Orthogonal Approximate Message Passing for Spiked Matrix Models under Rotationally-Invariant Noise

Haohua Chen* Junjie Ma†

April 3, 2026

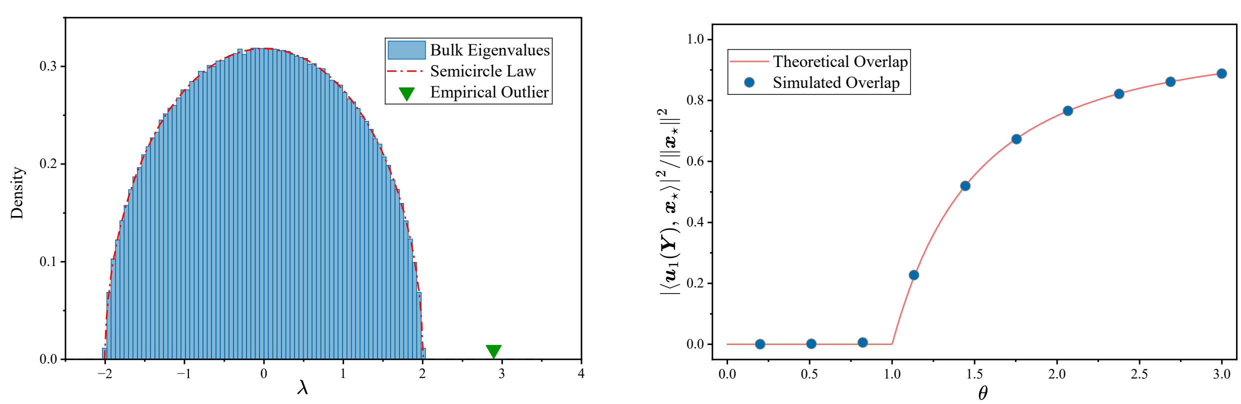
1 Introduction

We consider the problem of signal estimation in the classical *spiked matrix model* [1]:

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

where $\boldsymbol{x}_\star \in \mathbb{R}^N$ denotes an unknown rank-one signal, $\theta > 0$ is the signal-to-noise ratio, and $\boldsymbol{W}$ is a symmetric noise matrix. Despite its apparent simplicity, model (1) serves as a canonical framework for a wide range of high-dimensional inference problems, including sparse principal component analysis, community detection, group synchronization, and low-rank factorization with structured latent variables [2–4].

When $\boldsymbol{W}$ is a Wigner matrix (with i.i.d. Gaussian entries up to symmetry), a natural estimator of $\boldsymbol{x}_\star$ is given by the leading eigenvector of $\boldsymbol{Y}$. The asymptotic behavior of this spectral estimator is characterized by the celebrated Baik–Ben Arous–Péché (BBP) phase transition [5]: below a critical signal-to-noise ratio, the top eigenvector is asymptotically uncorrelated with $\boldsymbol{x}_\star$, whereas above the threshold it acquires a nontrivial overlap and becomes informative; see Fig. 1.



(a) Noise spectrum and an outlier created by the spike.

(b) Overlap between the top eigenvector and $\boldsymbol{x}_\star$ contrast by θ.

Figure 1: **Baik–Ben Arous–Péché phase transition [5] in the spiked Wigner model.**

In many applications, the signal $\boldsymbol{x}_\star$ is not an arbitrary direction but follows a known prior distribution encoding structural information such as sparsity, discreteness, or non-negativity. Leveraging this prior can strictly improve upon purely spectral estimators. *Approximate message passing* (AMP) [2, 6, 7] is a class of

*Academy of Mathematics and Systems Science, Chinese Academy of Sciences. Email: chenhaohua25@mails.ucas.ac.cn
†Academy of Mathematics and Systems Science, Chinese Academy of Sciences. Email: majunjie@lsec.cc.ac.cn

iterative algorithms designed precisely for this purpose:

$$\boldsymbol{r}_t = \boldsymbol{Y}\boldsymbol{x}_t - \langle \partial_{t-1}\boldsymbol{x}_t \rangle\, \boldsymbol{x}_{t-1}, \tag{2}$$

