

Orthogonal Approximate Message Passing with Optimal Spectral Initializations for Rectangular Spiked Matrix Models

Haohua Chen*

Songbin Liu[†]

Junjie Ma[‡]

December 20, 2025

Abstract

We propose an orthogonal approximate message passing (OAMP) algorithm for signal estimation in the rectangular spiked matrix model with general rotationally invariant (RI) noise. We establish a rigorous state evolution that precisely characterizes the algorithm's high-dimensional dynamics and enables the construction of iteration-wise optimal denoisers. Within this framework, we accommodate spectral initializations under minimal assumptions on the empirical noise spectrum. In the rectangular setting, where a single rank-one component typically generates multiple informative outliers, we further propose a procedure for combining these outliers under mild non-Gaussian signal assumptions. For general RI noise models, the predicted performance of the proposed optimal OAMP algorithm agrees with replica-symmetric predictions for the associated Bayes-optimal estimator, and we conjecture that it is statistically optimal within a broad class of iterative estimation methods.

Contents

1	Introduction	3
2	Preliminaries	4
2.1	Rectangular spiked model and assumptions	5
2.2	High-dimensional asymptotics and notation	5
2.3	Stieltjes Transform, Hilbert Transform, and C-transform	6
2.4	Signal-eigenspace Spectral Measures	6
2.5	The Master Equation and Outlier Location	7
2.6	Limiting Spectral Measures and Outlier Behavior	8
3	Orthogonal Approximate Message Passing Algorithms	10
3.1	Orthogonally AMP for Rectangular Spiked Models	11
3.2	The Optimal OAMP Algorithm	12
3.3	Example: I.I.D Gaussian Noise	14
4	Optimal Spectral Estimation Under Multiple Outliers	14
4.1	Optimal Oracle Spectral Estimators	15
4.2	Data-Driven Optimal Linear Spectral Estimators	16
4.3	Estimation of Relative Signs	18
5	OAMP Algorithm with Spectral Initialization	19
5.1	Spectrally-Initialized Optimal OAMP	19
5.2	State Evolution of Spectrally-Initialized OAMP	20

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

[†]Department of Electrical Engineering, Columbia University. Email: s15878@columbia.edu

[‡]Academy of Mathematics and Systems Science, Chinese Academy of Sciences. Email: majunjie@lsec.cc.ac.cn

6	Simulation Results	21
6.1	Spectral Estimator	22
6.2	Performance of OAMP	23
7	Conclusions and Future Work	24
A	Proofs for Preliminaries and Spectral Analysis	29
A.1	Proof of Lemma 1	29
A.2	Limits of Stieltjes Transforms	31
A.3	Proof of Lemma 2	33
A.4	Proof of Proposition 1	35
A.5	Integrals of Spectral Measures	38
B	General OAMP Algorithm with Rotationally-Invariant Matrices	39
B.1	General OAMP Iteration	39
B.2	State Evolution of General OAMP Iteration	40
B.3	Proof of Theorem 3	43
C	State Evolution of OAMP for Spiked Models (Theorem 1)	47
C.1	Proof of Lemma 5	49
C.2	Proof of Lemma 6	52
D	Derivations for Optimal Denoisers	54
D.1	Optimal Matrix Denoisers	54
D.2	Optimal Iterate Denoisers	57
E	Proof of Proposition 2	57
E.1	Proof of Lemma 7	59
F	Proof of Proposition 3 (I.I.D. Gaussian Noise Model)	61
F.1	Spectral Analysis of I.I.D. Gaussian Noise Model	61
F.2	Optimal Denoisers and its OAMP Recursion	61
F.3	Proof of Proposition 3	63
G	Proof of The Optimal Spectral Estimators	65
G.1	Proof of Proposition 4	65
G.2	Proof of Proposition 5	66
G.3	Proof of Proposition 6	69
G.4	Proof of Proposition 7	70
	G.4.1 Both Gaussian priors: impossibility of sign recovery	70
	G.4.2 Non-Gaussian Priors: Identifiability and MLE Consistency	71
	G.4.3 Inter-channel Sign Coupling	73
G.5	Proof of Proposition 8	74
H	Global Sign Detection	76
H.1	Global-Sign Maximum Likelihood Estimator (GSMLE) Scheme	77
H.2	Global Sign Odd-moment Contrast Scheme (GOMC) Scheme	77
H.3	Selection of the Signed DMMSE Estimators	78
I	Proof of Theorem 2	78
J	Some Miscellaneous Results	80

1 Introduction

We study the estimation of asymmetric rank-one signals $\mathbf{u}_* \in \mathbb{R}^M$ and $\mathbf{v}_* \in \mathbb{R}^N$ generated from the *rectangular spiked model*

$$\mathbf{Y} = \frac{\theta}{\sqrt{MN}} \mathbf{u}_* \mathbf{v}_*^\top + \mathbf{W} \in \mathbb{R}^{M \times N}, \quad (1)$$

where $\theta > 0$ is a signal-to-noise ratio (SNR) and $\mathbf{W}$ is a noise matrix. This model is widely used to analyze high-dimensional data in which the number of features and samples are of comparable scale, with applications ranging from financial data analysis [1, 2] to community detection [3, 4].

In the classical setting where $\mathbf{W}$ has i.i.d. Gaussian entries, the fundamental behavior of PCA is well understood. A sharp phase transition governs when the leading singular vectors correlate with the underlying signals [5–8]. These guarantees can be improved by incorporating structural priors such as sparsity [9–13] or Bayesian assumptions [14–16]. Approximate Message Passing (AMP) algorithms [17–34], play an important role in these settings: in the high-dimensional limit, their empirical performance is exactly described by a deterministic state evolution (SE) recursion. This property has enabled rigorous optimality guarantees among large classes of first order methods [28] and low-degree polynomial estimators [35], and agreement with replica predictions for the minimum mean square error (MMSE) in certain regimes [21, 26, 36, 37].

Naturally, in many practical high-dimensional settings, the noise structure deviates from the idealized i.i.d. Gaussian assumption. Rotationally invariant (RI) noise models, in which the singular vectors of $\mathbf{W}$ are Haar distributed and independent of its singular values, admit arbitrary limiting spectra and therefore form a broad and expressive class for high-dimensional noise. For such models, the PCA outlier behavior is well understood [38–41], and several AMP-type extensions have been proposed [30, 42, 43], though their MSE optimality remains unresolved. For the *symmetric* counterpart of (1), [44] introduced an orthogonal approximate message passing (OAMP) algorithm [22, 23] in which each iteration applies a matrix denoiser tailored to the limiting noise spectrum, and established a state evolution characterization together with optimality guarantees. The form of these matrix denoisers is closely related to classical matrix denoising and covariance shrinkage procedures [45–48], and the performance achieved in [44] was shown to match replica-symmetric predictions for the Bayes-optimal performance in certain regimes [43, 49].

When turning to rectangular models, an additional difficulty arises from the structure of spectral information itself. Unlike the i.i.d. Gaussian case, where a rank-one signal produces a single informative outlier (see Remark 2), rectangular RI models can generate multiple outlier singular values [39, 50], distributing the signal energy across them. Standard PCA is therefore suboptimal, as the principal components may not be the most informative [51], with the signal energy often spread across several singular directions [52, 53]. These works point to the need for combining all informative singular vectors, while the development of a practically executable aggregation scheme remains open.

A natural and effective initialization for AMP is based on the principal components (PCs) of the data [54–58]. However, this approach induces dependence on the noise matrix $\mathbf{W}$, violating a crucial assumption underlying standard state evolution analysis. In the i.i.d. Gaussian setting, this difficulty was resolved using a decoupling technique that separates the PCs from the spectral bulk [28]. This approach relies critically on the entrywise independence of $\mathbf{W}$ and does not directly extend to general rotationally invariant (RI) ensembles. For RI noise, subsequent work developed a different approach based on a two-phase artificial AMP construction [42, 59], in which an auxiliary AMP with a noise-independent initialization is designed to converge to the empirical PCA estimator. Variants of this method have also proved useful in Gaussian generalized linear models [55, 56], although the analyses in both the matrix and GLM settings rely on additional technical conditions, such as non-negative free cumulants [59] or sufficiently large SNR [42].

Our Contributions. This paper develops and analyzes an OAMP algorithm for the rectangular spiked model with rotationally invariant (RI) noise. Our main contributions are as follows.

- **Optimal OAMP for Rectangular RI Models:** We extend the OAMP framework [44] to the rectangular setting and establish a rigorous state evolution (SE) that characterizes the joint dynamics of the left and right singular vector estimates. This analysis allows us to derive iteration-wise Bayes-optimal matrix and scalar denoisers. We demonstrate that the algorithm’s fixed point aligns with replica symmetric predictions for the minimum mean square error (MMSE) [60] in the absence

of statistical-computational gaps, and in the specific case of i.i.d. Gaussian noise, it recovers the performance of standard AMP.

- **Optimal Spectral Initialization with Multiple Outliers:** A key challenge in rectangular RI models is that a single rank-one signal generally generates multiple informative outlier singular values. Standard PCA (using only the top singular vector) is therefore suboptimal. We characterize the theoretically optimal linear combination of all informative outliers. To implement this in practice, we solve the “relative sign alignment” problem, where the signs of the outliers are unknown, by proposing two methods: a Maximum Likelihood Estimator (MLE) and a computationally efficient Non-Gaussian Moment Contrast (NGMC) scheme. The NGMC method requires only a mild condition (the existence of an even moment distinct from the Gaussian) to asymptotically match oracle performance.
- **Spectrally-Initialized OAMP:** We integrate the optimal spectral estimator as a principled initialization for OAMP. Based on a resolvent reformulation of the singular equation, we show that the spectral step can be viewed as a single OAMP update, thereby incorporating spectral initialization seamlessly into the OAMP framework and eliminating the need for artificial two-phase constructions or additional assumptions (such as nonnegative free cumulants [59] or sufficiently large signal-to-noise ratios [42]) used in earlier analyses of spectrally initialized AMP. This formulation yields a state evolution characterization for spectrally initialized OAMP that explicitly accounts for the intrinsic global-sign ambiguity and applies to general rotationally invariant noise models.

We conclude this section by introducing the notations used in this paper.

Algebra. Let $\mathbb{N}$, $\mathbb{R}$, $\mathbb{R}_+$ and $\mathbb{C}$ denote the sets of positive integers, real numbers, positive real numbers and complex numbers, respectively. For any $N \in \mathbb{N}$, we define the set $[N] := \{1, 2, 3, \dots, N\}$, and let $\mathbb{O}(N)$ denote the set of $N \times N$ orthogonal matrices. We use bold-face font for vectors and matrices whose dimensions diverge, such as signal vectors $\mathbf{u}_* \in \mathbb{R}^M$, $\mathbf{v}_* \in \mathbb{R}^N$ or a noise matrix $\mathbf{W} \in \mathbb{R}^{M \times N}$. For objects in a fixed finite dimension k , we use regular font. Specifically, for vectors $x, y \in \mathbb{R}^k$, let $\|x\|_2$ denote the ℓ_2 norm, $\langle x, y \rangle = \sum_{i=1}^k x_i y_i$ be the standard inner product, and $\text{diag}(x)$ be the $k \times k$ diagonal matrix formed by the entries of x . For any vector $s \in \mathbb{R}^k$, we write $[s]_j$ for its j -th coordinate. For a matrix $M \in \mathbb{R}^{k \times k}$, we write $\text{Tr}(M)$, $\|M\|_{\text{op}}$, and $\|M\|_F$ for its trace, operator (spectral) norm, and Frobenius norm, respectively. If M is symmetric, its eigenvalues are ordered $\lambda_1(M) \geq \dots \geq \lambda_k(M)$, with corresponding eigenvectors $u_1(M), \dots, u_k(M)$. The spectrum of $\mathbf{M}$ is denoted $\text{sp}(\mathbf{M})$. $|\mathcal{A}|$ denotes the cardinality of a finite set $\mathcal{A}$. $\Im(z)$ denotes the imaginary part of a complex number $z \in \mathbb{C}$.

Probability and Analysis. We denote the Gaussian distribution on $\mathbb{R}^k$ with mean vector $\mu \in \mathbb{R}^k$ and covariance matrix $\Sigma \in \mathbb{R}^{k \times k}$ by $\mathcal{N}(\mu, \Sigma)$. For a finite set A , $\text{Unif}(A)$ represents the uniform distribution on A , and for any $x \in \mathbb{R}$, the measure $\delta_{\{x\}}$ denotes the point mass (or Dirac measure) at x . By extension, $\text{Unif}(\mathbb{O}(N))$ denotes the Haar measure on the orthogonal group $\mathbb{O}(N)$. Furthermore, for sequences of random variables, convergence almost surely and in distribution are denoted by $\xrightarrow{a.s.}$ and $\xrightarrow{d}$, respectively. For random variables X and Y , we write $X \perp\!\!\!\perp Y$ to denote that X and Y are independent. For a finite measure χ of bounded variation on a space $\mathcal{E}$ and any bounded, Borel-measurable function $f : \mathcal{E} \rightarrow \mathbb{R}$, we denote its integral by $\langle f(x) \rangle_\chi \stackrel{\text{def}}{=} \int_{\mathcal{E}} f(x) d\chi(x)$. The support of a measure χ is denoted as $\text{supp}(\chi)$. We use $\text{sign}(\cdot)$ to denote the signum function, which returns 1, -1 , or 0 if its argument is positive, negative, or zero, respectively.

2 Preliminaries

We collect here the main probabilistic and spectral tools used in our analysis of the rectangular spiked model. We first formalize the high-dimensional asymptotic regime and convergence notions, and then review the spectral transform machinery associated with the noise spectrum. We introduce signal-projected spectral measures that encode how the eigenspace of the observation aligns with the true signal directions, and summarize their limiting behavior and the resulting outlier eigenvalues. These results will be used in the spectral estimators and as initialization for the OAMP state evolution in Sections 4 and 5.

2.1 Rectangular spiked model and assumptions

We recall the rectangular spiked model introduced in (1):

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

and detail the asymptotic regime and structural assumptions on the signal and the noise.

Assumption 1. We make the following assumptions on the signal and noise in the model (1).

- (a) We consider the asymptotic regime where $M, N \rightarrow \infty$ such that the aspect ratio converges, $M/N \rightarrow \delta \in (0, 1]$.
- (b) The signal and side information, represented by random vector pairs, converge in Wasserstein distance:

$$(\mathbf{u}_*, \mathbf{a}) \xrightarrow{W} (\mathbf{U}_*, \mathbf{A}), \quad (\mathbf{v}_*, \mathbf{b}) \xrightarrow{W} (\mathbf{V}_*, \mathbf{B}),$$

where $(\mathbf{U}_*, \mathbf{A}, \mathbf{V}_*, \mathbf{B})$ have finite moments of all orders. Without loss of generality, we assume

$$\mathbb{E}[\mathbf{U}_*^2] = \mathbb{E}[\mathbf{V}_*^2] = 1.$$

- (c) The noise matrix $\mathbf{W} \in \mathbb{R}^{M \times N}$ is independent of $(\mathbf{u}_*, \mathbf{a}, \mathbf{v}_*, \mathbf{b})$ and is orthogonally invariant. Specifically, its singular value decomposition $\mathbf{W} = \mathbf{U}_{\mathbf{W}} \text{diag}(\boldsymbol{\sigma}) \mathbf{V}_{\mathbf{W}}^\top$. The matrices $\mathbf{U}_{\mathbf{W}} \in \mathbb{O}(M)$ and $\mathbf{V}_{\mathbf{W}} \in \mathbb{O}(N)$ are independent and Haar distributed on their respective orthogonal groups, and $\text{diag}(\boldsymbol{\sigma})$ is deterministic. We assume $\|\mathbf{W}\|_{\text{op}} \leq C$ for some constant C independent of (M, N) and that the empirical spectral distribution of $\mathbf{W}\mathbf{W}^\top$ converges weakly to a deterministic probability measure μ . We define the limiting spectral measure of $\mathbf{W}^\top \mathbf{W}$ as

$$\tilde{\mu} \triangleq \delta\mu + (1 - \delta)\delta_{\{0\}}.$$

We assume μ is absolutely continuous with a Hölder continuous density and has compact support $\text{supp}(\mu) \subset \mathbb{R}_+$.

- (d) Let $\text{sp}(\mathbf{W}\mathbf{W}^\top)$ denote the sets of empirical eigenvalues of $\mathbf{W}\mathbf{W}^\top$. We assume that the empirical spectrum of $\mathbf{W}\mathbf{W}^\top$ is asymptotically contained in any small neighborhood of the limiting support:

$$\lim_{M \rightarrow \infty} \sup_{\lambda \in \text{sp}(\mathbf{W}\mathbf{W}^\top)} d(\lambda, \text{supp}(\mu)) \xrightarrow{\text{a.s.}} 0,$$

where $d(\lambda, S) \stackrel{\text{def}}{=} \inf_{x \in S} |\lambda - x|$ denotes the distance from a point λ to a set S .

2.2 High-dimensional asymptotics and notation

We next specify the mode of convergence used throughout to describe limits of empirical distributions of vector entries and asymptotic equivalence of high-dimensional random vectors. More information can be found in [18, 30, 61].

Definition 1 (Wasserstein convergence). Let $(\mathbf{v}_1, \dots, \mathbf{v}_\ell)$ be a collection of random vectors in $\mathbb{R}^d$. We say that the empirical distribution of the entries of $(\mathbf{v}_1, \dots, \mathbf{v}_\ell)$ converges to random variables $(\mathbf{V}_1, \dots, \mathbf{V}_\ell)$ in the Wasserstein space of order p if for any test function $h : \mathbb{R}^\ell \rightarrow \mathbb{R}$ satisfying

$$|h(\mathbf{v}) - h(\mathbf{v}')| \leq L(1 + \|\mathbf{v}\|^{p-1} + \|\mathbf{v}'\|^{p-1})\|\mathbf{v} - \mathbf{v}'\|, \quad \forall \mathbf{v}, \mathbf{v}' \in \mathbb{R}^\ell, \quad (2)$$

for some $L < \infty$, we have

$$\frac{1}{d} \sum_{i=1}^d h(v_1[i], \dots, v_\ell[i]) \xrightarrow{\text{a.s.}} \mathbb{E}[h(\mathbf{V}_1, \dots, \mathbf{V}_\ell)], \quad \text{as } d \rightarrow \infty.$$

We denote convergence in this sense by $(\mathbf{v}_1, \dots, \mathbf{v}_\ell) \xrightarrow{W_p} (\mathbf{V}_1, \dots, \mathbf{V}_\ell)$. If this convergence holds for all $p \geq 1$, we write $(\mathbf{v}_1, \dots, \mathbf{v}_\ell) \xrightarrow{W} (\mathbf{V}_1, \dots, \mathbf{V}_\ell)$.

Following [44], we also introduce a notion of asymptotic equivalence for high-dimensional vectors.

Definition 2 (Asymptotic equivalence). Two d -dimensional random vectors $\mathbf{u}$ and $\mathbf{v}$ are asymptotically equivalent if

$$\frac{\|\mathbf{u} - \mathbf{v}\|^2}{d} \xrightarrow{\text{a.s.}} 0 \quad \text{as } d \rightarrow \infty.$$

We denote this by $\mathbf{u} \stackrel{d}{\rightrightarrows} \mathbf{v}$.

2.3 Stieltjes Transform, Hilbert Transform, and C-transform

The analysis of the singular spectrum of the rectangular spiked model is most naturally expressed in terms of transforms of spectral measures. We recall the Stieltjes transform and associated Hilbert transform for finite signed measures, and define the C-transform that will appear in the master equation governing outlier eigenvalues.

A *signed measure* χ on the Borel subsets of $\mathbb{R}$ generalizes a measure by allowing both positive and negative values. There is a well-known bijection between finite signed measures on $\mathbb{R}$ and right continuous functions of bounded variation [62, Proposition 4.4.3].

Definition 3 (Stieltjes transform of a finite signed measure). Let χ be a finite signed measure on $\mathbb{R}$, and let $F_\chi(x) \stackrel{\text{def}}{=} \chi((-\infty, x])$ be its right continuous function of bounded variation. The Stieltjes transform $\mathcal{S}_\chi$ is defined for $z \in \mathbb{C} \setminus \text{supp}(\chi)$ by

$$\mathcal{S}_\chi(z) \stackrel{\text{def}}{=} \int_{\mathbb{R}} \frac{1}{z - \lambda} dF_\chi(\lambda), \quad z \in \mathbb{C} \setminus \text{supp}(\chi). \quad (3)$$

This transform uniquely determines the measure χ (and hence F_χ); see, e.g., [8, Theorem B.8]. The Hilbert transform of χ , denoted by $\mathcal{H}_\chi$, is defined by the Cauchy principal value integral

$$\mathcal{H}_\chi(x) \stackrel{\text{def}}{=} \frac{1}{\pi} \text{P.V.} \int_{\mathbb{R}} \frac{1}{x - \lambda} dF_\chi(\lambda), \quad x \in \mathbb{R}. \quad (4)$$

When χ is absolutely continuous with respect to Lebesgue measure with a Hölder continuous density, the integral (4) exists and $\mathcal{H}_\chi$ is itself Hölder continuous [63, Section 2.1].

In what follows, we write $\mathcal{S}_\mu$ and $\mathcal{H}$ for the Stieltjes and Hilbert transforms associated with the spectral measure μ in Assumption 1. The Stieltjes and Hilbert transforms are linked on the real axis, and this relationship will allow us to express the densities of certain limiting measures in closed form.

The C-transform plays a central role in the master equation governing the emergence of outlier eigenvalues. Its structure is closely related to the D -transform appearing in the analysis of deformed random matrix models; see, e.g., [39, Section 2.3].

Definition 4 (C-transform). Let μ be the limiting spectral measure of the noise matrix in Assumption 1. The C-transform associated with μ is defined for $z \in \mathbb{C} \setminus \text{supp}(\mu)$ by

$$\mathcal{C}(z) \stackrel{\text{def}}{=} z \mathcal{S}_\mu(z) \left[\delta \mathcal{S}_\mu(z) + \frac{1 - \delta}{z} \right], \quad (5)$$

where $\mathcal{S}_\mu$ denotes the Stieltjes transform of μ .

2.4 Signal-eigenspace Spectral Measures

We now introduce spectral measures that project the eigenspace of the observed matrix onto the spans of the true signals $\mathbf{u}_*$ and $\mathbf{v}_*$. These measures encode how signal energy is distributed across the empirical spectrum and will be central in our state evolution analysis of OAMP.

Definition 5 (Signal-eigenspace spectral measures). Let $(\lambda_i(\cdot), \mathbf{u}_i(\cdot))$ be the eigenvalue/eigenvector pairs of a symmetric matrix.

- (a) **Parallel spectral measures.** To analyze quadratic forms in the signal directions, we define the parallel spectral measures $\nu_{M,1}$ and $\nu_{N,2}$ as the weighted empirical measures

$$\begin{aligned}\nu_{M,1} &\stackrel{\text{def}}{=} \frac{1}{M} \sum_{i=1}^M \langle \mathbf{u}_i(\mathbf{Y}\mathbf{Y}^\top), \mathbf{u}_* \rangle^2 \delta_{\lambda_i(\mathbf{Y}\mathbf{Y}^\top)}, \\ \nu_{N,2} &\stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N \langle \mathbf{u}_i(\mathbf{Y}^\top \mathbf{Y}), \mathbf{v}_* \rangle^2 \delta_{\lambda_i(\mathbf{Y}^\top \mathbf{Y})}.\end{aligned}$$

- (b) **Cross spectral measure.** To analyze bilinear forms coupling the two signal directions, we construct the symmetric dilation

$$\hat{\mathbf{Y}} \stackrel{\text{def}}{=} \begin{bmatrix} 0 & \mathbf{Y} \\ \mathbf{Y}^\top & 0 \end{bmatrix} \in \mathbb{R}^{L \times L}, \quad L = M + N,$$

and define the cross spectral measure $\nu_{L,3}$ by

$$\nu_{L,3} \stackrel{\text{def}}{=} \frac{1}{L} \sum_{i=1}^L \langle \mathbf{u}_i(\hat{\mathbf{Y}}), \hat{\mathbf{u}}_* \rangle \langle \mathbf{u}_i(\hat{\mathbf{Y}}), \hat{\mathbf{v}}_* \rangle \delta_{\lambda_i(\hat{\mathbf{Y}})},$$

where the zero-padded vectors are $\hat{\mathbf{u}}_* \stackrel{\text{def}}{=} [\mathbf{u}_*^\top, \mathbf{0}^\top]^\top$ and $\hat{\mathbf{v}}_* \stackrel{\text{def}}{=} [\mathbf{0}^\top, \mathbf{v}_*^\top]^\top$.

Shrinkage functions. In order to state the limiting characterization of the measures ν_i , we introduce shrinkage functions $\varphi_i : \mathbb{R} \rightarrow \mathbb{R}$ that describe their absolutely continuous components. Let $\mathcal{H}(\lambda)$ be the Hilbert transform of μ at λ . Using the Sokhotski–Plemelj formula [64], one can compute boundary values of $1 - \theta^2 \mathcal{C}(\lambda - i\epsilon)$ as $\epsilon \rightarrow 0^+$:

$$\lim_{\epsilon \rightarrow 0^+} |1 - \theta^2 \mathcal{C}(\lambda - i\epsilon)|^2 \tag{6a}$$

$$= \left\{ 1 - \delta \theta^2 \pi^2 \lambda \mathcal{H}^2(\lambda) + \delta \theta^2 \pi^2 \lambda \mu^2(\lambda) - (1 - \delta) \theta^2 \pi \mathcal{H}(\lambda) \right\}^2 \tag{6b}$$

$$+ \left\{ \pi \theta^2 \mu(\lambda) [(1 - \delta) + 2\delta \pi \lambda \mathcal{H}(\lambda)] \right\}^2. \tag{6c}$$

We then define the shrinkage functions $\varphi_1, \varphi_2, \varphi_3 : \mathbb{R} \rightarrow \mathbb{R}$ by

$$\varphi_1(\lambda) \stackrel{\text{def}}{=} \frac{1 + \delta \theta^2 \pi^2 \lambda (\mathcal{H}(\lambda)^2 + \mu(\lambda)^2)}{\lim_{\epsilon \rightarrow 0^+} |1 - \theta^2 \mathcal{C}(\lambda - i\epsilon)|^2}, \tag{7a}$$

$$\varphi_3(\lambda) \stackrel{\text{def}}{=} \frac{\theta (1 - \delta + 2\delta \pi \lambda \mathcal{H}(\lambda))}{\lim_{\epsilon \rightarrow 0^+} |1 - \theta^2 \mathcal{C}(\lambda - i\epsilon)|^2} \cdot \mathbf{1}_{\{\lambda \neq 0\}}, \tag{7b}$$

$$\varphi_2(\lambda) \stackrel{\text{def}}{=} \begin{cases} \delta \varphi_1(\lambda) + \frac{\theta(1 - \delta)}{\lambda} \varphi_3(\lambda), & \lambda > 0, \\ \frac{\delta}{1 - \theta^2(1 - \delta)\pi \mathcal{H}(0)}, & \lambda = 0. \end{cases} \tag{7c}$$

2.5 The Master Equation and Outlier Location

The emergence of outlier singular-values in the rectangular spiked model is governed by the *master equation*

$$\Gamma(z) \stackrel{\text{def}}{=} 1 - \theta^2 \mathcal{C}(z) = 0, \quad z \in \mathbb{C} \setminus \text{supp}(\mu), \tag{8}$$

where μ is the limiting spectral measure of $\mathbf{W}\mathbf{W}^\top$ and $\mathcal{C}$ is the C-transform defined in Definition 4. Real solutions of (8) outside the support of μ correspond to isolated spectral components, while the analytic properties of Γ determine where such solutions may occur and ensure that any such solution is isolated and simple. We therefore begin by establishing the basic analytic properties of Γ .

Lemma 1 (Analytic structure and zeros of the master equation). *Under Assumption 1(c), Γ has the following properties:*

1. Γ is holomorphic on $\mathbb{C} \setminus \text{supp}(\mu)$ and not identically zero. In particular, $\Gamma(z) \rightarrow 1$ as $|z| \rightarrow \infty$.
2. Every real zero of Γ lying in $\mathbb{R} \setminus \text{supp}(\mu)$ is isolated and simple; in particular, if $\Gamma(\lambda) = 0$ for some $\lambda \in \mathbb{R} \setminus \text{supp}(\mu)$, then $\Gamma'(\lambda) \neq 0$.
3. $\Gamma(\lambda) \neq 0$ for all λ in the interior of $\text{supp}(\mu)$, and moreover $\Gamma(0) \neq 0$. Consequently, any real solution of the master equation that produces an isolated spectral component must lie in $\mathbb{R} \setminus (\text{supp}(\mu) \cup \{0\})$.
4. Let $\lambda \in \mathbb{R} \setminus \text{supp}(\mu)$ such that $\Gamma(\lambda) = 0$. Then,

$$\text{sign}(\Gamma'(\lambda)) = \text{sign}(\mathcal{S}_\mu(\lambda))$$

Equivalently, $\text{sign}(\mathcal{C}'(\lambda)) = -\text{sign}(\mathcal{S}_\mu(\lambda))$. In particular, $\Gamma'(\lambda) \neq 0$.

Proof. See Appendix A.1. □

As an immediate consequence, any real solution of the master equation (8) that generates an isolated spectral component must lie in $\mathbb{R} \setminus (\text{supp}(\mu) \cup 0)$ and corresponds to a simple zero of Γ . These properties will be used repeatedly in the characterization of the singular parts of the limiting spectral measures.

2.6 Limiting Spectral Measures and Outlier Behavior

We are now ready to state the limiting behavior of the signal-eigenspace spectral measures, and to connect them with the isolated outlier eigenvalues and singular vectors of the rectangular spiked model.

Lemma 2. *Under Assumption 1, in the rectangular spiked model (1), the following hold.*

1. **Weak convergence.** *The measures $\nu_{M,1}, \nu_{N,2}$ and $\nu_{L,3}$ from Definition 5 converge weakly almost surely to deterministic, compactly supported measures ν_1, ν_2, ν_3 , respectively. The limiting measures ν_1 and ν_2 are probability measures on $\mathbb{R}_+$, and ν_3 is a finite signed measure on $\mathbb{R}$.*
2. **Stieltjes transforms.** *For $z \in \mathbb{C} \setminus \mathbb{R}$, their Stieltjes transforms are*

$$\mathcal{S}_{\nu_1}(z) = \frac{\mathcal{S}_\mu(z)}{1 - \theta^2 \mathcal{C}(z)}, \quad \mathcal{S}_{\nu_2}(z) = \frac{\delta \mathcal{S}_\mu(z) + \frac{1-\delta}{z}}{1 - \theta^2 \mathcal{C}(z)}, \quad \mathcal{S}_{\nu_3}(z) = \frac{\sqrt{\delta}}{1 + \delta} \cdot \frac{\theta \mathcal{C}(z^2)}{1 - \theta^2 \mathcal{C}(z^2)}. \quad (9)$$

3. **Absolutely continuous parts.** *Let the Lebesgue decomposition of each measure be $\nu_i = \nu_i^\parallel + \nu_i^\perp$ for $i \in \{1, 2, 3\}$, where $\nu_i^\parallel$ is the absolutely continuous component and $\nu_i^\perp$ is the singular component. Let $\varphi_1, \varphi_2, \varphi_3$ be the shrinkage functions in (7). Then*

$$\frac{d\nu_1^\parallel}{d\lambda} = \mu(\lambda)\varphi_1(\lambda), \quad \frac{d\nu_2^\parallel}{d\lambda} = \mu(\lambda)\varphi_2(\lambda), \quad \frac{d\nu_3^\parallel}{d\sigma} = \frac{\sqrt{\delta}}{1 + \delta} \text{sign}(\sigma) \mu(\sigma^2)\varphi_3(\sigma^2). \quad (10)$$

4. **Singular parts.** *Let*

$$\mathcal{K}^* \stackrel{\text{def}}{=} \{\lambda \in \mathbb{R} \setminus \text{supp}(\mu) : \Gamma(\lambda) = 0\},$$

and assume $\mathcal{K}^*$ is finite. Then the singular components are purely atomic and admit the representations

$$\nu_1^\perp = \sum_{\lambda_* \in \mathcal{K}^*} \nu_1(\{\lambda_*\}) \delta_{\lambda_*}, \quad (11a)$$

$$\nu_2^\perp = \sum_{\lambda_* \in \mathcal{K}^*} \nu_2(\{\lambda_*\}) \delta_{\lambda_*} + \mathbf{1}_{\{\delta < 1\}} \nu_2(\{0\}) \delta_0, \quad (11b)$$

$$\nu_3^\perp = \sum_{\lambda_* \in \mathcal{K}^*} (\nu_3(\{\sigma_*\}) \delta_{\sigma_*} + \nu_3(\{-\sigma_*\}) \delta_{-\sigma_*}), \quad \sigma_* \stackrel{\text{def}}{=} \sqrt{\lambda_*}. \quad (11c)$$

Moreover, for $\lambda_* \in \mathcal{K}^*$,

$$\nu_1(\{\lambda_*\}) = -\frac{\mathcal{S}_\mu(\lambda_*)}{\theta^2 \mathcal{C}'(\lambda_*)}, \quad \nu_2(\{\lambda_*\}) = -\frac{\delta \mathcal{S}_\mu(\lambda_*) + (1-\delta)/\lambda_*}{\theta^2 \mathcal{C}'(\lambda_*)},$$

and

$$\nu_3(\{\pm\sigma_*\}) = \mp \frac{\sqrt{\delta}}{1+\delta} \frac{1}{2\theta^3 \sigma_* \mathcal{C}'(\sigma_*^2)}.$$

If $\delta < 1$, ν_2 has an additional atom at 0 with mass

$$\nu_2(\{0\}) = \frac{1-\delta}{1-\theta^2(1-\delta)\pi\mathcal{H}(0)}.$$

Proof. See Appendix A.3. □

To connect the spectral measures in Definition 5 with the eigen-structure of the rectangular spiked model (1), we next characterize the limiting outlier eigenvalues and their associated overlaps. This extends the results of [39, Theorems 2.8–2.9] to the multi-outlier setting.

Proposition 1 (Outlier characterization). *Under Assumption 1, let $\mathcal{K}^*$ be the real zero set of the master equation $\Gamma(\lambda) = 0$ as in Lemma 2. Assume we are in a supercritical regime in which $\mathcal{K}^*$ is finite. Then the following hold.*

1. **Isolation of population outliers.** *There exists $\varepsilon > 0$ such that the intervals $I_k \stackrel{\text{def}}{=} (\lambda_k - \varepsilon, \lambda_k + \varepsilon)$, $\lambda_k \in \mathcal{K}^*$, are pairwise disjoint and satisfy*

$$I_k \cap \text{supp}(\mu) = \emptyset, \quad I_k \cap \mathcal{K}^* = \{\lambda_k\}, \quad \forall \lambda_k \in \mathcal{K}^*.$$

2. **Exact spectral separation.** *Let $\varepsilon > 0$ be as in item (1), and define*

$$\mathcal{K}_\varepsilon^* \stackrel{\text{def}}{=} \bigcup_{\lambda_k \in \mathcal{K}^*} (\lambda_k - \varepsilon, \lambda_k + \varepsilon), \quad \text{supp}_\varepsilon(\mu) \stackrel{\text{def}}{=} \{\lambda \in \mathbb{R} : d(\lambda, \text{supp}(\mu)) < \varepsilon\}.$$

Almost surely, there exists $M_0 < \infty$ such that for all $M \geq M_0$:

$$\text{sp}(\mathbf{W}\mathbf{W}^\top) \cap \mathcal{K}_\varepsilon^* = \emptyset, \tag{12}$$

$$|(\text{sp}(\mathbf{Y}\mathbf{Y}^\top) \cap (\lambda_k - \varepsilon, \lambda_k + \varepsilon))| = 1, \quad \forall \lambda_k \in \mathcal{K}^*, \tag{13}$$

$$\text{sp}(\mathbf{Y}\mathbf{Y}^\top) \subseteq \text{supp}_\varepsilon(\mu) \cup \mathcal{K}_\varepsilon^*. \tag{14}$$

In words, for all sufficiently large M , all eigenvalues of $\mathbf{Y}\mathbf{Y}^\top$ are within $\text{supp}_\varepsilon(\mu) \cup \mathcal{K}_\varepsilon^*$, and each outlier window $(\lambda_k - \varepsilon, \lambda_k + \varepsilon)$ contains exactly one eigenvalue.

3. **Convergence of empirical outliers.** *For each $\lambda_k \in \mathcal{K}^*$, let $\lambda_{k,M}$ denote the unique eigenvalue in $\text{sp}(\mathbf{Y}\mathbf{Y}^\top) \cap (\lambda_k - \varepsilon, \lambda_k + \varepsilon)$ (well-defined for all $M \geq M_0$ by (13)). Then*

$$\lambda_{k,M} \xrightarrow{a.s.} \lambda_k \quad \text{as } M \rightarrow \infty.$$

4. **Limiting overlaps.** *Let $(\sigma_{k,M}, \mathbf{u}_k(\mathbf{Y}), \mathbf{v}_k(\mathbf{Y}))$ be a singular value–vector triplet of $\mathbf{Y} \in \mathbb{R}^{M \times N}$ such that $\sigma_{k,M}^2 \rightarrow \lambda_k$, and write $\sigma_k \stackrel{\text{def}}{=} \sqrt{\lambda_k}$. Then*

$$\frac{1}{M} \langle \mathbf{u}_*, \mathbf{u}_k(\mathbf{Y}) \rangle^2 \xrightarrow{a.s.} \nu_1(\{\lambda_k\}), \tag{15}$$

$$\frac{1}{N} \langle \mathbf{v}_*, \mathbf{v}_k(\mathbf{Y}) \rangle^2 \xrightarrow{a.s.} \nu_2(\{\lambda_k\}), \tag{16}$$

$$\frac{1}{\sqrt{MN}} \langle \mathbf{u}_*, \mathbf{u}_k(\mathbf{Y}) \rangle \langle \mathbf{v}_*, \mathbf{v}_k(\mathbf{Y}) \rangle \xrightarrow{a.s.} 2 \frac{1+\delta}{\sqrt{\delta}} \nu_3(\{\sigma_k\}), \tag{17}$$

where ν_1, ν_2, ν_3 are the limiting spectral measures in Lemma 2.

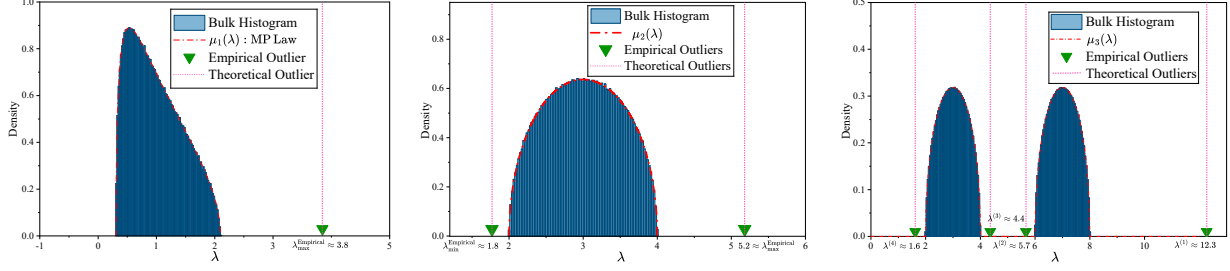


Figure 1: Spectral behavior of the rectangular spiked model in super critical θ -regime under different noise distributions. Left: Gaussian noise; the bulk follows the Marčenko–Pastur density $\mu_1(\lambda)$ and exhibits a single outlier. Center: Non-Gaussian noise with bulk $\mu_2(\lambda) = \frac{2}{\pi} \sqrt{(\lambda-2)(4-\lambda)} \mathbf{1}_{[2,4]}(\lambda)$, producing two outliers. Right: Non-Gaussian noise with bulk $\mu_3(\lambda) = \frac{1}{\pi} \sqrt{(\lambda-2)(4-\lambda)} \mathbf{1}_{[2,4]}(\lambda) + \frac{1}{\pi} \sqrt{(\lambda-6)(8-\lambda)} \mathbf{1}_{[6,8]}(\lambda)$, producing multiple outliers. Dashed vertical lines indicate the real roots of the master equation in Lemma 1.

Proof. See Appendix A.4. □

Remark 1 (Multiplicity of outliers and spectral behavior). In the rectangular rotationally invariant (RI) model, a single rank-one spike may give rise to *multiple* outlier singular values. This behavior can be understood through the analytic structure of the master equation $\mathcal{C}(\lambda) = 1/\theta^2$, where the associated C-transform $\mathcal{C}(\lambda)$ need not be monotone on $\mathbb{R} \setminus \text{supp}(\mu)$. The importance of possible non-monotonicity of such transforms has already been noted in the general theory of rectangular low-rank perturbations (e.g., in the work of Benaych-Georges and Nadakuditi [39]), and it naturally allows the outlier equation to admit multiple real solutions, each corresponding to a distinct outlier.

This phenomenon appears in two qualitatively different regimes, illustrated in Figure 1:

- *Single-interval bulk.* Even when the noise spectrum $\text{supp}(\mu)$ consists of a single connected interval, the effective rank-two structure of the perturbation $\mathbf{Y}\mathbf{Y}^\top$ can generate *two* distinct outliers, typically appearing on opposite sides of the bulk; see the center panel of Fig. 1.
- *Multi-interval bulk.* When $\text{supp}(\mu)$ consists of finitely many disjoint intervals, outliers typically emerge within the spectral gaps separating these intervals, as illustrated in the right panel of Fig. 1.

Remark 2 (MP law and absence of sub-bulk solutions). As noted in Remark 1, even when the noise spectrum consists of a single interval, a rank-one rectangular spike may generate *two* outliers, reflecting the effective rank-two structure of the perturbation $\mathbf{Y}\mathbf{Y}^\top$. In the classical spiked Marčenko–Pastur (MP) setting, however, the potential outlier below the lower edge of the MP law does not occur: when the spike is supercritical, the unique outlier lies strictly above the upper edge. See the left panel of Fig. 1 for illustration. While this conclusion follows immediately from the Marčenko–Pastur specialization of the general outlier equation, we are not aware of a reference where it is stated explicitly; for this reason, we record it as Lemma 8 (in Appendix F).

3 Orthogonal Approximate Message Passing Algorithms

This section introduces a family of Orthogonal Approximate Message Passing (OAMP) algorithms for rank-one rectangular matrix estimation. The construction relies on a set of spectral denoisers, iterate denoisers, and side information, together with *trace-free* and *divergence-free* conditions that ensure a closed-form state evolution description.

3.1 Orthogonal AMP for Rectangular Spiked Models

Definition 6 (OAMP algorithm). Given the observation matrix $\mathbf{Y} \in \mathbb{R}^{M \times N}$ and side information vectors $\mathbf{a} \in \mathbb{R}^{M \times k}$ and $\mathbf{b} \in \mathbb{R}^{N \times k}$, an OAMP algorithm generates iterates $(\mathbf{u}_t)_{t \geq 1}$ and $(\mathbf{v}_t)_{t \geq 1}$ through the updates

$$\mathbf{u}_t = F_t(\mathbf{Y}\mathbf{Y}^\top) f_t(\mathbf{u}_1, \dots, \mathbf{u}_{t-1}; \mathbf{a}) + \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} g_t(\mathbf{v}_1, \dots, \mathbf{v}_{t-1}; \mathbf{b}), \quad (18)$$

$$\mathbf{v}_t = G_t(\mathbf{Y}^\top \mathbf{Y}) g_t(\mathbf{v}_1, \dots, \mathbf{v}_{t-1}; \mathbf{b}) + \tilde{G}_t(\mathbf{Y}^\top \mathbf{Y}) \mathbf{Y}^\top f_t(\mathbf{u}_1, \dots, \mathbf{u}_{t-1}; \mathbf{a}), \quad (19)$$

for $t \geq 1$. Here $F_t, \tilde{F}_t, G_t, \tilde{G}_t$ are *spectral denoisers*: if $\mathbf{Y}\mathbf{Y}^\top = \mathbf{U} \text{diag}(\lambda_i) \mathbf{U}^\top$, then

$$F_t(\mathbf{Y}\mathbf{Y}^\top) = \mathbf{U} \text{diag}(F_t(\lambda_i)) \mathbf{U}^\top,$$

and similarly for the other spectral denoisers. The functions f_t, g_t are *iterate denoisers* applied entrywise to vector inputs. At iteration t , the estimates of the signals $\mathbf{u}_*$ and $\mathbf{v}_*$ are produced by postprocessing maps $\phi_{u,t}$ and $\phi_{v,t}$,

$$\hat{\mathbf{u}}_t = \phi_{u,t}(\mathbf{u}_1, \dots, \mathbf{u}_t; \mathbf{a}), \quad (20)$$

$$\hat{\mathbf{v}}_t = \phi_{v,t}(\mathbf{v}_1, \dots, \mathbf{v}_t; \mathbf{b}). \quad (21)$$

Regularity and orthogonality constraints. We require the spectral denoisers be dimension-independent and continuous on $\text{supp}(\mu)$ and F_t, G_t (but not $\tilde{F}_t$ and $\tilde{G}_t$) satisfy the *trace-free constraint*

$$\frac{1}{M} \text{Tr} F_t(\mathbf{Y}\mathbf{Y}^\top) \xrightarrow{\text{a.s.}} \langle F_t(\lambda) \rangle_\mu = 0, \quad (22)$$

$$\frac{1}{N} \text{Tr} G_t(\mathbf{Y}^\top \mathbf{Y}) \xrightarrow{\text{a.s.}} \langle G_t(\lambda) \rangle_{\tilde{\mu}} = 0. \quad (23)$$

The iterate denoisers f_t, g_t and the postprocessing functions must be dimension-independent, Lipschitz, and continuously differentiable. Furthermore, the sequence $(f_t)_{t \geq 1}$ and $(g_t)_{t \geq 1}$ satisfy the *divergence-free condition*

$$\mathbb{E}[\partial_s f_t(\mathbf{U}_1, \dots, \mathbf{U}_{t-1}; \mathbf{A})] = 0, \quad \mathbb{E}[\partial_s g_t(\mathbf{V}_1, \dots, \mathbf{V}_{t-1}; \mathbf{B})] = 0, \quad (24)$$

for every $s < t$, where the expectations are taken under the limiting joint laws $(\mathbf{U}_*, \mathbf{A}) \sim \pi_u$ and $(\mathbf{V}_*, \mathbf{B}) \sim \pi_v$.

The trace-free and divergence-free conditions in OAMP (and vector AMP) algorithms [22, 23] ensure that the effective noise in each update is asymptotically orthogonal to all past iterates, thereby removing the Onsager correction and enabling a valid state evolution characterization.

State Evolution Random Variables. Each OAMP algorithm is associated with a collection of state evolution random variables. It describes the joint asymptotic behavior of the signal, the iterates and the side information $\mathbf{a}, \mathbf{b}$. Let $\mathbf{Z}_{u,t}, \mathbf{Z}_{v,t}$ be Gaussian random variables, the distributions are given by

$$(\mathbf{U}_*, \mathbf{A}) \sim \pi_u, \quad \mathbf{U}_t = \mu_{u,t} \mathbf{U}_* + \mathbf{Z}_{u,t} \quad \forall t \in \mathbb{N}, \quad (25a)$$

$$(\mathbf{V}_*, \mathbf{B}) \sim \pi_v, \quad \mathbf{V}_t = \mu_{v,t} \mathbf{V}_* + \mathbf{Z}_{v,t} \quad \forall t \in \mathbb{N}. \quad (25b)$$

The random variables of the iterative denoisers with side information are formally defined as

$$\mathbf{F}_t = f_t(\mathbf{U}_1, \dots, \mathbf{U}_{t-1}; \mathbf{A}), \quad \mathbf{G}_t = g_t(\mathbf{V}_1, \dots, \mathbf{V}_{t-1}; \mathbf{B}). \quad (26a)$$

The alignment metrics between estimates and ground truth are characterized by

$$\alpha_t = \mathbb{E}[\mathbf{U}_*^\top \mathbf{F}_t], \quad \beta_t = \mathbb{E}[\mathbf{V}_*^\top \mathbf{G}_t]. \quad (26b)$$

Consequently, the residual covariances are defined as

$$\sigma_{f,st}^2 = \mathbb{E}[\mathbf{F}_s^\top \mathbf{F}_t] - \alpha_s \alpha_t, \quad \sigma_{g,st}^2 = \mathbb{E}[\mathbf{G}_s^\top \mathbf{G}_t] - \beta_s \beta_t. \quad (26c)$$

The coefficients in (25a) and (25b) are defined via the recursion

$$\mu_{u,t} \stackrel{\text{def}}{=} \alpha_t \langle F_t(\lambda) \rangle_{\nu_1} + \beta_t (1 + \delta^{-1}) \langle \sigma \tilde{F}_t(\sigma^2) \rangle_{\nu_3}, \quad (27a)$$

$$\mu_{v,t} \stackrel{\text{def}}{=} \beta_t \langle G_t(\lambda) \rangle_{\nu_2} + \alpha_t (1 + \delta) \langle \sigma \tilde{G}_t(\sigma^2) \rangle_{\nu_3}. \quad (27b)$$

The variables $(Z_{u,t}, Z_{v,t})_{t \in \mathbb{N}}$ are zero-mean jointly Gaussian random variables, sampled independently of the true signals. Their covariance matrix entries are given by the following recursions for any $s, t \in \mathbb{N}$

$$\begin{aligned} \Sigma_{u,st} \stackrel{\text{def}}{=} \mathbb{E}[Z_{u,s} Z_{u,t}] &= \alpha_s \alpha_t \langle F_s(\lambda) F_t(\lambda) \rangle_{\nu_1} + \beta_s \beta_t \delta^{-1} \langle \lambda \tilde{F}_s(\lambda) \tilde{F}_t(\lambda) \rangle_{\nu_3} - \mu_{u,s} \mu_{u,t} \\ &\quad + (1 + \delta^{-1}) \left(\alpha_s \beta_t \langle \sigma F_s(\sigma^2) \tilde{F}_t(\sigma^2) \rangle_{\nu_3} + \alpha_t \beta_s \langle \sigma F_t(\sigma^2) \tilde{F}_s(\sigma^2) \rangle_{\nu_3} \right) \\ &\quad + \sigma_{f,st}^2 \langle F_s(\lambda) F_t(\lambda) \rangle_{\tilde{\mu}} + \delta^{-1} \sigma_{g,st}^2 \langle \lambda \tilde{F}_s(\lambda) \tilde{F}_t(\lambda) \rangle_{\tilde{\mu}}, \end{aligned} \quad (28a)$$

$$\begin{aligned} \Sigma_{v,st} \stackrel{\text{def}}{=} \mathbb{E}[Z_{v,s} Z_{v,t}] &= \beta_s \beta_t \langle G_s(\lambda) G_t(\lambda) \rangle_{\nu_2} + \alpha_s \alpha_t \delta \langle \lambda \tilde{G}_s(\lambda) \tilde{G}_t(\lambda) \rangle_{\nu_1} - \mu_{v,s} \mu_{v,t} \\ &\quad + (1 + \delta) \left(\beta_s \alpha_t \langle \sigma G_s(\sigma^2) \tilde{G}_t(\sigma^2) \rangle_{\nu_3} + \beta_t \alpha_s \langle \sigma G_t(\sigma^2) \tilde{G}_s(\sigma^2) \rangle_{\nu_3} \right) \\ &\quad + \sigma_{g,st}^2 \langle G_s(\lambda) G_t(\lambda) \rangle_{\tilde{\mu}} + \delta \sigma_{f,st}^2 \langle \lambda \tilde{G}_s(\lambda) \tilde{G}_t(\lambda) \rangle_{\tilde{\mu}}. \end{aligned} \quad (28b)$$

Our first main result is the following theorem on the state evolution of the proposed OAMP algorithm for spiked matrix models.

Theorem 1 (State evolution). *Consider the OAMP algorithm in Definition 6, and let the state evolution random variables be defined as in (25). Then for each fixed $t \in \mathbb{N}$,*

$$(\mathbf{u}_*, \mathbf{u}_1, \dots, \mathbf{u}_t; \mathbf{a}) \xrightarrow{W_2} (\mathbf{U}_*, \mathbf{U}_1, \dots, \mathbf{U}_t; \mathbf{A}), \quad (29)$$

$$(\mathbf{v}_*, \mathbf{v}_1, \dots, \mathbf{v}_t; \mathbf{b}) \xrightarrow{W_2} (\mathbf{V}_*, \mathbf{V}_1, \dots, \mathbf{V}_t; \mathbf{B}), \quad (30)$$

where the convergence is in the Wasserstein sense of Definition 1.

Proof. See Appendix C. □

3.2 The Optimal OAMP Algorithm

We now specialize the OAMP framework to derive an algorithm that achieves the Bayes-optimal performance predicted by state evolution. The resulting procedure uses MMSE-based scalar denoisers, optimal spectral shrinkage functions derived from the limiting spectral measures, and cosine similarity parameters that track alignment with the true signals.

Final algorithm. The optimal OAMP iterates $(\mathbf{u}_t^*)_{t \geq 1}$ and $(\mathbf{v}_t^*)_{t \geq 1}$ are defined using the squared cosine similarities $w_{1,t}, w_{2,t} \in [0, 1)$ and take the form

$$\mathbf{u}_t^* = \frac{1}{\sqrt{w_{1,t}}} \left[F_t^*(\mathbf{Y} \mathbf{Y}^\top) \bar{\phi}(\mathbf{u}_{t-1}^*; \mathbf{a} | w_{1,t-1}) + \tilde{F}_t^*(\mathbf{Y} \mathbf{Y}^\top) \mathbf{Y} \bar{\phi}(\mathbf{v}_{t-1}^*; \mathbf{b} | w_{2,t-1}) \right], \quad (31)$$

$$\mathbf{v}_t^* = \frac{1}{\sqrt{w_{2,t}}} \left[G_t^*(\mathbf{Y}^\top \mathbf{Y}) \bar{\phi}(\mathbf{v}_{t-1}^*; \mathbf{b} | w_{2,t-1}) + \tilde{G}_t^*(\mathbf{Y}^\top \mathbf{Y}) \mathbf{Y}^\top \bar{\phi}(\mathbf{u}_{t-1}^*; \mathbf{a} | w_{1,t-1}) \right]. \quad (32)$$

The estimates at iteration t are

$$\hat{\mathbf{u}}_t^* = \phi(\mathbf{u}_t^*; \mathbf{a} | w_{1,t}), \quad \hat{\mathbf{v}}_t^* = \phi(\mathbf{v}_t^*; \mathbf{b} | w_{2,t}).$$

Remark 3. The prefactors $1/\sqrt{w_{1,t}}$ and $1/\sqrt{w_{2,t}}$ normalize the iterates so that the corresponding state-evolution variables satisfy $\mathbb{E}[(\mathbf{U}_t^*)^2] = \mathbb{E}[(\mathbf{V}_t^*)^2] = 1$. This is a convention.

1. Scalar MMSE and DMMSE denoisers

The scalar MMSE function ϕ and the divergence-free MMSE (DMMSE) function $\bar{\phi}$ follow [44, Definition 3]. For a scalar Gaussian channel

$$\mathbf{X} \mid (\mathbf{X}_*, \mathbf{C}) \sim \mathcal{N}(\sqrt{\omega} \mathbf{X}_*, 1 - \omega), \quad \omega \in [0, 1),$$

the denoisers are

$$\phi(x; c \mid \omega) \stackrel{\text{def}}{=} \mathbb{E}[\mathbf{X}_* \mid \mathbf{X} = x, \mathbf{C} = c], \quad (33)$$

$$\bar{\phi}(x; c \mid \omega) \stackrel{\text{def}}{=} \frac{\phi(x; c \mid \omega) - \frac{1}{\sqrt{1-\omega}} \mathbb{E}[Z \phi(\mathbf{X}; \mathbf{C} \mid \omega)] x}{1 - \frac{\sqrt{\omega}}{\sqrt{1-\omega}} \mathbb{E}[Z \phi(\mathbf{X}; \mathbf{C} \mid \omega)]}, \quad Z \sim \mathcal{N}(0, 1). \quad (34)$$

The DMMSE denoiser enforces the divergence-free condition required by OAMP.

Assumption 2. For every $\omega \in [0, 1)$, the MMSE estimator $\phi(\cdot \mid \omega)$ is continuously differentiable and Lipschitz.

2. Optimal spectral denoisers

Let the shrinkage functions $\varphi_1, \varphi_2, \varphi_3$ be given by (7a)–(7c). For SE parameters $\rho_{1,t}, \rho_{2,t} > 0$, define

$$P_t^*(\lambda) \stackrel{\text{def}}{=} \frac{\lambda(\rho_{2,t}\varphi_2(\lambda) + \delta)}{(\rho_{1,t}\varphi_1(\lambda) + 1)(\rho_{2,t}\varphi_2(\lambda) + \delta)\lambda - \rho_{1,t}\rho_{2,t}\varphi_3(\lambda)^2}, \quad (35)$$

$$\tilde{P}_t^*(\lambda) \stackrel{\text{def}}{=} \frac{\sqrt{\delta}\rho_{2,t}\varphi_3(\lambda)}{(\rho_{1,t}\varphi_1(\lambda) + 1)(\rho_{2,t}\varphi_2(\lambda) + \delta)\lambda - \rho_{1,t}\rho_{2,t}\varphi_3(\lambda)^2}, \quad (36)$$

$$Q_t^*(\lambda) \stackrel{\text{def}}{=} \frac{\delta\lambda(\rho_{1,t}\varphi_1(\lambda) + 1)}{(\rho_{1,t}\varphi_1(\lambda) + 1)(\rho_{2,t}\varphi_2(\lambda) + \delta)\lambda - \rho_{1,t}\rho_{2,t}\varphi_3(\lambda)^2}, \quad (37)$$

$$\tilde{Q}_t^*(\lambda) \stackrel{\text{def}}{=} \frac{\sqrt{\delta}\rho_{1,t}\varphi_3(\lambda)}{(\rho_{1,t}\varphi_1(\lambda) + 1)(\rho_{2,t}\varphi_2(\lambda) + \delta)\lambda - \rho_{1,t}\rho_{2,t}\varphi_3(\lambda)^2}. \quad (38)$$

The trace-free optimal matrix denoisers are then

$$F_t^*(\lambda) \stackrel{\text{def}}{=} \left(1 + \frac{1}{\rho_{1,t}}\right) \left(1 - \frac{P_t^*(\lambda)}{\langle P_t^* \rangle_\mu}\right), \quad \tilde{F}_t^*(\lambda) \stackrel{\text{def}}{=} \left(1 + \frac{1}{\rho_{2,t}}\right) \frac{\tilde{P}_t^*(\lambda)}{\langle \tilde{P}_t^* \rangle_\mu}, \quad (39)$$

$$G_t^*(\lambda) \stackrel{\text{def}}{=} \left(1 + \frac{1}{\rho_{2,t}}\right) \left(1 - \frac{Q_t^*(\lambda)}{\langle Q_t^* \rangle_{\tilde{\mu}}}\right), \quad \tilde{G}_t^*(\lambda) \stackrel{\text{def}}{=} \left(1 + \frac{1}{\rho_{1,t}}\right) \frac{\tilde{Q}_t^*(\lambda)}{\langle \tilde{Q}_t^* \rangle_{\tilde{\mu}}}. \quad (40)$$

3. Recursion for $(w_{i,t}, \rho_{i,t})$

Let $\text{mmse}_{\mathbf{X}}(w) \stackrel{\text{def}}{=} \mathbb{E}[(\mathbf{X}_* - \mathbb{E}[\mathbf{X}_* \mid \mathbf{X}])^2]$, where $\mathbf{X} = \sqrt{w}\mathbf{X}_* + \sqrt{1-w}\mathbf{Z}$. Then

$$\rho_{1,t} = \frac{1}{\text{mmse}_{\mathbf{U}}(w_{1,t-1})} - \frac{1}{1 - w_{1,t-1}}, \quad \rho_{2,t} = \frac{1}{\text{mmse}_{\mathbf{V}}(w_{2,t-1})} - \frac{1}{1 - w_{2,t-1}}, \quad (41)$$

$$w_{1,t} = 1 - \frac{1 - \langle P_t^* \rangle_\mu}{\langle P_t^* \rangle_\mu} \frac{1}{\rho_{1,t}}, \quad w_{2,t} = 1 - \frac{1 - \langle Q_t^* \rangle_{\tilde{\mu}}}{\langle Q_t^* \rangle_{\tilde{\mu}}} \frac{1}{\rho_{2,t}}. \quad (42)$$

The recursion is initialized via $w_{1,0}, w_{2,0} \in (0, 1)$.

