'=1}^L \delta E_\mu \delta E_{\mu'} \sum_{\nu \in \mathcal{S}_\mu \cap \mathcal{S}_{\mu'}} \Lambda_{\mu,\nu} \Lambda_{\mu',\nu} \frac{\partial [T_c(\mathbf{E})]_\nu}{\partial \Sigma_\nu^{-2}} \right| \\ & \leq \frac{K \Lambda^{*2} \delta E^*}{2\Delta^2} \left| \sum_{\mu=1}^L \delta E_\mu \sum_{\mu' \in \{\mu-2w:\mu+2w\}} \text{card}(\mathcal{S}_\mu \cap \mathcal{S}_{\mu'}) \right| \leq \mathcal{O}\left(\frac{1}{w}\right). \end{aligned} \quad (64)$$

The last inequality follows from the following facts: the sum over  $\mu = 1, \dots, L$  is telescopic,  $\Lambda^* = \mathcal{O}(1/w)$ , Lemma 24, and for any fixed  $\mu$  the following holds

$$\sum_{\mu' \in \{\mu-2w:\mu+2w\}} \text{card}(\mathcal{S}_\mu \cap \mathcal{S}_{\mu'}) = (2w+1)^2. \quad (65)$$

Finally, from (63), (64) and the triangle inequality we obtain

$$\frac{1}{2} \left| \sum_{\mu,\mu'=1}^L \delta E_\mu \delta E_{\mu'} \frac{\partial^2 f_{\text{RS}}^c(\mathbf{E})}{\partial E_\mu \partial E_{\mu'}} \Big|_{\mathbf{E}_\lambda} \right| = \mathcal{O}\left(\frac{1}{w}\right) \quad (66)$$

uniformly in  $L$ . Combining this result with Lemma 23 ends the proof.  $\blacksquare$

### 4.3.2 LOWER BOUND ON THE POTENTIAL VARIATION UNDER A SHIFT

The second step in the proof is based on a direct evaluation of  $f_{\text{RS}}^c(S(E)) - f_{\text{RS}}^c(E)$ . We first need the Lemma:

**Lemma 26** *Let  $\mathbf{E}$  be a saturated profile such that  $\mathbf{E} \succ \mathbf{E}_{\text{good}}(\Delta)$ . Then  $E_{\text{max}} \notin \mathcal{V}_{\text{good}}$ .*

**Proof** The fact that the error profile is non decreasing and the assumption that  $\mathbf{E} \succ \mathbf{E}_{\text{good}}(\Delta)$  imply that  $E_{\text{max}} > E_0 = E_{\text{good}}(\Delta)$ . Moreover,  $E_{\text{max}} \leq [T_c(\mathbf{E})]_{\mu_{\text{max}}} \leq T_u(E_{\text{max}})$  where the first inequality follows from  $\mathbf{E} \succ \mathbf{E}^*$  and the monotonicity of  $T_c$ , while the second comes from the fact that  $\mathbf{E}$  is non decreasing. Combining these with the monotonicity of  $T_u$  gives  $T_u(E_{\text{max}}) \geq E_{\text{max}}$  which implies  $T_u^{(\infty)}(E_{\text{max}}) \geq E_{\text{max}} > E_{\text{good}}(\Delta)$  which means  $E_{\text{max}} \notin \mathcal{V}_{\text{good}}$ .  $\blacksquare$

**Proposition 27** *Fix  $\Delta < \Delta_{\text{RS}}$  and let  $\mathbf{E}$  be a saturated profile such that  $\mathbf{E} \succ \mathbf{E}_{\text{good}}(\Delta)$ . Then*

$$|f_{\text{RS}}^c(S(\mathbf{E})) - f_{\text{RS}}^c(\mathbf{E})| \geq \delta f_{\text{RS}}^u(\Delta) \quad (67)$$

where  $\delta f_{\text{RS}}^u(\Delta)$  is the potential gap (Definition 15).

**Proof** Set

$$\mathcal{I}(\Sigma) := \mathbb{E}_{S,Z} \left[ \ln \left( \int dx P_0(x) e^{-\frac{1}{2\Sigma^2} (x^2 - 2Sx - 2\Sigma Zx)} \right) \right].$$

By (55)

$$\begin{aligned} f_{\text{RS}}^c(S(\mathbf{E})) - f_{\text{RS}}^c(\mathbf{E}) &= \sum_{\mu=-1}^{L-1} \sum_{\nu \in \mathcal{S}_\mu} \frac{\Lambda_{\mu+1\nu+1}}{4\Delta} (v - E_\mu)(v - E_\nu) - \sum_{\mu=0}^L \sum_{\nu \in \mathcal{S}_\mu} \frac{\Lambda_{\mu\nu}}{4\Delta} (v - E_\mu)(v - E_\nu) \\ &\quad - \sum_{\mu=0}^L \mathcal{I}(\Sigma_\mu(S(\mathbf{E}))) + \sum_{\mu=0}^L \mathcal{I}(\Sigma_\mu(\mathbf{E})) \\ &= \sum_{\nu \in \mathcal{S}_{-1}} \frac{\Lambda_{\mu\nu}}{4\Delta} (v - E_\mu)(v - E_\nu) - \sum_{\nu \in \mathcal{S}_L} \frac{\Lambda_{\mu\nu}}{4\Delta} (v - E_\mu)(v - E_\nu) \\ &\quad - \mathcal{I}(\Sigma_{-1}(\mathbf{E})) + \mathcal{I}(\Sigma_L(\mathbf{E})), \end{aligned} \quad (68)$$

where we used  $\Lambda_{\mu+1\nu+1} = \Lambda_{\mu\nu}$  implying also  $\Sigma_\mu(S(\mathbf{E})) = \Sigma_{\mu-1}(\mathbf{E})$  as seen from (50). Recall  $\Sigma(E)^{-2} = (v - E)/\Delta$ . Now looking at (50), one notices that thanks to the saturation of  $\mathbf{E}$ ,  $\Sigma_{-1}(\mathbf{E}) = \Sigma(E_0)$  where  $E_0 = E_{\text{good}}(\Delta)$  (see the green branch in Fig. 4), while  $\Sigma_L(\mathbf{E}) = \Sigma(E_L)$  where  $E_L = E_{\text{max}}$  (see the red branch Fig. 4). Finally from (68), using that the coupling matrix is (doubly) stochastic and the saturation of  $\mathbf{E}$

$$\begin{aligned} f_{\text{RS}}^c(S(\mathbf{E})) - f_{\text{RS}}^c(\mathbf{E}) &= \left[ \frac{(v - E_0)^2}{4\Delta} - \mathcal{I}(\Sigma(E_{\text{good}}(\Delta))) \right] - \left[ \frac{(v - E_L)^2}{4\Delta} - \mathcal{I}(\Sigma(E_L)) \right] \\ &= f_{\text{RS}}^u(E_{\text{good}}) - f_{\text{RS}}^u(E_{\text{max}}) \leq -\delta f_{\text{RS}}^u(\Delta), \end{aligned} \quad (69)$$

where we recognized the potential function of the underlying system  $f_{\text{RS}}^{\text{u}}(E; \Delta) = i_{\text{RS}}(E; \Delta) - \frac{v^2}{4\Delta}$  and the last inequality is a direct application of Lemma 26 and Definition 15. Finally, using the positivity of  $\delta f_{\text{RS}}^{\text{u}}(\Delta)$  for  $\Delta < \Delta_{\text{RS}}$ , we obtain the desired result.  $\blacksquare$

### 4.3.3 END OF PROOF OF THRESHOLD SATURATION

We now have the necessary ingredients in order to prove threshold saturation.

**Theorem 28 (Asymptotic performance of AMP for the coupled system)** *Fix  $\Delta < \Delta_{\text{RS}}$ . Take a spatially coupled system with  $w > C(\Delta)/\delta f_{\text{RS}}^{\text{u}}(\Delta)$  where  $C(\Delta)$  is the constant in Proposition 25. Then any fixed point profile  $\mathbf{E}^*$  of the coupled state evolution iteration (52) must satisfy  $\mathbf{E}^* \prec \mathbf{E}_{\text{good}}(\Delta)$ .*

**Proof** The proof is by contradiction. Fix  $\Delta < \Delta_{\text{RS}}$  and  $w \geq C(\Delta)/\delta f_{\text{RS}}^{\text{u}}(\Delta)$ . We assume there exists a fixed point profile which does not satisfy  $\mathbf{E}^* \prec \mathbf{E}_{\text{good}}(\Delta)$ . Then we construct the associated saturated profile  $\mathbf{E}$ . This profile satisfies both statements of Propositions 25 and 27. Therefore we must have  $\delta f_{\text{RS}}^{\text{u}}(\Delta) \leq C(\Delta)/w$  which contradicts the choice  $w > C(\Delta)/\delta f_{\text{RS}}^{\text{u}}(\Delta)$ . We conclude that  $\mathbf{E}^* \prec \mathbf{E}_{\text{good}}(\Delta)$  must be true.  $\blacksquare$

Theorem 11 is a direct corollary of Theorem 28 and Definition 18. Take some  $\Delta_* < \Delta_{\text{RS}}$  and choose  $w > C(\Delta_*)/\delta f_{\text{RS}}^{\text{u}}(\Delta_*)$ . Then we have  $\Delta_{\text{AMP},w,L} \geq \Delta_*$ . Note that  $\delta f_{\text{RS}}^{\text{u}}(\Delta_*) \rightarrow 0_+$  for  $\Delta_* \rightarrow \Delta_{\text{RS}}$ . Thus Taking  $L \rightarrow +\infty$  first and  $w \rightarrow +\infty$  second we can make  $\Delta_*$  as close to  $\Delta_{\text{RS}}$  as we wish. Therefore we obtain  $\Delta_{\text{AMP}}^{\text{c}} := \liminf_{L,w \rightarrow +\infty} \Delta_{\text{AMP},w,L} \geq \Delta_{\text{RS}}$  where the limit is taken in the specified order.

## 5. Invariance of the mutual information under spatial coupling

In this section we prove that the mutual information remains unchanged under spatial coupling in a suitable asymptotic limit (Theorem 10). We will compare the mutual informations of the four following variants of (12). In each case, the signal  $\mathbf{s}$  has  $n(L+1)$  i.i.d components.

- *The fully connected:* If we choose  $w = L/2$  and a homogeneous coupling matrix with elements  $\Lambda_{\mu,\nu} = (L+1)^{-1}$  in (12). This yields a homogeneous fully connected system equivalent to (1) with  $n(L+1)$  instead of  $n$  variables. The associated mutual information per variable for fixed  $L$  and  $n$  is denoted by  $i_{n,L}^{\text{con}}$ .
- *The SC pinned system:* This is the system studied in Section 4 to prove threshold saturation, with the pinning condition. In this case we choose  $0 < w < L/2$ . The coupling matrix  $\mathbf{\Lambda}$  is any matrix that fulfills the requirements in Section 2.3 (the concrete example given there will do). The associated mutual information per variable is here denoted  $i_{n,w,L}^{\text{cou}}$ . Note that  $i_{n,w,L}^{\text{cou}} = (n(L+1))^{-1} I_{w,L}(\mathbf{S}; \mathbf{W})$
- *The periodic SC system:* This is the same SC system (with same coupling window and coupling matrix) but without the pinning condition. The associated mutual information per variable at fixed  $L, w, n$  is denoted  $i_{n,w,L}^{\text{per}}$ .

- *The decoupled system:* This corresponds simply to  $L + 1$  identical and independent systems of the form (1) with  $n$  variables each. This is equivalent to periodic SC system with  $w = 0$ . The associated mutual information per variable is denoted  $i_{n,L}^{\text{dec}}$ . Note that  $i_{n,L}^{\text{dec}} = n^{-1}I(\mathbf{S}; \mathbf{W})$ .

Let us outline the proof strategy. In a first step, we use an interpolation method twice: first interpolating between the fully connected and periodic SC systems, and then between the decoupled and periodic SC systems. This will allow to sandwich the mutual information of the periodic SC system by those of the fully connected and decoupled systems respectively (see Lemma 31). In the second step, using again a similar interpolation and Fekete's theorem for superadditive sequences, we prove that the decoupled and fully connected systems have asymptotically the same mutual information (see Lemma 32 for the existence of the limit). From these results we deduce the proposition:

**Proposition 29** *For any  $0 \leq w \leq L/2$*

$$\lim_{n \rightarrow +\infty} i_{n,w,L}^{\text{per}} = \lim_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) \quad (70)$$

**Proof** Lemma 32 implies that  $\lim_{n \rightarrow +\infty} i_{n,L}^{\text{con}} = \lim_{n \rightarrow +\infty} i_{n,L}^{\text{dec}}$ . One also notes that  $i_{n,L}^{\text{dec}} = \frac{1}{n} I(\mathbf{S}; \mathbf{W})$ . Thus the result follows from Lemma 31.  $\blacksquare$

In a third step an easy argument shows

**Proposition 30** *Assume  $P_0$  has finite first four moments. For any  $0 \leq w \leq L/2$*

$$|i_{n,w,L}^{\text{per}} - i_{n,w,L}^{\text{cou}}| = \mathcal{O}\left(\frac{w}{L}\right) \quad (71)$$

**Proof** See Appendix G.  $\blacksquare$

Since  $i_{n,w,L}^{\text{cou}} = (n(L+1))^{-1} I_{w,L}(\mathbf{S}; \mathbf{W})$ , Theorem 10 is an immediate consequence of Propositions 29 and 30.

