Foundations of Computational Mathematics (2019) 19:703–773
https://doi.org/10.1007/s10208-018-9395-y

FOUNDATIONS OF COMPUTATIONAL MATHEMATICS
The Journal of the Society for the Foundations of Computational Mathematics

Fundamental Limits of Weak Recovery with Applications to Phase Retrieval

Marco Mondelli[1] · **Andrea Montanari**[2,3]

Received: 2 October 2017 / Revised: 12 July 2018 / Accepted: 18 July 2018 / Published online: 4 September 2018

Abstract
In phase retrieval, we want to recover an unknown signal $\boldsymbol{x} \in \mathbb{C}^d$ from n quadratic measurements of the form $y_i = |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle|^2 + w_i$, where $\boldsymbol{a}_i \in \mathbb{C}^d$ are known sensing vectors and w_i is measurement noise. We ask the following *weak recovery* question: What is the minimum number of measurements n needed to produce an estimator $\hat{\boldsymbol{x}}(\boldsymbol{y})$ that is positively correlated with the signal $\boldsymbol{x}$? We consider the case of Gaussian vectors $\boldsymbol{a}_i$. We prove that—in the high-dimensional limit—a sharp phase transition takes place, and we locate the threshold in the regime of vanishingly small noise. For $n \le d - o(d)$, no estimator can do significantly better than random and achieve a strictly positive correlation. For $n \ge d + o(d)$, a simple spectral estimator achieves a positive correlation. Surprisingly, numerical simulations with the same spectral estimator demonstrate promising performance with realistic sensing matrices. Spectral methods are used to initialize non-convex optimization algorithms in phase retrieval, and our approach can boost the performance in this setting as well. Our impossibility result is based on classical information-theoretic arguments. The spectral algorithm computes the leading eigenvector of a weighted empirical covariance matrix. We obtain a sharp characterization of the spectral properties of this random matrix using tools from free probability and generalizing a recent result by Lu and Li. Both the upper bound and lower bound generalize beyond phase retrieval to measurements y_i produced according to a generalized linear model. As a by-product of our analysis, we compare the threshold of the proposed spectral method with that of a message passing algorithm.

Keywords Spectral initialization · Phase transition · Mutual information · Second-moment method · Phase retrieval

Mathematics Subject Classification 68Q32 · 68T05 · 91E40

Communicated by Emmanuel Candes.

M. Mondelli was supported by an Early Postdoc.Mobility fellowship from the Swiss National Science Foundation. A. Montanari was partially supported by Grants NSF DMS-1613091 and NSF CCF-1714305.

Extended author information available on the last page of the article

1 Introduction

In this work, we consider the problem of recovering a signal $\boldsymbol{x}$ of dimension d, given n *generalized linear measurements*. More specifically, the measurements are taken independently according to the conditional distribution

$$y_i \sim p(y \mid \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle), \qquad i \in \{1, \dots, n\}, \tag{1}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product, $\{\boldsymbol{a}_i\}_{1 \le i \le n}$ is a set of known sensing vector, and $p(\cdot \mid \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle)$ is a known probability density function. This model appears in many problems in signal processing and statistical estimation, e.g., photon-limited imaging [72,81], signal recovery from quantized measurements [62], and phase retrieval [34, 65]. For the problem of *phase retrieval*, the model (1) is specialized to

$$y_i = |\langle \boldsymbol{x}, \boldsymbol{a}_i \rangle|^2 + w_i, \qquad i \in \{1, \dots, n\}, \tag{2}$$

where w_i is noise. Applications of phase retrieval arise in several areas of science and engineering, including X-ray crystallography [38,52], microscopy [51], astronomy [35], optics [76], acoustics [4], interferometry [25], and quantum mechanics [22].

Popular methods to solve the phase retrieval problem are based on semi-definite programming relaxations [14,15,17,75]. However, these algorithms rapidly become prohibitive from a computational point of view when the dimension d of the signal increases, which makes them impractical in most of the real-world applications. For this reason, several algorithms have been developed in order to solve directly the non-convex least-squares problem, including the error reduction schemes dating back to Gerchberg–Saxton and Fienup [34,36], alternating minimization [57], approximate message passing (AMP) [64], Wirtinger Flow [16], iterative projections [47], the Kaczmarz method [80], and a number of other approaches [13,20,33,68,77–79,83]. Furthermore, recently a convex relaxation that operates in the natural domain of the signal was independently proposed by two groups of authors [2,37]. All these techniques require an initialization step, whose goal is to provide a solution $\hat{\boldsymbol{x}}$ that is positively correlated with the unknown signal $\boldsymbol{x}$. To do so, spectral methods are widely employed: The estimate $\hat{\boldsymbol{x}}$ is given by the principal eigenvector of a suitable matrix constructed from the data. A similar strategy (initialization step followed by an iterative algorithm) has proved successful for many other estimation problems, e.g., matrix completion [40,44], blind deconvolution [46,48], sparse coding [1], and joint alignment from pairwise noisy observations [19].

We focus on a regime in which both the number of measurement n and the dimension of the signal d tend to infinity, but their ratio n/d tends to a positive constant δ. The *weak recovery* problem requires to provide an estimate $\hat{\boldsymbol{x}}(\boldsymbol{y})$ that has a positive correlation with the unknown vector $\boldsymbol{x}$:

$$\liminf_{n \to \infty} \mathbb{E}\left\{ \frac{|\langle \hat{\boldsymbol{x}}(\boldsymbol{y}), \boldsymbol{x} \rangle|}{\|\hat{\boldsymbol{x}}(\boldsymbol{y})\|_2 \, \|\boldsymbol{x}\|_2} \right\} > \epsilon, \tag{3}$$

for some $\epsilon > 0$.

In this paper, we consider either $\boldsymbol{x} \in \mathbb{R}^d$ or $\boldsymbol{x} \in \mathbb{C}^d$ and assume that the measurement vectors $\boldsymbol{a}_i$ are standard Gaussian (either real or complex). In the general setting of model (1), we present two types of results:

1. We develop an *information-theoretic lower bound* δ_ℓ: For $\delta < \delta_\ell$, no estimator can output non-trivial estimates. In other words, the weak recovery problem cannot be solved.
2. We establish an *upper bound* δ_{u} based on a *spectral algorithm*: For $\delta > \delta_{\mathrm{u}}$, we can achieve weak recovery [see (3)] by letting $\hat{\boldsymbol{x}}$ be the principal eigenvector of a matrix suitably constructed from the data. We also show that δ_{u} is the optimal threshold for spectral methods.

The values of the thresholds δ_ℓ and δ_{u} depend on the conditional distribution $p(\cdot \mid \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle)$. For the special case of phase retrieval [see (2)], we evaluate these bounds and we show that they coincide in the limit of vanishing noise.

Theorem *Let $\boldsymbol{x}$ be uniformly distributed on the d-dimensional complex sphere with radius $\sqrt{d}$ and assume that $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{i.i.d.} \mathsf{CN}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$. Let $\boldsymbol{y} \in \mathbb{R}^n$ be given by (2), with $\{w_i\}_{1\le i\le n} \sim \mathsf{N}(0, \sigma^2)$, and $n, d \to \infty$ with $n/d \to \delta \in (0, +\infty)$. Then,*

- *For $\delta < 1$, no algorithm can provide non-trivial estimates on $\boldsymbol{x}$;*
- *For $\delta > 1$, there exists $\sigma_0(\delta) > 0$ and a spectral algorithm that returns an estimate $\hat{\boldsymbol{x}}$ satisfying (3), for any $\sigma \in [0, \sigma_0(\delta)]$.*

The assumption that $\boldsymbol{x}$ is uniform on the sphere can be dropped for the upper bound part. We also show that $\sigma_0(\delta)$ scales as $\sqrt{\delta - 1}$ when δ is close to 1. In the 'real case' $\boldsymbol{x} \in \mathbb{R}^d$ with $\|\boldsymbol{x}\|_2^2 = d$ and $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{i.i.d.} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$, we prove that an analogous result holds and that the threshold moves from 1 to 1/2. This is reminiscent of how the injectivity thresholds are $\delta = 4$ and $\delta = 2$ in the complex and the real case, respectively [4,5,21]. A possible intuition for this halving phenomenon comes from the fact that the complex problem has twice as many variables but the same amount of equations of the real problem. Hence, it is reasonable that the complex case requires twice the amount of data with respect to the real case.

Let us emphasize that we are considering the problem of weak recovery. Therefore, we may need less than n samples in order to obtain positive correlation on n unknowns. For instance, in the linear case $y_i = \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle + w_i$, weak recovery is possible for any $\delta > 0$. Consequently, it is not surprising that for phase retrieval in the real case weak recovery can be achieved for δ below one.

Our information-theoretic lower bound is proved by estimating the conditional entropy via the second-moment method. In general, this might not match the spectral upper bound. We provide an example in which there is a strictly positive gap between δ_ℓ and δ_{u} in Remark 3 at the end of Sect. 3.

As in earlier work (see Sect. 1.1), we consider spectral algorithms that compute the eigenvector corresponding to the largest eigenvalue of a matrix of the form:

$$\boldsymbol{D}_n = \frac{1}{n} \sum_{i=1}^{n} \mathcal{T}(y_i) \boldsymbol{a}_i \boldsymbol{a}_i^* , \tag{4}$$

where $\mathcal{T} : \mathbb{R} \to \mathbb{R}$ is a pre-processing function. For δ large enough (and a suitable choice of $\mathcal{T}$), we expect the resulting eigenvector $\hat{\boldsymbol{x}}(\boldsymbol{y})$ to be positively correlated with the true signal $\boldsymbol{x}$. The recent paper [49] computed exactly the threshold value δ_{u}, under the assumption that the measurement vectors are real Gaussian, and $\mathcal{T}$ is nonnegative.

Here, we generalize the result of [49] by removing the assumption that $\mathcal{T}(y) \geq 0$ and by considering the complex case. The main technical lemma of this generalization consists in the computation of the largest eigenvalue of a matrix of the form $\boldsymbol{U}\boldsymbol{M}_n\boldsymbol{U}^*$, where the entries of $\boldsymbol{U}$ are $\sim_{i.i.d.}$ $\mathsf{CN}(0, 1)$ and $\boldsymbol{M}_n$ is independent of $\boldsymbol{U}$ and has known empirical spectral distribution. The case in which $\boldsymbol{M}_n$ is PSD is handled in [3]. In this paper, by using tools from free probability, we solve the case in which $\boldsymbol{M}_n$ is not necessarily PSD. To do so, it is not sufficient to compute the weak limit of the empirical spectral distribution of $\boldsymbol{U}\boldsymbol{M}_n\boldsymbol{U}^*$, but we also need to compute the almost sure limit of its principal eigenvalue. Armed with this result, we compute the optimal pre-processing function $\mathcal{T}^*_\delta(y)$ for the general model (1). This pre-processing function is optimal in the sense that it provides the smallest possible weak recovery threshold for the spectral method. Our upper bound δ_u is the phase transition location for this optimal spectral method. In the case of phase retrieval (as $\sigma \to 0$), the optimal pre-processing function is given by

$$\mathcal{T}^*_\delta(y) = \frac{y-1}{y+\sqrt{\delta}-1}, \tag{5}$$

and achieves weak recovery for any $\delta > \delta_u = 1$. In the limit $\delta \downarrow 1$, this converges to the limiting function $\mathcal{T}^*(y) = 1 - (1/y)$.

While expression (5) is remarkably simple, it is somewhat counterintuitive. Earlier methods [16,18,49] use $\mathcal{T}(y) \geq 0$ and try to extract information from the large values of y_i. Function (5) has a large negative part for small y, in particular when δ is close to 1. Furthermore, it extracts useful information from data points with y_i small. One possible interpretation is that the points in which the measurement vector is basically orthogonal to the unknown signal are not informative; hence, we penalize them.

Our analysis applies to Gaussian measurement matrices. However, the proposed spectral method works well also on real images and realistic measurement matrices. To illustrate this fact, in Fig. 1 we test our algorithm on a digital photograph of the painting "The birth of Venus" by Sandro Botticelli. We consider a type of measurements that falls under the category of coded diffraction patterns (CDP) [15,20]: The measurement matrix is given by the product of δ copies of a Fourier matrix and a diagonal matrix with entries i.i.d. and uniform in $\{1, -1, i, -i\}$, where i denotes the imaginary unit. We compare our method with the truncated spectral initialization proposed in [20], which consists in discarding the measurements larger than an assigned threshold and leaving the others untouched. The proposed choice of the pre-processing function allows to recover a good estimate of the original image already when $\delta = 4$, while the truncated spectral initialization of [20] requires $\delta = 12$ to obtain similar results.

In general, our proposed spectral method can be thought of as a first step of the following two-round algorithm: First, use spectral initialization to perform weak recovery and then improve the solution with an iterative algorithm, e.g., AMP or Wirtinger Flow. By using optimal truncation methods, the weak recovery threshold is smaller, which

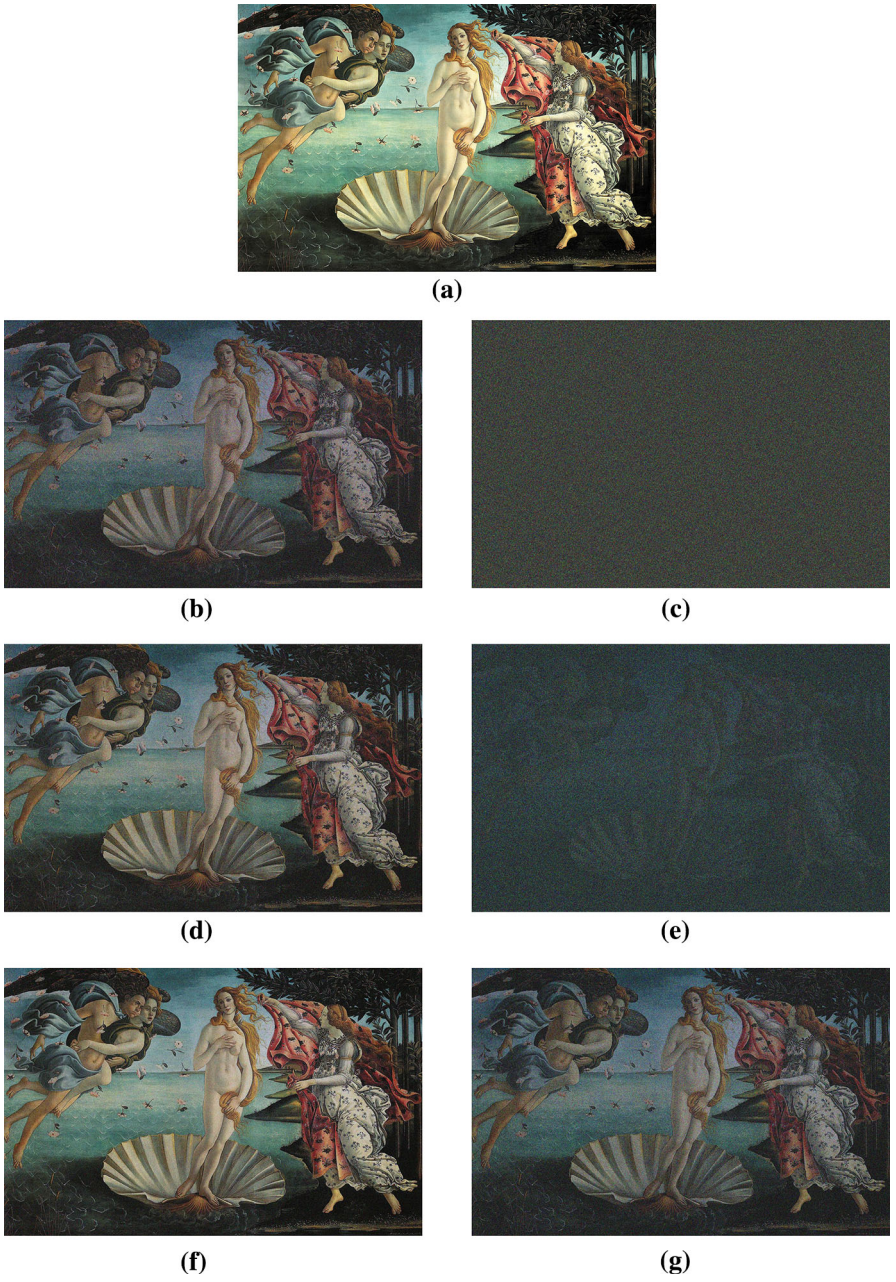

Fig. 1 Performance comparison between the proposed spectral method and the truncated spectral initialization of [20] for the recovery of a digital photograph from coded diffraction patterns. **a** Original image, **b** proposed—$\delta = 4$, **c** truncated—$\delta = 4$, **d** proposed—$\delta = 6$, **e** truncated—$\delta = 6$, **f** proposed—$\delta = 12$, **g** truncated—$\delta = 12$

means that less measurements are required in order to successfully complete the first step of the algorithm. If a different truncation is used, the resulting performances are limited by the corresponding weak recovery threshold.

Note that the pre-processing function (5) is optimal in the sense that it minimizes the weak recovery threshold associated with the spectral method. Hence, for a given correlation $\bar{\epsilon} \in (0, 1)$, the exact expression of the optimal pre-processing function that allows to obtain a correlation $\bar{\epsilon}$ between $\hat{\boldsymbol{x}}(\boldsymbol{y})$ and $\boldsymbol{x}$ might be different and it might depend on $\bar{\epsilon}$. However, we observe that (5) provides excellent empirical performance and outperforms state-of-the-art methods for a wide range of target correlations (see the simulation results of Sect. 7).

The rest of the paper is organized as follows. In Sect. 2, after introducing the necessary notation, we define formally the problem. We then state our general information-theoretic lower bound and our spectral upper bound for the case of complex signal $\boldsymbol{x}$ and complex measurement vectors $\boldsymbol{a}_i$. The main results for the real case are stated in Sect. 3. In Sects. 4 and 5, we present the proof of the information-theoretic lower bound and of the spectral upper bound, respectively. In Sect. 6, we compare the spectral approach to a message passing algorithm. In particular, we show that the latter cannot have a better threshold than δ_{u} and that δ_{u} is the threshold for a linearized version of message passing. In Sect. 7, we present some numerical simulations that illustrate the behavior of the proposed spectral method for the phase retrieval problem. The proofs of several results are deferred to the various appendices.

1.1 Related Work

Precise asymptotic information on high-dimensional regression problems has been obtained by several groups in recent years [8,10,31,32,43,58–60,70,71,82]. In particular, information-theoretically optimal estimation was considered for compressed sensing [29] and random linear estimation [7,63]. Minimax optimal estimation is considered, among others, in [31,70,73].

The performance of the spectral methods for phase retrieval was first considered in [57]. In the present notation, [57] uses $\mathcal{T}(y) = y$ and proves that there exists a constant c_1 such that weak recovery can be achieved for $n > c_1 \cdot d \cdot \log^3 d$. The same paper also gives an iterative procedure to improve over the spectral method, but the bottleneck is in the spectral step. The sample complexity of weak recovery using spectral methods was improved to $n > c_2 \cdot d \cdot \log d$ in [16] and then to $n > c_3 \cdot d$ in [20], for some constants c_2 and c_3. Both of these papers also prove guarantees for exact recovery by suitable descent algorithms. The guarantees on the spectral initialization are proved by matrix concentration inequalities, a technique that typically does not return exact threshold values.

In [37], the authors introduce the PhaseMax relaxation and prove an exact recovery result for phase retrieval, which depends on the correlation between the true signal and the initial estimate given to the algorithm. The same idea was independently proposed in [2]. Furthermore, the analysis in [2] allows to use the same set of measurements for both initialization and convex programming, whereas the analysis in [37] requires fresh extra measurements for convex programming. By using our spectral method to

obtain the initial estimate, it should be possible to improve the existing upper bounds on the number of samples needed for exact recovery.

As previously mentioned, our analysis of spectral methods builds on the recent work of Lu and Li [49] that compute the exact spectral threshold for a matrix of the form (4) with $\mathcal{T}(y) \geq 0$. Here, we generalize this result to signed pre-processing functions $\mathcal{T}(y)$ and construct a function of this type that achieves the information-theoretic threshold for phase retrieval. Our proof indeed implies that nonnegative pre-processing functions lead to an unavoidable gap with respect to the ideal threshold.

Finally, while this paper was under completion, two works appeared that address related problems. In [6], the authors characterize the information-theoretically optimal estimation error for a broad class of models of the form (1). However, note that this analysis does not prove—in general—the existence of an efficient estimation algorithm (for instance in the case of phase retrieval). The paper [27] studies the PhaseMax approach [2,37] to phase retrieval and uses the non-rigorous replica method from statistical physics to derive exact thresholds for this algorithm. The rigorous performance analysis of PhaseMax under Gaussian measurements in the large system limit is provided in [28].

2 Main Results: Complex Case

2.1 Notation and System Model

We use $[n]$ as a shortcut for $\{1, \dots, n\}$. We use uppercase letters (e.g., X, Y, Z, ...) to denote random variables when we are taking operators such as expectation, variance, or mutual information. We denote by $\mathbf{0}_n$ the vector consisting of n 0s. Given a vector $\boldsymbol{x}$, we denote by $\|\boldsymbol{x}\|_2$ its ℓ_2 norm. Given a matrix $\boldsymbol{A}$, we denote by $\|\boldsymbol{A}\|_F$ its Frobenius norm, by $\|\boldsymbol{A}\|_{\text{op}}$ its operator norm, by $\boldsymbol{A}^\mathsf{T}$ its transpose, and by $\boldsymbol{A}^*$ its conjugate transpose. Given two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^d$, we denote by $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^d x_i y_i^*$ their scalar product. We take logarithms in the natural basis and we measure entropies in nats. Given $c \in \mathbb{C}$, we denote by $\Re(c)$ and $\Im(c)$ its real and imaginary part, respectively. We use $\xrightarrow{\mathcal{P}}$ and $\xrightarrow{\text{a.s.}}$ to denote the convergence in probability and the almost sure convergence, respectively.

Let $\boldsymbol{x} \in \mathbb{C}^d$ be chosen uniformly at random on the d-dimensional complex sphere with radius $\sqrt{d}$, i.e.,

$$\boldsymbol{x} \sim \text{Unif}(\sqrt{d}\mathsf{S}_{\mathbb{C}}^{d-1}). \tag{6}$$

Let the sensing vectors $\{\boldsymbol{a}_i\}_{1 \leq i \leq n}$, with $\boldsymbol{a}_i \in \mathbb{C}^d$, be independent and identically distributed according to a circularly symmetric complex normal distribution with variance $1/d$, i.e.,

$$\{\boldsymbol{a}_i\}_{1 \leq i \leq n} \sim_{i.i.d.} \mathsf{CN}(\mathbf{0}_d, \boldsymbol{I}_d/d). \tag{7}$$

Given $g_i = \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle$, the vector of measurements $\boldsymbol{y} \in \mathbb{R}^n$ is obtained by drawing each component independently according to the following distribution:

$$y_i \sim p(y \mid |g_i|), \qquad i \in [n]. \tag{8}$$

For the special case of *phase retrieval*, the measurements are given by the squared scalar product corrupted by additive Gaussian noise with variance σ^2, i.e.,

$$p_{\mathrm{PR}}(y \mid |g_i|) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y - |g_i|^2)^2}{2\sigma^2}\right). \tag{9}$$

Let $\delta_n = n/d$ and assume that, as $n \to \infty$, $\delta_n \to \delta$ for some $\delta \in (0, \infty)$.

2.2 Information-Theoretic Lower Bound

The main result of this section establishes the following: There is a critical value δ_ℓ such that, for any $\delta < \delta_\ell$, the optimal estimator has the same performance as a trivial estimator that does not have access to any measurement. The value of δ_ℓ depends on the distribution (8) of the measurements, and we provide an expression to compute it.

In order to state formally the result, we need to introduce a few definitions. Consider the function $f : [0, 1] \to \mathbb{R}$, given by

$$f(m) = \int_{\mathbb{R}} \frac{\mathbb{E}_{G_1,G_2}\{p(y \mid |G_1|)p(y \mid |G_2|)\}}{\mathbb{E}_G\{p(y \mid |G|)\}}\, \mathrm{d}y, \tag{10}$$

with

$$G \sim \mathsf{CN}(0, 1), \qquad (G_1, G_2) \sim \mathsf{CN}\left(\mathbf{0}_2, \begin{bmatrix} 1 & c \\ c^* & 1 \end{bmatrix}\right), \tag{11}$$

and $m = |c|^2$. Note that the RHS of (10) depends only on $m = |c|^2$. Indeed, by applying the transformation $(G_1, G_2) \to (e^{i\theta_1}G_1, e^{i\theta_2}G_2)$, $f(m)$ does not change, but the correlation coefficient c is mapped into $ce^{i(\theta_1-\theta_2)}$. A more explicit formula for $f(m)$ is provided by Lemma 6 in Appendix A. The function $f(m)$ is related to the conditional entropy $H(Y_1, \ldots, Y_n \mid \boldsymbol{A}_1, \ldots, \boldsymbol{A}_n)$, as clarified in the proof of Lemma 1 in Sect. 4.1. Furthermore, set

$$F_\delta(m) = \delta \log f(m) + \log(1 - m). \tag{12}$$

Note that when $m = 0$, G_1 and G_2 are independent. Hence, $f(0) = 1$, which implies that $F_\delta(0) = 0$ for any $\delta > 0$. We define the information-theoretic threshold δ_ℓ as the largest value of δ such that the maximum of $F_\delta(m)$ is attained at $m = 0$, i.e.,

$$\delta_\ell = \sup\{\delta \mid F_\delta(m) < 0 \text{ for } m \in (0, 1]\}. \tag{13}$$

Let us now define the error metric. The setting is the following: We observe the vector of n measurements $\boldsymbol{y}$ and, given a new sensing vector $\boldsymbol{a}_{n+1}$, we want to estimate some function $\phi(|\langle \boldsymbol{x}, \boldsymbol{a}_{n+1}\rangle|)$ given by

$$\phi(|\langle \boldsymbol{x}, \boldsymbol{a}_{n+1}\rangle|) = \int_{\mathbb{R}} \varphi(y) p(y \mid |\langle \boldsymbol{x}, \boldsymbol{a}_{n+1}\rangle|)\, \mathrm{d}y. \tag{14}$$

Then, the minimum mean square error is defined as

$$\mathsf{MMSE}(\delta_n) = \mathbb{E}\left\{\left(\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|) - \mathbb{E}\left\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|) \,\middle|\, \boldsymbol{Y}, \{\boldsymbol{A}_i\}_{1\le i\le n}\right\}\right)^2\right\}, \tag{15}$$

where $\mathbb{E}\left\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|) \mid \boldsymbol{Y}, \{\boldsymbol{A}_i\}_{1\le i\le n}\right\}$ represents the optimal estimator of the quantity $\phi(|\langle \boldsymbol{x}, \boldsymbol{a}_{n+1}\rangle|)$ and the expectation of the square error is to be intended over all the randomness of the system, i.e., over $\boldsymbol{X}, \boldsymbol{A}_{n+1}, \boldsymbol{Y}$, and $\{\boldsymbol{A}_i\}_{1\le i\le n}$. Note that this error metric depends on the choice of the function ϕ. Furthermore, observe that if we do not have access to the vector of measurements $\boldsymbol{Y}$, the trivial estimator $\mathbb{E}\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|)\}$ has a mean square error given by

$$\mathbb{E}\left\{\left(\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|) - \mathbb{E}\left\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|)\right\}\right)^2\right\} = \mathrm{Var}\left\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|)\right\}. \tag{16}$$

At this point, we are ready to state our main result, which is proved in Sect. 4.1.

Theorem 1 (Information-theoretic lower bound for general complex sensing model) *Let $\boldsymbol{x}$, $\{\boldsymbol{a}_i\}_{1\le i\le n+1}$, and $\boldsymbol{y}$ be distributed according to (6), (7), and (8), respectively. Let $n/d \to \delta$ and define δ_ℓ as in (13). Furthermore, assume that the function φ that appears in (14) is bounded. Then, for any $\delta < \delta_\ell$, we have that*

$$\lim_{n\to\infty} \mathsf{MMSE}(\delta_n) = \mathrm{Var}\left\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|)\right\}. \tag{17}$$

Let us point out that the requirement that the function φ is bounded can be relaxed when the tails of the distribution of $\boldsymbol{Y}$ are sufficiently light (e.g., sub-Gaussian). Indeed, this is what happens for the special case of phase retrieval, which is considered immediately below.

For the special case of *phase retrieval*, a more explicit error metric is given by the matrix minimum mean square error, defined as

$$\mathsf{MMSE}_{\mathrm{PR}}(\delta_n) = \frac{1}{d^2}\mathbb{E}\left\{\left\|\boldsymbol{X}\boldsymbol{X}^* - \mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^* \mid \boldsymbol{Y}, \{\boldsymbol{A}_i\}_{1\le i\le n}\right\}\right\|_F^2\right\}. \tag{18}$$

Indeed, the vector $\boldsymbol{x}$ can be recovered only up to a sign change, since we observe a function of the scalar products $|\langle \boldsymbol{x}, \boldsymbol{a}_i\rangle|$. Clearly, $\mathsf{MMSE}(\delta_n) \in [0, 1]$ and $\mathsf{MMSE}(\delta_n) = 1$ implies that the optimal estimator coincides with the trivial estimator that outputs the all-0 vector.

The corollary below provides the exact value of δ_ℓ for the case of phase retrieval, and it is proved in Appendix A.

Corollary 1 (Information-theoretic lower bound for phase retrieval) *Let $\boldsymbol{x}$, $\{\boldsymbol{a}_i\}_{1\le i\le n}$, and $\boldsymbol{y}$ be distributed according to (6), (7), and (9), respectively. Let $n/d \to \delta$. Then, for any $\delta < 1$, we have that*

$$\lim_{\sigma\to 0}\lim_{n\to\infty} \mathsf{MMSE}_{\mathrm{PR}}(\delta_n) = 1. \tag{19}$$

2.3 Upper Bound via Spectral Method

The main result of this section establishes the following: There is a critical value δ_{u} such that, for any $\delta > \delta_{\mathrm{u}}$, the principal eigenvector of a suitably constructed matrix, call it $\boldsymbol{D}_n$, provides an estimate $\hat{\boldsymbol{x}}$ that satisfies (3). The threshold δ_{u} is defined as

$$\delta_{\mathrm{u}} = \frac{1}{\displaystyle\int_{\mathbb{R}} \frac{\left(\mathbb{E}_G\left\{p(y \mid |G|)(|G|^2-1)\right\}\right)^2}{\mathbb{E}_G\left\{p(y \mid |G|)\right\}}\,\mathrm{d}y}, \tag{20}$$

with $G \sim \mathsf{CN}(0,1)$. Given the measurements $\{y_i\}_{1\le i\le n}$, we construct the matrix $\boldsymbol{D}_n$ as

$$\boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^*, \tag{21}$$

where $\mathcal{T}:\mathbb{R}\to\mathbb{R}$ is a pre-processing function.

At this point, we are ready to state our main result, which is proved in Sect. 5.

Theorem 2 (Spectral upper bound for complex general sensing model) *Let $\boldsymbol{x}$, $\{\boldsymbol{a}_i\}_{1\le i\le n}$, and $\boldsymbol{y}$ be distributed according to (6), (7), and (8), respectively. Let $n/d \to \delta$ and define δ_{u} as in (20). Let $\hat{\boldsymbol{x}}$ be the principal eigenvector of the matrix $\boldsymbol{D}_n$ defined in (21). For any $\delta > \delta_{\mathrm{u}}$, set the pre-processing function $\mathcal{T}$ to the function $\mathcal{T}_\delta^*$ given by*

$$\mathcal{T}_\delta^*(y) = \frac{\sqrt{\delta_{\mathrm{u}}}\cdot\mathcal{T}^*(y)}{\sqrt{\delta} - (\sqrt{\delta}-\sqrt{\delta_{\mathrm{u}}})\mathcal{T}^*(y)}, \tag{22}$$

where

$$\mathcal{T}^*(y) = 1 - \frac{\mathbb{E}_G\left\{p(y \mid |G|)\right\}}{\mathbb{E}_G\left\{p(y \mid |G|)\cdot|G|^2\right\}}. \tag{23}$$

Then, we have that, almost surely,

$$\lim_{n\to\infty} \frac{|\langle\hat{\boldsymbol{x}},\boldsymbol{x}\rangle|}{\|\hat{\boldsymbol{x}}\|_2\,\|\boldsymbol{x}\|_2} > \epsilon, \tag{24}$$

for some $\epsilon > 0$. Furthermore, for any $\delta \le \delta_{\mathrm{u}}$, there is no pre-processing function $\mathcal{T}$ such that, almost surely, (24) holds.

