Optimality of Approximate Message Passing Algorithms for Spiked Matrix Models with Rotationally Invariant Noise

Rishabh Dudeja* Songbin Liu† Junjie Ma‡§

May 28, 2024

Abstract

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

Contents

*Department of Statistics, University of Wisconsin–Madison. Email: rdudeja@wisc.edu
†Academy of Mathematics and Systems Science, Chinese Academy of Sciences. Email: liusongbin@lsec.cc.ac.cn
‡Academy of Mathematics and Systems Science, Chinese Academy of Sciences. Email: majunjie@lsec.cc.ac.cn
§The authors are alphabetically ordered.

1 Introduction

We consider the problem of estimating a symmetric rank-one matrix from a noisy $N \times N$ observed matrix $\boldsymbol{Y}$ generated from the *spiked matrix model* [35]:

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

where $\boldsymbol{x}_\star \in \mathbb{R}^N$ is the N-dimensional unknown signal of interest, $\theta \geq 0$ is the signal-to-noise ratio (SNR) parameter, and $\boldsymbol{W}$ is a symmetric noise matrix. This model and its variants have been used to study a broad range of statistical inference problems, including sparse PCA [20, 36, 78], community detection [2, 21], and group synchronization [17, 30, 64, 73].

Wigner Noise Model. The most well-studied variant of the spiked model is the *Spiked Wigner Model* [6, 20, 27, 28, 54, 65], which assumes that the noise matrix has i.i.d. Gaussian entries. A rich line of work in high-dimensional statistics and random matrix theory has studied the problem from various perspectives.

- *Design and Analysis of Estimators.* A natural estimator for the signal is the PCA estimator or the leading eigenvector of the observed matrix. A line of work [4, 5, 12, 32, 39, 62, 63] initiated by Baik et al. [5] have obtained a sharp asymptotic characterization of the performance of this estimator in the high-dimensional limit and uncovered surprising properties of this estimator. One way to improve the performance of PCA is to exploit prior structural information about the signal, such as sparsity. Approximate Message-passing algorithms are a popular class of computationally efficient iterative algorithms designed to exploit such structural information. These algorithms were first discovered in the context of compressed sensing [9, 13, 23, 37] and have been adapted for the spiked Wigner model [38, 46, 47, 54, 60, 68]. A particularly attractive feature of these algorithms is that their performance in the high-dimensional limit is characterized by a simple deterministic recursion known as *state evolution*, which makes it possible to assess their information-theoretic optimality (or sub-optimality).

- *Information-theoretic Limits.* Under the assumption that the signal is drawn from a known prior distribution, the optimal estimator is the Bayes estimator $\mathbb{E}[\boldsymbol{x}_\star|\boldsymbol{Y}]$. Using the powerful but non-rigorous cavity method from statistical physics, Lesieur et al. [44] derived a conjecture for the Bayes risk (or the asymptotic performance of the Bayes estimator) for the Spiked Wigner model. This conjecture has been proved rigorously in increasing generality in a series of influential works [6, 7, 10, 20, 27, 40, 43, 52]. This formula for asymptotic Bayes risk provides a fundamental information-theoretic limit on the performance of any estimator for the problem.

- *Computational Limits.* In general, computing the Bayes-optimal estimator is computationally intractable in high dimensions. However, a suitably designed AMP algorithm called the Bayes-optimal AMP algorithm can attain the Bayes risk for the problem for sufficiently high signal-to-noise ratios. For lower signal-to-noise ratios, the Bayes-optimal AMP algorithm fails to do so, and no computationally efficient algorithm is known to achieve the Bayes risk [7, 40, 45, 52]. This has led to a popular conjecture that the Bayes-optimal AMP algorithm is the optimal polynomial time algorithm for this problem [45, 52]. Celentano et al. [16], Montanari and Wu [56], and Montanari and Wein [55] (see also the earlier work of Schramm and Wein [71]) have provided evidence for this conjecture by showing the Bayes-optimal AMP algorithm achieves the lowest possible estimation error among a broad class of iterative algorithms and low-degree polynomial estimators.

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

Rotationally Invariant Noise Model. In this paper, we study a natural generalization of the i.i.d. Gaussian noise model known as the rotationally invariant noise model where one posits that the eigenvectors of the noise matrix are uniformly random orthogonal matrices independent of the eigenvalues. The rotationally invariant noise model is intended to model noise matrices with strong statistical dependence whose eigenvectors are generic. Several works (*see e.g.,* [22, 48, 59]) have observed that this can be a good model in some applications. Benaych-Georges and Nadakuditi [12] analyzed the performance of PCA (or spectral estimators) for this noise model. A line of work [29, 53, 58, 76] initiated by Opper et al. [58] and Fan [29] has developed AMP algorithms for this problem, which can improve the performance of PCA by exploiting signal structure. However, our understanding of the fundamental information-theoretic and computational limits in this model is extremely limited. Recent work by Barbier et al. [8] studies the spiked matrix model with rotationally invariant noise under the assumption that the noise matrix is drawn from the *trace ensemble*. Under this model, the density of $\boldsymbol{W}$ is given by:

$$p(\boldsymbol{W}) \propto \exp\left(-\frac{N}{2}\sum_{i=1}^{N} V(\lambda_i(\boldsymbol{W}))\right) \quad \text{where } \lambda_1(\boldsymbol{W}), \ldots, \lambda_N(\boldsymbol{W}) \text{ denote the eigenvalues of } \boldsymbol{W}, \tag{2}$$

and the *potential function* $V : \mathbb{R} \mapsto \mathbb{R}$ is a functional parameter for the noise model. This assumption ensures that the problem has a well-defined likelihood, enabling the study of information-theoretic limits. Moreover, an appropriate choice of V can capture a wide range of noise eigenvalue spectrums in different applications. Barbier et al. [8] make progress towards understanding the information-theoretic limits of the problem and provide important insights regarding the structure of optimal computationally efficient algorithms.

- *Information-theoretic Limits.* Barbier et al. [8] develop a general (although non-rigorous) recipe based on the replica method to derive conjectured formulas for Bayes risk under the assumption that the potential V is a polynomial function, providing explicit conjectured formulas for the asymptotic Bayes risk when V is a quartic (degree four) or a sestic (degree six) polynomial. As the degree of V grows, the derivation and the resulting formulas become increasingly complex, and a general conjectured formula for the Bayes risk is still unavailable.

- *Optimal Computationally Efficient Algorithms.* Surprisingly, Barbier et al. [8] demonstrate that a natural generalization of the Bayes-optimal AMP algorithm (this is the optimal iterative algorithm for i.i.d. Gaussian noise and is conjectured to be the optimal polynomial-time algorithm) to the rotationally invariant noise model is sub-optimal. The authors develop AMP algorithms that achieve improved performance by applying a non-linear matrix denoiser to the eigenvalues of the observed

matrix $\boldsymbol{Y}$ and characterize their state evolution. Based on non-rigorous statistical physics techniques [57], the authors provide a procedure to derive good matrix denoisers for the problem. The authors propose matrix denoisers with explicit formulas when V is a quartic or sestic polynomial. However, the state evolution of the resulting AMP algorithm is quite complicated; hence, the resulting algorithm's optimality (or sub-optimality) properties are not understood.

Our Contributions. Taking inspiration from the insights of Barbier et al. [8], we study the spiked matrix model with rotationally invariant noise from an algorithmic point of view. Our main contributions are:

- We develop a new class of approximate message-passing algorithms for this problem and provide a state evolution result that characterizes their dynamics in the high-dimensional limit (Theorem 1 in Section 3.1). At each iteration, these algorithms exploit the noise structure by applying a non-linear matrix denoiser to the eigenvalues of the observed matrix and the signal structure by applying a non-linear iterate denoiser to the previous iterates generated by the algorithm. These algorithms can be viewed as natural analogs of orthogonal [50] or vector approximate message-passing algorithms [69] developed for compressed sensing.

- A key feature of the AMP algorithms proposed in this work is that their state evolution is significantly simpler than that of existing AMP algorithms for this problem. Consequently, we are able to exploit our result to derive the optimal choices for the matrix and iterate denoisers (Section 3.2). Interestingly, we find that the matrix denoisers that optimize the performance of the AMP algorithm are closely related to the eigenvalue shrinkage estimators discovered by Bun et al. [15] in a separate line of work on denoising high-rank and unstructured signal matrices corrupted with rotationally invariant noise [15, 41, 42, 49, 67, 72].

- Building on the techniques developed by Celentano et al. [16] and Montanari and Wu [56] (in the context of i.i.d. Gaussian noise), we show that the AMP algorithm with the optimal choices for the matrix and iterate denoisers achieves the smallest possible asymptotic estimation error among a broad class of iterative algorithms under a fixed iteration budget (Theorem 2 in Section 3.3). This suggests that this algorithm might be the natural candidate for the optimal polynomial-time estimator for this problem.

- Finally, our results also suggest a general and concise conjecture for the Bayes risk for the problem (Section 3.4), which we hope can be proved rigorously in the future.

Organization. This paper is organized as follows. We start with some preliminary results in Section 2. The main results of this paper are presented in Section 3. Section 4 highlights the key ideas behind our results. Numerical experiments are presented in Section 5. The complete proofs of our main results are provided in the appendices.

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

Some common sets: The sets $\mathbb{N}, \mathbb{R}, \mathbb{C}$ represent the set of positive integers, real numbers, and complex numbers, respectively. For $N \in \mathbb{N}$, $[N]$ is the set $\{1, 2, 3, \ldots, N\}$ and $\mathbb{O}(N)$ denotes the set of $N \times N$ orthogonal matrices.

Linear Algebra: For vectors $u, v \in \mathbb{R}^k$, $\|u\|$ is the ℓ_2 norm of u, $\langle u, v \rangle = \sum_{i=1}^k u_i v_i$ denotes the standard inner product on $\mathbb{R}^k$, and $\mathrm{diag}(u)$ represents the $k \times k$ diagonal matrix constructed by placing the entries of u along the diagonal. For a matrix $M \in \mathbb{R}^{k \times k}$, $\mathrm{Tr}[M], \|M\|_{\mathrm{op}}, \|M\|$ represent the trace, operator (spectral) norm, and Frobenius norm of M respectively. If M is symmetric, we denote the sorted eigenvalues of M by $\lambda_1(M) \geq \cdots \geq \lambda_k(M)$. We use 1_k to denote the vector $(1, 1, \ldots, 1)$ in $\mathbb{R}^k$, 0_k to denote the zero vector $(0, 0, \ldots, 0)$ in $\mathbb{R}^k$, I_k denotes the $k \times k$ identity matrix, and $e_1, e_2, \ldots, e_k$ to denote the standard basis vectors in $\mathbb{R}^k$. When the dimension is clear from the context, we will abbreviate $1_k, 0_k, I_k$ as $1, 0, I$. We use the bold-face font for vectors and matrices whose dimensions diverge as N (the dimension of the signal vector) grows to ∞. For example, the signal $\boldsymbol{x}_\star \in \mathbb{R}^N$, and the noise matrix $\boldsymbol{W} \in \mathbb{R}^{N \times N}$ are bold-faced.

Probability: We use $\mathbb{E}[\cdot], \mathrm{Var}[\cdot], \mathrm{Cov}[\cdot]$ to denote expectations, variances, and covariances of random variables. The Gaussian distribution on $\mathbb{R}^k$ with mean vector $\mu \in \mathbb{R}^k$ and covariance matrix $\Sigma \in \mathbb{R}^{k \times k}$ is denoted by $\mathcal{N}(\mu, \Sigma)$. For a finite set A, $\mathsf{Unif}(A)$ represents the uniform distribution on A. We will use $\mathsf{Unif}(\mathbb{O}(N))$ to denote the Haar measure on the orthogonal group $\mathbb{O}(N)$. For any $x \in \mathbb{R}$, the probability measure δ_x on $\mathbb{R}$

denotes the point mass at x. We use $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{d}$ to denote convergence in probability and distribution, respectively. For a sequence of real-valued random variables $(Y_N)_{N\in\mathbb{N}}$, we say that $\operatorname{plim} Y_N = y$ if $Y_N \xrightarrow{\mathbb{P}} y$, $\operatorname{plim\,sup} Y_N \leq y$ if for any $\epsilon > 0$, $\lim_{N\to\infty} \mathbb{P}(Y_N \geq y + \epsilon) = 0$, and $\operatorname{plim\,inf} Y_N \geq y$ if for any $\epsilon > 0$, $\lim_{N\to\infty} \mathbb{P}(Y_N \leq y - \epsilon) = 0$.

2 Preliminaries

We begin by introducing the assumptions under which we study the rank-1 matrix estimation problem in (1), along with some key concepts that play an important role in this paper.

Convergence and Asymptotic Equivalence of High-Dimensional Vectors. We rely on the following notion of convergence.

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

$$|h(v) - h(v')| \leq L\|v - v'\| \cdot (1 + \|v\| + \|v'\|) \quad \forall\, v, v' \in \mathbb{R}^\ell, \tag{3}$$

for some $L < \infty$, we have:

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

We denote convergence in this sense using the notation: $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_\ell) \xrightarrow{W_2} (\mathsf{V}_1, \ldots, \mathsf{V}_\ell)$.

The following definition introduces a related notion of asymptotic equivalence of high-dimensional vectors.

Definition 2. Two N-dimensional random vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are asymptotically equivalent if:

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

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

Signal and Noise Models. We posit the following assumptions on the signal $\boldsymbol{x}_\star$, the side information $\boldsymbol{a}$, and the noise $\boldsymbol{W}$.

Assumption 1 (Signal Model)**.** The signal vector and side information satisfy $(\boldsymbol{x}_\star; \boldsymbol{a}) \xrightarrow{W_2} (\mathsf{X}_\star; \mathsf{A})$ for some limiting random variables $(\mathsf{X}_\star, \mathsf{A})$ with joint distribution π which satisfies $\mathbb{E}[\mathsf{X}_\star^2] = 1$ and $\mathbb{E}[\|\mathsf{A}\|^2] < \infty$.

Assumption 2 (Noise Model)**.** We model the noise matrix $\boldsymbol{W}$ as a random matrix with eigen-decomposition:

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

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

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

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

$$|\mu(\lambda) - \mu(\lambda')| \leq L \cdot |\lambda - \lambda'|^\alpha \quad \forall\, \lambda, \lambda' \in \mathbb{R}. \tag{4}$$

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

Stieltjes and Hilbert Transform. For any probability measure (or more generally, finite measure) χ on $\mathbb{R}$, the Stieltjes Transform $\mathcal{S}_\chi : \mathbb{C}\backslash\mathbb{R} \mapsto \mathbb{C}$ of χ is defined as:

$$\mathcal{S}_\chi(z) \stackrel{\text{def}}{=} \int_{\mathbb{R}} \frac{\chi(\mathrm{d}\lambda)}{z-\lambda} \quad \forall\, z \,\in\, \mathbb{C}\backslash\mathbb{R}.$$

It is well-known that a probability measure is uniquely determined by its Stieltjes transform (see for e.g., [3, Theorem 2.4.3]). Suppose that χ is compactly supported and absolutely continuous with respect to the Lebesgue measure with density $\chi(\cdot)$, which is Holder continuous (cf. (4)). Then, the Hilbert transform $\mathscr{H}_\chi : \mathbb{R} \mapsto \mathbb{R}$ of χ is defined as the Cauchy principal value of the following singular integral:

$$\mathscr{H}_\chi(z) \stackrel{\text{def}}{=} \lim_{\epsilon\to 0} \frac{1}{\pi} \int_{|z-\lambda|\geq\epsilon} \frac{\chi(\lambda)}{z-\lambda}\, \mathrm{d}\lambda \quad \forall\, z \,\in\, \mathbb{R}. \tag{5}$$

Under the Holder continuity requirement (4) on χ, the above limit exists and $\mathscr{H}_\sigma : \mathbb{R} \mapsto \mathbb{R}$ is also Holder continuous in the sense of (4); see *e.g.*, [61, Section 2.1] and [34, Theorem 14.11a].

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

$$\nu_N \stackrel{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \langle \boldsymbol{u}_i(\boldsymbol{Y}), \boldsymbol{x}_\star \rangle^2 \cdot \delta_{\lambda_i(\boldsymbol{Y})}, \tag{6}$$

where $\lambda_1(\boldsymbol{Y}), \ldots, \lambda_N(\boldsymbol{Y})$ denote the eigenvalues of $\boldsymbol{Y}$ and $\boldsymbol{u}_1(\boldsymbol{Y}), \ldots, \boldsymbol{u}_N(\boldsymbol{Y})$ denote the corresponding eigenvectors. The measure ν_N plays an important role in our analysis. The following lemma collects some important results regarding the asymptotic behavior of ν_N as $N \to \infty$. In order to state the result, we introduce the function $\phi : \mathbb{R} \mapsto \mathbb{R}$, which will play an important role throughout the paper:

$$\phi(\lambda) \stackrel{\text{def}}{=} (1 - \pi\theta\mathscr{H}_\mu(\lambda))^2 + \pi^2\theta^2\mu^2(\lambda) \quad \lambda \,\in\, \mathbb{R}. \tag{7}$$

In the above display $\mathscr{H}_\mu$ denotes the Hilbert transform of μ.

Lemma 1. *We have,*

1. *The measure ν_N converges weakly in probability to a compactly supported probability measure ν on $\mathbb{R}$ with Stieltjes Transform:*

$$\mathcal{S}_\nu(z) = \frac{\mathcal{S}_\mu(z)}{1 - \theta\mathcal{S}_\mu(z)} \quad \forall\, z \,\in\, \mathbb{C}\backslash\mathbb{R},$$

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

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

2. *For Lebesgue-almost every λ, $\phi(\lambda) \neq 0$ and for $\nu_\perp$-almost every λ, $\phi(\lambda) = 0$.*
3. *The density of the absolutely continuous part of ν is given by $\mu(\cdot)/\phi(\cdot)$ where $\mu(\cdot)$ denotes the density of μ and the function $\phi(\cdot)$ is as defined in (7).*

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

Definition 3 (Gaussian Channel)**.** A Gaussian channel is a collection of real-valued random variables $(\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})$ where the signal $\mathsf{X}_\star$ and the side information A are drawn from the prior π fro Assumption 1: $(\mathsf{X}_\star, \mathsf{A}) \sim \pi$ and the observations $\mathsf{X}_1, \ldots, \mathsf{X}_t$ are given by: $\mathsf{X}_i = \alpha_i \mathsf{X}_\star + \mathsf{Z}_i \quad \forall\, i \in [t]$, where $\alpha_1, \ldots, \alpha_t$ are real numbers and $(\mathsf{Z}_1, \ldots, \mathsf{Z}_t) \sim \mathcal{N}(0, \Sigma)$ are zero mean and jointly Gaussian random variables, sampled independently from $(\mathsf{X}_\star, \mathsf{A})$. We introduce some important notions related to Gaussian channels.

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

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

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

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

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

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

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

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

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

MMSE Function and MMSE Estimator for Scalar Gaussian Channels. The function $\mathsf{mmse}_\pi : [0, 1] \mapsto [0, 1]$ represents MMSE of a scalar Gaussian channel as a function of the SNR ω:

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

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

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

DMMSE Function and DMMSE Estimator for Scalar Gaussian Channels. The function $\mathsf{dmmse}_\pi : [0, 1] \mapsto [0, 1]$ represents DMMSE of a scalar Gaussian channel as a function of the SNR ω:

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

The function $\bar{\varphi}(\cdot|\omega) : \mathbb{R} \times \mathbb{R}^k \mapsto \mathbb{R}$ denotes the DMMSE estimator for the scalar Gaussian channel at SNR ω. Ma and Ping [50] have shown that the DMMSE estimator of a scalar Gaussian channel at SNR ω is given by (see Lemma 2 in Appendix A.2):

$$\bar{\varphi}(x; a|\omega) \stackrel{\text{def}}{=} \begin{cases} \left(1 - \frac{\sqrt{\omega}}{\sqrt{1-\omega}} \cdot \mathbb{E}[\mathsf{Z}\varphi(\mathsf{X}; \mathsf{A}|\omega)]\right)^{-1} \cdot \left(\varphi(x; a|\omega) - \frac{\mathbb{E}[\mathsf{Z}\varphi(\mathsf{X}; \mathsf{A}|\omega)]}{\sqrt{1-\omega}} \cdot x\right) & : \omega < 1 \\ x & : \omega = 1. \end{cases} \tag{11}$$

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

Assumption 3. For any $\omega \in [0, 1]$, the MMSE estimator $\varphi(\cdot|\omega) : \mathbb{R} \times \mathbb{R}^k \mapsto \mathbb{R}$ for the scalar Gaussian channel $(\mathsf{X}_\star; \mathsf{A}) \sim \pi, \quad \mathsf{X}|\mathsf{A}, \mathsf{X}_\star \sim \mathcal{N}\left(\sqrt{\omega} \cdot \mathsf{X}_\star, 1 - \omega\right)$ is continuously differentiable and Lipschitz.

Remark 1. In the absence of any side information, a sufficient condition for Assumption 3 to hold is that the signal random variable $\mathsf{X}_\star$ is compactly supported (see [54, Remark 2.3] and [31, Lemma 3.8]).

3 Main results

We now present the main results obtained in this paper.

3.1 Orthogonal Approximate Message Passing Algorithms

We introduce a class of iterative methods for rank-1 matrix estimation problem, called *Orthogonal Approximate Message Passing* (OAMP) algorithms.

Definition 4 (OAMP algorithms)**.** An OAMP algorithm generates iterates $\boldsymbol{x}_1, \boldsymbol{x}_2, \dots$ in $\mathbb{R}^N$ according to the update rule:

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

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

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

In the above equations, for each $t \in \mathbb{N}$, the matrix denoiser $\Psi_t : \mathbb{R} \mapsto \mathbb{R}$ acts on the matrix $\boldsymbol{Y}$ using the usual convention: if $\boldsymbol{Y} = \boldsymbol{O}\,\mathrm{diag}(\lambda_1, \dots, \lambda_N)\boldsymbol{O}^\top$ is the eigen-decomposition of $\boldsymbol{Y}$, then, $\Psi_t(\boldsymbol{Y}) = \boldsymbol{O}\,\mathrm{diag}(\Psi_t(\lambda_1), \dots, \Psi_t(\lambda_N))\boldsymbol{O}^\top$. Likewise, the iterate denoisers $f_t : \mathbb{R}^{t-1} \times \mathbb{R}^k \mapsto \mathbb{R}$ and the post-processing function function $\psi_t : \mathbb{R}^t \times \mathbb{R}^k \mapsto \mathbb{R}$ act entry-wise on the N components of its vector inputs.

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

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

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

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

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

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

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

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

$$\operatorname*{plim}_{N\to\infty} \frac{\mathrm{Tr}[\Psi_t(\boldsymbol{Y})]}{N} = \mathbb{E}[\Psi_t(\mathsf{\Lambda})] = 0, \quad \mathsf{\Lambda} \sim \mu. \tag{14}$$

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

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

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

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

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

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

We present a heuristic derivation of this result in Section 4.1, and the complete proof in Appendix B.

Remark 2 (Connections to AMP Algorithms for Compressed Sensing). The OAMP algorithm introduced in Definition 4 can be viewed as a natural analog of the orthogonal AMP (OAMP) [50] or vector AMP (VAMP) algorithm [69, 74] developed for compressed sensing (or regularized linear regression). The key feature of these algorithms, which is shared by the algorithm in Definition 1, is the use of trace-free matrix denoisers (cf. (14)) and divergence-free iterate denoisers (cf. (15)). These features significantly simplify the algorithm's state evolution. However, unlike the OAMP algorithm introduced in this work, the compressed sensing algorithms compute a matrix-vector multiplication involving a rotationally invariant matrix at each iteration. In contrast, only the observed matrix $\boldsymbol{Y}$ is available in the spiked matrix model, not the rotationally invariant noise matrix $\boldsymbol{W}$. Although $\boldsymbol{Y}$ is a rank-1 perturbation of $\boldsymbol{W}$, $\Psi_t(\boldsymbol{Y})$, the matrix used by the OAMP algorithm at iteration t, cannot be expressed as a simple perturbation of $\boldsymbol{W}$ for a general matrix denoiser Ψ_t. This complicates the analysis of these algorithms. We refer the reader to Fan [29, Remark 3.3] for related discussions.

3.2 The Optimal OAMP Algorithm

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

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

The limiting value of the MSE depends implicitly on the functions $\{\Psi_t, f_t, \psi_t\}_{t\geq 1}$ used in the OAMP algorithm since these functions determine the joint distribution of $(\mathsf{X}_\star, (\mathsf{X}_t)_{t\in\mathbb{N}}; \mathsf{A})$. As our second contribution, we use the above characterization to derive the optimal choice for these functions which minimizes the MSE. We call the resulting algorithm the *optimal OAMP algorithm*, and introduce it below.

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

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

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

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

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

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

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

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

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

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

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

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

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

In the above equations,

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

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

  where $\Lambda \sim \mu$ and the function $\phi : \mathbb{R} \mapsto \mathbb{R}$ was introduced in (7).

- ω_t and ρ_t are computed using the recursion:

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

  initialized with $\omega_0 \stackrel{\text{def}}{=} 0$.

- $\mathsf{dmmse}_\pi(\omega)$ denotes the DMMSE for a scalar Gaussian channel with SNR ω, and $\varphi(\cdot|\omega), \bar{\varphi}(\cdot|\omega)$ denote the MMSE and DMMSE estimators for the Gaussian channel (Definition 3).

Remark 3. The additional scaling factor $1/\sqrt{\omega_t} \cdot (1 + 1/\rho_t)$ in (17c) is introduced to normalize the iterates so that $\|\boldsymbol{x}_t\|^2/N \xrightarrow{\mathbb{P}} 1$.

The following result characterizes the asymptotic MSE of the optimal OAMP algorithm in this case. Its proof can be found in Appendix C.

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

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

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

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

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

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

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

3.3 An Optimality Result

Our final main result shows that the optimal OAMP algorithm introduced in (17) attains the best possible asymptotic MSE within a broad class of iterative algorithms, which we define formally below.

Definition 5 (Iterative Algorithms)**.** An iterative algorithm generates iterates $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots$ in $\mathbb{R}^N$ according to the update rule:

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

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

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

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

The following theorem states our optimality result.

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

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

3.4 Information-Theoretic v.s. Computational Limits

A natural question is whether the optimal OAMP algorithm introduced in (17) achieves the smallest asymptotic estimation error among *all estimators*, not just estimators computable using efficient iterative algorithms. To address this question, we begin by recalling the conjecture of Barbier et al. [8] regarding the fundamental information-theoretic limits of this problem.

Replica conjecture for the Bayes risk. Under the assumption that the entries of $(\boldsymbol{x}_\star, \boldsymbol{a})$ are drawn i.i.d. from a prior π, the information-theoretically optimal estimator is the Bayes estimator $\mathbb{E}[\boldsymbol{x}_\star|\boldsymbol{Y}, \boldsymbol{a}]$. Barbier et al. [8] provide a conjecture for the asymptotic MSE of this estimator (also called the Bayes risk) assuming the noise matrix drawn from the *trace ensemble* [8, 61] with density:

$$p(\boldsymbol{W}) \propto \exp\left(-\frac{N}{2}\sum_{i=1}^{N} V(\lambda_i(\boldsymbol{W}))\right) \qquad \text{where } \lambda_1(\boldsymbol{W}), \ldots, \lambda_N(\boldsymbol{W}) \text{ denote the eigenvalues of } \boldsymbol{W}, \tag{20}$$

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

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

and $\gamma \in [0, 1]$ is a parameter for the noise model. The particular choice of κ in (21) ensures that the limiting spectral measure (μ) of $\boldsymbol{W}$ drawn from the trace ensemble (20) has unit variance [8].

Conjecture 1 (Replica Conjecture for Bayes risk, Barbier et al. 8)**.** *Suppose that the entries of $(\boldsymbol{x}_\star, \boldsymbol{a})$ are drawn i.i.d. from a prior π with $\mathsf{mmse}_\pi(0) \in (0, 1)$ and the noise matrix $\boldsymbol{W}$ is drawn from the trace ensemble (20) with the quartic potential function V from (21). Then,*

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

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

$$m = 1 - \mathsf{mmse}_\pi(\omega), \tag{22a}$$

$$\mathbb{E}[\mathsf{H}] = 1 - m, \tag{22b}$$

$$\chi = \mathbb{E}[\Lambda \mathsf{Q} \mathsf{H}], \tag{22c}$$

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

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

$$\mathsf{Q} \stackrel{def}{=} \kappa\theta^2 m \Lambda^2 + \kappa\theta^2 \chi \Lambda - \frac{\kappa\theta^2}{1-m} \mathbb{E}\left[m\Lambda^2 \mathsf{H} + \chi \Lambda \mathsf{H} \right] + \frac{m}{1-m}, \tag{22e}$$

$$\mathsf{H} \stackrel{def}{=} \left(\rho + \theta^2 a^2 (\gamma + 2a^2\kappa)^2 + 1 - \theta \left(\gamma\Lambda - \theta\kappa\Lambda^2 + \kappa\Lambda^3 \right) \right)^{-1}. \tag{22f}$$

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

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

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

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

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

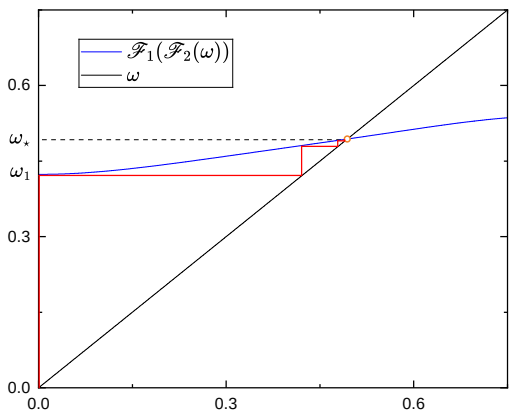
found by the recursion $\rho_t = \mathscr{F}_2(\omega_{t-1})$, $\omega_t = \mathscr{F}_1(\rho_t)$. The following proposition (proved in Appendix D.2) shows that for the quartic potential, the *replica fixed point equations* (22) and the *state evolution fixed point equations* (23) are equivalent.

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

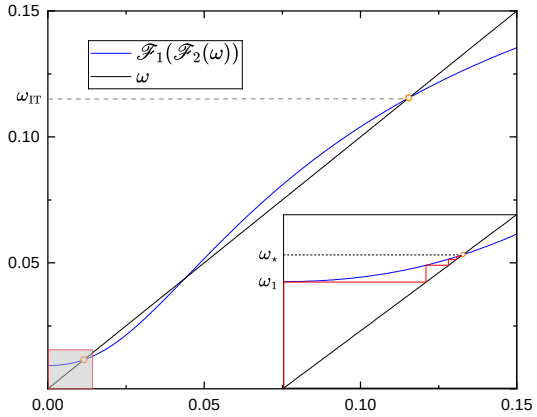
We remark that an analogous result holds for the sestic (degree 6) potential. Since the replica equations for the sestic ensemble in Barbier et al. [8] are even more involved, we do not provide the details here. We anticipate that the fixed point equations in (23) are a unified and concise reformulation of the replica fixed point equations and characterize the asymptotic Bayes risk for general potential functions V. This reformulation of the replica conjecture for the asymptotic Bayes risk may be more amenable to rigorous proof than the significantly more complicated replica formulas derived using the approach of Barbier et al. [8].

Information-Theoretic Optimality and Sub-Optimality of OAMP. We can simplify the fixed point equations in (23) (which are equivalent to the replica fixed point equations (22)) by eliminating ρ to obtain a single equation $\omega = \mathscr{F}_1(\mathscr{F}_2(\omega))$. In some cases, this equation has a unique solution, implying the optimal OAMP algorithm matches the asymptotic MSE of the Bayes estimator and is information-theoretically optimal (see Figure 1(a)). In other cases, the equation may have multiple solutions, as illustrated in Figure 1(b), where it has two stable fixed points. Here, the optimal OAMP algorithm converges to the inferior fixed point ω_*, while the Bayes risk corresponds to the superior fixed point $\omega_{\rm IT}$. Since $\omega_* < \omega_{\rm IT}$, the optimal OAMP or any iterative algorithm (in the form of (19a)) with a constant (N-independent) number of iterations fails to achieve the Bayes-optimal MSE. Such scenarios also occur in i.i.d. Gaussian noise models [16, 55, 56], and it is conjectured that no polynomial time algorithm can achieve the information-theoretically optimal MSE in these cases.

[1] When the fixed point equations in (22) have multiple solutions, the correct fixed point is the one that minimizes a certain free energy function calculated in [8].



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



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

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

4 Proof Ideas

This section presents some important intuitions and ideas used to obtain our main results.

4.1 Heuristic Derivation of State Evolution (Theorem 1)

We begin by an intuitive derivation of the state evolution result (Theorem 1) for OAMP algorithms, highlighting the key ideas involved in the proof. A rigorous proof of Theorem 1 is presented in Appendix B. Consider a general OAMP algorithm (Definition 4):

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

where $(\Psi_t)_{t\in\mathbb{N}}$ are trace-free (cf. (14)) and $(f_t)_{t\in\mathbb{N}}$ are divergence-free (cf. (15)). For ease of exposition, we present the key ideas assuming that the matrix denoisers $(\Psi_t)_{t\in\mathbb{N}}$ *are polynomials* (the general case follows by a polynomial approximation argument). Let us decompose $f_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{a})$ into a component along the direction of $\boldsymbol{x}_\star$ and a component perpendicular to it:

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

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

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

As pointed out in Remark 2, the main difficulty in analyzing the algorithm above is that the matrix $\Psi_t(\boldsymbol{Y})$ is not rotationally invariant. To address this, we express the update equation (25) in terms of the rotationally invariant noise matrix $\boldsymbol{W}$ by expanding the polynomial $\Psi_t(\boldsymbol{Y})$ in terms of $\boldsymbol{W}$ using the fact that $\boldsymbol{Y} = \frac{\theta}{N}\boldsymbol{x}_\star\boldsymbol{x}_\star^\mathsf{T} + \boldsymbol{W}$. Although this expansion can initially seem complicated, the key insight which makes it tractable is to show that the following approximations hold:

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

where $\widetilde{\Psi}_t : \mathbb{R} \mapsto \mathbb{R}$ is polynomial obtained by appropriately transforming Ψ_t; we refer the reader to Appendix B for details of the above approximations. The transformed matrix denoiser $\widetilde{\Psi}_t$ is determined by Ψ_t via a

complicated recursion and is not necessarily trace-free. Substituting (26) into (25), we get

$$\boldsymbol{x}_t \approx \alpha_t \cdot \widetilde{\Psi}_t(\boldsymbol{W})\boldsymbol{x}_\star + \Psi_t(\boldsymbol{W})\boldsymbol{f}_t^\perp = \alpha_t \cdot \frac{\mathrm{Tr}[\widetilde{\Psi}_t(\boldsymbol{W})]}{N} \cdot \boldsymbol{x}_\star + \alpha_t \cdot \left(\widetilde{\Psi}_t(\boldsymbol{W}) - \frac{\mathrm{Tr}[\widetilde{\Psi}_t(\boldsymbol{W})]}{N} \cdot \boldsymbol{I}_N\right)\boldsymbol{x}_\star + \Psi_t(\boldsymbol{W})\boldsymbol{f}_t^\perp$$

$$\stackrel{\text{def}}{=} \underbrace{\alpha_t \cdot \frac{\mathrm{Tr}[\widetilde{\Psi}_t(\boldsymbol{W})]}{N} \cdot \boldsymbol{x}_\star}_{\text{Signal Component}} + \underbrace{\boldsymbol{z}_t}_{\text{Eff. Noise}}, \tag{27a}$$

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

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

To derive the limiting distribution of $\boldsymbol{x}_t$, we analyze the signal and noise components separately.