Proposition 2 (State evolution: optimal OAMP). *Let $(\mathbf{U}_*, \mathbf{U}_t^*; \mathbf{A})$ and $(\mathbf{V}_*, \mathbf{V}_t^*; \mathbf{B})$ denote the state-evolution variables associated with the optimal OAMP iterates. Then:*

1. For $i \in \{1, 2\}$, we have $w_{i,t} \in (0, 1)$ and $\rho_{i,t} > 0$, and $(\mathbf{U}_*, \mathbf{U}_t^*)$ and $(\mathbf{V}_*, \mathbf{V}_t^*)$ form scalar Gaussian channels with similarities $w_{1,t}$ and $w_{2,t}$, respectively. Moreover,

$$\lim_{M \rightarrow \infty} \frac{\|\hat{\mathbf{u}}_t^* - \mathbf{u}_*\|^2}{M} \stackrel{\text{a.s.}}{=} \text{mmse}_{\mathbf{U}}(w_{1,t}), \quad \lim_{N \rightarrow \infty} \frac{\|\hat{\mathbf{v}}_t^* - \mathbf{v}_*\|^2}{N} \stackrel{\text{a.s.}}{=} \text{mmse}_{\mathbf{V}}(w_{2,t}).$$

2. The sequence $(\rho_{1,t}, \rho_{2,t}, w_{1,t}, w_{2,t})$ is monotone and converges to $(\rho_1^*, \rho_2^*, w_1^*, w_2^*) \in (0, \infty)^2 \times [0, 1]^2$ satisfying

$$\rho_1 = \frac{1}{\text{mmse}_U(w_1)} - \frac{1}{1 - w_1}, \quad \text{mmse}_U(w_1) = \frac{1}{\rho_1} (1 - \langle P^* \rangle_\mu), \quad (43a)$$

$$\rho_2 = \frac{1}{\text{mmse}_V(w_2)} - \frac{1}{1 - w_2}, \quad \text{mmse}_V(w_2) = \frac{1}{\rho_2} (1 - \langle Q^* \rangle_{\tilde{\mu}}). \quad (43b)$$

Consequently,

$$\lim_{t \rightarrow \infty} \lim_{M \rightarrow \infty} \frac{\|\hat{\mathbf{u}}_t^* - \mathbf{u}_*\|^2}{M} \stackrel{\text{a.s.}}{=} \text{mmse}_U(w_1^*), \quad \lim_{t \rightarrow \infty} \lim_{N \rightarrow \infty} \frac{\|\hat{\mathbf{v}}_t^* - \mathbf{v}_*\|^2}{N} \stackrel{\text{a.s.}}{=} \text{mmse}_V(w_2^*).$$

Proof. See Appendix E. □

Remark 4 (Connection with replica-symmetric Bayes-risk predictions). The fixed-point equations in (43) match the replica-symmetric characterization of the Bayes risk for the rectangular spiked rotationally-invariant model, whenever (43) has a unique solution. Further details will appear in a forthcoming paper [60].

3.3 Example: I.I.D Gaussian Noise

We now specialize our results to the noise matrix $\mathbf{W}$ with i.i.d. $\mathcal{N}(0, 1/N)$ entries. In this canonical setting, the limiting spectral measure μ of $\mathbf{W}\mathbf{W}^\top$ is the Marčenko–Pastur law [8] with aspect ratio $\delta \in (0, 1)$, whose density is

$$\mu_{\text{MP}}(\lambda) = \frac{\sqrt{(b_+ - \lambda)(\lambda - a_-)}}{2\pi\delta\lambda} \mathbf{1}_{[a_-, b_+]}(\lambda), \quad a_- \stackrel{\text{def}}{=} (1 - \sqrt{\delta})^2, \quad b_+ \stackrel{\text{def}}{=} (1 + \sqrt{\delta})^2.$$

As an application of Proposition 1, a detailed spectral analysis in such I.I.D. Gaussian noise model, which derives the phase transition and the outlier location, is provided in Appendix F.1.

We demonstrate that for this model, the fixed-point equations (43) governing our optimal OAMP algorithm coincides with that of the standard AMP [19, 28] up to a re-parameterization.

Proposition 3. *For the rectangular spiked model (1) with i.i.d. Gaussian noise matrix, the fixed point equations (43) can be simplified to*

$$\frac{w_1}{1 - w_1} = \frac{\theta^2}{\delta} (1 - \text{mmse}(w_2)), \quad (44)$$

$$\frac{w_2}{1 - w_2} = \theta^2 (1 - \text{mmse}(w_1)). \quad (45)$$

Proof. See Appendix F. □

4 Optimal Spectral Estimation Under Multiple Outliers

In the absence of a nonzero mean or side information, a random initialization fails for the OAMP algorithm: its state evolution converges to a trivial fixed point, as observed previously in phase retrieval [55, 65] and spiked models [28, 42, 59]. A spectral initialization is therefore required to produce a nontrivial estimate.

In this section, we study spectral estimation for the rectangular spiked model. As detailed in Remark 1 and Figure 1, a single rank-one signal in this setting typically generates *multiple* informative outlier singular values. In such regimes, relying solely on the leading singular vector (standard PCA) is suboptimal because it discards the signal energy carried by secondary outliers. Prior work [51] notes this phenomenon but does not provide an optimal method for combining the outlier components. Here, we develop a data-driven estimator that aggregates the informative outliers optimally under mild non-Gaussian assumptions on the signal.

4.1 Optimal Oracle Spectral Estimators

As established in Proposition 1, each outlying singular vector of $\mathbf{Y}$ retains a nonvanishing asymptotic correlation with the true signal directions. In the multiple-outlier regime, it is therefore natural to consider linear combinations of all informative components rather than relying on a single leading singular vector. This subsection characterizes the optimal such combination.

Let the singular-value decomposition of $\mathbf{Y}$ be

$$\mathbf{Y} = \sum_{i=1}^M \sigma_{i,M}(\mathbf{Y}) \mathbf{u}_i(\mathbf{Y}) \mathbf{v}_i(\mathbf{Y})^\top, \quad \sigma_{1,M} \geq \cdots \geq \sigma_{M,M},$$

and define $\lambda_{i,M} \triangleq \sigma_{i,M}(\mathbf{Y})^2$. Under Proposition 1 (supercritical θ), let μ denote the limiting spectral distribution of $\mathbf{W}\mathbf{W}^\top$, and let $\mathcal{K}^*$ be the finite set of population outliers.

Choose $\varepsilon > 0$ small enough so that the ε -neighborhood of points in $\mathcal{K}^*$ are disjoint and lie outside $\text{supp}(\mu)$; denote this union by $\mathcal{K}_\varepsilon^*$. We then define the *empirical outlier index set*

$$\mathcal{I}_M \stackrel{\text{def}}{=} \{i : \lambda_{i,M} \in \text{sp}(\mathbf{Y}\mathbf{Y}^\top) \cap \mathcal{K}_\varepsilon^*\}. \quad (46)$$

As ensured by Proposition 1, Claim (2), for all sufficiently large M , the empirical outliers in $\mathcal{I}_M$ correspond one-to-one with the population outliers in $\mathcal{K}^*$. With these informative components reliably identified, we consider linear spectral estimators supported on $\mathcal{I}_M$:

$$\mathbf{u}_{\text{PCA}}(\mathbf{c}_u) \stackrel{\text{def}}{=} \sqrt{M} \sum_{i \in \mathcal{I}_M} c_{u,i} \mathbf{u}_i(\mathbf{Y}), \quad \mathbf{v}_{\text{PCA}}(\mathbf{c}_v) \stackrel{\text{def}}{=} \sqrt{N} \sum_{i \in \mathcal{I}_M} c_{v,i} \mathbf{v}_i(\mathbf{Y}), \quad (47)$$

where $\mathbf{c}_u, \mathbf{c}_v \in \mathbb{R}^{|\mathcal{I}_M|}$ denote the combination coefficients. Since each outlying singular vector carries nonvanishing alignment with the true signal, an appropriate linear combination may improve the overall directional accuracy compared to using any single component.

The next proposition characterizes the *oracle* asymptotic squared cosine similarity achievable by this class of estimators, which equals the projection of the true signal onto the outlier eigenspace. The optimal coefficients attaining this limit depend on the unknown signal and are therefore not implementable in practice, but the result serves as the fundamental performance benchmark for all linear spectral methods.

Proposition 4. *Consider the class of estimators in (47). For any $\mathbf{c}_u, \mathbf{c}_v \in \mathbb{R}^{|\mathcal{I}_M|}$, almost surely,*

$$\lim_{M \rightarrow \infty} \frac{\langle \mathbf{u}_{\text{PCA}}(\mathbf{c}_u), \mathbf{u}_* \rangle^2}{\|\mathbf{u}_{\text{PCA}}(\mathbf{c}_u)\|^2 \|\mathbf{u}_*\|^2} \leq \sum_{\lambda_i \in \mathcal{K}^*} \nu_1(\{\lambda_i\}), \quad \lim_{N \rightarrow \infty} \frac{\langle \mathbf{v}_{\text{PCA}}(\mathbf{c}_v), \mathbf{v}_* \rangle^2}{\|\mathbf{v}_{\text{PCA}}(\mathbf{c}_v)\|^2 \|\mathbf{v}_*\|^2} \leq \sum_{\lambda_i \in \mathcal{K}^*} \nu_2(\{\lambda_i\}), \quad (48)$$

where $\nu_1(\{\lambda_k\})$ and $\nu_2(\{\lambda_k\})$ are defined in Lemma 2, Claim (4). Moreover, these upper bounds are asymptotically attained by the oracle combinations

$$\mathbf{u}_{\text{ora}}^* \stackrel{\text{def}}{=} \sqrt{M} \sum_{i \in \mathcal{I}_M} \langle \mathbf{u}_*, \mathbf{u}_i(\mathbf{Y}) \rangle \mathbf{u}_i(\mathbf{Y}), \quad (49)$$

$$\mathbf{v}_{\text{ora}}^* \stackrel{\text{def}}{=} \sqrt{N} \sum_{i \in \mathcal{I}_M} \langle \mathbf{v}_*, \mathbf{v}_i(\mathbf{Y}) \rangle \mathbf{v}_i(\mathbf{Y}). \quad (50)$$

Proof. See Appendix G.1. □

Remark 5 (Connection to RIE estimators [47]). Our construction of optimal spectral estimators is structurally related to the *rotationally invariant estimator* (RIE) framework developed for extensive-rank matrix denoising in [47] and for rectangular models in [48]. In both settings, one first characterizes an oracle estimator and then constructs a data-driven procedure that asymptotically attains the oracle performance.

There are, however, important differences. RIE operates in the extensive rank regime, where the signal information is distributed across the whole spectrum and the optimal estimator applies an eigenvalue-dependent shrinkage to all singular values. In contrast, our model is rank one, and all informative content is concentrated in a finite number of outlier singular components; optimal estimation thus requires combining only

these outliers. A second distinction concerns the estimation objective. Whereas RIE aims to reconstruct the underlying low-rank matrix, our goal is to recover the rank-one signal vectors. In this setting, the optimal linear combination of the outlier components involves relative signs that cannot be inferred from random matrix theory alone. As a result, accurate aggregation of multiple outliers requires an explicit sign-resolution procedure, addressed in Section 4.2.

4.2 Data-Driven Optimal Linear Spectral Estimators

The oracle estimator in (49) achieves the linear-spectral performance bound of Proposition 4, but its coefficients depend on the unknown signal through the outlier-signal overlaps. Thus Proposition 4 provides an *oracle* benchmark for what any linear spectral method based solely on $\mathbf{Y}$ can achieve. In this subsection, we construct a data-driven estimator that asymptotically attains this benchmark. As a first step, we derive a signal-plus-noise limit law for each outlier direction.

Proposition 5. *Under the assumptions of Proposition 1, let $\mathcal{K}^*$ denote the finite set of outlier eigenvalues. Let $\mathcal{K}^* = \{\lambda_1, \dots, \lambda_K\}$ and $\sigma_k = \sqrt{\lambda_k}$ for $1 \leq k \leq K$. For each $k \in \{1, \dots, K\}$ and all sufficiently large M , let $\lambda_{k,M}$ be the empirical outlier associated with $\lambda_k \in \mathcal{K}^*$ as in Proposition 1, and $\mathbf{u}_k(\mathbf{Y}), \mathbf{v}_k(\mathbf{Y})$ the corresponding left and right singular vectors of unit norm. As $M, N \rightarrow \infty$, we have the joint convergence*

$$(\langle \mathbf{u}_*, \mathbf{u}_1(\mathbf{Y}) \rangle \mathbf{u}_1(\mathbf{Y}), \dots, \langle \mathbf{u}_*, \mathbf{u}_K(\mathbf{Y}) \rangle \mathbf{u}_K(\mathbf{Y})) \xrightarrow{W} (\mathbf{U}_1^{\text{OUT}}, \dots, \mathbf{U}_K^{\text{OUT}})^\top, \quad (51)$$

$$(\langle \mathbf{v}_*, \mathbf{v}_1(\mathbf{Y}) \rangle \mathbf{v}_1(\mathbf{Y}), \dots, \langle \mathbf{v}_*, \mathbf{v}_K(\mathbf{Y}) \rangle \mathbf{v}_K(\mathbf{Y})) \xrightarrow{W} (\mathbf{V}_1^{\text{OUT}}, \dots, \mathbf{V}_K^{\text{OUT}})^\top, \quad (52)$$

where the random variables appearing on the RHS satisfy the following:

1 **Signal-plus-noise decomposition.** For every $k \in \{1, \dots, K\}$ we have

$$\mathbf{U}_k^{\text{OUT}} = \nu_1(\{\lambda_k\}) \mathbf{U}_* + \sqrt{\nu_1(\{\lambda_k\}) - \nu_1^2(\{\lambda_k\})} \mathbf{Z}_{u,k}, \quad (53a)$$

$$\mathbf{V}_k^{\text{OUT}} = \nu_2(\{\lambda_k\}) \mathbf{V}_* + \sqrt{\nu_2(\{\lambda_k\}) - \nu_2^2(\{\lambda_k\})} \mathbf{Z}_{v,k}, \quad (53b)$$

where $(\mathbf{U}_*, \mathbf{V}_*)$ are the limiting signal distributions and $\{\mathbf{Z}_{u,k}\}_{k=1}^K, \{\mathbf{Z}_{v,k}\}_{k=1}^K$ are Gaussian noise variables satisfying

$$(\mathbf{Z}_{u,1}, \dots, \mathbf{Z}_{u,K}) \perp \mathbf{U}_*, \quad (\mathbf{Z}_{v,1}, \dots, \mathbf{Z}_{v,K}) \perp \mathbf{V}_*.$$

2 **Gaussian noise and covariance.** The vectors $(\mathbf{Z}_{u,1}, \dots, \mathbf{Z}_{u,K})$ and $(\mathbf{Z}_{v,1}, \dots, \mathbf{Z}_{v,K})$ are centered jointly Gaussian. For each $k \in \{1, \dots, K\}$,

$$\mathbb{E}[\mathbf{Z}_{u,k}] = \mathbb{E}[\mathbf{Z}_{v,k}] = 0, \quad \mathbb{E}[\mathbf{Z}_{u,k}^2] = \mathbb{E}[\mathbf{Z}_{v,k}^2] = 1, \quad (54)$$

and for all $1 \leq k < \ell \leq K$,

$$\mathbb{E}[\mathbf{Z}_{u,k} \mathbf{Z}_{u,\ell}] = -\frac{\nu_1(\{\lambda_k\}) \nu_1(\{\lambda_\ell\})}{\sqrt{\nu_1(\{\lambda_k\}) - \nu_1^2(\{\lambda_k\})} \sqrt{\nu_1(\{\lambda_\ell\}) - \nu_1^2(\{\lambda_\ell\})}}, \quad (55)$$

$$\mathbb{E}[\mathbf{Z}_{v,k} \mathbf{Z}_{v,\ell}] = -\frac{\nu_2(\{\lambda_k\}) \nu_2(\{\lambda_\ell\})}{\sqrt{\nu_2(\{\lambda_k\}) - \nu_2^2(\{\lambda_k\})} \sqrt{\nu_2(\{\lambda_\ell\}) - \nu_2^2(\{\lambda_\ell\})}}, \quad (56)$$

In particular, the covariance matrices $\Sigma_u^{\text{OUT}} = (\mathbb{E}[\mathbf{Z}_{u,k} \mathbf{Z}_{u,\ell}])_{1 \leq k < \ell \leq K}$ and $\Sigma_v^{\text{OUT}} = (\mathbb{E}[\mathbf{Z}_{v,k} \mathbf{Z}_{v,\ell}])_{1 \leq k < \ell \leq K}$ are positive definite. For $i \in \{1, 2\}$, $\nu_i(\{\lambda_k\})$ denotes the point mass of the parallel spectral measure defined in Lemma 2.

Proof. See Appendix G.2. □

Remark 6 (Heuristic derivation of the outlier limit law). At a heuristic level, the decompositions (51)–(52) describe each projected outlier component as a deterministic multiple of the signal plus an asymptotically Gaussian noise term. In the i.i.d. Gaussian noise case, closely related eigenvector asymptotics are well known; see [28, Appendix C]. In our setting, a convenient starting point is the singular-equation for the outlying vector

$$\langle \mathbf{u}_*, \mathbf{u}_k \rangle \mathbf{u}_k = (\lambda_{k,M} \mathbf{I}_M - \mathbf{W} \mathbf{W}^\top)^{-1} \left(\frac{\theta \sigma_{k,M}}{\sqrt{MN}} \langle \mathbf{v}_*, \mathbf{v}_k \rangle \langle \mathbf{u}_*, \mathbf{u}_k \rangle \mathbf{u}_* + \frac{\theta}{\sqrt{MN}} \langle \mathbf{u}_*, \mathbf{u}_k \rangle^2 \mathbf{W} \mathbf{v}_* \right),$$

and the analogous equation for $\langle \mathbf{v}_*, \mathbf{v}_k \rangle \mathbf{v}_k$. The overlaps $\langle \mathbf{v}_*, \mathbf{v}_k \rangle \langle \mathbf{u}_*, \mathbf{u}_k \rangle$ and $\langle \mathbf{u}_*, \mathbf{u}_k \rangle^2$ can be shown, via standard resolvent and concentration arguments, to converge to deterministic limits determined by the spectral measures. Moreover, by Haar invariance and the independence between $\mathbf{W}$ and $(\mathbf{u}_*, \mathbf{v}_*)$, the full vector $\langle \mathbf{u}_*, \mathbf{u}_k \rangle \mathbf{u}_k$ can be analyzed using arguments similar to those in Appendix B, yielding convergence of the empirical law to the signal-plus-Gaussian form stated in Proposition 5. Its proof is provided in the appendix.

To construct a practical estimator that attains the oracle bound (49), it suffices to approximate the oracle linear combination of the informative outlier singular vectors. Since empirical singular vectors are defined only up to a global sign, we fix their orientation via a randomized sign convention (cf. [61, Remark 3.6]), which simplifies the theoretical analysis. For each $k \in \mathcal{I}_M$, let $\mathbf{u}_k(\mathbf{Y})$ and $\mathbf{v}_k(\mathbf{Y})$ be any choice of unit outlier singular vectors. Let $\{\xi_k\}_{k \in \mathcal{I}_M}$ be i.i.d. Rademacher random variables, independent of all other random elements in the model (1). Define the $M \times K$ matrix of randomized scaled singular vectors

$$\mathbf{U}^\# \stackrel{\text{def}}{=} [\mathbf{u}_1^\# \quad \cdots \quad \mathbf{u}_K^\#], \quad \mathbf{u}_k^\# \stackrel{\text{def}}{=} \sqrt{M} \xi_k \mathbf{u}_k(\mathbf{Y}), \quad k \in \mathcal{I}_M. \quad (57)$$

Proposition 5 shows that the associated asymptotic signal magnitudes $\sqrt{\nu_1(\{\lambda_k\})}$ and $\sqrt{\nu_2(\{\lambda_k\})}$ are deterministic functions of the noise spectrum (see Lemma 2). The only remaining unknowns are the *relative signs* of the overlaps

$$\{\langle \mathbf{u}_i^\#, \mathbf{u}_* \rangle\}_{i \in \mathcal{I}_M}, \quad \{\langle \mathbf{v}_i^\#, \mathbf{v}_* \rangle\}_{i \in \mathcal{I}_M},$$

which determine the alignment of the outlier directions with the signal. Let $s_i^u, s_i^v \in \{\pm 1\}$ denote sign variables (defined up to a global flip in each channel), and define the practical spectral estimators

$$\mathbf{u}_{\text{PCA}}^* \stackrel{\text{def}}{=} \sum_{i \in \mathcal{I}_M} s_i^u \sqrt{\nu_1(\{\lambda_i\})} \mathbf{u}_i^\#, \quad \mathbf{v}_{\text{PCA}}^* \stackrel{\text{def}}{=} \sum_{i \in \mathcal{I}_M} s_i^v \sqrt{\nu_2(\{\lambda_i\})} \mathbf{v}_i^\#. \quad (58a)$$

The next proposition shows that these estimators attain the oracle performance whenever the signs s_i^u, s_i^v are chosen consistently with the true overlaps (up to global sign flips). Its proof can be found in Appendix G.3.

Proposition 6 (Optimality via Consistent Sign Estimation). *Assume the setting of Proposition 1 with supercritical θ and Assumptions 1, and let $\mathbf{u}_{\text{PCA}}^*$ and $\mathbf{v}_{\text{PCA}}^*$ be defined in (58), with $\mathcal{I}_M$, λ_i , and $\nu_1(\{\lambda_i\}), \nu_2(\{\lambda_i\})$ as above. Suppose the signs satisfy*

$$s_i^u \stackrel{\text{a.s.}}{=} \text{sign} \langle \mathbf{u}_i^\#, \mathbf{u}_* \rangle, \quad s_i^v \stackrel{\text{a.s.}}{=} \text{sign} \langle \mathbf{v}_i^\#, \mathbf{v}_* \rangle, \quad \forall i \in \mathcal{I}_M,$$

up to a common global flip in each channel. Then

$$\lim_{M \rightarrow \infty} \frac{\langle \mathbf{u}_{\text{PCA}}^*, \mathbf{u}_* \rangle^2}{\|\mathbf{u}_{\text{PCA}}^*\|^2 \|\mathbf{u}_*\|^2} \stackrel{\text{a.s.}}{=} \sum_{\lambda_i \in \mathcal{K}^*} \nu_1(\{\lambda_i\}),$$

$$\lim_{N \rightarrow \infty} \frac{\langle \mathbf{v}_{\text{PCA}}^*, \mathbf{v}_* \rangle^2}{\|\mathbf{v}_{\text{PCA}}^*\|^2 \|\mathbf{v}_*\|^2} \stackrel{\text{a.s.}}{=} \sum_{\lambda_i \in \mathcal{K}^*} \nu_2(\{\lambda_i\}).$$

Thus, the final step is to find consistent estimators of the relative signs, which we address next.

4.3 Estimation of Relative Signs

This section addresses the estimation of the relative signs required for the spectral estimators in (58). We work under the setting of Proposition 5. Let $\mathcal{I}_M$ denote the set of empirical outlier indices in (46), with $K = |\mathcal{I}_M|$. We fix a reference index $r \in \mathcal{I}_M$ and encode the true relative sign by the vector $\mathbf{s}_{u,*}^R \in \{\pm 1\}^K$ defined, for $\ell \in [K]$, as

$$[\mathbf{s}_{u,*}^R]_\ell \stackrel{\text{def}}{=} \text{sign}(\langle \mathbf{u}_\ell^\#, \mathbf{u}_* \rangle) \text{sign}(\langle \mathbf{u}_r^\#, \mathbf{u}_* \rangle), \quad \text{so that } [\mathbf{s}_{u,*}^R]_r \equiv +1, \quad (59)$$

which is well-defined in the supercritical regime. Analogously we define $\mathbf{s}_{v,*}^R$ as the true relative sign vectors in the $\mathbf{v}$ -channel. We next characterize the row-wise limiting law of the randomized outlying singular vectors. It can be shown that conditioned on $\mathbf{s}_{u,*}^R \in \mathbb{R}^K$, the following convergence holds (see Appendix G.4.2)

$$(\mathbf{u}_1^\#, \dots, \mathbf{u}_K^\#) \xrightarrow{W} (\mathbf{U}_1^\#, \dots, \mathbf{U}_K^\#) \stackrel{\text{def}}{=} \mathbf{U}^\# \in \mathbb{R}^K, \quad (60)$$

with

$$\mathbf{U}_\ell^\# \stackrel{\text{def}}{=} [\mathbf{s}_{u,*}^R]_\ell \sqrt{\nu_1(\{\lambda_\ell\})} \mathbf{U}_* + \sqrt{1 - \nu_1(\{\lambda_\ell\})} \mathbf{Z}_\ell, \quad \ell \in \mathcal{I}_M, \quad (61)$$

with $\{\mathbf{Z}_\ell\}_{\ell \in \mathcal{I}_M}$ standard Gaussian independent of $\mathbf{U}_*$.

We consider two estimators of the relative signs: (i) a maximum likelihood estimator (MLE) based on the full prior, and (ii) a non-Gaussian moment contrast (NGMC) estimator which exploits higher order moments and is computationally simpler.

Proposition 7 (MLE for relative signs). *Let $\mathbf{s} \in \{\pm 1\}^K$ and $[\mathbf{s}]_r = 1$ be any fixed relative sign vector. Denote by P_s the joint probability density function of*

$$([\mathbf{s}]_\ell \sqrt{\nu_1(\{\lambda_\ell\})} \mathbf{U}_* + \sqrt{1 - \nu_1(\{\lambda_\ell\})} \mathbf{Z}_\ell)_{\ell \in \mathcal{I}_M}, \quad (62)$$

where $\mathbf{U}_*$ and $(\mathbf{Z}_\ell)_{\ell \in \mathcal{I}_M}$ are distributed as in (61). Denote the i -th row of the matrix $\mathbf{U}^\# \in \mathbb{R}^{M \times K}$ (61) by $\mathbf{U}_{i,:}^\#$. Let $\hat{\mathbf{s}}_u^{\text{MLE}}$ be the maximum likelihood estimator of $\mathbf{s}_{u,*}^R$

$$\hat{\mathbf{s}}_u^{\text{MLE}} \in \underset{\mathbf{s} \in \mathcal{S}_r}{\text{argmax}} \sum_{i=1}^M \log P_s(\mathbf{U}_{i,:}^\#), \quad (63)$$

and analogously $\hat{\mathbf{s}}_v^{\text{MLE}}$ the maximum likelihood estimator of $\mathbf{s}_{v,*}^R$. We have:

1. If either the law of $\mathbf{U}_*$ or $\mathbf{V}_*$ is not standard Gaussian, then

$$\hat{\mathbf{s}}_u^{\text{MLE}} \xrightarrow{a.s.} \mathbf{s}_{u,*}^R, \quad \hat{\mathbf{s}}_v^{\text{MLE}} \xrightarrow{a.s.} \mathbf{s}_{v,*}^R.$$

2. If both $\mathbf{U}_*$ and $\mathbf{V}_*$ are standard Gaussian, then $P_s \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_K)$ for any $\mathbf{s} \in \mathcal{S}_r$, and consistent estimation of the relative signs via MLE is impossible.

Proof. See Appendix G.4. □

Remark 7 (Well-posedness of the likelihood). The random vector $\mathbf{U}^\#$ is constructed by adding an independent Gaussian vector with non-degenerate covariance to the signal $\mathbf{U}_*$. Consequently, for any sign configuration $\mathbf{s}$, the joint law of $\mathbf{U}^\#$ is the convolution of the prior measure of $\mathbf{U}_*$ with a non-degenerate Gaussian distribution on $\mathbb{R}^K$. This ensures that the distribution admits a smooth, strictly positive density P_s with respect to the Lebesgue measure. Hence, the log-likelihood terms in (63) are well-defined.

Proposition 7 establishes that any non-Gaussianity in the prior $\mathbf{U}_*$ renders the relative signs identifiable, yielding an asymptotically consistent MLE. However, minimizing the objective over $\mathcal{S}_r$ can be computationally intensive when the prior lacks a closed-form Gaussian convolution. This motivates a simpler alternative that specifically exploits non-Gaussianity through an appropriate moment contrast, called the *non-Gaussian moment contrast* (NGMC) scheme.

Assumption 3 (Non-Gaussian even moment). There exists an even integer $k \geq 0$ such that

$$\mathbb{E}[\mathbf{U}_*^{k+2}] \neq (k+1)!!.$$

In other words, at least one even-order moment of $\mathbf{U}_*$ differs from the corresponding moment of a standard Gaussian random variable.

Proposition 8 (NGMC estimator for relative signs). *Assume the setting of Proposition 5, and suppose Assumption 3 holds. Let $k \geq 2$ be the smallest even integer admissible in Assumption 3, and let $f(x) \stackrel{\text{def}}{=} x^{k+1}$ be the corresponding entrywise moment-contrast function. Fix an arbitrary reference outlier index $r \in \mathcal{I}_M$, and for any other outlier $j \in \mathcal{I}_M \setminus \{r\}$ define*

$$\hat{s}_{u,j}^{\text{NGMC}} \stackrel{\text{def}}{=} \text{sign} \left(f(\mathbf{u}_r^\#)^\top \mathbf{u}_j^\# \right) \cdot \text{sign} \left(\mathbb{E}[\mathbf{U}_*^{k+2}] - (k+1)!! \right), \quad (64a)$$

$$\hat{s}_{v,j}^{\text{NGMC}} \stackrel{\text{def}}{=} \hat{s}_{u,j}^{\text{NGMC}} \cdot \text{sign} \left(\nu_3(\{\sigma_r\}) \nu_3(\{\sigma_j\}) \right). \quad (64b)$$

Then the NGMC estimators are consistent: $\hat{s}_{u,j}^{\text{NGMC}} \xrightarrow{a.s.} [\mathbf{s}_{u,*}^R]_j$ and $\hat{s}_{v,j}^{\text{NGMC}} \xrightarrow{a.s.} [\mathbf{s}_{v,*}^R]_j$.

Proof. See Appendix G.5. □

5 OAMP Algorithm with Spectral Initialization

The optimal spectral estimator developed in the previous section naturally suggests a way to initialize iterative algorithms. Here, we study OAMP when initialized using this spectral estimator, following the construction of Section 3.2.

5.1 Spectrally-Initialized Optimal OAMP

We use a tilde to distinguish the iterates of the spectrally initialized algorithm from those of the generic OAMP recursion in Section 3.2. Under the assumptions of Proposition 6, the initialization is given by the unit-variance normalized versions of the optimal spectral estimators (58). Specifically, at $t = 1$,

$$\tilde{\mathbf{u}}_1 = \left(\sum_{k \in \mathcal{I}_M} \nu_1(\{\lambda_k\}) \right)^{-1/2} \mathbf{u}_{\text{PCA}}^* \quad \text{and} \quad \tilde{\mathbf{v}}_1 = \left(\sum_{k \in \mathcal{I}_M} \nu_2(\{\lambda_k\}) \right)^{-1/2} \mathbf{v}_{\text{PCA}}^*, \quad (65)$$

where $\mathbf{u}_{\text{PCA}}^*$ and $\mathbf{v}_{\text{PCA}}^*$ are the optimal spectral estimates defined in (58).

For all subsequent iterations $t \geq 2$, and for fixed sign parameters $s_1, s_2 \in \{+1, -1\}$, the algorithm proceeds according to the standard optimal OAMP update rules in (31)–(32). The update rules are

$$\tilde{\mathbf{u}}_t^* = \frac{1}{\sqrt{\tilde{w}_{1,t}}} \left[F_t^*(\mathbf{Y}\mathbf{Y}^\top) \bar{\phi}(\tilde{\mathbf{u}}_{t-1}^* \mid \tilde{w}_{1,t-1}, s_1) + \tilde{F}_t^*(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \bar{\phi}(\tilde{\mathbf{v}}_{t-1}^* \mid \tilde{w}_{2,t-1}, s_2) \right], \quad (66a)$$

$$\tilde{\mathbf{v}}_t^* = \frac{1}{\sqrt{\tilde{w}_{2,t}}} \left[G_t^*(\mathbf{Y}^\top \mathbf{Y}) \bar{\phi}(\tilde{\mathbf{v}}_{t-1}^* \mid \tilde{w}_{2,t-1}, s_2) + \tilde{G}_t^*(\mathbf{Y}^\top \mathbf{Y}) \mathbf{Y}^\top \bar{\phi}(\tilde{\mathbf{u}}_{t-1}^* \mid \tilde{w}_{1,t-1}, s_1) \right], \quad (66b)$$

where $F_t^*, \tilde{F}_t^*, G_t^*, \tilde{G}_t^*$ (as functions of $\lambda, \tilde{\rho}_{1,t}, \tilde{\rho}_{2,t}$) denote the trace-free spectral matrix denoisers from (39)–(40), and $\bar{\phi}(\cdot \mid w, s)$ denotes the signed DMMSE denoiser associated with the scalar Gaussian channel

$$\mathbf{X} = s\sqrt{w}\mathbf{X}_* + \sqrt{1-w}\mathbf{Z}, \quad s \in \{\pm 1\},$$

defined by

$$\bar{\phi}(x \mid w, s) \stackrel{\text{def}}{=} \bar{\phi}(sx \mid w), \quad (67)$$

where $\bar{\phi}(\cdot \mid w)$ is the DMMSE denoiser for the standard scalar Gaussian channel in (33).

Update of state evolution parameters. The scalar parameters used in the denoisers are updated, for $t \geq 2$, by

$$\tilde{\rho}_{1,t} = \frac{1}{\text{mmse}_U(\tilde{w}_{1,t-1})} - \frac{1}{1 - \tilde{w}_{1,t-1}}, \quad \tilde{\rho}_{2,t} = \frac{1}{\text{mmse}_V(\tilde{w}_{2,t-1})} - \frac{1}{1 - \tilde{w}_{2,t-1}}, \quad (68a)$$

$$\tilde{w}_{1,t} = 1 - \frac{1 - \langle P_t^*(\lambda; \tilde{\rho}_{1,t}, \tilde{\rho}_{2,t}) \rangle_\mu}{\langle P_t^*(\lambda; \tilde{\rho}_{1,t}, \tilde{\rho}_{2,t}) \rangle_\mu} \cdot \frac{1}{\tilde{\rho}_{1,t}}, \quad \tilde{w}_{2,t} = 1 - \frac{1 - \langle Q_t^*(\lambda; \tilde{\rho}_{1,t}, \tilde{\rho}_{2,t}) \rangle_{\tilde{\mu}}}{\langle Q_t^*(\lambda; \tilde{\rho}_{1,t}, \tilde{\rho}_{2,t}) \rangle_{\tilde{\mu}}} \cdot \frac{1}{\tilde{\rho}_{2,t}}. \quad (68b)$$

The recursion is initialized with

$$\tilde{w}_{1,1} = \sum_{\lambda_i \in \mathcal{K}^*} \nu_1(\{\lambda_i\}), \quad \tilde{w}_{2,1} = \sum_{\lambda_i \in \mathcal{K}^*} \nu_2(\{\lambda_i\}).$$

Choice of global sign parameters. The spectral initializers $(\tilde{\mathbf{u}}_1, \tilde{\mathbf{v}}_1)$ inherit a single global Rademacher ambiguity from the randomization (57): their overlaps with the true signals, $\langle \tilde{\mathbf{u}}_1, \mathbf{u}_* \rangle$ and $\langle \tilde{\mathbf{v}}_1, \mathbf{v}_* \rangle$, are determined only up to a ± 1 factor. The sign parameters $s_1, s_2 \in \{+1, -1\}$ in (66) are introduced to resolve this ambiguity; they are chosen once according to the prior structure:

- *Asymmetric priors.* When the priors for $(\mathbf{U}_*, \mathbf{V}_*)$ are asymmetric, the global signs are statistically identifiable. We estimate them using the MLE or moment-based procedures in Appendix H.1 and Appendix H.2, and set (s_1, s_2) to these estimates. With this choice, the OAMP iterates have asymptotically positive overlap with the true signals; see Fact 6(i).
- *Symmetric priors.* For symmetric priors, the individual global signs cannot be identified (as noted in [61, Remark 3.6]). Nevertheless, Lemma 15 shows that one may, without loss of generality, adopt the convention

$$(s_1, s_2) = \left(1, \text{sign}(\nu_3(\{\sigma_r\}))\right),$$

where $\nu_3(\{\sigma_r\}) \neq 0$ denotes the point mass of the (signed) cross measure ν_3 associated with the reference outlier $\lambda_r \in \mathcal{K}^*$. This pair is determined only up to a common global flip; under such a flip, the state evolution is preserved in absolute value. Equivalently, the SE recursion for the squared overlaps (and hence the cosine similarities) is invariant; see Fact 6(ii).

Unless stated otherwise, all subsequent state evolution results are understood for the sign-resolved iterates obtained by the above choice of (s_1, s_2) . A detailed treatment of the global sign ambiguity is provided in Appendix H.

5.2 State Evolution of Spectrally-Initialized OAMP

Let $(\tilde{\mu}_{u,t}, \tilde{\mathbf{Z}}_{u,t})$ and $(\tilde{\mu}_{v,t}, \tilde{\mathbf{Z}}_{v,t})$ denote the SE parameters for the spectrally-initialized iterates. The associated scalar random variables are

$$(\mathbf{U}_*, \mathbf{A}) \sim \pi_u, \quad \tilde{\mathbf{U}}_t = \tilde{\mu}_{u,t} \mathbf{U}_* + \tilde{\mathbf{Z}}_{u,t}, \quad \forall t \in \mathbb{N}, \quad (69)$$

$$(\mathbf{V}_*, \mathbf{B}) \sim \pi_v, \quad \tilde{\mathbf{V}}_t = \tilde{\mu}_{v,t} \mathbf{V}_* + \tilde{\mathbf{Z}}_{v,t}, \quad \forall t \in \mathbb{N}, \quad (70)$$

where, for each t , the noise variables $\tilde{\mathbf{Z}}_{u,t}$ and $\tilde{\mathbf{Z}}_{v,t}$ are centered Gaussian and independent of $(\mathbf{U}_*, \mathbf{A})$ and $(\mathbf{V}_*, \mathbf{B})$, respectively.

Initialization and the global-sign issue. The SE is initialized at $t = 1$ using the spectral estimators from (65). As usual, spectral singular vectors are only defined up to a global sign, and (prior to any convention) this sign is not identifiable from the data under symmetric priors. To make the initialization amenable to state evolution, we adopt the randomized sign convention used in the construction of the spectral initializer (cf. (57)). Using this method, we may write the realized orientations as

$$\text{sign}(\langle \tilde{\mathbf{u}}_1, \mathbf{u}_* \rangle) =: S_u \in \{\pm 1\}, \quad \text{sign}(\langle \tilde{\mathbf{v}}_1, \mathbf{v}_* \rangle) =: S_v \in \{\pm 1\},$$

where (S_u, S_v) are Rademacher variables independent of $(\mathbf{u}_*, \mathbf{v}_*, \mathbf{W})$ and hence can be treated as fixed when conditioning. Hereafter, we work conditionally on (S_u, S_v) . Accordingly, we define the initial SE parameters in terms of the *realized* orientation:

$$\tilde{\mu}_{u,1} = S_u \cdot \left(\sum_{\lambda_i \in \mathcal{K}^*} \nu_1(\{\lambda_i\}) \right)^{1/2}, \quad \text{Var}(\tilde{Z}_{u,1}) = 1 - \sum_{\lambda_i \in \mathcal{K}^*} \nu_1(\{\lambda_i\}), \quad (71)$$

$$\tilde{\mu}_{v,1} = S_v \cdot \left(\sum_{\lambda_i \in \mathcal{K}^*} \nu_2(\{\lambda_i\}) \right)^{1/2}, \quad \text{Var}(\tilde{Z}_{v,1}) = 1 - \sum_{\lambda_i \in \mathcal{K}^*} \nu_2(\{\lambda_i\}), \quad (72)$$

where the point masses $\nu_1(\{\lambda\})$ and $\nu_2(\{\lambda\})$ on the limiting outliers $\lambda \in \mathcal{K}^*$ are characterized in Proposition 1.

Recursion for $t \geq 2$. For $t \geq 2$, the SE parameters $\{(\tilde{\mu}_{u,t}, \tilde{Z}_{u,t})\}$ and $\{(\tilde{\mu}_{v,t}, \tilde{Z}_{v,t})\}$ follow from the general update rules in (27) and (28). Substituting the denoisers in (66) yields the compact scalar-channel form

$$\tilde{\mu}_{u,t} = \text{sign} \left(\mathbb{E} \left[\mathbf{U}_* \bar{\phi}(\tilde{\mathbf{U}}_{t-1} \mid \tilde{w}_{1,t-1}, s_1) \right] \right) \tilde{w}_{1,t}^{1/2}, \quad \text{Var}(\tilde{Z}_{u,t}) = 1 - \tilde{w}_{1,t}, \quad (73)$$

$$\tilde{\mu}_{v,t} = \text{sign} \left(\mathbb{E} \left[\mathbf{V}_* \bar{\phi}(\tilde{\mathbf{V}}_{t-1} \mid \tilde{w}_{2,t-1}, s_2) \right] \right) \tilde{w}_{2,t}^{1/2}, \quad \text{Var}(\tilde{Z}_{v,t}) = 1 - \tilde{w}_{2,t}. \quad (74)$$

where $\bar{\phi}(\cdot \mid w, s)$ is the signed DMMSE denoiser (cf. (343)), and (s_1, s_2) encode the convention used to fix the global-sign ambiguity (see the discussion around the initialization).

Having defined the SE parameters in (71)–(73), we now state the state evolution characterization for spectrally-initialized OAMP.

Theorem 2 (State Evolution of Spectrally-Initialized OAMP). *Consider the rectangular spiked model $\mathbf{Y}$ in (1) with super-critical θ , satisfying Assumption 1 and 2 under the settings of Proposition 6. Let $\{(\tilde{\mathbf{u}}_t, \tilde{\mathbf{v}}_t)\}_{t \geq 1}$ be the iterates generated by (65)–(66), and let $\{(\tilde{\mathbf{U}}_t, \tilde{\mathbf{V}}_t)\}_{t \geq 1}$ be the SE variables defined by (69) with initialization (71)–(72) and recursion (73).*

Then, for any fixed $t \in \mathbb{N}$, conditionally on the realized phase variables (S_u, S_v) , the empirical joint distribution of the iterates converges in Wasserstein-2 distance to the law of the SE variables:

$$(\mathbf{u}_*, \tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_t; \mathbf{a}) \xrightarrow{W_2} (\mathbf{U}_*, \tilde{\mathbf{U}}_1, \dots, \tilde{\mathbf{U}}_t; \mathbf{A}), \quad (75)$$

$$(\mathbf{v}_*, \tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_t; \mathbf{b}) \xrightarrow{W_2} (\mathbf{V}_*, \tilde{\mathbf{V}}_1, \dots, \tilde{\mathbf{V}}_t; \mathbf{B}). \quad (76)$$

Proof. See Appendix I. $\square$

Remark 8 (Relation to existing spectral initialization results). Our approach differs from prior work on spectral initialization [42, 59] in the following respects. First, to accommodate the multi-outlier setting, we employ an optimally weighted combination of all informative outlier components, rather than relying solely on the leading eigenvector. Second, we formulate the initialization phase as a *one-shot* OAMP update based on singular-vector equations, avoiding the auxiliary iterative AMP constructions used in [42, 59]. Third, this direct formulation bypasses restrictive technical conditions required for the convergence of auxiliary AMPs to sample PCs (such as nonnegative free cumulants [59] or sufficiently large signal-to-noise ratios [42]), thereby establishing validity for general rotationally invariant models under Assumption 1.

6 Simulation Results

In this section, we provide numerical evidence to validate our theoretical results. We first evaluate the finite-sample performance of the proposed spectral estimators for relative-sign recovery. Subsequently, we investigate the dynamics of the spectrally-initialized OAMP algorithm under both Gaussian and general rotationally invariant noise models.

6.1 Spectral Estimator

A key feature of the rectangular RI model is that a single rank-one signal typically generates *multiple* informative outlier singular values once the signal strength exceeds a certain threshold. In such regimes, standard PCA (which relies solely on the top singular vector) is suboptimal because it discards the signal energy contained in secondary outliers. Our proposed method aims to remedy this by optimally combining all informative outliers.

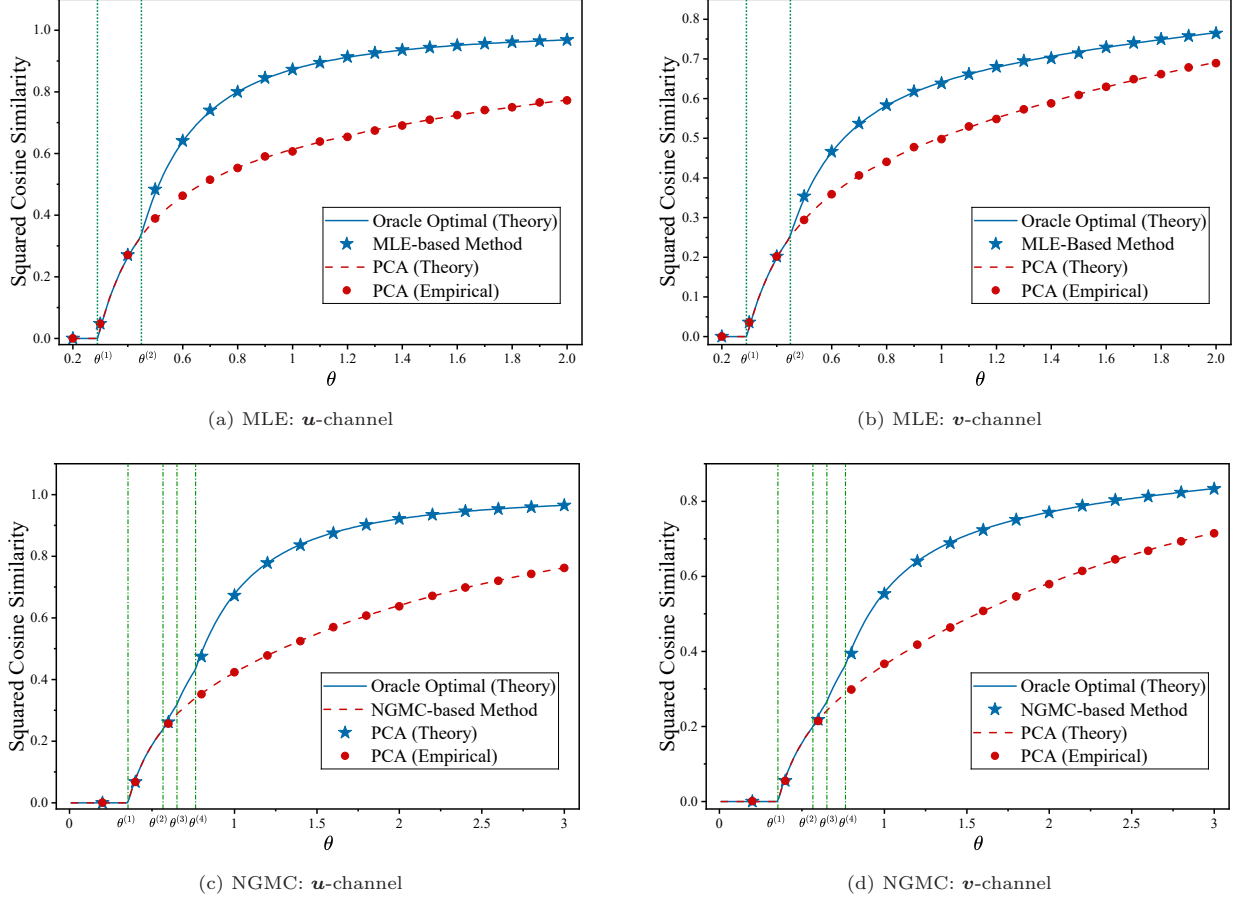


Figure 2: **Relative-sign estimation via MLE and NGMC.** (a)–(b): MLE with Rademacher $\mathbf{U}_*$ and Gaussian $\mathbf{V}_*$ under noise law μ_2 ($\delta = 0.7$). (c)–(d): NGMC with Student- t $\mathbf{U}_*$ ($\text{df} = 5$) under noise law μ_3 ($\delta = 0.8$). Vertical green lines indicate the SNR phase transitions where successive outliers detach from the bulk (cf. Fig. 1). Across all experiments, $N = 5000$ and results are averaged over 50 trials. Stars denote the proposed estimators; circles denote PCA; the solid line denotes the oracle bound.

We evaluate the maximum-likelihood (MLE) and non-Gaussian moment-contrast (NGMC) sign estimators under two representative prior settings; the results are summarized in Fig. 2.

- *MLE with a Rademacher prior.* We consider the setting of Fig. 1 with $\mathbf{U}_* \sim \text{Rad}(\pm 1)$, $\mathbf{V}_* \sim \mathcal{N}(0, 1)$, and noise law

$$\mu_2(\lambda) = \frac{2}{\pi} \sqrt{(\lambda - 2)(4 - \lambda)} \mathbf{1}_{[2, 4]}(\lambda).$$

The resulting MLE performance is reported in the top row of Fig. 2. We take the largest outlier (index 1) as the reference. For each $j \in \mathcal{I}_M \setminus \{1\}$, define the coordinate pairs $(x_i, y_i) \stackrel{\text{def}}{=} ([\mathbf{u}_1^\#]_i, [\mathbf{u}_j^\#]_i)$. The pairwise

relative-sign MLE is then

$$\hat{s}_{u,j}^{\text{MLE}} \in \operatorname{argmax}_{s \in \{\pm 1\}} \sum_{i=1}^M \log p_s(x_i, y_i), \quad (77)$$

where, in terms of the spectral atoms $\nu_1(\{\lambda\})$ from Lemma 2, the per-coordinate likelihood satisfies

$$p_s(x, y) \propto \cosh\left(\frac{\sqrt{\nu_1(\{\lambda_1\})}}{1 - \nu_1(\{\lambda_1\})} x + s \frac{\sqrt{\nu_1(\{\lambda_j\})}}{1 - \nu_1(\{\lambda_j\})} y\right). \quad (78)$$

- *NGMC with a Student- t prior.* We take $\mathbf{U}_*$ to be a rescaled Student- t prior (df= 5, unit variance) and $\mathbf{V}_* \sim \mathcal{N}(0, 1)$ under the noise law

$$\mu_3(\lambda) = \frac{1}{\pi} \sqrt{(\lambda - 2)(4 - \lambda)} \mathbf{1}_{[2,4]}(\lambda) + \frac{1}{\pi} \sqrt{(\lambda - 6)(8 - \lambda)} \mathbf{1}_{[6,8]}(\lambda),$$

as in Fig. 1. Since the Gaussian convolution of this prior is not available in closed form, we use the NGMC estimator in Proposition 8 with the cubic contrast $f(x) = x^3$. The bottom row of Fig. 2 reports the resulting performance.

Fig. 2 shows that aggregating informative outliers strictly improves upon standard PCA and closely tracks the oracle benchmark, indicating accurate recovery of the relative signs.

6.2 Performance of OAMP

We first validate the theory in the classical i.i.d. Gaussian setting. The true signals $\mathbf{u}_*$ and $\mathbf{v}_*$ have i.i.d. Rademacher entries. Figure 3 reports the squared cosine similarities achieved by PCA, AMP, and OAMP, together with the corresponding state evolution (SE) predictions. The OAMP iterates closely match SE and converge to the same fixed point as standard AMP, in agreement with the equivalence in Proposition 3.

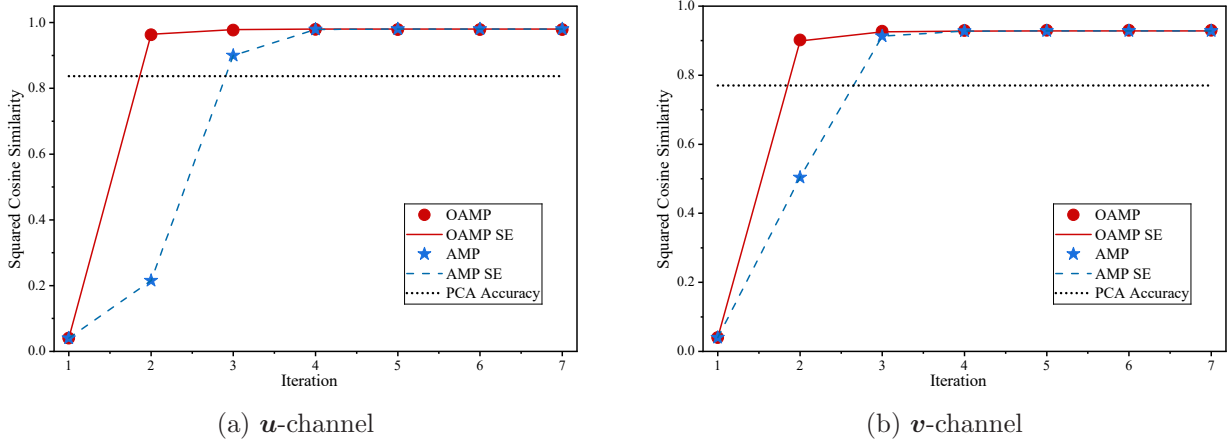


Figure 3: Performance under i.i.d. Gaussian noise ($N = 8000, \delta = 0.6, \theta = 2$). Markers are empirical averages over 50 trials. All methods are initialized with cosine similarity 0.2 in both channels.

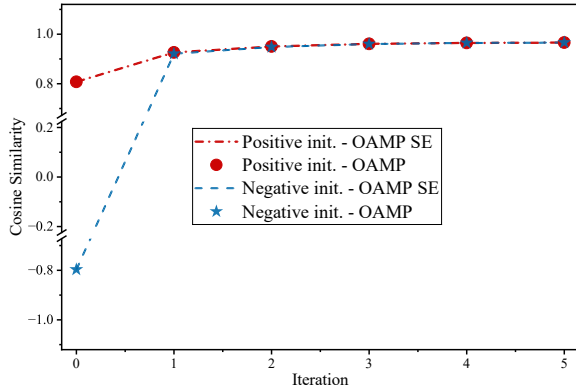
We next study spectrally-initialized OAMP under rotationally invariant (RI) noise with bulk density μ_2 in Fig. 1. Throughout, OAMP uses the optimal spectral initialization from Section 5. Figure 4 plots the *signed* cosine similarity to highlight the intrinsic global sign ambiguity of the spectral initializer. To illustrate the impact of global sign ambiguity, we explicitly realize two possible global orientations for the spectral initializer: one with a positive initial overlap with $(\mathbf{u}_*, \mathbf{v}_*)$, and one with a negative overlap.

We compare two signal priors that necessitate different strategies for determining the denoiser signs (s_1, s_2) , as discussed in Section 5.1:

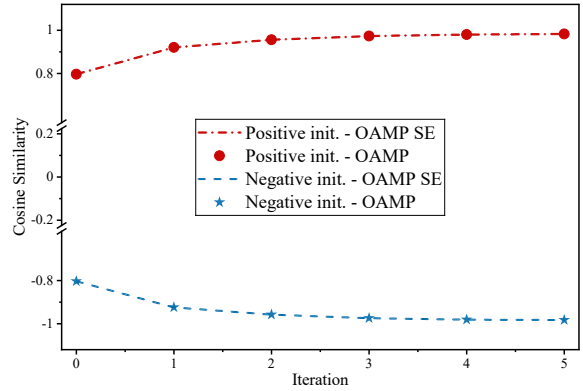
- *Asymmetric three-point prior.* Both $\mathbf{u}_*$ and $\mathbf{v}_*$ follow a three-point mixture supported on $\{-1, 1.5, -0.5\}$ with probabilities $(0.2, 0.3, 0.5)$. The asymmetry renders the realized global signs (s_u, s_v) identifiable. Using the global-sign estimators in Appendix H, we align the denoiser signs (s_1, s_2) with (s_u, s_v) at initialization. Consequently, the iterates rapidly correct the orientation: the trajectories started from the two initial orientations coincide for $t \geq 2$; see Fig. 4(a).
- *Symmetric Rademacher prior.* Both $\mathbf{u}_*$ and $\mathbf{v}_*$ are Rademacher. Here the individual global signs are unidentifiable and only the *relative* sign is recoverable. We adopt the convention

$$(s_1, s_2) = (1, \text{sign}(\nu_3(\{\sqrt{\lambda_r}\}))),$$

where r is the reference outlier index in Lemma 15. Under this convention, the signed overlap tracks the realized orientation of the initializer: a positive (resp. negative) initial overlap remains positive (resp. negative), and the two trajectories are exact sign-mirrors; see Fig. 4(b). In particular, the squared cosine similarity is invariant to the global sign.



(a) Three-point prior, $\mathbf{v}$ -channel



(b) Rademacher prior, $\mathbf{v}$ -channel

Figure 4: Signed cosine similarity of spectrally-initialized OAMP under RI noise with bulk density $\mu_2(\lambda) = \frac{2}{\pi} \sqrt{(\lambda-2)(4-\lambda)} \mathbf{1}_{[2,4]}(\lambda)$ (cf. Fig. 1), with $\delta = 0.7$ and $\theta = 1$. Left: asymmetric three-point prior. Right: symmetric Rademacher prior. In both cases $N = 20000$. Stars and circles represent the two global orientations of the spectral initializer (initial overlap positive vs. negative).

We finally compare OAMP to the RI-AMP framework of [30]. We report a single-iterate variant (AMP-S) and a multi-iterate variant (AMP-M). AMP-S applies the scalar MMSE denoiser to the current iterate, whereas AMP-M first forms the optimal linear combination of past signal-plus-noise observations using their covariance and then applies a single-iterate MMSE denoiser; see [30, Remark 3.3]. Figure 5 shows that the spectrally-initialized optimal OAMP consistently outperforms PCA and both RI-AMP variants.

7 Conclusions and Future Work

This paper established an optimal orthogonal approximate message passing (OAMP) framework for rectangular spiked matrix models under general rotationally invariant noise. We demonstrated that the proposed algorithm admits a rigorous state evolution characterization and incorporates an optimal spectral initialization that effectively aggregates information from multiple outlier eigenvalues. Furthermore, the algorithm achieves asymptotic performance consistent with replica-symmetric Bayes-optimal predictions [60], provided the model operates in a regime where there is no statistical-computational gap. These results provide a robust, computationally efficient approach to inference in high-dimensional settings where the classical i.i.d. noise assumption does not hold.

Several directions for future research emerge from this work:

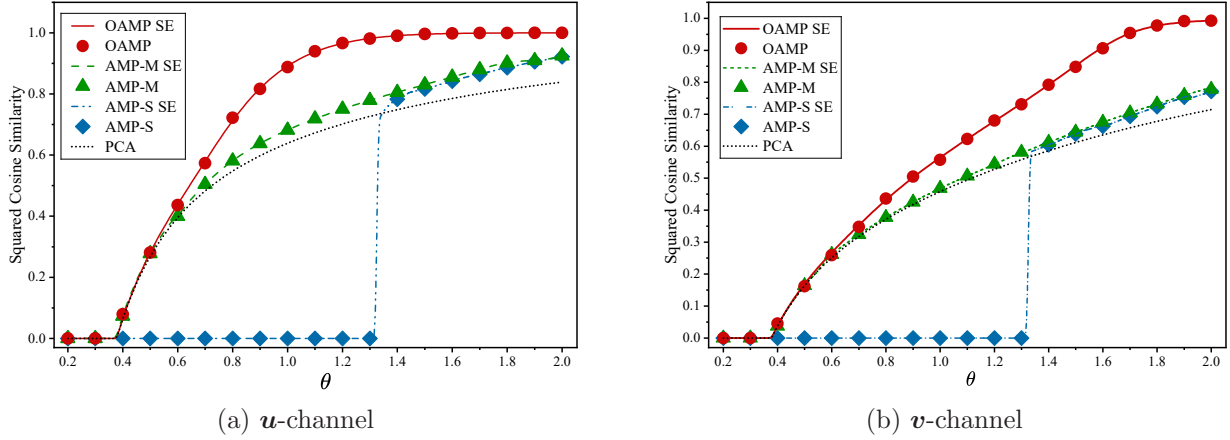


Figure 5: Non-Gaussian RI noise with bulk density $\mu_4(\lambda) = \frac{2}{\pi} \sqrt{(\lambda-1)(3-\lambda)} \mathbf{1}_{[1,3]}(\lambda)$ and Rademacher priors ($N = 10000$, $\delta = 0.5$). Markers are empirical averages over 50 trials. When outlier eigenvectors are present, OAMP denotes the spectrally-initialized optimal OAMP in Theorem 2, whereas AMP uses the spectral initialization of [42] based only on the top eigenvector. When no outlier eigenvectors are available, all methods are initialized with cosine similarity 0.1.

- **Optimality:** Establishing rigorous optimality within a general class of iterative algorithms, extending recent results for symmetric models [44] to the rectangular case.
- **Finite-Rank Generalizations:** Extending the framework to finite-rank spikes to characterize the algorithmic limits of multi-signal inference.
- **Rigorous Bayes Risk:** Providing a formal proof of the Bayes-optimal error under general rotationally invariant noise, potentially utilizing the adaptive interpolation method [66] or related techniques.

Acknowledgement

We are grateful to Rishabh Dudeja for numerous insightful comments on this work.

References

- [1] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley, “Random matrix approach to cross correlations in financial data,” *Phys. Rev. E*, vol. 65, no. 6, p. 066126, 2002.
- [2] J.-P. Bouchaud and M. Potters, “Financial applications of random matrix theory: a short review,” in *The Oxford Handbook of Random Matrix Theory* (G. Akemann, J. Baik, and P. Di Francesco, eds.), pp. 824–850, Oxford, U.K.: Oxford University Press, 2011.
- [3] F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborová, and P. Zhang, “Spectral redemption in clustering sparse networks,” *Proc. Natl. Acad. Sci. USA*, vol. 110, no. 52, pp. 20935–20940, 2013.
- [4] E. Abbe, A. S. Bandeira, and G. Hall, “Exact recovery in the stochastic block model,” *IEEE Trans. Inf. Theory*, vol. 62, pp. 471–487, Jan. 2016.
- [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. Féral and S. Péché, “The largest eigenvalue of rank one deformation of large Wigner matrices,” *Commun. Math. Phys.*, vol. 272, no. 1, pp. 185–228, 2007.