Let us highlight that the pre-processing function (22) provides the optimal threshold among spectral methods that use matrices of the form (4) in the sense that it achieves weak recovery for $\delta > \delta_{\mathrm{u}}$ and no function achieves weak recovery for $\delta \le \delta_{\mathrm{u}}$. Note

also that the assumption that $\boldsymbol{x}$ is uniform on the sphere can be dropped (see the beginning of the proof of Lemma 2 in Sect. 5).

As a by-product of our analysis, we also give guarantees on the value of δ sufficient to achieve an assigned correlation with the ground truth, using the spectral method, see (84) in the statement of Lemma 2 in Sect. 5. Hence, we can combine our upper bound with existing non-convex optimization algorithms, in order to obtain provable performance guarantees.

The corollary below provides the exact value of δ_{u} and an explicit expression for $\mathcal{T}_\delta^*(y)$ for the case of phase retrieval. Its proof is contained in Appendix B. Note that, for phase retrieval, $\delta_{\mathrm{u}} = \delta_\ell = 1$, i.e., the spectral upper bound matches the information-theoretic lower bound.

Corollary 2 (Spectral upper bound for phase retrieval) *Let $\boldsymbol{x}$, $\{\boldsymbol{a}_i\}_{1\le i\le n}$, and $\boldsymbol{y}$ be distributed according to (6), (7), and (9), respectively. Let $n/d \to \delta$. Let $\hat{\boldsymbol{x}}$ be the principal eigenvector of the matrix $\boldsymbol{D}_n$ defined in (21). For any $\delta > 1$, set the pre-processing function $\mathcal{T}$ to the function $\mathcal{T}_\delta^*$ given by (with $y_+ \equiv \max(0, y)$):*

$$\mathcal{T}_\delta^*(y) = \frac{y_+ - 1}{y_+ + \sqrt{\delta} - 1}. \tag{25}$$

Then, we have that, almost surely,

$$\lim_{\sigma\to 0}\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}, \boldsymbol{x}\rangle|}{\|\hat{\boldsymbol{x}}\|_2 \, \|\boldsymbol{x}\|_2} > \epsilon, \tag{26}$$

for some $\epsilon > 0$.

Notice that this statement is stronger than the claim that $\delta_{\mathrm{u}}(\sigma^2) \to 1$ as $\sigma^2 \to 0$, where $\delta_{\mathrm{u}}(\sigma^2)$ is the spectral threshold at noise level σ^2. Indeed, it requires proving that the scalar product $|\langle \hat{\boldsymbol{x}}, \boldsymbol{x}\rangle|$ stays bounded away from 0, as $\sigma^2 \to 0$. Furthermore, this is achieved with the pre-processing function (25) that does not require to estimate σ, which can be challenging with real data.

We also characterize the scaling between δ_{u} and σ^2 when σ^2 is close to 0: $\delta_{\mathrm{u}}(\sigma^2) = 1 + \sigma^2 + o(\sigma^2)$ (see Lemma 8 in Appendix B).

3 Main Results: Real Case

Let us now briefly discuss what happens in the real case. Let $\boldsymbol{x} \in \mathbb{R}^d$ be chosen uniformly at random on the d-dimensional real sphere with radius $\sqrt{d}$, i.e.,

$$\boldsymbol{x} \sim \mathrm{Unif}(\sqrt{d}\mathsf{S}_{\mathbb{R}}^{d-1}). \tag{27}$$

Let the sensing vectors $\{\boldsymbol{a}_i\}_{1\le i\le n}$, with $\boldsymbol{a}_i \in \mathbb{R}^d$ being independent and identically distributed according to a normal distribution with zero mean and variance $1/d$, i.e.,

$$\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{i.i.d.} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d). \tag{28}$$

Given $g_i = \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle$, the vector of measurements $\boldsymbol{y} \in \mathbb{R}^n$ is obtained by drawing each component independently according to the following distribution:

$$y_i \sim p(y \mid g_i), \qquad i \in [n]. \tag{29}$$

We can define the *"real" phase retrieval* model, whereby the measurements are given by the squared scalar product corrupted by additive Gaussian noise with variance σ^2, i.e.,

$$p_{\mathrm{PR}}(y \mid g_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(y - g_i^2)^2}{2\sigma^2}\right). \tag{30}$$

We first present the information-theoretic lower bound. Consider the function $f : [-1, 1] \to \mathbb{R}$, given by

$$f(m) = \int_{\mathbb{R}} \frac{\mathbb{E}_{G_1,G_2}\{p(y \mid G_1)p(y \mid G_2)\}}{\mathbb{E}_G\{p(y \mid G)\}}\,\mathrm{d}y, \tag{31}$$

with

$$G \sim \mathsf{N}(0, 1), \qquad (G_1, G_2) \sim \mathsf{N}\left(\mathbf{0}_2, \begin{bmatrix} 1 & m \\ m & 1 \end{bmatrix}\right). \tag{32}$$

Furthermore, set

$$F_\delta(m) = \delta \log f(m) + \frac{1}{2}\log(1 - m^2). \tag{33}$$

Again, $F_\delta(0) = 0$ for any $\delta > 0$. We define the information-theoretic threshold δ_ℓ as the largest value of δ such that the maximum of $F_\delta(m)$ is attained at $m = 0$, i.e.,

$$\delta_\ell = \sup\{\delta \mid F_\delta(m) < 0 \text{ for } m \in [-1, 1]\backslash\{0\}\}. \tag{34}$$

As for the error metric, we observe the vector of n measurements $\boldsymbol{y}$ and, given a new sensing vector $\boldsymbol{a}_{n+1}$, we want to estimate some function $\phi(\langle \boldsymbol{x}, \boldsymbol{a}_{n+1}\rangle)$ given by

$$\phi(\langle \boldsymbol{x}, \boldsymbol{a}_{n+1}\rangle) = \int_{\mathbb{R}} \varphi(y)p(y \mid \langle \boldsymbol{x}, \boldsymbol{a}_{n+1}\rangle)\mathrm{d}y. \tag{35}$$

Then, the minimum mean square error is defined as

$$\mathsf{MMSE}(\delta_n) = \mathbb{E}\left\{\left(\phi(\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle) - \mathbb{E}\left\{\phi(\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle) \,\middle|\, \boldsymbol{Y}, \{\boldsymbol{A}_i\}_{1\le i\le n}\right\}\right)^2\right\}. \tag{36}$$

Recall that, if we do not have access to the vector of measurements $\boldsymbol{y}$, the trivial estimator $\mathbb{E}\{\phi(\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle)\}$ has a mean square error given by

$$\mathbb{E}\left\{\left(\phi(\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle) - \mathbb{E}\left\{\phi(\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle)\right\}\right)^2\right\} = \mathrm{Var}\left\{\phi(\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle)\right\}. \tag{37}$$

At this point, we are ready to state the information-theoretic lower bound, which is proved in Sect. 4.2.

Theorem 3 (Information-theoretic lower bound for real general sensing model) *Let* $\boldsymbol{x}$, $\{\boldsymbol{a}_i\}_{1\le i\le n+1}$, *and* $\boldsymbol{y}$ *be distributed according to* (27), (28), *and* (29), *respectively. Let* $n/d \to \delta$ *and define* δ_ℓ *as in* (34). *Furthermore, assume that the function* φ *that appears in* (35) *is bounded. Then, for any* $\delta < \delta_\ell$, *we have that*

$$\lim_{n\to\infty} \mathsf{MMSE}(\delta_n) = \mathrm{Var}\left\{\phi(\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle)\right\}. \tag{38}$$

Remark 1 (Information-theoretic lower bound for real phase retrieval) For the special case of *phase retrieval*, a more explicit error metric is given by the matrix minimum mean square error, defined as

$$\mathsf{MMSE}_{\mathrm{PR}}(\delta_n) = \frac{1}{d^2}\mathbb{E}\left\{\left\|\boldsymbol{X}\boldsymbol{X}^{\mathsf{T}} - \mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^{\mathsf{T}} \mid \boldsymbol{Y}, \{\boldsymbol{A}_i\}_{1\le i\le n}\right\}\right\|_F^2\right\}. \tag{39}$$

By calculations similar to those in Lemma 7 contained in Appendix A, one can prove that, if the distribution $p(\cdot \mid G)$ appearing in (31) is given by (30), then

$$\lim_{\sigma\to 0} \delta_\ell(\sigma^2) = 1/2. \tag{40}$$

Consequently, by following a proof analogous to that of Corollary 1 in Appendix A, we conclude that, for any $\delta < 1/2$,

$$\lim_{\sigma\to 0}\lim_{n\to\infty} \mathsf{MMSE}_{\mathrm{PR}}(\delta_n) = 1. \tag{41}$$

Let us now move to the spectral upper bound. The threshold δ_{u} is defined as

$$\delta_{\mathrm{u}} = \frac{1}{\displaystyle\int_{\mathbb{R}} \frac{\left(\mathbb{E}_G\left\{p(y \mid G)(G^2-1)\right\}\right)^2}{\mathbb{E}_G\left\{p(y \mid G)\right\}}\,\mathrm{d}y}, \tag{42}$$

with $G \sim \mathsf{N}(0,1)$. Given the measurements $\{y_i\}_{1\le i\le n}$, we construct the matrix $\boldsymbol{D}_n$ as

$$\boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^{\mathsf{T}}, \tag{43}$$

where $\mathcal{T} : \mathbb{R} \to \mathbb{R}$ is a pre-processing function.

The proof of the following spectral upper bound is discussed in Remark 7 at the end of Sect. 5.

Theorem 4 (Spectral upper bound for real general sensing model) *Let $\boldsymbol{x}$, $\{\boldsymbol{a}_i\}_{1\le i\le n}$, and $\boldsymbol{y}$ be distributed according to (27), (28), and (29), respectively. Let $n/d \to \delta$ and define δ_{u} as in (42). Let $\hat{\boldsymbol{x}}$ be the principal eigenvector of the matrix $\boldsymbol{D}_n$ defined in (43). For any $\delta > \delta_{\mathrm{u}}$, set the pre-processing function $\mathcal{T}$ to the function $\mathcal{T}_\delta^*$ given by*

$$\mathcal{T}_\delta^*(y) = \frac{\sqrt{\delta_{\mathrm{u}}}\cdot \mathcal{T}^*(y)}{\sqrt{\delta} - (\sqrt{\delta} - \sqrt{\delta_{\mathrm{u}}})\mathcal{T}^*(y)}, \tag{44}$$

where

$$\mathcal{T}^*(y) = 1 - \frac{\mathbb{E}_G\{p(y \mid G)\}}{\mathbb{E}_G\left\{p(y \mid G)\cdot G^2\right\}}. \tag{45}$$

Then, we have that, almost surely,

$$\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}, \boldsymbol{x}\rangle|}{\|\hat{\boldsymbol{x}}\|_2\, \|\boldsymbol{x}\|_2} > \epsilon, \tag{46}$$

for some $\epsilon > 0$. Furthermore, for any $\delta \le \delta_{\mathrm{u}}$, there is no pre-processing function $\mathcal{T}$ such that, almost surely, (46) holds.

Remark 2 (Spectral upper bound for real phase retrieval) By calculations similar to those in Lemma 8 contained in Appendix B, one can prove that, if the distribution $p(\cdot \mid G)$ appearing in (31) is given by (30), then

$$\lim_{\sigma\to 0} \delta_{\mathrm{u}}(\sigma^2) = 1/2. \tag{47}$$

Furthermore, by following a proof analogous to that of Corollary 2 in Appendix B, one can prove the following result. For any $\delta > 1/2$, set the pre-processing function $\mathcal{T}$ to the function $\mathcal{T}_\delta^*$ given by (with $y_+ = \max(y, 0)$)

$$\mathcal{T}_\delta^*(y) = \frac{y_+ - 1}{y_+ + \sqrt{2\delta} - 1}. \tag{48}$$

Then, we have that, almost surely,

$$\lim_{\sigma\to 0}\lim_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}, \boldsymbol{x}\rangle|}{\|\hat{\boldsymbol{x}}\|_2\, \|\boldsymbol{x}\|_2} > \epsilon, \tag{49}$$

for some $\epsilon > 0$. Note that, for real phase retrieval, the spectral upper bound matches the information-theoretic lower bound.

In the following remark, we provide an example in which there is a strictly positive gap between δ_ℓ and δ_{u}.

Remark 3 (Gap between δ_ℓ and δ_u) Let us define

$$H(a) = \mathbb{E}_G \left\{ \tanh^2(a\, G)\, (G^2 - 1) \right\}, \tag{50}$$

where $G \sim \mathrm{N}(0, 1)$. Note that $H(0) = 0$ and $\lim_{a\to\infty} H(a) = 0$. Hence, there exists $a_2 > a_1$ such that $H(a_1) = H(a_2)$.

Consider the following distribution for the components of the vector of measurements $\boldsymbol{y}$:

$$p(y \mid g) = \begin{cases} \tanh^2(a_2\, g) - \tanh^2(a_1\, g), & \text{for } y \in [1, 2], \\ 1 - (\tanh^2(a_2\, g) - \tanh^2(a_1\, g)), & \text{for } y \in [-2, -1]. \end{cases} \tag{51}$$

Then, we have that, for any $y \in \mathbb{R}$,

$$\mathbb{E}_G \left\{ p(y \mid G)\, (G^2 - 1) \right\} = 0, \tag{52}$$

which, by definition (42), immediately implies that $\delta_u = \infty$. Note that this argument works when we substitute $\tanh^2(x)$ with any function which is even, increasing for $x \geq 0$ and bounded between 0 and 1.

Let us now show that δ_ℓ is finite. Consider the function $f(m)$ defined in (31). As previously mentioned, $f(0) = 1$. Furthermore,

$$\begin{aligned} f(1) &= \int_{\mathbb{R}} \frac{\mathbb{E}_G \left\{ (p(y \mid G))^2 \right\}}{\mathbb{E}_G \left\{ p(y \mid G) \right\}} \mathrm{d}y \\ &= \int_{\mathbb{R}} \frac{(\mathbb{E}_G \left\{ p(y \mid G) \right\})^2 + \mathrm{Var} \left\{ p(y \mid G) \right\}}{\mathbb{E}_G \left\{ p(y \mid G) \right\}} \mathrm{d}y \\ &= 1 + \int_{\mathbb{R}} \frac{\mathrm{Var} \left\{ p(y \mid G) \right\}}{\mathbb{E}_G \left\{ p(y \mid G) \right\}} \mathrm{d}y > 1. \end{aligned} \tag{53}$$

Consequently, there exists $m_* \in (0, 1)$ such that $f(m_*) > 1$. Set

$$\delta^* = -\frac{\log(1 - m_*^2)}{2 \log f(m_*)} + 1. \tag{54}$$

Then, we have that, for any $\delta \geq \delta^*$,

$$F_\delta(m_*) \geq F_{\delta_*}(m_*) = 1 > 0. \tag{55}$$

Hence, by definition (34), we conclude that $\delta_\ell < \delta^*$, which implies that δ_ℓ is finite. Note that this upper bound on δ_ℓ applies to any $p(y \mid G)$ which is not constant in G on a set of positive measure. As a result, there is a strictly positive gap between δ_ℓ and δ_u.[1]

[1] This gap is not due to the looseness of our lower bound. Indeed, by using the result of [6], one can show that the actual information-theoretic threshold is finite.

4 Proof of Theorems 1 and 3: Information-Theoretic Lower Bound

4.1 Complex Case

The crucial point of the proof consists in the computation of the conditional entropy $H(\boldsymbol{Y} \mid \boldsymbol{A})$, which is contained in Lemma 1. Then, we use this result to compute the mutual information for the considered model. Finally, we provide the proof of Theorem 1.

Lemma 1 (Conditional entropy) *Let* $\boldsymbol{x} \sim \mathrm{Unif}(\sqrt{d}\mathsf{S}_{\mathbb{C}}^{d-1})$, $\boldsymbol{A} = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)$ *with* $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{i.i.d.} \mathsf{CN}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$, *and* $\boldsymbol{y} = (y_1, \dots, y_n)$ *with* $y_i \sim p(\cdot \mid |g_i|)$ *and* $g_i = \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle$. *Let* $n/d \to \delta$ *and define* δ_ℓ *as in* (13). *Then, for any* $\delta < \delta_\ell$, *we have that*

$$\lim_{n\to\infty} \frac{1}{n} H(\boldsymbol{Y} \mid \boldsymbol{A}) = H(Y_1). \tag{56}$$

Proof We divide the proof into two steps. The *first step* consists in showing that

$$-\frac{1}{n}\left(\int_{\mathbb{R}^d} \frac{\mathbb{E}_{\boldsymbol{A}}\left\{(p(\boldsymbol{y} \mid \boldsymbol{A}))^2\right\}}{\mathbb{E}_{\boldsymbol{A}}\left\{p(\boldsymbol{y} \mid \boldsymbol{A})\right\}} \mathrm{d}\boldsymbol{y} - 1\right) \le \frac{1}{n} H(\boldsymbol{Y} \mid \boldsymbol{A}) - H(Y_1) \le 0, \tag{57}$$

which holds for all $n \in \mathbb{N}$ and for all $\delta > 0$. The proof of (57) does not require any assumption on the distribution of $\boldsymbol{x}$ and on the distribution of $\{\boldsymbol{a}_i\}_{1\le i\le n}$ (as long as the vectors $\{\boldsymbol{a}_i\}_{1\le i\le n}$ are independent).

The *second step* consists in showing that

$$\lim_{n\to+\infty} \frac{1}{n}\left(\int_{\mathbb{R}^d} \frac{\mathbb{E}_{\boldsymbol{A}}\left\{(p(\boldsymbol{y} \mid \boldsymbol{A}))^2\right\}}{\mathbb{E}_{\boldsymbol{A}}\left\{p(\boldsymbol{y} \mid \boldsymbol{A})\right\}} \mathrm{d}\boldsymbol{y} - 1\right) = 0. \tag{58}$$

It is clear that (57) and (58) imply the thesis.

First step. By definition of conditional entropy, we have that

$$\frac{1}{n} H(\boldsymbol{Y} \mid \boldsymbol{A}) = \frac{1}{n} \int \mathbb{E}_{\boldsymbol{A}}\left\{-p(\boldsymbol{y} \mid \boldsymbol{A}) \log p(\boldsymbol{y} \mid \boldsymbol{A})\right\} \mathrm{d}\boldsymbol{y}. \tag{59}$$

By using the definition of y_i and the fact that they are independent, we can rewrite $\mathbb{E}_{\boldsymbol{A}}\{p(\boldsymbol{y} \mid \boldsymbol{A})\}$ as follows

$$\begin{aligned} \mathbb{E}_{\boldsymbol{A}}\{p(\boldsymbol{y} \mid \boldsymbol{A})\} &= \mathbb{E}_{\boldsymbol{A},\boldsymbol{X}}\{p(\boldsymbol{y} \mid \boldsymbol{A}, \boldsymbol{X})\} \\ &= \mathbb{E}_{\boldsymbol{A},\boldsymbol{X}}\left\{\prod_{i=1}^{n} p(y_i \mid |\langle \boldsymbol{X}, \boldsymbol{A}_i \rangle|)\right\} \end{aligned} \tag{60}$$

$$= \prod_{i=1}^{n} \mathbb{E}_{G_i} \{p(y_i \mid |G_i|)\},$$

where we set $G_i = \langle \boldsymbol{X}, \boldsymbol{A}_i \rangle$.

Let us now give an upper bound on the RHS of (59):

$$\begin{aligned}
&\frac{1}{n} \int_{\mathbb{R}^d} \mathbb{E}_{\boldsymbol{A}} \{-p(\boldsymbol{y} \mid \boldsymbol{A}) \log p(\boldsymbol{y} \mid \boldsymbol{A})\} \, \mathrm{d}\boldsymbol{y} \\
&\overset{\text{(a)}}{\leq} \frac{1}{n} \int_{\mathbb{R}^d} -\mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\} \log \mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\} \, \mathrm{d}\boldsymbol{y} \\
&\overset{\text{(b)}}{=} \frac{1}{n} \int_{\mathbb{R}^d} -\prod_{i=1}^{n} \mathbb{E}_{G_i} \{p(y_i \mid |G_i|)\} \sum_{j=1}^{n} \log \mathbb{E}_{G_j} \left\{p(y_j \mid G_j)\right\} \, \mathrm{d}\boldsymbol{y} \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}} -\mathbb{E}_{G_i} \{p(y_i \mid |G_i|)\} \log \mathbb{E}_{G_i} \{p(y_i \mid |G_i|)\} \, \mathrm{d}y_i \\
&= H(Y_1),
\end{aligned}$$

where in (a) we apply Jensen's inequality as the function $g(x) = -x \log x$ is concave, and in (b) we use (60). This immediately implies that

$$\frac{1}{n} H(\boldsymbol{Y} \mid \boldsymbol{A}) - H(Y_1) \leq 0. \tag{61}$$

Note that the upper bound (61) is based on the inequality

$$\mathbb{E}_{\boldsymbol{A}} \{-p(\boldsymbol{y} \mid \boldsymbol{A}) \log p(\boldsymbol{y} \mid \boldsymbol{A})\} - (-\mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\} \log \mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\}) \leq 0.$$

Let us now find a lower bound to this quantity:

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{A}} \{-p(\boldsymbol{y} \mid \boldsymbol{A}) \log p(\boldsymbol{y} \mid \boldsymbol{A})\} - (-\mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\} \log \mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\}) \\
&= \mathbb{E}_{\boldsymbol{A}} \left\{-p(\boldsymbol{y} \mid \boldsymbol{A}) \log \frac{p(\boldsymbol{y} \mid \boldsymbol{A})}{\mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\}}\right\} \\
&\overset{\text{(a)}}{=} \mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\} \, \mathbb{E}_{\boldsymbol{Z}} \{-\boldsymbol{Z} \log \boldsymbol{Z}\} \\
&\overset{\text{(b)}}{=} \mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\} \, \mathbb{E}_{\boldsymbol{Z}} \{-\boldsymbol{Z} \log \boldsymbol{Z} + \boldsymbol{Z} - 1\} \\
&\overset{\text{(c)}}{\geq} -\mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\} \, \mathbb{E}_{\boldsymbol{Z}} \left\{(\boldsymbol{Z} - 1)^2\right\} \\
&\overset{\text{(d)}}{=} -\mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\} \left(\mathbb{E}_{\boldsymbol{Z}} \left\{\boldsymbol{Z}^2\right\} - 1\right) \\
&= -\left(\frac{\mathbb{E}_{\boldsymbol{A}} \left\{(p(\boldsymbol{y} \mid \boldsymbol{A}))^2\right\}}{\mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\}} - \mathbb{E}_{\boldsymbol{A}} \{p(\boldsymbol{y} \mid \boldsymbol{A})\}\right),
\end{aligned} \tag{62}$$

where in (a) we set $\boldsymbol{Z} = p(\boldsymbol{y} \mid \boldsymbol{A})/\mathbb{E}_{\boldsymbol{A}}\{p(\boldsymbol{y} \mid \boldsymbol{A})\}$, in (b) we use that $\mathbb{E}_{\boldsymbol{Z}}\{\boldsymbol{Z}\} = 1$, in (c) we use that $-z\log z + z - 1 \geq -(z-1)^2$ for any $z \geq 0$, and in (d) we use again that $\mathbb{E}_{\boldsymbol{Z}}\{\boldsymbol{Z}\} = 1$. Therefore,

$$\begin{aligned}\frac{1}{n}H(\boldsymbol{Y} \mid \boldsymbol{A}) - H(Y_1) &= \frac{1}{n}\int \left(\mathbb{E}_{\boldsymbol{A}}\{-p(\boldsymbol{y} \mid \boldsymbol{A})\log p(\boldsymbol{y} \mid \boldsymbol{A})\}\right.\\ &\qquad \left. - \left(-\mathbb{E}_{\boldsymbol{A}}\{p(\boldsymbol{y} \mid \boldsymbol{A})\}\log \mathbb{E}_{\boldsymbol{A}}\{p(\boldsymbol{y} \mid \boldsymbol{A})\}\right)\right)\mathrm{d}\boldsymbol{y}\\ &\overset{(a)}{\geq} -\frac{1}{n}\int_{\mathbb{R}^d}\left(\frac{\mathbb{E}_{\boldsymbol{A}}\left\{(p(\boldsymbol{y} \mid \boldsymbol{A}))^2\right\}}{\mathbb{E}_{\boldsymbol{A}}\{p(\boldsymbol{y} \mid \boldsymbol{A})\}} - \mathbb{E}_{\boldsymbol{A}}\{p(\boldsymbol{y} \mid \boldsymbol{A})\}\right)\mathrm{d}\boldsymbol{y},\\ &\overset{(b)}{=} -\frac{1}{n}\left(\int_{\mathbb{R}^d}\frac{\mathbb{E}_{\boldsymbol{A}}\left\{(p(\boldsymbol{y} \mid \boldsymbol{A}))^2\right\}}{\mathbb{E}_{\boldsymbol{A}}\{p(\boldsymbol{y} \mid \boldsymbol{A})\}}\mathrm{d}\boldsymbol{y} - 1\right),\end{aligned}$$

where in (a) we use (62) and in (b) we use that the integral of $p(\boldsymbol{y} \mid \boldsymbol{A})$ is 1. This concludes the proof of (57).

Second step. As $\boldsymbol{X} \sim \mathrm{Unif}(\sqrt{d}\mathsf{S}_{\mathbb{C}}^{d-1})$ and $\boldsymbol{A}_i \sim \mathsf{CN}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$, we have that

$$\{G_i\}_{1\leq i\leq n} \sim_{i.i.d.} \mathsf{CN}(0,1).$$

Let us rewrite the quantity $\mathbb{E}_{\boldsymbol{A}}\left\{(p(\boldsymbol{y} \mid \boldsymbol{A}))^2\right\}$ as follows:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{A}}\left\{(p(\boldsymbol{y} \mid \boldsymbol{A}))^2\right\} &= \mathbb{E}_{\boldsymbol{A}}\left\{\left(\mathbb{E}_{\boldsymbol{X}}\left\{\prod_{i=1}^{n} p(y_i \mid |\langle \boldsymbol{X}, \boldsymbol{A}_i\rangle|)\right\}\right)^2\right\}\\ &\overset{(a)}{=} \mathbb{E}_{\boldsymbol{A}}\left\{\mathbb{E}_{\boldsymbol{X}_1,\boldsymbol{X}_2}\left\{\prod_{i=1}^{n} p(y_i \mid |\langle \boldsymbol{X}_1, \boldsymbol{A}_i\rangle|)\cdot p(y_i \mid |\langle \boldsymbol{X}_2, \boldsymbol{A}_i\rangle|)\right\}\right\}\\ &\overset{(b)}{=} \mathbb{E}_{C}\left\{\prod_{i=1}^{n}\mathbb{E}_{G_{i,1},G_{i,2}}\left\{p(y_i \mid |G_{i,1}|)\cdot p(y_i \mid |G_{i,2}|)\right\}\right\},\end{aligned} \tag{63}$$

where in (a) $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent and in (b) we set $G_{i,1} = \langle \boldsymbol{X}_1, \boldsymbol{A}_i\rangle$, $G_{i,2} = \langle \boldsymbol{X}_2, \boldsymbol{A}_i\rangle$, and

$$C = \frac{\langle \boldsymbol{X}_1, \boldsymbol{X}_2\rangle}{\|\boldsymbol{X}_1\|_2\,\|\boldsymbol{X}_2\|_2}.$$

Then, given $C = c$, as $\boldsymbol{X}_1, \boldsymbol{X}_2 \sim_{i.i.d.} \mathrm{Unif}(\sqrt{d}\mathsf{S}_{\mathbb{C}}^{d-1})$ and $\boldsymbol{A}_i \sim \mathsf{CN}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$, we have that

$$\{(G_{i,1}, G_{i,2})\}_{1\leq i\leq n} \sim_{i.i.d.} \mathsf{CN}\left(\boldsymbol{0}_2, \begin{bmatrix}1 & c\\ c^* & 1\end{bmatrix}\right).$$

Hence,

$$\begin{aligned}\frac{1}{n}\int_{\mathbb{R}^d}\frac{\mathbb{E}_{\boldsymbol{A}}\left\{(p(\boldsymbol{y}\mid\boldsymbol{A}))^2\right\}}{\mathbb{E}_{\boldsymbol{A}}\left\{p(\boldsymbol{y}\mid\boldsymbol{A})\right\}}\,\mathrm{d}\boldsymbol{y} &\overset{(a)}{=} \frac{1}{n}\int_{\mathbb{R}^d}\mathbb{E}_C\left\{\prod_{i=1}^{n}\frac{\mathbb{E}_{G_{i,1},G_{i,2}}\left\{p(y\mid|G_{i,1}|)p(y\mid|G_{i,2}|)\right\}}{\mathbb{E}_{G_i}\left\{p(y\mid|G_i|)\right\}}\right\}\mathrm{d}\boldsymbol{y}\\ &= \frac{1}{n}\mathbb{E}_C\left\{\prod_{i=1}^{n}\int_{\mathbb{R}}\frac{\mathbb{E}_{G_{i,1},G_{i,2}}\left\{p(y\mid|G_{i,1}|)p(y\mid|G_{i,2}|)\right\}}{\mathbb{E}_{G_i}\left\{p(y\mid|G_i|)\right\}}\mathrm{d}y_i\right\}\\ &\overset{(b)}{=} \frac{1}{n}\mathbb{E}_M\left\{(f(M))^n\right\}\\ &\overset{(c)}{=} \frac{d-1}{n}\int_0^1 (f(m))^n(1-m)^{d-2}\,\mathrm{d}m,\end{aligned}$$

where in (a) we use (60) and (63); in (b) we use the fact that f depends only on $m=|c|^2$, which is clear from the explicit expression provided by Lemma 6 contained in Appendix A; and in (c) we use that $M\sim\mathrm{Beta}(1,d-1)$ by Lemma 9 contained in Appendix C.

Set $d'=d-2$ and $\delta'_n=n/d'$. Thus,

$$\int_0^1 (f(m))^n(1-m)^{d-2}\,\mathrm{d}m=\int_0^1\exp\left(n\cdot F_{\delta'_n}(m)\right)\,\mathrm{d}m, \tag{64}$$

where $F_{\delta'_n}(m)$ is given by (12). Define

$$\tilde{F}_\delta(m)=\delta\max(\log f(m),0)+\log(1-m). \tag{65}$$

As $\delta<\delta_\ell$ and $n/d'\to\delta$, there exists $\delta_*\in(\delta,\delta_\ell)$ such that $\delta'_n<\delta_*$ for n sufficiently large. As $F_{\delta'_n}(m)\le\tilde{F}_{\delta'_n}(m)$ and $\tilde{F}_\delta(m)$ is non-decreasing in δ, we have that

$$\int_0^1\exp\left(n\cdot F_{\delta'_n}(m)\right)\,\mathrm{d}m\le\int_0^1\exp\left(n\cdot\tilde{F}_{\delta_*}(m)\right)\,\mathrm{d}m. \tag{66}$$

Note that $\tilde{F}_{\delta_*}(m)<0$ if and only if $F_{\delta_*}(m)<0$. Thus, by definition of δ_ℓ, we have that $\tilde{F}_{\delta_*}(m)<0$ for $m\in(0,1]$ when n is sufficiently large. Furthermore, $\tilde{F}_{\delta_*}(0)=0$ and $\tilde{F}_{\delta_*}$ is a continuous function. As a result, by Lemma 11, the integral in (66) tends to 0 as $n\to\infty$ and the claim immediately follows. □

Remark 4 (Mutual information) An immediate consequence of Lemma 1 is that one can compute the mutual information $I(\boldsymbol{X};\boldsymbol{Y},\boldsymbol{A})$ for any $\delta<\delta_\ell$:

$$\lim_{n\to+\infty}\frac{1}{n}I(\boldsymbol{X};\boldsymbol{Y},\boldsymbol{A})=H\left(\mathbb{E}_G\left\{p(\cdot\mid|G|)\right\}\right)-\mathbb{E}_G\left\{H(p(\cdot\mid|G|))\right\}, \tag{67}$$

where $G\sim\mathsf{CN}(0,1)$.