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

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

where the approximation in (a) follows from standard concentration results for quadratic forms of rotationally invariant matrices (see Fact 2 in Appendix F), (b) follows from the defining property of $\widetilde{\Psi}_t(\boldsymbol{W})$ in (26), (c) follows by recalling the definition of the spectral measure in the signal direction (ν_N) from (6), and (d) follows from the weak convergence of the compactly supported measure ν_N to ν (Lemma 1, item (1)).

Noise Component. The update rule (27b) for the effective noise $\boldsymbol{z}_t$ is written in a form for which existing state evolution results apply [25, 26, 29, 50, 69, 74, 75]. Indeed, $\widehat{\Psi}_t(\boldsymbol{W})$ and $\Psi_t(\boldsymbol{W})$ are functions of the rotationally-invariant matrix $\boldsymbol{W}$ and have zero trace; and $\boldsymbol{f}_t^\perp$ is a divergence-free function of $\{\boldsymbol{x}_i\}_{i<t}$. By appealing to existing results on the dynamics of AMP algorithms driven by rotationally invariant matrices, we show that the effective noise component $\boldsymbol{z}_t$ converges to a centered Gaussian random variable Z_t. Moreover, the covariance of Z_t and Z_s can be heuristically calculated based on a convenient property of OAMP algorithms. Specifically, for any $G_1(\boldsymbol{W}), G_2(\boldsymbol{W}) \in \left\{\widehat{\Psi}_t(\boldsymbol{W}), \Psi_t(\boldsymbol{W}), \Psi_s(\boldsymbol{W})\right\}$ and $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \{\boldsymbol{x}_\star, \boldsymbol{f}_s^\perp, \boldsymbol{f}_t^\perp\}$, we have

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

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

$$\mathbb{E}[\mathsf{Z}_s\mathsf{Z}_t] \approx \frac{\langle \boldsymbol{z}_s, \boldsymbol{z}_t\rangle}{N} = \frac{\left\langle \alpha_t\widehat{\Psi}_t(\boldsymbol{W})\boldsymbol{x}_\star + \Psi_t(\boldsymbol{W})\boldsymbol{f}_t^\perp, \alpha_s\widehat{\Psi}_s(\boldsymbol{W})\boldsymbol{x}_\star + \Psi_s(\boldsymbol{W})\boldsymbol{f}_s^\perp \right\rangle}{N} \tag{30a}$$

$$\approx \alpha_s\alpha_t \cdot \frac{\mathrm{Tr}[\widehat{\Psi}_t(\boldsymbol{W}), \widehat{\Psi}_s(\boldsymbol{W})]}{N} \cdot \frac{\|\boldsymbol{x}_\star\|^2}{N} + \frac{\mathrm{Tr}[\Psi_t(\boldsymbol{W})\Psi_s(\boldsymbol{W})]}{N} \cdot \frac{\left\langle \boldsymbol{f}_t^\perp, \boldsymbol{f}_s^\perp\right\rangle}{N}, \tag{30b}$$

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

4.2 Derivation of the Optimal OAMP Algorithm

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

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

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

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

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

$$\beta_t = \mathbb{E}[\mathsf{X}_\star f_t(\mathsf{X}_{t-1}; \mathsf{A})] \cdot \mathbb{E}\big[\Psi_t(\mathsf{\Lambda}_\nu)\big], \tag{32b}$$

$$\mathbb{E}[\mathsf{Z}_t^2] = \big(\mathbb{E}[\mathsf{X}_\star f_t(\mathsf{X}_{t-1}; \mathsf{A})]\big)^2 \cdot \mathrm{Var}[\Psi_t(\mathsf{\Lambda}_\nu)] + \Big(\mathbb{E}[f_t(\mathsf{X}_{t-1}; \mathsf{A})^2] - \big(\mathbb{E}[\mathsf{X}_\star f_t(\mathsf{X}_{t-1}; \mathsf{A})]\big)^2\Big) \cdot \mathbb{E}[\Psi_t^2(\mathsf{\Lambda})]. \tag{32c}$$

In the above definitions, the random variables $\mathsf{\Lambda} \sim \mu$, $\mathsf{\Lambda}_\nu \sim \nu$ are independent of all other random variables. A natural design principle is to choose f_t, Ψ_t to maximize the SNR ω_t of the Gaussian channel $(\mathsf{X}_\star, \mathsf{X}_t; \mathsf{A})$, which is given by the squared cosine similarity between X_t and $\mathsf{X}_\star$:

$$\omega_t \overset{\text{def}}{=} \frac{\big(\mathbb{E}\left[\mathsf{X}_\star \mathsf{X}_t\right]\big)^2}{\mathbb{E}[\mathsf{X}_\star^2] \cdot \mathbb{E}[\mathsf{X}_t^2]} = \frac{\beta_t^2}{\beta_t^2 + \mathbb{E}[\mathsf{Z}_t^2]} = \frac{\big(\mathbb{E}\big[\Psi_t(\mathsf{\Lambda}_\nu)\big]\big)^2}{\mathbb{E}\big[\Psi_t^2(\mathsf{\Lambda}_\nu)\big] + \frac{\delta_t}{1-\delta_t} \cdot \mathbb{E}[\Psi_t^2(\mathsf{\Lambda})]} \quad \text{where} \quad \delta_t \overset{\text{def}}{=} 1 - \frac{\big(\mathbb{E}[\mathsf{X}_\star f_t(\mathsf{X}_{t-1}; \mathsf{A})]\big)^2}{\mathbb{E}[\mathsf{F}_t^2]}. \tag{33}$$

We consider the two maximization problems over Ψ_t and f_t one-by-one.

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

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

with the parameter $\rho \in (0, \infty)$ fixed at $\rho = \delta_t^{-1} - 1$ (notice δ_t does not depend on Ψ_t). The optimization problem in (34) can be written in the more convenient but equivalent form:

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

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

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

where step (a) can be verified by solving the inner quadratic minimization problem over c explicitly, and step (b) is due to a change of variable $\Psi \mapsto c\Psi$; note that $c\Psi$ also satisfies the zero-trace constraint. Note that any minimizer of (35c) is also a maximizer of (34) since the objective function of (34) is invariant to a rescaling of Ψ. Let $\nu = \nu_\perp + \nu_\parallel$ be the Lebesgue decomposition of of ν into the singular part $\nu_\perp$ and the absolutely continuous part $\nu_\parallel$. Let S denote the set:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Recalling the definition of the DMMSE estimator from Definition 3 (see also Lemma 2 in Appendix A.2), the minimizer of (39c) is:

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

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

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

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

4.3 Proof of the Optimality Result (Theorem 2)

We now present the proof of Theorem 2. We will introduce the key ideas involved in the form of some intermediate results, whose proofs are deferred to Appendix E.

Implementing Iterative Algorithms using OAMP Algorithms. Our proof builds on the techniques introduced in the work of Celentano et al. [16] and Montanari and Wu [56]. These works show that when the noise matrix $\boldsymbol{W}$ has i.i.d. Gaussian entries, any estimator computed using $t \in \mathbb{N}$ iterations of an iterative algorithm can also be computed using t iterations of a suitably designed AMP algorithm. Consequently, we can restrict the search space for the optimal t-iteration iterative algorithm to the space of t-iteration AMP algorithms, whose dynamics are significantly easier to analyze. In light of these results, it is natural to hope that any estimator computed using t iterations of a general iterative algorithm (as defined in Definition 5) can be computed using t iterations of a suitably designed OAMP algorithm (as defined in Definition 4). Unfortunately, the argument used by Celentano et al. [16] and Montanari and Wu [56] breaks down in our case because we require that the iterate denoisers used in the OAMP algorithm satisfy the divergence-free requirement in (15). To illustrate the difficulty caused by imposing the divergence-free requirement (15) and motivate our proof strategy, we consider the first two iterations of a simple iterative algorithm:

$$\boldsymbol{r}_1 = \Psi_1(\boldsymbol{Y}) \cdot f_1(\boldsymbol{a}), \quad \boldsymbol{r}_2 = \Psi_2(\boldsymbol{Y}) f_2(\boldsymbol{r}_1; \boldsymbol{a}), \tag{42}$$

which we refer to as the *target algorithm*, and try to design an OAMP algorithm, which can reconstruct the iterates of the target algorithm. For simplicity, we assume that the matrix denoisers used by the target algorithm respect the trace-free requirement $\mathbb{E}_{\Lambda\sim\mu}[\Psi_1(\Lambda)] = \mathbb{E}_{\Lambda\sim\mu}[\Psi_2(\Lambda)] = 0$ (requirement (14) in Definition 4). Since the divergence-free constraint (15) is inactive for the first iteration, we can construct a valid OAMP algorithm which simply copies the first iteration of the target algorithm $\boldsymbol{x}_1 = \Psi_1(\boldsymbol{Y}) f_1(\boldsymbol{a})$. Let $(\mathsf{X}_\star, \mathsf{X}_1; \mathsf{A})$ be the state evolution random variables corresponding to the first iteration of this OAMP algorithm. Next, we consider the second iteration of the target algorithm (42). Notice that the denoiser f_2 might not satisfy the divergence-free requirement $\mathbb{E}[\partial_1 f_2(\mathsf{X}_1; \mathsf{A})] = 0$ imposed in the definition of OAMP algorithms (Definition 4). To ensure this requirement is fulfilled, we are forced to design the second iteration of the OAMP algorithm as:

$$\boldsymbol{x}_2 = \Psi_2(\boldsymbol{Y}) \cdot \bar{f}_2(\boldsymbol{x}_1; \boldsymbol{a}) \quad \text{where} \quad \bar{f}_2(x; a) \stackrel{\text{def}}{=} f_2(x; a) - \mathbb{E}[\partial_1 f_2(\mathsf{X}_1; \mathsf{A})] \cdot x \quad \forall\, x \in \mathbb{R}, a \in \mathbb{R}^k.$$

By construction $\bar{f}_2$ satisfies the divergence-free constraint $\mathbb{E}[\partial_1 \bar{f}_2(\mathsf{X}_1; \mathsf{A})] = 0$. We can express the second iterate of the target algorithm (42) as:

$$\boldsymbol{r}_2 = \boldsymbol{x}_2 + \mathbb{E}[\partial_1 f_1(\mathsf{X}_1; \mathsf{A})] \cdot \Psi_2(\boldsymbol{Y}) \cdot \boldsymbol{x}_1 \stackrel{\text{(a)}}{=} \boldsymbol{x}_2 + \mathbb{E}[\partial_1 f_1(\mathsf{X}_1; \mathsf{A})] \cdot \underbrace{\Psi_2(\boldsymbol{Y}) \Psi_1(\boldsymbol{Y}) \cdot f_1(\boldsymbol{a})}_{\text{uncomputed}}, \tag{43}$$

where in step (a) we recalled that $\boldsymbol{x}_1 = \Psi_1(\boldsymbol{Y}) \cdot f_1(\boldsymbol{a})$. In particular, we are unable to reconstruct the second iterate of the target algorithm in (42) using the first two iterations of the OAMP algorithm we have designed since the matrix-vector product $\Psi_2(\boldsymbol{Y})\Psi_1(\boldsymbol{Y}) \cdot f_1(\boldsymbol{a})$ in (43) has not been computed by the OAMP algorithm in the sense that it cannot be expressed in terms of the OAMP iterates $\boldsymbol{x}_1, \boldsymbol{x}_2$.

Lifted OAMP Algorithms. To address the issue highlighted above, we introduce a more powerful class of algorithms called *lifted OAMP algorithms*, which are parameterized by a degree parameter $D \in \mathbb{N}$, which modulates the computational power of these algorithms. A degree-D OAMP algorithm can compute D matrix-vector products per iteration. For instance, in the first iteration, a degree-D lifted OAMP algorithm computes the matrix-vector products:

$$(\boldsymbol{Y} - \mathbb{E}[\Lambda] \cdot \boldsymbol{I}_N) \cdot f_1(\boldsymbol{a}),\ (\boldsymbol{Y}^2 - \mathbb{E}[\Lambda^2] \cdot \boldsymbol{I}_N) \cdot f_1(\boldsymbol{a}),\ \ldots,\ (\boldsymbol{Y}^D - \mathbb{E}[\Lambda^D] \cdot \boldsymbol{I}_N) \cdot f_1(\boldsymbol{a}). \tag{44}$$

Since continuous functions can be approximated arbitrarily well by polynomial functions on compact sets, for any continuous function $\Phi : \mathbb{R} \mapsto \mathbb{R}$, we can approximate the matrix-vector product $\Phi(\boldsymbol{Y}) \cdot f_1(\boldsymbol{a})$ using a linear combination of the matrix-vector products in (44). In particular, the missing vector-matrix product in (43) needed to reconstruct the second iteration of the target algorithm (42) can also be approximated using the matrix-vector products (44) computed by the lifted OAMP algorithm. The following definition formally introduces the class of lifted OAMP algorithms.

Definition 6 (Degree-D Lifted OAMP Algorithm)**.** For any $D \in \mathbb{N}$ (independent of N), a degree-D lifted OAMP algorithm maintains a sequence of iterates $(\boldsymbol{w}_{ti})_{t\in\mathbb{N},i\in[D]}$ indexed by a time index $t \in \mathbb{N}$ and a degree index $i \in [D]$. At step t, the lifted OAMP computes the iterates $\boldsymbol{w}_{t,1}, \dots, \boldsymbol{w}_{t,D}$ as follows:

$$\boldsymbol{w}_{t,i} = \big(\boldsymbol{Y}^i - \mathbb{E}[\Lambda^i] \cdot \boldsymbol{I}_N\big) \cdot f_t\big(\boldsymbol{w}_{1,\bullet}, \dots, \boldsymbol{w}_{t-1,\bullet}; \boldsymbol{a}\big) \quad \forall\, i \in [D],\ t \in \mathbb{N}. \tag{45}$$

The estimate of $\boldsymbol{x}_\star$ at iteration t is obtained by applying a post-processing function ψ_t to the iterates $\boldsymbol{w}_{1,\bullet}, \dots, \boldsymbol{w}_{t,\bullet}$ generated so far and the side-information $\boldsymbol{a}$:

$$\widehat{\boldsymbol{w}}_t = \psi_t(\boldsymbol{w}_{1,\bullet}, \dots, \boldsymbol{w}_{t-1,\bullet}; \boldsymbol{a}), \tag{46}$$

In the above equations, the notation $\boldsymbol{w}_{t,\bullet}$ is a shorthand for the collection of vectors $\boldsymbol{w}_{t,1}, \boldsymbol{w}_{t,2}, \dots, \boldsymbol{w}_{t,D}$. Furthermore for each $t \in \mathbb{N}$, the iterate denoiser $f_t : \mathbb{R}^{(t-1)D} \times \mathbb{R}^k \mapsto \mathbb{R}$ and the postprocessing function$\psi_t : \mathbb{R}^{tD} \times \mathbb{R}^k \mapsto \mathbb{R}$ are continuously differentiable, Lipschitz, act entry-wise on their vector inputs, and do not change with N. Each lifted OAMP algorithm is associated with a collection of *state evolution random variables* $(\mathsf{X}_\star, (\mathsf{W}_{t,i})_{t\in\mathbb{N},i\in[D]}; \mathsf{A})$, whose distribution is given by:

$$(\mathsf{X}_\star, \mathsf{A}) \sim \pi, \quad \mathsf{W}_{t,i} = \alpha_t q_i \cdot \mathsf{X}_\star + \mathsf{Z}_{ti} \quad \forall\, t \in \mathbb{N}, i \in [D], \tag{47a}$$

where $\{\mathsf{Z}_{ti}\}_{t\in\mathbb{N},i\in\mathbb{N}}$ are centered Gaussian random variables, sampled independently of $\mathsf{X}_\star, \mathsf{A})$ with covariance:

$$\mathbb{E}[\mathsf{Z}_{si}\mathsf{Z}_{tj}] = \alpha_s\alpha_t \cdot \mathrm{Cov}_{\Lambda_\nu\sim\nu}[\Lambda^i_\nu, \Lambda^j_\nu] + (\Sigma_{st} - \alpha_s\alpha_t) \cdot \mathrm{Cov}_{\Lambda\sim\mu}[\Lambda^i, \Lambda^j]. \tag{47b}$$

In the above equations, $(\alpha_t)_{t\in\mathbb{N}}$ and $(\Sigma_{st})_{s\in\mathbb{N},t\in\mathbb{N}}$ are given by the recursion:

$$\alpha_t = \mathbb{E}\left[\mathsf{X}_\star f_t\left(\mathsf{W}_{1,\bullet}, \dots, \mathsf{W}_{t-1,\bullet}; \mathsf{A}\right)\right], \quad \Sigma_{st} = \mathbb{E}\left[f_s\left(\mathsf{W}_{1,\bullet}, \dots, \mathsf{W}_{s-1,\bullet}; \mathsf{A}\right) f_t\left(\mathsf{W}_{1,\bullet}, \dots, \mathsf{W}_{t-1,\bullet}; \mathsf{A}\right)\right]. \tag{47c}$$

Finally, we require the iterate denoisers $(f_t)_{t\in\mathbb{N}}$ to satisfy the *divergence-free* constraint:

$$\mathbb{E}[\partial_{s,i} f_t(\mathsf{W}_{1,\bullet}, \dots, \mathsf{W}_{t-1,\bullet}; \mathsf{A})] = 0 \quad \forall\, s \in [t-1],\ i \in [D],\ t \in \mathbb{N}, \tag{48}$$

where $\partial_{s,i} f_t(w_{1,\bullet}, \dots, w_{s,\bullet}, \dots, w_{t-1,\bullet}; a)$ denotes the partial derivative of $f_t(w_{1,\bullet}, \dots, w_{s,\bullet}, \dots, w_{t-1,\bullet}; a)$ with respect to $w_{s,i}$.

A degree-D lifted OAMP algorithm which runs for t iterations can be viewed as an instance of an OAMP algorithm (Definition 4) which runs for tD iterations. Hence, the following result as an immediate corollary of Theorem 1.

Corollary 1 (State evolution of lifted OAMP algorithms)**.** *Consider a general degree-D lifted OAMP algorithm of the form (45) and let $(\mathsf{X}_\star, (\mathsf{W}_{t,i})_{t\in\mathbb{N},i\in[D]}; \mathsf{A})$ be the associated state evolution random variables. Then for any $t \in \mathbb{N}$,*

$$(\boldsymbol{x}_\star, \boldsymbol{w}_{1,\bullet}, \boldsymbol{w}_{2,\bullet}, \dots, \boldsymbol{w}_{t,\bullet}; \boldsymbol{a}) \xrightarrow{W_2} (\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \dots, \mathsf{W}_{t,\bullet}; \mathsf{A}).$$

The following proposition (proved in Appendix E.1), shows that any estimator that is computed by an iterative algorithm (Definition 5) can be approximated arbitrarily well by a suitably designed degree-D lifted OAMP algorithm in the limit $D \to \infty$.

Proposition 3. *Let $\widehat{\boldsymbol{r}}_t$ be the estimator returned by any iterative algorithm (Definition 5) after $t \in \mathbb{N}$ iterations. Then, for each $D \in \mathbb{N}$, there is a degree-D lifted OAMP algorithm which returns an estimator $\widetilde{\boldsymbol{w}}_t^{(D)}$ after t iterations, which satisfies:*

$$\lim_{D\to\infty} \operatorname*{plim\,sup}_{N\to\infty} \frac{\|\widehat{\boldsymbol{r}}_t - \widetilde{\boldsymbol{w}}_t^{(D)}\|^2}{N} = 0.$$

The Optimal Lifted OAMP Algorithm. Since Proposition 3 guarantees that any estimator computed using an iterative algorithm can be approximated by a suitably designed lifted OAMP algorithm, we focus on characterizing the optimal lifted OAMP algorithms. The optimal degree-D lifted OAMP algorithm takes the form:

$$\boldsymbol{w}_{t,i} = \big(\boldsymbol{Y}^i - \mathbb{E}[\Lambda^i] \cdot \boldsymbol{I}_N\big) \cdot f_t^\star\big(\boldsymbol{w}_{1,\bullet}, \ldots, \boldsymbol{w}_{t-1,\bullet}; \boldsymbol{a}\big) \quad \forall\, i \in [D],\ t \in \mathbb{N}, \tag{49a}$$

and returns the following estimator after $t \in \mathbb{N}$ iterations:

$$\widehat{\boldsymbol{w}}_t^{(D)} = h_t^\star(\boldsymbol{w}_{1,\bullet}, \ldots, \boldsymbol{w}_{t-1,\bullet}; \boldsymbol{a}). \tag{49b}$$

The description of the iterate denoisers $(f_t^\star)_{t\in\mathbb{N}}$ and the postprocessing functions $(h_t^\star)_{t\in\mathbb{N}}$ used by the optimal lifted OAMP algorithm is recursive. Suppose that the functions $f_1^\star, \ldots, f_t^\star$ have been specified. Let $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$ denote the state evolution random variables corresponding to the first t iterations of the resulting lifted OAMP algorithm. Then:

1. The post-processing function $h_t^\star$ used at iteration t is the MMSE estimator (recall Definition 3) for the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$.
2. $f_{t+1}^\star$, the iterate denoiser used by the lifted OAMP algorithm in the next iteration is the DMMSE estimator for the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$.

The following proposition (proved in Appendix E.2), shows that the estimator (49) achieves the lowest mean squared error among all estimators that can be computed using t iterations of a degree-D lifted OAMP algorithm.

Proposition 4. *Let $\widehat{\boldsymbol{w}}_t^{(D)}$ be the estimator described in (49). Let $\widetilde{\boldsymbol{w}}_t^{(D)}$ be any other estimator that can be computed using t iterations of some degree-D lifted OAMP algorithm. Then, we have:*

$$\operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|^2}{N} \le \operatorname*{plim}_{N\to\infty} \frac{\|\widetilde{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|^2}{N}.$$

Proof Sketch. Consider a general lifted OAMP algorithm:

$$\boldsymbol{w}_{t,i} = \big(\boldsymbol{Y}^i - \mathbb{E}[\Lambda^i] \cdot \boldsymbol{I}_N\big) \cdot f_t\big(\boldsymbol{w}_{1,\bullet}, \ldots, \boldsymbol{w}_{t-1,\bullet}; \boldsymbol{a}\big) \quad \forall\, i \in [D],\ t \in \mathbb{N},$$

which returns the estimator $\widehat{\boldsymbol{w}}_t^{(D)} = h_t(\boldsymbol{w}_{1,\bullet}, \ldots, \boldsymbol{w}_{t,\bullet}; \boldsymbol{a})$ after t iterations. By Corollary 1, the asymptotic mean squared error for this estimator is $\mathbb{E}|\mathsf{X}_\star - h_t(\mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})|^2$, where $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$ denote the state evolution random variables associated with the lifted OAMP algorithm. Recalling the definition of the MMSE denoiser for a Gaussian channel from Definition 3, we find that the optimal choice for the postprocessing function is $h_t = h_t^\star$, where $h_t^\star$ is the MMSE estimator for the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$. With this optimal choice of the post-processing function, the limiting MSE of the resulting estimator is $\mathrm{MMSE}(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$. This depends implicitly on the iterate denoisers $f_{1:t}$ used in the lifted OAMP algorithm since $f_{1:t}$ determine the joint distribution of the state evolution random variables. To make this dependence explicit, we define $\mathscr{M}_t(f_1, \ldots, f_t) \stackrel{\text{def}}{=} \mathrm{MMSE}(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$ for any $t \in \mathbb{N}$. Deriving the optimal choice of the iterate denoisers is equivalent to minimizing the functional $\mathscr{M}_t(f_1, \ldots, f_t)$ with respect to $f_{1:t}$. The crux of the argument is to show that the iterate denoisers $(f_t^\star)_{t\in\mathbb{N}}$ used by the optimal lifted OAMP algorithm (49) correspond to a *greedy approach* to minimizing $\mathscr{M}_t$:

$$f_1^\star \in \arg\min_{f_1} \mathscr{M}_1(f_1), \quad f_2^\star \in \arg\min_{f_2} \mathscr{M}_2(f_1^\star, f_2), \quad \ldots \quad f_t^\star \in \arg\min_{f_t} \mathscr{M}_t(f_1^\star, \ldots, f_{t-1}^\star, f_t) \quad \forall\, t \in \mathbb{N}. \tag{50}$$

Moreover, we show that the functional $\mathscr{M}_t$ has the property that for any choice of iterate denoisers $f_{1:t-1}$:

$$\min_{f_t} \mathscr{M}_t(f_1, \ldots, f_t) = \mathscr{A}(\mathscr{M}_{t-1}(f_1, \ldots, f_{t-1})), \tag{51}$$

for some explicit *non-decreasing* function $\mathscr{A} : [0,1] \mapsto [0,1]$. This property guarantees that the greedy approach (50) finds the global minimizer of $\mathscr{M}_t$. Indeed, if we assume as our induction hypothesis that $\mathscr{M}_{t-1}(f_{1:t-1}^\star) \le \mathscr{M}_{t-1}(f_{1:t-1})$ for any choice of $f_{1:t-1}$, we can conclude that:

$$\mathscr{M}_t(f_{1:t}) \ge \min_{f_t} \mathscr{M}_t(f_{1:t}) \overset{(51)}{=} \mathscr{A}(\mathscr{M}_{t-1}(f_{1:t-1})) \overset{(a)}{\ge} \mathscr{A}(\mathscr{M}_{t-1}(f_{1:t-1}^\star)) \overset{(51)}{=} \min_{f_t} \mathscr{M}_t(f_{1:t-1}^\star, f_t) \overset{(50)}{=} \mathscr{M}_t(f_{1:t}^\star),$$

where the inequality (a) follows from the induction hypothesis and the monotonicity of $\mathscr{A}$. □

Simplifying the optimal lifted OAMP algorithm. Observe that the optimal degree-D lifted OAMP algorithm introduced in (49) is not an iterative algorithm in the sense of Definition 5. Indeed, an iterative algorithm is allowed a single matrix-vector multiplication per iteration, whereas the optimal lifted OAMP computes D matrix-vector products per iteration. Fortunately, the optimal lifted OAMP algorithm (49) can be simplified to eliminate these extra matrix-vector multiplications. This simplification leads to the following result, which shows that the optimal OAMP algorithm introduced in (17) can match the infinite degree ($D \to \infty$) limit of the performance of the optimal lifted OAMP algorithm in (49).

Proposition 5. *Let $\widehat{\boldsymbol{x}}_t$ be the estimator returned by the optimal OAMP algorithm in (17) after t iterations. Let $\widehat{\boldsymbol{w}}_t^{(D)}$ be the estimator returned by optimal degree-D lifted OAMP algorithm in (49) after t iterations. Then, for any $t \in \mathbb{N}$:*

$$\operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_\star\|^2}{N} = \lim_{D\to\infty} \operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|^2}{N}.$$

Proof Sketch. We show that the estimator $\widehat{\boldsymbol{w}}_t^{(D)}$ returned by the optimal lifted OAMP algorithm (49) after t iterations depends only on a linear combination $\boldsymbol{x}_t^{(D)} = \sum_{i=1}^D v_i \cdot \boldsymbol{w}_{t,i}$ of the iterates $\boldsymbol{w}_{t,\bullet}$ generated at the last step of the algorithm. Thanks to this property, the $\boldsymbol{x}_t^{(D)}$ (and hence, the estimator $\widehat{\boldsymbol{w}}_t$) can be computed using a single (instead of D) matrix-vector multiplication(s):

$$\boldsymbol{x}_t^{(D)} = \sum_{i=1}^D v_i \cdot \boldsymbol{w}_{t,i}^{(D)} \overset{(49)}{=} \underbrace{\left(\sum_{i=1}^D v_i \cdot (\boldsymbol{Y}^i - \mathbb{E}_{\Lambda\sim\mu}[\Lambda^i] \cdot \boldsymbol{I}_N) \right)}_{\overset{\text{def}}{=} \Psi_t^{(D)}(\boldsymbol{Y})} \cdot f_t^\star(\boldsymbol{w}_{<t,\bullet}; \boldsymbol{a}) = \Psi_t^{(D)}(\boldsymbol{Y}) \cdot f_t^\star(\boldsymbol{w}_{<t,\bullet}; \boldsymbol{a}).$$

We argue that as $D \to \infty$ the matrix denoiser $\Psi_t^{(D)}$ that appears in the above equation converges to the optimal matrix denoiser used by the optimal OAMP algorithm (17) in an L^2 sense, and hence the claimed result follows. □

Theorem 2 is an immediate consequence of the three intermediate results introduced above.

Proof of Theorem 2. Let $\widehat{\boldsymbol{x}}_t$ be the estimator returned by the optimal OAMP algorithm from (17) and let $\widehat{\boldsymbol{r}}_t$ be any estimator that can be computed using t iterations of some iterative algorithm. Proposition 3 guarantees the existence of a sequence of estimators $(\widetilde{\boldsymbol{w}}_t^{(D)})_{D\in\mathbb{N}}$ that approximate $\widehat{\boldsymbol{r}}_t$ and can be computed using t iterations of a degree-D lifted OAMP algorithm. Finally, let $\widehat{\boldsymbol{w}}_t^{(D)}$ be the estimator returned by the optimal degree-D lifted OAMP algorithm from (49). By the reverse triangle inequality, for any $D \in \mathbb{N}$, $\|\widehat{\boldsymbol{r}}_t - \boldsymbol{x}_\star\| \geq \|\widetilde{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\| - \|\widetilde{\boldsymbol{w}}_t^{(D)} - \widehat{\boldsymbol{r}}_t\|$. We let $N \to \infty$ and then $D \to \infty$ to obtain:

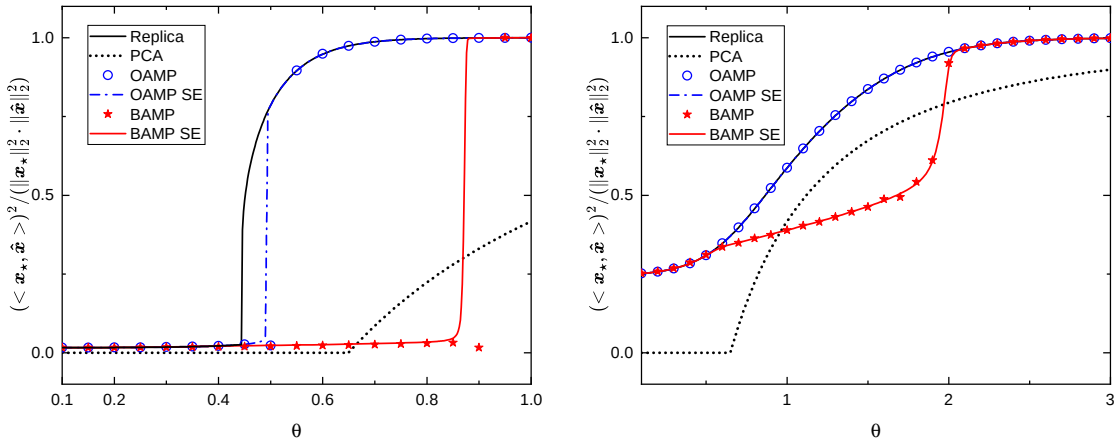
$$\begin{aligned}
\operatorname*{plim\,inf}_{N\to\infty} \frac{\|\widehat{\boldsymbol{r}}_t - \boldsymbol{x}_\star\|}{\sqrt{N}} &\geq \liminf_{D\to\infty} \operatorname*{plim\,inf}_{N\to\infty} \frac{\|\widetilde{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|}{\sqrt{N}} - \limsup_{D\to\infty} \operatorname*{plim\,sup}_{N\to\infty} \frac{\|\widetilde{\boldsymbol{w}}_t^{(D)} - \widehat{\boldsymbol{r}}_t\|}{\sqrt{N}} \\
&\overset{\text{Prop. 3}}{=} \liminf_{D\to\infty} \operatorname*{plim}_{N\to\infty} \frac{\|\widetilde{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|}{\sqrt{N}} \overset{\text{Prop. 4}}{\geq} \liminf_{D\to\infty} \operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|}{\sqrt{N}} \overset{\text{Prop. 5}}{=} \operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_\star\|}{\sqrt{N}}.
\end{aligned}$$

Comparing the first and final expressions in the above chain of inequalities gives us the claim of Theorem 2. □

5 Numerical Experiments

We conclude the paper with some numerical experiments to demonstrate the performance of the optimal OAMP (17) algorithm. The codes to reproduce the results in this section are publicly available [1].

Synthetic Noise Matrices. Following [50, 69], we generate the noise matrix $\boldsymbol{W}$ in a structured manner, making it feasible to efficiently simulate the dynamics of iterative algorithms on very high-dimensional matrices ($N = 2 \times 10^6$). Specifically, we use the noise matrix $\boldsymbol{W} = \boldsymbol{O}\mathrm{diag}(\lambda_1, \ldots, \lambda_N)\boldsymbol{O}^\mathsf{T}$ with $\boldsymbol{O} = \boldsymbol{S}_1\boldsymbol{F}\boldsymbol{S}_2\boldsymbol{F}^\mathsf{T}\boldsymbol{S}_3$, where $\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_3$ are independent random signed diagonal matrices, and $\boldsymbol{F}$ is the discrete cosine transform matrix. The eigenvalues $\{\lambda_i\}_{i\in[N]}$ are sampled independently from the spectral μ corresponding to the quartic noise model (21) with parameter $\gamma = 0$. Known universality results [24, 26, 75] guarantee that using this structured noise matrix (instead of random rotationally invariant noise drawn from the quartic model) does not change the asymptotic dynamics of AMP algorithms.



(a) Two-points prior with $\epsilon = \frac{1}{8}$.

