
Optimal Algorithms for the Inhomogeneous Spiked Wigner model

Justin Ko
ENS de Lyon
justin.ko@ens-lyon.fr

Florent Krzakala
EPFL
florent.krzakala@epfl.ch

Aleksandr Pak
ENS de Lyon, EPFL
aleksandr.pak@epfl.ch

Abstract

We study a spiked Wigner problem with an inhomogeneous noise profile. Our aim in this problem is to recover the signal passed through an inhomogeneous low-rank matrix channel. While the information-theoretic performances are well-known, we focus on the algorithmic problem. First, we derive an approximate message-passing algorithm (AMP) for the inhomogeneous problem and show that its rigorous state evolution coincides with the fixed-point equations satisfied by the Bayes-optimal estimator. Second, we deduce a simple and efficient spectral method that outperforms PCA and is shown to match the information-theoretic transition.

Low-rank information extraction from a noisy data matrix is a crucial statistical challenge. The spiked random matrix models have recently gained extensive interest in the fields of statistics, probability, and machine learning, serving as a valuable platform for exploring this issue [Donoho and Johnstone \[1995\]](#), [Péché \[2014\]](#), [Baik et al. \[2005\]](#). A prominent example is the spiked Wigner model, where a rank one matrix is observed through a component-wise homogeneous noise.

Here we consider the inhomogeneous version of the model which has been recently introduced in a series of papers [Barbier and Reeves \[2020\]](#), [Behne and Reeves \[2022\]](#), [Alberici et al. \[2021a, 2022\]](#) and naturally arises in different contexts such as community detection [Behne and Reeves \[2022\]](#), deep Boltzmann machines [Alberici et al. \[2021b\]](#) etc. In particular, it corresponds to the dense limit of the degree-corrected stochastic block model [Karrer and Newman \[2011\]](#) (see the explicit mapping in [Guionnet et al. \[2022\]](#)). In this model the signal is observed through an inhomogeneous Gaussian noise with block-constant variance profile. Such block-constant noise naturally arises in settings such as community detection and stochastic-block models. The assumption for the noise to be Gaussian is in no way restrictive as a universality result in [Guionnet et al. \[2022\]](#) shows that an entrywise transformation of data generated from a general inhomogeneous inference problem reduces any probabilistic noise to the spiked Wigner problem with Gaussian data. This generality is what makes the model appealing.

We now define the model: Consider a partition $\{1, \dots, N\} = [N]$ into q disjoint groups $C_1^N \cup \dots \cup C_q^N = [N]$. This partition is encoded by a function $g : [N] \mapsto [q]$ which maps each index $i \in [N]$ into its group $g(i) \in [q]$. Let $\tilde{\Delta} \in \mathbb{R}^{q \times q}$ be a symmetric matrix encoding a block-constant symmetric matrix $\Delta \in \mathbb{R}^{N \times N}$

$$\Delta_{ij} = \tilde{\Delta}_{g(i)g(j)}. \quad (1)$$

We assume both the noise profile $\tilde{\Delta}$ and the partition function g to be known. This is often the case in practical applications, such as the degree-corrected block model, [Karrer and Newman \[2011\]](#)

Additionally, in practical applications where the noise profile is not known, one can often empirically estimate the noise profile and assign group membership according to the empirical variances. Such an approach is likely to yield good results but lies outside the scope of the present work.

We observe the signal $\mathbf{x}^* \in \mathbb{R}^N$ which is assumed to have independent identically distributed coordinates generated from some prior distribution \mathbb{P}_0 (i.e. $\mathbb{P}(\mathbf{x}^* = \mathbf{x}) = \prod_{i=1}^N \mathbb{P}_0(x_i^* = x_i)$) through noisy measurements:

$$\mathbf{Y} = \sqrt{\frac{1}{N}} \mathbf{x}^* (\mathbf{x}^*)^T + \mathbf{A} \odot \sqrt{\mathbf{\Delta}}. \quad (2)$$

Here and throughout the article \odot denotes the Hadamard product, $\sqrt{\mathbf{\Delta}}$ is the Hadamard square-root of $\mathbf{\Delta}$ and \mathbf{A} is a real-valued symmetric GOE matrix with off-diagonal elements of unit variance. The Bayes-optimal performance of this model in the asymptotic limit $N \rightarrow \infty$ was studied rigorously in Guionnet et al. [2022], Behne and Reeves [2022], Chen and Xia [2022], Chen et al. [2021] who characterized the fundamental information-theoretic limit of reconstruction in this model.

Here we focus instead on the algorithmic problem of reconstructing the (hidden) spike. **Our contributions are many-fold:**

- We show how one can construct an Approximate Message Passing (AMP) algorithm for the inhomogeneous Wigner problem, whose asymptotic performance can be tracked by a rigorous state evolution, generalizing the homogeneous version of the algorithm for low-rank factorization [Bayati and Montanari, 2011, Deshpande et al., 2015, Lesieur et al., 2017].
- We derive a fixed-point equation for AMP and show that it coincides with the fixed-point equation for the Bayes-optimal estimator obtained in Guionnet et al. [2022]
- Finally, we present a linear version of AMP [Maillard et al., 2022], that is equivalent to a spectral method. We conjecture it to be optimal in the sense that it can detect the presence of the spike in the same region as AMP. This is quite remarkable since, as shown in [Guionnet et al., 2022, Section 2.4], the standard spectral method (PCA) fails to do so.

Related work — The class of approximate message passing algorithms (AMP) has attracted a lot of attention in the high-dimensional statistics and machine learning community, see e.g. [Donoho et al., 2009, Bayati and Montanari, 2011, Rangan, 2011, Deshpande et al., 2015, Lesieur et al., 2017, Gerbelot and Berthier, 2021, Feng et al., 2022]. The ideas behind this algorithm have roots in physics of spin glasses Mézard et al. [1987], Bolthausen [2014], Zdeborová and Krzakala [2016]. AMP algorithms are optimal among first order methods [Cecantano et al., 2020], thus their reconstruction threshold provides a bound on the algorithmic complexity in our model. Our approach to the inhomogeneous version of AMP relies on several refinements of AMP methods to handle the full complexity of the problem, notably the spatial coupling technique Krzakala et al. [2012], Donoho et al. [2013], Javanmard and Montanari [2013], Gerbelot and Berthier [2021], Rossetti and Reeves [2023].

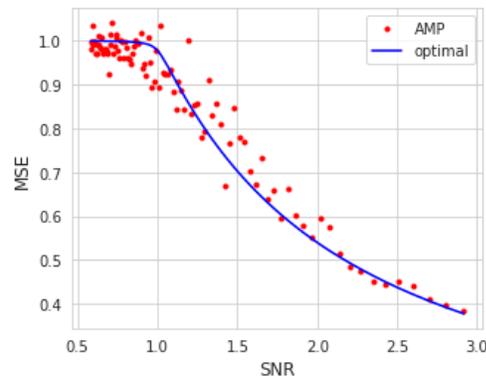


Figure 1: Performance of the inhomogeneous AMP algorithm against the information-theoretical optimal MMSE. The variance profile is proportional to $\tilde{\mathbf{\Delta}} = \begin{bmatrix} 1 & 3 \\ 3 & 2 \end{bmatrix}$ with two equally sized blocks with standard Gaussian prior when $N = 500$ at various snr.

Factorizing low-rank matrices is a ubiquitous problem with many applications in machine learning and statistics, ranging from sparse PCA to community detection and sub-matrix localization. Many variants of the homogeneous problem have been studied in the high-dimensional limit [Deshpande et al., 2015, Lesieur et al., 2017, Barbier et al., 2018, Lesieur et al., 2017, Alaoui et al., 2020, Lelarge and Miolane, 2019, Louart and Couillet, 2018, Barbier and Reeves, 2020]. The inhomogeneous version was discussed in detail in Guionnet et al. [2022], Behne and Reeves [2022], Alberici et al. [2021a]. Spectral methods are a very popular tool to solve rank-factorization problems [Donoho and Johnstone, 1995, P ech e, 2014, Baik et al., 2005]. Using AMP as an inspiration for deriving

new spectral methods was discussed, for instance, in [Saade et al. \[2014\]](#), [Lesieur et al. \[2017\]](#), [Aubin et al. \[2019\]](#), [Mondelli and Montanari \[2018\]](#), [Mondelli et al. \[2022\]](#), [Maillard et al. \[2022\]](#), [Venkataramanan et al. \[2022\]](#).