Proof (Proof of Theorem 1) Define $\boldsymbol{y}_{1:n}=(y_1,\ldots,y_n)$ and $\boldsymbol{a}_{1:n}=(\boldsymbol{a}_1,\ldots,\boldsymbol{a}_n)$. We divide the proof into two steps. The *first step* consists in showing that the mutual

information between the next observation y_{n+1} and the previous observations $\boldsymbol{y}_{1:n}$ tends to 0. More formally, we will prove that

$$I(Y_{n+1}; \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n} \mid \boldsymbol{A}_{n+1}) = o_n(1). \tag{68}$$

The *second step* consists in showing that the estimate obtained on $\phi(|\langle \boldsymbol{x}, \boldsymbol{a}_{n+1}\rangle|)$ given the observations $\boldsymbol{y}_{1:n}$ is similar to the estimate on $\phi(|\langle \boldsymbol{x}, \boldsymbol{a}_{n+1}\rangle|)$ when no observation is available. This means that the observations $\boldsymbol{y}_{1:n}$ do not provide any help. More formally, we will prove that

$$\mathbb{E}_{\boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n+1}} \left\{ \left(\mathbb{E}\left\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|)\right\} - \mathbb{E}\left\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|) \,\middle|\, \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n}\right\} \right)^2 \right\} = o_n(1), \tag{69}$$

where ϕ is defined in (14).

Furthermore, we have that

$$\begin{aligned} &\mathbb{E}\left\{ \left(\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|) - \mathbb{E}\left\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|) \,\middle|\, \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n}\right\} \right)^2 \right\} \\ &\quad - \left(\mathbb{E}\left\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|)\right\} - \mathbb{E}\left\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|) \,\middle|\, \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n}\right\} \right)^2 \\ &= \mathbb{E}\left\{ \left(\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|) - \mathbb{E}\left\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|)\right\} \right)^2 \right\} \\ &= \mathrm{Var}\left\{\phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1}\rangle|)\right\}. \end{aligned} \tag{70}$$

By applying (69) and (70), the proof of Theorem 1 follows.

First step. By using the chain rule of entropy and that y_i is independent from $\boldsymbol{a}_{i+1:n+1}$, we obtain that

$$\begin{aligned} \frac{1}{n+1} H(\boldsymbol{Y}_{1:n+1} \mid \boldsymbol{A}_{1:n+1}) &= \frac{1}{n+1} \sum_{i=1}^{n+1} H(Y_i \mid \boldsymbol{Y}_{1:i-1}, \boldsymbol{A}_{1:n+1}) \\ &= \frac{1}{n+1} \sum_{i=1}^{n+1} H(Y_i \mid \boldsymbol{Y}_{1:i-1}, \boldsymbol{A}_{1:i}). \end{aligned}$$

The sequence $s_n = H(Y_n \mid \boldsymbol{Y}_{1:n-1}, \boldsymbol{A}_{1:n})$ is decreasing, as conditioning reduces entropy. Hence, s_n has a limit, and this limit must be equal to $H(Y_1)$ by Lemma 1. Since the Y_i are i.i.d., we obtain that

$$H(Y_{n+1} \mid \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n+1}) = H(Y_{n+1}) + o_n(1).$$

By using again that conditioning reduces entropy, we also obtain that

$$H(Y_{n+1} \mid \boldsymbol{A}_{n+1}) = H(Y_{n+1}) + o_n(1).$$

By putting these last two equations together, we deduce that (68) holds.

Second step. Given two probability distributions p and q, let $D_{\mathrm{KL}}(p||q)$ and $\|p-q\|_{\mathrm{TV}}$ denote their Kullback–Leibler divergence and their total variation distance, respectively. Then,

$$
\begin{aligned}
&I(Y_{n+1}; \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n} \mid \boldsymbol{A}_{n+1}) \\
&\quad = \mathbb{E}_{\boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n+1}} \{ D_{\mathrm{KL}}(p(y_{n+1} \mid \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n+1}) || p(y_{n+1} \mid \boldsymbol{A}_{n+1})) \} \\
&\quad \overset{\text{(a)}}{\geq} \frac{1}{2} \cdot \mathbb{E}_{\boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n+1}} \left\{ \left(\| p(y_{n+1} \mid \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n+1}) - p(y_{n+1} \mid \boldsymbol{A}_{n+1}) \|_{\mathrm{TV}} \right)^2 \right\} \\
&\quad \overset{\text{(b)}}{\geq} \frac{1}{2K^2} \cdot \mathbb{E}_{\boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n+1}} \left\{ \left(\int_{\mathbb{R}} p(y_{n+1} \mid \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n+1}) \varphi(y_{n+1}) \, \mathrm{d}y_{n+1} \right. \right. \\
&\qquad \left. \left. - \int_{\mathbb{R}} p(y_{n+1} \mid \boldsymbol{A}_{n+1}) \varphi(y_{n+1}) \, \mathrm{d}y_{n+1} \right)^2 \right\} \qquad (71) \\
&\quad \overset{\text{(c)}}{=} \frac{1}{2K^2} \cdot \mathbb{E}_{\boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n+1}} \left\{ \left(\int_{\mathbb{C}^d} p(\boldsymbol{x} \mid \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n}) \int_{\mathbb{R}} p(y_{n+1} \mid \boldsymbol{x}, \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n+1}) \varphi(y_{n+1}) \, \mathrm{d}y_{n+1} \, \mathrm{d}\boldsymbol{x} \right. \right. \\
&\qquad \left. \left. - \int_{\mathbb{C}^d} p(\boldsymbol{x}) \int_{\mathbb{R}} p(y_{n+1} \mid \boldsymbol{x}, \boldsymbol{A}_{n+1}) \varphi(y_{n+1}) \, \mathrm{d}y_{n+1} \, \mathrm{d}\boldsymbol{x} \right)^2 \right\} \\
&\quad \overset{\text{(d)}}{=} \frac{1}{2K^2} \cdot \mathbb{E}_{\boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n+1}} \left\{ \left(\mathbb{E} \{ \phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1} \rangle|) \} - \mathbb{E} \{ \phi(|\langle \boldsymbol{X}, \boldsymbol{A}_{n+1} \rangle|) \mid \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n} \} \right)^2 \right\},
\end{aligned}
$$

where in (a) we use Pinsker's inequality; in (b) we use that φ is bounded and we set $\|\varphi\|_\infty = K$; in (c) we use that $\boldsymbol{X}$ and $\boldsymbol{A}_{n+1}$ are independent; and in (d) we use definition (14). By combining (68) and (71), (69) immediately follows. □

4.2 Real Case

The proof is very similar to the one provided in Sect. 4.1 for the complex case. In particular, the crucial point consists in showing that

$$
\lim_{n \to \infty} \frac{1}{n} H(\boldsymbol{Y} \mid \boldsymbol{A}) = H(Y_1), \qquad (72)
$$

where $\boldsymbol{x} \sim \mathrm{Unif}(\sqrt{d}\mathsf{S}_{\mathbb{R}}^{d-1})$, $\boldsymbol{a} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)$ with $\{\boldsymbol{a}_i\}_{1 \leq i \leq n} \sim_{i.i.d.} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$, and $\boldsymbol{y} = (y_1, \ldots, y_n)$ with $y_i \sim p(\cdot \mid g_i)$ and $g_i = \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle$. Then, the proof of Theorem 3 follows similar passages as the proof of Theorem 1.

In order to prove (72), we show that (57) and (58) hold. The proof of (57) follows the same passages as the first step of the proof of Lemma 1; hence, it is omitted. The proof of (58) is slightly different, and we detail what changes in the remaining part of this section.

Similarly to (60), we have that

$$
\mathbb{E}_A \{ p(\boldsymbol{y} \mid \boldsymbol{A}) \} = \prod_{i=1}^{n} \mathbb{E}_{G_i} \{ p(y_i \mid G_i) \},
$$

where $G_i = \langle \boldsymbol{X}, \boldsymbol{A}_i \rangle \sim \mathsf{N}(0, 1)$. Furthermore, similarly to (63), we also have that

$$\mathbb{E}_{\boldsymbol{A}} \left\{ (p(\boldsymbol{y} \mid \boldsymbol{A}))^2 \right\} = \mathbb{E}_M \left\{ \prod_{i=1}^{n} \mathbb{E}_{G_{i,1}, G_{i,2}} \left\{ p(y_i \mid G_{i,1}) \cdot p(y_i \mid G_{i,2}) \right\} \right\},$$

where $G_{i,1} = \langle \boldsymbol{X}_1, \boldsymbol{A}_i \rangle$, $G_{i,2} = \langle \boldsymbol{X}_2, \boldsymbol{A}_i \rangle$, and we define

$$M = \frac{\langle \boldsymbol{X}_1, \boldsymbol{X}_2 \rangle}{\|\boldsymbol{X}_1\|_2 \, \|\boldsymbol{X}_2\|_2}.$$

Then, given $M = m$, as $\boldsymbol{X}_1, \boldsymbol{X}_2 \sim_{i.i.d.} \mathrm{Unif}(\sqrt{d}\mathsf{S}_{\mathbb{R}}^{d-1})$ and $\boldsymbol{A}_i \sim \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$, we have that

$$\{(G_{i,1}, G_{i,2})\}_{1 \le i \le n} \sim_{i.i.d.} \mathsf{N}\left(\boldsymbol{0}_2, \begin{bmatrix} 1 & m \\ m & 1 \end{bmatrix} \right).$$

Hence,

$$\begin{aligned} \frac{1}{n} \int_{\mathbb{R}^d} \frac{\mathbb{E}_{\boldsymbol{A}} \left\{ (p(\boldsymbol{y} \mid \boldsymbol{A}))^2 \right\}}{\mathbb{E}_{\boldsymbol{A}} \left\{ p(\boldsymbol{y} \mid \boldsymbol{A}) \right\}} \, \mathrm{d}\boldsymbol{y} &\overset{\text{(a)}}{=} \frac{1}{n} \mathbb{E}_M \left\{ (f(M))^n \right\} \\ &\overset{\text{(b)}}{=} \frac{1}{n} \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi}\,\Gamma(\frac{d-1}{2})} \int_{-1}^{1} (f(m))^n (1 - m^2)^{\frac{d-3}{2}} \, \mathrm{d}m, \end{aligned} \tag{73}$$

where in (a) we use definition (33) of f and in (b) we plug in the distribution of M obtained from Lemma 10 contained in Appendix C. Note that

$$\lim_{d \to \infty} \frac{\Gamma(\frac{d}{2})}{\frac{d}{2} \cdot \Gamma(\frac{d-1}{2})} = 1.$$

Therefore, by showing that the integral in the RHS of (73) tends to 0, the claim immediately follows.

Set $d' = d - 3$ and $\delta_n' = n/d'$. Thus,

$$\int_{-1}^{1} (f(m))^n (1 - m^2)^{\frac{d-3}{2}} \, \mathrm{d}m = \int_{-1}^{1} \exp\left(n \cdot F_{\delta_n'}(m) \right) \, \mathrm{d}m, \tag{74}$$

where $F_{\delta_n'}(m)$ is defined in (33). Define

$$\tilde{F}_{\delta}(m) = \delta \max(\log f(m), 0) + \frac{1}{2} \log(1 - m^2). \tag{75}$$

As $\delta < \delta_\ell$ and $n/d' \to \delta$, there exists $\delta_* \in (\delta, \delta_\ell)$ such that $\delta'_n < \delta_*$ for n sufficiently large. As $F_{\delta'_n}(m) \le \tilde{F}_{\delta'_n}(m)$ and $\tilde{F}_\delta(m)$ is non-decreasing in δ, we have that

$$\int_0^1 \exp\left(n \cdot F_{\delta'_n}(m)\right) \mathrm{d}m \le \int_0^1 \exp\left(n \cdot \tilde{F}_{\delta_*}(m)\right) \mathrm{d}m. \tag{76}$$

Note that $\tilde{F}_{\delta_*}(m) < 0$ if and only if $F_{\delta_*}(m) < 0$. Thus, by definition of δ_ℓ, we have that $\tilde{F}_{\delta_*}(m) < 0$ for $m \neq 0$ when n is sufficiently large. Furthermore, $\tilde{F}_{\delta_*}(0) = 0$ and $\tilde{F}_{\delta_*}$ is a continuous function. As a result, by Lemma 11, the integral in (76) tends to 0 as $n \to \infty$ and the claim immediately follows.

5 Proof of Theorems 2 and 4: Spectral Upper Bound

We will consider the complex case. The proof for the real case is essentially the same, and it is briefly discussed in Remark 7 at the end of this section.

A crucial ingredient of the proof consists in Lemma 2, which is a generalization of Theorem 1 of [49]. Before stating this result, we need some definitions. Let $G \sim \mathsf{CN}(0, 1)$, $Y \sim p(\cdot \mid |G|)$, and $Z = \mathcal{T}(Y)$. Assume that Z has bounded support and let τ be the supremum of this support, i.e.,

$$\tau = \inf\{z : \mathbb{P}(Z \le z) = 1\}. \tag{77}$$

For $\lambda \in (\tau, \infty)$ and $\delta \in (0, \infty)$, define

$$\phi(\lambda) = \lambda \cdot \mathbb{E}\left\{\frac{Z \cdot |G|^2}{\lambda - Z}\right\}, \tag{78}$$

and

$$\psi_\delta(\lambda) = \lambda\left(\frac{1}{\delta} + \mathbb{E}\left\{\frac{Z}{\lambda - Z}\right\}\right). \tag{79}$$

Note that $\phi(\lambda)$ is a monotone non-increasing function and that $\psi_\delta(\lambda)$ is a convex function. Let $\bar{\lambda}_\delta$ be the point at which ψ_δ attains its minimum, i.e.,

$$\bar{\lambda}_\delta = \arg\min_{\lambda \ge \tau} \psi_\delta(\lambda). \tag{80}$$

For $\lambda \in (\tau, \infty)$, define also

$$\zeta_\delta(\lambda) = \psi_\delta(\max(\lambda, \bar{\lambda}_\delta)). \tag{81}$$

Lemma 2 (Generalization of Theorem 1 of [49]) *Let* $\boldsymbol{x} \sim \mathrm{Unif}(\mathsf{S}_{\mathbb{C}}^{d-1})$, $\{\boldsymbol{a}_i\}_{1 \le i \le n} \sim_{i.i.d.} \mathsf{CN}(\boldsymbol{0}_d, \boldsymbol{I}_d)$, *and* $\boldsymbol{y}$ *be distributed according to* (8). *Let* $n/d \to \delta$, $G \sim \mathsf{CN}(0, 1)$ *and define* $Z = \mathcal{T}(Y)$ *for* $Y \sim p(\cdot \,|\, |G|)$. *Assume that* Z *satisfies* $\mathbb{P}(Z = 0) < 1$ *and that*

it has bounded support. Let τ be defined in (77). Assume further that, as λ approaches τ from the right, we have

$$\lim_{\lambda\to\tau^+} \mathbb{E}\left\{\frac{Z}{(\lambda - Z)^2}\right\} = \lim_{\lambda\to\tau^+} \mathbb{E}\left\{\frac{Z\cdot|G|^2}{\lambda - Z}\right\} = \infty. \tag{82}$$

Let $\hat{\boldsymbol{x}}$ be the principal eigenvector of the matrix $\boldsymbol{D}_n$, defined as in (21). Then, the following results hold:

(1) The equation

$$\zeta_\delta(\lambda) = \phi(\lambda) \tag{83}$$

admits a unique solution, call it λ^*_δ, for $\lambda > \tau$.

(2) As $n \to \infty$,

$$\frac{|\langle\hat{\boldsymbol{x}}, \boldsymbol{x}\rangle|^2}{\|\hat{\boldsymbol{x}}\|_2^2\,\|\boldsymbol{x}\|_2^2} \xrightarrow{a.s.} \begin{cases} 0, & \text{if } \psi'_\delta(\lambda^*_\delta) \le 0, \\ \dfrac{\psi'_\delta(\lambda^*_\delta)}{\psi'_\delta(\lambda^*_\delta) - \phi'(\lambda^*_\delta)}, & \text{if } \psi'_\delta(\lambda^*_\delta) > 0, \end{cases} \tag{84}$$

where ψ'_δ and ϕ' denote the derivatives of these two functions.

(3) Let $\lambda_1^{\boldsymbol{D}_n} \ge \lambda_2^{\boldsymbol{D}_n}$ denote the two largest eigenvalues of $\boldsymbol{D}_n$. Then, as $n \to \infty$,

$$\begin{aligned} \lambda_1^{\boldsymbol{D}_n} &\xrightarrow{a.s.} \zeta_\delta(\lambda^*_\delta), \\ \lambda_2^{\boldsymbol{D}_n} &\xrightarrow{a.s.} \zeta_\delta(\bar{\lambda}_\delta). \end{aligned} \tag{85}$$

Before proceeding with the proof, we discuss these results in more detail and we describe in what sense Lemma 2 provides a generalization of Theorem 1 of [49].

Remark 5 (Two different regimes) The results of Lemma 2 imply that, according to the value of δ, we can distinguish between two possible regimes.

On the one hand, suppose that $\phi(\bar{\lambda}_\delta) > \psi_\delta(\bar{\lambda}_\delta)$. Recall that $\phi(\lambda)$ is non-increasing and that $\bar{\lambda}_\delta$ is the point in which $\psi_\delta(\lambda)$ attains its minimum. Thus, $\bar{\lambda}_\delta < \lambda^*_\delta$, which implies that $\psi'_\delta(\lambda^*_\delta) > 0$ and that $\zeta_\delta(\lambda^*_\delta) > \zeta_\delta(\bar{\lambda}_\delta)$. This means that the scalar product $|\langle\hat{\boldsymbol{x}}, \boldsymbol{x}\rangle|$ is bounded away from zero and that there is a strictly positive gap between the two largest eigenvalues of $\boldsymbol{D}_n$. In this regime, the spectral method that outputs $\hat{\boldsymbol{x}}$ solves the weak recovery problem and (24) holds for some $\epsilon > 0$.

On the other hand, suppose that $\phi(\bar{\lambda}_\delta) \le \psi_\delta(\bar{\lambda}_\delta)$. Thus, $\bar{\lambda}_\delta \ge \lambda^*_\delta$, which implies that $\psi'_\delta(\lambda^*_\delta) \le 0$ and that $\zeta_\delta(\lambda^*_\delta) = \zeta_\delta(\bar{\lambda}_\delta)$. In words, this means that the scalar product $|\langle\hat{\boldsymbol{x}}, \boldsymbol{x}\rangle|$ converges to zero and that there is no strictly positive gap between the two largest eigenvalues of $\boldsymbol{D}_n$. In this regime, the spectral method that outputs $\hat{\boldsymbol{x}}$ does not solve the weak recovery problem.

Remark 6 (Lemma 2 and Theorem 1 of [49]) Lemma 2 generalizes Theorem 1 of [49] in the following two regards:

– $\boldsymbol{x}$ and $\{\boldsymbol{a}_i\}_{1\le i\le n}$ are complex vectors, while Theorem 1 of [49] considers the real case;
– Z can also be negative, while Theorem 1 of [49] assumes that $Z \ge 0$.

The first generalization does not require additional work as the whole argument of [49] generalizes in the natural way to the complex case: Gaussian random variables become circularly symmetric complex Gaussian random variables, transposes of vectors and matrices become conjugate transposes, squares become modulus squares, and so on.

On the contrary, the second generalization is more challenging, as it requires the result of Lemma 3, which is stated below and proved in Appendix D.

As a final observation, let us point out that Theorem 1 of [49] assumes also that $\mathbb{E}\left\{Z\cdot|G|^2\right\} > \mathbb{E}\{Z\}$. A careful check shows that this hypothesis is never used in the proof of that theorem, but it is required only in the proof of some additional results of [49].

Lemma 3 (Generalization of [3] to non-PSD matrices) *Consider the random matrix*

$$\boldsymbol{S}_n = \frac{1}{n}\boldsymbol{U}\boldsymbol{M}_n\boldsymbol{U}^*, \tag{86}$$

where the entries of $\boldsymbol{U} \in \mathbb{C}^{(d-1)\times n}$ *are* $\sim_{i.i.d.} \mathsf{CN}(0,1)$, *and* $\boldsymbol{M}_n \in \mathbb{C}^{n\times n}$ *is independent of* $\boldsymbol{U}$. *Let* $\lambda_1^{\boldsymbol{M}_n}$ *denote the largest eigenvalue of* $\boldsymbol{M}_n$. *Assume that the empirical spectral measure of the eigenvalues of* $\boldsymbol{M}_n$ *almost surely converges weakly to the probability distribution* H, *where* H *is the law of the random variable* Z. *Let* Γ_H *be the support of* H *and let* τ *be the supremum of* Γ_H. *Assume also that, as* $n\to\infty$,

$$\lambda_1^{\boldsymbol{M}_n} \xrightarrow{a.s.} \alpha_* \notin \Gamma_H. \tag{87}$$

Let $n/d \to \delta$, *denote by* $\lambda_1^{\boldsymbol{S}_n}$ *the largest eigenvalue of the matrix* (86), *and define* ψ_δ *as in* (79). *Then, as* $n\to\infty$,

$$\begin{aligned} \lambda_1^{\boldsymbol{S}_n} &\xrightarrow{a.s.} \psi_\delta(\alpha_*), \qquad \text{if } \psi_\delta'(\alpha_*) > 0, \\ \lambda_1^{\boldsymbol{S}_n} &\xrightarrow{a.s.} \min_{\lambda>\tau} \psi_\delta(\lambda), \quad \text{if } \psi_\delta'(\alpha_*) \le 0. \end{aligned} \tag{88}$$

Proof (Proof of Lemma 2) In this proof, we follow closely the approach detailed in Section III of [49]. First of all, let us write the matrix $\boldsymbol{D}_n$ defined in (21) as

$$\boldsymbol{D}_n = \frac{1}{n}\boldsymbol{A}\boldsymbol{Z}\boldsymbol{A}^*, \tag{89}$$

where $\boldsymbol{A} = [\boldsymbol{a}_1,\dots,\boldsymbol{a}_n]$, $\boldsymbol{Z}$ is a diagonal matrix with entries $z_i = \mathcal{T}(y_i)$ for $i\in[n]$, the random variables y_i are independent and distributed according to $p(\cdot\mid|g_i|)$, and $\{g_i\}_{1\le i\le n} \sim_{i.i.d.} \mathsf{CN}(0,1)$. As the sensing vectors $\{\boldsymbol{a}_i\}_{1\le i\le n}$ are drawn from the circularly symmetric complex normal distribution, we can assume without loss of generality that $\boldsymbol{x} = \boldsymbol{e}_1$, where $\boldsymbol{e}_1$ is the first element of the canonical basis of $\mathbb{C}^d$.

Consider a matrix $\boldsymbol{U} \in \mathbb{C}^{(d-1)\times n}$ independent of $\{g_i\}_{1\le i\le n}$ and $\boldsymbol{Z}$. Let the elements of $\boldsymbol{U}$ be $\sim_{i.i.d.} \mathsf{CN}(0,1)$. Define

$$\boldsymbol{P}_n = \frac{1}{n}\boldsymbol{U}\boldsymbol{Z}\boldsymbol{U}^*, \tag{90}$$

and

$$\boldsymbol{q}_n = \frac{1}{n}\boldsymbol{U}\boldsymbol{v}, \tag{91}$$

where $\boldsymbol{v} = [z_1 g_1, \ldots, z_n g_n]^*$. Then, (89) can be rewritten as

$$\boldsymbol{D}_n = \begin{bmatrix} a_n & \boldsymbol{q}_n^* \\ \boldsymbol{q}_n & \boldsymbol{P}_n \end{bmatrix}, \tag{92}$$

where $a_n = \sum_{i=1}^n z_i |g_i|^2/n$ is a scalar that converges almost surely to $\mathbb{E}(Z\cdot|G|^2)$ as $n\to\infty$, with $G\sim\mathsf{CN}(0,1)$.

Next, consider a parametric family of matrices $\{\boldsymbol{P}_n + \mu\boldsymbol{q}_n\boldsymbol{q}_n^*\}$ and let $L_n(\mu)$ denote their largest eigenvalues, i.e.,

$$L_n(\mu) = \lambda_1(\boldsymbol{P}_n + \mu\boldsymbol{q}_n\boldsymbol{q}_n^*).$$

The idea is to compute the largest eigenvalue of $\boldsymbol{D}_n$, call it $\lambda_1^{\boldsymbol{D}_n}$, and the scalar product between $\hat{\boldsymbol{X}}$ and $\boldsymbol{e}_1$ via a fixed-point equation involving $L_n(\mu)$.

To do so, we first need an intermediate result holding for any matrix $\boldsymbol{D}$ that can be written in the form

$$\boldsymbol{D} = \begin{bmatrix} a & \boldsymbol{q}^* \\ \boldsymbol{q} & \boldsymbol{P} \end{bmatrix},$$

where $a\in\mathbb{R}$, $\boldsymbol{P}\in\mathbb{C}^{(d-1)\times(d-1)}$ is a Hermitian matrix and $\boldsymbol{q}\in\mathbb{C}^{d-1}$ is such that $\|\boldsymbol{q}\|\neq 0$. Note that the matrix $\boldsymbol{D}_n$ defined in (21) fulfills such requirements, since the matrix $\boldsymbol{P}_n$ defined in (90) is Hermitian and $\boldsymbol{q}_n$ defined in (91) is such that $\|\boldsymbol{q}_n\|\neq 0$ with high probability, as $\mathbb{P}(Z=0)<1$.

Let $\lambda_1^{\boldsymbol{P}} \ge \lambda_2^{\boldsymbol{P}} \ge \cdots \ge \lambda_{d-1}^{\boldsymbol{P}}$ be the set of eigenvalues of $\boldsymbol{P}$, and let $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_{d-1}$ be a corresponding set of eigenvectors. For $\lambda \in (\max\{\lambda_i^{\boldsymbol{P}} : \langle\boldsymbol{q},\boldsymbol{w}_i\rangle\neq 0\}, \infty)$, define

$$R(\lambda) = \boldsymbol{q}^*(\boldsymbol{P}-\lambda\boldsymbol{I})^{-1}\boldsymbol{q} = \sum_{i=1}^{d-1}\frac{|\langle\boldsymbol{q},\boldsymbol{w}_i\rangle|^2}{\lambda_i^{\boldsymbol{P}}-\lambda}. \tag{93}$$

Note that $R(\lambda)$ increases monotonically from $-\infty$ to 0. Hence, it admits an inverse, call it $R^{-1}(x)$, for $x<0$. Then, the maximum eigenvalue $L(\mu)=\lambda_1(\boldsymbol{P}+\mu\boldsymbol{q}\boldsymbol{q}^*)$ is given by

$$L(\mu) = \max(R^{-1}(-1/\mu), \lambda_1^{\boldsymbol{P}}). \tag{94}$$

The proof of (94) is standard, cf., e.g., Lemma 1 in [49]. Note that $L(\mu)$ is a non-decreasing function such that

$$\lim_{\mu\to\infty} L(\mu) = \infty.$$

Indeed, by construction, $R^{-1}(-1/\mu)$ is strictly increasing and

$$\lim_{\mu\to\infty} R^{-1}(-1/\mu) = \infty.$$

Furthermore, $L(\mu)$ is convex since it is the maximum of a set of linear functions, as

$$L(\mu) = \lambda_1(\boldsymbol{P} + \mu\boldsymbol{q}\boldsymbol{q}^*) = \max_{\boldsymbol{x}:\|\boldsymbol{x}\|=1} \boldsymbol{x}^*(\boldsymbol{P} + \mu\boldsymbol{q}\boldsymbol{q}^*)\boldsymbol{x}.$$

Let $\mu^* > 0$ be the solution to the fixed-point equation

$$\mu = (L(\mu) - a)^{-1}. \tag{95}$$

This solution is unique, since $L(\mu)$ is a non-decreasing function with $\lim_{\mu\to\infty} L(\mu) = \infty$. Then,

$$\lambda_1^{\boldsymbol{D}} = L(\mu^*), \tag{96}$$

and

$$|\langle \hat{\boldsymbol{x}}, \boldsymbol{e}_1\rangle|^2 \in \left[\frac{\partial_- L(\mu^*)}{\partial_- L(\mu^*) + (1/\mu^*)^2}, \frac{\partial_+ L(\mu^*)}{\partial_+ L(\mu^*) + (1/\mu^*)^2}\right], \tag{97}$$

where $\partial_- L(\mu^*)$ and $\partial_+ L(\mu^*)$ denote the left and right derivative of $L(\mu)$, respectively. In particular, if $L(\mu)$ is differentiable at μ^*, then

$$|\langle \hat{\boldsymbol{x}}, \boldsymbol{e}_1\rangle|^2 = \frac{L'(\mu^*)}{L'(\mu^*) + (1/\mu^*)^2}. \tag{98}$$

The proof of (96), (97), and (98) uses the characterization (94), and it is analogous to the proof of Proposition 2 in [49].

At this point, we need to compute $L_n(\mu)$ for the matrix $\boldsymbol{D}_n$ defined in (92). The eigenvalues of a low-rank perturbation of a random matrix are studied in [12]. However, we cannot apply those results, as $\boldsymbol{P}_n$ and $\boldsymbol{q}_n$ are dependent. Hence, we write

$$\boldsymbol{P}_n + \mu\boldsymbol{q}_n\boldsymbol{q}_n^* = \frac{1}{n}\boldsymbol{U}\boldsymbol{M}_n\boldsymbol{U}^*,$$

where $\boldsymbol{M}_n$ is independent of $\boldsymbol{U}$ with

$$\boldsymbol{M}_n = \boldsymbol{Z} + \frac{\mu}{n}\boldsymbol{v}\boldsymbol{v}^*.$$

We start by studying the spectrum of $\boldsymbol{M}_n$. Let $\lambda_1^{\boldsymbol{M}_n} \geq \lambda_2^{\boldsymbol{M}_n} \geq \cdots \geq \lambda_n^{\boldsymbol{M}_n}$ be the set of eigenvalues of $\boldsymbol{M}_n$ and let

$$f^{\boldsymbol{M}_n} = \frac{1}{n-1}\sum_{i=2}^{n} \delta_{\lambda_i^{\boldsymbol{M}_n}}$$

be the empirical spectral measure of the last $n-1$ eigenvalues.

Then, standard interlacing theorems (see [39, Section 4.3]) yield that $f^{\boldsymbol{M}_n}$ almost surely converges weakly to the probability law of Z. Furthermore, by using the characterization (94), we can show that

$$\lambda_1^{\boldsymbol{M}_n} \xrightarrow{\text{a.s.}} \lambda_\mu = Q^{-1}(1/\mu), \tag{99}$$

where Q^{-1} is the inverse of the function

$$Q(\lambda) = \mathbb{E}\left\{\frac{Z^2 \cdot |G|^2}{\lambda - Z}\right\}.$$

The proof of these results is the same as the proof of Proposition 3 in [49].

Note that $Q(\lambda)$ is defined for $\lambda \in (\tau, \infty)$, it is continuous and strictly decreasing with $Q(\infty) = 0$. Furthermore, by hypothesis (82), we have that $\lim_{\lambda \to \tau^+} Q(\lambda) = \infty$. Thus, $Q(\lambda)$ admits an inverse and $Q^{-1}(1/\mu)$ is well defined for all $\mu > 0$.

Let us now consider the matrix $\frac{1}{n}\boldsymbol{U}\boldsymbol{M}_n\boldsymbol{U}^*$. First, if $Z \geq 0$, then $\boldsymbol{M}_n$ is positive semi-definite (PSD) and we can apply results from [3] to compute the limit of $L_n(\mu)$. If $\boldsymbol{M}_n$ is not necessarily PSD, we use Lemma 3 with $\alpha_* = \lambda_\mu$ to conclude that

$$\begin{aligned} L_n(\mu) &\xrightarrow{\text{a.s.}} \psi_\delta(\lambda_\mu), \quad \text{if } \psi_\delta'(\lambda_\mu) > 0, \\ L_n(\mu) &\xrightarrow{\text{a.s.}} \min_{\lambda > \tau} \psi_\delta(\lambda), \quad \text{if } \psi_\delta'(\lambda_\mu) \leq 0. \end{aligned} \tag{100}$$

The remaining part of the proof follows the argument of Section III-D in [49]. For the sake of readability, we reproduce it below.