## 5.1 A generic interpolation

Let us consider two systems of same total size  $n(L+1)$  with coupling matrices  $\mathbf{\Lambda}^{(1)}$  and  $\mathbf{\Lambda}^{(0)}$  supported on coupling windows  $w_1$  and  $w_0$  respectively. Moreover, we assume that the observations associated with the first system are corrupted by an AWGN equals to  $\sqrt{\Delta/t}\mathbf{z}$  while the AWGN corrupting the second system is  $\sqrt{\Delta/(1-t)}\mathbf{z}'$ , where  $Z_{ij}$  and  $Z'_{ij}$  are two i.i.d. standard Gaussians and  $t \in [0, 1]$  is the *interpolation parameter*. The interpolating inference problem has the form

$$\begin{cases} w_{i_\mu j_\nu} &= s_{i_\mu} s_{j_\nu} \sqrt{\frac{\Lambda_{\mu\nu}^{(1)}}{n}} + z_{i_\mu j_\nu} \sqrt{\frac{\Delta}{t}}, \\ w_{i_\mu j_\nu} &= s_{i_\mu} s_{j_\nu} \sqrt{\frac{\Lambda_{\mu\nu}^{(0)}}{n}} + z'_{i_\mu j_\nu} \sqrt{\frac{\Delta}{1-t}} \end{cases} \quad (72)$$

In this setting, at  $t = 1$  the interpolated system corresponds to the first system as the noise is infinitely large in the second one and no information is available about it, while at  $t = 0$

the opposite happens. The associated interpolating posterior distribution can be expressed as

$$P_t(\mathbf{x}|\mathbf{s}, \mathbf{z}, \mathbf{z}') = \frac{1}{\mathcal{Z}_{\text{int}}(t)} e^{-\mathcal{H}(t, \mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(0)})} \prod_{\mu=0}^L \prod_{i_\mu=1}^n P_0(x_{i_\mu}) \quad (73)$$

where the ‘‘Hamiltonian’’ is  $\mathcal{H}_{\text{int}}(t, \mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(0)}) := \mathcal{H}(t, \mathbf{\Lambda}^{(1)}) + \mathcal{H}(1-t, \mathbf{\Lambda}^{(0)})$  with<sup>4</sup>

$$\begin{aligned} \mathcal{H}(t, \mathbf{\Lambda}) := & \frac{t}{\Delta} \sum_{\mu=0}^L \Lambda_{\mu, \mu} \sum_{i_\mu \leq j_\mu} \left( \frac{x_{i_\mu}^2 x_{j_\mu}^2}{2n} - \frac{s_{i_\mu} s_{j_\mu} x_{i_\mu} x_{j_\mu}}{n} - \frac{x_{i_\mu} x_{j_\mu} z_{i_\mu j_\mu} \sqrt{\Delta}}{\sqrt{nt} \Lambda_{\mu, \mu}} \right) \\ & + \frac{t}{\Delta} \sum_{\mu=0}^L \sum_{\nu=\mu+1}^{\mu+w} \Lambda_{\mu, \nu} \sum_{i_\mu, j_\nu=1}^n \left( \frac{x_{i_\mu}^2 x_{j_\nu}^2}{2n} - \frac{s_{i_\mu} s_{j_\nu} x_{i_\mu} x_{j_\nu}}{n} - \frac{x_{i_\mu} x_{j_\nu} z_{i_\mu j_\nu} \sqrt{\Delta}}{\sqrt{nt} \Lambda_{\mu, \nu}} \right). \end{aligned} \quad (74)$$

and  $\mathcal{Z}_{\text{int}}(t)$  is the obvious normalizing factor, the ‘‘partition function’’. The posterior average with respect to (73) is denoted by the bracket notation  $\langle - \rangle_t$ . It is easy to see that the mutual information per variable (for the interpolating inference problem) can be expressed as

$$i_{\text{int}}(t) := -\frac{1}{n(L+1)} \mathbb{E}_{\mathbf{s}, \mathbf{z}, \mathbf{z}'} [\ln \mathcal{Z}_{\text{int}}(t)] + \frac{v^2}{4\Delta} + \frac{1}{4\Delta n(L+1)} (2\mathbb{E}[S^4] - v^2) \quad (75)$$

The aim of the interpolation method in the present context is to compare the mutual informations of the systems at  $t=1$  and  $t=0$ . To do so, one uses the fundamental theorem of calculus

$$i_{\text{int}}(1) - i_{\text{int}}(0) = \int_0^1 dt \frac{di_{\text{int}}(t)}{dt}. \quad (76)$$

and tries to determine the sign of the integral term.

We first prove that

$$\begin{aligned} \frac{di_{\text{int}}(t)}{dt} = & \frac{1}{4\Delta(L+1)} \mathbb{E}_{\mathbf{s}, \mathbf{z}, \mathbf{z}'} \left[ \left\langle -\frac{1}{n^2} \left( \sum_{\mu=0}^L \sum_{\nu=\mu-w_1}^{\mu+w_1} \Lambda_{\mu\nu}^{(1)} \sum_{i_\mu, j_\nu=1}^n X_{i_\mu} X_{j_\nu} S_{i_\mu} S_{j_\nu} + \sum_{\mu=0}^L \Lambda_{\mu\mu}^{(1)} \sum_{i_\mu=1}^n X_{i_\mu}^2 S_{i_\mu}^2 \right) \right. \right. \\ & \left. \left. + \frac{1}{n^2} \left( \sum_{\mu=0}^L \sum_{\nu=\mu-w_0}^{\mu+w_0} \Lambda_{\mu\nu}^{(0)} \sum_{i_\mu, j_\nu=1}^n X_{i_\mu} X_{j_\nu} S_{i_\mu} S_{j_\nu} + \sum_{\mu=0}^L \Lambda_{\mu\mu}^{(0)} \sum_{i_\mu=1}^n X_{i_\mu}^2 S_{i_\mu}^2 \right) \right\rangle_t \right], \end{aligned} \quad (77)$$

where  $\langle - \rangle_t$  denotes the expectation over the posterior distribution associated with the interpolated Hamiltonian  $\mathcal{H}_{\text{int}}(t, \mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(0)})$ . We start with a simple differentiation of the Hamiltonian w.r.t.  $t$  which yields

$$\frac{d}{dt} \mathcal{H}_{\text{int}}(t, \mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(0)}) = \frac{1}{\Delta} (\mathcal{A}(t, \mathbf{\Lambda}^{(1)}) - \mathcal{B}(t, \mathbf{\Lambda}^{(0)})),$$

---

4. Note that since the SC system is defined on a ring, we can express the Hamiltonian in terms of forward coupling only.

where

$$\begin{aligned}
\mathcal{A}(t, \Lambda^{(1)}) &= \sum_{\mu=0}^L \Lambda_{\mu\mu}^{(1)} \sum_{i_\mu \leq j_\mu} \left( \frac{x_{i_\mu}^2 x_{j_\mu}^2}{2n} - \frac{s_{i_\mu} s_{j_\mu} x_{i_\mu} x_{j_\mu}}{n} - \frac{x_{i_\mu} x_{j_\mu} z_{i_\mu j_\mu} \sqrt{\Delta}}{2\sqrt{nt\Lambda_{\mu\mu}^{(1)}}} \right) \\
&\quad + \sum_{\mu=0}^L \sum_{\nu=\mu+1}^{\mu+w_1} \Lambda_{\mu\nu}^{(1)} \sum_{i_\mu, j_\nu=1}^n \left( \frac{x_{i_\mu}^2 x_{j_\nu}^2}{2n} - \frac{s_{i_\mu} s_{j_\nu} x_{i_\mu} x_{j_\nu}}{n} - \frac{x_{i_\mu} x_{j_\nu} z_{i_\mu j_\nu} \sqrt{\Delta}}{2\sqrt{nt\Lambda_{\mu\nu}^{(1)}}} \right) \\
\mathcal{B}(t, \Lambda^{(0)}) &= \sum_{\mu=0}^L \Lambda_{\mu\mu}^{(0)} \sum_{i_\mu \leq j_\mu} \left( \frac{x_{i_\mu}^2 x_{j_\mu}^2}{2n} - \frac{s_{i_\mu} s_{j_\mu} x_{i_\mu} x_{j_\mu}}{n} - \frac{x_{i_\mu} x_{j_\mu} z'_{i_\mu j_\mu} \sqrt{\Delta}}{2\sqrt{n(1-t)\Lambda_{\mu\mu}^{(0)}}} \right) \\
&\quad + \sum_{\mu=0}^L \sum_{\nu=\mu+1}^{\mu+w_0} \Lambda_{\mu\nu}^{(0)} \sum_{i_\mu, j_\nu=1}^n \left( \frac{x_{i_\mu}^2 x_{j_\nu}^2}{2n} - \frac{s_{i_\mu} s_{j_\nu} x_{i_\mu} x_{j_\nu}}{n} - \frac{x_{i_\mu} x_{j_\nu} z'_{i_\mu j_\nu} \sqrt{\Delta}}{2\sqrt{n(1-t)\Lambda_{\mu\nu}^{(0)}}} \right).
\end{aligned}$$

Using integration by parts with respect to the Gaussian variables  $Z_{ij}$ ,  $Z'_{ij}$ , one gets

$$\mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [Z_{i_\mu j_\nu} \langle X_{i_\mu} X_{j_\nu} \rangle_t] = \sqrt{\frac{t\Lambda_{\mu,\nu}}{n\Delta}} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle X_{i_\mu}^2 X_{j_\nu}^2 \rangle_t - \langle X_{i_\mu} X_{j_\nu} \rangle_t^2] \quad (78)$$

$$\mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [Z'_{i_\mu j_\nu} \langle X_{i_\mu} X_{j_\nu} \rangle_t] = \sqrt{\frac{(1-t)\Lambda_{\mu,\nu}^0}{n\Delta}} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle X_{i_\mu}^2 X_{j_\nu}^2 \rangle_t - \langle X_{i_\mu} X_{j_\nu} \rangle_t^2]. \quad (79)$$

Moreover an application of the Nishimori identity (164) shows

$$\mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle X_{i_\mu} X_{j_\nu} \rangle_t^2] = \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle X_{i_\mu} X_{j_\nu} S_{i_\mu} S_{j_\nu} \rangle_t]. \quad (80)$$

Combining (77)-(80) and using the fact that the SC system defined on a ring satisfies

$$\begin{aligned}
&\sum_{\mu=0}^L \Lambda_{\mu\mu} \sum_{i_\mu \leq j_\mu} x_{i_\mu} x_{j_\mu} s_{i_\mu} s_{j_\mu} + \sum_{\mu=0}^L \sum_{\nu=\mu+1}^{\mu+w} \Lambda_{\mu\nu} \sum_{i_\mu, j_\nu=1}^n x_{i_\mu} x_{j_\nu} s_{i_\mu} s_{j_\nu} = \\
&\frac{1}{2} \sum_{\mu=0}^L \sum_{\nu=\mu-w}^{\mu+w} \Lambda_{\mu\nu} \sum_{i_\mu, j_\nu=1}^n x_{i_\mu} x_{j_\nu} s_{i_\mu} s_{j_\nu} + \frac{1}{2} \sum_{\mu=0}^L \Lambda_{\mu\mu} x_{i_\mu}^2 s_{i_\mu}^2,
\end{aligned}$$

we obtain (77).

Now, define the *overlaps* associated to each block  $\mu$  as

$$q_\mu := \frac{1}{n} \sum_{i_\mu=1}^n X_{i_\mu} S_{i_\mu}, \quad \tilde{q}_\mu := \frac{1}{n} \sum_{i_\mu=1}^n X_{i_\mu}^2 S_{i_\mu}^2. \quad (81)$$

Hence, (77) can be rewritten as

$$\frac{d\dot{i}_{\text{int}}(t)}{dt} = \frac{1}{4\Delta(L+1)} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} \left[ \left\langle \mathbf{q}^\top \Lambda^{(0)} \mathbf{q} - \mathbf{q}^\top \Lambda^{(1)} \mathbf{q} + \frac{1}{n} \left( \tilde{\mathbf{q}}^\top \text{diag}(\Lambda^{(0)}) - \tilde{\mathbf{q}}^\top \text{diag}(\Lambda^{(1)}) \right) \right\rangle_t \right], \quad (82)$$

where  $\mathbf{q}^\top = [q_0 \cdots q_L]$ ,  $\tilde{\mathbf{q}}^\top = [\tilde{q}_0 \cdots \tilde{q}_L]$  are row vectors and  $\text{diag}(\mathbf{\Lambda})$  represents the column vector with entries  $\{\Lambda_{\mu\mu}\}_{\mu=0}^L$ . The coupling matrices  $\mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(0)}$  are real, symmetric, circulant (due to the periodicity of the ring) and thus can be diagonalized in the same Fourier basis. We have

$$\frac{di_{\text{int}}(t)}{dt} = \frac{1}{4\Delta(L+1)} \mathbb{E}_{\mathbf{s}, \mathbf{z}, \mathbf{z}'} \left[ \left\langle \hat{\mathbf{q}}^\top (\mathbf{D}^{(0)} - \mathbf{D}^{(1)}) \hat{\mathbf{q}} + \frac{1}{n} \left( \tilde{\mathbf{q}}^\top \text{diag}(\mathbf{\Lambda}^{(0)}) - \tilde{\mathbf{q}}^\top \text{diag}(\mathbf{\Lambda}^{(1)}) \right) \right\rangle_t \right], \quad (83)$$

where  $\hat{\mathbf{q}}$  is the discrete Fourier transform of  $\mathbf{q}$  and  $\mathbf{D}^{(1)}, \mathbf{D}^{(0)}$  are the diagonal matrices with the eigenvalues of  $\mathbf{\Lambda}^{(1)}, \mathbf{\Lambda}^{(0)}$ . Since the coupling matrices are stochastic with non-negative Fourier transform, their largest eigenvalue equals 1 (and is associated to the 0-th Fourier mode) while the remaining eigenvalues are non-negative. These properties will be essential in the following paragraphs.

