Construction of optimal spectral methods in phase retrieval

Antoine Maillard   ANTOINE.MAILLARD@ENS.FR
Laboratoire de Physique de l'ENS, PSL University, CNRS, Sorbonne Universités, Paris, France.

Florent Krzakala   FLORENT.KRZAKALA@EPFL.CH
IdePHICS laboratory, EPFL, Switzerland.

Yue M. Lu   YUELU@SEAS.HARVARD.EDU
John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA.

Lenka Zdeborová   LENKA.ZDEBOROVA@EPFL.CH
SPOC laboratory, EPFL, Switzerland.

Editors: Joan Bruna, Jan S Hesthaven, Lenka Zdeborova

Abstract

We consider the *phase retrieval* problem, in which the observer wishes to recover a n-dimensional real or complex signal $\mathbf{X}^\star$ from the (possibly noisy) observation of $|\boldsymbol{\Phi}\mathbf{X}^\star|$, in which $\boldsymbol{\Phi}$ is a matrix of size $m \times n$. We consider a *high-dimensional* setting where $n, m \to \infty$ with $m/n = \mathcal{O}(1)$, and a large class of (possibly correlated) random matrices $\boldsymbol{\Phi}$ and observation channels. Spectral methods are a powerful tool to obtain approximate observations of the signal $\mathbf{X}^\star$ which can be then used as initialization for a subsequent algorithm, at a low computational cost. In this paper, we extend and unify previous results and approaches on spectral methods for the phase retrieval problem. More precisely, we combine the linearization of message-passing algorithms and the analysis of the *Bethe Hessian*, a classical tool of statistical physics. Using this toolbox, we show how to derive optimal spectral methods for arbitrary channel noise and right-unitarily invariant matrix $\boldsymbol{\Phi}$, in an automated manner (i.e. with no optimization over any hyperparameter or preprocessing function).

Keywords: Phase retrieval, spectral methods, message-passing algorithms.

1. Introduction

1.1. Setting of the problem and related works

In the phase retrieval problem, one aims to recover an unknown *signal* $\mathbf{X}^\star \in \mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$) from m measurements $\{y_\mu\}$, which are noisy observations of $|\boldsymbol{\Phi}\mathbf{X}^\star|$ (the modulus is applied element-wise), with $\boldsymbol{\Phi} \in \mathbb{K}^{m \times n}$ a (random) sensing matrix. This model arises in a large set of problems ranging from signal processing Fienup (1982); Unser and Eden (1988); Drémeau et al. (2015) to statistical estimation Candès et al. (2015); Jaganathan et al. (2015), optics, X-ray crystallography, astronomy or microscopy Shechtman et al. (2015), where detectors can often only measure information about the amplitude of signals, and lose all information about its phase. Phase retrieval is also a textbook example of a learning problem with a highly non-convex loss landscape Netrapalli et al. (2015); Sun et al. (2018); Hand et al. (2018).

The majority of algorithms developed to solve this problem are based either on semi-definite programming relaxations Candes et al. (2015); Waldspurger (2018); Goldstein and Studer (2018) or on more direct non-convex optimization procedures, e.g. Wirtinger flow Candès et al. (2015) or

approximate message-passing Schniter and Rangan (2014); Mondelli and Venkataramanan (2021) to name a few. In general, these optimization methods require an "informed" initialization $\hat{\mathbf{X}}$, i.e. that is positively correlated with the signal $\mathbf{X}^\star$. The privileged class of algorithms to obtain such initializations in a computationally cheap manner are spectral methods, i.e. estimates given by the principal eigenvector of an appropriate matrix constructed from the sensing matrix $\mathbf{\Phi}$ and the observations $\{y_\mu\}$ Mondelli et al. (2020); Luo et al. (2019); Ma et al. (2021).

In the present work, we consider a *high-dimensional limit* (or thermodynamic limit in the statistical physics language), in which $n, m \to \infty$ with $\alpha \equiv m/n = \Theta(1)$. In this limit, a great amount of work is present both in the statistical physics and in the information theory literature for different assumptions on the matrix $\mathbf{\Phi}$. The asymptotic optimal performances in a large class of problems including phase retrieval were conjectured using the non-rigorous replica method of statistical physics in Kabashima (2008); Takahashi and Kabashima (2020), and these results were extended and partly proven in Barbier et al. (2019); Maillard et al. (2020). Specifically for the phase retrieval problem, the limits of weak-recovery were analyzed for Gaussian matrices $\mathbf{\Phi}$ in Lu and Li (2020); Mondelli and Montanari (2019); Luo et al. (2019), and for column-unitary $\mathbf{\Phi}$ in Ma et al. (2021); Dudeja et al. (2020b,a). In this work we derive the optimal spectral methods for a more generic assumption of right orthogonal (or unitary) invariance, that is we assume:

Hypothesis 1 (Matrix ensemble) *For every $\boldsymbol{O} \in \mathcal{U}(n)$ (or $\mathcal{O}(n)$ in the real case), the following equality holds in distribution $\mathbf{\Phi} \stackrel{\mathrm{d}}{=} \mathbf{\Phi}\boldsymbol{O}$. We assume that the spectral measure of $\mathbf{\Phi}^\dagger\mathbf{\Phi}/n$ weakly converges (a.s.) to a deterministic probability measure ν and we designate $\langle f(\lambda) \rangle_\nu \equiv \int \nu(\mathrm{d}\lambda) f(\lambda)$ the linear statistics of ν.*

We assume to have access to a factorized prior distribution P_0 used to generate $\mathbf{X}^\star$, with zero mean and variance $\rho = \mathbb{E}_{P_0}[|x|^2] > 0$, as well as the "channel" distribution $P_{\mathrm{out}}(y|z)$, giving the probability of the observations conditioned on the value of $\mathbf{\Phi}\mathbf{X}^\star$. The observations are therefore generated as:

$$Y_\mu \sim P_{\mathrm{out}}\Big(\cdot \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_{\mu i} X_i^\star \Big), \quad 1 \le \mu \le m. \tag{1}$$

Eq. (1) defines the very general class of *Generalized Linear Models* (GLMs). The present work covers a wide class of *phase retrieval* problems, in which we assume that $P_{\mathrm{out}}(y|z)$ is a function of $|z|^2$, and in which the prior distribution P_0 is also symmetric: $P_0(x) = P_0(|x|)$. The knowledge of P_0, P_{out} allows us to discuss the so-called "Bayes-optimal" estimator: although somewhat restrictive this knowledge allows for many insightful theoretical studies. The information-theoretic and algorithmic limits of the models described by eq. (1) have been rigorously analyzed in Barbier et al. (2019); Maillard et al. (2020). The Bayes-optimal estimation can be summarized in the study of the *posterior probability* of $\mathbf{x}$ given the observations $\mathbf{Y}$ and the sensing matrix $\mathbf{\Phi}$:

$$P(\mathbf{x}|\mathbf{Y}, \mathbf{\Phi}) \equiv \frac{1}{\mathcal{Z}_n(\mathbf{Y})} \prod_{i=1}^n P_0(x_i) \prod_{\mu=1}^m P_{\mathrm{out}}\Big(Y_\mu \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n \Phi_{\mu i} x_i \Big). \tag{2}$$

The logarithm of the normalization $(1/n) \ln \mathcal{Z}_n(\mathbf{Y})$ is usually called the *free entropy* in the statistical physics terminology. We will generically denote by $\langle \cdot \rangle$ the average with respect to the posterior probability (2) of $\mathbf{x}$. A key role in this paper will be played by the algorithmic *weak recovery* threshold, called $\alpha_{\mathrm{WR,Algo}}$, defined in such a way that for $\alpha < \alpha_{\mathrm{WR,Algo}}$ all known polynomial-time

estimators are uncorrelated with the signal $\mathbf{X}^\star$, while for $\alpha > \alpha_{\text{WR,Algo}}$, there exists estimators that recover a finite fraction of the signal in polynomial time. This algorithmic weak recovery threshold depends on the spectral distribution of the matrix $\boldsymbol{\Phi}$ and of the specific form of the output channel distribution. Interestingly, it only depends on the prior distribution P_0 via its variance ρ. Its derivation has been presented in Maillard et al. (2020), where it was shown that $\alpha_{\text{WR,Algo}}$ is the only solution to the equation:

$$\alpha = \frac{\langle\lambda\rangle_\nu^2}{\langle\lambda^2\rangle_\nu}\Big(1 + \Big[\int \mathrm{d}y \frac{\Big|\int_{\mathbb{K}} \mathcal{D}_\beta z\, (|z|^2 - 1)\, P_{\text{out}}\big(y\big|\sqrt{\tfrac{\rho\langle\lambda\rangle_\nu}{\alpha}}z\big)\Big|^2}{\int_{\mathbb{K}} \mathcal{D}_\beta z\, P_{\text{out}}\big(y\big|\sqrt{\tfrac{\rho\langle\lambda\rangle_\nu}{\alpha}}z\big)}\Big]^{-1}\Big). \tag{3}$$

In this equation, we let $\beta \in \{1, 2\}$, with $\mathbb{K} = \mathbb{R}$ if $\beta = 1$ and $\mathbb{K} = \mathbb{C}$ if $\beta = 2$[1]. We introduced the standard Gaussian measure on $\mathbb{K}$ as $\mathcal{D}_\beta z \equiv (2\pi/\beta)^{-\beta/2} \exp(-\beta|z|^2/2)\mathrm{d}z$. Note that in eq. (3), the integrated quantity and the averages $\langle\cdot\rangle_\nu$ depend on α, so that this is actually an implicit equation on $\alpha_{\text{WR,Algo}}$. An important algorithmic question is to characterize the class of polynomial-time algorithms that can achieve weak recovery above the predicted threshold.

The (generalized) *vector approximate message-passing* (G-VAMP) algorithm Rangan et al. (2017); Schniter et al. (2016) has been shown to achieve the threshold in Maillard et al. (2020). Furthermore, it has been conjectured to achieve the optimal polynomial-time recovery for rotationally (unitarily) invariant matrices, i.e. satisfying Hypothesis 1. However this algorithm is rather sensitive to the assumptions of the model, that often do not hold in real data: thus, its applications to real problems are somewhat limited. It is therefore of great interest to investigate more robust and computationally even cheaper algorithms that could achieve similar performances. A natural class of such algorithms are spectral methods. Their output can be used as informative initializations for local gradient-based optimization algorithms, and can induce a jump in the accuracy achieved at a reasonable computational cost. Such techniques have already been applied e.g. in optical systems Valzania et al. (2021). In the context of phase retrieval, the performance of these spectral methods has been rigorously analyzed for Gaussian Lu and Li (2020); Mondelli and Montanari (2019); Luo et al. (2019) and unitary Ma et al. (2021); Dudeja et al. (2020a); Dudeja and Bakhshizadeh (2020) sensing matrices. For Gaussian sensing matrices, Mondelli et al. (2020) also shows how to optimally combine such spectral methods with simple linear estimators, improving even further the performance.

The main goal of the present paper is to design *optimal* spectral methods for the phase retrieval problem in the aforementioned limit, for the very generic class of sensing matrices of Hypothesis 1. Most importantly, in contrast to the previous aforementioned works our approach is completely *automated*, in the sense the spectral methods we derive are (conjectured to be) optimal without the need for optimization over additional parameters. The constructiveness of our approach gives more weight to this optimality conjecture, as we do not restrict to a specific family of spectral methods.

We construct and unify three different approaches for the design of such algorithms, for any sensing matrix satisfying Hypothesis 1: (a) a "pedestrian" optimization of the preprocessing function (the approach of the aforementioned previous works), (b) the linearization of message-passing algorithms, and (c) a *Bethe Hessian* analysis. In short we show that (a) is just a shifted version of (c); (c) automatically uses the optimal preprocessing function in (a); and two eigenvalues of (b) (the dominant one and a peculiar one) have an exact correspondence with the top eigenvalue of (a).

---

[1] The integrals on $\mathbb{C}$ are effectively defined as integrals over $\mathbb{R}^2$.

1.2. Main results

In most previous approaches Lu and Li (2020); Mondelli and Montanari (2019); Luo et al. (2019); Ma et al. (2021), the design of spectral methods for the phase retrieval problem was restricted to consider spectra of matrices of the type:

$$\mathbf{M}(\mathcal{T}) \equiv \frac{1}{n} \sum_{\mu=1}^{m} \mathcal{T}(y_\mu) \overline{\Phi_{\mu i}} \Phi_{\mu j}. \tag{4}$$

These matrices are functions of a (bounded) *preprocessing* function $\mathcal{T}$. It was previously shown for Gaussian i.i.d. matrices $\mathbf{\Phi}$ Lu and Li (2020); Luo et al. (2019) and for random column-unitary matrices $\mathbf{\Phi}$ Ma et al. (2021); Dudeja et al. (2020a) that the optimal transition and reconstruction errors in the class of spectral methods described by eq. (4) is attained by the following functions:

$$\mathcal{T}_{\text{Gaussian}}^\star(y) \equiv \frac{\partial_\omega g_{\text{out}}(y_\mu, 0, \rho)}{1 + \rho \partial_\omega g_{\text{out}}(y_\mu, 0, \rho)}, \qquad \mathcal{T}_{\text{Unitary}}^\star(y) \equiv \frac{\partial_\omega g_{\text{out}}(y_\mu, 0, \rho/\alpha)}{1 + \frac{\rho}{\alpha} \partial_\omega g_{\text{out}}(y_\mu, 0, \rho/\alpha)}. \tag{5}$$

In eq. (5) we introduced the function g_{out}, defined as:

$$g_{\text{out}}(y_\mu, \omega, \sigma^2) \equiv \frac{1}{\sigma^2} \frac{\int_{\mathbb{K}} \mathrm{d}x \, e^{-\frac{\beta}{2\sigma^2}|x-\omega|^2} (x-\omega) \, P_{\text{out}}(y_\mu|x)}{\int_{\mathbb{K}} \mathrm{d}x \, e^{-\frac{\beta}{2\sigma^2}|x-\omega|^2} \, P_{\text{out}}(y_\mu|x)}. \tag{6}$$

In particular, this implies[1]:

$$\partial_\omega g_{\text{out}}(y_\mu, 0, \sigma^2) = -\frac{1}{\sigma^2} + \frac{1}{\sigma^4} \frac{\int_{\mathbb{K}} \mathrm{d}x \, e^{-\frac{\beta}{2\sigma^2}|x|^2} |x|^2 \, P_{\text{out}}(y_\mu|x)}{\int_{\mathbb{K}} \mathrm{d}x \, e^{-\frac{\beta}{2\sigma^2}|x|^2} \, P_{\text{out}}(y_\mu|x)}. \tag{7}$$

Our first result is a conjecture, that generalizes the above two results and gives the optimal spectral method for any phase retrieval problem of the type of eq. (1) which satisfies Hypothesis 1[2]:

Conjecture 2 *For any right-rotationally (or unitarily) invariant matrix $\mathbf{\Phi}$ satisfying Hypothesis 1, the optimal (in terms of both weak-recovery transition and achieved reconstruction error) spectral method belongs to the class of eq. (4), and is attained by:*

$$\mathcal{T}^\star(y) \equiv \frac{\partial_\omega g_{\text{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}{1 + \frac{\rho\langle\lambda\rangle_\nu}{\alpha} \partial_\omega g_{\text{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}.$$

Before detailing further our results, let us explicit two important consequences of Conjecture 2:

- Note that one can always assume the global scaling $\text{Tr}[\mathbf{\Phi}^\dagger\mathbf{\Phi}]/n^2 \to \langle\lambda\rangle_\nu = \alpha$, as it can be absorbed into the channel P_{out}[3]. The optimal spectral method (in terms of weak-recovery threshold and achieved correlation) is then given by $\mathcal{T}^\star(y) = \partial_\omega g_{\text{out}}(y_\mu, 0, \rho)/(1 + \rho \partial_\omega g_{\text{out}}(y_\mu, 0, \rho))$. Remarkably, this optimal function *does not depend on the spectrum of the sensing matrix $\mathbf{\Phi}$, nor*

---

[1]In the complex case, this is the "Wirtinger" derivative $\partial_z f(z) \equiv (\partial_x f(z) - i\partial_y f(z))/2$.