We start by proving the first claim of the lemma. For $n \geq 1$, let μ_n be the unique solution to the fixed-point equation (95). Then,

$$L_n(\mu_n) - 1/\mu_n = a_n.$$

Now, fix any $\mu > 0$. Then, by using definition (81) and the fact that $\lambda_\mu = Q^{-1}(1/\mu)$, (100) immediately implies that, as $n \to \infty$,

$$L_n(\mu) - 1/\mu \xrightarrow{\text{a.s.}} \zeta_\delta(Q^{-1}(1/\mu)) - 1/\mu. \tag{101}$$

Note that, as $n \to \infty$, $a_n \xrightarrow{\text{a.s.}} \mathbb{E}(Z \cdot |G|^2)$. Furthermore, as $L_n(\mu)$ and $\zeta_\delta(\mu)$ are non-decreasing, the two functions on both sides of (101) are strictly increasing. Consequently, by Lemma 3 in Appendix E of [49], we conclude that

$$\mu_n \xrightarrow{\text{a.s.}} \mu^*, \tag{102}$$

where μ^* is the unique fixed point such that

$$\zeta_\delta(Q^{-1}(1/\mu^*)) = \mathbb{E}(Z \cdot |G|^2) + 1/\mu^*. \tag{103}$$

Define

$$\lambda^* = Q^{-1}(1/\mu^*). \tag{104}$$

Then, (103) can be rewritten as

$$\zeta_\delta(\lambda^*) = \mathbb{E}(Z \cdot |G|^2) + Q(\lambda^*) = \phi(\lambda^*), \tag{105}$$

where ϕ is defined in (78). By construction, $\zeta_\delta(\lambda)$ is a non-decreasing continuous function on (τ, ∞) and $\phi(\lambda)$ is a strictly decreasing continuous function. Furthermore, by hypothesis (82), we have that $\lim_{\lambda \to \tau^+} \phi(\lambda) = \infty$. Hence, the existence and the uniqueness of λ^* satisfying (105) are guaranteed. This suffices to prove the first claim of the lemma.

Let us now move on to the proof of the second claim of the lemma. Suppose that $\zeta_\delta(Q^{-1}(1/\mu))$ is differentiable at $\mu = \mu^*$. Then, as $L_n(\mu)$ is convex for any $n \geq 1$, by Lemma 4 in Appendix E of [49], we have that

$$\partial_- L_n(\mu_n) \xrightarrow{\text{a.s.}} \left.\frac{\mathrm{d}\zeta_\delta(Q^{-1}(1/\mu))}{\mathrm{d}\mu}\right|_{\mu=\mu^*} = \frac{-\zeta_\delta'(Q^{-1}(1/\mu^*))}{Q'(Q^{-1}(1/\mu^*)) \cdot (\mu^*)^2}.$$

Similarly,

$$\partial_+ L_n(\mu_n) \xrightarrow{\text{a.s.}} \frac{-\zeta_\delta'(Q^{-1}(1/\mu^*))}{Q'(Q^{-1}(1/\mu^*)) \cdot (\mu^*)^2}.$$

By using (97), we obtain that

$$|\langle \hat{\boldsymbol{x}}, \boldsymbol{e}_1 \rangle|^2 \xrightarrow{\text{a.s.}} \frac{\zeta_\delta'(Q^{-1}(1/\mu^*))}{\zeta_\delta'(Q^{-1}(1/\mu^*)) - Q'(Q^{-1}(1/\mu^*))} = \frac{\zeta_\delta'(\lambda^*)}{\zeta_\delta'(\lambda^*) - \phi'(\lambda^*)},$$

where the equality follows from definition (104) of λ^* and from the fact that $Q'(\lambda) = \phi'(\lambda)$. In order to prove the second claim of the lemma, it suffices to note that, by its definition in (81), $\zeta_\delta'(\lambda) = \psi_\delta'(\lambda)$ if $\psi_\delta'(\lambda) > 0$, and $\zeta_\delta'(\lambda) = 0$ if $\psi_\delta'(\lambda) < 0$.

Finally, let us prove the third claim of the lemma. By using (96), we immediately obtain that $\lambda_1^{\boldsymbol{D}_n} = L_n(\mu_n)$. By applying (102) and Lemma 3 in Appendix E of [49], we conclude that

$$\lambda_1^{\boldsymbol{D}_n} \xrightarrow{\text{a.s.}} \zeta_\delta(\lambda^*).$$

As $\boldsymbol{P}_n$ is obtained by deleting the first row and column of $\boldsymbol{D}_n$, by applying Cauchy interlacing theorem (see, e.g., [39, Theorem 4.3.17]), we also have that

$$\lambda_2^{\boldsymbol{P}_n} \leq \lambda_2^{\boldsymbol{D}_n} \leq \lambda_1^{\boldsymbol{P}_n}.$$

Furthermore, the upper edge of the support of the limiting spectral distribution of $\boldsymbol{P}_n$ is given by [67, Section 4] and [3, Lemma 3.1]

$$\min_{\lambda > \tau} \psi_\delta(\lambda) = \zeta_\delta(\bar{\lambda}_\delta),$$

where $\bar{\lambda}_\delta$ is defined in (80). Therefore,

$$\lambda_2^{\boldsymbol{D}_n} \xrightarrow{\text{a.s.}} \zeta_\delta(\bar{\lambda}_\delta),$$

which concludes the proof. □

At this point, we are ready to prove our spectral upper bound.

Proof (Proof of Theorem 2) Note that the normalization of $\boldsymbol{x}$ and $\{\boldsymbol{a}_i\}_{1 \leq i \leq n}$ required in Lemma 2 is different from the normalization required in Theorem 2. However, the scalar product $\langle \boldsymbol{x}, \boldsymbol{a}_i \rangle$ is the same and the data matrix $\boldsymbol{D}_n$ changes by a factor d. Hence, the principal eigenvector $\hat{\boldsymbol{x}}$ is not affected by this change in the normalization.

Let $G \sim \mathsf{CN}(0, 1)$, $Y \sim p(\cdot \mid |G|)$ and $Z = \mathcal{T}(Y)$, where p is defined in (8) and $\mathcal{T}$ is some pre-processing function that we will choose later on. We will assume that the supremum τ of the support of Z is strictly positive and that conditions (82) are satisfied, and will verify later that our choice of the function $\mathcal{T}$ satisfies these requirements. Recall that the function $\psi_\delta(\lambda)$ defined in (79) is convex and that it attains its minimum at the point $\bar{\lambda}_\delta$. Since by condition (82) $\psi_\delta(\lambda) \uparrow \infty$ as $\lambda \downarrow 0$, we have $\bar{\lambda}_\delta \in (\tau, \infty)$. Hence, $\psi'_\delta(\bar{\lambda}_\delta) = 0$. By calculating the derivative of $\psi_\delta(\lambda)$ and setting it to 0, we have

$$\mathbb{E}\left\{\frac{Z^2}{(\bar{\lambda}_\delta - Z)^2}\right\} = \frac{1}{\delta}. \tag{106}$$

Furthermore, as pointed out in Remark 5, (24) holds for some $\epsilon > 0$ if and only if

$$\phi(\bar{\lambda}_\delta) > \psi_\delta(\bar{\lambda}_\delta). \tag{107}$$

As $\tau > 0$, we also have that $\bar{\lambda}_\delta > 0$. Consider now the matrix $\boldsymbol{D}'_n = \boldsymbol{D}_n/\alpha$ for some $\alpha > 0$. Then, the principal eigenvector of $\boldsymbol{D}'_n$ is equal to the principal eigenvector

of $\boldsymbol{D}_n$. Hence, we can assume without loss of generality that $\bar{\lambda}_\delta = 1$. Consequently, conditions (106) and (107) can be, respectively, rewritten as

$$\mathbb{E}\left\{\frac{Z^2}{(1-Z)^2}\right\} = \frac{1}{\delta}, \tag{108}$$

$$\mathbb{E}\left\{\frac{Z(|G|^2-1)}{1-Z}\right\} > \frac{1}{\delta}. \tag{109}$$

Furthermore, as $Z = \mathcal{T}(Y)$, we also obtain that

$$\begin{aligned}\mathbb{E}\left\{\frac{Z^2}{(1-Z)^2}\right\} &= \int_{\mathbb{R}} \left(\frac{\mathcal{T}(y)}{1-\mathcal{T}(y)}\right)^2 \mathbb{E}_G\left\{p(y \mid |G|)\right\} \mathrm{d}y,\\ \mathbb{E}\left\{\frac{Z(|G|^2-1)}{1-Z}\right\} &= \int_{\mathbb{R}} \frac{\mathcal{T}(y)}{1-\mathcal{T}(y)} \mathbb{E}_G\left\{p(y \mid |G|)\cdot(|G|^2-1)\right\} \mathrm{d}y.\end{aligned} \tag{110}$$

Let $\mathcal{T}^*(y)$ be defined in (23). Note that, if we substitute $\mathcal{T}(y) = \mathcal{T}^*(y)$ into the RHS of (110), then

$$\mathbb{E}\left\{\frac{Z^2}{(1-Z)^2}\right\} = \mathbb{E}\left\{\frac{Z(|G|^2-1)}{1-Z}\right\} = \frac{1}{\delta_{\mathrm{u}}},$$

where δ_{u} is defined in (20). Let $\mathcal{T}_\delta^*(y)$ be defined in (22). Then,

$$\frac{\mathcal{T}_\delta^*(y)}{1-\mathcal{T}_\delta^*(y)} = \sqrt{\frac{\delta_{\mathrm{u}}}{\delta}}\frac{\mathcal{T}^*(y)}{1-\mathcal{T}^*(y)},$$

which immediately implies that

$$\mathbb{E}\left\{\frac{(\mathcal{T}_\delta^*(Y))^2}{(1-\mathcal{T}_\delta^*(Y))^2}\right\} = \tfrac{1}{\delta}, \tag{111}$$

$$\mathbb{E}\left\{\frac{\mathcal{T}_\delta^*(Y)(|G|^2-1)}{1-\mathcal{T}_\delta^*(Y)}\right\} = \tfrac{1}{\sqrt{\delta\cdot\delta_{\mathrm{u}}}} > \tfrac{1}{\delta}. \tag{112}$$

As a result, we need to show that the function $\mathcal{T}_\delta^*(y)$ fulfills the following requirements:

(1) $\mathcal{T}_\delta^*(y)$ is bounded;
(2) $\mathbb{P}(\mathcal{T}_\delta^*(Y) = 0) < 1$;
(3) the supremum τ of the support of $\mathcal{T}_\delta^*(Y)$ is strictly positive;
(4) condition (82) holds.

Note that $\mathcal{T}_\delta^*(y)$ is bounded, as $\mathcal{T}^*(y) \le 1$. Furthermore, if

$$\mathbb{E}_G\left\{p(y \mid |G|)\right\} = \mathbb{E}_G\left\{p(y \mid |G|)|G|^2\right\}, \tag{113}$$

identically, then $\delta_{\mathrm{u}} = \infty$ and the claim of Theorem 2 trivially holds. Hence, we can assume that (113) does not hold, which implies that the function $\mathcal{T}^*$ is not equal to the constant value 0. Consequently, $\mathbb{P}(\mathcal{T}^*_\delta(Y) = 0) < 1$.

By definition (23) of $\mathcal{T}^*$, we have that

$$\mathbb{E}_Y\left\{\frac{1}{1-\mathcal{T}^*(Y)}\right\} = \int_{\mathbb{R}} \mathbb{E}_G\left\{p(y \mid |G|)\cdot |G|^2\right\}\mathrm{d}y = \mathbb{E}_G\left\{|G|^2\right\} = 1. \quad (114)$$

Hence, $\mathbb{P}(\mathcal{T}^*(Y) > 0) > 0$, which implies that $\mathbb{P}(\mathcal{T}^*_\delta(Y) > 0) > 0$. Consequently, the supremum τ of the support of $\mathcal{T}^*_\delta(Y)$ is strictly positive.

If $\mathbb{P}(\mathcal{T}^*_\delta(Y) = \tau) > 0$, then condition (82) is satisfied. Suppose now that $\mathbb{P}(\mathcal{T}^*_\delta(Y) = \tau) = 0$. Then, for any $\epsilon_1 > 0$, there exists $\Delta_1(\epsilon_1)$ such that

$$0 < \mathbb{P}\big(\mathcal{T}^*_\delta(Y) \in (\tau - \Delta_1(\epsilon_1), \tau)\big) \le \epsilon_1. \quad (115)$$

Define

$$\mathcal{T}^*_\delta(y, \epsilon_1) = \begin{cases} \mathcal{T}^*_\delta(y), & \text{if } \mathcal{T}^*_\delta(y) \le \tau - \Delta_1(\epsilon_1), \\ \tau - \Delta_1(\epsilon_1), & \text{otherwise.} \end{cases} \quad (116)$$

Clearly, the random variable $\mathcal{T}^*_\delta(Y, \epsilon_1)$ has a point mass; hence, condition (82) is satisfied.

As a final step, we show that we can take $\epsilon_1 = 0$. Define

$$\boldsymbol{D}_n(\epsilon_1) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}^*_\delta(y_i, \epsilon_1)\boldsymbol{a}_i\boldsymbol{a}_i^*.$$

Define also

$$\boldsymbol{D}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}^*_\delta(y_i)\boldsymbol{a}_i\boldsymbol{a}_i^*.$$

Let $\hat{\boldsymbol{x}}(\epsilon_1)$ and $\hat{\boldsymbol{x}}$ be the principal eigenvectors of $\boldsymbol{D}_n(\epsilon_1)$ and of $\boldsymbol{D}_n$, respectively. Then,

$$\|\boldsymbol{D}_n(\epsilon_1) - \boldsymbol{D}_n\|_{\mathrm{op}} \le C_1 \cdot \Delta_1(\epsilon_1), \quad (117)$$

where the constant C_1 depends only on n/d. By Lemma 2, there is a strictly positive gap, call it θ, between the first and the second eigenvalue of $\boldsymbol{D}_n(\epsilon_1)$. Consequently, by the Davis–Kahan theorem [23], we conclude that

$$\left\|\hat{\boldsymbol{x}}(\epsilon_1) - \hat{\boldsymbol{x}}\right\|_2 \le C_2 \cdot \Delta_1(\epsilon_1), \quad (118)$$

where the constant C_2 depends only on n/d and on θ. In words, for any n, as ϵ_1 tends to 0, the principal eigenvector of $\boldsymbol{D}_n(\epsilon_1)$ tends to the principal eigenvector of $\boldsymbol{D}_n$. This means that we can set $\mathcal{T} = \mathcal{T}^*_\delta$ and have that, almost surely, (24) holds.

In order to conclude the proof, it remains to show that δ_{u} is the optimal threshold for the spectral method, namely for any $\delta < \delta_{\mathrm{u}}$, there is no pre-processing function $\mathcal{T}$ such that (24) holds almost surely. To do so, note that (24) holds almost surely if and only if (108) and (109) are satisfied. By setting $u(y) = \mathcal{T}(y)/(1 - \mathcal{T}(y))$ and using (110), we have that these conditions can be rewritten as

$$\int_{\mathbb{R}} (u(y))^2 \mathbb{E}_G \{p(y \mid |G|)\} \, \mathrm{d}y = \frac{1}{\delta}, \tag{119}$$

$$\mathbb{E}\left\{\frac{Z(|G|^2-1)}{1-Z}\right\} = \int_{\mathbb{R}} u(y)\sqrt{\mathbb{E}_G \{p(y \mid |G|)\}}\frac{\mathbb{E}_G \left\{p(y \mid |G|) \cdot (|G|^2-1)\right\}}{\sqrt{\mathbb{E}_G \{p(y \mid |G|)\}}} \, \mathrm{d}y > \frac{1}{\delta}. \tag{120}$$

By Cauchy–Schwarz inequality, we also have that

$$\begin{aligned} &\int_{\mathbb{R}} u(y)\sqrt{\mathbb{E}_G \{p(y \mid |G|)\}}\frac{\mathbb{E}_G \left\{p(y \mid |G|) \cdot (|G|^2-1)\right\}}{\sqrt{\mathbb{E}_G \{p(y \mid |G|)\}}} \, \mathrm{d}y \\ &\leq \sqrt{\int_{\mathbb{R}} (u(y))^2 \mathbb{E}_G \{p(y \mid |G|)\} \, \mathrm{d}y} \sqrt{\int_{\mathbb{R}} \frac{\left(\mathbb{E}_G \left\{p(y \mid |G|) \cdot (|G|^2-1)\right\}\right)^2}{\mathbb{E}_G \{p(y \mid |G|)\}} \, \mathrm{d}y}. \end{aligned} \tag{121}$$

By combining (119), (120) and (121) with definition (20) of δ_{u}, we conclude that

$$\frac{1}{\sqrt{\delta_{\mathrm{u}}}}\frac{1}{\sqrt{\delta}} > \frac{1}{\delta}, \tag{122}$$

which implies that $\delta > \delta_{\mathrm{u}}$. Consequently, for $\delta \leq \delta_{\mathrm{u}}$, no pre-processing function achieves weak recovery and the proof is complete. □

Remark 7 (Proof of spectral upper bound for the real case) First, we need to prove a result analogous to that of Lemma 2, where $\boldsymbol{x} \sim \mathrm{Unif}(\mathsf{S}_{\mathbb{R}}^{d-1})$, $\{\boldsymbol{a}_i\}_{1\leq i\leq n} \sim_{i.i.d.} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d)$, $\boldsymbol{y}$ is distributed according to (29), and $G \sim \mathsf{N}(0, 1)$. To do so, one can follow the proof of Theorem 1 of [49]. The technical difficulty consists in the fact that the matrix $\boldsymbol{M}_n$ is not necessarily PSD. In order to solve this issue, we apply the version of Lemma 3 for the real case discussed in Remark 8 at the end of Appendix D. At this point, the proof of Theorem 4 follows from the same argument as the proof of Theorem 2.

6 Comparison with Message Passing Algorithms

6.1 Motivation and Background

Message passing algorithms have proved successful in a broad range of statistical estimation problems, including high-dimensional regression [10], robust regression [32], low-rank matrix estimation [26,42,45,53], and network structure estimation [24,55,

56]. A bold conjecture from statistical physics suggests that—for these and other problems—message passing approaches achieve optimal statistical performances among polynomial-time algorithms. In view of this conjecture, it is interesting to compare our spectral approach to message passing algorithms. We will present two types of results (with δ_{u} the spectral threshold defined in (42)):

1. We prove that, for $\delta < \delta_{\mathrm{u}}$ (i.e., in the regime in which the spectral approach fails), message passing converges to an un-informative fixed point, even if initialized in a state that is correlated with the true signal $\boldsymbol{x}$.
2. Vice versa, for $\delta > \delta_{\mathrm{u}}$ (when the spectral algorithm achieves weak recovery), we consider a linearized message passing algorithm and prove that the un-informative fixed point is unstable. The proof of this fact builds on the analysis contained in the previous pages.

Let us point out that the techniques described in Sect. 5 to compute the spectral threshold δ_{u} are different from those described in this section to analyze message passing algorithms. Hence, we find very interesting fact that the spectral threshold is closely related to the performance of message passing. In particular, our findings suggest the conjecture that δ_{u} represents the fundamental limit for all polynomial-time algorithms.

Note also that message passing often allows to further refine the spectral estimate, in order to provide an exact recovery of the signal. Hence, combining the analyses of message passing and of the spectral method to provide a threshold for exact recovery constitutes an interesting direction for future research (see [54] for an example in which this program is carried out).

For the sake of simplicity, we will assume that the signal $\boldsymbol{x}$ and the measurement matrix $\boldsymbol{A}$ are real. Of particular interest for the present setting is approximate message passing (AMP) [9,30]: This is a broad class of iterative methods that operates with dense random matrices (as the sensing matrix $\boldsymbol{A}$ in the present case). In particular, in [61] it was proposed a "generalized approximate message passing" (GAMP) scheme, which is an AMP algorithm for Bayesian estimation in nonlinear regression models. This approach was further developed in the context of phase retrieval in [64]. We will follow the same Bayesian formulation here, by considering an AMP algorithm that is equivalent to GAMP although somewhat simpler.

In order to minimize technical overhead, we assume throughout this section that the conditional density $p(y \mid g)$ is bounded and two times differentiable with respect to g. Denote by $\partial_g p(y \mid g)$ and $\partial_g^2 p(y \mid g)$ the first and the second derivative of $p(y \mid g)$, respectively. Let $G \sim \mathsf{N}(0, 1)$ and define the function

$$\mathsf{F}(x, y; \bar{q}) = \frac{\mathbb{E}_G\{\partial_g p(y \mid \bar{q}\, x + \sqrt{\bar{q}} G)\}}{\mathbb{E}_G\{p(y \mid \bar{q}\, x + \sqrt{\bar{q}} G)\}}. \tag{123}$$

We further define the following "state evolution" recursion:

$$\begin{aligned} \mu_{t+1} &= \delta \cdot h(q_t), \\ q_t &= \frac{\mu_t}{1 + \mu_t}, \end{aligned} \tag{124}$$

where

$$h(q) = \int_{\mathbb{R}} \mathbb{E}_{G_0} \left\{ \frac{\left(\mathbb{E}_{G_1}\{\partial_g p(y \mid \sqrt{q}G_0 + \sqrt{1-q}G_1)\}\right)^2}{\mathbb{E}_{G_1}\{p(y \mid \sqrt{q}G_0 + \sqrt{1-q}G_1)\}} \right\} \mathrm{d}y, \tag{125}$$

with $G_0, G_1 \sim_{i.i.d.} \mathsf{N}(0,1)$.

Given the sensing matrix $\boldsymbol{A} = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)^{\mathsf{T}} \in \mathbb{R}^{n\times d}$, and the vector of measurements $\boldsymbol{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$, the message passing algorithm updates iteratively the estimate $\boldsymbol{z}^t \in \mathbb{R}^d$ of the signal $\boldsymbol{x} \in \mathbb{R}^d$, with $\|\boldsymbol{x}\|_2 = \sqrt{d}$, according to the iteration

$$\begin{aligned} \boldsymbol{z}^{t+1} &= \boldsymbol{A}^{\mathsf{T}} f_t(\hat{\boldsymbol{z}}^t; \boldsymbol{y}) - \mathsf{b}_t \boldsymbol{z}^t, \\ \hat{\boldsymbol{z}}^t &= \boldsymbol{A}\boldsymbol{z}^t - f_{t-1}(\hat{\boldsymbol{z}}^{t-1}; \boldsymbol{y}). \end{aligned} \tag{126}$$

Here, the function $f_t(\hat{\boldsymbol{z}}; \boldsymbol{y}) = (f_t(\hat{z}_1; y_1), \dots, f_t(\hat{z}_n; y_n))$ is understood to be applied component-wise to its arguments and b_t is defined as

$$f_t(\hat{z}; y) = \mathsf{F}(\hat{z}, y; 1 - q_t), \tag{127}$$

and the "Onsager coefficient" b_t is defined as

$$\mathsf{b}_t = \delta \cdot \mathbb{E}\{f_t'(\mu_t G_0 + \sqrt{\mu_t} G_1; Y)\}, \tag{128}$$

where $f_t'(\hat{z}; y)$ denotes the derivative of $f_t(\hat{z}; y)$ with respect to $\hat{z}$, and the expectation is with respect to $G_0, G_1 \sim_{i.i.d.} \mathsf{N}(0,1)$ and $Y \sim p(\cdot \,|G_0)$. The recursion (126) is initialized with $\boldsymbol{z}^0 \in \mathbb{R}^d$ and it is understood that $f_{-1}(\cdot\,;\cdot) = \boldsymbol{0}_n$.

State evolution precisely tracks the asymptotics of AMP. The next statement is a consequence of [9,41]. We refer to Appendix E for its proof.

Lemma 4 (State evolution for AMP iteration (126)) *Let $\boldsymbol{x} \in \mathbb{R}^d$ denote the unknown signal such that $\|\boldsymbol{x}\|_2 = \sqrt{d}$, $\boldsymbol{A} = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)^T \in \mathbb{R}^{n\times d}$ with $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{i.i.d.} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$, and $\boldsymbol{y} = (y_1, \dots, y_n)$ with $y_i \sim p(\cdot \mid \langle \boldsymbol{x}, \boldsymbol{a}_i\rangle)$. Consider the AMP iterates $\boldsymbol{z}^t, \hat{\boldsymbol{z}}^t$ defined in (126), where $f_t(\hat{z}; y)$ and b_t are given by (127) and (128), respectively. Assume that the initialization $\boldsymbol{z}^0$ is independent of $\boldsymbol{A}$ and that, almost surely,*

$$\lim_{n\to\infty} \frac{1}{d}\langle \boldsymbol{x}, \boldsymbol{z}^0\rangle = \mu_0, \quad \lim_{n\to\infty} \frac{1}{d}\|\boldsymbol{z}^0\|^2 = \mu_0^2 + \mu_0. \tag{129}$$

Let the state evolution recursion q_t, μ_t be defined as in (124) with initialization μ_0. Then, for any t, and for any function $\psi : \mathbb{R}^2 \to \mathbb{R}$ such that $|\psi(\boldsymbol{u}) - \psi(\boldsymbol{v})| \le L(1 + \|\boldsymbol{u}\|_2 + \|\boldsymbol{v}\|_2)\|\boldsymbol{u} - \boldsymbol{v}\|_2$ for some $L \in \mathbb{R}$, we have that, almost surely,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \psi(x_i, z_i^t) = \mathbb{E}\left\{\psi(X_0, \mu_t X_0 + \sqrt{\mu_t} G)\right\}, \tag{130}$$

where the expectation is taken with respect to $X_0, G \sim_{i.i.d.} \mathsf{N}(0,1)$.

Informally, this lemma states that z^t is a noisy version of the signal $\boldsymbol{x}$, namely $z^t \approx \mu_t \boldsymbol{x} + \sqrt{\mu_t}\, \boldsymbol{g}$, with $\boldsymbol{g} \sim \mathsf{N}(\mathbf{0}_d, \boldsymbol{I}_d)$, and that this approximation holds for empirical averages.

6.2 Results

In order to obtain a nonvanishing weak recovery threshold, we assume that the observation model satisfies the condition

$$\mathbb{E}_G\{\partial_g p(y \mid G)\} = 0, \tag{131}$$

where the expectation is with respect to $G \sim \mathsf{N}(0, 1)$. Notice that this implies $h(0) = 0$; therefore, $\mu_t = q_t = 0$ is a fixed point of state evolution. Furthermore, $\mathsf{F}(0, y; 1) = 0$; therefore, $q_t = 0$, $z^t = \mathbf{0}_d$ is a fixed point of the message passing algorithm. We will refer to this as to the "un-informative fixed point." Note that condition (131) holds—among others—for the phase retrieval problem.

Vice versa, if $\mathbb{E}_G\{\partial_g p(y \mid G)\} \neq 0$, then $\mu_1 > 0$ even if $\mu_0 = 0$, for any $\delta > 0$. Thanks to Lemma 4, this implies that weak recovery is possible for all $\delta > 0$. Hence, we will assume that condition (131) holds.

The first result of this section establishes the following: For $\delta < \delta_{\mathrm{u}}$, the message passing algorithm fails even if the initial condition has a positive correlation with the unknown signal. We refer to Appendix E for its proof.

Theorem 5 (Message passing fails for $\delta < \delta_{\mathrm{u}}$) *Let $\boldsymbol{x} \in \mathbb{R}^d$ denote the unknown signal such that $\|\boldsymbol{x}\|_2 = \sqrt{d}$. Let $\boldsymbol{A} = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)^{\mathsf{T}} \in \mathbb{R}^{n\times d}$ with $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{i.i.d.} \mathsf{N}(\mathbf{0}_d, \boldsymbol{I}_d/d)$, and $\boldsymbol{y} = (y_1, \dots, y_n)$ with $y_i \sim p(\cdot \mid \langle \boldsymbol{x}, \boldsymbol{a}_i\rangle)$. Let $n/d \to \delta$ and define δ_{u} as in (42). Let $G \sim \mathsf{N}(0, 1)$ and assume that the condition (131) holds for any $y \in \mathbb{R}$.*

Consider the AMP algorithm defined in (126) and assume that the initial condition z^0 is such that

$$\lim_{n\to\infty} \frac{\langle z^0, \boldsymbol{x}\rangle}{\|z^0\|_2\, \|\boldsymbol{x}\|_2} = \epsilon. \tag{132}$$

Then, for any $\delta < \delta_{\mathrm{u}}$, there exists $\epsilon_0(\delta)$ such that for any $\epsilon \in (0, \epsilon_0(\delta))$, almost surely,

$$\lim_{t\to\infty} \lim_{n\to\infty} \frac{1}{d} \|z^t\|_2 = \mathbf{0}_d. \tag{133}$$

Next, we consider the case $\delta > \delta_{\mathrm{u}}$ and we linearize the iteration (126) around the non-informative fixed point.

Lemma 5 (Linearized AMP Equations) *Consider the one-step map defined in (126) and assume that condition (131) holds. Define $R_t = \|z^t\|_2 + \|\hat{z}^{t-1}\|_2$. Then, as $R_t \to 0$ and $q_t \to 0$, we obtain*

$$\begin{pmatrix} z^{t+1} \\ \hat{z}^t \end{pmatrix} = \boldsymbol{L}_n \begin{pmatrix} z^t \\ \hat{z}^{t-1} \end{pmatrix} + o(R_t) + R_t\, o_{q_t}(1), \tag{134}$$

where $\boldsymbol{L}_n \in \mathbb{R}^{(n+d)\times(n+d)}$ *is defined as*

$$\boldsymbol{L}_n = \begin{pmatrix} \boldsymbol{A}^\mathsf{T}\boldsymbol{J}\boldsymbol{A} & -\boldsymbol{A}^\mathsf{T}\boldsymbol{J}^2 \\ \boldsymbol{A} & -\boldsymbol{J} \end{pmatrix}, \tag{135}$$

and $\boldsymbol{J} \in \mathbb{R}^{n\times n}$ *is a diagonal matrix with entries* $j_i = \mathsf{F}'(0, y_i; 1)$ *for* $i \in [n]$*, with* F' *denoting the derivative of* F *with respect to the first argument.*

The second result of this section establishes the following: For $\delta > \delta_{\mathrm{u}}$, the un-informative fixed point is unstable for iteration (126), i.e., the matrix $\boldsymbol{L}_n$ has an eigenvalue that is larger than 1 in modulus. To do so, we will relate the matrix $\boldsymbol{J}$ appearing in (135) to the optimal pre-processing function defined in (45) [see Eq. (237) in Appendix F]. We refer to Appendix F for the complete proof.

Theorem 6 (Message passing escapes from un-informative fixed point for $\delta > \delta_{\mathrm{u}}$) *Let* $\boldsymbol{x} \in \mathbb{R}^d$ *denote the unknown signal such that* $\|\boldsymbol{x}\|_2 = \sqrt{d}$*. Let* $\boldsymbol{A} = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)^\mathsf{T} \in \mathbb{R}^{n\times d}$ *with* $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{i.i.d.} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$*, and* $\boldsymbol{y} = (y_1, \dots, y_n)$ *with* $y_i \sim p(\cdot \mid \langle \boldsymbol{x}, \boldsymbol{a}_i\rangle)$*. Let* $n/d \to \delta$ *and define* δ_{u} *as in* (42)*. Furthermore, assume that* (131) *holds for any* $y \in \mathbb{R}$*. Define* $\boldsymbol{L}_n \in \mathbb{R}^{(n+d)\times(n+d)}$ *as in* (135)*, where* $\boldsymbol{J} \in \mathbb{R}^{n\times n}$ *is a diagonal matrix with entries* $j_i = \mathsf{G}(0, y_i; 1)$ *for* $i \in [n]$*. Then, the eigenvalues of* $\boldsymbol{L}_n$ *are real and the largest of them, call it* $\lambda_1^{\boldsymbol{L}_n}$*, is such that, for any* $\delta > \delta_{\mathrm{u}}$*,*

$$\lim_{n\to\infty} \lambda_1^{\boldsymbol{L}_n} > 1. \tag{136}$$

7 Numerical Experiments

We focus on the phase retrieval problem and present some numerical results to illustrate the performance achieved by the proposed spectral method. First, we consider the case in which the unknown vector is chosen uniformly at random and the sensing vectors are Gaussian. Then, we consider the more practical scenario in which the unknown vector is an image and the sensing vectors come from a coded diffraction model.

7.1 Gaussian Sensing Vectors for Synthetic Data

Let us consider the complex case. In our experiments, the vector $\boldsymbol{x}$ is chosen uniformly at random on the d-dimensional complex sphere with radius $\sqrt{d}$, the sensing vectors $\{\boldsymbol{a}_i\}_{1\le i\le n}$ are i.i.d. circularly symmetric normal with variance $1/d$, and for $i \in [n]$, the measurement y_i is equal to $|\langle \boldsymbol{x}, \boldsymbol{a}_i\rangle|^2$. We take $d = 4096$ and the numerical simulations are averaged over $n_{\mathrm{sample}} = 40$ independent trials. The results are plotted in Fig. 2a.

