The Annals of Statistics
2026, Vol. 54, No. 1, 466–489
https://doi.org/10.1214/25-AOS2575

OPTIMALITY OF APPROXIMATE MESSAGE PASSING FOR SPIKED MATRIX MODELS WITH ROTATIONALLY INVARIANT NOISE

BY RISHABH DUDEJA[1,a], SONGBIN LIU[2,b] AND JUNJIE MA[2,c]

[1]*Department of Statistics, University of Wisconsin, Madison,* [a]*rdudeja@wisc.edu*
[2]*Academy of Mathematics and Systems Science, Chinese Academy of Sciences,* [b]*liusongbin@lsec.cc.ac.cn,* [c]*majunjie@lsec.cc.ac.cn*

We study the problem of estimating a rank-one signal matrix from a noisy observed matrix corrupted by additive rotationally invariant noise. We develop a new class of approximate message passing algorithms for this problem and provide a simple and concise characterization of their dynamics in the high-dimensional limit. At each iteration, these algorithms leverage prior knowledge about the noise structure by applying a nonlinear matrix denoiser to the eigenvalues of the observed matrix, and utilize prior information regarding the signal structure by applying a nonlinear iterate denoiser to the previous iterates generated by the algorithm. We derive the optimal choices for both the matrix and iterate denoisers and demonstrate that the resulting algorithm achieves the lowest possible asymptotic estimation error among a broad class of iterative algorithms under a fixed iteration budget.

1. Introduction. We consider the problem of estimating a symmetric rank-one matrix from a noisy $N \times N$ observed matrix $\boldsymbol{Y}$ generated from the *spiked matrix model* [33]:

$$\boldsymbol{Y} = \frac{\theta}{N}\boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T} + \boldsymbol{W}, \tag{1}$$

where $\boldsymbol{x}_\star \in \mathbb{R}^N$ is the N-dimensional unknown signal of interest, $\theta \geq 0$ is the signal-to-noise ratio (SNR) parameter and $\boldsymbol{W}$ is a symmetric noise matrix. This model and its variants have been used to study a broad range of statistical inference problems, including sparse PCA [18, 34, 74], community detection [1, 17] and group synchronization [13, 28, 61, 68].

Wigner noise model. The most well-studied variant of the spiked model is the *Spiked Wigner Model* [7, 18, 25, 26, 51, 62], which assumes that the noise matrix has i.i.d. Gaussian entries. A rich line of work in high-dimensional statistics and random matrix theory has studied the problem from various perspectives.

- *Design and analysis of estimators.* A natural estimator for the signal is the PCA estimator or the leading eigenvector of the observed matrix. A line of work [2, 3, 10, 30, 37, 59, 60] initiated by Baik, Ben Arous and Péché [2] has obtained a sharp asymptotic characterization of the performance of this estimator in the high-dimensional limit and uncovered surprising properties of this estimator. One way to improve the performance of PCA is to exploit prior structural information about the signal (e.g., sparsity). Approximate message passing (AMP) algorithms are a popular class of computationally efficient iterative algorithms designed to exploit such structural information. These algorithms were first discovered in the context of compressed sensing [8, 11, 20, 35] and have been adapted for the spiked Wigner model [31, 36, 44, 45, 51, 57]. A particularly attractive feature of these algorithms is that their performance in the high-dimensional limit is characterized by a

Received June 2024; revised August 2025.
MSC2020 subject classifications. Primary 62H99; secondary 62F15, 62E20, 68Q17.
Key words and phrases. Statistical-computational gap, low-rank matrix estimation, approximate message passing.

simple deterministic recursion known as *state evolution*, which makes it possible to assess their information-theoretic optimality (or sub-optimality).

- *Information-theoretic limits.* Under the assumption that the signal is drawn from a known prior distribution, the optimal estimator is the Bayes estimator $\mathbb{E}[\boldsymbol{x}_\star|\boldsymbol{Y}]$. Using the powerful cavity method from statistical physics, Lesieur, Krzakala and Zdeborová [42] derived a conjecture for the Bayes risk (or the asymptotic performance of the Bayes estimator) for the spiked Wigner model. This conjecture has been proved rigorously in increasing generality in a series of influential works [6, 7, 9, 18, 25, 38, 41, 49]. This formula for asymptotic Bayes risk provides a fundamental information-theoretic limit on the performance of any estimator for the problem.
- *Computational limits.* In general, computing the Bayes-optimal estimator is computationally intractable in high dimensions. However, a suitably designed AMP algorithm called Bayes-optimal AMP can attain the Bayes risk for the problem under sufficiently high signal-to-noise ratios. For lower signal-to-noise ratios, the Bayes-optimal AMP algorithm fails to do so, and no computationally efficient algorithm is known to achieve the Bayes risk [6, 38, 43, 49]. This has led to a popular conjecture that the Bayes-optimal AMP algorithm is the optimal polynomial time algorithm for this problem [43, 49]. Celentano, Montanari and Wu [14], Montanari and Wu [53] and Montanari and Wein [52] (see also the earlier work of Schramm and Wein [66]) have provided evidence for this conjecture by showing that the Bayes-optimal AMP algorithm achieves the lowest possible estimation error among a broad class of iterative algorithms and low-degree polynomial estimators.

Collectively, these works have enriched our understanding of the fundamental trade-offs between statistical optimality and computational efficiency in high-dimensional statistics.

Rotationally invariant noise model. In this paper, we study a natural generalization of the i.i.d. Gaussian noise model known as the rotationally invariant noise model where one posits that the eigenvectors of the noise matrix are given by a uniformly random orthogonal matrix independent of the eigenvalues. The rotationally invariant noise model is intended to model noise matrices with strong statistical dependence whose eigenvectors are generic [27]. Several works (see, e.g., [19, 46, 56]) have observed that this can be a good model in some applications. Benaych-Georges and Nadakuditi [10] analyzed the performance of PCA (or spectral estimators) for this noise model. A line of work [27, 50, 54, 73] initiated by Opper, Cakmak and Winther [54] and Fan [27] has developed AMP algorithms for this problem, which can improve the performance of PCA by exploiting signal structure. Unlike AMP algorithms for i.i.d. Gaussian noise, the dynamics of these algorithms are characterized by a significantly more intricate state evolution, suggesting that analyzing the rotationally invariant noise model requires new ideas beyond the i.i.d. Gaussian case. Our current understanding of the fundamental information-theoretic and computational limits for the rotationally invariant noise model is extremely limited. Recent work by Barbier et al. [4] studies the spiked matrix model with rotationally invariant noise under the assumption that the noise matrix is drawn from the *trace ensemble*. Under this model, the density of $\boldsymbol{W}$ is given by

$$p(\boldsymbol{W}) \propto \exp\left(-\frac{N}{2}\sum_{i=1}^{N} V\big(\lambda_i(\boldsymbol{W})\big)\right) \quad \text{where } (\lambda_i(\boldsymbol{W}))_{i\in[N]} \text{ denote the eigenvalues of } \boldsymbol{W},$$

and the *potential function* $V : \mathbb{R} \to \mathbb{R}$ is a functional parameter for the noise model. This assumption ensures that the problem has a well-defined likelihood, enabling the study of information-theoretic limits. Moreover, an appropriate choice of V can capture a wide range of noise eigenvalue spectrums in different applications. Barbier et al. [4] make progress towards understanding the information-theoretic limits of the problem and provide important insights regarding the structure of the optimal computationally efficient algorithms.

- *Information-theoretic limits.* Barbier et al. [4] develop a general (although nonrigorous) recipe based on the replica method to derive conjectured formulas for Bayes risk under the assumption that the potential V is a polynomial function, providing explicit conjectured formulas for the asymptotic Bayes risk when V is a quartic (degree four) or a sestic (degree six) polynomial. As the degree of V grows, the derivation and the resulting formulas become increasingly complex and a general conjecture for the Bayes risk is unavailable.
- *Optimal computationally efficient algorithms.* Surprisingly, Barbier et al. [4] demonstrate that a natural generalization of the Bayes-optimal AMP algorithm (this is the optimal iterative algorithm for i.i.d. Gaussian noise and is conjectured to be the optimal polynomial-time algorithm) to the rotationally invariant noise model is suboptimal. The authors develop AMP algorithms that achieve improved performance by applying a nonlinear matrix denoiser to the eigenvalues of the observed matrix $\boldsymbol{Y}$ and characterize their state evolution. This is in sharp contrast to the i.i.d. Gaussian noise model, where such a matrix denoising step is unnecessary. Based on nonrigorous statistical physics techniques [55], the authors provide a procedure to derive good matrix denoisers for the problem. The authors propose matrix denoisers with explicit formulas when V is a quartic or sestic polynomial (a general formula is not available). However, the state evolution of the resulting AMP algorithm is quite complicated; hence, the resulting algorithm's optimality (or suboptimality) properties are not understood.

Our contributions. We take inspiration from the insights of Barbier et al. [4] and study the spiked matrix model with rotationally invariant noise from an algorithmic point of view. Our results are not restricted to the trace ensemble with a polynomial potential V. Rather, they apply to all rotationally invariant noise matrices that meet mild regularity conditions. Our main contributions are:

- We develop a new class of AMP algorithms for this problem and provide a state evolution result that characterizes their dynamics in the high-dimensional limit (Theorem 1 in Section 3.1). At each iteration, these algorithms exploit the noise structure by applying a nonlinear matrix denoiser to the eigenvalues of the observed matrix and the signal structure by applying a nonlinear iterate denoiser to the previous iterates. These algorithms can be viewed as natural analogs of orthogonal [48] or vector AMP algorithms [64] developed for compressed sensing.
- A key feature of the AMP algorithms proposed in this work is that their state evolution is significantly simpler than that of existing AMP algorithms for this problem. Consequently, we are able to exploit our result to derive the optimal choices for the matrix and iterate denoisers (Section 3.2). Interestingly, we find that the matrix denoisers that optimize the performance of the AMP algorithm are closely related to the eigenvalue shrinkage estimators discovered by Bun et al. [12] in a separate line of work on denoising high-rank and unstructured signal matrices corrupted with rotationally invariant noise [12, 39, 40, 47, 63, 67].
- Building on the techniques developed by Celentano, Montanari and Wu [14] and Montanari and Wu [53] (in the context of i.i.d. Gaussian noise), we show that the AMP algorithm with the optimal choices for the matrix and iterate denoisers achieves the smallest possible asymptotic estimation error among a broad class of iterative algorithms under a fixed iteration budget (Theorem 2 in Section 3.3). This suggests that this algorithm might be the natural candidate for the optimal polynomial-time estimator for this problem.
- Finally, our results also suggest a general and concise conjecture for a set of fixed-point equations, which characterize the Bayes risk (Section 3.4) for the problem, which we hope can be proved rigorously in the future.

Related concurrent work. Independent work by Barbier et al. [5], which appeared shortly after our arXiv submission, also studies this problem from a different and complementary perspective. Using the replica method, the authors derive conjectures for the asymptotic Bayes risk and mutual information for this problem, generalizing their prior work [4]. The fixed-point equations characterizing the Bayes risk coincide with those derived in our paper, providing further evidence supporting the conjecture. The authors also derive the Thouless–Anderson–Palmer (TAP) equations for this problem. These are high-dimensional nonlinear fixed-point equations that characterize the Bayes estimator $\mathbb{E}[\boldsymbol{x}_\star|\boldsymbol{Y}]$. Using the TAP equations, Barbier et al. [5] propose a natural algorithm to compute the Bayes estimator. While a state evolution result for this algorithm is not available, based on simulations, the authors conjecture that this algorithm achieves the Bayes-optimal performance whenever it is computationally feasible to do so.

Organization. This paper is organized as follows. We start with some preliminary results in Section 2. The main results of this paper are presented in Section 3. Section 4 highlights the key ideas behind our results. Numerical experiments are presented in Section 5. The complete proofs of our main results and additional numerical results not presented in the main paper are provided in the Supplementary Material [22].

Notation. We conclude the Introduction by defining the notation used in this paper.
Some common sets. The sets $\mathbb{N}, \mathbb{R}, \mathbb{C}$ represent the set of positive integers, real numbers and complex numbers, respectively. For $N \in \mathbb{N}$, $[N]$ is the set $\{1, 2, 3, \dots, N\}$ and $\mathbb{O}(N)$ denotes the set of $N \times N$ orthogonal matrices.
Linear algebra. For vectors $u, v \in \mathbb{R}^k$, $\|u\|$ is the ℓ_2 norm of u, $\langle u, v\rangle = \sum_{i=1}^k u_i v_i$ denotes the standard inner product on $\mathbb{R}^k$, and $\operatorname{diag}(u)$ represents the $k \times k$ diagonal matrix constructed by placing the entries of u along the diagonal. For a matrix $M \in \mathbb{R}^{k\times k}$, $\operatorname{Tr}[M]$, $\|M\|_{\mathrm{op}}$, $\|M\|$ represent the trace, operator (spectral) norm and Frobenius norm of M, respectively. If M is symmetric, we denote the sorted eigenvalues of M by $\lambda_1(M) \geq \cdots \geq \lambda_k(M)$. We use 1_k to denote the vector $(1, 1, \dots, 1)$ in $\mathbb{R}^k$, 0_k to denote the zero vector $(0, 0, \dots, 0)$ in $\mathbb{R}^k$, I_k denotes the $k \times k$ identity matrix and $e_1, e_2, \dots, e_k$ to denote the standard basis vectors in $\mathbb{R}^k$. When the dimension is clear from the context, we will abbreviate $1_k, 0_k, I_k$ as $1, 0, I$. We use the bold-face font for vectors and matrices whose dimensions diverge as N (the dimension of the signal vector) grows to ∞. For example, the signal $\boldsymbol{x}_\star \in \mathbb{R}^N$, and the noise matrix $\boldsymbol{W} \in \mathbb{R}^{N\times N}$ are bold-faced.
Probability. We use $\mathbb{E}[\cdot]$, $\operatorname{Var}[\cdot]$, $\operatorname{Cov}[\cdot]$ to denote expectations, variances and covariances of random variables. The Gaussian distribution on $\mathbb{R}^k$ with mean vector $\mu \in \mathbb{R}^k$ and covariance matrix $\Sigma \in \mathbb{R}^{k\times k}$ is denoted by $\mathcal{N}(\mu, \Sigma)$. For a finite set A, $\mathsf{Unif}(A)$ represents the uniform distribution on A. We will use $\mathsf{Unif}(\mathbb{O}(N))$ to denote the Haar measure on the orthogonal group $\mathbb{O}(N)$. For any $x \in \mathbb{R}$, the probability measure δ_x on $\mathbb{R}$ denotes the point mass at x. We use $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{d}$ to denote convergence in probability and distribution, respectively. For a sequence of real-valued random variables $(Y_N)_{N\in\mathbb{N}}$, we say that $\operatorname{plim} Y_N = y$ if $Y_N \xrightarrow{\mathbb{P}} y$, $\operatorname{plim\,sup} Y_N \leq y$ if for any $\epsilon > 0$, $\lim_{N\to\infty} \mathbb{P}(Y_N \geq y + \epsilon) = 0$ and $\operatorname{plim\,inf} Y_N \geq y$ if for any $\epsilon > 0$, $\lim_{N\to\infty} \mathbb{P}(Y_N \leq y - \epsilon) = 0$.