[2]Note that Conjecture 2 is compatible with the results of eq. (5). Indeed, for Gaussian i.i.d. matrices, one has $\langle\lambda\rangle_\nu = \alpha$, while for random column-unitary matrices, $\langle\lambda\rangle_\nu = 1$.

[3]This scaling is chosen to match the one of Gaussian i.i.d. matrices.

on the sampling ratio α. The universality of the method is striking when one compares the optimal performances achievable both information-theoretically and by message-passing algorithms that are both heavily dependent on the spectrum of the sensing matrix and the sampling ratio α, as analyzed in Maillard et al. (2020). Universality also has deep consequences for phase retrieval practitioners: when using a spectral initialization for a non-convex optimization algorithm, she/he does not have to take into account the details of the correlations in $\boldsymbol{\Phi}$ to construct an optimal spectral method. Although our conjecture requires Hypothesis 1, this assumption can possibly be partially loosened as numerically explored in Section 3.

- Importantly, Conjecture 2 claims optimality of our method among *all spectral methods* that one can construct from the data $\boldsymbol{\Phi}$ and the observations $\{y_\mu\}$. As we will see, it turns out that this optimal method belongs to the class of eq. (4), but our derivation is fully constructive and did not assume anything on the form of the spectral method. We believe this is an important improvement of our work with respect to the previous analysis of spectral methods in phase retrieval, which always assumed the method to be in the class of eq. (4). In this sense, our work also confirms the validity of this restriction.

Our second main result, which is linked to Conjecture 2, is the reconciliation of different constructions of spectral methods. In particular, we develop two *automated* approaches to design optimal spectral methods for the phase retrieval problem.

- The first approach arises as a linearization of the Generalized *Vector Approximate Message Passing* (G-VAMP) algorithm Schniter et al. (2016); Rangan et al. (2017). Similar techniques to obtain efficient spectral methods were already investigated in community detection Krzakala et al. (2013), phase retrieval with Gaussian and column-unitary matrices Mondelli and Montanari (2019); Ma et al. (2021), and in the spiked matrix problem Aubin et al. (2020b) to name a few. Here we extend this method to real and complex phase retrieval with a sensing matrix satisfying Hypothesis 1. We call $\mathbf{M}^{(\mathrm{LAMP})}$ (for Linearized-AMP) the corresponding matrix. It is given by:

$$\mathbf{M}^{(\mathrm{LAMP})} \equiv \frac{\rho \langle\lambda\rangle_\nu}{\alpha} \Big(\frac{\alpha}{\langle\lambda\rangle_\nu} \frac{\boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger}{n} - \mathbb{1}_m \Big) \mathrm{Diag}(\partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)). \tag{8}$$

The aforementioned existing works used the *principal eigenvector* $\hat{\mathbf{u}}$ of this matrix to construct the spectral estimator as $\hat{\mathbf{x}}_{\mathrm{LAMP}} \propto \boldsymbol{\Phi}^\dagger \mathrm{Diag}(\partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha))\hat{\mathbf{u}}$. Interestingly, we will see that this estimator achieves the optimal recovery threshold but *sub-optimal performance*. In Section 2.3, we show that the optimal estimator can also be derived from the spectrum of $\mathbf{M}^{(\mathrm{LAMP})}$ but that it is "hidden" inside the bulk of $\mathbf{M}^{(\mathrm{LAMP})}$.

- Our second approach leverages the Thouless-Anderson-Palmer (TAP) formalism of statistical physics Thouless et al. (1977), using the results of Maillard et al. (2019). The TAP approach consists in studying the posterior distribution of eq. (2) by "tilting" it in a controllable manner: this allows to study a modified posterior distribution in which the first and second moments of all x_i are fixed. These moments become then variables of the free energy associated with this modified posterior distribution: this is called the *TAP free energy* in the statistical physics language. When weak recovery of the signal is impossible, this free energy possesses a global minimum in the completely uninformative point in which the estimator is the vector $\mathbf{m} = 0$. On the other hand, when weak recovery is possible, the optimal estimator corresponds to the global minimum of the

TAP free energy with $\mathbf{m} \neq 0$. However ws we will see the point $\mathbf{m} = 0$ always remains a stationary point of the TAP free energy. Moreover, a spectral method used for initializing a non-convex optimization algorithm can be based solely on the observations (i.e. on $\boldsymbol{\Phi}$ and $\{y_\mu\}$), and therefore can not exploit any physical information other than the one present in the uninformative point. When this point is locally stable, we therefore expect all polynomial-time algorithms not to be able to achieve weak recovery. This conjecture has been proven in some cases, e.g. in Mondelli et al. (2020) for Gaussian $\boldsymbol{\Phi}$, in Dudeja et al. (2020b) for unitary $\boldsymbol{\Phi}$, and in Maillard et al. (2020) for a large class of right-rotationally invariant $\boldsymbol{\Phi}$. On the other hand, linear instability of the $\mathbf{m} = 0$ point implies that there should exist a minimum of the TAP free entropy with positive correlation with the signal, and which corresponds to the optimal estimator. With this picture in mind, it is natural to conjecture that the optimal spectral estimator is the dominant unstable direction of the uninformative fixed point, i.e. the smallest eigenvalue of the Hessian. Indeed, this is the most informative direction that one can obtain solely by a local analysis of the $\mathbf{m} = 0$ point. The Hessian of the TAP free energy at the uninformative point is also denoted *Bethe Hessian*. Notably, this Bethe Hessian has been investigated in the context of community detection Saade et al. (2014). This leads to another method, called $\mathbf{M}^{(\mathrm{TAP})}$, which is (up to a shift) the method $\mathbf{M}(\mathcal{T}^\star)$ given in Conjecture 2:

$$\mathbf{M}^{(\mathrm{TAP})} \equiv -\frac{1}{\rho}\mathbb{1}_n + \frac{1}{n}\sum_{\mu=1}^{m} \frac{\partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}{1 + \frac{\rho\langle\lambda\rangle_\nu}{\alpha}\partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)} \overline{\Phi_{\mu i}}\Phi_{\mu j}. \tag{9}$$

Let us now briefly outline the structure of the paper. In Section 2, we unify three different approaches to construct optimal spectral methods for the phase retrieval problem. The first one, based on linearizing the vector approximate message passing is studied in Section 2.1, In Section 2.2 we consider a second approach, based on the Bethe Hessian. Remarkably, as we show in Section 2.3, for any channel distribution and sensing matrix $\boldsymbol{\Phi}$, this method coincides exactly with the third approach, which consists in simply generalizing a spectral method that has been proven to be optimal for Gaussian Luo et al. (2019) and unitary Dudeja et al. (2020b) sensing matrices, see eq. (5). We relate the performance of these different approaches, and show that they allow to conjecture the optimal spectral method, summarized in Conjecture 2. In Section 3, we give numerical evidence to support our claims. We give the performance of the spectral methods we derived in phase retrieval, for noiseless and Poisson-noisy observations. We also show that our methods perform very well even by allowing more structure in the sensing matrix than assumed in Hypothesis 1, by considering for example randomly subsampled DFT, Hadamard or DCT matrices[1].

Notations - Before presenting the technical aspects of our work, we introduce some notations. Recall that $\beta = 1, 2$ for respectively real and complex variables. $\mathcal{U}_\beta(n)$ denotes the orthogonal (or unitary) group. For $x, y \in \mathbb{K}$, we define $x \cdot y \equiv xy$ if $\mathbb{K} = \mathbb{R}$ and $x \cdot y \equiv \mathrm{Re}[\overline{x}y]$ if $\mathbb{K} = \mathbb{C}$.

2. Spectral methods, message-passing algorithms and TAP approach

2.1. Linearized vector approximate message passing

In this section, we describe the vector approximate message-passing algorithm for the phase retrieval problem with sensing matrices satisfying Hypothesis 1. The algorithm was first stated in Rangan

---

[1]Note that the universality of linearized approximate message passing algorithms for a Gaussian prior and different ensembles of column-orthogonal matrices was analyzed recently in Dudeja and Bakhshizadeh (2020).

Algorithm 1: Generalized Vector Approximate Message Passing

Data: The sensing matrix $\mathbf{\Phi}/\sqrt{n} = \mathbf{USV}^\dagger$, the outputs $\{y_\mu\}_{\mu=1}^m$, a number of iterations T.

Result: An estimate $\hat{\mathbf{x}}$ of $\mathbf{X}^\star$.

Randomly initialize all variables.

for $t = 1, \cdots, T$ **do**

    (Denoising x)   *(Denoising $z \equiv \frac{1}{\sqrt{n}}\mathbf{\Phi}x$)*

    $\hat{\mathbf{x}}_1^t = g_{x1}(\mathbf{T}_1^t, \gamma_1^t)$ $\hat{\mathbf{z}}_1^t = g_{z1}(\mathbf{R}_1^t, \tau_1^t)$

    $v_1^t = \frac{1}{\beta n}\sum_{i=1}^n \partial_{T_i} g_{x1}(\mathbf{T}_1^t, \gamma_1^t)$ $c_1^t = \frac{1}{\beta m}\sum_{\mu=1}^m \partial_{R_\mu} g_{z1}(\mathbf{R}_1^t, \tau_1^t)$

    $\mathbf{T}_2^t = \frac{1}{v_1^t}\hat{\mathbf{x}}_1^t - \mathbf{T}_1^t$ $\mathbf{R}_2^t = \frac{1}{c_1^t}\hat{z}_1^t - \mathbf{R}_1^t$

    $\gamma_2^t = \frac{1}{v_1^t} - \gamma_1^t$ $\tau_2^t = \frac{1}{c_1^t} - \tau_1^t$

    (Estimation of x)   *(Estimation of z)*

    $\hat{\mathbf{x}}_2^t = g_{x2}(\mathbf{T}_2^t, \mathbf{R}_2^t, \gamma_2^t, \tau_2^t)$ $\hat{\mathbf{z}}_2^t = g_{z2}(\mathbf{T}_2^t, \mathbf{R}_2^t, \gamma_2^t, \tau_2^t)$

    $v_2^t = \left\langle \frac{1}{\tau_2^t\lambda + \gamma_2^t}\right\rangle_\nu$ $c_2^t = \frac{1}{\alpha}\left\langle \frac{\lambda}{\tau_2^t\lambda + \gamma_2^t}\right\rangle_\nu$

    $\mathbf{T}_1^{t+1} = \frac{1}{v_2^t}\hat{\mathbf{x}}_2^t - \mathbf{T}_2^t$ $\mathbf{R}_1^{t+1} = \frac{1}{c_2^t}\hat{\mathbf{z}}_2^t - \mathbf{R}_2^t$

    $\gamma_1^{t+1} = \frac{1}{v_2^t} - \gamma_2^t$ $\tau_1^{t+1} = \frac{1}{c_2^t} - \tau_2^t$

end

return $\hat{\mathbf{x}}_1^T$

---

et al. (2017) for the compressed sensing problem, and later generalized in Schniter et al. (2016) to any GLM described by eq. (1). It makes use of the SVD decomposition of $\mathbf{\Phi}$, that we write as $\mathbf{\Phi}/\sqrt{n} = \mathbf{USV}^\dagger$. The full iterations of the algorithm are detailed in Algorithm 1. We used some auxiliary functions, defined below:

$$\begin{cases} g_{x1}(\mathbf{T}, \gamma)_i \equiv \mathbb{E}_{P_0(\gamma, -T_i)}[x], & g_{x2}(\mathbf{T}, \mathbf{R}, \gamma, \tau) \equiv \frac{\mathbf{T}}{\gamma} + \mathbf{VS}^\mathsf{T}\left(\frac{\gamma}{\tau} + \mathbf{SS}^\mathsf{T}\right)^{-1}\left(\frac{\mathbf{U}^\dagger \mathbf{R}}{\tau} - \frac{\mathbf{SV}^\dagger \mathbf{T}}{\gamma}\right), \\ g_{z1}(\mathbf{R}, \tau)_\mu \equiv \mathbb{E}_{P_{\text{out}}\left(y_\mu, \frac{R_\mu}{\tau}, \frac{1}{\tau}\right)}[z], & g_{z2}(\mathbf{T}, \mathbf{R}, \gamma, \tau) \equiv \mathbf{USV}^\dagger g_{x2}(\mathbf{T}, \mathbf{R}, \gamma, \tau). \end{cases} \tag{10}$$

We denoted $P_0(\gamma, \lambda)$ the probability distribution with density proportional to $P_0(x)e^{-\frac{\beta\gamma}{2}|x|^2 - \beta\lambda_i \cdot x}$, and $P_{\text{out}}(y_\mu, \omega_\mu, b)$ the one with density proportional to $P_{\text{out}}(y_\mu|z)e^{-\frac{\beta|z - \omega_\mu|^2}{2b}}$.