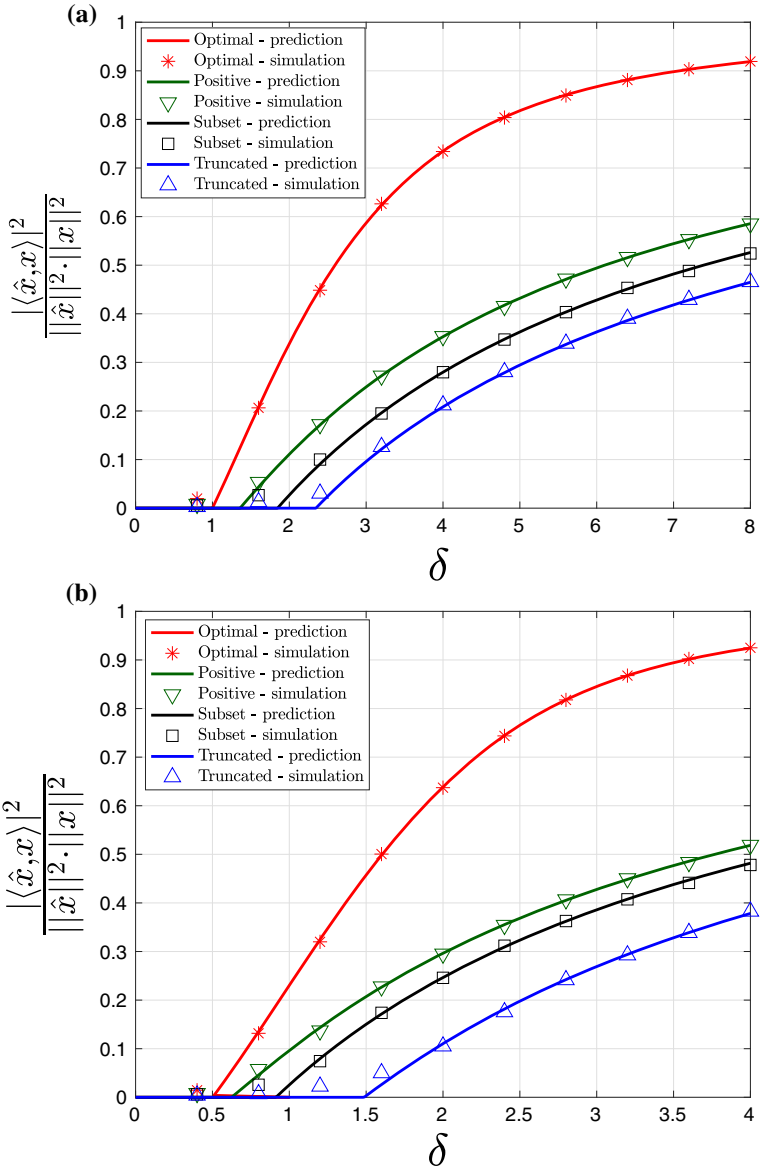

Fig. 2 Performance of the spectral method for the phase retrieval problem where the unknown vector is uniformly random on the sphere and the sensing vectors are Gaussian. On the x-axis, we have the ratio δ between the number of samples and the dimension of the signal; on the y-axis, we have the square of the normalized scalar product between the unknown signal $\boldsymbol{x}$ and the estimate $\hat{\boldsymbol{x}}$. Note that the proposed choice of the pre-processing function (red curve) provides a significant performance improvement with respect to the subset algorithm considered in [49,78] (black curve) and the truncated spectral initialization considered in [20,49] (blue curve). **a** Complex case, **b** real case (Color figure online)

The red curve corresponds to the proposed pre-processing function given by

$$\mathcal{T}(y) = \frac{y-1}{y+\sqrt{\tilde{\delta}}-1}. \tag{137}$$

We pick $\tilde{\delta} = 1.001$ and, as shown by the figure, weak recovery is possible for values of δ very close to 1.

The green curve corresponds to the pre-processing function given by

$$\mathcal{T}(y) = \max\left(\frac{y-1}{y+\sqrt{\tilde{\delta}}-1}, 0\right). \tag{138}$$

We add this plot in order to show that, by enforcing nonnegativity of the pre-processing function, we incur in a degradation of the performance of the spectral method.

The black curve corresponds to the pre-processing function given by

$$\mathcal{T}(y) = \begin{cases} 1, & \text{for } y > t, \\ 0, & \text{otherwise.} \end{cases} \tag{139}$$

This choice was proposed in [78] and it is also considered in [49], where the authors refer to it as the "subset algorithm." For each value of t, we can compute the smallest value of δ, call it $\delta^*(t)$, that yields a strictly positive scalar product according to the result of Lemma 2. Hence, we pick $t = 2$ that corresponds to the smallest value of $\delta^*(t)$ over $t \in \{0.25, 0.5, 0.75, \ldots, 10\}$.

The blue curve corresponds to the pre-processing function given by

$$\mathcal{T}(y) = \begin{cases} y, & \text{for } y \leq t, \\ 0, & \text{otherwise.} \end{cases} \tag{140}$$

This choice corresponds to the truncated spectral initialization proposed in [20] and it is also considered in [49], where the authors refer to it as the "trimming algorithm." For each value of t, we can compute the smallest value of δ, call it $\delta^*(t)$, that yields a strictly positive scalar product according to the result of Lemma 2. Hence, we pick $t = 5.25$ that corresponds to the smallest value of $\delta^*(t)$ over $t \in \{0.25, 0.5, 0.75, \ldots, 10\}$.

Note that the numerical simulations follow closely the theoretical prediction given by (84). Furthermore, the choice of the pre-processing function (137) yields a remarkable performance gain with respect to both the subset algorithm and the trimming algorithm.

Similar considerations apply to the real case. Here, the vector $\boldsymbol{x}$ is chosen uniformly at random on the d-dimensional real sphere with radius $\sqrt{d}$ and the sensing vectors $\{\boldsymbol{a}_i\}_{1\leq i\leq n}$ are i.i.d. normal with zero mean and variance $1/d$. We pick $d = 4096$ and $n_{\text{sample}} = 40$. The results are plotted in Fig. 2b. Again, the numerical simulations follow closely the theoretical prediction. The red curve corresponds to the pre-processing function given by (137), where we pick $\tilde{\delta} = 1.001$. Note that weak recovery is possible for values of δ very close to $1/2$. The green curve corresponds to the pre-processing

function given by (138). The blue curve corresponds to the pre-processing function given by (139), where we pick $t = 2$ which yields the smallest value of $\delta^*(t)$ over $t \in \{0.25, 0.5, 0.75, \ldots, 10\}$. The black curve corresponds to the pre-processing function given by (140), where we pick $t = 7$ which yields the smallest value of $\delta^*(t)$ over $t \in \{0.25, 0.5, 0.75, \ldots, 10\}$.

7.2 Coded Diffraction Model for Natural Images

We consider a model of coded diffraction patterns in which the sensing vectors $\{\boldsymbol{a}_r\}_{1 \le r \le n}$ are obtained as follows. For $t_1 \in [d_1]$ and $t_2 \in [d_2]$, denote by $a_r(t_1, t_2)$ the (t_1, t_2)th component of the vector $\boldsymbol{a}_r \in \mathbb{C}^d$, with $d = d_1 \cdot d_2$. Then,

$$a_r(t_1, t_2) = d_\ell(t_1, t_2) \cdot e^{i2\pi k_1 t_1/d_1} \cdot e^{i2\pi k_2 t_2/d_2}, \tag{141}$$

where i denotes the imaginary unit. The index $r \in [n]$ is associated with a pair (ℓ, k), with $\ell \in [L]$; the index $k \in [d]$ is associated with a pair (k_1, k_2) with $k_1 \in [d_1]$ and $k_2 \in [d_2]$. As usual, the measurement y_r of an unknown d-dimensional vector $\boldsymbol{x}$ is equal to $|\langle \boldsymbol{x}, \boldsymbol{a}_r \rangle|^2$. As an immediate consequence, the number of measurements n is equal to $L \cdot d$; therefore, $\delta = L \in \mathbb{N}$. In words, for a fixed ℓ, we collect the magnitude of the diffraction pattern of $\boldsymbol{x}$ modulated by $\boldsymbol{d}_\ell$. By varying ℓ and changing the modulation pattern $\boldsymbol{d}_\ell$, we generate L distinct views. The vectors $\{\boldsymbol{d}_\ell\}_{1 \le \ell \le L}$ are i.i.d. and their entries are also i.i.d. drawn uniformly from the set $\{1, -1, i, -i\}$.

We test the spectral method on the digital photograph represented in Fig. 1a. Each color image can be viewed as a $d_1 \times d_2 \times 3$ array. We run the spectral algorithm separately on the vectors $\boldsymbol{x}_j \in \mathbb{R}^d$, where $j \in \{1, 2, 3\}$. In our example, $d_1 = 820$ and $d_2 = 1280$. Let $\hat{\boldsymbol{x}}_j$ be the estimate of $\boldsymbol{x}_j$ provided by the spectral method. Then, we employ as a performance metric the average squared normalized scalar product

$$\frac{1}{3} \sum_{j=1}^{3} \frac{|\langle \hat{\boldsymbol{x}}_j, \boldsymbol{x}_j \rangle|^2}{\|\hat{\boldsymbol{x}}_j\|^2 \, \|\boldsymbol{x}_j\|^2}. \tag{142}$$

Note that the scalar product between the input and the measurement vectors can be interpreted as a two-dimensional Fourier transform; hence, it can be computed with an FFT algorithm. In order to evaluate the principal eigenvector of the data matrix, we use the power method with a random initialization, as described in Appendix B of [16]. As a stopping criterion, we require that one of the following two conditions is fulfilled: Either the number of iterations reaches the maximum value of 10000, or the modulus of the scalar product between the estimate at the current iteration T and at the iteration $T - 10$ is larger than $1 - 10^{-7}$.

The results are summarized in Fig. 3. The red curve corresponds to the proposed pre-processing function. In this case, the eigenvalues of the data matrix can be negative. Recall that the power method outputs the eigenvector associated with the largest eigenvalue *in modulus*, while we are interested in the eigenvector associated with the largest eigenvalue. To address this issue, we add to the data matrix a multiple α of the identity matrix. However, as α grows, the convergence of the power method becomes

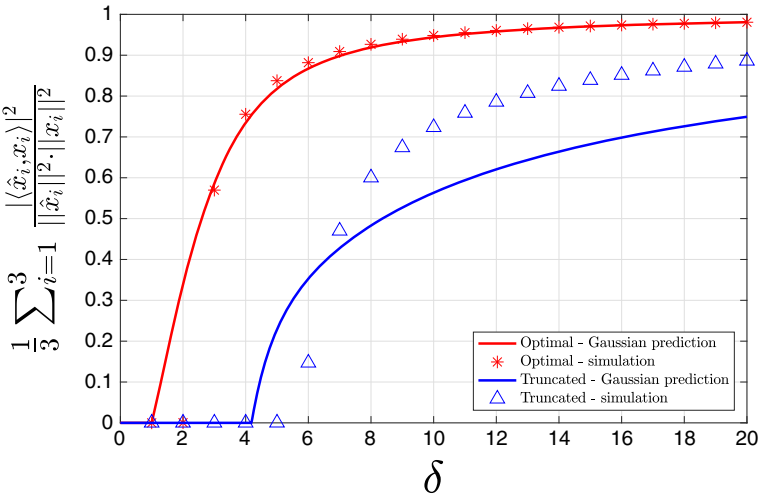

Fig. 3 Performance of the spectral method for the phase retrieval problem where the unknown vector is a digital photograph and the sensing vectors are obtained from a coded diffraction model. On the x-axis, we have the ratio δ between the number of samples and the dimension of the signal; on the y-axis, we have the square of the normalized scalar product between the unknown signal $\boldsymbol{x}$ and the estimate $\hat{\boldsymbol{x}}$ (averaged on the three RGB components of the image). Note that the proposed choice of the pre-processing function (red curve) provides a significant performance improvement with respect to the truncated spectral initialization considered in [20] (blue curve) (Color figure online)

slower and slower. In order to reduce the negative tail of the distribution of eigenvalues and, consequently, the value of α, we pick the pre-processing function given by

$$\mathcal{T}_1(y) = \max(\mathcal{T}(y), -M), \tag{143}$$

where $\mathcal{T}(y)$ is defined in (137), $\tilde{\delta} = 1.001$, and $M = 40$. In this way, by taking $\alpha = 100$, the largest eigenvalue in modulus has positive sign.

The blue curve corresponds to the truncated spectral initialization in [20], i.e., the pre-processing function is given by (140) with $t = 9$.

The numerical simulations for the optimal pre-processing function follow closely the theoretical predictions (84) obtained for a Gaussian measurement matrix, with the exception of the point $\delta = 2$. On the contrary, the numerical simulations for the truncated spectral initialization show a different behavior with respect to the Gaussian model. Our algorithm provides weak recovery of the original image for $\delta \geq 3$, while the truncated spectral initialization requires $\delta \geq 6$. Furthermore, for any value of δ, the proposed choice of the pre-processing function yields a better performance than the choice in [20]. For a visual representation of these results, see Fig. 1.

Acknowledgements We would like to thank the anonymous reviewers for their helpful comments.

A Proof of Corollary 1

We start by providing in Lemma 6 a less compact, but more explicit form of expression (10). This more explicit expression is employed to prove Lemma 7, which yields the value of δ_ℓ for the case of phase retrieval. Finally, we provide the proof of Corollary 1.

Lemma 6 (Explicit formula for $f(m)$—complex case) *Consider the function* $f : [0,1] \to \mathbb{R}$ *defined in* (10)*. Then,* $f(m)$ *is given by the following expression:*

$$f(m) = \int_{\mathbb{R}} \frac{\frac{1}{1-m}\int_0^{+\infty}\int_0^{+\infty} 4r_1 r_2 \cdot p(y \mid r_1)p(y \mid r_2) \cdot \exp\left(-\frac{r_1^2+r_2^2}{1-m}\right) \cdot I_0\left(\frac{2r_1 r_2\sqrt{m}}{1-m}\right) \mathrm{d}r_1 \mathrm{d}r_2}{\int_0^{+\infty} 2r \cdot p(y \mid r) \cdot \exp\left(-r^2\right) \mathrm{d}r} \mathrm{d}y, \tag{144}$$

where I_0 *denotes the modified Bessel function of the first kind, given by*

$$I_0(x) = \frac{1}{\pi}\int_0^{\pi} \exp\left(x\cos\theta\right) \mathrm{d}\theta. \tag{145}$$

Proof Let us rewrite G as

$$G = G^{(\mathrm{R})} + jG^{(\mathrm{I})}, \quad \text{with } (G^{(\mathrm{R})}, G^{(\mathrm{I})}) \sim \mathsf{N}\left(\mathbf{0}_d, \frac{1}{2}\boldsymbol{I}_2\right),$$

i.e., $G^{(\mathrm{R})}$ and $G^{(\mathrm{I})}$ are i.i.d. Gaussian random variables with mean 0 and variance $1/2$. Set

$$R = \sqrt{\left(G^{(\mathrm{R})}\right)^2 + \left(G^{(\mathrm{I})}\right)^2}.$$

Then, R follows a Rayleigh distribution with scale parameter $1/\sqrt{2}$, and hence

$$\mathbb{E}_G\left\{p(y \mid |G|)\right\} = \mathbb{E}_R\left\{p(y \mid R)\right\} = \int_0^{+\infty} 2r \cdot p(y \mid r) \cdot \exp\left(-r^2\right) \mathrm{d}r. \tag{146}$$

Let us rewrite (G_1, G_2) as

$$(G_1, G_2) = (G_1^{(\mathrm{R})} + jG_1^{(\mathrm{I})}, G_2^{(\mathrm{R})} + jG_2^{(\mathrm{I})}),$$

with

$$(G_1^{(\mathrm{R})}, G_2^{(\mathrm{R})}, G_1^{(\mathrm{I})}, G_2^{(\mathrm{I})}) \sim \mathsf{N}\left(\mathbf{0}_d, \frac{1}{2}\begin{bmatrix} 1 & \Re(c) & 0 & -\Im(c) \\ \Re(c) & 1 & \Im(c) & 0 \\ 0 & \Im(c) & 1 & \Re(c) \\ -\Im(c) & 0 & \Re(c) & 1 \end{bmatrix}\right),$$

and consider the following change in variables:

$$\begin{cases} G_1^{(\mathrm{R})} = R_1 \cos\theta_1 \\ G_2^{(\mathrm{R})} = R_2 \cos\theta_2 \\ G_1^{(\mathrm{I})} = R_1 \sin\theta_1 \\ G_2^{(\mathrm{I})} = R_2 \sin\theta_2 \end{cases}.$$

Then, after some algebra, we have that

$$\begin{aligned} &\mathbb{E}_{G_1,G_2}\left\{p(y \mid |G_1|) \cdot p(y \mid |G_2|)\right\} \\ &= \frac{1}{\pi^2(1-|c|^2)} \int_0^{+\infty}\int_0^{+\infty}\int_0^{2\pi}\int_0^{2\pi} r_1 r_2 \cdot p(y \mid r_1)p(y \mid r_2) \cdot \\ &\exp\left(-\frac{r_1^2 + r_2^2 - 2r_1 r_2\left(\Re(c)\cos(\theta_2-\theta_1) - \Im(c)\sin(\theta_2-\theta_1)\right)}{1-|c|^2}\right) \mathrm{d}r_1\,\mathrm{d}r_2\,\mathrm{d}\theta_1\,\mathrm{d}\theta_2. \end{aligned} \tag{147}$$

By writing $(\Re(c), \Im(c)) = (|c|\cos\theta_c, |c|\sin\theta_c)$ and by using definition (145), we can further simplify the RHS of (147) as

$$\begin{aligned} &\frac{1}{1-|c|^2}\int_0^{+\infty}\int_0^{+\infty} 4r_1 r_2 \cdot p(y \mid r_1)p(y \mid r_2) \cdot \exp\left(-\frac{r_1^2+r_2^2}{1-|c|^2}\right) \\ &\cdot I_0\left(\frac{2r_1 r_2|c|}{1-|c|^2}\right)\mathrm{d}r_1\mathrm{d}r_2. \end{aligned} \tag{148}$$

From (146) and (148), the claim easily follows. □

Lemma 7 (Computation of δ_ℓ for phase retrieval) *Computation of δ_ℓ for Phase Retrieval Let $\delta_\ell(\sigma^2)$ be defined as in (13) and assume that the distribution $p(\cdot \mid |G|)$ appearing in (10) is given by (9). Then,*

$$\lim_{\sigma\to 0}\delta_\ell(\sigma^2) = 1. \tag{149}$$

Proof For the special case of phase retrieval, it is possible to compute explicitly the function $f(m)$ defined in (10) and simplified in Lemma 6. Indeed,

$$\begin{aligned} &\int_0^{+\infty} 2r \cdot p_{\mathrm{PR}}(y \mid r) \cdot \exp\left(-r^2\right)\mathrm{d}r \\ &\overset{(a)}{=} \int_0^{+\infty} p_{\mathrm{PR}}(y \mid \sqrt{z}) \cdot \exp(-z)\,\mathrm{d}z \\ &\overset{(b)}{=} \int_{-\infty}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(y-z)^2}{2\sigma^2}\right) \cdot \exp(-z) \cdot H(z)\,\mathrm{d}z \\ &\overset{(c)}{=} \mathbb{E}_Z\left\{\exp(-Z)H(Z)\right\}, \end{aligned} \tag{150}$$

where in (a) we do the change in variables $z = r^2$; in (b) we use definition (9) and we define $H(x) = 1$ if $x \geq 0$ and $H(x) = 0$ otherwise; and in (c) we define $Z \sim \mathsf{N}(y, \sigma^2)$. In the limit $\sigma^2 \to 0$, we have that

$$\mathbb{E}_Z\{\exp(-Z)H(Z)\} \to \exp(-y)\cdot H(y), \tag{151}$$

by Lebesgue's dominated convergence theorem. Similarly,

$$\begin{aligned}
&\int_0^{+\infty}\int_0^{+\infty} 4r_1r_2\cdot p_{\mathrm{PR}}(y\mid r_1)p_{\mathrm{PR}}(y\mid r_2)\cdot\exp\left(-\frac{r_1^2+r_2^2}{1-m}\right)\cdot I_0\left(\frac{2r_1r_2\sqrt{m}}{1-m}\right)\mathrm{d}r_1\mathrm{d}r_2\\
&\overset{(a)}{=}\int_0^{+\infty}\int_0^{+\infty} p_{\mathrm{PR}}(y\mid\sqrt{z_1})p_{\mathrm{PR}}(y\mid\sqrt{z_2})\cdot\exp\left(-\frac{z_1+z_2}{1-m}\right)\cdot I_0\left(\frac{2\sqrt{z_1z_2}\sqrt{m}}{1-m}\right)\mathrm{d}z_1\mathrm{d}z_2\\
&\overset{(b)}{=}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^2\exp\left(-\frac{(y-z_1)^2+(y-z_2)^2}{2\sigma^2}\right)\\
&\quad\cdot\exp\left(-\frac{z_1+z_2}{1-m}\right)\cdot I_0\left(\frac{2\sqrt{z_1z_2}\sqrt{m}}{1-m}\right)\cdot H(z_1)H(z_2)\mathrm{d}z_1\mathrm{d}z_2\\
&\overset{(c)}{=}\mathbb{E}_{Z_1,Z_2}\left\{\exp\left(-\frac{Z_1+Z_2}{1-m}\right)\cdot I_0\left(\frac{2\sqrt{Z_1Z_2}\sqrt{m}}{1-m}\right)\cdot H(Z_1)H(Z_2)\right\},
\end{aligned}$$

where in (a) we do the change in variables $z_1=r_1^2$ and $z_2=r_2^2$; in (b) we use definition (9) and we define $H(x)=1$ if $x\geq 0$ and $H(x)=0$ otherwise; and in (c) we define $(Z_1,Z_2)\sim_{i.i.d.}\mathsf{N}(y,\sigma^2)$. In the limit $\sigma^2\to 0$, we have that

$$\begin{aligned}
&\mathbb{E}_{Z_1,Z_2}\left\{\exp\left(-\frac{Z_1+Z_2}{1-m}\right)\cdot I_0\left(\frac{2\sqrt{Z_1Z_2}\sqrt{m}}{1-m}\right)\cdot H(Z_1)H(Z_2)\right\}\\
&\quad\to\exp\left(-\frac{2y}{1-m}\right)\cdot I_0\left(\frac{2y\sqrt{m}}{1-m}\right)\cdot H(y),
\end{aligned}$$

by Lebesgue's dominated convergence theorem. As a result, by using (145), we obtain that

$$\begin{aligned}
f(m)&\overset{\sigma^2\to 0}{\longrightarrow}\frac{1}{1-m}\int_0^{+\infty}\exp\left(-\frac{2y}{1-m}\right)\cdot I_0\left(\frac{2y\sqrt{m}}{1-m}\right)\cdot\exp(y)\,\mathrm{d}y\\
&=\frac{1}{\pi(1-m)}\int_0^{\pi}\int_0^{+\infty}\exp\left(-y\left(\frac{1+m-2\sqrt{m}\cos\theta}{1-m}\right)\right)\mathrm{d}y\,\mathrm{d}\theta\\
&=\frac{1}{\pi}\int_0^{\pi}\frac{1}{1+m-2\sqrt{m}\cos\theta}\,\mathrm{d}\theta=\frac{1}{1-m}.
\end{aligned}$$

Consequently,

$$F_\delta(m)\overset{\sigma^2\to 0}{\longrightarrow}(1-\delta)\log(1-m),$$

which implies the desired result. □

Proof (Proof of Corollary 1) We follow the proof of Theorem 1 presented in Sect. 4. The first step is exactly the same, i.e., by applying Lemma 1, we show that (68) holds.

On the contrary, the second step requires some modifications, since the definition of the error metric is different. In particular, we will prove that

$$\frac{1}{d^2}\mathbb{E}_{\boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n}}\left\{\left\|\mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^*\right\}-\mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^*\mid \boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n}\right\}\right\|_F^2\right\}=o_n(1). \tag{152}$$

Furthermore, we have that

$$\begin{aligned}
&\left\|\mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^*\right\}-\mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^*\mid \boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n}\right\}\right\|_F^2+\mathbb{E}\left\{\left\|\boldsymbol{X}\boldsymbol{X}^*-\mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^*\mid \boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n}\right\}\right\|_F^2\right\}\\
&\overset{(a)}{\geq}\mathbb{E}\left\{\left\|\mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^*\right\}-\boldsymbol{X}\boldsymbol{X}^*\right\|_F^2\right\}\\
&\overset{(b)}{=}\mathbb{E}\left\{\left\|\boldsymbol{I}_d-\boldsymbol{X}\boldsymbol{X}^*\right\|_F^2\right\}\\
&\overset{(c)}{=}\mathbb{E}\left\{\operatorname{trace}\left(\boldsymbol{I}_d-2\boldsymbol{X}\boldsymbol{X}^*+\boldsymbol{X}\boldsymbol{X}^*\boldsymbol{X}\boldsymbol{X}^*\right)\right\}\\
&\overset{(d)}{=}d-2d+d^2=d^2-d,
\end{aligned} \tag{153}$$

where in (a) we use the triangle inequality; in (b) we use that $\mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^*\right\}=\boldsymbol{I}_d$ by Lemma 12; in (c) we use that, for any matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|_F=\sqrt{\operatorname{trace}(\boldsymbol{A}\boldsymbol{A}^*)}$; and in (d) we use that $\mathbb{E}\left\{\operatorname{trace}\left(\boldsymbol{X}\boldsymbol{X}^*\boldsymbol{X}\boldsymbol{X}^*\right)\right\}=\mathbb{E}\left\{\|\boldsymbol{X}\|^4\right\}=d^2$. By applying (152) and (153), the proof of Corollary 1 is complete.

Let us now give the proof of (152). Similarly to (71), we have that

$$\begin{aligned}
&I(Y_{n+1};\boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n}|\boldsymbol{A}_{n+1})\\
&\quad\geq\frac{1}{2K^2}\cdot\mathbb{E}_{\boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n+1}}\left\{\left|\int_{\mathbb{C}^d}p(\boldsymbol{x}\mid\boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n})\int_{\mathbb{R}}p_{\mathrm{PR}}(y_{n+1}\mid\boldsymbol{x},\right.\right.\\
&\qquad\left.\boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n+1})\varphi_{\mathrm{PR}}(y_{n+1})\mathrm{d}y_{n+1}\mathrm{d}\boldsymbol{x}\right.\\
&\qquad\left.\left.-\int_{\mathbb{C}^d}p(\boldsymbol{x})\int_{\mathbb{R}}p_{\mathrm{PR}}(y_{n+1}\mid\boldsymbol{x},\boldsymbol{A}_{n+1})\varphi_{\mathrm{PR}}(y_{n+1})\,\mathrm{d}y_{n+1}\,\mathrm{d}\boldsymbol{x}\right|^2\right\},
\end{aligned} \tag{154}$$

where we define $\varphi_{\mathrm{PR}}(x)=x$ for $|x|\leq M$, and $\varphi_{\mathrm{PR}}(x)=M\cdot\operatorname{sign}(x)$ otherwise. Then,

$$\begin{aligned}
&\int_{\mathbb{C}^d}p(\boldsymbol{x}\mid\boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n})\int_{\mathbb{R}}p_{\mathrm{PR}}(y_{n+1}\mid\boldsymbol{x},\boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n+1})\cdot\varphi_{\mathrm{PR}}(y_{n+1})\,\mathrm{d}y_{n+1}\,\mathrm{d}\boldsymbol{x}\\
&\overset{(a)}{=}\int_{\mathbb{C}^d}p(\boldsymbol{x}\mid\boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n})\int_{\mathbb{R}}p_{\mathrm{PR}}(y_{n+1}\mid\boldsymbol{x},\boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n+1})\cdot y_{n+1}\,\mathrm{d}y_{n+1}\,\mathrm{d}\boldsymbol{x}+E_1\\
&\overset{(b)}{=}\int_{\mathbb{C}^d}p(\boldsymbol{x}\mid\boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n})\cdot|\langle\boldsymbol{A}_{n+1},\boldsymbol{x}\rangle|^2\,\mathrm{d}\boldsymbol{x}+E_1\\
&\overset{(c)}{=}\langle\boldsymbol{A}_{n+1},\left(\int_{\mathbb{C}^d}p(\boldsymbol{x}\mid\boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n})\cdot\boldsymbol{x}\boldsymbol{x}^*\,\mathrm{d}\boldsymbol{x}\right)\boldsymbol{A}_{n+1}\rangle+E_1\\
&=\langle\boldsymbol{A}_{n+1},\mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^*\mid\boldsymbol{Y}_{1:n},\boldsymbol{A}_{1:n}\right\}\boldsymbol{A}_{n+1}\rangle+E_1,
\end{aligned} \tag{155}$$

where in (a) we set

$$E_1 = \int_{\mathbb{C}^d} p(\boldsymbol{x} \mid \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n}) \int_{\mathbb{R}} p_{\mathrm{PR}}(y_{n+1} \mid \boldsymbol{x}, \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n+1}) \cdot (\varphi_{\mathrm{PR}}(y_{n+1}) - y_{n+1}) \,\mathrm{d}y_{n+1} \,\mathrm{d}\boldsymbol{x},$$

in (b) we use definition (9), and in (c) we use that $|\langle \boldsymbol{A}_{n+1}, \boldsymbol{x}\rangle|^2 = \langle \boldsymbol{A}_{n+1}, \boldsymbol{x}\boldsymbol{x}^* \boldsymbol{A}_{n+1}\rangle$. Similarly, we have that

$$\begin{aligned} &\int_{\mathbb{C}^d} p(\boldsymbol{x}) \int_{\mathbb{R}} p_{\mathrm{PR}}(y_{n+1} \mid \boldsymbol{x}, \boldsymbol{A}_{n+1}) \varphi_{\mathrm{PR}}(y_{n+1}) \,\mathrm{d}y_{n+1} \,\mathrm{d}\boldsymbol{x} \\ &\quad = \langle \boldsymbol{A}_{n+1}, \mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^*\right\} \boldsymbol{A}_{n+1}\rangle + E_2, \end{aligned} \tag{156}$$

with

$$E_2 = \int_{\mathbb{C}^d} p(\boldsymbol{x}) \int_{\mathbb{R}} p_{\mathrm{PR}}(y_{n+1} \mid \boldsymbol{x}, \boldsymbol{A}_{n+1}) \cdot (\varphi_{\mathrm{PR}}(y_{n+1}) - y_{n+1}) \,\mathrm{d}y_{n+1} \,\mathrm{d}\boldsymbol{x}.$$

By applying (155) and (156), we can rewrite the RHS of (154) as

$$\begin{aligned} &\frac{1}{2K^2} \cdot \mathbb{E}_{\boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n}} \mathbb{E}_{\boldsymbol{A}_{n+1}} \left\{ \left| \langle \boldsymbol{A}_{n+1}, \left(\mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^* \mid \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n}\right\} - \mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^*\right\}\right) \boldsymbol{A}_{n+1}\rangle + E_1 - E_2 \right|^2 \right\} \\ &\quad \geq \frac{1}{2K^2} \cdot \mathbb{E}_{\boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n}} \left(\mathbb{E}_{\boldsymbol{A}_{n+1}} \left\{ |\langle \boldsymbol{A}_{n+1}, \boldsymbol{M}\boldsymbol{A}_{n+1}\rangle|^2 \right\} - \mathbb{E}_{\boldsymbol{A}_{n+1}} \left\{ |E_1|^2 \right\} - \mathbb{E}_{\boldsymbol{A}_{n+1}} \left\{ |E_2|^2 \right\} \right), \end{aligned} \tag{157}$$

where we define $\boldsymbol{M} = \mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^* \mid \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n}\right\} - \mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^*\right\}$. As K goes large, $\mathbb{E}_{\boldsymbol{A}_{n+1}} \left\{ |E_i|^2 \right\}$ tends to 0, for $i \in \{1, 2\}$. Furthermore, we have that

$$\begin{aligned} \mathbb{E}_{\boldsymbol{A}_{n+1}} \left\{ |\langle \boldsymbol{A}_{n+1}, \boldsymbol{M}\boldsymbol{A}_{n+1}\rangle|^2 \right\} &\overset{\text{(a)}}{=} \sum_{i,j,k,l=1}^{d} M_{ij} M_{kl}^* \cdot \frac{1}{d^2} (\delta_{ij} \cdot \delta_{kl} + \delta_{il} \cdot \delta_{jk}) \\ &= \frac{1}{d^2} \left(|\mathrm{trace}(\boldsymbol{M})|^2 + \|\boldsymbol{M}\|_F^2 \right) \\ &\overset{\text{(b)}}{=} \frac{1}{d^2} \|\boldsymbol{M}\|_F^2 , \end{aligned} \tag{158}$$

where in (a) we use the following definition of the Kronecker delta:

$$\delta_{ab} = \begin{cases} 1, & \text{if } a = b, \\ 0, & \text{otherwise}, \end{cases} \tag{159}$$

and in (b) we use that

$$\begin{aligned}\operatorname{trace}(\boldsymbol{M}) &= \sum_{i=1}^{d} \left(\mathbb{E}\left\{ |X_i|^2 \mid \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n} \right\} - \mathbb{E}\left\{ |X_i|^2 \right\} \right) \\ &= \mathbb{E}\left\{ \sum_{i=1}^{d} |X_i|^2 \mid \boldsymbol{Y}_{1:n}, \boldsymbol{A}_{1:n} \right\} - \mathbb{E}\left\{ \sum_{i=1}^{d} |X_i|^2 \right\} = 0. \end{aligned} \tag{160}$$

As a result, we conclude that (152) holds. □

B Proof of Corollary 2

First, we evaluate the RHS of (20), as well as the scaling between δ_{u} and σ^2 when $\sigma^2 \to 0$. Then, we give the proof of Corollary 2.