2. Preliminaries.

We begin by introducing our assumptions along with some concepts that play an important role in this paper.

2.1. *Convergence and asymptotic equivalence of high-dimensional vectors.*

We will rely on the following notions of convergence and equivalence of high-dimensional vectors.

DEFINITION 1. Let $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_\ell)$ be a collection of random vectors in $\mathbb{R}^N$. We say that the empirical distribution of the entries of the vectors $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_\ell)$ converges to random variables $(\mathsf{V}_1, \dots, \mathsf{V}_\ell)$ as $N \to \infty$ if, for any test function $h : \mathbb{R}^\ell \to \mathbb{R}$ which satisfies

$$|h(v) - h(v')| \le L\|v - v'\| \cdot (1 + \|v\| + \|v'\|) \quad \forall v, v' \in \mathbb{R}^\ell, \tag{2}$$

for some $L < \infty$, we have

$$\frac{1}{N}\sum_{i=1}^{N} h\big(v_1[i], \dots, v_\ell[i]\big) \xrightarrow{\mathbb{P}} \mathbb{E}\big[h(\mathsf{V}_1, \dots, \mathsf{V}_\ell)\big] \quad \text{as } N \to \infty.$$

We denote convergence in this sense using the notation: $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_\ell) \xrightarrow{W_2} (\mathsf{V}_1, \dots, \mathsf{V}_\ell)$. We refer the reader to [8, 27, 29] for more information on this type of convergence. We say that two N-dimensional random vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are asymptotically equivalent if

$$\frac{\|\boldsymbol{u} - \boldsymbol{v}\|^2}{N} \xrightarrow{\mathbb{P}} 0 \quad \text{as } N \to \infty.$$

We denote equivalence in this sense using the notation: $\boldsymbol{u} \overset{N\to\infty}{\simeq} \boldsymbol{v}$.

2.2. *Signal and noise models.* Recall that our goal is to recover the signal vector $\boldsymbol{x}_\star \in \mathbb{R}^N$ from the noisy observation $\boldsymbol{Y} = (\theta/N) \cdot \boldsymbol{x}_\star \boldsymbol{x}_\star^\top + \boldsymbol{W}$. We will allow our estimators the flexibility to exploit any side information $\boldsymbol{a} \in \mathbb{R}^{N\times k}$ available regarding the signal. We posit the following assumptions on the signal $\boldsymbol{x}_\star$, the side information $\boldsymbol{a}$, and the noise $\boldsymbol{W}$.

ASSUMPTION 1 (Signal and noise models). The signal vector and side information satisfy $(\boldsymbol{x}_\star; \boldsymbol{a}) \xrightarrow{W_2} (\mathsf{X}_\star; \mathsf{A})$ for some limiting random variables $(\mathsf{X}_\star, \mathsf{A})$ with joint distribution π, which satisfies $\mathbb{E}[\mathsf{X}_\star^2] = 1$ and $\mathbb{E}[\|\mathsf{A}\|^2] < \infty$. We model the noise matrix $\boldsymbol{W}$ as a random matrix sampled independently of $(\boldsymbol{x}_\star; \boldsymbol{a})$, with eigendecomposition:

$$\boldsymbol{W} = \boldsymbol{U} \cdot \operatorname{diag}\big(\lambda_1(\boldsymbol{W}), \dots, \lambda_N(\boldsymbol{W})\big) \cdot \boldsymbol{U}^\top,$$

where the matrix of eigenvectors $\boldsymbol{U} \sim \mathsf{Unif}(\mathbb{O}(N))$ is a Haar-distributed random orthogonal matrix and the eigenvalues $\lambda_1(\boldsymbol{W}), \dots, \lambda_N(\boldsymbol{W})$ are deterministic. We assume that $\|\boldsymbol{W}\|_{\mathrm{op}}$ is bounded by a N-independent constant C and the spectral measure μ_N of $\boldsymbol{W}$:

$$\mu_N \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(\boldsymbol{W})}$$

converges weakly to a compactly supported distribution μ on $\mathbb{R}$. We require the limiting spectral measure μ to be absolutely continuous with respect to the Lebesgue measure. Furthermore, the density of μ, which we denote using the same symbol $\mu : \mathbb{R} \to \mathbb{R}$, is assumed to be Hölder continuous in the following sense:

$$|\mu(\lambda) - \mu(\lambda')| \le L \cdot |\lambda - \lambda'|^\alpha \quad \forall \lambda, \lambda' \in \mathbb{R}, \tag{3}$$

where $L < \infty$ and $\alpha > 0$ are some constants.

2.3. *Spectral measure in the signal direction.* Let ν_N denote the spectral measure of the observed matrix $\boldsymbol{Y}$ in the direction of the signal:

$$\nu_N \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^{N} \langle \boldsymbol{u}_i(\boldsymbol{Y}), \boldsymbol{x}_\star\rangle^2 \cdot \delta_{\lambda_i(\boldsymbol{Y})}, \tag{4}$$

where $\lambda_1(\boldsymbol{Y}), \dots, \lambda_N(\boldsymbol{Y})$ denote the eigenvalues of $\boldsymbol{Y}$ and $\boldsymbol{u}_1(\boldsymbol{Y}), \dots, \boldsymbol{u}_N(\boldsymbol{Y})$ denote the corresponding eigenvectors. The measure ν_N and its weak limit ν play a key role in our analysis. Expressing the state evolution of the proposed AMP algorithm in terms of ν leads to a concise formula. The following lemma (proved in the Supplementary Material [22], Appendix A.1.1) provides a formula for the density of ν in terms of the function $\phi : \mathbb{R} \to \mathbb{R}$:

$$\phi(\lambda) \stackrel{\text{def}}{=} \big(1 - \pi\theta \mathscr{H}_\mu(\lambda)\big)^2 + \pi^2\theta^2\mu^2(\lambda) \quad \lambda \in \mathbb{R}. \tag{5}$$

In the above display, $\mathscr{H}_\mu : \mathbb{R} \to \mathbb{R}$ denotes the Hilbert transform of μ, which is defined as the Cauchy principal value of the following singular integral:[1]

$$\mathscr{H}_\mu(z) \stackrel{\text{def}}{=} \lim_{\epsilon \to 0} \frac{1}{\pi} \int_{|z-\lambda| \geq \epsilon} \frac{\mu(\lambda)}{z - \lambda} \, \mathrm{d}\lambda \quad \forall z \in \mathbb{R}. \tag{6}$$

The measure ν and the function ϕ will be frequently referenced in this paper.

LEMMA 1. *We have*:

1. *The measure ν_N converges weakly in probability to a compactly supported probability measure ν on $\mathbb{R}$ as $N \to \infty$.*

Let $\nu = \nu_\parallel + \nu_\perp$ denote the Lebesgue decomposition of ν into the absolutely continuous part $\nu_\parallel$ and the singular part $\nu_\perp$. Then, we have the following:

2. *For Lebesgue-almost every λ, $\phi(\lambda) \neq 0$ and for $\nu_\perp$-almost every λ, $\phi(\lambda) = 0$.*
3. *The density of the absolutely continuous part of ν is given by $\mu(\cdot)/\phi(\cdot)$ where $\mu(\cdot)$ denotes the density of μ and the function $\phi(\cdot)$ is as defined in (5).*

REMARK 1. Consider a typical situation when μ is supported on a single interval $[\lambda_-, \lambda_+]$ with a positive density on (λ_-, λ_+). The results of Benaych-Georges and Nadakuditi [10] show that when the SNR θ exceeds the critical threshold $\theta_c \stackrel{\text{def}}{=} 1/(\pi \mathscr{H}_\mu(\lambda_+))$, $\lambda_1(\boldsymbol{Y})$, the largest eigenvalue of $\boldsymbol{Y}$, separates from the support of μ and converges to a limit $\lambda_c \in (\lambda_+, \infty)$ which is the unique root of the equation (in λ):

$$\mathscr{H}_\mu(\lambda) = \frac{1}{\pi\theta} \stackrel{(5)}{\Leftrightarrow} \phi(\lambda) = 0 \quad \text{on the domain } \lambda \in \big(-\infty, \lambda_-\big] \cup \big[\lambda_+, \infty\big). \tag{7}$$

Moreover, the corresponding eigenvector satisfies

$$\frac{\langle \boldsymbol{u}_1(\boldsymbol{Y}), \boldsymbol{x}_\star \rangle^2}{N} \xrightarrow{\mathbb{P}} -\frac{1}{\theta^2 \pi \mathscr{H}'_\mu(\lambda_c)}.$$

In light of the definition of ν_N (4), we expect that ν should have a point mass at λ_c with weight $-\frac{1}{\theta^2\pi\mathscr{H}'_\mu(\lambda_c)}$. This is consistent with Lemma 1, which shows that the singular part of ν concentrates on the set $\{\lambda \in \mathbb{R} : \phi(\lambda) = 0\}$ (cf. (7)), and hence consists of a point mass at λ_c.

2.4. *Gaussian channels*. We will also use some basic notions regarding Gaussian channels, introduced in the definition below.

DEFINITION 2 (Gaussian channel). A Gaussian channel is a collection of real-valued random variables $(\mathsf{X}_\star, \mathsf{X}_1, \dots, \mathsf{X}_t; \mathsf{A})$ where the signal $\mathsf{X}_\star$ and the side information A are drawn from the prior π from Assumption 1. The observations $\mathsf{X}_1, \dots, \mathsf{X}_t$ are given by $\mathsf{X}_i = \alpha_i \mathsf{X}_\star + \mathsf{Z}_i \quad \forall i \in [t]$, where $\alpha_1, \dots, \alpha_t$ are real numbers and $(\mathsf{Z}_1, \dots, \mathsf{Z}_t) \sim \mathcal{N}(0, \Sigma)$ are zero mean and jointly Gaussian random variables, which are independent of $(\mathsf{X}_\star, \mathsf{A})$. We introduce some important notions related to Gaussian channels.

[1] Under the Hölder continuity requirement (3) on μ, the limit in (6) exists and the Hilbert transform $\mathscr{H}_\mu : \mathbb{R} \to \mathbb{R}$ is also Hölder continuous in the sense of (3); see [58], Section 2.1, and [32], Theorem 14.11a.

MMSE and MMSE estimator. The minimum mean squared error (MMSE) for estimating the signal $\mathsf{X}_\star$ based on the observations $(\mathsf{X}_1, \ldots, \mathsf{X}_t)$ and the side information A, denoted by $\mathrm{MMSE}(\mathsf{X}_\star|\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})$, is defined as

$$\mathrm{MMSE}(\mathsf{X}_\star|\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A}) \stackrel{\mathrm{def}}{=} \min_{f \in L^2(\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})} \mathbb{E}\big[\{\mathsf{X}_\star - f(\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})\}^2\big], \tag{8}$$

where the minimum is over $L^2(\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})$, which denotes the set of all measurable functions $f : \mathbb{R}^{t+k} \to \mathbb{R}$ satisfying $\mathbb{E}[f^2(\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})] < \infty$. The function $f \in L^2(\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})$ that minimizes the RHS in (8) is called the MMSE estimator for $\mathsf{X}_\star$.

DMMSE and DMMSE estimator. The divergence-free minimum mean squared error (DMMSE) for estimating the signal $\mathsf{X}_\star$ based on $(\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})$, denoted by $\mathrm{DMMSE}(\mathsf{X}_\star|\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})$, is defined as

$$\begin{aligned} &\mathrm{DMMSE}(\mathsf{X}_\star|\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A}) \\ &\stackrel{\mathrm{def}}{=} \min_{f \in L^2(\mathsf{X}_{1:t}; \mathsf{A})} \mathbb{E}\big[\{\mathsf{X}_\star - f(\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})\}^2\big] \\ &\text{subject to } \mathbb{E}\big[\mathsf{Z}_i f(\mathsf{X}_{1:t}; \mathsf{A})\big] = 0, \forall i \in [t], \end{aligned} \tag{9}$$

where $\mathsf{X}_{1:t}$ is a shorthand for the collection of random variables $(\mathsf{X}_1, \ldots, \mathsf{X}_t)$. The constraints $\mathbb{E}[\mathsf{Z}_i f(\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})] = 0$, $\forall i \in [t]$ in (9) require the estimator to be uncorrelated with the noise and are called *divergence-free* constraints. The function f, which minimizes the RHS in (9), is called the DMMSE estimator for $\mathsf{X}_\star$.

Scalar Gaussian channels. A important role is played by *scalar* Gaussian channels, which refers to a collection of real-valued random variables $(\mathsf{X}_\star, \mathsf{X}; \mathsf{A})$ generated as follows:

$$(\mathsf{X}_\star; \mathsf{A}) \sim \pi, \quad \mathsf{X}|\mathsf{X}_\star; \mathsf{A} \sim \mathcal{N}(\sqrt{\omega} \cdot \mathsf{X}_\star, 1 - \omega),$$

where $\omega \in [0, 1]$ is called the signal-to-noise ratio (SNR) of the channel. Notice that a scalar Gaussian channel is a Gaussian channel with a single observation X normalized to satisfy $\mathbb{E}[\mathsf{X}^2] = 1$. For a scalar Gaussian channel $(\mathsf{X}_\star, \mathsf{X}; \mathsf{A})$, we define several important functions, which will play a key role in this paper.