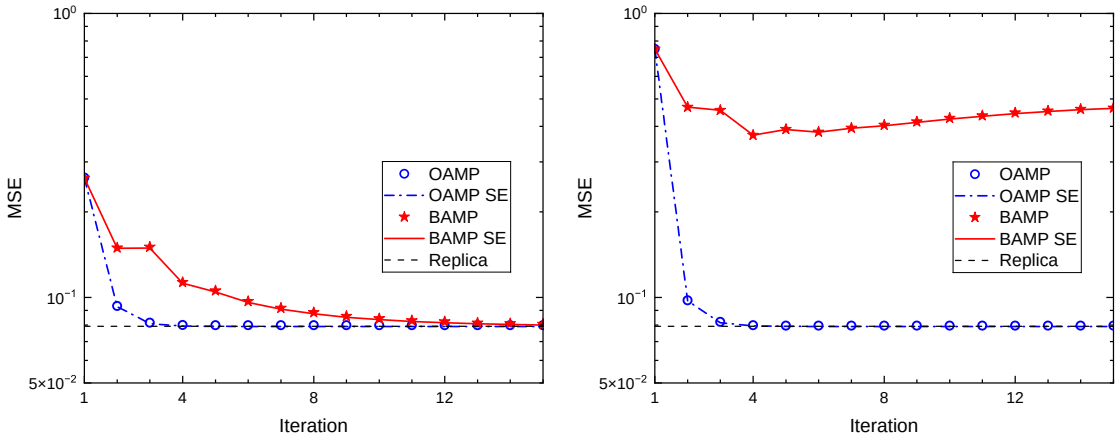
(b) Two-points prior with $\epsilon = \frac{1}{2}$.

Figure 2: Performance (as measured by the normalized overlap with signal) of the PCA estimator, the Bayes-optimal estimator, the optimal OAMP algorithm (15 iterations), and the BAMP algorithm of Barbier et al. [8] (15 iterations) as a function of the SNR θ.

Fig. 2 compares the performance of optimal OAMP with existing methods, including vanilla PCA and the BAMP algorithm introduced by Barbier et al. [8], along with conjectured replica formula (Conjecture 1) for the Bayes-optimal estimator (marked "Replica"). We sampled the signal $\boldsymbol{x}_\star$ from an i.i.d. two-point prior $\epsilon^2\delta_{1/\epsilon} + (1-\epsilon^2)\delta_0$ (no side information is provided). Fig. 2 shows that the performances of the optimal OAMP algorithm and the BAMP algorithm match well with their corresponding state evolution (SE) predictions. The optimal OAMP algorithm outperforms vanilla PCA and the BAMP algorithm in both cases. We also see a statistical-computational gap for this problem for small ϵ. Namely, there is a performance gap between the conjectured Bayes-optimal performance and the best achievable performance by iterative algorithms (attained by the optimal OAMP algorithm). In contrast, no statistical-computational gap exists for the two-point prior with large enough ϵ. A similar phenomenon occurs when the noise matrix is i.i.d. Gaussian [54].

The performance of BAMP depends on the quality of initialization. Fig. 3 compares the dynamics of the optimal OAMP algorithm and BAMP when initialized at $\boldsymbol{x}_0 = \sqrt{\omega_0}\boldsymbol{x}_\star + \sqrt{1-\omega_0}\boldsymbol{z}$ with $\boldsymbol{z} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_N)$. We use the same noise model as before and sample the signal from an i.i.d. two-point prior with $\epsilon = 0.5$ and $\theta = 1.8$. For a good initialization (Fig. 3(a)), the optimal OAMP algorithm and BAMP achieve the same MSE on convergence. However, with uninformative initialization (Fig. 3(b)), the performance of BAMP degrades after a few iterations, and the optimal OAMP algorithm achieves a superior MSE on convergence. Notice that the performance of BAMP is well predicted by its asymptotic state evolution. Hence, the observed behavior of BAMP is not caused by finite-N effects or numerical stability issues and seems to be an inherent limitation of the algorithm.

Realistic Noise Matrices. The theoretical prediction of the asymptotic performance of OAMP requries the noise matrix to be rotationally-invaraint. The noise matrix used in the experiments for Fig. 2 and Fig. 3 is structured but still synthetic. We now consider a more realistic model where the noise matrix is obtained from real datasets. Taking inspiration from [77], we generate the noise matrix from the covariance matrix of

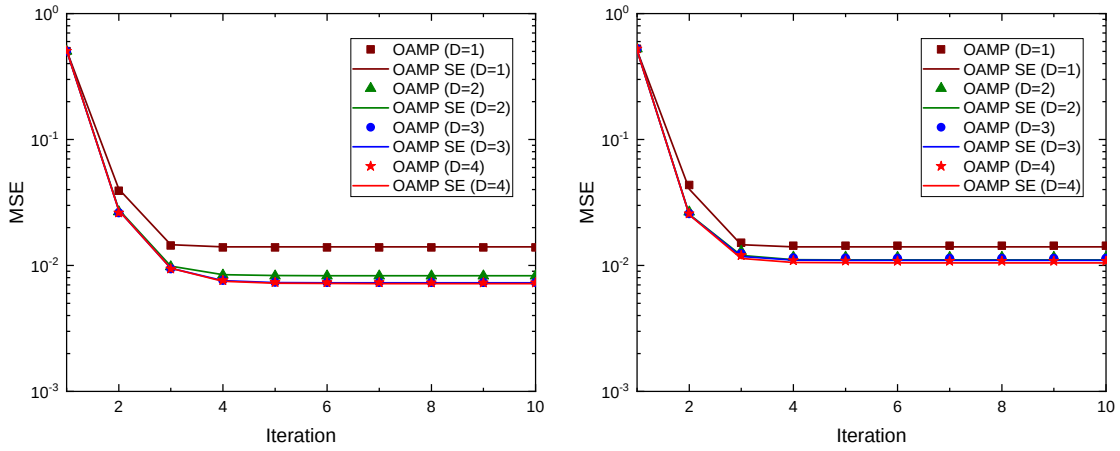


(a) Good initialization: $\omega_0 = 0.6$.

(b) Worse initialization: $\omega_0 = 0$.

Figure 3: MSE performance of the optimal OAMP algorithm and the BAMP algorithm of Barbier et al. [8] as a function of the iteration number compared with the replica conjecture for the Bayes-optimal performance.

real datasets in bioinformatics, namely, the 1000 Genomes Project (1000G) [18] and International HapMap Project (Hapmap3) [19]. Both datasets undergo the preprocessing steps outlined in [77], producing a dataset containing 100,000 common single nucleotide polymorphisms (SNPs) for 2504 individuals for 1000G and a dataset comprising 142,186 SNPs for 1397 individuals for Hapmap3. In our experiments, we randomly select 3000 SNPs and compute the corresponding covariance matrix, resulting in a 2054×2054 symmetric matrix for 1000G and a 1397×1397 symmetric matrix for Hapmap3. We then obtain the noise matrix by further removing the "ground truth" signal components. Following [77], we estimate the "ground truth" signal components by forming covariance matrices using all SNPs from the original datasets (instead of the sub-sampled ones). The noise matrix is obtained by the following pre-processing steps: (1) we remove the "ground truth" signal components; (2) we further extract the principal components that correspond to outlying eigenvalues; (3) We center and scale the data properly.



(a) Real datasets from 1000G [18].

(b) Real datasets from Hapmap3 [19].

Figure 4: Performance of a data-driven implementation of the optimal degree-D lifted OAMP algorithm from (49) on noise matrices derived from real datasets and signals randomly drawn from $\mathsf{X}_\star \sim 0.1\mathcal{N}(0, 10) + 0.9\delta_0$.

Running the optimal OAMP algorithm requires estimating the spectral measure μ and the associated optimal matrix denoiser (17e) from the observed data. While feasible, we found it more convenient to design a data-driven implementation of the optimal degree-D lifted OAMP algorithm introduced in (49). Fig. 4 displays the MSE performance of this algorithm using the above-described realistic noise matrices and a randomly generated signal. The performance of the lifted OAMP improves as the degree D increases, saturating around $D \approx 4$, indicating that this algorithm can match the performance of the optimal OAMP algorithm in a data-driven way at the cost of 3 additional matrix multiplications per iteration. Fig. 4 also displays the asymptotic MSE predicted by the state evolution result for the corresponding rotationally invariant noise model with a matching spectrum. The performance of the lifted OAMP algorithm closely aligns with the state evolution prediction, suggesting an underlying universality phenomenon. We have observed that the pre-processing steps used are critical for universality. Understanding the precise conditions for universality (or its absence) is an interesting future direction.

References

[1] Accompanying code for optimality of message passing algorithms for spiked matrix models with rotationally invariant noise. `https://github.com/songIce/OAMP`.

[2] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

[3] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Cambridge University Press, 2010.

[4] Jinho Baik and Jack W Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.

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

[6] Jean Barbier and Nicolas Macris. The adaptive interpolation method for proving replica formulas. applications to the Curie–Weiss and Wigner spike models. *Journal of Physics A: Mathematical and Theoretical*, 52(29):294002, 2019.

[7] Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, and Lenka Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. *Advances in Neural Information Processing Systems*, 29, 2016.

[8] Jean Barbier, Francesco Camilli, Marco Mondelli, and Manuel Sáenz. Fundamental limits in structured principal component analysis and how to reach them. *Proceedings of the National Academy of Sciences*, 120(30):e2302028120, 2023.

[9] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

[10] Joshua K Behne and Galen Reeves. Fundamental limits for rank-one matrix estimation with groupwise heteroskedasticity. In *International Conference on Artificial Intelligence and Statistics*, pages 8650–8672. PMLR, 2022.

[11] Serban Teodor Belinschi. The Lebesgue decomposition of the free additive convolution of two probability distributions. *Probability Theory and Related Fields*, 142(1):125–150, 2008.

[12] Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.

[13] Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model. *Communications in Mathematical Physics*, 325(1):333–366, 2014.

[14] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford University Press, 2013.

[15] Joël Bun, Romain Allez, Jean-Philippe Bouchaud, and Marc Potters. Rotational invariant estimator for general noisy matrices. *IEEE Transactions on Information Theory*, 62(12):7475–7490, 2016.

[16] Michael Celentano, Andrea Montanari, and Yuchen Wu. The estimation error of general first order methods. In *Conference on Learning Theory*, pages 1078–1141. PMLR, 2020.

[17] Michael Celentano, Zhou Fan, and Song Mei. Local convexity of the TAP free energy and AMP convergence for $\mathbb{Z}_2$-synchronization. *The Annals of Statistics*, 51(2):519–546, 2023.

[18] 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*, 526(7571): 68, 2015.

[19] International HapMap 3 Consortium. Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311):52, 2010.

[20] Yash Deshpande and Andrea Montanari. Information-theoretically optimal sparse PCA. In *2014 IEEE International Symposium on Information Theory*, pages 2197–2201. IEEE, 2014.

[21] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the balanced binary stochastic block model. *Information and Inference: A Journal of the IMA*, 6(2):125–170, 2017.

[22] David Donoho and Jared Tanner. Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1906):4273–4293, 2009.

[23] David L Donoho, Arian Maleki, and Andrea Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[24] Rishabh Dudeja and Milad Bakhshizadeh. Universality of linearized message passing for phase retrieval with structured sensing matrices. *IEEE Transactions on Information Theory*, 68(11):7545–7574, 2022.

[25] Rishabh Dudeja, Subhabrata Sen, and Yue M Lu. Spectral universality of regularized linear regression with nearly deterministic sensing matrices. *arXiv preprint arXiv:2208.02753*, 2022.

[26] Rishabh Dudeja, Yue M. Lu, and Subhabrata Sen. Universality of approximate message passing with semirandom matrices. *The Annals of Probability*, 51(5):1616–1683, 2023.

[27] Ahmed El Alaoui and Florent Krzakala. Estimation in the spiked Wigner model: a short proof of the replica formula. In *2018 IEEE International Symposium on Information Theory (ISIT)*, pages 1874–1878. IEEE, 2018.

[28] Ahmed El Alaoui, Florent Krzakala, and Michael Jordan. Fundamental limits of detection in the spiked Wigner model. *The Annals of statistics*, 48(2):863–885, 2020.

[29] Zhou Fan. Approximate message passing algorithms for rotationally invariant matrices. *The Annals of Statistics*, 50(1):197–224, 2022.

[30] Zhou Fan, Song Mei, and Andrea Montanari. TAP free energy, spin glasses and variational inference. *The Annals of Probability*, 49(1), 2021.

[31] Oliver Y Feng, Ramji Venkataramanan, Cynthia Rush, and Richard J Samworth. A unifying tutorial on approximate message passing. *Foundations and Trends® in Machine Learning*, 15(4):335–536, 2022.

[32] Delphine Féral and Sandrine Péché. The largest eigenvalue of rank one deformation of large Wigner matrices. *Communications in Mathematical Physics*, 272:185–228, 2007.

[33] Dongning Guo, Yihong Wu, Shlomo S Shitz, and Sergio Verdú. Estimation in Gaussian noise: Properties of the minimum mean-square error. *IEEE Transactions on Information Theory*, 57(4):2371–2385, 2011.

[34] Peter Henrici. *Applied and computational complex analysis, Volume 3: Discrete Fourier analysis, Cauchy integrals, construction of conformal maps, univalent functions*, volume 41. John Wiley & Sons, 1993.

[35] Iain M Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics*, 29(2):295–327, 2001.

[36] Iain M Johnstone and Arthur Yu Lu. On consistency and sparsity for principal components analysis in high dimensions. *Journal of the American Statistical Association*, 104(486):682–693, 2009.

[37] Yoshiyuki Kabashima. A cdma multiuser detection algorithm on the basis of belief propagation. *Journal of Physics A: Mathematical and General*, 36(43):11111, 2003.

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

[39] Antti Knowles and Jun Yin. The isotropic semicircle law and deformation of Wigner matrices. *Communications on Pure and Applied Mathematics*, 66(11):1663–1749, 2013.

[40] Florent Krzakala, Jiaming Xu, and Lenka Zdeborová. Mutual information in rank-one matrix estimation. In *2016 IEEE Information Theory Workshop (ITW)*, pages 71–75. IEEE, 2016.

[41] Olivier Ledoit and Sandrine Péché. Eigenvectors of some large sample covariance matrix ensembles. *Probability Theory and Related Fields*, 151(1):233–264, 2011.

[42] Olivier Ledoit and Michael Wolf. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics*, 40(2):1024–1060, 2012.

[43] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. In *Conference on Learning Theory*, pages 1297–1301. PMLR, 2017.

[44] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Mmse of probabilistic low-rank matrix estimation: Universality with respect to the output channel. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 680–687. IEEE, 2015.

[45] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Phase transitions in sparse PCA. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 1635–1639. IEEE, 2015.

[46] Gen Li and Yuting Wei. A non-asymptotic framework for approximate message passing in spiked models. *arXiv preprint arXiv:2208.03313*, 2022.

[47] Gen Li, Wei Fan, and Yuting Wei. Approximate message passing from random initialization with applications to z2 synchronization. *Proceedings of the National Academy of Sciences*, 120(31):e2302930120, 2023.

[48] Yufan Li and Pragya Sur. Spectrum-aware adjustment: A new debiasing framework with applications to principal components regression. *arXiv preprint arXiv:2309.07810*, 2023.

[49] Panagiotis Lolas and Lexing Ying. Shrinkage estimation of functions of large noisy symmetric matrices. *arXiv preprint arXiv:2106.05183*, 2021.

[50] Junjie Ma and Li Ping. Orthogonal AMP. *IEEE Access*, 5:2020–2033, 2017.

[51] James A Mingo and Roland Speicher. *Free probability and random matrices*, volume 35. Springer, 2017.

[52] Léo Miolane. Fundamental limits of low-rank matrix estimation: the non-symmetric case. *arXiv preprint arXiv:1702.00473*, 2017.

[53] Marco Mondelli and Ramji Venkataramanan. PCA initialization for approximate message passing in rotationally invariant models. *Advances in Neural Information Processing Systems*, 34:29616–29629, 2021.

[54] Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. *The Annals of Statistics*, 49(1), 2021.

[55] Andrea Montanari and Alexander S Wein. Equivalence of approximate message passing and low-degree polynomials in rank-one matrix estimation. *arXiv preprint arXiv:2212.06996*, 2022.

[56] Andrea Montanari and Yuchen Wu. Statistically optimal first order algorithms: A proof via orthogonalization. *arXiv preprint arXiv:2201.05101*, 2022.

[57] Manfred Opper and Ole Winther. Adaptive and self-averaging Thouless-Anderson-Palmer mean-field theory for probabilistic modeling. *Physical Review E*, 64(5):056131, 2001.

[58] Manfred Opper, Burak Cakmak, and Ole Winther. A theory of solving TAP equations for Ising models with general invariant random matrices. *Journal of Physics A: Mathematical and Theoretical*, 49(11): 114002, 2016.

[59] Samet Oymak and Babak Hassibi. A case for orthogonal measurements in linear inverse problems. In *2014 IEEE International Symposium on Information Theory*, pages 3175–3179. IEEE, 2014.

[60] Jason T Parker, Philip Schniter, and Volkan Cevher. Bilinear generalized approximate message passing—part i: Derivation. *IEEE Transactions on Signal Processing*, 62(22):5839–5853, 2014.

[61] Leonid Andreevich Pastur and Mariya Shcherbina. *Eigenvalue distribution of large random matrices*. American Mathematical Soc., 2011.

[62] Debashis Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

[63] Sandrine Péché. The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probability Theory and Related Fields*, 134:127–173, 2006.

[64] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Message-passing algorithms for synchronization problems over compact groups. *Communications on Pure and Applied Mathematics*, 71(11):2275–2322, 2018.

[65] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra. Optimality and sub-optimality of PCA i: Spiked random matrix models. *The Annals of Statistics*, 46(5):2416–2451, 2018.

[66] Marc Potters and Jean-Philippe Bouchaud. *A First Course in Random Matrix Theory: For Physicists, Engineers and Data Scientists*. Cambridge University Press, 2020.

[67] Farzad Pourkamali and Nicolas Macris. Rectangular rotational invariant estimator for general additive noise matrices. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 2081–2086. IEEE, 2023.

[68] Sundeep Rangan and Alyson K Fletcher. Iterative estimation of constrained rank-one matrices in noise. In *2012 IEEE International Symposium on Information Theory Proceedings*, pages 1246–1250. IEEE, 2012.

[69] Sundeep Rangan, Philip Schniter, and Alyson K Fletcher. Vector approximate message passing. *IEEE Transactions on Information Theory*, 65(10):6664–6684, 2019.

[70] Konrad Schmüdgen. *The moment problem*, volume 9. Springer, 2017.

[71] Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree polynomials. *The Annals of Statistics*, 50(3):1833–1858, 2022.

[72] Guilhem Semerjian. Matrix denoising: Bayes-optimal estimators via low-degree polynomials. *arXiv preprint arXiv:2402.16719*, 2024.

[73] Amit Singer. Angular synchronization by eigenvectors and semidefinite programming. *Applied and computational harmonic analysis*, 30(1):20–36, 2011.

[74] Keigo Takeuchi. Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements. *IEEE Transactions on Information Theory*, 66(1):368–386, 2019.

[75] Tianhao Wang, Xinyi Zhong, and Zhou Fan. Universality of approximate message passing algorithms and tensor networks. *arXiv preprint arXiv:2206.13037*, 2022.

[76] Xinyi Zhong, Tianhao Wang, and Zhou Fan. Approximate message passing for orthogonally invariant ensembles: Multivariate non-linearities and spectral initialization. *arXiv preprint arXiv:2110.02318*, 2021.

[77] Xinyi Zhong, Chang Su, and Zhou Fan. Empirical Bayes PCA in high dimensions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(3):853–878, 2022.

[78] Hui Zou, Trevor Hastie, and Robert Tibshirani. Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2):265–286, 2006.

A Proofs for Preliminary Results

A.1 Proofs for Random Matrix Theory Results

This section is devoted to the proof of Lemma 1.

Proof of Lemma 1. We consider each of the claims made in the lemma one by one.

Weak convergence of ν_N (Claim (1)). Let $\widetilde{\nu}_N$ be a slightly modified version of ν_N:

$$\widetilde{\nu}_N = \frac{1}{N}\sum_{i=1}^{N} \left(\widetilde{\boldsymbol{x}}_\star^\mathsf{T}\boldsymbol{u}_i\right)^2 \cdot \delta_{\lambda_i(\boldsymbol{Y})}$$

where $\widetilde{\boldsymbol{x}}_\star \stackrel{\text{def}}{=} \frac{\sqrt{N}\boldsymbol{x}_\star}{\|\boldsymbol{x}_\star\|}$. This modification makes $\widetilde{\nu}_N$ a probability measure, i.e., $\widetilde{\nu}_N(\mathbb{R}) = 1$. Notice that for any bounded, continuous function $f : \mathbb{R} \mapsto \mathbb{R}$, we have:

$$\int_{\mathbb{R}} f(\lambda)\, \widetilde{\nu}_N(\lambda) = \frac{N}{\|\boldsymbol{x}_\star\|^2} \cdot \int_{\mathbb{R}} f(\lambda)\, \nu_N(\lambda).$$

By Assumption 1, $\|\boldsymbol{x}_\star\|^2/N \xrightarrow{\mathbb{P}} 1$. Hence, it suffices to show that $\widetilde{\nu}_N$ converges weakly to a probability measure ν in probability. To do so, it suffices to verify that the Stieltjes transform $\mathcal{S}_{\widetilde{\nu}_N}(z)$ of $\widetilde{\nu}_N$ converges pointwise to the Stieltjes transform of a probability measure ν for any $z \in \mathbb{C}\backslash\mathbb{R}$ (see for e.g., [3, Theorem 2.4.4]). To this end, we note that:

$$\mathcal{S}_{\widetilde{\nu}_N}(z) = \frac{1}{\|\boldsymbol{x}_\star\|^2}\sum_{i=1}^{N}\frac{\langle \boldsymbol{u}_i(\boldsymbol{Y}), \boldsymbol{x}_\star\rangle^2}{z - \lambda_i(\boldsymbol{Y})} = \frac{1}{\|\boldsymbol{x}_\star\|^2}\cdot \boldsymbol{x}_\star^\mathsf{T}\left(z\boldsymbol{I}_N - \boldsymbol{Y}\right)^{-1}\boldsymbol{x}_\star,$$

where $\lambda_{1:N}(\boldsymbol{Y})$ and $\boldsymbol{u}_{1:N}(\boldsymbol{Y})$ denote the eigenvalues and eigenvectors of $\boldsymbol{Y}$. Hence,

$$\mathcal{S}_{\widetilde{\nu}_N}(z) = \frac{1}{\|\boldsymbol{x}_\star\|^2}\boldsymbol{x}_\star^\mathsf{T}\left(z\boldsymbol{I}_N - \boldsymbol{Y}\right)^{-1}\boldsymbol{x}_\star = \frac{1}{\|\boldsymbol{x}_\star\|^2}\boldsymbol{x}_\star^\mathsf{T}\left(z\boldsymbol{I}_N - \frac{\theta\boldsymbol{x}_\star\boldsymbol{x}_\star^\mathsf{T}}{N} - \boldsymbol{W}\right)^{-1}\boldsymbol{x}_\star \stackrel{\text{(a)}}{=} \frac{\frac{1}{\|\boldsymbol{x}_\star\|^2}\boldsymbol{x}_\star^\mathsf{T}\left(z\boldsymbol{I}_N - \boldsymbol{W}\right)^{-1}\boldsymbol{x}_\star}{1 - \frac{\theta}{N}\boldsymbol{x}_\star^\mathsf{T}\left(z\boldsymbol{I}_N - \boldsymbol{W}\right)^{-1}\boldsymbol{x}_\star}, \tag{52}$$

where step (a) uses the Sherman-Morrison formula for the rank-one update to the matrix inverse. Using standard concentration results regarding quadratic forms of rotationally invariant matrices (stated as Fact 2 in Appendix F for convenience), we find that:

$$\frac{\boldsymbol{x}_\star^\mathsf{T}\left(z\boldsymbol{I}_N-\boldsymbol{W}\right)^{-1}\boldsymbol{x}_\star}{\|\boldsymbol{x}_\star\|^2}\xrightarrow{\mathbb{P}}\int_{\mathbb{R}}\frac{\mu(\mathrm{d}\lambda)}{z-\lambda}=\mathcal{S}_\mu(z)\quad\forall\, z\in\mathbb{C}\backslash\mathbb{R}. \tag{53}$$

By Assumption 1, $\|\boldsymbol{x}_\star\|^2/N\xrightarrow{\mathbb{P}}1$. Hence,

$$1-\frac{\theta}{N}\boldsymbol{x}_\star^\mathsf{T}\left(z\boldsymbol{I}_N-\boldsymbol{W}\right)^{-1}\boldsymbol{x}_\star=1-\theta\cdot\frac{\|\boldsymbol{x}_\star\|^2}{N}\cdot\frac{\boldsymbol{x}_\star^\mathsf{T}\left(z\boldsymbol{I}_N-\boldsymbol{W}\right)^{-1}\boldsymbol{x}_\star}{\|\boldsymbol{x}_\star\|^2}\cdot\xrightarrow{\mathbb{P}}1-\theta\mathcal{S}_\mu(z)\overset{\text{(a)}}{\neq}0\quad\forall\, z\in\mathbb{C}\backslash\mathbb{R}, \tag{54}$$

where the non-equality marked (a) follows by observing that $\Im(\mathcal{S}_\mu(z))\neq 0$ whenever $\mathrm{Im}(z)\neq 0$, and so $1-\theta\mathcal{S}_\mu(z)\neq 0$. Plugging (53) and (54) into (52), we obtain:

$$\mathcal{S}_{\widetilde{\nu}_N}(z)\xrightarrow{\mathbb{P}}\frac{\mathcal{S}_\mu(z)}{1-\theta\mathcal{S}_\mu(z)}\quad\forall\, z\in\mathbb{C}\backslash\mathbb{R}. \tag{55}$$

To prove the weak convergence of $\widetilde{\nu}_N$ to a *probability measure* ν, we need to show additionally that the RHS of the above equation is the Stieltjes transform of some probability measure. By [51, Theorem 10, Chapter 3], it is sufficient to verify that:

$$\limsup_{y\to\infty}y\left|\frac{\mathcal{S}_\mu(\imath y)}{1-\theta\mathcal{S}_\mu(\imath y)}\right|=1. \tag{56}$$

Since μ is a probability measure, we have $y|\mathcal{S}_\mu(\imath y)|\to 1$ and $\mathcal{S}_\mu(\imath y)\to 0$ as $y\to\infty$ (see [51, Lemma 3, Chapter 3]). Hence, (56) holds. Consequently, the RHS of (55) is a Stieltjes transform of a probability measure ν and $\widetilde{\nu}_N$ and ν_N converge weakly to ν in probability.

Lebesgue Decomposition of ν (Claims (2) and (3)). Belinschi [11, Lemma 2.17] has shown that for any probability measure χ on $\mathbb{R}$ with Lebesgue decomposition $\chi=\chi_\parallel+\chi_\perp$, the Stieltjes transform of ν satisfies:

$$\lim_{\epsilon\downarrow 0}\mathcal{S}_\chi(\lambda+\imath\epsilon)=-\infty\quad\text{for }\chi_\perp\text{-almost every }\lambda\in\mathbb{R}, \tag{57}$$

$$\lim_{\epsilon\downarrow 0}\Im[\mathcal{S}_\chi(x+\imath\epsilon)]=-\pi\frac{\mathrm{d}\chi_\parallel}{\mathrm{d}\lambda}(\lambda)\quad\text{for Lebesgue-almost every }\lambda\in\mathbb{R}. \tag{58}$$

In the display above, $\frac{\mathrm{d}\chi_\parallel}{\mathrm{d}\lambda}$ denotes the density of $\chi_\parallel$ with respect to the Lebesgue measure and the limit on the LHS of (58) is guaranteed to exist and is finite for Lebesgue-almost every λ. In light of this result, we study the limit:

$$\lim_{\epsilon\downarrow 0}\Im[\mathcal{S}_\nu(\lambda+\imath\epsilon)]$$

for an arbitrary $\lambda\in\mathbb{R}$. For any $\epsilon>0$, we have:

$$\Im[\mathcal{S}_\nu(\lambda+\imath\epsilon)]=\Im\left[\frac{\mathcal{S}_\mu(\lambda+\imath\epsilon)}{1-\theta\mathcal{S}_\mu(\lambda+\imath\epsilon)}\right]=\frac{\Im[\mathcal{S}_\mu(\lambda+\imath\epsilon)]}{(1-\theta\Re[\mathcal{S}_\mu(\lambda+\imath\epsilon)])^2+\theta^2\Im^2[\mathcal{S}_\mu(\lambda+\imath\epsilon)]}.$$

Since μ is absolutely continuous with respect to the Lebesgue measure (Assumption 2), (58) implies:

$$\Im[\mathcal{S}_\mu(\lambda+\imath\epsilon)]\to-\pi\mu(\lambda)\quad\text{as }\epsilon\downarrow 0.$$

where $\mu(\cdot)$ denotes the density of μ. Moreover, by the Sohotsky-Plemelj formula (see for instance, [61, Section 2.1]),

$$\Re[\mathcal{S}_\mu(\lambda+\imath\epsilon)]\to\pi\mathscr{H}_\mu(\lambda)\quad\text{as }\epsilon\downarrow 0.$$

Recall that:

$$\phi(\lambda) \stackrel{\text{def}}{=} (1 - \theta\pi\mathscr{H}_\mu(\lambda))^2 + \theta^2\pi^2 \cdot \mu(\lambda).$$

Hence,

$$\lim_{\epsilon\downarrow 0} \Im[\mathscr{H}_\nu(\lambda + \imath\epsilon)] = -\infty \quad \text{iff } \phi(\lambda) = 0,$$
$$\lim_{\epsilon\downarrow 0} \Im[\mathscr{H}_\nu(\lambda + \imath\epsilon)] = -\frac{\pi\mu(\lambda)}{\phi(\lambda)} \in (-\infty, 0] \quad \text{iff } \phi(\lambda) \neq 0.$$

Combining the above results with (57) and (58), we conclude that:

$$\begin{aligned} \phi(\lambda) &\neq 0 \quad \text{for Lebesgue-almost every } \lambda \in \mathbb{R}., \\ \phi(\lambda) &= 0 \quad \text{for } \nu_\perp\text{-almost every } \lambda \in \mathbb{R}, \\ \frac{\mathrm{d}\nu_\parallel}{\mathrm{d}\lambda}(\lambda) &= \frac{\mu(\lambda)}{\phi(\lambda)} \quad \text{for Lebesgue-almost every } \lambda \in \mathbb{R}, \end{aligned}$$

as claimed. This concludes the proof of the lemma. □

A.2 Preliminaries on Gaussian Channels

We state and prove some important properties of Gaussian channels (Definition 3) in this appendix.

The DMMSE Estimator for Scalar Gaussian Channels. The following lemma, due to Ma and Ping [50] provides a formula for the DMMSE estimator for a scalar Gaussian channel with SNR ω. We provide a proof of this result in Section A.2.1 for completeness.

Lemma 2 (Ma and Ping [50])**.** *Let $(\mathsf{X}_\star, \mathsf{X}; \mathsf{A})$ be a scalar Gaussian channel with SNR $\omega \in [0, 1]$. Then,*

1. *The function $\bar{\varphi}(\cdot|\omega) : \mathbb{R}^{1+k} \mapsto \mathbb{R}$:*

$$\bar{\varphi}(x; a|\omega) \stackrel{def}{=} \begin{cases} \left(1 - \frac{\sqrt{\omega}}{\sqrt{1-\omega}} \cdot \mathbb{E}[\mathsf{Z}\varphi(\mathsf{X}; \mathsf{A}|\omega)]\right)^{-1} \cdot \left(\varphi(x; a|\omega) - \frac{\mathbb{E}[\mathsf{Z}\varphi(\mathsf{X};\mathsf{A}|\omega)]}{\sqrt{1-\omega}} \cdot x\right) & : \omega < 1 \\ x & : \omega = 1 \end{cases} \quad \forall\, x \in \mathbb{R},\ a \in \mathbb{R}^k. \tag{59}$$

   is the DMMSE estimator for $(\mathsf{X}_\star, \mathsf{X}; \mathsf{A})$.

2. *The DMMSE estimator satisfies the identities:*

$$\mathbb{E}[\mathsf{X}_\star\bar{\varphi}(\mathsf{X}; \mathsf{A}|\omega)] = \mathbb{E}[\varphi(\mathsf{X}; \mathsf{A}|\omega)\bar{\varphi}(\mathsf{X}; \mathsf{A}|\omega)] = \mathbb{E}[\bar{\varphi}(\mathsf{X}; \mathsf{A}|\omega)^2] = 1 - \mathsf{dmmse}_\pi(\omega).$$

3. *When $\omega < 1$, the mmse_π and dmmse_π functions are related by the identity:*

$$\frac{1}{\mathsf{dmmse}_\pi(\omega)} = \frac{1}{\mathsf{mmse}_\pi(\omega)} - \frac{\omega}{1-\omega}.$$

MMSE and DMMSE Estimators for General Gaussian Channels. Next, we provide formulas for the MMSE and DMMSE estimators for general (multivariate) Gaussian channels. Consider a Gaussian channel $(\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})$:

$$(\mathsf{X}_\star, \mathsf{A}) \sim \pi, \quad (Z_1, \ldots, Z_t) \sim \mathcal{N}(0, \Sigma), \quad \mathsf{X}_i = \alpha_i \cdot \mathsf{X}_\star + Z_i \quad \forall\, i \in [t],$$

where the Gaussian noise $(Z_1, \ldots, Z_t)$ is sampled independently of the signal and side information $(\mathsf{X}_\star; \mathsf{A})$. Observe that for any vector $v \in \mathbb{R}^t$ which satisfies:

$$\langle v, \alpha\rangle^2 + v^\top \Sigma v = 1, \tag{60}$$

$(\mathsf{X}_\star, \langle v, \mathsf{X}_{\le t}\rangle\,; \mathsf{A})$ forms a scalar Gaussian channel with SNR $\omega = \langle v, \alpha\rangle^2$. Among all vectors $v \in \mathbb{R}^t$ which satisfy (60), the maximum SNR for the corresponding scalar channel is achieved by the vector:

$$v_{\mathrm{opt}}(\mathsf{X}_\star|\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A}) \stackrel{\mathrm{def}}{=} \begin{cases} \left(\alpha^\top\Sigma^\dagger\alpha + (\alpha^\top\Sigma^\dagger\alpha)^2\right)^{-\frac{1}{2}} \cdot \Sigma^\dagger\alpha & \text{if } \alpha \in \mathrm{Range}(\Sigma), \\ \langle P_\perp[\alpha], \alpha\rangle^{-1} \cdot P_\perp[\alpha] & \text{if } \alpha \notin \mathrm{Range}(\Sigma), \end{cases} \tag{61a}$$

where $\Sigma^\dagger$ denotes the pseudo-inverse of Σ, $\mathrm{Range}(\Sigma)$ is the range space of Σ, and $P_\perp[\cdot]$ denotes the projector on the orthogonal complement of $\mathrm{Range}(\Sigma)$. We will refer to $v_{\mathrm{opt}}(\mathsf{X}_\star|\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})$ as the *optimal linear combination* for the Gaussian channel $(\mathsf{X}_\star, \mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})$. The maximum SNR is given by:

$$\omega_{\mathrm{eff}}(\mathsf{X}_\star|\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A}) \stackrel{\mathrm{def}}{=} \langle \alpha, v_{\mathrm{opt}}(\mathsf{X}_\star|\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})\rangle^2 = \begin{cases} \left(1 + \alpha^\top\Sigma^\dagger\alpha\right)^{-1} \cdot \alpha^\top\Sigma^\dagger\alpha & \text{if } \alpha \in \mathrm{Range}(\Sigma), \\ 1 & \text{if } \alpha \notin \mathrm{Range}(\Sigma), \end{cases} \tag{61b}$$

which we call the *effective SNR* of the Gaussian channel. The following lemma shows that the MMSE and DMMSE of the Gaussian channel $(\mathsf{X}_\star, \mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})$ is same as the MMSE and DMMSE of the scalar Gaussian channel $(\mathsf{X}_\star, \langle \mathsf{X}_{\le t}, v_{\mathrm{opt}}\rangle\,;\mathsf{A})$, which operates at the SNR ω_{eff}.