1 Main results

Message passing algorithm — For each $t \geq 0$, let $(f_t^a)_{a \in [q]}$ be a collection of Lipschitz functions from $\mathbb{R} \rightarrow \mathbb{R}$, and define for $t \in \mathbb{N}$ $f_t : \mathbb{R}^N \mapsto \mathbb{R}^N$ by

$$f_t(\mathbf{x}) := (f_t^{g(i)}(x_i))_{i \in [N]} \in \mathbb{R}^N.$$

These functions are often called denoiser functions and can be chosen amongst several options, such as the Bayes optimal denoisers for practical applications (see Section 2), or even linear denoisers (see Section 3). We shall consider the following AMP recursion

$$\mathbf{x}^{t+1} = \left(\frac{1}{\sqrt{N\Delta}} \odot \mathbf{Y} \right) f_t(\mathbf{x}^t) - \mathbf{b}_t \odot f_{t-1}(\mathbf{x}^{t-1}) \quad (3)$$

with the so-called Onsager term $\mathbf{b}_t = \frac{1}{N\Delta} f'_t(x^t) \in \mathbb{R}^N$ where $\frac{1}{\Delta}$ is the Hadamard inverse of Δ and f'_t is the vector of coordinate wise derivatives.

In practical implementations, we initialize the algorithm with some non-null \mathbf{x}^0 and let it run for a certain number of iterations. One efficient way to do this is the spectral initialization [[Mondelli and Venkataramanan, 2021](#)] with the method described in sec. 3. In Figure 1 we provide an example of the performance of the AMP together with the Bayes-optimal estimator predicted by the asymptotic theory. Even at very moderate sizes, the agreement between theory and simulation is clear.

State evolution — AMP was the basis of many works for the homogeneous case e.g. [Deshpande et al. \[2015\]](#), [Lesieur et al. \[2017\]](#). Our first contribution is the introduction of an AMP and of its rigorous state evolution [[Javanmard and Montanari, 2013](#), Theorem 1] in the inhomogeneous setting. To state a well-defined limit of the AMP, we have the following assumptions.

Assumption 1.1. *To ensure that our inhomogeneous AMP has a well-defined limit, we assume that*

1. *For each $a \in [q]$, we have*

$$\lim_{N \rightarrow \infty} \frac{|C_a^N|}{N} \rightarrow c_a \in (0, 1).$$

2. *For each $t \in [N]$ and $a \in [q]$, both (f_t^a) and $(f_t^a)'$ are Lipschitz.*

3. *For each $a \in [q]$, there exists $(\sigma_a^0)^2 \in \mathbb{R}$ such that, in probability,*

$$\lim_{N \rightarrow \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} f_0^a(x_i^0) f_0^a(x_i^0) = (\sigma_a^0)^2.$$

Our first result describes the distribution of the iterates in the limit. Our mode of convergence will be with respect to L -pseudo-Lipschitz test functions $\phi : \mathbb{R}^M \rightarrow \mathbb{R}$ satisfying

$$|\phi(x) - \phi(y)| \leq L(1 + \|x\| + \|y\|)\|x - y\| \quad \text{for all } x, y \in \mathbb{R}^M. \quad (4)$$

We define the following state evolution parameters μ_b^t and σ_b^t for $b \in [q]$ through the recursion

$$\begin{aligned} \mu_b^{t+1} &= \sum_{a \in [q]} \frac{c_a}{\Delta_{ab}} \mathbb{E}_{x_0^*, Z} [x_0^* f_t^a(\mu_a^t x_0^* + \sigma_a^t Z)] \text{ with } x_0^* \sim \mathbb{P}_0, Z \sim \mathcal{N}(0, 1) \\ (\sigma_b^{t+1})^2 &= \sum_{a=1}^q \frac{c_a}{\Delta_{ab}} \mathbb{E}_{x_0^*, Z} [(f_t^a(\mu_a^t x_0^* + \sigma_a^t Z))^2] \text{ with } x_0^* \sim \mathbb{P}_0, Z \sim \mathcal{N}(0, 1), \end{aligned} \quad (5)$$

where x_0^* and Z are independent. We prove that the iterates \mathbf{x}_i^t are asymptotically equal in distribution to $\mu_{g(i)}^t x_0^* + \sigma_{g(i)}^t Z$ where $x_0^* \sim \mathbb{P}_0$ and Z is an independent standard Gaussian.

Theorem 1.2 (State evolution of AMP iterates in the inhomogeneous setting). *Suppose that Assumption 1.1 holds, and that \mathbb{P}_0 has bounded second moment. Let $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$ be a 2-pseudo-Lipschitz test functions satisfying (4). For any $a \in [q]$, the following limit holds almost surely*

$$\lim_{N \rightarrow \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \phi(x_i^t, x_i^*) = \mathbb{E}_{x_0^*, Z} \phi(\mu_a^t x_0^* + \sigma_a^t Z, x_0^*)$$

where Z is a standard Gaussian independent from all other variables.

Remark 1.3. *The notion of convergence under the L -pseudo-Lipschitz test functions induces a topology that is equivalent to the one generated by the L -Wasserstein topology [Feng et al., 2022, Remark 7.18]. We can strengthen the second moment assumption on \mathbb{P}_0 to finite k th moment, but the induced topology will then change to the k -Wasserstein topology, see [Javanmard and Montanari, 2013, Theorem 1].*

Even though the theoretical result above applies in the high-dimensional limit, numerical simulations show that even for medium-sized N (around 500), the behaviour of the iterates is well described by the state evolution parameters. Through the state evolution equations (5) we are able to track the iterates of the AMP iteration with just two vectors of parameters obeying the state evolution recursion: the overlap with the true signal $(\mu_a^t)_{a \in [q]}$ and its variance $(\sigma_a^t)_{a \in [q]}$. We next obtain the following necessary and sufficient condition for the overlaps of a fixed point of the iteration (2):

Theorem 1.4 (Bayes-Optimal fixed point). *Assume AMP satisfies Assumption 1.1 and let the denoising functions be the Bayes ones (32). Then the overlaps $\mu = (\mu_a)_{a \in [q]}$ in (5) satisfy the following fixed point equation*

$$\mu_b = \sum_{a \in [q]} \frac{c_a}{\tilde{\Delta}_{ab}} \mathbb{E}_{x_0^*, Z} [x_0^* \mathbb{E}_{\text{posterior}} [x_0^* | \mu_a x_0^* + \sqrt{\mu_a} Z]]. \quad (6)$$

Remark 1.5. *The state evolution fixed point equation above coincides with the fixed point equation satisfied by the Bayes optimal estimator in [Guionnet et al., 2022, Equation 2.14].*

A spectral method — Given the matrix \mathbf{Y} defined in (2) we consider the transformed matrix

$$\tilde{\mathbf{Y}} := \frac{\mathbb{E}_{x_0^*} [(x_0^*)^2]}{\sqrt{N\Delta}} \odot \mathbf{Y} - \mathbb{E}_{x_0^*} [(x_0^*)^2]^2 \text{diag} \left(\frac{1}{N\Delta} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right). \quad (7)$$

Let $\mathbf{c} = (c_a)_{a \in [q]}$. We define the inhomogeneous signal-to-noise (SNR) ratio of such a model by

$$\text{SNR}(\Delta) := \lambda(\Delta) = \mathbb{E}_{x_0^*} [(x_0^*)^2]^2 \left\| \text{diag}(\sqrt{\mathbf{c}}) \frac{1}{\Delta} \text{diag}(\sqrt{\mathbf{c}}) \right\|_{op}. \quad (8)$$

Conjecture 1.6. *The top eigenvalue of $\tilde{\mathbf{Y}}$ separates from the bulk if and only if the signal to noise ratio $\lambda(\Delta) > 1$. In particular, if $\hat{\mathbf{x}}$ is the top eigenvector of $\tilde{\mathbf{Y}}$, then if and only if $\lambda(\Delta) < 1$ we have:*

$$\lim_{N \rightarrow \infty} \frac{|\hat{\mathbf{x}} \cdot \mathbf{x}^*|}{\|\hat{\mathbf{x}}\| \|\mathbf{x}^*\|} = 0.$$

The conjecture is based on heuristic arguments identifying the fixed point of AMP with a spectral method. This matches precisely the recovery transition in [Guionnet et al., 2022, Lemma 2.15 Part (b)]. In this paper, we rigorously show that with $\text{SNR}(\Delta) < 1$ our proposed spectral method fails to recover the signal. We illustrate the eigenvalue BBP-like transition in Fig.2.