MMSE function and MMSE estimator for scalar Gaussian channels. The function $\mathsf{mmse}_\pi : [0, 1] \to [0, 1]$ represents MMSE of a scalar Gaussian channel as a function of the SNR ω:

$$\mathsf{mmse}_\pi(\omega) \stackrel{\mathrm{def}}{=} \mathrm{MMSE}(\mathsf{X}_\star|\mathsf{X}, \mathsf{A}).$$

The function $\varphi(\cdot|\omega) : \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}$ denotes the MMSE estimator for the scalar Gaussian channel at SNR ω:

$$\varphi(x; a|\omega) \stackrel{\mathrm{def}}{=} \mathbb{E}[\mathsf{X}_\star|\mathsf{X} = x, \mathsf{A} = a] \quad \forall x \in \mathbb{R}, a \in \mathbb{R}^k. \tag{10}$$

DMMSE function and DMMSE estimator for scalar Gaussian channels. The function $\mathsf{dmmse}_\pi : [0, 1] \to [0, 1]$ represents DMMSE of a scalar Gaussian channel as a function of the SNR ω:

$$\mathsf{dmmse}_\pi(\omega) \stackrel{\mathrm{def}}{=} \mathrm{DMMSE}(\mathsf{X}_\star|\mathsf{X}, \mathsf{A}).$$

The function $\bar{\varphi}(\cdot|\omega) : \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}$ denotes the DMMSE estimator for the scalar Gaussian channel at SNR ω. Ma and Ping [48] have shown that the DMMSE estimator of a scalar Gaussian channel at SNR ω is given by (see the Supplementary Material [22], Lemma 2, Appendix A, for a self-contained proof):

$$\bar{\varphi}(x; a|\omega) \stackrel{\mathrm{def}}{=} \begin{cases} \left(1 - \dfrac{\sqrt{\omega}}{\sqrt{1-\omega}} \mathbb{E}\big[\mathsf{Z}\varphi(\mathsf{X}; \mathsf{A}|\omega)\big]\right)^{-1} \\ \quad \times \left(\varphi(x; a|\omega) - \dfrac{\mathbb{E}[\mathsf{Z}\varphi(\mathsf{X}; \mathsf{A}|\omega)]}{\sqrt{1-\omega}} x\right) & \omega < 1, \\ x & \omega = 1. \end{cases} \tag{11}$$

Finally, we will impose the following regularity condition in our analysis.

ASSUMPTION 2. For any $\omega \in [0,1]$, the MMSE estimator $\varphi(\cdot|\omega) : \mathbb{R} \times \mathbb{R}^k \to \mathbb{R}$ for the scalar Gaussian channel $(\mathsf{X}_\star; \mathsf{A}) \sim \pi, \quad \mathsf{X}|\mathsf{A}, \mathsf{X}_\star \sim \mathcal{N}(\sqrt{\omega} \cdot \mathsf{X}_\star, 1-\omega)$ is continuously differentiable and Lipschitz.[2]

3. Main results.

We now present the main results obtained in this paper.

3.1. *Orthogonal approximate message passing algorithms.*

We introduce a class of iterative methods to estimate the signal $\boldsymbol{x}_\star$ using observation $\boldsymbol{Y}$ generated from the spiked matrix model (1), called *Orthogonal Approximate Message Passing* (OAMP) algorithms.

DEFINITION 3 (OAMP algorithms). An OAMP algorithm generates iterates $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ in $\mathbb{R}^N$ according to the update rule:

$$\boldsymbol{x}_t = \Psi_t(\boldsymbol{Y}) \cdot f_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{a}) \quad \forall t \in \mathbb{N}, \tag{12a}$$

where $\boldsymbol{Y} \in \mathbb{R}^{N \times N}$ is the observed noisy matrix and $\boldsymbol{a} \in \mathbb{R}^{N \times k}$ denotes the side information available for estimating $\boldsymbol{x}_\star$. The estimate of $\boldsymbol{x}_\star$ at iteration t is obtained by applying a post-processing function ψ_t to the iterates $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t$ and the side-information $\boldsymbol{a}$:

$$\widehat{\boldsymbol{x}}_t = \psi_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t; \boldsymbol{a}). \tag{12b}$$

In the above equations, for each $t \in \mathbb{N}$, the matrix denoiser $\Psi_t : \mathbb{R} \to \mathbb{R}$ acts on the matrix $\boldsymbol{Y}$ in the following way: if $\boldsymbol{Y} = \boldsymbol{O}\,\mathrm{diag}(\lambda_1, \ldots, \lambda_N)\boldsymbol{O}^\top$ is the eigendecomposition of $\boldsymbol{Y}$, then $\Psi_t(\boldsymbol{Y}) = \boldsymbol{O}\,\mathrm{diag}(\Psi_t(\lambda_1), \ldots, \Psi_t(\lambda_N))\boldsymbol{O}^\top$. Likewise, the iterate denoisers $f_t : \mathbb{R}^{t-1} \times \mathbb{R}^k \to \mathbb{R}$ and the post-processing function $\psi_t : \mathbb{R}^t \times \mathbb{R}^k \to \mathbb{R}$ act entrywise on the N components of its vector inputs. We also require that the matrix denoisers $(\Psi_t)_{t\in\mathbb{N}}$ and iterate denoisers $(f_t)_{t\in\mathbb{N}}$ satisfy certain constraints, which we will introduce in equations (14) and (15) below, after defining the notion of state evolution random variables associated with an OAMP algorithm.

State evolution random variables. Each OAMP algorithm is associated with a collection of state evolution random variables $(\mathsf{X}_\star, (\mathsf{X}_t)_{t\in\mathbb{N}}; \mathsf{A})$, which describes the joint asymptotic behavior of the signal $\boldsymbol{x}_\star$, the iterates $(\boldsymbol{x}_t)_{t\in\mathbb{N}}$, and the side information $\boldsymbol{a}$. The distribution of these random variables is given by

$$(\mathsf{X}_\star, \mathsf{A}) \sim \pi, \quad \mathsf{X}_t = \beta_t \mathsf{X}_\star + \mathsf{Z}_t \quad \forall t \in \mathbb{N}, \tag{13a}$$

where $(\beta_t)_{t\in\mathbb{N}}$ is defined via the recursion

$$\beta_t \overset{\text{def}}{=} \mathbb{E}\big[\mathsf{X}_\star f_t(\mathsf{X}_1, \ldots, \mathsf{X}_{t-1}; \mathsf{A})\big] \cdot \mathbb{E}_{\Lambda_\nu \sim \nu}\big[\Psi_t(\Lambda_\nu)\big], \tag{13b}$$

and $(\mathsf{Z}_t)_{t\in\mathbb{N}}$ are zero mean jointly Gaussian random variables, independent of $(\mathsf{X}_\star; \mathsf{A})$, whose covariance matrix is given by the recursion

$$\begin{aligned}\mathbb{E}[\mathsf{Z}_s\mathsf{Z}_t] &= \mathbb{E}[\mathsf{X}_\star\mathsf{F}_s]\mathbb{E}[\mathsf{X}_\star\mathsf{F}_t] \cdot \mathrm{Cov}_{\Lambda_\nu\sim\nu}\big[\Psi_s(\Lambda_\nu), \Psi_t(\Lambda_\nu)\big] \\ &\quad + \big(\mathbb{E}[\mathsf{F}_s\mathsf{F}_t] - \mathbb{E}[\mathsf{X}_\star\mathsf{F}_s]\mathbb{E}[\mathsf{X}_\star\mathsf{F}_t]\big) \cdot \mathrm{Cov}_{\Lambda\sim\mu}\big[\Psi_s(\Lambda), \Psi_t(\Lambda)\big].\end{aligned} \tag{13c}$$

In the above display, $\mathsf{F}_s \overset{\text{def}}{=} f_s(\mathsf{X}_1, \ldots, \mathsf{X}_{s-1}; \mathsf{A})$ and $\mathsf{F}_t \overset{\text{def}}{=} f_t(\mathsf{X}_1, \ldots, \mathsf{X}_{t-1}; \mathsf{A})$.

[2] In the absence of any side information, a sufficient condition for Assumption 2 to hold is that the signal random variable $\mathsf{X}_\star$ is compactly supported (see [51], Remark 2.3, and [29], Lemma 3.8).

Requirements on matrix denoisers. The matrix denoisers $(\Psi_t)_{t\in\mathbb{N}}$ used in the OAMP algorithm should be continuous functions, which do not change with N, and are required to satisfy the *trace-free constraint*

$$\mathbb{E}\big[\Psi_t(\Lambda)\big] = 0, \quad \Lambda \sim \mu. \tag{14}$$

Requirements on iterate denoisers and post-processing functions. For each $t \in \mathbb{N}$, the iterate denoiser $f_t : \mathbb{R}^{t-1} \times \mathbb{R}^k \to \mathbb{R}$ and the post-processing function $\psi_t : \mathbb{R}^t \times \mathbb{R}^k \to \mathbb{R}$ used in the OAMP algorithm should be continuously differentiable and Lipschitz functions, which do not change with N. In addition, the iterate denoisers $(f_t)_{t\in\mathbb{N}}$ are required to satisfy the *divergence-free* constraint

$$\mathbb{E}\big[\partial_s f_t(\mathsf{X}_1, \ldots, \mathsf{X}_{t-1}; \mathsf{A})\big] = 0 \quad \forall s \in [t-1], t \in \mathbb{N}, \tag{15}$$

where $\partial_s f_t$ denotes the partial derivative of $f_t(x_1, \ldots, x_s, \ldots, x_{t-1}; a)$ with respect to x_s.

Our first main result is the following theorem, which provides a characterization of the dynamics of an OAMP algorithm, in the high-dimensional limit, in terms of the associated state evolution random variables.

THEOREM 1 (State evolution of OAMP). *Consider a general OAMP algorithm of the form* (12) *that satisfies the requirements stated in Definition* 3, *and let* $\{\mathsf{X}_\star, (\mathsf{X}_t)_{t\in\mathbb{N}}, \mathsf{A}\}$ *be the associated state evolution random variables. Then for any* $t \in \mathbb{N}$,

$$(\boldsymbol{x}_\star, \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_t; \boldsymbol{a}) \xrightarrow{W_2} (\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A}). \tag{16}$$

We present a proof sketch of this result highlighting the key ideas in Section 4.1, and defer the complete proof to the Supplementary Material [22], Appendix B.

REMARK 2 (Designing divergence-free denoisers). One can construct iterate denoisers that satisfy the divergence-free constraint (15) by designing them adaptively: once divergence-free iterate denoisers $f_1, \ldots, f_t$ have been specified, any candidate denoiser $\widetilde{f}_{t+1} : \mathbb{R}^t \times \mathbb{R}^k \to \mathbb{R}$ can be corrected to obtain a divergence-free denoiser,

$$f_{t+1}(x; a) \stackrel{\text{def}}{=} \widetilde{f}_{t+1}(x; a) - \sum_{i=1}^{t} \mathbb{E}\big[\partial_i \widetilde{f}_{t+1}(\mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})\big] \cdot x_i \quad \forall x \in \mathbb{R}^t, a \in \mathbb{R}^k,$$

which satisfies the divergence-free constraint (15) by construction. Here, $(\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})$ are the state evolution random variables associated with the first t iterations of the algorithm.

REMARK 3 (Connections to AMP algorithms for compressed sensing). The OAMP algorithm introduced in Definition 3 can be viewed as a natural analog of the orthogonal AMP (OAMP) [48] or vector AMP (VAMP) algorithm [64] developed for compressed sensing (or regularized linear regression). The key feature of these algorithms, which is shared by the algorithm in Definition 1, is the use of trace-free matrix denoisers (cf. (14)) and divergence-free iterate denoisers (cf. (15)). These features significantly simplify the algorithm's state evolution. However, unlike the OAMP algorithm introduced in this work, the compressed sensing algorithms compute a matrix-vector multiplication involving a rotationally invariant matrix at each iteration. In contrast, only the observed matrix $\boldsymbol{Y}$ is available in the spiked matrix model, not the rotationally invariant noise matrix $\boldsymbol{W}$. Although $\boldsymbol{Y}$ is a rank-1 perturbation of $\boldsymbol{W}$, $\Psi_t(\boldsymbol{Y})$, the matrix used by the OAMP algorithm at iteration t cannot be expressed as a simple perturbation of $\boldsymbol{W}$ for a general matrix denoiser Ψ_t. This complicates the analysis of these algorithms. We refer the reader to Fan [27], Remark 3.3, for related discussions.

3.2. *The optimal OAMP algorithm.* Theorem 1 provides a characterization of the asymptotic mean squared error (MSE) of the estimator $\widehat{\boldsymbol{x}}_t$ computed by a general OAMP algorithm (12) in terms of the associated state evolution random variables

$$\operatorname*{plim}_{N\to\infty} \frac{\|\boldsymbol{x}_\star - \widehat{\boldsymbol{x}}_t\|^2}{N} = \mathbb{E}\big|\mathsf{X}_\star - \psi_t(\mathsf{X}_1, \dots, \mathsf{X}_t; \mathsf{A})\big|^2.$$

The limiting value of the MSE depends implicitly on the functions $(\Psi_t, f_t, \psi_t)_{t\in\mathbb{N}}$ used in the OAMP algorithm since these functions determine the joint distribution of $(\mathsf{X}_\star, (\mathsf{X}_t)_{t\in\mathbb{N}}; \mathsf{A})$. As our second contribution, we use the above characterization to derive the optimal choice for these functions, which minimizes the MSE. We call the resulting algorithm the *optimal OAMP algorithm*, and introduce it below.

The optimal OAMP algorithm. We first introduce the optimal OAMP algorithm in some simple corner cases, and then consider the typical case.