- [7] D. Paul, “Asymptotics of sample eigenstructure for a large dimensional spiked covariance model,” *Stat. Sin.*, vol. 17, no. 4, pp. 1617–1642, 2007.
- [8] Z. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*. Springer Series in Statistics, New York, NY, USA: Springer, 2010.
- [9] H. Zou, T. Hastie, and R. Tibshirani, “Sparse principal component analysis,” *J. Comput. Graph. Stat.*, vol. 15, no. 2, pp. 265–286, 2006.
- [10] Y. Deshpande and A. Montanari, “Information-theoretically optimal sparse PCA,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2197–2201, June 2014.
- [11] T. Lesieur, F. Krzakala, and L. Zdeborová, “Phase transitions in sparse PCA,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1635–1639, June 2015.
- [12] A. Montanari and E. Richard, “Non-negative principal component analysis: Message passing algorithms and sharp asymptotics,” *IEEE Trans. Inf. Theory*, vol. 62, pp. 1458–1484, Mar. 2016.
- [13] T. Lesieur, F. Krzakala, and L. Zdeborová, “Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications,” *J. Stat. Mech.*, vol. 2017, no. 7, p. 073403, 2017.
- [14] C. M. Bishop, “Bayesian PCA,” in *Adv. Neural Inf. Process. Syst.*, vol. 11, pp. 382–388, 1998.
- [15] N. D. Lawrence and R. Urtasun, “Non-linear matrix factorization with Gaussian processes,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, pp. 601–608, 2009.
- [16] A. K. Fletcher and S. Rangan, “Iterative reconstruction of rank-one matrices in noise,” *Inf. Inference*, vol. 7, no. 3, pp. 531–562, 2018.
- [17] 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.
- [18] 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.
- [19] S. Rangan and A. K. Fletcher, “Iterative estimation of constrained rank-one matrices in noise,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1246–1250, July 2012.
- [20] 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.
- [21] J. Barbier, M. Dia, N. Macris, F. Krzakala, T. Lesieur, and L. Zdeborová, “Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula,” in *Adv. Neural Inf. Process. Syst.*, vol. 29, pp. 424–432, 2016.
- [22] J. Ma and L. Ping, “Orthogonal AMP,” *IEEE Access*, vol. 5, pp. 2020–2033, 2017.
- [23] S. Rangan, P. Schniter, and A. K. Fletcher, “Vector approximate message passing,” *IEEE Trans. Inf. Theory*, vol. 65, pp. 6664–6684, Oct. 2019.
- [24] A. Maillard, L. Foini, A. L. Castellanos, F. Krzakala, M. Mézard, and L. Zdeborová, “High-temperature expansions and message passing algorithms,” *J. Stat. Mech.*, vol. 2019, no. 11, p. 113301, 2019.
- [25] K. Takeuchi, “Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements,” *IEEE Trans. Inf. Theory*, vol. 66, pp. 368–386, Jan. 2020.
- [26] L. Miolane, “Fundamental limits of low-rank matrix estimation: The non-symmetric case,” *arXiv preprint*, vol. arXiv:1702.00473, 2018.
- [27] K. Takeuchi, “Convolutional approximate message-passing,” *IEEE Signal Process. Lett.*, vol. 27, pp. 416–420, 2020.

- [28] A. Montanari and R. Venkataramanan, “Estimation of low-rank matrices via approximate message passing,” *Ann. Stat.*, vol. 49, no. 1, pp. 321–345, 2021.
- [29] K. Takeuchi, “Bayes-optimal convolutional AMP,” *IEEE Trans. Inf. Theory*, vol. 67, pp. 4405–4428, July 2021.
- [30] Z. Fan, “Approximate message passing algorithms for rotationally invariant matrices,” *Ann. Stat.*, vol. 50, no. 1, pp. 197–224, 2022.
- [31] L. Liu, S. Huang, and B. M. Kurkoski, “Memory AMP,” *IEEE Trans. Inf. Theory*, vol. 68, pp. 8015–8039, Dec. 2022.
- [32] C. Gerbelot and R. Berthier, “Graph-based approximate message passing iterations,” *Inf. Inference*, vol. 12, no. 4, pp. 2562–2628, 2023.
- [33] R. Dudeja, Y. M. Lu, and S. Sen, “Universality of approximate message passing with semirandom matrices,” *Ann. Probab.*, vol. 51, no. 5, pp. 1616–1683, 2023.
- [34] M. Lovig, T. Wang, and Z. Fan, “On universality of non-separable approximate message passing algorithms,” *arXiv preprint*, vol. arXiv:2506.23010, 2025.
- [35] 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.
- [36] A. El Alaoui and F. Krzakala, “Estimation in the spiked Wigner model: A short proof of the replica formula,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 1874–1878, June 2018.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] M. Capitaine and C. Donati-Martin, “Spectrum of deformed random matrices and free probability,” in *Advanced Topics in Random Matrices*, Panoramas et Synthèses 53, Soc. Math. France, 2018.
- [41] N. Noiry, “Spectral measures of spiked random matrices,” *J. Theor. Probab.*, vol. 34, pp. 923–952, June 2021.
- [42] 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. iaee024, 2024.
- [43] 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.
- [44] R. Dudeja, S. Liu, and J. Ma, “Optimality of Approximate Message Passing Algorithms for Spiked Matrix Models with Rotationally Invariant Noise,” *arXiv preprint arXiv:2405.18081*, 2024.
- [45] O. Ledoit and M. Wolf, “Nonlinear shrinkage estimation of large-dimensional covariance matrices,” *Ann. Stat.*, vol. 40, no. 2, pp. 1024–1060, 2012.
- [46] R. R. Nadakuditi, “OptShrink: An algorithm for improved low-rank signal matrix denoising by optimal, data-driven singular value shrinkage,” *IEEE Trans. Inf. Theory*, vol. 60, pp. 3002–3018, May 2014.
- [47] J. Bun, R. Allez, J.-P. Bouchaud, and M. Potters, “Rotational invariant estimator for general noisy matrices,” *IEEE Trans. Inf. Theory*, vol. 62, pp. 7475–7490, Dec. 2016.

- [48] F. Pourkamali and N. Macris, “Rectangular rotational invariant estimator for general additive noise matrices,” in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2081–2086, June 2023.
- [49] 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.
- [50] S. T. Belinschi, H. Bercovici, M. Capitaine, and M. Février, “Outliers in the spectrum of large deformed unitarily invariant models,” *Ann. Probab.*, vol. 45, no. 6A, pp. 3571–3625, 2017.
- [51] R. R. Nadakuditi, “When are the most informative components for inference also the principal components?,” *arXiv preprint*, vol. arXiv:1302.1232, 2013.
- [52] R. Monasson and D. Villamaina, “Estimating the principal components of correlation matrices from all their empirical eigenvectors,” *EPL (Europhys. Lett.)*, vol. 112, no. 6, p. 60001, 2015.
- [53] Z. Fan, Y. Sun, and Z. Wang, “Principal components in linear mixed models with general bulk,” *Ann. Stat.*, vol. 49, no. 3, pp. 1489–1513, 2021.
- [54] M. Mondelli and A. Montanari, “Fundamental limits of weak recovery with applications to phase retrieval,” *Found. Comput. Math.*, vol. 19, no. 3, pp. 703–773, 2019.
- [55] M. Mondelli and R. Venkataramanan, “Approximate message passing with spectral initialization for generalized linear models,” in *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS)*, vol. 130, pp. 397–405, 2021.
- [56] R. Venkataramanan, K. Kögler, and M. Mondelli, “Estimation in rotationally invariant generalized linear models via approximate message passing,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, vol. 162, pp. 22120–22144, 2022.
- [57] Y. M. Lu and G. Li, “Phase transitions of spectral initialization for high-dimensional non-convex estimation,” *Inf. Inference*, vol. 9, no. 3, pp. 507–541, 2020.
- [58] Y. Chen, Y. Chi, J. Fan, and C. Ma, “Spectral methods for data science: A statistical perspective,” *Found. Trends Mach. Learn.*, vol. 14, no. 5, pp. 566–806, 2021.
- [59] M. Mondelli and R. Venkataramanan, “PCA initialization for approximate message passing in rotationally invariant models,” in *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 29616–29629, 2021.
- [60] Q. Chen, J. Ma, and H. Zhang, “Bayes-optimal inference in rectangular spiked matrix models with rotationally invariant noise: A replica analysis.” In preparation, 2025.
- [61] O. Y. Feng, R. Venkataramanan, C. Rush, and R. J. Samworth, “A unifying tutorial on approximate message passing,” *Found. Trends Mach. Learn.*, vol. 15, no. 4, pp. 335–536, 2022.
- [62] D. L. Cohn, *Measure Theory*. Birkhäuser Advanced Texts Basler Lehrbücher, Boston, MA, USA: Birkhäuser, 2nd ed., 2013.
- [63] L. Pastur and M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices*, vol. 171 of *Mathematical Surveys and Monographs*. Providence, RI, USA: American Mathematical Society, 2011.
- [64] P. Blanchard and E. Brüning, *Mathematical Methods in Physics: Distributions, Hilbert Space Operators and Variational Methods*. Boston, MA, USA: Birkhäuser, 2003.
- [65] J. Ma, J. Xu, and A. Maleki, “Optimization-based AMP for phase retrieval: The impact of initialization and ℓ_2 regularization,” *IEEE Trans. Inf. Theory*, vol. 65, pp. 3600–3629, June 2019.
- [66] 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.

- [67] R. Berthier, A. Montanari, and P.-M. Nguyen, “State evolution for approximate message passing with non-separable functions,” *Inf. Inference*, vol. 9, no. 1, pp. 33–79, 2020.
- [68] R. Dudeja, S. Sen, and Y. M. Lu, “Spectral universality in regularized linear regression with nearly deterministic sensing matrices,” *IEEE Trans. Inf. Theory*, vol. 70, pp. 6851–6881, Oct. 2024.
- [69] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford, UK: Oxford University Press, 2013.
- [70] A. W. van der Vaart, *Asymptotic Statistics*. Cambridge, UK: Cambridge University Press, 1998.
- [71] C. Villani, *Optimal Transport: Old and New*, vol. 338 of *Grundlehren der mathematischen Wissenschaften*. Berlin, Germany: Springer, 2009.
- [72] J. A. Mingo and R. Speicher, *Free Probability and Random Matrices*, vol. 35 of *Fields Institute Monographs*. New York, NY, USA: Springer, 2017.
- [73] N. A. Watson, “Applications of geometric measure theory to the study of Gauss-Weierstrass and Poisson integrals,” *Ann. Acad. Sci. Fenn. Ser. A I Math.*, vol. 19, pp. 115–132, 1994.

A Proofs for Preliminaries and Spectral Analysis

This section is dedicated to proving the characterizations of the signal-eigenspace spectral measures and illustrating their applications. We first prove Lemma 1 in Section A.1, which is concerned with the analytic aspects of the master equation. We then address the properties of the measures in the signal direction in Section A.4. Finally, Section A.5 provides applications of these measures to the spectral analysis of the rectangular spiked model.

A.1 Proof of Lemma 1

Proof. We prove the four claims in order.

1. Analyticity and behavior at infinity. By Definition 4, the function $\mathcal{C}(z)$ is a polynomial combination of $\mathcal{S}_\mu(z)$ and z^{-1} . Since $\mathcal{S}_\mu$ is holomorphic on $\mathbb{C} \setminus \text{supp}(\mu)$, it follows that $\mathcal{C}$ (and hence $\Gamma(z) = 1 - \theta^2 \mathcal{C}(z)$) is holomorphic on $\mathbb{C} \setminus \text{supp}(\mu)$. Moreover, as $|z| \rightarrow \infty$ we have the standard asymptotic expansion for the Stieltjes transform of a probability measure,

$$\mathcal{S}_\mu(z) = \frac{1}{z} + O\left(\frac{1}{z^2}\right), \quad |z| \rightarrow \infty,$$

so $z \mathcal{S}_\mu(z) \rightarrow 1$ and

$$\delta \mathcal{S}_\mu(z) + \frac{1 - \delta}{z} = \frac{1}{z} + O\left(\frac{1}{z^2}\right).$$

Therefore

$$\mathcal{C}(z) = z \mathcal{S}_\mu(z) \left(\delta \mathcal{S}_\mu(z) + \frac{1 - \delta}{z} \right) = O\left(\frac{1}{z}\right), \quad |z| \rightarrow \infty,$$

and hence $\Gamma(z) \rightarrow 1$ as $|z| \rightarrow \infty$. In particular, Γ is not identically zero.

2. Isolated and simple real zeros outside $\text{supp}(\mu)$. Let $\lambda_* \in \mathbb{R} \setminus \text{supp}(\mu)$ satisfy $\Gamma(\lambda_*) = 0$. Since Γ is holomorphic and not identically zero, its zeros are isolated; hence λ_* is isolated.

To prove simplicity, recall from Lemma 2 that the limiting Stieltjes transform of ν_1 is

$$\mathcal{S}_{\nu_1}(z) = \frac{\mathcal{S}_\mu(z)}{\Gamma(z)}.$$

The function $\mathcal{S}_{\nu_1}$ is the Stieltjes transform of the finite measure ν_1 . In particular, all poles of $\mathcal{S}_{\nu_1}$ are simple: a point mass $m \delta_{\lambda_*}$ contributes exactly the term $m/(z - \lambda_*)$, and higher-order poles cannot arise from a finite measure. Since $\mathcal{S}_\mu(\lambda_*) \in \mathbb{R}$ is finite for $\lambda_* \notin \text{supp}(\mu)$, the only possible singularity of $\mathcal{S}_{\nu_1}$ at λ_* is via the denominator $\Gamma(z)$. Therefore the pole of $\mathcal{S}_{\nu_1}$ at $z = \lambda_*$ must be simple, which forces $\Gamma'(\lambda_*) \neq 0$.

3. Absence of real zeros in the interior of $\text{supp}(\mu)$ and $\Gamma(0) \neq 0$. Let λ lie in the interior of $\text{supp}(\mu)$. Since μ has a Hölder continuous density, the boundary value $\mathcal{S}_\mu(\lambda - i0^+)$ exists for Lebesgue-a.e. such λ and satisfies the Sokhotski–Plemelj formula (see, e.g., [63, Section 2.1])

$$\mathcal{S}_\mu(\lambda - i0^+) = \pi\mathcal{H}(\lambda) + i\pi\mu(\lambda). \quad (79)$$

Write

$$\Gamma(\lambda - i0^+) = \mathcal{A}(\lambda) - i\mathcal{B}(\lambda),$$

where $\mathcal{A}(\lambda)$ and $\mathcal{B}(\lambda)$ are the real and imaginary parts of $1 - \theta^2\mathcal{C}(\lambda - i0^+)$. Substituting the boundary value of $\mathcal{S}_\mu$ into (5) and separating real and imaginary parts yields

$$\mathcal{B}(\lambda) = \pi\theta^2\mu(\lambda) \left[(1 - \delta) + 2\delta\pi\lambda\mathcal{H}(\lambda) \right],$$

and $\mathcal{A}(\lambda)$ is the corresponding real-part expression appearing in (6).

Suppose $\mu(\lambda) > 0$ and $\Gamma(\lambda - i0^+) = 0$. Then $\mathcal{B}(\lambda) = 0$, which implies

$$\pi\mathcal{H}(\lambda) = -\frac{1 - \delta}{2\delta\lambda}.$$

Substituting this identity into $\mathcal{A}(\lambda)$ gives

$$\mathcal{A}(\lambda) = 1 + \theta^2 \frac{(1 - \delta)^2}{4\delta\lambda} + \delta\theta^2\pi^2\lambda\mu(\lambda)^2 > 0,$$

a contradiction. Hence $\Gamma(\lambda - i0^+) \neq 0$ at every interior point for which $\mu(\lambda) > 0$ (in particular, for Lebesgue-a.e. λ in the interior of $\text{supp}(\mu)$). Moreover, we have $\Gamma(0) \neq 0$ since for $z = -i\varepsilon$:

$$\lim_{\varepsilon \downarrow 0} \Gamma(-i\varepsilon) \stackrel{(a)}{=} \lim_{\varepsilon \downarrow 0} \left\{ 1 - \theta^2(1 - \delta)\mathcal{S}_\mu(-i\varepsilon) \right\} \stackrel{(b)}{=} 1 - \theta^2(1 - \delta)\pi\mathcal{H}(0) \stackrel{(c)}{>} 0, \quad (80)$$

where (a) uses the definition of Γ in (8) and the non-tangential limit $z\mathcal{S}_\mu(z) \rightarrow 0$ as $z \rightarrow 0$ for μ with absolutely continuous density (cf. Fact 1); (b) follows from the Sokhotski–Plemelj formula (79) at $\lambda = 0$; and (c) holds since $\pi\mathcal{H}(0) = \text{P.V.} \int_{\mathbb{R}_+} \frac{1}{-t} \mu(t) dt < 0$. Consequently, any real solution of the master equation that produces an isolated spectral component must lie in $\mathbb{R} \setminus (\text{supp}(\mu) \cup \{0\})$.

4. Sign of the derivative at real zeros. Let λ satisfy the stated conditions. Differentiating $\Gamma(\lambda) = 1 - \theta^2\mathcal{C}(\lambda)$ yields

$$\Gamma'(\lambda) = -\theta^2\mathcal{C}'(\lambda), \quad \mathcal{C}'(z) = \delta\mathcal{S}_\mu^2(z) + (2\delta z\mathcal{S}_\mu(z) + 1 - \delta)\mathcal{S}'_\mu(z). \quad (81)$$

If $\lambda \in \mathbb{R} \setminus \text{supp}(\mu)$ satisfies $\Gamma(\lambda) = 0$, then the master equation gives

$$\mathcal{C}(\lambda) = \frac{1}{\theta^2} = \mathcal{S}_\mu(\lambda)(\delta\lambda\mathcal{S}_\mu(\lambda) + 1 - \delta).$$

Since $\mathcal{C}(\lambda) \neq 0$, we have $\mathcal{S}_\mu(\lambda) \neq 0$, and thus

$$\delta\lambda\mathcal{S}_\mu(\lambda) + 1 - \delta = \frac{1}{\theta^2\mathcal{S}_\mu(\lambda)}. \quad (82)$$

Substituting (82) into (81) yields

$$\mathcal{C}'(\lambda) = \frac{1}{\mathcal{S}_\mu(\lambda)} \left(\delta\mathcal{S}_\mu^2(\lambda)[\mathcal{S}_\mu(\lambda) + \lambda\mathcal{S}'_\mu(\lambda)] + \frac{\mathcal{S}'_\mu(\lambda)}{\theta^2} \right). \quad (83)$$

For any $\lambda \notin \text{supp}(\mu)$,

$$\mathcal{S}'_\mu(\lambda) = \int \frac{-1}{(\lambda - t)^2} d\mu(t) < 0, \quad (84)$$

$$\mathcal{S}_\mu(\lambda) + \lambda\mathcal{S}'_\mu(\lambda) = \int \frac{-t}{(\lambda - t)^2} d\mu(t) < 0. \quad (85)$$

Hence the parenthetical term in (83) is strictly negative. Since $\mathcal{S}_\mu(\lambda) \neq 0$, it follows that

$$\text{sign}(\mathcal{C}'(\lambda)) = -\text{sign}(\mathcal{S}_\mu(\lambda)), \quad \text{sign}(\Gamma'(\lambda)) = \text{sign}(-\theta^2 \mathcal{C}'(\lambda)) = \text{sign}(\mathcal{S}_\mu(\lambda)).$$

In particular, $\Gamma'(\lambda) \neq 0$.

This completes the proof. $\square$

A.2 Limits of Stieltjes Transforms

Before we present the proof of Lemma 2 (which we collect in Section A.3), we first derive the limiting Stieltjes transforms that will be used throughout the proof.

Define the normalized vectors

$$\bar{\mathbf{u}}_* \stackrel{\text{def}}{=} \frac{\mathbf{u}_*}{\sqrt{M}} \in \mathbb{R}^M, \quad \bar{\mathbf{v}}_* \stackrel{\text{def}}{=} \frac{\mathbf{v}_*}{\sqrt{N}} \in \mathbb{R}^N.$$

By Assumption 1(b), $\|\bar{\mathbf{u}}_*\|^2 \rightarrow 1$ and $\|\bar{\mathbf{v}}_*\|^2 \rightarrow 1$ almost surely. With this notation, the model reads

$$\mathbf{Y} = \theta \bar{\mathbf{u}}_* \bar{\mathbf{v}}_*^\top + \mathbf{W}.$$

Moreover, the Stieltjes transforms of the empirical measures in Definition 5 admit the resolvent representations

$$\mathcal{S}_{\nu_{M,1}}(z) = \frac{1}{M} \mathbf{u}_*^\top (z \mathbf{I}_M - \mathbf{Y} \mathbf{Y}^\top)^{-1} \mathbf{u}_* = \bar{\mathbf{u}}_*^\top (z \mathbf{I}_M - \mathbf{Y} \mathbf{Y}^\top)^{-1} \bar{\mathbf{u}}_*, \quad (86)$$

$$\mathcal{S}_{\nu_{N,2}}(z) = \frac{1}{N} \mathbf{v}_*^\top (z \mathbf{I}_N - \mathbf{Y}^\top \mathbf{Y})^{-1} \mathbf{v}_* = \bar{\mathbf{v}}_*^\top (z \mathbf{I}_N - \mathbf{Y}^\top \mathbf{Y})^{-1} \bar{\mathbf{v}}_*, \quad (87)$$

$$\mathcal{S}_{\nu_{L,3}}(z) = \frac{1}{L} \hat{\mathbf{u}}_*^\top (z \mathbf{I}_L - \hat{\mathbf{Y}})^{-1} \hat{\mathbf{v}}_*, \quad L = M + N. \quad (88)$$

We prove the claimed limit for $\mathcal{S}_{\nu_{N,2}}$; the derivation for $\mathcal{S}_{\nu_{M,1}}$ is completely analogous. Expanding $\mathbf{Y}^\top \mathbf{Y}$ gives

$$\begin{aligned} \mathbf{Y}^\top \mathbf{Y} &= \mathbf{W}^\top \mathbf{W} + \theta \bar{\mathbf{v}}_* \bar{\mathbf{u}}_*^\top \mathbf{W} + \theta \mathbf{W}^\top \bar{\mathbf{u}}_* \bar{\mathbf{v}}_*^\top + \theta^2 \|\bar{\mathbf{u}}_*\|^2 \bar{\mathbf{v}}_* \bar{\mathbf{v}}_*^\top \\ &= \mathbf{W}^\top \mathbf{W} + \tilde{\mathbf{S}}_e + \mathbf{R}_M, \end{aligned}$$

where

$$\tilde{\mathbf{S}}_e \stackrel{\text{def}}{=} \theta^2 \bar{\mathbf{v}}_* \bar{\mathbf{v}}_*^\top + \theta \bar{\mathbf{v}}_* \bar{\mathbf{u}}_*^\top \mathbf{W} + \theta \mathbf{W}^\top \bar{\mathbf{u}}_* \bar{\mathbf{v}}_*^\top, \quad \mathbf{R}_M \stackrel{\text{def}}{=} \theta^2 (\|\bar{\mathbf{u}}_*\|^2 - 1) \bar{\mathbf{v}}_* \bar{\mathbf{v}}_*^\top.$$

Note that $\mathbf{R}_M$ is rank one and

$$\|\mathbf{R}_M\|_{\text{op}} \leq \theta^2 \|\bar{\mathbf{u}}_*\|^2 - 1 \|\bar{\mathbf{v}}_*\|^2 \xrightarrow{\text{a.s.}} 0.$$

Let $\mathbf{C} \stackrel{\text{def}}{=} z \mathbf{I}_N - \mathbf{W}^\top \mathbf{W}$. Then

$$z \mathbf{I}_N - \mathbf{Y}^\top \mathbf{Y} = \mathbf{C} - \tilde{\mathbf{S}}_e - \mathbf{R}_M.$$

The resolvent identity gives

$$(\mathbf{C} - \tilde{\mathbf{S}}_e - \mathbf{R}_M)^{-1} = (\mathbf{C} - \tilde{\mathbf{S}}_e)^{-1} + (\mathbf{C} - \tilde{\mathbf{S}}_e - \mathbf{R}_M)^{-1} \mathbf{R}_M (\mathbf{C} - \tilde{\mathbf{S}}_e)^{-1}.$$

Taking the quadratic form in direction $\bar{\mathbf{v}}_*$ and using the resolvent bounds,

$$\begin{aligned} \left| \mathcal{S}_{\nu_{N,2}}(z) - \bar{\mathbf{v}}_*^\top (\mathbf{C} - \tilde{\mathbf{S}}_e)^{-1} \bar{\mathbf{v}}_* \right| &\leq \|\bar{\mathbf{v}}_*\|^2 \|(\mathbf{C} - \tilde{\mathbf{S}}_e - \mathbf{R}_M)^{-1}\|_{\text{op}} \|\mathbf{R}_M\|_{\text{op}} \|(\mathbf{C} - \tilde{\mathbf{S}}_e)^{-1}\|_{\text{op}} \\ &\leq \frac{\|\bar{\mathbf{v}}_*\|^2}{\eta^2} \|\mathbf{R}_M\|_{\text{op}} \xrightarrow{\text{a.s.}} 0. \end{aligned} \quad (89)$$

Hence it suffices to analyze $\bar{\mathbf{v}}_*^\top (\mathbf{C} - \tilde{\mathbf{S}}_e)^{-1} \bar{\mathbf{v}}_*$. Write $\tilde{\mathbf{S}}_e = \mathbf{A} \mathbf{B}$ with $\mathbf{A} \in \mathbb{R}^{N \times 2}$ and $\mathbf{B} \in \mathbb{R}^{2 \times N}$ defined by

$$\mathbf{A} \stackrel{\text{def}}{=} [\theta \bar{\mathbf{v}}_*, \mathbf{W}^\top \bar{\mathbf{u}}_*], \quad \mathbf{B} \stackrel{\text{def}}{=} \begin{bmatrix} \theta \bar{\mathbf{v}}_*^\top + \bar{\mathbf{u}}_*^\top \mathbf{W} \\ \theta \bar{\mathbf{v}}_*^\top \end{bmatrix}.$$

Then $\mathbf{AB} = \tilde{\mathbf{S}}_e$ is verified by direct multiplication. The Sherman–Morrison–Woodbury formula gives

$$(\mathbf{C} - \mathbf{AB})^{-1} = \mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{A}(\mathbf{I}_2 - \mathbf{BC}^{-1}\mathbf{A})^{-1}\mathbf{BC}^{-1}. \quad (90)$$

Consequently,

$$\bar{\mathbf{v}}_*^\top (\mathbf{C} - \tilde{\mathbf{S}}_e)^{-1} \bar{\mathbf{v}}_* = \bar{\mathbf{v}}_*^\top \mathbf{C}^{-1} \bar{\mathbf{v}}_* + \bar{\mathbf{v}}_*^\top \mathbf{C}^{-1} \mathbf{A}(\mathbf{I}_2 - \mathbf{BC}^{-1}\mathbf{A})^{-1} \mathbf{BC}^{-1} \bar{\mathbf{v}}_*. \quad (91)$$

By rotational invariance of $\mathbf{W}$ and $(\bar{\mathbf{u}}_*, \bar{\mathbf{v}}_*) \perp \mathbf{W}$, the limits follow by the same Haar-rotation/continuous-mapping argument as in (129)–(131) (cf. [30, Proposition E.2]); in particular, for each fixed $z \in \mathbb{C} \setminus \mathbb{R}_+$,

$$\bar{\mathbf{v}}_*^\top (z\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1} \bar{\mathbf{v}}_* \xrightarrow{\text{a.s.}} \delta \mathcal{S}_\mu(z) + \frac{1 - \delta}{z}, \quad (92)$$

$$\bar{\mathbf{u}}_*^\top \mathbf{W} (z\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top \bar{\mathbf{u}}_* \xrightarrow{\text{a.s.}} \int_{\mathbb{R}_+} \frac{\lambda}{z - \lambda} d\mu(\lambda) = z \mathcal{S}_\mu(z) - 1, \quad (93)$$

$$\bar{\mathbf{u}}_*^\top \mathbf{W} (z\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1} \bar{\mathbf{v}}_* \xrightarrow{\text{a.s.}} 0, \quad \bar{\mathbf{v}}_*^\top (z\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top \bar{\mathbf{u}}_* \xrightarrow{\text{a.s.}} 0. \quad (94)$$

Define the deterministic quantities

$$d_1(z) \stackrel{\text{def}}{=} \delta \mathcal{S}_\mu(z) + \frac{1 - \delta}{z}, \quad d_2(z) \stackrel{\text{def}}{=} z \mathcal{S}_\mu(z) - 1.$$

Using (92)–(94), we obtain

$$\begin{aligned} \mathbf{BC}^{-1}\mathbf{A} &= \begin{bmatrix} (\theta \bar{\mathbf{v}}_*^\top + \bar{\mathbf{u}}_*^\top \mathbf{W}) \mathbf{C}^{-1} (\theta \bar{\mathbf{v}}_*) & (\theta \bar{\mathbf{v}}_*^\top + \bar{\mathbf{u}}_*^\top \mathbf{W}) \mathbf{C}^{-1} (\mathbf{W}^\top \bar{\mathbf{u}}_*) \\ \theta \bar{\mathbf{v}}_*^\top \mathbf{C}^{-1} (\theta \bar{\mathbf{v}}_*) & \theta \bar{\mathbf{v}}_*^\top \mathbf{C}^{-1} (\mathbf{W}^\top \bar{\mathbf{u}}_*) \end{bmatrix} \\ &\xrightarrow{\text{a.s.}} \begin{bmatrix} \theta^2 d_1(z) & d_2(z) \\ \theta^2 d_1(z) & 0 \end{bmatrix}. \end{aligned}$$

Likewise,

$$\bar{\mathbf{v}}_*^\top \mathbf{C}^{-1} \mathbf{A} \xrightarrow{\text{a.s.}} [\theta d_1(z), 0], \quad \mathbf{BC}^{-1} \bar{\mathbf{v}}_* \xrightarrow{\text{a.s.}} \begin{bmatrix} \theta d_1(z) \\ \theta d_1(z) \end{bmatrix}.$$

A direct 2×2 calculation then yields

$$\bar{\mathbf{v}}_*^\top (\mathbf{C} - \tilde{\mathbf{S}}_e)^{-1} \bar{\mathbf{v}}_* \xrightarrow{\text{a.s.}} \frac{d_1(z)}{1 - \theta^2 d_1(z)(1 + d_2(z))}. \quad (95)$$

Since $1 + d_2(z) = z \mathcal{S}_\mu(z)$, we recognize

$$d_1(z)(1 + d_2(z)) = \left(\delta \mathcal{S}_\mu(z) + \frac{1 - \delta}{z} \right) \cdot z \mathcal{S}_\mu(z) = z \mathcal{S}_\mu(z) \left(\delta \mathcal{S}_\mu(z) + \frac{1 - \delta}{z} \right) = \mathcal{C}(z).$$

Therefore, combining (89) and (95),

$$\mathcal{S}_{\nu_{N,2}}(z) \xrightarrow{\text{a.s.}} \mathcal{S}_{\nu_2}(z) \stackrel{\text{def}}{=} \frac{\delta \mathcal{S}_\mu(z) + (1 - \delta)/z}{1 - \theta^2 \mathcal{C}(z)}, \quad z \in \mathbb{C} \setminus \mathbb{R}. \quad (96)$$

The proof for $\mathcal{S}_{\nu_{M,1}}(z)$ is analogous and gives

$$\mathcal{S}_{\nu_{M,1}}(z) \xrightarrow{\text{a.s.}} \mathcal{S}_{\nu_1}(z) \stackrel{\text{def}}{=} \frac{\mathcal{S}_\mu(z)}{1 - \theta^2 \mathcal{C}(z)}. \quad (97)$$

This establishes the first two displays in (9).

Recall the symmetric dilation

$$\hat{\mathbf{Y}} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{0} & \mathbf{Y} \\ \mathbf{Y}^\top & \mathbf{0} \end{bmatrix} \in \mathbb{R}^{L \times L}, \quad L = M + N,$$

and define $\widehat{\mathbf{W}}$ analogously. Define the embedded normalized signal directions

$$\bar{\mathbf{u}}_*^{(L)} \stackrel{\text{def}}{=} \begin{bmatrix} \bar{\mathbf{u}}_* \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^L, \quad \bar{\mathbf{v}}_*^{(L)} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{0} \\ \bar{\mathbf{v}}_* \end{bmatrix} \in \mathbb{R}^L.$$

Then $\widehat{\mathbf{u}}_* = \sqrt{M} \bar{\mathbf{u}}_*^{(L)}$ and $\widehat{\mathbf{v}}_* = \sqrt{N} \bar{\mathbf{v}}_*^{(L)}$, and

$$\widehat{\mathbf{Y}} = \widehat{\mathbf{W}} + \theta(\bar{\mathbf{u}}_*^{(L)}(\bar{\mathbf{v}}_*^{(L)})^\top + \bar{\mathbf{v}}_*^{(L)}(\bar{\mathbf{u}}_*^{(L)})^\top).$$

Let $\mathbf{C}_b \stackrel{\text{def}}{=} z\mathbf{I}_L - \widehat{\mathbf{W}}$ and define

$$\mathbf{X} \stackrel{\text{def}}{=} [\bar{\mathbf{u}}_*^{(L)}, \bar{\mathbf{v}}_*^{(L)}], \quad \mathbf{J} \stackrel{\text{def}}{=} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Applying Sherman–Morrison–Woodbury to this rank-two perturbation yields

$$(z\mathbf{I}_L - \widehat{\mathbf{Y}})^{-1} = \mathbf{C}_b^{-1} + \mathbf{C}_b^{-1} \mathbf{X} (\mathbf{I}_2 - \theta \mathbf{J} \mathbf{X}^\top \mathbf{C}_b^{-1} \mathbf{X})^{-1} \theta \mathbf{J} \mathbf{X}^\top \mathbf{C}_b^{-1}. \quad (98)$$

By (88) and the relations above,

$$\mathcal{S}_{\nu_{L,3}}(z) = \frac{1}{L} \widehat{\mathbf{u}}_*^\top (z\mathbf{I}_L - \widehat{\mathbf{Y}})^{-1} \widehat{\mathbf{v}}_* = \frac{\sqrt{MN}}{L} (\bar{\mathbf{u}}_*^{(L)})^\top (z\mathbf{I}_L - \widehat{\mathbf{Y}})^{-1} \bar{\mathbf{v}}_*^{(L)}. \quad (99)$$

A Schur complement calculation gives the standard identity

$$\mathbf{C}_b^{-1} = \begin{bmatrix} z^{-1}(\mathbf{I}_M + \mathbf{W}(z^2\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top) & \mathbf{W}(z^2\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1} \\ (z^2\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1} \mathbf{W}^\top & z(z^2\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1} \end{bmatrix}.$$

Hence

$$(\bar{\mathbf{u}}_*^{(L)})^\top \mathbf{C}_b^{-1} \bar{\mathbf{v}}_*^{(L)} = \bar{\mathbf{u}}_*^\top \mathbf{W} (z^2\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1} \bar{\mathbf{v}}_* \xrightarrow{\text{a.s.}} 0,$$

by the same cross-term argument as (94). Define

$$e_1(z) \stackrel{\text{def}}{=} z \left(\delta \mathcal{S}_\mu(z^2) + \frac{1-\delta}{z^2} \right), \quad e_2(z) \stackrel{\text{def}}{=} z \mathcal{S}_\mu(z^2).$$

The same quadratic-form concentration used above yields almost surely

$$(\bar{\mathbf{v}}_*^{(L)})^\top \mathbf{C}_b^{-1} \bar{\mathbf{v}}_*^{(L)} \rightarrow e_1(z), \quad (\bar{\mathbf{u}}_*^{(L)})^\top \mathbf{C}_b^{-1} \bar{\mathbf{u}}_*^{(L)} \rightarrow e_2(z), \quad (\bar{\mathbf{u}}_*^{(L)})^\top \mathbf{C}_b^{-1} \bar{\mathbf{v}}_*^{(L)} \rightarrow 0.$$

A direct 2×2 computation in (98) then gives

$$(\bar{\mathbf{u}}_*^{(L)})^\top (z\mathbf{I}_L - \widehat{\mathbf{Y}})^{-1} \bar{\mathbf{v}}_*^{(L)} \xrightarrow{\text{a.s.}} \frac{\theta e_1(z) e_2(z)}{1 - \theta^2 e_1(z) e_2(z)}.$$

Finally,

$$e_1(z) e_2(z) = z^2 \mathcal{S}_\mu(z^2) \left(\delta \mathcal{S}_\mu(z^2) + \frac{1-\delta}{z^2} \right) = \mathcal{C}(z^2).$$

Since $\sqrt{MN}/L \rightarrow \sqrt{\delta}/(1+\delta)$, (99) yields

$$\mathcal{S}_{\nu_{L,3}}(z) \xrightarrow{\text{a.s.}} \mathcal{S}_{\nu_3}(z) \stackrel{\text{def}}{=} \frac{\sqrt{\delta}}{1+\delta} \cdot \frac{\theta \mathcal{C}(z^2)}{1 - \theta^2 \mathcal{C}(z^2)}, \quad z \in \mathbb{C} \setminus \mathbb{R}, \quad (100)$$

which is the last identity in (9).

A.3 Proof of Lemma 2

We prove each claim separately.

1. Weak convergence. We have pointwise a.s. convergence of Stieltjes transforms on $\mathbb{C} \setminus \mathbb{R}$. Moreover, by Assumption 1(c), $\|\mathbf{W}\|_{\text{op}}$ is uniformly bounded. As a consequence, $\{\nu_{M,1}\}_M$, $\{\nu_{N,2}\}_N$, and $\{\nu_{L,3}\}_L$ are uniformly compactly supported and (in particular, tight). It follows from the Stieltjes continuity theorem (and its signed-measure analogue) that $\nu_{M,1}, \nu_{N,2}, \nu_{L,3}$ converges weakly (almost surely) to ν_1, ν_2, ν_3 , respectively. Finally, ν_1 and ν_2 are probability measures since $\lim_{y \rightarrow \infty} (-iy)\mathcal{S}_{\nu_1}(iy) = 1$ and $\lim_{y \rightarrow \infty} (-iy)\mathcal{S}_{\nu_2}(iy) = 1$ (using $y\mathcal{S}_{\mu}(iy) \rightarrow -i$ and $\mathcal{C}(iy) = O(1/y)$). Likewise, ν_3 is finite since $\limsup_{y \rightarrow \infty} y|\mathcal{S}_{\nu_3}(iy)| < \infty$.

2. Limiting Stieltjes transform. They have been established in Section A.2, specifically, (97), (96), and (100).

3. Absolutely continuous parts. Let $\nu_i = \nu_i^{\parallel} + \nu_i^{\perp}$ be the Lebesgue decomposition for $i \in \{1, 2, 3\}$. We use the Stieltjes inversion principle for finite signed measures (cf. Fact 4): if $\chi = \chi^{\parallel} + \chi^{\perp}$, then

$$\begin{aligned} \lim_{\epsilon \downarrow 0} \Im \{ \mathcal{S}_{\chi}(\lambda - i\epsilon) \} &= \pi \frac{d\chi^{\parallel}}{d\lambda}(\lambda), \quad \text{for Lebesgue-a.e. } \lambda \in \mathbb{R}, \\ \lim_{\epsilon \downarrow 0} \left| \Im \{ \mathcal{S}_{\chi}(\lambda - i\epsilon) \} \right| &= +\infty, \quad \text{for } |\chi^{\perp}| \text{-a.e. } \lambda \in \mathbb{R}. \end{aligned} \tag{101}$$

Moreover, since μ has a Hölder continuous density (Assumption 1(c)), its Stieltjes transform admits boundary values given by the Sokhotski–Plemelj formula: for Lebesgue-a.e. $\lambda \in \mathbb{R}$,

$$\lim_{\epsilon \downarrow 0} \mathcal{S}_{\mu}(\lambda - i\epsilon) = \pi \mathcal{H}(\lambda) + i\pi \mu(\lambda), \tag{102}$$

where $\mathcal{H}$ is the Hilbert transform of μ . Applying (101) to $\chi = \nu_1$ and $\chi = \nu_2$, and using (9), we obtain for Lebesgue-a.e. $\lambda \in \mathbb{R}$,

$$\frac{d\nu_1^{\parallel}}{d\lambda}(\lambda) = \frac{1}{\pi} \lim_{\epsilon \downarrow 0} \Im \{ \mathcal{S}_{\nu_1}(\lambda - i\epsilon) \}, \quad \frac{d\nu_2^{\parallel}}{d\lambda}(\lambda) = \frac{1}{\pi} \lim_{\epsilon \downarrow 0} \Im \{ \mathcal{S}_{\nu_2}(\lambda - i\epsilon) \}.$$

Substituting the boundary value (102) into (9) and separating real and imaginary parts yields the expressions in (10) with the shrinkage functions φ_1 and φ_2 defined in (7). The required algebra is exactly the one encoded in (6) and (7).

We apply (101) to the signed measure $\chi = \nu_3$ and use the closed-form transform $\mathcal{S}_{\nu_3}$ in (9). For $\sigma \neq 0$, the boundary value of $\mathcal{S}_{\mu}((\sigma - i0^+)^2)$ is given by Fact 3, which yields

$$\frac{d\nu_3^{\parallel}}{d\sigma}(\sigma) = \frac{1}{\pi} \lim_{\epsilon \downarrow 0} \Im \{ \mathcal{S}_{\nu_3}(\sigma - i\epsilon) \} = \frac{\sqrt{\delta}}{1 + \delta} \text{sign}(\sigma) \mu(\sigma^2) \varphi_3(\sigma^2),$$

for Lebesgue-a.e. $\sigma \in \mathbb{R}$, and this is precisely the third identity in (10).

4. Singular parts and atomic masses. Let

$$\mathcal{K}_* \stackrel{\text{def}}{=} \{ \lambda \in \mathbb{R} \setminus \text{supp}(\mu) : \Gamma(\lambda) = 0 \}, \quad \Gamma(z) \stackrel{\text{def}}{=} 1 - \theta^2 \mathcal{C}(z).$$

Notice that $\mathcal{K}_*$ is assumed a *finite* set. By Fact 4, the singular component $\chi^{\perp}$ of a finite (signed) measure $\chi = \chi^{\parallel} + \chi^{\perp}$ can only assign mass to points where the boundary imaginary part of $\mathcal{S}_{\chi}(\lambda - i\epsilon)$ diverges as $\epsilon \downarrow 0$. We apply this criterion to ν_1, ν_2, ν_3 using the explicit Stieltjes transforms (9).

Fix $\lambda_0 \in \mathbb{R} \setminus \text{supp}(\mu)$. Since $\mathcal{S}_{\mu}$ is analytic on $\mathbb{C} \setminus \text{supp}(\mu)$, the functions $\mathcal{C}$ and Γ are analytic at λ_0 . If $\Gamma(\lambda_0) \neq 0$, then by continuity there exists $\epsilon_0 > 0$ such that $\inf_{0 < \epsilon < \epsilon_0} |\Gamma(\lambda_0 - i\epsilon)| > 0$. Moreover, $\mathcal{S}_{\mu}(\lambda_0 - i\epsilon)$ remains bounded for $0 < \epsilon < \epsilon_0$. Hence,

$$\mathcal{S}_{\nu_1}(\lambda_0 - i\epsilon) = \frac{\mathcal{S}_{\mu}(\lambda_0 - i\epsilon)}{\Gamma(\lambda_0 - i\epsilon)} \quad \text{remains bounded as } \epsilon \downarrow 0,$$

so $\Im \mathcal{S}_{\nu_1}(\lambda_0 - i\epsilon)$ cannot diverge at such a point. Therefore, outside $\text{supp}(\mu)$, any singular mass of ν_1 can only occur at real zeros of Γ , i.e., at points in $\mathcal{K}_*$. The same conclusion holds for ν_2 at any $\lambda_0 \neq 0$, since

$$\mathcal{S}_{\nu_2}(z) = \frac{\delta \mathcal{S}_{\mu}(z) + (1 - \delta)/z}{\Gamma(z)}$$

is analytic at λ_0 whenever $\lambda_0 \in \mathbb{R} \setminus \text{supp}(\mu)$ and $\Gamma(\lambda_0) \neq 0$. When $\delta < 1$, the term $(1 - \delta)/z$ may create an additional pole at $z = 0$, yielding a possible atom at the origin even if $0 \notin \mathcal{K}_*$. For ν_3 , we use

$$\mathcal{S}_{\nu_3}(z) = \frac{\sqrt{\delta}}{1 + \delta} \cdot \frac{\theta \mathcal{C}(z^2)}{1 - \theta^2 \mathcal{C}(z^2)} = \frac{\sqrt{\delta}}{1 + \delta} \cdot \frac{\theta \mathcal{C}(z^2)}{\Gamma(z^2)}.$$

Thus any singularity of $\mathcal{S}_{\nu_3}$ away from the bulk can only occur when $\Gamma(z^2) = 0$, i.e., at $z = \pm\sqrt{\lambda_*}$ with $\lambda_* \in \mathcal{K}_*$.

Since $\mathcal{K}_*$ is finite, the preceding localization implies that the singular components are purely atomic and admit the representations

$$\begin{aligned} \nu_1^\perp &= \sum_{\lambda_* \in \mathcal{K}_*} \nu_1(\{\lambda_*\}) \delta_{\lambda_*}, \\ \nu_2^\perp &= \sum_{\lambda_* \in \mathcal{K}_*} \nu_2(\{\lambda_*\}) \delta_{\lambda_*} + \mathbf{1}_{\{\delta < 1\}} \nu_2(\{0\}) \delta_0, \\ \nu_3^\perp &= \sum_{\lambda_* \in \mathcal{K}_*} \left(\nu_3(\{\sigma_*\}) \delta_{\sigma_*} + \nu_3(\{-\sigma_*\}) \delta_{-\sigma_*} \right), \quad \sigma_* \stackrel{\text{def}}{=} \sqrt{\lambda_*}. \end{aligned}$$

Fix $\lambda_* \in \mathcal{K}_*$. Then $\lambda_* \notin \text{supp}(\mu)$, so $\mathcal{S}_\mu$ and $\mathcal{C}$ are analytic at λ_* . By Lemma 1, λ_* is a simple zero of Γ , hence $\Gamma'(\lambda_*) \neq 0$. Therefore $\mathcal{S}_{\nu_1}$ and $\mathcal{S}_{\nu_2}$ have simple poles at $z = \lambda_*$, and since a Dirac mass $m \delta_{\lambda_*}$ contributes $m/(z - \lambda_*)$ to the Stieltjes transform, the corresponding atom masses are precisely the residues:

$$\nu_1(\{\lambda_*\}) = \text{Res}_{z=\lambda_*} \mathcal{S}_{\nu_1}(z) = \frac{\mathcal{S}_\mu(\lambda_*)}{\Gamma'(\lambda_*)} = -\frac{\mathcal{S}_\mu(\lambda_*)}{\theta^2 \mathcal{C}'(\lambda_*)},$$

and

$$\nu_2(\{\lambda_*\}) = \text{Res}_{z=\lambda_*} \mathcal{S}_{\nu_2}(z) = \frac{\delta \mathcal{S}_\mu(\lambda_*) + (1 - \delta)/\lambda_*}{\Gamma'(\lambda_*)} = -\frac{\delta \mathcal{S}_\mu(\lambda_*) + (1 - \delta)/\lambda_*}{\theta^2 \mathcal{C}'(\lambda_*)}.$$

These residues are nonzero: otherwise $\mathcal{C}(\lambda_*) = 0$, contradicting $\Gamma(\lambda_*) = 0$. For ν_3 , the poles occur at $z = \pm\sigma_*$ with $\sigma_* = \sqrt{\lambda_*} > 0$. A direct residue calculation from $\mathcal{S}_{\nu_3}(z) = \frac{\sqrt{\delta}}{1 + \delta} \frac{\theta \mathcal{C}(z^2)}{\Gamma(z^2)}$ yields

$$\nu_3(\{\pm\sigma_*\}) = \text{Res}_{z=\pm\sigma_*} \mathcal{S}_{\nu_3}(z) = \mp \frac{\sqrt{\delta}}{1 + \delta} \frac{1}{2\theta^3 \sigma_* \mathcal{C}'(\sigma_*^2)}.$$

Finally, when $\delta < 1$, $\mathcal{S}_{\nu_2}$ has a simple pole at $z = 0$ due to the term $(1 - \delta)/z$, and its residue is

$$\nu_2(\{0\}) = \text{Res}_{z=0} \mathcal{S}_{\nu_2}(z) = \lim_{z \rightarrow 0} z \mathcal{S}_{\nu_2}(z) = \frac{1 - \delta}{1 - \theta^2(1 - \delta)\pi \mathcal{H}(0)}.$$

This completes the proof.

A.4 Proof of Proposition 1

Proof. We prove Claims 1–4 in order.

Proof of Claim 1 (Isolation of population outliers). Since $\mathcal{K}_* \subset \mathbb{R} \setminus \text{supp}(\mu)$ is finite, define

$$d_\mu \stackrel{\text{def}}{=} \min_{\lambda \in \mathcal{K}_*} d(\lambda, \text{supp}(\mu)) > 0, \quad d_K \stackrel{\text{def}}{=} \min_{\substack{\lambda, \lambda' \in \mathcal{K}_* \\ \lambda \neq \lambda'}} |\lambda - \lambda'| \in (0, \infty],$$

with $d_K = +\infty$ if $|\mathcal{K}_*| = 1$. Set

$$\varepsilon \stackrel{\text{def}}{=} \min \left\{ \frac{d_\mu}{4}, \frac{d_K}{4} \right\} > 0. \tag{103}$$

Then for each $\lambda_k \in \mathcal{K}_*$,

$$(\lambda_k - \varepsilon, \lambda_k + \varepsilon) \cap \text{supp}(\mu) = \emptyset, \quad (\lambda_k - \varepsilon, \lambda_k + \varepsilon) \cap \mathcal{K}_* = \{\lambda_k\},$$

and the intervals are pairwise disjoint. Define the outlier neighborhood

$$\mathcal{K}^{*\epsilon} \stackrel{\text{def}}{=} \bigcup_{\lambda_k \in \mathcal{K}_*} (\lambda_k - \epsilon, \lambda_k + \epsilon).$$

By construction, $d(\mathcal{K}^{*\epsilon}, \text{supp}(\mu)) \geq 3\epsilon$, hence $\mathcal{K}^{*\epsilon} \cap \text{supp}_{\epsilon/2}(\mu) = \emptyset$. This proves Claim 1.

General setup: confinement of roots and uniform convergence. Let

$$\mathbf{S}_1(z) \stackrel{\text{def}}{=} (z\mathbf{I}_M - \mathbf{W}\mathbf{W}^\top)^{-1}, \quad \mathbf{S}_2(z) \stackrel{\text{def}}{=} (z\mathbf{I}_N - \mathbf{W}^\top\mathbf{W})^{-1},$$

and recall the empirical master function $\Gamma_M(z)$.

Fix $\eta > 0$ and a compact set $\mathcal{E} \subset \mathbb{C} \setminus \text{supp}_\eta(\mu)$. By Assumption 1(d), almost surely for all M large enough, $\text{sp}(\mathbf{W}\mathbf{W}^\top) \subset \text{supp}_{\eta/2}(\mu)$, hence

$$\|\mathbf{S}_1(z)\|_{\text{op}}, \|\mathbf{S}_2(z)\|_{\text{op}} \leq 2/\eta \quad \forall z \in \mathcal{E}.$$

Therefore the quadratic forms appearing in Γ_M define uniformly bounded holomorphic families on $\mathcal{E}$.

For each fixed $z \in \mathcal{E}$, standard quadratic-form limits for Haar singular vectors (e.g., [39, Proposition 8.2]) yield almost surely

$$\frac{\theta}{M} \mathbf{u}_*^\top \mathbf{S}_1(z) \mathbf{u}_* \rightarrow \theta \mathcal{S}_\mu(z), \tag{104a}$$

$$\frac{\theta}{N} \mathbf{v}_*^\top \mathbf{S}_2(z) \mathbf{v}_* \rightarrow \theta \left(\delta \mathcal{S}_\mu(z) + \frac{1-\delta}{z} \right), \tag{104b}$$

$$\frac{\theta}{\sqrt{MN}} \mathbf{v}_*^\top \mathbf{W}^\top \mathbf{S}_1(z) \mathbf{u}_* \rightarrow 0, \tag{104c}$$

$$\frac{\theta}{\sqrt{MN}} \mathbf{u}_*^\top \mathbf{W} \mathbf{S}_2(z) \mathbf{v}_* \rightarrow 0. \tag{104d}$$

By Montel's theorem and uniqueness of the holomorphic limit, the convergence is uniform on $\mathcal{E}$. Substituting into Γ_M yields

$$\Gamma_M(z) \rightarrow \Gamma(z) \stackrel{\text{def}}{=} 1 - \theta^2 \mathcal{C}(z) \quad \text{uniformly on } \mathcal{E}.$$

Since $\eta > 0$ and $\mathcal{E}$ are arbitrary, $\Gamma_M \rightarrow \Gamma$ almost surely locally uniformly on $\mathbb{C} \setminus \text{supp}(\mu)$.

Moreover, since $\mathcal{S}_\mu(z) \sim 1/z$ as $|z| \rightarrow \infty$, we have $\Gamma(z) \rightarrow 1$. Thus there exists $R > 0$ such that $|\Gamma(z)| > 0$ for $|z| \geq R$. By local uniform convergence, the same holds for Γ_M for all large M . Hence all real zeros of Γ and Γ_M lie in $(-R, R)$ eventually.

Proof of Claim 3 (Existence and convergence of empirical outliers). Fix $\lambda_k \in \mathcal{K}_*$. Since λ_k is an isolated and simple real zero of Γ , there exists $\rho \in (0, \epsilon)$ such that the closed disk

$$D_k \stackrel{\text{def}}{=} \{z \in \mathbb{C} : |z - \lambda_k| \leq \rho\}$$

satisfies

$$D_k \cap \text{supp}(\mu) = \emptyset, \quad D_k \cap \mathcal{K}_* = \{\lambda_k\}, \quad \inf_{z \in \partial D_k} |\Gamma(z)| > 0.$$

Define the contour

$$\gamma_k \stackrel{\text{def}}{=} \partial D_k = \{z \in \mathbb{C} : |z - \lambda_k| = \rho\},$$

which is the positively oriented circle of radius ρ centered at λ_k .

Since $\gamma_k \subset \mathbb{C} \setminus \text{supp}(\mu)$, the function Γ is holomorphic in a neighborhood of γ_k . By the argument principle,

$$\frac{1}{2\pi i} \oint_{\gamma_k} \frac{\Gamma'(z)}{\Gamma(z)} dz = \text{Wind}(\Gamma(\gamma_k), 0) = 1, \tag{105}$$

where the equality follows from the simplicity of the zero at λ_k .

By the local uniform convergence $\Gamma_M \rightarrow \Gamma$ on $\mathbb{C} \setminus \text{supp}(\mu)$, there exists $M_k < \infty$ such that for all $M \geq M_k$,

$$\inf_{z \in \gamma_k} |\Gamma_M(z)| \geq \frac{1}{2} \inf_{z \in \gamma_k} |\Gamma(z)| > 0.$$

Moreover, by Assumption 1(d), $\text{sp}(\mathbf{W}\mathbf{W}^\top) \subset \text{supp}_{\varepsilon/2}(\mu)$ for all large M , hence $\gamma_k \cap \text{sp}(\mathbf{W}\mathbf{W}^\top) = \emptyset$ and Γ_M is holomorphic in a neighborhood of γ_k . Uniform convergence of Γ_M and Γ'_M on γ_k then implies

$$\frac{\Gamma'_M(z)}{\Gamma_M(z)} \rightarrow \frac{\Gamma'(z)}{\Gamma(z)}, \quad \text{uniformly in } z \in \gamma_k.$$

Therefore,

$$\lim_{M \rightarrow \infty} \frac{1}{2\pi i} \oint_{\gamma_k} \frac{\Gamma'_M(z)}{\Gamma_M(z)} dz = \frac{1}{2\pi i} \oint_{\gamma_k} \frac{\Gamma'(z)}{\Gamma(z)} dz = 1. \quad (106)$$

Since the left-hand side is integer-valued for each M , it follows that for all sufficiently large M ,

$$\frac{1}{2\pi i} \oint_{\gamma_k} \frac{\Gamma'_M(z)}{\Gamma_M(z)} dz = 1.$$

Hence Γ_M has exactly one zero (counted with multiplicity) inside D_k .

Proof of Claim 2 (Exact spectral separation). Let $\epsilon > 0$ be fixed as in Claim 1, and recall the definitions of the outlier neighborhood $\mathcal{K}_\epsilon^*$ and the bulk neighborhood $\text{supp}_\epsilon(\mu)$. By Assumption 1(d), we have $\text{sp}(\mathbf{W}\mathbf{W}^\top) \subset \text{supp}_{\epsilon/2}(\mu)$ almost surely for large M . Since Claim 1 ensures $\mathcal{K}_\epsilon^* \cap \text{supp}_{\epsilon/2}(\mu) = \emptyset$, it follows that $\text{sp}(\mathbf{W}\mathbf{W}^\top) \cap \mathcal{K}_\epsilon^* = \emptyset$, which proves (12).

We next exclude spurious roots of the empirical master equation. Since $\Gamma(z) \rightarrow 1$ as $|z| \rightarrow \infty$, there exists $R > 0$ such that $|\Gamma(z)| > 0$ for $|z| \geq R$. By definition of $\mathcal{K}_*$, the function Γ has no real zeros on $[-R, R] \setminus (\text{supp}_\epsilon(\mu) \cup \mathcal{K}_\epsilon^*)$. Hence there exists $\alpha > 0$ such that $|\Gamma(x)| \geq \alpha$, $\forall x \in [-R, R] \setminus (\text{supp}_\epsilon(\mu) \cup \mathcal{K}_\epsilon^*)$. By the almost sure local uniform convergence $\Gamma_M \rightarrow \Gamma$ on $\mathbb{C} \setminus \text{supp}(\mu)$, for all sufficiently large M , $|\Gamma_M(x)| \geq \alpha/2$ on the same set, and thus Γ_M has no real zeros outside $\text{supp}_\epsilon(\mu) \cup \mathcal{K}_\epsilon^*$.

Now fix $\lambda \in \text{sp}(\mathbf{Y}\mathbf{Y}^\top) \setminus \text{supp}_\epsilon(\mu)$. We first claim that $\lambda > 0$ almost surely. If $0 \in \text{supp}(\mu)$, then $0 \in \text{supp}_\epsilon(\mu)$, and the assumption $\lambda \notin \text{supp}_\epsilon(\mu)$ immediately implies $\lambda \neq 0$. If instead $0 \notin \text{supp}(\mu)$, then $\mathbf{W}\mathbf{W}^\top \succ 0$ almost surely. Let $\mathbf{W} = \mathbf{U}_W \mathbf{\Sigma}_W \mathbf{V}_W^\top$ be the SVD of $\mathbf{W}$. In this case one can also show that $\mathbb{P}(0 \in \text{sp}(\mathbf{Y}\mathbf{Y}^\top)) = 0$. Indeed, conditional on all randomness other than the Haar matrix $\mathbf{V}_W$, the event $\det(\mathbf{Y}\mathbf{Y}^\top) = 0$ is characterized by the vanishing of a nontrivial polynomial in the entries of $\mathbf{V}_W$; equivalently, it defines a proper algebraic subset of the orthogonal group, which has Haar measure zero. Consequently, $\lambda > 0$ almost surely. Moreover, since $\lambda \notin \text{supp}_\epsilon(\mu)$, we also have $\lambda \notin \text{sp}(\mathbf{W}\mathbf{W}^\top)$. Fact 2 therefore applies and yields $\Gamma_M(\lambda) = 0$. On the other hand, we have shown in the above that Γ_M has no real zeros outside $\text{supp}_\epsilon(\mu) \cup \mathcal{K}_\epsilon^*$, this forces $\lambda \in \mathcal{K}_\epsilon^*$, which proves (14).

Finally, fix $\lambda_k \in \mathcal{K}_*$. Since $0 \notin \mathcal{K}_*$, we have $\lambda_k > 0$. Choose $\rho \in (0, \epsilon)$ such that $\rho < \lambda_k/2$, and let $D_k = \{z \in \mathbb{C} : |z - \lambda_k| \leq \rho\}$. By Claim 3, for all sufficiently large M the equation $\Gamma_M(z) = 0$ has exactly one zero (counted with multiplicity) in D_k . Since $\Gamma_M(\bar{z}) = \overline{\Gamma_M(z)}$, this zero must be real; denote it by $\hat{\lambda}_{k,M} \in (\lambda_k - \rho, \lambda_k + \rho) \subset (0, \infty)$. Moreover, $D_k \cap \text{sp}(\mathbf{W}\mathbf{W}^\top) = \emptyset$ for all large M . Hence Fact 2 implies that $\mathbf{Y}\mathbf{Y}^\top$ has exactly one eigenvalue in $(\lambda_k - \rho, \lambda_k + \rho)$ (counted with multiplicity). This proves (13) and completes the proof.