Lemma 8 (Computation of δ_{u} for phase retrieval) *Let $\delta_{\mathrm{u}}(\sigma^2)$ be defined as in (20) and assume that the distribution $p(\cdot \mid |g|)$ is given by (9). Then,*

$$\delta_{\mathrm{u}}(\sigma^2) = 1 + \sigma^2 + o(\sigma^2). \tag{161}$$

Proof The proof boils down to computing expected values and integrals. By using (146) and (150), we immediately obtain that

$$\begin{aligned}\mathbb{E}_G\left\{ p_{\mathrm{PR}}(y \mid |G|) \right\} &= \int_0^{+\infty} 2r \cdot p_{\mathrm{PR}}(y \mid r) \cdot \exp\left(-r^2\right) \mathrm{d}r \\ &= \exp(-y)\, \mathbb{E}_X\left\{ \exp(-\sigma X) H(y + \sigma X) \right\}, \end{aligned}$$

where $X \sim \mathsf{N}(0, 1)$. By computing explicitly the expectation, we deduce that

$$\mathbb{E}_G\left\{ p_{\mathrm{PR}}(y \mid |G|) \right\} = \frac{1}{2} \exp\left(-y + \frac{\sigma^2}{2}\right) \operatorname{erfc}\left(\frac{1}{\sqrt{2}} \left(-\frac{y}{\sigma} + \sigma \right) \right), \tag{162}$$

where $\operatorname{erfc}(\cdot)$ is the complimentary error function. Similarly, we have that

$$\begin{aligned}\mathbb{E}_G\left\{ p_{\mathrm{PR}}(y \mid |G|)(|G|^2 - 1) \right\} &= \int_0^{+\infty} 2(r^3 - r) p_{\mathrm{PR}}(y \mid r) \cdot \exp\left(-r^2\right) \mathrm{d}r \\ &= \exp(-y)\, \mathbb{E}_X\left\{ \exp(-\sigma X) H(y + \sigma X)(y - 1 + \sigma X) \right\} \\ &= \frac{\sigma}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2\sigma^2} \right) + \frac{1}{2}(y - 1 \\ &\quad - \sigma^2) \exp\left(-y + \frac{\sigma^2}{2}\right) \operatorname{erfc}\left(\frac{1}{\sqrt{2}} \left(-\frac{y}{\sigma} + \sigma \right) \right). \end{aligned} \tag{163}$$

Thus, by using (162) and (163), after some manipulations, we obtain that

$$\begin{aligned}\frac{1}{\delta_{\mathrm{u}}} &= \int_{\mathbb{R}} \frac{\left(\mathbb{E}_G\left\{p_{\mathrm{PR}}(y \mid |G|)(|G|^2-1)\right\}\right)^2}{\mathbb{E}_G\left\{p(y_{\mathrm{PR}} \mid |G|)\right\}}\,\mathrm{d}y \\ &= \int_{\mathbb{R}} \frac{\sigma^2}{2\pi}\exp\left(y-\frac{\sigma^2}{2}-\frac{y^2}{\sigma^2}\right)\frac{2}{\operatorname{erfc}\left(\frac{1}{\sqrt{2}}\left(-\frac{y}{\sigma}+\sigma\right)\right)}\mathrm{d}y \\ &\quad + \int_{\mathbb{R}} \frac{2\sigma}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\sigma^2}\right)(y-1-\sigma^2)\mathrm{d}y \\ &\quad + \int_{\mathbb{R}} \frac{1}{2}\exp\left(-y+\frac{\sigma^2}{2}\right)(y-1-\sigma^2)^2\operatorname{erfc}\left(\frac{1}{\sqrt{2}}\left(-\frac{y}{\sigma}+\sigma\right)\right)\mathrm{d}y.\end{aligned} \tag{164}$$

By performing the change in variables $t = -y/\sigma + \sigma$, we simplify the first integral in the RHS of (164) as

$$\begin{aligned}&\int_{\mathbb{R}} \frac{\sigma^2}{2\pi}\exp\left(y-\frac{\sigma^2}{2}-\frac{y^2}{\sigma^2}\right)\frac{2}{\operatorname{erfc}\left(\frac{1}{\sqrt{2}}\left(-\frac{y}{\sigma}+\sigma\right)\right)}\mathrm{d}y \\ &\quad = \frac{2\sigma^3}{2\pi}\int_{\mathbb{R}} \frac{\exp\left(-t^2\right)}{\operatorname{erfc}\left(\frac{t}{\sqrt{2}}\right)}\exp\left(\sigma t-\frac{\sigma^2}{2}\right)\mathrm{d}t = o(\sigma^2),\end{aligned} \tag{165}$$

where in the last equality we use that the integral

$$\int_{\mathbb{R}} \frac{\exp\left(-t^2\right)}{\operatorname{erfc}\left(\frac{t}{\sqrt{2}}\right)}\mathrm{d}t$$

is finite. The other two integrals in the RHS of (164) can be expressed in closed form as

$$\int_{\mathbb{R}} \frac{2\sigma}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2\sigma^2}\right)(y-1-\sigma^2)\mathrm{d}y = -2\sigma^2(1+\sigma^2), \tag{166}$$

$$\begin{aligned}&\int_{\mathbb{R}} \frac{1}{2}\exp\left(-y+\frac{\sigma^2}{2}\right)(y-1-\sigma^2)^2\operatorname{erfc}\left(\frac{1}{\sqrt{2}}\left(-\frac{y}{\sigma}+\sigma\right)\right)\mathrm{d}y \\ &\quad = \frac{\sigma}{2}\exp\left(-\frac{\sigma^2}{2}\right)\int_{\mathbb{R}} \exp\left(\sigma t\right)(\sigma t+1)^2\operatorname{erfc}\left(\frac{t}{\sqrt{2}}\right)\mathrm{d}t = 1+\sigma^2+\sigma^4.\end{aligned} \tag{167}$$

By combining (164), (165), (166), and (167), the result follows. □

Proof (Proof of Corollary 2) Pick σ sufficiently small. Let $G \sim \mathsf{CN}(0,1)$, $Y \sim p_{\mathrm{PR}}(\cdot \mid |G|)$ and $Z = \mathcal{T}(Y)$, where p_{PR} is defined in (9) and $\mathcal{T}$ is a pre-processing function (possibly dependent on σ) that we will choose later on. Assume that

(1) $\mathcal{T}(y)$ is upper and lower bounded by constants independent of σ;

(2) $\mathbb{P}(Z = 0) \leq c_1 < 1$ and c_1 is independent of σ;
(3) condition (82) holds.

Then, by Lemma 2, we have that, as $n \to \infty$,

$$\frac{|\langle \hat{\boldsymbol{x}}, \boldsymbol{x} \rangle|^2}{\|\hat{\boldsymbol{x}}\|_2^2 \|\boldsymbol{x}\|^2} \xrightarrow{\text{a.s.}} \rho = \begin{cases} 0, & \text{if } \psi_\delta'(\lambda_\delta^*) \leq 0, \\ \dfrac{\psi_\delta'(\lambda_\delta^*)}{\psi_\delta'(\lambda_\delta^*) - \phi'(\lambda_\delta^*)}, & \text{if } \psi_\delta'(\lambda_\delta^*) > 0, \end{cases} \tag{168}$$

where λ_δ^* is the unique solution of the equation $\zeta_\delta(\lambda) = \phi(\lambda)$, and ϕ, ψ_δ, and ζ_δ are defined in (78), (79), and (81), respectively.

Let τ be the supremum of the support of Z. Assume also that, for $\bar{\lambda}_\delta < \lambda_\delta^*$,

(4) $\tau \geq c_2 > 0$ and c_2 is independent of σ;
(5) $\phi'(\lambda_\delta^*)$ is lower bounded by a constant independent of σ;
(6) $\min_{\lambda \in (\min(\bar{\lambda}_\delta, \lambda_\delta^*))} \psi_\delta''(\lambda)$ is lower bounded by a strictly positive constant independent of σ.

Let $\bar{\lambda}_\delta$ be the point at which ψ_δ attains its minimum. Then,

$$\begin{aligned} \phi(\bar{\lambda}_\delta) - \psi_\delta(\bar{\lambda}_\delta) &\overset{(a)}{=} \phi(\bar{\lambda}_\delta) - \phi(\lambda_\delta^*) + \zeta_\delta(\lambda_\delta^*) - \psi_\delta(\bar{\lambda}_\delta) \\ &\overset{(b)}{=} \phi(\bar{\lambda}_\delta) - \phi(\lambda_\delta^*) + \zeta_\delta(\lambda_\delta^*) - \zeta_\delta(\bar{\lambda}_\delta) \\ &\overset{(c)}{=} \big(\zeta_\delta'(x_1) - \phi'(x_1)\big) \cdot (\lambda_\delta^* - \bar{\lambda}_\delta) \\ &\overset{(d)}{=} \frac{\big(\zeta_\delta'(x_1) - \phi'(x_1)\big)}{\psi_\delta''(x_2)} \cdot \big(\psi_\delta'(\lambda_\delta^*) - \psi_\delta'(\bar{\lambda}_\delta)\big) \\ &\overset{(e)}{=} \frac{\big(\zeta_\delta'(x_1) - \phi'(x_1)\big)}{\psi_\delta''(x_2)} \cdot \psi_\delta'(\lambda_\delta^*) \\ &\overset{(f)}{\leq} c_3 \cdot \psi_\delta'(\lambda_\delta^*), \end{aligned} \tag{169}$$

where in (a) we use that $\zeta_\delta(\lambda_\delta^*) = \phi(\lambda_\delta^*)$; in (b) we use that $\zeta_\delta(\bar{\lambda}_\delta) = \psi_\delta(\bar{\lambda}_\delta)$; (c) holds for some $x_1 \in (\bar{\lambda}_\delta, \lambda_\delta^*)$ by the mean value theorem; (d) holds for some $x_2 \in (\bar{\lambda}_\delta, \lambda_\delta^*)$ by the mean value theorem; and in (e) we use that $\psi_\delta'(\bar{\lambda}_\delta) = 0$. Note that (f) holds for some constant c_3 independent of σ, as $\zeta_\delta'(x_1) \geq 0$, $\psi_\delta''(x_2)$ is bounded, and $\phi'(x_2) < 0$ since $\mathbb{P}(Z = 0) < 1$.

As $\phi'(\lambda_\delta^*)$ is bounded, from (168) and (169) we deduce that

$$\rho \geq c_4 \cdot \big(\phi(\bar{\lambda}_\delta) - \psi_\delta(\bar{\lambda}_\delta)\big), \tag{170}$$

for some constant c_4 independent of σ. Notice that, if $\lambda_\delta^* \leq \bar{\lambda}_\delta$, then the right-hand side is non-positive and hence the lower bound still holds.

As $\tau > 0$, we also have that $\bar{\lambda}_\delta > 0$. Consider now the matrix $\boldsymbol{D}_n' = \boldsymbol{D}_n/\alpha$ for some $\alpha > 0$. Then, the principal eigenvector of $\boldsymbol{D}_n'$ is equal to the principal eigenvector of

$\boldsymbol{D}_n$. Hence, we can assume without loss of generality that $\bar{\lambda}_\delta = 1$. This condition can be rewritten as

$$\mathbb{E}\left\{\frac{Z^2}{(1-Z)^2}\right\} = \frac{1}{\delta}, \tag{171}$$

and (170) can be rewritten as

$$\rho \geq c_4 \cdot \left(\mathbb{E}\left\{\frac{Z(|G|^2-1)}{1-Z}\right\} - \frac{1}{\delta}\right). \tag{172}$$

We set

$$\mathcal{T}(y) = \mathcal{T}^*_\delta(y,\sigma) \triangleq \frac{y_+ - 1}{y_+ + \sqrt{\delta\, c(\sigma)} - 1}, \tag{173}$$

where $y_+ = \max(y, 0)$ and $c(\sigma)$ is a function of σ to be set as to satisfy Eq. (171). By substituting (173) into (171), we get

$$\mathbb{E}\left\{\frac{Z^2}{(1-Z)^2}\right\} = \frac{1}{\delta c(\sigma)}\mathbb{E}\big\{(Y_+ - 1)^2\big\}. \tag{174}$$

Hence, Eq. (171) is satisfied by

$$c(\sigma) = \mathbb{E}\big\{(Y_+ - 1)^2\big\} = \mathbb{E}\Big\{\big((|G|^2 + \sigma W)_+ - 1\big)^2\Big\}, \tag{175}$$

where $W \sim \mathsf{N}(0,1)$ is independent of G. Therefore, $c(\sigma)$ is always well defined and, by dominated convergence, $c(\sigma) \to c(0) = 1$, as $\sigma \to 0$. Furthermore,

$$\mathbb{E}\left\{\frac{Z(|G|^2-1)}{1-Z}\right\} = \frac{1}{\sqrt{\delta c(\sigma)}}\mathbb{E}\big\{(Y_+ - 1)(|G|^2 - 1)\big\}. \tag{176}$$

By applying again dominated convergence, we get

$$\lim_{\sigma\to 0}\mathbb{E}\left\{\frac{Z(|G|^2-1)}{1-Z}\right\} = \frac{1}{\sqrt{\delta}}\mathbb{E}\big\{(|G|^2-1)^2\big\} = \frac{1}{\sqrt{\delta}}. \tag{177}$$

Hence, by using (170), we get that, for $\delta > 1$ and $\sigma \leq \sigma_1(\delta)$,

$$\liminf_{n\to\infty} \frac{|\langle \hat{\boldsymbol{x}}_\sigma, \boldsymbol{x}\rangle|^2}{\|\hat{\boldsymbol{x}}_\sigma\|_2^2\, \|\boldsymbol{x}\|_2^2} \geq c_5\Big(\frac{1}{\sqrt{\delta}} - \frac{1}{\delta}\Big) > 0, \tag{178}$$

where $\hat{\boldsymbol{x}}_\sigma$ denotes the spectral estimator corresponding to the pre-processing function (173). Let us now verify that, by setting $\mathcal{T} = \mathcal{T}^*_\delta$, the requirements stated above are fulfilled. As $\delta > 1$, the function $\mathcal{T}$ is bounded by constants independent of σ. It is also

clear that conditions (2) and (4) hold. Furthermore, conditions (5) and (6) follow by showing that $\phi(\lambda)$, $\psi_\delta(\lambda)$ have well-defined uniform limits as $\sigma \to 0$ that satisfy those conditions: This can be proved by one more application of dominated convergence.

In order to show that condition (3) holds, we follow the argument presented at the end of the proof of Lemma 2. First, we add a point mass with associated probability at most ϵ_1, which immediately implies that (82) is satisfied. Then, by applying the Davis–Kahan theorem [23], we show that we can take $\epsilon_1 = 0$.

This proves the claim of the corollary for the pre-processing function $\mathcal{T}^*_\delta(y, \sigma)$, defined in (173). Let us now prove that the same conclusion holds for $\mathcal{T}^*_\delta(y)$ defined in (25). Let

$$f_a(x) = \frac{x+1}{x+a}. \tag{179}$$

Then, for any $x, a \in \mathbb{R}_{\geq 0}$,

$$|f_a'(x)| = \frac{|a-1|}{(x+a)^2} \leq \max\left(1, \frac{1}{x^2}\right). \tag{180}$$

Therefore, since $\mathcal{T}^*_\delta(y, \sigma) = 1 - f_{y_+}(\sqrt{\delta c(\sigma)} - 1)$, we have that

$$\sup_{y \in \mathbb{R}} \left|\mathcal{T}^*_\delta(y, \sigma) - \mathcal{T}^*_\delta(y)\right| \leq \frac{1}{\left(\min(\sqrt{\delta c(\sigma)} - 1, \sqrt{\delta} - 1, 1)\right)^2} \sqrt{\delta} \cdot \left|\sqrt{c(\sigma)} - 1\right|. \tag{181}$$

Denote by $\boldsymbol{D}_n(\sigma)$ and $\boldsymbol{D}_n$ the matrices constructed with the pre-processing functions $\mathcal{T}^*_\delta(y, \sigma)$ and $\mathcal{T}^*_\delta(y)$, respectively. It follows that, for any $\delta > 1$, there exists a function $\Delta(\sigma)$ with $\Delta(\sigma) \to 0$ as $\sigma \to 0$ such that

$$\|\boldsymbol{D}_n(\sigma) - \boldsymbol{D}_n\|_{\mathrm{op}} \leq \Delta(\sigma). \tag{182}$$

Hence, by applying again the Davis–Kahan theorem, we conclude that, for all $\delta > 1$ and $\sigma \leq \sigma_2(\delta)$,

$$\liminf_{n \to \infty} \frac{|\langle \hat{\boldsymbol{x}}, \boldsymbol{x} \rangle|^2}{\|\hat{\boldsymbol{x}}\|_2^2 \, \|\boldsymbol{x}\|_2^2} \geq c_5\left(\frac{1}{\sqrt{\delta}} - \frac{1}{\delta}\right) > 0, \tag{183}$$

where $\hat{\boldsymbol{x}}$ is the estimator corresponding to the pre-processing function $\mathcal{T}^*_\delta(y)$. □

C Auxiliary Lemmas

Lemma 9 (Distribution of scalar product of two unit complex vectors) *Let* $\boldsymbol{x}_1, \boldsymbol{x}_2 \sim_{i.i.d.} \mathrm{Unif}(\mathsf{S}_{\mathbb{C}}^{d-1})$ *and define* $M = |\langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle|^2$. *Then,*

$$M \sim \mathrm{Beta}(1, d-1). \tag{184}$$

Proof Without loss of generality, we can pick $\boldsymbol{x}_2$ to be the first element of the canonical base of $\mathbb{C}^d$. Thus, M is equal to the squared modulus of the first component of $\boldsymbol{x}_1$. Furthermore, we can think to $\boldsymbol{x}_1$ as being chosen uniformly at random on the $2d$-dimensional real sphere with radius 1. Note that, by taking a vector of i.i.d. standard Gaussian random variables and dividing it by its norm, we obtain a vector uniformly random on the sphere of radius 1. Hence,

$$M = \frac{U_1^2 + U_2^2}{\sum_{i=1}^{2d} U_i^2}, \quad \text{with } \{U_i\}_{1 \le i \le 2d} \sim_{i.i.d.} \mathsf{N}(0,1).$$

Set $A = U_1^2 + U_2^2$ and $B = \sum_{i=3}^{2d} U_i^2$. Then, A and B are independent, A follows a gamma distribution with shape 1 and scale 2, i.e., $A \sim \Gamma(1,2)$, and B follows a Gamma distribution with shape $d-1$ and scale 2, i.e., $B \sim \Gamma(d-1,2)$. Thus, we conclude that

$$M = \frac{A}{A+B} \sim \mathrm{Beta}(1, d-1),$$

which proves the claim. □

Lemma 10 (Distribution of scalar product of two unit real vectors) *Let* $\boldsymbol{x}_1, \boldsymbol{x}_2 \sim_{i.i.d.} \mathrm{Unif}(\mathsf{S}_{\mathbb{R}}^{d-1})$ *and define* $M = \langle \boldsymbol{x}_1, \boldsymbol{x}_2 \rangle$. *Then, the distribution of* M *is given by*

$$p(m) = \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi}\,\Gamma(\frac{d-1}{2})}(1-m^2)^{\frac{d-3}{2}}, \qquad m \in [-1,1]. \tag{185}$$

Proof Without loss of generality, we can pick $\boldsymbol{x}_2$ to be the first element of the canonical base of $\mathbb{R}^d$. Thus, M is equal to the first component of $\boldsymbol{x}_1$. Note that, by taking a vector of i.i.d. standard Gaussian random variables and dividing it by its norm, we obtain a vector uniformly random on the sphere of radius 1. Hence,

$$M^2 = \frac{U_1^2}{\sum_{i=1}^{d} U_i^2}, \quad \text{with } \{U_i\}_{1 \le i \le d} \sim_{i.i.d.} \mathsf{N}(0,1).$$

Set $A = U_1^2$ and $B = \sum_{i=2}^{d} U_i^2$. Then, A and B are independent, A follows a gamma distribution with shape $1/2$ and scale 2, i.e., $A \sim \Gamma(1/2,2)$, and B follows a Gamma distribution with shape $(d-1)/2$ and scale 2, i.e., $B \sim \Gamma((d-1)/2,2)$. Thus, we obtain that

$$M^2 = \frac{A}{A+B} \sim \mathrm{Beta}(1/2, (d-1)/2).$$

A change in variable and the observation that the distribution of M is symmetric around 0 immediately let us conclude that, for $m \in [-1,1]$,

$$p(m) = c \cdot (1-m^2)^{\frac{d-3}{2}}, \tag{186}$$

where the normalization constant c is given by

$$c = \left(\int_{-1}^{1} (1 - m^2)^{\frac{d-3}{2}} \mathrm{d}m \right)^{-1} = \frac{\Gamma(\frac{d}{2})}{\sqrt{\pi}\,\Gamma(\frac{d-1}{2})}.$$

□

Lemma 11 (Laplace's method) *Let* $F : [0, 1] \to \mathbb{R}$ *be such that*

- F *is continuous;*
- $F(x) < 0$ *for* $x \in (0, 1]$*;*
- $F(0) = 0$.

Then,

$$\lim_{n \to +\infty} \int_0^1 \exp\left(n \cdot F(x)\right) \mathrm{d}x = 0. \tag{187}$$

Proof Pick $\epsilon > 0$ and separate the integral into two parts:

$$\int_0^1 \exp\left(n \cdot F(x)\right) \mathrm{d}x = \int_0^\epsilon \exp\left(n \cdot F(x)\right) \mathrm{d}x + \int_\epsilon^1 \exp\left(n \cdot F(x)\right) \mathrm{d}x.$$

Now, the first integral is at most ϵ since $F(x) \leq 0$ for any $x \in [0, 1]$, and the second integral tends to 0 as $n \to +\infty$ since $F(x) < 0$ for $x \in (0, 1]$. Thus, the claim immediately follows. □

Lemma 12 (Second moment of uniform vector on complex sphere) *Let* $\boldsymbol{x} \sim \mathrm{Unif}(\sqrt{d}\mathbb{S}_{\mathbb{C}}^{d-1})$. *Then,*

$$\mathbb{E}\left\{\boldsymbol{X}\boldsymbol{X}^*\right\} = \boldsymbol{I}_d. \tag{188}$$

Proof Let $\boldsymbol{z} \sim \mathsf{CN}(\boldsymbol{0}_d, \boldsymbol{I}_d)$ and note that, by taking a vector of i.i.d. standard complex normal random variables and dividing it by its norm, we obtain a vector uniformly random on the complex sphere of radius 1. Then, $\boldsymbol{x} = \sqrt{d}\boldsymbol{z}/\left\|\boldsymbol{z}\right\|$.

For $i \in [d]$, denote by x_i and by z_i the ith component of $\boldsymbol{x}$ and $\boldsymbol{z}$, respectively. Then, for $i \neq j$,

$$\mathbb{E}\left\{X_i X_j^*\right\} = d \cdot \mathbb{E}\left\{\frac{Z_i Z_j^*}{\|Z\|^2}\right\} = 0,$$

where the last equality holds by symmetry. Furthermore,

$$\mathbb{E}\left\{|X_i|^2\right\} = d \cdot \mathbb{E}\left\{\frac{|Z_i|^2}{\|Z\|^2}\right\} = 1,$$

as $|Z_i|^2/\left\|Z\right\|^2 \sim \mathrm{Beta}(1, d-1)$ by the argument of Lemma 9. As a result, the thesis is readily proved. □

D Proof of Lemma 3

Before presenting the proof of the lemma, let us introduce some basic definitions and well-known results. Let H be a probability measure on $[0, +\infty)$. Denote by Γ_H the support of H and by τ the supremum of Γ_H. Let $s_H(g)$ denote the Stieltjes transform of H, which is defined as

$$s_H(g) = \int \frac{1}{t-g}\,\mathrm{d}H(t), \tag{189}$$

and let $g_H(s)$ denote its inverse.

Consider a matrix

$$\boldsymbol{S}_n = \frac{1}{d}\boldsymbol{U}\boldsymbol{M}_n\boldsymbol{U}^*, \tag{190}$$

and assume that

(1) $\boldsymbol{M}_n$ is PSD for all $n \in \mathbb{N}$;
(2) $\boldsymbol{U} \in \mathbb{C}^{d\times n}$ is a random matrix whose entries $\{u_{i,j}\}_{1\le i\le d, 1\le j\le n}$ are i.i.d. such that $\mathbb{E}\left\{U_{i,j}\right\} = 0$, $\mathbb{E}\left\{|U_{i,j}|^2\right\} = 1$, and $\mathbb{E}\left\{|U_{i,j}|^4\right\} < \infty$ (this includes the cases in which the entries are $\sim_{i.i.d.} \mathsf{CN}(0,1)$ or are $\sim_{i.i.d.} \mathsf{N}(0,1)$);
(3) The sequence of empirical spectral distributions of $\boldsymbol{M}_n \in \mathbb{C}^{n\times n}$ converges weakly to a probability distribution H, as $n \to +\infty$;
(4) $n/d \to \delta \in (0, +\infty)$, as $n \to \infty$;
(5) The sequence of spectral norms of $\boldsymbol{M}_n$ is bounded.

Note that the normalization of (190) differs from the normalization of (86) by a factor of δ. However, since the form (190) is more common in the literature, we will stick to it for the rest of this section. In order to obtain the desired result for the matrix (86), it suffices to incorporate a factor $1/\delta$ in the definition of the function ψ_δ.

Let $F_{\delta,H}$ be the probability measure on $[0, +\infty)$ such that the inverse $g_{F_{\delta,H}}$ of its Stieltjes transform $s_{F_{\delta,H}}$ is given by

$$g_{F_{\delta,H}}(s) = -\frac{1}{s} + \delta\int \frac{t}{1+ts}\,\mathrm{d}H(t), \quad s \in \{z \in \mathbb{C} : \Im(z) > 0\}. \tag{191}$$

Then, the sequence of empirical spectral distributions of $\boldsymbol{S}_n$ converges weakly to $F_{\delta,H}$ [50], [66, Chapter 4].

For $\alpha \notin \Gamma_H$ and $\alpha \neq 0$, let us also define

$$\psi_{F_{\delta,H}}(\alpha) = g_{F_{\delta,H}}\left(-\frac{1}{\alpha}\right). \tag{192}$$

The function $\psi_{F_{\delta,H}}$ links the support of $F_{\delta,H}$ with the support of the generating measure H (see [67, Section 4] and [3, Lemma 3.1]). In particular, if $\lambda \notin \Gamma_{F_{\delta,H}}$, then $s_{F_{\delta,H}}(\lambda) \neq 0$ and $\alpha = -1/s_{F_{\delta,H}}(\lambda)$ satisfies

(1) $\alpha \notin \Gamma_H$ and $\alpha \neq 0$ (so that $\psi_{F_{\delta,H}}(\alpha)$ is well defined);
(2) $\psi'_{F_{\delta,H}}(\alpha) > 0$.

Conversely, if α satisfies (1) and (2), then $\lambda = \psi_{F_{\delta,H}}(\alpha) \notin \Gamma_{F_{\delta,H}}$.

Let $\lambda_1^{\boldsymbol{M}_n}$ denote the largest eigenvalue of $\boldsymbol{M}_n$ and assume that, as $n \to \infty$,

$$\lambda_1^{\boldsymbol{M}_n} \xrightarrow{\text{a.s.}} \alpha_* \notin \Gamma_H. \tag{193}$$

Denote by $\lambda_1^{\boldsymbol{S}_n}$ the largest eigenvalue of $\boldsymbol{S}_n$. Then, the results in [3] prove that

$$\begin{aligned} \lambda_1^{\boldsymbol{S}_n} &\xrightarrow{\text{a.s.}} \lambda_* = \psi_{F_{\delta,H}}(\alpha_*), \quad \text{if } \psi'_{F_{\delta,H}}(\alpha_*) > 0, \\ \lambda_1^{\boldsymbol{S}_n} &\xrightarrow{\text{a.s.}} \min_{\alpha > \tau} \psi_{F_{\delta,H}}(\alpha), \qquad \text{if } \psi'_{F_{\delta,H}}(\alpha_*) \le 0. \end{aligned} \tag{194}$$

Informally, the eigenvalue $\lambda_1^{\boldsymbol{M}_n}$ is mapped into the point $\psi_{F_{\delta,H}}(\alpha_*)$, where $\alpha_* = -1/s_{F_{\delta,H}}(\lambda_*)$. This point emerges from the support of $F_{\delta,H}$ if and only if $\psi'_{F_{\delta,H}}(\alpha_*) > 0$.

In what follows, we relax the first hypothesis, i.e., we consider the case in which the matrix $\boldsymbol{M}_n$ is not PSD. We will show that (194) still holds, which implies the claim of Lemma 3.

Proof (Proof of Lemma 3) As $\boldsymbol{U}$ is drawn from a rotationally invariant distribution, we can assume without loss of generality that $\boldsymbol{M}_n$ is diagonal. Then, we have that

$$\begin{aligned} \boldsymbol{S}_n &= (\boldsymbol{U}_+, \boldsymbol{U}_-) \begin{pmatrix} \boldsymbol{M}_n^+ & \boldsymbol{0}_k \\ \boldsymbol{0}_{n-k} & -\boldsymbol{M}_n^- \end{pmatrix} \begin{pmatrix} \boldsymbol{U}_+^* \\ \boldsymbol{U}_-^* \end{pmatrix} \\ &= \frac{1}{d} \boldsymbol{U}_+ \boldsymbol{M}_n^+ \boldsymbol{U}_+^* - \frac{1}{d} \boldsymbol{U}_- \boldsymbol{M}_n^- \boldsymbol{U}_-^*, \end{aligned} \tag{195}$$

where $\boldsymbol{M}_n^+ \in \mathbb{R}^{k \times k}$ is the diagonal matrix containing the positive eigenvalues of $\boldsymbol{M}_n$, $\boldsymbol{M}_n^- \in \mathbb{R}^{(n-k)\times(n-k)}$ is the diagonal matrix containing the negative eigenvalues of $\boldsymbol{M}_n$ with the sign changed, $\boldsymbol{U}_+$ contains the first k columns of $\boldsymbol{U}$, and $\boldsymbol{U}_-$ contains the remaining $n-k$ columns of $\boldsymbol{U}$.

Note that $\boldsymbol{U}_+$ and $\boldsymbol{U}_-$ are independent. Furthermore, if $\boldsymbol{H}$ is a unitary matrix, then $\boldsymbol{U}_-$ and $\boldsymbol{H}\boldsymbol{U}_-$ have the same distribution. Hence, we can rewrite the matrix $\boldsymbol{S}_n$ as

$$\boldsymbol{S}_n = \frac{1}{d} \boldsymbol{U}_1 \boldsymbol{M}_n^+ \boldsymbol{U}_1^* - \frac{1}{d} \boldsymbol{H} \boldsymbol{U}_2 \boldsymbol{M}_n^- \boldsymbol{U}_2^* \boldsymbol{H}^*, \tag{196}$$

where $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are independent with entries $\sim_{i.i.d.} \mathsf{CN}(0,1)$, and $\boldsymbol{H}$ is a random unitary matrix distributed according to the Haar measure.