$$\boldsymbol{x}_{t+1} = f_{t+1}(\boldsymbol{r}_t), \tag{3}$$

where $\langle \partial_{t-1}\boldsymbol{x}_t \rangle \stackrel{\text{def}}{=} \frac{1}{N}\sum_{n=1}^{N} f_t'(\boldsymbol{r}_{t-1}[n])$. A key feature of AMP is that its high-dimensional dynamics are sharply characterized by a scalar *state evolution* recursion, playing a role analogous to *density evolution* in belief propagation for LDPC codes [8]. AMP enjoys strong optimality properties. For a fixed number of iterations, suitably designed AMP algorithms are optimal within broad classes of estimators, including general first-order methods [9, 10] and low-degree polynomial estimators [11]. For certain priors and noise models, AMP provably attains the Bayes-optimal risk in the high-dimensional limit [12]. For example, for discrete priors such as BPSK, replica predictions, together with available rigorous results, characterize a regime with no statistical–computational gap, in which suitably tuned AMP algorithms achieve the Bayes-optimal risk [13–15].

2 AMP Algorithms for RI Noise

Much of the classical theory of spiked matrix models is developed under the assumption of Gaussian orthogonal ensemble (GOE) noise. While this setting is analytically convenient, it represents a highly idealized model in which the noise spectrum is fixed and semicircular. In many applications, however, the noise exhibits rich spectral structure arising from correlations, rendering the GOE assumption inadequate. A natural and widely studied extension is to consider *rotationally invariant* (RI) noise models, in which the noise matrix $\boldsymbol{W}$ is invariant in distribution under orthogonal conjugation. This class retains the isotropy of the eigenvectors while allowing for an essentially arbitrary limiting spectral distribution, thereby providing a flexible and realistic framework that strictly generalizes the GOE setting.

From an algorithmic perspective, moving beyond Gaussian noise changes the behavior of iterative estimators. The central intuition behind approximate message passing (AMP) is that each iterate should behave as the true signal observed through an effective Gaussian channel. In the GOE setting, this Gaussian decoupling is preserved by a remarkably simple mechanism: a single scalar Onsager correction cancels the correlations induced across iterations, yielding an exact scalar state evolution. This structural simplicity, however, is specific to Gaussian noise. For general rotationally invariant noise with a nontrivial spectrum, the standard AMP decoupling mechanism breaks down, and the Onsager correction must explicitly account for the spectral structure of the noise matrix $\boldsymbol{W}$. This perspective underlies the development of Fan's rotationally-invariant AMP algorithm [16]:

$$\boldsymbol{r}_t = \boldsymbol{Y}\boldsymbol{x}_t - \sum_{i=1}^{t} b_{t,i}\, \boldsymbol{x}_i, \tag{4}$$

$$\boldsymbol{x}_{t+1} = f_{t+1}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_t), \tag{5}$$

where the coefficients $\{b_{t,i}\}$ depend explicitly on the *free cumulants* of the noise spectrum. While this framework yields a rigorous state evolution description, it does not by itself provide a principled prescription for choosing the denoisers f_{t+1}. In the spiked matrix model, optimizing these nonlinear updates is highly nontrivial: unlike in classical linear or generalized linear models, the canonical choice of the pointwise MMSE denoiser is generally not optimal. Understanding how to design denoisers that achieve optimal performance in this setting remains an open and largely unexplained issue.

Along a related line of work, Barbier *et al.* proposed augmenting the linear step of AMP with a nonlinear spectral transformation [17, 18]. In their approach, the data matrix $\boldsymbol{Y}$ is replaced by a nonlinear eigenvalue map $J(\boldsymbol{Y})$, leading to the iteration

$$\boldsymbol{r}_t = J(\boldsymbol{Y})\boldsymbol{x}_t - \sum_{i=1}^{t} c_{t,i}\, \boldsymbol{x}_i, \tag{6}$$

$$\boldsymbol{x}_{t+1} = f_{t+1}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_t), \tag{7}$$

where the choice of $J(\cdot)$ is motivated by a fixed-point characterization of Bayes-optimal estimators arising from the Thouless–Anderson–Palmer (TAP) equations in statistical physics. It is observed in simulations that this approach could (but not always) lead to better reconstruction performance. However, the resulting dynamics are considerably more complex: general convergence guarantees are not yet available, and the algorithm's optimality properties remain largely open.

3 OAMP Algorithms and Optimality Results

These developments naturally motivate *orthogonal approximate message passing* (OAMP). Rather than explicitly constructing Onsager corrections tailored to the noise spectrum, OAMP enforces asymptotic decorrelation through trace-free and divergence-free constraints on its linear and nonlinear steps [19]. This perspective can be viewed as the natural analogue, for spiked matrix models, of OAMP/VAMP [19, 20] algorithms developed in the compressed sensing literature.