Proof of Claim 4 (Limiting overlaps). We prove the first convergence; the others are analogous.

Fix $\lambda_k \in \mathcal{K}_*$. By Claims 1–3, λ_k is isolated and $\lambda_{k,M} \rightarrow \lambda_k$. Let h be continuous, supported near λ_k , with $h(\lambda_k) = 1$. By weak convergence of $\nu_{M,1}$,

$$\int h d\nu_{M,1} \rightarrow \nu_1(\{\lambda_k\}),$$

and for all large M exactly one term contributes, yielding

$$\frac{1}{M} \langle \mathbf{u}_k(\mathbf{Y}), \mathbf{u}_* \rangle^2 \rightarrow \nu_1(\{\lambda_k\}).$$

For the mixed overlap, apply the same localization argument to the symmetric dilation of $\mathbf{Y}$, whose spectrum contains the pair $\pm\sigma_{k,M} = \pm\sqrt{\lambda_{k,M}}$. The signal component splits equally between the symmetric and anti-symmetric modes, producing an additional factor of 2. After normalization by $L = M + N$ and using $N/M \rightarrow \delta$, this yields

$$\frac{1}{\sqrt{MN}} \langle \mathbf{u}_k(\mathbf{Y}), \mathbf{u}_* \rangle \langle \mathbf{v}_k(\mathbf{Y}), \mathbf{v}_* \rangle \rightarrow 2 \frac{1+\delta}{\sqrt{\delta}} \nu_3(\{\sigma_k\}),$$

as claimed. $\square$

A.5 Integrals of Spectral Measures

The parallel and cross-spectral measures are instrumental in characterizing the asymptotic behavior of certain quadratic and bilinear forms that appear in algorithmic analysis. The following proposition unifies these results, showing that such forms converge to expectations with respect to the corresponding spectral measures, followed by the established weak convergence in Lemma 2.

Proposition 9. *Let $\mathbf{Y}$ be the data matrix from the model in (1), and let ν_1, ν_2, ν_3 be the limiting spectral measures Definition 5.*

- (1) *Let $h : \mathbb{R}_+ \rightarrow \mathbb{R}$ be a function that is bounded and continuous on $\text{supp}(\mu)$. The function h acts on a symmetric positive semi-definite matrix via its eigenvalues. Then, the following quadratic forms converge almost surely:*

$$\frac{1}{M} \mathbf{u}_*^\top h(\mathbf{Y}\mathbf{Y}^\top) \mathbf{u}_* \xrightarrow{a.s.} \int h(\lambda) d\nu_1(\lambda) = \langle h(\lambda), 1 \rangle_{\nu_1}, \quad (107a)$$

$$\frac{1}{N} \mathbf{v}_*^\top h(\mathbf{Y}^\top \mathbf{Y}) \mathbf{v}_* \xrightarrow{a.s.} \int h(\lambda) d\nu_2(\lambda) = \langle h(\lambda), 1 \rangle_{\nu_2}. \quad (107b)$$

- (2) *Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be an odd function that is bounded and continuous on $\text{supp}(\mu)$. The function f acts on $\mathbf{Y}$ via its singular values, i.e., $f(\mathbf{Y}) \stackrel{\text{def}}{=} \mathbf{U}_\mathbf{Y} \text{diag}(f(\sigma_i)) \mathbf{V}_\mathbf{Y}^\top$. Then, the following bilinear form converges almost surely with $L = M + N$:*

$$\frac{1}{L} \mathbf{u}_*^\top f(\mathbf{Y}) \mathbf{v}_* \xrightarrow{a.s.} \int f(\sigma) d\nu_3(\sigma) = \langle f(\sigma), 1 \rangle_{\nu_3}. \quad (108)$$

Proof. (1) We will prove only (107a) for brevity, as the proof for the other case is analogous. Let the eigen-decomposition of $\mathbf{Y}\mathbf{Y}^\top$ be $\sum_{i=1}^M \lambda_i(\mathbf{Y}\mathbf{Y}^\top) \mathbf{u}_i \mathbf{u}_i^\top$. We can express the quadratic form as:

$$\begin{aligned} \frac{1}{M} \mathbf{u}_*^\top h(\mathbf{Y}\mathbf{Y}^\top) \mathbf{u}_* &= \frac{1}{M} \sum_{i=1}^M h(\lambda_i(\mathbf{Y}\mathbf{Y}^\top)) (\mathbf{u}_*^\top \mathbf{u}_i(\mathbf{Y}\mathbf{Y}^\top))^2 \stackrel{(a)}{=} \int h(\lambda) d \left(\frac{1}{M} \sum_{i=1}^M \langle \mathbf{u}_i(\mathbf{Y}\mathbf{Y}^\top), \mathbf{u}_* \rangle^2 \delta_{\lambda_i(\mathbf{Y}\mathbf{Y}^\top)} \right) \\ &= \int h(\lambda) d\nu_{M,1}(\lambda) \stackrel{(b)}{\rightarrow} \int h(\lambda) d\nu_1(\lambda) \stackrel{(c)}{=} \langle h(\lambda), 1 \rangle_{\nu_1}. \end{aligned}$$

Here, step (a) rewrites the sum as a Stieltjes integral with respect to an empirical measure, which is precisely the definition of the finite-sample parallel measure $\nu_{M,1}$ from Definition Definition 5. Step (b) follows from the weak convergence of measures, $\nu_{M,1} \rightarrow \nu_1$, established in Lemma 2, combined with the fact that h is bounded and continuous, allowing the interchange of limit and integral. Step (c) simply expresses the integral with respect to the probability measure ν_1 as an expectation.

(2) The eigendecomposition of the symmetric dialated matrix, $\hat{\mathbf{Y}} = \mathbf{Q}_{\hat{\mathbf{Y}}} \mathbf{\Lambda}_{\hat{\mathbf{Y}}} \mathbf{Q}_{\hat{\mathbf{Y}}}^\top$, is directly related to the singular value decomposition of $\mathbf{Y} = \mathbf{U}_\mathbf{Y} [\mathbf{\Sigma}_\mathbf{Y} | \mathbf{0}] [\mathbf{V}_{\mathbf{Y},1} | \mathbf{V}_{\mathbf{Y},2}]^\top$. The eigenvalue and eigenvector matrices have the following block forms

$$\mathbf{\Lambda}_{\hat{\mathbf{Y}}} = \text{diag}(\mathbf{\Sigma}_\mathbf{Y}, -\mathbf{\Sigma}_\mathbf{Y}, \mathbf{0}_{N-M}), \quad \mathbf{Q}_{\hat{\mathbf{Y}}} = \left[\begin{array}{c|c|c} \frac{1}{\sqrt{2}} \mathbf{U}_\mathbf{Y} & -\frac{1}{\sqrt{2}} \mathbf{U}_\mathbf{Y} & \mathbf{0}_{M \times (N-M)} \\ \hline \frac{1}{\sqrt{2}} \mathbf{V}_{\mathbf{Y},1} & \frac{1}{\sqrt{2}} \mathbf{V}_{\mathbf{Y},1} & \mathbf{V}_{\mathbf{Y},2} \end{array} \right].$$

Substituting this decomposition into the definition of ν_3 with augmented vectors $\hat{\mathbf{u}}_* = [\mathbf{u}_*^\top, \mathbf{0}^\top]^\top$ and $\hat{\mathbf{v}}_* = [\mathbf{0}^\top, \mathbf{v}_*^\top]^\top$ yields the structure of the measure:

$$\nu_3 \stackrel{w}{=} \lim_{L \rightarrow \infty} \frac{1}{2L} \sum_{i=1}^M \langle \mathbf{u}_i(\mathbf{Y}), \mathbf{u}_* \rangle \langle \mathbf{v}_i(\mathbf{Y}), \mathbf{v}_* \rangle (\delta_{\sigma_i(\mathbf{Y})} - \delta_{-\sigma_i(\mathbf{Y})}). \quad (110)$$

The contribution from the null space of $\hat{\mathbf{Y}}$ vanishes because the corresponding eigenvectors, having the block form $[\mathbf{0}^\top; \mathbf{v}_k^\top]^\top$, are orthogonal to $\hat{\mathbf{u}}_*$. Now, let the singular value decomposition of $\mathbf{Y}$ be $\sum_{i=1}^M \sigma_i(\mathbf{Y}) \mathbf{u}_i(\mathbf{Y}) \mathbf{v}_i(\mathbf{Y})^\top$. A function of $\mathbf{Y}$ is defined as $f(\mathbf{Y}) = \sum_{i=1}^M f(\sigma_i(\mathbf{Y})) \mathbf{u}_i(\mathbf{Y}) \mathbf{v}_i(\mathbf{Y})^\top$. We express the bilinear form as:

$$\begin{aligned} \frac{1}{L} \mathbf{u}_*^\top f(\mathbf{Y}) \mathbf{v}_* &= \frac{1}{L} \sum_{i=1}^M f(\sigma_i(\mathbf{Y})) \langle \mathbf{u}_i(\mathbf{Y}), \mathbf{u}_* \rangle \langle \mathbf{v}_i(\mathbf{Y}), \mathbf{v}_* \rangle \\ &\stackrel{(a)}{=} \frac{1}{L} \sum_{i=1}^M \frac{\langle \mathbf{u}_i, \mathbf{u}_* \rangle \langle \mathbf{v}_i, \mathbf{v}_* \rangle}{2} (f(\sigma_i(\mathbf{Y})) - f(-\sigma_i(\mathbf{Y}))) \\ &\stackrel{(b)}{=} \int f(\sigma) d\nu_L(\sigma) \stackrel{(c)}{\rightarrow} \int f(\sigma) d\nu_3(\sigma) = \langle f(\sigma), 1 \rangle_{\nu_3}. \end{aligned}$$

Here, step (a) utilizes the fact that f is an odd function, so $f(\sigma_i) = \frac{1}{2}(f(\sigma_i) - f(-\sigma_i))$. Step (b) rewrites the sum as a Stieltjes integral with respect to the finite-sample empirical measure ν_L , as defined by the structure in (110). Step (c) follows from the weak convergence of measures, $\nu_L \rightarrow \nu_3$, established in Lemma 2. Since f is bounded and continuous, the limit and integral can be interchanged. This completes the proof. $\square$

B General OAMP Algorithm with Rotationally-Invariant Matrices

The proof of the main result (Theorem 1) follows a reduction strategy similar to that in [44]. Specifically, we transform the OAMP iteration (6), which depends on the signal matrix $\mathbf{Y}$, into an asymptotically equivalent iteration that depends only on the random matrix $\mathbf{W}$. The resulting algorithm admits a state evolution characterization, which can be established using standard conditioning techniques [23, 25, 30].

As we are not aware of prior work that directly addresses our specific setting, we include in this appendix the formulation of the general OAMP iteration with random $\mathbf{W}$ and its associated state evolution for completeness. Our derivation closely follows [30], and we therefore omit many technical details, emphasizing instead the key differences from that work.

B.1 General OAMP Iteration

We introduce a general OAMP iteration with bi-rotationally-invariant random matrix $\mathbf{W}$.

Definition 7 (General OAMP algorithm). For $t \in \mathbb{N}$, the general OAMP algorithm generates the iterates $(\mathbf{x}_t)_{t \in \mathbb{N}}$ and $(\mathbf{z}_t)_{t \in \mathbb{N}}$ via

$$\mathbf{x}_t = \Psi_t(\mathbf{W}\mathbf{W}^\top) m_t(\mathbf{x}_{\leq t-1}; \mathbf{a}) + \tilde{\Psi}_t(\mathbf{W}\mathbf{W}^\top) \mathbf{W} q_t(\mathbf{z}_{\leq t-1}; \mathbf{b}), \quad (112a)$$

$$\mathbf{z}_t = \Phi_t(\mathbf{W}^\top \mathbf{W}) q_t(\mathbf{z}_{\leq t-1}; \mathbf{b}) + \tilde{\Phi}_t(\mathbf{W}^\top \mathbf{W}) \mathbf{W}^\top m_t(\mathbf{x}_{\leq t-1}; \mathbf{a}), \quad (112b)$$

where the matrix denoising functions Ψ_t and Φ_t satisfy (while $\tilde{\Psi}_t$ and $\tilde{\Phi}_t$ do not) the following *trace-free* conditions as $M, N \rightarrow \infty$ with $M/N \rightarrow \delta \in (0, 1]$

$$\mathbb{E}[\Psi_t(\mathbf{D}_M^2)] = 0, \quad \mathbf{D}_M^2 \sim \mu, \quad (113)$$

$$\mathbb{E}[\Phi_t(\mathbf{D}_N^2)] = 0, \quad \mathbf{D}_N^2 \sim \tilde{\mu}. \quad (114)$$

and the signal denoisers $(m_t)_{t \geq 1}$ and $(q_t)_{t \geq 1}$ are *divergence-free*:

$$\mathbb{E}[\partial_i m_t(\mathbf{X}_{\leq t-1}; \mathbf{A})] = 0, \quad \forall t \in \mathbb{N}, i \in [t], \quad (115)$$

$$\mathbb{E}[\partial_i q_t(\mathbf{Z}_{\leq t-1}; \mathbf{B})] = 0, \quad \forall t \in \mathbb{N}, i \in [t]. \quad (116)$$

The random variables $(\mathbf{D}_M, \mathbf{D}_N)$ and $(\mathbf{X}_t, \mathbf{Z}_t)_{t \in \mathbb{N}}$ are to be defined in Definition 9.

Let the singular value decomposition of $\mathbf{W}$ be $\mathbf{W} = \mathbf{U}_W \mathbf{\Sigma} \mathbf{V}_W^\top$. We make a change of variables:

$$\tilde{\mathbf{x}}_t \stackrel{\text{def}}{=} \mathbf{U}_W^\top \mathbf{x}_t \quad \text{and} \quad \tilde{\mathbf{z}}_t \stackrel{\text{def}}{=} \mathbf{V}_W^\top \mathbf{z}_t. \quad (117)$$

Using the new variables, we can write the OAMP iteration into the following factorized form (see Fig. 6 for an illustration).

Definition 8 (General OAMP algorithm: factorized form). The factorized form OAMP algorithm proceeds as follows ($\forall t \in \mathbb{N}$):

$$\text{(Orthogonal transform)} \quad \mathbf{r}_t = \mathbf{U}_W^\top \mathbf{m}_t, \quad \mathbf{s}_t = \mathbf{V}_W^\top \mathbf{q}_t, \quad (118a)$$

$$\text{(Matrix denoising)} \quad \tilde{\mathbf{x}}_t = \psi_t(\mathbf{r}_t, \mathbf{D} \mathbf{s}_t | \mathbf{D} \mathbf{D}^\top), \quad \tilde{\mathbf{z}}_t = \phi_t(\mathbf{s}_t, \mathbf{D}^\top \mathbf{r}_t | \mathbf{D}^\top \mathbf{D}), \quad (118b)$$

$$\text{(Orthogonal transform)} \quad \mathbf{x}_t = \mathbf{U}_W \tilde{\mathbf{x}}_t, \quad \mathbf{z}_t = \mathbf{V}_W \tilde{\mathbf{z}}_t, \quad (118c)$$

$$\text{(Iterate denoising)} \quad \mathbf{m}_{t+1} = m_{t+1}(\mathbf{x}_{\leq t}; \mathbf{a}), \quad \mathbf{q}_{t+1} = q_{t+1}(\mathbf{z}_{\leq t}; \mathbf{b}), \quad (118d)$$

where ψ_t and ϕ_t are defined as

$$\psi_t(\mathbf{r}_t, \mathbf{D} \mathbf{s}_t | \mathbf{D} \mathbf{D}^\top) \stackrel{\text{def}}{=} \Psi_t(\mathbf{D} \mathbf{D}^\top) \mathbf{r}_t + \tilde{\Psi}_t(\mathbf{D} \mathbf{D}^\top) \mathbf{D} \mathbf{s}_t, \quad (118e)$$

$$\phi_t(\mathbf{s}_t, \mathbf{D}^\top \mathbf{r}_t | \mathbf{D}^\top \mathbf{D}) \stackrel{\text{def}}{=} \Phi_t(\mathbf{D}^\top \mathbf{D}) \mathbf{s}_t + \tilde{\Phi}_t(\mathbf{D}^\top \mathbf{D}) \mathbf{D}^\top \mathbf{r}_t. \quad (118f)$$

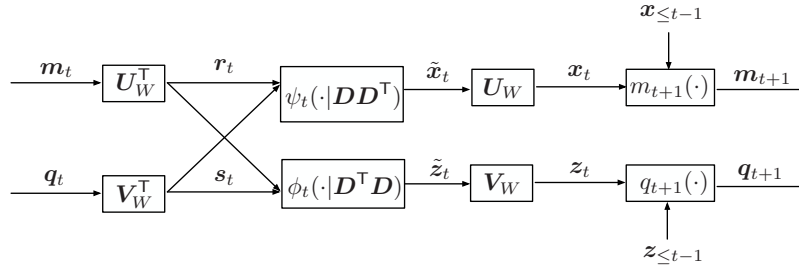


Figure 6: Diagram of the OAMP algorithm (118). m_t and q_t are divergence-free. ψ_t and ϕ_t are divergence-free with respect to the direct inputs (but not necessarily with respect to the cross input terms).

B.2 State Evolution of General OAMP Iteration

To establish a high-dimensional asymptotic characterization of the OAMP algorithm, we impose the following assumptions.

Assumption 4. The following conditions hold for the factorized OAMP algorithm defined in Def. 8:

- (1) The matrix $\mathbf{W}$ satisfies Assumption 1-(c).
- (2) The initialization vectors $\mathbf{m}_1 \in \mathbb{R}^N$ and $\mathbf{q}_1 \in \mathbb{R}^N$ are independent of $\mathbf{U}_W$ and $\mathbf{V}_W$, respectively. Moreover, $(\mathbf{m}_1, \mathbf{a}) \xrightarrow{W} (\mathbf{X}_1, \mathbf{A})$ and $(\mathbf{q}_1, \mathbf{b}) \xrightarrow{W} (\mathbf{Q}_1, \mathbf{B})$, where the limit random variables possess finite moments of all orders.

- (3) The matrix denoisers ψ_t and ϕ_t are continuous, and the iterate denoisers m_t and q_t are continuously differentiable and Lipschitz.

The iterates of the general OAMP algorithm introduced in Definition 8 admit an exact asymptotic characterization via a state evolution. Before presenting the formal result, we first describe the corresponding state evolution recursion.

Definition 9 (State evolution of OAMP: factorized form). Set $\mathbf{\Omega}_1^u = \mathbb{E}[\mathbf{M}_1^2]$ and $\mathbf{\Omega}_1^v = \mathbb{E}[\mathbf{Q}_1^2]$. Iterate the following steps for $t = 1, 2, \dots$

- (i) *Gaussian random variables:*

$$\mathbf{R}_{\leq t} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_t^u), \quad \mathbf{S}_{\leq t} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_t^v). \quad (119a)$$

- (ii) *Matrix denoising:*

$$\tilde{\mathbf{X}}_t = \psi_t(\mathbf{R}_t, \mathbf{D}_M \mathbf{S}_t | \mathbf{D}_M^2), \quad \tilde{\mathbf{Z}}_t = \phi_t(\mathbf{S}_t, \mathbf{D}_N \mathbf{R}_t | \mathbf{D}_N^2), \quad (119b)$$

where $\mathbf{D}_M$, $\mathbf{R}_{\leq t}$ and $\mathbf{S}_{\leq t}$ are mutually independent. Moreover, $\mathbf{D}_N \stackrel{d}{=} \mathbf{H} \mathbf{D}_M$, where $\mathbf{H} \sim \text{Ber}(\delta)$ is independent of other random variables.

- (iii) *Covariance update:*

$$\mathbf{\Sigma}_t^u = \mathbb{E}[\tilde{\mathbf{X}}_{\leq t} \tilde{\mathbf{X}}_{\leq t}^\top], \quad \mathbf{\Sigma}_t^v = \mathbb{E}[\tilde{\mathbf{Z}}_{\leq t} \tilde{\mathbf{Z}}_{\leq t}^\top]. \quad (119c)$$

- (iv) *Gaussian random variables:*

$$\mathbf{X}_{\leq t} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_t^u), \quad \mathbf{Z}_{\leq t} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_t^v). \quad (119d)$$

- (v) *Iterate denoising:*

$$\mathbf{M}_{t+1} = m_{t+1}(\mathbf{X}_{\leq t}; \mathbf{A}), \quad \mathbf{Q}_{t+1} = q_{t+1}(\mathbf{Z}_{\leq t}; \mathbf{B}), \quad (119e)$$

where $\mathbf{X}_{\leq t} \sim \mathcal{N}(\mathbf{0}_t, \mathbf{\Sigma}_t^u) \perp\!\!\!\perp (\mathbf{A}, \mathbf{M}_1)$ and $\mathbf{Z}_{\leq t} \sim \mathcal{N}(\mathbf{0}_t, \mathbf{\Sigma}_t^v) \perp\!\!\!\perp (\mathbf{B}, \mathbf{Q}_1)$.

- (vi) *Covariance update:*

$$\mathbf{\Omega}_{t+1}^u = \mathbb{E}[\mathbf{M}_{\leq t+1} \mathbf{M}_{\leq t+1}^\top], \quad \mathbf{\Omega}_{t+1}^v = \mathbb{E}[\mathbf{Q}_{\leq t+1} \mathbf{Q}_{\leq t+1}^\top]. \quad (119f)$$

In the above equations, with slight abuse of notations, ψ_t and ϕ_t are defined as

$$\psi_t(\mathbf{R}_t, \mathbf{D}_M \mathbf{S}_t | \mathbf{D}_M^2) \stackrel{\text{def}}{=} \Psi_t(\mathbf{D}_M^2) \mathbf{R}_t + \tilde{\Psi}_t(\mathbf{D}_M^2) \mathbf{D}_M \mathbf{S}_t, \quad (120a)$$

$$\phi_t(\mathbf{S}_t, \mathbf{D}_N \mathbf{R}_t | \mathbf{D}_N^2) \stackrel{\text{def}}{=} \Phi_t(\mathbf{D}_N^2) \mathbf{S}_t + \tilde{\Phi}_t(\mathbf{D}_N^2) \mathbf{D}_N \mathbf{R}_t. \quad (120b)$$

Remark 9. Below are some remarks about Definition 9:

- Whenever the collections of random variables $\mathbf{R}_{\leq t}$, $\mathbf{S}_{\leq t}$ and $(\mathbf{D}_M, \mathbf{D}_N)$ appear jointly, they are understood as mutually independent. This independence is used in the definitions of the covariance matrices $\mathbf{\Sigma}_t^u$ and $\mathbf{\Sigma}_t^v$.
- Explicit formulas for $\{\mathbf{\Omega}_{t+1}^u[i, j], i \leq j\}$ are given below:

$$\mathbf{\Omega}_{t+1}^u[i, j] = \begin{cases} \mathbb{E}[\mathbf{M}_1^2] & \text{for } i = j = 1, \\ \mathbb{E}[\mathbf{M}_1 \cdot f_{j-1}(\mathbf{X}_{\leq j-1}; \mathbf{A})] & \text{for } i = 1, j > 1, \\ \mathbb{E}[f_{i-1}(\mathbf{X}_{\leq i-1}; \mathbf{A}) f_{j-1}(\mathbf{X}_{\leq j-1}; \mathbf{A})] & \text{for } 1 < i \leq j \leq t+1. \end{cases} \quad (121a)$$

In the second expectation, $(\mathbf{M}_1, \mathbf{A}) \perp\!\!\!\perp \mathbf{X}_{\leq t}$. Similar formulas apply analogously to $\mathbf{\Omega}_{t+1}^v$.

The orthogonality property stated below explains the term ‘‘Orthogonal AMP’’ and serves as a key ingredient in the proof of Theorem B.3. This property arises from the divergence-free and trace-free constraints imposed on the OAMP denoisers.

Lemma 3 (Orthogonality). *Suppose that the covariance matrices $(\mathbf{\Sigma}_t^u, \mathbf{\Sigma}_t^v)_{t \in \mathbb{N}}$ in Definition 9 are non-singular. Then, the state evolution random variables in Definition 9 satisfy*

$$\mathbb{E}[\mathbf{R}_i \tilde{\mathbf{X}}_j] = 0, \quad \mathbb{E}[\mathbf{S}_i \tilde{\mathbf{Z}}_j] = 0, \quad \forall i, j \in \mathbb{N}, \quad (122a)$$

$$\mathbb{E}[\mathbf{X}_i \mathbf{M}_j] = 0, \quad \mathbb{E}[\mathbf{Z}_i \mathbf{Q}_j] = 0, \quad \forall i, j \in \mathbb{N}. \quad (122b)$$

Proof. For $\mathbb{E}[\mathbf{R}_i \tilde{\mathbf{X}}_j] = 0$, we substitute the definition of $\tilde{\mathbf{X}}_j$:

$$\begin{aligned} \mathbb{E}[\mathbf{R}_i \tilde{\mathbf{X}}_j] &= \mathbb{E}\left[\mathbf{R}_i \left(\Psi_j(\mathbf{D}_M^2) \mathbf{R}_j + \tilde{\Psi}_j(\mathbf{D}_M^2) \mathbf{D}_M \mathbf{S}_j\right)\right], \quad \forall i, j \in \mathbb{N}, \\ &\stackrel{a}{=} \mathbb{E}[\mathbf{R}_i \mathbf{R}_j] \cdot \mathbb{E}[\Psi_j(\mathbf{D}_M^2)] + \mathbb{E}[\mathbf{R}_i] \cdot \mathbb{E}[\mathbf{S}_j] \cdot \mathbb{E}[\Psi_j(\mathbf{D}_M^2) \mathbf{D}_M] \\ &\stackrel{b}{=} 0, \end{aligned}$$

where step (a) is due to the independence of $(\mathbf{R}_i, \mathbf{R}_j)$, $\mathbf{S}_j$ and $\mathbf{D}_M$, and step (b) is due to the trace-free property $\mathbb{E}[\Psi_i(\mathbf{D}_M^2)] = 0$ and the fact that $\mathbf{R}_i, \mathbf{S}_j$ have zero mean. The term $\mathbb{E}[\mathbf{X}_i \mathbf{M}_j]$ can be computed as

$$\mathbb{E}[\mathbf{X}_i \mathbf{M}_j] = \begin{cases} \mathbb{E}[\mathbf{X}_i] \cdot \mathbb{E}[\mathbf{M}_1] \stackrel{a}{=} 0, & \text{if } j = 1, \\ \sum_{k=1}^j \mathbb{E}[\mathbf{X}_i \mathbf{X}_k] \cdot \mathbb{E}[\partial_k m_j(\mathbf{X}_{\leq j-1}; \mathbf{A})] \stackrel{b}{=} 0, & \text{if } j > 1, \end{cases}$$

where step (a) is due to the independence between $\mathbf{X}_i$ and $\mathbf{M}_1$, and step (b) is a consequence of the multi-variate Stein's lemma and the divergence-free properties of $m_j(\cdot)$. The other two properties can be proved analogously and omitted for brevity. $\square$

The following theorem shows that, under Assumption 4, the performance of the OAMP algorithm is governed by the corresponding state evolution equations. All convergence statements are understood in the limit $M, N \rightarrow \infty$ with $M/N \rightarrow \delta \in (0, 1]$. Its proof is deferred to Section B.3.

Theorem 3 (State evolution characterization of OAMP: factorized form). *Consider the OAMP algorithm in Definition 8 with initialization $\mathbf{m}_1 \in \mathbb{R}^M$ and $\mathbf{q}_t \in \mathbb{R}^N$. Suppose Assumption 4 holds. Let the covariance matrices $(\mathbf{\Omega}_t^u, \mathbf{\Omega}_t^v, \mathbf{\Sigma}_{t+1}^u, \mathbf{\Sigma}_{t+1}^v)_{t \in \mathbb{N}}$ be defined as in Definition 9. Assume additionally that these covariance matrices are non-singular for all fixed $t \in \mathbb{N}$. The following hold for all fixed $t \in \mathbb{N}$:*

(a)

$$(\mathbf{r}_{\leq t}, \mathbf{D} \mathbf{s}_{\leq t}, \mathbf{d}_M) \xrightarrow{W} (\mathbf{R}_{\leq t}, \mathbf{D}_M \mathbf{S}_{\leq t}, \mathbf{D}_M), \quad (125a)$$

$$(\mathbf{s}_{\leq t}, \mathbf{D}^\top \mathbf{z}_{\leq t}, \mathbf{d}_N) \xrightarrow{W} (\mathbf{S}_{\leq t}, \mathbf{D}_N \mathbf{Z}_{\leq t}, \mathbf{D}_N), \quad (125b)$$

where $\mathbf{R}_{\leq t} \sim \mathcal{N}(\mathbf{0}_t, \mathbf{\Omega}_t^u)$, $\mathbf{S}_{\leq t} \sim \mathcal{N}(\mathbf{0}_t, \mathbf{\Omega}_t^v)$; and $\mathbf{R}_{\leq t}$, $\mathbf{S}_{\leq t}$ and $(\mathbf{D}_M, \mathbf{D}_N)$ are mutually independent.

(b)

$$(\mathbf{r}_{\leq t}, \tilde{\mathbf{x}}_{\leq t}, \mathbf{d}_M) \xrightarrow{W} (\mathbf{R}_{\leq t}, \tilde{\mathbf{X}}_{\leq t}, \mathbf{D}_M), \quad (126a)$$

$$(\mathbf{s}_{\leq t}, \tilde{\mathbf{z}}_{\leq t}, \mathbf{d}_N) \xrightarrow{W} (\mathbf{S}_{\leq t}, \tilde{\mathbf{Z}}_{\leq t}, \mathbf{D}_N), \quad (126b)$$

where $\mathbf{R}_{\leq t} \sim \mathcal{N}(\mathbf{0}_t, \mathbf{\Omega}_t^u)$, $\mathbf{S}_{\leq t} \sim \mathcal{N}(\mathbf{0}_t, \mathbf{\Omega}_t^v)$; $\tilde{\mathbf{X}}_t = \psi_t(\mathbf{R}_t, \mathbf{D}_M \mathbf{S}_t)$, $\tilde{\mathbf{Z}}_t = \phi_t(\mathbf{S}_t, \mathbf{D}_N \mathbf{R}_t)$ with $\mathbf{R}_{\leq t}$, $\mathbf{S}_{\leq t}$ and $(\mathbf{D}_M, \mathbf{D}_N)$ mutually independent.

(c)

$$(\mathbf{x}_{\leq t}, \mathbf{m}_{\leq t+1}; \mathbf{a}) \xrightarrow{W} (\mathbf{X}_{\leq t}, \mathbf{M}_{\leq t+1}, \mathbf{A}), \quad (127a)$$

$$(\mathbf{z}_{\leq t}, \mathbf{q}_{\leq t+1}; \mathbf{b}) \xrightarrow{W} (\mathbf{Z}_{\leq t}, \mathbf{Q}_{\leq t+1}, \mathbf{B}), \quad (127b)$$

where $\mathbf{X}_{\leq t} \sim \mathcal{N}(\mathbf{0}_t, \mathbf{\Sigma}_t^u)$, $\mathbf{M}_{t+1} = m_{t+1}(\mathbf{X}_{\leq t}; \mathbf{A})$, $\mathbf{Z}_{\leq t} \sim \mathcal{N}(\mathbf{0}_t, \mathbf{\Sigma}_t^v)$, $\mathbf{Q}_{t+1} = q_{t+1}(\mathbf{Z}_{\leq t}; \mathbf{B})$. Moreover, $\mathbf{X}_{\leq t} \perp\!\!\!\perp (\mathbf{M}_1, \mathbf{A})$, and $\mathbf{Z}_{\leq t} \perp\!\!\!\perp (\mathbf{Q}_1, \mathbf{B})$.

Remark 10 (Additional asymptotic independence). The proof of Theorem 3 actually shows that for each fixed t ,

$$(\mathbf{r}_t, \tilde{\mathbf{x}}_t, \mathbf{x}_t, \mathbf{m}_{t+1}) \xrightarrow{W} (\mathbf{R}_t, \tilde{\mathbf{X}}_t, \mathbf{X}_t, \mathbf{M}_{t+1}),$$

with $(\mathbf{R}_t, \tilde{\mathbf{X}}_t) \perp (\mathbf{X}_t, \mathbf{M}_{t+1})$, and analogously for $(\mathbf{s}_t, \tilde{\mathbf{z}}_t, \mathbf{z}_t, \mathbf{q}_{t+1})$. This refinement is not used in this paper and is therefore omitted from Theorem 3.

Theorem 3 relies on the assumption that the covariance matrices $(\boldsymbol{\Omega}_t^u, \boldsymbol{\Omega}_t^v, \boldsymbol{\Sigma}_{t+1}^u, \boldsymbol{\Sigma}_{t+1}^v)_{t \in \mathbb{N}}$ are non-singular, which can be hard to check. However, by the perturbation argument in [67] (see also [30, Corollary 4.4] and [68, Appendix D.1]), one may drop this assumption at the cost of weakening convergence from W_p for all $p \geq 1$ to W_2 . We state this result in the next corollary; its proof is almost identical to that of [30, Corollary 4.4] and is omitted.

Corollary 1 (Removing non-degeneracy assumption). *Under Assumption 4, the statements (125)–(127) continue to hold when convergence is replaced by W_2 .*

Finally, the state evolution associated with the factorized-form OAMP (Definition 8) can be transformed back to the original OAMP algorithm (Definition 7). For ease of reference, we record it below.

Theorem 4 (State evolution characterization of general OAMP: original form). *Consider the OAMP algorithm in Definition 7 with initialization $\mathbf{m}_1 := m_1(\mathbf{x}_0; \mathbf{a})$ and $\mathbf{q}_1 := q_1(\mathbf{z}_0; \mathbf{b})$. Suppose Assumption 4 holds. The following hold for all fixed $t \in \mathbb{N}$:*

$$(\mathbf{x}_1, \dots, \mathbf{x}_t, \mathbf{a}) \xrightarrow{W_2} (\mathbf{X}_1, \dots, \mathbf{X}_t, \mathbf{A}), \quad (128a)$$

$$(\mathbf{z}_1, \dots, \mathbf{z}_t, \mathbf{b}) \xrightarrow{W_2} (\mathbf{Z}_1, \dots, \mathbf{Z}_t, \mathbf{B}), \quad (128b)$$

where $(\mathbf{X}_1, \dots, \mathbf{X}_t) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t^u)$ is independent of $\mathbf{A}$, $(\mathbf{Z}_1, \dots, \mathbf{Z}_t) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_t^v)$ is independent of $\mathbf{B}$. Define $\Sigma_{u,st} \stackrel{\text{def}}{=} \mathbb{E}[\mathbf{X}_s \mathbf{X}_t^T]$ and $\Sigma_{v,st} \stackrel{\text{def}}{=} \mathbb{E}[\mathbf{Z}_s \mathbf{Z}_t^T]$, for $s, t \in \mathbb{N}$. Then,

$$\Sigma_{u,st} = \mathbb{E}[\Psi_s(\mathbf{D}_M^2) \Psi_t(\mathbf{D}_M^2)] \cdot \mathbb{E}[\mathbf{M}_s \mathbf{M}_t^T] + \mathbb{E}[\tilde{\Psi}_s(\mathbf{D}_M^2) \tilde{\Psi}_t(\mathbf{D}_M^2) \mathbf{D}_M^2] \cdot \mathbb{E}[\mathbf{Q}_s \mathbf{Q}_t^T], \quad (128c)$$

$$\Sigma_{v,st} = \mathbb{E}[\Phi_s(\mathbf{D}_N^2) \Phi_t(\mathbf{D}_N^2)] \cdot \mathbb{E}[\mathbf{Q}_s \mathbf{Q}_t^T] + \mathbb{E}[\tilde{\Phi}_s(\mathbf{D}_N^2) \tilde{\Phi}_t(\mathbf{D}_N^2) \mathbf{D}_N^2] \cdot \mathbb{E}[\mathbf{M}_s \mathbf{M}_t^T], \quad (128d)$$

where the random variables $(\mathbf{D}_M, \mathbf{D}_N)$ are defined as in Definition 9, and $\forall s \in \mathbb{N}$

$$\mathbf{M}_{s+1} \stackrel{\text{def}}{=} m_{s+1}(\mathbf{X}_1, \dots, \mathbf{X}_s; \mathbf{A}), \quad (128e)$$

$$\mathbf{Q}_{s+1} \stackrel{\text{def}}{=} q_{s+1}(\mathbf{Z}_1, \dots, \mathbf{Z}_s; \mathbf{B}). \quad (128f)$$

Theorem 4 follows directly from Theorem 3 and Corollary 1, and its proof is therefore omitted.

B.3 Proof of Theorem 3

Let $(t.a)$ and $(t.b)$ denote the claims (a) and (b) for iteration t . We prove by induction on $t = 1, 2, \dots$

Base case: proof of claim (1.a) – (1.c) Recall that $\mathbf{r}_1 = \mathbf{U}_W^\top \mathbf{m}_1$, $\mathbf{s}_1 = \mathbf{V}_W^\top \mathbf{q}_1$, where $\mathbf{U}_W \in \mathbb{O}(M)$ and $\mathbf{V}_W \in \mathbb{O}(N)$ are independent Haar random matrices. Based on standard properties of W convergence of empirical probability measures and Haar random matrix (see [30, Appendix E and Appendix F]), together with the fact that $\mathbf{U}_W$ and $\mathbf{V}_W$ are independent and the entries of $(\mathbf{d}_M, \mathbf{d}_N)$ are bounded by dimension-independent constants, we obtain

$$(\mathbf{r}_1, \mathbf{D} \mathbf{s}_1, \mathbf{d}_M) \xrightarrow{W} (\mathbf{R}_1, \mathbf{D}_M \mathbf{S}_1, \mathbf{D}_M), \quad (129a)$$

$$(\mathbf{s}_1, \mathbf{D}^\top \mathbf{r}_1, \mathbf{d}_N) \xrightarrow{W} (\mathbf{S}_1, \mathbf{D}_N \mathbf{R}_1, \mathbf{D}_N), \quad (129b)$$

where $\mathbf{R}_1 \sim \mathcal{N}(\mathbf{0}, \mathbb{E}[\mathbf{M}_1^2])$, $\mathbf{S}_1 \sim \mathcal{N}(\mathbf{0}, \mathbb{E}[\mathbf{Q}_1^2])$ and $(\mathbf{D}_M, \mathbf{D}_N)$ are mutually independent. Recall that

$$\tilde{\mathbf{x}}_1 = \Psi_1(\mathbf{d}_M^2) \circ \mathbf{r}_1 + \tilde{\Psi}_1(\mathbf{d}_M^2) \circ (\mathbf{D} \mathbf{s}_1), \quad (130a)$$

$$\tilde{\mathbf{z}}_1 = \Phi_1(\mathbf{d}_N^2) \circ \mathbf{s}_1 + \tilde{\Phi}_1(\mathbf{d}_N^2) \circ (\mathbf{D}^\top \mathbf{r}_1), \quad (130b)$$

where $\circ$ denotes Hadamard product. Combining (129) and (130), and further noting that $(\Psi_1, \tilde{\Psi}_1, \Phi_1, \tilde{\Phi}_1)$ are continuous and the entries of $(\mathbf{d}_M, \mathbf{d}_N)$ are uniformly bounded, it is straightforward to show that (cf. [30, Proposition E.2])

$$(\mathbf{r}_1, \tilde{\mathbf{x}}_1, \mathbf{d}_M) \xrightarrow{W} (\mathbf{R}_1, \tilde{\mathbf{X}}_1, \mathbf{D}_M), \quad (131a)$$

$$(\mathbf{s}_1, \tilde{\mathbf{z}}_1, \mathbf{d}_N) \xrightarrow{W} (\mathbf{S}_1, \tilde{\mathbf{Z}}_1, \mathbf{D}_N), \quad (131b)$$

where the joint distributions of the random variables appeared above are described in Definition 9. This proves claim (1.a) and (1.b).

To prove claim (1.c), we identify the conditional distribution of $(\mathbf{U}_W, \mathbf{V}_W)$ given $(\mathbf{r}_1, \mathbf{s}_1)$, equivalently given the linear constraints $\mathbf{r}_1 = \mathbf{U}_W^\top \mathbf{m}_1$ and $\mathbf{s}_1 = \mathbf{V}_W^\top \mathbf{q}_1$. Let $\mathcal{G}_1 \stackrel{\text{def}}{=} \sigma(\mathbf{r}_1, \mathbf{s}_1)$ be the σ -algebra generated by $(\mathbf{r}_1, \mathbf{s}_1)$. For all sufficiently large M, N , by [23, Lemma 4] [25] there exist independent Haar matrices $\tilde{\mathbf{U}}_W \in \mathbb{O}(M-1)$ and $\tilde{\mathbf{V}}_W \in \mathbb{O}(N-1)$, independent of $\mathcal{G}_1$, such that under the conditional law $\mathbb{P}(\cdot \mid \mathcal{G}_1)$ we have

$$\mathbf{U}_W \stackrel{d}{=} \frac{\mathbf{m}_1 \mathbf{r}_1^\top}{\|\mathbf{m}_1\|^2} + \Pi_{\mathbf{m}_1}^\perp \tilde{\mathbf{U}}_W \Pi_{\mathbf{r}_1}^\perp, \quad (132)$$

$$\mathbf{V}_W \stackrel{d}{=} \frac{\mathbf{q}_1 \mathbf{s}_1^\top}{\|\mathbf{q}_1\|^2} + \Pi_{\mathbf{q}_1}^\perp \tilde{\mathbf{V}}_W \Pi_{\mathbf{s}_1}^\perp, \quad (133)$$

where $\Pi_x^\perp$ denotes the orthogonal projector onto $\mathbf{x}^\perp$. Consequently, still under $\mathbb{P}(\cdot \mid \mathcal{G}_1)$, the iterates $\mathbf{x}_1 = \mathbf{U}_W \tilde{\mathbf{x}}_1$ and $\mathbf{z}_1 = \mathbf{V}_W \tilde{\mathbf{z}}_1$ admit the decompositions

$$\mathbf{x}_1 \stackrel{d}{=} \frac{\mathbf{m}_1 \mathbf{r}_1^\top \tilde{\mathbf{x}}_1}{\|\mathbf{m}_1\|^2} + \Pi_{\mathbf{m}_1}^\perp \tilde{\mathbf{U}}_W \Pi_{\mathbf{r}_1}^\perp \tilde{\mathbf{x}}_1, \quad (134)$$

$$\mathbf{z}_1 \stackrel{d}{=} \frac{\mathbf{q}_1 \mathbf{s}_1^\top \tilde{\mathbf{z}}_1}{\|\mathbf{q}_1\|^2} + \Pi_{\mathbf{q}_1}^\perp \tilde{\mathbf{V}}_W \Pi_{\mathbf{s}_1}^\perp \tilde{\mathbf{z}}_1. \quad (135)$$

From claim (1.a), we have

$$\frac{1}{M} \mathbf{r}_1^\top \tilde{\mathbf{x}}_1 \xrightarrow{\text{a.s.}} \mathbb{E}[\mathbf{R}_1 \tilde{\mathbf{X}}_1] \stackrel{a}{=} 0, \quad (136a)$$

$$\frac{1}{M} \mathbf{s}_1^\top \tilde{\mathbf{z}}_1 \xrightarrow{\text{a.s.}} \mathbb{E}[\mathbf{S}_1 \tilde{\mathbf{Z}}_1] \stackrel{b}{=} 0, \quad (136b)$$

where step (a) and step (b) are a consequence of Lemma 3. Hence, by arguments analogous to those employed in the proof of [30, Lemma A.4], we obtain:

$$(\mathbf{x}_1, \mathbf{m}_1, \mathbf{m}_2, \mathbf{a}) \xrightarrow{W} (\mathbf{X}_1, \mathbf{M}_1, \mathbf{M}_2, \mathbf{A}), \quad (137a)$$

$$(\mathbf{z}_1, \mathbf{q}_1, \mathbf{q}_2, \mathbf{b}) \xrightarrow{W} (\mathbf{Z}_1, \mathbf{Q}_1, \mathbf{Q}_2, \mathbf{B}), \quad (137b)$$

where $\mathbf{X}_1 \sim \mathcal{N}(0, \mathbb{E}[\tilde{\mathbf{X}}_1^2])$, $\mathbf{M}_2 = m_2(\mathbf{X}_1; \mathbf{A})$, $\mathbf{X}_1 \perp (\mathbf{M}_1, \mathbf{A})$, and $\mathbf{Z}_1 \sim \mathcal{N}(0, \mathbb{E}[\tilde{\mathbf{Z}}_1^2])$, $\mathbf{Q}_2 = q_2(\mathbf{Z}_1; \mathbf{B})$, $\mathbf{Z}_1 \perp (\mathbf{Q}_1, \mathbf{B})$.

Induction step: proof of claim $(t+1.a)$ We shall assume that the claims hold up to $(t.e)$. In what follows, we analyze the distribution of $\mathbf{r}_{t+1}$. We introduce the following matrix notations for the iterates:

$$\mathbf{M}_t \stackrel{\text{def}}{=} [\mathbf{m}_1, \dots, \mathbf{m}_t].$$

We define the matrices $\mathbf{X}_t, \mathbf{R}_t, \tilde{\mathbf{X}}_t, \mathbf{Q}_t, \mathbf{Z}_t, \mathbf{S}_t$ and $\tilde{\mathbf{Z}}_t$ analogously. Using the matrix notations, the OAMP iterates (118) can be written as follows

$$[\mathbf{M}_t, \mathbf{X}_t] = \mathbf{U}_W [\mathbf{R}_t, \tilde{\mathbf{X}}_t], \quad (138a)$$

$$[\mathbf{Q}_t, \mathbf{Z}_t] = \mathbf{V}_W [\mathbf{S}_t, \tilde{\mathbf{Z}}_t]. \quad (138b)$$

Let $\mathcal{G}_t$ be the σ -algebra generated by the iterates up to $\mathbf{x}_t$. Note that by claim (t.a) and Lemma 3,

$$\frac{1}{M} \begin{bmatrix} \mathbf{M}_t^\top \mathbf{M}_t & \mathbf{M}_t^\top \mathbf{X}_t \\ \mathbf{X}_t^\top \mathbf{M}_t & \mathbf{X}_t^\top \mathbf{X}_t \end{bmatrix} \xrightarrow{\text{a.s.}} \begin{bmatrix} \boldsymbol{\Omega}_t^u & \mathbf{0}_{t \times t} \\ \mathbf{0}_{t \times t} & \boldsymbol{\Sigma}_t^u \end{bmatrix}, \quad \frac{1}{M} \begin{bmatrix} \mathbf{M}_t^\top \mathbf{m}_{t+1} \\ \mathbf{X}_t^\top \mathbf{m}_{t+1} \end{bmatrix} \xrightarrow{\text{a.s.}} \begin{bmatrix} \boldsymbol{\omega}_{t+1}^u \\ \mathbf{0}_{t \times 1} \end{bmatrix}, \quad (139a)$$

where the covariance matrices $\boldsymbol{\Omega}_t^u$ and $\boldsymbol{\Sigma}_t^u$ are defined in Definition 9, and $\boldsymbol{\omega}_{t+1}^u$ is defined by

$$\boldsymbol{\omega}_{t+1}^u[i] \stackrel{\text{def}}{=} \mathbb{E}[\mathbf{M}_i \mathbf{M}_{t+1}], \quad \forall i \in [t]. \quad (139b)$$

Moreover, $\boldsymbol{\Omega}_t^u$ and $\boldsymbol{\Sigma}_t^u$ are invertible by assumption. Hence, the following matrix is invertible for all sufficiently large M, N :

$$\begin{bmatrix} \mathbf{M}_t^\top \mathbf{M}_t & \mathbf{M}_t^\top \mathbf{X}_t \\ \mathbf{X}_t^\top \mathbf{M}_t & \mathbf{X}_t^\top \mathbf{X}_t \end{bmatrix}.$$

By Lemma 4 in [23] and [25], the conditional laws of $\mathbf{U}_W$ and $\mathbf{V}_W$ given $\mathcal{G}_t$ can be represented as

$$\mathbf{U}_W \stackrel{\text{d}}{=} [\mathbf{M}_t, \mathbf{X}_t] \begin{bmatrix} \mathbf{M}_t^\top \mathbf{M}_t & \mathbf{M}_t^\top \mathbf{X}_t \\ \mathbf{X}_t^\top \mathbf{M}_t & \mathbf{X}_t^\top \mathbf{X}_t \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_t^\top \\ \widetilde{\mathbf{X}}_t^\top \end{bmatrix} + \Pi_{[\mathbf{M}_t, \mathbf{X}_t]}^\perp \widetilde{\mathbf{U}}_W \Pi_{[\mathbf{R}_t, \widetilde{\mathbf{X}}_t]}^\perp, \quad (140a)$$

$$\mathbf{V}_W \stackrel{\text{d}}{=} [\mathbf{Q}_t, \mathbf{Z}_t] \begin{bmatrix} \mathbf{Q}_t^\top \mathbf{Q}_t & \mathbf{Q}_t^\top \mathbf{Z}_t \\ \mathbf{Z}_t^\top \mathbf{Q}_t & \mathbf{Z}_t^\top \mathbf{Z}_t \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{S}_t^\top \\ \widetilde{\mathbf{Z}}_t^\top \end{bmatrix} + \Pi_{[\mathbf{Q}_t, \mathbf{Z}_t]}^\perp \widetilde{\mathbf{V}}_W \Pi_{[\mathbf{S}_t, \widetilde{\mathbf{Z}}_t]}^\perp, \quad (140b)$$

where $\widetilde{\mathbf{U}}_W \in \mathbb{O}(M-2t)$ and $\widetilde{\mathbf{V}}_W \in \mathbb{O}(N-2t)$ are Haar-distributed orthogonal matrices, which are mutually independent and independent of $\mathcal{G}_t$. $\Pi_{[\mathbf{M}_t, \mathbf{X}_t]}^\perp \in \mathbb{R}^{M \times (M-2t)}$ is a matrix whose columns form an orthonormal basis for $(\text{col}[\mathbf{M}_t, \mathbf{X}_t])^\perp$. Other projection matrices are defined analogously. Hence, conditional on $\mathcal{G}_t$, the iterate $\mathbf{r}_{t+1} = \mathbf{U}_W^\top \mathbf{m}_{t+1}$ can be written as

$$\mathbf{r}_{t+1} \stackrel{\text{d}}{=} \mathbf{r}_{t+1}^\parallel + \mathbf{r}_{t+1}^\perp, \quad (141a)$$

with

$$\mathbf{r}_{t+1}^\parallel \stackrel{\text{def}}{=} [\mathbf{R}_t, \widetilde{\mathbf{X}}_t] \begin{bmatrix} \mathbf{M}_t^\top \mathbf{M}_t & \mathbf{M}_t^\top \mathbf{X}_t \\ \mathbf{X}_t^\top \mathbf{M}_t & \mathbf{X}_t^\top \mathbf{X}_t \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{M}_t^\top \mathbf{m}_{t+1} \\ \mathbf{X}_t^\top \mathbf{m}_{t+1} \end{bmatrix}, \quad (141b)$$

$$\mathbf{r}_{t+1}^\perp \stackrel{\text{def}}{=} \Pi_{[\mathbf{R}_t, \widetilde{\mathbf{X}}_t]}^\perp \widetilde{\mathbf{U}}_W^\top \Pi_{[\mathbf{M}_t, \mathbf{X}_t]}^\perp \mathbf{m}_{t+1}. \quad (141c)$$

From (139) and the asymptotic orthogonality stated in Lemma 3, and by arguments analogous to those in the proof of [30, Lemma A.4], one obtains:

$$\mathbf{r}_{t+1}^\parallel \xrightarrow{W} \mathbf{R}_{t+1}^\parallel \stackrel{\text{def}}{=} [\mathbf{R}_1, \dots, \mathbf{R}_t] (\boldsymbol{\Omega}_t^u)^{-1} \boldsymbol{\omega}_{t+1}^u, \quad (142a)$$

$$\mathbf{r}_{t+1}^\perp \xrightarrow{W} \mathbf{R}_{t+1}^\perp \sim \mathcal{N}(0, \mathbb{E}[(\mathbf{M}_{t+1}^\perp)^2]), \quad (142b)$$

with $\mathbf{R}_{t+1}^\perp$ is independent of $(\mathbf{R}_1, \dots, \mathbf{R}_t)$, and $\mathbf{M}_{t+1}^\perp$ denotes the projection (in the L_2 sense) of $\mathbf{M}_{t+1}$ onto the orthogonal space of $\text{span}(\mathbf{M}_1, \dots, \mathbf{M}_t)$: $\mathbf{M}_{t+1}^\perp \stackrel{\text{def}}{=} \Pi_{(\mathbf{M}_1, \dots, \mathbf{M}_t)}^\perp (\mathbf{M}_{t+1})$. The variance of $\mathbf{R}_{t+1}^\perp$ can be further expressed as $\mathbb{E}[(\mathbf{R}_{t+1}^\perp)^2] = \mathbb{E}[(\mathbf{M}_{t+1}^\perp)^2] = \mathbb{E}[\mathbf{M}_{t+1}^2] - (\boldsymbol{\omega}_{t+1}^u)^\top (\boldsymbol{\Omega}_t^u)^{-1} \boldsymbol{\omega}_{t+1}^u$. One can further obtain the convergence of the joint empirical law of $(\mathbf{r}_{\leq t}, \mathbf{r}_{t+1}, \mathbf{d}_M)$ based on the same reasoning as those in [30]. First, by induction hypothesis

$$(\mathbf{r}_{\leq t}, \mathbf{d}_M) \xrightarrow{W} (\mathbf{R}_{\leq t}, \mathbf{D}_M). \quad (143)$$

Note that $\mathbf{r}_{t+1}^\parallel$ is a linear transform of $(\mathbf{r}, \dots, \mathbf{r}_t, \mathbf{d}_M)$ up to an error term that vanish in W_p for every $p \geq 1$. Applying [30, Proposition E.4] yields

$$(\mathbf{r}_{\leq t}, \mathbf{r}_{t+1}^\parallel, \mathbf{d}_M) \xrightarrow{W} (\mathbf{R}_{\leq t}, \mathbf{R}_{t+1}^\parallel, \mathbf{D}_M). \quad (144)$$

Using [30, Proposition F.2], we obtain

$$(\mathbf{r}_{\leq t}, \mathbf{r}_{t+1}^\parallel + \mathbf{r}_{t+1}^\perp, \mathbf{d}_M) \xrightarrow{W} (\mathbf{R}_{\leq t}, \mathbf{R}_{t+1}^\parallel + \mathbf{R}_{t+1}^\perp, \mathbf{D}_M). \quad (145)$$

Let $\mathbf{R}_{t+1} \stackrel{\text{def}}{=} \mathbf{R}_{t+1}^{\parallel} + \mathbf{R}_{t+1}^{\perp}$. It is straightforward to check that $\mathbf{R}_{\leq t+1} \sim \mathcal{N}(\mathbf{0}_{t+1}, \mathbf{\Omega}_{t+1}^u)$. We can apply exactly the same arguments to $\mathbf{s}_{t+1}$. Note that the fresh Haar random matrix $\tilde{\mathbf{V}}_W$ is independent of $\tilde{\mathbf{U}}_W$. We can repeat the above reasoning to conclude that

$$(\mathbf{r}_{\leq t}, \mathbf{r}_{t+1}, \mathbf{D}\mathbf{s}_{t+1}, \mathbf{d}_M) \xrightarrow{W} (\mathbf{R}_{\leq t}, \mathbf{R}_{t+1}, \mathbf{D}_M \mathbf{s}_{t+1}, \mathbf{D}_M). \quad (146)$$

The analysis of the dimension- N vectors are similar. The proof of claim $(t+1.a)$ is complete.

Induction step: proof of claim $(t+1.b)$ To prove claim $(t+1.b)$, we note that the map from the rows of $(\mathbf{r}_{\leq t+1}, \mathbf{D}\mathbf{s}_{\leq t+1}, \mathbf{d}_N)$ to those of $(\mathbf{r}_{\leq t+1}, \tilde{\mathbf{x}}_{\leq t+1}, \mathbf{d}_M)$ is polynomially bounded; cf. (130a). Then, applying [30, Proposition E.2] together with claim $(t+1.a)$ shows that the joint empirical law of $(\mathbf{r}_{\leq t+1}, \tilde{\mathbf{x}}_{\leq t+1}, \mathbf{d}_M)$ converges. The analysis of $(\mathbf{s}_{\leq t+1}, \tilde{\mathbf{z}}_{\leq t+1}, \mathbf{d}_N)$ is similar.

Induction step: proof of claim $(t+1.c)$ To analyze $\mathbf{x}_{t+1}$ and $\mathbf{z}_{t+1}$, we derive the law of $\mathbf{U}_W$ and $\mathbf{V}_W$ conditional on the OAMP iterates up to $\mathbf{r}_{t+1}$ and $\mathbf{s}_{t+1}$, i.e.,

$$[\mathbf{M}_{t+1}, \mathbf{X}_t] = \mathbf{U}_W[\mathbf{R}_{t+1}, \tilde{\mathbf{X}}_t], \quad (147a)$$

$$[\mathbf{Q}_{t+1}, \mathbf{Z}_t] = \mathbf{V}_W[\mathbf{S}_{t+1}, \tilde{\mathbf{Z}}_t]. \quad (147b)$$

Let $\mathcal{G}_t^+$ be the σ -algebra generated by the iterates up to $\mathbf{r}_{t+1}$ and $\mathbf{s}_{t+1}$. The conditional law of $\mathbf{U}_W$ and $\mathbf{V}_W$ for large M, N are given by

$$\mathbf{U}_W \stackrel{\text{d}}{=} [\mathbf{M}_{t+1}, \mathbf{X}_t] \begin{bmatrix} \mathbf{M}_{t+1}^{\top} \mathbf{M}_{t+1} & \mathbf{M}_{t+1}^{\top} \mathbf{X}_t \\ \mathbf{X}_t^{\top} \mathbf{M}_{t+1} & \mathbf{X}_t^{\top} \mathbf{X}_t \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_{t+1}^{\top} \\ \tilde{\mathbf{X}}_t^{\top} \end{bmatrix} + \Pi_{[\mathbf{M}_{t+1}, \mathbf{X}_t]}^{\perp} \tilde{\mathbf{U}}_W \Pi_{[\mathbf{R}_{t+1}, \tilde{\mathbf{X}}_t]}^{\perp}, \quad (148a)$$

$$\mathbf{V}_W \stackrel{\text{d}}{=} [\mathbf{Q}_{t+1}, \mathbf{Z}_t] \begin{bmatrix} \mathbf{Q}_{t+1}^{\top} \mathbf{Q}_{t+1} & \mathbf{Q}_{t+1}^{\top} \mathbf{Z}_t \\ \mathbf{Z}_t^{\top} \mathbf{Q}_{t+1} & \mathbf{Z}_t^{\top} \mathbf{Z}_t \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{S}_{t+1}^{\top} \\ \tilde{\mathbf{Z}}_t^{\top} \end{bmatrix} + \Pi_{[\mathbf{Q}_{t+1}, \mathbf{Z}_t]}^{\perp} \tilde{\mathbf{V}}_W \Pi_{[\mathbf{S}_{t+1}, \tilde{\mathbf{Z}}_t]}^{\perp}, \quad (148b)$$

Hence, the conditional law of $\mathbf{x}_{t+1} = \mathbf{U}_W \tilde{\mathbf{x}}_{t+1}$ is

$$\mathbf{x}_{t+1} \stackrel{\text{d}}{=} \mathbf{x}_{t+1}^{\parallel} + \mathbf{x}_{t+1}^{\perp}, \quad (149a)$$

with

$$\mathbf{x}_{t+1}^{\parallel} \stackrel{\text{def}}{=} [\mathbf{M}_{t+1}, \mathbf{X}_t] \begin{bmatrix} \mathbf{M}_{t+1}^{\top} \mathbf{M}_{t+1} & \mathbf{M}_{t+1}^{\top} \mathbf{X}_t \\ \mathbf{X}_t^{\top} \mathbf{M}_{t+1} & \mathbf{X}_t^{\top} \mathbf{X}_t \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_{t+1}^{\top} \tilde{\mathbf{x}}_{t+1} \\ \tilde{\mathbf{X}}_t^{\top} \tilde{\mathbf{x}}_{t+1} \end{bmatrix}, \quad (149b)$$

$$\mathbf{x}_{t+1}^{\perp} \stackrel{\text{def}}{=} \Pi_{[\mathbf{M}_{t+1}, \mathbf{X}_t]}^{\perp} \tilde{\mathbf{U}}_W \Pi_{[\mathbf{R}_{t+1}, \tilde{\mathbf{X}}_t]}^{\perp} \tilde{\mathbf{x}}_{t+1}. \quad (149c)$$

From claim $(t+1.b)$, and appealing to the orthogonality properties in Lemma 3, we obtain

$$\begin{bmatrix} \mathbf{M}_{t+1}^{\top} \mathbf{M}_{t+1} & \mathbf{M}_{t+1}^{\top} \mathbf{X}_t \\ \mathbf{X}_t^{\top} \mathbf{M}_{t+1} & \mathbf{X}_t^{\top} \mathbf{X}_t \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{R}_{t+1}^{\top} \tilde{\mathbf{x}}_{t+1} \\ \tilde{\mathbf{X}}_t^{\top} \tilde{\mathbf{x}}_{t+1} \end{bmatrix} \xrightarrow{\text{a.s.}} \begin{bmatrix} \mathbf{\Omega}_{t+1}^u & \mathbf{0}_{(t+1) \times t} \\ \mathbf{0}_{(t+1) \times t} & \mathbf{\Sigma}_t^u \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0}_{(t+1) \times 1} \\ \boldsymbol{\sigma}_{t+1}^u \end{bmatrix} \quad (150a)$$

$$= \begin{bmatrix} \mathbf{0}_{(t+1) \times 1} \\ (\mathbf{\Sigma}_t^u)^{-1} \boldsymbol{\sigma}_{t+1}^u \end{bmatrix}, \quad (150b)$$

where $\boldsymbol{\sigma}_{t+1}^u[i] \stackrel{\text{def}}{=} \mathbb{E}[\tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_{t+1}]$ and $\mathbf{\Sigma}_t^u[i, j] \stackrel{\text{def}}{=} \mathbb{E}[\tilde{\mathbf{X}}_i \tilde{\mathbf{X}}_j]$, $\forall i, j \in [t]$. Based on claim $(t.c)$ and similar to the analysis of $\mathbf{r}_{t+1}$, we obtain

$$(\mathbf{x}_{\leq t}, \mathbf{x}_{t+1}^{\parallel}, \mathbf{x}_{t+1}^{\perp}, \mathbf{a}) \xrightarrow{W} (\mathbf{X}_{\leq t}, \mathbf{X}_{t+1}^{\parallel}, \mathbf{X}_{t+1}^{\perp}, \mathbf{A}), \quad (151a)$$

where

$$\mathbf{X}_{t+1}^{\parallel} = [\mathbf{X}_1, \dots, \mathbf{X}_t] (\mathbf{\Sigma}_t^u)^{-1} \boldsymbol{\sigma}_{t+1}^u, \quad (151b)$$

$$\mathbf{X}_{t+1}^{\perp} \sim \mathcal{N}\left(0, \mathbb{E}[\mathbf{X}_{t+1}^2] - (\boldsymbol{\sigma}_{t+1}^u)^{\top} (\mathbf{\Sigma}_t^u)^{-1} \boldsymbol{\sigma}_{t+1}^u\right). \quad (151c)$$

Moreover, $\mathbf{X}_{t+1}^\perp$ is independent of $(\mathbf{X}_1, \dots, \mathbf{X}_t, \mathbf{A})$. Finally, since $\mathbf{m}_{t+1}$ is a Lipschitz continuous function of $(\mathbf{x}_{\leq t}, \mathbf{a})$, applying [30, Proposition E.2] and using the induction hypothesis (t.c) yields

$$(\mathbf{x}_{\leq t+1}, \mathbf{m}_{\leq t+2}, \mathbf{a}) \xrightarrow{W} (\mathbf{X}_{\leq t+1}, \mathbf{M}_{\leq t+2}, \mathbf{A}), \quad (152)$$

where the state evolution random variable appeared on the above equation are distributed as described in Definition 9. The analysis of $(\mathbf{z}_{\leq t+1}, \mathbf{q}_{t+2}; \mathbf{b})$ is similar and omitted.