## 5.2 Applications

Our first application is

**Lemma 31** *Let the coupling matrix  $\Lambda$  verify the requirements (i)-(v) in Sec. 2.3. The mutual informations of the decoupled, periodic SC and fully connected systems verify*

$$i_{n,L}^{\text{dec}} \leq i_{n,w,L}^{\text{per}} \leq i_{n,L}^{\text{con}}. \quad (84)$$

**Proof** We start with the second inequality. We choose  $\Lambda_{\mu\nu}^{(1)} = (L+1)^{-1}$  for the fully connected system at  $t = 1$ . This matrix has a unique eigenvalue equal to 1 and  $L$  degenerate eigenvalues equal to 0. Therefore it is clear that  $\mathbf{D}^{(0)} - \mathbf{D}^{(1)}$  is positive semi-definite and  $\hat{\mathbf{q}}^\top (\mathbf{D}^{(0)} - \mathbf{D}^{(1)}) \hat{\mathbf{q}} \geq 0$ . Moreover notice that  $\Lambda_{\mu\mu}^{(0)} = \Lambda_{00}$  is independent of  $L$ . Therefore for  $L$  large enough

$$\tilde{\mathbf{q}}^\top \text{diag}(\mathbf{\Lambda}^{(0)}) - \tilde{\mathbf{q}}^\top \text{diag}(\mathbf{\Lambda}^{(1)}) = \left( \Lambda_{00} - \frac{1}{L+1} \right) \sum_{\mu=0}^L \tilde{q}_\mu \geq 0. \quad (85)$$

Therefore we conclude that (83) is positive and from (76)  $i_{n,L}^{\text{con}} - i_{n,w,L}^{\text{per}} \geq 0$ . For the first inequality we proceed similarly, but this time we choose  $\Lambda_{\mu\nu}^{(1)} = \delta_{\mu\nu}$  for the decoupled system which has all eigenvalues equal to 1. Therefore  $\mathbf{D}^{(0)} - \mathbf{D}^{(1)}$  is negative semidefinite so  $\hat{\mathbf{q}}^\top (\mathbf{D}^{(0)} - \mathbf{D}^{(1)}) \hat{\mathbf{q}} \leq 0$ . Moreover this time

$$\tilde{\mathbf{q}}^\top \text{diag}(\mathbf{\Lambda}^{(0)}) - \tilde{\mathbf{q}}^\top \text{diag}(\mathbf{\Lambda}^{(1)}) = \left( \Lambda_{00} - 1 \right) \sum_{\mu=0}^L \tilde{q}_\mu \leq 0 \quad (86)$$

because we necessarily have  $0 \leq \Lambda_{00}^{(0)} \leq 1$ . We conclude that (83) is negative and from (76)  $i_{n,L}^{\text{dec}} - i_{n,w,L}^{\text{per}} \leq 0$ . ■

The second application is

**Lemma 32** Consider the mutual information of system (1) and set  $i_n = n^{-1}I(\mathbf{S}; \mathbf{W})$ . Consider also  $i_{n_1}$  and  $i_{n_2}$  the mutual informations of two systems of size  $n_1$  and  $n_2$  with  $n = n_1 + n_2$ . The sequence  $ni_n$  is superadditive in the sense that

$$n_1 i_{n_1} + n_2 i_{n_2} \leq n i_n. \quad (87)$$

Fekete's lemma then implies that  $\lim_{n \rightarrow +\infty} i_n$  exists.

**Proof** The proof is easily obtained by following the generic interpolation method of Sec. 5.1 for a coupled system with two spatial positions (i.e.  $L + 1 = 2$ ). We choose  $\Lambda_{\mu\nu}^{(0)} = \delta_{\mu\nu}$ ,  $\mu, \nu \in 0, 1$  for the "decoupled" system and  $\Lambda_{\mu\nu}^{(1)} = 1/2$  for  $\mu, \nu \in 0, 1$  for the "fully connected" system. This analysis is essentially identical to [Guerra and Toninelli (2002)] where the existence of the thermodynamic limit of the free energy for the Sherrington-Kirkpatrick mean field spin glass is proven.  $\blacksquare$

## 6. Proof of the replica symmetric formula (Theorem 4)

In this section we provide the proof of the RS formula for the mutual information of the underlying model (Theorem 4) for  $0 < \Delta \leq \Delta_{\text{opt}}$  (Proposition 39) and then for  $\Delta \geq \Delta_{\text{opt}}$  (Proposition 41). For  $0 < \Delta \leq \Delta_{\text{opt}}$  the proof directly follows from the I-MMSE relation Lemma 33, the replica bound (16) and the suboptimality of the AMP algorithm. In this interval the proof doesn't require spatial coupling. For  $\Delta \geq \Delta_{\text{opt}}$  the proof uses the results of Sections 4 and 5 on the spatially coupled model.

Let us start with two preliminary lemmas. The first is an I-MMSE relation [Guo et al. (2005)] adapted to the current matrix estimation problem.

**Lemma 33** Let  $P_0$  has finite first four moments. The mutual information and the matrix-MMSE are related by

$$\frac{1}{n} \frac{dI(\mathbf{S}; \mathbf{W})}{d\Delta^{-1}} = \frac{1}{4} \text{Mmmse}_n(\Delta^{-1}) + \mathcal{O}(1/n). \quad (88)$$

**Proof**

$$\begin{aligned} \frac{1}{n} \frac{dI(\mathbf{S}; \mathbf{W})}{d\Delta^{-1}} &= \frac{1}{2n^2} \mathbb{E}_{\mathbf{S}, \mathbf{W}} \left[ \sum_{i \leq j} (S_i S_j - \mathbb{E}[X_i X_j | \mathbf{W}])^2 \right] \\ &= \frac{1}{4n^2} \mathbb{E}_{\mathbf{S}, \mathbf{W}} \left[ \|\mathbf{S}\mathbf{S}^\top - \mathbb{E}[\mathbf{X}\mathbf{X}^\top | \mathbf{W}]\|_{\text{F}}^2 \right] + \frac{1}{4n^2} \sum_{i=1}^n \mathbb{E}_{\mathbf{S}, \mathbf{W}} [(S_i^2 - \mathbb{E}[X_i^2 | \mathbf{W}])^2] \\ &= \frac{1}{4} \text{Mmmse}_n(\Delta^{-1}) + \mathcal{O}(1/n), \end{aligned} \quad (89)$$

The proof details for first equality are in Appendix C. The second equality is obtained by completing the sum and accounting for the diagonal terms. The last equality is obtained from

$$\begin{aligned} \mathbb{E}_{\mathbf{S}, \mathbf{W}} [(S_i^2 - \mathbb{E}[X_i^2 | \mathbf{W}])^2] &= \mathbb{E}[S_i^4] - 2\mathbb{E}_{S_i, \mathbf{W}} [S_i^2 \mathbb{E}[X_i^2 | \mathbf{W}]] + \mathbb{E}_{\mathbf{W}} [\mathbb{E}[X_i^2 | \mathbf{W}]^2] \\ &= \mathbb{E}[S_i^4] - \mathbb{E}_{\mathbf{W}} [\mathbb{E}[X_i^2 | \mathbf{W}]^2] \\ &\leq \mathbb{E}[S_i^4]. \end{aligned} \quad (90)$$



where we have used the Nishimori identity  $\mathbb{E}_{S_i, \mathbf{W}}[S_i^2 \mathbb{E}[X_i^2 | \mathbf{W}]] = \mathbb{E}_{\mathbf{W}}[\mathbb{E}[X_i^2 | \mathbf{W}]^2]$  in the second equality (Appendix D).  $\blacksquare$

**Lemma 34** *The limit  $\lim_{n \rightarrow +\infty} n^{-1} I(\mathbf{S}; \mathbf{W})$  exists and is a concave, continuous, function of  $\Delta$ .*

**Proof** The existence of the limit is the statement of Lemma 32 in Sec. 5. The continuity follows from the concavity of the mutual information with respect to  $\Delta^{-1}$ : because the limit of a sequence of concave functions remains concave, and thus it is continuous. To see the concavity notice that the first derivative of the mutual information w.r.t  $\Delta^{-1}$  equals the matrix-MMSE (Lemma 33) and that the later cannot increase as a function of  $\Delta^{-1}$ .  $\blacksquare$

### 6.1 Proof of Theorem 4 for $0 < \Delta \leq \Delta_{\text{opt}}$

**Lemma 35** *Assume  $P_0$  is a discrete distribution. Fix  $\Delta < \Delta_{\text{AMP}}$ . The mutual information per variable is asymptotically given by the RS formula (15).*

**Proof** By the suboptimality of the AMP algorithm we have

$$\text{Mmse}_{n, \text{AMP}}^{(t)}(\Delta^{-1}) \geq \text{Mmmse}_n(\Delta^{-1}). \quad (91)$$

Taking limits in the order  $\lim_{t \rightarrow +\infty} \limsup_{n \rightarrow +\infty}$  and using (11) we find

$$v^2 - (v - E^{(\infty)})^2 \geq \limsup_{n \rightarrow +\infty} \text{Mmmse}_n(\Delta^{-1}). \quad (92)$$

Furthermore, by applying Lemma 33 we obtain

$$\frac{v^2 - (v - E^{(\infty)})^2}{4} \geq \limsup_{n \rightarrow +\infty} \frac{1}{n} \frac{dI(\mathbf{S}; \mathbf{W})}{d\Delta^{-1}}. \quad (93)$$

Now, for  $\Delta < \Delta_{\text{AMP}}$  we have  $E^{(\infty)} = E_{\text{good}}(\Delta)$  which is the unique and hence *global* minimum of  $i_{\text{RS}}(E; \Delta)$  over  $E \in [0, v]$ . Moreover, for  $\Delta < \Delta_{\text{AMP}}$  we have that  $E^{(\infty)}(\Delta)$  is continuously differentiable  $\Delta^{-1}$  with locally bounded derivative. Thus

$$\begin{aligned} \frac{d}{d\Delta^{-1}} \left( \min_{E \in [0, v]} i_{\text{RS}}(E; \Delta) \right) &= \frac{di_{\text{RS}}}{d\Delta^{-1}}(E^{(\infty)}; \Delta) \\ &= \frac{\partial i_{\text{RS}}}{\partial E}(E^{(\infty)}; \Delta) \frac{dE^{(\infty)}}{d\Delta^{-1}} + \frac{\partial i_{\text{RS}}}{\partial \Delta^{-1}}(E^{(\infty)}; \Delta) \\ &= \frac{\partial i_{\text{RS}}}{\partial \Delta^{-1}}(E^{(\infty)}; \Delta) \\ &= \frac{(v - E^{(\infty)})^2 + v^2}{4} - \frac{\partial \mathbb{E}_{S, Z}[\dots]}{\partial \Sigma^{-2}} \Big|_{E^{(\infty)}} \frac{\partial \Sigma^{-2}}{\partial \Delta^{-1}} \Big|_{E^{(\infty)}} \\ &= \frac{v^2 - (v - E^{(\infty)})^2}{4}, \end{aligned} \quad (94)$$

where  $\mathbb{E}_{S,Z}[\dots]$  is the expectation that appears in the RS potential (14). The third equality is obtained from

$$\left. \frac{\partial \Sigma^{-2}}{\partial \Delta^{-1}} \right|_{E^{(\infty)}} = v - E^{(\infty)} \quad (95)$$

and

$$\left. \frac{\partial \mathbb{E}_{S,Z}[\dots]}{\partial \Sigma^{-2}} \right|_{E^{(\infty)}} = \frac{1}{2}(v - E^{(\infty)}). \quad (96)$$