Lemma 3. *Let $(\mathsf{X}_\star, \mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})$ be a Gaussian channel. Let v_{opt} and ω_{eff} denote the optimal linear combination and effective SNR for the Gaussian channel, as defined in (61). Then:*

1. $\mathrm{MMSE}(\mathsf{X}_\star|\mathsf{X}_1,\dots,\mathsf{X}_t,\mathsf{A}) = \mathsf{mmse}_\pi(\omega_{\mathrm{eff}})$ *and the function:*

$$f_\star(x;a) \stackrel{def}{=} \varphi(\langle x, v_{\mathrm{opt}}\rangle\,;a|\omega_{\mathrm{eff}}) \quad \forall\, x \in \mathbb{R}^t, \quad a \in \mathbb{R}^k,$$

is the MMSE estimator:

$$f_\star \in \arg\min_{f \in L^2(\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})} \mathbb{E}[\{\mathsf{X}_\star - f(\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})\}^2].$$

2. $\mathrm{DMMSE}(\mathsf{X}_\star|\mathsf{X}_1,\dots,\mathsf{X}_t,\mathsf{A}) = \mathsf{dmmse}_\pi(\omega_{\mathrm{eff}})$ *and the function:*

$$\bar{f}_\star(x;a) \stackrel{def}{=} \bar{\varphi}(\langle x, v_{\mathrm{opt}}\rangle\,;a|\omega_{\mathrm{eff}}) \quad \forall\, x \in \mathbb{R}^t, \quad a \in \mathbb{R}^k,$$

is the DMMSE estimator:

$$\bar{f}_\star \in \arg\min_{f \in L^2(\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})} \mathbb{E}[\{\mathsf{X}_\star - f(\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})\}^2] \quad \textit{subject to } \mathbb{E}[\mathsf{Z}_i f(\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})] = 0\ \forall\, i \in [t].$$

Proof. The proof is deferred to Section A.2.2. □

Monotonicity of MMSE and DMMSE. Lastly, we will rely on the following natural monotonicity property of the mmse_π function.

Fact 1 (Guo et al. 33, Proposition 9). The function $\mathsf{mmse}_\pi : [0,1] \mapsto [0,1]$ is non-increasing. If $\mathsf{mmse}_\pi(0) = \mathbb{E}[\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}]] \neq 0$, $\mathsf{mmse}_\pi : [0,1] \mapsto [0,1]$ is strictly decreasing.

The following lemma, whose proof is provided in Section A.2.3, shows that the dmmse_π function is also non-decreasing.

Lemma 4. *The function $\mathsf{dmmse}_\pi : [0,1] \mapsto [0,1]$ is non-increasing.*

The remainder of this section is devoted to the proofs of Lemma 2, Lemma 3, and Lemma 4 introduced above.

A.2.1 Proof of Lemma 2

Proof of Lemma 2. This result is due to Ma and Ping [50], we provide a proof here for completeness. We consider a scalar Gaussian channel at SNR ω:

$$(\mathsf{X}_\star;\mathsf{A}) \sim \pi, \quad \mathsf{Z} \sim \mathcal{N}(0,1), \quad \mathsf{X} = \sqrt{\omega}\cdot \mathsf{X}_\star + \sqrt{1-\omega}\cdot \mathsf{Z}, \tag{62}$$

where the Gaussian noise Z is sampled independently of the signal and side information $(\mathsf{X}_\star;\mathsf{A})$. Recall from Definition 3 that:

$$\mathsf{dmmse}_\pi(\omega) \stackrel{\text{def}}{=} \mathrm{DMMSE}(\mathsf{X}_\star|\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A}) \stackrel{\text{def}}{=} \min_{f\in L^2(\mathsf{X};\mathsf{A})} \mathbb{E}[\{\mathsf{X}_\star - f(\mathsf{X};\mathsf{A})\}^2] \quad \text{subject to } \mathbb{E}[\mathsf{Z}f(\mathsf{X};\mathsf{A})]=0. \tag{63}$$

Moreover, the function that achieves the lowest MSE in (63) is the DMMSE estimator for the channel. Let f be any function in $L^2(\mathsf{X};\mathsf{A})$ which satisfies the divergence-free requirement $\mathbb{E}[\mathsf{Z}f(\mathsf{X};\mathsf{A})]=0$. We can expand the mean squared error (MSE) of f as follows:

$$\mathbb{E}[(\mathsf{X}_\star - f(\mathsf{X};\mathsf{A}))^2] = 1 + \mathbb{E}[f^2(\mathsf{X};\mathsf{A})] - 2\mathbb{E}[\mathsf{X}_\star f(\mathsf{X};\mathsf{A})] = 1 + \mathbb{E}[f^2(\mathsf{X};\mathsf{A})] - 2\mathbb{E}[\varphi(\mathsf{X};\mathsf{A})f(\mathsf{X};\mathsf{A})], \tag{64}$$

where we used the tower property in the last step. For convenience, we define:

$$\beta \stackrel{\text{def}}{=} \frac{\mathbb{E}[\mathsf{Z}\varphi(\mathsf{X};\mathsf{A}|\omega)]}{\sqrt{1-\omega}}, \quad \widehat{\varphi}(x;a|\omega) \stackrel{\text{def}}{=} \varphi(x;a|\omega) - \beta x. \tag{65}$$

Observe that $\widehat{\varphi}$ satisfies the divergence-free requirement:

$$\mathbb{E}[\mathsf{Z}\widehat{\varphi}(\mathsf{X};\mathsf{A}|\omega)] = 0. \tag{66}$$

We consider the cross-term in (64):

$$\begin{aligned}
\mathbb{E}[\varphi(\mathsf{X};\mathsf{A}|\omega)f(\mathsf{X};\mathsf{A})] &= \mathbb{E}[\widehat{\varphi}(\mathsf{X};\mathsf{A}|\omega)f(\mathsf{X};\mathsf{A})] + \beta\mathbb{E}[\mathsf{X}f(\mathsf{X};\mathsf{A})] \\
&\stackrel{(a)}{=} \mathbb{E}[\widehat{\varphi}(\mathsf{X};\mathsf{A}|\omega)f(\mathsf{X};\mathsf{A})] + \beta\sqrt{\omega}\cdot\mathbb{E}[\mathsf{X}_\star f(\mathsf{X};\mathsf{A})] \\
&\stackrel{(b)}{=} \mathbb{E}[\widehat{\varphi}(\mathsf{X};\mathsf{A}|\omega)f(\mathsf{X};\mathsf{A})] + \beta\sqrt{\omega}\cdot\mathbb{E}[\varphi(\mathsf{X};\mathsf{A}|\omega)f(\mathsf{X};\mathsf{A})].
\end{aligned} \tag{67}$$

In the above display step (a) follows from the fact that $\mathbb{E}[\mathsf{Z}f(\mathsf{X};\mathsf{A})]=0$ and step (b) uses the tower property. We claim that:

$$\omega < 1 \implies \beta\sqrt{\omega} \neq 1. \tag{68}$$

Assuming the above claim, we can split our analysis into two cases.

Case 1: $\omega < 1$. We first prove each of the claims made in the lemma assuming $\omega<1$.

Proof of Claim 1. Notice that (68) allows us to rearrange (67) to obtain:

$$\mathbb{E}[\varphi(\mathsf{X};\mathsf{A}|\omega)f(\mathsf{X};\mathsf{A})] = \frac{\mathbb{E}[\widehat{\varphi}(\mathsf{X};\mathsf{A}|\omega)f(\mathsf{X};\mathsf{A})]}{1-\beta\sqrt{\omega}} = \mathbb{E}[\bar{\varphi}(\mathsf{X};\mathsf{A}|\omega)f(\mathsf{X};\mathsf{A})], \tag{69}$$

where we recalled from the statement of the lemma that $\bar{\varphi}$ was defined as:

$$\begin{aligned}
\bar{\varphi}(x;a|\omega) &\stackrel{\text{def}}{=} \left(1 - \frac{\sqrt{\omega}}{\sqrt{1-\omega}}\cdot\mathbb{E}[\mathsf{Z}\varphi(\mathsf{X};\mathsf{A}|\omega)]\right)^{-1}\cdot\left(\varphi(x;a|\omega) - \frac{\mathbb{E}[\mathsf{Z}\varphi(\mathsf{X};\mathsf{A}|\omega)]}{\sqrt{1-\omega}}\cdot x\right) \\
&\stackrel{(65)}{=} \frac{\varphi(x;a|\omega)-\beta x}{1-\beta\sqrt{\omega}} \stackrel{(65)}{=} \frac{\widehat{\varphi}(x;a|\omega)}{1-\beta\sqrt{\omega}}
\end{aligned} \tag{70}$$

Substituting the formula (69) in (64) and completing the square yields the following MSE decomposition:

$$\mathbb{E}[(\mathsf{X}_\star - f(\mathsf{X};\mathsf{A}))^2] = 1 - \mathbb{E}[\bar{\varphi}(\mathsf{X};\mathsf{A})^2] + \mathbb{E}[\{f(\mathsf{X};\mathsf{A}) - \bar{\varphi}(\mathsf{X};\mathsf{A})\}^2] \tag{71}$$

Since $\bar{\varphi}$ satisfies the divergence-free requirement $\mathbb{E}[\mathsf{Z}\bar{\varphi}(\mathsf{X};\mathsf{A})]=0$ (cf. (66)), the above decomposition shows that $\bar{\varphi}$ is the DMMSE estimator.

Proof of Claim 2. Setting $f = \bar{\varphi}(\cdot\,;\cdot|\omega)$ in (69) and (71) yields:

$$\mathbb{E}[\varphi(\mathsf{X};\mathsf{A}|\omega)\bar{\varphi}(\mathsf{X};\mathsf{A}|\omega)] = \mathbb{E}[\bar{\varphi}^2(\mathsf{X};\mathsf{A}|\omega)] = 1 - \mathbb{E}[(\mathsf{X}_\star - \bar{\varphi}(\mathsf{X};\mathsf{A}|\omega))^2] = 1 - \mathsf{dmmse}_\pi(\omega).$$

By the Tower property, $\mathbb{E}[\mathsf{X}_\star\bar{\varphi}(\mathsf{X};\mathsf{A}|\omega)] = \mathbb{E}[\varphi(\mathsf{X};\mathsf{A}|\omega)\bar{\varphi}(\mathsf{X};\mathsf{A}|\omega)]$. Hence,

$$\mathbb{E}[\mathsf{X}_\star\bar{\varphi}(\mathsf{X};\mathsf{A}|\omega)] = \mathbb{E}[\varphi(\mathsf{X};\mathsf{A}|\omega)\bar{\varphi}(\mathsf{X};\mathsf{A}|\omega)] = \mathbb{E}[\bar{\varphi}^2(\mathsf{X};\mathsf{A}|\omega)] = 1 - \mathsf{dmmse}_\pi(\omega), \tag{72}$$

which is the second claim made in the lemma.

Proof of Claim 3. Notice from (70) that the MMSE and DMMSE estimators are related by:

$$\varphi(x;a|\omega) = (1 - \beta\sqrt{\omega}) \cdot \bar{\varphi}(x;a|\omega) + \beta x. \tag{73}$$

Hence, we can relate $\mathsf{mmse}_\pi(\omega)$ and $\mathsf{dmmse}_\pi(\omega)$ as follows:

$$\begin{aligned}
\mathsf{mmse}_\pi(\omega) = \mathbb{E}[(\mathsf{X}_\star - \varphi(\mathsf{X};\mathsf{A}|\omega))^2] &= \mathbb{E}[\mathsf{X}_\star^2] - 2\mathbb{E}[\mathsf{X}_\star\varphi(\mathsf{X};\mathsf{A}|\omega)] + \mathbb{E}[\varphi(\mathsf{X};\mathsf{A}|\omega)^2] \\
&\overset{\text{(a)}}{=} 1 - \mathbb{E}[\mathsf{X}_\star\varphi(\mathsf{X};\mathsf{A}|\omega)] \qquad (74)\\
&\overset{(73)}{=} 1 - (1 - \beta\sqrt{\omega}) \cdot \mathbb{E}[\mathsf{X}_\star\bar{\varphi}(\mathsf{X};\mathsf{A}|\omega)] - \beta\mathbb{E}[\mathsf{X}_\star\mathsf{X}] \\
&= (1 - \beta\sqrt{\omega}) \cdot (1 - \mathbb{E}[\mathsf{X}_\star\bar{\varphi}(\mathsf{X};\mathsf{A}|\omega)]) \\
&\overset{(72)}{=} (1 - \beta\sqrt{\omega}) \cdot \mathsf{dmmse}_\pi(\omega). \qquad (75)
\end{aligned}$$

In the above display, step (a) relies on the assumption $\mathbb{E}[\mathsf{X}_\star^2] = 1$ (cf. Assumption 1) and the tower property $\mathbb{E}[\mathsf{X}_\star\varphi(\mathsf{X};\mathsf{A}|\omega)] = \mathbb{E}[\varphi(\mathsf{X};\mathsf{A}|\omega)^2]$. Moreover, notice that:

$$\begin{aligned}
\beta &\overset{\text{def}}{=} \frac{\mathbb{E}[\mathsf{Z}\varphi(\mathsf{X};\mathsf{A}|\omega)]}{\sqrt{1-\omega}} \overset{(62)}{=} \frac{\mathbb{E}[\mathsf{X}\varphi(\mathsf{X};\mathsf{A}|\omega)] - \sqrt{\omega}\mathbb{E}[\mathsf{X}_\star\varphi(\mathsf{X};\mathsf{A}|\omega)]}{1-\omega} \\
&\overset{\text{(a)}}{=} \frac{\mathbb{E}[\mathsf{X}\mathsf{X}_\star] - \sqrt{\omega}\mathbb{E}[\mathsf{X}_\star\varphi(\mathsf{X};\mathsf{A}|\omega)]}{1-\omega} = \frac{\sqrt{\omega}(1 - \mathbb{E}[\mathsf{X}_\star\varphi(\mathsf{X};\mathsf{A}|\omega)])}{1-\omega} \overset{(74)}{=} \frac{\sqrt{\omega}\cdot\mathsf{mmse}_\pi(\omega)}{1-\omega}.
\end{aligned}$$

In the above display, step (a) again uses the tower property $\mathbb{E}[\mathsf{X}_\star\varphi(\mathsf{X};\mathsf{A}|\omega)] = \mathbb{E}[\varphi(\mathsf{X};\mathsf{A}|\omega)^2]$. Finally, we substitute the formula for β obtained in the above display in (75) to obtain:

$$\frac{1}{\mathsf{dmmse}_\pi(\omega)} = \frac{1}{\mathsf{mmse}_\pi(\omega)} - \frac{\omega}{1-\omega}$$

after rearrangement. This is exactly the third claim made in the lemma.

Case 2: $\omega = 1$. When $\omega = 1$, $\mathsf{X} = \mathsf{X}_\star$. The estimator $f(\mathsf{X};\mathsf{A}) = \mathsf{X}$ has zero MSE and satisfies the divergence-free requirement $\mathbb{E}[\mathsf{Z}\mathsf{X}] = \mathbb{E}[\mathsf{Z}]\mathbb{E}[\mathsf{X}_\star] = 0$. Hence, $\bar{\varphi}(x;a) \overset{\text{def}}{=} x$ is the DMMSE estimator. Moreover, the identity (which is the second claim made in the lemma):

$$\mathbb{E}[\mathsf{X}_\star\bar{\varphi}(\mathsf{X};\mathsf{A}|\omega)] = \mathbb{E}[\varphi(\mathsf{X};\mathsf{A}|\omega)\bar{\varphi}(\mathsf{X};\mathsf{A}|\omega)] = \mathbb{E}[\bar{\varphi}^2(\mathsf{X};\mathsf{A}|\omega)] = 1 - \mathsf{dmmse}_\pi(\omega)$$

holds trivially in this case since $\mathsf{X}_\star = \bar{\varphi}(\mathsf{X};\mathsf{A}|\omega) = \mathsf{X}$, $\mathsf{dmmse}_\pi(1) = 0$, and $\mathbb{E}[\mathsf{X}_\star^2] = 1$.

To finish the proof of the lemma, we need to prove the claim in (68).

Proof of (68). We prove the contrapositive of (68): $\beta\sqrt{\omega} = 1 \implies \omega = 1$. When $\beta\sqrt{\omega} = 1$, (67) shows that for any $f \in L^2(\mathsf{X};\mathsf{A})$ which satisfies the divergence-free requirement $\mathbb{E}[\mathsf{Z}f(\mathsf{X};\mathsf{A})] = 0$, we have $\mathbb{E}[\widehat{\varphi}(\mathsf{X};\mathsf{A})f(\mathsf{X};\mathsf{A})] = 0$. Taking $f = \widehat{\varphi}$ (cf. (66)), we conclude that $\mathbb{E}[\widehat{\varphi}^2] = 0$. Recalling the definition of $\widehat{\varphi}$, we obtain the following formula for the MMSE estimator for the scalar Gaussian channel:

$$\varphi(\mathsf{X};\mathsf{A}) = \beta\mathsf{X} = \frac{\mathsf{X}}{\sqrt{\omega}},$$

which gives us the following formula for the MMSE:

$$\mathrm{MMSE}(\mathsf{X}_\star|\mathsf{X};\mathsf{A}) = \frac{1-\omega}{\omega}.$$

Since $\mathrm{MMSE}(\mathsf{X}_\star|\mathsf{X};\mathsf{A})$ is bounded by the MSE of the linear estimator $\sqrt{\omega}\cdot\mathsf{X}$ (this is the linear estimator with the lowest MSE),

$$\mathrm{MMSE}(\mathsf{X}_\star|\mathsf{X};\mathsf{A}) = \frac{1-\omega}{\omega} \le \mathbb{E}[(\sqrt{\omega}\mathsf{X}-\mathsf{X}_\star)^2] = 1-\omega.$$

Since $\omega \in [0,1]$, by rearranging the above inequality we conclude that $\omega = 1$. □

A.2.2 Proof of Lemma 3

Proof of Lemma 3. It will be convenient to define the random vector $\mathsf{X}_{\le t} \stackrel{\text{def}}{=} (\mathsf{X}_1,\dots,\mathsf{X}_t)^\top$ and the random variable $\mathsf{S} \stackrel{\text{def}}{=} \langle \mathsf{X}_{\le t}, v_{\mathrm{opt}}\rangle$. Observe that the conditional joint distribution of $(\mathsf{X}_{\le t},\mathsf{S})$ given $(\mathsf{X}_\star;\mathsf{A})$ is:

$$\begin{bmatrix}\mathsf{X}_{\le t}\\ \mathsf{S}\end{bmatrix} \;\Big|\; (\mathsf{X}_\star;\mathsf{A}) \sim \mathcal{N}\left(\mathsf{X}_\star\cdot\begin{bmatrix}\alpha\\ \sqrt{\omega_{\mathrm{eff}}}\end{bmatrix}, \begin{bmatrix}\Sigma & v_{\mathrm{opt}}^\top\Sigma\\ \Sigma v_{\mathrm{opt}} & 1-\omega_{\mathrm{eff}}\end{bmatrix}\right).$$

Using Gaussian conditioning, we find that the conditional distribution of $\mathsf{X}_{\le t}$ given $(\mathsf{S},\mathsf{X}_\star;\mathsf{A})$ is given by:

$$\mathsf{X}_{\le t} \mid (\mathsf{S},\mathsf{X}_\star;\mathsf{A}) \sim \mathcal{N}\left(\frac{\mathsf{S}\alpha}{\sqrt{\omega_{\mathrm{eff}}}}, \Sigma - \frac{1-\omega_{\mathrm{eff}}}{\omega_{\mathrm{eff}}}\cdot\alpha\alpha^\top\right).$$

Since the RHS does not have any dependence on $(\mathsf{X}_\star;\mathsf{A})$, this implies that $\mathsf{X}_{\le t}$ and $(\mathsf{X}_\star;\mathsf{A})$ are conditionally independent given S:

$$\mathsf{X}_{\le t} \perp (\mathsf{X}_\star;\mathsf{A}) \mid \mathsf{S}. \tag{76}$$

Analysis of MMSE. We first prove the MMSE formula and identify the MMSE estimator. To do so, we begin by showing the lower bound:

$$\mathrm{MMSE}(\mathsf{X}_\star|\mathsf{X}_{\le t};\mathsf{A}) = \min_{f\in L^2(\mathsf{X}_{\le t};\mathsf{A})} \mathbb{E}[\{\mathsf{X}_\star - f(\mathsf{X}_{\le t};\mathsf{A})\}^2] \ge \mathsf{mmse}_\pi(\omega_{\mathrm{eff}}).$$

To do so, we consider any $f \in L^2(\mathsf{X}_{\le t};\mathsf{A})$:

$$\begin{aligned}
\mathbb{E}[\{\mathsf{X}_\star - f(\mathsf{X}_{\le t};\mathsf{A})\}^2] &\stackrel{\text{(a)}}{\ge} \mathbb{E}[\{\mathsf{X}_\star - \mathbb{E}[f(\mathsf{X}_{\le t};\mathsf{A})|\mathsf{S},\mathsf{X}_\star,\mathsf{A}]\}^2]\\
&\stackrel{(76)}{=} \mathbb{E}[\{\mathsf{X}_\star - \mathbb{E}[f(\mathsf{X}_{\le t};\mathsf{A})|\mathsf{S},\mathsf{A}]\}^2]\\
&\stackrel{\text{(b)}}{\ge} \min_{g\in L^2(\mathsf{S};\mathsf{A})} \mathbb{E}[\{\mathsf{X}_\star - g(\mathsf{S};\mathsf{A})\}^2]\\
&= \mathrm{MMSE}(\mathsf{X}_\star|\mathsf{S};\mathsf{A})\\
&\stackrel{\text{(c)}}{=} \mathsf{mmse}_\pi(\omega_{\mathrm{eff}}),
\end{aligned}$$

where step (a) used Jensen's inequality, step (b) follows by observing $\mathbb{E}[f(\mathsf{X}_{\le t};\mathsf{A})|\mathsf{S},\mathsf{A}] \in L^2(\mathsf{S};\mathsf{A})$, and step (c) is justified by noticing that $(\mathsf{X}_\star,\mathsf{S};\mathsf{A})$ forms a scalar Gaussian channel with SNR ω_{eff}. Since $f \in L^2(\mathsf{X}_{\le t};\mathsf{A})$ was arbitrary, we can minimize the LHS of the above display with respect to f, which yields:

$$\mathrm{MMSE}(\mathsf{X}_\star|\mathsf{X}_{\le t};\mathsf{A}) \ge \mathsf{mmse}_\pi(\omega_{\mathrm{eff}}). \tag{77}$$

To obtain the upper bound, we observe that:

$$\mathsf{mmse}_\pi(\omega_{\mathrm{eff}}) \stackrel{\text{(a)}}{=} \mathbb{E}[\{\mathsf{X}_\star - \varphi(\mathsf{S};\mathsf{A}|\omega_{\mathrm{eff}})\}^2] \stackrel{\text{(b)}}{\ge} \min_{f\in L^2(\mathsf{X}_{\le t};\mathsf{A})} \mathbb{E}[\{\mathsf{X}_\star - f(\mathsf{X}_{\le t};\mathsf{A})\}^2] \stackrel{\text{def}}{=} \mathrm{MMSE}(\mathsf{X}_\star|\mathsf{X}_{\le t};\mathsf{A}), \tag{78}$$

where the equality in (a) follows by recalling the formula for the MMSE estimator in a scalar Gaussian channel and the inequality in (b) follows by observing that:

$$\varphi(\mathsf{S};\mathsf{A}|\omega_{\mathrm{eff}}) = \varphi(\langle \mathsf{X}_{\le t}, v_{\mathrm{opt}}\rangle;\mathsf{A}|\omega_{\mathrm{eff}}) \stackrel{\text{def}}{=} f_\star(\mathsf{X}_{\le t};\mathsf{A}) \in L^2(\mathsf{X}_{\le t};\mathsf{A}).$$

Combining the conclusions of (77) and (78) yields:

$$\mathrm{MMSE}(\mathsf{X}_\star|\mathsf{X}_{\le t};\mathsf{A}) = \mathbb{E}[\{\mathsf{X}_\star - f_\star(\mathsf{X}_{\le t};\mathsf{A})\}^2] = \mathsf{mmse}_\pi(\omega_{\mathrm{eff}}),$$

which gives us the claimed formula for the MMSE and the MMSE estimator.

Analysis of DMMSE. We begin by observing that since $(\mathsf{X}_\star, \mathsf{S} \stackrel{\text{def}}{=} \langle v_{\text{opt}}, \mathsf{X}_{\le t} \rangle ; \mathsf{A})$ form a scalar Gaussian channel with SNR ω_{eff}, we can write S as:

$$\mathsf{S} \stackrel{\text{def}}{=} \langle v_{\text{opt}}, \mathsf{X}_{\le t} \rangle = \sqrt{\omega_{\text{eff}}}\mathsf{X}_\star + \sqrt{1-\omega_{\text{eff}}}\mathsf{W}, \quad \mathsf{W} \stackrel{\text{def}}{=} \frac{\langle v_{\text{opt}}, \mathsf{Z}_{\le t} \rangle}{\sqrt{1-\omega_{\text{eff}}}} \sim \mathcal{N}(0,1). \tag{79}$$

As before, we begin by showing the lower bound:

$$\begin{aligned}\text{DMMSE}(\mathsf{X}_\star|\mathsf{X}_{\le t};\mathsf{A}) &\stackrel{\text{def}}{=} \left(\min_{f \in L^2(\mathsf{X}_{\le t};\mathsf{A})} \mathbb{E}[\{\mathsf{X}_\star - f(\mathsf{X}_{\le t};\mathsf{A})\}^2] \text{ subject to } \mathbb{E}[\mathsf{Z}_i f(\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})] = 0 \; \forall \; i \; \in \; [t] \right) \\ &\ge \mathsf{dmmse}_\pi(\omega_{\text{eff}}).\end{aligned}$$

For any $f \in L^2(\mathsf{X}_{\le t};\mathsf{A})$ which satisfies the divergence-free constraints $\mathbb{E}[\mathsf{Z}_i f(\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})] = 0 \; \forall \; i \; \in \; [t]$, we have:

$$\begin{aligned}\mathbb{E}[\{\mathsf{X}_\star - f(\mathsf{X}_{\le t};\mathsf{A})\}^2] &\stackrel{\text{(a)}}{\ge} \mathbb{E}[\{\mathsf{X}_\star - \mathbb{E}[f(\mathsf{X}_{\le t};\mathsf{A})|\mathsf{S},\mathsf{X}_\star,\mathsf{A}]\}^2] \\ &\stackrel{(76)}{=} \mathbb{E}[\{\mathsf{X}_\star - \mathbb{E}[f(\mathsf{X}_{\le t};\mathsf{A})|\mathsf{S},\mathsf{A}]\}^2] \\ &\stackrel{\text{(b)}}{\ge} \min_{g \in L^2(\mathsf{S};\mathsf{A})} \left(\mathbb{E}[\{\mathsf{X}_\star - g(\mathsf{S};\mathsf{A})\}^2] \text{ subject to } \mathbb{E}[\mathsf{W}g(\mathsf{S};\mathsf{A})] = 0 \right) \\ &= \text{DMMSE}(\mathsf{X}_\star|\mathsf{S};\mathsf{A}) \\ &= \mathsf{dmmse}_\pi(\omega_{\text{eff}}).\end{aligned}$$

In the above display:

1. Step (a) uses Jensen's inequality.
2. In step (b) we observed that $\mathbb{E}[f(\mathsf{X}_{\le t};\mathsf{A})|\mathsf{S},\mathsf{A}] \in L^2(\mathsf{S};\mathsf{A})$. Moreover, the estimator $\mathbb{E}[f(\mathsf{X}_{\le t};\mathsf{A})|\mathsf{S},\mathsf{A}]$ satisfies the divergence-free constraint $\mathbb{E}[\mathsf{W}\mathbb{E}[f(\mathsf{X}_{\le t};\mathsf{A})|\mathsf{S},\mathsf{A}]] = 0$. This can be verified as follows:

   $$\mathbb{E}[\mathsf{W}\mathbb{E}[f(\mathsf{X}_{\le t};\mathsf{A})|\mathsf{S},\mathsf{A}]] \stackrel{(76)}{=} \mathbb{E}[\mathsf{W}\mathbb{E}[f(\mathsf{X}_{\le t};\mathsf{A})|\mathsf{S},\mathsf{X}_\star\mathsf{A}]] \stackrel{\text{(i)}}{=} \mathbb{E}[\mathbb{E}[\mathsf{W}f(\mathsf{X}_{\le t};\mathsf{A})|\mathsf{S},\mathsf{X}_\star\mathsf{A}]] = \mathbb{E}[\mathsf{W}f(\mathsf{X}_{\le t};\mathsf{A})] \stackrel{\text{(ii)}}{=} 0,$$

   where we used the fact that W is measurable with respect to $\mathsf{X}_\star, \mathsf{S}$ (recall (79)) in step (i). To obtain the equality marked (ii), we expressed W as a linear combination of $\mathsf{Z}_{\le t}$ (recall (79)) and used the fact that f satisfies the divergence-free constraints.

Since $f \in L^2(\mathsf{X}_{\le t};\mathsf{A})$ was an arbitrary function that satisfies the divergence-free constraints, we can minimize the LHS of the above display with respect to f and obtain:

$$\text{DMMSE}(\mathsf{X}_\star|\mathsf{X}_{\le t};\mathsf{A}) \ge \mathsf{dmmse}_\pi(\omega_{\text{eff}}). \tag{80}$$

To obtain the upper bound, we observe that:

$$\begin{aligned}\mathsf{dmmse}_\pi(\omega_{\text{eff}}) &\stackrel{\text{(a)}}{=} \mathbb{E}[\{\mathsf{X}_\star - \bar{\varphi}(\mathsf{S};\mathsf{A}|\omega_{\text{eff}})\}^2] \\ &\stackrel{\text{(b)}}{\ge} \left(\min_{f \in L^2(\mathsf{X}_{\le t};\mathsf{A})} \mathbb{E}[\{\mathsf{X}_\star - f(\mathsf{X}_{\le t};\mathsf{A})\}^2] \text{ subject to } \mathbb{E}[\mathsf{Z}_i f(\mathsf{X}_1,\dots,\mathsf{X}_t;\mathsf{A})] = 0 \; \forall \; i \; \in \; [t] \right) \\ &\stackrel{\text{def}}{=} \text{DMMSE}(\mathsf{X}_\star|\mathsf{X}_{\le t};\mathsf{A}).\end{aligned} \tag{81}$$

In the above display:

1. The equality in (a) follows by recalling the formula for the DMMSE estimator in a scalar Gaussian channel (Lemma 2).
2. To obtain the inequality in (b), we observed that:

   $$\bar{\varphi}(\mathsf{S};\mathsf{A}|\omega_{\text{eff}}) = \bar{\varphi}(\langle \mathsf{X}_{\le t}, v_{\text{opt}} \rangle ; \mathsf{A}|\omega_{\text{eff}}) \stackrel{\text{def}}{=} \bar{f}_\star(\mathsf{X}_{\le t};\mathsf{A}) \in L^2(\mathsf{X}_{\le t};\mathsf{A}).$$

Moreover, $\bar{f}_\star(\mathsf{X}_{\le t};\mathsf{A}) \stackrel{\text{def}}{=} \bar{\varphi}(\langle \mathsf{X}_{\le t}, v_{\text{opt}}\rangle;\mathsf{A}|\omega_{\text{eff}})$ satisfies the divergence-free requirements:

$$\mathbb{E}[\mathsf{Z}_i\bar{\varphi}(\langle \mathsf{X}_{\le t}, v_{\text{opt}}\rangle;\mathsf{A}|\omega_{\text{eff}})] = 0 \quad \forall\, i \in [t].$$

This can be verified as follows:

$$\begin{aligned}
\mathbb{E}[\mathsf{Z}_i\bar{\varphi}(\langle \mathsf{X}_{\le t}, v_{\text{opt}}\rangle;\mathsf{A}|\omega_{\text{eff}})] &\stackrel{(79)}{=} \mathbb{E}\left[\mathsf{Z}_i\bar{\varphi}\left(\sqrt{\omega_{\text{eff}}}\mathsf{X}_\star + \sqrt{1-\omega_{\text{eff}}}\mathsf{W};\mathsf{A}|\omega_{\text{eff}}\right)\right] \\
&= \mathbb{E}[\bar{\varphi}\left(\sqrt{\omega_{\text{eff}}}\mathsf{X}_\star + \sqrt{1-\omega_{\text{eff}}}\mathsf{W};\mathsf{A}|\omega_{\text{eff}}\right)\cdot \mathbb{E}[\mathsf{Z}_i|\mathsf{X}_\star,\mathsf{A},\mathsf{W}]] \\
&\stackrel{(i)}{=} c\mathbb{E}[\bar{\varphi}\left(\sqrt{\omega_{\text{eff}}}\mathsf{X}_\star + \sqrt{1-\omega_{\text{eff}}}\mathsf{W};\mathsf{A}|\omega_{\text{eff}}\right)\cdot \mathsf{W}] \\
&\stackrel{(ii)}{=} 0,
\end{aligned}$$

where the equality marked (i) follows by observing that $(\mathsf{Z}_i,\mathsf{W})$ are centered Gaussian random variables independent of $(\mathsf{X}_\star,\mathsf{A})$ and hence $\mathbb{E}[\mathsf{Z}_i|\mathsf{X}_\star,\mathsf{A},\mathsf{W}] = \mathbb{E}[\mathsf{Z}_i|\mathsf{W}] = c\mathsf{W}$ for some constant c (the exact formula for c is not important for the argument). The equality (ii) follows because $\bar{\varphi}(\mathsf{S};\mathsf{A})$ satisfies the divergence-free requirement $\mathbb{E}[\mathsf{W}\bar{\varphi}(\mathsf{S};\mathsf{A})] = 0$.