C State Evolution of OAMP for Spiked Models (Theorem 1)

Recall that the OAMP algorithm for spiked matrix models consists of the following iterations (Definition 6)

$$\mathbf{u}_t = F_t(\mathbf{Y}\mathbf{Y}^\top) f_t(\mathbf{u}_{<t}; \mathbf{a}) + \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} g_t(\mathbf{v}_{<t}; \mathbf{b}), \quad (153a)$$

$$\mathbf{v}_t = G_t(\mathbf{Y}^\top \mathbf{Y}) g_t(\mathbf{v}_{<t}; \mathbf{b}) + \tilde{G}_t(\mathbf{Y}^\top \mathbf{Y}) \mathbf{Y}^\top f_t(\mathbf{u}_{<t}; \mathbf{a}). \quad (153b)$$

A major difficulty in analyzing the above OAMP algorithm is that the matrix denoisers act on the observation matrix $\mathbf{Y}$ rather than the random matrix $\mathbf{W}$. Our strategy for proving Theorem 1 parallels that of [44, Theorem 1] and proceeds through the following steps:

1. We approximate the matrix denoisers in the OAMP algorithm by polynomial functions, which is justified by the Weierstrass approximation theorem.
2. The OAMP algorithm with polynomial matrix denoisers acting on $\mathbf{Y}$ can be reformulated as an auxiliary OAMP algorithm that depends only on $\mathbf{W}$, whose dynamics are characterized by existing results (cf. Theorem 4).

Polynomial Approximation Following the approach in [44, Lemma 5], we can assume that the matrix denoisers $F_t, \tilde{F}_t, G_t, \tilde{G}_t$ are polynomial functions, which is justified by the Weierstrass approximation theorem. The result is formalized in the following lemma, whose proof—being analogous to that of [44, Lemma 5]—is omitted.

Lemma 4. *It is sufficient to prove Theorem 1 under the additional assumption that for each $t \in \mathbb{N}$, the matrix denoisers $F_t(\cdot), \tilde{F}_t(\cdot), G_t(\cdot), \tilde{G}_t(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ are polynomials.*

To analyze the behavior of these iterations, we decompose the functions f_t and g_t into two components: one that is aligned with the ground-truth signal and an orthogonal residual. Specifically, we write:

$$f_t(\mathbf{u}_{<t}; \mathbf{a}) = \alpha_t \mathbf{u}_* + \mathbf{f}_t^\perp, \quad g_t(\mathbf{v}_{<t}; \mathbf{b}) = \beta_t \mathbf{v}_* + \mathbf{g}_t^\perp. \quad (154a)$$

The signal alignment parameters, α_t and β_t , and the residual vectors, $\mathbf{f}_t^\perp$ and $\mathbf{g}_t^\perp$, are defined as follows:

$$\alpha_t \stackrel{\text{def}}{=} \mathbb{E}[\mathbf{U}_* f_t(\mathbf{U}_{<t}; \mathbf{A})], \quad \mathbf{f}_t^\perp \stackrel{\text{def}}{=} f_t(\mathbf{u}_{<t}; \mathbf{a}) - \alpha_t \mathbf{u}_*, \quad (154b)$$

$$\beta_t \stackrel{\text{def}}{=} \mathbb{E}[\mathbf{V}_* g_t(\mathbf{V}_{<t}; \mathbf{B})], \quad \mathbf{g}_t^\perp \stackrel{\text{def}}{=} g_t(\mathbf{v}_{<t}; \mathbf{b}) - \beta_t \mathbf{v}_*, \quad (154c)$$

where $(\mathbf{U}_*, \mathbf{V}_*, \mathbf{U}_{<t}, \mathbf{V}_{<t})$ are state evolution random variables defined in Section 3. Substituting these decompositions into the update rules yields the following expressions for $\mathbf{u}_t$ and $\mathbf{v}_t$:

$$\mathbf{u}_t = \alpha_t F_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{u}_* + \beta_t \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \mathbf{v}_* + F_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{f}_t^\perp + \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \mathbf{g}_t^\perp, \quad (155a)$$

$$\mathbf{v}_t = \beta_t G_t(\mathbf{Y}^\top \mathbf{Y}) \mathbf{v}_* + \alpha_t \tilde{G}_t(\mathbf{Y}^\top \mathbf{Y}) \mathbf{Y}^\top \mathbf{u}_* + G_t(\mathbf{Y}^\top \mathbf{Y}) \mathbf{g}_t^\perp + \tilde{G}_t(\mathbf{Y}^\top \mathbf{Y}) \mathbf{Y}^\top \mathbf{f}_t^\perp. \quad (155b)$$

A key challenge in analyzing this expression is that the matrix $\mathbf{Y}\mathbf{Y}^\top$ is not rotationally invariant. The following lemma, which parallels [44, Lemma 6], provides a crucial tool for addressing this issue by relating the terms to expressions involving the rotationally invariant matrix $\mathbf{W}\mathbf{W}^\top$. Its proof is deferred to Section C.1.

Lemma 5. *Let $F, \tilde{F}, G, \tilde{G} : \mathbb{R} \rightarrow \mathbb{R}$ be dimension-independent polynomials.*

1. There exist polynomial functions $\Psi^u, \tilde{\Psi}^u, \Psi^v, \tilde{\Psi}^v : \mathbb{R} \mapsto \mathbb{R}$ associated with $F, \tilde{F}, G, \tilde{G} : \mathbb{R} \rightarrow \mathbb{R}$ such that the following asymptotic equivalences hold:

$$F(\mathbf{Y}\mathbf{Y}^\top)\mathbf{u}_* \stackrel{M}{\simeq} \Psi^u(\mathbf{W}\mathbf{W}^\top)\mathbf{u}_* + \Psi^v(\mathbf{W}\mathbf{W}^\top)\mathbf{W}\mathbf{v}_*, \quad (156a)$$

$$\tilde{F}(\mathbf{Y}\mathbf{Y}^\top)\mathbf{Y}\mathbf{v}_* \stackrel{M}{\simeq} \tilde{\Psi}^v(\mathbf{W}\mathbf{W}^\top)\mathbf{W}\mathbf{v}_* + \tilde{\Psi}^u(\mathbf{W}\mathbf{W}^\top)\mathbf{u}_*, \quad (156b)$$

$$G(\mathbf{Y}^\top\mathbf{Y})\mathbf{v}_* \stackrel{N}{\simeq} \Phi^v(\mathbf{W}^\top\mathbf{W})\mathbf{v}_* + \Phi^u(\mathbf{W}^\top\mathbf{W})\mathbf{W}^\top\mathbf{u}_*, \quad (156c)$$

$$\tilde{G}(\mathbf{Y}^\top\mathbf{Y})\mathbf{Y}^\top\mathbf{u}_* \stackrel{N}{\simeq} \tilde{\Phi}^u(\mathbf{W}^\top\mathbf{W})\mathbf{W}^\top\mathbf{u}_* + \tilde{\Phi}^v(\mathbf{W}^\top\mathbf{W})\mathbf{v}_*, \quad (156d)$$

where $\stackrel{M}{\simeq}$ denotes asymptotic equivalence between random vectors as defined in Definition 2.

2. Let $\mathbf{u} \in \mathbb{R}^M$ and $\mathbf{v} \in \mathbb{R}^N$ be two random vectors such that the following hold for all $i \in \mathbb{N} \cup \{0\}$:

$$\frac{\langle \mathbf{u}_*, (\mathbf{W}\mathbf{W}^\top)^i \mathbf{u} \rangle}{M} \xrightarrow{a.s.} 0, \quad \frac{\langle \mathbf{u}_*, (\mathbf{W}\mathbf{W}^\top)^i \mathbf{W}\mathbf{v} \rangle}{M} = \frac{\langle \mathbf{u}_*, \mathbf{W}(\mathbf{W}^\top\mathbf{W})^i \mathbf{v} \rangle}{M} \xrightarrow{a.s.} 0, \quad (157a)$$

$$\frac{\langle \mathbf{v}_*, (\mathbf{W}^\top\mathbf{W})^i \mathbf{v} \rangle}{N} \xrightarrow{a.s.} 0, \quad \frac{\langle \mathbf{v}_*, (\mathbf{W}^\top\mathbf{W})^i \mathbf{W}^\top\mathbf{u} \rangle}{N} = \frac{\langle \mathbf{v}_*, \mathbf{W}^\top(\mathbf{W}\mathbf{W}^\top)^i \mathbf{u} \rangle}{N} \xrightarrow{a.s.} 0. \quad (157b)$$

Then, the following asymptotic equivalence holds

$$F(\mathbf{Y}\mathbf{Y}^\top)\mathbf{u} \stackrel{M}{\simeq} F(\mathbf{W}\mathbf{W}^\top)\mathbf{u}, \quad \tilde{F}(\mathbf{Y}\mathbf{Y}^\top)\mathbf{Y}\mathbf{v} \stackrel{M}{\simeq} \tilde{F}(\mathbf{W}\mathbf{W}^\top)\mathbf{W}\mathbf{v}, \quad (158a)$$

$$G(\mathbf{Y}^\top\mathbf{Y})\mathbf{v} \stackrel{N}{\simeq} G(\mathbf{W}^\top\mathbf{W})\mathbf{v}, \quad \tilde{G}(\mathbf{Y}^\top\mathbf{Y})\mathbf{Y}^\top\mathbf{u} \stackrel{N}{\simeq} \tilde{G}(\mathbf{W}^\top\mathbf{W})\mathbf{W}^\top\mathbf{u}. \quad (158b)$$

Importantly, Lemma 5 shows that as long as the random vectors $\mathbf{u} \in \mathbb{R}^M$ and $\mathbf{v} \in \mathbb{R}^N$ satisfy the orthogonality conditions (157), then they do not interact with the signal components in $\mathbf{Y}$. Later we shall show that the component $\mathbf{f}_t^\perp \stackrel{\text{def}}{=} f_t(\mathbf{u}_{<t}; \mathbf{a}) - \mathbb{E}[\mathbf{U}_* f_t(\mathbf{U}_{<t}; \mathbf{A})] \cdot \mathbf{u}_*$ and $\mathbf{g}_t^\perp \stackrel{\text{def}}{=} g_t(\mathbf{v}_{<t}; \mathbf{b}) - \mathbb{E}[\mathbf{V}_* g_t(\mathbf{V}_{<t}; \mathbf{B})] \cdot \mathbf{v}_*$ satisfy these orthogonality conditions. Combined with Lemma 5, this would yield the following asymptotic equivalence:

$$F_t(\mathbf{Y}\mathbf{Y}^\top)\mathbf{f}_t^\perp(\mathbf{u}_{<t}; \mathbf{a}) \stackrel{M}{\simeq} F_t(\mathbf{W}\mathbf{W}^\top)\mathbf{f}_t^\perp(\mathbf{u}_{<t}; \mathbf{a}), \quad (159a)$$

$$\tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top)\mathbf{Y}\mathbf{g}_t^\perp(\mathbf{v}_{<t}; \mathbf{b}) \stackrel{M}{\simeq} \tilde{F}_t(\mathbf{W}\mathbf{W}^\top)\mathbf{W}\mathbf{g}_t^\perp(\mathbf{v}_{<t}; \mathbf{b}), \quad (159b)$$

$$G_t(\mathbf{Y}^\top\mathbf{Y})\mathbf{g}_t^\perp(\mathbf{v}_{<t}; \mathbf{b}) \stackrel{N}{\simeq} G_t(\mathbf{W}^\top\mathbf{W})\mathbf{g}_t^\perp(\mathbf{v}_{<t}; \mathbf{b}), \quad (159c)$$

$$\tilde{G}_t(\mathbf{Y}^\top\mathbf{Y})\mathbf{Y}^\top\mathbf{f}_t^\perp(\mathbf{u}_{<t}; \mathbf{a}) \stackrel{M}{\simeq} \tilde{G}_t(\mathbf{W}^\top\mathbf{W})\mathbf{W}^\top\mathbf{f}_t^\perp(\mathbf{u}_{<t}; \mathbf{a}). \quad (159d)$$

Auxiliary OAMP Algorithm By replacing the terms in the original OAMP algorithm (155) using the corresponding asymptotic equivalence as established in (156) and (159), we introduce the following auxiliary OAMP algorithm:

$$\tilde{\mathbf{u}}_t = \Psi_t(\mathbf{W}\mathbf{W}^\top)\mathbf{u}_* + \tilde{\Psi}_t(\mathbf{W}\mathbf{W}^\top)\mathbf{W}\mathbf{v}_* + F_t(\mathbf{W}\mathbf{W}^\top)\mathbf{f}_t^\perp(\tilde{\mathbf{u}}_{<t}; \mathbf{a}) + \tilde{F}_t(\mathbf{W}\mathbf{W}^\top)\mathbf{W}\mathbf{g}_t^\perp(\tilde{\mathbf{v}}_{<t}; \mathbf{b}), \quad (160a)$$

$$\tilde{\mathbf{v}}_t = \Phi_t(\mathbf{W}^\top\mathbf{W})\mathbf{v}_* + \tilde{\Phi}_t(\mathbf{W}^\top\mathbf{W})\mathbf{W}^\top\mathbf{u}_* + G_t(\mathbf{W}^\top\mathbf{W})\mathbf{g}_t^\perp(\tilde{\mathbf{v}}_{<t}; \mathbf{b}) + \tilde{G}_t(\mathbf{W}^\top\mathbf{W})\mathbf{W}^\top\mathbf{f}_t^\perp(\tilde{\mathbf{u}}_{<t}; \mathbf{a}), \quad (160b)$$

where the matrix denoisers $\{\Psi_t, \tilde{\Psi}_t, \Phi_t, \tilde{\Phi}_t\}$ are defined as linear combinations of the transformed polynomials $\{\Psi_t^u, \tilde{\Psi}_t^u, \Phi_t^u, \tilde{\Phi}_t^u\}$ introduced in Lemma 5¹:

$$\Psi_t(\lambda) \stackrel{\text{def}}{=} \alpha_t \Psi_t^u(\lambda) + \beta_t \tilde{\Psi}_t^u(\lambda), \quad \tilde{\Psi}_t(\lambda) \stackrel{\text{def}}{=} \alpha_t \Psi_t^v(\lambda) + \beta_t \tilde{\Psi}_t^v(\lambda), \quad (161a)$$

$$\Phi_t(\lambda) \stackrel{\text{def}}{=} \beta_t \Phi_t^v(\lambda) + \alpha_t \tilde{\Phi}_t^v(\lambda), \quad \tilde{\Phi}_t(\lambda) \stackrel{\text{def}}{=} \beta_t \Phi_t^u(\lambda) + \alpha_t \tilde{\Phi}_t^u(\lambda). \quad (161b)$$

The advantage of this auxiliary algorithm is that its dynamics are governed by the rotationally invariant matrix $\mathbf{W}$ instead of the observation matrix $\mathbf{Y}$. The dynamics of this system is tractable using existing techniques (cf. Theorem 3), which we summarize in the following lemma, whose proof is deferred to Section C.2.

¹Note that our convention slightly differs from that in [44]: the deterministic scalars α_t and β_t , which appear in the orthogonal decomposition of the iterates (cf. (154)), are absorbed into the definitions of $\{\Psi_t, \tilde{\Psi}_t, \Phi_t, \tilde{\Phi}_t\}$.

Lemma 6. *The following hold for any $t \in \mathbb{N}$:*

1. *The iterates generated by the auxiliary OAMP algorithm in (160) satisfy*

$$(\mathbf{u}_*, \tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_t; \mathbf{a}) \xrightarrow{W_2} (\mathbf{U}_*, \mathbf{U}_1, \dots, \mathbf{U}_t; \mathbf{A}), \quad (162a)$$

$$(\mathbf{v}_*, \tilde{\mathbf{v}}_1, \dots, \tilde{\mathbf{v}}_t; \mathbf{b}) \xrightarrow{W_2} (\mathbf{V}_*, \mathbf{V}_1, \dots, \mathbf{V}_t; \mathbf{B}), \quad (162b)$$

where $(\mathbf{U}_*, \mathbf{U}_1, \dots, \mathbf{U}_t; \mathbf{A})$ and $(\mathbf{V}_*, \mathbf{V}_1, \dots, \mathbf{V}_t; \mathbf{B})$ are the state evolution random variables defined for the original OAMP algorithm in (25a) and (25b).

2. Denote $\mathbf{f}_t^\perp \stackrel{\text{def}}{=} \mathbf{f}_t^\perp(\tilde{\mathbf{u}}_{<t}; \mathbf{a})$ and $\mathbf{g}_t^\perp \stackrel{\text{def}}{=} \mathbf{g}_t^\perp(\tilde{\mathbf{v}}_{<t}; \mathbf{b})$. The follow holds for all $i \in \mathbb{N}$:

$$\frac{\langle \mathbf{u}_*, (\mathbf{W}\mathbf{W}^\top)^i \mathbf{f}_t^\perp \rangle}{M} \xrightarrow{a.s.} 0, \quad \frac{\langle \mathbf{u}_*, (\mathbf{W}\mathbf{W}^\top)^i \mathbf{W} \mathbf{g}_t^\perp \rangle}{M} = \frac{\langle \mathbf{u}_*, \mathbf{W} (\mathbf{W}^\top \mathbf{W})^i \mathbf{g}_t^\perp \rangle}{M} \xrightarrow{a.s.} 0, \quad (163a)$$

$$\frac{\langle \mathbf{v}_*, (\mathbf{W}^\top \mathbf{W})^i \mathbf{g}_t^\perp \rangle}{N} \xrightarrow{a.s.} 0, \quad \frac{\langle \mathbf{v}_*, (\mathbf{W}^\top \mathbf{W})^i \mathbf{W}^\top \mathbf{f}_t^\perp \rangle}{N} = \frac{\langle \mathbf{v}_*, \mathbf{W}^\top (\mathbf{W}\mathbf{W}^\top)^i \mathbf{f}_t^\perp \rangle}{N} \xrightarrow{a.s.} 0. \quad (163b)$$

Remark 11. Note that a direct application of Theorem 3 would yield a state evolution expressed in terms of the transformed polynomials $\{\Psi^u, \Psi^v, \tilde{\Psi}^u, \tilde{\Psi}^v, \Phi^u, \Phi^v, \tilde{\Phi}^u, \tilde{\Phi}^v\}$, which themselves depend on the original functions $\{F, \tilde{F}, G, \tilde{G}\}$ in an implicit and complicated manner. Fortunately, by invoking the asymptotic equivalence established in Lemma 5, the state evolution can be reformulated directly in terms of the original functions $\{F, \tilde{F}, G, \tilde{G}\}$ and the probability measures $\{\nu_1, \nu_2, \nu_3\}$, as presented in the original OAMP algorithm in (25a) and (25b).

Proof of Theorem 1 Building on the preceding results, we are now ready to prove Theorem 1. Invoking the state evolution of the auxiliary OAMP algorithm in (162), it suffices to show that the auxiliary OAMP algorithm approximates the original OAMP algorithm in the following sense:

$$\mathbf{u}_t \stackrel{N}{\simeq} \tilde{\mathbf{u}}_t \quad \text{and} \quad \mathbf{v}_t \stackrel{N}{\simeq} \tilde{\mathbf{v}}_t, \quad \forall t \in \mathbb{N}. \quad (164)$$

We prove this via induction on t . The base case is trivial. Suppose (164) holds up to iteration $t-1$. We next show (164) holds for t . We have

$$\begin{aligned} \mathbf{u}_t &\stackrel{(155)}{=} \alpha_t F_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{u}_* + \beta_t \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \mathbf{v}_* + F_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{f}_t^\perp(\mathbf{u}_{<t}; \mathbf{a}) + \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \mathbf{g}_t^\perp(\mathbf{v}_{<t}; \mathbf{b}) \\ &\stackrel{N}{\simeq} \alpha_t F_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{u}_* + \beta_t \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \mathbf{v}_* + F_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{f}_t^\perp(\tilde{\mathbf{u}}_{<t}; \mathbf{a}) + \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \mathbf{g}_t^\perp(\tilde{\mathbf{v}}_{<t}; \mathbf{b}) \\ &\stackrel{N}{\simeq} \Psi_t(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* + \tilde{\Psi}_t(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_* + F_t(\mathbf{W}\mathbf{W}^\top) \mathbf{f}_t^\perp(\tilde{\mathbf{u}}_{<t}; \mathbf{a}) + \tilde{F}_t(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{g}_t^\perp(\tilde{\mathbf{v}}_{<t}; \mathbf{b}) \\ &\stackrel{(160)}{=} \tilde{\mathbf{u}}_t, \end{aligned}$$

where the second step follows from the inductive hypothesis that $\mathbf{u}_s \stackrel{N}{\simeq} \tilde{\mathbf{u}}_s$ and $\mathbf{v}_s \stackrel{N}{\simeq} \tilde{\mathbf{v}}_s$ for all $s < t$ as well as the Lipschitz continuity of the iterate denoiser and the operator norm bound on matrix denoisers (cf. (169)), and the third step is due to Lemma 5 and Lemma 6 (which guarantees that the orthogonality conditions (157) are met).

The analysis of $\mathbf{v}_t$ is completely analogous and omitted. The proof is complete.

C.1 Proof of Lemma 5

The proof is presented in two parts, corresponding to the two claims in the lemma statement.

Proof of Claim (1). In what follows, we prove (156a), which we recall below for convenience:

$$F(\mathbf{Y}\mathbf{Y}^\top) \mathbf{u}_* \stackrel{M}{\simeq} \Psi^u(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* + \Psi^v(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_*. \quad (165)$$

Since F is a polynomial, it suffices to consider a monomial term and show that, for all $d \in \mathbb{N}$, there exist polynomials (Q_u^d, Q_v^d) such that the following asymptotic equivalence holds:

$$(\mathbf{Y}\mathbf{Y}^\top)^d \mathbf{u}_* \stackrel{M}{\simeq} Q_u^d(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* + Q_v^d(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_*. \quad (166)$$

We prove (166) via induction on d . The base case $d = 0$ is immediate. We now consider the induction step. Assume that (166) holds for some integer $d \geq 0$; we will show that it also holds for $d + 1$. Left-multiplying the expression for $(\mathbf{Y}\mathbf{Y}^\top)^d \mathbf{u}_*$ by $\mathbf{Y}\mathbf{Y}^\top$ gives

$$(\mathbf{Y}\mathbf{Y}^\top)^{d+1} \mathbf{u}_* = (\mathbf{Y}\mathbf{Y}^\top)((\mathbf{Y}\mathbf{Y}^\top)^d \mathbf{u}_*) \quad (167a)$$

$$\stackrel{\text{def}}{=} (\mathbf{Y}\mathbf{Y}^\top) \left(Q_u^d(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* + Q_v^d(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_* + \boldsymbol{\epsilon} \right) \quad (167b)$$

$$= (\mathbf{Y}\mathbf{Y}^\top) \left(Q_u^d(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* + Q_v^d(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_* \right) + (\mathbf{Y}\mathbf{Y}^\top) \boldsymbol{\epsilon}. \quad (167c)$$

By the induction hypothesis (166), we have $\|\boldsymbol{\epsilon}\|^2/M \rightarrow 0$ almost surely. Using the definition of $\mathbf{Y}$, we see that $\mathbf{Y}\mathbf{Y}^\top$ is a rank-two perturbation of $\mathbf{W}\mathbf{W}^\top$:

$$\mathbf{Y}\mathbf{Y}^\top = \frac{\theta^2}{M} \mathbf{u}_* \mathbf{u}_*^\top + \frac{\theta}{\sqrt{MN}} (\mathbf{u}_* \mathbf{v}_*^\top \mathbf{W}^\top + \mathbf{W} \mathbf{v}_* \mathbf{u}_*^\top) + \mathbf{W}\mathbf{W}^\top. \quad (168)$$

Since $\mathbf{W}$ has bounded operator norm, as shown in [44, Appendix B.1], we have

$$\limsup_{N \rightarrow \infty} \|\mathbf{Y}\mathbf{Y}^\top\|_{\text{op}} < \infty, \quad \text{almost surely.} \quad (169)$$

Combining (167)–(169) yields

$$\begin{aligned} (\mathbf{Y}\mathbf{Y}^\top)^{d+1} \mathbf{u}_* &\stackrel{M \rightarrow \infty}{\simeq} (\mathbf{Y}\mathbf{Y}^\top) \left(Q_u^d(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* + Q_v^d(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_* \right) \\ &\stackrel{(a)}{=} \left(\frac{\theta^2}{M} \mathbf{u}_* \mathbf{u}_*^\top + \frac{\theta}{\sqrt{MN}} (\mathbf{u}_* \mathbf{v}_*^\top \mathbf{W}^\top + \mathbf{W} \mathbf{v}_* \mathbf{u}_*^\top) + \mathbf{W}\mathbf{W}^\top \right) \left(Q_u^d(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* + Q_v^d(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_* \right), \end{aligned} \quad (170)$$

where step (a) uses (168). Expanding the product yields six terms, which we analyze below. For this purpose, define the following scalar quantities arising from inner products that converge almost surely in the same spirit of (92)–(94):

$$\alpha_{uu} \triangleq \lim_{M \rightarrow \infty} \frac{1}{M} \mathbf{u}_*^\top Q_u^d(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* = \langle Q_u^d(\lambda) \rangle_\mu, \quad (171a)$$

$$\alpha_{uv} \triangleq \lim_{M \rightarrow \infty} \frac{1}{M} \mathbf{u}_*^\top Q_v^d(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_* = 0, \quad (171b)$$

$$\beta_{uv} \triangleq \lim_{M \rightarrow \infty} \frac{1}{N} \mathbf{v}_*^\top \mathbf{W}^\top Q_u^d(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* = 0, \quad (171c)$$

$$\beta_{vv} \triangleq \lim_{M \rightarrow \infty} \frac{1}{N} \mathbf{v}_*^\top \mathbf{W}^\top Q_v^d(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_* = \langle \lambda Q_v^d(\lambda) \rangle_{\tilde{\mu}}, \quad (171d)$$

where the spectral measures μ and $\tilde{\mu}$ are defined in Proposition 9. Using (170) and (171), it follows that

$$\begin{aligned} (\mathbf{Y}\mathbf{Y}^\top)^{d+1} \mathbf{u}_* &\stackrel{M \rightarrow \infty}{\simeq} \left(\mathbf{W}\mathbf{W}^\top Q_u^d(\mathbf{W}\mathbf{W}^\top) + \theta^2 \alpha_{uu} \mathbf{I} + \frac{\theta}{\sqrt{\delta}} \beta_{vv} \mathbf{I} \right) \mathbf{u}_* \\ &\quad + \left(\mathbf{W}\mathbf{W}^\top Q_v^d(\mathbf{W}\mathbf{W}^\top) + \theta \sqrt{\delta} \alpha_{uv} \mathbf{I} \right) \mathbf{W} \mathbf{v}_* \end{aligned} \quad (172a)$$

$$= Q_u^{d+1}(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* + Q_v^{d+1}(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_*, \quad (172b)$$

where the degree- $(d+1)$ polynomials (Q_u^{d+1}, Q_v^{d+1}) are defined as

$$Q_u^{d+1}(\lambda) \triangleq \lambda Q_u^d(\lambda) + \theta^2 \langle Q_u^d(\lambda) \rangle_\mu + \frac{\theta}{\sqrt{\delta}} \langle \lambda Q_v^d(\lambda) \rangle_{\tilde{\mu}}, \quad (173)$$

$$Q_v^{d+1}(\lambda) \triangleq \lambda Q_v^d(\lambda) + \theta \sqrt{\delta} \langle Q_u^d(\lambda) \rangle_\mu. \quad (174)$$

Hence, (172b) establishes the induction step for (166), thereby completing the proof of (156a) in Claim (1).

The analyses of (156b)–(156d) in Claim (1) are analogous and are therefore omitted.

Proof of Claim (2). Again, it suffices to prove that the following holds for all $d \in \mathbb{N} \cup \{0\}$:

$$(\mathbf{Y}\mathbf{Y}^\top)^d \mathbf{u} \stackrel{M}{\simeq}^\infty (\mathbf{W}\mathbf{W}^\top)^d \mathbf{u}, \quad (175a)$$

$$(\mathbf{Y}\mathbf{Y}^\top)^d \mathbf{Y} \mathbf{v} \stackrel{M}{\simeq}^\infty (\mathbf{W}\mathbf{W}^\top)^d \mathbf{W} \mathbf{v}, \quad (175b)$$

$$(\mathbf{Y}^\top \mathbf{Y})^d \mathbf{v} \stackrel{N}{\simeq}^\infty (\mathbf{W}^\top \mathbf{W})^d \mathbf{v}, \quad (175c)$$

$$(\mathbf{Y}^\top \mathbf{Y})^d \mathbf{Y}^\top \mathbf{u} \stackrel{N}{\simeq}^\infty (\mathbf{W}^\top \mathbf{W})^d \mathbf{W}^\top \mathbf{u}. \quad (175d)$$

Analysis of $(\mathbf{Y}\mathbf{Y}^\top)^d \mathbf{u}$: We prove by induction on d . Assume that the claim is true for an integer $d \geq 0$ and proceed to the inductive step for $d + 1$. We have:

$$(\mathbf{Y}\mathbf{Y}^\top)^{d+1} \mathbf{u} = (\mathbf{Y}\mathbf{Y}^\top) ((\mathbf{Y}\mathbf{Y}^\top)^d \mathbf{u}) \stackrel{M}{\simeq}^\infty (\mathbf{Y}\mathbf{Y}^\top) ((\mathbf{W}\mathbf{W}^\top)^d \mathbf{u}),$$

where the second step follows from the induction hypothesis together with (169). Next, we substitute the decomposition of $\mathbf{Y}\mathbf{Y}^\top$ in (168) and expand the product:

$$\begin{aligned} (\mathbf{Y}\mathbf{Y}^\top)^{d+1} \mathbf{u} &\stackrel{M}{\simeq}^\infty \left(\frac{\theta^2}{M} \mathbf{u}_* \mathbf{u}_*^\top + \frac{\theta}{\sqrt{MN}} (\mathbf{u}_* \mathbf{v}_*^\top \mathbf{W}^\top + \mathbf{W} \mathbf{v}_* \mathbf{u}_*^\top) + \mathbf{W}\mathbf{W}^\top \right) ((\mathbf{W}\mathbf{W}^\top)^d \mathbf{u}) \\ &= \underbrace{\theta^2 \left(\frac{1}{M} \mathbf{u}_*^\top (\mathbf{W}\mathbf{W}^\top)^d \mathbf{u} \right) \mathbf{u}_*}_{\text{Term (i)}} + \underbrace{\frac{\theta\sqrt{N}}{\sqrt{M}} \left(\frac{1}{N} \mathbf{v}_*^\top \mathbf{W}^\top (\mathbf{W}\mathbf{W}^\top)^d \mathbf{u} \right) \mathbf{u}_*}_{\text{Term (ii)}} \\ &\quad + \underbrace{\frac{\theta\sqrt{M}}{\sqrt{N}} \left(\frac{1}{M} \mathbf{u}_*^\top (\mathbf{W}\mathbf{W}^\top)^d \mathbf{u} \right) \mathbf{W} \mathbf{v}_*}_{\text{Term (iii)}} + \underbrace{(\mathbf{W}\mathbf{W}^\top)^{d+1} \mathbf{u}}_{\text{Term (iv)}}. \end{aligned}$$

Based on the assumptions (157), and using the facts $\|\mathbf{u}_*\|^2/M \xrightarrow{a.s.} 1$ and $\|\mathbf{W} \mathbf{v}_*\|^2/M \xrightarrow{a.s.} C < \infty$ (where C is a constant), it is straightforward to show that

$$\frac{\|\text{Term (i)}\|^2 + \|\text{Term (ii)}\|^2 + \|\text{Term (iii)}\|^2}{N} \xrightarrow{a.s.} 0.$$

Hence, only Term (iv) survives, yielding the desired result:

$$(\mathbf{Y}\mathbf{Y}^\top)^{d+1} \mathbf{u} \stackrel{M}{\simeq}^\infty (\mathbf{W}\mathbf{W}^\top)^{d+1} \mathbf{u}. \quad (176)$$

Analysis of $(\mathbf{Y}\mathbf{Y}^\top)^d \mathbf{Y} \mathbf{v}$: We first note that

$$\mathbf{Y} \mathbf{v} = \left(\frac{\theta}{\sqrt{MN}} \mathbf{u}_* \mathbf{v}_*^\top + \mathbf{W} \right) \mathbf{v} \quad (177a)$$

$$= \left(\frac{\theta}{\sqrt{MN}} \mathbf{v}_*^\top \mathbf{v} \right) \mathbf{u}_* + \mathbf{W} \mathbf{v} \quad (177b)$$

$$\stackrel{M}{\simeq}^\infty \mathbf{W} \mathbf{v}. \quad (177c)$$

Note that the assumptions in (157) imply the asymptotic orthogonality $\mathbf{v}_*^\top \mathbf{v}/N \xrightarrow{a.s.} 0$. Using (177) and (169), we obtain

$$(\mathbf{Y}\mathbf{Y}^\top)^d \mathbf{Y} \mathbf{v} \stackrel{M}{\simeq}^\infty (\mathbf{Y}\mathbf{Y}^\top)^d \mathbf{W} \mathbf{v}. \quad (178)$$

A careful inspection shows that the vector $\hat{\mathbf{u}} \stackrel{\text{def}}{=} \mathbf{W} \mathbf{v}$ satisfies the requirements on $\mathbf{u}$ in (157). Hence, applying (175a), which we have just established above, yields

$$(\mathbf{Y}\mathbf{Y}^\top)^d \mathbf{W} \mathbf{v} \stackrel{M}{\simeq}^\infty (\mathbf{W}\mathbf{W}^\top)^d \mathbf{W} \mathbf{v}. \quad (179)$$

Combining the above two results gives

$$(\mathbf{Y}\mathbf{Y}^\top)^d \mathbf{Y} \mathbf{v} \stackrel{M}{\simeq}^\infty (\mathbf{W}\mathbf{W}^\top)^d \mathbf{W} \mathbf{v}. \quad (180)$$

The proofs of (175c) and (175d) are analogous and are therefore omitted.

C.2 Proof of Lemma 6

Proof of Claim (1): We will prove the convergence for the $\mathbf{u}$ -channel only, as the argument for the $\mathbf{v}$ -channel is analogous. Our goal is to show that for any $t \in \mathbb{N}$, the iterates of the auxiliary OAMP algorithm converge weakly to the state evolution random variables:

$$(\mathbf{u}_*, \tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_t; \mathbf{a}) \xrightarrow{W_2} (\mathbf{U}_*, \mathbf{U}_1, \dots, \mathbf{U}_t; \mathbf{A}).$$

The strategy is to demonstrate that the auxiliary OAMP algorithm can be rewritten in a canonical *signal plus noise* form, whose state evolution has already been characterized in Theorem 3.

$$\begin{aligned} \tilde{\mathbf{u}}_t &= \Psi_t(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* + \tilde{\Psi}_t(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_* + F_t(\mathbf{W}\mathbf{W}^\top) \mathbf{f}_t^\perp(\tilde{\mathbf{u}}_{<t}; \mathbf{a}) + \tilde{F}_t(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{g}_t^\perp(\tilde{\mathbf{v}}_{<t}; \mathbf{b}) \\ &= \mathbb{E} [\Psi_t(\mathbf{D}_M^2)] \cdot \mathbf{u}_* + \mathbf{e}_t + \mathbf{h}_t, \end{aligned} \quad (181a)$$

where $\mathbf{D}_M^2 \sim \mu$ and

$$\mathbf{e}_t \stackrel{\text{def}}{=} (\Psi_t(\mathbf{W}\mathbf{W}^\top) - \mathbb{E} [\Psi_t(\mathbf{D}_M^2)] \cdot \mathbf{I}_M) \mathbf{u}_*, \quad (181b)$$

$$\mathbf{h}_t \stackrel{\text{def}}{=} \tilde{\Psi}_t(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_* + F_t(\mathbf{W}\mathbf{W}^\top) \mathbf{f}_t^\perp(\tilde{\mathbf{u}}_{<t}; \mathbf{a}) + \tilde{F}_t(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{g}_t^\perp(\tilde{\mathbf{v}}_{<t}; \mathbf{b}). \quad (181c)$$

By viewing $\mathbf{u}_*$ as a side information, the above iteration an instance of the general OAMP algorithm introduced in Definition 7. By Theorem 4, we have

$$(\mathbf{e}_{\leq t}, \mathbf{h}_{\leq t}; \mathbf{a}, \mathbf{u}_*) \xrightarrow{W_2} (\mathbf{E}_{\leq t}, \mathbf{H}_{\leq t}; \mathbf{A}, \mathbf{U}_*), \quad (182)$$

where

1. The random variables $\mathbf{E}_{\leq t}$, $\mathbf{H}_{\leq t}$ and $(\mathbf{A}, \mathbf{U}_*)$ are mutually independent;

2. $\mathbf{E}_{\leq t}$, $\mathbf{W}_{\leq t}$ are zero-mean Gaussian with

$$\begin{aligned} \mathbb{E} [\mathbf{E}_s \mathbf{E}_t] &= \mathbb{E} [\Psi_s(\mathbf{D}_M^2) \Psi_t(\mathbf{D}_M^2)] - \mathbb{E} [\Psi_s(\mathbf{D}_M^2)] \cdot \mathbb{E} [\Psi_t(\mathbf{D}_M^2)], \quad \forall s, t \in \mathbb{N}, \\ \mathbb{E} [\mathbf{H}_s \mathbf{H}_t] &= \mathbb{E} [\tilde{\Psi}_s(\mathbf{D}_M^2) \tilde{\Psi}_t(\mathbf{D}_M^2) \mathbf{D}_M^2] + \mathbb{E} [F_s(\mathbf{D}_M^2) F_t(\mathbf{D}_M^2)] \cdot \mathbb{E} [\mathbf{F}_s^\perp \mathbf{F}_t^\perp] + \mathbb{E} [\tilde{F}_s(\mathbf{D}_M^2) \tilde{F}_t(\mathbf{D}_M^2) \mathbf{D}_M^2] \cdot \mathbb{E} [\mathbf{G}_s^\perp \mathbf{G}_t^\perp] \end{aligned} \quad (183)$$

where

$$\begin{aligned} \mathbf{F}_t &\stackrel{\text{def}}{=} f_t \left(\mathbb{E} [\Psi_1(\mathbf{D}_M^2)] \mathbf{U}_* + \mathbf{E}_1 + \mathbf{H}_1, \dots, \mathbb{E} [\Psi_{t-1}(\mathbf{D}_M^2)] \mathbf{U}_* + \mathbf{E}_{t-1} + \mathbf{H}_{t-1}; \mathbf{A} \right), \\ \mathbf{F}_t^\perp &\stackrel{\text{def}}{=} \mathbf{F}_t - \mathbb{E} [\mathbf{U}_* \mathbf{F}_t] \cdot \mathbf{U}_*. \end{aligned}$$

The random variables $(\mathbf{G}_t^\perp)_{t \geq 1}$ are similarly defined.

The proof for the above claims are similar to those in [44, Appendix B.4] and hence omitted.

Next, we express the inner products involving $(\Psi_t, \tilde{\Psi}_t)$ in terms of $(F, \tilde{F}, G, \tilde{G})$, in the same spirit as the treatment in [44, Appendix B.4]. As an initial step, we identify the coefficient of the signal components in (181). Recall

$$\begin{aligned} \mathbb{E} [\Psi_t(\mathbf{D}_M^2)] &\stackrel{(a)}{=} \lim_{N \rightarrow \infty} \frac{\langle \Psi_t(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_*, \mathbf{u}_* \rangle}{M} \\ &\stackrel{(b)}{=} \lim_{N \rightarrow \infty} \frac{\langle \alpha_t F_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{u}_* + \beta_t \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \mathbf{v}_*, \mathbf{u}_* \rangle}{M} \\ &= \alpha_t \cdot \lim_{N \rightarrow \infty} \frac{\langle F_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{u}_*, \mathbf{u}_* \rangle}{M} + \beta_t \cdot \lim_{N \rightarrow \infty} \frac{\langle \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \mathbf{v}_*, \mathbf{u}_* \rangle}{M} \\ &\stackrel{(c)}{=} \alpha_t \cdot \langle F_t(\lambda) \rangle_{\nu_1} + \beta_t \cdot (1 + \delta^{-1}) \langle \sigma \tilde{F}_t(\sigma^2) \rangle_{\nu_3}, \end{aligned} \quad (184)$$

where step (a) follows from (129), and all limits are taken in the almost sure sense; step (b) substitutes the definition of the transformed polynomial Ψ_t (cf. (161)); step (c) is due to Proposition 9 in the following manner

- The first term, $\frac{1}{M} \mathbf{u}_*^\top F_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{u}_*$, is a direct application of Proposition 9 Claim (1) with $h(\cdot) = F_t(\cdot)$.
- The second term involves the bilinear form $\frac{1}{M} \mathbf{u}_*^\top \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \mathbf{v}_*$. To use Proposition 9, we define an operator $f(\mathbf{Y}) = \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y}$. The function f acts on the singular values σ of $\mathbf{Y}$ as $f(\sigma) = \sigma \tilde{F}_t(\sigma^2)$, which is odd by construction. We can therefore apply Proposition 9 Claim (2), which gives the limit: $\lim_{M \rightarrow \infty} \frac{L}{M} \cdot \frac{1}{L} \mathbf{u}_*^\top \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \mathbf{v}_* \xrightarrow{\text{a.s.}} (1 + \delta^{-1}) \int \sigma \tilde{F}_t(\sigma^2) d\nu_3(\sigma)$.

We now examine the covariance structure (183) in the sequel.

$$\begin{aligned}
& \mathbb{E}[\Psi_s(\mathbf{D}_M^2) \Psi_t(\mathbf{D}_M^2)] + \mathbb{E}[\tilde{\Psi}_s(\mathbf{D}_M^2) \tilde{\Psi}_t(\mathbf{D}_M^2) \mathbf{D}_M^2] \\
& \stackrel{(d)}{=} \lim_{M \rightarrow \infty} \frac{1}{M} \left(\langle \Psi_s(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_*, \Psi_t(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* \rangle + \langle \tilde{\Psi}_s(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_*, \tilde{\Psi}_t(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_* \rangle \right) \\
& \stackrel{(e)}{=} \lim_{M \rightarrow \infty} \frac{1}{M} \left\langle (\alpha_s \Psi_s^u(\mathbf{W}\mathbf{W}^\top) + \beta_s \tilde{\Psi}_s^u(\mathbf{W}\mathbf{W}^\top)) \mathbf{u}_* + (\alpha_s \Psi_s^v(\mathbf{W}\mathbf{W}^\top) + \beta_s \tilde{\Psi}_s^v(\mathbf{W}\mathbf{W}^\top)) \mathbf{W} \mathbf{v}_*, \right. \\
& \quad \left. (\alpha_t \Psi_t^u(\mathbf{W}\mathbf{W}^\top) + \beta_t \tilde{\Psi}_t^u(\mathbf{W}\mathbf{W}^\top)) \mathbf{u}_* + (\alpha_t \Psi_t^v(\mathbf{W}\mathbf{W}^\top) + \beta_t \tilde{\Psi}_t^v(\mathbf{W}\mathbf{W}^\top)) \mathbf{W} \mathbf{v}_* \right\rangle \\
& \stackrel{(f)}{=} \lim_{M \rightarrow \infty} \frac{1}{M} \left\langle \alpha_s F_s(\mathbf{Y}\mathbf{Y}^\top) \mathbf{u}_* + \beta_s \tilde{F}_s(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \mathbf{v}_*, \alpha_t F_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{u}_* + \beta_t \tilde{F}_t(\mathbf{Y}\mathbf{Y}^\top) \mathbf{Y} \mathbf{v}_* \right\rangle \\
& \stackrel{(g)}{=} \alpha_s \alpha_t \langle F_s(\lambda) F_t(\lambda) \rangle_{\nu_1} + \beta_s \beta_t \delta^{-1} \langle \lambda \tilde{F}_s(\lambda) \tilde{F}_t(\lambda) \rangle_{\nu_2} + (1 + \delta^{-1}) \left(\alpha_s \beta_t \langle \sigma F_s(\sigma^2) \tilde{F}_t(\sigma^2) \rangle_{\nu_3} + \alpha_t \beta_s \langle \sigma F_t(\sigma^2) \tilde{F}_s(\sigma^2) \rangle_{\nu_3} \right),
\end{aligned} \tag{185}$$

where step (d) follows from (129)-(131), and all limits are taken in the almost sure sense; step (e) substitutes the expansion of the transformed polynomial $\Psi_t, \tilde{\Psi}_t$ (cf. (161)) and uses the independence between $\mathbf{u}_*, \mathbf{v}_*$ to absorb the cross term; step (f) utilizes the asymptotical equivalence (156) in Lemma 5; step (g) repeats the same procedure as in step (c) in the light of Proposition 9. Next, we investigate

$$\begin{aligned}
& \mathbb{E}[F_s(\mathbf{D}_M^2) F_t(\mathbf{D}_M^2)] \cdot \mathbb{E}[F_s^\perp F_t^\perp] = \mathbb{E}[(F_s - \alpha_s \mathbf{U}_*)(F_t - \alpha_t \mathbf{U}_*)] \cdot \lim_{M \rightarrow \infty} \frac{\langle F_s(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_*, F_t(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* \rangle}{M} \\
& = (\mathbb{E}[F_s F_t] - \alpha_s \alpha_t) \cdot \langle F_s(\lambda) F_t(\lambda) \rangle_\mu \stackrel{\text{def}}{=} \sigma_{f,st}^2 \cdot \langle F_s(\lambda) F_t(\lambda) \rangle_\mu.
\end{aligned} \tag{186}$$

Similarly we have

$$\begin{aligned}
& \mathbb{E}[\tilde{F}_s(\mathbf{D}_M^2) \tilde{F}_t(\mathbf{D}_M^2) \mathbf{D}_M^2] \cdot \mathbb{E}[\tilde{F}_s^\perp \tilde{F}_t^\perp] = \mathbb{E}[(G_s - \beta_s \mathbf{V}_*)(G_t - \beta_t \mathbf{V}_*)] \cdot \lim_{M \rightarrow \infty} \frac{\langle \tilde{F}_s(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_*, \tilde{F}_t(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_* \rangle}{M} \\
& = (\mathbb{E}[G_s G_t] - \beta_s \beta_t) \cdot \delta^{-1} \langle \lambda \tilde{F}_s(\lambda) \tilde{F}_t(\lambda) \rangle_{\tilde{\mu}} \stackrel{\text{def}}{=} \sigma_{g,st}^2 \cdot \delta^{-1} \langle \lambda \tilde{F}_s(\lambda) \tilde{F}_t(\lambda) \rangle_{\tilde{\mu}}.
\end{aligned} \tag{187}$$

Finally, let us compute the total covariance (183) by gathering (184) to (187)

$$\Sigma_{u,st} \stackrel{\text{def}}{=} \mathbb{E}[\mathbf{Z}_{u,s} \mathbf{Z}_{u,t}] = \mathbb{E}[\mathbf{E}_s \mathbf{E}_t] + \mathbb{E}[\mathbf{H}_s \mathbf{H}_t] \tag{188}$$

$$\begin{aligned}
& = \alpha_s \alpha_t \langle F_s(\lambda) F_t(\lambda) \rangle_{\nu_1} + \beta_s \beta_t \delta^{-1} \langle \lambda \tilde{F}_s(\lambda) \tilde{F}_t(\lambda) \rangle_{\nu_2} - \mu_{u,s} \mu_{u,t} \\
& \quad + (1 + \delta^{-1}) \left(\alpha_s \beta_t \langle \sigma F_s(\sigma^2) \tilde{F}_t(\sigma^2) \rangle_{\nu_3} + \alpha_t \beta_s \langle \sigma F_t(\sigma^2) \tilde{F}_s(\sigma^2) \rangle_{\nu_3} \right) \\
& \quad + \sigma_{f,st}^2 \langle F_s(\lambda) F_t(\lambda) \rangle_\mu + \delta^{-1} \sigma_{g,st}^2 \langle \lambda \tilde{F}_s(\lambda) \tilde{F}_t(\lambda) \rangle_{\tilde{\mu}},
\end{aligned} \tag{189}$$

which is precisely the claimed covariance structure in (28).

Proof of Claim (2). The proof follows the same strategy as in [44, Appendix B.4]. We briefly outline the argument for (163a); the proof of (163b) is analogous and therefore omitted.

For any fixed $t \in \mathbb{N}$ and $i \in \mathbb{N}$, the terms

$$(\mathbf{W}\mathbf{W}^\top)^i \mathbf{f}_t^\perp \quad \text{and} \quad (\mathbf{W}\mathbf{W}^\top)^i \mathbf{W} \mathbf{g}_t^\perp \tag{190}$$

may be interpreted as post-processing steps of an OAMP algorithm (cf. (181)), whose dynamics are characterized by state evolution. By (1) a simple re-indexing, (2) viewing $\mathbf{u}_*$ and $\mathbf{v}_*$ as side information, and (3)

an appropriate specification of the matrix-denoising functions, these post-processed terms can be identified with the iterates of another OAMP algorithm of the form (7). Consequently, the state-evolution results in Theorem 4 apply.

In particular, we establish that the empirical distributions of these terms converge to Gaussian random variables that are independent of the side information (in this case, $\mathbf{U}_*$ and $\mathbf{V}_*$, which represent the underlying signals). The claimed asymptotic orthogonality therefore follows.

D Derivations for Optimal Denoisers

We adopt a greedy per-iteration design for the denoisers, focusing on the $\mathbf{v}$ -channel (the $\mathbf{u}$ -channel is analogous). At iteration t , the state evolution in (25b) yields a scalar Gaussian channel $\mathbf{V}_t = \mu_{v,t}\mathbf{V}_* + \mathbf{Z}_{v,t}$ with variance $\sigma_{v,t}^2 \stackrel{\text{def}}{=} \text{Var}(\mathbf{Z}_{v,t})$. We design the denoisers to locally maximize the squared cosine similarity

$$w_{2,t} \stackrel{\text{def}}{=} \frac{\mu_{v,t}^2}{\mu_{v,t}^2 + \sigma_{v,t}^2}. \quad (191)$$

The design at iteration t is decoupled into two conditional optimizations:

- With (f_t, g_t) fixed, the pair $(\mu_{v,t}, \sigma_{v,t}^2)$ depends on the matrix denoisers $(G_t, \tilde{G}_t)$. Optimizing $(G_t, \tilde{G}_t)$ under the trace-free constraint (22) yields the optimal spectral denoisers (Appendix D.1).
- With $(G_t^*, \tilde{G}_t^*)$ fixed, the pair $(\mu_{v,t}, \sigma_{v,t}^2)$ depends on (f_t, g_t) only through the residual covariances in state evolution. Optimizing (f_t, g_t) under the divergence-free constraints (24) yields the DMMSE iterate denoisers (Appendix D.2).

D.1 Optimal Matrix Denoisers

At iteration t , we treat the iterate denoisers (f_t, g_t) as given and optimize the spectral denoisers $(G, \tilde{G})$ in the $\mathbf{v}$ -update. Recall the SE variables and suppress t for simplicity (26a)-(26c)

$$\mathbf{F} \stackrel{\text{def}}{=} f_t(\mathbf{U}_1, \dots, \mathbf{U}_{t-1}; \mathbf{A}), \quad \mathbf{G} \stackrel{\text{def}}{=} g_t(\mathbf{V}_1, \dots, \mathbf{V}_{t-1}; \mathbf{B}), \quad \alpha \stackrel{\text{def}}{=} \mathbb{E}[\mathbf{U}_* \mathbf{F}], \quad \beta \stackrel{\text{def}}{=} \mathbb{E}[\mathbf{V}_* \mathbf{G}],$$

$$\sigma_f^2 \stackrel{\text{def}}{=} \mathbb{E}(\mathbf{F}^2) - \alpha^2, \quad \sigma_g^2 \stackrel{\text{def}}{=} \mathbb{E}(\mathbf{G}^2) - \beta^2.$$

Introduce the effective precisions

$$\rho_1 \stackrel{\text{def}}{=} \frac{\alpha^2}{\sigma_f^2}, \quad \rho_2 \stackrel{\text{def}}{=} \frac{\beta^2}{\sigma_g^2}, \quad (192)$$

With (f_t, g_t) fixed (hence $(\alpha, \beta, \rho_1, \rho_2)$ fixed), we choose $(G, \tilde{G})$ to maximize (191) subject to the trace-free constraint. Writing $(\mu_{v,t}, \sigma_{v,t}^2)$ as functionals of $(G, \tilde{G})$, we consider

$$\max_{G, \tilde{G}} \frac{[\mu_{v,t}(G, \tilde{G})]^2}{[\mu_{v,t}(G, \tilde{G})]^2 + \sigma_{v,t}^2(G, \tilde{G})} \quad \text{s.t.} \quad \langle G \rangle_{\tilde{\mu}} = 0 \quad (193)$$

$$= \min_{G, \tilde{G}} \left\{ 1 - \frac{[\mu_{v,t}(G, \tilde{G})]^2}{[\mu_{v,t}(G, \tilde{G})]^2 + \sigma_{v,t}^2(G, \tilde{G})} \right\} \quad \text{s.t.} \quad \langle G \rangle_{\tilde{\mu}} = 0$$

$$\stackrel{(a)}{=} \min_{G, \tilde{G}} \min_{c \in \mathbb{R}} [1 - c \mu_{v,t}(G, \tilde{G})]^2 + c^2 \sigma_{v,t}^2(G, \tilde{G}) \quad \text{s.t.} \quad \langle cG \rangle_{\tilde{\mu}} = 0$$

$$\stackrel{(b)}{=} \min_{G, \tilde{G}} [1 - \mu_{v,t}(G, \tilde{G})]^2 + \sigma_{v,t}^2(G, \tilde{G}) \quad \text{s.t.} \quad \langle G \rangle_{\tilde{\mu}} = 0$$

$$\stackrel{(c)}{=} \min_{G, \tilde{G}} \left\langle (\beta G - 1)^2 \right\rangle_{\nu_2} + \delta \left\langle \lambda (\alpha \tilde{G})^2 \right\rangle_{\nu_1} + \frac{1}{\rho_2} \left\langle (\beta G)^2 \right\rangle_{\tilde{\mu}} \quad (194)$$

$$+ \frac{\delta}{\rho_1} \left\langle \lambda (\alpha \tilde{G})^2 \right\rangle_{\mu} + 2(1 + \delta) \left\langle (\beta G(\sigma^2) - 1)(\alpha \tilde{G}(\sigma^2)) \sigma \right\rangle_{\nu_3} \quad \text{s.t.} \quad \langle G \rangle_{\tilde{\mu}} = 0.$$

Here $\tilde{\mu} \stackrel{\text{def}}{=} \delta \cdot \mu + (1 - \delta) \cdot \delta_{\{0\}}$ is the noise measure in (22). The reductions are:

- (a) Apply $1 - a^2/(a^2 + b) = \min_{c \in \mathbb{R}} (1 - ca)^2 + c^2 b$ and use $\langle cG \rangle_{\tilde{\mu}} = c \langle G \rangle_{\tilde{\mu}}$.
- (b) Use the homogeneity of $(\mu_{v,t}, \sigma_{v,t}^2)$ to absorb c into $(G, \tilde{G})$.
- (c) Substitute (25b) and (155b), and eliminate (σ_f^2, σ_g^2) via (192).

The variational problem (194) is convex. We introduce a Lagrange multiplier ξ_v for the trace-free constraint:

$$\begin{aligned} \mathcal{L}_v &\stackrel{\text{def}}{=} \langle (\beta G - 1)^2 \rangle_{\nu_2} + \delta \langle \lambda (\alpha \tilde{G})^2 \rangle_{\nu_1} + \frac{1}{\rho_2} \langle (\beta G)^2 \rangle_{\tilde{\mu}} + \frac{\delta}{\rho_1} \langle \lambda (\alpha \tilde{G})^2 \rangle_{\mu} + 2(1 + \delta) \langle (\beta G(\sigma^2) - 1)(\alpha \tilde{G}(\sigma^2)) \sigma \rangle_{\nu_3} - 2\xi_v \langle G \rangle_{\tilde{\mu}} \\ &= \mathcal{L}_v^{(0)} + \mathcal{L}_v^{(\text{bulk})} + \mathcal{L}_v^{(\text{out})}, \end{aligned}$$

where in spirit of Lemma 2 we decompose $\mathcal{L}_v$ into the contributions from the atom at 0, the bulk $\text{supp}(\mu)$, and the limiting outliers $\mathcal{K}_*$ as:

$$\mathcal{L}_v^{(0)} \stackrel{\text{def}}{=} \nu_2(\{0\}) (\beta G(0) - 1)^2 + (1 - \delta) \frac{1}{\rho_2} (\beta G(0))^2 - 2\xi_v (1 - \delta) G(0). \quad (195)$$

$$\begin{aligned} \mathcal{L}_v^{(\text{out})} &\stackrel{\text{def}}{=} \sum_{\lambda_* \notin \text{supp}(\mu) \cup \{0\}} \left\{ \nu_2(\{\lambda_*\}) (\beta G(\lambda_*) - 1)^2 + \delta \nu_1(\{\lambda_*\}) \lambda_* (\alpha \tilde{G}(\lambda_*))^2 \right. \\ &\quad \left. + 4(1 + \delta) \nu_3(\{\sigma_*\}) \sigma_* (\beta G(\lambda_*) - 1) (\alpha \tilde{G}(\lambda_*)) \right\}. \end{aligned} \quad (196)$$

$$\begin{aligned} \mathcal{L}_v^{(\text{bulk})} &\stackrel{\text{def}}{=} \int_{\text{supp}(\mu)} (\beta G(\lambda) - 1)^2 d\nu_2(\lambda) + \delta \int_{\text{supp}(\mu)} \lambda (\alpha \tilde{G}(\lambda))^2 d\nu_1(\lambda) \\ &\quad + \frac{1}{\rho_2} \int_{\text{supp}(\mu)} (\beta G(\lambda))^2 d\{\delta \cdot \mu(\lambda)\} + \frac{\delta}{\rho_1} \int_{\text{supp}(\mu)} \lambda (\alpha \tilde{G}(\lambda))^2 d\mu(\lambda) \\ &\quad + 2(1 + \delta) \int_{\{\sigma: \sigma^2 \in \text{supp}(\mu)\}} (\beta G(\sigma^2) - 1) (\alpha \tilde{G}(\sigma^2)) \sigma d\nu_3(\sigma) - 2\xi_v \int_{\text{supp}(\mu)} G(\lambda) d\{\delta \cdot \mu(\lambda)\}. \end{aligned} \quad (197)$$

Optimal Denoisers in $\text{supp}(\mu)$. We minimize $\mathcal{L}_v^{(\text{bulk})}$ pointwise over $\text{supp}(\mu)$. Using Lemma 2 and the characterization of the absolutely continuous parts $\nu_i^\parallel$ via the shrinkage functions φ_i , we have

$$\begin{aligned} \mathcal{L}_v^{(\text{bulk})} &= \left\langle (\beta G(\lambda) - 1)^2 \varphi_2(\lambda) + \frac{\delta}{\rho_2} (\beta G(\lambda))^2 - 2\xi_v \delta G(\lambda) \right. \\ &\quad \left. + \delta \lambda (\alpha \tilde{G}(\lambda))^2 \left(\varphi_1(\lambda) + \frac{1}{\rho_1} \right) + 2\sqrt{\delta} (\beta G(\lambda) - 1) (\alpha \tilde{G}(\lambda)) \varphi_3(\lambda) \right\rangle_{\mu, \text{supp}(\mu)}. \end{aligned} \quad (198)$$

Here we use the oddness of $\nu_3^\parallel$ and the change of variables $\lambda = \sigma^2$ on $\{\sigma: \sigma^2 \in \text{supp}(\mu)\}$.