2.1.1. THE TRIVIAL FIXED POINT

In Algorithm 1, one can use the Bayes-optimality hypothesis to derive the following relation (see for instance eq. (107) of Kabashima et al. (2016)):

$$\frac{1}{m}\sum_{\mu=1}^m \mathbb{E}_{P_{\text{out}}\left(y_\mu, \frac{(R_1^t)_\mu}{\tau_1^t}, \frac{1}{\tau_1^t}\right)}\left[\left|z - \frac{(R_1^t)_\mu}{\tau_1^t}\right|^2\right] = \frac{1}{\tau_1^t}. \tag{11}$$

Informally, eq. (11) expresses that the estimated variance of $\mathbf{z}$, defined as τ_1^t, is equal to the mean square difference between $\mathbf{z}$ and the estimation of $\mathbf{z}$ (being $\mathbf{R}_1^t/\tau_1^t$) under the estimated posterior. Recall that we assumed that P_0 is symmetric with $\rho \equiv \mathbb{E}_{P_0}[|x|^2]$ and that $P_{\text{out}}(y|z)$ only depends on $|z|^2$. Using eq. (11) along with this hypothesis, it is easy to see that Algorithm 1 admits the following

fixed point, that we call "trivial" as it is completely uninformative:

$$
\begin{cases}
\gamma_1 = 0, & \gamma_2 = \rho^{-1}, & v_1 = \rho, & v_2 = \rho \\
\hat{\mathbf{x}}_1 = \mathbf{T}_1 = 0, & \hat{\mathbf{x}}_2 = \mathbf{T}_2 = 0, & \tau_1 = \alpha/(\rho\langle\lambda\rangle_\nu), & \tau_2 = 0 \\
c_1 = \rho\langle\lambda\rangle_\nu/\alpha, & c_2 = \rho\langle\lambda\rangle_\nu/\alpha, & \hat{\mathbf{z}}_1 = \mathbf{R}_1 = 0, & \hat{\mathbf{z}}_2 = \mathbf{R}_2 = 0.
\end{cases}
\tag{12}
$$

2.1.2. LINEARIZATION AROUND THE FIXED POINT

We can now linearize Algorithm 1 around the fixed point given by eq. (12). We begin by showing that the first order variations of all the variances and inverse variances parameters are negligible, and we detail this derivation in Appendix D.1. This will greatly simplify our linearization around the trivial fixed point, as we can focus solely on the vector parameters. For clarity, we restrict here to the real case $\beta = 1$, while the derivation in the complex case is provided in Appendix A. We write the linearization of Algorithm 1 as (all derivatives are taken at the fixed point of eq. (12)):

$$
\begin{aligned}
\delta\hat{\mathbf{x}}_1^t &= \nabla_{\mathbf{T}} g_{x1} \delta\mathbf{T}_1^t, & \delta\hat{\mathbf{z}}_1^t &= \nabla_{\mathbf{R}} g_{z1} \delta\mathbf{R}_1^t, & \delta\mathbf{T}_2^t &= \tfrac{1}{\rho}\delta\hat{\mathbf{x}}_1^t - \delta\mathbf{T}_1^t, \\
\delta\hat{\mathbf{x}}_2^t &= \nabla_{\mathbf{T}} g_{x2} \delta\mathbf{T}_2^t + \nabla_{\mathbf{R}} g_{x2} \delta\mathbf{R}_2^t, & & & \delta\hat{\mathbf{z}}_2^t &= \nabla_{\mathbf{T}} g_{z2} \delta\mathbf{T}_2^t + \nabla_{\mathbf{R}} g_{z2} \delta\mathbf{R}_2^t, \\
\delta\mathbf{R}_2^t &= \tfrac{\alpha}{\rho\langle\lambda\rangle_\nu}\delta\hat{\mathbf{z}}_1^t - \delta\mathbf{R}_1^t, & & & \delta\mathbf{T}_1^{t+1} &= \tfrac{1}{\rho}\delta\hat{\mathbf{x}}_2^t - \delta\mathbf{T}_2^t, & \delta\mathbf{R}_1^{t+1} &= \tfrac{\alpha}{\rho\langle\lambda\rangle_\nu}\delta\hat{\mathbf{z}}_2^t - \delta\mathbf{R}_2^t.
\end{aligned}
\tag{13}
$$

The derivatives of the auxiliary functions of eq. (10) at the trivial fixed point of eq. (12) are:

$$
\begin{cases}
\partial_{T_j}[(g_{x1})_i] = \rho\,\delta_{ij}, & \partial_{T_j}[(g_{x2})_i] = \rho\,\delta_{ij}, & \partial_{R_\nu}[(g_{z1})_\mu] = \delta_{\mu\nu}\mathbb{E}_{P_{\text{out}}\left(y_\mu, 0, \frac{\rho\langle\lambda\rangle_\nu}{\alpha}\right)}[z^2], \\
\partial_{R_\mu}[(g_{x2})_i] = \rho\frac{(\boldsymbol{\Phi}^\dagger)_{i\mu}}{\sqrt{n}}, & \partial_{T_i}[(g_{z2})_\mu] = \rho\frac{\boldsymbol{\Phi}_{\mu i}}{\sqrt{n}}, & \partial_{R_\nu}[(g_{z2})_\mu] = \rho\frac{(\boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger)_{\mu\nu}}{n}.
\end{cases}
\tag{14}
$$

Plugging eq. (14) in eq. (13) yields, with $v(y_\mu) \equiv \mathbb{E}_{P_{\text{out}}\left(y_\mu, 0, \frac{\rho\langle\lambda\rangle_\nu}{\alpha}\right)}[z^2]$:

$$
\begin{cases}
\delta\hat{\mathbf{x}}_1^t = \rho\delta\mathbf{T}_1^t, & \delta\hat{\mathbf{z}}_1^t = \text{Diag}(\{v(y_\mu)\})\delta\mathbf{R}_1^t, & \delta\mathbf{T}_2^t = \tfrac{1}{\rho}\delta\hat{\mathbf{x}}_1^t - \delta\mathbf{T}_1^t, \\
\delta\mathbf{R}_2^t = \tfrac{\alpha}{\rho\langle\lambda\rangle_\nu}\delta\hat{\mathbf{z}}_1^t - \delta\mathbf{R}_1^t, & \delta\hat{\mathbf{x}}_2^t = \rho\delta\mathbf{T}_2^t + \rho\frac{\boldsymbol{\Phi}^\dagger}{\sqrt{n}}\delta\mathbf{R}_2^t, & \delta\hat{\mathbf{z}}_2^t = \rho\frac{\boldsymbol{\Phi}}{\sqrt{n}}\delta\mathbf{T}_2^t + \rho\frac{\boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger}{n}\delta\mathbf{R}_2^t, \\
\delta\mathbf{T}_1^{t+1} = \tfrac{1}{\rho}\delta\hat{\mathbf{x}}_2^t - \delta\mathbf{T}_2^t, & \delta\mathbf{R}_1^{t+1} = \tfrac{\alpha}{\rho\langle\lambda\rangle_\nu}\delta\hat{\mathbf{z}}_2^t - \delta\mathbf{R}_2^t.
\end{cases}
\tag{15}
$$

These equations imply $\delta\mathbf{T}_2^t = 0$. The equations can then simply be closed on $\delta\mathbf{R}_1^t$:

$$
\delta\mathbf{R}_1^{t+1} = \left(\frac{\alpha}{\langle\lambda\rangle_\nu}\frac{\boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger}{n} - \mathbb{1}_m\right)\left[\frac{\alpha}{\rho\langle\lambda\rangle_\nu}\text{Diag}(\{v(y_\mu)\}) - \mathbb{1}_m\right]\delta\mathbf{R}_1^t.
\tag{16}
$$

As shown in Appendix A, we obtain the same equation in the complex case. Interestingly, $v(y_\mu)$ can be linked to the function $\partial_\omega g_{\text{out}}$, simply by eq. (7): $\partial_\omega g_{\text{out}}(y_\mu, 0, \sigma^2) = -\sigma^{-2} + \sigma^{-4} v(y_\mu)$.

2.1.3. THE LAMP SPECTRAL METHOD

The Linearized-AMP (LAMP) spectral method is based on eq. (16), and consists in taking the largest eigenvalue and corresponding eigenvector of the $m \times m$ matrix:

$$
\mathbf{M}^{(\text{LAMP})} \equiv \frac{\rho\langle\lambda\rangle_\nu}{\alpha}\left(\frac{\alpha}{\langle\lambda\rangle_\nu}\frac{\boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger}{n} - \mathbb{1}_m\right)\text{Diag}(\partial_\omega g_{\text{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)).
$$

Note that $\mathbf{M}^{(\mathrm{LAMP})}$ is not a Hermitian matrix, so "largest" eigenvalue means here eigenvalue of largest real part. If $\hat{\mathbf{u}}$ is the eigenvector of $\mathbf{M}^{(\mathrm{LAMP})}$ associated to this largest eigenvalue, then one can construct a corresponding estimate $\hat{\mathbf{x}}$ using the relations of eq. (15), as:

$$\hat{\mathbf{x}} \equiv \frac{\mathbf{\Phi}^{\dagger}\left[\frac{\alpha}{\rho\langle\lambda\rangle_{\nu}}\mathrm{Diag}(\{v(y_{\mu})\}) - \mathbb{1}_m\right]\hat{\mathbf{u}}}{\left\|\mathbf{\Phi}^{\dagger}\left[\frac{\alpha}{\rho\langle\lambda\rangle_{\nu}}\mathrm{Diag}(\{v(y_{\mu})\}) - \mathbb{1}_m\right]\hat{\mathbf{u}}\right\|}\sqrt{n\rho}. \tag{17}$$

Surprisingly, and as we will see in more details in Sections 2.3 and 3, this spectral method achieves the optimal weak-recovery threshold but only sub-optimal performance compared to $\mathbf{M}(\mathcal{T}^{\star})$. There is, however, a way to recover the optimal performance from $\mathbf{M}^{(\mathrm{LAMP})}$ by considering an eigenvalue equal to 1 (and thus "hidden" inside the bulk) which appears when weak recovery is possible.

2.2. The Bethe Hessian approach

2.2.1. THE TAP FREE ENTROPY

In this section we detail a statistical-physics based constructive approach to derive the optimal spectral method for the phase retrieval problem. We consider the so-called *Thouless-Anderson-Palmer* (TAP) Thouless et al. (1977) free entropy of the system, that we denote $f_{\mathrm{TAP}}(\mathbf{Y}, \mathbf{\Phi}, \mathbf{m}, \sigma)$. The idea of this approach is to constrain the posterior probability of eq. (2) to satisfy the first and second moment constraints $\langle\mathbf{x}\rangle = \mathbf{m}$, $\langle\|\mathbf{x} - \langle\mathbf{x}\rangle\|^2\rangle = n\sigma^2$, and to study the free entropy of this "tilted" probability distribution. This provides a dual perspective on the posterior distribution (also called Gibbs measure), by considering the landscape of f_{TAP}. For clarity we will drop the dependency of f_{TAP} on $\mathbf{Y}, \mathbf{\Phi}$. Of particular interest are the maxima of this free entropy, corresponding to *pure states* in the statistical physics language. It is known that the fixed points of the optimal algorithm for this problem, i.e. generalized *vector approximate message-passing* (see Section 2.1), are in exact correspondence with the local maxima of the TAP free entropy. This is shown in Maillard et al. (2019), in which the TAP free entropy for rotationally-invariant generalized linear models is also derived[1]. By maximizing as well on the variance parameter σ^2, it yields, up to $\mathcal{O}_n(1)$ terms:

$$f_{\mathrm{TAP}}(\mathbf{m}) = \sup_{\substack{\sigma\geq 0 \\ }}\sup_{\substack{\mathbf{g}\in\mathbb{K}^m \\ r\geq 0}}\underset{\substack{\boldsymbol{\omega}\in\mathbb{K}^m \\ b\geq 0}}{\mathrm{extr}}\underset{\substack{\boldsymbol{\lambda}\in\mathbb{K}^n \\ \gamma\geq 0}}{\mathrm{extr}}\left[\frac{\beta}{n}\sum_{i=1}^{n}\lambda_i\cdot m_i + \frac{\beta\gamma}{2n}\left(n\sigma^2 + \sum_{i=1}^{n}|m_i|^2\right) - \frac{\beta}{n}\sum_{\mu=1}^{m}\omega_\mu\cdot g_\mu\right.$$

$$- \frac{\beta b}{2n}\left(\sum_{\mu=1}^{m}|g_\mu|^2 - \alpha nr\right) + \frac{1}{n}\sum_{i=1}^{n}\ln\int_{\mathbb{K}}P_0(\mathrm{d}x)e^{-\frac{\beta\gamma}{2}|x|^2 - \beta\lambda_i\cdot x} \tag{18}$$

$$\left.+ \frac{\alpha}{m}\sum_{\mu=1}^{m}\ln\int_{\mathbb{K}}\frac{\mathrm{d}h}{\left(\frac{2\pi b}{\beta}\right)^{\beta/2}}P_{\mathrm{out}}(y_\mu|h)e^{-\frac{\beta|h-\omega_\mu|^2}{2b}} + \frac{\beta}{n}\sum_{i=1}^{n}\sum_{\mu=1}^{m}g_\mu\cdot\left(\frac{\Phi_{\mu i}}{\sqrt{n}}m_i\right) + \beta F(\sigma^2, r)\right].$$

Here the notation $\mathrm{extr}_\gamma f(\gamma)$ means that one should solve the corresponding saddle-point equation $\partial_\gamma f(\gamma) = 0$, and the function F is defined as:

$$F(x, y) \equiv \inf_{\zeta_x, \zeta_y > 0}\left[\frac{\zeta_x x}{2} + \frac{\alpha\zeta_y y}{2} - \frac{\alpha - 1}{2}\ln\zeta_y - \frac{1}{2}\langle\ln(\zeta_x\zeta_y + \lambda)\rangle_\nu\right] - \frac{1}{2}\ln x - \frac{\alpha}{2}\ln y - \frac{1 + \alpha}{2}.$$

---

[1] The results of Maillard et al. (2019) stand in the real case, but can be straightforwardly generalized to complex variables.

One can write the saddle-point equations associated to eq. (18), called the *TAP equations*:

$$
\begin{aligned}
m_i &= \mathbb{E}_{P_0(\gamma,\lambda_i)}[x], & \sigma^2 &= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{P_0(\gamma,\lambda_i)}[|x - m_i|^2], \\
g_\mu &= g_{\text{out}}(y_\mu,\omega_\mu,b), & r &= \frac{1}{m}\sum_{\mu=1}^{m}\left\{|g_\mu|^2 + \tfrac{1}{b} - \mathbb{E}_{P_{\text{out}}(y_\mu,\omega_\mu,b)}\left[\left|\tfrac{h-\omega_\mu}{b}\right|^2\right]\right\}, \\
\omega_\mu + bg_\mu &= \sum_{i=1}^{n}\frac{\Phi_{\mu i}}{\sqrt{n}}m_i, & b &= -\tfrac{2}{\alpha}\partial_r F(\sigma^2,r), \quad \gamma = -2\partial_{\sigma^2} F(\sigma^2,r).
\end{aligned}
\tag{19}
$$

2.2.2. THE TRIVIAL FIXED POINT

It is easy to see that the TAP equations (19) admits a trivial fixed point at $\mathbf{m} = 0$ (corresponding to a local maximum of f_{TAP}). At this point, the parameters are $\sigma^2 = \rho$, $\mathbf{g} = \boldsymbol{\omega} = \boldsymbol{\lambda} = 0$, $\gamma = r = 0$, $b = \rho\langle\lambda\rangle_\nu/\alpha$. This uses in particular a known consequence of the Bayes-optimality, that relates the variance parameter b to the mean squared difference between the true $\boldsymbol{\Phi}\mathbf{X}^\star$ and its estimate, see Kabashima et al. (2016)[1]: $\frac{1}{m}\sum_{\mu=1}^{m}\mathbb{E}_{P_{\text{out}}(y_\mu,\omega_\mu,b)}[|h - \omega_\mu|^2] = b$. The derivation of the fixed point also uses the behavior of $F(\sigma^2,r)$ at small r, computed in Appendix D.2:

$$
F(\sigma^2,r) = -\frac{\langle\lambda\rangle_\nu r\sigma^2}{2} + \frac{\sigma^4 r^2}{4\alpha}[\alpha\langle\lambda^2\rangle_\nu - (1+\alpha)\langle\lambda\rangle_\nu^2] + \sigma^6 r^3 G(r\sigma^2),
\tag{20}
$$

with $G(x)$ a continuous bounded function in $x = 0$.

2.2.3. THE SPECTRAL METHOD

A natural way to design a spectral method for this inference problem is to consider the Hessian of $-f_{\text{TAP}}$ at this trivial fixed point, as we expect a descending informative direction to appear in its spectrum at the weak recovery threshold. As we show in Appendix B, this procedure leads to consider the $n \times n$ matrix:

$$
\mathbf{M}^{(\text{TAP})} \equiv -\frac{n}{\beta}\nabla^2 f_{\text{TAP}}(0) = -\frac{1}{\rho}\mathbb{1}_n + \frac{1}{n}\sum_{\mu=1}^{m}\frac{\partial_\omega g_{\text{out}}(y_\mu,0,\rho\langle\lambda\rangle_\nu/\alpha)}{1 + \frac{\rho\langle\lambda\rangle_\nu}{\alpha}\partial_\omega g_{\text{out}}(y_\mu,0,\rho\langle\lambda\rangle_\nu/\alpha)}\overline{\Phi_{\mu i}}\Phi_{\mu j}.
$$

2.3. Unification of the approaches

We now detail our main claims and results concerning the spectral methods we just derived.