2 The inhomogeneous AMP algorithm

In this section, we derive the formula for the inhomogeneous AMP iteration (25). We first recall the general matrix framework of AMP from [Javanmard and Montanari, 2013]:

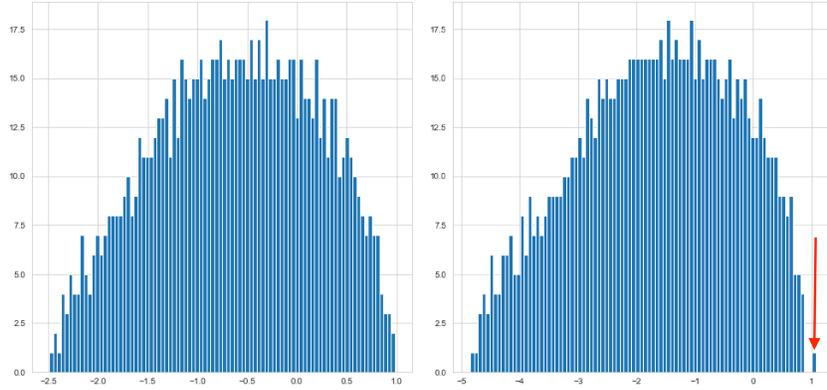


Figure 2: Illustration of the spectrum of $\tilde{\mathbf{Y}} \in \mathbb{R}^{10^3 \times 10^3}$ evaluated at noise profiles with snr $\lambda(\Delta) = 0.7$ (left, before the transition) and 1.8 (right, after the transition), with the outlying eigenvector correlated with the spike arises at eigenvalue one.

Matrix AMP — In the matrix setting an AMP algorithm operates on the vector space $\mathcal{V}_{q,N} \equiv (\mathbb{R}^q)^N \simeq \mathbb{R}^{N \times q}$. Each element of $\mathbf{v} = (v_1, \dots, v_N) \in \mathcal{V}_{q,N}$ will be regarded as N -vector with entries $v_i \in \mathbb{R}^q$.

Definition 2.1 (AMP). A matrix AMP acting on this space is represented by $(\mathbf{A}, \mathcal{F}, \mathbf{v}^0)$, where:

1. $\mathbf{A} = \mathbf{G} + \mathbf{G}^\top$, where $\mathbf{G} \in \mathbb{R}^{N \times N}$ has iid entries $G_{ij} \sim N(0, \frac{1}{2})$.
2. \mathcal{F} is a family of N Lipschitz functions $f_t^i : \mathbb{R}^q \mapsto \mathbb{R}^q$, $i \in [N]$ indexed by time $t \in \mathbb{N}$. The family \mathcal{F} encodes a function $f_t : \mathcal{V}_{q,N} \rightarrow \mathcal{V}_{q,N}$ that acts separately on each coordinate $v_j \in \mathbb{R}^q$,

$$f_t(\mathbf{v}) = (f_t^1(v_1), \dots, f_t^N(v_N)) \in \mathcal{V}_{q,N}. \quad (9)$$

3. $\mathbf{v}^0 \in \mathcal{V}_{q,N}$ is a starting condition.

The algorithm itself is a sequence of iterates generated by:

$$\mathbf{v}^{t+1} = \frac{\mathbf{A}}{\sqrt{N}} f_t(\mathbf{v}^t) - f_{t-1}(\mathbf{v}^{t-1}) \mathbf{B}_t^T \quad (10)$$

where \mathbf{B}_t is the $q \times q$ Onsager matrix given by

$$\mathbf{B}_t = \frac{1}{N} \sum_{j=1}^N \partial f_t^j(\mathbf{v}_j^t). \quad (11)$$

where ∂f_t^j denotes the Jacobian matrix of f_t^j . The limiting properties of the AMP sequences are well known and can be found in [Javanmard and Montanari, 2013, Theorem 1].

The inhomogeneous AMP — We now define an inhomogeneous AMP iteration which takes into account the block-constant structure of the noise:

Definition 2.2. An inhomogeneous AMP on $\mathcal{V}_{1,N} = \mathbb{R}^N$ is represented by $(\mathbf{A}, \mathcal{F}, \mathbf{x}^0, \Delta)$, where the terms $\mathbf{A}, \mathcal{F}, \mathbf{x}^0$ are defined in Definition 2.1 and Δ is the $N \times N$ variance profile encoded by $\tilde{\Delta} \in \mathbb{R}^{q \times q}$ and grouping $g : [N] \rightarrow [q]$ defined by (1). We further assume that the family of functions \mathcal{F} is encoded by functions $f_t^a : \mathbb{R} \mapsto \mathbb{R}$ for $a \in [q]$ which define the group dependent function

$$f_t(\mathbf{x}) = (f_t^{g(1)}(x_1), \dots, f_t^{g(N)}(x_N)) \in \mathbb{R}^N. \quad (12)$$

The sequence of iterates $\mathbf{x}^t \in \mathbb{R}^N$ of the $(\mathbf{A}, \mathcal{F}, \mathbf{v}^0, \Delta)$ are defined as follows:

$$\mathbf{x}^{t+1} = \left(\frac{1}{\sqrt{N}\sqrt{\Delta}} \odot \mathbf{A} \right) f_t(\mathbf{x}^t) - \mathbf{b}_t \odot f_{t-1}(\mathbf{x}^{t-1}), \quad (13)$$

where $\frac{1}{\sqrt{\Delta}}$ is the Hadamard inverse square root of the noise, and the Onsager term \mathbf{b}_t has the following form

$$\mathbf{b}_t = \frac{1}{N} \begin{pmatrix} \frac{1}{\Delta_{11}}(f_t^{g(1)})'(x_1^t) & + & \dots & + & \frac{1}{\Delta_{1N}}(f_t^{g(N)})'(x_N^t) \\ \vdots & & \vdots & & \vdots \\ \frac{1}{\Delta_{N1}}(f_t^{g(1)})'(x_1^t) & + & \dots & + & \frac{1}{\Delta_{NN}}(f_t^{g(N)})'(x_N^t) \end{pmatrix} = \frac{1}{N\Delta} f_t'(x^t) \in \mathbb{R}^N. \quad (14)$$

State evolution of the inhomogeneous AMP — Through a continuous embedding, we will reduce our inhomogeneous AMP to the matrix AMP framework, and recover the state evolution of the inhomogeneous AMP. We define the diagonal matrix operator $\text{blockdiag} : \mathbb{R}^N \mapsto \mathcal{V}_{q,N}$ which outputs a block diagonal matrix according to the block structure of our discretization of $[N]$:

$$\text{blockdiag}(\mathbf{v}) = \mathbf{M} \quad \text{where} \quad M_{ij} = \begin{cases} v_j & g(j) = i \\ 0 & \text{otherwise.} \end{cases}$$

Likewise, we define the projection operator $\text{blockproj} : \mathcal{V}_{q,N} \mapsto \mathbb{R}^N$ which extracts a vector of size N from a $N \times q$ according to the block structure of $[N]$ by

$$\text{blockproj}(\mathbf{M}) = (M_{ig(i)})_{i \leq N} \in \mathbb{R}^N.$$

Under these changes of variables, we define

$$\mathbf{r}^t = \text{blockdiag}(\mathbf{x}^t) \in \mathcal{V}_{q,N} \quad \text{for } t \geq 0$$

and $\tilde{f}_t : (\mathbb{R}^q)^N \mapsto (\mathbb{R}^q)^N$ by

$$\left(\tilde{f}_t(\mathbf{r}^t)\right)_{ij} = \frac{1}{\sqrt{\tilde{\Delta}_{g(i)j}}} f_t^{g(i)}(x_i^t) \quad \text{for } i, j \in [N] \times [q]. \quad (15)$$

We encode the family of functions \tilde{f}_t by $\tilde{\mathcal{F}}(\Delta)$.