OAMP algorithm. In the symmetric rotationally invariant spiked model, OAMP alternates a spectral step that exploits the noise spectrum and a coordinatewise step that exploits the signal prior. The iterates $\{\boldsymbol{r}_t\}_{t\geq 1}$ and $\{\widehat{\boldsymbol{x}}_t\}_{t\geq 1}$ are generated by

$$\boldsymbol{r}_t = \boldsymbol{\Psi}_t(\boldsymbol{Y})\, \boldsymbol{f}_t(\boldsymbol{r}_1, \dots, \boldsymbol{r}_{t-1}), \qquad \widehat{\boldsymbol{x}}_t = \boldsymbol{\psi}_t(\boldsymbol{r}_1, \dots, \boldsymbol{r}_t), \tag{8}$$

where $\boldsymbol{f}_t$ and $\boldsymbol{\psi}_t$ act coordinatewise. The matrix denoiser $\boldsymbol{\Psi}_t(\boldsymbol{Y})$ is defined by eigenvalue-wise application: if $\boldsymbol{Y} = \boldsymbol{O}\mathrm{diag}(\lambda_1, \dots, \lambda_n)\boldsymbol{O}^\top$, then

$$\boldsymbol{\Psi}_t(\boldsymbol{Y}) = \boldsymbol{O}\mathrm{diag}\big(\Psi_t(\lambda_1), \dots, \Psi_t(\lambda_N)\big)\boldsymbol{O}^\top, \tag{9}$$

for a scalar function $\Psi_t : \mathbb{R} \to \mathbb{R}$. The spectral step is *trace-free* with respect to the noise spectrum, and the coordinatewise nonlinearity is *divergence-free* along every memory direction. Concretely,

$$\text{trace-free:} \qquad \mathbb{E}_{\mathsf{\Lambda}\sim\mu}\big[\Psi_t(\mathsf{\Lambda})\big] = 0, \tag{10}$$

$$\text{divergence-free:} \qquad \mathbb{E}\big[\partial_s f_t(\mathsf{R}_1, \dots, \mathsf{R}_{t-1})\big] = 0, \qquad \forall\, s \in \{1, \dots, t-1\}. \tag{11}$$

where $(\mathsf{R}_1, \dots, \mathsf{R}_{t-1})$ are the state-evolution variables. Under these constraints, the messages decouple: for each fixed t, the empirical law of the iterate converges (in Wasserstein distance):

$$\boldsymbol{r}_t \xrightarrow{\mathsf{W}_2} \mathsf{R}_t = \beta_t\, \mathsf{X}_\star + \tau_t\, \mathsf{Z}, \qquad \mathsf{Z} \sim \mathcal{N}(0,1),\ \mathsf{Z} \perp\!\!\!\perp \mathsf{X}_\star,$$

where the scalars (β_t, τ_t) satisfy a state evolution equation:

$$\beta_t = \mathbb{E}\Big[\mathsf{X}_\star\, \mathsf{F}_t\Big]\, \mathbb{E}_{\mathsf{\Lambda}_\nu\sim\nu}\big[\Psi_t(\mathsf{\Lambda}_\nu)\big], \quad \mathsf{F}_t \overset{\text{def}}{=} f_t(\mathsf{R}_1, \dots, \mathsf{R}_{t-1}) \tag{12}$$

$$\tau_t^2 = \Big(\mathbb{E}[\mathsf{X}_\star \mathsf{F}_t]\Big)^2\, \mathbb{E}_{\mathsf{\Lambda}_\nu\sim\nu}\big[\Psi_t(\mathsf{\Lambda}_\nu)^2\big] \;+\; \Big(\mathbb{E}\big[\mathsf{F}_t^2\big] - \Big(\mathbb{E}[\mathsf{X}_\star \mathsf{F}_t]\Big)^2\Big)\, \mathbb{E}_{\mathsf{\Lambda}\sim\mu}\big[\Psi_t(\mathsf{\Lambda})^2\big] \;-\; \beta_t^2. \tag{13}$$

Here ν is the weak limit of the overlap-weighted spectral measure

$$\nu_N = \frac{1}{N}\sum_{i=1}^{N} \langle \boldsymbol{u}_i(\boldsymbol{Y}), \boldsymbol{x}_\star\rangle^2\, \delta_{\lambda_i(\boldsymbol{Y})}, \qquad \nu_N \Rightarrow \nu, \tag{14}$$

where $\{(\lambda_i(\boldsymbol{Y}), \boldsymbol{u}_i(\boldsymbol{Y}))\}_{i\leq N}$ are eigenpairs of $\boldsymbol{Y}$.