The optimal spectral method and the Bethe Hessian

Very importantly, as opposed to previous approaches, our derivation is *constructive*: we start from the fully-explicit expression of the TAP free entropy given in eq. (18) and simply compute its Hessian at the trivial fixed point. From the statistical physics literature (as we detailed in Section 1.2), we expect that the optimal spectral method will be given by the largest eigenvalue (and associated eigenvector) of this Hessian. The result of our computation of this Hessian was given in eq. (9). Importantly, this implies that the optimal spectral method that can be built from the data $\boldsymbol{\Phi}$ and the observations $\{y_\mu\}$ belongs to the class of methods given by eq. (4). Our conjecture therefore also gives weight to many previous analysis of spectral methods for phase retrieval, which only studied spectral methods of the type of eq. (4) Lu and Li (2020); Mondelli and Montanari (2019); Luo et al. (2019); Ma et al. (2021).

---

[1]This relation is equivalent to eq. (11), which states it for AMP iterations rather than the solutions of the TAP equations.

Relating linearized-AMP and the Bethe Hessian

Our derivation of $\mathbf{M}^{(\mathrm{LAMP})}$ is *constructive* as well, and in this sense fundamentally differs from the L-AMP algorithms designed in Ma et al. (2021) to assess the performance of other spectral methods. We start by a proposition, proven in Appendix D.3, which relates the eigenpairs of the two methods.

Proposition 3 *Without loss of generality, we assume $\langle\lambda\rangle_\nu = \alpha$. Let $z_\mu \equiv \partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)$, and $(\lambda_{\mathrm{LAMP}}, \boldsymbol{v})$ be an eigenpair of $\boldsymbol{M}^{(\mathrm{LAMP})}$. Assume that $\lambda_{\mathrm{LAMP}} + \rho z_\mu \neq 0$ for all $\mu = 1, \cdots, m$. Then $\boldsymbol{\Phi}^\dagger \mathrm{Diag}(z_\mu)\boldsymbol{v} \neq 0$, and we let $\hat{\boldsymbol{x}} \propto \boldsymbol{\Phi}^\dagger \mathrm{Diag}(z_\mu)\boldsymbol{v}$ with $\|\hat{\boldsymbol{x}}\|^2 = n$. Moreover:*

$$\Big\{ \frac{1}{m} \sum_{\mu=1}^m \frac{\rho z_\mu}{\lambda_{\mathrm{LAMP}} + \rho z_\mu} \boldsymbol{\Phi}_\mu \boldsymbol{\Phi}_\mu^\dagger \Big\} \hat{\boldsymbol{x}} = \hat{\boldsymbol{x}}.$$

Conversely, let $\boldsymbol{x}$ be an eigenvector of $\boldsymbol{M}^{(\mathrm{TAP})}$ with norm $\|\boldsymbol{x}\|^2 = n$, with associated eigenvalue λ_{TAP}. We define $\boldsymbol{u} \equiv \mathrm{Diag}[(1 + \rho z_\mu)^{-1}]\boldsymbol{\Phi}\boldsymbol{x}/\sqrt{n}$. Then one has:

$$\boldsymbol{M}^{(\mathrm{LAMP})}\boldsymbol{u} = \boldsymbol{u} + \rho\lambda_{\mathrm{TAP}}\mathrm{Diag}(1 + \rho\partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho))\boldsymbol{u}.$$

Moreover, if $\lambda_{\mathrm{TAP}} = 0$, eq. (17) applied to $\boldsymbol{u}$ yields the same performance as the TAP estimator.

By considering $\lambda_{\mathrm{LAMP}} = 1$ and $\lambda_{\mathrm{TAP}} = 0$, one immediately deduces two important consequences of Proposition 3 and the definitions of the methods (cf. eqs. (8),(9)):

- The appearance of an unstable direction, in the spectrum of $\mathbf{M}^{(\mathrm{TAP})}$ (i.e. a positive eigenvalue) and of $\mathbf{M}^{(\mathrm{LAMP})}$ (i.e. an eigenvalue with real part greater than 1), occurs at a common threshold (i.e. the *weak-recovery* threshold, given by eq. (3)).

- An eigenvalue 0 appears in the spectrum of $\mathbf{M}^{(\mathrm{TAP})}$ if and only if an eigenvalue 1 appears in the spectrum of $\mathbf{M}^{(\mathrm{LAMP})}$. These two eigenvalues therefore correspond to *marginal stability* of the linear dynamics. Moreover, the two estimators associated to these eigenvalues are identical, i.e. $\mathbf{M}^{(\mathrm{LAMP})}$ *contains the optimal estimator*. Importantly, this estimator is different from the largest eigenvector of $\mathbf{M}^{(\mathrm{LAMP})}$, which reaches only suboptimal performance as we will see in Section 3.

3. Numerical experiments and perspectives

In this section, we numerically assess our predictions and compare the performance of the spectral methods on various problems. In Section 3.1, we consider the recovery of a randomly generated signal with different right-rotationally invariant sensing matrix ensembles. In Sec. 3.2, we illustrate the transition phenomena in the spectra of $\mathbf{M}^{(\mathrm{TAP})}$ and $\mathbf{M}^{(\mathrm{LAMP})}$, which raise interesting random matrix theory questions. Finally, in Section 3.3, we validate our predictions for the recovery of a natural image with various matrix ensembles. We numerically verify that all our conclusions derived for random signals still hold in this setting. The numerical code used to generate all figures is available in the supplementary material.

Another spectral method – In the figures, we sometimes consider another spectral method, called $\mathbf{M}^{(\mathrm{MM})}$. It is obtained by naively considering the preprocessing function of Mondelli and Montanari (2019), which was shown to achieve the optimal transition for Gaussian sensing matrices. More precisely, we have (assuming $\rho = 1$ and $\langle\lambda\rangle_\nu = \alpha$): $\mathcal{T}_{\mathrm{MM}}(y) \equiv \partial_\omega g_{\mathrm{out}}(y, 0, 1)/\Big[\sqrt{\frac{2\alpha}{\beta}} + \partial_\omega g_{\mathrm{out}}(y, 0, 1)\Big]$. In particular note that at $\alpha = \beta/2$, we have $\mathcal{T}_{\mathrm{MM}} = \mathcal{T}^\star$, so that $\mathcal{T}_{\mathrm{MM}}$ indeed achieves the optimal weak-recovery transition for Gaussian matrices, for which $\alpha_{\mathrm{WR,Algo}} = \beta/2$.
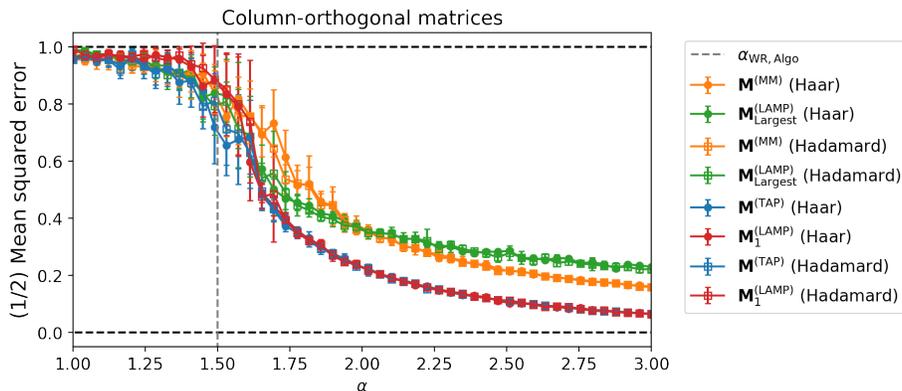
Figure 1: Mean squared error achieved by our spectral methods and a naive version of the spectral method of Mondelli and Montanari (2019) for real column-orthogonal sensing matrices and a noiseless channel. We give the performance on uniformly sampled column-orthogonal matrices as well as randomly subsampled Hadamard matrices. The simulations were done using $m = 8192$, and the error bars are taken over 10 instances.

3.1. Performance of the spectral methods

We show the performance of the spectral methods to recover a random signal in three different cases, that we briefly describe:

- In Fig. 1, we consider noiseless real phase retrieval (i.e. sign retrieval), with uniformly sampled column-unitary sensing matrices. We also show that our conclusions transfer to randomly subsampled Hadamard matrices, validating the conclusions of Dudeja and Bakhshizadeh (2020).

- In Fig. 2(a)subfigure we consider noiseless real phase retrieval when the sensing matrix is a product of two Gaussian i.i.d. matrices. This setup can for instance be interpreted as Gaussian phase retrieval in which the signal is drawn from a known generative prior, similarly to the analysis of Aubin et al. (2020a). Importantly, it is not covered by any previous analysis of the spectral methods, emphasizing the generality of the framework of Hypothesis 1.

- In Fig. 2(b)subfigure, we compare our results in noiseless and noisy settings. More precisely, we consider complex phase retrieval with a Gaussian sensing matrix, and either a noiseless channel or a Poisson observation channel with intensity $\Lambda > 0$:

$$P_{\text{out}}(y|z) = e^{-\Lambda|z|^2} \sum_{k=0}^{\infty} \delta(y-k) \frac{\Lambda^k |z|^{2k}}{k!}.$$

This latter channel is particularly relevant for optical applications, in which the detector can be modeled as being affected by a Poisson noise. In both cases, we find that all our conclusions on the optimality of the $\mathbf{M}^{(\text{TAP})}$, and on the link between $\mathbf{M}^{(\text{LAMP})}$ and $\mathbf{M}^{(\text{TAP})}$, still hold.
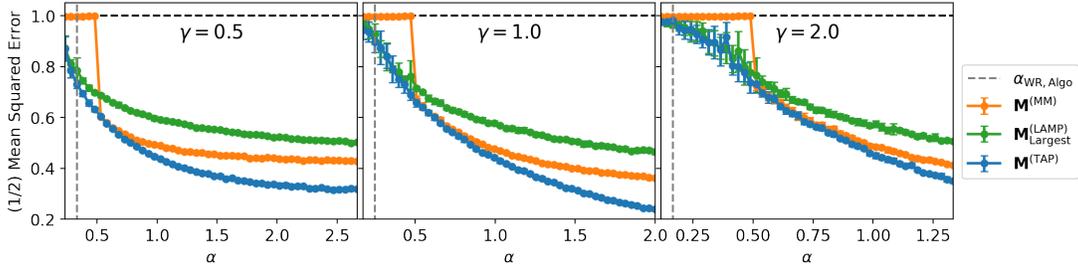
(*a*) Product of two real i.i.d. Gaussian sensing matrices with a size ratio $\gamma \in \{0.5, 1.0, 2.0\}$. The simulations were done using $m = 10000$, and error bars are taken over 10 instances.
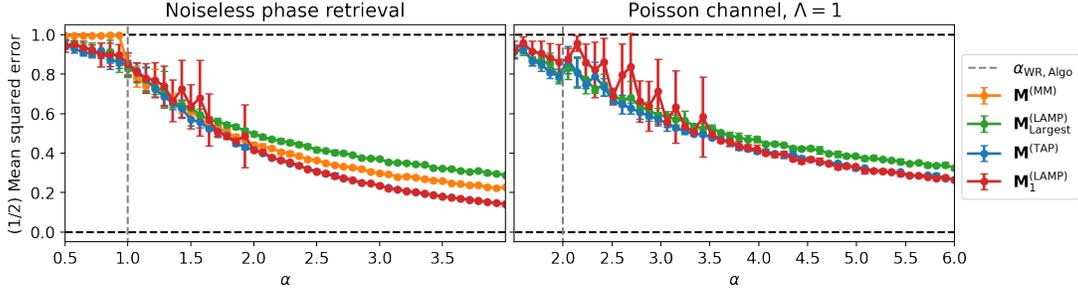
(*b*) Complex Gaussian matrix, in noiseless phase retrieval and in Poisson-noise phase retrieval with $\Lambda = 1$. The simulations were done using $m = 10000$ (noiseless case), 12000 (Poisson case), and the error bars are taken over 10 (noiseless case), 5 (Poisson case) instances.

Figure 2: Mean squared error achieved by the different spectral methods in two different settings.

3.2. Transition phenomena in the spectra

We illustrate the weak-recovery transition in the spectra of the different methods. Precisely, we confirm the following claims of Section 2.3:

- Both $\mathbf{M}^{(\mathrm{LAMP})}$ and $\mathbf{M}^{(\mathrm{TAP})}$ have a largest eigenvalue (in real part) that detaches from the bulk for $\alpha > \alpha_{\mathrm{WR,Algo}}$, given by eq. (3).

- In the regime in which weak-recovery is possible, the largest eigenvalue of $\mathbf{M}^{(\mathrm{TAP})}$ approaches 0 as $n \to \infty$. The associated eigenvector achieves optimal correlation with the signal (among spectral methods) as $n \to \infty$.

- $\mathbf{M}^{(\mathrm{LAMP})}$ gives *two* estimators that are positively correlated with the signal for $\alpha > \alpha_{\mathrm{WR,Algo}}$. The first one corresponds to its largest eigenvalue in real part, and achieves worse correlation than the largest eigenvector of $\mathbf{M}^{(\mathrm{TAP})}$. The second one corresponds to an eigenvalue inside the bulk (but isolated from the other eigenvalues) that approaches 1 as $n \to \infty$, and achieves the same optimal performance as the estimator given by $\mathbf{M}^{(\mathrm{TAP})}$.

We verify these claims for different values of α, below and above the weak-recovery threshold, in complex Gaussian phase retrieval with Poisson-noise, in Fig. 3. We complete this analysis in Appendix C, by considering noiseless phase retrieval and more values of α in Fig. 6, and product of complex Gaussian matrices and structured signals in Fig. 7.
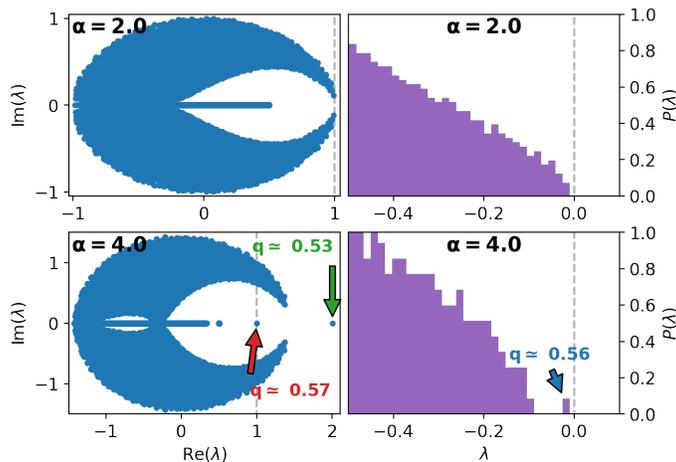
Figure 3: Transition in the spectra of $\mathbf{M}^{(\mathrm{LAMP})}$ (left) and $\mathbf{M}^{(\mathrm{TAP})}$ (right) for a complex Gaussian $\mathbf{\Phi}$ and a Poisson channel with $\Lambda = 1$. For $\alpha > \alpha_{\mathrm{WR,Algo}} = 2$, we indicate the approximate overlap q corresponding to the the relevant eigenvalues.