Taking pointwise first-order conditions yields, for μ -a.e. $\lambda \in \text{supp}(\mu)$, the following equations hold:

$$\begin{aligned} (\rho_2 \varphi_2(\lambda) + \delta) (\beta G(\lambda) - 1) + \sqrt{\delta} \rho_2 \varphi_3(\lambda) (\alpha \tilde{G}(\lambda)) &= -\delta \left(1 - \frac{\xi_v \rho_2}{\beta} \right), \\ \sqrt{\delta} \rho_1 \varphi_3(\lambda) (\beta G(\lambda) - 1) + \delta \lambda (\rho_1 \varphi_1(\lambda) + 1) (\alpha \tilde{G}(\lambda)) &= 0. \end{aligned} \quad (199)$$

Solving (199) yields the bulk minimizers

$$\begin{aligned} \beta G_{\text{bulk}}^*(\lambda) &= 1 - \left(1 - \frac{\xi_v \rho_2}{\beta} \right) \frac{\delta [\rho_1 \varphi_1(\lambda) + 1] \lambda}{[\rho_1 \varphi_1(\lambda) + 1] [\rho_2 \varphi_2(\lambda) + \delta] \lambda - \rho_1 \rho_2 \varphi_3^2(\lambda)} \stackrel{\text{def}}{=} 1 - \left(1 - \frac{\xi_v \rho_2}{\beta} \right) Q^*(\lambda), \\ \alpha \tilde{G}_{\text{bulk}}^*(\lambda) &= \left(1 - \frac{\xi_v \rho_2}{\beta} \right) \frac{\sqrt{\delta} \rho_1 \varphi_3(\lambda)}{[\rho_1 \varphi_1(\lambda) + 1] [\rho_2 \varphi_2(\lambda) + \delta] \lambda - \rho_1 \rho_2 \varphi_3^2(\lambda)} \stackrel{\text{def}}{=} \left(1 - \frac{\xi_v \rho_2}{\beta} \right) \tilde{Q}^*(\lambda). \end{aligned} \quad (200)$$

Optimal denoisers at the origin. We next minimize the contribution of the atom at $\{0\}$. From (195),

$$\mathcal{L}_v^{(0)} = \left[\nu_2(\{0\}) + (1 - \delta) \frac{1}{\rho_2} \right] (\beta G(0))^2 - 2 \left[\nu_2(\{0\}) + (1 - \delta) \frac{\xi_v}{\beta} \right] (\beta G(0)) + \nu_2(\{0\}). \quad (201)$$

Minimizing this quadratic gives

$$\beta G_0^*(0) \stackrel{(a)}{=} 1 - \left(1 - \frac{\xi_v \rho_2}{\beta} \right) \frac{1 - \delta}{\rho_2 \nu_2(\{0\}) + (1 - \delta)} \stackrel{(b)}{=} 1 - \left(1 - \frac{\xi_v \rho_2}{\beta} \right) \frac{1 - \theta^2(1 - \delta) \pi \mathcal{H}(0)}{\rho_2 + 1 - \theta^2(1 - \delta) \pi \mathcal{H}(0)}. \quad (202)$$

Here (a) is the closed-form minimizer of (201), and (b) substitutes $\nu_2(\{0\})$ from Lemma 2.

To match the bulk formula (200) at the origin (202), note that $\beta G_0^*(0)$ is exactly the $\lambda \downarrow 0$ limit of (200) by the explicit form of φ_2 in (7c):

$$\beta G_0^*(0) = \lim_{\lambda \rightarrow 0} \beta G_{\text{bulk}}^*(\lambda) \quad \text{since} \quad \lim_{\lambda \rightarrow 0} Q^*(\lambda) = \frac{\delta}{\rho_2 \varphi_2(0) + \delta} = \frac{1 - \theta^2(1 - \delta) \pi \mathcal{H}(0)}{\rho_2 + 1 - \theta^2(1 - \delta) \pi \mathcal{H}(0)}.$$

To enforce the trace-free constraint $\langle G \rangle_{\tilde{\mu}} = 0$ with $\tilde{\mu} = \delta \mu + (1 - \delta) \delta_{\{0\}}$, we expand

$$0 = \langle G \rangle_{\tilde{\mu}} = \left\langle \frac{1}{\beta} \left(1 - \left(1 - \frac{\xi_v \rho_2}{\beta} \right) Q^*(\lambda) \right) \right\rangle_{\tilde{\mu}} \implies 1 - \frac{\xi_v \rho_2}{\beta} = \left\langle Q^*(\lambda) \right\rangle_{\tilde{\mu}}^{-1}, \quad (203)$$

Substituting into (200) gives the unified bulk/zero expressions via $Q^*(\lambda), \tilde{Q}^*(\lambda)$ in (200)

$$\beta G^*(\lambda) = 1 - \left\langle Q^*(\lambda) \right\rangle_{\tilde{\mu}}^{-1} Q^*(\lambda), \quad \alpha \tilde{G}^*(\lambda) = \left\langle Q^*(\lambda) \right\rangle_{\tilde{\mu}}^{-1} \tilde{Q}^*(\lambda),$$

for all $\lambda \in \text{supp}(\mu) \cup \{0\}$.

Optimal denoisers at non-zero outliers. Fix a nonzero outlier $\lambda_* \in \mathcal{K}_*$ and let $\sigma_* = \sqrt{\lambda_*}$. The outlier contribution to the Lagrangian is

$$\begin{aligned} \mathcal{L}_v^{(\{\lambda_*\})} &= \nu_2(\{\lambda_*\}) (\beta G(\lambda_*) - 1)^2 + \delta \nu_1(\{\lambda_*\}) \lambda_* (\alpha \tilde{G}(\lambda_*))^2 \\ &\quad + 4(1 + \delta) \nu_3(\{\sigma_*\}) \sigma_* (\beta G(\lambda_*) - 1) (\alpha \tilde{G}(\lambda_*)), \end{aligned} \quad (204)$$

where we use the atomic characterizations of ν_1, ν_2, ν_3 in Lemma 2 and the oddness $\nu_3(\{\sigma_*\}) = -\nu_3(\{-\sigma_*\})$. The first-order conditions of (204) are

$$\nu_2(\{\lambda_*\}) (\beta G(\lambda_*) - 1) + 2(1 + \delta) \nu_3(\{\sigma_*\}) \sigma_* (\alpha \tilde{G}(\lambda_*)) = 0, \quad (205)$$

$$\delta \nu_1(\{\lambda_*\}) \lambda_* (\alpha \tilde{G}(\lambda_*)) + 2(1 + \delta) \nu_3(\{\sigma_*\}) \sigma_* (\beta G(\lambda_*) - 1) = 0, \quad (206)$$

and substituting the point masses from Lemma 2 shows this linear system is underdetermined. Hence all solutions admit the parametrization

$$(\beta G(\lambda_*) - 1, \alpha \tilde{G}(\lambda_*)) = \tau \left(-2(1 + \delta) \nu_3(\{\sigma_*\}) \sigma_*, \nu_2(\{\lambda_*\}) \right), \quad \tau \in \mathbb{R}. \quad (207)$$

To match the bulk minimizer at λ_* , we verify that the bulk ratio satisfies (205); indeed, taking $\lambda \rightarrow \lambda_*$ in the first equation of (199) gives

$$\lim_{\lambda \rightarrow \lambda_*} \frac{1 - \beta G_{\text{bulk}}^*(\lambda)}{\alpha \tilde{G}_{\text{bulk}}^*(\lambda)} = \frac{2(1 + \delta) \nu_3(\{\sigma_*\}) \sigma_*}{\nu_2(\{\lambda_*\})} = \sqrt{\delta} \theta \lambda_* \mathcal{S}_\mu(\lambda_*), \quad (208)$$

and the second equation in (199) yields the analogous consistency for (206).

Combining (200) and (208) yields the unified form for $\lambda \in \text{supp}(\mu) \cup \{0\} \cup \mathcal{K}_*$,

$$\beta G^*(\lambda) = 1 - \left\langle Q^*(\lambda; \rho_1, \rho_2) \right\rangle_{\tilde{\mu}}^{-1} Q^*(\lambda), \quad \alpha \tilde{G}^*(\lambda) = \left\langle Q^*(\lambda; \rho_1, \rho_2) \right\rangle_{\tilde{\mu}}^{-1} \tilde{Q}^*(\lambda), \quad (209)$$

where $(\alpha, \beta, \rho_1, \rho_2)$ are induced by the iterate denoisers (f_t, g_t) and will be optimized in the next subsection.

D.2 Optimal Iterate Denoisers

With the optimal spectral denoisers $(G_t^*, \tilde{G}_t^*)$ fixed at iteration t , the objective in (193)–(194) depends on the iterate denoisers (f_t, g_t) only through the (σ_f^2, σ_g^2) in (26c) and, for fixed $(G_t^*, \tilde{G}_t^*)$, only via

$$\sigma_g^2 \int [G_t^*(\lambda)]^2 d\{\delta\mu(\lambda) + (1-\delta)\delta_0\}, \quad \delta\sigma_f^2 \int \lambda [\tilde{G}_t^*(\lambda)]^2 d\mu(\lambda). \quad (210)$$

Therefore, maximizing the squared cosine similarity is equivalent to minimizing (σ_f^2, σ_g^2) subject to the divergence-free constraints (24). This yields two decoupled scalar programs:

$$\min_{f_t} \mathbb{E} \left[(\mathbf{U}_* - f_t(\mathbf{U}_{t-1}; \mathbf{A}))^2 \right] \quad \text{s.t.} \quad \mathbb{E}[f_t'(\mathbf{U}_{t-1}; \mathbf{A})] = 0, \quad \min_{g_t} \mathbb{E} \left[(\mathbf{V}_* - g_t(\mathbf{V}_{t-1}; \mathbf{B}))^2 \right] \quad \text{s.t.} \quad \mathbb{E}[g_t'(\mathbf{V}_{t-1}; \mathbf{B})] = 0,$$

where $(\mathbf{U}_*, \mathbf{A}) \sim \pi_u$ and $(\mathbf{V}_*, \mathbf{B}) \sim \pi_v$ are the scalar priors induced by state evolution. By [44, Definition 3], each problem is solved by the corresponding DMMSE estimator, hence

$$f_t^*(\mathbf{u}; \mathbf{a}) = \bar{\phi} \left(\mathbf{u} / \sqrt{\mu_{u,t-1}^2 + \text{Var}(\mathbf{Z}_{u,t-1})}; \mathbf{a} \mid w_{1,t-1} \right), \quad w_{1,t-1} = \frac{\mu_{u,t-1}^2}{\mu_{u,t-1}^2 + \text{Var}(\mathbf{Z}_{u,t-1})}, \quad (211)$$

$$g_t^*(\mathbf{v}; \mathbf{b}) = \bar{\phi} \left(\mathbf{v} / \sqrt{\mu_{v,t-1}^2 + \text{Var}(\mathbf{Z}_{v,t-1})}; \mathbf{b} \mid w_{2,t-1} \right), \quad w_{2,t-1} = \frac{\mu_{v,t-1}^2}{\mu_{v,t-1}^2 + \text{Var}(\mathbf{Z}_{v,t-1})}. \quad (212)$$

To express the SE effective precision in closed form, define the standardized observation

$$\mathbf{X}_{v,t-1} \stackrel{\text{def}}{=} \frac{\mathbf{V}_{t-1}}{\sqrt{\mu_{v,t-1}^2 + \text{Var}(\mathbf{Z}_{v,t-1})}} = \sqrt{w_{2,t-1}} \mathbf{V}_* + \sqrt{1 - w_{2,t-1}} \mathbf{Z}, \quad \mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}).$$

Since $\mathbf{G}_t = g_t^*(\mathbf{V}_{t-1}; \mathbf{B}) = \bar{\phi}(\mathbf{X}_{v,t-1}; \mathbf{B} \mid w_{2,t-1})$, [44, Lemma 2] yields $\mathbb{E}[\mathbf{G}_t^2] = \mathbb{E}[\mathbf{V}_* \mathbf{G}_t] = \beta_t$ and hence

$$\rho_{2,t} = \frac{\beta_t}{1 - \beta_t} = \frac{1}{\text{mmse}_{\pi_v}(w_{2,t-1})} - \frac{1}{1 - w_{2,t-1}}, \quad \beta_t = \frac{\rho_{2,t}}{1 + \rho_{2,t}}. \quad (213)$$

The same identities hold for the $\mathbf{u}$ -channel with similar forms for $(\mathbf{U}_*, \mathbf{X}_{u,t-1}, \mathbf{A}, w_{1,t-1}, \alpha_t, \rho_{1,t}, \pi_u)$.

E Proof of Proposition 2

The state evolution parameters are defined by the following recursion, initialized with $w_{1,0} = w_{2,0} \in (0, 1)$

$$\rho_{1,t} = \mathcal{F}_1(w_{1,t-1}) \stackrel{\text{def}}{=} \frac{1}{\text{mmse}_u(w_{1,t-1})} - \frac{1}{1 - w_{1,t-1}}, \quad (214a)$$

$$\rho_{2,t} = \mathcal{F}_2(w_{2,t-1}) \stackrel{\text{def}}{=} \frac{1}{\text{mmse}_v(w_{2,t-1})} - \frac{1}{1 - w_{2,t-1}}, \quad (214b)$$

$$w_{1,t} = \mathcal{F}_3(\rho_{1,t}, \rho_{2,t}) \stackrel{\text{def}}{=} 1 - \frac{1 - \langle P_t^*(\lambda; \rho_{1,t}, \rho_{2,t}) \rangle_\mu}{\langle P_t^*(\lambda; \rho_{1,t}, \rho_{2,t}) \rangle_\mu} \cdot \frac{1}{\rho_{1,t}}, \quad (214c)$$

$$w_{2,t} = \mathcal{F}_4(\rho_{1,t}, \rho_{2,t}) \stackrel{\text{def}}{=} 1 - \frac{1 - \langle Q_t^*(\lambda; \rho_{1,t}, \rho_{2,t}) \rangle_{\tilde{\mu}}}{\langle Q_t^*(\lambda; \rho_{1,t}, \rho_{2,t}) \rangle_{\tilde{\mu}}} \cdot \frac{1}{\rho_{2,t}}, \quad (214d)$$

The proof relies on the properties of these recursive functions, summarized in the following lemma which we defer its proof in Section E.1.

Lemma 7. *The functions defining the state evolution recursion satisfy:*

1. $\mathcal{F}_1(w)$ and $\mathcal{F}_2(w)$ are continuous and non-decreasing functions mapping $[0, 1)$ to $[0, \infty)$, with $\lim_{w \rightarrow 1^-} \mathcal{F}_{1,2}(w) = \infty$.

2. $\mathcal{F}_3(\rho_1, \rho_2)$ and $\mathcal{F}_4(\rho_1, \rho_2)$ are continuous and non-decreasing in both arguments on $(0, \infty)^2$, and map to $[0, 1)$, with $\lim_{\rho_1, \rho_2 \rightarrow \infty} \mathcal{F}_{3,4}(\rho_1, \rho_2) < 1$.

Proof. Proof of Claim (1). We prove Claim (1) for the $\mathbf{v}$ -channel; the $\mathbf{u}$ -channel is analogous. We show by induction that, for every $t \geq 1$,

$$w_{2,t} \in (0, 1), \quad \rho_{2,t} > 0, \quad \mathbf{V}_t^* | \mathbf{V}_* \sim \mathcal{N}(\sqrt{w_{2,t}} \mathbf{V}_*, 1 - w_{2,t}). \quad (215)$$

Assuming (215), the MMSE postprocessing $\hat{\mathbf{v}}_t^* = \phi(\mathbf{v}_t^*; \mathbf{b} | w_{2,t})$ satisfies, by Theorem 1,

$$\frac{\|\hat{\mathbf{v}}_t^* - \mathbf{v}_*^*\|^2}{N} \xrightarrow{\text{a.s.}} \mathbb{E} \left[(\phi(\mathbf{V}_t^*; \mathbf{B} | w_{2,t}) - \mathbf{V}_*)^2 \right] = \text{mmsev}(w_{2,t}), \quad (216)$$

Hence it remains to prove (215) by induction.

Induction Steps. Assume (215) holds at step $t - 1$, i.e.

$$\mathbf{V}_{t-1}^* | \mathbf{V}_* \sim \mathcal{N}(\sqrt{w_{2,t-1}} \mathbf{V}_*, 1 - w_{2,t-1}). \quad (217)$$

At iteration t , state evolution yields the scalar representation

$$\mathbf{V}_t^* = \mu_{v,t} \mathbf{V}_* + \mathbf{Z}_{v,t}, \quad \mathbf{Z}_{v,t} \sim \mathcal{N}(0, \sigma_{v,t}^2) \perp\!\!\!\perp \mathbf{V}_*,$$

and our goal is to verify that this channel has squared cosine similarity $w_{2,t}$, i.e.

$$\mu_{v,t} = \sqrt{w_{2,t}}, \quad \sigma_{v,t}^2 = 1 - w_{2,t}. \quad (218)$$

We only prove the mean identity $\mu_{v,t} = \sqrt{w_{2,t}}$; the variance identity follows analogously from (28b).

Under the induction hypothesis (217), the iterate denoiser equals $g_t^*(\cdot; \mathbf{b}) = \bar{\phi}(\cdot; \mathbf{b} | w_{2,t-1})$. Hence the DMMSE precision identities (213) hold; see Appendix D.2. Substituting the explicit forms of G_t^* and $\tilde{G}_t^*$ in (40) into the general SE formula (27) yields the following representation

$$\mu_{v,t} = \frac{1}{\sqrt{w_{2,t}}} \left[1 - \frac{1}{\langle Q_t^* \rangle_{\tilde{\mu}}} \left\{ \langle Q_t^*(\lambda) \rangle_{\nu_2} - (1 + \delta) \langle \sigma \tilde{Q}_t^*(\sigma^2) \rangle_{\nu_3} \right\} \right]. \quad (219)$$

We claim that for every $t \geq 1$,

$$\left\{ \langle Q_t^*(\lambda) \rangle_{\nu_2} - (1 + \delta) \langle \sigma \tilde{Q}_t^*(\sigma^2) \rangle_{\nu_3} \right\} = \frac{1}{\rho_{2,t}} \left(1 - \langle Q_t^*(\lambda) \rangle_{\tilde{\mu}} \right). \quad (220)$$

Assuming (220), (219) and recalling the definition of $w_{2,t}$ in the recursion (41) give

$$\mu_{v,t} = \frac{1}{\sqrt{w_{2,t}}} \left(1 - \frac{1 - \langle Q_t^* \rangle_{\tilde{\mu}}}{\langle Q_t^* \rangle_{\tilde{\mu}}} \cdot \frac{1}{\rho_{2,t}} \right) = \frac{w_{2,t}}{\sqrt{w_{2,t}}} = \sqrt{w_{2,t}}.$$

The variance identity $\sigma_{v,t}^2 = 1 - w_{2,t}$ follows analogously from (28b), completing the induction step and hence Claim (1).

It remains to prove (220). Fix t and abbreviate $\rho_2 = \rho_{2,t}$, $Q = Q_t^*$, $\tilde{Q} = \tilde{Q}_t^*$. Then

$$\begin{aligned} \langle Q(\lambda) \rangle_{\nu_2} - (1 + \delta) \langle \sigma \tilde{Q}(\sigma^2) \rangle_{\nu_3} &\stackrel{(a)}{=} \underbrace{\left\langle \varphi_2(\lambda) Q(\lambda) \right\rangle_{\mu} - \sqrt{\delta} \left\langle \varphi_3(\lambda) \tilde{Q}(\lambda) \right\rangle_{\mu}}_{\text{a.c. part}} \\ &\quad + \underbrace{\sum_{\lambda_* \in \mathcal{K}_*} \left\{ \nu_2(\{\lambda_*\}) Q(\lambda_*) - 2(1 + \delta) \nu_3(\{\sigma_*\}) \sigma_* \tilde{Q}(\lambda_*) \right\}}_{\text{outliers}} + \underbrace{\nu_2(\{0\}) Q(0)}_{\text{atom at 0}} \\ &\stackrel{(b)}{=} \frac{\delta}{\rho_2} \left(1 - \langle Q(\lambda) \rangle_{\mu} \right) + \nu_2(\{0\}) Q(0) \stackrel{(c)}{=} \frac{1}{\rho_2} \left(1 - \langle Q(\lambda) \rangle_{\tilde{\mu}} \right). \end{aligned} \quad (221)$$

where: (a) applies Lemma 2 for Lebesgue decomposition of ν_2, ν_3 and the change of variables $\lambda = \sigma^2$ on $\nu_3^\parallel$. Step (b) simplifies the a.c. part by substituting the explicit formulas (200) of $Q, \tilde{Q}$ and integrating the resulting pointwise identity against μ , and uses (208) to cancel the outlier terms for each $\lambda_* \in \mathcal{K}_*$. Step (c) uses (202) to eliminate $\nu_2(\{0\})$ and rewrites $\delta\langle Q \rangle_\mu + (1 - \delta)Q(0) = \langle Q \rangle_{\tilde{\mu}}$.

Proof of Claim (2). With respect to Claim (2), we now prove that the sequences $\{w_{2,t}\}$ and $\{\rho_{2,t}\}$ (and their $\mathbf{u}$ -channel counterparts) are non-decreasing and converge to the specified fixed point.

- **Monotonicity:** We proceed by induction. We initialize with $w_{1,0} = w_{2,0} = 0$. Then $\rho_{1,1} = \mathcal{F}_1(w_{1,0})$ and $\rho_{2,1} = \mathcal{F}_2(w_{2,0})$. The next iterate is $w_{1,1} = \mathcal{F}_3(\rho_{1,1}, \rho_{2,1})$. Since $\mathcal{F}_3 \geq 0$, we have $w_{1,1} \geq w_{1,0} = 0$. Now, assume $w_{1,t-1} \leq w_{1,t}$ and $w_{2,t-1} \leq w_{2,t}$. The monotonicity of the functions $\mathcal{F}_i$ given by Lemma 7 ensures that $\rho_{1,t} \leq \rho_{1,t+1}$ and $\rho_{2,t} \leq \rho_{2,t+1}$. Consequently, $w_{2,t+1} = \mathcal{F}_4(\rho_{1,t+1}, \rho_{2,t+1}) \geq \mathcal{F}_4(\rho_{1,t}, \rho_{2,t}) = w_{2,t}$. The same logic applies to $w_{1,t}$. Therefore, all four sequences are non-decreasing.
- **Convergence:** The sequences $\{w_{1,t}\}$ and $\{w_{2,t}\}$ are non-decreasing and bounded above by 1. By the monotone convergence theorem, they must converge to limits w_1^* and w_2^* . Lemma 7 guarantees that the limits are strictly less than 1. Consequently, the sequences $\{\rho_{1,t}\}$ and $\{\rho_{2,t}\}$ also converge to finite limits $\rho_1^* = \mathcal{F}_1(w_1^*)$ and $\rho_2^* = \mathcal{F}_2(w_2^*)$. By the continuity of all functions, the limit point $(\rho_1^*, \rho_2^*, w_1^*, w_2^*)$ must be a solution to the fixed-point equations given in the proposition. $\square$

E.1 Proof of Lemma 7

The properties of $\mathcal{F}_1$ and $\mathcal{F}_2$ follow from standard scalar Gaussian MMSE arguments; see [44, Lemma 4]. We focus on $\mathcal{F}_3$ and $\mathcal{F}_4$, and only treat $\mathcal{F}_3(\rho_1, \rho_2)$, as the argument for $\mathcal{F}_4$ is identical.

Recall that $\mathcal{F}_3$ is defined via $P^*(\lambda; \rho_1, \rho_2)$. For $\rho_1, \rho_2 > 0$ and $\lambda \in \text{supp}(\mu)$, set

$$a(\rho_2; \lambda) \stackrel{\text{def}}{=} \rho_2 \lambda \varphi_2(\lambda) + \delta \lambda, \quad (222a)$$

$$b(\rho_2; \lambda) \stackrel{\text{def}}{=} \rho_2 [\lambda \varphi_1(\lambda) \varphi_2(\lambda) - \varphi_3^2(\lambda)] + \delta \lambda \varphi_1(\lambda), \quad (222b)$$

$$D(\rho_1, \rho_2; \lambda) \stackrel{\text{def}}{=} (\rho_1 \varphi_1(\lambda) + 1)(\rho_2 \varphi_2(\lambda) + \delta) \lambda - \rho_1 \rho_2 \varphi_3^2(\lambda) = \rho_1 b(\rho_2; \lambda) + a(\rho_2; \lambda). \quad (222c)$$

By the explicit formulas for $\varphi_1, \varphi_2, \varphi_3$ in (7a)–(7c), one verifies that

$$a(\rho_2; \lambda) > 0, \quad b(\rho_2; \lambda) > 0, \quad D(\rho_1, \rho_2; \lambda) > 0 \quad \text{for all } \lambda \in \text{supp}(\mu), \rho_1, \rho_2 > 0.$$

Hence

$$P^*(\lambda) = \frac{a(\rho_2; \lambda)}{D(\rho_1, \rho_2; \lambda)}, \quad \frac{1}{\rho_1} (1 - P^*(\lambda)) = \frac{b(\rho_2; \lambda)}{D(\rho_1, \rho_2; \lambda)}.$$

We define

$$N(\rho_1, \rho_2) \stackrel{\text{def}}{=} \left\langle \frac{b(\rho_2; \lambda)}{D(\rho_1, \rho_2; \lambda)} \right\rangle_\mu, \quad M(\rho_1, \rho_2) \stackrel{\text{def}}{=} \left\langle \frac{a(\rho_2; \lambda)}{D(\rho_1, \rho_2; \lambda)} \right\rangle_\mu,$$

so that

$$\mathcal{F}_3(\rho_1, \rho_2) = 1 - \frac{\langle b(\rho_2; \lambda) / D(\rho_1, \rho_2; \lambda) \rangle_\mu}{\langle a(\rho_2; \lambda) / D(\rho_1, \rho_2; \lambda) \rangle_\mu} = 1 - \frac{N(\rho_1, \rho_2)}{M(\rho_1, \rho_2)}.$$

Continuity. The functions $a(\rho_2; \lambda)$, $b(\rho_2; \lambda)$, and $D(\rho_1, \rho_2; \lambda)$ are continuous in (ρ_1, ρ_2) for $\rho_1, \rho_2 > 0$, and $D(\rho_1, \rho_2; \lambda) > 0$ by construction. Since $\text{supp}(\mu)$ is compact, the ratios $a(\rho_2; \lambda) / D(\rho_1, \rho_2; \lambda)$ and $b(\rho_2; \lambda) / D(\rho_1, \rho_2; \lambda)$ are uniformly bounded on compact subsets of $(0, \infty)^2$. Dominated convergence then yields continuity of M and N , and hence $\mathcal{F}_3$ is continuous on $(0, \infty)^2$.

Monotonicity. We show that $\mathcal{F}_3$ is non-decreasing in each coordinate.

(i) *Monotonicity in ρ_2 .* Recalling $\mathcal{F}_3 = 1 - N/M$, we obtain

$$\frac{\partial \mathcal{F}_3}{\partial \rho_2} = \frac{N(\rho_1, \rho_2) M'_{\rho_2}(\rho_1, \rho_2) - N'_{\rho_2}(\rho_1, \rho_2) M(\rho_1, \rho_2)}{M(\rho_1, \rho_2)^2}.$$

Differentiating b/D and a/D with respect to ρ_2 , we obtain

$$N(\rho_1, \rho_2) M'_{\rho_2}(\rho_1, \rho_2) - N'_{\rho_2}(\rho_1, \rho_2) M(\rho_1, \rho_2) = \left\langle \frac{\delta \lambda \varphi_3^2(\lambda)}{D^2(\rho_1, \rho_2; \lambda)} \right\rangle_{\mu} \geq 0,$$

since $b(\rho_2; \lambda) a'(\rho_2; \lambda) - a(\rho_2; \lambda) b'(\rho_2; \lambda) = \delta \lambda \varphi_3^2(\lambda)$ and $D(\rho_1, \rho_2; \lambda) > 0$ on $\text{supp}(\mu)$. Consequently,

$$\frac{\partial \mathcal{F}_3}{\partial \rho_2} \geq 0,$$

and $\mathcal{F}_3$ is non-decreasing in ρ_2 .

(ii) *Monotonicity in ρ_1 .* Again writing $\mathcal{F}_3 = 1 - N/M$, we have

$$\frac{\partial \mathcal{F}_3}{\partial \rho_1} = \frac{N(\rho_1, \rho_2) M'_{\rho_1}(\rho_1, \rho_2) - N'_{\rho_1}(\rho_1, \rho_2) M(\rho_1, \rho_2)}{M(\rho_1, \rho_2)^2}.$$

Differentiating b/D and a/D with respect to ρ_1 and taking $\langle \cdot \rangle_{\mu}$ yields

$$\begin{aligned} & N(\rho_1, \rho_2) M'_{\rho_1}(\rho_1, \rho_2) - N'_{\rho_1}(\rho_1, \rho_2) M(\rho_1, \rho_2) \\ &= \left\langle \frac{b^2(\rho_2; \lambda)}{D^2(\rho_1, \rho_2; \lambda)} \right\rangle_{\mu} \left\langle \frac{a(\rho_2; \lambda)}{D(\rho_1, \rho_2; \lambda)} \right\rangle_{\mu} - \left\langle \frac{a(\rho_2; \lambda) b(\rho_2; \lambda)}{D^2(\rho_1, \rho_2; \lambda)} \right\rangle_{\mu} \left\langle \frac{b(\rho_2; \lambda)}{D(\rho_1, \rho_2; \lambda)} \right\rangle_{\mu}. \end{aligned}$$

To sign this quantity, set

$$X \stackrel{\text{def}}{=} \frac{a(\rho_2; \lambda)}{b(\rho_2; \lambda)}, \quad Z \stackrel{\text{def}}{=} \frac{b^2(\rho_2; \lambda)}{D^2(\rho_1, \rho_2; \lambda)} \geq 0,$$

and take the non-decreasing functions $f(x) = x$ and $g(x) = \rho_1 + x$. Chebyshev's association inequality [69, Theorem 2.14] states that for non-decreasing f, g and a non-negative random variable Z ,

$$\mathbb{E}[Z] \cdot \mathbb{E}[f(X)g(X)Z] \geq \mathbb{E}[f(X)Z] \cdot \mathbb{E}[g(X)Z].$$

Applied to the law of X induced by μ and the above choice of Z , this gives exactly

$$\left\langle \frac{b^2(\rho_2; \lambda)}{D^2(\rho_1, \rho_2; \lambda)} \right\rangle_{\mu} \left\langle \frac{a(\rho_2; \lambda)}{D(\rho_1, \rho_2; \lambda)} \right\rangle_{\mu} \geq \left\langle \frac{a(\rho_2; \lambda) b(\rho_2; \lambda)}{D^2(\rho_1, \rho_2; \lambda)} \right\rangle_{\mu} \left\langle \frac{b(\rho_2; \lambda)}{D(\rho_1, \rho_2; \lambda)} \right\rangle_{\mu}.$$

Consequently, $\mathcal{F}_3$ is non-decreasing in ρ_1 since

$$\frac{\partial \mathcal{F}_3}{\partial \rho_1} = \frac{N M'_{\rho_1} - N'_{\rho_1} M}{M^2} \geq 0.$$

Range. By definition $M(\rho_1, \rho_2) > 0$, hence $\mathcal{F}_3(\rho_1, \rho_2) = 1 - N/M < 1$ for all $\rho_1 < \infty$. For the lower bound, using the monotonicity of $\mathcal{F}_3$ in (ρ_1, ρ_2) ,

$$\begin{aligned} \mathcal{F}_3(\rho_1, \rho_2) &\geq \lim_{\rho_2 \rightarrow 0} \lim_{\rho_1 \rightarrow 0} \mathcal{F}_3(\rho_1, \rho_2) \\ &\stackrel{(a)}{=} 1 - \lim_{\rho_2 \rightarrow 0} \left\langle \frac{b(\rho_2; \lambda)}{a(\rho_2; \lambda)} \right\rangle_{\mu} \stackrel{(b)}{=} 1 - \langle \varphi_1(\lambda) \rangle_{\mu} \stackrel{(c)}{=} 1 - \nu_1^{\parallel}(\mathbb{R}) \geq 0. \end{aligned} \tag{223}$$

where (a) lets $\rho_1 \rightarrow 0$ so that $D(\rho_1, \rho_2; \lambda) \rightarrow a(\rho_2; \lambda)$ and applies dominated convergence; (b) uses $a(\rho_2; \lambda) \rightarrow \delta \lambda$ and $b(\rho_2; \lambda) \rightarrow \delta \lambda \varphi_1(\lambda)$ as $\rho_2 \rightarrow 0$; (c) uses $d\nu_1^{\parallel}/d\lambda = \varphi_1(\lambda)\mu(\lambda)$ from Lemma 2 and $\nu_1^{\parallel}(\mathbb{R}) \leq 1$ since ν_1 is a probability measure.

Finally, we compute the limit at infinity:

$$\begin{aligned} \lim_{\rho_2 \rightarrow \infty} \lim_{\rho_1 \rightarrow \infty} \mathcal{F}_3(\rho_1, \rho_2) &\stackrel{(a)}{=} \lim_{\rho_2 \rightarrow \infty} \lim_{\rho_1 \rightarrow \infty} 1 - \left\langle \frac{a(\rho_2; \lambda)}{D(\rho_1, \rho_2; \lambda)} \right\rangle_{\mu}^{-1} \left\langle \frac{b(\rho_2; \lambda)}{D(\rho_1, \rho_2; \lambda)} \right\rangle_{\mu} \\ &\stackrel{(b)}{=} \lim_{\rho_2 \rightarrow \infty} 1 - \left\langle \frac{a(\rho_2; \lambda)}{b(\rho_2; \lambda)} \right\rangle_{\mu}^{-1} \\ &\stackrel{(c)}{=} 1 - \left\langle \frac{\lambda \varphi_2(\lambda)}{\lambda \varphi_1(\lambda) \varphi_2(\lambda) - \varphi_3^2(\lambda)} \right\rangle_{\mu}^{-1} < 1. \end{aligned} \tag{224}$$

where (a) expands $\mathcal{F}_3 = 1 - N/M$ with $N = \langle b/D \rangle_{\mu}$ and $M = \langle a/D \rangle_{\mu}$; (b) lets $\rho_1 \rightarrow \infty$ in $D = \rho_1 b + a$, so by dominated convergence $\langle b/D \rangle_{\mu} / \langle a/D \rangle_{\mu} \rightarrow \langle a/b \rangle_{\mu}^{-1}$; (c) lets $\rho_2 \rightarrow \infty$ and uses the fact that $\lambda \varphi_1 \varphi_2 - \varphi_3^2 > 0$.

F Proof of Proposition 3 (I.I.D. Gaussian Noise Model)

This section analyzes a special case of the model from (1), where the noise matrix entries are IID Gaussian random variables: $W_{i,j} \sim \mathcal{N}(0, 1/N)$, for any $1 \leq i \leq M, 1 \leq j \leq N$, with an aspect ratio $\lim_{M,N \rightarrow \infty} \delta = \frac{M}{N} \in (0, 1)$.

F.1 Spectral Analysis of I.I.D. Gaussian Noise Model

We first recall that when $\mathbf{W}$ has i.i.d. $\mathcal{N}(0, 1/N)$ entries and $M/N \rightarrow \delta \in (0, 1)$, the empirical spectral distribution of $\mathbf{W}\mathbf{W}^\top$ converges almost surely to the Marčenko–Pastur law with density

$$\mu_{\text{MP}}(\lambda) = \frac{\sqrt{(b_+ - \lambda)(\lambda - a_-)}}{2\pi\delta\lambda} \mathbf{1}_{[a_-, b_+]}(\lambda), \quad a_- \stackrel{\text{def}}{=} (1 - \sqrt{\delta})^2, \quad b_+ \stackrel{\text{def}}{=} (1 + \sqrt{\delta})^2,$$

see, e.g., [8, Theorem 3.6]. In the rectangular spiked model (1) with i.i.d. Gaussian noise $\mathbf{W}_{ij} \sim \mathcal{N}(0, 1/N)$, the largest eigenvalue of $\mathbf{Y}\mathbf{Y}^\top$ is known to exhibit a phase transition (cf. [7, 8]). There is a critical value $\theta^2 = \sqrt{\delta}$ such that

$$\begin{aligned} \theta^2 \leq \sqrt{\delta} &\Rightarrow \lambda_1(\mathbf{Y}\mathbf{Y}^\top) \xrightarrow{\text{a.s.}} b_+, \\ \theta^2 > \sqrt{\delta} &\Rightarrow \lambda_1(\mathbf{Y}\mathbf{Y}^\top) \xrightarrow{\text{a.s.}} \lambda_* \stackrel{\text{def}}{=} 1 + \delta + \theta^2 + \frac{\delta}{\theta^2} \geq b_+, \end{aligned} \quad (225)$$

where equality holds if and only if $\theta^2 = \delta^{1/2}$. We next examine the behavior of the master equation (8) on the left of the Marčenko–Pastur bulk.

Lemma 8. *Let $\mathbf{Y}$ follow the rectangular spiked model (1) with $\mathbf{W}$ having i.i.d. $\mathcal{N}(0, 1/N)$ entries, and let $M, N \rightarrow \infty$ with $M/N \rightarrow \delta \in (0, 1)$. Then the master equation*

$$\Gamma(\lambda) = 1 - \theta^2 \mathcal{C}(\lambda) = 0$$

admits no real solution on $[0, a_-)$.

Proof. It suffices to show that $\Gamma(\lambda) \neq 0$ for all $\lambda < a_-$. For the Marčenko–Pastur law, the Stieltjes transform $\mathcal{S}_\mu$ satisfies (cf. [8, Lemma 3.11]),

$$\mathcal{S}_\mu(z) = \frac{1}{2\delta z} \left[z + \delta - 1 - \sqrt{(z - \delta - 1)^2 - 4\delta} \right], \quad \forall z \notin [a_-, b_+]$$

substituting into the definition of $\mathcal{C}(\lambda)$ (5) yields:

$$\mathcal{C}(\lambda) = \lambda \mathcal{S}_\mu(\lambda) - 1 = \int \frac{t}{\lambda - t} \mu_{\text{MP}}(dt), \quad \lambda \notin [a_-, b_+].$$

If $\lambda < a_-$, then $t \in [a_-, b_+]$ implies $\lambda - t < 0$, so the integrand $t/(\lambda - t)$ is strictly negative and hence $\mathcal{C}(\lambda) < 0$. Since $\theta^2 > 0$, it follows that

$$\Gamma(\lambda) = 1 - \theta^2 \mathcal{C}(\lambda) > 1,$$

and therefore $\Gamma(\lambda) \neq 0$ for all $\lambda < a_-$. This proves that the master equation has no real solution on $[0, a_-)$. $\square$

F.2 Optimal Denoisers and its OAMP Recursion

The spectral measures admit an outlier atom at λ_* in (225) and, for the $\mathbf{v}$ -channel, an additional atom at 0 (cf. Lemma 2). Under Lemma 8, there is no atom on $[0, a_-)$, and the point masses are

$$\nu_1(\{\lambda_*\}) = 1 - \frac{\delta(1 + \theta^2)}{\theta^2(\delta + \theta^2)}, \quad \nu_2(\{\lambda_*\}) = 1 - \frac{\delta + \theta^2}{\theta^2(1 + \theta^2)}, \quad \nu_2(\{0\}) = \frac{1 - \delta}{1 + \theta^2}. \quad (226)$$

The absolutely continuous parts follow directly from Lemma 2:

$$\nu_1^\parallel(\lambda) = \mu(\lambda)\varphi_1(\lambda) = \frac{\delta + \frac{\delta}{\theta^2}}{\lambda_* - \lambda}\mu(\lambda), \quad (227a)$$

$$\nu_2^\parallel(\lambda) = \mu(\lambda)\varphi_2(\lambda) = \frac{\delta + \frac{\delta^2}{\theta^2}}{\lambda_* - \lambda}\mu(\lambda), \quad (227b)$$

$$\nu_3^\parallel(\sigma) = \text{sign}(\sigma) \frac{\sqrt{\delta}}{1 + \delta} \cdot \mu(\sigma^2)\varphi_3(\sigma^2) = \text{sign}(\sigma) \frac{\sqrt{\delta}}{1 + \delta} \cdot \frac{\frac{\delta}{\theta}\sigma^2}{\lambda_* - \sigma^2}\mu(\sigma^2). \quad (227c)$$

Substituting (226)–(227) into (39) yields the $\mathbf{u}$ -channel denoisers

$$P_t^*(\lambda; \rho_{1,t}, \rho_{2,t}) = \frac{\rho_{2,t}(1 + \frac{\delta}{\theta^2}) + \lambda_* - \lambda}{\rho_{1,t}(\delta + \frac{\delta}{\theta^2}) + \rho_{2,t}(1 + \frac{\delta}{\theta^2}) + \rho_{1,t}\rho_{2,t} \cdot \frac{\delta}{\theta^2} + \lambda_* - \lambda},$$

$$\tilde{P}_t^*(\lambda; \rho_{1,t}, \rho_{2,t}) = \frac{\frac{\sqrt{\delta}}{\theta}\rho_{2,t}}{\rho_{1,t}(\delta + \frac{\delta}{\theta^2}) + \rho_{2,t}(1 + \frac{\delta}{\theta^2}) + \rho_{1,t}\rho_{2,t} \cdot \frac{\delta}{\theta^2} + \lambda_* - \lambda},$$

and, analogously, (40) gives the $\mathbf{v}$ -channel denoisers

$$Q_t^*(\lambda; \rho_{1,t}, \rho_{2,t}) = \frac{\rho_{1,t}(\delta + \frac{\delta}{\theta^2}) + \lambda_* - \lambda}{\rho_{1,t}(\delta + \frac{\delta}{\theta^2}) + \rho_{2,t}(1 + \frac{\delta}{\theta^2}) + \rho_{1,t}\rho_{2,t} \cdot \frac{\delta}{\theta^2} + \lambda_* - \lambda}, \quad (229a)$$

$$\tilde{Q}_t^*(\lambda; \rho_{1,t}, \rho_{2,t}) = \frac{\frac{\sqrt{\delta}}{\theta}\rho_{1,t}}{\rho_{1,t}(\delta + \frac{\delta}{\theta^2}) + \rho_{2,t}(1 + \frac{\delta}{\theta^2}) + \rho_{1,t}\rho_{2,t} \cdot \frac{\delta}{\theta^2} + \lambda_* - \lambda}, \quad (229b)$$

Having established the close form of optimal denoisers, we have the following representation of limit squared cosine similarities.

Lemma 9. *In the rectangular spiked matrix model (1) with I.I.D. Gaussian noise $\mathbf{W} \sim \mathcal{N}(0, 1/N)$, let $T(\lambda_*) = \rho_1(\delta + \frac{\delta}{\theta^2}) + \rho_2(1 + \frac{\delta}{\theta^2}) + \rho_1\rho_2 \cdot \frac{\delta}{\theta^2} + \lambda_*$, and let $\mathcal{S}_\mu(\cdot)$ be the Stieltjes transform of the Marčenko-Pastur law. The limit squared cosine similarities w_1 and w_2 are determined by the fixed-point equations*

$$w_1 = 1 - \frac{(\frac{\delta}{\theta^2}\rho_2 + \frac{\delta}{\theta^2} + \delta)\mathcal{S}_\mu(T(\lambda_*))}{1 - \rho_1(\frac{\delta}{\theta^2}\rho_2 + \frac{\delta}{\theta^2} + \delta)\mathcal{S}_\mu(T(\lambda_*))}, \quad (230a)$$

$$w_2 = 1 - \frac{(\frac{\delta}{\theta^2}\rho_1 + \frac{\delta}{\theta^2} + 1)[\delta\mathcal{S}_\mu(T(\lambda_*)) + (1 - \delta)/T(\lambda_*)]}{1 - \rho_2(\frac{\delta}{\theta^2}\rho_1 + \frac{\delta}{\theta^2} + 1)[\delta\mathcal{S}_\mu(T(\lambda_*)) + (1 - \delta)/T(\lambda_*)]}. \quad (230b)$$

Proof. We prove the $\mathbf{v}$ -channel identity (230b); the proof for w_1 is analogous. By the fixed-point update (214d),

$$w_2 \stackrel{(c)}{=} 1 - \frac{1 - \langle Q^*(\lambda; \rho_1, \rho_2) \rangle_{\tilde{\mu}}}{\rho_2 \langle Q^*(\lambda; \rho_1, \rho_2) \rangle_{\tilde{\mu}}}.$$

Thus it suffices to compute $\langle Q^*(\lambda; \rho_1, \rho_2) \rangle_{\tilde{\mu}}$. Recalling $T(\lambda_*)$ and $\tilde{\mu}$ from Lemma 2, we have

$$\begin{aligned} \langle Q^*(\lambda; \rho_1, \rho_2) \rangle_{\tilde{\mu}} &\stackrel{(a)}{=} \left\langle 1 - \frac{\rho_2(1 + \frac{\delta}{\theta^2}) + \rho_1\rho_2\frac{\delta}{\theta^2}}{T(\lambda_*) - \lambda} \right\rangle_{\tilde{\mu}} \\ &= 1 - \rho_2 \left(\frac{\delta}{\theta^2}\rho_1 + \frac{\delta}{\theta^2} + 1 \right) \left\langle \frac{1}{T(\lambda_*) - \lambda} \right\rangle_{\tilde{\mu}} \\ &\stackrel{(b)}{=} 1 - \rho_2 \left(\frac{\delta}{\theta^2}\rho_1 + \frac{\delta}{\theta^2} + 1 \right) \left[\delta\mathcal{S}_\mu(T(\lambda_*)) + \frac{1 - \delta}{T(\lambda_*)} \right]. \end{aligned} \quad (231)$$

where (a) uses (229); and (b) uses $\tilde{\mu} = \delta\mu + (1 - \delta)\delta_0$ from Lemma 2 and $\langle (T(\lambda_*) - \lambda)^{-1} \rangle_{\mu} = \mathcal{S}_\mu(T(\lambda_*))$, hence $\langle (T(\lambda_*) - \lambda)^{-1} \rangle_{\tilde{\mu}} = \delta\mathcal{S}_\mu(T(\lambda_*)) + \frac{1 - \delta}{T(\lambda_*)}$. Substituting (231) into (214d) yields (230b). $\square$

For a clear comparison, the next Lemma 10 reformulates the fixed point equation of AMP [56, Theorem 3] for the rectangular spiked model with I.I.D. Gaussian noise.

Lemma 10. *Under Assumption 1, run the AMP iteration of [56, Alg. (4.2)–(4.3)] with posterior-mean iterative denoisers for the two channels. Let $(\bar{\mu}_{u,t}, \bar{\sigma}_{u,t}, \bar{\mu}_{v,t}, \bar{\sigma}_{v,t})_{t \geq 0}$ be the SE parameters, and define the squared cosine similarities*

$$\bar{w}_{1,t} \stackrel{\text{def}}{=} \frac{\bar{\mu}_{u,t}^2}{\bar{\mu}_{u,t}^2 + \bar{\sigma}_{u,t}^2}, \quad \bar{w}_{2,t} \stackrel{\text{def}}{=} \frac{\bar{\mu}_{v,t}^2}{\bar{\mu}_{v,t}^2 + \bar{\sigma}_{v,t}^2}.$$

Assume $\bar{w}_{1,t} \rightarrow \bar{w}_1$ and $\bar{w}_{2,t} \rightarrow \bar{w}_2$ as $t \rightarrow \infty$. Then the limiting overlaps satisfy the coupled fixed-point equations

$$\frac{\bar{w}_2}{1 - \bar{w}_2} = \theta^2 [1 - \text{mmse}_U(\bar{w}_1)], \quad \frac{\bar{w}_1}{1 - \bar{w}_1} = \frac{\theta^2}{\delta} [1 - \text{mmse}_V(\bar{w}_2)]. \quad (232)$$

Proof. By [56, Theorem. 3], the SE gives

$$\bar{\mu}_{v,t+1} = \theta \mathbb{E}[\mathbf{U}_* g_t(\bar{\mu}_{u,t} \mathbf{U}_* + \bar{\sigma}_{u,t} G)], \quad \bar{\sigma}_{v,t+1}^2 = \mathbb{E}[g_t(\bar{\mu}_{u,t} \mathbf{U}_* + \bar{\sigma}_{u,t} G)^2], \quad (233)$$

$$\bar{\mu}_{u,t} = \frac{\theta}{\delta} \mathbb{E}[\mathbf{V}_* f_t(\bar{\mu}_{v,t} \mathbf{V}_* + \bar{\sigma}_{v,t} G)], \quad \bar{\sigma}_{u,t}^2 = \frac{1}{\delta} \mathbb{E}[f_t(\bar{\mu}_{v,t} \mathbf{V}_* + \bar{\sigma}_{v,t} G)^2], \quad (234)$$

where $G \sim \mathcal{N}(0, 1)$ is independent of $(\mathbf{U}_*, \mathbf{V}_*)$, and the denoisers are the posterior means

$$g_t(x) = \mathbb{E}[\mathbf{U}_* | \bar{\mu}_{u,t} \mathbf{U}_* + \bar{\sigma}_{u,t} G = x], \quad f_t(y) = \mathbb{E}[\mathbf{V}_* | \bar{\mu}_{v,t} \mathbf{V}_* + \bar{\sigma}_{v,t} G = y]. \quad (235)$$

By the tower property applied to (235),

$$\mathbb{E}[\mathbf{U}_* g_t(\bar{\mu}_{u,t} \mathbf{U}_* + \bar{\sigma}_{u,t} G)] = \mathbb{E}[g_t(\bar{\mu}_{u,t} \mathbf{U}_* + \bar{\sigma}_{u,t} G)^2]. \quad (236)$$

Using $\mathbb{E}[\mathbf{U}_*^2] = 1$ from Assumption 1, the scalar-channel identity gives

$$\text{mmse}_U(\bar{w}_{1,t}) = \mathbb{E}[(\mathbf{U}_* - \mathbb{E}[\mathbf{U}_* | \bar{\mu}_{u,t} \mathbf{U}_* + \bar{\sigma}_{u,t} G])^2] = 1 - \mathbb{E}[g_t(\bar{\mu}_{u,t} \mathbf{U}_* + \bar{\sigma}_{u,t} G)^2]. \quad (237)$$

Combining (233) with (236)–(237) yields

$$\bar{\mu}_{v,t+1} = \theta (1 - \text{mmse}_U(\bar{w}_{1,t})), \quad \bar{\sigma}_{v,t+1}^2 = 1 - \text{mmse}_U(\bar{w}_{1,t}). \quad (238)$$

Hence, by definition of $\bar{w}_{2,t}$,

$$\frac{\bar{w}_{2,t+1}}{1 - \bar{w}_{2,t+1}} = \frac{\bar{\mu}_{v,t+1}^2}{\bar{\sigma}_{v,t+1}^2} = \theta^2 (1 - \text{mmse}_U(\bar{w}_{1,t})). \quad (239)$$

An analogous application of the tower property to (235) gives

$$\frac{\bar{w}_{1,t}}{1 - \bar{w}_{1,t}} = \frac{\bar{\mu}_{u,t}^2}{\bar{\sigma}_{u,t}^2} = \frac{\theta^2}{\delta} (1 - \text{mmse}_V(\bar{w}_{2,t})). \quad (240)$$

Finally, letting $t \rightarrow \infty$ in (239)–(240) under the assumed convergence $\bar{w}_{1,t} \rightarrow \bar{w}_1$ and $\bar{w}_{2,t} \rightarrow \bar{w}_2$ gives exactly (232). $\square$

F.3 Proof of Proposition 3

We now prove the main proposition regarding the equivalence of the OAMP and AMP state evolution equations. The goal is to show that, by Lemma 10, the fixed-point equations of the OAMP algorithm (41) can be simplified into

$$\frac{w_1}{1 - w_1} = \frac{\theta^2}{\delta} (1 - \text{mmse}_V(w_2)), \quad (241a)$$

$$\frac{w_2}{1 - w_2} = \theta^2 (1 - \text{mmse}_U(w_1)). \quad (241b)$$

Our proof relies on the following lemma.

Lemma 11. *Define the auxiliary parameters*

$$a(\rho_1) \stackrel{\text{def}}{=} \frac{\delta}{\theta^2}(1 + \rho_1) + 1, \quad b(\rho_2) \stackrel{\text{def}}{=} \frac{\delta}{\theta^2}(1 + \rho_2) + \delta, \quad (242a)$$

$$T(\lambda_*) = \rho_2 \left(1 + \frac{\delta}{\theta^2}\right) + \rho_1 \left(\delta + \frac{\delta}{\theta^2}\right) + \rho_1 \rho_2 \frac{\delta}{\theta^2} + \lambda_*. \quad (242b)$$

Let $\mathcal{S}_\mu(z)$ denote the Stieltjes transform of the MP law

$$\mathcal{S}_\mu(z) = \frac{1}{2\delta z} \left[z + \delta - 1 - \sqrt{(z - \delta - 1)^2 - 4\delta} \right]. \quad (242c)$$

Then the resolvent $\mathcal{S}_\mu$ satisfies the coupled system for $T(\lambda_*)$ outside the bulk

$$a(\rho_1) [\delta \mathcal{S}_\mu(T(\lambda_*)) + \frac{1-\delta}{T(\lambda_*)}] = 1 + \frac{\delta}{\theta^2} \left(1 - \frac{1 - \rho_1 b(\rho_2) \mathcal{S}_\mu(T(\lambda_*))}{b(\rho_2) \mathcal{S}_\mu(T(\lambda_*))} \right), \quad (242d)$$

$$b(\rho_2) \mathcal{S}_\mu(T(\lambda_*)) = 1 + \frac{1}{\theta^2} \left(1 - \frac{1 - \rho_2 a(\rho_1) [\delta \mathcal{S}_\mu(T(\lambda_*)) + \frac{1-\delta}{T(\lambda_*)}]}{a(\rho_1) [\delta \mathcal{S}_\mu(T(\lambda_*)) + \frac{1-\delta}{T(\lambda_*)}]} \right). \quad (242e)$$

Proof. We show that (242d)–(242e) are equivalent to the MP self-consistent equation at $z = T(\lambda_*)$. Throughout, set

$$\mathcal{S}_\mu \stackrel{\text{def}}{=} \mathcal{S}_\mu(T(\lambda_*)), \quad T \stackrel{\text{def}}{=} T(\lambda_*), \quad \mathcal{S}_{\tilde{\mu}} \stackrel{\text{def}}{=} \delta \mathcal{S}_\mu + \frac{1-\delta}{T}.$$

Start from (242e) and isolate $a(\rho_1) \mathcal{S}_{\tilde{\mu}}$:

$$b(\rho_2) \mathcal{S}_\mu = 1 + \frac{1}{\theta^2} \left(1 - \frac{1 - \rho_2 a(\rho_1) \mathcal{S}_{\tilde{\mu}}}{a(\rho_1) \mathcal{S}_{\tilde{\mu}}} \right) = 1 + \frac{1}{\theta^2} \left(1 + \rho_2 - \frac{1}{a(\rho_1) \mathcal{S}_{\tilde{\mu}}} \right), \quad (243)$$

$$\theta^2 (b(\rho_2) \mathcal{S}_\mu - 1) = 1 + \rho_2 - \frac{1}{a(\rho_1) \mathcal{S}_{\tilde{\mu}}} \implies a(\rho_1) \mathcal{S}_{\tilde{\mu}} = \frac{1}{1 + \rho_2 + \theta^2 - \theta^2 b(\rho_2) \mathcal{S}_\mu}. \quad (244)$$

Substitute (244) into (242d) and simplify the right-hand side:

$$1 + \frac{\delta}{\theta^2} \left(1 - \frac{1 - \rho_1 b(\rho_2) \mathcal{S}_\mu}{b(\rho_2) \mathcal{S}_\mu} \right) = 1 + \frac{\delta}{\theta^2} \left(\frac{(1 + \rho_1) b(\rho_2) \mathcal{S}_\mu - 1}{b(\rho_2) \mathcal{S}_\mu} \right) = a(\rho_1) - \frac{\delta}{\theta^2 b(\rho_2) \mathcal{S}_\mu}. \quad (245)$$

Equating (244) and (245) gives a single identity in $\mathcal{S}_\mu$:

$$\frac{1}{1 + \rho_2 + \theta^2 - \theta^2 b(\rho_2) \mathcal{S}_\mu} = a(\rho_1) - \frac{\delta}{\theta^2 b(\rho_2) \mathcal{S}_\mu}. \quad (246)$$

Clearing denominators and rearranging yields the quadratic

$$(a(\rho_1) \theta^2 b(\rho_2)) \mathcal{S}_\mu^2 - [a(\rho_1) (1 + \rho_2 + \theta^2) + \delta - 1] \mathcal{S}_\mu + \frac{\delta(1 + \rho_2 + \theta^2)}{\theta^2 b(\rho_2)} = 0. \quad (247)$$

Using $a(\rho_1) = 1 + \frac{\delta}{\theta^2}(1 + \rho_1)$ and $b(\rho_2) = \delta + \frac{\delta}{\theta^2}(1 + \rho_2)$, one directly computes

$$\theta^2 a(\rho_1) b(\rho_2) = \delta T, \quad \frac{\delta(1 + \rho_2 + \theta^2)}{\theta^2 b(\rho_2)} = 1, \quad a(\rho_1)(1 + \rho_2 + \theta^2) + \delta - 1 = T + \delta - 1. \quad (248)$$

Substituting these into (247) gives

$$\delta T \mathcal{S}_\mu^2 - (T + \delta - 1) \mathcal{S}_\mu + 1 = 0, \quad (249)$$

which is precisely the MP self-consistent quadratic at $z = T(\lambda_*)$. This completes the proof. $\square$

Proof of Proposition 3. We prove the identity for the $\mathbf{u}$ -channel; the proof for the symmetric identity is analogous. The derivation proceeds by simplifying the LHS and showing its equivalence to the RHS via the resolvent system in Lemma 11.

$$\begin{aligned} \text{mmse}_U(w_1) &\stackrel{(a)}{=} \frac{1}{\rho_1 + \frac{1}{1-w_1}} \stackrel{(b)}{=} b(\rho_2) \mathcal{S}_\mu(T(\lambda_*)) \stackrel{(c)}{=} 1 + \frac{1}{\theta^2} \left(1 - \frac{1 - \rho_2 a(\rho_1) \left[\delta \mathcal{S}_\mu(T(\lambda_*)) + \frac{1-\delta}{T(\lambda_*)} \right]}{a(\rho_1) \left[\delta \mathcal{S}_\mu(T(\lambda_*)) + \frac{1-\delta}{T(\lambda_*)} \right]} \right) \\ &= 1 - \frac{1}{\theta^2} \left(\frac{1}{a(\rho_1) \left[\delta \mathcal{S}_\mu(T(\lambda_*)) + \frac{1-\delta}{T(\lambda_*)} \right]} - 1 - \rho_2 \right) \stackrel{(d)}{=} 1 - \frac{1}{\theta^2} \frac{w_2}{1 - w_2}. \end{aligned}$$

Here, step (a) follows from the definitions of the parameters. Step (b) is a simplification using the fixed-point equation for w_1 (230a). Step (c) applies the resolvent system identity (242e) from Lemma 11. Step (d) uses the expression for effective SNR derived from the fixed-point equation for w_2 (230b). $\square$

G Proof of The Optimal Spectral Estimators

This appendix provides a detailed analysis comparing the performance of optimal OAMP with that of optimal linear spectral methods. The core of the analysis is to first establish the theoretical performance limit of any estimator based on linear combinations of outlying singular vectors (cf. Proposition 1). We then demonstrate how to construct estimators that achieve this optimal performance in Proposition 6. A crucial element of this construction is a procedure for determining the relative signs of the different outlying singular vector components, which is achievable under a non-Gaussian signal assumption.