Lemma 2.3. *Let \mathbf{x}^t be iterates from the AMP $(\mathbf{A}, \mathcal{F}, \mathbf{v}^0, \Delta)$. Then the iterates $\mathbf{r}^t := \text{blockdiag}(\mathbf{x}^t)$ follow the generalized matrix AMP $(\mathbf{A}, \tilde{\mathcal{F}}(\Delta), \mathbf{r}^0)$.*

Proof. We will show that the projection of the iterates \mathbf{r}^t from $(\mathbf{A}, \tilde{\mathcal{F}}(\Delta), \mathbf{r}^0)$ are the iterates from $(\mathbf{A}, \mathcal{F}, \mathbf{v}^0, \Delta)$. It is easy to check that

$$\frac{\mathbf{A}}{\sqrt{N}\sqrt{\Delta}} f_t(\mathbf{x}^t) = \text{blockproj} \left(\frac{\mathbf{A}}{\sqrt{N}} \tilde{f}_t(\mathbf{r}^t) \right). \quad (16)$$

Next, notice that Jacobian is given by

$$\partial \tilde{f}_t^i(\mathbf{r}_i^t) = \begin{bmatrix} 0 & \dots & \frac{1}{\sqrt{\tilde{\Delta}_{g(i)1}}} (f_t^{g(i)})'(x_i^t) & \dots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \dots & \frac{1}{\sqrt{\tilde{\Delta}_{g(i)q}}} (f_t^{g(i)})'(x_i^t) & \dots & 0 \end{bmatrix},$$

which is a matrix where only the column number $g(i)$ has non-zero elements. Applying (11), we thus get for $a, b \in [q] \times [q]$

$$(\mathbf{B}_t)_{ab} = \frac{1}{N\sqrt{\tilde{\Delta}_{ab}}} \sum_{i:g(i)=a} (f_t^a)'(x_i^t). \quad (17)$$

It follows that

$$\mathbf{b}_t \odot f_{t-1}(\mathbf{x}^{t-1}) = \text{blockproj} \left(\tilde{f}_{t-1}(\mathbf{r}^{t-1}) \mathbf{B}_t^T \right).$$

□

As a consequence, the inhomogeneous state evolution equations in Theorem 1.2 follow immediately from the state evolution equations of $(\mathbf{A}, \tilde{\mathcal{F}}(\Delta), \mathbf{r}^0)$ discussed in [Javanmard and Montanari, 2013, Section 2.1]. This follows from the observation that given the law of \mathbf{r}^t in the high dimensional limit, the law of $\mathbf{x}^t = \text{blockproj}(\mathbf{r}^t)$ is straightforward to compute. We define

$$(\sigma_b^{t+1})^2 := \sum_{a=1}^q \frac{c_a}{\tilde{\Delta}_{ab}} \mathbb{E} [(f_t^b(Z_b^t))^2]. \quad (18)$$

We will show that the distribution of the iterate \mathbf{x}_i^t is asymptotically normal with mean 0 and variance $(\sigma_{g(i)}^t)^2$.

Lemma 2.4 (Behavior of the inhomogeneous AMP iterates). *Suppose that Assumption 1.1 holds, and that \mathbb{P}_0 has a bounded second moment. Let $\phi: \mathbb{R}^2 \rightarrow \mathbb{R}$ be a L -pseudo-Lipschitz test function satisfying (4). For any $a \in [q]$, then the following limit holds almost surely*

$$\lim_{N \rightarrow \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \phi(x_i^t, x_i^*) = \mathbb{E}_{x_0^*, Z} \phi(\sigma_a^t Z, x_0^*)$$

where Z is an independent standard Gaussian.

Proof. In matrix-AMP [Javanmard and Montanari, 2013, Th. 1], the marginals of the iterates \mathbf{r}^t from $(\mathbf{A}, \tilde{\mathcal{F}}(\Delta), \mathbf{r}^0)$ are approximately Gaussian and encoded by the positive definite matrices

$$\widehat{\Sigma}_a^t = \mathbb{E} \left[\tilde{f}_t^i(\mathbf{Z}^t) \tilde{f}_t^i(\mathbf{Z}^t)^\top \right] \quad \text{for all } i \in C_a^N, \quad (19)$$

$$\text{with } \mathbf{Z}^t \sim N(0, \Sigma^t) \text{ and } \Sigma^{t+1} = \sum_{a=1}^q c_a \widehat{\Sigma}_a^t. \quad (20)$$

We now show that $a \in [q]$ and $i \in C_a^N$, $\widehat{\Sigma}_a^t$ depends only on $(\sigma_a^t)^2 = \Sigma_{aa}^t$. Indeed, by the definition of \tilde{f}_t^i we have

$$\widehat{\Sigma}_a^t(k, l) = \mathbb{E} \left[(\tilde{f}_t^i(\mathbf{Z}^t))_k (\tilde{f}_t^i(\mathbf{Z}^t))_l \right] = \mathbb{E} \left[\frac{1}{\sqrt{\Delta_{ak}}} \frac{1}{\sqrt{\Delta_{al}}} (f_t^a(Z_a^t))^2 \right], \quad (21)$$

where $Z_a^t \sim N(0, (\sigma_a^t)^2)$ is the a th component of the Gaussian vector \mathbf{Z}^t . The key observation here is that by construction our function $\tilde{f}_t^i: \mathbb{R}^q \mapsto \mathbb{R}^q$ depends only on the i th component Z_i^t of the Gaussian vector \mathbf{Z}^t . To characterize the limiting distribution of $\mathbf{x}^t = \text{blockproj}(\mathbf{r}^t)$, we only need to keep track of the variances $(\sigma_j^t)^2, j \in [N]$. Using (21), for a given $a \in [q]$ and $i \in C_a^N$ we write

$$\widehat{\Sigma}_a^t(g(j), g(j)) = \mathbb{E} \left[(\tilde{f}_t^i(\mathbf{Z}^t))_{g(j)} (\tilde{f}_t^i(\mathbf{Z}^t))_{g(j)} \right] = \frac{1}{\Delta_{ag(j)}} \mathbb{E} [(f_t^i(Z_i^t))^2]. \quad (22)$$

Finally, with (20) we get that for any $b \in [q]$ and any $j \in C_b^N$, using $Z_a^t \sim N(0, (\sigma_a^t)^2)$,

$$(\sigma_j^{t+1})^2 = (\sigma_b^{t+1})^2 = \Sigma_{bb}^{t+1} = \sum_{a=1}^q c_a \widehat{\Sigma}_a^t(g(j), g(j)) = \sum_{a=1}^q \frac{c_a}{\tilde{\Delta}_{ag(j)}} \mathbb{E} [(f_t^a(Z_a^t))^2] \quad (23)$$

□

The inhomogeneous spiked Wigner model in the light of the AMP approach — We now generalize the state evolution equations from Lemma 2.4 to spiked matrices with an inhomogeneous noise profile as was stated in Theorem 1.2. This reduction via a change of variables is standard, see for example [Deshpande et al., 2015, Lemma 4.4]. Remember that in the inhomogeneous version of the spiked Wigner model we observe the signal \mathbf{x}^* through an inhomogeneous channel:

$$\mathbf{Y} = \sqrt{\frac{1}{N}} \mathbf{x}^* (\mathbf{x}^*)^\top + \mathbf{A} \odot \sqrt{\Delta}. \quad (24)$$

Our AMP algorithm is defined with the following recursion:

$$\mathbf{x}^{t+1} = \left(\frac{1}{\sqrt{N\Delta}} \odot \mathbf{Y} \right) f_t(\mathbf{x}^t) - \mathbf{b}_t \odot f_{t-1}(\mathbf{x}^{t-1}) \quad (25)$$

where $\mathbf{b}_t = \frac{1}{N\Delta} f'_t(\mathbf{x}^t)$ and f is encoded by the family of functions in Definition 2.2. The main difference in contrast to the iteration (13) is that our data matrix \mathbf{Y} is no longer a centered matrix, while $\frac{1}{\sqrt{\Delta}} \odot \mathbf{A}$ is. We would like to reduce (25) to an iteration of the form (13) with respect to a different parameter \mathbf{s}^t which is uniquely determined by \mathbf{x}^t

$$\mathbf{s}^{t+1} = \left(\frac{1}{\sqrt{N}\sqrt{\Delta}} \odot \mathbf{A} \right) f_t(\mathbf{s}^t) - \mathbf{b}_t \odot f_{t-1}(\mathbf{s}^{t-1}).$$

Doing so will allow us to recover the limiting laws of the iterates from Lemma 2.4. This is done via a standard change of variables to recenter \mathbf{Y} . We will sketch the argument in this section, and defer the full proof of Theorem 1.2 to the Appendix A in the Supplementary Material.

To simplify notation, let us denote $f_t(\mathbf{x}^t) := \hat{\mathbf{x}}^t$. We proceed following the approach of [Deshpande et al., 2015, Lemma 4.4]. We rewrite (25) using the definition of \mathbf{Y} to get

$$\begin{aligned} \mathbf{x}^{t+1} &= \left(\frac{1}{\sqrt{N\Delta}} \odot \mathbf{Y} \right) \hat{\mathbf{x}}^t - \mathbf{b}_t \odot \hat{\mathbf{x}}^{t-1} \\ &= \left(\frac{1}{N\Delta} \odot \mathbf{x}^*(\mathbf{x}^*)^T \right) \hat{\mathbf{x}}^t + \left(\frac{1}{\sqrt{N}\sqrt{\Delta}} \odot \mathbf{A} \right) \hat{\mathbf{x}}^t - \mathbf{b}_t \odot \hat{\mathbf{x}}^{t-1}. \end{aligned} \quad (26)$$

If indices are independent, then by the strong law of large numbers one would expect that

$$\left(\left(\frac{1}{N\Delta} \odot \mathbf{x}^*(\mathbf{x}^*)^T \right) \hat{\mathbf{x}}^t \right)_j = x_j^* \sum_{a \in [q]} \sum_{i \in C_a^N} \frac{1}{N} \frac{x_i^* \hat{x}_i^t}{\Delta_{ji}} \rightarrow x_j^* \sum_{a \in [q]} \frac{c_a}{\Delta_{ji_a}} \mathbb{E}[x_0^* \hat{x}_{i_a}^t], \quad (27)$$

where i_a is some index belonging to the group C_a^N and x_0^* is a random variable distributed according to the prior distribution \mathbb{P}_0 . For $b \in [q]$ and $i \in C_b^N$, we define the block overlap μ_b^t using the recursion

$$\mu_b^{t+1} = \mu_b^t = \sum_{a \in [q]} \frac{c_a}{\Delta_{ab}} \mathbb{E}_{x_0^*, Z} [x_0^* f_t^a(\mu_a^t x_0^* + \sigma_a^t Z)], \quad (28)$$

where Z is a standard Gaussian random variable independent from all others sources of randomness. Notice that (28) is precisely the asymptotic behavior of the summation appearing in (27).