Remark – In the shown figures there is a very small discrepancy between the overlaps achieved by the principal eigenvector of $\mathbf{M}^{(\mathrm{TAP})}$ and the eigenvector of $\mathbf{M}^{(\mathrm{LAMP})}$ with eigenvalue 1. This is due to the fact that the subplots of Fig. 3 (and similarly for Fig. 6) are generated with different instances of the matrix $\mathbf{\Phi}$ and signal $\mathbf{X}^\star$.

On the performance of the spectral methods – When weak recovery is possible the largest eigenvalue of $\mathbf{M}^{(\mathrm{TAP})}$ concentrates on 0 as we noticed. However, the spectrum of $\mathbf{M}^{(\mathrm{TAP})}$ also contains many very large negative eigenvalues. In practice, we use an inverse iteration method to quickly estimate the associated eigenvector. We use a similar approach for $\mathbf{M}^{(\mathrm{LAMP})}$, using inverse iterations to estimate the eigenvector with eigenvalue 1, and usual power iterations for the largest eigenvalue.

3.3. Real image reconstruction

As a final analysis, we numerically investigate our predictions for the reconstruction of a natural image. For comparability, we consider the image of *The Birth of Venus* already used in Mondelli and Montanari (2019); Ma et al. (2021). Although this signal is not i.i.d., we will see that all our previous conclusions, numerically investigated in Sections 3.1, 3.2, transfer to this case. We consider a noiseless phase retrieval channel and different sensing matrices $\mathbf{\Phi}$: multiple ensembles of column-unitary matrices (which partly reproduces the analysis of Ma et al. (2021)) and a product of two complex Gaussian matrices with aspect ratio $\gamma = 1$. In particular we consider partial DFT matrices, introduced in Ma et al. (2014, 2021), which are an ensemble of column-unitary matrices obtained from the usual DFT matrices. Namely, there are defined for $m \geq n$ as $\mathbf{\Phi}/\sqrt{n} = \mathbf{FSP}$, with $\mathbf{F} \in \mathbb{C}^{m \times m}$ a DFT matrix, $\mathbf{S} \in \mathbb{R}^{m \times n}$ containing n columns (randomly taken) of the identity matrix $\mathbb{1}_m$, and $\mathbf{P}$ a diagonal of random phases. In Fig. 4, we give the MSE obtained by the different spectral methods and these two matrix ensembles. We also give examples of the images recovered by the algorithms. Eventually, despite the fact that the signal (and possibly the matrix as well) is structured, we still observe the same transition phenomenon in the spectra of $\mathbf{M}^{(\mathrm{TAP})}$ and $\mathbf{M}^{(\mathrm{LAMP})}$, as shown
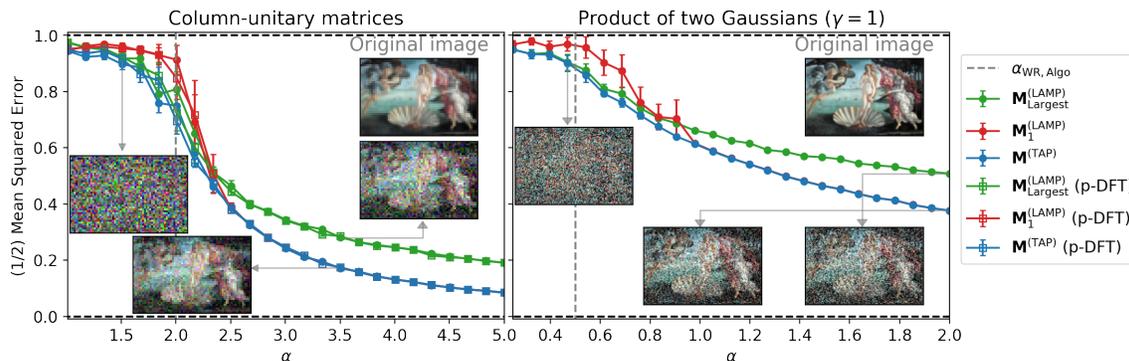
Figure 4: Mean squared error achieved by the different spectral methods for the recovery of a natural image in noiseless phase retrieval. We consider column-unitary matrices $\mathbf{\Phi}$ (both uniformly sampled and partial DFT matrices, left) and the product of two complex Gaussian matrices with aspect ratio $\gamma = 1$ (right). We reduced each dimension of the original 1280×820 image by a factor 20 (left) or 10 (right), and we average the MSE over 5 instances and the 3 RGB channels (which are recovered independently).

in the supplementary material, in Fig. 7. Namely, we still observe that the optimal estimator is associated with marginal stability of both spectral methods, while the largest eigenvalue of $\mathbf{M}^{(\mathrm{LAMP})}$ is associated to a non-optimal estimator.

Let us also illustrate how this spectral method can be combined with a subsequent local optimization algorithm. We use the spectral estimator as the initialization point to running vanilla gradient descent on the square loss $L(\mathbf{x}) \equiv \frac{1}{2m} \sum_{\mu=1}^{m} \left\{ \left| \frac{(\mathbf{\Phi}\mathbf{x})_{\mu}}{\sqrt{n}} \right|^2 - \left| \frac{(\mathbf{\Phi}\mathbf{X}^\star)_{\mu}}{\sqrt{n}} \right|^2 \right\}^2$. This allows to already obtain a perfect recovery of the image for $\alpha = 4$, as shown in Fig. 5. In Appendix C.2 we expand this analysis by showing the MSE achieved by the gradient descent procedure. In particular, we confirm that combining the gradient descent with the spectral initialization allows to reach perfect recovery at finite α, which is not possible with the "vanilla" spectral methods.

3.4. Perspectives

Our analysis raises interesting open questions, both from the random matrix theory and the statistical physics viewpoint.

- First, we notice that the optimal estimator is always associated with *marginal stability*, both in $\mathbf{M}^{(\mathrm{LAMP})}$ and $\mathbf{M}^{(\mathrm{TAP})}$. A clear understanding of this marginal stability (which was already observed in Ma et al. (2021) in the column-unitary case) is still lacking. Moreover, the principal eigenvector of the matrix $\mathbf{M}^{(\mathrm{LAMP})}$ is associated to an *unstable* direction, thus dominating the dynamics of the linearized-AMP. However its achieved correlation is smaller than the one achieved by the marginally stable, optimal, eigenvector. We also noticed that the eigenvectors of $\mathbf{M}^{(\mathrm{TAP})}$ *do not contain any information about this suboptimal estimator*[1]. This blindness of $\mathbf{M}^{(\mathrm{TAP})}$ to the

---

[1]In particular, this is an important distinction between our L-AMP constructive derivation and the L-AMP algorithms of Ma et al. (2021), which are *designed* to match the spectral methods of the type $\mathbf{M}(\mathcal{T})$: in the latter, it was shown that the L-AMP estimator always matched the one of the spectral method.
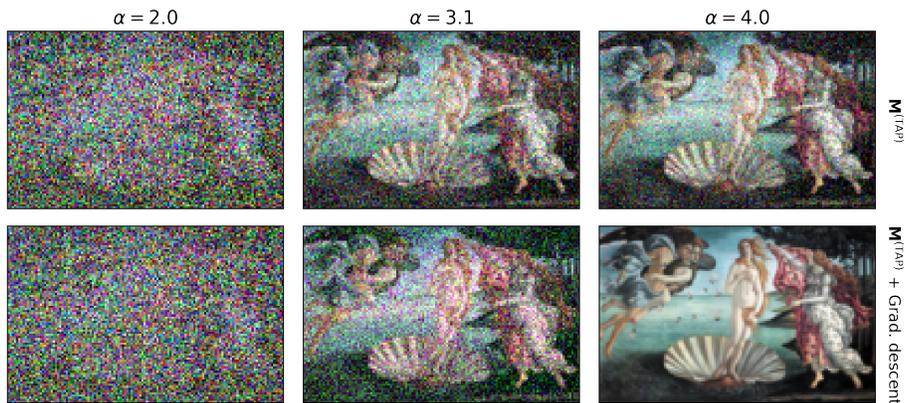
Figure 5: Reconstruction of a real image in noiseless phase retrieval with partial DFT matrices. We reduce the image size from 1280×820 to 128×82. We compare, for three different values of α, the estimators of $\mathbf{M}(\mathcal{T}^\star)$ (top line) and the estimator obtained by running a gradient descent procedure starting from the estimator of $\mathbf{M}(\mathcal{T}^\star)$ (bottom line). We recover the 3 RGB channels with independent instances of the sensing matrix.

principal eigenvector of $\mathbf{M}^{(\mathrm{LAMP})}$ is very puzzling from a theoretical point of view. Indeed, as shown in Maillard et al. (2019) and reminded in Section 2.2.1, the stationary limit of G-VAMP (Algorithm 1) is in exact correspondence with the stationary point equations of the TAP free entropy. One would therefore expect the two spectral methods $\mathbf{M}^{(\mathrm{LAMP})}$ and $\mathbf{M}^{(\mathrm{TAP})}$ to contain the same physical information on the system. Finally, the different qualitative behaviors of the two methods (instability of $\mathbf{M}^{(\mathrm{LAMP})}$ a opposed to marginal stability of $\mathbf{M}^{(\mathrm{TAP})}$) only deepens this puzzle, and understanding this disparity is an interesting open problem.

- Importantly, our analysis is essentially not rigorous (hence the use of conjectures). An interesting perspective would be to establish rigorously our statements, in similarity with what is proven in Dudeja et al. (2020a) on the analysis of Ma et al. (2021) for column-unitary matrices. This would require a random matrix theory analysis of the "BBP"[1] transition in matrices of the form of eq. (9), which is, to the best of our knowledge, lacking in the generic rotationally-invariant case. Another approach would be to use the (rigorously known) *state evolution* (SE) of AMP, which allows to track its asymptotic performance. This approach was considered in Ma et al. (2021); Dudeja and Bakhshizadeh (2020): importantly, this method also provides analytically the asymptotic performance of the spectral method, which is not derived in the present work.

- Another important perspective is to apply our methods in real-world settings in which the way the data and the signal are generated is not necessarily known. Our analysis of a real image (cf Fig. 4) suggests that having a structured prior distribution does not harm our conclusions. The influence of a so-called "mismatched" setting in the channel distribution (i.e. when the data is generated with a distribution P_{out}^0 and inferred with a different distribution P_{out}) is however less clear, and we leave it for future work.

---

[1] i.e. the appearance of a largest eigenvalue detached from the bulk of the other eigenvalues, as α increases. It was first rigorously analyzed in Baik et al. (2005) for spiked Gaussian matrices.

Acknowledgments

References

Benjamin Aubin, Bruno Loureiro, Antoine Baker, Florent Krzakala, and Lenka Zdeborová. Exact asymptotics for phase retrieval and compressed sensing with random generative priors. In *Mathematical and Scientific Machine Learning*, pages 55–73. PMLR, 2020a.

Benjamin Aubin, Bruno Loureiro, Antoine Maillard, Florent Krzakala, and Lenka Zdeborová. The spiked matrix model with generative priors. *IEEE Transactions on Information Theory*, 2020b.

Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

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

Emmanuel J Candès, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval via wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, 2015.

Emmanuel J Candes, Xiaodong Li, and Mahdi Soltanolkotabi. Phase retrieval from coded diffraction patterns. *Applied and Computational Harmonic Analysis*, 39(2):277–299, 2015.

Angélique Drémeau, Antoine Liutkus, David Martina, Ori Katz, Christophe Schülke, Florent Krzakala, Sylvain Gigan, and Laurent Daudet. Reference-less measurement of the transmission matrix of a highly scattering material using a dmd and phase retrieval techniques. *Optics express*, 23(9): 11898–11911, 2015.

Rishabh Dudeja and Milad Bakhshizadeh. Universality of linearized message passing for phase retrieval with structured sensing matrices. *arXiv preprint arXiv:2008.10503*, 2020.

Rishabh Dudeja, Milad Bakhshizadeh, Junjie Ma, and Arian Maleki. Analysis of spectral methods for phase retrieval with random orthogonal matrices. *IEEE Transactions on Information Theory*, 2020a.

Rishabh Dudeja, Junjie Ma, and Arian Maleki. Information theoretic limits for phase retrieval with subsampled haar sensing matrices. *IEEE Transactions on Information Theory*, 2020b.

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

Tom Goldstein and Christoph Studer. Phasemax: Convex phase retrieval via basis pursuit. *IEEE Transactions on Information Theory*, 64(4):2675–2689, 2018.

Paul Hand, Oscar Leong, and Vladislav Voroninski. Phase retrieval under a generative prior. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pages 9154–9164, 2018.

Kishore Jaganathan, Yonina C Eldar, and Babak Hassibi. Phase retrieval: An overview of recent developments. *arXiv preprint arXiv:1510.07713*, 2015.

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

Yoshiyuki Kabashima, Florent Krzakala, Marc Mézard, Ayaka Sakata, and Lenka Zdeborová. Phase transitions and sample complexity in bayes-optimal matrix factorization. *IEEE Transactions on information theory*, 62(7):4228–4265, 2016.

Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013.

Yue M Lu and Gen Li. Phase transitions of spectral initialization for high-dimensional non-convex estimation. *Information and Inference: A Journal of the IMA*, 9(3):507–541, 2020.

Wangyu Luo, Wael Alghamdi, and Yue M Lu. Optimal spectral initialization for signal recovery with applications to phase retrieval. *IEEE Transactions on Signal Processing*, 67(9):2347–2356, 2019.

Junjie Ma, Xiaojun Yuan, and Li Ping. Turbo compressed sensing with partial dft sensing matrix. *IEEE Signal Processing Letters*, 22(2):158–161, 2014.

Junjie Ma, Rishabh Dudeja, Ji Xu, Arian Maleki, and Xiaodong Wang. Spectral method for phase retrieval: an expectation propagation perspective. *IEEE Transactions on Information Theory*, 67 (2):1332–1355, 2021.

Antoine Maillard, Laura Foini, Alejandro Lage Castellanos, Florent Krzakala, Marc Mézard, and Lenka Zdeborová. High-temperature expansions and message passing algorithms. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(11):113301, 2019.

Antoine Maillard, Bruno Loureiro, Florent Krzakala, and Lenka Zdeborová. Phase retrieval in high dimensions: Statistical and computational phase transitions. *Advances in Neural Information Processing Systems*, 33, 2020.

Marco Mondelli and Andrea Montanari. Fundamental limits of weak recovery with applications to phase retrieval. *Foundations of Computational Mathematics*, 19(3):703–773, 2019.

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.

Marco Mondelli, Christos Thrampoulidis, and Ramji Venkataramanan. Optimal combination of linear and spectral estimators for generalized linear models. *arXiv preprint arXiv:2008.03326*, 2020.

Praneeth Netrapalli, Prateek Jain, and Sujay Sanghavi. Phase retrieval using alternating minimization. *IEEE Transactions on Signal Processing*, 63(18):4814–4826, 2015.

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

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.

Philip Schniter and Sundeep Rangan. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing*, 63(4):1043–1055, 2014.

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

Yoav Shechtman, Yonina C Eldar, Oren Cohen, Henry Nicholas Chapman, Jianwei Miao, and Mordechai Segev. Phase retrieval with application to optical imaging: a contemporary overview. *IEEE signal processing magazine*, 32(3):87–109, 2015.