G.1 Proof of Proposition 4

We prove the claim for $\mathbf{u}$; the argument for $\mathbf{v}$ is identical. Using the empirical outlier index set $\mathcal{I}_M$ from (46), write

$$\mathbf{u}_{\text{PCA}}(\mathbf{c}_u) = \sqrt{M} \sum_{i \in \mathcal{I}_M} c_{u,i} \mathbf{u}_i(\mathbf{Y}),$$

and fix any coefficient vector $\mathbf{c}_u \neq \mathbf{0}$. By Proposition 1(2)–(3), almost surely for all sufficiently large M , each $i \in \mathcal{I}_M$ corresponds to a unique $\lambda_i \in \mathcal{K}_*$, and we use this identification below. Then

$$\begin{aligned} \limsup_{M \rightarrow \infty} \frac{\langle \mathbf{u}_{\text{PCA}}(\mathbf{c}_u), \mathbf{u}_* \rangle^2}{\|\mathbf{u}_{\text{PCA}}(\mathbf{c}_u)\|^2 \|\mathbf{u}_*\|^2} &\stackrel{(a)}{=} \limsup_{M \rightarrow \infty} \frac{\left(\sum_{i \in \mathcal{I}_M} c_{u,i} \langle \mathbf{u}_i(\mathbf{Y}), \mathbf{u}_* \rangle \right)^2}{\left\| \sum_{i \in \mathcal{I}_M} c_{u,i} \mathbf{u}_i(\mathbf{Y}) \right\|^2 \|\mathbf{u}_*\|^2} \\ &\stackrel{(b)}{=} \limsup_{M \rightarrow \infty} \frac{\left(\sum_{i \in \mathcal{I}_M} c_{u,i} \langle \mathbf{u}_i(\mathbf{Y}), \mathbf{u}_* / \sqrt{M} \rangle \right)^2}{(\|\mathbf{u}_*\|^2 / M) \sum_{i \in \mathcal{I}_M} c_{u,i}^2} \\ &\stackrel{(c)}{\leq} \limsup_{M \rightarrow \infty} \frac{1}{\|\mathbf{u}_*\|^2 / M} \sum_{i \in \mathcal{I}_M} \langle \mathbf{u}_i(\mathbf{Y}), \mathbf{u}_* / \sqrt{M} \rangle^2 \\ &\stackrel{(d)}{=} \sum_{\lambda \in \mathcal{K}_*} \nu_1(\{\lambda\}), \quad \text{a.s.} \end{aligned} \tag{250}$$

where (a) substitutes $\mathbf{u}_{\text{PCA}}(\mathbf{c}_u)$ and cancels the common factor M ; (b) uses the orthonormality of $\{\mathbf{u}_i(\mathbf{Y})\}_{i=1}^M$ and $\langle \mathbf{u}_i(\mathbf{Y}), \mathbf{u}_* \rangle = \sqrt{M} \langle \mathbf{u}_i(\mathbf{Y}), \mathbf{u}_* / \sqrt{M} \rangle$; (c) follows from Cauchy–Schwarz in $\mathbb{R}^{|\mathcal{I}_M|}$; and (d) uses $M^{-1} \|\mathbf{u}_*\|^2 \rightarrow 1$ and Proposition 1(4) together with Claim (2) and (3). Moreover, (c) is tight if and only if $(c_{u,i})_{i \in \mathcal{I}_M}$ is proportional to $(\langle \mathbf{u}_i(\mathbf{Y}), \mathbf{u}_* / \sqrt{M} \rangle)_{i \in \mathcal{I}_M}$, i.e., the oracle choice in (49).

G.2 Proof of Proposition 5

To justify the empirical procedure for relative sign detection, we must understand the *joint* distribution of the singular vectors of $\mathbf{Y}$ associated with all outlying singular values in $\mathcal{K}^*$. This can be obtained from a self-consistent representation of these singular vectors, in parallel with the OAMP state evolution in Theorem 1.

Proof of Proposition 5. We focus on the left singular vectors; the proof for the right singular vectors is entirely symmetric.

Proof of Claim (1). Fix the finite set of population outlying squared singular values $\mathcal{K}^* = \{\lambda_1, \dots, \lambda_K\}$, and write $\sigma_k \stackrel{\text{def}}{=} \sqrt{\lambda_k}$ for $1 \leq k \leq K$. For each k , let $\lambda_{k,M}$ denote the corresponding empirical outlying squared singular value of $\mathbf{Y}$, with associated singular value $\sigma_{k,M} \stackrel{\text{def}}{=} \sqrt{\lambda_{k,M}}$, and let $(\mathbf{u}_k, \mathbf{v}_k)$ be the associated left and right singular vectors. By Proposition 1, for all sufficiently large M , the indexing is well defined and $\lambda_{k,M} \xrightarrow{\text{a.s.}} \lambda_k$ (equivalently, $\sigma_{k,M} \xrightarrow{\text{a.s.}} \sigma_k$); in particular, $\sigma_{k,M} \neq 0$ almost surely for all large M (cf. Lemma 1, Claim (3)).

Since $\sigma_{k,M} \neq 0$, for each k we can write the singular vector into the following resolvent-based representation (cf. Fact 5):

$$\mathbf{u}_k = (\lambda_{k,M} \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1} \left(\frac{\theta \sigma_{k,M}}{\sqrt{MN}} \langle \mathbf{v}_*, \mathbf{v}_k \rangle \mathbf{u}_* + \frac{\theta}{\sqrt{MN}} \langle \mathbf{u}_*, \mathbf{u}_k \rangle \mathbf{W} \mathbf{v}_* \right), \quad (251a)$$

where the inverse exists almost surely for all sufficiently large M , since $\lambda_{k,M}$ lies outside the spectrum of $\mathbf{W} \mathbf{W}^\top$. We are interested in the projections of $\mathbf{u}_*$ along the outlying left singular vectors $\mathbf{u}_k$. Multiplying both sides by $\langle \mathbf{u}_*, \mathbf{u}_k \rangle$ yields

$$\begin{aligned} \mathbf{u}_k^{\text{OUT}} \stackrel{\text{def}}{=} \langle \mathbf{u}_*, \mathbf{u}_k \rangle \mathbf{u}_k &= (\lambda_{k,M} \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1} \left(\frac{\theta \sigma_{k,M}}{\sqrt{MN}} \langle \mathbf{v}_*, \mathbf{v}_k \rangle \langle \mathbf{u}_*, \mathbf{u}_k \rangle \mathbf{u}_* + \frac{\theta}{\sqrt{MN}} \langle \mathbf{u}_*, \mathbf{u}_k \rangle^2 \mathbf{W} \mathbf{v}_* \right) \\ &\stackrel{M \rightarrow \infty}{\simeq} (\lambda_k \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1} \left(\frac{\theta \sigma_k}{\sqrt{MN}} \langle \mathbf{v}_*, \mathbf{v}_k \rangle \langle \mathbf{u}_*, \mathbf{u}_k \rangle \mathbf{u}_* + \frac{\theta}{\sqrt{MN}} \langle \mathbf{u}_*, \mathbf{u}_k \rangle^2 \mathbf{W} \mathbf{v}_* \right) \\ &\stackrel{\text{def}}{=} \mathbf{p}_{u,k}. \end{aligned}$$

The approximation error above is due to the replacement of the empirical eigenvalue $\lambda_{k,M}$ and singular value $\sigma_{k,M}$ by their population limits λ_k and σ_k . To justify this replacement, note first that $\sigma_{k,M} \xrightarrow{\text{a.s.}} \sigma_k$, so the difference between the corresponding scalar prefactors vanishes. For the resolvent term, we use the resolvent identity

$$(\lambda_{k,M} \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1} - (\lambda_k \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1} = (\lambda_k - \lambda_{k,M}) (\lambda_{k,M} \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1} (\lambda_k \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1}. \quad (253)$$

Since $\lambda_{k,M}$ and λ_k remain at a strictly positive distance from $\text{sp}(\mathbf{W} \mathbf{W}^\top)$ almost surely for all sufficiently large M , the standard resolvent bound $\|(\lambda \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1}\|_{\text{op}} = \text{d}(\lambda, \text{sp}(\mathbf{W} \mathbf{W}^\top))^{-1}$ implies that both $\|(\lambda_{k,M} \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1}\|_{\text{op}}$ and $\|(\lambda_k \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1}\|_{\text{op}}$ are uniformly bounded. Together with $\lambda_{k,M} \xrightarrow{\text{a.s.}} \lambda_k$, this shows that the error incurred by replacing $(\lambda_{k,M}, \sigma_{k,M})$ with (λ_k, σ_k) is asymptotically negligible under any of the convergence notions used below.

We next analyze the asymptotic distributions of $\mathbf{p}_{u,k}$:

$$\mathbf{p}_{u,k} = (\lambda_k \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1} \left(\frac{\theta \sigma_k}{\sqrt{MN}} \langle \mathbf{v}_*, \mathbf{v}_k \rangle \langle \mathbf{u}_*, \mathbf{u}_k \rangle \mathbf{u}_* + \frac{\theta}{\sqrt{MN}} \langle \mathbf{u}_*, \mathbf{u}_k \rangle^2 \mathbf{W} \mathbf{v}_* \right). \quad (254)$$

To isolate the deterministic signal-aligned contribution, we subtract and add the averaged resolvent trace acting on $\mathbf{u}_*$. Specifically, write

$$(\lambda_k \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1} \mathbf{u}_* = \frac{1}{M} \text{Tr}(\lambda_k \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1} \mathbf{u}_* + \left[(\lambda_k \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1} - \frac{1}{M} \text{Tr}(\lambda_k \mathbf{I} - \mathbf{W} \mathbf{W}^\top)^{-1} \mathbf{I}_M \right] \mathbf{u}_*.$$

Accordingly, we decompose $\mathbf{p}_{u,k}$ as

$$\mathbf{p}_{u,k} = \mathbf{s}_{u,k} + \mathbf{n}_{u,k}^{(1)} + \mathbf{n}_{u,k}^{(2)}, \quad (255a)$$

where the signal component $\mathbf{s}_{u,k}$ and the noise components $\mathbf{n}_{u,k}^{(1)}$ and $\mathbf{n}_{u,k}^{(2)}$ are respectively given by

$$\mathbf{s}_{u,k} \stackrel{\text{def}}{=} \left(\frac{\theta\sigma_k}{\sqrt{MN}} \langle \mathbf{v}_*, \mathbf{v}_k \rangle \langle \mathbf{u}_*, \mathbf{u}_k \rangle \right) \left(\frac{1}{M} \text{Tr}(\lambda_k \mathbf{I}_M - \mathbf{W}\mathbf{W}^\top)^{-1} \right) \mathbf{u}_*, \quad (255b)$$

$$\mathbf{n}_{u,k}^{(1)} \stackrel{\text{def}}{=} \left(\frac{\theta\sigma_k}{\sqrt{MN}} \langle \mathbf{v}_*, \mathbf{v}_k \rangle \langle \mathbf{u}_*, \mathbf{u}_k \rangle \right) \left[(\lambda_k \mathbf{I}_M - \mathbf{W}\mathbf{W}^\top)^{-1} - \frac{1}{M} \text{Tr}(\lambda_k \mathbf{I}_M - \mathbf{W}\mathbf{W}^\top)^{-1} \mathbf{I}_M \right] \mathbf{u}_*, \quad (255c)$$

$$\mathbf{n}_{u,k}^{(2)} \stackrel{\text{def}}{=} \left(\frac{\theta}{\sqrt{MN}} \langle \mathbf{u}_*, \mathbf{u}_k \rangle^2 \right) (\lambda_k \mathbf{I}_M - \mathbf{W}\mathbf{W}^\top)^{-1} \mathbf{W} \mathbf{v}_*. \quad (255d)$$

Collecting these vectors columnwise, define the matrices

$$\mathbf{P}_u \stackrel{\text{def}}{=} [\mathbf{p}_{u,1} \ \cdots \ \mathbf{p}_{u,K}], \quad \mathbf{S}_u \stackrel{\text{def}}{=} [\mathbf{s}_{u,1} \ \cdots \ \mathbf{s}_{u,K}], \quad \mathbf{N}_u \stackrel{\text{def}}{=} [\mathbf{n}_{u,1} \ \cdots \ \mathbf{n}_{u,K}],$$

so that

$$\mathbf{P}_u = \mathbf{S}_u + \mathbf{N}_u \in \mathbb{R}^{M \times K}.$$

We now analyze the joint limit of $(\mathbf{P}_u, \mathbf{S}_u, \mathbf{N}_u)$, where each $\mathbf{n}_{u,k} = \mathbf{n}_{u,k}^{(1)} + \mathbf{n}_{u,k}^{(2)}$ is the sum of the centered (resolvent- $\mathbf{u}_*$) fluctuation and the transverse $(\mathbf{W}\mathbf{v}_*)$ fluctuation.

Signal component. Recall that

$$\mathbf{s}_{u,k} = \left(\frac{\theta\sigma_k}{\sqrt{MN}} \langle \mathbf{v}_*, \mathbf{v}_k \rangle \langle \mathbf{u}_*, \mathbf{u}_k \rangle \right) \left(\frac{1}{M} \text{Tr}(\lambda_k \mathbf{I} - \mathbf{W}\mathbf{W}^\top)^{-1} \right) \mathbf{u}_*.$$

By Proposition 1 (Claim 2) and the definition of $\nu_3(\{\sigma_k\})$, the scalar prefactor converges almost surely to

$$\frac{\theta\sigma_k}{\sqrt{MN}} \langle \mathbf{v}_*, \mathbf{v}_k \rangle \langle \mathbf{u}_*, \mathbf{u}_k \rangle \xrightarrow{\text{a.s.}} 2\theta\sigma_k \frac{1+\delta}{\sqrt{\delta}} \nu_3(\{\sigma_k\}).$$

Moreover, by the standard trace convergence for resolvents:

$$\frac{1}{M} \text{Tr}(\lambda_k \mathbf{I} - \mathbf{W}\mathbf{W}^\top)^{-1} \xrightarrow{\text{a.s.}} \mathcal{S}_\mu(\lambda_k).$$

Combining these and substituting the explicit expression the point mass $\nu_3(\{\sigma_k\})$ from Lemma 2 yields

$$\begin{aligned} \left(2\theta\sigma_k \frac{1+\delta}{\sqrt{\delta}} \nu_3(\{\sigma_k\}) \right) \mathcal{S}_\mu(\lambda_k) &= 2\theta\sigma_k \frac{1+\delta}{\sqrt{\delta}} \left(-\frac{\sqrt{\delta}}{1+\delta} \frac{1}{2\theta^3\sigma_k C'(\lambda_k)} \right) \mathcal{S}_\mu(\lambda_k) \\ &= \frac{\mathcal{S}_\mu(\lambda_k)}{-\theta^2 C'(\lambda_k)} = \nu_1(\{\lambda_k\}). \end{aligned}$$

Hence, in the sense of empirical row laws,

$$\mathbf{S}_u \xrightarrow{W} \mathbf{S}_u \stackrel{\text{def}}{=} (\nu_1(\{\lambda_1\})\mathbf{U}_*, \dots, \nu_1(\{\lambda_K\})\mathbf{U}_*)^\top,$$

where $\mathbf{U}_*$ is the limiting signal coordinate distribution.

Noise component: joint Gaussianity of $\mathbf{n}_{u,k}^{(1)}$ and $\mathbf{n}_{u,k}^{(2)}$. Let $\mathbf{W} = \mathbf{U}_W \mathbf{\Sigma}_W \mathbf{V}_W^\top$ be the singular value decomposition of $\mathbf{W}$. Write

$$\tilde{\mathbf{u}} \stackrel{\text{def}}{=} \mathbf{U}_W^\top \mathbf{u}_*, \quad \tilde{\mathbf{v}} \stackrel{\text{def}}{=} \mathbf{V}_W^\top \mathbf{v}_*.$$

Under Assumption 1, $\mathbf{U}_W$ and $\mathbf{V}_W$ are independent Haar matrices and are independent of $(\mathbf{u}_*, \mathbf{v}_*)$, so $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$ have asymptotically standard normal coordinates and are independent of $\mathbf{\Sigma}_W$ in the sense of empirical laws (cf. [30, Appendix E–F]).

Define the diagonal matrix

$$\mathbf{G}_k \stackrel{\text{def}}{=} (\lambda_k \mathbf{I} - \mathbf{\Sigma}_W \mathbf{\Sigma}_W^\top)^{-1}, \quad \bar{g}_{k,M} \stackrel{\text{def}}{=} \frac{1}{M} \text{Tr}(\mathbf{G}_k).$$

Then the centered resolvent term can be written as

$$\mathbf{n}_{u,k}^{(1)} = \alpha_{k,M} \mathbf{U}_W (\mathbf{G}_k - \bar{g}_{k,M} \mathbf{I}) \tilde{\mathbf{u}}, \quad \alpha_{k,M} \stackrel{\text{def}}{=} \frac{\theta \sigma_k}{\sqrt{MN}} \langle \mathbf{v}_*, \mathbf{v}_k \rangle \langle \mathbf{u}_*, \mathbf{u}_k \rangle,$$

and the transverse term can be written as

$$\mathbf{n}_{u,k}^{(2)} = \beta_{k,M} \mathbf{U}_W \mathbf{D}_k \tilde{\mathbf{v}}, \quad \mathbf{D}_k \stackrel{\text{def}}{=} (\lambda_k \mathbf{I} - \Sigma_W \Sigma_W^\top)^{-1} \Sigma_W, \quad \beta_{k,M} \stackrel{\text{def}}{=} \frac{\theta}{\sqrt{MN}} \langle \mathbf{u}_*, \mathbf{u}_k \rangle^2.$$

Both $\alpha_{k,M}$ and $\beta_{k,M}$ converge almost surely to deterministic limits (jointly over k) by Proposition 1.

For each k , set

$$\mathbf{q}_k^{(1)} \stackrel{\text{def}}{=} \alpha_{k,M} (\mathbf{G}_k - \bar{g}_{k,M} \mathbf{I}) \tilde{\mathbf{u}}, \quad \mathbf{q}_k^{(2)} \stackrel{\text{def}}{=} \beta_{k,M} \mathbf{D}_k \tilde{\mathbf{v}}, \quad \mathbf{q}_k \stackrel{\text{def}}{=} \mathbf{q}_k^{(1)} + \mathbf{q}_k^{(2)},$$

so that

$$\mathbf{n}_{u,k} = \mathbf{n}_{u,k}^{(1)} + \mathbf{n}_{u,k}^{(2)} = \mathbf{U}_W \mathbf{q}_k.$$

Collecting $\mathbf{q}_k$ columnwise yields an $M \times K$ matrix

$$\mathbf{Q} \stackrel{\text{def}}{=} [\mathbf{q}_1 \ \cdots \ \mathbf{q}_K], \quad \text{so that} \quad \mathbf{N}_u = \mathbf{U}_W \mathbf{Q}.$$

Step 1 (Gaussian limit for the coefficient rows). Each coordinate of $(\mathbf{q}_1, \dots, \mathbf{q}_K)$ is an affine function of $(\tilde{u}_i, \tilde{v}_i)$ with coefficients given by bounded continuous functions of the singular values of $\mathbf{W}$ (through $\mathbf{G}_k$ and $\mathbf{D}_k$). By [30, Propositions E.2 and E.4] and [42, Lemma G.4], for fixed K , the empirical joint law of the rows of $\mathbf{Q}$ converges to a centered Gaussian vector $\mathbf{Q} \in \mathbb{R}^K$ with deterministic covariance matrix Σ_Q , i.e.,

$$\mathbf{Q} \xrightarrow{W} \mathbf{Q}, \quad \mathbf{Q} \sim \mathcal{N}(\mathbf{0}, \Sigma_Q).$$

Step 2 (Haar mixing). Since $\mathbf{Q} \perp \mathbf{U}_W$ under Assumption 1, rank- K Haar mixing (cf. [42, Lemma G.5]) implies that the empirical row law of $\mathbf{N}_u = \mathbf{U}_W \mathbf{Q}$ converges to a centered Gaussian vector $\mathbf{N}_u \in \mathbb{R}^K$ with covariance Σ_Q , independent of $\mathbf{U}_*$, namely,

$$(\mathbf{S}_u, \mathbf{N}_u) \xrightarrow{W} (\mathbf{S}_u, \mathbf{N}_u), \quad \mathbf{N}_u \sim \mathcal{N}(\mathbf{0}, \Sigma_Q), \quad \mathbf{N}_u \perp \mathbf{U}_*.$$

Finally, since addition is Lipschitz on $\mathbb{R}^K$, applying similar arguments as in [42, Lemma G.4] once more yields

$$\mathbf{P}_u = \mathbf{S}_u + \mathbf{N}_u \xrightarrow{W} \mathbf{S}_u + \mathbf{N}_u \stackrel{\text{def}}{=} (\mathbf{U}_1^{\text{OUT}}, \dots, \mathbf{U}_K^{\text{OUT}})^\top,$$

where

$$\mathbf{U}_k^{\text{OUT}} = \nu_1(\{\lambda_k\}) \mathbf{U}_* + \mathbf{N}_{u,k}, \quad 1 \leq k \leq K, \quad (\mathbf{N}_{u,1}, \dots, \mathbf{N}_{u,K}) \sim \mathcal{N}(\mathbf{0}, \Sigma_Q).$$

Identification of Σ_Q via orthogonality. For $k \neq \ell$, the vectors $\mathbf{p}_{u,k} = \langle \mathbf{u}_*, \mathbf{u}_k \rangle \mathbf{u}_k$ are orthogonal for each finite M , hence $\langle \mathbf{p}_{u,k}, \mathbf{p}_{u,\ell} \rangle = 0$. Passing to the row-law limit gives

$$\mathbb{E}[\mathbf{U}_k^{\text{OUT}} \mathbf{U}_\ell^{\text{OUT}}] = 0, \quad k \neq \ell.$$

Since $\mathbf{U}_k^{\text{OUT}} = \nu_1(\{\lambda_k\}) \mathbf{U}_* + \mathbf{N}_{u,k}$ with $\mathbf{N}_u \perp \mathbf{U}_*$ and $\mathbb{E}[\mathbf{U}_*^2] = 1$, this determines the off-diagonal entries of Σ_Q and yields the stated formula for Σ_u^{OUT} . Moreover, $\mathbb{E}[(\mathbf{U}_k^{\text{OUT}})^2] = \nu_1(\{\lambda_k\})$ and therefore

$$\text{Var}(\mathbf{N}_{u,k}) = \mathbb{E}[(\mathbf{U}_k^{\text{OUT}})^2] - \nu_1(\{\lambda_k\})^2 = \nu_1(\{\lambda_k\}) - \nu_1^2(\{\lambda_k\}).$$

Thus we have

$$\mathbf{U}_k^{\text{OUT}} = \nu_1(\{\lambda_k\}) \mathbf{U}_* + \sqrt{\nu_1(\{\lambda_k\}) - \nu_1^2(\{\lambda_k\})} \mathbf{Z}_{u,k}, \quad \mathbf{Z}_{u,k} \sim \mathcal{N}(0, 1) \perp \mathbf{U}_*. \quad (256)$$

The proof for the right singular vectors is identical, replacing $\mathbf{W}\mathbf{W}^\top$ by $\mathbf{W}^\top\mathbf{W}$ and interchanging the roles of $(\mathbf{u}_*, M)$ and $(\mathbf{v}_*, N)$ throughout, so we omit the details.

Proof of Claim (2). For $k \neq \ell$, the orthogonality of the empirical projected vectors $\mathbf{p}_{u,k} = \langle \mathbf{u}_*, \mathbf{u}_k(\mathbf{Y}) \rangle \mathbf{u}_k(\mathbf{Y})$ implies $\langle \mathbf{p}_{u,k}, \mathbf{p}_{u,\ell} \rangle = 0$ for each finite M . Passing to the row–law limit in (51) gives

$$\mathbb{E}[\mathbf{U}_k^{\text{OUT}} \mathbf{U}_\ell^{\text{OUT}}] = 0, \quad k \neq \ell.$$

Using the independence $(Z_{u,1}, \dots, Z_{u,K}) \perp\!\!\!\perp \mathbf{U}_*$ with $\mathbb{E}[\mathbf{U}_*^2] = 1$,

$$0 = \mathbb{E}[\mathbf{U}_k^{\text{OUT}} \mathbf{U}_\ell^{\text{OUT}}] = \nu_1(\{\lambda_k\})\nu_1(\{\lambda_\ell\}) + \sqrt{\nu_1(\{\lambda_k\}) - \nu_1^2(\{\lambda_k\})} \sqrt{\nu_1(\{\lambda_\ell\}) - \nu_1^2(\{\lambda_\ell\})} \mathbb{E}[Z_{u,k} Z_{u,\ell}],$$

since by Lemma 2 we have $\nu_1(\{\lambda_k\}) \in (0, 1)$, so for $k \neq \ell$,

$$\mathbb{E}[Z_{u,k} Z_{u,\ell}] = -\frac{\nu_1(\{\lambda_k\})\nu_1(\{\lambda_\ell\})}{\sqrt{\nu_1(\{\lambda_k\}) - \nu_1^2(\{\lambda_k\})} \sqrt{\nu_1(\{\lambda_\ell\}) - \nu_1^2(\{\lambda_\ell\})}}.$$

This yields the stated covariance entries for Σ_u^{OUT} . The same argument with ν_2 and $(\mathbf{V}_k^{\text{OUT}})$ gives the expression for Σ_v^{OUT} . It remains to justify that Σ_u^{OUT} is positive definite. Since ν_1 is a probability measure and $\nu_1^\parallel(\mathbb{R}_+) > 0$, the total atomic mass on the outliers satisfies

$$\sum_{k=1}^K \nu_1(\{\lambda_k\}) = 1 - \nu_1^\parallel(\mathbb{R}_+) \in (0, 1).$$

For any $\mathbf{x} = (x_1, \dots, x_K)^\top \in \mathbb{R}^K$, define

$$y_k \stackrel{\text{def}}{=} \frac{x_k}{\sqrt{\nu_1(\{\lambda_k\}) - \nu_1^2(\{\lambda_k\})}}, \quad k = 1, \dots, K.$$

A direct computation using the explicit entries of Σ_u^{OUT} shows that

$$\mathbf{x}^\top \Sigma_u^{\text{OUT}} \mathbf{x} = \sum_{k=1}^K \nu_1(\{\lambda_k\}) y_k^2 - \left(\sum_{k=1}^K \nu_1(\{\lambda_k\}) y_k \right)^2. \quad (257)$$

By the Cauchy–Schwarz inequality with weights $\{\nu_1(\{\lambda_k\})\}_{k=1}^K$,

$$\left(\sum_{k=1}^K \nu_1(\{\lambda_k\}) y_k \right)^2 \leq \left(\sum_{k=1}^K \nu_1(\{\lambda_k\}) \right) \left(\sum_{k=1}^K \nu_1(\{\lambda_k\}) y_k^2 \right),$$

and hence, combining with (257),

$$\mathbf{x}^\top \Sigma_u^{\text{OUT}} \mathbf{x} \geq \left(1 - \sum_{k=1}^K \nu_1(\{\lambda_k\}) \right) \sum_{k=1}^K \nu_1(\{\lambda_k\}) y_k^2.$$

As observed above, $1 - \sum_{k=1}^K \nu_1(\{\lambda_k\}) > 0$ and $\nu_1(\{\lambda_k\}) > 0$ for all k , so the right-hand side is strictly positive whenever $\mathbf{x} \neq \mathbf{0}$ (equivalently, $(y_1, \dots, y_K) \neq \mathbf{0}$). Thus Σ_u^{OUT} is positive definite. The same reasoning, with ν_2 in place of ν_1 , shows that Σ_v^{OUT} is also positive definite. $\square$

G.3 Proof of Proposition 6

We treat the $\mathbf{u}$ –channel; the $\mathbf{v}$ –channel is identical. Recall from (58) that

$$\mathbf{u}_{\text{PCA}}^* = \sum_{i \in \mathcal{I}_M} s_i^u \sqrt{\nu_1(\{\lambda_i\})} \mathbf{u}_i^\#, \quad \mathbf{u}_i^\# = \sqrt{M} \xi_i \mathbf{u}_i(\mathbf{Y}).$$

By Proposition 1(2)–(3), almost surely for all sufficiently large M , each outlier window contains exactly one eigenvalue of $\mathbf{Y}\mathbf{Y}^\top$. Hence, for each $i \in \mathcal{I}_M$ we may associate a unique population outlier $\lambda_i \in \mathcal{K}_*$. Then

$$\begin{aligned} \lim_{M \rightarrow \infty} \frac{\langle \mathbf{u}_{\text{PCA}}^*, \mathbf{u}_* \rangle^2}{\|\mathbf{u}_{\text{PCA}}^*\|^2 \|\mathbf{u}_*\|^2} &\stackrel{(a)}{=} \lim_{M \rightarrow \infty} \frac{\left(\sum_{i \in \mathcal{I}_M} s_i^u \sqrt{\nu_1(\{\lambda_i\})} \langle \mathbf{u}_i^\#, \mathbf{u}_* \rangle \right)^2}{\left(M \sum_{i \in \mathcal{I}_M} \nu_1(\{\lambda_i\}) \right) \|\mathbf{u}_*\|^2} \\ &\stackrel{(b)}{=} \lim_{M \rightarrow \infty} \frac{\left(\sum_{i \in \mathcal{I}_M} s_i^u \sqrt{\nu_1(\{\lambda_i\})} \left(\frac{1}{M} \langle \mathbf{u}_i^\#, \mathbf{u}_* \rangle \right) \right)^2}{\left(\sum_{i \in \mathcal{I}_M} \nu_1(\{\lambda_i\}) \right) \left(\frac{1}{M} \|\mathbf{u}_*\|^2 \right)} \\ &\stackrel{(c)}{\xrightarrow{\text{a.s.}}} \sum_{\lambda \in \mathcal{K}_*} \nu_1(\{\lambda\}). \end{aligned} \quad (258)$$

Here (a) expands $\mathbf{u}_{\text{PCA}}^*$ and uses $\langle \mathbf{u}_i^\#, \mathbf{u}_j^\# \rangle = M \mathbf{1}\{i = j\}$ to evaluate $\|\mathbf{u}_{\text{PCA}}^*\|^2 = M \sum_{i \in \mathcal{I}_M} \nu_1(\{\lambda_i\})$. Step (b) divides the numerator and denominator by M^2 . Step (c) uses $\frac{1}{M} \|\mathbf{u}_*\|^2 \xrightarrow{\text{a.s.}} 1$ and, for each fixed $i \in \mathcal{I}_M$,

$$\frac{1}{M} \langle \mathbf{u}_i^\#, \mathbf{u}_* \rangle \xrightarrow{\text{a.s.}} [\mathbf{s}_{u,*}^R]_i \sqrt{\nu_1(\{\lambda_i\})} \quad \text{and} \quad s_i^u \xrightarrow{\text{a.s.}} [\mathbf{s}_{u,*}^R]_i,$$

which follow from Proposition 5 together with (60), and from Proposition 1(4) for the limiting overlaps. Finally, by Proposition 1(2) the correspondence $i \in \mathcal{I}_M \leftrightarrow \lambda_i \in \mathcal{K}_*$ for all large M , so $\sum_{i \in \mathcal{I}_M} \nu_1(\{\lambda_i\}) = \sum_{\lambda \in \mathcal{K}_*} \nu_1(\{\lambda\})$.

G.4 Proof of Proposition 7

We work throughout with the limit model (264) and the notation introduced in Section 4.3. The goal is to characterize when the global sign vector is identifiable from the rows of $\mathbf{U}^\#$, and to show that the MLE is asymptotically consistent whenever identifiability holds. We first analyze the Gaussian case, then the non-Gaussian case in the $\mathbf{u}$ -channel, and finally use the inter-channel coupling to transfer the result to the $\mathbf{v}$ -channel.

G.4.1 Both Gaussian priors: impossibility of sign recovery

In this part we show that, when both $\mathbf{U}_*$ and $\mathbf{V}_*$ are standard Gaussian, the distribution of the observed rows is independent of the sign vector. Consequently, the relative signs are not identifiable and no estimator can be consistent.

Lemma 12 (Gaussian non-identifiability). *Assume the setting of Proposition 5, and suppose $\mathbf{U}_* \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Let P_s be defined as in Proposition 7. Then $P_s = \mathcal{N}(\mathbf{0}, \mathbf{I}_K)$; in particular, its law does not depend on $\mathbf{s}$.*

Proof. By definition, P_s is the joint density of

$$([\mathbf{s}]_\ell \sqrt{\nu_1(\{\lambda_\ell\})} \mathbf{U}_* + \sqrt{1 - \nu_1(\{\lambda_\ell\})} \mathbf{Z}_\ell)_{\ell \in \mathcal{I}_M},$$

where $\mathbf{U}_*$ is independent of

$$(\sqrt{1 - \nu_1(\{\lambda_\ell\})} \mathbf{Z}_\ell)_{\ell \in \mathcal{I}_M} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_K - \gamma(\mathbf{s})\gamma(\mathbf{s})^\top), \quad \gamma(\mathbf{s}) \stackrel{\text{def}}{=} ([\mathbf{s}]_\ell \sqrt{\nu_1(\{\lambda_\ell\})})_{\ell \in \mathcal{I}_M}.$$

If $\mathbf{U}_* \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, then

$$([\mathbf{s}]_\ell \sqrt{\nu_1(\{\lambda_\ell\})} \mathbf{U}_*)_{\ell \in \mathcal{I}_M} \sim \mathcal{N}(\mathbf{0}, \gamma(\mathbf{s})\gamma(\mathbf{s})^\top),$$

and hence, by independence,

$$P_s = \mathcal{N}(\mathbf{0}, \gamma(\mathbf{s})\gamma(\mathbf{s})^\top + \mathbf{I}_K - \gamma(\mathbf{s})\gamma(\mathbf{s})^\top) = \mathcal{N}(\mathbf{0}, \mathbf{I}_K),$$

which does not depend on $\mathbf{s}$. □

G.4.2 Non-Gaussian Priors: Identifiability and MLE Consistency

We now assume that $\mathbf{U}_*$ is non-Gaussian and focus on the $\mathbf{u}$ -channel. For each $\mathbf{s} \in \mathcal{S}_r$, let $P_{\mathbf{s}}$ be as defined in Proposition 7. The proof mostly follows the classical MLE scheme [70, Chapter 5], where the only caveat being that we obtain uniform convergence not from a direct law of large numbers but from the Wasserstein convergence of the empirical row measure to $P_{\mathbf{s}_{u,*}^R}$ in (60). We proceed in the following steps:

Uniform convergence of empirical log-likelihoods. Throughout this section we reindex the empirical outlier index set as $\mathcal{I}_M = \{1, \dots, K\}$, where $K = |\mathcal{I}_M|$. For each $\mathbf{s} \in \mathcal{S}_r$, let $P_{\mathbf{s}}$ be as defined in Proposition 7. Denote the M rows of the singular vectors matrix $\mathbf{U}^\sharp$ by $(\mathbf{U}_{i,:}^\sharp)_{i \leq M}$. Define the sample and population log-likelihoods, respectively, by

$$L_M(\mathbf{s}; \mathbf{U}^\sharp) \stackrel{\text{def}}{=} \frac{1}{M} \sum_{i=1}^M \log P_{\mathbf{s}}(\mathbf{U}_{i,:}^\sharp), \quad (259)$$

$$L(\mathbf{s}; \mathbf{U}^\sharp) \stackrel{\text{def}}{=} \mathbb{E}_{P_{\mathbf{s}_{u,*}^R}} [\log P_{\mathbf{s}}(\mathbf{U}^\sharp)], \quad (260)$$

i.e., $L(\mathbf{s}; \mathbf{U}^\sharp)$ is computed under the true law $P_{\mathbf{s}_{u,*}^R}$ (equivalently, conditioning on the ground truth sign vector $\mathbf{s}_{u,*}^R$). The following lemma establishes uniform convergence of the empirical criterion $L_M(\mathbf{s}; \mathbf{U}^\sharp)$ to $L(\mathbf{s}; \mathbf{U}^\sharp)$ over the finite set $\mathcal{S}_r$.

Lemma 13 (Uniform convergence of empirical log-likelihoods). *Let $L_M(\mathbf{s}; \mathbf{U}^\sharp)$ and $L(\mathbf{s}; \mathbf{U}^\sharp)$ be defined in (259)–(260). Then*

$$\sup_{\mathbf{s} \in \mathcal{S}_r} |L_M(\mathbf{s}; \mathbf{U}^\sharp) - L(\mathbf{s}; \mathbf{U}^\sharp)| \xrightarrow{\text{a.s.}} 0, \quad (261)$$

where it is understood that the same ground truth sign vector $\mathbf{s}_{u,*}^R$ is shared by both $\mathbf{U}^\sharp$ and $\mathbf{U}^\sharp$.

Proof. By Proposition 5, the convergence

$$(\langle \mathbf{u}_*, \mathbf{u}_1(\mathbf{Y}) \rangle \mathbf{u}_1(\mathbf{Y}), \dots, \langle \mathbf{u}_*, \mathbf{u}_K(\mathbf{Y}) \rangle \mathbf{u}_K(\mathbf{Y})) \xrightarrow{W} (\mathbf{U}_1^{\text{OUT}}, \dots, \mathbf{U}_K^{\text{OUT}})^\top \quad (262)$$

holds in the sense of Wasserstein convergence of the empirical row measure. Moreover, by Proposition 1, $|\langle \mathbf{u}_*, \mathbf{u}_k(\mathbf{Y}) \rangle|/\sqrt{M} \xrightarrow{\text{a.s.}} \sqrt{\nu_1(\{\lambda_k\})}$ for each $k \in [K]$. Together with the sign randomization in (60), this implies that the relative sign vector $\mathbf{s}_{u,*}^R$ is independent of the remaining limit randomness, and hence the empirical row measure of $(\mathbf{u}_1^\sharp, \dots, \mathbf{u}_K^\sharp)$ converges in Wasserstein distance, under the conditional law given $\mathbf{s}_{u,*}^R$, i.e.,

$$(\mathbf{u}_1^\sharp, \dots, \mathbf{u}_K^\sharp) \xrightarrow{W} \mathbf{U}^\sharp(\mathbf{s}_{u,*}^R) \in \mathbb{R}^K, \quad (263)$$

where the ℓ -th marginal is

$$\mathbf{U}_\ell^\sharp([\mathbf{s}_{u,*}^R]_\ell) = [\mathbf{s}_{u,*}^R]_\ell \sqrt{\nu_1(\{\lambda_\ell\})} \mathbf{U}_* + \sqrt{1 - \nu_1(\{\lambda_\ell\})} \mathbf{Z}_\ell, \quad \ell \in [K]. \quad (264)$$

We first show that for any fixed $\mathbf{s} \in \mathcal{S}_r$,

$$L_M(\mathbf{s}; \mathbf{U}^\sharp) \xrightarrow{\text{a.s.}} L(\mathbf{s}; \mathbf{U}^\sharp). \quad (265)$$

To invoke the Wasserstein convergence (263), it suffices to verify the quadratic growth condition (cf. [71, Definition 6.8])

$$|\log P_{\mathbf{s}}(\mathbf{x})| \leq C(1 + \|\mathbf{x}\|^2), \quad \forall \mathbf{x} \in \mathbb{R}^K, \forall \mathbf{s} \in \mathcal{S}_r. \quad (266)$$

Conditioning (62) on a fixed $\mathbf{s} \in \mathcal{S}_r$ gives

$$P_{\mathbf{s}}(\mathbf{x}) = c(\mathbf{s}) \mathbb{E}_{\mathbf{U}_*} \left[\exp \left(-\frac{1}{2} (\mathbf{x} - \gamma(\mathbf{s}) \mathbf{U}_*)^\top \Sigma(\mathbf{s})^{-1} (\mathbf{x} - \gamma(\mathbf{s}) \mathbf{U}_*) \right) \right], \quad (267)$$

where

$$\boldsymbol{\gamma}(\mathbf{s}) \stackrel{\text{def}}{=} ([\mathbf{s}]_k \sqrt{\nu_1(\{\lambda_k\})})_{k \in [K]} \in \mathbb{R}^K, \quad (268a)$$

$$\boldsymbol{\Sigma}(\mathbf{s}) \stackrel{\text{def}}{=} \mathbf{I}_K - \boldsymbol{\gamma}(\mathbf{s})\boldsymbol{\gamma}(\mathbf{s})^\top, \quad (268b)$$

$$c(\mathbf{s}) \stackrel{\text{def}}{=} (2\pi)^{-K/2} (\det \boldsymbol{\Sigma}(\mathbf{s}))^{-1/2}. \quad (268c)$$

By Lemma 2, $\sum_{k=1}^K \nu_1(\{\lambda_k\}) < 1$, hence $\boldsymbol{\Sigma}(\mathbf{s}) \succ 0$ for all $\mathbf{s} \in \mathcal{S}_r$. Since $\mathcal{S}_r$ is finite, the eigenvalues of $\boldsymbol{\Sigma}(\mathbf{s})$ are uniformly bounded away from 0 and ∞ , and there exist $0 < c_{\min} \leq c_{\max} < \infty$ such that

$$c_{\min} \leq c(\mathbf{s}) \leq c_{\max}, \quad \mathbf{s} \in \mathcal{S}_r. \quad (269)$$

Taking logarithms in (267) yields

$$\log P_s(\mathbf{x}) = \log c(\mathbf{s}) + \log \mathbb{E}_{\mathbf{U}_*} \left[\exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\gamma}(\mathbf{s})\mathbf{U}_*)^\top \boldsymbol{\Sigma}(\mathbf{s})^{-1} (\mathbf{x} - \boldsymbol{\gamma}(\mathbf{s})\mathbf{U}_*) \right) \right]. \quad (270)$$

For the upper bound, the quadratic form in (270) is nonnegative, hence

$$\log P_s(\mathbf{x}) \leq \log c(\mathbf{s}) \leq \log c_{\max}, \quad \forall \mathbf{x} \in \mathbb{R}^K, \forall \mathbf{s} \in \mathcal{S}_r. \quad (271)$$

For the lower bound we apply Jensen, expand the quadratic form, and then use uniform spectral bounds:

$$\begin{aligned} \log P_s(\mathbf{x}) &\stackrel{(a)}{\geq} \log c(\mathbf{s}) - \frac{1}{2} \mathbb{E} \left[(\mathbf{x} - \boldsymbol{\gamma}(\mathbf{s})\mathbf{U}_*)^\top \boldsymbol{\Sigma}(\mathbf{s})^{-1} (\mathbf{x} - \boldsymbol{\gamma}(\mathbf{s})\mathbf{U}_*) \right] \\ &\stackrel{(b)}{=} \log c(\mathbf{s}) - \frac{1}{2} \left(\mathbf{x}^\top \boldsymbol{\Sigma}(\mathbf{s})^{-1} \mathbf{x} - 2m \boldsymbol{\gamma}(\mathbf{s})^\top \boldsymbol{\Sigma}(\mathbf{s})^{-1} \mathbf{x} + \boldsymbol{\gamma}(\mathbf{s})^\top \boldsymbol{\Sigma}(\mathbf{s})^{-1} \boldsymbol{\gamma}(\mathbf{s}) \right) \\ &\stackrel{(c)}{\geq} \log c_{\min} - \frac{1}{2} (C_1 \|\mathbf{x}\|^2 + C_2) \stackrel{(d)}{\geq} -C(1 + \|\mathbf{x}\|^2), \quad \forall \mathbf{x} \in \mathbb{R}^K, \forall \mathbf{s} \in \mathcal{S}_r, \end{aligned} \quad (272)$$

where $m \stackrel{\text{def}}{=} \mathbb{E}[\mathbf{U}_*]$ and $\mathbb{E}[\mathbf{U}_*^2] = 1$. Here (a) applies Jensen's inequality to (270); (b) expands the expectation; (c) uses (269), uniform spectral bounds on $\boldsymbol{\Sigma}(\mathbf{s})^{-1}$, and boundedness of $\boldsymbol{\gamma}(\mathbf{s})$ to obtain constants $C_1, C_2 > 0$; and (d) absorbs constants into a single $C > 0$. Combining (271) and (272) yields (266).

By the conditional Wasserstein convergence (263) and the growth bound (266), for each fixed $\mathbf{s} \in \mathcal{S}_r$ we obtain

$$L_M(\mathbf{s}; \mathbf{U}^\sharp) = \frac{1}{M} \sum_{i=1}^M \log P_s(\mathbf{U}_{i,\cdot}^\sharp) \xrightarrow{\text{a.s.}} L(\mathbf{s}; \mathbf{U}^\sharp). \quad (273)$$

Since $\mathcal{S}_r$ is finite, pointwise almost-sure convergence implies the uniform convergence (261). $\square$

Consistency of MLE. With these ingredients in place, standard results of MLE consistency imply that any maximizer of the empirical log-likelihood over the finite set $\mathcal{S}_r$ converges almost surely to the unique maximizer of L , namely the true sign vector $\mathbf{s}_{u,*}^R$.

Lemma 14 (Consistency of the MLE under a non-Gaussian prior). *Assume the setting of Proposition 5 and Lemma 13, and suppose that the scalar signal $\mathbf{U}_*$ in (264) is not standard Gaussian with $\mathbb{E}[\mathbf{U}_*^2] = 1$. Let $\mathcal{I}_M$ be the outlier index set with $K = |\mathcal{I}_M|$, and let*

$$\mathcal{S}_r \stackrel{\text{def}}{=} \{\mathbf{s} \in \{\pm 1\}^K : [\mathbf{s}]_r = +1\}$$

and $\mathbf{s}_{u,*}^R \in \mathcal{S}_r$ be the ground truth sign vector as in Section 4.3. Let $\hat{\mathbf{s}}_u^{\text{MLE}}$ be an estimator defined in (63). Then

$$\hat{\mathbf{s}}_u^{\text{MLE}} \xrightarrow{\text{a.s.}} \mathbf{s}_{u,*}^R. \quad (274)$$

Proof. Throughout we reindex $\mathcal{I}_M = \{1, \dots, K\}$. We work under the conditional law given the ground-truth sign vector $\mathbf{s}_{u,*}^R$. Let $L_M(\mathbf{s}; \mathbf{U}^\#)$ and $L(\mathbf{s}; \mathbf{U}^\#)$ be as in Lemma 13. To prove (274), it suffices to show that $\mathbf{s}_{u,*}^R$ is the unique maximizer of $L(\cdot; \mathbf{U}^\#)$ on $\mathcal{S}_r$ and to invoke the uniform convergence (261).

We first establish identifiability on $\mathcal{S}_r$, namely

$$\mathbf{s} \neq \mathbf{t} \implies P_{\mathbf{s}} \neq P_{\mathbf{t}}, \quad \mathbf{s}, \mathbf{t} \in \mathcal{S}_r, \quad (275)$$

where $P_{\mathbf{s}}$ denotes the law induced by (62). Let $\phi_{U_*}(\omega) \stackrel{\text{def}}{=} \mathbb{E}[e^{i\omega U_*}]$ and define $\Psi(\omega) \stackrel{\text{def}}{=} \phi_{U_*}(\omega)e^{\omega^2/2}$. By (62), for any $\mathbf{w} \in \mathbb{R}^K$, the characteristic function under $P_{\mathbf{s}}$ yields

$$\Phi_{\mathbf{s}}(\mathbf{w}) = \mathbb{E}[e^{i\langle \mathbf{w}, \mathbf{U}^\# \rangle}] = e^{-\|\mathbf{w}\|^2/2} \Psi(\langle \mathbf{w}, \gamma(\mathbf{s}) \rangle), \quad (276)$$

with $\gamma(\mathbf{s})$ defined in (268a). To prove (275), it suffices to show that

$$\forall \mathbf{s} \neq \mathbf{t} \in \mathcal{S}_r, \exists \mathbf{w}_0 \in \mathbb{R}^K \text{ such that } \Phi_{\mathbf{s}}(\mathbf{w}_0) \neq \Phi_{\mathbf{t}}(\mathbf{w}_0). \quad (277)$$

Since U_* is not standard Gaussian, Ψ is non-constant (otherwise $\phi_{U_*}(\omega) = e^{-\omega^2/2}$). Choose $\omega_1 \neq \omega_2$ such that $\Psi(\omega_1) \neq \Psi(\omega_2)$. Fix $\mathbf{s} \neq \mathbf{t}$ in $\mathcal{S}_r$. If $\gamma(\mathbf{t}) = c\gamma(\mathbf{s})$ for some scalar c , then (268a) yields $c = [\mathbf{t}]_k/[\mathbf{s}]_k \in \{\pm 1\}$ for every $k \in \mathcal{I}_M$. Evaluating this identity at $k = r$ gives $c = 1$ (since $[\mathbf{s}]_r = [\mathbf{t}]_r = +1$), whereas evaluating it at an index j with $[\mathbf{s}]_j \neq [\mathbf{t}]_j$ gives $c = -1$, a contradiction. Assume $K \geq 2$ (the case $K = 1$ is trivial since $\mathcal{S}_r$ is a singleton, no relative sign needed). Hence the map $T(\mathbf{w}) \stackrel{\text{def}}{=} (\langle \mathbf{w}, \gamma(\mathbf{s}) \rangle, \langle \mathbf{w}, \gamma(\mathbf{t}) \rangle)$ is surjective, so there exists $\mathbf{w}_0$ such that $\langle \mathbf{w}_0, \gamma(\mathbf{s}) \rangle = \omega_1$ and $\langle \mathbf{w}_0, \gamma(\mathbf{t}) \rangle = \omega_2$. Plugging into (276) yields (277), and thus $P_{\mathbf{s}} \neq P_{\mathbf{t}}$, proving (275).

Next, identifiability implies that $\mathbf{s}_{u,*}^R$ uniquely maximizes the population log-likelihood. Since $L(\mathbf{s}; \mathbf{U}^\#)$ is computed under the true law $P_{\mathbf{s}_{u,*}^R}$, the definition of KL divergence gives

$$L(\mathbf{s}_{u,*}^R; \mathbf{U}^\#) - L(\mathbf{s}; \mathbf{U}^\#) = \text{KL}(P_{\mathbf{s}_{u,*}^R} \parallel P_{\mathbf{s}}) \geq 0,$$

with equality iff $\mathbf{s} = \mathbf{s}_{u,*}^R$ by (275). Since $\mathcal{S}_r$ is finite, the following gap is strictly positive:

$$\Delta \stackrel{\text{def}}{=} \min_{\mathbf{s} \in \mathcal{S}_r, \mathbf{s} \neq \mathbf{s}_{u,*}^R} \left(L(\mathbf{s}_{u,*}^R; \mathbf{U}^\#) - L(\mathbf{s}; \mathbf{U}^\#) \right) > 0.$$

Finally, Lemma 13 implies that, for any $\varepsilon > 0$, almost surely for all sufficiently large M ,

$$\sup_{\mathbf{s} \in \mathcal{S}_r} |L_M(\mathbf{s}; \mathbf{U}^\#) - L(\mathbf{s}; \mathbf{U}^\#)| \leq \varepsilon. \quad (278)$$

Set $\varepsilon = \Delta/3$. Then for such M , the triangle inequality yields

$$L_M(\mathbf{s}_{u,*}^R; \mathbf{U}^\#) \geq L(\mathbf{s}_{u,*}^R; \mathbf{U}^\#) - \varepsilon,$$

and for any $\mathbf{s} \neq \mathbf{s}_{u,*}^R$,

$$L_M(\mathbf{s}; \mathbf{U}^\#) \leq L(\mathbf{s}; \mathbf{U}^\#) + \varepsilon \leq L(\mathbf{s}_{u,*}^R; \mathbf{U}^\#) - \Delta + \varepsilon = L(\mathbf{s}_{u,*}^R; \mathbf{U}^\#) - 2\varepsilon.$$

Hence $L_M(\mathbf{s}_{u,*}^R; \mathbf{U}^\#) > L_M(\mathbf{s}; \mathbf{U}^\#)$ for all $\mathbf{s} \neq \mathbf{s}_{u,*}^R$, and therefore any maximizer $\hat{\mathbf{s}}_u^{\text{MLE}} \in \arg\max_{\mathbf{s} \in \mathcal{S}_r} L_M(\mathbf{s}; \mathbf{U}^\#)$ satisfies $\hat{\mathbf{s}}_u^{\text{MLE}} = \mathbf{s}_{u,*}^R$ for all sufficiently large M almost surely, i.e., $\hat{\mathbf{s}}_u^{\text{MLE}} \xrightarrow{\text{a.s.}} \mathbf{s}_{u,*}^R$. $\square$

G.4.3 Inter-channel Sign Coupling

Finally, we couple the $\mathbf{u}$ - and $\mathbf{v}$ -channel signs through the cross spectral measure ν_3 in Definition 5. The key point is that the outlier point mass $\nu_3(\{\sigma_k\})$ is nonzero, which yields an asymptotically deterministic inter-channel sign relation.

Lemma 15 (Inter-channel sign coupling). *Under the assumptions of Lemma 2 and Proposition 5, for any outlier index $k \in \mathcal{I}_M$ with limiting eigenvalue $\lambda_k \in \mathcal{K}^*$ and singular value $\sigma_k = \sqrt{\lambda_k}$,*

$$\text{sign} \left(\langle \mathbf{u}_k(\mathbf{Y}), \mathbf{u}_* \rangle \langle \mathbf{v}_k(\mathbf{Y}), \mathbf{v}_* \rangle \right) \xrightarrow{\text{a.s.}} \text{sign} \left(\nu_3(\{\sigma_k\}) \right), \quad (279)$$

and $\nu_3(\{\sigma_k\}) \neq 0$.

Proof. By the definition of $\nu_{L,3}$ in Definition 5 and the overlap convergence in Proposition 1,

$$\frac{1}{2(M+N)} \langle \mathbf{u}_k(\mathbf{Y}), \mathbf{u}_* \rangle \langle \mathbf{v}_k(\mathbf{Y}), \mathbf{v}_* \rangle \xrightarrow{\text{a.s.}} \nu_3(\{\sigma_k\}). \quad (280)$$

Writing $\lambda_k = \sigma_k^2$, Lemma 2 gives

$$\nu_3(\{\sigma_k\}) = \frac{\sqrt{\delta}}{1+\delta} \cdot \frac{\theta \mathcal{C}(\lambda_k)}{2\sigma_k \Gamma'(\lambda_k)} = \frac{\sqrt{\delta}}{1+\delta} \cdot \frac{1}{2\theta \sigma_k \Gamma'(\lambda_k)}, \quad (281)$$

where the second equality uses the master equation $1 - \theta^2 \mathcal{C}(\lambda_k) = 0$. Lemma 1 ensures $\Gamma'(\lambda_k) \neq 0$, hence $\nu_3(\{\sigma_k\}) \neq 0$. Since the limit in (280) is nonzero, taking signs in (280) yields (279). $\square$

Proof of Proposition 7. For Claim (1), assume without loss of generality that $\mathbf{U}_*$ is non-Gaussian. Then Lemma 14 gives $\hat{\mathbf{s}}_u^{\text{MLE}} \xrightarrow{\text{a.s.}} \mathbf{s}_{u,*}^{\text{R}}$. Lemma 15 yields an asymptotically deterministic relation between the channel-wise relative sign vectors; in particular, with the reference index r defining $\mathcal{S}_r$, set

$$[\hat{\mathbf{s}}_v^{\text{MLE}}]_j \stackrel{\text{def}}{=} [\hat{\mathbf{s}}_u^{\text{MLE}}]_j \text{sign}(\nu_3(\{\sigma_r\}) \nu_3(\{\sigma_j\})), \quad j \in \mathcal{I}_M.$$

Then $\hat{\mathbf{s}}_v^{\text{MLE}} \xrightarrow{\text{a.s.}} \mathbf{s}_{v,*}^{\text{R}}$. By symmetry, the same conclusion holds when only $\mathbf{V}_*$ is non-Gaussian.

For Claim (2), when both $\mathbf{U}_*$ and $\mathbf{V}_*$ are standard Gaussian, Lemma 12 (applied separately in each channel) shows that the row law of the scaled outlier singular vectors is $\mathcal{N}(\mathbf{0}, \mathbf{I}_K)$ and does not depend on the sign vector. Consequently, the likelihood is invariant over $\mathcal{S}_r$. $\square$

G.5 Proof of Proposition 8

Before proving Proposition 8, we explain heuristically why NGMC identifies the relative sign $[\mathbf{s}_{u,*}^{\text{R}}]_r [\mathbf{s}_{u,*}^{\text{R}}]_j$. Fix distinct outlier indices $r, j \in \mathcal{I}_M$. Then the corresponding empirical singular-vector coordinates admit the limit representation by (264)

$$\mathbf{U}_r^\# = [\mathbf{s}_{u,*}^{\text{R}}]_r \sqrt{\nu_1(\{\lambda_r\})} \mathbf{U}_* + \sqrt{1 - \nu_1(\{\lambda_r\})} \mathbf{Z}_r, \quad \mathbf{U}_j^\# = [\mathbf{s}_{u,*}^{\text{R}}]_j \sqrt{\nu_1(\{\lambda_j\})} \mathbf{U}_* + \sqrt{1 - \nu_1(\{\lambda_j\})} \mathbf{Z}_j, \quad (282)$$

where $\mathbf{U}_*$ is common and $(\mathbf{Z}_r, \mathbf{Z}_j)$ is a standard Gaussian pair independent of $\mathbf{U}_*$. For a contrast $f: \mathbb{R} \rightarrow \mathbb{R}$, set

$$T_f(a, b) \stackrel{\text{def}}{=} \mathbb{E} \left[f(\mathbf{U}_r^\#) \mathbf{U}_j^\# \right], \quad a, b \in \{\pm 1\},$$

where $\mathbf{U}_r^\#, \mathbf{U}_j^\#$ are formed with $a = [\mathbf{s}_{u,*}^{\text{R}}]_r$ and $b = [\mathbf{s}_{u,*}^{\text{R}}]_j$. If f is odd, then by (282) we have $T_f(-a, b) = -T_f(a, b)$ and $T_f(a, -b) = -T_f(a, b)$, hence T_f depends on (a, b) only through the product ab :

$$T_f(a, b) = ab C_f, \quad C_f \stackrel{\text{def}}{=} T_f(1, 1).$$

Therefore, whenever $C_f \neq 0$,

$$\text{sign } T_f([\mathbf{s}_{u,*}^{\text{R}}]_r, [\mathbf{s}_{u,*}^{\text{R}}]_j) = \text{sign}([\mathbf{s}_{u,*}^{\text{R}}]_r [\mathbf{s}_{u,*}^{\text{R}}]_j) \text{sign}(C_f),$$

so the sign of T_f recovers the relative sign up to a fixed global orientation. In NGMC we take $f(x) = x^{k+1}$ with even k from Assumption 3; the ensuing calculation shows $C_f \propto \mathbb{E}[\mathbf{U}_*^{k+2}] - (k+1)!! \neq 0$, ensuring non-degeneracy.

Proof of Proposition 8. We treat the $\mathbf{u}$ -channel; the $\mathbf{v}$ -channel follows by the same argument together with the inter-channel sign relation in Proposition 5. Fix distinct outlier indices $r, j \in \mathcal{I}_M$. Define the deterministic alignments

$$\gamma_k \stackrel{\text{def}}{=} \sqrt{\nu_1(\{\lambda_k\})} \in (0, 1), \quad \tilde{\gamma}_k \stackrel{\text{def}}{=} \sqrt{1 - \gamma_k^2}, \quad k \in \mathcal{I}_M,$$

and the limiting pair

$$\mathbf{U}_r^\# = [\mathbf{s}_{u,*}^R]_r \gamma_r \mathbf{U}_* + \tilde{\gamma}_r \mathbf{Z}_r, \quad \mathbf{U}_j^\# = [\mathbf{s}_{u,*}^R]_j \gamma_j \mathbf{U}_* + \tilde{\gamma}_j \mathbf{Z}_j, \quad (283)$$

where $\mathbb{E}[\mathbf{U}_*^2] = 1$ and $(\mathbf{Z}_r, \mathbf{Z}_j)$ is a standard Gaussian pair independent of $\mathbf{U}_*$. By Proposition 5 and (264), the joint law of empirical rows $(\mathbf{u}_r^\#, \mathbf{u}_j^\#) \xrightarrow{W} (\mathbf{U}_r^\#, \mathbf{U}_j^\#)$ by (62). Let $f(x) \stackrel{\text{def}}{=} x^{k+1}$ and $g(x_1, x_2) \stackrel{\text{def}}{=} x_1^{k+1} x_2$. Using the Wasserstein convergence above, the polynomial growth of g , we have

$$\frac{1}{M} f(\mathbf{u}_r^\#)^\top \mathbf{u}_j^\# = \frac{1}{M} \sum_{m=1}^M g([\mathbf{u}_r^\#]_m, [\mathbf{u}_j^\#]_m) \xrightarrow{\text{a.s.}} \mathbb{E}[(\mathbf{U}_r^\#)^{k+1} \mathbf{U}_j^\#]. \quad (284)$$

We will show that

$$\mathbb{E}[(\mathbf{U}_r^\#)^{k+1} \mathbf{U}_j^\#] = [\mathbf{s}_{u,*}^R]_r [\mathbf{s}_{u,*}^R]_j \nu_1(\{\lambda_r\})^{(k+1)/2} \sqrt{\nu_1(\{\lambda_j\})} (\mathbb{E}[\mathbf{U}_*^{k+2}] - (k+1)!!), \quad (285)$$

and therefore (284) implies

$$\frac{1}{M} f(\mathbf{u}_r^\#)^\top \mathbf{u}_j^\# \xrightarrow{\text{a.s.}} [\mathbf{s}_{u,*}^R]_r [\mathbf{s}_{u,*}^R]_j \nu_1(\{\lambda_r\})^{(k+1)/2} \sqrt{\nu_1(\{\lambda_j\})} (\mathbb{E}[\mathbf{U}_*^{k+2}] - (k+1)!!). \quad (286)$$

Acknowledging (286) and taking sign of both sides, the NGMC sign recovers the relative sign $[\mathbf{s}_{u,*}^R]_r [\mathbf{s}_{u,*}^R]_j$ up to the deterministic orientation sign $(\mathbb{E}[\mathbf{U}_*^{k+2}] - (k+1)!!) \neq 0$ (Assumption 3); with the convention $[\mathbf{s}_{u,*}^R]_r = +1$, this yields $\hat{s}_{u,j}^{\text{NGMC}} \xrightarrow{\text{a.s.}} [\mathbf{s}_{u,*}^R]_j$.