Combining the conclusions of (80) and (81) yields:

$$\text{DMMSE}(\mathsf{X}_\star|\mathsf{X}_{\le t};\mathsf{A}) = \mathbb{E}[\{\mathsf{X}_\star - \bar{f}_\star(\mathsf{X}_{\le t};\mathsf{A})\}^2] = \mathsf{dmmse}_\pi(\omega_{\text{eff}}),$$

which gives us the claimed formula for the DMMSE and the DMMSE estimator. This concludes the proof of this lemma. □

A.2.3 Proof of Lemma 4

Proof of Lemma 4. It is sufficient to show that for any $\rho \in [0,1]$ and any $\omega \in [0,1]$, $\mathsf{dmmse}_\pi(\omega) \le \mathsf{dmmse}_\pi(\rho\omega)$. Consider a scalar Gaussian channel $(\mathsf{X}_\star,\mathsf{X}_0;\mathsf{A})$ at SNR ω:

$$(\mathsf{X}_\star;\mathsf{A}) \sim \pi, \quad \mathsf{X}_0 = \sqrt{\omega}\cdot\mathsf{X}_\star + \sqrt{1-\omega}\cdot\mathsf{Z}_0, \quad \mathsf{Z}_0 \sim \mathcal{N}(0,1),$$

where $(\mathsf{X}_\star;\mathsf{A})$ and Z_0 are sampled independently. Let W be a $\mathcal{N}(0,1)$ random variable, sampled independently of $(\mathsf{X}_\star;\mathsf{A})$ and Z_0. We introduce the random variable $\mathsf{X}_1 = \sqrt{\rho}\cdot\mathsf{X} + \sqrt{1-\rho}\cdot\mathsf{W}$. This construction ensures that $(\mathsf{X}_\star,\mathsf{X}_1;\mathsf{A})$ is a scalar Gaussian channel with SNR $\rho\omega$:

$$\mathsf{X}_1 = \sqrt{\rho\omega}\cdot\mathsf{X}_\star + \sqrt{1-\rho\omega}\cdot\mathsf{Z}_1, \quad \mathsf{Z}_1 \stackrel{\text{def}}{=} \frac{\sqrt{\rho(1-\omega)}\cdot\mathsf{Z}_0 + \sqrt{1-\rho}\cdot\mathsf{W}}{\sqrt{1-\rho\omega}} \sim \mathcal{N}(0,1). \tag{82}$$

To obtain the claim of the lemma, we observe that:

$$\begin{aligned}
\mathsf{dmmse}_\pi(\omega) &\stackrel{(a)}{=} \text{DMMSE}(\mathsf{X}_\star|\mathsf{X}_0,\mathsf{X}_1;\mathsf{A}) \\
&\stackrel{\text{def}}{=} \left(\min_{f\in L^2(\mathsf{X}_0,\mathsf{X}_1;\mathsf{A})} \mathbb{E}[\{\mathsf{X}_\star - f(\mathsf{X}_0,\mathsf{X}_1;\mathsf{A})\}^2] \text{ subject to } \mathbb{E}[\mathsf{Z}_i f(\mathsf{X}_0,\mathsf{X}_1;\mathsf{A})] = 0\ \forall\, i \in \{0,1\}\right) \\
&\stackrel{(b)}{\le} \left(\min_{f\in L^2(\mathsf{X}_1;\mathsf{A})} \mathbb{E}[\{\mathsf{X}_\star - f(\mathsf{X}_1;\mathsf{A})\}^2] \text{ subject to } \mathbb{E}[\mathsf{Z}_1 f(\mathsf{X}_1;\mathsf{A})] = 0\right) \\
&\stackrel{\text{def}}{=} \text{DMMSE}(\mathsf{X}_\star|\mathsf{X}_1;\mathsf{A}) \\
&= \mathsf{dmmse}_\pi(\rho\omega).
\end{aligned}$$

In the above display, the equality in step (a) follows from Lemma 3 (the effective SNR of the Gaussian channel $(\mathsf{X}_\star,\mathsf{X}_0,\mathsf{X}_1;\mathsf{A})$ is ω) and the inequality in step (b) follows by observing that the feasible set in the definition of $\text{DMMSE}(\mathsf{X}_\star|\mathsf{X}_1;\mathsf{A})$ is a subset of the feasible set in the definition of $\text{DMMSE}(\mathsf{X}_\star|\mathsf{X}_0,\mathsf{X}_1;\mathsf{A})$. Indeed if $f \in L^2(\mathsf{X}_1;\mathsf{A})$ satisfies $\mathbb{E}[\mathsf{Z}_1 f(\mathsf{X}_1;\mathsf{A})] = 0$, then $f \in L^2(\mathsf{X}_0,\mathsf{X}_1;\mathsf{A})$, and:

$$\mathbb{E}[\mathsf{Z}_0 f(\mathsf{X}_1;\mathsf{A})] = \mathbb{E}[f(\mathsf{X}_1;\mathsf{A})\mathbb{E}[\mathsf{Z}_0|\mathsf{X}_\star,\mathsf{Z}_1,\mathsf{A}]] \stackrel{(i)}{=} \mathbb{E}[f(\mathsf{X}_1;\mathsf{A})\mathbb{E}[\mathsf{Z}_0|\mathsf{Z}_1]] \stackrel{(82)}{=} \frac{\sqrt{1-\rho\omega}}{\sqrt{\rho(1-\omega)}}\cdot\mathbb{E}[f(\mathsf{X}_1;\mathsf{A})\mathsf{Z}_1] = 0,$$

where the equality marked (i) follows by observing that $(\mathsf{Z}_0,\mathsf{Z}_1)$ are centered Gaussian random variables independent of $(\mathsf{X}_\star;\mathsf{A})$ (cf. (82)). This completes the proof of this lemma. □

B State Evolution for OAMP Algorithms (Theorem 1)

In this appendix, we present the proof of Theorem 1. We begin by introducing the key ideas involved in the proof in the form of some intermediate results.

Polynomial Approximation. Consider a general OAMP algorithm:

$$\boldsymbol{x}_t = \Psi_t(\boldsymbol{Y}) \cdot f_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N}. \tag{83}$$

By a polynomial approximation argument, we may assume that the matrix denoisers $\{\Psi_t\}_{t\in\mathbb{N}}$ are polynomials. This argument is summarized in the following lemma, whose proof is deferred to Appendix B.1.

Lemma 5. *It is sufficient to prove Theorem 1 under the additional assumption that for each $t \in \mathbb{N}$, the matrix denoiser $\Psi_t : \mathbb{R} \mapsto \mathbb{R}$ is a polynomial.*

Orthogonal Decomposition. Let $(\mathsf{X}_\star, (\mathsf{X}_t)_{t\in\mathbb{N}}; \mathsf{A})$ denote the state evolution random variables associated with the OAMP algorithm in (83). A key idea involved involved in the proof of Theorem 1 is to consider the following decomposition of the functions $(f_t)_{t\in\mathbb{N}}$:

$$f_t(x_1, \ldots, x_{t-1}; a) = \alpha_t x_\star + f_t^\perp(x_1, \ldots, x_{t-1}; x_\star, a), \tag{84}$$

where:

$$\alpha_t \stackrel{\text{def}}{=} \mathbb{E}[\mathsf{X}_\star f_t(\mathsf{X}_1, \ldots, \mathsf{X}_{t-1}; \mathsf{A})], \quad f_t^\perp(x_1, \ldots, x_{t-1}; x_\star, a) \stackrel{\text{def}}{=} f_t(x_1, \ldots, x_{t-1}; a) - \alpha_t x_\star.$$

Notice that by construction, $f_t^\perp$ is orthogonal to the signal state evolution random variable $\mathsf{X}_\star$:

$$\mathbb{E}[\mathsf{X}_\star f_t^\perp(\mathsf{X}_1, \ldots, \mathsf{X}_{t-1}; \mathsf{X}_\star, \mathsf{A})] = 0. \tag{85}$$

Using this decomposition, we have:

$$\boldsymbol{x}_t = \alpha_t \cdot \Psi_t(\boldsymbol{Y}) \cdot \boldsymbol{x}_\star + \Psi_t(\boldsymbol{Y}) \cdot f_t^\perp(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}). \tag{86}$$

Recall that the observed matrix $\boldsymbol{Y}$ consists of a signal part and a noise part: $\boldsymbol{Y} = \frac{\theta}{N}\boldsymbol{x}_\star\boldsymbol{x}_\star^\top + \boldsymbol{W}$. The decomposition above allows us to disentangle the action of the signal component and the noise component using the following lemma, whose proof is deferred to Appendix B.2

Lemma 6. *Let $\Psi : \mathbb{R} \mapsto \mathbb{R}$ be a polynomial function.*

1. *There exists a polynomial function $\widetilde{\Psi} : \mathbb{R} \mapsto \mathbb{R}$ associated with Ψ such that $\boldsymbol{\Psi}(\boldsymbol{Y}) \cdot \boldsymbol{x}_\star \overset{N \to \infty}{\simeq} \widetilde{\Psi}(\boldsymbol{W}) \cdot \boldsymbol{x}_\star$, where $\overset{N \to \infty}{\simeq}$ denotes asymptotic equivalence of random vectors (Definition 2).*
2. *Let $\boldsymbol{v}$ be a N-dimensional random vector with the property that:*

$$\frac{\langle \boldsymbol{W}^i \boldsymbol{v}, \boldsymbol{x}_\star \rangle}{N} \xrightarrow{\mathbb{P}} 0 \quad \forall\, i \in \mathbb{N}. \tag{87}$$

Then, $\boldsymbol{\Psi}(\boldsymbol{Y}) \cdot \boldsymbol{v} \overset{N \to \infty}{\simeq} \Psi(\boldsymbol{W}) \cdot \boldsymbol{v}$.

We will use the lemma above to show that $f_t^\perp(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t; \boldsymbol{x}_\star, \boldsymbol{a})$ does not interact with the signal component of $\Psi_t(\boldsymbol{Y})$, thanks to the orthogonality property in (85), which will guarantee that $\boldsymbol{v} = f_t^\perp(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t; \boldsymbol{x}_\star, \boldsymbol{a})$ satisfies the orthogonality condition in (87). Formally, we will show that:

$$\Psi_t(\boldsymbol{Y}) \cdot f_t^\perp(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}) \overset{N \to \infty}{\simeq} \Psi_t(\boldsymbol{W}) \cdot f_t^\perp(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}), \tag{88a}$$

On the other hand, $\boldsymbol{x}_\star$ does interact with the signal component in $\Psi_t(\boldsymbol{Y})$ and this interaction is captured by the *transformed polynomial* $\widetilde{\Psi}_t$ associated with Ψ_t, whose existence is guaranteed by the lemma above:

$$\Psi_t(\boldsymbol{Y}) \cdot \boldsymbol{x}_\star \overset{N \to \infty}{\simeq} \widetilde{\Psi}_t(\boldsymbol{W}) \cdot \boldsymbol{x}_\star. \tag{88b}$$

Auxiliary OAMP Algorithm. Using the approximations from (88) in the update equation of the OAMP algorithm given in (86), leads us to introduce an auxiliary OAMP algorithm which generates iterates $\widetilde{\boldsymbol{x}}_1, \widetilde{\boldsymbol{x}}_2, \dots$ via the update rule:

$$\widetilde{\boldsymbol{x}}_t = \alpha_t \cdot \widetilde{\Psi}_t(\boldsymbol{W}) \cdot \boldsymbol{x}_\star + \Psi_t(\boldsymbol{W}) \cdot f_t^{\perp}(\widetilde{\boldsymbol{x}}_1, \dots, \widetilde{\boldsymbol{x}}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}) \quad \forall\, t \in \mathbb{N}. \tag{89}$$

This auxiliary OAMP algorithm will serve as an easy-to-analyze approximation to the original OAMP algorithm in (83). Since the auxiliary OAMP algorithm uses the rotationally invariant noise matrix $\boldsymbol{W}$ in its iterations (rather than the non-rotationally invariant matrix $\boldsymbol{Y}$ used by the original OAMP algorithm), its dynamics can be easily analyzed using known results on the state evolution of AMP algorithms for rotationally invariant matrices and the associated universality class [25, 26, 29, 50, 69, 74, 75]. Applying these results requires us to compute:

$$\operatorname*{plim}_{N\to\infty} \frac{\operatorname{Tr}[\widetilde{\Psi}_t(\boldsymbol{W})]}{N}, \quad \operatorname*{plim}_{N\to\infty} \frac{\operatorname{Tr}[\widetilde{\Psi}_s(\boldsymbol{W})\widetilde{\Psi}_t(\boldsymbol{W})]}{N} \quad \forall\, s, t \in \mathbb{N}.$$

At first glance, computing the above limits appears challenging as the transformed polynomials $(\widetilde{\Psi}_t)_{t\in\mathbb{N}}$ constructed in Lemma 6 have a complicated recursive characterization. Fortunately, the above limits have a simple formula in terms of ν, the limiting spectral measure of $\boldsymbol{Y}$ in the direction of the signal (recall Lemma 1).

Lemma 7. *For any $s, t \in \mathbb{N}$, we have:*

$$\frac{\operatorname{Tr}[\widetilde{\Psi}_t(\boldsymbol{W})]}{N} \xrightarrow{\mathbb{P}} \mathbb{E}[\widetilde{\Psi}_t(\Lambda)] = \mathbb{E}[\Psi_t(\Lambda_\nu)], \quad \frac{\operatorname{Tr}[\widetilde{\Psi}_s(\boldsymbol{W})\widetilde{\Psi}_t(\boldsymbol{W})]}{N} \xrightarrow{\mathbb{P}} \mathbb{E}[\widetilde{\Psi}_s(\Lambda)\widetilde{\Psi}_t(\Lambda)] = \mathbb{E}[\Psi_s(\Lambda_\nu)\Psi_t(\Lambda_\nu)],$$

where $\Lambda \sim \mu$ and $\Lambda_\nu \sim \nu$.

The proof of this lemma is presented in Appendix B.3. An immediate consequence of this lemma and an existing result on the dynamics of AMP algorithms driven by rotationally invariant matrices [25, Theorem 2] is the following characterization of the dynamics of the auxiliary OAMP algorithm.

Lemma 8. *1. For any $t \in \mathbb{N}$, the iterates generated by the auxiliary OAMP algorithm in (89) satisfy*

$$(\boldsymbol{x}_\star, \widetilde{\boldsymbol{x}}_1, \dots, \widetilde{\boldsymbol{x}}_t; \boldsymbol{a}) \xrightarrow{W_2} (\mathsf{X}_\star, \mathsf{X}_1, \dots, \mathsf{X}_t; \mathsf{A}),$$

where $(\mathsf{X}_\star, \mathsf{X}_1, \dots, \mathsf{X}_t; \mathsf{A})$ are the state evolution random variables associated with the original OAMP algorithm in (83).

2. Moreover, for any $t \in \mathbb{N}$ and any $i \in \mathbb{N}$,

$$\frac{\left\langle \boldsymbol{W}^i f_t^{\perp}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}), \boldsymbol{x}_\star \right\rangle}{N} \xrightarrow{\mathbb{P}} 0.$$

The proof of the above lemma is deferred to Appendix B.4. We have now introduced all the key ideas involved in the proof of Theorem 1 and are in a position to present its proof.

Proof of Theorem 1. Our goal is to show that for any $t \in \mathbb{N}$,

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

In light of Lemma 8, it suffices to show that the auxiliary OAMP algorithm in (89) approximates the given OAMP algorithm in the sense:

$$\boldsymbol{x}_t \overset{N \to \infty}{\simeq} \widetilde{\boldsymbol{x}}_t \quad \forall\, t \in \mathbb{N}. \tag{90}$$

We show the above claim by induction. As our induction hypothesis, we assume that $\boldsymbol{x}_s \overset{N \to \infty}{\simeq} \widetilde{\boldsymbol{x}}_s$ for all $s < t$, and verify the claim at iteration t:

$$\begin{aligned}
\boldsymbol{x}_t &\overset{(86)}{=} \alpha_t \cdot \Psi_t(\boldsymbol{Y}) \cdot \boldsymbol{x}_\star + \Psi_t(\boldsymbol{Y}) \cdot f_t^\perp(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}) \\
&\overset{N \to \infty}{\overset{(a)}{\simeq}} \alpha_t \cdot \Psi_t(\boldsymbol{Y}) \cdot \boldsymbol{x}_\star + \Psi_t(\boldsymbol{Y}) \cdot f_t^\perp(\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}) \\
&\overset{N \to \infty}{\overset{(b)}{\simeq}} \alpha_t \cdot \widetilde{\Psi}_t(\boldsymbol{W}) \cdot \boldsymbol{x}_\star + \Psi_t(\boldsymbol{W}) \cdot f_t^\perp(\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}) \\
&\overset{(89)}{=} \widetilde{\boldsymbol{x}}_t
\end{aligned}$$

In the above display, step (a) follows from the induction hypothesis and the Lipschitz continuity of $f_t^\perp$ (which is implied by the Lipschitz continuity of f_t; see Definition 4 and (84)). In step (b), we appealed to Lemma 6. Indeed, Lemma 8 (claim (2)) guarantees that the orthogonality requirement (87) imposed in Lemma 6 is met. This proves the claim (90) by induction and concludes the proof of Theorem 1. □

B.1 Proof of Lemma 5

Proof of Lemma 5. To prove Lemma 5, we assume that the claim of Theorem 1 holds under the additional assumption that the matrix denoisers used in the OAMP algorithm are polynomial functions. We will show that Theorem 1 continues to hold even without this additional assumption. To this end, we consider a general OAMP algorithm from Definition 4:

$$\boldsymbol{x}_t = \Psi_t(\boldsymbol{Y}) \cdot f_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N}, \tag{91}$$

where the matrix denoisers $(\Psi_t)_{t \in \mathbb{N}}$ are possibly non-polynomial functions. Let $(\mathsf{X}_\star, \{\mathsf{X}_t\}_{t \in \mathbb{N}}; \mathsf{A})$ denote the state evolution random variables associated with the OAMP algorithm above. Our goal is to show that:

$$(\boldsymbol{x}_\star, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_t; \boldsymbol{a}) \xrightarrow{W_2} (\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A}) \quad \forall\, t \in \mathbb{N}.$$

Polynomial Approximation. We begin by approximating the matrix denoisers $\Psi_t : \mathbb{R} \mapsto \mathbb{R}$ by polynomials. Recall from Assumption 2, that there is an N-independent finite constant C such that $\|\boldsymbol{W}\|_{\text{op}} \leq C$. Furthermore, thanks to Assumption 1,

$$\|\boldsymbol{Y}\|_{\text{op}} = \left\| \frac{\theta}{n} \boldsymbol{x}_\star \boldsymbol{x}_\star^\top + \boldsymbol{W} \right\|_{\text{op}} \leq C + \frac{\theta \cdot \|\boldsymbol{x}_\star\|^2}{N} \xrightarrow{\mathbb{P}} C + \theta.$$

Hence, the event:

$$\mathcal{E}_N \overset{\text{def}}{=} \{\|\boldsymbol{Y}\|_{\text{op}} \leq K\} \quad \text{with} \quad K \overset{\text{def}}{=} C + \theta + 1$$

occurs with probability tending to 1. We restrict ourselves to this good event in the remainder of the proof. For each $t \in \mathbb{N}$, the Weierstrass Approximation Theorem guarantees the existence of a sequence of approximating polynomial functions $(\Psi_t^{(D)})_{D \in \mathbb{N}}$ of increasing degree, where $\Psi_t^{(D)}$ is a degree D polynomial, with the approximation guarantee:

$$\lim_{D \to \infty} \sup_{|\lambda| \leq K} \left| \Psi_t^{(D)}(\lambda) - \Psi_t(\lambda) \right| = 0, \tag{92}$$

Consider the OAMP algorithm which uses the degree-D approximations $(\Psi_t^{(D)})_{t \in \mathbb{N}}$ as the matrix denoisers:

$$\boldsymbol{x}_t^{(D)} = \left(\Psi_t^{(D)}(\boldsymbol{Y}) - \mathbb{E}[\Psi_t^{(D)}(\Lambda)] \cdot \boldsymbol{I}_N \right) \cdot \left(f_t(\boldsymbol{x}_{<t}^{(D)}; \boldsymbol{a}) - \sum_{s=1}^{t-1} \mathbb{E}[\partial_s f_t(\mathsf{X}_{<t}^{(D)}; \mathsf{A})] \cdot \boldsymbol{x}_s^{(D)} \right) \quad \forall\, t \in \mathbb{N}, \tag{93}$$

where $(\mathsf{X}_\star, (\mathsf{X}_t^{(D)})_{t \in \mathbb{N}}; \mathsf{A})$ are the state evolution random variables associated with the OAMP algorithm above. Since we assume that Theorem 1 holds for OAMP algorithms with polynomial matrix denoisers, we have that:

$$(\boldsymbol{x}_\star, \boldsymbol{x}_1^{(D)}, \ldots, \boldsymbol{x}_t^{(D)}; \boldsymbol{a}) \xrightarrow{W_2} (\mathsf{X}_\star, \mathsf{X}_1^{(D)}, \ldots, \mathsf{X}_t^{(D)}; \mathsf{A}) \quad \forall\, t \in \mathbb{N},\ D \in \mathbb{N}.$$

We make the following claims regarding the polynomial approximation to the original OAMP algorithm and its state evolution random variables.

Claim 1 (Convergence of State Evolution Random Variables)**.** For any $t \in \mathbb{N}$,

$$(\mathsf{X}_\star, \mathsf{X}_1^{(D)}, \ldots, \mathsf{X}_t^{(D)}; \mathsf{A}) \xrightarrow{W_2} (\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A}) \quad \text{as } D \to \infty,$$

where $\xrightarrow{W_2}$ denotes Wasserstein-2 convergence for random variables. Formally, this means that for any test function $h : \mathbb{R}^{t+1} \times \mathbb{R}^k$, independent of N, which satisfies the smoothness condition:

$$|h(x;a) - h(x';a')| \leq L \cdot (\|x - x'\| + \|a - a'\|) \cdot (1 + \|x\| + \|x'\| + \|a\| + \|a'\|) \quad \forall\, x, x' \in \mathbb{R}^{t+1},\ a, a' \in \mathbb{R}^k, \tag{94}$$

for some $L < \infty$, we have:

$$\lim_{D\to\infty} \mathbb{E}[h(\mathsf{X}_1^{(D)}, \ldots, \mathsf{X}_t^{(D)}; \mathsf{A})] = \mathbb{E}[h(\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})].$$

Claim 2 (Convergence of Iterates)**.** For any $t \in \mathbb{N}$,

$$\limsup_{D\to\infty} \operatorname*{plim\,sup}_{N\to\infty} \frac{\|\boldsymbol{x}_t - \boldsymbol{x}_t^{(D)}\|^2}{N} = 0.$$

We first prove Lemma 5 assuming the above results. To do so, we consider a test function $h : \mathbb{R}^t \times \mathbb{R}^k \mapsto \mathbb{R}$, which satisfies the smoothness hypothesis in (94) and show that:

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N h(x_\star[i], x_1[i], \ldots, x[t]; a[i]) = \mathbb{E}h(\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A}).$$

Notice that for any $D \in \mathbb{N}$, we have the decomposition:

$$\left| \frac{1}{N} \sum_{i=1}^N h(x_\star[i], x_1[i], \ldots, x[t]; a[i]) - \mathbb{E}h(\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A}) \right| \leq (i) + (ii) + (iii), \tag{95}$$

where,

$$\begin{aligned}
(i) &\stackrel{\text{def}}{=} \left| \frac{1}{N} \sum_{i=1}^N h(x_\star[i], x_1[i], \ldots, x[t]; a[i]) - \frac{1}{N} \sum_{i=1}^N h(x_\star[i], x_1^{(D)}[i], \ldots, x_t^{(D)}[i]; a[i]) \right| \\
(ii) &\stackrel{\text{def}}{=} \left| \frac{1}{N} \sum_{i=1}^N h(x_\star[i], x_1^{(D)}[i], \ldots, x_t^{(D)}[i]; a[i]) - \mathbb{E}[h(\mathsf{X}_\star, \mathsf{X}_1^{(D)}, \ldots, \mathsf{X}_t^{(D)}; \mathsf{A})] \right| \\
(iii) &\stackrel{\text{def}}{=} \left| \mathbb{E}[h(\mathsf{X}_\star, \mathsf{X}_1^{(D)}, \ldots, \mathsf{X}_t^{(D)}; \mathsf{A})] - \mathbb{E}[h(\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})] \right|.
\end{aligned}$$

We consider each of the terms above. From Claim 2 and (94), we conclude that:

$$\lim_{D\to\infty} \operatorname*{plim\,sup}_{N\to\infty} (i) = 0.$$

Since we assume that Theorem 1 holds for OAMP algorithms with polynomial matrix denoisers,

$$\operatorname*{plim}_{N\to\infty} (ii) = 0.$$

Finally, Claim 1 implies that:

$$\lim_{D\to\infty} (iii) = 0.$$

We let $N \to \infty$ and $D \to \infty$ in (95) and use the results above to conclude that:

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N} \sum_{i=1}^N h(x_\star[i], x_1[i], \ldots, x[t]; a[i]) = \mathbb{E}h(\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A}),$$

as desired. To finish the proof of the lemma, we need to prove Claim 1 and Claim 2.

Convergence of State Evolution Random Variables (Proof of Claim 1). Our goal is to show that:

$$(\mathsf{X}_\star, \mathsf{X}_1^{(D)}, \dots, \mathsf{X}_t^{(D)}; \mathsf{A}) \xrightarrow{W_2} (\mathsf{X}_\star, \mathsf{X}_1, \dots, \mathsf{X}_t; \mathsf{A}) \quad \text{as } D \to \infty \quad \forall\, t \in \mathbb{N}, \tag{96}$$

We show this by induction. As our induction hypothesis, we assume that (96) holds for some $t \in \mathbb{N}$ and show that:

$$(\mathsf{X}_\star, \mathsf{X}_1^{(D)}, \dots, \mathsf{X}_{t+1}^{(D)}; \mathsf{A}) \xrightarrow{W_2} (\mathsf{X}_\star, \mathsf{X}_1, \dots, \mathsf{X}_{t+1}; \mathsf{A}) \quad \text{as } D \to \infty.$$

To this end, we consider a test function $h : \mathbb{R}^{t+2} \times \mathbb{R}^k \mapsto \mathbb{R}$ which satisfies the smoothness requirement in (94) and verify that:

$$\lim_{D\to\infty} \mathbb{E}[h(\mathsf{X}_\star, \mathsf{X}_1^{(D)}, \dots, \mathsf{X}_{t+1}^{(D)}; \mathsf{A})] = \mathbb{E}[h(\mathsf{X}_\star, \mathsf{X}_1, \dots, \mathsf{X}_{t+1}; \mathsf{A})]. \tag{97}$$

Recall from Definition 4 that the joint distribution of the state evolution random variables is given by:

$$(\mathsf{X}_\star; \mathsf{A}) \sim \pi, \quad \mathsf{X}_i = \beta_i \mathsf{X}_\star + \mathsf{Z}_i, \quad \mathsf{X}_i^{(D)} = \beta_i^{(D)} \cdot \mathsf{X}_\star + \mathsf{Z}_i^{(D)} \quad \forall\, i \in [t+1],$$

where $(\mathsf{Z}_1, \dots, \mathsf{Z}_{t+1})$ and $(\mathsf{Z}_1^{(D)}, \dots, \mathsf{Z}_{t+1}^{(D)})$ are Gaussian random vectors, sampled independently of $(\mathsf{X}_\star; \mathsf{A})$:

$$(\mathsf{Z}_1, \dots, \mathsf{Z}_{t+1}) \sim \mathcal{N}(0, \Sigma_{t+1}), \quad (\mathsf{Z}_1^{(D)}, \dots, \mathsf{Z}_{t+1}^{(D)}) \sim \mathcal{N}\left(0, \Sigma_{t+1}^{(D)}\right).$$

In the above equations, coefficients $(\beta_i)_{i\in[t+1]}$ and $(\beta_i^{(D)})_{i\in[t+1]}$ are given by:

$$\beta_i \stackrel{\text{def}}{=} \mathbb{E}[\mathsf{X}_\star \mathsf{F}_i] \cdot \mathbb{E}\big[\Psi_i(\Lambda_\nu)\big], \quad \beta_i^{(D)} \stackrel{\text{def}}{=} \mathbb{E}[\mathsf{X}_\star \mathsf{F}_i^{(D)}] \cdot \Big(\mathbb{E}\big[\Psi_i^{(D)}(\Lambda_\nu)\big] - \mathbb{E}\big[\Psi_i^{(D)}(\Lambda)\big]\Big) \quad \forall\, i \in [t+1],$$

and the entries of the covariance matrices $\Sigma_{t+1}, \Sigma_{t+1}^{(D)}$ are given by:

$$\begin{aligned}
(\Sigma_{t+1})_{i,j} &= \mathbb{E}[\mathsf{X}_\star \mathsf{F}_i] \cdot \mathbb{E}[\mathsf{X}_\star \mathsf{F}_j] \cdot \mathrm{Cov}[\Psi_i(\Lambda_\nu), \Psi_j(\Lambda_\nu)] \\
&\qquad + (\mathbb{E}[\mathsf{F}_i \mathsf{F}_j] - \mathbb{E}[\mathsf{X}_\star \mathsf{F}_i]\mathbb{E}[\mathsf{X}_\star \mathsf{F}_j]) \cdot \mathrm{Cov}[\Psi_i(\Lambda), \Psi_j(\Lambda)] \quad \forall\, i, j \in [t+1], \\
\left(\Sigma_{t+1}^{(D)}\right)_{i,j} &= \mathbb{E}[\mathsf{X}_\star \mathsf{F}_i^{(D)}] \cdot \mathbb{E}[\mathsf{X}_\star \mathsf{F}_j^{(D)}] \cdot \mathrm{Cov}[\Psi_i^{(D)}(\Lambda_\nu), \Psi_j^{(D)}(\Lambda_\nu)] \\
&\qquad + (\mathbb{E}[\mathsf{F}_i^{(D)} \mathsf{F}_j^{(D)}] - \mathbb{E}[\mathsf{X}_\star \mathsf{F}_i]\mathbb{E}[\mathsf{X}_\star \mathsf{F}_j]) \cdot \mathrm{Cov}[\Psi_i^{(D)}(\Lambda), \Psi_j^{(D)}(\Lambda)] \quad \forall\, i, j \in [t+1],
\end{aligned}$$

where $\Lambda \sim \mu$ and $\Lambda_\nu \sim \nu$ and:

$$\mathsf{F}_i \stackrel{\text{def}}{=} f_i(\mathsf{X}_1, \dots, \mathsf{X}_{i-1}; \mathsf{A}), \quad \mathsf{F}_i^{(D)} \stackrel{\text{def}}{=} f_i(\mathsf{X}_1^{(D)}, \dots, \mathsf{X}_{i-1}^{(D)}; \mathsf{A}) - \sum_{j=1}^{i-1} \mathbb{E}[\partial_j f_i(\mathsf{X}_1^{(D)}, \dots, \mathsf{X}_{i-1}^{(D)}; \mathsf{A})] \cdot \mathsf{X}_j^{(D)}.$$

By a weak convergence and uniform integrability argument, (97) follows if we can show that:

$$\lim_{D\to\infty} \beta_i^{(D)} = \beta_i, \quad \lim_{D\to\infty} \left(\Sigma_{t+1}^{(D)}\right)_{i,j} = (\Sigma_{t+1})_{i,j} \quad \forall\, i, j \in [t+1]. \tag{98}$$

Indeed, from (92) we deduce that:

$$\begin{aligned}
&\lim_{D\to\infty} \mathbb{E}\big[\Psi_t^{(D)}(\Lambda)\big] = \mathbb{E}\big[\Psi_t(\Lambda)\big], \quad \lim_{D\to\infty} \mathbb{E}\big[\Psi_t^{(D)}(\Lambda_\nu)\big] = \mathbb{E}\big[\Psi_t(\Lambda_\nu)\big], \\
&\lim_{D\to\infty} \mathrm{Cov}[\Psi_i^{(D)}(\Lambda), \Psi_j^{(D)}(\Lambda)] = \mathrm{Cov}[\Psi_i(\Lambda), \Psi_j(\Lambda)], \quad \lim_{D\to\infty} \mathrm{Cov}[\Psi_i^{(D)}(\Lambda_\nu), \Psi_j^{(D)}(\Lambda_\nu)] = \mathrm{Cov}[\Psi_i(\Lambda_\nu), \Psi_j(\Lambda_\nu)].
\end{aligned}$$

Furthermore, using the induction hypothesis $(\mathsf{X}_\star, \mathsf{X}_1^{(D)}, \dots, \mathsf{X}_t^{(D)}; \mathsf{A}) \xrightarrow{W_2} (\mathsf{X}_\star, \mathsf{X}_1, \dots, \mathsf{X}_t; \mathsf{A})$ and the fact that the functions $f_1, \dots, f_t$ are continuously differentiable and Lipschitz, we conclude that:

$$\lim_{D\to\infty} \mathbb{E}[\mathsf{X}_\star \mathsf{F}_i^{(D)}] = \mathbb{E}[\mathsf{X}_\star \mathsf{F}_i], \quad \lim_{D\to\infty} \mathbb{E}[\mathsf{F}_i^{(D)} \mathsf{F}_j^{(D)}] = \mathbb{E}[\mathsf{F}_i \mathsf{F}_j].$$

Plugging in these limits into the formulae for $(\beta_i^{(D)})_{i\in[t+1]}$ and $\Sigma_{t+1}, \Sigma_{t+1}^{(D)}$, immediately yields the desired conclusion (98) and hence proves (97). This completes the proof of the claim (96) by induction.