Corner cases. If $\mathsf{mmse}_\pi(0) = \mathbb{E}\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}] = 0$, then it is possible to reconstruct the signal perfectly from the side information alone. Hence, the optimal OAMP algorithm outputs the estimator

$$\widehat{\boldsymbol{x}}_t = \varphi(\boldsymbol{a}|0) \quad \forall t \in \mathbb{N}, \tag{17a}$$

where $\varphi(\cdot|0)$ is the MMSE estimator for the Gaussian channel $(\mathsf{X}_\star; \mathsf{A})$ (which operates at SNR $\omega = 0$). In this case, the optimal OAMP algorithm achieves zero asymptotic mean squared error

$$\operatorname*{plim}_{N\to\infty} \frac{\|\boldsymbol{x}_\star - \widehat{\boldsymbol{x}}_t\|^2}{N} \overset{\text{Thm. 1}}{=} \mathbb{E}\big|\mathsf{X}_\star - \varphi(\mathsf{A}|0)\big|^2 \overset{(10)}{=} \mathbb{E}\big|\mathsf{X}_\star - \mathbb{E}[\mathsf{X}_\star|\mathsf{A}]\big|^2 = \mathbb{E}\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}] = 0 \quad \forall t \in \mathbb{N}.$$

In the other extreme, if $\mathsf{mmse}_\pi(0) = \mathbb{E}\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}] = 1$ (recall $\mathbb{E}[\mathsf{X}_\star^2] = 1$ from Assumption 1), the optimal OAMP algorithm returns the trivial estimator

$$\widehat{\boldsymbol{x}}_t = \boldsymbol{0} \quad \forall t \in \mathbb{N}. \tag{17b}$$

In this case, the asymptotic MSE of the optimal OAMP algorithm is

$$\operatorname*{plim}_{N\to\infty} \frac{\|\boldsymbol{x}_\star - \widehat{\boldsymbol{x}}_t\|^2}{N} = \operatorname*{plim}_{N\to\infty} \frac{\|\boldsymbol{x}_\star\|^2}{N} = \mathbb{E}\mathsf{X}_\star^2 = 1 \quad \text{by Assumption 1.}$$

Typical case. In the typical situation when $\mathsf{mmse}_\pi(0) = \mathbb{E}\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}] \in (0, 1)$, the optimal OAMP algorithm generates a sequence of iterates $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots$ using the update rule:

$$\boldsymbol{x}_t = \frac{1}{\sqrt{\omega_t}}\left(1 + \frac{1}{\rho_t}\right) \cdot \Psi_\star(\boldsymbol{Y}; \rho_t) \cdot \bar{\varphi}(\boldsymbol{x}_{t-1}; \boldsymbol{a}|\omega_{t-1}). \tag{17c}$$

The estimator returned by the optimal OAMP algorithm at iteration t is

$$\widehat{\boldsymbol{x}}_t \overset{\text{def}}{=} \varphi(\boldsymbol{x}_t; \boldsymbol{a}|\omega_t). \tag{17d}$$

In the above equations:

- The matrix denoiser $\Psi_\star$ used by the optimal OAMP algorithm is given by

$$\Psi_\star(\lambda; \rho) = 1 - \left(\mathbb{E}\left[\frac{\phi(\Lambda)}{\phi(\Lambda) + \rho}\right]\right)^{-1} \cdot \frac{\phi(\lambda)}{\phi(\lambda) + \rho} \quad \forall \lambda \in \mathbb{R}, \rho \in (0, \infty), \tag{17e}$$

  where $\Lambda \sim \mu$ and the function $\phi : \mathbb{R} \to \mathbb{R}$ was introduced in (5).

- ω_t and ρ_t are computed using the recursion

$$\rho_t = \frac{1}{\mathsf{dmmse}_\pi(\omega_{t-1})} - 1, \quad \omega_t = 1 - \left(\mathbb{E}\left[\frac{\phi(\Lambda)}{\phi(\Lambda)+\rho_t}\right]\right)^{-1} \cdot \mathbb{E}\left[\frac{1}{\phi(\Lambda)+\rho_t}\right] \tag{17f}$$

  initialized with $\omega_0 \stackrel{\text{def}}{=} 0$.
- $\mathsf{dmmse}_\pi(\omega)$ denotes the DMMSE for a scalar Gaussian channel with SNR ω, and $\varphi(\cdot|\omega)$, $\bar{\varphi}(\cdot|\omega)$ denote the MMSE and DMMSE estimators (Definition 2).

REMARK 4. The additional scaling factor $1/\sqrt{\omega_t} \cdot (1 + 1/\rho_t)$ in (17c) is introduced to normalize the iterates so that $\|\boldsymbol{x}_t\|^2/N \xrightarrow{\mathbb{P}} 1$.

The following result characterizes the asymptotic MSE of the optimal OAMP algorithm in this case. Its proof can be found in the Supplementary Material [22], Appendix C.

PROPOSITION 1. *Assume that* $\mathsf{mmse}_\pi(0) = \mathbb{E}\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}] \in (0,1)$. *Let* $(\mathsf{X}_\star, (\mathsf{X}_t)_{t\in\mathbb{N}}; \mathsf{A})$ *denote the state evolution random variables associated with the optimal OAMP algorithm* (17). *Then*:

1. *For each* $t \in \mathbb{N}$, $\omega_t \in [0,1)$ *and* $\rho_t \in (0,\infty)$. *Moreover, the state evolution random variables* $(\mathsf{X}_\star, \mathsf{X}_t; \mathsf{A})$ *form a scalar Gaussian channel with SNR* ω_t *and, the asymptotic MSE of the estimator* $\widehat{\boldsymbol{x}}_t$ *returned by the optimal OAMP algorithm in* (17) *is given by*

$$\operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_\star\|^2}{N} = \mathsf{mmse}_\pi(\omega_t).$$

2. *The sequences* $(\omega_t)_{t\in\mathbb{N}}$ *and* $(\rho_t)_{t\in\mathbb{N}}$ *from* (17f) *are nondecreasing and converge to limit points* $\omega_* \in [0,1)$ *and* $\rho_* \in (0,\infty)$ *as* $t \to \infty$. *The limit points* (ω_*, ρ_*) *solve the following fixed-point equation in* (ω, ρ):

$$\omega = 1 - \left(\mathbb{E}\left[\frac{\phi(\Lambda)}{\phi(\Lambda)+\rho}\right]\right)^{-1} \cdot \mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda)}\right], \quad \rho = \frac{1}{\mathsf{dmmse}_\pi(\omega)} - 1, \Lambda \sim \mu. \tag{18}$$

Consequently, as $t \to \infty$, *the asymptotic MSE of* $\widehat{\boldsymbol{x}}_t$ *converges to*

$$\lim_{t\to\infty} \operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_\star\|^2}{N} = \mathsf{mmse}_\pi(\omega_*).$$

REMARK 5. In the absence of side information, if the signal is drawn from a zero-mean prior, $\mathsf{mmse}_\pi(0) = 1$ and the optimal OAMP algorithm in (17) returns the trivial estimator $\widehat{\boldsymbol{x}}_t = \boldsymbol{0}$. Our optimality result (introduced as Theorem 2 in the following section) shows that no iterative algorithm that runs for a constant (N-independent) number of iterations can achieve a better performance. An interesting direction for future work is to analyze OAMP algorithms with spectral initialization [50, 70, 73] or randomly initialized iterative algorithms that run for $T \gtrsim \ln(N)$ iterations [44, 45, 65] in this situation.

3.3. *An optimality result.* Next, to discuss the optimality properties of the OAMP algorithm introduced in (17), we introduce the following broad class of iterative algorithms.

DEFINITION 4 (Iterative algorithms). An iterative algorithm is any procedure, which generates iterates $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots$ in $\mathbb{R}^N$ according to an update rule of the form

$$\boldsymbol{r}_t = \Psi_t(\boldsymbol{Y}) \cdot f_t(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{t-1}; \boldsymbol{a}) + g_t(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{t-1}; \boldsymbol{a}) \quad \forall t \in \mathbb{N}. \tag{19a}$$

At iteration t, an estimate $\widehat{\boldsymbol{r}}_t$ of $\boldsymbol{x}_\star$ is obtained by applying a post-processing function ψ_t to the iterates $\boldsymbol{r}_1, \dots, \boldsymbol{r}_t$ and the side-information $\boldsymbol{a}$:

$$\widehat{\boldsymbol{r}}_t = \psi_t(\boldsymbol{r}_1, \dots, \boldsymbol{r}_t; \boldsymbol{a}). \tag{19b}$$

For each $t \in \mathbb{N}$, the matrix denoiser $\Psi_t : \mathbb{R} \to \mathbb{R}$ is required to be continuous and the functions $f_t : \mathbb{R}^{t-1} \times \mathbb{R}^k \to \mathbb{R}$, $g_t : \mathbb{R}^{t-1} \times \mathbb{R}^k \to \mathbb{R}$, and $\psi_t : \mathbb{R}^t \times \mathbb{R}^k \to \mathbb{R}$ are required to be continuously differentiable and Lipschitz. Furthermore, $(\Psi_t)_{t\in\mathbb{N}}$, $(f_t)_{t\in\mathbb{N}}$, $(g_t)_{t\in\mathbb{N}}$ and $(\psi_t)_{t\in\mathbb{N}}$ do not change with the dimension N.

The definition above extends the notion of general first-order methods (GFOMs) introduced by Celentano, Montanari and Wu [14] (for the i.i.d. Gaussian noise model) to include a matrix denoising step at each iteration. The class of iterative algorithms defined above includes many commonly used estimators, such as those computed using gradient descent or power method, and their proximal and projected generalizations. The following theorem (proved in the Supplementary Material [22], Appendix D) shows that the optimal OAMP algorithm achieves the minimum possible estimation error among all algorithms in this class under a given iteration budget.

THEOREM 2. *Let $\widehat{\boldsymbol{r}}_t$ be the estimator returned by any iterative algorithm of the form* (19) *after $t \in \mathbb{N}$ iterations. Let $\widehat{\boldsymbol{x}}_t$ be the estimator returned by the optimal OAMP algorithm in* (17) *after t iterations. Then*

$$\operatorname*{plim\,inf}_{N\to\infty} \frac{\|\widehat{\boldsymbol{r}}_t - \boldsymbol{x}_\star\|^2}{N} \geq \operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_\star\|^2}{N}.$$

3.4. *Information-theoretic vs. computational limits.* A natural question is whether the optimal OAMP algorithm introduced in (17) achieves the smallest asymptotic estimation error among *all estimators*, not just estimators computable using efficient iterative algorithms. To address this question, we begin by recalling the conjecture of Barbier et al. [4] regarding the fundamental information-theoretic limits of this problem.

Replica conjecture for the Bayes risk. Under the assumption that the entries of $(\boldsymbol{x}_\star, \boldsymbol{a})$ are drawn i.i.d. from a prior π, the information-theoretically optimal estimator is the Bayes estimator $\mathbb{E}[\boldsymbol{x}_\star|\boldsymbol{Y}, \boldsymbol{a}]$. Barbier et al. [4] provide a conjecture for the asymptotic MSE of this estimator (also called the Bayes risk) assuming the noise matrix drawn from the *trace ensemble* [4, 58] with density

$$p(\boldsymbol{W}) \propto \exp\left(-\frac{N}{2}\sum_{i=1}^{N} V\big(\lambda_i(\boldsymbol{W})\big)\right) \quad \text{where } \lambda_{1:N}(\boldsymbol{W}) \text{ denote the eigenvalues of } \boldsymbol{W}, \tag{20}$$

and the *potential function* $V : \mathbb{R} \to \mathbb{R}$ is a functional parameter for the noise model. The authors develop a recipe to derive conjectured formulas for the asymptotic Bayes risk based on the nonrigorous replica method for polynomial potentials V, providing explicit formulas for quartic and sestic polynomials. As the degree of V increases, the complexity of the derivation and resulting formulas also increases, with no general formula available. We state the replica conjecture for the Bayes risk when the potential $V : \mathbb{R} \to \mathbb{R}$ is a quartic polynomial

$$V(\lambda) = \frac{\gamma\lambda^2}{2} + \frac{\kappa\lambda^4}{4} \quad \forall \lambda \in \mathbb{R} \text{ where } \kappa = \kappa(\gamma) = \frac{8 - 9\gamma + \sqrt{64 - 144\gamma + 108\gamma^2 - 27\gamma^3}}{27}, \tag{21}$$

and $\gamma \in [0, 1]$ is a parameter for the noise model. The choice of κ in (21) ensures that the limiting spectral measure (μ) of $\boldsymbol{W}$ drawn from the trace ensemble has unit variance [4].

CONJECTURE 1 (Replica conjecture for Bayes risk [4]). *Suppose that the entries of* $(\boldsymbol{x}_\star, \boldsymbol{a})$ *are drawn i.i.d. from a prior* π *with* $\mathsf{mmse}_\pi(0) \in (0, 1)$ *and the noise matrix* $\boldsymbol{W}$ *is drawn from the trace ensemble* (20) *with the quartic potential function* V *from* (21). *Then*

$$\operatorname*{plim}_{N\to\infty} \frac{\|\mathbb{E}[\boldsymbol{x}_\star|\boldsymbol{Y}, \boldsymbol{a}] - \boldsymbol{x}_\star\|^2}{N} = \mathsf{mmse}_\pi(\omega_{\mathrm{IT}}),$$

for some $\omega_{\mathrm{IT}} \in (0, 1)$ *and* $\rho_{\mathrm{IT}} \in (0, \infty)$, *which solve the following system of fixed-point equations*[3] (*in* ω, ρ):

$$m = 1 - \mathsf{mmse}_\pi(\omega), \qquad \mathbb{E}[\mathsf{H}] = 1 - m, \qquad \chi = \mathbb{E}[\Lambda\mathsf{Q}\mathsf{H}], \tag{22a}$$

$$\widehat{m} \equiv \frac{\omega}{1-\omega} = \kappa\theta^2\Big(\frac{m}{1-m}\mathbb{E}\big[\Lambda^2\mathsf{H}\big] + \frac{\chi}{1-m}\mathbb{E}[\Lambda\mathsf{H}] + \mathbb{E}\big[\Lambda^2\mathsf{Q}\mathsf{H}\big]\Big) + \gamma\theta^2 m, \tag{22b}$$

where $m \in \mathbb{R}$, $\chi \in \mathbb{R}$ *are two intermediate variables,* $\Lambda \sim \mu$, *and the random variables* $\mathsf{Q} = \mathsf{Q}(\Lambda, m, \chi)$, $\mathsf{H} = \mathsf{H}(\Lambda, \rho)$ *are defined by*

$$\mathsf{Q} \overset{\text{def}}{=} \kappa\theta^2 m\Lambda^2 + \kappa\theta^2\chi\Lambda - \frac{\kappa\theta^2}{1-m}\mathbb{E}\big[m\Lambda^2\mathsf{H} + \chi\Lambda\mathsf{H}\big] + \frac{m}{1-m}, \tag{22c}$$

$$\mathsf{H} \overset{\text{def}}{=} \big(\rho + \theta^2 a^2\big(\gamma + 2a^2\kappa\big)^2 + 1 - \theta\big(\gamma\Lambda - \theta\kappa\Lambda^2 + \kappa\Lambda^3\big)\big)^{-1}. \tag{22d}$$

In the above equations, (γ, κ) *are the parameters for the quartic potential function* (21), θ *is the SNR for the spiked matrix model* (1) *and* $a^2 \overset{\text{def}}{=} (\sqrt{\gamma^2 + 12\kappa} - \gamma)/(6\kappa)$.