Within this framework, the state evolution formalism enables a principled optimization of the matrix and iterative denoisers [21]. For a local parameter $\rho_t > 0$, the optimal matrix denoiser is

$$\Psi_t^\star(\lambda) = \left(1 - \mathbb{E}_{\mathsf{\Lambda}\sim\mu}\left[\frac{\phi(\mathsf{\Lambda})}{\phi(\mathsf{\Lambda}) + \rho_t}\right]\right)^{-1} \frac{\phi(\lambda)}{\phi(\lambda) + \rho_t}, \tag{15}$$

and the optimal iterate denoiser is the divergence-free minimum mean-square error denoiser [19, 21]. The associated recursion converges to a stationary point (ρ, ω) satisfying

$$\rho = \frac{1}{\mathrm{mmse}_\pi(\omega)} - \frac{1}{1-\omega}, \tag{16a}$$

$$\omega = 1 - \left(\mathbb{E}_{\Lambda\sim\mu}\left[\frac{\phi(\Lambda)}{\phi(\Lambda)+\rho}\right]\right)^{-1} \mathbb{E}_{\Lambda\sim\mu}\left[\frac{1}{\phi(\Lambda)+\rho}\right], \tag{16b}$$

where ω admits the interpretation of the asymptotic squared cosine similarity between $\boldsymbol{r}_t$ and the true signal $\boldsymbol{x}_\star$, and $\phi(\cdot)$ is a nonlinearity that depends on the noise spectrum. (The explicit formula of $\phi(\cdot)$ can be found in [21].)

Optimality of OAMP. The fixed-point equations (16) are consistent with the replica predictions for Bayes-optimal inference under rotationally invariant noise [18]. Under the widely accepted validity of the replica description, this consistency suggests that whenever the fixed point equations admit a unique solution, the corresponding algorithmic dynamics attain Bayes-optimal performance. In such regimes, OAMP attains this solution and is therefore optimal. When multiple solutions coexist, however, OAMP may converge to a sub-optimal one. A natural question then arises: can any other polynomial-time algorithm do better?

Following the framework introduced for Gaussian orthogonal ensemble noise via *general first-order methods* (GFOM) [9, 10], one can formalize a broad comparison class of iterative algorithms. In the GOE setting, this class includes spectral methods, power iterations, and message-passing–type schemes, and AMP is known to be optimal within GFOM at any fixed number of iterations. For rotationally invariant noise, the analogous comparison class replaces the linear operator $\boldsymbol{Y}$ with trace-free spectral denoisers constructed from $\boldsymbol{Y}$, reflecting the necessity of explicitly exploiting spectral information.

This perspective leads naturally to the following abstract formulation: a generalized iterative algorithm (GIA) generates a sequence of iterates $\{\boldsymbol{x}_t\}_{t\geq 1}$ and estimates $\{\widehat{\boldsymbol{x}}_t\}_{t\geq 1}$ via

$$\boldsymbol{x}_{t+1} = g_t(\boldsymbol{Y})\, f_t(\boldsymbol{x}_1, \dots, \boldsymbol{x}_t) + F_t(\boldsymbol{x}_1, \dots, \boldsymbol{x}_t), \qquad \widehat{\boldsymbol{x}}_{t+1} = \psi_{t+1}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t+1}),$$

where $g_t(\boldsymbol{Y})$ is a trace-free matrix denoiser and f_t, F_t, ψ_{t+1} act coordinatewise with finite memory. Within this class, the optimal OAMP algorithm achieves the smallest asymptotic mean-square error at any fixed iteration budget: for every fixed t and under regularity assumptions, we have in probability

$$\lim_{N\to\infty} \frac{1}{N}\|\boldsymbol{x}_t^{\mathrm{GIA}} - \boldsymbol{x}_\star\|_2^2 \;\geq\; \lim_{N\to\infty} \frac{1}{N}\|\boldsymbol{x}_t^{\mathrm{OAMP}} - \boldsymbol{x}_\star\|_2^2.$$

This is the rotationally invariant analogue of finite-iteration optimality of AMP under Gaussian orthogonal ensemble noise [9, 10], with the important modification that the comparison class is allowed to adapt to the spectrum through trace-free spectral denoisers.