This last identity immediately follows from  $\left. \frac{\partial i_{\text{RS}}}{\partial E} \right|_{E^{(\infty)}} = 0$ . From (93) and (94)

$$\frac{d}{d\Delta^{-1}} \left( \min_{E \in [0,v]} i_{\text{RS}}(E; \Delta) \right) \geq \limsup_{n \rightarrow +\infty} \frac{1}{n} \frac{dI(\mathbf{S}; \mathbf{W})}{d\Delta^{-1}}, \quad (97)$$

which is equivalent to

$$\frac{d}{d\Delta} \left( \min_{E \in [0,v]} i_{\text{RS}}(E; \Delta) \right) \leq \liminf_{n \rightarrow +\infty} \frac{1}{n} \frac{dI(\mathbf{S}; \mathbf{W})}{d\Delta}. \quad (98)$$

We now integrate inequality (98) over an interval  $[0, \Delta] \subset [0, \Delta_{\text{AMP}}[$

$$\begin{aligned} \min_{E \in [0,v]} i_{\text{RS}}(E; \Delta) - \min_{E \in [0,v]} i_{\text{RS}}(E; 0) &\leq \int_0^\Delta d\tilde{\Delta} \liminf_{n \rightarrow +\infty} \frac{1}{n} \frac{dI(\mathbf{S}; \mathbf{W})}{d\tilde{\Delta}} \\ &\leq \liminf_{n \rightarrow +\infty} \int_0^\Delta d\tilde{\Delta} \frac{1}{n} \frac{dI(\mathbf{S}; \mathbf{W})}{d\tilde{\Delta}} \\ &= \liminf_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) - H(S). \end{aligned} \quad (99)$$

The second inequality uses Fatou's Lemma and the last equality uses that for a discrete prior

$$\lim_{\Delta \rightarrow 0_+} I(\mathbf{S}; \mathbf{W}) = H(\mathbf{S}) - \lim_{\Delta \rightarrow 0_+} H(\mathbf{S}|\mathbf{W}) = nH(S). \quad (100)$$

In Appendix F an explicit calculation shows that  $\min_E i_{\text{RS}}(E; 0) = H(S)$ . Therefore

$$\min_{E \in [0,v]} i_{\text{RS}}(E; \Delta) \leq \liminf_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}). \quad (101)$$

The final step combines inequality (101) with the replica bound (16) to obtain

$$\min_{E \in [0,v]} i_{\text{RS}}(E; \Delta) \leq \liminf_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) \leq \limsup_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) \leq \min_{E \in [0,v]} i_{\text{RS}}(E; \Delta). \quad (102)$$

This shows that the limit of the mutual information exists and is equal to the RS formula for  $\Delta < \Delta_{\text{AMP}}$ . Note that in this proof we did not need the a-priori existence of the limit.  $\blacksquare$

**Remark 36** One can try to apply the same proof idea to the regime  $\Delta > \Delta_{\text{RS}}$ . Equations (91)-(98) work out exactly in the same way because the AMP fixed point  $E^{(\infty)}$  is a global minimum of  $i_{\text{RS}}(E; \Delta)$ . Then when integrating on  $] \Delta, +\infty[ \subset [ \Delta_{\text{RS}}, +\infty[$ , one finds

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) \leq \min_{E \in [0, v]} i_{\text{RS}}(E; \Delta). \quad (103)$$

This essentially gives an alternative proof of (16) for  $\Delta > \Delta_{\text{RS}}$ .

**Lemma 37** We necessarily have  $\Delta_{\text{AMP}} \leq \Delta_{\text{opt}}$ .

**Proof** Notice first that it not possible to have  $\Delta_{\text{RS}} < \Delta_{\text{AMP}}$  because in the range  $]0, \Delta_{\text{AMP}}[$ , as a function of  $E$ , the function  $i_{\text{RS}}(E; \Delta)$  has a unique stationary point. Since  $\min_{E \in [0, v]} i_{\text{RS}}(E; \Delta)$  is analytic for  $\Delta < \Delta_{\text{RS}}$ , it is analytic for  $\Delta < \Delta_{\text{AMP}}$ . Now we proceed by contradiction: suppose we would have  $\Delta_{\text{AMP}} \geq \Delta_{\text{opt}}$ . Lemma 35 asserts that  $\lim_{n \rightarrow +\infty} n^{-1} I(\mathbf{S}; \mathbf{W}) = \min_{E \in [0, v]} i_{\text{RS}}(E; \Delta)$  for  $\Delta < \Delta_{\text{AMP}}$  thus we would have  $\lim_{n \rightarrow +\infty} n^{-1} I(\mathbf{S}; \mathbf{W})$  analytic at  $\Delta_{\text{opt}}$ . This is a contradiction by definition of  $\Delta_{\text{opt}}$ . ■

**Lemma 38** We necessarily have  $\Delta_{\text{RS}} \geq \Delta_{\text{opt}}$ .

**Proof** If  $\Delta_{\text{RS}} = +\infty$  then we are done, so we suppose it is finite. The proof proceeds by contradiction: suppose  $\Delta_{\text{RS}} < \Delta_{\text{opt}}$ . So we assume  $\Delta_{\text{RS}} \in [ \Delta_{\text{AMP}}, \Delta_{\text{opt}}[$  (in the previous lemma we showed that this must be the case). For  $\Delta \in ]0, \Delta_{\text{RS}}[$  we have  $\min_{E \in [0, v]} i_{\text{RS}}(E; \Delta) = i_{\text{RS}}(E_{\text{good}}(\Delta); \Delta)$  which is an analytic function in this interval. By definition of  $\Delta_{\text{opt}}$ , the function  $\lim_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W})$  is analytic in  $]0, \Delta_{\text{opt}}[$ . Therefore both functions are analytic on  $]0, \Delta_{\text{RS}}[$  and since by Lemma 35 they are equal for  $]0, \Delta_{\text{AMP}}[ \subset ]0, \Delta_{\text{RS}}[$ , they must be equal on the whole range  $]0, \Delta_{\text{RS}}[$ . This implies that the two functions are equal at  $\Delta_{\text{RS}}$  because they are continuous. Explicitly,

$$\min_E i_{\text{RS}}(E; \Delta) = \lim_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W})|_{\Delta} \quad \forall \Delta \in ]0, \Delta_{\text{RS}}]. \quad (104)$$

Now, fix some  $\Delta \in ] \Delta_{\text{RS}}, \Delta_{\text{opt}}[$ . Since this  $\Delta$  is greater than  $\Delta_{\text{RS}}$  the fixed point of state evolution  $E^{(\infty)}$  is also the global minimum of  $i_{\text{RS}}(E; \Delta)$ . Hence exactly as in (91)-(98) we can show that for  $\Delta \in ] \Delta_{\text{RS}}, \Delta_{\text{opt}}[$ , (98) is verified. This time, combining (16), (98) and the assumption  $\Delta_{\text{RS}} \in [ \Delta_{\text{AMP}}, \Delta_{\text{opt}}[$ , leads to a contradiction, and hence we must have  $\Delta_{\text{RS}} \geq \Delta_{\text{opt}}$ . To see explicitly how the contradiction appears, integrate (98) on  $] \Delta_{\text{RS}}, \Delta[ \subset ] \Delta_{\text{RS}}, \Delta_{\text{opt}}[$ , and use Fatou's Lemma, to obtain

$$\begin{aligned} \min_E i_{\text{RS}}(E; \Delta) - \min_E i_{\text{RS}}(E; \Delta_{\text{RS}}) &\leq \liminf_{n \rightarrow +\infty} \left( \frac{1}{n} I(\mathbf{S}; \mathbf{W})|_{\Delta} - \frac{1}{n} I(\mathbf{S}; \mathbf{W})|_{\Delta_{\text{RS}}} \right) \\ &= \lim_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W})|_{\Delta} - \lim_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W})|_{\Delta_{\text{RS}}}. \end{aligned} \quad (105)$$

From (104) and (16) we obtain  $\min_E i_{\text{RS}}(E; \Delta) = \lim_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W})$  when  $\Delta_{\text{AMP}} \leq \Delta_{\text{RS}} < \Delta < \Delta_{\text{opt}}$ . But from (104), this equality is also true for  $0 < \Delta \leq \Delta_{\text{RS}}$ . So the equality is valid in the whole interval  $]0, \Delta_{\text{opt}}[$  and therefore  $\min_E i_{\text{RS}}(E; \Delta)$  is analytic at  $\Delta_{\text{RS}}$ . But this is impossible by the definition of  $\Delta_{\text{RS}}$ . ■

**Proposition 39** *Assume  $P_0$  is a discrete distribution. Fix  $\Delta \leq \Delta_{\text{opt}}$ . The mutual information per variable is asymptotically given by the RS formula (15).*

**Proof** Lemma 35 says that the two functions,  $\lim_{n \rightarrow +\infty} n^{-1} I(\mathbf{S}; \mathbf{W})$  and  $\min_{E \in [0, v]} i_{\text{RS}}(E; \Delta)$ , are equal for  $\Delta < \Delta_{\text{AMP}}$  and Lemma 38 implies that *both* functions are analytic for  $\Delta < \Delta_{\text{opt}}$ . Thus they must be equal on the whole range  $\Delta < \Delta_{\text{opt}}$ . Since we also know they are continuous, then they are equal also at  $\Delta = \Delta_{\text{opt}}$ . ■

## 6.2 Proof of Theorem 4 for $\Delta \geq \Delta_{\text{opt}}$

We first need the following lemma where spatial coupling comes into the play.

**Lemma 40** *The optimal threshold is given by the potential threshold:  $\Delta_{\text{opt}} = \Delta_{\text{RS}}$ .*

**Proof** It suffices to see that

$$\Delta_{\text{RS}} \leq \Delta_{\text{AMP}}^c \leq \Delta_{\text{opt}}^c = \Delta_{\text{opt}} \leq \Delta_{\text{RS}}. \quad (106)$$

The first inequality is the threshold saturation result of Theorem 11 in Section 4. The second inequality is due the suboptimality of the AMP algorithm.<sup>5</sup> The equality is a consequence of Theorem 10 in Section 5. Indeed, equality of asymptotic mutual informations of the coupled and underlying system implies that they must be non-analytic at the same value of  $\Delta$ . Finally, the last inequality is the statement of Lemma 38 in Section 6.1. ■

**Proposition 41** *Assume  $P_0$  is a discrete distribution. Fix  $\Delta \geq \Delta_{\text{opt}}$ . The mutual information per variable is asymptotically given by the RS formula (15).*

**Proof** We already remarked in section 6.1 that for  $\Delta > \Delta_{\text{RS}}$ ,

$$\frac{d}{d\Delta} (\min_E i_{\text{RS}}(E; \Delta)) \leq \liminf_{n \rightarrow +\infty} \frac{1}{n} \frac{dI(\mathbf{S}; \mathbf{W})}{d\Delta}. \quad (107)$$

Now we integrate on an interval  $]\Delta_{\text{RS}}, \Delta]$  both sides of the inequality. Since from Lemma 40 we have that  $\Delta_{\text{RS}} = \Delta_{\text{opt}}$ , it is equivalent to integrate from  $\Delta_{\text{opt}}$  upwards<sup>6</sup>

$$\int_{\Delta_{\text{opt}}}^{\Delta} d\tilde{\Delta} \frac{d}{d\tilde{\Delta}} (\min_E i_{\text{RS}}(E; \tilde{\Delta})) \leq \int_{\Delta_{\text{opt}}}^{\Delta} d\tilde{\Delta} \liminf_{n \rightarrow +\infty} \frac{1}{n} \frac{dI(\mathbf{S}; \mathbf{W})}{d\tilde{\Delta}}. \quad (108)$$

By Fatou's lemma the inequality is preserved if we bring the  $\liminf$  outside of the integral, thus

$$\begin{aligned} \min_E i_{\text{RS}}(E; \Delta) - \min_E i_{\text{RS}}(E; \Delta_{\text{RS}}) &\leq \liminf_{n \rightarrow +\infty} \left\{ \frac{1}{n} I(\mathbf{S}; \mathbf{W}) \Big|_{\Delta} - \frac{1}{n} I(\mathbf{S}; \mathbf{W}) \Big|_{\Delta_{\text{opt}}} \right\} \\ &= \lim_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) \Big|_{\Delta} - \lim_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) \Big|_{\Delta_{\text{opt}}}. \end{aligned} \quad (109)$$

5. More precisely one shows by the same methods Lemmas 35 and 37 for the spatially coupled system.

6. This is the point we did not yet know in section 6.1.

To get the last line we have used the existence of the thermodynamic limit (see Lemma 34). But we already know from Proposition 39 that  $\min_E i_{\text{RS}}(E; \Delta_{\text{opt}}) = \lim_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) \Big|_{\Delta_{\text{opt}}}$ . Therefore

$$\min_E i_{\text{RS}}(E; \Delta) \leq \lim_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}), \quad (110)$$

which together with (16) ends the proof.  $\blacksquare$

## 7. Proof of corollaries 7 and 8

In this section, we provide the proofs of Corollary 7 and Corollary 8 concerning the MMSE formulae and the optimality of the AMP algorithm. We first show the following result about the matrix and vector MMSE's in Definition 6.