Ju Sun, Qing Qu, and John Wright. A geometric analysis of phase retrieval. *Foundations of Computational Mathematics*, 18(5):1131–1198, 2018.

Takashi Takahashi and Yoshiyuki Kabashima. Macroscopic analysis of vector approximate message passing in a model mismatch setting. In *2020 IEEE International Symposium on Information Theory (ISIT)*, pages 1403–1408. IEEE, 2020.

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

Michael Unser and Murray Eden. Maximum likelihood estimation of liner signal parameters for poisson processes. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 36(6): 942–945, 1988.

Lorenzo Valzania, Jonathan Dong, and Sylvain Gigan. Accelerating ptychographic reconstructions using spectral initializations. *Optics Letters*, 46(6):1357–1360, 2021.

Irene Waldspurger. Phase retrieval with random gaussian sensing vectors by alternating projections. *IEEE Transactions on Information Theory*, 64(5):3301–3312, 2018.

Appendix A. Linearized Approximate Message Passing in the complex case

In the complex case, we write the linearization of Algorithm 1 as:

$$
\begin{cases}
\delta\hat{\mathbf{x}}_1^t = \nabla_{\mathbf{T}}g_{x1}(0,0)\delta\mathbf{T}_1^t + \nabla_{\bar{\mathbf{T}}}g_{x1}(0,0)\overline{\delta\mathbf{T}_1^t}, \\
\delta\hat{\mathbf{z}}_1^t = \nabla_{\mathbf{R}}g_{z1}(0,\rho^{-1})\delta\mathbf{R}_1^t + \nabla_{\bar{\mathbf{R}}}g_{z1}(0,\rho^{-1})\overline{\delta\mathbf{R}_1^t}, \\
\delta\mathbf{T}_2^t = \frac{1}{\rho}\delta\hat{\mathbf{x}}_1^t - \delta\mathbf{T}_1^t, \qquad\qquad\qquad\qquad\quad \delta\mathbf{R}_2^t = \frac{\alpha}{\rho\langle\lambda\rangle_\nu}\delta\hat{\mathbf{z}}_1^t - \delta\mathbf{R}_1^t, \\
\delta\hat{\mathbf{x}}_2^t = \nabla_{\mathbf{T}}g_{x2}(0,0,\rho^{-1},0)\delta\mathbf{T}_2^t + \nabla_{\mathbf{R}}g_{x2}(0,0,\rho^{-1},0)\delta\mathbf{R}_2^t, \\
\delta\hat{\mathbf{z}}_2^t = \nabla_{\mathbf{T}}g_{z2}(0,0,\rho^{-1},0)\delta\mathbf{T}_2^t + \nabla_{\mathbf{R}}g_{z2}(0,0,\rho^{-1},0)\delta\mathbf{R}_2^t, \\
\delta\mathbf{T}_1^{t+1} = \frac{1}{\rho}\delta\hat{\mathbf{x}}_2^t - \delta\mathbf{T}_2^t, \qquad\qquad\qquad\qquad\; \delta\mathbf{R}_1^{t+1} = \frac{\alpha}{\rho\langle\lambda\rangle_\nu}\delta\hat{\mathbf{z}}_2^t - \delta\mathbf{R}_2^t.
\end{cases} \tag{21}
$$

Recall that here $\partial_z, \partial_{\bar{z}}$ are the usual Wirtinger derivatives. Since the functions g_{x2}, g_{z2}, defined in eq. (10), are obviously holomorphic, we did not include their derivative $\partial_{\bar{z}}$ as it is trivially zero. Moreover, we assumed that $P_0(z), P_{\mathrm{out}}(y|z)$ are functions of $|z|^2$ (i.e. *spherical symmetry*), which defined our phase retrieval problem. Starting from the definition of eq. (10), this implies that

$$
\partial_{\overline{T_i}}g_{x1}(0,0) = 2\big(\mathbb{E}_{P_0}[z^2] - \mathbb{E}_{P_0}[z]\big) = 0,
$$

in which the last equality is a consequence of the spherical symmetry. In the same way, one obtains $\nabla_{\bar{\mathbf{R}}}g_{z1}(0,\rho^{-1}) = 0$. We can then compute, as in the real case (cf eq. (14)):

$$
\begin{cases}
\partial_{T_j}[(g_{x1}(0,0)_i] & = \rho\,\delta_{ij}, \\
\partial_{R_\nu}[g_{z1}(0,\rho^{-1})_\mu] & = \delta_{\mu\nu}\mathbb{E}_{P_{\mathrm{out}}(y_\mu,0,\rho\langle\lambda\rangle_\nu/\alpha)} \equiv \delta_{\mu\nu}v(y_\mu), \\
\partial_{T_j}[g_{x2}(0,0,\rho^{-1},0)] & = \rho\,\delta_{ij}, \\
\partial_{R_\mu}[g_{x2}(0,0,\rho^{-1},0)_i] & = \rho(\mathbf{V}\mathbf{S}^{\mathsf{T}}\mathbf{U}^\dagger)_{i\mu} = \rho\frac{(\mathbf{\Phi}^\dagger)_{i\mu}}{\sqrt{n}}, \\
\partial_{T_i}[g_{z2}(0,0,\rho^{-1},0)_\mu] & = \rho\frac{\mathbf{\Phi}_{\mu i}}{\sqrt{n}}, \\
\partial_{R_\nu}[g_{z2}(0,0,\rho^{-1},0)_\mu] & = \rho\frac{(\mathbf{\Phi}\mathbf{\Phi}^\dagger)_{\mu\nu}}{n}.
\end{cases} \tag{22}
$$

The derivation of the real case then straightforwardly transfers to the complex case, and we reach eq. (16) in the complex case, as claimed.

Appendix B. The Hessian of the TAP free entropy

B.1. The derivatives of the parameters at the trivial fixed point

We start from the relations of eq. (19). Let us differentiate them with respect to $m_i^{(a)}$, for any $a \in \{1,\beta\}$ and $i \in \{1,\cdots,n\}$. We denote $P_{\mathrm{out}}^\mu \equiv P_{\mathrm{out}}(y_\mu,\omega_\mu,b)$. We get after tedious calculations

the cumbersome equations (valid for any **m**) :

$$
\left\{
\begin{aligned}
&\delta_{ij}\mathbf{e}_a = \beta\Big\{ -\frac{1}{2}\frac{\partial\gamma}{\partial m_i^{(a)}}(\mathbb{E}_{P_0(\gamma,\lambda_j)}[x|x|^2] - \mathbb{E}_{P_0(\gamma,\lambda_j)}[x]\mathbb{E}_{P_0(\gamma,\lambda_j)}[|x|^2]) && \text{(23a)}\\
&\quad -\mathbb{E}_{P_0(\gamma,\lambda_j)}\Big[x\big(x\cdot\frac{\partial\lambda_j}{\partial m_i^{(a)}}\big)\Big] + \mathbb{E}_{P_0(\gamma,\lambda_j)}[x]\mathbb{E}_{P_0(\gamma,\lambda_j)}\Big[x\cdot\frac{\partial\lambda_j}{\partial m_i^{(a)}}\Big]\Big\},\\[4pt]
&\frac{\partial\sigma^2}{\partial m_i^{(a)}} = \frac{1}{n}\sum_{j=1}^{n}\Big[-2m_j^{(a)}\delta_{ij} + \beta\Big\{ -\frac{1}{2}\frac{\partial\gamma}{\partial m_i^{(a)}}(\mathbb{E}_{P_0(\gamma,\lambda_j)}[|x|^4] - (\mathbb{E}_{P_0(\gamma,\lambda_j)}[|x|^2])^2) && \text{(23b)}\\
&\quad +\mathbb{E}_{P_0(\gamma,\lambda_j)}\Big[|x|^2\big(x\cdot\frac{\partial\lambda_j}{\partial m_i^{(a)}}\big)\Big] - \mathbb{E}_{P_0(\gamma,\lambda_j)}[|x|^2]\mathbb{E}_{P_0(\gamma,\lambda_j)}\Big[x\cdot\frac{\partial\lambda_j}{\partial m_i^{(a)}}\Big]\Big\}\Big],\\[4pt]
&\frac{\partial g_\mu}{\partial m_i^{(a)}} = \frac{1}{b}\frac{\partial\omega_\mu}{\partial m_i^{(a)}} + \frac{\partial b}{\partial m_i^{(a)}}\Big\{ b^{-2}\mathbb{E}_{P_{\text{out}}^\mu}[h-\omega_\mu] && \text{(23c)}\\
&\quad +\frac{\beta}{2b^3}\big(\mathbb{E}_{P_{\text{out}}^\mu}[(h-\omega_\mu)|h-\omega_\mu|^2] - \mathbb{E}_{P_{\text{out}}^\mu}[h-\omega_\mu]\mathbb{E}_{P_{\text{out}}^\mu}[|h-\omega_\mu|^2]\big)\Big\}\\
&\quad -\frac{\beta}{b^2}\Big(\mathbb{E}_{P_{\text{out}}^\mu}\Big[(h-\omega_\mu)(h-\omega_\mu)\cdot\frac{\partial\omega_\mu}{\partial m_i^{(a)}}\Big] - \mathbb{E}_{P_{\text{out}}^\mu}[h-\omega_\mu]\mathbb{E}_{P_{\text{out}}^\mu}\Big[(h-\omega_\mu)\cdot\frac{\partial\omega_\mu}{\partial m_i^{(a)}}\Big]\Big),\\[4pt]
&\frac{\partial r}{\partial m_i^{(a)}} = \frac{1}{m}\sum_{\mu=1}^{m}\Big\{ 2g_\mu\cdot\frac{\partial g_\mu}{\partial m_i^{(a)}} - b^{-2}\frac{\partial b}{\partial m_i^{(a)}} + \frac{1}{b^2}\mathbb{E}_{P_{\text{out}}^\mu}\Big[(h-\omega_\mu)\cdot\frac{\partial\omega_\mu}{\partial m_i^{(a)}}\Big] && \text{(23d)}\\
&\quad +\frac{\partial b}{\partial m_i^{(a)}}\Big\{ 2b^{-3}\mathbb{E}_{P_{\text{out}}^\mu}[|h-\omega_\mu|^2] + \frac{\beta}{2b^4}\big(\mathbb{E}_{P_{\text{out}}^\mu}[|h-\omega_\mu|^4] - (\mathbb{E}_{P_{\text{out}}^\mu}[|h-\omega_\mu|^2])^2\big)\Big\}\\
&\quad +\frac{\beta}{b^3}\Big(\mathbb{E}_{P_{\text{out}}^\mu}\Big[|h-\omega_\mu|^2(h-\omega_\mu)\cdot\frac{\partial\omega_\mu}{\partial m_i^{(a)}}\Big] - \mathbb{E}_{P_{\text{out}}^\mu}[|h-\omega_\mu|^2]\mathbb{E}_{P_{\text{out}}^\mu}\Big[(h-\omega_\mu)\cdot\frac{\partial\omega_\mu}{\partial m_i^{(a)}}\Big]\Big)\Big\},\\[4pt]
&\frac{\partial\gamma}{\partial m_i^{(a)}} = -2\Big[\frac{\partial\sigma^2}{\partial m_i^{(a)}}\partial_{\sigma^2}^2 F(\sigma^2,r) + \frac{\partial r}{\partial m_i^{(a)}}\partial_{\sigma^2,r}^2 F(\sigma^2,r)\Big], && \text{(23e)}\\[4pt]
&\frac{\partial\omega_\mu}{\partial m_i^{(a)}} - \frac{\partial b}{\partial m_i^{(a)}}g_\mu - b\frac{\partial g_\mu}{\partial m_i^{(a)}} = \frac{\Phi_{\mu i}}{\sqrt{n}}\mathbf{e}_a, && \text{(23f)}\\[4pt]
&\frac{\partial b}{\partial m_i^{(a)}} = -\frac{2}{\alpha}\Big[\frac{\partial\sigma^2}{\partial m_i^{(a)}}\partial_{\sigma^2,r}^2 F(\sigma^2,r) + \frac{\partial r}{\partial m_i^{(a)}}\partial_r^2 F(\sigma^2,r)\Big]. && \text{(23g)}
\end{aligned}
\right.
$$

Here we denoted $\mathbf{e}_a = 1$ if $\mathbb{K} = \mathbb{R}$, and $(\mathbf{e}_a)_b = \delta_{ab}$ if $\mathbb{K} = \mathbb{C}$. In particular, taken at the trivial fixed point, these equations can be greatly simplified, using the value of the parameters at this point, the

symmetries of the channel and prior, and the development of the F function, cf eq. (20):

$$
\begin{cases}
\delta_{ij}\mathbf{e}_a = -\beta\mathbb{E}_{P_0}\Big[x\big(x\cdot\frac{\partial\lambda_j}{\partial m_i^{(a)}}\big)\Big], & \text{(24a)} \\[2ex]
\dfrac{\partial\sigma^2}{\partial m_i^{(a)}} = \dfrac{-\beta}{2}\dfrac{\partial\gamma}{\partial m_i^{(a)}}(\mathbb{E}_{P_0}[|x|^4]-(\mathbb{E}_{P_0}[|x|^2])^2), & \text{(24b)} \\[2ex]
\dfrac{\partial g_\mu}{\partial m_i^{(a)}} = \dfrac{\alpha}{\rho\langle\lambda\rangle_\nu}\dfrac{\partial\omega_\mu}{\partial m_i^{(a)}} - \dfrac{\beta\alpha^2}{\rho^2\langle\lambda\rangle_\nu^2}\mathbb{E}_{P_{\text{out}}(y_\mu,0,\rho\langle\lambda\rangle_\nu/\alpha)}\Big[h\big(h\cdot\frac{\partial\omega_\mu}{\partial m_i^{(a)}}\big)\Big], & \text{(24c)} \\[2ex]
\dfrac{\partial r}{\partial m_i^{(a)}} = \dfrac{2\alpha^2}{\rho^2\langle\lambda\rangle_\nu^2}\dfrac{\partial b}{\partial m_i^{(a)}} & \text{(24d)} \\[2ex]
\dfrac{\partial\gamma}{\partial m_i^{(a)}} = \langle\lambda\rangle_\nu\dfrac{\partial r}{\partial m_i^{(a)}}, & \text{(24e)} \\[2ex]
\dfrac{\partial\omega_\mu}{\partial m_i^{(a)}} - \dfrac{\rho\langle\lambda\rangle_\nu}{\alpha}\dfrac{\partial g_\mu}{\partial m_i^{(a)}} = \dfrac{\Phi_{\mu i}}{\sqrt{n}}\mathbf{e}_a, & \text{(24f)} \\[2ex]
\dfrac{\partial b}{\partial m_i^{(a)}} = \dfrac{\langle\lambda\rangle_\nu}{\alpha}\dfrac{\partial\sigma^2}{\partial m_i^{(a)}} - \dfrac{\rho^2}{2\alpha^2}[\alpha\langle\lambda^2\rangle_\nu-(1+\alpha)\langle\lambda\rangle_\nu^2]\dfrac{\partial r}{\partial m_i^{(a)}}. & \text{(24g)}
\end{cases}
$$

We used eq. (29) and eq. (11) (from the derivation of $\mathbf{M}^{(\text{LAMP})}$) to simplify the equation involving the derivative of r. One can already notice the very interesting fact that the variance scalar parameters and the vector parameters are decoupled ! Moreover, it is easy to see that the equations on the variance parameters can be closed to:

$$
\frac{\partial\sigma^2}{\partial m_i^{(a)}} = -\frac{\rho^2(1+\alpha)(\langle\lambda^2\rangle_\nu-\langle\lambda\rangle_\nu^2)}{\beta\langle\lambda\rangle_\nu^2\text{Var}_{P_0}[|X|^2]}\frac{\partial\sigma^2}{\partial m_i^{(a)}}.
$$