We now make a change of variables and track the iterates

$$\mathbf{s}^0 = \mathbf{x}^0 - \boldsymbol{\mu}^0 \odot \mathbf{x}^* \quad \mathbf{s}^t = \mathbf{x}^t - \boldsymbol{\mu}^t \odot \mathbf{x}^*, \quad t \geq 1 \quad (29)$$

where $\boldsymbol{\mu}^0$ is the vector of block overlaps of the initial condition \mathbf{x}^0 with the truth. We reduced the (25) iteration to the following iteration in which we easily recognize a version of (13):

$$\mathbf{s}^{t+1} = \left(\frac{1}{\sqrt{N}\sqrt{\Delta}} \odot \mathbf{A} \right) f_t(\mathbf{s}^t + \boldsymbol{\mu}^t \odot \mathbf{x}^*) - \mathbf{b}_t \odot f_{t-1}(\mathbf{s}^{t-1} + \boldsymbol{\mu}^{t-1} \odot \mathbf{x}^*) \quad (30)$$

with the initial condition $\mathbf{s}^0 = \mathbf{x}^0 - \boldsymbol{\mu}^0 \odot \mathbf{x}^*$ and the Onsager term taken from (14) is given by

$$\mathbf{b}_t = \frac{1}{N\Delta} f'_t(\mathbf{s}^t + \boldsymbol{\mu}^t \odot \mathbf{x}^*). \quad (31)$$

Using Lemma 2.4, we can recover the asymptotic behavior of the iterates \mathbf{x}^t given in (25) by computing the iterates $\mathbf{s}^t + \boldsymbol{\mu}^t \odot \mathbf{x}^*$ where \mathbf{s}^t follows (30) and $\boldsymbol{\mu}^t$ satisfies (28). From this reduction we obtain the following state evolution equations describing the behaviour of (25):

1. $x_j^t \cong \mu_{g(j)}^t x_0^* + \sigma_{g(j)}^t Z$ for $j \in [N]$, where $Z \sim \mathcal{N}(0, 1)$
2. $\mu_b^{t+1} = \sum_{a \in [q]} \frac{c_a}{\Delta_{ab}} \mathbb{E}_{x_0^*, Z} [x_0^* f_t^a(\mu_a^t x_0^* + \sigma_a^t Z)]$ with $x_0^* \sim \mathbb{P}_0, Z \sim \mathcal{N}(0, 1)$
3. $(\sigma_b^{t+1})^2 = \sum_{a=1}^q \frac{c_a}{\Delta_{ab}} \mathbb{E}_{x_0^*, Z} [(f_t^a(\mu_a^t x_0^* + \sigma_a^t Z))^2]$ with $x_0^* \sim \mathbb{P}_0, Z \sim \mathcal{N}(0, 1)$.

This (informally) characterizes the limiting distribution of the state evolution of the iterates from the inhomogeneous AMP stated in Theorem 1.2.

Fixed-point equation of state evolution in the Bayes-optimal setting — Suppose that we know the prior distribution \mathbb{P}_0 of x_0^* . The Bayes-optimal choice for the denoising functions $f_t^j, j \in [N]$ is simply the expectation of x_0^* with respect to the posterior distribution,

$$f_t^j(r) = f_t^{g(j)}(r) = \mathbb{E}_{\text{posterior}}[x_0^* | \mu_{g(j)}^t x_0^* + \sigma_{g(j)}^t Z = r]. \quad (32)$$

Under this Bayes-optimal setting, we can simplify the equations obtained in the previous section and see that AMP estimator is indeed an optimal one by studying its fixed point.

Proof of Theorem 1.4. For this choice of f_t^j the Nishimori identity (see for example [Lelarge and Miolane, 2019, Proposition 16]) states that for $a \in [q]$ and $j \in C_a^N$,

$$\tilde{\mu}_a^t := \mathbb{E}_{x_0^*, Z}[x_0^* f_t^a(\mu_a^t x_0^* + \sigma_a^t Z)] = \mathbb{E}[(f_t^a(\mu_a^t x_0^* + \sigma_a^t Z))^2]. \quad (33)$$

In this setting, the state evolution equations from Theorem 1.2 reduce to

$$\begin{cases} \tilde{\mu}_a^t = \mathbb{E}_{x_0^*, Z}[x_0^* f_t^a(\mu_a^t x_0^* + \sigma_a^t Z)] \\ \mu_b^{t+1} = \sum_{a \in [q]} \frac{c_a}{\Delta_{ab}} \tilde{\mu}_a^t, b \in [q] \\ (\sigma_b^{t+1})^2 = \sum_{a \in [q]} \frac{c_a}{\Delta_{ab}} \tilde{\mu}_a^t, b \in [q]. \end{cases} \quad (34)$$

Remarkably with the Bayes-optimal choice of the denoising functions we have that for $t \geq 1$ for each block $b \in [q]$, $\mu_b^{t+1} = (\sigma_b^{t+1})^2$. Therefore a necessary and sufficient condition for an estimator to be a fixed point of the state evolution is to simply have its overlaps μ_b^t unchanged by an iteration of the state evolution. This translates into the following equation for the overlaps $\mu_b, b \in [q]$

$$\mu_b = \sum_{a \in [q]} \frac{c_a}{\Delta_{ab}} \mathbb{E}_{x_0^*, Z}[x_0^* \mathbb{E}_{\text{posterior}}[x_0^* | \mu_a x_0^* + \sqrt{\mu_a} Z]]. \quad (35)$$

The result of Theorem 1.4 now follows immediately. \square

3 A spectral method adapted to the inhomogeneous spiked Wigner model

From AMP to a spectral method — Remarkably, AMP and the state evolution machinery associated with it can help us design a simple spectral algorithm that matches the information-theoretic phase transition [Guionnet et al., 2022, Remark 2.16]. Recall that Theorem 1.2 does not require the denoising functions f_t to be Bayes-optimal, but can be applied to any Lipschitz family of functions. In this section, we analyze the state evolution for the family of identity functions, $f_t(x) = x$. We can gain some intuition motivating such a choice by considering simple priors, such as the Rademacher prior. In the case of the Rademacher prior a simple computation shows that the Bayes-optimal choice of the denoising functions yields $f_t^j(\star) = \tanh(\star)$. In the first order approximation we have $\tanh(x) \approx x$. Thus, at least for the Rademacher prior, the choice of identity functions as denoising functions corresponds to the first order approximation of the Bayes-optimal choice. By Remark 3.1, we can assume that the entries of x^* have unit variance. With this choice of denoising functions the AMP iteration will simply become:

$$x^{t+1} = \left(\frac{1}{\sqrt{N\Delta}} \odot Y \right) x^t - \mathbf{b}_t \odot x^{t-1} \quad \text{where} \quad \mathbf{b}_t = \frac{1}{N\Delta} f'_t = \frac{1}{N\Delta} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}. \quad (36)$$

If we denote $B_t = \text{diag}(\mathbf{b}_t)$, it is easy to see that the fixed point of this iteration yields

$$x = \left(\frac{1}{\sqrt{N\Delta}} \odot Y \right) x - B_t x \quad (37)$$

so any x fixed by the AMP iteration (36) must be an eigenvector of the matrix

$$\tilde{Y} = \left(\frac{1}{\sqrt{N\Delta}} \odot Y \right) - B_t = \left(\frac{1}{\sqrt{N\Delta}} \odot Y \right) - \text{diag} \left(\frac{1}{N\Delta} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right). \quad (38)$$

A simple spectral method consists in taking the principal eigenvector (associated to the largest eigenvalue) of the matrix $\tilde{Y} = \frac{Y}{\sqrt{N\Delta}} - B_t$.