On the other hand, from Proposition 1 that the asymptotic MSE of the estimator $\widehat{\boldsymbol{x}}_t$ returned by the optimal OAMP algorithm in (17) satisfies

$$\lim_{t\to\infty} \operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_\star\|^2}{N} = \mathsf{mmse}_\pi(\omega_*),$$

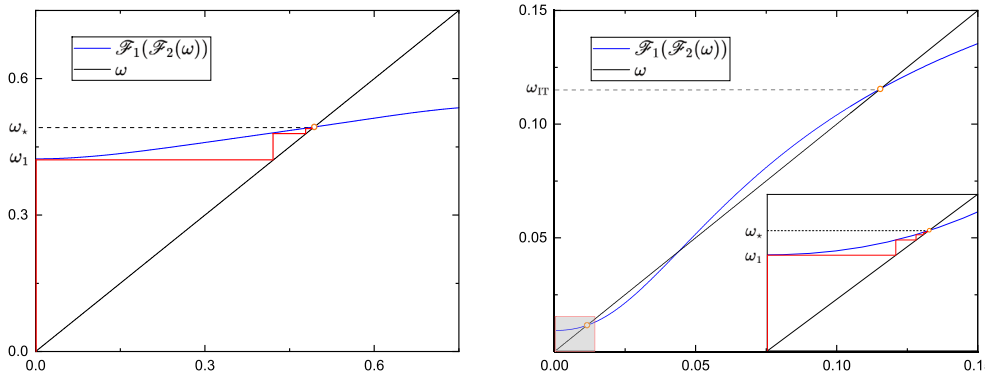
where (ω_*, ρ_*) denote the solution of the state evolution fixed-point equations (in $\omega \in (0, 1)$, $\rho \in (0, \infty)$)

$$\omega = \mathscr{F}_1(\rho),\ \rho = \mathscr{F}_2(\omega) \text{ with}$$
$$\mathscr{F}_1(\rho) \overset{\text{def}}{=} 1 - \frac{\mathbb{E}_{\Lambda\sim\mu}[\frac{1}{\phi(\Lambda)+\rho}]}{\mathbb{E}_{\Lambda\sim\mu}[\frac{\phi(\Lambda)}{\phi(\Lambda)+\rho}]}, \qquad \mathscr{F}_2(\omega) \overset{\text{def}}{=} \frac{1}{\mathsf{dmmse}_\pi(\omega)} - 1, \tag{23}$$

found by the recursion $\rho_t = \mathscr{F}_2(\omega_{t-1})$, $\omega_t = \mathscr{F}_1(\rho_t)$. The following proposition (proved in the Supplementary Material [22], Appendix E) shows that for the quartic potential, the *replica fixed-point equations* (22) and the *state evolution fixed-point equations* (23) are equivalent.

PROPOSITION 2. *Assume that the prior* π *satisfies* $\mathsf{mmse}_\pi(0) \in (0, 1)$. *Any solution* (ω, ρ) *to the state evolution fixed-point equations* (23) *with* $\omega \in (0, 1)$, $\rho \in (0, \infty)$ *is also a solution to the replica fixed-point equations* (22), *and vice versa.*

[3] When the fixed-point equations in (22) have multiple solutions, the correct fixed point is the one that minimizes a certain free energy function calculated in [4].



(a) Three-points prior with $\epsilon_1 = \frac{1}{3}$ and $\epsilon_2 = \frac{1}{5}$. (b) Three-points prior with $\epsilon_1 = \frac{1}{6}$ and $\epsilon_2 = \frac{1}{10}$.

FIG. 1. *Plot of the fixed-point equation* $\omega = \mathscr{F}_1(\mathscr{F}_2(\omega))$ *for SNR* $\theta = 0.46$, *quartic noise model* (21) *with* $\gamma = 0$ *and a three-point signal prior* $\mathsf{X}_\star \sim \frac{\epsilon_1^2}{2}\delta_{\frac{1}{\epsilon_1}} + \frac{\epsilon_2^2}{2}\delta_{\frac{1}{\epsilon_2}} + (1 - \frac{\epsilon_1^2}{2} - \frac{\epsilon_2^2}{2})\delta_0$.

We remark that an analogous result holds for the sestic (degree 6) potential. Since the replica equations for the sestic ensemble in Barbier et al. [4] are even more involved, we do not provide the details here. We anticipate that the fixed-point equations in (23) are a unified and concise reformulation of the replica fixed-point equations and characterize the asymptotic Bayes risk for general potential functions V. This reformulation of the replica conjecture for the asymptotic Bayes risk may be more amenable to rigorous proof than the significantly more complicated replica formulas derived using the approach of Barbier et al. [4].

Information-theoretic optimality and suboptimality of OAMP. We can simplify the fixed-point equations in (23) (which are equivalent to the replica fixed-point equations (22)) by eliminating ρ to obtain a single equation $\omega = \mathscr{F}_1(\mathscr{F}_2(\omega))$. In some cases, this equation has a unique solution, implying the optimal OAMP algorithm matches the conjectured asymptotic MSE of the Bayes estimator and is expected to be information-theoretically optimal (see Figure 1(a)). In other cases, the equation may have multiple solutions, as illustrated in Figure 1(b), where it has two stable fixed points. Here, the optimal OAMP algorithm converges to the inferior fixed-point ω_*, while the Bayes risk corresponds to the superior fixed-point ω_{IT}. Since $\omega_* < \omega_{\mathrm{IT}}$, the optimal OAMP or any iterative algorithm (in the form of (19a)) with a constant (N-independent) number of iterations fails to achieve the Bayes-optimal MSE. Such scenarios also occur in i.i.d. Gaussian noise models [14, 52, 53], and it is conjectured that no polynomial time algorithm can achieve the information-theoretically optimal MSE in these cases.

4. Proof ideas.

We now present some important ideas used to obtain our main results.

4.1. *Heuristic derivation of state evolution* (*Theorem* 1). We begin by an intuitive derivation of the state evolution result (Theorem 1) for OAMP algorithms, highlighting the key ideas involved in the proof. A formal proof of Theorem 1 is presented in the Supplementary Material [22], Appendix B. Consider a general OAMP algorithm (Definition 3)

$$\boldsymbol{x}_t = \Psi_t(\boldsymbol{Y}) \cdot f_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{a}) \quad \forall t \in \mathbb{N}, \tag{24}$$

where $(\Psi_t)_{t\in\mathbb{N}}$ are trace-free (cf. (14)) and $(f_t)_{t\in\mathbb{N}}$ are divergence-free (cf. (15)). For ease of exposition, we present the key ideas assuming that the matrix denoisers $(\Psi_t)_{t\in\mathbb{N}}$ *are polynomials* (the general case follows by a polynomial approximation argument). Let us decompose

$f_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{a})$ into a component along the direction of $\boldsymbol{x}_\star$ and a component perpendicular to it:

$$f_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{a}) = \alpha_t \boldsymbol{x}_\star + \boldsymbol{f}_t^\perp$$
$$\text{where } \alpha_t \stackrel{\text{def}}{=} \frac{\langle \boldsymbol{x}_\star, f_t(\boldsymbol{x}_{<t}; \boldsymbol{a})\rangle}{\|\boldsymbol{x}_\star\|^2}, \boldsymbol{f}_t^\perp \stackrel{\text{def}}{=} f_t(\boldsymbol{x}_{<t}; \boldsymbol{a}) - \alpha_t \boldsymbol{x}_\star.$$

Using this decomposition, we can write the new iterate $\boldsymbol{x}_t$ in (24) as

$$\boldsymbol{x}_t = \alpha_t \cdot \Psi_t(\boldsymbol{Y})\boldsymbol{x}_\star + \Psi_t(\boldsymbol{Y})\boldsymbol{f}_t^\perp. \tag{25}$$

As pointed out in Remark 3, the main difficulty in analyzing the algorithm above is that the matrix $\Psi_t(\boldsymbol{Y})$ is not rotationally invariant. To address this, we express the update equation (25) in terms of the rotationally invariant noise matrix $\boldsymbol{W}$ by expanding the polynomial $\Psi_t(\boldsymbol{Y})$ in terms of $\boldsymbol{W}$ using the fact that $\boldsymbol{Y} = (\theta/N) \cdot \boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T} + \boldsymbol{W}$. Although this expansion can initially seem complicated, the key insight that makes it tractable is to show that the following approximations hold:

$$\Psi_t(\boldsymbol{Y})\boldsymbol{x}_\star \overset{N \to \infty}{\simeq} \widetilde{\Psi}_t(\boldsymbol{W})\boldsymbol{x}_\star, \qquad \Psi_t(\boldsymbol{Y})\boldsymbol{f}_t^\perp \overset{N \to \infty}{\simeq} \Psi_t(\boldsymbol{W})\boldsymbol{f}_t^\perp, \tag{26}$$

where $\widetilde{\Psi}_t : \mathbb{R} \to \mathbb{R}$ is polynomial obtained by appropriately transforming Ψ_t; we refer the reader to the Supplementary Material [22], Appendix B, for additional details regarding these approximations. The transformed matrix denoiser $\widetilde{\Psi}_t$ is determined by Ψ_t via a complicated recursion and is not necessarily trace-free. Substituting (26) into (25), we get

$$\begin{aligned}
\boldsymbol{x}_t &\overset{N \to \infty}{\simeq} \alpha_t \cdot \widetilde{\Psi}_t(\boldsymbol{W})\boldsymbol{x}_\star + \Psi_t(\boldsymbol{W})\boldsymbol{f}_t^\perp \\
&= \alpha_t \cdot \frac{\operatorname{Tr}[\widetilde{\Psi}_t(\boldsymbol{W})]}{N} \cdot \boldsymbol{x}_\star + \alpha_t \cdot \left(\widetilde{\Psi}_t(\boldsymbol{W}) - \frac{\operatorname{Tr}[\widetilde{\Psi}_t(\boldsymbol{W})]}{N} \cdot \boldsymbol{I}_N\right)\boldsymbol{x}_\star + \Psi_t(\boldsymbol{W})\boldsymbol{f}_t^\perp \qquad \text{(27a)} \\
&\stackrel{\text{def}}{=} \underbrace{\alpha_t \cdot \frac{\operatorname{Tr}[\widetilde{\Psi}_t(\boldsymbol{W})]}{N} \cdot \boldsymbol{x}_\star}_{\text{Signal Component}} + \underbrace{\boldsymbol{z}_t}_{\text{Eff. Noise}},
\end{aligned}$$

where the effective noise $\boldsymbol{z}_t$ is defined as

$$\begin{aligned}
\boldsymbol{z}_t &\stackrel{\text{def}}{=} \alpha_t \cdot \widehat{\Psi}_t(\boldsymbol{W})\boldsymbol{x}_\star + \Psi_t(\boldsymbol{W})\boldsymbol{f}_t^\perp \\
&\text{where } \widehat{\Psi}_t(\boldsymbol{W}) \stackrel{\text{def}}{=} \widetilde{\Psi}_t(\boldsymbol{W}) - \frac{\operatorname{Tr}[\widetilde{\Psi}_t(\boldsymbol{W})]}{N} \cdot \boldsymbol{I}_N.
\end{aligned} \tag{27b}$$

We analyze the signal and noise components separately.

Signal component. Notice that the signal component in (27a) involves the factor $\operatorname{Tr}[\widetilde{\Psi}_t(\boldsymbol{W})]/N$. Computing the limiting value of this factor is challenging because the transformed matrix denoiser $\widetilde{\Psi}_t$ does not have a convenient formula and is determined by Ψ_t via a complicated recursion (see the Supplementary Material [22], Appendix B, for further details). The key idea, which yields a concise formula for this factor, is to observe

$$\begin{aligned}
\frac{\operatorname{Tr}[\widetilde{\Psi}_t(\boldsymbol{W})]}{N} &\overset{\text{(a)}}{\approx} \frac{\boldsymbol{x}_\star^\mathsf{T}\widetilde{\Psi}_t(\boldsymbol{W})\boldsymbol{x}_\star}{N} \overset{\text{(b)}}{\approx} \frac{\boldsymbol{x}_\star^\mathsf{T}\Psi_t(\boldsymbol{Y})\boldsymbol{x}_\star}{N} \\
&\overset{\text{(c)}}{=} \int_{\mathbb{R}} \Psi_t(\lambda)\nu_N(\mathrm{d}\lambda) \overset{\text{(d)}}{\approx} \mathbb{E}_{\Lambda_\nu \sim \nu}\big[\Psi_t(\Lambda_\nu)\big],
\end{aligned} \tag{28}$$

where the approximation in (a) follows from standard concentration results for quadratic forms of rotationally invariant matrices (see Fact A.1 in the Supplementary Material [22],

Appendix A.1.2), (b) follows from the defining property of $\widetilde{\Psi}_t(\boldsymbol{W})$ in (26), (c) follows by recalling the definition of the spectral measure in the signal direction (ν_N) from (4) and (d) follows from the weak convergence of the compactly supported measure ν_N to ν (Lemma 1, item (1)).