This equation is of the type $\partial_{m_i^{(a)}}\sigma^2 = -t\partial_{m_i^{(a)}}\sigma^2$, with $t>0$, and thus we have

$$
\frac{\partial\sigma^2}{\partial m_i^{(a)}} = \frac{\partial\gamma}{\partial m_i^{(a)}} = \frac{\partial r}{\partial m_i^{(a)}} = \frac{\partial b}{\partial m_i^{(a)}} = 0. \tag{25}
$$

Moreover, from eq. (24), we can obtain as well the derivatives of the vector parameters at the trivial fixed point:

$$
\begin{cases}
\dfrac{\partial\lambda_j}{\partial m_i^{(a)}} = -\dfrac{\delta_{ij}}{\rho}\mathbf{e}_a, & \text{(26a)} \\[2ex]
\dfrac{\partial g_\mu}{\partial m_i^{(a)}} = \dfrac{\alpha}{\rho\langle\lambda\rangle_\nu}\Big[1 - \dfrac{\alpha}{\rho\langle\lambda\rangle_\nu}\mathbb{E}_{P_{\text{out}}(y_\mu,0,\rho\langle\lambda\rangle_\nu/\alpha)}[|h|^2]\Big]\dfrac{\partial\omega_\mu}{\partial m_i^{(a)}} & \\[2ex]
\qquad = -\partial_\omega g_{\text{out}}(y_\mu,0,\rho\langle\lambda\rangle_\nu/\alpha)\dfrac{\partial\omega_\mu}{\partial m_i^{(a)}}, & \text{(26b)} \\[2ex]
\dfrac{\partial\omega_\mu}{\partial m_i^{(a)}} = \dfrac{\rho\langle\lambda\rangle_\nu}{\alpha}\dfrac{\partial g_\mu}{\partial m_i^{(a)}} + \dfrac{\Phi_{\mu i}}{\sqrt{n}}\mathbf{e}_a. & \text{(26c)}
\end{cases}
$$

These equations can easily be solved as:

$$
\begin{cases}
\dfrac{\partial \lambda_j}{\partial m_i^{(a)}} = -\dfrac{\delta_{ij}}{\rho}\mathbf{e}_a, & \text{(27a)} \\[3mm]
\dfrac{\partial \omega_\mu}{\partial m_i^{(a)}} = \dfrac{\Phi_{\mu i}}{\sqrt{n}} \dfrac{1}{1 + \frac{\rho\langle\lambda\rangle_\nu}{\alpha}\partial_\omega g_{\text{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}\mathbf{e}_a, & \text{(27b)} \\[3mm]
\dfrac{\partial g_\mu}{\partial m_i^{(a)}} = -\dfrac{\Phi_{\mu i}}{\sqrt{n}} \dfrac{\partial_\omega g_{\text{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}{1 + \frac{\rho\langle\lambda\rangle_\nu}{\alpha}\partial_\omega g_{\text{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}\mathbf{e}_a. & \text{(27c)}
\end{cases}
$$

B.2. The expansion of the free entropy

We start from eq. (18):

$$
\begin{aligned}
f_{\text{TAP}}(\mathbf{m}) =\ & \frac{\beta}{n}\sum_{i=1}^{n}\lambda_i \cdot m_i + \frac{\beta\gamma}{2n}\Big(n\sigma^2 + \sum_{i=1}^{n}|m_i|^2\Big) + \frac{\alpha\beta}{m}\sum_{\mu=1}^{m}\omega_\mu \cdot g_\mu \\
& - \frac{\beta b}{2n}\Big(\sum_{\mu=1}^{m}|g_\mu|^2 - \alpha n r\Big) + \frac{1}{n}\sum_{i=1}^{n}\ln \int_{\mathbb{K}}P_0(\mathrm{d}x)e^{-\frac{\beta\gamma}{2}|x|^2 - \beta\lambda_i \cdot x} \\
& + \frac{\alpha}{m}\sum_{\mu=1}^{m}\ln \int_{\mathbb{K}}\frac{\mathrm{d}h}{\left(\frac{2\pi b}{\beta}\right)^{\beta/2}}P_{\text{out}}(y_\mu|h)e^{-\frac{\beta|h-\omega_\mu|^2}{2b}} - \frac{\beta}{n}\sum_{i=1}^{n}\sum_{\mu=1}^{m}g_\mu \cdot \Big(\frac{\Phi_{\mu i}}{\sqrt{n}}m_i\Big) + \beta F(\sigma^2, r).
\end{aligned}
$$

At the trivial fixed point, we obtain by differentiating this expression twice (using the form of the trivial fixed point and eq. (25)):

$$
\begin{aligned}
\frac{\partial^2 f_{\text{TAP}}}{\partial m_i^{(a)}\partial m_j^{(b)}} =\ & \\
& \frac{2\beta}{n}\delta_{ij}\delta_{ab}\frac{\partial \lambda_i^{(a)}}{\partial m_i^{(a)}} + \frac{\alpha\beta}{m}\sum_{\mu=1}^{m}\Big[\frac{\partial \omega_\mu}{\partial m_i^{(a)}}\cdot\frac{\partial g_\mu}{\partial m_j^{(b)}} + \frac{\partial \omega_\mu}{\partial m_j^{(b)}}\cdot\frac{\partial g_\mu}{\partial m_i^{(a)}}\Big] + \beta\rho\delta_{ij}\delta_{ab}\Big(\frac{\partial \lambda_i^{(a)}}{\partial m_i^{(a)}}\Big)^2 \\
& - \frac{\beta\rho\langle\lambda\rangle_\nu}{m}\sum_{\mu=1}^{m}\frac{\partial g_\mu}{\partial m_i^{(a)}}\cdot\frac{\partial g_\mu}{\partial m_j^{(b)}} - \frac{\beta}{n}\sum_{\mu=1}^{m}\Big\{\frac{1}{\sqrt{n}}\frac{\partial(\overline{\Phi_{\mu i}}g_\mu)^{(a)}}{\partial m_j^{(b)}} + \frac{1}{\sqrt{n}}\frac{(\overline{\Phi_{\mu j}}\partial g_\mu)^{(b)}}{\partial m_i^{(a)}}\Big\} \\
& + \frac{\alpha}{m}\sum_{\mu=1}^{m}\Big[-\frac{\beta\alpha}{2\rho\langle\lambda\rangle_\nu}\frac{\partial^2 b}{\partial m_i^{(a)}\partial m_j^{(b)}} + \frac{\beta\alpha^2}{2\rho^2\langle\lambda\rangle_\nu^2}\frac{\partial^2 b}{\partial m_i^{(a)}\partial m_j^{(b)}}\mathbb{E}_{P_{\text{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}[|h|^2]\Big] \\
& + \frac{\beta\alpha^2}{m\rho\langle\lambda\rangle_\nu}\sum_{\mu=1}^{m}\Big(\frac{\partial \omega_\mu}{\partial m_i^{(a)}}\Big)\cdot\Big(\frac{\partial \omega_\mu}{\partial m_j^{(b)}}\Big)\Big\{\frac{\alpha}{\rho\langle\lambda\rangle_\nu}\mathbb{E}_{P_{\text{out}}(y_\mu, 0, \rho\langle\lambda\rangle_\nu/\alpha)}[|h|^2] - 1\Big\}.
\end{aligned}
$$

We then use eq. (27) and eq. (11), to simplify slightly the result:

$$
\frac{n}{\beta} \frac{\partial^2 f_{\mathrm{TAP}}}{\partial m_i^{(a)} \partial m_j^{(b)}} = \frac{-1}{\rho} \delta_{ij} \delta_{ab} + \sum_{\mu=1}^{m} \left[\frac{\partial \omega_\mu}{\partial m_i^{(a)}} \cdot \frac{\partial g_\mu}{\partial m_j^{(b)}} + \frac{\partial \omega_\mu}{\partial m_j^{(b)}} \cdot \frac{\partial g_\mu}{\partial m_i^{(a)}} \right]
$$

$$
- \frac{\rho \langle \lambda \rangle_\nu}{\alpha} \sum_{\mu=1}^{m} \frac{\partial g_\mu}{\partial m_i^{(a)}} \cdot \frac{\partial g_\mu}{\partial m_j^{(b)}} - \sum_{\mu=1}^{m} \left\{ \frac{1}{\sqrt{n}} \frac{\partial (\overline{\Phi_{\mu i}} g_\mu)^{(a)}}{\partial m_j^{(b)}} + \frac{1}{\sqrt{n}} \frac{(\overline{\Phi_{\mu j}} \partial g_\mu)^{(b)}}{\partial m_i^{(a)}} \right\}
$$

$$
+ \sum_{\mu=1}^{m} \left(\frac{\partial \omega_\mu}{\partial m_i^{(a)}} \right) \cdot \left(\frac{\partial \omega_\mu}{\partial m_j^{(b)}} \right) \partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho \langle \lambda \rangle_\nu / \alpha).
$$

We also used eq. (7) to make $\partial_\omega g_{\mathrm{out}}$ appear in the last term. As is clear from this last equation and eq. (27), the dependency on a, b of the result will fully be determined by the quantity $(\Phi_{\mu i} \mathbf{e}_a) \cdot (\Phi_{\mu j} \mathbf{e}_b)$. For $\beta = 1$, this is simply equal to $\Phi_{\mu i} \Phi_{\mu j}$. For $\beta = 2$, this can be represented as a 2×2 matrix:

$$
\left\{ (\Phi_{\mu i} \mathbf{e}_a) \cdot (\Phi_{\mu j} \mathbf{e}_b) \right\}_{a,b=1,2} = \begin{pmatrix} \mathrm{Re}[\overline{\Phi_{\mu i}} \Phi_{\mu j}] & -\mathrm{Im}[\overline{\Phi_{\mu i}} \Phi_{\mu j}] \\ \mathrm{Im}[\overline{\Phi_{\mu i}} \Phi_{\mu j}] & \mathrm{Re}[\overline{\Phi_{\mu i}} \Phi_{\mu j}] \end{pmatrix}.
$$

This is just the usual matrix representation of the complex number $\overline{\Phi_{\mu i}} \Phi_{\mu j}$. Following this representation, we can formally write $n \frac{\partial^2 f_{\mathrm{TAP}}}{\partial m_i \partial m_j}$ as an element of $\mathbb{K}$! This yields:

$$
\frac{n}{\beta} \frac{\partial^2 f_{\mathrm{TAP}}}{\partial m_i \partial m_j} = \frac{-1}{\rho} \delta_{ij} + \sum_{\mu=1}^{m} \frac{\overline{\Phi_{\mu i}} \Phi_{\mu j}}{n} \frac{\partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho \langle \lambda \rangle_\nu / \alpha)}{1 + \frac{\rho \langle \lambda \rangle_\nu}{\alpha} \partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho \langle \lambda \rangle_\nu / \alpha)}.
$$

Appendix C. Additional numerical experiments
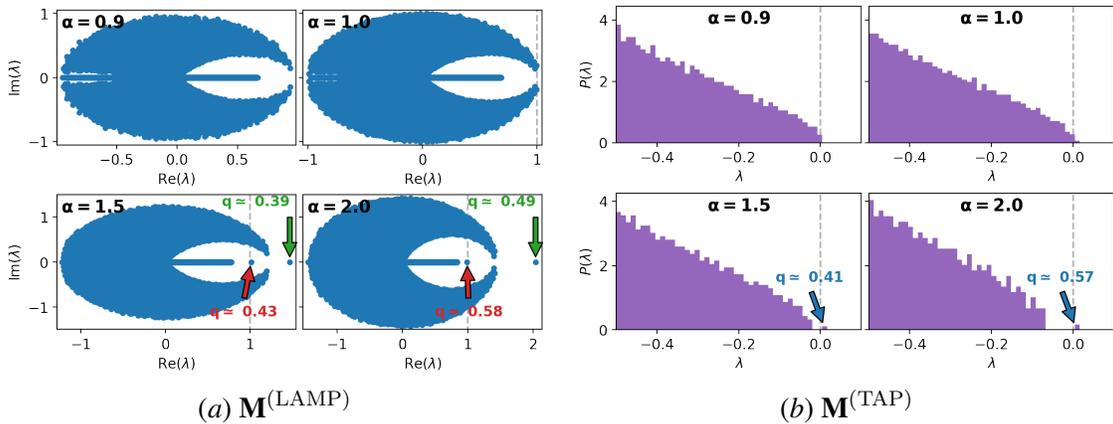
C.1. The transition in the spectra

(a) $\mathbf{M}^{(\mathrm{LAMP})}$ (b) $\mathbf{M}^{(\mathrm{TAP})}$

Figure 6: Transition in the spectrum of $\mathbf{M}^{(\mathrm{LAMP})}$ and $\mathbf{M}^{(\mathrm{TAP})}$ for a complex Gaussian $\boldsymbol{\Phi}$ and a noiseless phase retrieval channel. For $\alpha > \alpha_{\mathrm{WR,Algo}} = 1$, we indicate the approximate overlap q corresponding to the relevant eigenvalues.
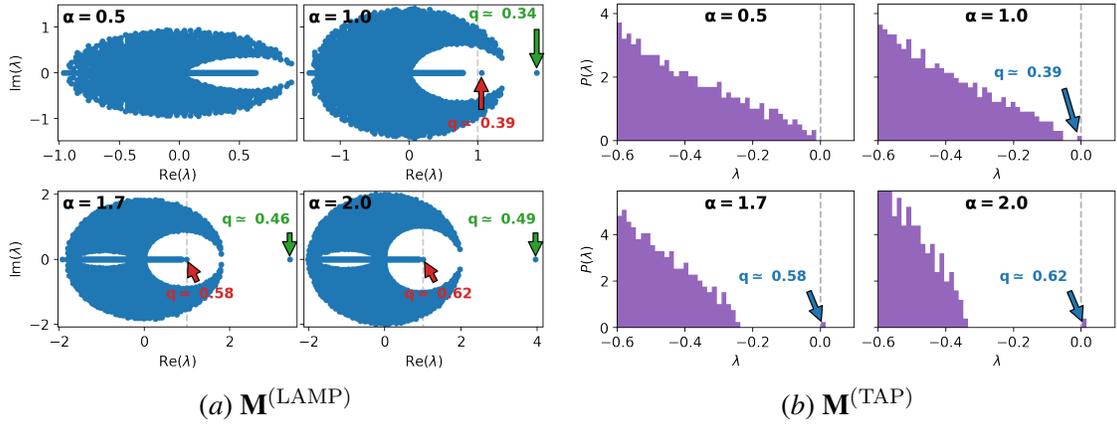
Figure 7: Transition in the spectra of $\mathbf{M}^{(\mathrm{LAMP})}$ and $\mathbf{M}^{(\mathrm{TAP})}$ for $\mathbf{\Phi}$ being the product of two complex Gaussian matrices, and a noiseless phase retrieval channel, for the recovery of a natural image. For $\alpha > \alpha_{\mathrm{WR,Algo}} = 0.5$, we indicate the approximate overlap q corresponding to the relevant eigenvalues.