Recall that, by hypothesis, the sequence of empirical spectral distributions of $\boldsymbol{M}_n$ converges weakly to the probability distribution H, where H is the law of the random variable Z. Then, the sequence of empirical spectral distributions of $\boldsymbol{M}_n^+$ converges weakly to the probability distribution H^+, where H^+ is the law of $Z^+ = \max(Z, 0)$. Let F_{δ,H^+} be the probability measure on $[0, +\infty)$ such that the inverse $g_{F_{\delta,H^+}}$ of its Stieltjes transform $s_{F_{\delta,H^+}}$ is given by

$$g_{F_{\delta,H^+}}(s) = -\frac{1}{s} + \delta \int \frac{t}{1+ts}\,\mathrm{d}H^+(t). \tag{197}$$

Define $\boldsymbol{S}_n^+ = \frac{1}{d}\boldsymbol{U}_1\boldsymbol{M}_n^+\boldsymbol{U}_1^*$. Then, as $\boldsymbol{M}_n^+$ is PSD, the sequence of empirical spectral distributions of $\boldsymbol{S}_n^+$ converges weakly to F_{δ,H^+} [50], [66, Chapter 4].

Similarly, the sequence of empirical spectral distributions of $\boldsymbol{M}_n^-$ converges weakly to the probability distribution H^-, where H^- is the law of $Z^- = -\min(Z,0)$. Let F_{δ,H^-} be the probability measure on $[0,+\infty)$ such that the inverse $g_{F_{\delta,H^-}}$ of its Stieltjes transform $s_{F_{\delta,H^-}}$ is given by

$$g_{F_{\delta,H^-}}(s) = -\frac{1}{s} + \delta \int \frac{t}{1+ts}\,\mathrm{d}H^-(t). \tag{198}$$

Define $\boldsymbol{S}_n^- = \frac{1}{d}\boldsymbol{U}_2\boldsymbol{M}_n^-\boldsymbol{U}_2^*$. Then, as $\boldsymbol{M}_n^-$ is PSD, the sequence of empirical spectral distributions of $\boldsymbol{S}_n^-$ converges weakly to F_{δ,H^-} [50], [66, Chapter 4]. Furthermore, the sequence of empirical spectral distributions of $-\boldsymbol{S}_n^-$ converges weakly to the probability measure $F_{\delta,H_{\mathrm{inv}}^-}$ such that

$$g_{F_{\delta,H_{\mathrm{inv}}^-}}(s) = -g_{F_{\delta,H^-}}(-s), \tag{199}$$

where $g_{F_{\delta,H_{\mathrm{inv}}^-}}$ denotes the inverse of the Stieltjes transform $s_{F_{\delta,H_{\mathrm{inv}}^-}}$ of $F_{\delta,H_{\mathrm{inv}}^-}$.

Define

$$F_{\delta,H} = F_{\delta,H^+} \boxplus F_{\delta,H_{\mathrm{inv}}^-}, \tag{200}$$

where $\boxplus$ denotes the free additive convolution. Recall the decomposition (196). Then, the sequence of empirical spectral distributions of $\boldsymbol{S}_n$ converges weakly to $F_{\delta,H}$ [69, 74]. Consequently, the inverse $g_{F_{\delta,H}}$ of the Stieltjes transform $s_{F_{\delta,H}}$ of $F_{\delta,H}$ can be computed as

$$\begin{aligned}
g_{F_{\delta,H}}(s) &\overset{(a)}{=} g_{F_{\delta,H^+}\boxplus F_{\delta,H_{\mathrm{inv}}^-}}(s)\\
&\overset{(b)}{=} g_{F_{\delta,H^+}}(s) + g_{F_{\delta,H_{\mathrm{inv}}^-}}(s) + \frac{1}{s}\\
&\overset{(c)}{=} -\frac{1}{s} + \delta\int \frac{t}{1+ts}\,\mathrm{d}H^+(t) - \delta\int \frac{t}{1-ts}\,\mathrm{d}H^-(t)\\
&\overset{(d)}{=} -\frac{1}{s} + \delta\int \frac{t}{1+ts}\,\mathrm{d}H^+(t) + \delta\int \frac{t}{1+ts}\,\mathrm{d}H^-(-t)\\
&\overset{(e)}{=} -\frac{1}{s} + \delta\int \frac{t}{1+ts}\,\mathrm{d}H(t),
\end{aligned} \tag{201}$$

where in (a) we use (200); in (b) we use that the $\mathcal{R}$-transform of the free convolution is the sum of the $\mathcal{R}$-transforms of the addends; in (c) we use (197), (198), and (199); in (d) we perform the change in variable $t \to -t$ in the second integral; and in (e) we use the fact that $H^+(t)$ is the law of $\max(Z,0)$, $H^-(-t)$ is the law of $\min(Z,0)$, and that $t/(1+ts) = 0$ for $t = 0$.

By hypothesis, $\lambda_1^{\boldsymbol{M}_n} \xrightarrow{\text{a.s.}} \alpha_* \notin \Gamma_H$. First, we establish under what condition the largest eigenvalue of $\boldsymbol{S}_n^+$, call it $\lambda_1^{\boldsymbol{S}_n^+}$, converges to a point outside the support of F_{δ,H^+}. To do so, define $\psi_{F_{\delta,H^+}}(\alpha) = g_{F_{\delta,H^+}}(-1/\alpha)$. Then, $\lambda_1^{\boldsymbol{S}_n^+} \xrightarrow{\text{a.s.}} \psi_{F_{\delta,H^+}}(\alpha_*)$, if $\psi'_{F_{\delta,H^+}}(\alpha_*) > 0$; and $\lambda_1^{\boldsymbol{S}_n^+}$ converges almost surely to a point inside the support of F_{δ,H^+}, otherwise [3].

For the moment, assume that $\psi'_{F_{\delta,H^+}}(\alpha_*) > 0$. We now establish under what condition the largest eigenvalue of $\boldsymbol{S}_n$, call it $\lambda_1^{\boldsymbol{S}_n}$, converges to a point outside the support of $F_{\delta,H}$. To do so, let ω_1 and ω_2 denote the subordination functions corresponding to the free convolution $F_{\delta,H^+} \boxplus F_{\delta,H_{\text{inv}}^-}$. These functions satisfy the following analytic subordination property:

$$s_{F_{\delta,H^+}\boxplus F_{\delta,H_{\text{inv}}^-}}(z) = s_{F_{\delta,H^+}}(\omega_1(z)) = s_{F_{\delta,H_{\text{inv}}^-}}(\omega_2(z)). \tag{202}$$

Then, by Theorem 2.1 of [11], we have that the spike $\psi_{F_{\delta,H^+}}(\alpha_*)$ is mapped into $\omega_1^{-1}(\psi_{F_{\delta,H^+}}(\alpha_*))$. The Stieltjes transform at this point is given by

$$\begin{aligned} s_{F_{\delta,H^+}\boxplus F_{\delta,H_{\text{inv}}^-}}(\omega_1^{-1}(\psi_{F_{\delta,H^+}}(\alpha_*))) &\overset{\text{(a)}}{=} s_{F_{\delta,H^+}}(\psi_{F_{\delta,H^+}}(\alpha_*)) \\ &\overset{\text{(b)}}{=} s_{F_{\delta,H^+}}(g_{F_{\delta,H^+}}(-1/\alpha_*)) \\ &\overset{\text{(c)}}{=} -1/\alpha_*, \end{aligned}$$

where in (a) we use (202); in (b) we use the definition of $\psi_{F_{\delta,H^+}}$; and in (c) we use that $g_{F_{\delta,H^+}}$ is the functional inverse of the Stieltjes transform $s_{F_{\delta,H^+}}$. As a result, by [67, Section 4], we conclude that $\omega_1^{-1}(\psi_{F_{\delta,H^+}}(\alpha_*)) \notin \Gamma_{F_{\delta,H}}$ if and only if $\psi'_{F_{\delta,H}}(\alpha_*) > 0$. Furthermore, the condition $\psi'_{F_{\delta,H}}(\alpha_*) > 0$ is more restrictive than the condition $\psi'_{F_{\delta,H^+}}(\alpha_*) > 0$ since

$$\psi'_{F_{\delta,H^+}}(\alpha_*) = 1 - \delta \int \left(\frac{t}{\alpha_* - t}\right)^2 \mathrm{d}H^+ \geq 1 - \delta \int \left(\frac{t}{\alpha_* - t}\right)^2 \mathrm{d}H = \psi'_{F_{\delta,H}}(\alpha_*).$$

Hence, $\lambda_1^{\boldsymbol{S}_n}$ converges to a point outside the support of $F_{\delta,H}$ if and only if $\psi'_{F_{\delta,H}}(\alpha_*) > 0$ and the proof is complete. □

Remark 8 (Lemma 3 for the real case) Consider the random matrix $\frac{1}{n}\boldsymbol{U}\boldsymbol{M}_n\boldsymbol{U}^\mathsf{T}$, where $\boldsymbol{U} \in \mathbb{R}^{(d-1)\times n}$ is a random matrix whose entries are $\sim_{i.i.d.} \mathsf{N}(0,1)$ and $\boldsymbol{M}_n \in \mathbb{R}^{n\times n}$. Then, the claim of Lemma 3 still holds. Let us briefly explain why this is the case.

If $\boldsymbol{M}_n$ is PSD, then the results of [3] allow us to conclude. If $\boldsymbol{M}_n$ is not PSD, we can write an expression analogous to (196):

$$\frac{1}{d}\boldsymbol{U}\boldsymbol{M}_n\boldsymbol{U}^\mathsf{T} = \frac{1}{d}\boldsymbol{U}_1\boldsymbol{M}_n^+\boldsymbol{U}_1^\mathsf{T} - \frac{1}{d}\boldsymbol{H}\boldsymbol{U}_2\boldsymbol{M}_n^-\boldsymbol{U}_2^\mathsf{T}\boldsymbol{H}^*, \tag{203}$$

where $\boldsymbol{M}_n^+$ is the diagonal matrix containing the positive eigenvalues of $\boldsymbol{M}_n$, $\boldsymbol{M}_n^-$ is the diagonal matrix containing the negative eigenvalues of $\boldsymbol{M}_n$ with the sign changed, $\boldsymbol{U}_1$ and $\boldsymbol{U}_2$ are independent with entries $\sim_{i.i.d.} \mathsf{N}(0,1)$, $\boldsymbol{H}$ is a random unitary matrix distributed according to the Haar measure, and we have used the fact that the eigenvalues of $\boldsymbol{U}_2\boldsymbol{M}_n^-\boldsymbol{U}_2^\mathsf{T}$ are the same as the eigenvalues of $\boldsymbol{H}\boldsymbol{U}_2\boldsymbol{M}_n^-\boldsymbol{U}_2^\mathsf{T}\boldsymbol{H}^*$ since $\boldsymbol{H}$ is unitary. Hence, the proof follows from the same argument of Lemma 3.

E Proof of Lemma 4 and Theorem 5

We start by proving a result similar to Lemma 4 for a general AMP iteration, where the function $f_t(\hat{z}; y)$ is generic.

Lemma 13 (State evolution for general AMP iteration) *Let $\boldsymbol{x} \in \mathbb{R}^d$ denote the unknown signal such that $\|\boldsymbol{x}\|_2 = \sqrt{d}$, $\boldsymbol{A} = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_n)^\mathsf{T} \in \mathbb{R}^{n\times d}$ with $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{i.i.d.} \mathsf{N}(\boldsymbol{0}_d, \boldsymbol{I}_d/d)$, and $\boldsymbol{y} = (y_1, \dots, y_n)$ with $y_i \sim p(\cdot \mid \langle \boldsymbol{x}, \boldsymbol{a}_i\rangle)$. Consider the AMP iterates $\boldsymbol{z}^t, \hat{\boldsymbol{z}}^t$ defined in (126) for some function $f_t(\hat{z}; y)$, with b_t given by*

$$\mathsf{b}_t = \delta \cdot \mathbb{E}\{f_t'(\mu_t G_0 + \tau_t G_1; Y)\}, \tag{204}$$

where the expectation is with respect to $G_0, G_1 \sim_{i.i.d.} \mathsf{N}(0,1)$ and $Y \sim p(\cdot \mid G_0)$. Assume that the initialization $\boldsymbol{z}^0$ is independent of $\boldsymbol{A}$ and that, almost surely,

$$\lim_{n\to\infty} \frac{1}{d}\langle \boldsymbol{x}, \boldsymbol{z}^0\rangle = \mu_0, \quad \lim_{n\to\infty} \frac{1}{d}\|\boldsymbol{z}^0\|^2 = \mu_0^2 + \tau_0^2. \tag{205}$$

Let the state evolution recursion τ_t, μ_t be defined as

$$\begin{aligned} \mu_{t+1} &= \delta \int_{\mathbb{R}} \mathbb{E}\big\{\partial_g p(y \mid X_0) f_t(\mu_t X_0 + \tau_t G; y)\big\}\, \mathrm{d}y, \\ \tau_{t+1}^2 &= \delta \cdot \mathbb{E}\Big\{\big(f_t(\mu_t X_0 + \tau_t G; Y)\big)^2\Big\}, \end{aligned} \tag{206}$$

with initialization μ_0 and τ_0, where the expectation is taken with respect to $X_0, G \sim_{i.i.d.} \mathsf{N}(0,1)$. Then, for any t, and for any function $\psi : \mathbb{R}^2 \to \mathbb{R}$ such that $|\psi(\boldsymbol{u}) - \psi(\boldsymbol{v})| \le L(1 + \|\boldsymbol{u}\|_2 + \|\boldsymbol{v}\|_2)\|\boldsymbol{u} - \boldsymbol{v}\|_2$ for some $L \in \mathbb{R}$, we have that, almost surely,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \psi(x_i, z_i^t) = \mathbb{E}\{\psi(X_0, \mu_t X_0 + \tau_t G)\}. \tag{207}$$

Proof For $g \in \mathbb{R}$, let $\mathcal{H}(\cdot; g) : [0,1] \to \mathbb{R} \cup \{+\infty, -\infty\}$ be the generalized inverse of

$$\mathcal{F}(y \mid g) \equiv \int_{-\infty}^{y} p(y' \mid g)\, \mathrm{d}y',$$

namely,

$$\mathcal{H}(w; g) \equiv \inf \left\{ y \in \mathbb{R} : \quad \mathcal{F}(y \mid g) \geq w \right\}. \tag{208}$$

With this definition, the model $y_i \sim p(\cdot \mid \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle)$ is equivalent to $y_i = \mathcal{H}(w_i; \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle)$ for $\{w_i\}_{1 \leq i \leq n} \sim_{i.i.d.} \mathrm{Unif}([0, 1])$ independent of $\boldsymbol{A}$ and $\boldsymbol{x}$. Let $\boldsymbol{w} = (w_1, \ldots, w_n) \in \mathbb{R}^n$ and denote by $[\boldsymbol{v}_1 \mid \cdots \mid \boldsymbol{v}_k] \in \mathbb{R}^{m \times k}$ the matrix obtained by stacking column vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k \in \mathbb{R}^m$.

For $t \geq 0$, define $\boldsymbol{r}^t = \boldsymbol{0}_d$, $\hat{\boldsymbol{r}}^t = \boldsymbol{A}\boldsymbol{x}$, and introduce the extended state variables $\boldsymbol{s}^t \in \mathbb{R}^{d \times 2}$ and $\hat{\boldsymbol{s}}^t \in \mathbb{R}^{n \times 2}$, defined as

$$\begin{aligned} \boldsymbol{s}^t &= [\boldsymbol{z}^t \mid \boldsymbol{r}^t], \\ \hat{\boldsymbol{s}}^t &= [\hat{\boldsymbol{z}}^t \mid \hat{\boldsymbol{r}}^t]. \end{aligned} \tag{209}$$

We further define the functions $h_t = [h_{t,1} \mid h_{t,2}] : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ and $\hat{h}_t = [\hat{h}_{t,1} \mid \hat{h}_{t,2}] : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}^2$ by setting

$$\begin{aligned} h_t(s_1, s_2; x) &\equiv [s_1 \mid x\,], \\ \hat{h}_t(\hat{s}_1, \hat{s}_2; w) &\equiv [\, f_t(\hat{s}_1; \mathcal{H}(w; \hat{s}_2)) \mid 0\,]. \end{aligned} \tag{210}$$

With these notations, the iteration (126) is equivalent to

$$\begin{aligned} \boldsymbol{s}^{t+1} &= \boldsymbol{A}^{\mathsf{T}} \hat{h}_t(\hat{\boldsymbol{s}}^t; \boldsymbol{w}) - h_t(\boldsymbol{s}^t; \boldsymbol{x}) \hat{\mathsf{B}}_t, \\ \hat{\boldsymbol{s}}^t &= \boldsymbol{A} h_t(\boldsymbol{s}^t; \boldsymbol{x}) - \hat{h}_{t-1}(\hat{\boldsymbol{s}}^{t-1}; \boldsymbol{w}) \mathsf{B}_{t-1}, \end{aligned} \tag{211}$$

where the functions $h_t(\boldsymbol{s}^t; \boldsymbol{x})$ and $\hat{h}_t(\hat{\boldsymbol{s}}^t; \boldsymbol{w})$ are understood to be applied component-wise to their arguments and $\mathsf{B}_t, \hat{\mathsf{B}}_t \in \mathbb{R}^{2 \times 2}$ are defined by

$$\begin{aligned} (\hat{\mathsf{B}}_t)_{j,k} &= \delta \cdot \mathbb{E}\left\{ \frac{\partial \hat{h}_{t,k}}{\partial \hat{s}_j}(\mu_t X_0 + \tau_t G, X_0; W) \right\}, \\ (\mathsf{B}_t)_{j,k} &= \delta \cdot \mathbb{E}\left\{ \frac{\partial h_{t,k}}{\partial s_j}(\mu_t X_0 + \tau_t G, 0; X_0) \right\}. \end{aligned} \tag{212}$$

The iteration (211) satisfies the assumptions of [41][Proposition 5]. By applying that result, the claim follows. □

At this point, first we present the proof of Lemma 4 and then of Theorem 5.

Proof (Proof of Lemma 4) Consider the state evolution recursion defined in (124) with initialization μ_0. Let f_t be defined as in (127) with F given by (123). Suppose that, for any t, (206) holds with $\mu_t = \tau_t^2$. Then, by Lemma 13, the claim immediately follows.

The remaining part of the proof is devoted to show that (206) holds with $\mu_t = \tau_t^2$, for $t \geq 0$. First, we prove by induction that $\mu_t = \tau_t^2$, for $t \geq 0$. The basis of the induction, i.e., $\mu_0 = \tau_0^2$, is true by the hypothesis of the Lemma. Now, we assume that $\mu_t = \tau_t^2$ and we show that $\mu_{t+1} = \tau_{t+1}^2$. Set

$$Z = \mu_t X_0 + \tau_t G, \tag{213}$$

and note that $Z \sim \mathsf{N}(0, \mu_t^2 + \tau_t^2)$. Then, we can rewrite X_0 as

$$X_0 = aZ + b\widetilde{G},$$

for some $a, b \in \mathbb{R}$, where $\widetilde{G} \sim \mathsf{N}(0, 1)$ and independent from Z. In order to compute the coefficients a and b, we evaluate $\mathbb{E}\{X_0^2\}$ and $\mathbb{E}\{X_0 \cdot Z\}$, thus obtaining the equations

$$\begin{aligned} a^2(\mu_t^2 + \tau_t^2) + b^2 &= 1, \\ a(\mu_t^2 + \tau_t^2) &= \mu_t, \end{aligned}$$

which can be simplified as

$$\begin{aligned} a &= \frac{\mu_t}{\mu_t^2 + \tau_t^2}, \\ b &= \frac{\tau_t}{\sqrt{\mu_t^2 + \tau_t^2}}. \end{aligned}$$

Furthermore, by using the inductive hypothesis $\mu_t = \tau_t^2$ and that $q_t = \mu_t/(1 + \mu_t)$, we obtain that

$$X_0 = (1 - q_t)\, Z + \sqrt{1 - q_t}\, \widetilde{G}. \tag{214}$$

Hence, the following chain of equalities holds:

$$\begin{aligned} \tau_{t+1}^2 &\overset{(a)}{=} \delta \int_{\mathbb{R}} \mathbb{E}\Big\{ p(y \mid X_0) \cdot \big(f_t(\mu_t X_0 + \tau_t G; y)\big)^2 \Big\}\, \mathrm{d}y \\ &\overset{(b)}{=} \delta \int_{\mathbb{R}} \mathbb{E}\Big\{ p\big(y \mid (1 - q_t)\, Z + \sqrt{1 - q_t}\, \widetilde{G}\big) \cdot \big(f_t(Z; y)\big)^2 \Big\}\, \mathrm{d}y \\ &\overset{(c)}{=} \delta \int_{\mathbb{R}} \mathbb{E}\Big\{ \big(f_t(Z; y)\big)^2 \cdot \mathbb{E}\big\{ p\big(y \mid (1 - q_t)\, Z + \sqrt{1 - q_t}\, \widetilde{G}\big) \mid Z \big\} \Big\}\, \mathrm{d}y \\ &\overset{(d)}{=} \delta \int_{\mathbb{R}} \mathbb{E}\Bigg\{ \bigg(\frac{\mathbb{E}\{\partial_g p(y \mid (1 - q_t)\, Z + \sqrt{1 - q_t}\, \widetilde{G}) \mid Z\}}{\mathbb{E}\{p(y \mid (1 - q_t)\, Z + \sqrt{1 - q_t}\, \widetilde{G}) \mid Z\}} \bigg)^2 \cdot \mathbb{E}\big\{ p\big(y \mid (1 - q_t)\, Z \\ &\qquad + \sqrt{1 - q_t}\, \widetilde{G}\big) \mid Z \big\} \Bigg\}\, \mathrm{d}y \qquad (215) \\ &= \delta \int_{\mathbb{R}} \mathbb{E}\Bigg\{ \frac{\mathbb{E}\{\partial_g p(y \mid (1 - q_t)\, Z + \sqrt{1 - q_t}\, \widetilde{G}) \mid Z\}}{\mathbb{E}\{p(y \mid (1 - q_t)\, Z + \sqrt{1 - q_t}\, \widetilde{G}) \mid Z\}} \cdot \mathbb{E}\big\{ \partial_g p\big(y \mid (1 - q_t)\, Z \\ &\qquad + \sqrt{1 - q_t}\, \widetilde{G}\big) \mid Z \big\} \Bigg\}\, \mathrm{d}y \\ &\overset{(e)}{=} \delta \int_{\mathbb{R}} \mathbb{E}\Big\{ f_t(Z; y) \cdot \mathbb{E}\big\{ \partial_g p\big(y \mid (1 - q_t)\, Z + \sqrt{1 - q_t}\, \widetilde{G}\big) \mid Z \big\} \Big\}\, \mathrm{d}y \\ &\overset{(f)}{=} \delta \int_{\mathbb{R}} \mathbb{E}\Big\{ f_t(\mu_t X_0 + \tau_t G; y) \cdot \partial_g p(y \mid X_0) \Big\}\, \mathrm{d}y = \mu_{t+1}, \end{aligned}$$

where in (a) we use that $Y \sim p(\cdot \mid X_0)$; in (b) we use (213) and (214); in (c) we condition with respect to Z; in (d) we use definition (127) of f_t; in (e) we use again definition (127) of f_t; and in (f) we use again (213) and (214).

Finally, we prove that μ_{t+1} satisfies (206). Indeed, the following chain of equalities holds:

$$\begin{aligned}
\mu_{t+1} &\overset{(a)}{=} \delta \int_{\mathbb{R}} \mathbb{E}\left\{ \frac{\left(\mathbb{E}\{\partial_g p(y \mid (1-q_t)\, Z + \sqrt{1-q_t}\, \widetilde{G}) \mid Z\}\right)^2}{\mathbb{E}\{p(y \mid (1-q_t)\, Z + \sqrt{1-q_t}\, \widetilde{G}) \mid Z\}} \right\} \mathrm{d}y \\
&\overset{(b)}{=} \delta \int_{\mathbb{R}} \mathbb{E}\left\{ \frac{\left(\mathbb{E}\{\partial_g p(y \mid (1-q_t)\sqrt{\mu_t^2+\tau_t^2}\, G_0 + \sqrt{1-q_t}\, G_1) \mid G_0\}\right)^2}{\mathbb{E}\{p(y \mid (1-q_t)\sqrt{\mu_t^2+\tau_t^2}\, G_0 + \sqrt{1-q_t}\, G_1) \mid G_0\}} \right\} \mathrm{d}y \\
&\overset{(c)}{=} \delta \int_{\mathbb{R}} \mathbb{E}\left\{ \frac{\left(\mathbb{E}\{\partial_g p(y \mid \sqrt{q_t}\, G_0 + \sqrt{1-q_t}\, G_1) \mid G_0\}\right)^2}{\mathbb{E}\{p(y \mid \sqrt{q_t}\, G_0 + \sqrt{1-q_t}\, G_1) \mid G_0\}} \right\} \mathrm{d}y \; = \delta \cdot h(q_t),
\end{aligned} \tag{216}$$

where in (a) we use (215); in (b) we set $G_1 = \widetilde{G}$ and $G_0 = Z/\sqrt{\mu_t^2+\tau_t^2}$; and in (c) we use that $\mu_t = \tau_t^2$ and that $q_t = \mu_t/(1+\mu_t)$. □

Proof (Proof of Theorem 5) In view of Lemma 4, it is sufficient to show that $(q, \mu) = (0, 0)$ is an attractive fixed point of the recursion (124).

First of all, let us check that $(q, \mu) = (0, 0)$ is a fixed point. This happens if and only if

$$h(0) = \int_{\mathbb{R}} \frac{\left(\mathbb{E}_{G_1}\{\partial_g p(y \mid G_1)\}\right)^2}{\mathbb{E}_{G_1}\{p(y \mid G_1)\}} \mathrm{d}y = 0, \tag{217}$$

which holds because of condition (131).

Let us now prove that this fixed point is stable. We start by rewriting the function $h(q)$ defined in (125) as

$$h(q) = \int_{\mathbb{R}} \mathbb{E}_{G_0}\left\{ \frac{\left(h_{\mathrm{num}}(\sqrt{q}, y)\right)^2}{h_{\mathrm{den}}(\sqrt{q}, y)} \right\} \mathrm{d}y, \tag{218}$$

where

$$\begin{aligned}
h_{\mathrm{num}}(x, y) &= \mathbb{E}_{G_1}\left\{ \partial_g p(y \mid x \cdot G_0 + \sqrt{1-x^2}\, G_1) \right\}, \\
h_{\mathrm{den}}(x, y) &= \mathbb{E}_{G_1}\left\{ p(y \mid x \cdot G_0 + \sqrt{1-x^2}\, G_1) \right\}.
\end{aligned} \tag{219}$$

Note that $h_{\text{num}}(0, y) = 0$ by assumption (131). Then,

$$\begin{aligned} h_{\text{num}}(\sqrt{q}, y) &= \sqrt{q}\frac{\partial h_{\text{num}}(x, y)}{\partial x}\bigg|_{x=0} + \frac{q}{2}\frac{\partial^2 h_{\text{num}}(x, y)}{\partial^2 x}\bigg|_{x=x_1}, \\ h_{\text{den}}(x, y) &= h_{\text{den}}(0, y) + \sqrt{q}\frac{\partial h_{\text{den}}(x, y)}{\partial x}\bigg|_{x=x_2}, \end{aligned} \tag{220}$$

for some $x_1, x_2 \in [0, \sqrt{q}]$. Furthermore, by applying Stein's lemma, we have that

$$h_{\text{num}}(x, y) = \frac{1}{\sqrt{1-x^2}}\mathbb{E}_{G_1}\left\{G_1 \cdot p(y \mid x \cdot G_0 + \sqrt{1-x^2}\, G_1)\right\}. \tag{221}$$

By using (221), we can rewrite (220) as

$$\begin{aligned} h_{\text{num}}(\sqrt{q}, y) &= \sqrt{q}\, G_0 \cdot \mathbb{E}_{G_1}\{G_1 \cdot \partial_g p(y \mid G_1)\} \\ &\quad + \frac{q}{2}\frac{1}{(1-x_1^2)^{5/2}}\mathbb{E}_{G_1}\left\{f_{\text{num}}(G_0, G_1, x_1)\right\}, \\ h_{\text{den}}(x, y) &= \mathbb{E}_{G_1}\{p(y \mid G_1)\} + \sqrt{q}\,\mathbb{E}_{G_1}\left\{f_{\text{den}}(G_0, G_1, x_2)\right\}, \end{aligned}$$

where

$$\begin{aligned} f_{\text{num}}(G_0, G_1, x_1) &= G_1\Big((1+2x_1^2)\, p(y \mid x_1 \cdot G_0 + \sqrt{1-x_1^2}\, G_1) \\ &\quad - \big(2G_0 \cdot x_1(x_1^2-1) + G_1\sqrt{1-x_1^2}\,(1+2x_1^2)\big)\partial_g p(y \mid x_1 \cdot G_0 + \sqrt{1-x_1^2}\, G_1) \\ &\quad + (x_1^2-1)(G_0^2\,(x_1^2-1) - G_1^2\, x_1^2 + 2G_0 G_1 x_1\sqrt{1-x_1^2})\partial_g^2 p(y \mid x_1 \cdot G_0 \\ &\quad + \sqrt{1-x_1^2}\, G_1)\Big), \\ f_{\text{den}}(G_0, G_1, x_2) &= \left(G_0 - \frac{x_2}{\sqrt{1-x_2^2}}G_1\right) \cdot \partial_g p(y \mid x_2 \cdot G_0 + \sqrt{1-x_2^2}\, G_1). \end{aligned} \tag{222}$$

By applying again Stein's lemma and by using that the conditional density $p(y \mid g)$ is bounded, we note that $\mathbb{E}_{G_1}\{f_{\text{num}}(G_0, G_1, x_1)\}$ and $\mathbb{E}_{G_1}\{f_{\text{den}}(G_0, G_1, x_2)\}$ are bounded. Hence, by dominated convergence, we obtain that

$$h(q) = q \cdot \int_{\mathbb{R}} \frac{\left(\mathbb{E}_{G_1}\{\partial_g^2 p(y \mid G_1)\}\right)^2}{\mathbb{E}_{G_1}\{p(y \mid G_1)\}}\,\mathrm{d}y + o(q).$$

Therefore, in a neighborhood of the fixed point we have

$$\begin{aligned} q_t &= \mu_t + o(\mu_t), \\ \mu_{t+1} &= \delta \cdot q_t \cdot \int_{\mathbb{R}} \frac{\left(\mathbb{E}_{G_1}\{\partial_g^2 p(y \mid G_1)\}\right)^2}{\mathbb{E}_{G_1}\{p(y \mid G_1)\}} \, \mathrm{d}y + o(q_t). \end{aligned} \tag{223}$$

Furthermore, by applying twice Stein's lemma, we also have that

$$\mathbb{E}_{G_1}\{\partial_g^2 p(y \mid G_1)\} = \mathbb{E}_{G_1}\{p(y \mid G_1)(G_1^2 - 1)\}. \tag{224}$$

By using (223), (224) and by recalling definition (42) of δ_{u}, we conclude that

$$\begin{aligned} q_t &= \mu_t + o(\mu_t), \\ \mu_{t+1} &= \frac{\delta}{\delta_{\mathrm{u}}} q_t + o(q_t). \end{aligned} \tag{225}$$

As $\delta < \delta_{\mathrm{u}}$, the fixed point is stable. □

F Proof of Lemma 5 and Theorem 6

For the proofs in this section, it is convenient to introduce the function

$$\mathsf{G}(x, y; \bar{q}) = \frac{\mathbb{E}_G\{\partial_g^2 p(y \mid \bar{q}\, x + \sqrt{\bar{q}} G)\}}{\mathbb{E}_G\{p(y \mid \bar{q}\, x + \sqrt{\bar{q}} G)\}} - \left(\frac{\mathbb{E}_G\{\partial_g p(y \mid \bar{q}\, x + \sqrt{\bar{q}} G)\}}{\mathbb{E}_G\{p(y \mid \bar{q}\, x + \sqrt{\bar{q}} G)\}}\right)^2. \tag{226}$$

First, we present the proof of Lemma 5 and then of Theorem 6.