Convergence of Iterates (Proof of Claim 2). We need to show that the iterates generated by the algorithm in (93) approximate the iterates generated by the original OAMP algorithm (91) in the sense that,

$$\limsup_{D\to\infty} \operatorname*{plim\,sup}_{N\to\infty} \frac{\|\boldsymbol{x}_t - \boldsymbol{x}_t^{(D)}\|^2}{N} = 0 \quad \forall\, t \in \mathbb{N}.$$

We will shorthand asymptotic approximation statements like the above using the notation:

$$\boldsymbol{x}_t^{(D)} \overset{N,\,D\to\infty}{\simeq} \boldsymbol{x}_t \quad \forall\, t \in \mathbb{N}. \tag{99}$$

We will show the above claim using induction. As our induction hypothesis, we will assume that the claim holds for all the iterates generated before step t of the OAMP algorithm:

$$\boldsymbol{x}_s^{(D)} \overset{N,\,D\to\infty}{\simeq} \boldsymbol{x}_s \quad \forall\, s < t. \tag{100}$$

To complete the inductive proof, we need to show that the claim applies to the iterate generated at step t:

$$\boldsymbol{x}_t^{(D)} \overset{N,\,D\to\infty}{\simeq} \boldsymbol{x}_t.$$

Recalling the update equation for $\boldsymbol{x}_t$ from (91), we have:

$$\begin{aligned}
\boldsymbol{x}_t &= \Psi_t(\boldsymbol{Y}) \cdot f_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{a}) \\
&\overset{N,\,D\to\infty}{\underset{(a)}{\simeq}} \left(\Psi_t^{(D)}(\boldsymbol{Y}) - \mathbb{E}[\Psi_t^{(D)}(\Lambda)] \cdot \boldsymbol{I}_N\right) \cdot f_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{a}) \\
&\overset{N,\,D\to\infty}{\underset{(b)}{\simeq}} \left(\Psi_t^{(D)}(\boldsymbol{Y}) - \mathbb{E}[\Psi_t^{(D)}(\Lambda)] \cdot \boldsymbol{I}_N\right) \cdot \left(f_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{a}) - \sum_{s=1}^{t-1} \mathbb{E}[\partial_s f_t(\mathsf{X}_1^{(D)}, \ldots, \mathsf{X}_{t-1}^{(D)}; \mathsf{A})] \cdot \boldsymbol{x}_s\right) \\
&\overset{N,\,D\to\infty}{\underset{(c)}{\simeq}} \left(\Psi_t^{(D)}(\boldsymbol{Y}) - \mathbb{E}[\Psi_t^{(D)}(\Lambda)] \cdot \boldsymbol{I}_N\right) \cdot \left(f_t(\boldsymbol{x}_1^{(D)}, \ldots, \boldsymbol{x}_{t-1}^{(D)}; \boldsymbol{a}) - \sum_{s=1}^{t-1} \mathbb{E}[\partial_s f_t(\mathsf{X}_1^{(D)}, \ldots, \mathsf{X}_{t-1}^{(D)}; \mathsf{A})] \cdot \boldsymbol{x}_s^{(D)}\right) \\
&\overset{(93)}{=} \boldsymbol{x}_t^{(D)}.
\end{aligned}$$

In the above display:

1. The approximation in step (a) relies on the observations:

$$\limsup_{D\to\infty} \operatorname*{plim\,sup}_{N\to\infty} \|\Psi_t(\boldsymbol{Y}) - \Psi_t^{(D)}(\boldsymbol{Y})\|_{\text{op}} \leq \lim_{D\to\infty} \sup_{|\lambda|\leq K} \left|\Psi_t^{(D)}(\lambda) - \Psi_t(\lambda)\right| \overset{(92)}{=} 0,$$
$$\lim_{D\to\infty} \mathbb{E}[\Psi_t^{(D)}(\Lambda)] = \mathbb{E}[\Psi_t(\Lambda)] \overset{\text{Def. 4}}{=} 0.$$

2. The approximation in step (b) uses Claim 1 and the fact that f_t is Lipschitz to conclude that:

$$\lim_{D\to\infty} \mathbb{E}[\partial_s f_t(\mathsf{X}_1^{(D)}, \ldots, \mathsf{X}_{t-1}^{(D)}; \mathsf{A})] = \mathbb{E}[\partial_s f_t(\mathsf{X}_1, \ldots, \mathsf{X}_{t-1}; \mathsf{A})] = 0,$$

   where the last equality follows by recalling that the iterate denoisers satisfy the divergence-free requirement $\mathbb{E}[\partial_s f_t(\mathsf{X}_1, \ldots, \mathsf{X}_{t-1}; \mathsf{A})] = 0$ (recall Definition 4).

3. The approximation in step (c) relies on the induction hypothesis (100) and the fact that the functions f_t are Lipschitz.

Hence, (99) holds by induction. This concludes the proof of Lemma 5. □

B.2 Proof of Lemma 6

Proof of Lemma 6. . Since any polynomial can be written as a linear combination of monomials, it suffices to prove the claims of the lemma when $\Psi(\lambda) = \lambda^d \;\forall\; \lambda \;\in \mathbb{R}$ for some $d \in \mathbb{N}$.

Proof of Claim (1). We need to show that there exists a sequence of polynomials $(Q_d)_{d\in\mathbb{N}} : \mathbb{R} \mapsto \mathbb{R}$ such that:

$$\boldsymbol{Y}^d \cdot \boldsymbol{x}_\star \overset{N \to \infty}{\simeq} Q_d(\boldsymbol{W}) \cdot \boldsymbol{x}_\star \quad \forall\, d \in \mathbb{N}. \tag{101}$$

We will show that (101) holds by induction on d. We assume that (101) holds for some $d \in \mathbb{N}$ and show that the claim also holds for $d+1$. Recalling that $\boldsymbol{Y} = \theta \cdot \frac{\boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T}}{N} + \boldsymbol{W}$, we have:

$$\begin{aligned}
\boldsymbol{Y}^{d+1} \cdot \boldsymbol{x}_\star &= \left(\theta \cdot \frac{\boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T}}{N} + \boldsymbol{W}\right) \cdot \boldsymbol{Y}^d \cdot \boldsymbol{x}_\star \\
&\overset{N \to \infty}{\simeq}{}^{(a)} \left(\theta \cdot \frac{\boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T}}{N} + \boldsymbol{W}\right) \cdot Q_d(\boldsymbol{W}) \cdot \boldsymbol{x}_\star \\
&= \theta \cdot \frac{\boldsymbol{x}_\star^\top Q_d(\boldsymbol{W}) \boldsymbol{x}_\star}{N} + \boldsymbol{W} Q_d(\boldsymbol{W}) \cdot \boldsymbol{x}_\star \\
&\overset{N \to \infty}{\simeq}{}^{(b)} \theta \cdot \mathbb{E}[Q_d(\Lambda)] + \boldsymbol{W} Q_d(\boldsymbol{W}) \cdot \boldsymbol{x}_\star.
\end{aligned}$$

In the above display, step (a) follows from the induction hypothesis $\boldsymbol{Y}^d \cdot \boldsymbol{x}_\star \overset{N \to \infty}{\simeq} Q_d(\boldsymbol{W}) \cdot \boldsymbol{x}_\star$. and step (b) uses a standard result on the concentration of quadratic forms of rotationally invariant matrices (see Fact 2 in Appendix F). We define the polynomial $Q_{d+1} : \mathbb{R} \mapsto \mathbb{R}$ as:

$$Q_{d+1}(\lambda) = \theta \cdot \mathbb{E}[Q_d(\Lambda)] + \lambda Q_d(\lambda) \quad \forall\, \lambda \in \mathbb{R}.$$

Hence, $\boldsymbol{Y}^{d+1} \cdot \boldsymbol{x}_\star \overset{N \to \infty}{\simeq} Q_{d+1}(\boldsymbol{W}) \cdot \boldsymbol{x}_\star$, as desired. This proves the first claim made in the lemma by induction.

Proof of Claim (2). The proof of the second claim follows from an analogous induction argument. Specifically, we show by induction that:

$$\boldsymbol{Y}^d \cdot \boldsymbol{v} \overset{N \to \infty}{\simeq} \boldsymbol{W}^d \cdot \boldsymbol{v} \quad \forall\, d \in \mathbb{N}.$$

As before, we assume the claim holds for some $d \in \mathbb{N}$, and we verify it for $d+1$:

$$\begin{aligned}
\boldsymbol{Y}^{d+1} \cdot \boldsymbol{v} &= \left(\theta \cdot \frac{\boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T}}{N} + \boldsymbol{W}\right) \cdot \boldsymbol{Y}^d \cdot \boldsymbol{v} \\
&\overset{N \to \infty}{\simeq}{}^{(a)} \left(\theta \cdot \frac{\boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T}}{N} + \boldsymbol{W}\right) \cdot \boldsymbol{W}^d \cdot \boldsymbol{v} \\
&= \theta \cdot \frac{\boldsymbol{x}_\star^\top \boldsymbol{W}^d \boldsymbol{v}}{N} + \boldsymbol{W}^{d+1} \cdot \boldsymbol{x}_\star \\
&\overset{N \to \infty}{\simeq}{}^{(b)} \boldsymbol{W}^{d+1} \cdot \boldsymbol{v},
\end{aligned}$$

where step (a) follows from the induction hypothesis and step (b) uses the assumption $\boldsymbol{x}_\star^\top \boldsymbol{W}^d \boldsymbol{v}/N \xrightarrow{\mathbb{P}} 0$. This concludes the proof of this lemma. □

B.3 Proof of Lemma 7

Proof of Lemma 7. We prove the second claim, the proof of the first claim is analogous. Since the spectral measure of $\boldsymbol{W}$ converges to μ (Assumption 2),

$$\frac{\mathrm{Tr}[\widetilde{\Psi}_s(\boldsymbol{W})\widetilde{\Psi}_t(\boldsymbol{W})]}{N} \xrightarrow{\mathbb{P}} \mathbb{E}[\widetilde{\Psi}_s(\Lambda)\widetilde{\Psi}_t(\Lambda)], \quad \frac{\mathrm{Tr}[\widetilde{\Psi}_s(\boldsymbol{W})\widetilde{\Psi}_t(\boldsymbol{W})]}{N} \xrightarrow{\mathbb{P}} \mathbb{E}[\widetilde{\Psi}_s(\Lambda)\widetilde{\Psi}_t(\Lambda)], \quad \Lambda \sim \mu.$$

To obtain a more convenient formula for the limit which does not involve the transformed polynomials $(\widetilde{\Psi}_t)_{t\in\mathbb{N}}$, we observe that:

$$\mathbb{E}[\widetilde{\Psi}_s(\Lambda)\widetilde{\Psi}_t(\Lambda)] \overset{\text{(a)}}{=} \operatorname*{plim}_{N\to\infty} \frac{\boldsymbol{x}_\star^\top \widetilde{\Psi}_s(\boldsymbol{W})\widetilde{\Psi}_t(\boldsymbol{W})\boldsymbol{x}_\star}{N} \overset{\text{Lem. 6}}{=} \operatorname*{plim}_{N\to\infty} \frac{\boldsymbol{x}_\star^\top \Psi_s(\boldsymbol{Y})\Psi_t(\boldsymbol{Y})\boldsymbol{x}_\star}{N}.$$

where the equality in step (a) follows from standard concentration results for quadratic forms of rotationally invariant matrices (see Fact 2 in Appendix F). Notice that the RHS of the above display can be written as an expectation with respect to ν_N, the spectral measure of $\boldsymbol{Y}$ in the direction of $\boldsymbol{x}_\star$. Indeed, if $\lambda_{1:N}(\boldsymbol{Y})$ and $\boldsymbol{u}_{1:N}(\boldsymbol{Y})$ denote the eigenvalues and corresponding eigenvectors of $\boldsymbol{Y}$:

$$\frac{\boldsymbol{x}_\star^\top \Psi_s(\boldsymbol{Y})\Psi_t(\boldsymbol{Y})\boldsymbol{x}_\star}{N} = \frac{1}{N}\sum_{i=1}^N \langle \boldsymbol{x}_\star, \boldsymbol{u}_i(\boldsymbol{Y})\rangle^2 \cdot \Psi_s(\lambda_i(\boldsymbol{Y}))\Psi_t(\lambda_i(\boldsymbol{Y})) = \int_{\mathbb{R}} \Psi_s(\lambda)\Psi_t(\lambda)\ \nu_N(\mathrm{d}\lambda).$$

By Lemma 1 (item (1)), we know that ν_N converges weakly to ν. Hence,

$$\operatorname*{plim}_{N\to\infty} \frac{\operatorname{Tr}[\widetilde{\Psi}_s(\boldsymbol{W})\widetilde{\Psi}_t(\boldsymbol{W})]}{N} = \operatorname*{plim}_{N\to\infty} \frac{\boldsymbol{x}_\star^\top \widetilde{\Psi}_s(\boldsymbol{W})\widetilde{\Psi}_t(\boldsymbol{W})\boldsymbol{x}_\star}{N} = \mathbb{E}[\Psi_s(\Lambda_\nu)\Psi_t(\Lambda_\nu)], \quad \Lambda_\nu \sim \nu,$$

as claimed. □

B.4 Proof of Lemma 8

Proof. Recall that we started with a general OAMP algorithm (cf. Definition 4) with polynomial matrix denoising functions:

$$\boldsymbol{x}_t = \Psi_t(\boldsymbol{Y}) \cdot f_t(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N}, \tag{102}$$

with associated state evolution random variables

$$(\mathsf{X}_\star, (\mathsf{X}_t)_{t\in\mathbb{N}}; \mathsf{A}),$$

and constructed an auxiliary OAMP algorithm:

$$\widetilde{\boldsymbol{x}}_t = \alpha_t \cdot \widetilde{\Psi}_t(\boldsymbol{W}) \cdot \boldsymbol{x}_\star + \Psi_t(\boldsymbol{W}) \cdot f_t^\perp(\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}) \quad \forall\, t \in \mathbb{N}, \tag{103}$$

which is intended as an easy-to-analyze approximation to the original OAMP algorithm in (102). In the above display:

$$\alpha_t \overset{\text{def}}{=} \mathbb{E}[\mathsf{X}_\star f_t(\mathsf{X}_1, \ldots, \mathsf{X}_{t-1}; \mathsf{A})], \quad f_t^\perp(x_1, \ldots, x_{t-1}; x_\star, a) \overset{\text{def}}{=} f_t(x_1, \ldots, x_{t-1}; a) - \alpha_t x_\star, \tag{104}$$

and $(\widetilde{\Psi}_t)_{t\in\mathbb{N}}$ are the transformed polynomials associated with matrix denoisers $(\Psi_t)_{t\in\mathbb{N}}$, which were constructed in Lemma 6. Our goal is to show that for any $t \in \mathbb{N}$ and any $i \in \mathbb{N}$,

$$(\boldsymbol{x}_\star, \widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_t; \boldsymbol{a}) \xrightarrow{W_2} (\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A}), \quad \frac{\left\langle \boldsymbol{W}^i f_t^\perp(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}), \boldsymbol{x}_\star \right\rangle}{N} \xrightarrow{\mathbb{P}} 0.$$

We will show that the auxiliary OAMP algorithm can be re-written as an algorithm whose dynamics have characterized in prior work [25, Theorem 2]. To this end, we fix a $i \in \mathbb{N}$ and a $t \in \mathbb{N}$. Introduce the random variables:

$$\Lambda \sim \mu, \quad \Lambda_\nu \sim \nu.$$

For any $s \le t$, we define:

$$\boldsymbol{v}_s \overset{\text{def}}{=} \left(\widetilde{\Psi}_s(\boldsymbol{W}) - \mathbb{E}[\widetilde{\Psi}_s(\Lambda)] \cdot \boldsymbol{I}_N\right) \cdot \boldsymbol{x}_\star, \quad \boldsymbol{w}_s \overset{\text{def}}{=} \Psi_s(\boldsymbol{W}) \cdot f_s^\perp(\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_{s-1}; \boldsymbol{x}_\star, \boldsymbol{a}), \tag{105a}$$

$$\boldsymbol{w}_{t+1} \overset{\text{def}}{=} (\boldsymbol{W}^i - \mathbb{E}[\Lambda^{\mathsf{i}}] \cdot \boldsymbol{I}_N) \cdot f_t^\perp(\widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}). \tag{105b}$$

Comparing (103) and (105), we conclude that:

$$\widetilde{\boldsymbol{x}}_s = \beta_s \cdot \boldsymbol{x}_\star + \alpha_s \cdot \boldsymbol{v}_s + \boldsymbol{w}_s \quad \forall\, s \;\le\; t, \tag{106}$$

where we defined:

$$\beta_s \overset{\text{def}}{=} \alpha_s \cdot \mathbb{E}[\widetilde{\Psi}_s(\Lambda)] \overset{\text{Lem. 7}}{=} \alpha_s \cdot \mathbb{E}[\Psi_s(\Lambda_\nu)] \quad \forall\, s \le t.$$

Hence, (105) can be re-expressed as:

$$\begin{aligned}
\boldsymbol{v}_s &= \Big(\widetilde{\Psi}_s(\boldsymbol{W}) - \mathbb{E}[\widetilde{\Psi}_s(\Lambda)] \cdot \boldsymbol{I}_N\Big) \cdot \boldsymbol{x}_\star, \\
\boldsymbol{w}_s &\overset{(106)}{=} \Psi_s(\boldsymbol{W}) \cdot f_s^\perp(\beta_1 \boldsymbol{x}_\star + \alpha_1 \boldsymbol{v}_1 + \boldsymbol{w}_1, \ldots, \beta_{s-1}\boldsymbol{x}_\star + \alpha_{s-1}\boldsymbol{v}_{s-1} + \boldsymbol{w}_{s-1}; \boldsymbol{x}_\star, \boldsymbol{a}).
\end{aligned}$$

The above algorithm is an instance of a Vector Approximate Message Passing (VAMP) algorithm, as defined in [25, Section 4.1.2]. Moreover, [25, Theorem 2] shows that:

$$(\boldsymbol{x}_\star, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_t, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_{t+1}; \boldsymbol{a}) \xrightarrow{W_2} (\mathsf{X}_\star, \mathsf{V}_1, \ldots, \mathsf{V}_t, \mathsf{W}_1, \ldots, \mathsf{W}_{t+1}; \mathsf{A}), \tag{107}$$

where the collection of random variables:

$$(\mathsf{X}_\star; \mathsf{A}), \quad (\mathsf{V}_1, \ldots, \mathsf{V}_t), \quad (\mathsf{W}_1, \ldots, \mathsf{W}_{t+1})$$

are mutually independent. The random variables $(\mathsf{V}_1, \ldots, \mathsf{V}_t)$ have zero mean and are jointly Gaussian, and the same is true for the random variables $(\mathsf{W}_1, \ldots, \mathsf{W}_{t+1})$. The covariance matrices of $(\mathsf{V}_1, \ldots, \mathsf{V}_t)$ and $(\mathsf{W}_1, \ldots, \mathsf{W}_{t+1})$ are given by the recursion:

$$\mathbb{E}[\mathsf{V}_s \mathsf{V}_\tau] = \mathrm{Cov}[\widetilde{\Psi}_s(\Lambda), \widetilde{\Psi}_\tau(\Lambda)] \overset{\text{Lem. 7}}{=} \mathrm{Cov}[\Psi_s(\Lambda_\nu), \Psi_\tau(\Lambda_\nu)] \quad \forall\, s, \tau \le t, \tag{108a}$$

$$\mathbb{E}[\mathsf{W}_s \mathsf{W}_\tau] = \mathrm{Cov}[\Psi_s(\Lambda), \Psi_\tau(\Lambda)] \cdot (\mathbb{E}[\mathsf{F}_s \mathsf{F}_t] - \alpha_s \alpha_t) \quad \forall\, s, \tau \le t. \tag{108b}$$

In the above display, for any $s \le t$, $\mathsf{F}_s \overset{\text{def}}{=} f_s(\beta_1 \mathsf{X}_\star + \alpha_1 \mathsf{V}_1 + \mathsf{W}_1, \ldots, \beta_{s-1}\mathsf{X}_\star + \alpha_{s-1}\mathsf{V}_{s-1} + \mathsf{W}_{s-1}; \mathsf{A})$. Consequently,

$$\begin{aligned}
(\boldsymbol{x}_\star, \widetilde{\boldsymbol{x}}_1, \ldots, \widetilde{\boldsymbol{x}}_t; \boldsymbol{a}) &\overset{(106)}{=} (\boldsymbol{x}_\star, \beta_1 \cdot \boldsymbol{x}_\star + \alpha_1 \cdot \boldsymbol{v}_1 + \boldsymbol{w}_1, \ldots, \beta_t \cdot \boldsymbol{x}_\star + \alpha_t \cdot \boldsymbol{v}_t + \boldsymbol{w}_t; \boldsymbol{a}) \\
&\xrightarrow{W_2} (\mathsf{X}_\star, \beta_1 \cdot \mathsf{X}_\star + \alpha_1 \cdot \mathsf{V}_1 + \mathsf{W}_1, \ldots, \beta_t \cdot \mathsf{X}_\star + \alpha_t \cdot \mathsf{V}_t + \mathsf{W}_t; \mathsf{A}) \quad [\text{Using (107)}] \\
&\overset{\text{d}}{=} (\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A}),
\end{aligned} \tag{109}$$

where the distributional equality in the last step follows by comparing (108) and the description of the joint distribution of the state evolution random variables $(\mathsf{X}_\star, \mathsf{X}_1, \ldots, \mathsf{X}_t; \mathsf{A})$ corresponding to the original OAMP algorithm (102) provided in Definition 4. This proves the first claim made in the lemma. To prove the second claim, we observe that:

$$\begin{aligned}
\frac{\big\langle \boldsymbol{W}^i f_t^\perp(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}), \boldsymbol{x}_\star \big\rangle}{N} &\overset{(105)}{=} \frac{\langle \boldsymbol{w}_{t+1}, \boldsymbol{x}_\star \rangle}{N} + \mathbb{E}[\Lambda^i] \cdot \frac{\big\langle f_t^\perp(\widetilde{\boldsymbol{x}}_1, \ldots \widetilde{\boldsymbol{x}}_{t-1}; \boldsymbol{x}_\star, \boldsymbol{a}), \boldsymbol{x}_\star \big\rangle}{N} \\
&\xrightarrow{\mathbb{P}} \mathbb{E}[\mathsf{W}_{t+1} \mathsf{X}_\star] + \mathbb{E}[\Lambda^i] \cdot \mathbb{E}[\mathsf{X}_\star f_t^\perp(\mathsf{X}_1, \ldots, \mathsf{X}_{t-1}; \mathsf{X}_\star, \mathsf{A})] \\
&= 0.
\end{aligned}$$

In the above display, the convergence in the second step follows from (107) and (109). The equality in the last step follows by recalling that W_{t+1} is a mean-zero Gaussian random variable independent of $\mathsf{X}_\star$ and the definition of $f_t^\perp$ in (104) ensures $\mathbb{E}[\mathsf{X}_\star f_t^\perp(\mathsf{X}_1, \ldots, \mathsf{X}_{t-1}; \mathsf{X}_\star, \mathsf{A})] = 0$. □

C State Evolution for Optimal OAMP (Proposition 1)

Under the assumption $\mathbb{E}\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}] \in (0, 1)$, for the readers convenience, we recall that the optimal OAMP algorithm is given by (cf. (17)):

$$\boldsymbol{x}_t = \frac{1}{\sqrt{\omega_t}} \left(1 + \frac{1}{\rho_t}\right) \cdot \Psi_\star(\boldsymbol{Y}; \rho_t) \cdot \bar{\varphi}(\boldsymbol{x}_{t-1}; \boldsymbol{a} | \omega_{t-1}). \tag{110}$$

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

$$\widehat{\boldsymbol{x}}_t \stackrel{\text{def}}{=} \varphi(\boldsymbol{x}_t; \boldsymbol{a}|\omega_t). \tag{111}$$

In the above equations, the matrix denoiser $\Psi_\star$ used by the optimal OAMP algorithm is given by:

$$\Psi_\star(\lambda;\rho) = 1 - \left(\mathbb{E}\left[\frac{\phi(\Lambda)}{\phi(\Lambda)+\rho}\right]\right)^{-1} \cdot \frac{\phi(\lambda)}{\phi(\lambda)+\rho} \quad \forall\, \lambda \in \mathbb{R},\ \rho \in (0,\infty), \tag{112}$$

where $\Lambda \sim \mu$ and the function $\phi : \mathbb{R} \mapsto \mathbb{R}$ was introduced in (7). The parameters ω_t and ρ_t are computed using the following recursion, initialized with $\omega_0 \stackrel{\text{def}}{=} 0$:

$$\rho_t = \mathscr{F}_2(\omega_{t-1}),\ \omega_t = \mathscr{F}_1(\rho_t) \quad \text{where} \quad \mathscr{F}_2(\omega) \stackrel{\text{def}}{=} \frac{1}{\mathsf{dmmse}_\pi(\omega)} - 1, \quad \mathscr{F}_1(\rho) \stackrel{\text{def}}{=} 1 - \frac{\mathbb{E}\left[\frac{1}{\phi(\Lambda)+\rho}\right]}{\mathbb{E}\left[\frac{\phi(\Lambda)}{\phi(\Lambda)+\rho}\right]}. \tag{113}$$

Let $(\mathsf{X}_\star, (\mathsf{X}_t)_{t\in\mathbb{N}}; \mathsf{A})$ denote the state evolution random variables associated with the above algorithm. We will rely on the following intermediate lemma.

Lemma 9. *We have,*

1. *The function $\mathscr{F}_1$ is a continuous non-decreasing function which maps $(0,\infty)$ to $[0,1)$ and satisfies $\lim_{\rho\to\infty} \mathscr{F}_1(\rho) < 1$.*
2. *The function $\mathscr{F}_2$ is a continuous non-decreasing function which maps $[0,1)$ to $(0,\infty)$ and and satisfies $\lim_{\omega\to 1} \mathscr{F}_2(\omega) = \infty$.*

We defer the proof of this lemma to the end of this appendix (Appendix C.1), and present the proof of Proposition 1.

Proof of Proposition 1. We prove each of the two claims below.

Proof of Claim (1). We will show by induction that:

$$\omega_t \in [0,1), \quad \rho_t \in (0,\infty), \quad (\mathsf{X}_\star, \mathsf{X}_t; \mathsf{A}) \text{ form a scalar Gaussian channel with SNR } \omega_t. \tag{114}$$

Assuming this claim, we can compute the asymptotic MSE of the estimator $\widehat{\boldsymbol{x}}_t = \varphi(\boldsymbol{x}_t; \boldsymbol{a}|\omega_t)$ returned by the OAMP algorithm:

$$\operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_\star\|^2}{N} \stackrel{\text{Thm. 1}}{=} \mathbb{E}|\varphi(\mathsf{X}_t; \mathsf{A}|\omega_t) - \mathsf{X}_\star|^2 \stackrel{\text{(a)}}{=} \mathsf{mmse}_\pi(\omega_t).$$

In the above display, step (a) follows by observing that $\varphi(\mathsf{X}_t; \mathsf{A}|\omega_t)$ is the MMSE estimator for the scalar Gaussian channel $(\mathsf{X}_\star, \mathsf{X}_t; \mathsf{A})$ (which has SNR ω_t). We now focus on proving (114) by induction. We assume that the claim (114) holds at step t as the induction hypothesis and show that it also holds at step $t+1$. Indeed, since $\omega_t \in [0,1)$, Lemma 9 guarantees that $\rho_{t+1} = \mathscr{F}_2(\omega_t) \in (0,\infty)$ and $\omega_{t+1} = \mathscr{F}_1(\rho_{t+1}) \in [0,1)$. Next, we verify that $(\mathsf{X}_\star, \mathsf{X}_{t+1}; \mathsf{A})$ form a scalar Gaussian channel with SNR ω_{t+1}. From Definition 4, we know that:

$$\mathsf{X}_{t+1}|(\mathsf{X}_\star; \mathsf{A}) \sim \mathcal{N}\left(\beta_{t+1}\mathsf{X}_\star, \sigma^2_{t+1}\right),$$

where:

$$\beta_{t+1} = \frac{\mathbb{E}[\mathsf{X}_\star \cdot \bar{\varphi}(\mathsf{X}_t; \mathsf{A}|\omega_t)] \cdot \mathbb{E}[\Psi_\star(\Lambda_\nu\,; \rho_{t+1})]}{\sqrt{\omega_{t+1}}(1 - \mathsf{dmmse}_\pi(\omega_{t+1}))},$$

$$\sigma^2_{t+1} = \frac{\{\mathbb{E}[\mathsf{X}_\star \cdot \bar{\varphi}(\mathsf{X}_t; \mathsf{A}|\omega_t)]\}^2 \cdot \mathrm{Var}[\Psi_\star(\Lambda_\nu\,; \rho_{t+1})]}{\omega_{t+1}(1 - \mathsf{dmmse}_\pi(\omega_{t+1}))^2} + \frac{\{\mathbb{E}[\bar{\varphi}^2(\mathsf{X}_t; \mathsf{A}|\omega_t)] - (\mathbb{E}[\mathsf{X}_\star \cdot \bar{\varphi}(\mathsf{X}_t; \mathsf{A}|\omega_t)])^2\} \cdot \mathbb{E}[\Psi^2_\star(\Lambda\,; \rho_{t+1})]}{\omega_{t+1}(1 - \mathsf{dmmse}_\pi(\omega_{t+1}))^2}$$

In the above equations $\mathsf{\Lambda} \sim \mu$ and $\mathsf{\Lambda}_\nu \sim \nu$. We begin by simplifying β_{t+1}. By the induction hypothesis, $(\mathsf{X}_\star, \mathsf{X}_t; \mathsf{A})$ forms a Gaussian channel with SNR ω_t. A convenient property of the DMMSE estimator for a scalar Gaussian channel is the following identity (see Lemma 2 in Appendix A.2):

$$\mathbb{E}[\mathsf{X}_\star \cdot \bar{\varphi}(\mathsf{X}_t; \mathsf{A}|\omega_t)] = \mathbb{E}[\bar{\varphi}^2(\mathsf{X}_t; \mathsf{A}|\omega_t)] = 1 - \mathsf{dmmse}_\pi(\omega_t). \tag{115}$$

To compute $\mathbb{E}[\Psi_\star(\mathsf{\Lambda}_\nu\,; \rho_{t+1})]$, we consider the Lebesgue decomposition of ν into the absolutely continuous part $\nu_\parallel$ and the singular part $\nu_\perp$:

$$\begin{aligned}
\mathbb{E}[\Psi_\star(\mathsf{\Lambda}_\nu\,; \rho_{t+1})] &\overset{(112)}{=} 1 - \left(\mathbb{E}\left[\frac{\phi(\mathsf{\Lambda})}{\phi(\mathsf{\Lambda}) + \rho}\right]\right)^{-1} \cdot \mathbb{E}\left[\frac{\phi(\mathsf{\Lambda}_\nu)}{\phi(\mathsf{\Lambda}_\nu) + \rho_{t+1}}\right] \\
&\overset{(a)}{=} 1 - \left(\mathbb{E}\left[\frac{\phi(\mathsf{\Lambda})}{\phi(\mathsf{\Lambda}) + \rho}\right]\right)^{-1} \cdot \int_{\mathbb{R}} \frac{\phi(\lambda)}{\phi(\lambda) + \rho_{t+1}} \nu_\parallel(\mathrm{d}\lambda) \\
&\overset{(b)}{=} 1 - \frac{\mathbb{E}\left[\frac{1}{\phi(\mathsf{\Lambda}) + \rho_{t+1}}\right]}{\mathbb{E}\left[\frac{\phi(\mathsf{\Lambda})}{\phi(\mathsf{\Lambda}) + \rho_{t+1}}\right]} \overset{(113)}{=} \omega_{t+1}
\end{aligned} \tag{116}$$

In the above display, step (a) follows by recalling that $\phi(\lambda) = 0$ for $\nu_\perp$-a.e. λ (Lemma 1, item (2)) and step (b) follows from recalling that the density of $\nu_\parallel$ is given by $\mu(\cdot)/\phi(\cdot)$ (Lemma 1, item (3)). Plugging in the identities (115) and (116) into the expression for β_{t+1}, we obtain:

$$\beta_{t+1} = \frac{(1 - \mathsf{dmmse}_\pi(\omega_t)) \cdot \omega_{t+1}}{\sqrt{\omega_{t+1}}(1 - \mathsf{dmmse}_\pi(\omega_{t+1}))} = \sqrt{\omega_{t+1}}. \tag{117}$$

Similarly we can also simplify σ^2_{t+1} by exploiting the identities (115) and (116):

$$\sigma^2_{t+1} \overset{(115)}{=} \frac{\mathrm{Var}[\Psi_\star(\mathsf{\Lambda}_\nu\,; \rho_{t+1})] + \frac{\mathsf{dmmse}_\pi(\omega_{t+1})}{1 - \mathsf{dmmse}_\pi(\omega_{t+1})} \cdot \mathbb{E}[\Psi^2_\star(\mathsf{\Lambda}\,; \rho_{t+1})]}{\omega_{t+1}} \tag{118a}$$

$$\overset{(113)}{=} \frac{\mathbb{E}[\Psi^2_\star(\mathsf{\Lambda}_\nu\,; \rho_{t+1})] - (\mathbb{E}[\Psi_\star(\mathsf{\Lambda}_\nu\,; \rho_{t+1})])^2 + \frac{1}{\rho_{t+1}} \cdot \mathbb{E}[\Psi^2_\star(\mathsf{\Lambda}\,; \rho_{t+1})]}{\omega_{t+1}} \tag{118b}$$

$$\overset{(116)}{=} \frac{\mathbb{E}[\Psi^2_\star(\mathsf{\Lambda}_\nu\,; \rho_{t+1})] + \frac{1}{\rho_{t+1}} \cdot \mathbb{E}[\Psi^2_\star(\mathsf{\Lambda}\,; \rho_{t+1})] - \omega^2_{t+1}}{\omega_{t+1}}. \tag{118c}$$

After a straightforward (but tedious) calculation analogous to the one used to derive (116), we obtain the identity:

$$\mathbb{E}[\Psi^2_\star(\mathsf{\Lambda}_\nu\,; \rho_{t+1})] + \tfrac{1}{\rho_{t+1}} \cdot \mathbb{E}[\Psi^2_\star(\mathsf{\Lambda}\,; \rho_{t+1})] = \omega_{t+1}, \tag{118d}$$

which implies that $\sigma^2_{t+1} = 1 - \omega_{t+1}$. Hence,

$$\mathsf{X}_{t+1}|(\mathsf{X}_\star; \mathsf{A}) \sim \mathcal{N}\left(\sqrt{\omega_{t+1}} \cdot \mathsf{X}_\star, 1 - \omega_{t+1}\right), \tag{119}$$

which verifies that $(\mathsf{X}_\star, \mathsf{X}_{t+1}; \mathsf{A})$ forms a Gaussian channel with SNR ω_{t+1}.