**Lemma 42** *Assume the prior  $P_0$  has finite first four moments and recall the second moment is called  $v$ . The matrix and vector MMSE verify*

$$\text{Mmmse}_n \leq (v^2 - (v - \text{Vmmse}_n)^2) + \mathcal{O}\left(\frac{1}{n}\right). \quad (111)$$

**Proof** For this proof we denote  $\langle \cdot \rangle$  the expectation w.r.t the posterior distribution (20). The matrix and vector MMSE then read

$$\text{Mmmse}_n := \frac{1}{n^2} \mathbb{E}_{\mathbf{S}, \mathbf{W}} \left[ \|\mathbf{S}\mathbf{S}^\top - \langle \mathbf{X}\mathbf{X}^\top \rangle\|_{\text{F}}^2 \right], \quad (112)$$

$$\text{Vmmse}_n := \frac{1}{n} \mathbb{E}_{\mathbf{S}, \mathbf{W}} \left[ \|\mathbf{S} - \langle \mathbf{X} \rangle\|_2^2 \right]. \quad (113)$$

Expanding the Frobenius norm in (112) yields

$$\begin{aligned} \text{Mmmse}_n &= \frac{1}{n^2} \mathbb{E}_{\mathbf{S}, \mathbf{W}} \left[ \sum_{i,j=1}^n (S_i S_j - \langle X_i X_j \rangle)^2 \right] = \frac{1}{n^2} \mathbb{E}_{\mathbf{S}, \mathbf{W}} \left[ \sum_{i,j=1}^n S_i^2 S_j^2 - \langle X_i X_j \rangle^2 \right] \\ &= \mathbb{E}_{\mathbf{S}} \left[ \left( \frac{1}{n} \sum_{i=1}^n S_i^2 \right)^2 \right] - \frac{1}{n^2} \sum_{i,j=1}^n \mathbb{E}_{\mathbf{S}, \mathbf{W}} [\langle X_i X_j \rangle^2], \end{aligned} \quad (114)$$

where the second equality follows from  $\mathbb{E}_{\mathbf{S}, \mathbf{W}} [\langle X_i X_j \rangle^2] = \mathbb{E}_{\mathbf{S}, \mathbf{W}} [S_i S_j \langle X_i X_j \rangle]$ , implied by the Nishimori identity (164). Similarly, using  $\mathbb{E}_{\mathbf{S}, \mathbf{W}} [\langle X_i \rangle^2] = \mathbb{E}_{\mathbf{S}, \mathbf{W}} [S_i \langle X_i \rangle]$  implied by the Nishimori identity, (113) simplifies to

$$\text{Vmmse}_n = v - \frac{1}{n} \sum_{i=1}^n \mathbb{E}_{\mathbf{S}, \mathbf{W}} [\langle X_i \rangle^2]. \quad (115)$$

Hence,

$$\text{Mmmse}_n - (v^2 - (v - \text{Vmmse}_n)^2) = \mathcal{A}_n - \mathcal{B}_n, \quad (116)$$

with

$$\mathcal{A}_n := \mathbb{E}_{\mathbf{S}} \left[ \left( \frac{1}{n} \sum_{i=1}^n S_i^2 \right)^2 \right] - v^2, \quad (117)$$

$$\mathcal{B}_n := \frac{1}{n^2} \sum_{i,j=1}^n \left( \mathbb{E}_{\mathbf{S}, \mathbf{W}}[\langle X_i X_j \rangle^2] - \mathbb{E}_{\mathbf{S}, \mathbf{W}}[\langle X_i \rangle^2] \mathbb{E}_{\mathbf{S}, \mathbf{W}}[\langle X_j \rangle^2] \right). \quad (118)$$

Since the signal components  $\{S_i\}$  are i.i.d and  $P_0$  has finite first four moments,  $\mathcal{A}_n = \mathcal{O}(1/n)$ . It remains to show that  $\mathcal{B}_n \geq 0$ . This is most easily seen as follows. By defining the *overlap*

$$q(\mathbf{X}, \mathbf{S}) := \frac{1}{n} \sum_{i=1}^n S_i X_i \quad (119)$$

and using the Nishimori identities  $\mathbb{E}_{\mathbf{S}, \mathbf{W}}[\langle X_i \rangle^2] = \mathbb{E}_{\mathbf{S}, \mathbf{W}}[S_i \langle X_i \rangle]$  and  $\mathbb{E}_{\mathbf{S}, \mathbf{W}}[\langle X_i X_j \rangle^2] = \mathbb{E}_{\mathbf{S}, \mathbf{W}}[S_i S_j \langle X_i X_j \rangle]$ , we observe that

$$\begin{aligned} \mathcal{B}_n &= \mathbb{E}_{\mathbf{S}, \mathbf{W}}[\langle q^2 \rangle] - \mathbb{E}_{\mathbf{S}, \mathbf{W}}[\langle q \rangle]^2 \\ &= \mathbb{E}_{\mathbf{S}, \mathbf{W}}[(q - \mathbb{E}_{\mathbf{S}, \mathbf{W}}[\langle q \rangle])^2] \end{aligned} \quad (120)$$

which is non-negative. ■

**Remark 43** *Using ideas similar to [Korada and Macris (2009)] to prove concentration of overlaps in inference problems suggest that Lemma 42 holds with an equality when suitable “side observations” are added.*

### 7.1 Proof of Corollary 7

We first show how to prove the expression (22) for the asymptotic Mmmse $_n$  by taking the limit  $n \rightarrow +\infty$  on both sides of (88). First notice that since  $n^{-1}I(\mathbf{S}; \mathbf{W})$  is a sequence of concave functions with respect to  $\Delta^{-1}$ , the limit when  $n \rightarrow +\infty$  is also concave and differentiable for almost all  $\Delta^{-1}$  and at all differentiability points we have (by a standard theorem of real analysis on convex functions)

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \frac{d}{d\Delta^{-1}} I(\mathbf{S}; \mathbf{W}) = \frac{1}{n} \frac{d}{d\Delta^{-1}} \lim_{n \rightarrow +\infty} I(\mathbf{S}; \mathbf{W}). \quad (121)$$

Thus from Lemma 33 and Theorem 4 we have for all  $\Delta \neq \Delta_{\text{RS}}$

$$\lim_{n \rightarrow +\infty} \text{Mmmse}_n(\Delta^{-1}) = 4 \frac{d}{d\Delta^{-1}} \min_{E \in [0, v]} i_{\text{RS}}(E; \Delta). \quad (122)$$

It remains to compute the right hand side. Let  $E_0(\Delta)$  denote the (global) minimum of  $i_{\text{RS}}(E; \Delta)$ . For  $\Delta \neq \Delta_{\text{RS}}$  this is a differentiable function of  $\Delta$  with locally bounded deriva-

tive. Hence using a similar calculation to the one done in (94), we obtain

$$\begin{aligned}
\frac{d}{d\Delta^{-1}} \left( \min_{E \in [0, v]} i_{\text{RS}}(E; \Delta) \right) &= \frac{di_{\text{RS}}}{d\Delta^{-1}}(E_0; \Delta) \\
&= \frac{\partial i_{\text{RS}}}{\partial E}(E_0; \Delta) \frac{dE_0}{d\Delta^{-1}} + \frac{\partial i_{\text{RS}}}{\partial \Delta^{-1}}(E_0; \Delta) \\
&= \frac{\partial i_{\text{RS}}}{\partial \Delta^{-1}}(E_0; \Delta).
\end{aligned} \tag{123}$$

To compute the partial derivative with respect to  $\Delta$  we first note that  $\left. \frac{\partial i_{\text{RS}}}{\partial E} \right|_{E_0} = 0$  implies

$$0 = -\frac{v - E_0}{2\Delta} - \left. \frac{\partial \mathbb{E}_{S,Z}[\dots]}{\partial \Sigma^{-2}} \right|_{E_0} \left. \frac{\partial \Sigma^{-2}}{\partial E} \right|_{E_0}, \tag{124}$$

where  $\mathbb{E}_{S,Z}[\dots]$  is the expectation that appears in the RS potential (14). This immediately gives

$$\left. \frac{\partial \mathbb{E}_{S,Z}[\dots]}{\partial \Sigma^{-2}} \right|_{E_0} = \frac{1}{2}(v - E_0) \tag{125}$$

Thus

$$\begin{aligned}
\frac{\partial i_{\text{RS}}}{\partial \Delta^{-1}}(E_0; \Delta) &= \frac{(v - E_0)^2 + v^2}{4} - \left. \frac{\partial \mathbb{E}_{S,Z}[\dots]}{\partial \Sigma^{-2}} \right|_{E_0} \left. \frac{\partial \Sigma^{-2}}{\partial \Delta^{-1}} \right|_{E_0} \\
&= \frac{v^2 - (v - E_0)^2}{4}.
\end{aligned} \tag{126}$$

From (122), (123), (126) we obtain the desired result, formula (22).

We now turn to the proof of (23) for the expression of the asymptotic vector-MMSE. From Lemma 42 and the suboptimality of the AMP algorithm (here  $E_0(\Delta)$  is the global minimum of  $i_{\text{RS}}(E; \Delta)$  and  $E^{(\infty)}$  the fixed point of state evolution)

$$\begin{aligned}
v^2 - (v - E_0)^2 &= \lim_{n \rightarrow \infty} \text{Mmmse}_n \\
&\leq \liminf_{n \rightarrow \infty} (v^2 - (v - \text{Vmmse}_n)^2) \\
&\leq \limsup_{n \rightarrow \infty} (v^2 - (v - \text{Vmmse}_n)^2) \\
&\leq v^2 - (v - E^{(\infty)})^2,
\end{aligned} \tag{127}$$

For  $\Delta \notin [\Delta_{\text{AMP}}, \Delta_{\text{RS}}]$ , we have that  $E_0 = E^{(\infty)}$  which ends the proof.

## 7.2 Proof of Corollary 8

In view of (11) we have

$$\lim_{t \rightarrow +\infty} \lim_{n \rightarrow +\infty} \text{Vmse}_{n, \text{AMP}}^{(t)}(\Delta^{-1}) = E^{(\infty)}, \tag{128}$$

$$\lim_{t \rightarrow +\infty} \lim_{n \rightarrow +\infty} \text{Mmse}_{n, \text{AMP}}^{(t)}(\Delta^{-1}) = v^2 - (v - E^{(\infty)})^2. \tag{129}$$

For  $\Delta \notin [\Delta_{\text{AMP}}, \Delta_{\text{RS}}]$  we have  $E^{(\infty)} = \operatorname{argmin}_{E \in [0, v]} i_{\text{RS}}(E; \Delta)$  and also the two formulas of Corollary 7 hold. This directly implies the formulas (24) and (25).

On the other hand for  $\Delta \in [\Delta_{\text{AMP}}, \Delta_{\text{RS}}]$  we have  $E^{(\infty)} > \operatorname{argmin}_{E \in [0, v]} i_{\text{RS}}(E; \Delta)$ , so using the monotonicity of  $E^{(t)}$  leads to strict inequalities in (128) and (129) and thus to (26) and (27).

## Appendix A. Upper bound on the mutual information

For the completeness of this work we revisit the proof of the upper bound (16) on mutual information. This result was already obtained by [Krzakala et al. (2016)] using a Toninelli-Guerra type interpolation and is used in this paper so we only sketch the main steps.