Noise component. The update rule (27b) for the effective noise $\boldsymbol{z}_t$ is written in a form for which existing state evolution results apply [23, 24, 27, 48, 64, 69, 71]. Indeed, $\widehat{\Psi}_t(\boldsymbol{W})$ and $\Psi_t(\boldsymbol{W})$ are trace-free functions of the rotationally-invariant matrix $\boldsymbol{W}$; and $\boldsymbol{f}_t^{\perp}$ is a divergence-free function of $\{\boldsymbol{x}_i\}_{i<t}$. By appealing to existing results on the dynamics of AMP algorithms driven by rotationally invariant matrices, we show that the effective noise component $\boldsymbol{z}_t$ converges to a centered Gaussian random variable Z_t. Moreover, the covariance of Z_t and Z_s can be heuristically calculated based on a convenient property of OAMP algorithms. Specifically, for $G_1(\boldsymbol{W}), G_2(\boldsymbol{W}) \in \{\widehat{\Psi}_t(\boldsymbol{W}), \Psi_t(\boldsymbol{W}), \Psi_s(\boldsymbol{W})\}$ and $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \{\boldsymbol{x}_\star, \boldsymbol{f}_s^{\perp}, \boldsymbol{f}_t^{\perp}\}$, we have

$$\frac{\boldsymbol{v}_1^{\mathsf{T}} G_1(\boldsymbol{W}) G_2(\boldsymbol{W}) \boldsymbol{v}_2}{N} \approx \frac{\mathrm{Tr}[G_1(\boldsymbol{W}) G_2(\boldsymbol{W})]}{N} \cdot \frac{\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle}{N}. \tag{29}$$

The intuition for the above property is that the random vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$ behave as if they are independent of the noise matrix $\boldsymbol{W}$, and the divergence-free and trace-free requirements imposed on OAMP algorithms are crucial for the validity of the above approximation. Using the above property, we immediately have that

$$\begin{aligned} \mathbb{E}[\mathsf{Z}_s \mathsf{Z}_t] &\approx \frac{\langle z_s, z_t \rangle}{N} = \frac{\langle \alpha_t \widehat{\Psi}_t(\boldsymbol{W}) \boldsymbol{x}_\star + \Psi_t(\boldsymbol{W}) \boldsymbol{f}_t^{\perp}, \alpha_s \widehat{\Psi}_s(\boldsymbol{W}) \boldsymbol{x}_\star + \Psi_s(\boldsymbol{W}) \boldsymbol{f}_s^{\perp} \rangle}{N} \\ &\approx \alpha_s \alpha_t \cdot \frac{\mathrm{Tr}[\widehat{\Psi}_t(\boldsymbol{W}) \widehat{\Psi}_s(\boldsymbol{W})]}{N} \cdot \frac{\|\boldsymbol{x}_\star\|^2}{N} \\ &\quad + \frac{\mathrm{Tr}[\Psi_t(\boldsymbol{W}) \Psi_s(\boldsymbol{W})]}{N} \cdot \frac{\langle \boldsymbol{f}_t^{\perp}, \boldsymbol{f}_s^{\perp} \rangle}{N}, \end{aligned} \tag{30}$$

where the cross-terms vanish due to the orthogonality of $\boldsymbol{x}_\star$ and $\boldsymbol{f}_t^{\perp}$ (and $\boldsymbol{f}_s^{\perp}$). The limiting value of $\mathrm{Tr}[\widehat{\Psi}_t(\boldsymbol{W}), \widehat{\Psi}_s(\boldsymbol{W})]/N$ can be expressed in terms of ν (the spectral measure in the signal direction) using the argument from (28). Replacing the various normalized inner products by their limiting values eventually leads to the claimed state evolution (13).

4.2. *Derivation of the optimal OAMP algorithm.* Next, we present an intuitive derivation of the optimal OAMP algorithm introduced in (17). While this derivation does not show that this algorithm attains the smallest estimation error among all iterative algorithms (as claimed in Theorem 2), it provides a simple and natural approach to derive the matrix denoisers and the iterate denoisers used by the algorithm. For simplicity, we consider the class of simple memory-free OAMP algorithms in which the iterate denoiser f_t only depends on the most recent iterate (cf. (12)):

$$\boldsymbol{x}_t = \Psi_t(\boldsymbol{Y}) \cdot f_t(\boldsymbol{x}_{t-1}; \boldsymbol{a}) \quad \forall t \in \mathbb{N}. \tag{31}$$

We will use a natural greedy heuristic to design the matrix and iterate denoisers. Specifically, we derive a good choice of Ψ_t and f_t, assuming we have already specified the iterate and matrix denoisers for the first $t-1$ iterations. For any candidate Ψ_t and f_t, the distribution of the state evolution random variable X_t associated with the $\boldsymbol{x}_t$ is given by (see Definition 3)

$$(\mathsf{X}_\star, \mathsf{A}) \sim \pi, \quad \mathsf{X}_t = \beta_t \mathsf{X}_\star + \mathsf{Z}_t, \tag{32a}$$

where Z_t is a centered Gaussian random variable independent of $(\mathsf{X}_\star, \mathsf{A})$, and

$$\beta_t = \mathbb{E}\big[\mathsf{X}_\star f_t(\mathsf{X}_{t-1}; \mathsf{A})\big] \cdot \mathbb{E}_{\Lambda_\nu \sim \nu}\big[\Psi_t(\Lambda_\nu)\big], \qquad \alpha_t = \mathbb{E}\big[\mathsf{X}_\star f_t(\mathsf{X}_{t-1}; \mathsf{A})\big],$$
$$\mathbb{E}\big[\mathsf{Z}_t^2\big] = \alpha_t^2 \cdot \mathrm{Var}_{\Lambda_\nu \sim \nu}\big[\Psi_t(\Lambda_\nu)\big] + \big(\mathbb{E}\big[f_t(\mathsf{X}_{t-1}; \mathsf{A})^2\big] - \alpha_t^2\big) \cdot \mathbb{E}_{\Lambda \sim \mu}\big[\Psi_t^2(\Lambda)\big].$$

A natural design principle is to choose f_t, Ψ_t to maximize the SNR ω_t of the Gaussian channel $(\mathsf{X}_\star, \mathsf{X}_t = \beta_t \mathsf{X}_\star + \mathsf{Z}_t; \mathsf{A})$, which is given by the squared cosine similarity between X_t, $\mathsf{X}_\star$,

$$\omega_t \overset{\text{def}}{=} \frac{(\mathbb{E}[\mathsf{X}_\star \mathsf{X}_t])^2}{\mathbb{E}[\mathsf{X}_\star^2] \cdot \mathbb{E}[\mathsf{X}_t^2]} = \frac{\beta_t^2}{\beta_t^2 + \mathbb{E}[\mathsf{Z}_t^2]} = \frac{(\mathbb{E}[\Psi_t(\Lambda_\nu)])^2}{\mathbb{E}[\Psi_t^2(\Lambda_\nu)] + \frac{\delta_t}{1-\delta_t} \cdot \mathbb{E}[\Psi_t^2(\Lambda)]}, \tag{33}$$

where $\delta_t \overset{\text{def}}{=} 1 - (\mathbb{E}[\mathsf{X}_\star f_t(\mathsf{X}_{t-1}; \mathsf{A})])^2 / \mathbb{E}[f_t(\mathsf{X}_{t-1}; \mathsf{A})^2]$. We consider the two maximization problems over Ψ_t and f_t one-by-one.

Optimal choice of Ψ_t. Choosing Ψ_t to maximize ω_t is equivalent to solving the following optimization problem:

$$\max_{\Psi} \frac{(\mathbb{E}[\Psi(\Lambda_\nu)])^2}{\mathbb{E}[\Psi^2(\Lambda_\nu)] + \rho^{-1} \cdot \mathbb{E}[\Psi^2(\Lambda)]} \quad \text{subject to } \mathbb{E}\big[\Psi(\Lambda)\big] = 0 \tag{34}$$

with the parameter $\rho \in (0, \infty)$ fixed at $\rho = \delta_t^{-1} - 1$ (notice δ_t does not depend on Ψ_t). We transform the optimization problem in (34) into a simple quadratic optimization problem,

$$\min_{\Psi} 1 - \frac{(\mathbb{E}[\Psi(\Lambda_\nu)])^2}{\mathbb{E}[\Psi^2(\Lambda_\nu)] + \rho^{-1} \cdot \mathbb{E}[\Psi^2(\Lambda)]} \quad \text{subject to } \mathbb{E}\big[\Psi(\Lambda)\big] = 0 \tag{35a}$$

$$\overset{\text{(a)}}{=} \min_{\Psi} \min_{c \in \mathbb{R}} \mathbb{E}\big[\big(1 - c\Psi(\Lambda_\nu)\big)^2\big] + \rho^{-1} \cdot \mathbb{E}\big[c^2 \Psi^2(\Lambda)\big] \quad \text{subject to } \mathbb{E}\big[c\Psi(\Lambda)\big] = 0 \tag{35b}$$

$$\overset{\text{(b)}}{=} \min_{\Psi} \mathbb{E}\big[\big(1 - \Psi(\Lambda_\nu)\big)^2\big] + \rho^{-1} \cdot \mathbb{E}\big[\Psi^2(\Lambda)\big] \quad \text{subject to } \mathbb{E}\big[\Psi(\Lambda)\big] = 0, \tag{35c}$$

where step (a) can be verified by solving the inner quadratic minimization problem over c explicitly, and step (b) is due to a change of variable $\Psi \mapsto c\Psi$; note that $c\Psi$ also satisfies the trace-free constraint. Note that any minimizer of (35c) is also a maximizer of (34) since the objective function of (34) is invariant to a rescaling of Ψ. To solve this optimization problem, we consider the Lebesgue decomposition $\nu = \nu_\parallel + \nu_\perp$, where $\nu_\perp$ is the singular part and $\nu_\parallel$ is the absolutely continuous part. Recall the definition of ϕ from (5) and let S denote the set

$$S \overset{\text{def}}{=} \{\lambda \in \mathbb{R} : \phi(\lambda) \neq 0\} \quad \text{where } \phi(\lambda) \overset{\text{def}}{=} \big(1 - \pi\theta \mathscr{H}_\mu(\lambda)\big)^2 + \pi^2 \theta^2 \mu^2(\lambda). \tag{36}$$

Notice that $\mu(S^c) = 0$. Lemma 1 (items (2–3)) shows that $\nu_\perp(S) = 0$ and the density of $\nu_\parallel$ with respect to the Lebesgue measure is given by $\mu(\cdot)/\phi(\cdot)$ where $\mu(\cdot)$ is the density of μ (recall Assumption 1). Hence, the Lagrangian of the minimization problem (35c) is given by

$$\begin{aligned}
&\mathbb{E}\big[\big|\Psi(\Lambda_\nu) - 1\big|^2\big] + \frac{1}{\rho} \cdot \mathbb{E}\big[\Psi^2(\Lambda)\big] - 2\gamma \mathbb{E}\big[\Psi(\Lambda)\big] \\
&\quad = \int_{\mathbb{R}} \big|\Psi(\lambda) - 1\big|^2 \nu(\mathrm{d}\lambda) + \int_{\mathbb{R}} \Big(\frac{\Psi^2(\lambda)}{\rho} - 2\gamma \Psi(\lambda)\Big) \mu(\lambda)\, \mathrm{d}\lambda \\
&\quad = \int_{S^c} \big(\Psi(\lambda) - 1\big)^2 \nu_\perp(\mathrm{d}\lambda) + \int_S \Big(\big(\Psi(\lambda) - 1\big)^2 \cdot \frac{1}{\phi(\lambda)} + \frac{\Psi^2(\lambda)}{\rho} - 2\gamma \Psi(\lambda)\Big) \mu(\mathrm{d}\lambda),
\end{aligned}$$

where $\gamma \in \mathbb{R}$ is the Lagrange multiplier. The Lagrangian is minimized by minimizing the integrand pointwise, leading to the following formula for the minimizer Ψ_γ:

$$\Psi_\gamma(\lambda) \stackrel{\text{def}}{=} \frac{\gamma + \frac{1}{\phi(\lambda)}}{\frac{1}{\phi(\lambda)} + \frac{1}{\rho}} = 1 + \frac{(\rho\gamma - 1)\cdot\phi(\lambda)}{\rho + \phi(\lambda)} \quad \forall \lambda \in S \text{ and } \Psi_\gamma(\lambda) = 1 \ \forall \lambda \in S^c.$$

The solution of the constrained minimization problem (35c) is obtained by choosing γ so that the constraint $\mathbb{E}_{\Lambda\sim\mu}[\Psi_\gamma(\Lambda)] = 0$ is fulfilled. Hence, the optimizer of (35c) and (34) is

$$\Psi(\lambda) = 1 - \left(\mathbb{E}_{\Lambda\sim\mu}\left[\frac{\phi(\Lambda)}{\phi(\Lambda)+\rho}\right]\right)^{-1} \cdot \frac{\phi(\lambda)}{\phi(\lambda)+\rho} \stackrel{\text{def}}{=} \Psi_\star(\lambda;\rho) \quad \forall \lambda \in \mathbb{R}. \tag{37}$$

This is precisely the matrix denoiser used by the optimal OAMP algorithm in (17).