Hence it remains to compute $\mathbb{E}[(\mathbf{U}_r^\#)^{k+1} \mathbf{U}_j^\#]$ and verify (285). We first determine $\text{Cov}(\mathbf{Z}_r, \mathbf{Z}_j)$ from the limiting orthogonality $\mathbb{E}[\mathbf{U}_r^\# \mathbf{U}_j^\#] = 0$:

$$\begin{aligned} 0 &= \mathbb{E}[\mathbf{U}_r^\# \mathbf{U}_j^\#] \stackrel{(a)}{=} [\mathbf{s}_{u,*}^R]_r [\mathbf{s}_{u,*}^R]_j \gamma_r \gamma_j \mathbb{E}[\mathbf{U}_*^2] + \tilde{\gamma}_r \tilde{\gamma}_j \mathbb{E}[\mathbf{Z}_r \mathbf{Z}_j] \\ &\stackrel{(b)}{=} [\mathbf{s}_{u,*}^R]_r [\mathbf{s}_{u,*}^R]_j \gamma_r \gamma_j + \tilde{\gamma}_r \tilde{\gamma}_j \mathbb{E}[\mathbf{Z}_r \mathbf{Z}_j], \end{aligned} \quad (287)$$

hence

$$\mathbb{E}[\mathbf{Z}_r \mathbf{Z}_j] = -[\mathbf{s}_{u,*}^R]_r [\mathbf{s}_{u,*}^R]_j \frac{\gamma_r \gamma_j}{\tilde{\gamma}_r \tilde{\gamma}_j}. \quad (288)$$

Here (a) substitutes (283) and uses $(\mathbf{Z}_r, \mathbf{Z}_j) \perp \mathbf{U}_*$; (b) uses $\mathbb{E}[\mathbf{U}_*^2] = 1$.

Next, expand

$$\mathbb{E}[(\mathbf{U}_r^\#)^{k+1} \mathbf{U}_j^\#] = [\mathbf{s}_{u,*}^R]_j \gamma_j \mathbb{E}[(\mathbf{U}_r^\#)^{k+1} \mathbf{U}_*] + \tilde{\gamma}_j \mathbb{E}[(\mathbf{U}_r^\#)^{k+1} \mathbf{Z}_j] \stackrel{\text{def}}{=} \text{(I)} + \text{(II)}. \quad (289)$$

Term (I). Using the binomial expansion and $\mathbf{Z}_r \perp \mathbf{U}_*$,

$$\begin{aligned} \text{(I)} &= [\mathbf{s}_{u,*}^R]_j \gamma_j \mathbb{E}\left([\mathbf{s}_{u,*}^R]_r \gamma_r \mathbf{U}_* + \tilde{\gamma}_r \mathbf{Z}_r\right)^{k+1} \mathbf{U}_* \\ &\stackrel{(a)}{=} [\mathbf{s}_{u,*}^R]_j \gamma_j \sum_{m=0}^{k+1} \binom{k+1}{m} ([\mathbf{s}_{u,*}^R]_r \gamma_r)^m \tilde{\gamma}_r^{k+1-m} \mathbb{E}[\mathbf{U}_*^{m+1}] \mathbb{E}[\mathbf{Z}_r^{k+1-m}] \\ &\stackrel{(b)}{=} [\mathbf{s}_{u,*}^R]_r [\mathbf{s}_{u,*}^R]_j \gamma_j \sum_{p=0}^{k/2} \binom{k+1}{2p} \gamma_r^{k+1-2p} \tilde{\gamma}_r^{2p} (2p-1)!! \mathbb{E}[\mathbf{U}_*^{k+2-2p}], \end{aligned} \quad (290)$$

where (b) retains only even Gaussian moments $\mathbb{E}[\mathbf{Z}_r^{2p}] = (2p-1)!!$ and uses $[\mathbf{s}_{u,*}^R]_r^{k+1-2p} = [\mathbf{s}_{u,*}^R]_r$ for even k .

Term (II). Let $h_u(z) \stackrel{\text{def}}{=} ([\mathbf{s}_{u,*}^R]_r \gamma_r u + \tilde{\gamma}_r z)^{k+1}$ so that $(\mathbf{U}_r^\#)^{k+1} = h_{\mathbf{U}_*}(Z_r)$. Conditioning on $\mathbf{U}_*$ and applying Stein's identity for the Gaussian pair (Z_r, Z_j) yields

$$\begin{aligned}
(\text{II}) &= \tilde{\gamma}_j \mathbb{E}[h_{\mathbf{U}_*}(Z_r) Z_j] \\
&\stackrel{(a)}{=} \tilde{\gamma}_j \text{Cov}(Z_r, Z_j) \mathbb{E}\left[\frac{\partial}{\partial Z_r} (\mathbf{U}_r^\#)^{k+1}\right] \\
&\stackrel{(b)}{=} \tilde{\gamma}_j \text{Cov}(Z_r, Z_j) (k+1) \tilde{\gamma}_r \mathbb{E}[(\mathbf{U}_r^\#)^k] \\
&\stackrel{(c)}{=} -[\mathbf{s}_{u,*}^R]_r [\mathbf{s}_{u,*}^R]_j \gamma_r \gamma_j (k+1) \mathbb{E}[(\mathbf{U}_r^\#)^k] \\
&\stackrel{(d)}{=} -[\mathbf{s}_{u,*}^R]_r [\mathbf{s}_{u,*}^R]_j \gamma_r \gamma_j (k+1) \sum_{p=0}^{k/2} \binom{k}{2p} \gamma_r^{k-2p} \tilde{\gamma}_r^{2p} (2p-1)!! \mathbb{E}[\mathbf{U}_*^{k-2p}], \tag{291}
\end{aligned}$$

where (a) is Stein, (b) differentiates with respect to Z_r , (c) substitutes (288), and (d) expands $(\mathbf{U}_r^\#)^k$ and retains even Gaussian moments.

Finally, substituting (290)–(291) into (289) and using $(k+1)\binom{k}{2p} = (k+1-2p)\binom{k+1}{2p}$ gives

$$\mathbb{E}[(\mathbf{U}_r^\#)^{k+1} \mathbf{U}_j^\#] = [\mathbf{s}_{u,*}^R]_r [\mathbf{s}_{u,*}^R]_j \gamma_j \gamma_r^{k+1} \sum_{p=0}^{k/2} \binom{k+1}{2p} (2p-1)!! \gamma_r^{-2p} \tilde{\gamma}_r^{2p} \left(\mathbb{E}[\mathbf{U}_*^{k+2-2p}] - (k+1-2p) \mathbb{E}[\mathbf{U}_*^{k-2p}] \right). \tag{292}$$

By the minimality of k in Proposition 8, $\mathbb{E}[\mathbf{U}_*^{2\ell}] = (2\ell-1)!!$ for $2\ell \leq k$ while $\mathbb{E}[\mathbf{U}_*^{k+2}] \neq (k+1)!!$, which forces every summand in (292) with $p \geq 1$ to vanish; hence

$$\mathbb{E}[(\mathbf{U}_r^\#)^{k+1} \mathbf{U}_j^\#] = [\mathbf{s}_{u,*}^R]_r [\mathbf{s}_{u,*}^R]_j \gamma_r^{k+1} \gamma_j (\mathbb{E}[\mathbf{U}_*^{k+2}] - (k+1)!!).$$

Using $\gamma_r^{k+1} \gamma_j = \nu_1(\{\lambda_r\})^{(k+1)/2} \sqrt{\nu_1(\{\lambda_j\})}$ yields (285), and hence (286) by (284). This completes the $\mathbf{u}$ -channel.

For the $\mathbf{v}$ -channel, define $\hat{s}_{v,j}^{\text{NGMC}} \stackrel{\text{def}}{=} \hat{s}_{u,j}^{\text{NGMC}} \text{sign}(\nu_3(\{\sigma_r\}) \nu_3(\{\sigma_j\}))$; the inter-channel sign coupling then yields $\hat{s}_{v,j}^{\text{NGMC}} \xrightarrow{\text{a.s.}} [\mathbf{s}_{v,*}^R]_j$. $\square$

H Global Sign Detection

The spectral initializer (65) is correlated with the ground truth only up to an *unknown global sign* (cf. [28, 59]). Under the auxiliary randomization in (57), this sign is Rademacher distributed (cf. [61, Remark 3.6]). The purpose of this appendix is to estimate this global sign from the observed initializer (when it is identifiable) and to use the resulting estimate to select the signs (s_1, s_2) in the signed DMMSE family (67) used by (66).

Concretely, the normalized spectral initializers in (65) admit the scalar Gaussian-channel limits

$$\tilde{\mathbf{u}}_1 \xrightarrow{W} \tilde{\mathbf{U}}_1(s_{u,*}^G) \stackrel{d}{=} s_{u,*}^G \sqrt{w_{1,1}} \mathbf{U}_* + \sqrt{1-w_{1,1}} \mathbf{Z}_u, \quad w_{1,1} = \sum_{\lambda_i \in \mathcal{K}^*} \nu_1(\{\lambda_i\}), \tag{293}$$

$$\tilde{\mathbf{v}}_1 \xrightarrow{W} \tilde{\mathbf{V}}_1(s_{v,*}^G) \stackrel{d}{=} s_{v,*}^G \sqrt{w_{2,1}} \mathbf{V}_* + \sqrt{1-w_{2,1}} \mathbf{Z}_v, \quad w_{2,1} = \sum_{\lambda_i \in \mathcal{K}^*} \nu_2(\{\lambda_i\}). \tag{294}$$

where $s_{u,*}^G, s_{v,*}^G \in \{\pm 1\}$ are the realized global signs and all other variables are defined as in (5).

This representation determines when $s_{u,*}^G$ is statistically identifiable from $\tilde{\mathbf{u}}_1$. If $\mathbf{U}_* \stackrel{d}{=} -\mathbf{U}_*$, then the marginal law of $\tilde{\mathbf{U}}_1$ is invariant under $s_{u,*}^G \mapsto -s_{u,*}^G$, and the global sign cannot be recovered from $\tilde{\mathbf{u}}_1$ alone. Conversely, under asymmetric priors the two induced laws are distinct, and $s_{u,*}^G$ can be estimated consistently from the empirical distribution of $\tilde{\mathbf{u}}_1$.

Under the asymmetric regime, two constructions of consistent estimators of the true global signs are recorded below (GSMLE and odd-moment contrast), which manifest global-sign counterparts of the relative-sign procedures in Proposition 8; their consistency proofs repeat the similar arguments and are omitted.

H.1 Global-Sign Maximum Likelihood Estimator (GSMLE) Scheme

We describe a two-hypothesis likelihood test induced by the scalar-channel limits (293)–(294). For $s \in \{\pm 1\}$, let p_s^u denote the density of

$$s\sqrt{w_{1,1}}\mathbf{U}_* + \sqrt{1-w_{1,1}}Z_u, \quad Z_u \sim \mathcal{N}(0,1), \quad Z_u \perp \mathbf{U}_*,$$

and define p_s^v analogously from (294). Given $\tilde{\mathbf{u}}_1 \in \mathbb{R}^M$ and $\tilde{\mathbf{v}}_1 \in \mathbb{R}^N$, set

$$\mathcal{L}_{u,M}(s) \stackrel{\text{def}}{=} \frac{1}{M} \sum_{i=1}^M \log p_s^u([\tilde{\mathbf{u}}_1]_i), \quad \mathcal{L}_{v,N}(s) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^N \log p_s^v([\tilde{\mathbf{v}}_1]_i), \quad (295)$$

and define the GSMLEs

$$\hat{s}_u^{\text{GSMLE}} \in \operatorname{argmax}_{s \in \{\pm 1\}} \mathcal{L}_{u,M}(s), \quad \hat{s}_v^{\text{GSMLE}} \in \operatorname{argmax}_{s \in \{\pm 1\}} \mathcal{L}_{v,N}(s). \quad (296)$$

As $M, N \rightarrow \infty$, the following regimes hold:

(i) *Both priors asymmetric.*

$$\hat{s}_u^{\text{GSMLE}} \xrightarrow{\text{a.s.}} s_{u,*}^G, \quad \hat{s}_v^{\text{GSMLE}} \xrightarrow{\text{a.s.}} s_{v,*}^G.$$

(ii) *Both priors symmetric.* Then $p_{+1}^u = p_{-1}^u$ and $p_{+1}^v = p_{-1}^v$. Hence neither channel-wise global sign is identifiable from its initializer. However, the relative global sign is identifiable: for any baseline outlier $r \in \mathcal{I}_M$ with limit $\lambda_r = \sigma_r^2$, Lemma 15 implies

$$s_{u,*}^G s_{v,*}^G \xrightarrow{\text{a.s.}} \operatorname{sign}(\nu_3(\{\sigma_r\})), \quad (297)$$

where ν_3 is the cross spectral measure in Lemma 2.

(iii) *Exactly one prior symmetric.* Suppose $\mathbf{U}_*$ is asymmetric, then

$$\hat{s}_u^{\text{GSMLE}} \xrightarrow{\text{a.s.}} s_{u,*}^G, \quad \hat{s}_v^{\text{GSMLE}} \stackrel{\text{def}}{=} \hat{s}_u^{\text{GSMLE}} \cdot \operatorname{sign}(\nu_3(\{\sigma_r\})) \xrightarrow{\text{a.s.}} s_{v,*}^G.$$

The converse case is analogous.

H.2 Global Sign Odd-moment Contrast Scheme (GOMC) Scheme

In analogy with the NGMC rule for relative-sign alignment in Proposition 8, we record a simpler method-of-moments estimator for the global sign in the $\mathbf{u}$ -channel.

Assumption 5 (Odd-moment asymmetry). There exists an odd integer $j \geq 1$ such that $\mathbb{E}[\mathbf{U}_*^j] \neq 0$.

Assume the setting of Proposition 6 and Assumption 5, and let j be the smallest odd integer such that $\mathbb{E}[\mathbf{U}_*^j] \neq 0$. Define

$$\hat{s}_u^{\text{GOMC}} \stackrel{\text{def}}{=} \operatorname{sign}\left(\frac{1}{M} \sum_{i=1}^M [\tilde{\mathbf{u}}_1]_i^j\right) \operatorname{sign}(\mathbb{E}[\mathbf{U}_*^j]). \quad (298)$$

Under the scalar-channel convergence (293), following similar steps in (284)–(292), it can be shown that

$$\frac{1}{M} \sum_{i=1}^M [\tilde{\mathbf{u}}_1]_i^j \xrightarrow{\text{a.s.}} s_{u,*}^G w_{1,1}^{j/2} \mathbb{E}[\mathbf{U}_*^j],$$

and hence $\hat{s}_u^{\text{GOMC}} \xrightarrow{\text{a.s.}} s_{u,*}^G$ as $M \rightarrow \infty$.

For the $\mathbf{v}$ -channel, fix the same baseline outlier index $r \in \mathcal{I}_M$ as in (297), and set

$$\hat{s}_v^{\text{GOMC}} \stackrel{\text{def}}{=} \hat{s}_u^{\text{GOMC}} \cdot \operatorname{sign}(\nu_3(\{\sigma_r\})). \quad (299)$$

By the inter-channel coupling in (297), $\hat{s}_v^{\text{GOMC}} \xrightarrow{\text{a.s.}} s_{v,*}^G$.

H.3 Selection of the Signed DMMSE Estimators

We select the denoiser signs (s_1, s_2) in (66) by combining a global-sign estimate from (293)–(294) with the sign behavior of the signed DMMSE family (67); see Fact 6.

- *At least one asymmetric prior.* Without loss of generality, assume $\mathbf{U}_*$ is asymmetric. Let $\hat{s}_u(\tilde{\mathbf{u}}_1)$ be any consistent estimator of $s_{u,*}^G$, e.g., $\hat{s}_u^{\text{GSMLE}}$ in (296) or $\hat{s}_u^{\text{GOMC}}$ in (298). If $\mathbf{V}_*$ is asymmetric, define $\hat{s}_v(\tilde{\mathbf{v}}_1)$ analogously; otherwise, set

$$\hat{s}_v \stackrel{\text{def}}{=} \hat{s}_u \cdot \text{sign}(\nu_3(\{\sigma_r\})), \quad (300)$$

with any fixed baseline outlier $r \in \mathcal{I}_M$ (cf. (297)). We then choose

$$s_1 \stackrel{\text{def}}{=} \hat{s}_u(\tilde{\mathbf{u}}_1), \quad s_2 \stackrel{\text{def}}{=} \hat{s}_v(\tilde{\mathbf{v}}_1). \quad (301)$$

With this choice, the signed DMMSE update is aligned with the realized global signs, and hence preserves a positive correlation with each signal; see Fact 6.

- *Both priors symmetric.* Neither channel-wise global sign is identifiable from its initializer, but the relative global sign is fixed by (297). We adopt the convention

$$s_1 \stackrel{\text{def}}{=} +1, \quad s_2 \stackrel{\text{def}}{=} \text{sign}(\nu_3(\{\sigma_r\})), \quad (302)$$

(cf. (297)) where $r \in \mathcal{I}_M$ is the reference outlier index (cf. (297)). This convention is consistent with the asymptotic inter-channel sign coupling:

$$\lim_{M \rightarrow \infty} \text{sign}(\langle \tilde{\mathbf{u}}_1, \mathbf{u}_* \rangle \langle \tilde{\mathbf{v}}_1, \mathbf{v}_* \rangle) \stackrel{\text{a.s.}}{=} \lim_{M \rightarrow \infty} \text{sign}(\langle \mathbf{u}_r^\#, \mathbf{u}_* \rangle \langle \mathbf{v}_r^\#, \mathbf{v}_* \rangle) \stackrel{\text{Lemma 15}}{=} \text{sign}(\nu_3(\{\sigma_r\})) \quad (303)$$

By Fact 6, the signed DMMSE is odd under symmetric priors, so any mismatch with the realized global signs induces only a *coherent* global flip of the iterates across both channels. Consequently, the SE recursion for the squared overlaps (and hence the cosine similarities) is invariant under this flip.

I Proof of Theorem 2

Convention and asymptotic equivalence. Throughout this proof we work under Assumption 1. We use the asymptotic vector equivalence notation introduced in Definition 2: for sequences $\mathbf{x} \in \mathbb{R}^d$ and $\mathbf{y} \in \mathbb{R}^d$,

$$\mathbf{x} \stackrel{d \rightarrow \infty}{\simeq} \mathbf{y} \quad \Longleftrightarrow \quad \frac{\|\mathbf{x} - \mathbf{y}\|^2}{d} \xrightarrow{\text{a.s.}} 0 \quad \text{as } d \rightarrow \infty.$$

When applying this notation below, d is the ambient dimension of the corresponding vectors (e.g., $d = M$ for $\mathbb{R}^M$ -valued iterates and $d = N$ for $\mathbb{R}^N$ -valued iterates).

Proof strategy. We recall the spiked model

$$\mathbf{Y} = \frac{\theta}{\sqrt{MN}} \mathbf{u}_* \mathbf{v}_*^\top + \mathbf{W} \in \mathbb{R}^{M \times N}, \quad (304)$$

with $\|\mathbf{u}_*\|^2/M \rightarrow 1$ and $\|\mathbf{v}_*\|^2/N \rightarrow 1$ almost surely. The proof has two components. First, we show that the (optimally) combined spectral initializer is asymptotically equivalent to a single **W**-driven OAMP step in the sense of Definition 7. Second, we invoke the spiked-to-noise reduction developed in Appendix C—in particular the construction of the **W**-driven auxiliary recursion around (160) and the induction establishing (164)—to transfer the state evolution from the auxiliary recursion to the original **Y**-driven recursion.

Spectral initializer as a $\mathbf{W}$ -driven OAMP step. Let $(\sigma_{k,M}, \mathbf{u}_k, \mathbf{v}_k)_{k \in \mathcal{I}_M}$ denote the outlier singular triplets of $\mathbf{Y}$ (with $\|\mathbf{u}_k\| = \|\mathbf{v}_k\| = 1$), where $\mathcal{I}_M$ is the (finite) outlier index set and $\sigma_{k,M}^2 \rightarrow \lambda_k \in \mathcal{K}_*$. We adopt the randomized orientation convention

$$\mathbf{u}_k^\# \stackrel{\text{def}}{=} \sqrt{M} \xi_k \mathbf{u}_k, \quad \mathbf{v}_k^\# \stackrel{\text{def}}{=} \sqrt{N} \xi_k \mathbf{v}_k, \quad k \in \mathcal{I}_M, \quad (305)$$

where $(\xi_k)_{k \in \mathcal{I}_M}$ are i.i.d. Rademacher, independent of all other randomness. Let $(s_k^u)_{k \in \mathcal{I}_M}$ and $(s_k^v)_{k \in \mathcal{I}_M}$ be consistent relative-sign estimators. Define the *normalized* combined spectral initializer by

$$\tilde{\mathbf{u}}_1 \stackrel{\text{def}}{=} \left(\sum_{k \in \mathcal{I}_M} \nu_1(\{\lambda_k\}) \right)^{-1/2} \sum_{k \in \mathcal{I}_M} s_k^u \sqrt{\nu_1(\{\lambda_k\})} \mathbf{u}_k^\#, \quad (306)$$

$$\tilde{\mathbf{v}}_1 \stackrel{\text{def}}{=} \left(\sum_{k \in \mathcal{I}_M} \nu_2(\{\lambda_k\}) \right)^{-1/2} \sum_{k \in \mathcal{I}_M} s_k^v \sqrt{\nu_2(\{\lambda_k\})} \mathbf{v}_k^\#. \quad (307)$$

Lemma 16 (Optimal spectral estimate as a $\mathbf{W}$ -driven OAMP step). *Assume Assumption 1 and that θ is super-critical so that $\mathcal{K}_* = \{\lambda \in \mathbb{R} \setminus \text{supp}(\mu) : \Gamma(\lambda) = 0\}$ is finite and nonempty. Then there exist deterministic (dimension-independent) functions $\Psi_1, \tilde{\Psi}_1, \Phi_1, \tilde{\Phi}_1 : \mathbb{R} \rightarrow \mathbb{R}$, whose restrictions to the (compact) support $\text{supp}(\mu)$ are continuous, such that*

$$\tilde{\mathbf{u}}_1 \stackrel{N}{\rightrightarrows} \Psi_1(\mathbf{W}\mathbf{W}^\top) \mathbf{u}_* + \tilde{\Psi}_1(\mathbf{W}\mathbf{W}^\top) \mathbf{W} \mathbf{v}_*, \quad (308)$$

$$\tilde{\mathbf{v}}_1 \stackrel{N}{\rightrightarrows} \Phi_1(\mathbf{W}^\top \mathbf{W}) \mathbf{v}_* + \tilde{\Phi}_1(\mathbf{W}^\top \mathbf{W}) \mathbf{W}^\top \mathbf{u}_*. \quad (309)$$

In particular, $(\tilde{\mathbf{u}}_1, \tilde{\mathbf{v}}_1)$ is (up to $\stackrel{N}{\rightrightarrows}$) a valid (one-shot) $\mathbf{W}$ -driven step in the sense of Definition 7.

Proof. Fix $k \in \mathcal{I}_M$ and define the normalized overlaps

$$\alpha_{k,M} \stackrel{\text{def}}{=} \frac{1}{\sqrt{M}} \langle \mathbf{u}_k, \mathbf{u}_* \rangle, \quad \beta_{k,M} \stackrel{\text{def}}{=} \frac{1}{\sqrt{N}} \langle \mathbf{v}_k, \mathbf{v}_* \rangle.$$

By Lemma 5, whenever $\sigma_{k,M}^2 \notin \text{sp}(\mathbf{W}\mathbf{W}^\top)$ and $\sigma_{k,M}^2 \notin \text{sp}(\mathbf{W}^\top \mathbf{W})$,

$$\mathbf{u}_k = \mathbf{R}_M(\sigma_{k,M}^2) \left(\theta \frac{\alpha_{k,M}}{\sqrt{N}} \mathbf{W} \mathbf{v}_* + \theta \frac{\sigma_{k,M} \beta_{k,M}}{\sqrt{M}} \mathbf{u}_* \right), \quad (310)$$

$$\mathbf{v}_k = \tilde{\mathbf{R}}_N(\sigma_{k,M}^2) \left(\theta \frac{\beta_{k,M}}{\sqrt{M}} \mathbf{W}^\top \mathbf{u}_* + \theta \frac{\sigma_{k,M} \alpha_{k,M}}{\sqrt{N}} \mathbf{v}_* \right), \quad (311)$$

where $\mathbf{R}_M(x) \stackrel{\text{def}}{=} (x\mathbf{I}_M - \mathbf{W}\mathbf{W}^\top)^{-1}$ and $\tilde{\mathbf{R}}_N(x) \stackrel{\text{def}}{=} (x\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1}$.

Uniform resolvent control and replacement of $\sigma_{k,M}^2$ by λ_k . Since $\lambda_k \in \mathbb{R} \setminus \text{supp}(\mu)$ and Assumption 1(d) enforces spectral containment of $\mathbf{W}\mathbf{W}^\top$ and $\mathbf{W}^\top \mathbf{W}$, there is an a.s. event on which, for all large M ,

$$\text{d}(\sigma_{k,M}^2, \text{sp}(\mathbf{W}\mathbf{W}^\top)) \wedge \text{d}(\sigma_{k,M}^2, \text{sp}(\mathbf{W}^\top \mathbf{W})) \geq c_k > 0,$$

and hence $\|\mathbf{R}_M(\sigma_{k,M}^2)\|_{\text{op}} \vee \|\tilde{\mathbf{R}}_N(\sigma_{k,M}^2)\|_{\text{op}} \leq c_k^{-1}$. Moreover, by the resolvent identity (cf. (253)),

$$\|\mathbf{R}_M(\sigma_{k,M}^2) - \mathbf{R}_M(\lambda_k)\|_{\text{op}} \leq \frac{|\sigma_{k,M}^2 - \lambda_k|}{c_k^2} \xrightarrow{\text{a.s.}} 0,$$

and similarly for $\tilde{\mathbf{R}}_N(\sigma_{k,M}^2)$. Since $\|\mathbf{u}_*\|^2/M$ and $\|\mathbf{v}_*\|^2/N$ stay bounded and $\|\mathbf{W}\|_{\text{op}} = O(1)$, replacing $\mathbf{R}_M(\sigma_{k,M}^2)$ by $\mathbf{R}_M(\lambda_k)$ (and similarly for $\tilde{\mathbf{R}}_N$) incurs negligible normalized MSE errors.

Replacement of scalar coefficients by deterministic limits. Proposition 1 yields $\alpha_{k,M}^2 \rightarrow \nu_1(\{\lambda_k\})$ and $\beta_{k,M}^2 \rightarrow \nu_2(\{\lambda_k\})$ almost surely, and $\sigma_{k,M} \rightarrow \sqrt{\lambda_k}$. Because $|\mathcal{I}_M| < \infty$, we may replace the finitely many scalar coefficients $\alpha_{k,M}, \beta_{k,M}, \sigma_{k,M}$ in (310)–(311) by their almost sure limits at the cost of an overall $o_{\text{a.s.}}(1)$ normalized MSE error in the resulting sums.

Handling the (relative) sign convention. By consistency of $(s_k^u)_{k \in \mathcal{I}_M}$ and $(s_k^v)_{k \in \mathcal{I}_M}$ and finiteness of $\mathcal{I}_M$, there exists an almost sure event on which, for all large M , the effective orientations in (306)–(307) agree across all outliers up to a single global sign. Equivalently, the only remaining ambiguity is the global orientation of $(\tilde{\mathbf{u}}_1, \tilde{\mathbf{v}}_1)$; this ambiguity is already accounted for in the signed denoiser convention and the definition of the state evolution variables used in Theorem 2 (see discussions in Section 5.1).

Putting these points together, substituting (310)–(311) into (306)–(307) shows that $\tilde{\mathbf{u}}_1$ and $\tilde{\mathbf{v}}_1$ admit representations of the form (308)–(309), with $\Psi_1(x)$, $\tilde{\Psi}_1(x)$, $\Phi_1(x)$, $\tilde{\Phi}_1(x)$ related to finite linear combinations of the resolvent kernels $x \mapsto \frac{1}{\lambda_k - x}$ and hence continuous on $\text{supp}(\mu)$ because $\text{dist}(\lambda_k, \text{sp}(\mu)) > 0$ for each outlier λ_k . This proves the claim. $\square$

Reduction to the $\mathbf{W}$ -driven auxiliary recursion. Let $(\tilde{\mathbf{u}}_t, \tilde{\mathbf{v}}_t)_{t \geq 1}$ denote the spectrally-initialized $\mathbf{Y}$ -driven OAMP iterates defined in Theorem 2, with initialization $(\tilde{\mathbf{u}}_1, \tilde{\mathbf{v}}_1)$ given by (306)–(307). Let $(\hat{\mathbf{u}}_t, \hat{\mathbf{v}}_t)_{t \geq 1}$ denote the $\mathbf{W}$ -driven auxiliary recursion constructed as (160) (in Appendix C), using the same iterate denoisers as the original recursion, and initialized by

$$(\hat{\mathbf{u}}_1, \hat{\mathbf{v}}_1) \stackrel{\text{def}}{=} (\tilde{\mathbf{u}}_1, \tilde{\mathbf{v}}_1).$$

Lemma 17 (Spiked-to-noise reduction for spectrally-initialized OAMP). *For every fixed $T < \infty$,*

$$(\tilde{\mathbf{u}}_t, \tilde{\mathbf{v}}_t) \stackrel{d}{\rightharpoonup} (\hat{\mathbf{u}}_t, \hat{\mathbf{v}}_t), \quad \forall t \leq T. \quad (312)$$

Proof. The proof is an iteration-by-iteration comparison, and is identical in structure to the induction carried out in Appendix C leading to (164). For completeness, we isolate the only inputs used at each inductive step.

First, decompose the iterate denoiser output into its signal-aligned and orthogonal parts, as in (155) of Appendix C; the induction maintains the orthogonality relations required to apply Lemma 5 there (the “transfer lemma” controlling the difference between applying polynomial spectral denoisers to $\mathbf{Y}\mathbf{Y}^\top$ versus $\mathbf{W}\mathbf{W}^\top$, and similarly on the N -side). Concretely, Lemma 5 is invoked exactly as in the derivation around (160): it converts each matrix-denoiser action on $\mathbf{Y}\mathbf{Y}^\top$ (resp. $\mathbf{Y}^\top\mathbf{Y}$) applied to an orthogonal residual into the corresponding $\mathbf{W}$ -counterpart, plus explicit $\mathbf{u}_*$ or $\mathbf{v}_*$ aligned forcing terms; the remainder is $o_{\text{a.s.}}(1)$ in normalized MSE. The error control uses only (i) the uniform operator-norm bounds on the involved matrices (cf. (169) in Appendix C), (ii) the regularity of the matrix and iterate denoisers and (iii) the propagated orthogonality relations (cf. (163) there).

The base case $t = 1$ is valid by construction (since $\hat{\mathbf{u}}_1 = \tilde{\mathbf{u}}_1$ and $\hat{\mathbf{v}}_1 = \tilde{\mathbf{v}}_1$), and Lemma 16 ensures that this common initialization already has the $\mathbf{W}$ -driven OAMP form (with initialization independent of $\mathbf{W}$) required by the transfer lemma and the auxiliary construction. The inductive step then proceeds verbatim as in Appendix C, yielding (312) for any fixed T . $\square$

State evolution and conclusion. By construction, the auxiliary recursion (160) is a $\mathbf{W}$ -driven OAMP algorithm in the sense of Definition 7, with initialization satisfying the admissibility requirements thanks to Lemma 16 (and with the usual global-sign convention handled exactly as in Theorem 2). Therefore, the general OAMP state evolution theorem (Theorem 4 and its specialization used in Appendix C) applies to $(\hat{\mathbf{u}}_t, \hat{\mathbf{v}}_t)$ and yields the state evolution limits claimed in Theorem 2 for the auxiliary iterates.

Finally, Lemma 17 transfers these limits from the auxiliary recursion back to the original $\mathbf{Y}$ -driven recursion. This completes the proof of Theorem 2. $\square$

J Some Miscellaneous Results

Fact 1 (Non-tangential limit as Point Mass). The following result, a variant of [72, Proposition 8], shows that the point mass of a finite measure ν on $\mathbb{R}$ can be recovered from its Stieltjes transform $\mathcal{S}(z)$. For any $a \in \mathbb{R}$:

$$\lim_{z \rightarrow a} (z - a) \mathcal{S}(z) = \nu(\{a\}),$$

where the limit $z \rightarrow a$ is non-tangential, meaning it is restricted to any cone of the form $\{x + iy \in \mathbb{C}^+ \mid y > 0 \text{ and } |x - a| < \gamma y\}$ for any $\gamma > 0$.

Fact 2 (Location of Empirical Outliers). Any positive empirical eigenvalues of $\mathbf{Y}\mathbf{Y}^\top$ which are not eigenvalues of $\mathbf{W}\mathbf{W}^\top$ are solution to the following equation for $z \in \mathbb{R}$

$$\Gamma_M(z) = (1 - \frac{\theta}{\sqrt{MN}} \mathbf{v}_*^\top \mathbf{W}^\top \mathbf{S}_1(z) \mathbf{u}_*) (1 - \frac{\theta}{\sqrt{MN}} \mathbf{u}_*^\top \mathbf{W} \mathbf{S}_2(z) \mathbf{v}_*) - z (\frac{\theta}{M} \mathbf{u}_*^\top \mathbf{S}_1(z) \mathbf{u}_*) (\frac{\theta}{N} \mathbf{v}_*^\top \mathbf{S}_2(z) \mathbf{v}_*) = 0, \quad (313)$$

where $\mathbf{S}_1(z) \stackrel{\text{def}}{=} (z\mathbf{I}_M - \mathbf{W}\mathbf{W}^\top)^{-1}$, $\mathbf{S}_2(z) \stackrel{\text{def}}{=} (z\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1}$.

Proof of Fact 2. A positive number z is an eigenvalue of $\mathbf{Y}\mathbf{Y}^\top$ if and only if its square root, $\sigma = \sqrt{z}$, is a singular value of $\mathbf{Y}$. This is equivalent to the augmented matrix $\begin{pmatrix} \mathbf{0} & \mathbf{Y} \\ \mathbf{Y}^\top & \mathbf{0} \end{pmatrix}$ having σ as an eigenvalue. The characteristic equation for this condition is

$$\det \begin{pmatrix} \sigma \mathbf{I}_M & -\mathbf{Y} \\ -\mathbf{Y}^\top & \sigma \mathbf{I}_N \end{pmatrix} = 0. \quad (314)$$

Substituting the spiked model for $\mathbf{Y}$ from (1), we get

$$\det \left(\begin{pmatrix} \sigma \mathbf{I}_M & -\mathbf{W} \\ -\mathbf{W}^\top & \sigma \mathbf{I}_N \end{pmatrix} - \begin{pmatrix} \mathbf{0} & \frac{\theta}{\sqrt{MN}} \mathbf{u}_* \mathbf{v}_*^\top \\ \frac{\theta}{\sqrt{MN}} \mathbf{v}_* \mathbf{u}_*^\top & \mathbf{0} \end{pmatrix} \right) = 0. \quad (315)$$

Let us define the unperturbed matrix $\widehat{\mathbf{W}} \stackrel{\text{def}}{=} \begin{pmatrix} \sigma \mathbf{I}_M & -\mathbf{W} \\ -\mathbf{W}^\top & \sigma \mathbf{I}_N \end{pmatrix}$. The perturbation is a rank-2 matrix which can be factored as $\widehat{\mathbf{U}} \widehat{\boldsymbol{\Theta}} \widehat{\mathbf{U}}^\top$, where

$$\widehat{\mathbf{U}} \stackrel{\text{def}}{=} \begin{pmatrix} \mathbf{u}_* & \mathbf{0} \\ \mathbf{0} & \mathbf{v}_* \end{pmatrix}, \quad \text{and} \quad \widehat{\boldsymbol{\Theta}} \stackrel{\text{def}}{=} \frac{\theta}{\sqrt{MN}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (316)$$

The characteristic equation is now $\det(\widehat{\mathbf{W}} - \widehat{\mathbf{U}} \widehat{\boldsymbol{\Theta}} \widehat{\mathbf{U}}^\top) = 0$. By the Weinstein-Aronszajn formula, this is equivalent to

$$\det(\widehat{\mathbf{W}} - \widehat{\mathbf{U}} \widehat{\boldsymbol{\Theta}} \widehat{\mathbf{U}}^\top) = \det(\widehat{\mathbf{W}}) \det(\mathbf{I}_2 - \widehat{\boldsymbol{\Theta}} \widehat{\mathbf{U}}^\top \widehat{\mathbf{W}}^{-1} \widehat{\mathbf{U}}) = 0. \quad (317)$$

From Assumption 1(c), the Lemma's premise is that z is not an eigenvalue of $\mathbf{W}\mathbf{W}^\top$. This ensures that $\widehat{\mathbf{W}}$ is invertible, so $\det(\widehat{\mathbf{W}}) \neq 0$. The condition thus simplifies to the singularity of the 2×2 matrix

$$\det(\mathbf{I}_2 - \widehat{\boldsymbol{\Theta}} \widehat{\mathbf{U}}^\top \widehat{\mathbf{W}}^{-1} \widehat{\mathbf{U}}) = 0. \quad (318)$$

The inverse of $\widehat{\mathbf{W}}$ is given by the Shur complement formula

$$\widehat{\mathbf{W}}^{-1} = \begin{pmatrix} \sigma(z\mathbf{I}_M - \mathbf{W}\mathbf{W}^\top)^{-1} & \mathbf{W}(z\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1} \\ \mathbf{W}^\top(z\mathbf{I}_M - \mathbf{W}\mathbf{W}^\top)^{-1} & \sigma(z\mathbf{I}_N - \mathbf{W}^\top \mathbf{W})^{-1} \end{pmatrix} \stackrel{\text{def}}{=} \begin{pmatrix} \sigma \mathbf{S}_1(z) & \mathbf{W} \mathbf{S}_2(z) \\ \mathbf{W}^\top \mathbf{S}_1(z) & \sigma \mathbf{S}_2(z) \end{pmatrix}, \quad (319)$$

where we have used $z = \sigma^2$ and the definitions of $\mathbf{S}_1(z)$ and $\mathbf{S}_2(z)$. The 2×2 core matrix is then

$$\widehat{\mathbf{U}}^\top \widehat{\mathbf{W}}^{-1} \widehat{\mathbf{U}} = \begin{pmatrix} \mathbf{u}_*^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{v}_*^\top \end{pmatrix} \begin{pmatrix} \sigma \mathbf{S}_1(z) & \mathbf{W} \mathbf{S}_2(z) \\ \mathbf{W}^\top \mathbf{S}_1(z) & \sigma \mathbf{S}_2(z) \end{pmatrix} \begin{pmatrix} \mathbf{u}_* & \mathbf{0} \\ \mathbf{0} & \mathbf{v}_* \end{pmatrix} = \begin{pmatrix} \sigma \mathbf{u}_*^\top \mathbf{S}_1(z) \mathbf{u}_* & \mathbf{u}_*^\top \mathbf{W} \mathbf{S}_2(z) \mathbf{v}_* \\ \mathbf{v}_*^\top \mathbf{W}^\top \mathbf{S}_1(z) \mathbf{u}_* & \sigma \mathbf{v}_*^\top \mathbf{S}_2(z) \mathbf{v}_* \end{pmatrix}. \quad (320a)$$

Substituting this into the determinant condition gives

$$\det \left(\mathbf{I}_2 - \frac{\theta}{\sqrt{MN}} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sigma \mathbf{u}_*^\top \mathbf{S}_1(z) \mathbf{u}_* & \mathbf{u}_*^\top \mathbf{W} \mathbf{S}_2(z) \mathbf{v}_* \\ \mathbf{v}_*^\top \mathbf{W}^\top \mathbf{S}_1(z) \mathbf{u}_* & \sigma \mathbf{v}_*^\top \mathbf{S}_2(z) \mathbf{v}_* \end{pmatrix} \right) = 0. \quad (321)$$

This expands to the determinant of the following 2×2 matrix

$$\det \begin{pmatrix} 1 - \frac{\theta}{\sqrt{MN}} \mathbf{v}_*^\top \mathbf{W}^\top \mathbf{S}_1(z) \mathbf{u}_* & -\frac{z\theta}{\sqrt{MN}} \mathbf{v}_*^\top \mathbf{S}_2(z) \mathbf{v}_* \\ -\frac{z\theta}{\sqrt{MN}} \mathbf{u}_*^\top \mathbf{S}_1(z) \mathbf{u}_* & 1 - \frac{\theta}{\sqrt{MN}} \mathbf{u}_*^\top \mathbf{W} \mathbf{S}_2(z) \mathbf{v}_* \end{pmatrix} = 0. \quad (322)$$

Evaluating the determinant and recalling that $\sigma^2 = z$ yields (313). $\square$

Fact 3 (Special Form of the Sokhotski–Plemelj Formula). Let ν be a finite signed measure on $\mathbb{R}$, with its Stieltjes transform $\mathcal{S}_\nu$ and Hilbert transform $\mathcal{H}_\nu$ given by (3) and (4), respectively. For any non-zero point $x \in \mathbb{R} \setminus \{0\}$ where the measure ν possesses a density at x^2 , denoted by $\nu(x^2)$, the following boundary limit for the Stieltjes transform holds

$$\lim_{\epsilon \rightarrow 0^+} \mathcal{S}_\nu((x - i\epsilon)^2) = \pi \mathcal{H}_\nu(x^2) + i\pi \operatorname{sign}(x) \nu(x^2). \quad (323)$$

Proof. Fix $x \in \mathbb{R} \setminus \{0\}$ such that the density $\frac{d\nu}{d\lambda}(x^2)$ exists, and denote this value by $\nu(x^2)$. Since ν is a finite signed measure, there exist finite positive measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$ by Jordan decomposition (cf. [62, Theorem 4.1.5]). By linearity of the Stieltjes transform, the Hilbert transform and the Radon–Nikodym derivative, this implies

$$\mathcal{S}_\nu = \mathcal{S}_{\nu^+} - \mathcal{S}_{\nu^-}, \quad \mathcal{H}_\nu = \mathcal{H}_{\nu^+} - \mathcal{H}_{\nu^-}, \quad \frac{d\nu}{d\lambda} = \frac{d\nu^+}{d\lambda} - \frac{d\nu^-}{d\lambda}. \quad (324)$$

For a finite positive measure χ and for Lebesgue-almost every $t \in \mathbb{R}$ such that $\frac{d\chi}{d\lambda}(t)$ exists, the Sokhotski–Plemelj boundary value theorem (cf. [50, Section 3.1]) yields

$$\lim_{\epsilon \downarrow 0} \mathcal{S}_\chi(t - i\epsilon) = \pi \mathcal{H}_\chi(t) + i\pi \frac{d\chi}{d\lambda}(t), \quad (325)$$

$$\lim_{\epsilon \downarrow 0} \mathcal{S}_\chi(t + i\epsilon) = \pi \mathcal{H}_\chi(t) - i\pi \frac{d\chi}{d\lambda}(t). \quad (326)$$

Applying (325)–(326) to ν^+ and ν^- at $t = x^2$, and using the relations (324), we obtain

$$\lim_{\epsilon \downarrow 0} \mathcal{S}_\nu(x^2 - i\epsilon) = \pi \mathcal{H}_\nu(x^2) + i\pi \nu(x^2), \quad (327)$$

$$\lim_{\epsilon \downarrow 0} \mathcal{S}_\nu(x^2 + i\epsilon) = \pi \mathcal{H}_\nu(x^2) - i\pi \nu(x^2). \quad (328)$$

Now consider the path

$$w(\epsilon) = (x - i\epsilon)^2 = (x^2 - \epsilon^2) - i2x\epsilon, \quad \epsilon > 0. \quad (329)$$

Then $w(\epsilon) \rightarrow x^2$ as $\epsilon \downarrow 0$. Moreover,

$$\Im w(\epsilon) = -2x\epsilon,$$

so for $x > 0$ the points $w(\epsilon)$ lie in the lower half-plane, while for $x < 0$ they lie in the upper half-plane.

Case $x > 0$. Here $\Im w(\epsilon) < 0$ for all $\epsilon > 0$, and $w(\epsilon) \rightarrow x^2$ from the lower half-plane. Since $\mathcal{S}_\nu$ is analytic on $\mathbb{C} \setminus \mathbb{R}$, the boundary value in (327) is independent of the particular approach within the lower half-plane. Hence

$$\lim_{\epsilon \downarrow 0} \mathcal{S}_\nu(w(\epsilon)) = \lim_{\epsilon \downarrow 0} \mathcal{S}_\nu(x^2 - i\epsilon) = \pi \mathcal{H}_\nu(x^2) + i\pi \nu(x^2). \quad (330)$$

Case $x < 0$. Now $\Im w(\epsilon) > 0$ for all $\epsilon > 0$, and $w(\epsilon) \rightarrow x^2$ from the upper half-plane. Using (328) and the same reasoning,

$$\lim_{\epsilon \downarrow 0} \mathcal{S}_\nu(w(\epsilon)) = \lim_{\epsilon \downarrow 0} \mathcal{S}_\nu(x^2 + i\epsilon) = \pi \mathcal{H}_\nu(x^2) - i\pi \nu(x^2). \quad (331)$$

Combining (330) and (331), and recalling $w(\epsilon) = (x - i\epsilon)^2$, we obtain for all $x \neq 0$ with density at x^2 :

$$\lim_{\epsilon \downarrow 0} \mathcal{S}_\nu((x - i\epsilon)^2) = \pi \mathcal{H}_\nu(x^2) + i\pi \operatorname{sign}(x) \nu(x^2),$$

which is exactly (323). □

Fact 4 (Boundary values for the Stieltjes transform of a signed measure). Let χ be a finite signed Borel measure on $\mathbb{R}$, and let

$$\mathcal{S}_\chi(z) \stackrel{\text{def}}{=} \int_{\mathbb{R}} \frac{1}{z-x} d\chi(x), \quad z \in \mathbb{C} \setminus \mathbb{R},$$

denote its Stieltjes transform. Let

$$\chi = \chi_{\parallel} + \chi_{\perp}$$

be the Lebesgue decomposition of χ with respect to Lebesgue measure λ , where $\chi_{\parallel} \ll \lambda$ and $\chi_{\perp} \perp \lambda$. Then the boundary values of $\mathcal{S}_\chi$ satisfy

$$\lim_{\epsilon \downarrow 0} \Im \mathcal{S}_\chi(x - i\epsilon) = \pi \frac{d\chi_{\parallel}}{d\lambda}(x), \quad \text{for Lebesgue-a.e. } x \in \mathbb{R}, \quad (332)$$

$$\lim_{\epsilon \downarrow 0} |\Im \mathcal{S}_\chi(x - i\epsilon)| = \infty, \quad \text{for } |\chi_{\perp}| \text{-almost every } x \in \mathbb{R}. \quad (333)$$

Here $|\chi_{\perp}|$ denotes the total variation measure of the signed measure $\chi_{\perp}$. If $\chi_{\perp} = \chi_{\perp}^+ - \chi_{\perp}^-$ is the Jordan decomposition of $\chi_{\perp}$ into mutually singular finite positive measures, then

$$|\chi_{\perp}| = \chi_{\perp}^+ + \chi_{\perp}^-. \quad (334)$$

In particular, (333) is equivalent to the existence of a Borel set $N \subset \mathbb{R}$ such that

$$|\chi_{\perp}|(N) = 0 \quad \text{and} \quad \lim_{\epsilon \downarrow 0} |\Im \mathcal{S}_\chi(x - i\epsilon)| = \infty \quad \text{for all } x \in \mathbb{R} \setminus N.$$

Proof. Let $\chi = \chi^+ - \chi^-$ be the Jordan decomposition, where $\chi^{\pm}$ are finite positive Borel measures with $\chi^+ \perp \chi^-$. For each $\chi^{\pm}$, let

$$\chi^{\pm} = \chi_{\parallel}^{\pm} + \chi_{\perp}^{\pm}$$

be the Lebesgue decomposition with respect to λ . By uniqueness of the Lebesgue decomposition (see, e.g., [62, Thm. 4.3.2]),

$$\chi_{\parallel} = \chi_{\parallel}^+ - \chi_{\parallel}^-, \quad \chi_{\perp} = \chi_{\perp}^+ - \chi_{\perp}^-.$$

We first establish (332). For each finite positive measure $\chi^{\pm}$, the boundary-value formula for its Stieltjes transform (see [50, Section 3.1], with the convention $x - i\epsilon$) yields

$$\lim_{\epsilon \downarrow 0} \Im \mathcal{S}_{\chi^{\pm}}(x - i\epsilon) = \pi \frac{d\chi_{\parallel}^{\pm}}{d\lambda}(x), \quad \text{for Lebesgue-a.e. } x \in \mathbb{R}. \quad (335)$$

Since $\mathcal{S}_\chi = \mathcal{S}_{\chi^+} - \mathcal{S}_{\chi^-}$ and both the Stieltjes transform and the Radon–Nikodým derivative are linear in the measure, (335) implies, for Lebesgue-a.e. $x \in \mathbb{R}$,

$$\lim_{\epsilon \downarrow 0} \Im \mathcal{S}_\chi(x - i\epsilon) = \pi \left(\frac{d\chi_{\parallel}^+}{d\lambda}(x) - \frac{d\chi_{\parallel}^-}{d\lambda}(x) \right) = \pi \frac{d\chi_{\parallel}}{d\lambda}(x),$$

which is (332).

We now prove (333). Let the Jordan decomposition of the singular part be

$$\chi_{\perp} = \chi_{\perp}^+ - \chi_{\perp}^-,$$

where $\chi_{\perp}^{\pm}$ are finite positive measures with $\chi_{\perp}^+ \perp \chi_{\perp}^-$. Then $|\chi_{\perp}| = \chi_{\perp}^+ + \chi_{\perp}^-$ by (334).

For any finite signed Borel measure η on $\mathbb{R}$ and $\epsilon > 0$,

$$\Im \mathcal{S}_\eta(x - i\epsilon) = \epsilon \int_{\mathbb{R}} \frac{1}{(x-t)^2 + \epsilon^2} d\eta(t).$$

In particular, for $\eta \geq 0$ this is (up to the factor π) the Poisson integral of η . Applying [50, Section 3.1] to the purely singular finite positive measures $\chi_{\perp}^{\pm}$ yields

$$\Im \mathcal{S}_{\chi_{\perp}^+}(x - i\epsilon) \xrightarrow{\epsilon \downarrow 0} +\infty \quad \text{for } \chi_{\perp}^+ \text{-a.e. } x \in \mathbb{R}, \quad \Im \mathcal{S}_{\chi_{\perp}^-}(x - i\epsilon) \xrightarrow{\epsilon \downarrow 0} +\infty \quad \text{for } \chi_{\perp}^- \text{-a.e. } x \in \mathbb{R}.$$

Define

$$u^+(x, \epsilon) \stackrel{\text{def}}{=} \pi^{-1} \Im \mathcal{S}_{\chi_{\perp}^+}(x - i\epsilon), \quad u^-(x, \epsilon) \stackrel{\text{def}}{=} \pi^{-1} \Im \mathcal{S}_{\chi_{\perp}^-}(x - i\epsilon).$$

By [73, Lemma 1], applied to the mutually singular finite positive measures $\chi_{\perp}^+$ and $\chi_{\perp}^-$, as $\epsilon \downarrow 0$ we have

$$\begin{aligned} u^-(x, \epsilon) &= o(u^+(x, \epsilon)), & \text{for } \chi_{\perp}^+ \text{-a.e. } x \in \mathbb{R}, \\ u^+(x, \epsilon) &= o(u^-(x, \epsilon)), & \text{for } \chi_{\perp}^- \text{-a.e. } x \in \mathbb{R}. \end{aligned}$$

Since the Poisson kernel is nonnegative, for all $x \in \mathbb{R}$ and $\epsilon > 0$ we have the domination

$$|\Im \mathcal{S}_{\chi_{\parallel}}(x - i\epsilon)| \leq \Im \mathcal{S}_{|\chi_{\parallel}|}(x - i\epsilon). \quad (336)$$

Set

$$u^{\parallel}(x, \epsilon) \stackrel{\text{def}}{=} \pi^{-1} \Im \mathcal{S}_{|\chi_{\parallel}|}(x - i\epsilon) \geq 0.$$

Moreover, $|\chi_{\parallel}| \ll \lambda$ while $\chi_{\perp}^{\pm} \perp \lambda$, hence $|\chi_{\parallel}| \perp \chi_{\perp}^{\pm}$. Another application of [73, Lemma 1] therefore yields, as $\epsilon \downarrow 0$,

$$\begin{aligned} u^{\parallel}(x, \epsilon) &= o(u^+(x, \epsilon)), & \text{for } \chi_{\perp}^+ \text{-a.e. } x \in \mathbb{R}, \\ u^{\parallel}(x, \epsilon) &= o(u^-(x, \epsilon)), & \text{for } \chi_{\perp}^- \text{-a.e. } x \in \mathbb{R}. \end{aligned}$$

Combining (336) with the preceding little- o relations, we obtain

$$\begin{aligned} \Im \mathcal{S}_{\chi}(x - i\epsilon) &= \Im \mathcal{S}_{\chi_{\parallel}}(x - i\epsilon) + \pi u^+(x, \epsilon) - \pi u^-(x, \epsilon) \\ &\sim \pi u^+(x, \epsilon) \xrightarrow{\epsilon \downarrow 0} +\infty, & \text{for } \chi_{\perp}^+ \text{-a.e. } x \in \mathbb{R}, \end{aligned} \quad (337)$$

$$\Im \mathcal{S}_{\chi}(x - i\epsilon) \sim -\pi u^-(x, \epsilon) \xrightarrow{\epsilon \downarrow 0} -\infty, \quad \text{for } \chi_{\perp}^- \text{-a.e. } x \in \mathbb{R}. \quad (338)$$

In particular,

$$\lim_{\epsilon \downarrow 0} |\Im \mathcal{S}_{\chi}(x - i\epsilon)| = \infty \quad \text{for } \chi_{\perp}^+ \text{-a.e. and } \chi_{\perp}^- \text{-a.e. } x \in \mathbb{R}.$$

Since $|\chi_{\perp}| = \chi_{\perp}^+ + \chi_{\perp}^-$, the same property holds for $|\chi_{\perp}|$ -almost every $x \in \mathbb{R}$, which is (333). $\square$

Fact 5 (Resolvent representation of singular vectors). Let $(\sigma, \mathbf{u}, \mathbf{v})$ be a singular value–vector triplet of

$$\mathbf{Y} = \frac{\theta}{\sqrt{MN}} \mathbf{u}_* \mathbf{v}_*^{\top} + \mathbf{W} \in \mathbb{R}^{M \times N}.$$

Assume that $\sigma^2 \notin \text{spec}(\mathbf{W}\mathbf{W}^{\top}) \cup \{0\}$. Then,

$$\mathbf{u} = \mathbf{R}_M(\sigma^2) \left(\theta \frac{\langle \mathbf{u}, \mathbf{u}_* \rangle}{\sqrt{MN}} \mathbf{W} \mathbf{v}_* + \theta \sigma \frac{\langle \mathbf{v}, \mathbf{v}_* \rangle}{\sqrt{MN}} \mathbf{u}_* \right), \quad (339)$$

$$\mathbf{v} = \tilde{\mathbf{R}}_N(\sigma^2) \left(\theta \frac{\langle \mathbf{v}, \mathbf{v}_* \rangle}{\sqrt{MN}} \mathbf{W}^{\top} \mathbf{u}_* + \theta \sigma \frac{\langle \mathbf{u}, \mathbf{u}_* \rangle}{\sqrt{MN}} \mathbf{v}_* \right), \quad (340)$$

where

$$\mathbf{R}_M(x) \stackrel{\text{def}}{=} (x \mathbf{I}_M - \mathbf{W}\mathbf{W}^{\top})^{-1}, \quad \tilde{\mathbf{R}}_N(x) \stackrel{\text{def}}{=} (x \mathbf{I}_N - \mathbf{W}^{\top} \mathbf{W})^{-1}.$$

Proof of Fact 5. This follows directly from the singular equation and we omit the details. $\square$

Fact 6 (Properties of the signed DMMSE). Let

$$\mathbf{X}^{\text{sgn}} \stackrel{\text{def}}{=} s_* \sqrt{w} \mathbf{X}_* + \sqrt{1-w} \mathbf{Z}, \quad (341)$$

$$\mathbf{X}^{\text{std}} \stackrel{\text{def}}{=} s_* \mathbf{X}^{\text{sgn}} \stackrel{d}{=} \sqrt{w} \mathbf{X}_* + \sqrt{1-w} \mathbf{Z}, \quad (342)$$

be scalar Gaussian channels with $w \in [0, 1)$, $s_* \in \{\pm 1\}$, $\mathbb{E}[\mathbf{X}_*^2] = 1$, and $\mathbf{Z} \sim \mathcal{N}(0, 1) \perp\!\!\!\perp \mathbf{X}_*$. Under Assumption 2, let $\bar{\phi}(\cdot | w)$ be the DMMSE estimator associated with (342) defined in (33), and define the signed DMMSE estimator associated with (341) by

$$\bar{\phi}(x | w, s) \stackrel{\text{def}}{=} \bar{\phi}(sx | w), \quad s \in \{\pm 1\}. \quad (343)$$

Then:

(i) (Matched Sign.) If $s = s_*$, then

$$\mathbb{E}[\mathbf{X}_* \bar{\phi}(\mathbf{X}^{\text{sgn}} | w, s_*)] \geq 0, \quad (344)$$

with strict positivity for non-degenerate $\mathbf{X}_*$ and $w \in (0, 1)$.

(ii) (Possible Mismatched Sign.) If $\mathbf{X}_* \stackrel{d}{=} -\mathbf{X}_*$, then $\bar{\phi}(\cdot | w)$ is odd and, for any $s \in \{\pm 1\}$,

$$\mathbb{E}[\mathbf{X}_* \bar{\phi}(\mathbf{X}^{\text{sgn}} | w, s)] = ss_* \mathbb{E}[\bar{\phi}(\mathbf{X}^{\text{std}} | w)^2]. \quad (345)$$

Hence the correlation flips sign with ss_* while its magnitude is unchanged.

Proof of Fact 6. We use the DMMSE projection identity (see [44, Appendix A.2]): for the standard channel (342),

$$\mathbb{E}[\mathbf{X}_* \bar{\phi}(\mathbf{X}^{\text{std}} | w)] = \mathbb{E}[\bar{\phi}(\mathbf{X}^{\text{std}} | w)^2]. \quad (346)$$

Proof of (i). If $s = s_*$, then $\bar{\phi}(\mathbf{X}^{\text{sgn}} | w, s_*) = \bar{\phi}(s_* \mathbf{X}^{\text{sgn}} | w) = \bar{\phi}(\mathbf{X}^{\text{std}} | w)$, and thus

$$\mathbb{E}[\mathbf{X}_* \bar{\phi}(\mathbf{X}^{\text{sgn}} | w, s_*)] = \mathbb{E}[\bar{\phi}(\mathbf{X}^{\text{std}} | w)^2] \geq 0,$$

where the equality follows from (346).

Proof of (ii). If $\mathbf{X}_* \stackrel{d}{=} -\mathbf{X}_*$, then $\bar{\phi}(\cdot | w)$ is odd by symmetry of (342). Moreover, $\bar{\phi}(\cdot | w)$ is odd as well, since by (33) it is an affine combination of $\phi(\cdot | w)$ and x with coefficients depending only on w . By (342) and (343), and using the oddness of $\bar{\phi}(\cdot | w)$, we have

$$\bar{\phi}(\mathbf{X}^{\text{sgn}} | w, s) = \bar{\phi}(ss_* \mathbf{X}^{\text{std}} | w) = ss_* \bar{\phi}(\mathbf{X}^{\text{std}} | w).$$

Therefore, applying (346),

$$\mathbb{E}[\mathbf{X}_* \bar{\phi}(\mathbf{X}^{\text{sgn}} | w, s)] = ss_* \mathbb{E}[\mathbf{X}_* \bar{\phi}(\mathbf{X}^{\text{std}} | w)] = ss_* \mathbb{E}[\bar{\phi}(\mathbf{X}^{\text{std}} | w)^2],$$

which is (345). □

Remark 12. Fact 6 is used to select s in the signed DMMSE (67) from a global-sign estimate (cf. Appendix H):

- *Asymmetric priors.* The ground truth s_* is consistently estimable. Choosing s accordingly yields a positively correlated signed DMMSE update, as in (344).
- *Symmetric priors.* The ground truth s_* is not identifiable from a single channel. By (345), a sign mismatch ($s = -s_*$) induces only a global flip of the output, while leaving its squared magnitude unchanged.