In this section, we present two additional numerical experiments illustrating the weak-recovery transition in the spectra of $\mathbf{M}^{(\mathrm{TAP})}$ and $\mathbf{M}^{(\mathrm{LAMP})}$. These figures are very similar to Fig. 3 in the main text. Namely, in Fig. 6, we consider noiseless phase retrieval with a complex Gaussian matrix, and in Fig. 7 we consider noiseless phase retrieval with a product of two complex Gaussian matrices, and a real image signal, detailed in Section 3.3.

C.2. The performance of the spectral initialization used in gradient descent

In this section, we show the MSE achieved by a combination of our spectral methods and a gradient descent procedure for the recovery of the real image shown in Fig. 5. The results are given in Fig. 8. The gradient descent procedure allows a significant improvement of the performance when the spectral method already achieves reasonably low error. In particular, it is able to reach perfect recovery at finite α, which is not possible via the vanilla spectral methods.

Appendix D. Some technicalities

D.1. The linear variations of the scalar parameters

For any quantity r, we write δr its linear variation around the trivial fixed point. One obtains the following set of equations, using the symmetry of P_0 and P_{out}:

$$
\begin{cases}
\delta v_1^t = -\frac{\beta}{2}\delta\gamma_1^t \int_{\mathbb{K}} P_0(\mathrm{d}x)\,[|x|^4 - \rho^2], & \delta\gamma_2^t = -\frac{1}{\rho^2}\delta v_1^t - \delta\gamma_1^t, \\[2mm]
\delta c_1^t = \frac{\beta}{2}\left[\frac{\rho^2\langle\lambda\rangle_\nu^2}{\alpha^2} - \frac{1}{m}\sum_{\mu=1}^m \frac{\int_{\mathbb{K}}\mathrm{d}z\,|z|^4\,P_{\mathrm{out}}(y_\mu|z)\,e^{-\frac{\beta\alpha|z|^2}{2\rho\langle\lambda\rangle_\nu}}}{\int_{\mathbb{K}}\mathrm{d}z\,P_{\mathrm{out}}(y_\mu|z)\,e^{-\frac{\beta\alpha|z|^2}{2\rho\langle\lambda\rangle_\nu}}}\right]\delta\tau_1^t, & \delta\tau_2^t = -\frac{\alpha^2}{\rho^2\langle\lambda\rangle_\nu^2}\delta c_1^t - \delta\tau_1^t, \\[2mm]
\delta v_2^t = -\rho^2\delta\gamma_2^t - \rho^2\langle\lambda\rangle_\nu\delta\tau_2^t, & \delta\gamma_1^{t+1} = -\frac{1}{\rho^2}\delta v_2^t - \delta\gamma_2^t, \\[2mm]
\delta c_2^t = -\frac{\langle\lambda\rangle_\nu}{\alpha}\rho^2\delta\gamma_2^t - \frac{\rho^2}{\alpha}\langle\lambda^2\rangle_\nu\delta\tau_2^t, & \delta\tau_1^{t+1} = -\frac{\alpha^2}{\rho^2\langle\lambda\rangle_\nu^2}\delta c_2^t - \delta\tau_2^t.
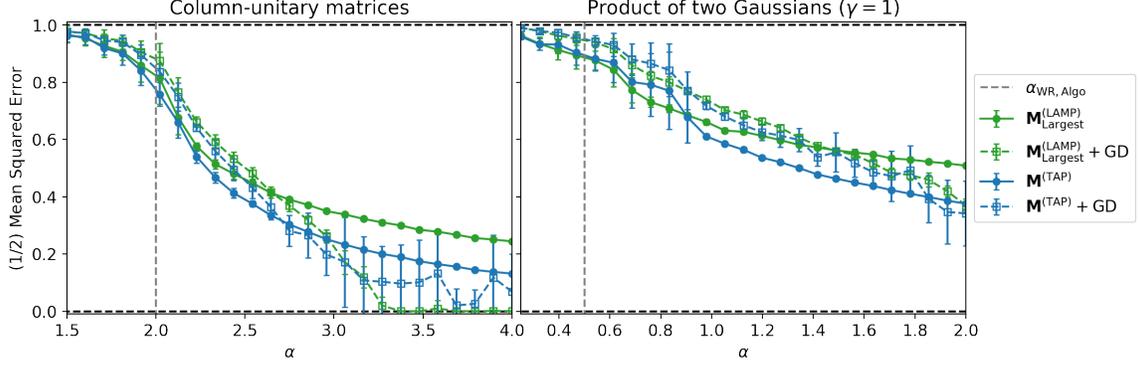\end{cases}
\tag{28}
$$

Figure 8: Mean squared error achieved for the reconstruction of a real image in noiseless phase retrieval with partial DFT (left) and product of Gaussians (right) sensing matrices. We compare the performance of the vanilla spectral methods and of a gradient descent procedure initialized at the spectral estimator. The image size was reduced from 1280×820 to 128×82. The error bars are taken over 3 instances for each of the 3 RGB channels.

Note that the linear variations of these scalar variance parameters do not depend on the variations of the vector parameters of Algorithm 1. Differentiating eq. (11) with respect to τ_1^t and taking it at the trivial fixed point implies:

$$\frac{1}{m}\sum_{\mu=1}^{m}\left[\frac{\int_{\mathbb{K}}\mathrm{d}z\,|z|^4\,P_{\mathrm{out}}(y_\mu|z)\,e^{-\frac{\beta\alpha}{2\rho\langle\lambda\rangle_\nu}|z|^2}}{\int_{\mathbb{K}}\mathrm{d}z\,P_{\mathrm{out}}(y_\mu|z)\,e^{-\frac{\beta\alpha}{2\rho\langle\lambda\rangle_\nu}|z|^2}}-\left\{\frac{\int_{\mathbb{K}}\mathrm{d}z\,|z|^2\,P_{\mathrm{out}}(y_\mu|z)\,e^{-\frac{\beta\alpha}{2\rho\langle\lambda\rangle_\nu}|z|^2}}{\int_{\mathbb{K}}\mathrm{d}z\,P_{\mathrm{out}}(y_\mu|z)\,e^{-\frac{\beta\alpha}{2\rho\langle\lambda\rangle_\nu}|z|^2}}\right\}^2\right]=\frac{2\rho^2\langle\lambda\rangle_\nu^2}{\beta\alpha^2}. \quad (29)$$

Using this relation, one obtains from eq. (28) that $\delta c_1^t = -\delta\tau_1^t\rho^2\langle\lambda\rangle_\nu^2/\alpha^2$, which then implies $\delta\tau_2^t = 0$. Similarly, it follows easily by the remaining equations that all the variations in eq. (28) must be zero.

D.2. The expansion of $F(x,y)$ around $y=0$

We describe here the behavior of $F(x,y)$ as $x>0$ and $y\to 0^+$. Let us write the equations satisfied by ζ_x,ζ_y:

$$\begin{cases}\left\langle\dfrac{\zeta_y}{\zeta_x\zeta_y+\lambda}\right\rangle_\nu = x, & (30a)\\[4mm]\dfrac{\alpha-1}{\zeta_y}+\left\langle\dfrac{\zeta_x}{\zeta_x\zeta_y+\lambda}\right\rangle_\nu = \alpha y. & (30b)\end{cases}$$

As $y\to 0^+$, this implies necessarily that $\zeta_y\to+\infty$, and one finds easily that $\zeta_y\sim 1/y$, $\zeta_x\sim 1/x$. We now turn to the next order variations, that we write as:

$$\zeta_y = y^{-1} + c_1 + \mathcal{O}(y),$$
$$\zeta_x = \frac{1}{x} + c_2 y + \mathcal{O}(y^2).$$

We use eq. (30) to compute $c_1 = -x \langle \lambda \rangle_\nu / \alpha$ and $c_2 = - \langle \lambda \rangle_\nu$. We can then develop the logarithmic potential:

$$\frac{1}{2} \langle \log(\zeta_x \zeta_y + \lambda) \rangle_\nu = -\frac{1}{2} \log y - \frac{1}{2} \log x - \frac{x}{2\alpha} \langle \lambda \rangle_\nu \, y + \mathcal{O}(y^2).$$

Developing the other terms involved in $F(x, y)$ is straightforward and yields:

$$F(x, y) = -\frac{xy}{2} \langle \lambda \rangle_\nu + \mathcal{O}(y^2). \tag{31}$$

One can push this analysis to the next order, and finds in the exact same way, from eq. (30):

$$\zeta_y = \frac{1}{y} - \frac{x \langle \lambda \rangle_\nu}{\alpha} + \frac{x^2}{\alpha^2} \left[\alpha \langle \lambda^2 \rangle_\nu - (1 + \alpha) \langle \lambda \rangle_\nu^2 \right] y + \mathcal{O}(y^2),$$

$$\zeta_x = \frac{1}{x} - \langle \lambda \rangle_\nu \, y + \frac{x}{\alpha} \left[\alpha \langle \lambda^2 \rangle_\nu - (1 + \alpha) \langle \lambda \rangle_\nu^2 \right] y^2 + \mathcal{O}(y^3).$$

This yields for $F(x, y)$:

$$F(x, y) = -\frac{\langle \lambda \rangle_\nu}{2} xy + \frac{x^2}{4\alpha} \left[\alpha \langle \lambda^2 \rangle_\nu - (1 + \alpha) \langle \lambda \rangle_\nu^2 \right] y^2 + \mathcal{O}(y^3),$$

which concludes our analysis.

D.3. Proof of Proposition 3

Let us recall the two spectral methods $\mathbf{M}^{(\mathrm{TAP})}$, $\mathbf{M}^{(\mathrm{LAMP})}$. Without loss of generality, we assume $\langle \lambda \rangle_\nu = \alpha$. Recall that we defined $z_\mu \equiv \partial_\omega g_{\mathrm{out}}(y_\mu, 0, \rho)$. We let $\mathbf{Z} = \mathrm{Diag}(z_\mu)$. We can thus write:

$$\begin{cases} \mathbf{M}^{(\mathrm{LAMP})} = \rho \Big(\dfrac{\mathbf{\Phi}\mathbf{\Phi}^\dagger}{n} - \mathbb{1}_m \Big) \mathbf{Z}, & (32a) \\[2ex] \mathbf{M}^{(\mathrm{TAP})} = -\dfrac{1}{\rho} \mathbb{1}_n + \dfrac{1}{n} \mathbf{\Phi}^\dagger \dfrac{\mathbf{Z}}{\mathbb{1}_m + \rho \mathbf{Z}} \mathbf{\Phi}. & (32b) \end{cases}$$

We start by the first claim. By definition of $(\lambda_{\mathrm{LAMP}}, \mathbf{v})$, we have

$$\rho \frac{\mathbf{\Phi}\mathbf{\Phi}^\dagger}{n} \mathbf{Z}\mathbf{v} = (\rho \mathbf{Z} + \lambda_{\mathrm{LAMP}}) \mathbf{v}. \tag{33}$$

Since we assumed $\lambda_{\mathrm{LAMP}} + \rho z_\mu \neq 0$ for all μ, this implies that $\mathbf{\Phi}^\dagger \mathbf{Z}\mathbf{v} \neq 0$, and we thus let

$$\hat{\mathbf{x}} \equiv \frac{\mathbf{\Phi}^\dagger \mathbf{Z}\mathbf{v}}{\| \mathbf{\Phi}^\dagger \mathbf{Z}\mathbf{v} \|} \sqrt{n}.$$

Multiplying eq. (33) by $\mathbf{\Phi}^\dagger \mathbf{Z} (\lambda_{\mathrm{LAMP}} + \rho \mathbf{Z})^{-1}$ on both sides, we directly reach the sought result:

$$\Big\{ \frac{1}{n} \mathbf{\Phi}^\dagger \frac{\rho \mathbf{Z}}{\lambda_{\mathrm{LAMP}} + \rho \mathbf{Z}} \mathbf{\Phi} \Big\} \hat{\mathbf{x}} = \hat{\mathbf{x}}.$$

We move on to the second claim. Let $\mathbf{x} \in \mathbb{K}^n$ be an eigenvector of $\mathbf{M}^{(\text{TAP})}$ with norm $\|\mathbf{x}\|^2 = n$, with associated eigenvalue λ_{TAP}. We let:

$$\mathbf{u} \equiv \frac{\mathbb{1}_m}{\mathbb{1}_m + \rho \mathbf{Z}} \frac{\boldsymbol{\Phi}}{\sqrt{n}} \mathbf{x}.$$

And we can then easily compute:

$$
\begin{aligned}
\mathbf{M}^{(\text{LAMP})}\mathbf{u} &= \rho\Big(\frac{\boldsymbol{\Phi}\boldsymbol{\Phi}^\dagger}{n} - \mathbb{1}_m\Big) \frac{\mathbf{Z}}{\mathbb{1}_m + \rho \mathbf{Z}} \frac{\boldsymbol{\Phi}}{\sqrt{n}} \mathbf{x}, \\
&= \frac{\rho \boldsymbol{\Phi}}{\sqrt{n}} \Big[\mathbf{M}^{(\text{TAP})} + \frac{1}{\rho}\mathbb{1}_n\Big]\mathbf{x} - \rho \mathbf{Z}\mathbf{u}, \\
&= \rho \lambda_{\text{TAP}} \frac{\boldsymbol{\Phi}}{\sqrt{n}}\mathbf{x} + \frac{\boldsymbol{\Phi}}{\sqrt{n}}\mathbf{x} - \rho \mathbf{Z}\mathbf{u}, \\
\mathbf{M}^{(\text{LAMP})}\mathbf{u} &= \mathbf{u} + \rho \lambda_{\text{TAP}}(\mathbb{1}_m + \rho \mathbf{Z})\mathbf{u}. &\text{(34)}
\end{aligned}
$$

At $\alpha = \alpha_{\text{WR,Algo}}$, the largest eigenvalue of $\mathbf{M}^{(\text{TAP})}$ concentrates on 0, which corresponds to the onset of marginal instability of the trivial local maximum. As one can see from eq. (34), this implies that $\mathbf{M}^{(\text{LAMP})}$ also possesses an eigenvalue equal to 1 at $\alpha = \alpha_{\text{WR,Algo}}$, indicating marginal instability as well. To put it shortly, *the two spectral methods have the same weak recovery threshold.* Moreover, eq. (34) implies that for any $\alpha \geq \alpha_{\text{WR,Algo}}$, if $\mathbf{M}^{(\text{TAP})}$ has en eigenvalue that concentrates on 0 as $n \to \infty$, then $\mathbf{M}^{(\text{LAMP})}$ has a corresponding eigenvalue concentrating on 1, and *with the same performance.* Indeed, as described in eq. (17), the estimator associated to $\mathbf{M}^{(\text{LAMP})}$ will be given by:

$$\hat{\mathbf{x}}_{\text{LAMP}} \propto \frac{\boldsymbol{\Phi}^\dagger}{\sqrt{n}} \mathbf{Z}\mathbf{u} = \frac{\boldsymbol{\Phi}^\dagger}{\sqrt{n}} \frac{\mathbf{Z}}{\mathbb{1}_m + \rho \mathbf{Z}} \frac{\boldsymbol{\Phi}}{\sqrt{n}} \hat{\mathbf{x}}_{\text{TAP}},$$

in which $\hat{\mathbf{x}}$ is an eigenvector of $\mathbf{M}^{(\text{TAP})}$ with eigenvalue 0. Therefore, we reach that $\hat{\mathbf{x}}_{\text{LAMP}} \propto \hat{\mathbf{x}}_{\text{TAP}}$, and these two vectors are thus equal as they are both normalized.