Analysis of the spectral method using state evolution — The spectral algorithm described above behaves like the AMP iteration with identity denoising functions around its fixed point. Therefore we can analyze this spectral algorithm using state evolution machinery for the AMP iteration. In the case of identity functions we have $f_t^a(x) = x$ for all $a \in [q]$, so

$$\mathbb{E}_{x_0^*, Z}[x_0^* f_t^a(\mu_a^t x_0^* + \sigma_a^t Z)] = \mathbb{E}_{x_0^*, Z}[x_0^*(\mu_a^t x_0^* + \sigma_a^t Z)] = \mu_a^t \quad (39)$$

$$\mathbb{E}_{x_0^*, Z}[(f_t^a(\mu_a^t x_0^* + \sigma_a^t Z))^2] = \mathbb{E}_{x_0^*, Z}[(\mu_a^t x_0^* + \sigma_a^t Z)^2] = (\mu_a^t)^2 + (\sigma_a^t)^2 \quad (40)$$

which transforms state evolution equations (5) into the following simple form:

$$\mu_b^{t+1} = \sum_{a \in [q]} \frac{c_a}{\Delta_{ba}} \mu_a^t \quad \text{and} \quad (\sigma_b^{t+1})^2 = \sum_{a \in [q]} \frac{c_a}{\Delta_{ba}} ((\mu_a^t)^2 + (\sigma_a^t)^2). \quad (41)$$

Rewriting the overlap state evolution in a matrix form we get for $\mathbf{c} = (c_a)_{a \in [q]}$ that

$$\text{diag}(\sqrt{\mathbf{c}}) \boldsymbol{\mu}^{t+1} = \text{diag}(\sqrt{\mathbf{c}}) \frac{1}{\Delta} \text{diag}(\sqrt{\mathbf{c}}) (\text{diag}(\sqrt{\mathbf{c}}) \boldsymbol{\mu}^t). \quad (42)$$

The special form of state evolution above is rather informative and we can interpret it as follows. First of all, in the regime where $\lambda(\Delta) = \|\text{diag}(\sqrt{\mathbf{c}}) \frac{1}{\Delta} \text{diag}(\sqrt{\mathbf{c}})\|_{op} < 1$, we can see from (42) that any iteration of the AMP recursion (36) contracts the vector overlap (multiplying it by a matrix with the operator norm smaller than 1). Thus the only possible overlap of a fixed point of (36) is 0. Moreover, we have defined the matrix $\tilde{\mathbf{Y}}$ in a way that any eigenvector of this matrix is a fixed point of the AMP recursion (36). Thus any eigenvector of $\tilde{\mathbf{Y}}$ must have zero expected overlap with the signal, meaning that the spectral method is uninformative in the regime $\lambda(\Delta) < 1$.

Second of all, even though we cannot say anything rigorous in the regime $\lambda(\Delta) > 1$, looking at state evolution equations of the AMP recursion (36), we see instability (overlap blows up but so does the variance). We conjecture that the principal eigenvector of $\tilde{\mathbf{Y}}$ correlates with the signal exactly in the regime $\lambda(\Delta) > 1$ which would imply that the BBP transition happens precisely at $\lambda(\Delta) = 1$.

Remark 3.1. In general, we can let $\gamma = \mathbb{E}_{x_0^*}[(x_0^*)^2]$ and consider the normalized matrix

$$\bar{\mathbf{Y}} = \frac{\mathbf{Y}}{\gamma} = \sqrt{\frac{1}{N}} \frac{\mathbf{x}^* (\mathbf{x}^*)^T}{\gamma} + \mathbf{A} \odot \frac{\sqrt{\Delta}}{\gamma} = \sqrt{\frac{1}{N}} \bar{\mathbf{x}}^* (\bar{\mathbf{x}}^*)^T + \mathbf{A} \odot \sqrt{\bar{\Delta}}$$

for $\bar{\mathbf{x}} = \frac{\mathbf{x}}{\sqrt{\gamma}}$ and $\bar{\Delta} = \frac{\Delta}{\gamma^2}$. Notice that the entries of $\bar{\mathbf{x}}$ now have unit variance. Under this setting, the transition of the transformation in (38) applied to $\bar{\mathbf{Y}}$, which appears in (7), has transition at

$$\lambda(\bar{\Delta}) = \left\| \text{diag}(\sqrt{\mathbf{c}}) \frac{1}{\bar{\Delta}} \text{diag}(\sqrt{\mathbf{c}}) \right\|_{op} = \mathbb{E}_{x_0^*}[(x_0^*)^2]^2 \left\| \text{diag}(\sqrt{\mathbf{c}}) \frac{1}{\Delta} \text{diag}(\sqrt{\mathbf{c}}) \right\|_{op} < 1$$

which is the generalized SNR defined in (8).

Acknowledgments and Disclosure of Funding

We thank Alice Guionnet and Lenka Zdeborova for valuable discussions. We acknowledge funding from the ERC Project LDRAM: ERC-2019-ADG Project 884584, and by the Swiss National Science Foundation grant SNFS OperaGOST, 200021 200390

References

- David L Donoho and Iain M Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the american statistical association*, 90(432):1200–1224, 1995.
- Sandrine Péché. Deformed ensembles of random matrices. In *Proceedings of the International Congress of Mathematicians, Seoul*, volume 3, pages 1059–1174, 2014.
- Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.*, 33(5):1643–1697, 2005. ISSN 0091-1798. doi: 10.1214/009117905000000233. URL <http://dx.doi.org/10.1214/009117905000000233>.
- Jean Barbier and Galen Reeves. Information-theoretic limits of a multiview low-rank symmetric spiked matrix model. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 2771–2776. IEEE, 2020.
- Joshua K Behne and Galen Reeves. Fundamental limits for rank-one matrix estimation with groupwise heteroskedasticity. In *International Conference on Artificial Intelligence and Statistics*, pages 8650–8672. PMLR, 2022.
- Diego Alberici, Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione. The multi-species mean-field spin-glass on the nishimori line. *Journal of Statistical Physics*, 182:1–20, 2021a.
- Diego Alberici, Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione. A statistical physics approach to a multi-channel wigner spiked model. *Europhysics Letters*, 136(4):48001, 2022.
- Diego Alberici, Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione. The solution of the deep Boltzmann machine on the Nishimori line. *Comm. Math. Phys.*, 387(2):1191–1214, 2021b. ISSN 0010-3616. doi: 10.1007/s00220-021-04165-0. URL <https://doi.org/10.1007/s00220-021-04165-0>.
- Brian Karrer and Mark EJ Newman. Stochastic blockmodels and community structure in networks. *Physical review E*, 83(1):016107, 2011.
- Alice Guionnet, Justin Ko, Florent Krzakala, and Lenka Zdeborová. Low-rank matrix estimation with inhomogeneous noise. *arXiv preprint arxiv:2208.05918*, 2022.
- Hong-Bin Chen and Jiaming Xia. Hamilton-Jacobi equations for inference of matrix tensor products. *Ann. Inst. Henri Poincaré Probab. Stat.*, 58(2):755–793, 2022. ISSN 0246-0203. doi: 10.1214/21-aihp1183. URL <https://doi.org/10.1214/21-aihp1183>.
- Hong-Bin Chen, Jean-Christophe Mourrat, and Jiaming Xia. Statistical inference of finite-rank tensors, 2021. URL <https://arxiv.org/abs/2104.05360>.
- Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011. doi: 10.1109/TIT.2010.2094817.
- Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the two-groups stochastic block model. *CoRR*, abs/1507.08685, 2015. URL <http://arxiv.org/abs/1507.08685>.
- Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(7):073403, 2017.
- Antoine Maillard, Florent Krzakala, Yue M Lu, and Lenka Zdeborová. Construction of optimal spectral methods in phase retrieval. In *Mathematical and Scientific Machine Learning*, pages 693–720. PMLR, 2022.
- David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

- Sundeeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2168–2172. IEEE, 2011.
- Cédric Gerbelot and Raphaël Berthier. Graph-based approximate message passing iterations. *arXiv preprint arXiv:2109.11905*, 2021.
- Oliver Y Feng, Ramji Venkataramanan, Cynthia Rush, Richard J Samworth, et al. A unifying tutorial on approximate message passing. *Foundations and Trends® in Machine Learning*, 15(4):335–536, 2022.
- Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.
- Erwin Bolthausen. An iterative construction of solutions of the tap equations for the sherrington–kirkpatrick model. *Communications in Mathematical Physics*, 325(1):333–366, 2014.
- Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016.
- Michael Celentano, Andrea Montanari, and Yuchen Wu. The estimation error of general first order methods. In *Conference on Learning Theory*, pages 1078–1141. PMLR, 2020.
- Florent Krzakala, Marc Mézard, Francois Sausset, Yifan Sun, and Lenka Zdeborová. Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(08):P08009, 2012.
- David L Donoho, Adel Javanmard, and Andrea Montanari. Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. *IEEE transactions on information theory*, 59(11):7434–7464, 2013.
- Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115–144, 2013. doi: 10.1093/imaia/iat004.
- Riccardo Rossetti and Galen Reeves. Approximate message passing for the matrix tensor product model, 2023.
- Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, and Lenka Zdeborová. Rank-one matrix estimation: analysis of algorithmic and information theoretic limits by the spatial coupling method. *arXiv preprint arXiv:1812.02537*, 2018.
- Ahmed El Alaoui, Florent Krzakala, and Michael Jordan. Fundamental limits of detection in the spiked Wigner model. *The Annals of Statistics*, 48(2):863 – 885, 2020. doi: 10.1214/19-AOS1826. URL <https://doi.org/10.1214/19-AOS1826>.
- Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probab. Theory Related Fields*, 173(3-4):859–929, 2019. ISSN 0178-8051. doi: 10.1007/s00440-018-0845-x. URL <https://doi.org/10.1007/s00440-018-0845-x>.
- Cosme Louart and Romain Couillet. Concentration of measure and large random matrices with an application to sample covariance matrices. *arXiv preprint arXiv:1805.08295*, 2018.
- Alaa Saade, Florent Krzakala, and Lenka Zdeborová. Spectral clustering of graphs with the bethe hessian. *Advances in Neural Information Processing Systems*, 27, 2014.
- Benjamin Aubin, Bruno Loureiro, Antoine Maillard, Florent Krzakala, and Lenka Zdeborová. The spiked matrix model with generative priors. *Advances in Neural Information Processing Systems*, 32, 2019.
- Marco Mondelli and Andrea Montanari. Fundamental limits of weak recovery with applications to phase retrieval. In *Conference On Learning Theory*, pages 1445–1450. PMLR, 2018.

- Marco Mondelli, Christos Thrampoulidis, and Ramji Venkataramanan. Optimal combination of linear and spectral estimators for generalized linear models. *Foundations of Computational Mathematics*, 22(5):1513–1566, 2022.
- Ramji Venkataramanan, Kevin Kögler, and Marco Mondelli. Estimation in rotationally invariant generalized linear models via approximate message passing. In *International Conference on Machine Learning*, pages 22120–22144. PMLR, 2022.
- Marco Mondelli and Ramji Venkataramanan. Approximate message passing with spectral initialization for generalized linear models. In *International Conference on Artificial Intelligence and Statistics*, pages 397–405. PMLR, 2021.
- Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. ISBN 978-0-521-19452-5.

A Proof of Theorem 1.2

We now provide a rigorous proof of the result that was sketched in Section 2. This proof is essentially identical to [Deshpande et al., 2015, Lemma 4.4]. Recall the iterates (30) given by $\mathbf{s}^0 = \mathbf{x}^0 - \boldsymbol{\mu}^0 \odot \mathbf{x}^*$ and

$$\mathbf{s}^{t+1} = \left(\frac{1}{\sqrt{N}\sqrt{\Delta}} \odot \mathbf{A} \right) f_t(\mathbf{s}^t + \boldsymbol{\mu}^t \odot \mathbf{x}^*) - \mathbf{b}_t \odot f_{t-1}(\mathbf{s}^{t-1} + \boldsymbol{\mu}^{t-1} \odot \mathbf{x}^*), \quad (43)$$

where $\boldsymbol{\mu}^t = (\mu_i^t)_{i \leq N}$ is given by the recursion

$$\mu_i^{t+1} = \mu_{g(i)}^{t+1} = \sum_{a \in [q]} \frac{c_a}{\Delta_{ag(i)}} \mathbb{E}_{x_0^*, Z} [x_0^* f_t^a(\mu_a^t x_0^* + \sigma_a^t Z)]$$

and $\mathbf{x}^* = (x_i^*)_{i \in [N]}$ is a vector with independent coordinates distributed according to \mathbb{P}_0 . By Lemma 2.4, for each $a \in [q]$, and any pseudo-Lipschitz function $\phi : \mathbb{R} \rightarrow \mathbb{R} \mapsto \mathbb{R}$ we have that almost surely

$$\lim_{N \rightarrow \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \phi(s_i^t, x_i^*) = \mathbb{E}_{x_0^*, Z} \phi(\sigma_a^t Z, x_0^*) \quad (44)$$

where

$$(\sigma_b^{t+1})^2 := \sum_{a=1}^q \frac{c_a}{\Delta_{ab}} \mathbb{E}_Z [(f_t^b(Z_b^t))^2].$$

as was defined in (18). For any pseudo-Lipschitz function $\psi : \mathbb{R} \rightarrow \mathbb{R}$, we have $\phi(x, y) = \phi(x - \mu_a^t y)$ is also pseudo-Lipschitz, so (44) implies that

$$\lim_{N \rightarrow \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} \psi(s_i^t + \mu_a^t x_i^*) = \mathbb{E}_{x_0^*, Z} \psi(\sigma_a^t Z + \mu_a^t x_0^*) \quad (45)$$

almost surely.

Now let $\hat{\mathbf{x}}^t$ be the iterates from the spiked AMP iteration for the inhomogeneous Wigner matrix (24) we derived in (26)

$$\mathbf{x}^{t+1} = \left(\frac{1}{N\Delta} \odot \mathbf{x}^* (\mathbf{x}^*)^T \right) \hat{\mathbf{x}}^t + \left(\frac{1}{\sqrt{N}\sqrt{\Delta}} \odot \mathbf{A} \right) \hat{\mathbf{x}}^t - \mathbf{b}_t \odot \hat{\mathbf{x}}^{t-1}. \quad (46)$$

It now suffices to show that for fixed t and all $a \in [q]$ that

$$\lim_{N \rightarrow \infty} \frac{1}{|C_a^N|} \sum_{i \in C_a^N} (\psi(s_i^t + \mu_a^t x_i^*) - \psi(x_i^t)) = 0 \quad (47)$$

almost surely. This will imply that $s_i^t + \mu_a^t x_i^*$ and x_i^t have the same asymptotic distribution which finish the proof of Theorem 1.2 by (45).

We now prove (47). Since ψ is L -pseudo-Lipschitz we have

$$\begin{aligned} |\psi(s_i^t + \mu_a^t x_i^*) - \psi(x_i^t)| &\leq L(1 + |s_i^t + \mu_a^t x_i^*| + |x_i^t|) |s_i^t + \mu_a^t x_i^* - x_i^t| \\ &\leq 2L |s_i^t + \mu_a^t x_i^* - x_i^t| (1 + |s_i^t + \mu_a^t x_i^*| + |s_i^t + \mu_a^t x_i^* - x_i^t|). \end{aligned}$$

The Cauchy–Schwarz inequality implies that

$$\begin{aligned} &\left| \frac{1}{|C_a^N|} \sum_{i \in C_a^N} (\psi(s_i^t + \mu_a^t x_i^*) - \psi(x_i^t)) \right| \\ &\leq \frac{2L}{C_a^N} (\sqrt{C_a^N} \|\mathbf{s}_a^t + \mu_a^t \mathbf{x}_a^* - \mathbf{x}_a^t\|_2 + \|\mathbf{s}_a^t + \mu_a^t \mathbf{x}_a^*\|_2 \|\mathbf{s}_a^t + \mu_a^t \mathbf{x}_a^* - \mathbf{x}_a^t\|_2 + \|\mathbf{s}_a^t + \mu_a^t \mathbf{x}_a^* - \mathbf{x}_a^t\|_2^2) \end{aligned}$$

where $\mathbf{s}_a^t = (s_i^t)_{i \in C_a^N} \in \mathbb{R}^{|C_a^N|}$, $\mathbf{x}_a^t = (x_i^t)_{i \in C_a^N} \in \mathbb{R}^{|C_a^N|}$. Therefore, to prove (47) it suffices to prove that for all $t \geq 0$,

$$\lim_{N \rightarrow \infty} \frac{1}{|C_a^N|} \|\mathbf{s}_a^t + \mu_a^t \mathbf{x}_a^* - \mathbf{x}_a^t\|_2^2 \rightarrow 0 \quad (48)$$

$$\limsup_{N \rightarrow \infty} \frac{1}{|C_a^N|} \|\mathbf{s}_a^t + \mu_a^t \mathbf{x}_a^*\|_2^2 \rightarrow 0 \quad (49)$$

Clearly, if we initialize $\mathbf{x}^0, \mathbf{s}^0$ at zero then (48) and (49) are satisfied by our state evolution equations (5). Notice that (49) follows directly from (45) applied to the square function. We use here that we assumed that the second moment of \mathbf{x}^* is finite.