Optimal choice of f_t. Notice from (33) that the iterate denoiser f_t, which maximizes the SNR ω_t can be derived by minimizing δ_t with respect to f_t, leading to the minimization problem

$$\min_f 1 - \frac{(\mathbb{E}[\mathsf{X}_\star f(\mathsf{X}_{t-1};\mathsf{A})])^2}{\mathbb{E}[f^2(\mathsf{X}_{t-1};\mathsf{A})]} \quad \text{subject to } \mathbb{E}\big[f'(\mathsf{X}_{t-1};\mathsf{A})\big] = 0. \tag{38}$$

In the above definition, $\mathsf{X}_{t-1} = \beta_{t-1}\mathsf{X}_\star + \sigma_{t-1}\mathsf{Z}$ is the state evolution random variable for $\boldsymbol{x}_{t-1}$, $\mathsf{Z} \sim \mathcal{N}(0,1)$ is independent of $(\mathsf{X}_\star, \mathsf{A}) \sim \pi$, and $\beta_{t-1} \in \mathbb{R}$ and $\sigma_{t-1} > 0$ are fixed parameters (determined by the matrix and iterate denoisers used in the first $t-1$ iterations). Repeating the arguments used in (35), we obtain

$$\min_f 1 - \frac{(\mathbb{E}[\mathsf{X}_\star f(\mathsf{X}_{t-1};\mathsf{A})])^2}{\mathbb{E}[f^2(\mathsf{X}_{t-1};\mathsf{A})]} \quad \text{subject to } \mathbb{E}\big[f'(\mathsf{X}_{t-1};\mathsf{A})\big] = 0 \tag{39a}$$

$$= \min_f \min_{c\in\mathbb{R}} \mathbb{E}\big[\big(\mathsf{X}_\star - cf(\mathsf{X}_{t-1};\mathsf{A})\big)^2\big] \quad \text{subject to } \mathbb{E}\big[cf'(\mathsf{X}_{t-1};\mathsf{A})\big] = 0 \tag{39b}$$

$$= \min_f \mathbb{E}\big[\big(\mathsf{X}_\star - f(\mathsf{X}_{t-1};\mathsf{A})\big)^2\big] \quad \text{subject to } \mathbb{E}\big[f'(\mathsf{X}_{t-1};\mathsf{A})\big] = 0. \tag{39c}$$

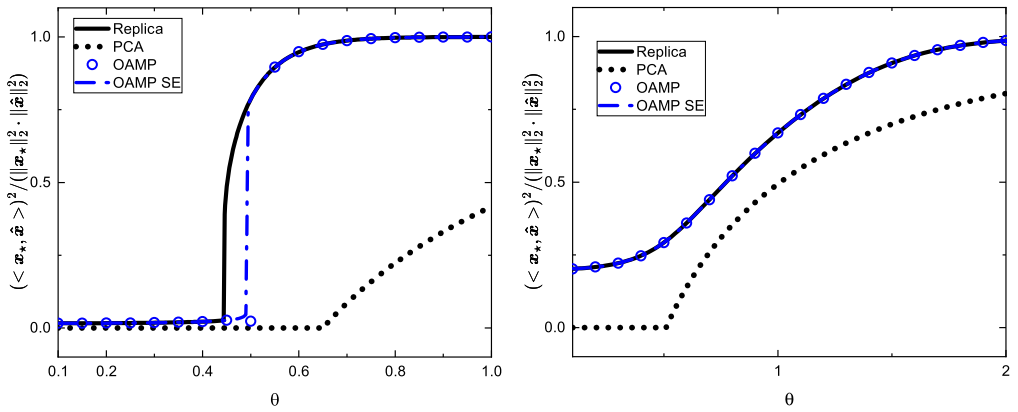
Recalling the definition of the DMMSE estimator from Definition 2 (see also Lemma A.2 in the Supplementary Material [22], Appendix A.2), the minimizer of (39c) is

$$f_t(x;a) = \bar{\varphi}\big((\beta_{t-1}^2 + \sigma_{t-1}^2)^{-\frac{1}{2}} \cdot x; a|\omega_{t-1}\big) \quad \text{where } \omega_{t-1} \stackrel{\text{def}}{=} \frac{\beta_{t-1}^2}{\beta_{t-1}^2 + \sigma_{t-1}^2}. \tag{40}$$

Optimal OAMP algorithm. Using the matrix and iterate denoisers derived above, we obtain the following memory-free OAMP algorithm (cf. (31)):

$$\boldsymbol{x}_t = c_t \cdot \Psi_\star(\boldsymbol{Y};\rho_t) \cdot \bar{\varphi}_t(\boldsymbol{x}_{t-1};\boldsymbol{a}|\omega_{t-1}), \tag{41}$$

where we defined $c_t \stackrel{\text{def}}{=} 1/\sqrt{\beta_t^2 + \sigma_t^2}$ and made a slight change of variable by absorbing the normalization factor (acting on the x input of $\bar{\varphi}$) in (40) into the definition of $\boldsymbol{x}_t$. This scaling parameter c_t ensures state evolution random variable X_t corresponding to $\boldsymbol{x}_t$ satisfies $\mathbb{E}[\mathsf{X}_t^2] = 1$. The resulting algorithm is precisely the optimal OAMP algorithm introduced in (17). The recursions for ω_t, ρ_t and the alternative formula $c_t = 1/\sqrt{w_t}\cdot(1 + 1/\rho_t)$ is derived by specializing the general state evolution equations in (32) to the optimal OAMP algorithm; see the proof of Proposition 1 in the Supplementary Material [22], Appendix C, for details.



(a) Statistical-computational gap exists. (b) No statistical-computational gap.

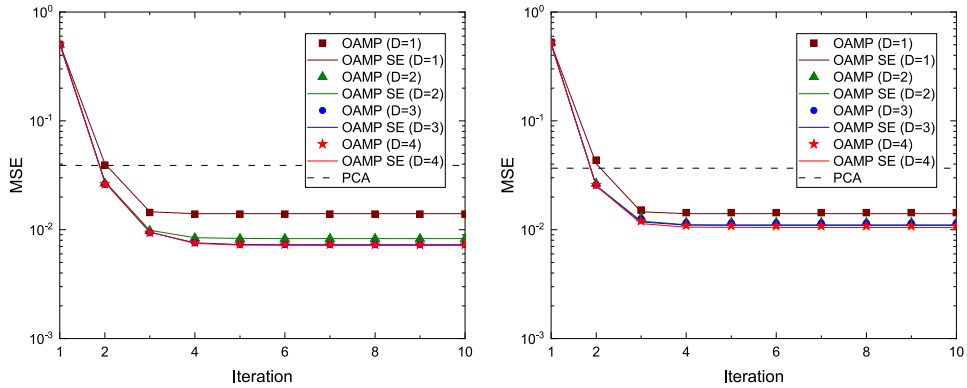
FIG. 2. *Performance (measured by the normalized overlap with signal) of PCA, the Bayes estimator and the optimal OAMP algorithm in (17). (a) The spectral μ corresponds to the quartic noise model in (21) with parameter $\gamma = 0$. The signal $\boldsymbol{x}_\star$ is sampled from an i.i.d. two-point prior $\epsilon^2\delta_{1/\epsilon} + (1-\epsilon^2)\delta_0$, where $\epsilon = 1/8$. (b) The spectral μ corresponds to the purely sestic noise model (see the Supplementary Material [22], (E.12)) and the the signal is sampled from the two-point prior with $\epsilon = 1/\sqrt{5}$.*

5. Numerical experiments. We conclude the paper with some numerical experiments. The Supplementary Material [22], Appendix F, provides additional numerical results.

Following [48, 64], we generate the noise matrix $\boldsymbol{W}$ in a structured manner, making it feasible to efficiently simulate the dynamics of iterative algorithms on very high-dimensional matrices ($N = 2 \times 10^6$). Specifically, we use the noise matrix $\boldsymbol{W} = \boldsymbol{O}\mathrm{diag}(\lambda_1, \dots, \lambda_N)\boldsymbol{O}^\mathsf{T}$ with $\boldsymbol{O} = \boldsymbol{S}_1\boldsymbol{F}\boldsymbol{S}_2\boldsymbol{F}^\mathsf{T}\boldsymbol{S}_3$, where $\boldsymbol{S}_1$, $\boldsymbol{S}_2$, $\boldsymbol{S}_3$ are independent random signed diagonal matrices, and $\boldsymbol{F}$ is the discrete cosine transform matrix. The eigenvalues $\{\lambda_i\}_{i\in[N]}$ are sampled independently from a spectral μ. We remark that known universality results [21, 23, 71] guarantee that using this structured noise matrix (instead of random rotationally invariant noise drawn from the quartic model) does not change the asymptotic dynamics of AMP algorithms. The signal $\boldsymbol{x}_\star$ is sampled from an i.i.d. two-point prior $\epsilon^2\delta_{1/\epsilon} + (1-\epsilon^2)\delta_0$, where $0 < \epsilon < 1$.

Figures 2a and 2b compare the performance of the optimal OAMP algorithm with the state evolution prediction from Theorem 1. In our simulations, the OAMP algorithm is initialized by $f_1(\boldsymbol{x}_0; \boldsymbol{a}) = \epsilon \cdot \mathbf{1}$ (which is the mean of the two-point prior) and no side information is provided. We also plot the performance of PCA and the Bayes-optimal estimator. In both Figure 2a and Figure 2b, OAMP achieves much better estimation accuracy than the PCA estimator and matches the Bayes-optimal performance at high SNR. Moreover, the performance of OAMP closely matches its state evolution. Notice that, for Figure 2a, there exists a performance gap between the conjectured Bayes-optimal performance (as predicted by the replica method) and the performance achieved by optimal OAMP (which has been shown to be optimal among a broad class of iterative algorithms). On the other hand, no such statistical-computational gap exists for Figure 2b where the signal is relatively "dense" (namely, the fraction of the nonzero components is relatively large). Similar phenomena have been observed in the context of spiked Wigner models [18, 51].

The theoretical prediction of the asymptotic performance of OAMP requires the noise matrix to be rotationally-invariant. The noise matrix used in the experiments for Figure 2 is structured but still synthetic. We now consider a more realistic model where the noise matrix is obtained from real datasets (while the signal is still randomly generated). Taking inspiration from [72], we generate the noise matrix from the covariance matrix of real datasets in bioinformatics, namely the 1000 Genomes Project (1000G) [15] and International HapMap Project (Hapmap3) [16]. Both datasets undergo the preprocessing steps, producing a 2054 $\times$



(a) Noise matrix derived from 1000G [16]. (b) Noise matrix derived from Hapmap3 [15].

FIG. 3. *Performance of a data-driven implementation of the optimal degree-D lifted OAMP algorithm (introduced in the Supplementary Material* [22], *Appendix D) on noise matrices derived from real datasets and signals randomly drawn from* $\mathsf{X}_\star \sim 0.1\mathcal{N}(0, 10) + 0.9\delta_0$. *The results are averaged over* 30 *random realizations.*

2054 symmetric matrix for 1000G and a 1397 × 1397 symmetric matrix for Hapmap3; see the Supplementary Material for details of the data processing procedure.

Running the optimal OAMP algorithm requires estimating the spectral measure μ and the associated optimal matrix denoiser (17e) from the observed data. While feasible, we found it more convenient to design a data-driven implementation for an approximation to the optimal OAMP algorithm, called the optimal degree-D lifted OAMP algorithm. This algorithm is introduced in the Supplementary Material [22], Appendix D, and plays an important role in the proof of our optimality result. At each iteration, the optimal degree-D lifted OAMP algorithm computes D matrix-vector multiplications. Moreover, as $D \to \infty$, this algorithm matches the performance of the optimal OAMP algorithm (see Proposition D.3 in the Supplementary Material [22], Appendix D). Figure 3 displays the MSE performance of this algorithm using the above described realistic noise matrices and a randomly generated signal. We initialize the algorithm by the signal mean as in Figure 2. The performance of the lifted OAMP improves as the degree D increases, saturating around $D \approx 4$, indicating that this algorithm can match the performance of the optimal OAMP algorithm in a data-driven way at the cost of 3 additional matrix multiplications per iteration. Figure 3 also displays the asymptotic MSE predicted by the state evolution result for the corresponding rotationally invariant noise model with a matching spectrum. The performance of the lifted OAMP algorithm closely aligns with the state evolution prediction, suggesting an underlying universality phenomenon.

6. Conclusion and future work. In this paper, we introduced an OAMP algorithm for spiked matrix models with rotationally-invariant noise and demonstrated that it obeys a simple state evolution characterization in the high-dimensional limit. We showed that the optimal OAMP algorithm achieves the best performance among a broad class of iterative algorithms.

Several promising directions for future research remain. As a first step, our analysis focused on a stylized symmetric rank-one model, and it would be valuable to extend these results to more practical settings, including asymmetric and multirank models. As discussed in Remark 1, in the absence of side information, if the signal is drawn from a zero-mean prior, no iterative algorithm that runs for a constant (N-independent) number of iterations can achieve nontrivial estimation error. An interesting direction for future work is to analyze OAMP algorithms with spectral initialization [50, 70, 73], or randomly initialized iterative algorithms that run for a diverging (N-dependent) number of iterations [44, 45, 65]. Finally, since the optimal OAMP algorithm relies on the knowledge of the signal prior and noise

spectrum, it would be interesting to develop and analyze a practical procedure to estimate these parameters from the data [72].

SUPPLEMENTARY MATERIAL

Supplement to "Optimality of approximate message passing for spiked matrix models with rotationally invariant noise" (DOI: 10.1214/25-AOS2575SUPP; .pdf). The supplement [22] contains the omitted proofs and additional numerical experiments.

REFERENCES

[1] ABBE, E. (2018). Community detection and stochastic block models: Recent developments. *J. Mach. Learn. Res.* **18** Paper No. 177, 86. MR3827065

[2] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. MR2165575 https://doi.org/10.1214/009117905000000233

[3] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. MR2279680 https://doi.org/10.1016/j.jmva.2005.08.003

[4] BARBIER, J., CAMILLI, F., MONDELLI, M. and SÁENZ, M. (2023). Fundamental limits in structured principal component analysis and how to reach them. *Proc. Natl. Acad. Sci. USA* **120** Paper No. e2302028120, 7. MR4637889 https://doi.org/10.1073/pnas.2302028120

[5] BARBIER, J., CAMILLI, F., XU, Y. and MONDELLI, M. (2025). Information limits and Thouless–Anderson–Palmer equations for spiked matrix models with structured noise. *Phys. Rev. Res.* **7** 013081.

[6] BARBIER, J., DIA, M., MACRIS, N., KRZAKALA, F., LESIEUR, T. and ZDEBOROVÁ, L. (2016). Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. *Adv. Neural Inf. Process. Syst.* **29**.