Proof of Claim (2). Recall that the sequences $\{\omega_t\}_{t\in\mathbb{N}}$ and $\{\rho_t\}_{t\in\mathbb{N}}$ are generated via the recursion:

$$\rho_t = \mathscr{F}_2(\omega_{t-1}) \overset{\text{def}}{=} \frac{1}{\mathsf{dmmse}_\pi(\omega_{t-1})} - 1, \tag{120a}$$

$$\omega_t = \mathscr{F}_1(\rho_t) \overset{\text{def}}{=} 1 - \left(\mathbb{E}\left[\frac{\phi(\mathsf{\Lambda})}{\phi(\mathsf{\Lambda}) + \rho_t}\right]\right)^{-1} \cdot \mathbb{E}\left[\frac{1}{\phi(\mathsf{\Lambda}) + \rho_t}\right], \tag{120b}$$

where $\omega_0 = 0$. We first prove the monotonicity of $\{\omega_t\}_{t\in\mathbb{N}}$ using an induction argument. Note that $\omega_t = \mathscr{F}_1 \circ \mathscr{F}_2(\omega_{t-1})$. First, note that $\omega_1 \geq \omega_0 = 0$ (recall we have already shown $\omega_t \geq 0$ for all $t \in \mathbb{N}$). Since the composite function $\mathscr{F}_1 \circ \mathscr{F}_2$ is non-decreasing (Lemma 9), it follows that $\mathscr{F}_1 \circ \mathscr{F}_2(\omega_1) \geq \mathscr{F}_1 \circ \mathscr{F}_2(\omega_0)$,

namely, $\omega_2 \geq \omega_1$. Repeating this argument proves that $\{\omega_t\}_{t\in\mathbb{N}}$ is non-decreasing. Moreover, $\{\omega_t\}_{t\in\mathbb{N}}$ is bounded sequence taking values in $[0,1)$ (proved in the first claim). Hence, $\{\omega_t\}_{t\in\mathbb{N}}$ converges to a limit point $\omega_\star \in [0,1]$ which satisfies the equation $\omega_* = \mathscr{F}_1 \circ \mathscr{F}_2(\omega_*)$. We eliminate the possibility that $\omega_\star = 1$ by observing that Lemma 9 guarantees that:

$$\lim_{\omega\to 1} \mathscr{F}_1 \circ \mathscr{F}_2(\omega) = \lim_{\rho\to\infty} \mathscr{F}_1(\rho) < 1.$$

Since the sequence $\{\rho_t\}_{t\in\mathbb{N}}$ is obtained by applying a continuous and non-decreasing function $\mathscr{F}_2$ to a non-decreasing and convergent sequence $\{\omega_t\}_{t\in\mathbb{N}}$, we conclude that $\{\rho_t\}_{t\in\mathbb{N}}$ is non-decreasing and converges to a limit $\rho_* \stackrel{\text{def}}{=} \mathscr{F}_2(\omega_*) \in (0,\infty)$. Observe that the limit points satisfy the equations: $\omega_* = \mathscr{F}_1 \circ \mathscr{F}_2(\omega_*),\ \rho_* = \mathscr{F}_2(\omega_*)$, which can be expressed as $\omega_* = \mathscr{F}_1(\rho_*),\ \rho_* = \mathscr{F}_2(\omega_*)$ as claimed. This concludes the proof of the proposition.

□

C.1 Proof of Lemma 9

Proof of Lemma 9. We prove the claims made about $\mathscr{F}_1$ and $\mathscr{F}_2$ below.

Properties of $\mathscr{F}_1$. Recall that:

$$\mathscr{F}_1(\rho) \stackrel{\text{def}}{=} 1 - \frac{\mathbb{E}\left[\frac{1}{\phi(\Lambda)+\rho}\right]}{\mathbb{E}\left[\frac{\phi(\Lambda)}{\phi(\Lambda)+\rho}\right]} \quad \forall\ \rho\ \in\ (0,\infty).$$

The continuity of $\mathscr{F}_1(\rho)$ follows from the Dominated Convergence Theorem. To verify the monotonicity of $\mathscr{F}_1$, we calculate its derivative:

$$\mathscr{F}_1'(\rho) = \frac{\mathbb{E}\left[\left(\frac{1}{\phi(\Lambda)+\rho}\right)^2\right] \cdot \mathbb{E}\left[\frac{\phi(\Lambda)}{\phi(\Lambda)+\rho}\right] - \mathbb{E}\left[\frac{1}{\phi(\Lambda)+\rho}\right] \cdot \mathbb{E}\left[\frac{\phi(\Lambda)}{(\phi(\Lambda)+\rho)^2}\right]}{\left(\mathbb{E}\left[\frac{\phi(\Lambda)}{\phi(\Lambda)+\rho}\right]\right)^2} \geq 0,$$

where the inequality follows from Chebyshev's association inequality [14, Theorem 2.14], which states that for any two random variables (X,Z) such that $Z \geq 0$ and any two non-decreasing functions f, g we have:

$$\mathbb{E}[Z]\mathbb{E}[f(X)g(X)Z] \geq \mathbb{E}[f(X)Z] \cdot \mathbb{E}[g(X)Z].$$

Indeed, applying this inequality with $X = \phi(\Lambda)$, $f(X) = X$, $g(X) = X + \rho$ and $Z = 1/(\phi(\Lambda)+\rho)^2$ shows that the numerator of $\mathscr{F}_1'(\rho)$ is non-negative. Finally, we check that $\mathscr{F}_1 : (0,\infty) \mapsto [0,1)$. Notice from (C.1) that for any $\rho \in (0,\infty)$, $\mathscr{F}_1(\rho) < 1$. On the other hand, since $\mathscr{F}_1(\rho) \geq 0$ since, thanks to the monotonicity of $\mathscr{F}_1$ we have:

$$\mathscr{F}_1(\rho) \geq \mathscr{F}_1(0) = 1 - \mathbb{E}\left[\frac{1}{\phi(\Lambda)}\right] \stackrel{\text{(a)}}{=} 1 - \nu_{\|}(\mathbb{R}) \geq 0.$$

In the above display, the step (a) follows by recalling from Lemma 1 (item (3)) that the density of $\nu_{\|}$, the absolutely continuous part of ν is given by $\mu(\cdot)/\phi(\cdot)$ and by observing that $\nu_{\|}(\mathbb{R}) \leq \nu(\mathbb{R}) = 1$. Last we compute:

$$\lim_{\rho\to\infty} \mathscr{F}_1(\rho) = 1 - \lim_{\rho\to\infty} \frac{\mathbb{E}\left[\frac{1}{\phi(\Lambda)+\rho}\right]}{\mathbb{E}\left[\frac{\phi(\Lambda)}{\phi(\Lambda)+\rho}\right]} = 1 - \lim_{\rho\to\infty} \frac{\mathbb{E}\left[\frac{1}{\rho^{-1}\cdot\phi(\Lambda)+1}\right]}{\mathbb{E}\left[\frac{\phi(\Lambda)}{\rho^{-1}\cdot\phi(\Lambda)+1}\right]} = 1 - \frac{1}{\mathbb{E}[\phi(\Lambda)]} < 1,$$

as claimed.

Properties of $\mathscr{F}_2$. Recall that:

$$\mathscr{F}_2(\omega) \stackrel{\text{def}}{=} \frac{1}{\mathsf{dmmse}_\pi(\omega)} - 1.$$

The continuity and monotonicity of the DMMSE function $\mathsf{dmmse}_\pi(\cdot)$ (see From Lemma 4 in Appendix A.2) implies that $\mathscr{F}_2$ continuous and non-decreasing. Next, we verify that $\mathscr{F}_2 : [0,1) \mapsto (0,\infty)$. Notice that for any $\omega \in [0,1)$, $\mathscr{F}_2(\omega) \geq \mathscr{F}_2(0) = (\mathsf{dmmse}_\pi(0))^{-1} - 1 = (\mathsf{mmse}_\pi(0))^{-1} - 1 > 0$ (recall that $\mathsf{mmse}_\pi(0) \in (0,1)$ was one of the hypothesis assumed in Proposition 1). Moreover, for any $\omega < 1$, $\mathscr{F}_2(\omega) < \infty$ since:

$$\mathsf{dmmse}_\pi(\omega) \geq \mathsf{mmse}_\pi(\omega) \stackrel{\text{(a)}}{>} \mathsf{mmse}_\pi(1) = 0,$$

where (a) follows from the fact that $\mathsf{mmse}_\pi(\cdot)$ is strictly decreasing if $\mathsf{mmse}_\pi(0) \in (0,1)$ (see Fact 1 in Appendix A.2). Hence, $\mathscr{F}_2 : [0,1) \mapsto (0,\infty)$. This concludes the proof of the lemma. □

D Replica Predictions for Quartic Potential (Proposition 2)

We begin with some preliminary results about the trace ensemble. The Proof of Proposition 2 will be presented in Appendix D.2.

D.1 Preliminaries of Trace Ensemble

When the potential function $V : \mathbb{R} \mapsto \mathbb{R}$ is a polynomial, the function $\phi(\lambda)$ in (7) corresponding to the trace ensemble is a polynomial with degree one less than V, for $\lambda \in \text{Supp}(\mu)$. Lemma 10 below summarizes some useful properties of ϕ.

Lemma 10 (The function $\phi(\lambda)$ for trace ensemble)**.** *Consider the trace ensemble defined in (20). Assume that $V : \mathbb{R} \mapsto \mathbb{R}$ is a degree k polynomial and independent of N, where $k \geq 2$. Let μ be the limiting spectral measure corresponding to the trace ensemble with V. Let $\phi : \mathbb{R} \mapsto \mathbb{R}$ be defined as in (7). Then, the following holds:*

1. *The spectral measure μ is absolutely continuous with density function:*

$$\mu(\lambda) = \frac{1}{2\pi}\sqrt{\left(4Q(\lambda) - (V'(\lambda))^2\right)_+}, \quad \forall \lambda \in \mathbb{R}, \tag{121}$$

where $(x)_+ = \max\{x, 0\}$ and $Q(\lambda)$ satisfies

$$Q(\lambda) = \int \frac{V'(\lambda) - V'(t)}{\lambda - t}\mu(t)\mathrm{d}t. \tag{122}$$

The support of μ is a finite union of finite intervals, and the density function $\mu(\lambda)$ is Holder continuous.

2. *Define $\phi_{poly} : \mathbb{R} \mapsto \mathbb{R}$ as*

$$\phi_{poly}(\lambda) \stackrel{def}{=} 1 - \theta \cdot V'(\lambda) + \theta^2 \cdot Q(\lambda), \quad \forall \lambda \in \mathbb{R}. \tag{123}$$

Then, ϕ_{poly} is a degree $k-1$ polynomial and

$$\phi(\lambda) = \phi_{poly}(\lambda), \quad \forall \lambda \in Supp(\mu), \tag{124}$$

where $\phi(\lambda)$ is defined in (7).

Proof. The expression of the density function of μ can be found in, e.g., [61, Theorem 11.2.1]. Since $V(x)$ is a degree-k polynomial, $V'(\lambda)$ is a degree $k-1$ polynomial and $Q(\lambda)$ defined in (122) is a degree $k-2$ polynomial. Hence, the support $\text{Supp}(\mu) = \left\{\lambda | 4Q(\lambda) - (V'(\lambda))^2 \geq 0\right\}$ must be a union of finite intervals [61, Eq. (11.2.14)]. Furthermore, $\mu(\lambda)$ is a composition of Holder continuous functions and is therefore Holder

continuous. To prove the second claim, we note that the Hilbert transform of $\mu(\lambda)$ is related to the potential function via [61, Theorem 11.2.1]:

$$\pi \cdot \mathscr{H}_\mu(\lambda) = \frac{1}{2} V'(\lambda), \quad \forall \lambda \in \mathrm{Supp}(\mu). \tag{125}$$

Substituting (121) and (125) into (7), we get

$$\phi(\lambda) \overset{\text{def}}{=} \left[1 - \theta\pi\mathscr{H}_\mu(\lambda)\right]^2 + \pi^2\theta^2\mu^2\left(\lambda\right) = 1 - \theta \cdot V'(\lambda) + \theta^2 \cdot Q(\lambda), \quad \forall \lambda \in \mathrm{Supp}(\mu). \tag{126}$$

To check that ϕ_{poly} is a degree $k-1$ polynomial, we note that $V'(\lambda)$ is a degree $k-1$ polynomial and $Q(\lambda)$ in (121) is a degree $k-2$ polynomial. □

Lemma 10 shows that $\phi(\lambda)$ agrees with a polynomial $\phi_{\text{poly}}(\lambda)$ for $\lambda \in \mathrm{Supp}(\mu)$, but they are generally different outside of the support. The following proposition shows that we can replace $\phi(\cdot)$ by the polynomial $\phi_{\text{poly}}(\cdot)$ in the optimal OAMP algorithm with no performance loss asymptotically.

Proposition 6 (A variant of optimal OAMP)**.** *Assume that $\boldsymbol{W}$ is drawn from the trace ensemble with a polynomial potential. Consider a variant of the optimal OAMP algorithm in (17c) where $\phi(\cdot)$ in the matrix denoiser is replaced by $\phi_{poly}(\cdot)$. Then, Proposition 1 still holds for this OAMP algorithm.*

Proof. First of all, the variant of the optimal OAMP algorithm satisfies the required trace-free condition: $\mathbb{E}\left[\Psi_{\text{poly}}(\Lambda)\right] = 0$, where

$$\Psi_{\text{poly}}(\lambda) \overset{\text{def}}{=} 1 - \left(\mathbb{E}\left[\frac{\phi_{\text{poly}}(\Lambda)}{\phi_{\text{poly}}(\Lambda) + \rho}\right]\right)^{-1} \cdot \frac{\phi_{\text{poly}}(\lambda)}{\phi_{\text{poly}}(\lambda) + \rho}, \quad \forall\, \lambda \in \mathbb{R},\ \rho \in (0, \infty).$$

Therefore, the state evolution framework applies to this variant of optimal OAMP. To prove that the state evolution equation for this new variant of OAMP is identical to the original one, it is not difficult to show that it suffices to check the following identity:

$$\frac{\left(\mathbb{E}\big[\Psi_{\text{poly}}(\Lambda_\nu)\big]\right)^2}{\mathbb{E}\big[\Psi^2_{\text{poly}}(\Lambda_\nu)\big] + \rho^{-1} \cdot \mathbb{E}[\Psi^2_{\text{poly}}(\Lambda)]} = \frac{\left(\mathbb{E}\big[\Psi(\Lambda_\nu)\big]\right)^2}{\mathbb{E}\big[\Psi^2(\Lambda_\nu)\big] + \rho^{-1} \cdot \mathbb{E}[\Psi^2(\Lambda)]}.$$

To this end, we shall verify $\mathbb{E}\big[\Psi_{\text{poly}}(\Lambda_\nu)\big] = \mathbb{E}\big[\Psi(\Lambda_\nu)\big]$, $\mathbb{E}\big[\Psi^2_{\text{poly}}(\Lambda_\nu)\big] = \mathbb{E}\big[\Psi^2(\Lambda_\nu)\big]$, and $\mathbb{E}\big[\Psi_{\text{poly}}(\Lambda)\big] = \mathbb{E}\big[\Psi(\Lambda)\big]$. From item 3 of Lemma 10, $\phi(\lambda) = \phi_{\text{poly}}(\lambda)$ for $\lambda \in \mathrm{Supp}(\mu)$. Hence, $\mathbb{E}\big[\Psi_{\text{poly}}(\Lambda)\big] = \mathbb{E}\big[\Psi(\Lambda)\big]$. It remains to prove $\mathbb{E}\big[\Psi_{\text{poly}}(\Lambda_\nu)\big] = \mathbb{E}\big[\Psi(\Lambda_\nu)\big]$. (The proof for $\mathbb{E}\big[\Psi^2_{\text{poly}}(\Lambda_\nu)\big] = \mathbb{E}\big[\Psi^2(\Lambda_\nu)\big]$ is similar.) From Lemma 1,

$$\begin{aligned}
\mathbb{E}\big[\Psi_{\text{poly}}(\Lambda_\nu)\big] &= \int_{\mathbb{R}} \Psi_{\text{poly}}(\lambda)\nu_\perp(\mathrm{d}\lambda) + \int_{\mathbb{R}} \Psi_{\text{poly}}(\lambda)\nu_\parallel(\mathrm{d}\lambda) \\
&= \int_{\mathbb{R}} \Psi_{\text{poly}}(\lambda)\nu_\perp(\mathrm{d}\lambda) + \int_{\mathbb{R}} \Psi(\lambda)\nu_\parallel(\mathrm{d}\lambda),
\end{aligned}$$

where the second equality is from the fact that $\nu_\parallel$ has the same support as μ, and $\phi(\lambda) = \phi_{\text{poly}}(\lambda)$ for $\lambda \in \mathrm{Supp}(\mu)$. It remains to prove

$$\int_{\mathbb{R}} \Psi_{\text{poly}}(\lambda)\nu_\perp(\mathrm{d}\lambda) = \int_{\mathbb{R}} \Psi(\lambda)\nu_\perp(\mathrm{d}\lambda).$$

Lemma 1 shows that the measure $\nu_\perp$ concentrates on the set $S \overset{\text{def}}{=} \{x \in \mathbb{R} : \phi(x) = 0\}$. It is therefore sufficient to prove $\phi_{\text{poly}}(\lambda) = \phi(\lambda)$ for $\lambda \in S$. From (7), we have

$$\pi\mathscr{H}_\mu(\lambda) = \frac{1}{\theta} \quad \text{and} \quad \mu(\lambda) = 0, \quad \forall \lambda \in S.$$

On the other hand, the Stieltjes transform of μ for the trace ensemble with a polynomial potential V satisfies [61, Theorem 11.2.1]

$$\mathcal{S}^2_\mu(z) - V'(z)\mathcal{S}_\mu(z) + Q(z) = 0, \quad \forall z \in \mathbb{C}\backslash\mathbb{R},$$

where the function Q is defined in (122). By the Sohotsky-Plemelj formula [61, Section 2.1],

$$\lim_{\epsilon \downarrow 0} \mathcal{S}_\mu(\lambda + \imath\epsilon) = \pi \mathscr{H}_\mu(\lambda) - \imath \cdot \pi\mu(\lambda), \quad \forall \lambda \in \mathbb{R}.$$

Combining the above three equations leads to the following equation:

$$\frac{1}{\theta^2} - V'(\lambda) \cdot \frac{1}{\theta} + Q(\lambda) = 0, \quad \forall \lambda \in S.$$

Therefore,

$$\phi_{\text{poly}}(\lambda) \stackrel{\text{def}}{=} 1 - \theta \cdot V'(\lambda) + \theta^2 \cdot Q(\lambda) = 0, \quad \forall \lambda \in S.$$

Hence, $\phi_{\text{poly}}(\lambda) = \phi(\lambda)$ for $\lambda \in S$, which completes the proof. □

We next define a function closely related to ϕ_{poly}, which recovers the matrix processing function used in the Bayes-optimal AMP (BAMP) algorithm proposed in Barbier et al. [8].

Definition 7 (Definition of $J(\lambda)$)**.** Let $k \geq 2$ and $V : \mathbb{R} \mapsto \mathbb{R}$ be a degree k polynomial. Let the degree $k-1$ polynomial ϕ_{poly} in (123) be represented as

$$\phi_{\text{poly}}(\lambda) = C_\phi + \sum_{i=1}^{k-1} C_{\phi,i} \lambda^i, \quad \forall \lambda \in \mathbb{R}. \tag{127a}$$

We define the function $J : \mathbb{R} \mapsto \mathbb{R}$ as

$$J(\lambda) \stackrel{\text{def}}{=} -\sum_{i=1}^{k-1} C_{\phi,i} \lambda^i, \quad \forall \lambda \in \mathbb{R}. \tag{127b}$$

Examples of $\phi_{\text{poly}}(\lambda)$ and $J(\lambda)$. We now provide explicit formulas for the functions $\phi_{\text{poly}}(\lambda)$ and $J(\lambda)$ for three polynomial potential functions considered in Barbier et al. [8].

1. *Quadratic potential*:

$$V(\lambda) = \frac{\lambda^2}{2}. \tag{128a}$$

This corresponds to the Gaussian orthogonal ensemble (GOE). The density function and its Hilbert transform are $\mu(\lambda) = \frac{1}{2\pi}\sqrt{4-\lambda^2}$ and $\mathscr{H}_\mu(\lambda) = V'(\lambda)/2\pi$, $\forall \lambda \in [-2, 2]$. From (123) and (127b), we have

$$\phi_{\text{poly}}(\lambda) = 1 + \theta^2 - \theta \cdot \lambda, \tag{128b}$$
$$J(\lambda) = \theta \cdot \lambda. \tag{128c}$$

2. *Quartic potential*:

$$V(\lambda) = \frac{\gamma\lambda^2}{2} + \frac{\kappa\lambda^4}{4}, \tag{129a}$$

where $\gamma \in [0, 1]$ is a parameter and $\kappa = \kappa(\gamma) = \left(8 - 9\gamma + \sqrt{64 - 144\gamma + 108\gamma^2 - 27\gamma^3}\right)/27 \in [0, \frac{16}{27}]$. This choice of κ ensures that the limiting eigenvalue distribution μ has unit variance. Its density function is given by $\mu(\lambda) = (\gamma + 2a^2\kappa + \kappa\lambda^2)\sqrt{4a^2 - \lambda^2}/(2\pi)$, $\forall \lambda \in [-2a, 2a]$, where $a^2 := \left(\sqrt{\gamma^2 + 12\kappa} - \gamma\right)/(6\kappa)$. The functions ϕ_{poly} and J read [66]

$$\phi_{\text{poly}}(\lambda) = \theta^2 a^2 (\gamma + 2a^2\kappa)^2 + 1 - \theta\left(\gamma\lambda - \theta\kappa\lambda^2 + \kappa\lambda^3\right), \tag{129b}$$
$$J(\lambda) = \theta\left(\gamma\lambda - \theta\kappa\lambda^2 + \kappa\lambda^3\right). \tag{129c}$$

3. *Purely sestic potential*:

$$V(\lambda) = \frac{\xi\lambda^6}{6}, \tag{130a}$$

where $\xi = \frac{27}{80}$. Again, this value of ξ ensures that the spectral measure μ has unit variance. Its density function is given by $\mu(\lambda) = (6b^4\xi + 2b^2\xi\lambda^2 + \xi\lambda^4)\sqrt{4b^2 - \lambda^2}/(2\pi)$, $\forall \lambda \in [-2b, 2b]$, where $b^2 := 2/3$. The functions ϕ_{poly} and J are given by [8]

$$\phi_{\text{poly}}(\lambda) = \frac{27}{50}\theta^2 + 1 - \theta\left(-\theta\xi\lambda^2 - \theta\xi\lambda^4 + \xi\lambda^5\right), \tag{130b}$$

$$J(\lambda) = \theta\left(-\theta\xi\lambda^2 - \theta\xi\lambda^4 + \xi\lambda^5\right). \tag{130c}$$

Fig. 5 plots the functions $\phi(\cdot)$ and $\phi_{\text{poly}}(\cdot)$ corresponding to the quartic potential. In this setup, the measure μ is supported on the interval $[-2a, 2a]$ with $2a \approx 1.7$, and it can be shown that ν has a point mass (corresponding to the outlying eigenvalue of $\boldsymbol{Y}$) at $\lambda_o \approx 2.3$. We see from Fig. 5 that $\phi(\cdot)$ and $\phi_{\text{poly}}(\cdot)$ coincide inside the support of μ but differ outside. Moreover, the curves of $\phi(\cdot)$ and $\phi_{\text{poly}}(\cdot)$ intersect precisely at the location of the outlying eigenvalue: $\phi(\lambda_o) = \phi_{\text{poly}}(\lambda_o) = 0$; see the proof of Proposition 6.

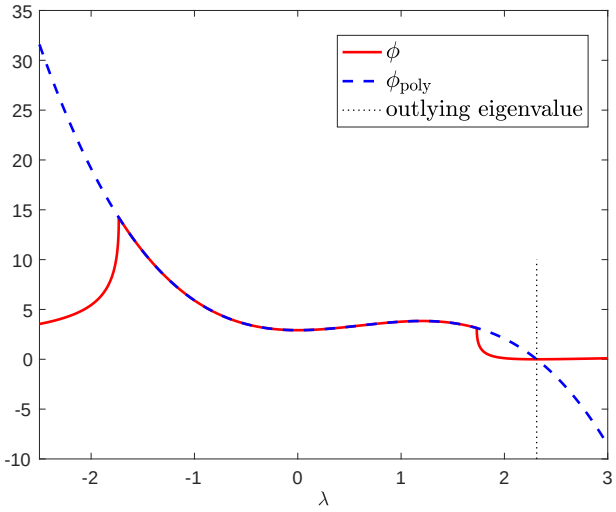


Figure 5: Plot of ϕ and ϕ_{poly} for the quartic ensemble with $\gamma = 0$. $\theta = 1.8$. $\text{Supp}(\mu) = [-2a, 2a]$ with $2a \approx 1.7$. The singular part of the measure ν contains a single point mass at $\lambda_o \approx 2.3$.

D.2 Proof of Proposition 2

We will show that any solution (ρ, ω) to the state evolution (SE) fixed point equation (17f) is a solution to the replica fixed point equation of (22). The proof for the other direction, namely, each solution to the replica fixed point equation is also a solution to the SE equations is similar and thus omitted.

We will introduce a series of manipulations of the SE equations (17f) that eventually lead to the replica equations (22), for the *quartic potential* case. Our proof consists of the following steps:

1. In Lemma 11, we rewrite (17f) as (131). Note that unlike the original SE equation (17f) which only involves the measure μ, the new formulation (131) involves both μ and ν. The introduction of the new formulation (131) is somewhat ad hoc, and is only for the purpose of deriving the replica equations as presented in [8].

2. We recognize that the term in (131a) that involves ν is a resolvent operator for $J(\Lambda_\nu)$. It can be shown that the expectation of this term is the limit of $\frac{1}{N}\boldsymbol{x}_\star^\mathsf{T}\left(\rho\boldsymbol{I}_N + C_\phi\boldsymbol{I}_N - J(\boldsymbol{Y})\right)^{-1}\boldsymbol{x}_\star$ as $N \to \infty$. In the quartic potential case, the function $J(\cdot)$ is a degree three polynomial; see (129c). It is possible

to invoke the matrix inversion lemma to express the term into a resolvent in terms of $J(\boldsymbol{W})$ plus a low-rank perturbation term, and subsequently calculate the asymptotics of each term involved. This step is summarized in Lemma 13.

3. Using Lemma 13 and some algebraic manipulations, we finally arrive at the replica equations we aim to prove in Proposition 2. The derivations are postponed to Section D.2.2.

D.2.1 Auxiliary results

To compare with the replica calculations in [8], we rewrite the fixed point equation (18) into an equivalent form in the following lemma.

Lemma 11 (Alternative fixed point equations for SE)**.** *Assume that* $\mathbb{E}\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}] \in (0,1)$*. Let* (ω,ρ) *be defined as in* (18)*. Then,* (ω,ρ) *satisfies the following equations:*

$$\frac{\omega}{1-\omega} = \left(1-\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda)}\right]\right)\cdot\left(\left(\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda)}\right]\right)^{-1} - \left(\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda_\nu)}\right]\right)^{-1}\right), \tag{131a}$$

$$\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda)}\right] = \mathsf{mmse}_\pi(\omega), \tag{131b}$$

where $\Lambda\sim\mu$ *and* $\Lambda_\nu\sim\nu$*.*

Proof. We prove the two equations seperately.

Proof of (131a)**.** We rewrite the first equation of (18) as

$$\frac{\omega}{1-\omega} = \left(\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda)}\right]\right)^{-1} - (1+\rho) \tag{132a}$$

$$= \left(1-\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda)}\right]\right)\cdot\left(\left(\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda)}\right]\right)^{-1} - \rho\cdot\left(1-\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda)}\right]\right)^{-1}\right). \tag{132b}$$

Comparing (131a) and (132a), we see that it suffices to prove the following

$$\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda_\nu)}\right] = \frac{1}{\rho}\cdot\left(1-\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda)}\right]\right), \quad \forall\rho>0. \tag{133}$$

To prove (133), we note that

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda_\nu)}\right] &= \int_{\mathbb{R}}\frac{1}{\rho+\phi(\lambda)}\nu(\mathrm{d}\lambda)\\
&= \int_{\mathbb{R}}\frac{1}{\rho+\phi(\lambda)}\nu_\perp(\mathrm{d}\lambda) + \int_{\mathbb{R}}\frac{1}{\rho+\phi(\lambda)}\nu_\parallel(\mathrm{d}\lambda)\\
&\overset{(a)}{=} \frac{1}{\rho}\cdot\nu_\perp(\mathbb{R}) + \int_{\mathbb{R}}\frac{1}{\rho+\phi(\lambda)}\nu_\parallel(\mathrm{d}\lambda)\\
&= \frac{1}{\rho}\cdot\big(1-\nu_\parallel(\mathbb{R})\big) + \int_{\mathbb{R}}\frac{1}{\rho+\phi(\lambda)}\nu_\parallel(\mathrm{d}\lambda)\\
&\overset{(b)}{=} \frac{1}{\rho}\cdot\left(1-\int_{\mathbb{R}}\frac{1}{\phi(\lambda)}\mu(\mathrm{d}\lambda)\right) + \int_{\mathbb{R}}\frac{1}{\rho+\phi(\lambda)}\cdot\frac{1}{\phi(\lambda)}\mu(\mathrm{d}\lambda)\\
&= \frac{1}{\rho}\cdot\left(1-\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda)}\right]\right), \quad \Lambda\sim\mu,
\end{aligned}$$

where step (a) and (b) are due to items 2 and 3 of Lemma 1, respectively.

Proof of (131b)**.** From the second equation of (18), we have

$$\rho = \frac{1}{\mathsf{dmmse}_\pi(\omega)} - 1 \overset{(a)}{=} \frac{1}{\mathsf{mmse}_\pi(\omega)} - \frac{1}{1-\omega},$$

where step (a) follows from Lemma 2. Hence,

$$\left(\rho + \frac{1}{1-\omega}\right)^{-1} = \mathsf{mmse}_\pi(\omega). \tag{134}$$

Now recall the first equation of (18):

$$\omega = 1 - \left(\mathbb{E}\left[\frac{\phi(\Lambda)}{\phi(\Lambda)+\rho}\right]\right)^{-1} \cdot \mathbb{E}\left[\frac{1}{\phi(\Lambda)+\rho}\right].$$

By simple algebra, we have

$$\left(\rho + \frac{1}{1-\omega}\right)^{-1} = \mathbb{E}\left[\frac{1}{\phi(\Lambda)+\rho}\right]. \tag{135}$$

Combining (134) and (135) leads to (131b).

□

Note that (131a) involves the term $\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda_\nu)}\right]$, which depends on the function $\phi(\cdot)$. To facilitate further analysis, we rewrite this term using the function $\phi_{\text{poly}}(\lambda) \overset{\text{def}}{=} C_\phi(\lambda) + J(\lambda)$ defined in Appendix D.1; see Lemma 10 and Definition 7. Lemma 12 below summarizes this result. Its proof is similar to the proof of Proposition 6 and omitted.

Lemma 12. *Let μ be the limiting spectral measure associated with the quartic potential. Let C_ϕ and $J(\cdot)$ be defined as in Definition (7). Then, the following holds for any $\rho > 0$:*

$$\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda_\nu)}\right] = \mathbb{E}\left[\frac{1}{\rho+C_\phi-J(\Lambda_\nu)}\right], \quad \Lambda_\nu \sim \nu. \tag{136}$$

The following lemma reformulate the term $\mathbb{E}\left[\frac{1}{\rho+C_\phi-J(\Lambda_\nu)}\right]$, which involves the measure $\Lambda_\nu \sim \nu$, into a (complicated) expression that only depends on $\Lambda \sim \mu$.