We consider the following interpolating inference problem

$$\begin{cases} w_{ij} = \frac{s_i s_j}{\sqrt{n}} + \sqrt{\frac{\Delta}{t}} z_{ij}, \\ y_i = s_i + \sqrt{\frac{\Delta}{m(1-t)}} z'_i, \end{cases} \quad (130)$$

with  $m := v - E \in [0, v]$ , and  $Z'_i \sim \mathcal{N}(0, 1)$ . For  $t = 1$  we find back the original problem (1) since the  $y_i$  observations become useless and for  $t = 0$  we have a set of decoupled observations from a Gaussian channel. The interpolating posterior distribution associated to this set of observations is

$$P_t(\mathbf{x} | \mathbf{s}, \mathbf{z}, \mathbf{z}') := \frac{e^{-\mathcal{H}(t)} \prod_{i=1}^n P_0(x_i)}{\int \{ \prod_{i=1}^n dx_i P_0(x_i) \} e^{-\mathcal{H}(t)}} := \frac{1}{\mathcal{Z}(t)} e^{-\mathcal{H}(t)} \prod_{i=1}^n P_0(x_i) \quad (131)$$

where

$$\begin{aligned} \mathcal{H}(t) = & \sum_{i \leq j=1}^n \left( \frac{t}{2\Delta n} x_i^2 x_j^2 - \frac{t}{\Delta n} x_i x_j s_i s_j - \sqrt{\frac{t}{n\Delta}} x_i x_j z_{ij} \right) \\ & + \sum_{i=1}^n \left( \frac{m(1-t)}{2\Delta} x_i^2 - \frac{m(1-t)}{\Delta} x_i s_i + \sqrt{\frac{m(1-t)}{\Delta}} x_i z'_i \right). \end{aligned}$$

can be interpreted as a ‘‘Hamiltonian’’ and the normalizing factor  $\mathcal{Z}(t)$  is interpreted as a ‘‘partition function’’. We adopt the Gibbs ‘‘bracket’’ notation  $\langle - \rangle_t$  for the expectation with respect to the posterior (131). The mutual information associated to interpolating inference problem is

$$i(t) = -\frac{1}{n} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\ln \mathcal{Z}(t)] + \frac{v^2}{4\Delta} + \frac{1}{4\Delta n} (2\mathbb{E}[S^4] - v^2) \quad (132)$$

Note that on one hand  $i(1) = \frac{1}{n} I(\mathbf{S}; \mathbf{W})$  the mutual information of the original matrix factorization problem and on the other hand

$$\begin{aligned} i_0 &= -\frac{1}{n} \mathbb{E}_{\mathbf{S}, \mathbf{Z}'} \left[ \ln \left( \int \left\{ \prod_{i=1}^n dx_i P_0(x_i) \right\} e^{-\frac{m \|\mathbf{x}\|_2^2}{2\Delta} + \mathbf{x}^\top \left( \frac{m\mathbf{S}}{\Delta} + \sqrt{\frac{m}{\Delta}} \mathbf{Z}' \right)} \right) \right] + \frac{v^2}{4\Delta} + \frac{1}{4\Delta n} (2\mathbb{E}[S^4] - v^2) \\ &= -\mathbb{E}_{S, Z'} \left[ \ln \left( \int dx P_0(x) e^{-\frac{mx^2}{2\Delta} + x \left( \frac{mS}{\Delta} + \sqrt{\frac{m}{\Delta}} Z' \right)} \right) \right] + \frac{v^2}{4\Delta} + \frac{1}{4\Delta n} (2\mathbb{E}[S^4] - v^2) \\ &= i_{\text{RS}}(E; \Delta) - \frac{m^2}{4\Delta} + \frac{1}{4\Delta n} (2\mathbb{E}[S^4] - v^2), \end{aligned} \quad (133)$$



From the fundamental theorem of calculus

$$i(1) - i(0) = -\frac{1}{n} \int_0^1 dt \frac{d}{dt} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\ln \mathcal{Z}(t)] \quad (134)$$

so we get

$$\frac{1}{n} I(\mathbf{S}; \mathbf{W}) = i_{\text{RS}}(E; \Delta) - \frac{m^2}{4\Delta} + \frac{1}{4\Delta n} (2\mathbb{E}[S^4] - v^2) - \frac{1}{n} \int_0^1 dt \frac{d}{dt} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\ln \mathcal{Z}(t)]. \quad (135)$$

We proceed to the computation of the derivative under the integral over  $t$ . Denoting by  $\langle - \rangle_t$  the expectation with respect to the posterior (131), we have

$$\frac{d}{dt} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\ln (\mathcal{Z}(t))] = \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} \left[ - \left\langle \frac{d\mathcal{H}(t)}{dt} \right\rangle_t \right]. \quad (136)$$

Hence, a simple differentiation of the Hamiltonian w.r.t.  $t$  yields

$$\begin{aligned} \frac{d}{dt} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\ln (\mathcal{Z}(t))] &= \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} \left[ \sum_{i \leq j=1}^n \left( -\frac{\langle X_i^2 X_j^2 \rangle_t}{2\Delta n} + \frac{\langle X_i X_j \rangle_t S_i S_j}{\Delta n} + \frac{Z_{ij} \langle X_i X_j \rangle_t}{2\sqrt{n\Delta t}} \right) \right. \\ &\quad \left. + \sum_{i=1}^n \left( m \frac{\langle X_i^2 \rangle_t}{2\Delta} - m \frac{\langle X_i \rangle_t S_i}{\Delta} - \frac{Z'_i \langle X_i \rangle_t}{2} \sqrt{\frac{m}{\Delta(1-t)}} \right) \right]. \end{aligned} \quad (137)$$

We now simplify this expression using integration by parts with respect to the Gaussian noises and the Nishimori identity (164) in Appendix D. Integration by parts with respect to  $Z_{ij}$  and  $Z'_i$  yields

$$\mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [Z_{ij} \langle X_i X_j \rangle_t] = \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\partial_{Z_{ij}} \langle X_i X_j \rangle_t] = \sqrt{\frac{t}{n\Delta}} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} \left[ \langle X_i^2 X_j^2 \rangle_t - \langle X_i X_j \rangle_t^2 \right] \quad (138)$$

and

$$\mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [Z'_i \langle X_i \rangle_t] = \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\partial_{Z'_i} \langle X_i \rangle_t] = \sqrt{\frac{m(1-t)}{\Delta}} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} \left[ \langle X_i^2 \rangle_t - \langle X_i \rangle_t^2 \right]. \quad (139)$$

An application of the Nishimori identity yields

$$\mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle X_i X_j \rangle_t S_i S_j] = \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle X_i X_j \rangle_t^2] \quad (140)$$

and

$$\mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle X_i \rangle_t S_i] = \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle X_i \rangle_t^2] \quad (141)$$

Combining (137) - (141) we get

$$\begin{aligned} \frac{1}{n} \frac{d}{dt} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\ln (\mathcal{Z}(t))] &= \frac{1}{2\Delta n^2} \sum_{i \leq j=1}^n \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle X_i X_j S_i S_j \rangle_t] - \frac{m}{2\Delta n} \sum_{i=1}^n \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle X_i S_i \rangle_t] \\ &= \frac{1}{4\Delta} \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle q(\mathbf{S}, \mathbf{X}) \rangle_t - 2\langle q(\mathbf{S}, \mathbf{X}) \rangle_t m] + \frac{1}{4\Delta n^2} \sum_{i=1}^n \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle X_i^2 \rangle_t S_i^2] \end{aligned}$$

where we have introduced the ‘‘overlap’’  $q(\mathbf{S}, \mathbf{X}) := n^{-1} \sum_{i=1}^n S_i X_i$ . Replacing this result in (135) we obtain the remarkable sum rule (recall  $m := v - E$ )

$$\begin{aligned} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) &= i_{\text{RS}}(E; \Delta) - \frac{1}{4\Delta} \int_0^1 dt \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle (q(\mathbf{S}, \mathbf{X}) - m)^2 \rangle_t] \\ &\quad - \frac{1}{4\Delta n^2} \sum_{i=1}^n \mathbb{E}_{\mathbf{S}, \mathbf{Z}, \mathbf{Z}'} [\langle X_i^2 \rangle_t S_i^2] + \frac{1}{4\Delta n} (2\mathbb{E}[S^4] - v^2). \end{aligned} \quad (142)$$

Thus for any  $E \in [0, v]$  we have

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} I(\mathbf{S}; \mathbf{W}) \leq i_{\text{RS}}(E; \Delta) \quad (143)$$

and (16) follows by optimizing the right hand side over  $E$ .

## Appendix B. Relating the mutual information to the free energy

The mutual information between  $\mathbf{S}$  and  $\mathbf{W}$  is defined as  $I(\mathbf{S}; \mathbf{W}) = H(\mathbf{S}) - H(\mathbf{S}|\mathbf{W})$  with

$$H(\mathbf{S}|\mathbf{W}) = -\mathbb{E}_{\mathbf{S}, \mathbf{W}} [\ln P(\mathbf{S}|\mathbf{W})] = -\mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\langle \ln P(\mathbf{X}|\mathbf{S}, \mathbf{Z}) \rangle]. \quad (144)$$

By substituting the posterior distribution in (33), one obtains

$$H(\mathbf{S}|\mathbf{W}) = \mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\ln \mathcal{Z}] + \mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\langle \mathcal{H}(\mathbf{X}|\mathbf{S}, \mathbf{Z}) \rangle] + H(\mathbf{S}). \quad (145)$$

Furthermore, using the Gaussian integration by part as in (138) and the Nishimori identity (140) yield

$$\mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\langle \mathcal{H}(\mathbf{X}|\mathbf{S}, \mathbf{Z}) \rangle] = -\frac{1}{2\Delta n} \mathbb{E}_{\mathbf{S}} \left[ \sum_{i \leq j=1}^n (S_i^2 S_j^2) \right] = -\frac{1}{4\Delta} (v^2(n-1) + 2\mathbb{E}[S^4]). \quad (146)$$

Hence, the normalized mutual information is given by (38), which we repeat here for better referencing

$$\frac{1}{n} I(\mathbf{S}; \mathbf{W}) = -\frac{1}{n} \mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\ln \mathcal{Z}] + \frac{v^2}{4\Delta} + \frac{1}{4\Delta n} (2\mathbb{E}[S^4] - v^2). \quad (147)$$

Alternatively, one can define the mutual information as  $I(\mathbf{S}; \mathbf{W}) = H(\mathbf{W}) - H(\mathbf{W}|\mathbf{S})$ . For the AWGN, it is easy to show that

$$H(\mathbf{W}|\mathbf{S}) = \frac{n(n+1)}{4} \ln(2\pi\Delta e). \quad (148)$$

Furthermore,  $H(\mathbf{W}) = -\mathbb{E}_{\mathbf{W}} [\ln P(\mathbf{W})]$  with

$$\begin{aligned} P(\mathbf{w}) &= \int \left\{ \prod_{i=1}^n dx_i P_0(x_i) \right\} P(\mathbf{w}|\mathbf{x}) = \frac{1}{(2\pi\Delta)^{\frac{n(n+1)}{4}}} \int \left\{ \prod_{i=1}^n dx_i P_0(x_i) \right\} e^{-\frac{1}{2\Delta} \sum_{i \leq j} \left( \frac{x_i x_j}{\sqrt{n}} - w_{ij} \right)^2} \\ &= \frac{\tilde{\mathcal{Z}}}{(2\pi\Delta)^{\frac{n(n+1)}{4}}}, \end{aligned} \quad (149)$$

where  $\tilde{\mathcal{Z}}$  is the partition function with complete square (20). Hence,  $H(\mathbf{W})$  reads

$$\begin{aligned} H(\mathbf{W}) &= \frac{n(n+1)}{4} \ln(2\pi\Delta) - \mathbb{E}_{\mathbf{W}}[\ln \tilde{\mathcal{Z}}] \\ &= \frac{n(n+1)}{4} \ln(2\pi\Delta) - \mathbb{E}_{\mathbf{S}, \mathbf{Z}}[\ln \mathcal{Z}] + \frac{1}{2\Delta} \mathbb{E}_{\mathbf{S}, \mathbf{Z}}\left[\sum_{i \leq j} \left(\frac{S_i S_j}{\sqrt{n}} + \sqrt{\Delta} Z_{ij}\right)^2\right], \end{aligned} \quad (150)$$

with  $\mathcal{Z}$  the simplified partition function obtained after expanding the square (33). A straightforward calculation yields

$$\begin{aligned} \frac{1}{2\Delta} \mathbb{E}_{\mathbf{S}, \mathbf{Z}}\left[\sum_{i \leq j} \left(\frac{S_i S_j}{\sqrt{n}} + \sqrt{\Delta} Z_{ij}\right)^2\right] &= \frac{n(n+1)}{4} + \frac{1}{2\Delta n} \mathbb{E}_{\mathbf{S}}\left[\sum_{i \leq j} (S_i^2 S_j^2)\right] \\ &= \frac{n(n+1)}{4} \ln(e) + \frac{1}{4\Delta} (v^2(n-1) + 2\mathbb{E}[S^4]). \end{aligned} \quad (151)$$

Finally, combining (148), (150) and (151) yields the same identity (147).

### Appendix C. Proof of the I-MMSE relation

For completeness, we give a detailed proof for the I-MMSE relation of Lemma 33 following the lines of [Guo et al. (2005)]. In the calculations below differentiations, expectations and integrations commute (see Lemma 8 in [Guo et al. (2005)]). All the matrices are symmetric and  $Z_{ij} \sim \mathcal{N}(0, 1)$  for  $i \leq j$ .