Extension to asymmetric models. The framework described above naturally extends to the rectangular rotationally invariant spiked model, in which one observes

$$\boldsymbol{Y} = \frac{\theta}{\sqrt{MN}}\boldsymbol{u}_\star \boldsymbol{v}_\star^\top + \boldsymbol{W} \in \mathbb{R}^{M\times N}, \tag{17}$$

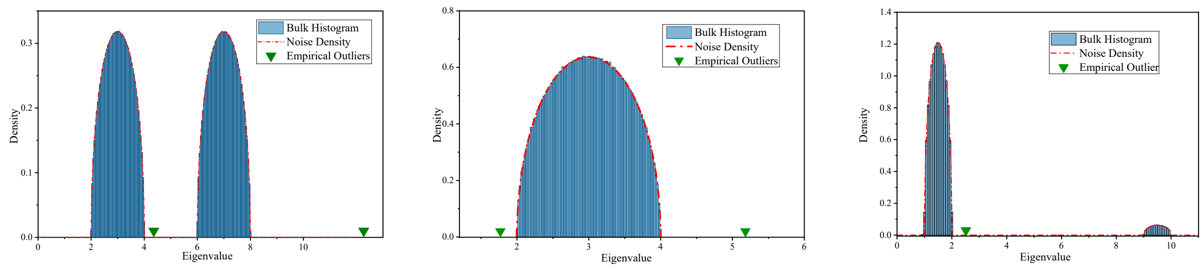
with $\boldsymbol{u}_\star \in \mathbb{R}^M$ and $\boldsymbol{v}_\star \in \mathbb{R}^N$, and where the noise is bi-orthogonally invariant. Rectangular OAMP produces iterates $\{\boldsymbol{u}_t\}_{t\geq 1} \subset \mathbb{R}^M$ and $\{\boldsymbol{v}_t\}_{t\geq 1} \subset \mathbb{R}^N$ by [22]

$$\boldsymbol{u}_t = \boldsymbol{F}_t(\boldsymbol{Y}\boldsymbol{Y}^\top)\, \boldsymbol{f}_t(\boldsymbol{u}_1, \dots, \boldsymbol{u}_{t-1}) \;+\; \widetilde{\boldsymbol{F}}_t(\boldsymbol{Y}\boldsymbol{Y}^\top)\, \boldsymbol{Y}\, \boldsymbol{g}_t(\boldsymbol{v}_1, \dots, \boldsymbol{v}_{t-1}), \tag{18}$$

$$\boldsymbol{v}_t = \boldsymbol{G}_t(\boldsymbol{Y}^\top\boldsymbol{Y})\, \boldsymbol{g}_t(\boldsymbol{v}_1, \dots, \boldsymbol{v}_{t-1}) \;+\; \widetilde{\boldsymbol{G}}_t(\boldsymbol{Y}^\top\boldsymbol{Y})\, \boldsymbol{Y}^\top\, \boldsymbol{f}_t(\boldsymbol{u}_1, \dots, \boldsymbol{u}_{t-1}). \tag{19}$$

The iterate denoisers f_t and g_t satisfy divergence-free constraints (the analogue of the symmetric condition), while the matrix denoisers $\boldsymbol{F}_t$ and $\boldsymbol{G}_t$ satisfy trace-free constraints with respect to the limiting empirical spectral distributions μ_1 and μ_2 of $\boldsymbol{W}\boldsymbol{W}^\top$ and $\boldsymbol{W}^\top\boldsymbol{W}$, respectively. The cross denoisers $\widetilde{\boldsymbol{F}}_t$ and $\widetilde{\boldsymbol{G}}_t$ are not required to be trace-free.

State evolution for the rectangular iterates must account for the coupling between the left and right messages induced by bi-orthogonal invariance; this coupling is captured by an additional signed cross measure



(a) Symmetric, two bulks.

(b) Rectangular, one bulk.

(c) Rectangular, two bulks.

Figure 2: **Spectral behaviors of spike-induced outliers under RI noise.**

(see Definition 5, [22]) that encodes the asymptotic effect of the cross terms. With this characterization, the optimal design again selects divergence-free minimum mean-square error iterate denoisers and matrix denoisers tuned to the local signal-to-noise ratios. The resulting state evolution converges to a stationary point $(\rho_1, \rho_2, w_1, w_2)$ satisfying

$$\rho_1 = \frac{1}{\mathrm{mmse}_{\mathsf{U}}(w_1)} - \frac{1}{1-w_1}, \qquad \mathrm{mmse}_{\mathsf{U}}(w_1) = \frac{1}{\rho_1}\Big[1 - \int P^\star(\lambda;\rho_1,\rho_2)\, d\mu_1(\lambda)\Big], \tag{20}$$