We now focus on proving (48) through strong induction. By definition of the iterates (43) and (46),

$$\begin{aligned} & (\mathbf{s}_a^t + \mu_a^t \mathbf{x}_a^* - \mathbf{x}_a^t) \\ &= \left[\left(\frac{1}{\sqrt{N}\sqrt{\Delta}} \odot \mathbf{A} \right) f_{t-1}(\mathbf{s}^{t-1} + \boldsymbol{\mu}^{t-1} \odot \mathbf{x}^*) - \left(\frac{1}{\sqrt{N}\sqrt{\Delta}} \odot \mathbf{A} \right) f_{t-1}(\mathbf{x}^{t-1}) \right. \\ & \quad + \mu^t \odot \mathbf{x}^* - \left(\frac{1}{N\Delta} \odot \mathbf{x}^* (\mathbf{x}^*)^T \right) f_{t-1}(\mathbf{x}^{t-1}) \\ & \quad \left. + \mathbf{b}_{t-1}^x \odot f_{t-2}(\mathbf{x}^{t-2}) - \mathbf{b}_{t-1}^s \odot f_{t-2}(\mathbf{s}^{t-2} + \boldsymbol{\mu}^{t-2} \odot \mathbf{x}^*) \right]_{i \in C_a^N} \end{aligned}$$

where $[\cdot]_i$ corresponds to the i th row of a vector and \mathbf{b}_{t-1}^x and \mathbf{b}_t^s are the Onsager terms defined in (14) with respect to \mathbf{x}^{t-1} and \mathbf{s}^{t-1} respectively. The Cauchy–Schwarz inequality and Jensen’s inequality imply that there exists some universal constant C such that

$$\begin{aligned} & \frac{1}{|C_a^N|} \|\mathbf{s}_a^t + \mu_a^t \mathbf{x}_a^* - \mathbf{x}_a^t\|_2^2 \\ & \leq \frac{C}{|C_a^N|} \sum_{i \in C_a^N} \frac{1}{N} \left\| \left[\frac{1}{\sqrt{\Delta}} \odot \mathbf{A} \right]_i \right\|_2^2 \| [f_{t-1}(\mathbf{s}^{t-1} + \boldsymbol{\mu}^{t-1} \odot \mathbf{x}^*) - f_{t-1}(\mathbf{x}^{t-1})]_i \|_2^2 \\ & \quad + \frac{C}{|C_a^N|} \sum_{i \in C_a^N} \left(\mu_a^t - \left[\frac{1}{N\Delta} (f_{t-1}(\mathbf{x}^{t-1}) \odot \mathbf{x}^*) \right]_i \right)^2 (x_i^*)^2 \\ & \quad + \frac{C}{|C_a^N|} \sum_{i \in C_a^N} ([\mathbf{b}_{t-1}^x]_i - [\mathbf{b}_{t-1}^s]_i)^2 [f_{t-2}(\mathbf{s}^{t-2} + \boldsymbol{\mu}^{t-2} \odot \mathbf{x}^*)]_i^2 \\ & \quad + \frac{C}{|C_a^N|} \sum_{i \in C_a^N} [\mathbf{b}_{t-1}^x]_i^2 ([f_{t-2}(\mathbf{x}^{t-2})]_i - [f_{t-2}(\mathbf{s}^{t-2} + \boldsymbol{\mu}^{t-2} \odot \mathbf{x}^*)]_i)^2 \end{aligned}$$

We now control each term separately.

1. To control the first term, notice that the matrix $\frac{1}{N} \left[\frac{1}{\sqrt{\Delta}} \odot \mathbf{A} \right]$ has iid entries within blocks and the sizes of the blocks diverge with the dimension, so we can control the sums of the squares of within each block using standard operator norm bounds Anderson et al. [2010]. The first term vanishes in the limit because f is pseudo-Lipschitz so we can apply the induction hypothesis bound which controls (48) at time $t - 1$.
2. To control the second term, notice that for $i \in C_a^N$ by Lemma 2.4 applied to the pseudo-Lipschitz function $y f_{t-1}(x)$ that

$$\left[\frac{1}{N\Delta} (f_{t-1}(\mathbf{x}^{t-1}) \odot \mathbf{x}^*) \right]_i \rightarrow \mu_a$$

almost surely. This implies that the average of such terms vanishes since we assumed that the second moment $\mathbb{E}[x_0^*]^2$ is finite.

3. To control the third and fourth terms, we can expand the definition of the Onsager terms and use the assumption that f' is pseudo-Lipschitz and almost surely bounded. Both terms vanish because our strong induction hypothesis gives us control of (48) at time $t - 2$.

Since all terms vanish in the limit, we have proven (48) for all $a \in [q]$, which finishes the proof of statement (47) and the proof of Theorem 1.2.

B Comparison with a naive PCA spectral method

In this appendix, we wish to show how the spectral method we propose differs, in practice, from a naive PCA. We provide an example of the spectrums of \mathbf{Y} and $\tilde{\mathbf{Y}}$ before and after the transition at

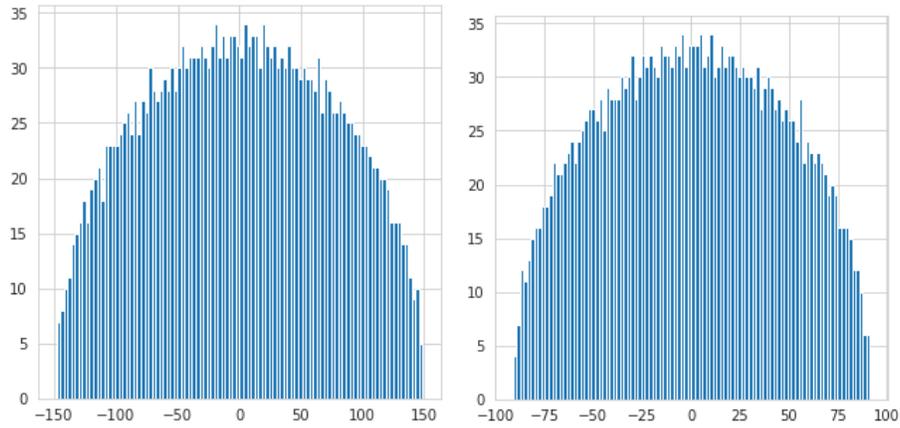


Figure 3: Illustration of the spectrum of $Y \in \mathbb{R}^{2500 \times 2500}$ evaluated at noise profiles with $\text{snr} \lambda(\Delta) = 0.7$ (left, before the transition) and on the left and 1.8 on the right (after the transition). There is no outlying eigenvalue in contrast to the transformed matrix: the transition for a naive spectral method is sub-optimal.

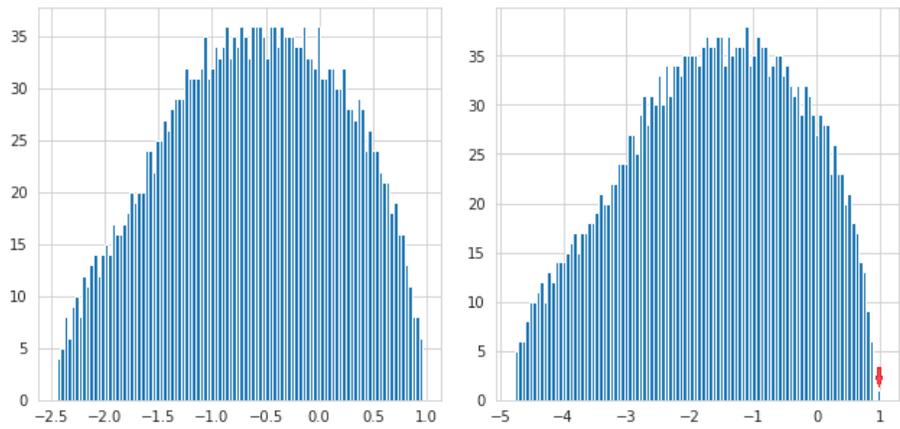


Figure 4: Illustration of the spectrum of $\tilde{Y} \in \mathbb{R}^{2500 \times 2500}$ evaluated at noise profiles with $\text{snr} \lambda(\Delta) = 0.7$ (left, before the transition) and on the left and 1.8 on the right (after the transition), with the outlying eigenvector correlated with the spike arises at eigenvalue one. This is at variance with the results of the naive method in Fig. 3

$\text{SNR}(\Delta) = 1$. In Figure 3 there is no clear separation of the extremal eigenvalue of Y from the bulk around this transition. This is in contrast to Figure 4 where there is an extremal eigenvalue of \tilde{Y} appearing at value one.