[7] BARBIER, J. and MACRIS, N. (2019). The adaptive interpolation method for proving replica formulas. Applications to the Curie–Weiss and Wigner spike models. *J. Phys. A: Math. Theor.* **52** 294002, 26. MR3983946 https://doi.org/10.1088/1751-8121/ab2735

[8] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785. MR2810285 https://doi.org/10.1109/TIT.2010.2094817

[9] BEHNE, J. K. and REEVES, G. (2022). Fundamental limits for rank-one matrix estimation with groupwise heteroskedasticity. In *International Conference on Artificial Intelligence and Statistics* 8650–8672. PMLR.

[10] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. MR2782201 https://doi.org/10.1016/j.aim.2011.02.007

[11] BOLTHAUSEN, E. (2014). An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Comm. Math. Phys.* **325** 333–366. MR3147441 https://doi.org/10.1007/s00220-013-1862-3

[12] BUN, J., ALLEZ, R., BOUCHAUD, J.-P. and POTTERS, M. (2016). Rotational invariant estimator for general noisy matrices. *IEEE Trans. Inf. Theory* **62** 7475–7490. MR3599095 https://doi.org/10.1109/TIT.2016.2616132

[13] CELENTANO, M., FAN, Z. and MEI, S. (2023). Local convexity of the TAP free energy and AMP convergence for $\mathbb{Z}_2$-synchronization. *Ann. Statist.* **51** 519–546. MR4600991 https://doi.org/10.1214/23-aos2257

[14] CELENTANO, M., MONTANARI, A. and WU, Y. (2020). The estimation error of general first order methods. In *Conference on Learning Theory* 1078–1141. PMLR.

[15] CONSORTIUM, G. P. (2015). A global reference for human genetic variation. *Nature* **526** 68.

[16] CONSORTIUM, I. H. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature* **467** 52.

[17] DESHPANDE, Y., ABBE, E. and MONTANARI, A. (2017). Asymptotic mutual information for the balanced binary stochastic block model. *Inf. Inference* **6** 125–170. MR3671474 https://doi.org/10.1093/imaiai/iaw017

[18] DESHPANDE, Y. and MONTANARI, A. (2014). Information-theoretically optimal sparse PCA. In 2014 *IEEE International Symposium on Information Theory* 2197–2201. IEEE Press, New York.

[19] DONOHO, D. and TANNER, J. (2009). Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **367** 4273–4293. With electronic supplementary materials available online. MR2546388 https://doi.org/10.1098/rsta.2009.0152

[20] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.

[21] DUDEJA, R. and BAKHSHIZADEH, M. (2022). Universality of linearized message passing for phase retrieval with structured sensing matrices. *IEEE Trans. Inf. Theory* **68** 7545–7574. MR4524656 https://doi.org/10.1109/tit.2022.3182018

[22] DUDEJA, R., LIU, S. and MA, J. (2026). Supplement to "Optimality of Approximate Message Passing for Spiked Matrix Models with Rotationally Invariant Noise." https://doi.org/10.1214/25-AOS2575SUPP

[23] DUDEJA, R., LU, Y. M. and SEN, S. (2023). Universality of approximate message passing with semirandom matrices. *Ann. Probab.* **51** 1616–1683. MR4642220 https://doi.org/10.1214/23-aop1628

[24] DUDEJA, R., SEN, S. and LU, Y. M. (2024). Spectral universality in regularized linear regression with nearly deterministic sensing matrices. *IEEE Trans. Inf. Theory* **70** 7923–7951. MR4818369 https://doi.org/10.1109/tit.2024.3458953

[25] EL ALAOUI, A. and KRZAKALA, F. (2018). Estimation in the spiked Wigner model: A short proof of the replica formula. In 2018 *IEEE International Symposium on Information Theory* (*ISIT*) 1874–1878. IEEE Press, New York.

[26] EL ALAOUI, A., KRZAKALA, F. and JORDAN, M. (2020). Fundamental limits of detection in the spiked Wigner model. *Ann. Statist.* **48** 863–885. MR4102679 https://doi.org/10.1214/19-AOS1826

[27] FAN, Z. (2022). Approximate message passing algorithms for rotationally invariant matrices. *Ann. Statist.* **50** 197–224. MR4382014 https://doi.org/10.1214/21-aos2101

[28] FAN, Z., MEI, S. and MONTANARI, A. (2021). TAP free energy, spin glasses and variational inference. *Ann. Probab.* **49** 1–45. MR4203332 https://doi.org/10.1214/20-AOP1443

[29] FENG, O. Y., VENKATARAMANAN, R., RUSH, C. and SAMWORTH, R. J. (2022). A unifying tutorial on approximate message passing. *Found. Trends Mach. Learn.* **15** 335–536.

[30] FÉRAL, D. and PÉCHÉ, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. MR2291807 https://doi.org/10.1007/s00220-007-0209-3

[31] FLETCHER, A. K. and RANGAN, S. (2018). Iterative reconstruction of rank-one matrices in noise. *Inf. Inference* **7** 531–562. MR3858334 https://doi.org/10.1093/imaiai/iax014

[32] HENRICI, P. (1993). *Applied and Computational Complex Analysis*, *Volume* 3: *Discrete Fourier Analysis*, *Cauchy Integrals*, *Construction of Conformal Maps*, *Univalent Functions* **41**. Wiley, New York.

[33] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. MR1863961 https://doi.org/10.1214/aos/1009210544

[34] JOHNSTONE, I. M. and LU, A. Y. (2009). On consistency and sparsity for principal components analysis in high dimensions. *J. Amer. Statist. Assoc.* **104** 682–693. MR2751448 https://doi.org/10.1198/jasa.2009.0121

[35] KABASHIMA, Y. (2003). A CDMA multiuser detection algorithm on the basis of belief propagation. *J. Phys. A: Math. Gen.* **36** 11111–11121. MR2025247 https://doi.org/10.1088/0305-4470/36/43/030

[36] KABASHIMA, Y., KRZAKALA, F., MÉZARD, M., SAKATA, A. and ZDEBOROVÁ, L. (2016). Phase transitions and sample complexity in Bayes-optimal matrix factorization. *IEEE Trans. Inf. Theory* **62** 4228–4265. MR3515748 https://doi.org/10.1109/TIT.2016.2556702

[37] KNOWLES, A. and YIN, J. (2013). The isotropic semicircle law and deformation of Wigner matrices. *Comm. Pure Appl. Math.* **66** 1663–1749. MR3103909 https://doi.org/10.1002/cpa.21450

[38] KRZAKALA, F., XU, J. and ZDEBOROVÁ, L. (2016). Mutual information in rank-one matrix estimation. In 2016 *IEEE Information Theory Workshop* (*ITW*) 71–75. IEEE Press, New York.

[39] LEDOIT, O. and PÉCHÉ, S. (2011). Eigenvectors of some large sample covariance matrix ensembles. *Probab. Theory Related Fields* **151** 233–264. MR2834718 https://doi.org/10.1007/s00440-010-0298-3

[40] LEDOIT, O. and WOLF, M. (2012). Nonlinear shrinkage estimation of large-dimensional covariance matrices. *Ann. Statist.* **40** 1024–1060. MR2985942 https://doi.org/10.1214/12-AOS989

[41] LELARGE, M. and MIOLANE, L. (2019). Fundamental limits of symmetric low-rank matrix estimation. *Probab. Theory Related Fields* **173** 859–929. MR3936148 https://doi.org/10.1007/s00440-018-0845-x

[42] LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2015). MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel. In 2015 53*rd Annual Allerton Conference on Communication*, *Control*, *and Computing* (*Allerton*) 680–687. IEEE Press, New York.

[43] LESIEUR, T., KRZAKALA, F. and ZDEBOROVÁ, L. (2015). Phase transitions in sparse PCA. In 2015 *IEEE International Symposium on Information Theory* (*ISIT*) 1635–1639. IEEE Press, New York.

[44] LI, G., FAN, W. and WEI, Y. (2023). Approximate message passing from random initialization with applications to $\mathbb{Z}_2$ synchronization. *Proc. Natl. Acad. Sci. USA* **120** Paper No. e2302930120, 7. MR4637851

[45] Li, G. and Wei, Y. (2022). A non-asymptotic framework for approximate message passing in spiked models. arXiv preprint. Available at arXiv:2208.03313.
[46] Li, Y. and Sur, P. (2023). Spectrum-aware adjustment: a new debiasing framework with applications to principal components regression. arXiv preprint. Available at arXiv:2309.07810.
[47] Lolas, P. and Ying, L. (2021). Shrinkage Estimation of Functions of Large Noisy Symmetric Matrices. arXiv preprint. Available at arXiv:2106.05183.
[48] Ma, J. and Ping, L. (2017). Orthogonal AMP. *IEEE Access* **5** 2020–2033.
[49] Miolane, L. (2017). Fundamental limits of low-rank matrix estimation: the non-symmetric case. arXiv preprint. Available at arXiv:1702.00473.
[50] Mondelli, M. and Venkataramanan, R. (2021). PCA initialization for approximate message passing in rotationally invariant models. *Adv. Neural Inf. Process. Syst.* **34** 29616–29629.
[51] Montanari, A. and Venkataramanan, R. (2021). Estimation of low-rank matrices via approximate message passing. *Ann. Statist.* **49** 321–345. MR4206680 https://doi.org/10.1214/20-AOS1958
[52] Montanari, A. and Wein, A. S. (2025). Equivalence of approximate message passing and low-degree polynomials in rank-one matrix estimation. *Probab. Theory Related Fields* **191** 181–233. MR4869255 https://doi.org/10.1007/s00440-024-01322-z
[53] Montanari, A. and Wu, Y. (2024). Statistically optimal firstorder algorithms: A proof via orthogonalization. *Inf. Inference* **13** Paper No. iaae027, 24. MR4806510 https://doi.org/10.1093/imaiai/iaae027
[54] Opper, M., Çakmak, B. and Winther, O. (2016). A theory of solving TAP equations for Ising models with general invariant random matrices. *J. Phys. A: Math. Theor.* **49** 114002, 24. MR3462332 https://doi.org/10.1088/1751-8113/49/11/114002
[55] Opper, M. and Winther, O. (2001). Adaptive and self-averaging Thouless–Anderson–Palmer mean-field theory for probabilistic modeling. *Phys. Rev. E* **64** 056131.
[56] Oymak, S. and Hassibi, B. (2014). A case for orthogonal measurements in linear inverse problems. In 2014 *IEEE International Symposium on Information Theory* 3175–3179. IEEE Press, New York.
[57] Parker, J. T., Schniter, P. and Cevher, V. (2014). Bilinear generalized approximate message passing—Part I: Derivation. *IEEE Trans. Signal Process.* **62** 5839–5853. MR3281527 https://doi.org/10.1109/TSP.2014.2357776
[58] Pastur, L. and Shcherbina, M. (2011). *Eigenvalue Distribution of Large Random Matrices. Mathematical Surveys and Monographs* **171**. Amer. Math. Soc., Providence, RI. MR2808038 https://doi.org/10.1090/surv/171
[59] Paul, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. MR2399865
[60] Péché, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. MR2221787 https://doi.org/10.1007/s00440-005-0466-z
[61] Perry, A., Wein, A. S., Bandeira, A. S. and Moitra, A. (2018). Message-passing algorithms for synchronization problems over compact groups. *Comm. Pure Appl. Math.* **71** 2275–2322. MR3862091 https://doi.org/10.1002/cpa.21750
[62] Perry, A., Wein, A. S., Bandeira, A. S. and Moitra, A. (2018). Optimality and sub-optimality of PCA I: Spiked random matrix models. *Ann. Statist.* **46** 2416–2451. MR3845022 https://doi.org/10.1214/17-AOS1625
[63] Pourkamali, F. and Macris, N. (2023). Rectangular rotational invariant estimator for general additive noise matrices. In 2023 *IEEE International Symposium on Information Theory* (*ISIT*) 2081–2086. IEEE Press, New York.
[64] Rangan, S., Schniter, P. and Fletcher, A. K. (2019). Vector approximate message passing. *IEEE Trans. Inf. Theory* **65** 6664–6684. MR4009222 https://doi.org/10.1109/TIT.2019.2916359
[65] Rush, C. and Venkataramanan, R. (2018). Finite sample analysis of approximate message passing algorithms. *IEEE Trans. Inf. Theory* **64** 7264–7286. MR3876443 https://doi.org/10.1109/TIT.2018.2816681
[66] Schramm, T. and Wein, A. S. (2022). Computational barriers to estimation from low-degree polynomials. *Ann. Statist.* **50** 1833–1858. MR4441142 https://doi.org/10.1214/22-aos2179
[67] Semerjian, G. (2024). Matrix denoising: Bayes-optimal estimators via low-degree polynomials. *J. Stat. Phys.* **191** Paper No. 139, 56. MR4813692 https://doi.org/10.1007/s10955-024-03359-9
[68] Singer, A. (2011). Angular synchronization by eigenvectors and semidefinite programming. *Appl. Comput. Harmon. Anal.* **30** 20–36. MR2737931 https://doi.org/10.1016/j.acha.2010.02.001
[69] Takeuchi, K. (2020). Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements. *IEEE Trans. Inf. Theory* **66** 368–386. MR4053400 https://doi.org/10.1109/TIT.2019.2947058

[70] VENKATARAMANAN, R., KÖGLER, K. and MONDELLI, M. (2022). Estimation in rotationally invariant generalized linear models via approximate message passing. In *International Conference on Machine Learning* 22120–22144. PMLR.

[71] WANG, T., ZHONG, X. and FAN, Z. (2024). Universality of approximate message passing algorithms and tensor networks. *Ann. Appl. Probab.* **34** 3943–3994. MR4783034 https://doi.org/10.1214/24-aap2056

[72] ZHONG, X., SU, C. and FAN, Z. (2022). Empirical Bayes PCA in high dimensions. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **84** 853–878. MR4460578

[73] ZHONG, X., WANG, T. and FAN, Z. (2024). Approximate message passing for orthogonally invariant ensembles: Multivariate non-linearities and spectral initialization. *Inf. Inference* **13** Paper No. iaae024, 22. MR4797647 https://doi.org/10.1093/imaiai/iaae024

[74] ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2006). Sparse principal component analysis. *J. Comput. Graph. Statist.* **15** 265–286. MR2252527 https://doi.org/10.1198/106186006X113430