$$\rho_2 = \frac{1}{\mathrm{mmse}_{\mathsf{V}}(w_2)} - \frac{1}{1-w_2}, \qquad \mathrm{mmse}_{\mathsf{V}}(w_2) = \frac{1}{\rho_2}\Big[1 - \int Q^\star(\lambda;\rho_1,\rho_2)\, d\mu_2(\lambda)\Big], \tag{21}$$

where w_1 and w_2 are interpreted as the limiting squared cosine similarities between the OAMP iterates and the corresponding ground-truth signals. The stationary relations above coincide with the replica-symmetric predictions for Bayes-optimal inference in rectangular rotationally invariant spiked models, as will be detailed in a forthcoming paper.

4 Optimal Spectral Estimators and Spectral Initialization

Spectral methods are often the first line of attack for spiked matrix models, relying on the leading eigenvalue or singular value to extract signal information. However, beyond the classical Gaussian setting, the structure of rotationally invariant (RI) noise leads to markedly richer spectral phenomena.

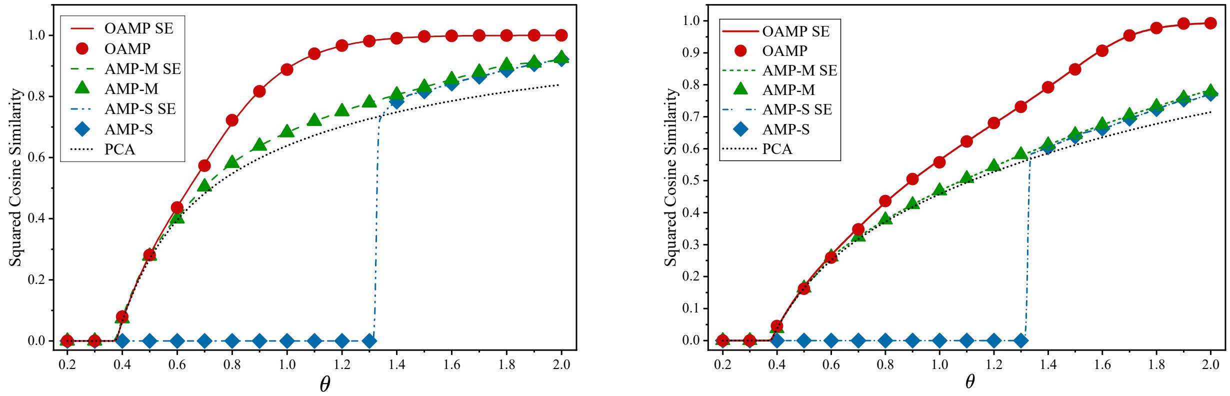
In the symmetric RI model, a single spike can produce *multiple* outlier eigenvalues when the noise spectrum has multiple bulks; see Fig. 2(a). In the rectangular RI model, the same multiplicity can occur even with a single-bulk noise spectrum; see Fig. 2(b). In both settings, certain spectral shapes can cause a *non-leading* outlier to separate from the bulk before the largest one, implying that the earliest detectable signal need not align with the top eigenvector or singular vector; see Fig. 2(c). As a consequence, standard PCA—which relies solely on the leading spectral component—can be strictly suboptimal. This observation motivates the study of *optimal spectral estimators* that aggregate information from *all* informative outliers in order to maximize alignment with the true underlying signal.

In the rectangular model (17), suppose $\boldsymbol{Y}\boldsymbol{Y}^\top$ exhibits a set of outliers indexed by $\mathcal{I}_M$. A general spectral estimator supported on outlier directions takes the form

$$\boldsymbol{u}_{\mathrm{PCA}}(\boldsymbol{c}) = \sqrt{M} \sum_{i \in \mathcal{I}_M} c_i\, \boldsymbol{u}_i(\boldsymbol{Y}\boldsymbol{Y}^\top),$$

and similarly for $\boldsymbol{v}_\star$ using the corresponding singular vectors of $\boldsymbol{Y}$ or eigenvectors of $\boldsymbol{Y}^\top\boldsymbol{Y}$. If one had oracle access to the overlaps $\langle \boldsymbol{u}_\star, \boldsymbol{u}_i \rangle$, then the best combination within this outlier-supported class is

$$\boldsymbol{u}^\star_{\mathrm{ora}} = \sqrt{M} \sum_{i \in \mathcal{I}_M} \langle \boldsymbol{u}_\star, \boldsymbol{u}_i(\boldsymbol{Y}\boldsymbol{Y}^\top) \rangle\, \boldsymbol{u}_i(\boldsymbol{Y}\boldsymbol{Y}^\top), \tag{22}$$