Instead of (1) it is convenient to work with the equivalent model  $w_{ij} = \frac{s_i s_j}{\sqrt{n\Delta}} + z_{ij}$  and set  $s_i s_j = u_{ij}$ . In fact, all subsequent calculations do not depend on the rank of the matrix  $\mathbf{u}$  and are valid for any finite rank matrix estimation problem as long as the noise is Gaussian. The mutual information is  $I(\mathbf{S}; \mathbf{W}) = H(\mathbf{W}) - H(\mathbf{W}|\mathbf{S})$  and  $H(\mathbf{W}|\mathbf{S}) = \frac{n(n+1)}{2} \ln(\sqrt{2\pi e})$ . Thus

$$\frac{1}{n} \frac{dI(\mathbf{S}; \mathbf{W})}{d\Delta^{-1}} = \frac{1}{n} \frac{dH(\mathbf{W})}{d\Delta^{-1}}. \quad (152)$$

We have  $H(\mathbf{W}) = -\mathbb{E}_{\mathbf{W}}[\ln P(\mathbf{W})]$  where

$$P(\mathbf{w}) = \mathbb{E}_{\mathbf{U}}[P(\mathbf{w}|\mathbf{U})] = \mathbb{E}_{\mathbf{U}}\left[(2\pi)^{-\frac{n(n+1)}{4}} e^{-\frac{1}{2} \sum_{i \leq j} \left(\frac{U_{ij}}{\sqrt{n\Delta}} - w_{ij}\right)^2}\right]. \quad (153)$$

Differentiating w.r.t  $\Delta^{-1}$

$$\frac{dH(\mathbf{W})}{d\Delta^{-1}} = -\mathbb{E}_{\mathbf{U}}\left[\int d\mathbf{w} (1 + \ln P(\mathbf{w})) \frac{dP(\mathbf{w}|\mathbf{U})}{d\Delta^{-1}}\right] \quad (154)$$

and

$$\frac{dP(\mathbf{w}|\mathbf{U})}{d\Delta^{-1}} = \sqrt{\frac{\Delta}{4n}} \sum_{k \leq l} U_{kl} \left(w_{kl} - \frac{U_{kl}}{\sqrt{\Delta n}}\right) e^{-\frac{1}{2} \sum_{i \leq j} \left(\frac{U_{ij}}{\sqrt{n\Delta}} - w_{ij}\right)^2} (2\pi)^{-\frac{n(n+1)}{2}} \quad (155)$$

$$= -\sqrt{\frac{\Delta}{4n}} \sum_{k \leq l} U_{kl} \frac{d}{dw_{kl}} e^{-\frac{1}{2} \sum_{i \leq j} \left(\frac{U_{ij}}{\sqrt{n\Delta}} - w_{ij}\right)^2} (2\pi)^{-\frac{n(n+1)}{2}}. \quad (156)$$

Replacing this last expression in (154), using an integration by part w.r.t  $w_{kl}$  (the boundary terms can be shown to vanish), then Bayes formula, and finally (153), one obtains

$$\begin{aligned}
\sqrt{\frac{4n}{\Delta}} \frac{dH(\mathbf{W})}{d\Delta^{-1}} &= \sum_{k \leq l} \mathbb{E}_{\mathbf{U}} \left[ U_{kl} \int d\mathbf{w} (1 + \ln P(\mathbf{w})) \frac{d}{dw_{kl}} e^{-\frac{1}{2} \sum_{i \leq j} \left( \frac{U_{ij}}{\sqrt{n\Delta}} - w_{ij} \right)^2} (2\pi)^{-\frac{n(n+1)}{2}} \right] \\
&= - \sum_{k \leq l} \int d\mathbf{w} \mathbb{E}_{\mathbf{U}} \left[ U_{kl} \frac{P(\mathbf{w}|\mathbf{U})}{P(\mathbf{w})} \right] \frac{dP(\mathbf{w})}{dw_{kl}} \\
&= - \sum_{k \leq l} \int d\mathbf{w} \mathbb{E}_{\mathbf{U}|\mathbf{w}} [U_{kl}] \mathbb{E}_{\mathbf{U}} \left[ \frac{dP(\mathbf{w}|\mathbf{U})}{dw_{kl}} \right] \\
&= \sum_{k \leq l} \int d\mathbf{w} \mathbb{E}_{\mathbf{U}|\mathbf{w}} [U_{kl}] \mathbb{E}_{\mathbf{U}} \left[ \left( w_{kl} - \frac{U_{kl}}{\sqrt{n\Delta}} \right) P(\mathbf{w}|\mathbf{U}) \right] \\
&= \mathbb{E}_{\mathbf{W}} \left[ \sum_{k \leq l} \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}] \left( W_{kl} - \frac{1}{\sqrt{n\Delta}} \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}] \right) \right]. \tag{157}
\end{aligned}$$

Now we replace  $\mathbf{w} = \frac{\mathbf{u}^0}{\sqrt{n\Delta}} + \mathbf{z}$ , where  $\mathbf{u}^0$  is an independent copy of  $\mathbf{u}$ . We denote  $\mathbb{E}_{\mathbf{W}}[\cdot] = \mathbb{E}_{\mathbf{U}^0, \mathbf{z}}[\cdot]$  the joint expectation. The last result then reads

$$\frac{dH(\mathbf{W})}{d\Delta^{-1}} = \frac{1}{2n} \mathbb{E}_{\mathbf{W}} \left[ \sum_{k \leq l} \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}] \left( U_{kl}^0 - \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}] + Z_{kl} \sqrt{n\Delta} \right) \right]. \tag{158}$$

Now note the two Nishimori identities (see Appendix D)

$$\mathbb{E}_{\mathbf{W}} \left[ \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}] U_{kl}^0 \right] = \mathbb{E}_{\mathbf{W}} \left[ \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}]^2 \right], \tag{159}$$

$$\mathbb{E}_{\mathbf{W}} \left[ (U_{kl}^0)^2 \right] = \mathbb{E}_{\mathbf{W}} \left[ \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}^2] \right], \tag{160}$$

and the following one obtained by a Gaussian integration by parts

$$\sqrt{n\Delta} \mathbb{E}_{\mathbf{W}} \left[ \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}] Z_{kl} \right] = \mathbb{E}_{\mathbf{W}} \left[ \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}^2] - \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}]^2 \right]. \tag{161}$$

Using the last three identities, equation (158) becomes

$$\begin{aligned}
\frac{1}{n} \frac{dH(\mathbf{W})}{d\Delta^{-1}} &= \frac{1}{2n^2} \mathbb{E}_{\mathbf{W}} \left[ \sum_{k \leq l} \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}^2] - \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}]^2 \right] \\
&= \frac{1}{2n^2} \mathbb{E}_{\mathbf{W}} \left[ \sum_{k \leq l} \left( U_{kl}^0 - \mathbb{E}_{\mathbf{U}|\mathbf{W}} [U_{kl}] \right)^2 \right], \tag{162}
\end{aligned}$$

which, in view of (152), ends the proof.

## Appendix D. Nishimori identity

Take a random vector  $\mathbf{S}$  distributed according to some known prior  $P_0^{\otimes n}$  and an observation  $\mathbf{W}$  is drawn from some known conditional distribution  $P_{\mathbf{W}|\mathbf{S}}(\mathbf{w}|\mathbf{s})$ . Take  $\mathbf{X}$  drawn from a

posterior distribution (for example this may be (20))

$$P(\mathbf{x}|\mathbf{w}) = \frac{P_0^{\otimes n}(\mathbf{x})P_{\mathbf{W}|\mathbf{S}}(\mathbf{w}|\mathbf{x})}{P(\mathbf{w})}.$$

Then for any (integrable) function  $g(\mathbf{s}, \mathbf{x})$  the Bayes formula implies

$$\mathbb{E}_{\mathbf{S}}\mathbb{E}_{\mathbf{W}|\mathbf{S}}\mathbb{E}_{\mathbf{X}|\mathbf{W}}[g(\mathbf{S}, \mathbf{X})] = \mathbb{E}_{\mathbf{W}}\mathbb{E}_{\mathbf{X}'|\mathbf{W}}\mathbb{E}_{\mathbf{X}|\mathbf{W}}[g(\mathbf{X}', \mathbf{X})] \quad (163)$$

where  $\mathbf{X}, \mathbf{X}'$  are independent random vectors distributed according to the posterior distribution. Therefore

$$\mathbb{E}_{\mathbf{S}, \mathbf{W}}\mathbb{E}_{\mathbf{X}|\mathbf{W}}[g(\mathbf{S}, \mathbf{X})] = \mathbb{E}_{\mathbf{W}}\mathbb{E}_{\mathbf{X}'|\mathbf{W}}\mathbb{E}_{\mathbf{X}|\mathbf{W}}[g(\mathbf{X}', \mathbf{X})]. \quad (164)$$

In the statistical mechanics literature this identity is sometimes called the Nishimori identity and we adopt this language here. For model (1) for example we can express  $\mathbf{W}$  in the posterior in terms of  $\mathbf{S}$  and  $\mathbf{Z}$  which are independent and  $\mathbb{E}_{\mathbf{X}|\mathbf{W}}[-] = \langle - \rangle$ . Then the Nishimori identity reads

$$\mathbb{E}_{\mathbf{S}, \mathbf{Z}}[\langle g(\mathbf{S}, \mathbf{X}) \rangle] = \mathbb{E}_{\mathbf{S}, \mathbf{Z}}[\langle g(\mathbf{X}', \mathbf{X}) \rangle]. \quad (165)$$

An important case for  $g$  depending only on the first argument is  $\mathbb{E}_{\mathbf{S}}[g(\mathbf{S})] = \mathbb{E}_{\mathbf{S}, \mathbf{Z}}[\langle g(\mathbf{X}) \rangle]$ .

Special cases that are often used in this paper are

$$\begin{cases} \mathbb{E}_{\mathbf{S}, \mathbf{Z}}[S_i \langle X_i \rangle] = \mathbb{E}_{\mathbf{S}, \mathbf{Z}}[\langle X_i^2 \rangle] \\ \mathbb{E}_{\mathbf{S}, \mathbf{Z}}[S_i S_j \langle X_i X_j \rangle] = \mathbb{E}_{\mathbf{S}, \mathbf{Z}}[\langle X_i X_j^2 \rangle] \\ \mathbb{E}[S^2] = \mathbb{E}_{\mathbf{S}, \mathbf{Z}}[\langle X_i^2 \rangle]. \end{cases} \quad (166)$$

A mild generalization of (165) which is also used is

$$\mathbb{E}_{\mathbf{S}, \mathbf{Z}}[S_i S_j \langle X_i \rangle \langle X_j \rangle] = \mathbb{E}_{\mathbf{S}, \mathbf{Z}}[\langle X_i X_j \rangle \langle X_i \rangle \langle X_j \rangle]. \quad (167)$$

We remark that these identities are used with brackets  $\langle - \rangle$  corresponding to various “interpolating” posteriors.

## Appendix E. Proof of Lemmas 13 and 21

We show the details for Lemma 13. The proof of Lemma 21 follows the same lines. A straightforward differentiation of  $f_{\text{RS}}^{\text{u}}$  w.r.t.  $E$  gives

$$\frac{df_{\text{RS}}^{\text{u}}(E; \Delta)}{dE} = \frac{E - v}{2\Delta} + \frac{1}{2\Delta} \mathbb{E}_{Z, S} \left[ \left\langle -X^2 + 2XS + ZX \sqrt{\frac{\Delta}{v - E}} \right\rangle \right]. \quad (168)$$

Recall that here the posterior expectation  $\langle \cdot \rangle$  is defined by (40). A direct application of the Nishimori condition gives

$$v := \mathbb{E}_S[S^2] = \mathbb{E}_{S, Z}[\langle X^2 \rangle], \quad (169)$$

$$\mathbb{E}_{S, Z}[S \langle X \rangle] = \mathbb{E}_{S, Z}[\langle X \rangle^2], \quad (170)$$

which implies

$$\mathbb{E}_{S,Z}[(S - \langle X \rangle_E)^2] = \mathbb{E}_{S,Z}[\langle X^2 \rangle] - \mathbb{E}_{S,Z}[\langle X \rangle^2]. \quad (171)$$

Thus from (168) we see that stationary points of  $f_{\text{RS}}^{\text{u}}$  satisfy

$$E = 2v - 2\mathbb{E}_{S,Z}[\langle X \rangle^2] - \sqrt{\frac{\Delta}{v-E}} \mathbb{E}_{S,Z}[Z\langle X \rangle]. \quad (172)$$

Now using an integration by part w.r.t  $Z$ , one gets

$$\sqrt{\frac{\Delta}{v-E}} \mathbb{E}_{S,Z}[Z\langle X \rangle] = v - \mathbb{E}_{S,Z}[\langle X \rangle^2], \quad (173)$$

which allows to rewrite (172) as

$$E = v - \mathbb{E}_{S,Z}[\langle X \rangle^2] = \mathbb{E}_{S,Z}[(S - \langle X \rangle)^2] \quad (174)$$

where the second equality follows from (169) and (171). Recalling the expression (41) of the state evolution operator we recognize the equation  $E = T_{\text{u}}(E)$ .

### Appendix F. Analysis of $i_{\text{RS}}(E; \Delta)$ for $\Delta \rightarrow 0$

In this appendix, we prove that  $\lim_{\Delta \rightarrow 0} \min_E i_{\text{RS}}(E; \Delta) = H(S)$ . First, a simple calculation leads to the following relation between  $i_{\text{RS}}$  and the mutual information of the scalar denoising problem for  $E \in [0, v]$

$$i_{\text{RS}}(E; \Delta) = I(S; S + \Sigma(E)Z) + \frac{E^2}{4\Delta}, \quad (175)$$

where  $Z \sim \mathcal{N}(0, 1)$  and  $\Sigma(E)^2 := \Delta/(v-E)$ . Note that as  $\Delta \rightarrow 0$ ,  $\Sigma(E) \rightarrow 0$  (for  $E \neq v$ ). Therefore,  $\lim_{\Delta \rightarrow 0} I(S; S + \Sigma(E)Z) = H(S)$ . Now let  $E_0$  be the global minimum of  $i_{\text{RS}}(E; \Delta)$ . By evaluating both sides of (175) at  $E_0$  and taking the limit  $\Delta \rightarrow 0$ , it remains to show that  $E_0^2/(4\Delta) \rightarrow 0$  as  $\Delta \rightarrow 0$  (i.e.  $E_0^2 \rightarrow 0$  faster than  $\Delta$ ). Since  $E_0$  is the global minimum of the RS potential, then  $E_0 = T_{\text{u}}(E_0) = \text{mmse}(\Sigma(E_0)^{-2})$  by Lemma 13. Moreover, one can show, under our assumptions on  $P_0$ , that the scalar MMSE function scales as

$$\text{mmse}(\Sigma^{-2}) = \mathcal{O}(e^{-c\Sigma^{-2}}), \quad (176)$$

with  $c$  a non-negative constant that depends on  $P_0$  [Barbier et al. (2017b)]. Hence,  $E_0^2/(4\Delta) \rightarrow 0$  as  $\Delta \rightarrow 0$ , which ends the proof.