Proof (Proof of Lemma 5) Condition (131) implies that

$$\mathsf{F}(0, y; 1) = 0. \tag{227}$$

Furthermore, we have that

$$\begin{aligned} \mathbb{E}_Y\{\mathsf{G}(0, Y; 1)\} &\overset{(a)}{=} \mathbb{E}_Y \left\{ \frac{\mathbb{E}_G\{\partial_g^2 p(Y \mid G)\}}{\mathbb{E}_G\{p(Y \mid G)\}} \right\} \\ &\overset{(b)}{=} \int_{\mathbb{R}} \mathbb{E}_G\{\partial_g^2 p(y \mid G)\} \, \mathrm{d}y \\ &= \mathbb{E}_G \left\{ \partial_g^2 \int_{\mathbb{R}} p(y \mid G) \, \mathrm{d}y \right\} = 0, \end{aligned} \tag{228}$$

where in (a) we use (227) and definition (226) of $\mathsf{G}(0, y; 1)$ and in (b) we use the fact that y has density $\mathbb{E}_G\{p(y \mid G)\}$.

Denote by $\mathsf{F}'(x, y; \bar{q})$ the derivative of F with respect to its first argument. Then, we have

$$\mathsf{F}'(x, y; \bar{q}) = \bar{q}\mathsf{G}(x, y; \bar{q}). \tag{229}$$

Hence,

$$\begin{aligned}\mathsf{b}_t &= \delta \cdot (1 - q_t) \cdot \mathbb{E}\{\mathsf{G}(\mu_t G_0 + \sqrt{\mu_t} G_1, Y; 1 - q_t)\} = \delta \cdot \mathbb{E}\{\mathsf{G}(0; Y; 1)\} + o_{q_t}(1) \\ &= o_{q_t}(1).\end{aligned} \tag{230}$$

By using (229) and (230), we linearize the recursion (126) around the fixed point $z^t = \mathbf{0}_d$ and $\hat{z}^{t-1} = \mathbf{0}_n$ as

$$z^{t+1} = \boldsymbol{A}^{\mathsf{T}} \boldsymbol{J} \hat{z}^t + o_{q_t}(1)(\|z^t\|_2 + \|\hat{z}^t\|_2) + o(\|\hat{z}^t\|_2)\,, \tag{231}$$

$$\hat{z}^t = \boldsymbol{A} z^t - \boldsymbol{J} \hat{z}^{t-1} + o_{q_t}(1)\, \|\hat{z}^{t-1}\|_2 + o(\|\hat{z}^{t-1}\|_2), \tag{232}$$

where $\boldsymbol{J} \in \mathbb{R}^{n \times n}$ is a diagonal matrix with entries $j_i = \mathsf{F}'(0, y_i; 1)$ for $i \in [n]$. By substituting expression (232) for $\hat{z}^t$ into the RHS of (231), the result follows. □

Proof (Proof of Theorem 6) By definition, α is an eigenvalue of $\boldsymbol{L}_n$ if and only if

$$\det(\boldsymbol{L}_n - \alpha \boldsymbol{I}_{n+d}) = 0. \tag{233}$$

Recall that, when $\boldsymbol{D}$ is invertible,

$$\det \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{D} \end{pmatrix} = \det(\boldsymbol{D}) \cdot \det(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{D}^{-1}\boldsymbol{C}). \tag{234}$$

Then, after some calculations, we obtain that (233) is equivalent to

$$\alpha^d \cdot \det(-\boldsymbol{J} - \alpha \boldsymbol{I}_n) \cdot \det(\boldsymbol{I}_d - \boldsymbol{A}^{\mathsf{T}}(\boldsymbol{I}_n + \alpha \boldsymbol{J}^{-1})^{-1} \boldsymbol{A}) = 0. \tag{235}$$

From (235), we immediately deduce that the eigenvalues of $\boldsymbol{L}_n$ are real if and only if all the solutions to

$$\det(\boldsymbol{I}_d - \boldsymbol{A}^{\mathsf{T}}(\boldsymbol{I}_n + \alpha \boldsymbol{J}^{-1})^{-1} \boldsymbol{A}) = 0 \tag{236}$$

are real. We will prove that in fact this equation does not have any solution for $\alpha \in \mathbb{C} \backslash \mathbb{R}$.

Let $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathsf{T}}$ be the SVD of $\boldsymbol{A}$. Then, (236) is equivalent to

$$\det(\boldsymbol{U}\boldsymbol{\Sigma}^{-2} - (\boldsymbol{I}_n + \alpha \boldsymbol{J}^{-1})^{-1} \boldsymbol{U}) = 0.$$

Using the fact that $\det(\boldsymbol{\Sigma}) \neq 0$, and $\det(\boldsymbol{I}_n + \alpha \boldsymbol{J}^{-1}) \neq 0$ for $\alpha \in \mathbb{C} \backslash \mathbb{R}$, Eq. (236) is equivalent to

$$\det((\boldsymbol{I}_n + \alpha \boldsymbol{J}^{-1})\boldsymbol{U} - \boldsymbol{U}\boldsymbol{\Sigma}^2) = 0,$$

or equivalently

$$\det(\boldsymbol{I}_n + \alpha \boldsymbol{J}^{-1} - \boldsymbol{A}\boldsymbol{A}^{\mathsf{T}}) = \det(\boldsymbol{I}_n + \alpha \boldsymbol{J}^{-1} - \boldsymbol{U}\boldsymbol{\Sigma}^2\boldsymbol{U}^{\mathsf{T}}) = 0.$$

Given that the solutions of this equations are generalized eigenvalues for the pairs of symmetric matrices $\boldsymbol{A}\boldsymbol{A}^{\mathsf{T}} - \boldsymbol{I}_n$ and $\boldsymbol{J}^{-1}$, they must be real. We conclude that the eigenvalues of $\boldsymbol{L}_n$ are real.

Note that

$$\begin{aligned} \mathsf{G}(0, y; 1) &\overset{\text{(a)}}{=} \frac{\mathbb{E}_G\{\partial_g^2 p(y \mid G)\}}{\mathbb{E}_G\{p(y \mid G)\}} \\ &\overset{\text{(b)}}{=} \frac{\mathbb{E}_G\{p(y \mid G)(G^2 - 1)\}}{\mathbb{E}_G\{p(y \mid G)\}} \\ &\overset{\text{(c)}}{=} \frac{\mathcal{T}^*(y)}{1 - \mathcal{T}^*(y)}, \end{aligned} \tag{237}$$

where in (a) we use that $\mathsf{F}(0, 1; y) = 0$ as (131) holds; in (b) we apply twice Stein's lemma; and in (c) we use definition (45) of $\mathcal{T}^*$. Then, (236) can be rewritten as

$$\det\left(\boldsymbol{I}_d - \sum_{i=1}^{n} \frac{\mathcal{T}^*(y_i)}{\mathcal{T}^*(y_i) + \alpha(1 - \mathcal{T}^*(y_i))} \boldsymbol{a}_i \boldsymbol{a}_i^{\mathsf{T}}\right) = 0. \tag{238}$$

Let $\lambda_1^{\boldsymbol{D}_n^*}(\alpha)$ be the largest eigenvalue of the matrix $\boldsymbol{D}_n^*(\alpha)$ defined as

$$\boldsymbol{D}_n^*(\alpha) = \sum_{i=1}^{n} \frac{\mathcal{T}^*(y_i)}{\mathcal{T}^*(y_i) + \alpha(1 - \mathcal{T}^*(y_i))} \boldsymbol{a}_i \boldsymbol{a}_i^{\mathsf{T}}. \tag{239}$$

Note that, as $\alpha \to +\infty$, the entries of $\boldsymbol{D}_n^*(\alpha)$ tend to 0 with high probability. Since the eigenvalues of a matrix are continuous functions of the elements of the matrix, we also obtain that

$$\lim_{\alpha \to +\infty} \lambda_1^{\boldsymbol{D}_n^*}(\alpha) = 0.$$

Hence, if there exists $\bar{\alpha} > 1$ such that $\lambda_1^{\boldsymbol{D}_n^*}(\bar{\alpha}) > 1$, then there exists also $\bar{\alpha}_0 > \bar{\alpha} > 1$ such that $\lambda_1^{\boldsymbol{D}_n^*}(\bar{\alpha}_0) = 1$. Consequently, there exists $\alpha > 1$ that satisfies (238), which implies the result of the theorem.

The rest of the proof consists in showing that $\bar{\alpha} = \sqrt{\delta/\delta_{\mathrm{u}}}$ satisfies the desired requirements. First of all, note that $\sqrt{\delta/\delta_{\mathrm{u}}} > 1$, as $\delta > \delta_{\mathrm{u}}$. Furthermore, we have that

$$\boldsymbol{D}_n^*(\bar{\alpha}) = \sum_{i=1}^{n} \mathcal{T}_\delta^*(y_i) \boldsymbol{a}_i \boldsymbol{a}_i^{\mathsf{T}}, \tag{240}$$

where $\mathcal{T}_\delta^*$ is defined in (44). Recall that, by hypothesis, $\boldsymbol{x}$ is such that $\|\boldsymbol{x}\|_2 = \sqrt{d}$ and $\{\boldsymbol{a}_i\}_{1\le i\le n} \sim_{i.i.d.} \mathrm{N}(\mathbf{0}_d, \boldsymbol{I}_d/d)$. Let $\tilde{\boldsymbol{x}} = \boldsymbol{x}/\sqrt{d}$ and $\tilde{\boldsymbol{a}}_i = \sqrt{d}\cdot\boldsymbol{a}_i$. Then, $\langle \boldsymbol{x}, \boldsymbol{a}_i\rangle = \langle \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{a}}_i\rangle$. Let $\lambda_1^{\widetilde{\boldsymbol{D}}_n}$ be the largest eigenvalue of the matrix $\widetilde{\boldsymbol{D}}_n$ defined as

$$\widetilde{\boldsymbol{D}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}_\delta^*(y_i)\tilde{\boldsymbol{a}}_i\tilde{\boldsymbol{a}}_i^{\mathsf{T}}. \tag{241}$$

Since $\widetilde{\boldsymbol{D}}_n = \boldsymbol{D}_n^*(\bar{\alpha})/\delta$, it remains to prove that

$$\lambda_1^{\widetilde{\boldsymbol{D}}_n} \xrightarrow{\text{a.s.}} \tilde{\lambda} > \frac{1}{\delta}. \tag{242}$$

To do so, we apply a result analogous to that of Lemma 2 for the real case with $\mathcal{T} = \mathcal{T}_\delta^*$. For the moment, assume that $\mathcal{T}_\delta^*$ fulfills the hypotheses of Lemma 2 (we will prove later that this is the case). Then, $\lambda_1^{\widetilde{\boldsymbol{D}}_n}$ converges almost surely to $\zeta_\delta(\lambda_\delta^*)$.

Recall that

$$\zeta_\delta(\lambda) = \psi_\delta(\max(\lambda, \bar{\lambda}_\delta)),$$

where $\bar{\lambda}_\delta$ is the point of minimum of the convex function $\psi_\delta(\lambda)$ defined as

$$\psi_\delta(\lambda) = \lambda\left(\frac{1}{\delta} + \mathbb{E}\left\{\frac{\mathcal{T}_\delta^*(Y)}{\lambda - \mathcal{T}_\delta^*(Y)}\right\}\right).$$

Notice also that this minimum is the unique local minimizer since ψ_δ is convex and analytic.

Furthermore, λ_δ^* is the unique solution to the equation $\zeta_\delta(\lambda_\delta^*) = \phi(\lambda_\delta^*)$, where $\phi(\lambda)$ is defined as

$$\phi(\lambda) = \lambda\cdot\mathbb{E}\left\{\frac{\mathcal{T}_\delta^*(Y)\cdot G^2}{\lambda - \mathcal{T}_\delta^*(Y)}\right\}.$$

By setting the derivative of $\psi_\delta(\lambda)$ to 0, we have that

$$\mathbb{E}\left\{\frac{(\mathcal{T}_\delta^*(Y))^2}{(\bar{\lambda}_\delta - \mathcal{T}_\delta^*(Y))^2}\right\} = \frac{1}{\delta}.$$

By using definition (44) of $\mathcal{T}_\delta^*$ and definition (45) of $\mathcal{T}^*$, we verify that

$$\frac{\mathcal{T}_\delta^*(Y)}{1 - \mathcal{T}_\delta^*(Y)} = \sqrt{\frac{\delta_{\mathrm{u}}}{\delta}}\,\frac{\mathbb{E}_G\{p(y\mid G)(G^2-1)\}}{\mathbb{E}_G\{p(y\mid G)\}}. \tag{243}$$

Hence, by using definition (42) of δ_{u}, we obtain that

$$\mathbb{E}\left\{\frac{(\mathcal{T}_\delta^*(Y))^2}{(1-\mathcal{T}_\delta^*(Y))^2}\right\} = \frac{\delta_{\mathrm{u}}}{\delta}\int_{\mathbb{R}} \frac{\left(\mathbb{E}_G\{p(y \mid G)(G^2-1)\}\right)^2}{\mathbb{E}_G\{p(y \mid G)\}}\,\mathrm{d}y = \frac{1}{\delta},$$

which immediately implies that

$$\bar{\lambda}_\delta = 1. \tag{244}$$

By using (243), one also obtains that

$$\mathbb{E}\left\{\frac{\mathcal{T}_\delta^*(Y)}{1-\mathcal{T}_\delta^*(Y)}\right\} = \sqrt{\frac{\delta_{\mathrm{u}}}{\delta}}\int_{\mathbb{R}} \mathbb{E}_G\{p(y \mid G)(G^2-1)\}\,\mathrm{d}y = \sqrt{\frac{\delta_{\mathrm{u}}}{\delta}}\,\mathbb{E}_G\{G^2-1\} = 0,$$

which implies that

$$\psi_\delta(1) = \frac{1}{\delta}. \tag{245}$$

Furthermore, we have that

$$\mathbb{E}\left\{\frac{\mathcal{T}_\delta^*(Y)(G^2-1)}{1-\mathcal{T}_\delta^*(Y)}\right\} = \sqrt{\frac{\delta_{\mathrm{u}}}{\delta}}\int_{\mathbb{R}} \frac{\left(\mathbb{E}_G\{p(y \mid G)(G^2-1)\}\right)^2}{\mathbb{E}_G\{p(y \mid G)\}}\,\mathrm{d}y = \frac{1}{\sqrt{\delta\cdot\delta_{\mathrm{u}}}} > \frac{1}{\delta},$$

which implies that

$$\phi(1) = \frac{1}{\sqrt{\delta\cdot\delta_{\mathrm{u}}}} > \frac{1}{\delta}, \tag{246}$$

as $\delta > \delta_{\mathrm{u}}$. By putting (244), (245), and (246) together, we obtain that

$$\phi(\bar{\lambda}_\delta) > \zeta_\delta(\bar{\lambda}_\delta). \tag{247}$$

Recall that $\zeta_\delta(\lambda)$ is monotone non-decreasing and $\phi(\lambda)$ is monotone non-increasing. Consequently, (247) implies that $\lambda_\delta^* > \bar{\lambda}_\delta$. Thus, we conclude that

$$\begin{aligned}\lim_{n\to\infty} \lambda_1^{\widetilde{\boldsymbol{D}}_n} &= \zeta_\delta(\lambda_\delta^*) = \psi_\delta(\lambda_\delta^*)\\ &> \psi_\delta(\bar{\lambda}_\delta) = \psi_\delta(1) = \frac{1}{\delta}.\end{aligned} \tag{248}$$

Now, we show that $\mathcal{T}_\delta^*$ fulfills the hypotheses of Lemma 2 by using arguments similar to those at the end of the proof of Theorem 2. First of all, since $\mathcal{T}^*(y) \le 1$, we have that $\mathcal{T}_\delta^*(y)$ is bounded. Furthermore, if $\mathcal{T}_\delta^*(y)$ is equal to the constant value 0, then $\delta_{\mathrm{u}} = \infty$ and the claim of Theorem 6 trivially holds. Hence, we can assume that $\mathbb{P}(\mathcal{T}_\delta^*(Y) = 0) < 1$. Let τ be the supremum of the support of $\mathcal{T}_\delta^*(Y)$. If $\mathbb{P}(\mathcal{T}_\delta^*(Y) =$

$\tau) > 0$, then the condition (82) is satisfied and the proof is complete. Otherwise, for any $\epsilon_1 > 0$, there exists $\Delta_1(\epsilon_1)$ such that Eq. (115) holds. Define $\mathcal{T}_\delta^*(y, \epsilon_1)$ as in (116). Clearly, the random variable $\mathcal{T}_\delta^*(Y, \epsilon_1)$ has a point mass at δ; hence, condition (82) is satisfied. As a final step, we show that we can take $\epsilon_1 \downarrow 0$. Define

$$\widetilde{\boldsymbol{D}}_n(\epsilon_1) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{T}_\delta^*(y_i, \epsilon_1)\tilde{\boldsymbol{a}}_i\tilde{\boldsymbol{a}}_i^*.$$

Then,

$$\left\|\widetilde{\boldsymbol{D}}_n(\epsilon_1) - \widetilde{\boldsymbol{D}}_n\right\|_{\text{op}} \le C_1 \cdot \Delta_1(\epsilon_1), \tag{249}$$

where the constant C_1 depends only on n/d. Consequently, by using (249) and Weyl's inequality, we conclude that

$$|\lambda_1^{\widetilde{\boldsymbol{D}}_n(\epsilon_1)} - \lambda_1^{\widetilde{\boldsymbol{D}}_n}| \le C_1 \cdot \Delta_1(\epsilon_1). \tag{250}$$

Hence, for any n, as ϵ_1 tends to 0, the largest eigenvalue of $\widetilde{\boldsymbol{D}}_n(\epsilon_1)$ tends to the largest eigenvalue of $\widetilde{\boldsymbol{D}}_n$, which concludes the proof. □

References

1. Arora, S., Ge, R., Ma, T., Moitra, A.: Simple, efficient, and neural algorithms for sparse coding. In: Conference on Learning Theory (COLT), pp. 113–149. Paris, France (2015)
2. Bahmani, S., Romberg, J.: Phase retrieval meets statistical learning theory: A flexible convex relaxation. In: Proc. of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 252–260. Fort Lauderdale, FL (2017)
3. Bai, Z., Yao, J.: On sample eigenvalues in a generalized spiked population model. Journal of Multivariate Analysis **106**, 167–177 (2012)
4. Balan, R., Casazza, P., Edidin, D.: On signal reconstruction without phase. Applied and Computational Harmonic Analysis **20**(3), 345–356 (2006)
5. Bandeira, A.S., Cahill, J., Mixon, D.G., Nelson, A.A.: Saving phase: Injectivity and stability for phase retrieval. Applied and Computational Harmonic Analysis **37**(1), 106–125 (2014)
6. Barbier, J., Krzakala, F., Macris, N., Miolane, L., Zdeborová, L.: Phase transitions, optimal errors and optimality of message-passing in generalized linear models (2017). arXiv:1708.03395
7. Barbier, J., Macris, N., Dia, M., Krzakala, F.: Mutual information and optimality of approximate message-passing in random linear estimation (2017). arXiv:1311.2445
8. Bayati, M., Lelarge, M., Montanari, A.: Universality in polytope phase transitions and message passing algorithms. Annals of Applied Probability **25**(2), 753–822 (2015)
9. Bayati, M., Montanari, A.: The dynamics of message passing on dense graphs, with applications to compressed sensing. IEEE Trans. Inform. Theory **57**, 764–785 (2011)
10. Bayati, M., Montanari, A.: The LASSO risk for Gaussian matrices. IEEE Trans. Inform. Theory **58**(4), 1997–2017 (2012)
11. Belinschi, S.T., Bercovici, H., Capitaine, M., Février, M.: Outliers in the spectrum of large deformed unitarily invariant models (2015). arXiv:1412.4916
12. Benaych-Georges, F., Nadakuditi, R.R.: The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. Advances in Mathematics **227**(1), 494–521 (2011)
13. Cai, T.T., Li, X., Ma, Z.: Optimal rates of convergence for noisy sparse phase retrieval via thresholded Wirtinger flow. The Annals of Statistics **44**(5), 2221–2251 (2016)

14. Candès, E.J., Eldar, Y.C., Strohmer, T., Voroninski, V.: Phase retrieval via matrix completion. SIAM Review **57**(2), 225–251 (2015)
15. Candès, E.J., Li, X., Soltanolkotabi, M.: Phase retrieval from coded diffraction patterns. Applied and Computational Harmonic Analysis **39**(2), 277–299 (2015)
16. Candès, E.J., Li, X., Soltanolkotabi, M.: Phase retrieval via Wirtinger flow: Theory and algorithms. IEEE Trans. Inform. Theory **61**(4), 1985–2007 (2015)
17. Candès, E.J., Strohmer, T., Voroninski, V.: Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. Communications on Pure and Applied Mathematics **66**(8), 1241–1274 (2013)
18. Chen, Y., Candès, E.J.: Solving random quadratic systems of equations is nearly as easy as solving linear systems. In: Advances in Neural Information Processing Systems, pp. 739–747 (2015)
19. Chen, Y., Candès, E.J.: The projected power method: An efficient algorithm for joint alignment from pairwise differences (2016). arXiv:1609.05820
20. Chen, Y., Candès, E.J.: Solving random quadratic systems of equations is nearly as easy as solving linear systems. Communications on Pure and Applied Mathematics **70**, 0822–0883 (2017)
21. Conca, A., Edidin, D., Hering, M., Vinzant, C.: An algebraic characterization of injectivity in phase retrieval. Applied and Computational Harmonic Analysis **38**(2), 346–356 (2015)
22. Corbett, J.V.: The Pauli problem, state reconstruction and quantum-real numbers. Reports on Mathematical Physics **57**(1), 53–68 (2006)
23. Davis, C., Kahan, W.M.: The rotation of eigenvectors by a perturbation. III. SIAM Journal on Numerical Analysis **7**(1), 1–46 (1970)
24. Decelle, A., Krzakala, F., Moore, C., Zdeborová, L.: Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. Physical Review E **84**(6), 066,106 (2011)
25. Demanet, L., Jugnon, V.: Convex recovery from interferometric measurements. IEEE Trans. Computational Imaging **3**(2), 282–295 (2017)
26. Deshpande, Y., Montanari, A.: Finding hidden cliques of size $\sqrt{N/e}$ in nearly linear time. Foundations of Computational Mathematics pp. 1–60 (2013)
27. Dhifallah, O., Lu, Y.M.: Fundamental limits of PhaseMax for phase retrieval: A replica analysis (2017). arXiv:1708.03355
28. Dhifallah, O., Thrampoulidis, C., Lu, Y.M.: Phase retrieval via linear programming: Fundamental limits and algorithmic improvements. In: 55th Annual Allerton Conference on Communication, Control, and Computing (2017). arXiv:1710.05234
29. Donoho, D.L., Javanmard, A., Montanari, A.: Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing. IEEE Trans. Inform. Theory **59**(11), 7434–7464 (2013)
30. Donoho, D.L., Maleki, A., Montanari, A.: Message Passing Algorithms for Compressed Sensing. Proceedings of the National Academy of Sciences **106**, 18,914–18,919 (2009)
31. Donoho, D.L., Maleki, A., Montanari, A.: The noise-sensitivity phase transition in compressed sensing. IEEE Trans. Inform. Theory **57**(10), 6920–6941 (2011)
32. Donoho, D.L., Montanari, A.: High dimensional robust M-estimation: Asymptotic variance via approximate message passing. Probability Theory and Related Fields **166**(3–4), 935–969 (2016)
33. Duchi, J.C., Ruan, F.: Solving (most) of a set of quadratic equalities: Composite optimization for robust phase retrieval (2017). arXiv:1705.02356
34. Fienup, J.R.: Phase retrieval algorithms: A comparison. Applied Optics **21**(15), 2758–2769 (1982)
35. Fienup, J.R., Dainty, J.C.: Phase retrieval and image reconstruction for astronomy. Image Recovery: Theory and Application pp. 231–275 (1987)
36. Gerchberg, R.W.: A practical algorithm for the determination of the phase from image and diffraction plane pictures. Optik **35**, 237–246 (1972)
37. Goldstein, T., Studer, C.: Phasemax: Convex phase retrieval via basis pursuit (2016). arXiv:1610.07531
38. Harrison, R.W.: Phase problem in crystallography. J. Optical Soc. America A **10**(5), 1046–1055 (1993)
39. Horn, R.A., Johnson, C.R.: Matrix analysis. Cambridge University Press (2012)
40. Jain, P., Netrapalli, P., Sanghavi, S.: Low-rank matrix completion using alternating minimization. In: Proc. of the 45th Ann. ACM Symp. on Theory of Computing (STOC), pp. 665–674. ACM, Palo Alto, CA (2013)
41. Javanmard, A., Montanari, A.: State evolution for general approximate message passing algorithms, with applications to spatial coupling. Information and Inference pp. 115–144 (2013)

42. Kabashima, Y., Krzakala, F., Mézard, M., Sakata, A., Zdeborová, L.: Phase transitions and sample complexity in bayes-optimal matrix factorization. IEEE Trans. Inform. Theory **62**(7), 4228–4265 (2016)
43. Karoui, N.E.: Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: Rigorous results (2013). arXiv:1311.2445
44. Keshavan, R.H., Montanari, A., Oh, S.: Matrix completion from a few entries. IEEE Trans. Inform. Theory **56**(6), 2980–2998 (2010)
45. Krzakala, F., Mézard, M., Zdeborová, L.: Phase diagram and approximate message passing for blind calibration and dictionary learning. In: Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT), pp. 659–663. Istanbul, Turkey (2013)
46. Lee, K., Li, Y., Junge, M., Bresler, Y.: Blind recovery of sparse signals from subsampled convolution. IEEE Trans. Inform. Theory **63**(2), 802–821 (2017)
47. Li, G., Gu, Y., Lu, Y.M.: Phase retrieval using iterative projections: Dynamics in the large systems limit. In: Proc. of the 53rd Annual Allerton Conf. on Commun., Control, and Computing (Allerton), pp. 1114–1118. Monticello, IL (2015)
48. Li, X., Ling, S., Strohmer, T., Wei, K.: Rapid, robust, and reliable blind deconvolution via nonconvex optimization (2016). arXiv:1606.04933
49. Lu, Y.M., Li, G.: Phase transitions of spectral initialization for high-dimensional nonconvex estimation (2017). arXiv:1702.06435
50. Marčenko, V.A., Pastur, L.A.: Distribution of eigenvalues in certain sets of random matrices. Mat. Sb. (N.S.) **72**, 457–483 (in Russian) (1967)
51. Miao, J., Ishikawa, T., Shen, Q., Earnest, T.: Extending X-ray crystallography to allow the imaging of noncrystalline materials, cells, and single protein complexes. Annu. Rev. Phys. Chem. **59**, 387–410 (2008)
52. Millane, R.P.: Phase retrieval in crystallography and optics. J. Optical Soc. America A **7**(3), 394–411 (1990)
53. Montanari, A., Richard, E.: Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. IEEE Trans. Inform. Theory **62**(3), 1458–1484 (2016)
54. Montanari, A., Venkataramanan, R.: Estimation of low-rank matrices via approximate message passing (2017). arXiv:1711.01682
55. Mossel, E., Neeman, J., Sly, A.: Belief propagation, robust reconstruction and optimal recovery of block models. In: Conference on Learning Theory (COLT), pp. 356–370. Barcelona, Spain (2014)
56. Mossel, E., Xu, J.: Density evolution in the degree-correlated stochastic block model. In: Conference on Learning Theory (COLT), pp. 1319–1356. New York, NY (2016)
57. Netrapalli, P., Jain, P., Sanghavi, S.: Phase retrieval using alternating minimization. In: Advances in Neural Information Processing Systems, pp. 2796–2804 (2013)
58. Neykov, M., Wang, Z., Liu, H.: Agnostic estimation for misspecified phase retrieval models. In: Advances in Neural Information Processing Systems, pp. 4089–4097 (2016)
59. Oymak, S., Thrampoulidis, C., Hassibi, B.: The squared-error of generalized LASSO: A precise analysis. In: Proc. of the 51st Annual Allerton Conf. on Commun., Control, and Computing (Allerton), pp. 1002–1009. Monticello, IL (2013)
60. Plan, Y., Vershynin, R.: The generalized lasso with non-linear observations. IEEE Transactions on information theory **62**(3), 1528–1537 (2016)
61. Rangan, S.: Generalized Approximate Message Passing for Estimation with Random Linear Mixing. In: Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT), pp. 2168–2172. St. Petersburg (2011)
62. Rangan, S., Goyal, V.K.: Recursive consistent estimation with bounded noise. IEEE Trans. Inform. Theory **47**(1), 457–464 (2001)
63. Reeves, G., Pfister, H.D.: The replica-symmetric prediction for compressed sensing with Gaussian matrices is exact. In: Proc. of the IEEE Int. Symposium on Inform. Theory (ISIT), pp. 665–669. Barcelona, Spain (2016)
64. Schniter, P., Rangan, S.: Compressive phase retrieval via generalized approximate message passing. IEEE Transactions on Signal Processing **63**(4), 1043–1055 (2015)
65. Shechtman, Y., Eldar, Y.C., Cohen, O., Chapman, H.N., Miao, J., Segev, M.: Phase retrieval with application to optical imaging: a contemporary overview. IEEE Signal Processing Magazine **32**(3), 87–109 (2015)
66. Silverstein, J.W., Bai, Z.: Spectral Analysis of Large Dimensional Random Matrices. (2^{nd} edition) Springer, (2010)

67. Silverstein, J.W., Choi, S.I.: Analysis of the limiting spectral distribution of large-dimensional random matrices. Journal of Multivariate Analysis **54**(2), 295–309 (1995)
68. Soltanolkotabi, M.: Structured signal recovery from quadratic measurements: Breaking sample complexity barriers via nonconvex optimization (2017). arXiv:1702.06175
69. Speicher, R.: Free convolution and the random sum of matrices. Publ. Res. Inst. Math. Sci. **29**, 731–744 (1993)
70. Su, W., Candès, E.J.: Slope is adaptive to unknown sparsity and asymptotically minimax. Annals of Statistics **44**(3), 1038–1068 (2016)
71. Thrampoulidis, C., Abbasi, E., Hassibi, B.: Lasso with non-linear measurements is equivalent to one with linear measurements. In: Advances in Neural Information Processing Systems, pp. 3420–3428 (2015)
72. Unser, M., Eden, M.: Maximum likelihood estimation of linear signal parameters for Poisson processes. IEEE Trans. Acoust., Speech, and Signal Process. **36**(6), 942–945 (1988)
73. Venkataramanan, R., Johnson, O.: Strong converse bounds for high-dimensional estimation (2017). arXiv:1706.04410
74. Voiculescu, D.: Limit laws for random matrices and free products. Inventiones Mathematicae **104**, 201–220 (1991)
75. Waldspurger, I., d'Aspremont, A., Mallat, S.: Phase recovery, maxcut and complex semidefinite programming. Mathematical Programming **149**(1-2), 47–81 (2015)
76. Walther, A.: The question of phase retrieval in optics. Journal of Modern Optics **10**(1), 41–49 (1963)
77. Wang, G., Giannakis, G.B.: Solving random systems of quadratic equations via truncated generalized gradient flow. In: Advances in Neural Information Processing Systems, pp. 568–576 (2016)
78. Wang, G., Giannakis, G.B., Eldar, Y.C.: Solving systems of random quadratic equations via truncated amplitude flow (2016). arXiv:1605.08285
79. Wang, G., Giannakis, G.B., Saad, Y., Chen, J.: Solving almost all systems of random quadratic equations (2017). arXiv:1705.10407
80. Wei, K.: Solving systems of phaseless equations via Kaczmarz methods: A proof of concept study. Inverse Problems **31**(12) (2015)
81. Yang, F., Lu, Y.M., Sbaiz, L., Vetterli, M.: Bits from photons: Oversampled image acquisition using binary Poisson statistics. IEEE Trans. Image Process. **21**(4), 1421–1436 (2012)
82. Zdeborová, L., Krzakala, F.: Statistical physics of inference: Thresholds and algorithms. Advances in Physics **65**(5), 453–552 (2016)
83. Zhang, H., Liang, Y.: Reshaped Wirtinger Flow for solving quadratic system of equations. In: Advances in Neural Information Processing Systems, pp. 2622–2630 (2016)

Affiliations

Marco Mondelli[1] · Andrea Montanari[2,3]

✉ Marco Mondelli
mondelli@stanford.edu

Andrea Montanari
montanari@stanford.edu

[1] Department of Electrical Engineering, Stanford University, Packard Building 239, 350 Serra Mall, Stanford, CA 94305, USA

[2] Department of Electrical Engineering, Stanford University, Packard Building 272, 350 Serra Mall, Stanford, CA 94305, USA

[3] Department of Statistics, Stanford University, 390 Serra Mall, Stanford, CA 94305, USA