(a) $\boldsymbol{u}$-channel (b) $\boldsymbol{v}$-channel

Figure 3: Non-Gaussian RI noise with bulk density $\mu(\lambda) = \frac{2}{\pi}\sqrt{(\lambda-1)(3-\lambda)}\,\mathbf{1}_{[1,3]}(\lambda)$ and Rademacher priors. Markers are empirical averages over 50 trials. When outlier eigenvectors are present, OAMP denotes the spectrally-initialized optimal OAMP in [22, Theorem 2], whereas AMP uses the spectral initialization of [25] based only on the top eigenvector. When no outlier eigenvectors are available, all methods are initialized with cosine similarity 0.1.

which maximizes the squared cosine similarity among all outlier-supported linear combinations. Random matrix theory (cf. [23,24]) predicts the *magnitudes* of these overlaps through signal-direction spectral measures. The remaining difficulty is the *relative signs*: each outlier overlap is only identifiable up to an unknown ± 1, and combining several outliers requires these signs to be aligned, otherwise different directions can partially cancel.

A convenient way to think about the sign issue is through the decoupling picture: each outlier eigenvector induces an effective scalar signal-plus-Gaussian channel in the high-dimensional limit. For exactly Gaussian priors, the model is symmetric under sign flips and the relative sign is not identifiable. For *non-Gaussian* priors, the symmetry breaks and the relative signs become identifiable. Two consistent and practical approaches are developed in [22]:

- **MLE (maximum likelihood estimation).** When the non-Gaussian prior is known, one can compare the joint likelihood of two outlier channels under the two choices $s \in \{\pm 1\}$ and select the maximizing sign.
- **NGMC (non-Gaussian moment contrast).** When certain non-Gaussian moments are available, a lightweight contrast statistic can consistently recover the relative signs at low computational cost.

With signs in hand, one can implement an outlier-aggregating estimator that matches the oracle performance asymptotically. Let

$$w_i = \lim_{M\to\infty} \left|\langle \boldsymbol{u}_\star, \boldsymbol{u}_i(\boldsymbol{Y}\boldsymbol{Y}^\top)\rangle\right|, \qquad i \in \mathcal{I}_M,$$

be the deterministic weights predicted by random matrix theory for the observed outliers, and let s_i be the estimated relative signs (via MLE or NGMC). Define the normalized initialization

$$\boldsymbol{u}_{\text{init}} = \Big(\sum_{i\in\mathcal{I}_M} w_i\Big)^{-1/2} \sqrt{M} \sum_{i\in\mathcal{I}_M} s_i \sqrt{w_i}\, \boldsymbol{u}_i(\boldsymbol{Y}\boldsymbol{Y}^\top), \tag{23}$$

with an analogous construction for $\boldsymbol{v}_{\text{init}}$. Whenever outliers exist and the relative signs are recovered consistently, (23) achieves the same asymptotic squared cosine similarity as the oracle estimator (22).

This initialization (23) is directly compatible with OAMP (cf. [22, Theorem 2]). Under the optimal denoiser configuration, the SE recursion itself is unchanged; only the initial overlap parameters are reset to the values achieved by the spectral estimator. Equivalently, one replaces the default initial cosine similarity in the SE updates by the cosine similarity induced by (23); all subsequent SE iterations remain the same, now starting from the strongest available spectral alignment.

To assess the practical impact of spectral initialization and SE-optimal design, we compare OAMP with the RI-AMP framework of [16] in a general RI setting. Following [16, Remark 3.3], we report two baselines: a single-iterate scheme dubbed AMP-S, which applies a scalar MMSE denoiser directly to the current iterate, and a multi-iterate scheme dubbed AMP-M, which first forms the optimal linear combination of all past effective signal-plus-noise observations (using their state-evolution covariances) and then applies a single scalar MMSE denoiser to the combined statistic. Figure 3 summarizes the outcome: spectrally-initialized SE-optimal OAMP consistently improves upon PCA as well as both RI-AMP baselines over the tested SNRs.

5 Acknowledgement

The framework and results presented in this article are primarily based on two recent works. The theoretical analysis regarding the optimality of Approximate Message Passing (AMP) under rotationally invariant noise is established by Dudeja, Liu, and Ma (2024). The algorithmic extension to rectangular models, along with the derivation of optimal spectral initializations, is detailed by Chen, Liu, and Ma (2025). This article serves as a summary of the key methodologies and findings from these studies.