### Appendix G. Proof of Proposition 30

Call  $\mathcal{H}^{\text{per}}$  and  $\langle - \rangle_{\text{per}}$  the Hamiltonian and posterior average associated to the periodic SC system with mutual information  $i_{n,w,L}^{\text{per}}$ . Similarly call  $\mathcal{H}^{\text{cou}}$  and  $\langle - \rangle_{\text{cou}}$  the Hamiltonian and

posterior average associated to the pinned SC system with mutual information  $i_{n,w,L}^{\text{cou}}$ . The Hamiltonians satisfy the identity  $\mathcal{H}^{\text{cou}} - \mathcal{H}^{\text{per}} = \delta\mathcal{H}$  with

$$\begin{aligned} \delta\mathcal{H} &= \sum_{\mu \in \mathcal{B}} \Lambda_{\mu,\mu} \sum_{i_\mu \leq j_\mu} \left[ \frac{x_{i_\mu}^2 x_{j_\mu}^2 + s_{i_\mu}^2 s_{j_\mu}^2}{2n\Delta} - \frac{s_{i_\mu} s_{j_\mu} x_{i_\mu} x_{j_\mu}}{n\Delta} - \frac{(x_{i_\mu} x_{j_\mu} - s_{i_\mu} s_{j_\mu}) z_{i_\mu j_\mu}}{\sqrt{n\Delta\Lambda_{\mu,\mu}}} \right] \\ &+ \sum_{\mu \in \mathcal{B}} \sum_{\nu \in \{\mu+1:\mu+w\} \cap \mathcal{B}} \Lambda_{\mu,\nu} \sum_{i_\mu \leq j_\nu} \left[ \frac{x_{i_\mu}^2 x_{j_\nu}^2 + s_{i_\mu}^2 s_{j_\nu}^2}{2n\Delta} - \frac{s_{i_\mu} s_{j_\nu} x_{i_\mu} x_{j_\nu}}{n\Delta} - \frac{(x_{i_\mu} x_{j_\nu} - s_{i_\mu} s_{j_\nu}) z_{i_\mu j_\nu}}{\sqrt{n\Delta\Lambda_{\mu,\nu}}} \right] \\ &+ \sum_{\mu \in \mathcal{B}} \sum_{\nu \in \{\mu-w:\mu-1\} \cap \mathcal{B}} \Lambda_{\mu,\nu} \sum_{i_\mu > j_\nu} \left[ \frac{x_{i_\mu}^2 x_{j_\nu}^2 + s_{i_\mu}^2 s_{j_\nu}^2}{2n\Delta} - \frac{s_{i_\mu} s_{j_\nu} x_{i_\mu} x_{j_\nu}}{n\Delta} - \frac{(x_{i_\mu} x_{j_\nu} - s_{i_\mu} s_{j_\nu}) z_{i_\mu j_\nu}}{\sqrt{n\Delta\Lambda_{\mu,\nu}}} \right]. \end{aligned}$$

It is easy to see that

$$i_{n,w,L}^{\text{per}} - i_{n,w,L}^{\text{cou}} = \frac{1}{n(L+1)} \mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\ln \langle e^{-\delta\mathcal{H}} \rangle_{\text{cou}}], \quad (177)$$

$$i_{n,w,L}^{\text{cou}} - i_{n,w,L}^{\text{per}} = \frac{1}{n(L+1)} \mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\ln \langle e^{\delta\mathcal{H}} \rangle_{\text{per}}]. \quad (178)$$

and using the convexity of the exponential, we get

$$i_{n,w,L}^{\text{cou}} + \frac{\mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\langle \delta\mathcal{H} \rangle_{\text{per}}]}{n(L+1)} \leq i_{n,w,L}^{\text{per}} \leq i_{n,w,L}^{\text{cou}} + \frac{\mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\langle \delta\mathcal{H} \rangle_{\text{cou}}]}{n(L+1)}. \quad (179)$$

Due to the pinning condition we have  $\mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\langle \delta\mathcal{H}(\mathbf{X}) \rangle_{\text{cou}}] = 0$ , and thus we get the upper bound  $i_{n,w,L}^{\text{per}} \leq i_{n,w,L}^{\text{cou}}$ . Let us now look at the lower bound. We note that by the Nishimori identity in Appendix D, as long as  $P_0$  has finite first four moments, we can find constants  $K_1, K_2$  independent of  $n, w, L$  such that  $\mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\langle X_{i_\mu}^2 X_{j_\nu}^2 \rangle_{\text{per}}] \leq K_1$  and  $\mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\langle X_{i_\mu}^4 \rangle_{\text{per}}] \leq K_2$ . First we use Gaussian integration by parts to eliminate  $z_{i_\mu j_\nu}$  from the brackets, the Cauchy-Schwartz inequality, and the Nishimori identity of Appendix D, to get an upper bound where only fourth order moments of signal are involved. Thus as long as  $P_0$  has finite first four moments we find

$$\frac{|\mathbb{E}_{\mathbf{S}, \mathbf{Z}} [\langle \delta\mathcal{H}(\mathbf{X}) \rangle_{\text{c}}]}{n(L+1)} \leq C \frac{\Lambda^* (2w+1)^2}{L+1} = \mathcal{O}\left(\frac{w}{L}\right). \quad (180)$$

for some constant  $C > 0$  independent of  $n, w, L$  and we recall  $\Lambda_* := \sup \Lambda_{\mu\nu} = \mathcal{O}(w^{-1})$ . Thus we get the lower bound  $i_{n,w,L}^{\text{cou}} - \mathcal{O}\left(\frac{w}{L}\right) \leq i_{n,w,L}^{\text{per}}$ . This completes the proof of Proposition 30.

## Acknowledgments

The work of Jean Barbier and Mohamad Dia was supported by the Swiss National Foundation for Science grant number 200021-156672. We thank Thibault Lesieur for help with the phase diagrams.

## References

- Arash A Amini and Martin J Wainwright. High-dimensional analysis of semidefinite relaxations for sparse principal components. In *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pages 2454–2458. IEEE, 2008.
- Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697, 2005.
- Boaz Barak, Samuel B Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *arXiv preprint arXiv:1604.03084*, 2016.
- Jean Barbier and Nicolas Macris. The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference. *Probability Theory and Related Fields*, Oct 2018. ISSN 1432-2064. doi: 10.1007/s00440-018-0879-0. URL <https://doi.org/10.1007/s00440-018-0879-0>.
- Jean Barbier, Mohamad Dia, Nicolas Macris, and Florent Krzakala. The mutual information in random linear estimation. In *2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 625–632, 2016a.
- Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, and Lenka Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Advances in Neural Information Processing Systems 29*, pages 424–432. 2016b.
- Jean Barbier, Mohamad Dia, and Nicolas Macris. Universal sparse superposition codes with spatial coupling and gamp decoding. *arXiv preprint arXiv:1707.04203*, 2017a.
- Jean Barbier, Nicolas Macris, Mohamad Dia, and Florent Krzakala. Mutual information and optimality of approximate message-passing in random linear estimation. *arXiv preprint arXiv:1701.05823*, 2017b.
- Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. on Information Theory*, 57(2):764–785, 2011.
- Peter J Bickel and Aiyou Chen. A nonparametric view of network models and newman-girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073, 2009.
- Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.
- Francesco Caltagirone, Silvio Franz, Richard G. Morris, and Lenka Zdeborová. Dynamics and termination cost of spatially coupled mean-field models. *Phys. Rev. E*, 89:012102, Jan 2014.



- Emmanuel J Candès and Benjamin Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.
- Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *arXiv preprint arXiv:1402.1267*, 2014.
- Alexandre d’Aspremont, Laurent El Ghaoui, Michael I Jordan, and Gert RG Lanckriet. A direct formulation for sparse pca using semidefinite programming. *SIAM review*, 49(3):434–448, 2007.
- Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Physical Review E*, 84(6):066106, 2011.
- Yash Deshpande and Andrea Montanari. Information-theoretically optimal sparse pca. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 2197–2201, June 2014. doi: 10.1109/ISIT.2014.6875223.
- Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the two-groups stochastic block model. *arXiv:1507.08685*, 2015.
- Mohamad Dia. *High-Dimensional Inference on Dense Graphs with Applications to Coding Theory and Machine Learning*. PhD thesis, EPFL IC School, Lausanne, 2018.
- Silvio Franz and Fabio Lucio Toninelli. Finite-range spin glasses in the kac limit: free energy and local observables. *Journal of Physics A: Mathematical and General*, 37(30):7433, 2004.
- Andrei Giurgiu, Nicolas Macris, and Rüdiger Urbanke. Spatial coupling as a proof technique and three applications. *IEEE Transactions on Information Theory*, 62(10):5281–5295, Oct 2016.
- Francesco Guerra. An introduction to mean field spin glass theory: methods and results. *Mathematical Statistical Physics*, pages 243–271, 2005.
- Francesco Guerra and Fabio Lucio Toninelli. The thermodynamic limit in mean field spin glass models. *Commun. Math. Phys.*, 230(1):71–79, 2002.
- Dongning Guo, Shlomo Shamai, and Sergio Verdú. Mutual information and minimum mean-square error in gaussian channels. *IEEE Trans. on Information Theory*, 51, 2005.
- Bruce Hajek, Yihong Wu, and Jiaming Xu. Submatrix localization via message passing. *arXiv preprint arXiv:1510.09219*, 2015.
- S Hamed Hassani, Nicolas Macris, and Rüdiger Urbanke. Coupled graphical models and their thresholds. In *IEEE Information Theory Workshop (ITW)*, 2010.
- Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Journal of Information and Inference*, 2(2):115–144, 2013.

- Iain M Johnstone and Arthur Yu Lu. Sparse principal components analysis. *Unpublished manuscript*, 7, 2004.
- Iain M Johnstone and Arthur Yu Lu. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 2012.
- Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical Review E*, 83(1):016107, 2011.
- Raghuveer H Keshavan, Sewoong Oh, and Andrea Montanari. Matrix completion from a few entries. In *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, pages 324–328. IEEE, 2009.
- Satish Babu Korada and Nicolas Macris. Exact solution of the gauge symmetric p-spin glass model on a complete graph. *Journal of Statistical Physics*, 136(2):205–230, 2009.
- Florent Krzakala, Jiaming Xu, and Lenka Zdeborová. Mutual information in rank-one matrix estimation. *arXiv preprint arXiv:1603.08447*, 2016.
- Shrinivas Kudekar, Thomas J Richardson, and Rüdiger Urbanke. Threshold saturation via spatial coupling: Why convolutional ldpc ensembles perform so well over the bec. *IEEE Trans. on Inf. Th.*, 57, 2011.
- Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *arXiv preprint arXiv:1611.03888*, 2017.
- Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Mmse of probabilistic low-rank matrix estimation: Universality with respect to the output channel. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 680–687, Sept 2015a. doi: 10.1109/ALLERTON.2015.7447070.
- Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Phase transitions in sparse pca. In *Information Theory (ISIT), 2015 IEEE International Symposium on*, pages 1635–1639. IEEE, 2015b.
- Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, pages 694–703. ACM, 2014.
- Ryosuke Matsushita and Toshiyuki Tanaka. Low-rank matrix reconstruction and clustering via approximate message passing. In *Advances in Neural Information Processing Systems*, pages 917–925, 2013.
- Sundeeep Rangan and Alyson K Fletcher. Iterative estimation of constrained rank-one matrices in noise. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 1246–1250. IEEE, 2012.
- Federico Ricci-Tersenghi, Adel Javanmard, and Andrea Montanari. Performance of a community detection algorithm based on semidefinite programming. In *Journal of Physics: Conference Series*, volume 699, page 012015. IOP Publishing, 2016.

Alaa Saade, Florent Krzakala, and Lenka Zdeborová. Spectral clustering of graphs with the bethe hessian. In *Advances in Neural Information Processing Systems*, pages 406–414, 2014.

Alaa Saade, Florent Krzakala, and Lenka Zdeborová. Matrix completion from fewer entries: Spectral detectability and rank estimation. In *Advances in Neural Information Processing Systems*, pages 1261–1269, 2015.

Arvind Yedla, Yung-Yih Jian, Phong S Nguyen, and Henry D Pfister. A simple proof of maxwell saturation for coupled scalar recursions. *IEEE Trans. on Inf. Theory*, 60(11): 6943–6965, 2014.

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