References

[1] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Stat.*, vol. 29, no. 2, pp. 295–327, 2001.

[2] Y. Deshpande and A. Montanari, "Information-theoretically optimal sparse PCA," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2197–2201, June 2014.

[3] E. Abbe, "Community detection and stochastic block models: Recent developments," *J. Mach. Learn. Res.*, vol. 18, no. 177, pp. 1–86, 2018.

[4] A. Singer, "Angular synchronization by eigenvectors and semidefinite programming," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 1, pp. 20–36, 2011.

[5] J. Baik, G. Ben Arous, and S. Péché, "Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices," *Ann. Probab.*, vol. 33, no. 5, pp. 1643–1697, 2005.

[6] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci. USA*, vol. 106, no. 45, pp. 18914–18919, 2009.

[7] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 57, pp. 764–785, Feb. 2011.

[8] T. Richardson and R. Urbanke, *Modern coding theory.* Cambridge university press, 2008.

[9] M. Celentano, A. Montanari, and Y. Wu, "The estimation error of general first order methods," in *Proc. Conf. Learn. Theory (COLT)*, vol. 125, pp. 1078–1141, 2020.

[10] A. Montanari and Y. Wu, "Statistically optimal first order algorithms: A proof via orthogonalization," *Inf. Inference*, vol. 13, p. iaae027, Oct. 2024.

[11] A. Montanari and A. S. Wein, "Equivalence of approximate message passing and low-degree polynomials in rank-one matrix estimation," *Probab. Theory Relat. Fields*, vol. 191, no. 1, pp. 181–233, 2025.

[12] M. Lelarge and L. Miolane, "Fundamental limits of symmetric low-rank matrix estimation," *Probab. Theory Relat. Fields*, vol. 173, no. 3, pp. 859–929, 2019.

[13] T. Lesieur, F. Krzakala, and L. Zdeborová, "MMSE of probabilistic low-rank matrix estimation: Universality with respect to the output channel," in *Proc. Annu. Allerton Conf. Commun., Control, Comput.*, pp. 680–687, 2015.

[14] F. Krzakala, J. Xu, and L. Zdeborová, "Mutual information in rank-one matrix estimation," in *Proc. IEEE Inf. Theory Workshop (ITW)*, pp. 71–75, Sept. 2016.

[15] J. Barbier and N. Macris, "The adaptive interpolation method for proving replica formulas. Applications to the Curie-Weiss and Wigner spike models," *J. Phys. A: Math. Theor.*, vol. 52, no. 29, p. 294002, 2019.

[16] Z. Fan, "Approximate message passing algorithms for rotationally invariant matrices," *Ann. Stat.*, vol. 50, no. 1, pp. 197–224, 2022.

[17] J. Barbier, F. Camilli, M. Mondelli, and M. Sáenz, "Fundamental limits in structured principal component analysis and how to reach them," *Proc. Natl. Acad. Sci. USA*, vol. 120, no. 30, p. e2302028120, 2023.

[18] J. Barbier, F. Camilli, Y. Xu, and M. Mondelli, "Information limits and Thouless-Anderson-Palmer equations for spiked matrix models with structured noise," *Phys. Rev. Res.*, vol. 7, no. 1, p. 013081, 2025.

[19] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[20] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Trans. Inf. Theory*, vol. 65, pp. 6664–6684, Oct. 2019.

[21] R. Dudeja, S. Liu, and J. Ma, "Optimality of Approximate Message Passing Algorithms for Spiked Matrix Models with Rotationally Invariant Noise," *to appear in Annals of Statistics*, 2024.

[22] H. Chen, S. Liu, and J. Ma, "Orthogonal approximate message passing with optimal spectral initializations for rectangular spiked matrix models," *arXiv preprint arXiv:2512.19334*, 2025.

[23] F. Benaych-Georges and R. R. Nadakuditi, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *Adv. Math.*, vol. 227, no. 1, pp. 494–521, 2011.

[24] F. Benaych-Georges and R. R. Nadakuditi, "The singular values and vectors of low rank perturbations of large rectangular random matrices," *J. Multivariate Anal.*, vol. 111, pp. 120–135, 2012.

[25] X. Zhong, T. Wang, and Z. Fan, "Approximate message passing for orthogonally invariant ensembles: Multivariate non-linearities and spectral initialization," *Inf. Inference*, vol. 13, no. 3, p. iaae024, 2024.