Lemma 13. *Consider the quartic potential for which the function $J(\lambda)$ is given in (129c). Then, the following identity holds for any $\rho > 0$:*

$$\mathbb{E}\left[\frac{1}{\rho+C_\phi-J(\Lambda_\nu)}\right] = -\frac{1}{(\kappa\theta^2)^2 d_4 + \gamma\theta^2 + \frac{3\kappa\theta^2 d_2 + (\kappa\theta^2)^2(2d_1d_3-3d_2^2) + (\kappa\theta^2)^3(d_2^3+d_0d_3^2-2d_1d_2d_3)-1}{d_0+\kappa\theta^2 d_1^2 - \kappa\theta^2 d_0 d_2}}, \tag{137}$$

where $\Lambda_\nu \sim \nu$, and κ and γ are parameters for the quartic potential function and θ is the parameter for the observation model, and

$$d_i \overset{def}{=} \mathbb{E}\left[\frac{\Lambda^i}{\rho+\phi(\Lambda)}\right], \quad \forall i \in \mathbb{N} \cup \{0\}. \tag{138}$$

Proof. Recall that the probability measure ν is the limit of the empirical measure $\nu_N \overset{\text{def}}{=} \frac{1}{N}\sum_{i=1}^N \langle \boldsymbol{u}_i(\boldsymbol{Y}), \boldsymbol{x}_\star\rangle^2 \cdot \delta_{\lambda_i(\boldsymbol{Y})}$, where $\{\boldsymbol{u}_i(\boldsymbol{Y})\}_{i\in[N]}$ and $\{\lambda_i(\boldsymbol{Y})\}_{i\in[N]}$ are the eigenvectors and eigenvalues of $\boldsymbol{Y}$. By the convergence of ν_N (Lemma 1), we have

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\boldsymbol{x}_\star^\mathsf{T}\left(c_*\boldsymbol{I}_N - J(\boldsymbol{Y})\right)^{-1}\boldsymbol{x}_\star = \mathbb{E}\left[\frac{1}{c_*-J(\Lambda_\nu)}\right]. \tag{139}$$

where $c_* \overset{\text{def}}{=} \rho + C_\phi$. Recall that $J(\cdot)$ is a degree three polynomial for quartic potential:

$$J(\lambda) = \theta\left(\gamma\lambda - \theta\kappa\lambda^2 + \kappa\lambda^3\right). \tag{140}$$

Based on the above definition of $J(\cdot)$, we can decompose $J(\boldsymbol{Y})$ as $J(\boldsymbol{W})$ plus a low-rank perturbation term:

$$\begin{aligned}
J(\boldsymbol{Y}) &= J\left(\frac{\theta \boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T}}{N} + \boldsymbol{W}\right) \\
&= \gamma\theta \boldsymbol{W} - \kappa\theta^2 \boldsymbol{W}^2 + \kappa\theta \boldsymbol{W}^3 + \frac{\gamma\theta^2}{N}\boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T} + \frac{1}{N}\kappa\theta^3 \boldsymbol{x}_\star^\mathsf{T} \boldsymbol{W} \boldsymbol{x}_\star \cdot \frac{\boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T}}{N} + \frac{1}{N}\kappa\theta^2 \left(\boldsymbol{W}^2 \boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T} + \boldsymbol{x}\boldsymbol{x}^\mathsf{T} \boldsymbol{W}^2 + \boldsymbol{W}\boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T} \boldsymbol{W}\right) \\
&= J(\boldsymbol{W}) + \frac{1}{N}\boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T}\left(\gamma\theta^2 \boldsymbol{I}_N + \kappa\theta^2 \boldsymbol{W}\right) + \frac{1}{N}\kappa\theta^2 \boldsymbol{W}\boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T} \boldsymbol{W} + \frac{1}{N}\kappa\theta^2 \boldsymbol{W}^2 \boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T} + \frac{1}{N}\kappa\theta^3 \boldsymbol{x}_\star^\mathsf{T} \boldsymbol{W}\boldsymbol{x}_\star \cdot \frac{\boldsymbol{x}_\star \boldsymbol{x}_\star^\mathsf{T}}{N} \\
&= J(\boldsymbol{W}) + \frac{1}{N}\boldsymbol{P}\boldsymbol{Q}^\mathsf{T},
\end{aligned}$$

where

$$\boldsymbol{P} \overset{\text{def}}{=} \left[\boldsymbol{x}_\star, \boldsymbol{W}\boldsymbol{x}_\star, \boldsymbol{W}^2 \boldsymbol{x}_\star\right] \in \mathbb{R}^{N\times 3}, \tag{141}$$

$$\boldsymbol{Q} \overset{\text{def}}{=} \left[\left(\left(\gamma\theta^2 + \kappa\theta^3 \frac{\boldsymbol{x}_\star^\mathsf{T} \boldsymbol{W}\boldsymbol{x}_\star}{N}\right)\boldsymbol{I}_N + \kappa\theta^2 \boldsymbol{W}^2\right)\boldsymbol{x}_\star, \kappa\theta^2 \boldsymbol{W}\boldsymbol{x}_\star, \kappa\theta^2 \boldsymbol{x}_\star\right] \in \mathbb{R}^{N\times 3}. \tag{142}$$

Hence,

$$\begin{aligned}
(c_* \boldsymbol{I} - J(\boldsymbol{Y}))^{-1} &= \left(c_* \boldsymbol{I} - J(\boldsymbol{W}) - \frac{1}{N}\boldsymbol{P}\boldsymbol{Q}^\mathsf{T}\right)^{-1} \\
&\overset{(a)}{=} \left(\boldsymbol{A} - \frac{1}{N}\boldsymbol{P}\boldsymbol{Q}^\mathsf{T}\right)^{-1} \\
&\overset{(b)}{=} \boldsymbol{A}^{-1} + \frac{1}{N}\boldsymbol{A}^{-1}\boldsymbol{P}\left(\boldsymbol{I}_3 - \frac{\boldsymbol{Q}^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{P}}{N}\right)^{-1} \boldsymbol{Q}^\mathsf{T}\boldsymbol{A}^{-1},
\end{aligned}$$

where the matrix $\boldsymbol{A}$ in step (a) denotes

$$\boldsymbol{A} \overset{\text{def}}{=} c_* \boldsymbol{I} - J(\boldsymbol{W}), \tag{143}$$

and step (b) is from the Morrison–Woodbury formula. Using the above identity, we obtain the following representation of the LHS of (139):

$$\frac{1}{N}\boldsymbol{x}^\mathsf{T}\left(c_* \boldsymbol{I} - J(\boldsymbol{Y})\right)^{-1}\boldsymbol{x} = \frac{\boldsymbol{x}_\star^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{x}_\star}{N} + \frac{1}{N^2}\boldsymbol{x}_\star^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{P}\left(\boldsymbol{I}_3 - \frac{\boldsymbol{Q}^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{P}}{N}\right)^{-1}\boldsymbol{Q}^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{x}_\star. \tag{144}$$

Therefore, to calculate the limit of $\frac{1}{N}\boldsymbol{x}^\mathsf{T}\left(c_* \boldsymbol{I} - J(\boldsymbol{Y})\right)^{-1}$ as $N \to \infty$, it suffices to derive the limit of the terms $\frac{1}{N}\boldsymbol{x}_\star^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{x}_\star$, $\frac{1}{N}\boldsymbol{x}_\star^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{P} \in \mathbb{R}^{1\times 3}$, $\frac{1}{N}\boldsymbol{Q}^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{P} \in \mathbb{R}^{3\times 3}$ and $\frac{1}{N}\boldsymbol{Q}^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{x}_\star \in \mathbb{R}^{3\times 1}$, which essentially involves computing the limit of $\frac{1}{N}\boldsymbol{x}_\star^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{W}^i\boldsymbol{x}_\star$, for $i = 1, 2$; cf. (141) and (142). For this purpose, we can apply standard results in random matrix theory (e.g., [12, Proposition 9.3]) to obtain

$$\begin{aligned}
\operatorname*{plim}_{N\to\infty} \frac{1}{N}\boldsymbol{x}_\star^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{W}^i\boldsymbol{x}_\star &= \operatorname*{plim}_{N\to\infty} \frac{1}{N}\boldsymbol{x}_\star^\mathsf{T}\left(c_* \boldsymbol{I} - J(\boldsymbol{W})\right)^{-1}\boldsymbol{W}^i\boldsymbol{x}_\star, \quad \forall i \in \mathbb{N}\cup\{0\} \\
&= \mathbb{E}\left[\frac{\Lambda^i}{c_* - J(\Lambda)}\right] \\
&\overset{(a)}{=} \mathbb{E}\left[\frac{\Lambda^i}{\rho + C_\phi - J(\Lambda)}\right] \\
&\overset{\text{def}}{=} d_i,
\end{aligned} \tag{145}$$

where $\Lambda \sim \mu$ denotes the limiting eigenvalue distribution of $\boldsymbol{W}$, and step (a) is from the definition of $c_* \overset{\text{def}}{=} \rho + C_\phi$. Similarly,

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\boldsymbol{x}_\star^\mathsf{T}\boldsymbol{W}\boldsymbol{x}_\star = \mathbb{E}[\Lambda] = 0. \tag{146}$$

Using (145) and (146), and recalling the definitions of the matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$ in (141) and (142), we arrive at

$$\begin{aligned}
&\operatorname*{plim}_{N\to\infty} \frac{1}{N}\boldsymbol{x}_\star^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{P} = [d_0, d_1, d_2],\\
&\operatorname*{plim}_{N\to\infty} \frac{1}{N}\boldsymbol{x}_\star^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{Q} = [\gamma\theta^2 d_0 + \kappa\theta^2 d_2, \kappa\theta^2 d_1, \kappa\theta^2 d_2],\\
&\operatorname*{plim}_{N\to\infty} \frac{1}{N}\boldsymbol{Q}^\mathsf{T}\boldsymbol{A}^{-1}\boldsymbol{P} = \begin{bmatrix} \kappa\theta^2 d_2 + \gamma\theta^2 d_0 & \kappa\theta^2 d_3 + \gamma\theta^2 d_2 & \kappa\gamma^2 d_4 + \gamma\theta^2 d_2 \\ \kappa\theta^2 d_1 & \kappa\theta^2 d_2 & \kappa\theta^2 d_3 \\ \kappa\theta^2 d_0 & \kappa\theta^2 d_1 & \kappa\theta^2 d_2 \end{bmatrix}.
\end{aligned} \tag{147}$$

Combining (144), (145) and (147), and after lengthy but straightforward calculations, we obtain

$$\operatorname*{plim}_{N\to\infty} \frac{1}{N}\boldsymbol{x}_\star^\mathsf{T}\left(c_*\boldsymbol{I} - J(\boldsymbol{Y})\right)^{-1}\boldsymbol{x}_\star = -\frac{1}{(\kappa\theta^2)^2 d_4 + \gamma\theta^2 + \frac{3\kappa\theta^2 d_2 + (\kappa\theta^2)^2(2d_1d_3 - 3d_2^2) + (\kappa\theta^2)^3(d_2^3 + d_0d_3^2 - 2d_1d_2d_3) - 1}{d_0 + \kappa\theta^2 d_1^2 - \kappa\theta^2 d_0 d_2}}. \tag{148}$$

Combining (148) and (139) yields the desired identity in (137). □

D.2.2 Proof of Proposition 2

In what follows, we prove that any solution (ρ, ω) to the state evolution (SE) fixed point equation (17f), with $\omega \in (0,1), \rho \in (0,\infty)$, is a solution to the replica fixed point equation of (22). This is essentially achieved by combining Lemma 11 and Lemma 13. More specifically, the replica fixed point equations in (22), which we aim to prove, can be obtained by replacing the term $\mathbb{E}\left[\frac{1}{\rho+\phi(\Lambda_\nu)}\right]$ in (131a) by the alternative representation of $\mathbb{E}\left[\frac{1}{\rho+C_\phi-J(\Lambda_\nu)}\right]$ in (137), which is valid because of Lemma 12, together with additional algebraic manipulations.

Before we present the detailed proof, we first introduce a few definitions. Let

$$\chi \equiv \chi(\rho) \overset{\text{def}}{=} \frac{(1-d_0)(d_1 + \kappa\theta^2 d_0 d_3 - \kappa\theta^2 d_1 d_2)}{d_0 - \kappa\theta^2 d_0 d_2 + \kappa\theta^2 d_1^2}, \tag{149}$$

where

$$\begin{aligned}
&d_i \overset{\text{def}}{=} \mathbb{E}[\mathsf{H}\Lambda^i] \quad \forall i \in \mathbb{N}\cup\{0\},\\
&\mathsf{H} \overset{\text{def}}{=} \mathsf{H}(\Lambda,\rho) = \frac{1}{\rho + \phi_{\text{poly}}(\Lambda)} = \frac{1}{\rho + C_\phi - J(\Lambda)}\\
&\qquad \overset{(129b)}{=} \frac{1}{\rho + \theta^2 a^2(\gamma + 2a^2\kappa)^2 + 1 - \theta\left(\gamma\Lambda - \theta\kappa\Lambda^2 + \kappa\Lambda^3\right)}.
\end{aligned}$$

The rationale for the above definition of χ will be clear from subsequent derivations. The notations $\chi(\rho)$ and $\mathsf{H}(\Lambda,\rho)$ above highlight the fact that χ and H both depend on the variable ρ, which is defined to be a solution to the SE fixed point equation. We further define

$$m \equiv m(\omega) \overset{\text{def}}{=} 1 - \mathsf{mmse}_\pi(\omega), \tag{150}$$

$$\mathsf{Q} \equiv \mathsf{Q}(\Lambda,\mathsf{H},\rho,\omega) \overset{\text{def}}{=} \kappa\theta^2 m\Lambda^2 + \kappa\theta^2\chi\Lambda - \frac{\kappa\theta^2}{1-m}\mathbb{E}\left[m\Lambda^2\mathsf{H} + \chi\Lambda\mathsf{H}\right] + \frac{m}{1-m}. \tag{151}$$

Again, the notation $m(\omega)$ and $\mathsf{Q}(\Lambda,\mathsf{H},\rho,\omega)$ indicate that m depends on ω, and Q depends on the random variables Λ and H (which is a function of ρ) and the variables (ρ,ω).

To finish the proof, it remains to verify that the variables χ, m, H and Q as defined above satisfy the replica fixed point equations (22).

Proof of (22b). By the definition of H:

$$\begin{aligned}\mathbb{E}[\mathsf{H}] &\overset{(22f)}{=} \mathbb{E}\left[\frac{1}{\rho + C_\phi - J(\Lambda)}\right] \\ &\overset{(124)}{=} \mathbb{E}\left[\frac{1}{\rho + \phi(\Lambda)}\right] \\ &\overset{(131b)}{=} \mathsf{mmse}_\pi(\omega) \\ &\overset{(150)}{=} 1 - m.\end{aligned}$$

This proves (22b).

Proof of (22c). We first verify equation (22b): $\chi = \mathbb{E}[\Lambda\mathsf{Q}\mathsf{H}]$. To this end, we will substitute the definitions of Q and χ into $\mathbb{E}[\Lambda\mathsf{Q}\mathsf{H}]$ and then show that it is equal to χ:

$$\begin{aligned}\mathbb{E}[\Lambda\mathsf{Q}\mathsf{H}] &\overset{(a)}{=} \mathbb{E}\left[\Lambda\left(\kappa\theta^2 m\Lambda^2 + \kappa\theta^2\chi\Lambda - \frac{\kappa\theta^2}{1-m}\mathbb{E}\left[m\Lambda^2\mathsf{H} + \chi\Lambda\mathsf{H}\right] + \frac{m}{1-m}\right)\mathsf{H}\right] \\ &= \kappa\theta^2\left(m\mathbb{E}[\Lambda^3\mathsf{H}] + \chi\mathbb{E}[\Lambda^2\mathsf{H}] - \frac{1}{1-m}\mathbb{E}[\Lambda\mathsf{H}]\left(m\mathbb{E}[\Lambda^2\mathsf{H}] + \chi\mathbb{E}[\Lambda\mathsf{H}]\right)\right) + \frac{m}{1-m}\mathbb{E}[\Lambda\mathsf{H}] \\ &\overset{(b)}{=} \kappa\theta^2\left(md_3 + \chi d_2 - \frac{1}{1-m}d_1(md_2 + \chi d_1)\right) + \frac{m}{1-m}d_1 \\ &\overset{(c)}{=} \kappa\theta^2\left((1-d_0)d_3 + \chi d_2 - \frac{1}{d_0}(1-d_0)d_1d_2 + \chi\frac{1}{d_0}d_1^2\right) + \frac{1-d_0}{d_0}d_1 \\ &\overset{(d)}{=} \frac{(1-d_0)(d_1 + \kappa\theta^2 d_0 d_3 - \kappa\theta^2 d_1 d_2)}{d_0 - \kappa\theta^2 d_0 d_2 + \kappa\theta^2 d_1^2} \\ &\overset{(e)}{=} \chi,\end{aligned} \tag{152}$$

where step (a) is from the definition of Q in (151); step (b) is from the definition of the shorthand d_i: $d_i = \mathbb{E}[\Lambda^i\mathsf{H}]$; step (c) is due to the identity $d_0 = 1 - m$, which further follows from definition $d_0 = \mathbb{E}[\mathsf{H}]$ and the equation $\mathbb{E}[\mathsf{H}] = 1 - m$ (see (22b)); step (d) follows from the definition of χ in (149) and straightforward calculations; step (e) is from the definition of χ in (149). This completes the proof of (22c).

Proof of (22d). The term $\mathbb{E}[\Lambda^2\mathsf{Q}\mathsf{H}]$ appears on the RHS of (22d). We first rewrite it as an explicit function of (d_0, d_1, d_2, d_3) and χ:

$$\begin{aligned}\mathbb{E}[\Lambda^2\mathsf{Q}\mathsf{H}] &= \mathbb{E}\left[\Lambda^2\left(\kappa\theta^2 m\Lambda^2 + \kappa\theta^2\chi\Lambda - \frac{\kappa\theta^2}{1-m}\mathbb{E}\left[m\Lambda^2\mathsf{H} + \chi\Lambda\mathsf{H}\right] + \frac{m}{1-m}\right)\mathsf{H}\right] \\ &\overset{(a)}{=} \kappa\theta^2 m d_4 + \kappa\theta^2 d_3\chi - \frac{\kappa\theta^2}{1-m}\left(md_2^2 + d_1d_2\chi\right) + \frac{m}{1-m}d_2 \\ &\overset{(b)}{=} (1-d_0)\left(\kappa\theta^2 d_4 + \frac{\kappa\theta^2 d_3\chi}{1-d_0} - \frac{\kappa\theta^2}{d_0}\left(d_2^2 + \frac{\chi}{1-d_0}d_1d_2\right) + \frac{1}{d_0}d_2\right),\end{aligned} \tag{153}$$

where step (a) is from the definition of Q in (151); step (b) uses the identity $d_0 = 1 - m$, similar to step (c) of (152). Using (153), we can continue to rewrite the RHS of (22d) as a function of (d_0, d_1, d_2, d_3) and χ:

$$\begin{aligned}\text{RHS of (22d)} &= \kappa\theta^2\left(\frac{m}{1-m}\mathbb{E}[\Lambda^2\mathsf{H}] + \frac{\chi}{1-m}\mathbb{E}[\Lambda\mathsf{H}] + \mathbb{E}[\Lambda^2\mathsf{Q}\mathsf{H}]\right) + \gamma\theta^2 m \\ &= \kappa\theta^2\left(\frac{1}{d_0}(1-d_0)d_2 + \frac{1}{d_0}d_1\chi + \mathbb{E}[\Lambda^2\mathsf{Q}\mathsf{H}]\right) + (1-d_0)\gamma\theta^2 \\ &= (1-d_0)\left((\kappa\theta^2)^2 d_4 + \gamma\theta^2 + \frac{\kappa\theta^2}{d_0(1-d_0)}\left(d_1 + \kappa\theta^2(d_0d_3 - d_1d_2)\right)\chi + \frac{\kappa\theta^2}{d_0}(2d_2 - \kappa\theta^2 d_2^2)\right).\end{aligned} \tag{154}$$

Note that χ is itself a function of (d_0, d_1, d_2, d_3); see (149). Substituting (149) into (154) eliminates the variable χ. After some algebra we finally obtain

$$\begin{aligned}
\text{RHS of (22d)} &\overset{\text{(a)}}{=} (1-d_0)\left((\kappa\theta^2)^2 d_4 + \gamma\theta^2 + \frac{3\kappa\theta^2 d_2 + (\kappa\theta^2)^2(2d_1d_3 - 3d_2^2) + (\kappa\theta^2)^3(d_2^3 + d_0d_3^2 - 2d_1d_2d_3) - 1}{d_0 + \kappa\theta^2 d_1^2 - \kappa\theta^2 d_0 d_2} + \frac{1}{d_0}\right) \\
&\overset{\text{(b)}}{=} (1-d_0)\left(-\left(\mathbb{E}\left[\frac{1}{\rho + C_\phi - J(\Lambda_\nu)}\right]\right)^{-1} + \frac{1}{d_0}\right) \\
&\overset{\text{(c)}}{=} \left(1 - \mathbb{E}\left[\frac{1}{\rho + \phi(\Lambda)}\right]\right)\left\{-\left(\mathbb{E}\left[\frac{1}{\rho + C_\phi - J(\Lambda_\nu)}\right]\right)^{-1} + \left(\mathbb{E}\left[\frac{1}{\rho + \phi(\Lambda)}\right]\right)^{-1}\right\} \\
&\overset{\text{(d)}}{=} \frac{\omega}{1-\omega},
\end{aligned} \tag{155}$$

where step (a) is a consequence of (149) and (154); step (b) is from Lemma 13; step (c) is from the definition of d_0 (see (138)); and step (d) is due to (131) (Lemma 11). This proves (22d).

To conclude, we have verified that the stationary point of the state evolution of OAMP (ρ, ω) satisfies equations (22a)-(22d). The proof for the other direction, namely, each fixed point of (22) such that $\rho > 0$ and $\omega \in (0, 1)$ also satisfies the SE equations (17f) is similar and thus omitted.

E Omitted Proofs for the Optimality Result (Theorem 2)

E.1 Proof of Proposition 3

We consider a given iterative algorithm (Definition 5) of the form:

$$\boldsymbol{r}_t = \Psi_t(\boldsymbol{Y}) \cdot f_t(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{t-1}; \boldsymbol{a}) + g_t(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_{t-1}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N}, \tag{156}$$

which returns the following estimator after t iterations:

$$\widehat{\boldsymbol{r}}_t = \psi_t(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_t; \boldsymbol{a}), \tag{157}$$

Our goal will be to construct a lifted OAMP algorithm (Definition 6), which can approximate the estimator above. Throughout the proof, we will refer to the given iterative algorithm (156) as the *target algorithm*. Our construction of the desired lifted OAMP algorithm proceeds in two steps.

Step 1 (Divergence Removal). We first implement the target algorithm (156) using an *intermediate OAMP algorithm* which takes the form:

$$\boldsymbol{v}_{t,S} = \bar{\Psi}_S(\boldsymbol{Y}) \cdot \bar{F}_t(\boldsymbol{v}_{<t,\bullet}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N},\ S \subset \mathbb{N},\ 1 \le |S| < \infty. \tag{158}$$

In the above display,

1. The iterates generated by the algorithm are indexed by two indices, a time index t, which takes values in $\mathbb{N}$, and a set index S, which can be any finite and non-empty subset of $\mathbb{N}$.
2. For any finite subset $S = \{s_1, s_2, \ldots, s_k\}$ with sorted elements $s_1 < s_2 < \cdots < s_k$, we define the matrix denoiser:
$$\Psi_S(\lambda) \overset{\text{def}}{=} \Psi_{s_k}(\lambda) \cdot \Psi_{s_{k-1}}(\lambda) \cdot \cdots \cdot \Psi_{s_1}(\lambda), \quad \bar{\Psi}_S(\lambda) \overset{\text{def}}{=} \Psi_S(\lambda) - \mathbb{E}_{\Lambda\sim\mu}\left[\Psi_S(\Lambda)\right] \quad \forall\, \lambda \in \mathbb{R}.$$
3. The iterates are generated in the following sequence:
   (a) In step (1), we generate the iterate $\boldsymbol{v}_{1,\{1\}}$.
   (b) In step (2), we generate the new iterates $\boldsymbol{v}_{1,\{1,2\}}$ and $\boldsymbol{v}_{2,\{1\}}, \boldsymbol{v}_{2,\{1,2\}}$.

(c) At step (t), we generate the new iterates:

$$\{\boldsymbol{v}_{1,S} : S \subset [t], t \in S\},\ \{\boldsymbol{v}_{2,S} : S \subset [t], t \in S\}, \ldots, \{\boldsymbol{v}_{t-1,S} : S \subset [t], t \in S\} \text{ and } \{\boldsymbol{v}_{t,S} : S \subset [t]\}.$$

We use the notation $\boldsymbol{v}_{\le t,\bullet}$ to denote the collection of iterates that have been generated at the end of t steps:

$$\boldsymbol{v}_{\le t,\bullet} = \{\boldsymbol{v}_{s,S} : s \le t,\ S \subset [t]\}.$$

The notation $\boldsymbol{v}_{<t,\bullet}$ denotes the collection of iterates that have been generated before the tth step, which is the same as the collection $\boldsymbol{v}_{\le (t-1),\bullet}$.

4. The functions $(F_t)_{t\in\mathbb{N}}$ are continuously differentiable, Lipschitz and satisfy the divergence-free requirement:

$$\mathbb{E}[\partial_{s,S}\bar{F}_t(\mathsf{V}_{<t,\bullet};\mathsf{A})] = 0 \quad \forall\, S \subset [t-1],\ s < t,$$

where $(\mathsf{X}_\star, \{\mathsf{V}_{t,S} : t \in \mathbb{N}, S \subset \mathbb{N}\}; \mathsf{A})$ are the state evolution random variables associated with the intermediate OAMP algorithm in (158).

5. Furthermore, the iterates generated by the intermediate OAMP algorithm in (158) can approximate the iterates of the target algorithm in (156). Specifically, for each $t \in \mathbb{N}$, there is a postprocessing function $H_t : \mathbb{R}^{t\cdot 2^t+k} \mapsto \mathbb{R}$, which is continuously differentiable and Lipschitz, which satisfies:

$$\operatorname*{plim}_{N\to\infty} \frac{\|\boldsymbol{r}_t - H_t(\boldsymbol{v}_{\le t,\bullet}, \boldsymbol{a})\|^2}{N} = 0. \tag{159}$$

The construction of the intermediate OAMP algorithm with the properties stated above is presented in the proof of the following Lemma 14, whose proof is deferred to Section E.1.1.

Lemma 14. *For any target algorithm of the form (156), there is an intermediate OAMP algorithm of the form (158), whose iterates can approximate the iterates generated by the target algorithm in the sense of (159).*

Step 2: Polynomial Approximation. Next, by approximating the functions $(\Psi_S)_{S\subset\mathbb{N}}$ by polynomials, we construct a lifted OAMP algorithm, whose iterates can approximate the iterates generated by the intermediate OAMP algorithm (158). This construction is presented in the following Lemma 5, whose proof is deferred to Section E.1.2.

Lemma 15. *For each $D \in \mathbb{N}$, there is a degree-D lifted AMP algorithm (Definition 6):*

$$\boldsymbol{w}_{t,i}^{(D)} = \left(\boldsymbol{Y}^i - \mathbb{E}_{\Lambda\sim\mu}[\Lambda^i]\cdot \boldsymbol{I}_N\right)\cdot \bar{F}_t^{(D)}(\boldsymbol{w}_{1,\bullet}^{(D)}, \ldots, \boldsymbol{w}_{t-1,\bullet}^{(D)}; \boldsymbol{a}) \quad t \in \mathbb{N},\ i \in [D], \tag{160}$$

which can approximate the iterates produced by the intermediate OAMP algorithm (158) in the following sense: for any $D \in \mathbb{N}$, $t \in \mathbb{N}$, and any finite subset $S \subset \mathbb{N}$, there is a homogeneous linear postprocessing function $h_{t,S}^{(D)} : \mathbb{R}^D \mapsto \mathbb{R}$ which has the property that the vector $h_{t,S}^{(D)}(\boldsymbol{w}_{t,\bullet}^{(D)})$ approximates the intermediate OAMP iterate $\boldsymbol{v}_{t,S}$ in the sense:

$$\lim_{D\to\infty} \operatorname*{plim\,sup}_{N\to\infty} \frac{\|h_{t,S}^{(D)}(\boldsymbol{w}_{t,\bullet}^{(D)}) - \boldsymbol{v}_{t,S}\|^2}{N} = 0 \tag{161}$$

With these preliminary results, we can now present the proof of Proposition 3.

Proof of Proposition 3. Throughout the proof, we shorthand an approximation result like (159) using the notation:

$$\boldsymbol{r}_t \overset{N\to\infty}{\simeq} H_t(\boldsymbol{v}_{\le t,\bullet}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N}. \tag{162}$$

Likewise, we shorthand an approximation result like (161) using:

$$\boldsymbol{v}_{t,S} \overset{N,\,D\to\infty}{\simeq} \boldsymbol{v}_{t,S}^{(D)} \quad \forall\, t \in \mathbb{N},\ S \subset \mathbb{N},\ 1 \le |S| < \infty. \tag{163}$$

Lemma 14 and Lemma 15 together show that the lifted OAMP algorithm on (160) can approximate the iterates of the target algorithm in (156):

$$\boldsymbol{r}_t \overset{(a)}{\overset{N\to\infty}{\simeq}} H_t(\boldsymbol{v}_{1,\bullet}, \dots, \boldsymbol{v}_{t,\bullet}; \boldsymbol{a}) \overset{(b)}{\overset{N,\,D\to\infty}{\simeq}} H_t(\{h_{s,R}^{(D)}(\boldsymbol{w}_{s,\bullet}^{(D)}) : s \le t, R \subset [t]\}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N}. \tag{164}$$

In the above display, the approximation in step (a) follows from (159) and the approximation in step (b) follows by replacing each iterate $\boldsymbol{v}_{s,R},\ s \in [t], R \subset [t]$ of the intermediate OAMP algorithm by its approximation in (161), and exploiting the fact that H_t is a Lipschitz function. For convenience, we define the composite postprocessing functions $\{H_t^{(D)} : t \in \mathbb{N}, D \in \mathbb{N}\}$:

$$H_t^{(D)}(w_{\le t,\bullet}; a) \overset{\text{def}}{=} H_t(\{h_{s,R}^{(D)}(w_{s,\bullet}) : s \le t, R \subset [t]\}; a) \quad \forall\, w_{\le t,\bullet} \in \mathbb{R}^{tD},\ a \in \mathbb{R}^k.$$

With this definition, (164) can be expressed as:

$$\boldsymbol{r}_t \overset{N,\,D\to\infty}{\simeq} H_t^{(D)}(\boldsymbol{w}_{\le t,\bullet}^{(D)}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N}.$$

Plugging in the above approximation in (157), we have the following approximation for the estimator $\widehat{\boldsymbol{r}}_t$ returned by the target algorithm:

$$\widehat{\boldsymbol{r}}_t \overset{N,\,D\to\infty}{\simeq} \widetilde{\boldsymbol{w}}_t^{(D)} \overset{\text{def}}{=} \psi_t(H_1^{(D)}(\boldsymbol{w}_{1,\bullet}^{(D)}; \boldsymbol{a}), H_2^{(D)}(\boldsymbol{w}_{\le 2,\bullet}^{(D)}; \boldsymbol{a}), \dots, H_t^{(D)}(\boldsymbol{w}_{\le t,\bullet}^{(D)}; \boldsymbol{a}); \boldsymbol{a}). \tag{165}$$

Hence, we have constructed an estimator $\widetilde{\boldsymbol{w}}_t^{(D)}$ which can be computed by running a degree-D lifted OAMP algorithm for t steps, and approximates the estimator $\widehat{\boldsymbol{r}}_t$ returned by the target iterative algorithm after t steps. This proves the claim of Proposition 3. □

We now present the proofs of Lemma 14 and Lemma 15.

E.1.1 Divergence Removal (Proof of Lemma 14)

Proof of Lemma 14. Consider a target iterative algorithm of the form:

$$\boldsymbol{r}_t = \Psi_t(\boldsymbol{Y}) \cdot f_t(\boldsymbol{r}_1, \dots, \boldsymbol{r}_{t-1}; \boldsymbol{a}) + g_t(\boldsymbol{r}_1, \dots, \boldsymbol{r}_{t-1}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N}. \tag{166}$$

Our goal is to construct an intermediate OAMP algorithm of the form:

$$\boldsymbol{v}_{t,S} = \bar{\Psi}_S(\boldsymbol{Y}) \cdot \bar{F}_t(\boldsymbol{v}_{<t,\bullet}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N},\ S \subset \mathbb{N},\ 1 \le |S| < \infty. \tag{167}$$

For each $t \in \mathbb{N}$, we wish to design a postprocessing function $H_t : \mathbb{R}^{t \cdot 2^t + k} \mapsto \mathbb{R}$ which allows us to approximate the iterates of the target algorithm with the iterates of the intermediate OAMP algorithm in the sense:

$$\operatorname*{plim}_{N\to\infty} \frac{\|\boldsymbol{r}_t - H_t(\boldsymbol{v}_{\le t,\bullet}, \boldsymbol{a})\|^2}{N} \to 0 \quad \forall\, t \in \mathbb{N}. \tag{168}$$

We will shorthand approximation statements of the form (159) using the notation:

$$\boldsymbol{r}_t \overset{N\to\infty}{\simeq} H_t(\boldsymbol{v}_{\le t,\bullet}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N}. \tag{169}$$

The construction proceeds by induction. Suppose that we have specified the first t functions $(F_s)_{s\le t}$ and the post-processing functions $(H_s)_{s\le t}$ for t iterations. We now extend our construction to the $(t+1)$th iteration. Recall the update rule for $\boldsymbol{r}_{t+1}$:

$$\begin{aligned}\boldsymbol{r}_{t+1} &= \Psi_{t+1}(\boldsymbol{Y}) \cdot f_{t+1}(\boldsymbol{r}_{\le t}; \boldsymbol{a}) + g_{t+1}(\boldsymbol{r}_{\le t}; \boldsymbol{a}) \\ &\overset{N\to\infty}{\simeq} \Psi_{t+1}(\boldsymbol{Y}) \cdot f_{t+1}(H_1(\boldsymbol{v}_{\le 1,\bullet}; \boldsymbol{a}), \dots, H_t(\boldsymbol{v}_{\le t,\bullet}; \boldsymbol{a})) + g_{t+1}(H_1(\boldsymbol{v}_{\le 1,\bullet}; \boldsymbol{a}), \dots, H_t(\boldsymbol{v}_{\le t,\bullet}; \boldsymbol{a})),\end{aligned}$$

where the approximation in the last step follows from the induction hypothesis $\boldsymbol{r}_s \overset{N \to \infty}{\simeq} H_s(\boldsymbol{v}_{\leq s,\bullet}; \boldsymbol{a})$ for any $s \leq t$, along with the fact that $\|\Psi_t(\boldsymbol{Y})\|_{\mathrm{op}}$ is uniformly bounded as $N \to \infty$ and the functions f_{t+1}, g_{t+1} are Lipschitz. We introduce the composite functions:

$$F_{t+1}(v_{\leq t,\bullet}; a) \overset{\text{def}}{=} f_{t+1}(H_1(v_{\leq 1,\bullet}; a), H_2(v_{\leq 2,\bullet}; a), \ldots, H_t(v_{\leq t,\bullet}; a); a) \quad v_{\leq t,\bullet} \in \mathbb{R}^{t2^t},\ a \in \mathbb{R}^k,$$
$$G_{t+1}(v_{\leq t,\bullet}; a) \overset{\text{def}}{=} g_{t+1}(H_1(v_{\leq 1,\bullet}; a), H_2(v_{\leq 2,\bullet}; a), \ldots, H_t(v_{\leq t,\bullet}; a); a) \quad v_{\leq t,\bullet} \in \mathbb{R}^{t2^t},\ a \in \mathbb{R}^k.$$

With these definitions we have that,

$$\boldsymbol{r}_{t+1} \overset{N \to \infty}{\simeq} \Psi_{t+1}(\boldsymbol{Y}) \cdot F_{t+1}(\boldsymbol{v}_{\leq t,\bullet}; \boldsymbol{a}) + G_{t+1}(\boldsymbol{v}_{\leq t,\bullet}; \boldsymbol{a}). \tag{170}$$

The above expression suggests that we should use the function F_{t+1} in the $t+1$ step of the intermediate OAMP algorithm. However, since F_{t+1} is not divergence-free (cf. Definition 4), we define:

$$\bar{F}_{t+1}(v_{\leq t,\bullet}; a) \overset{\text{def}}{=} F_t(v_{\leq t,\bullet}; a) - \sum_{s=1}^{t} \sum_{R \subset [t]} b_{s,R} \cdot v_{s,R} \quad \forall\ v_{\leq t,\bullet} \in \mathbb{R}^{t \cdot 2^t},\ a \in \mathbb{R}^k.$$

In the above display, we introduced the scalars:

$$b_{s,R} \overset{\text{def}}{=} \mathbb{E}\left[\partial_{s,R} F_{t+1}(\mathsf{V}_{\leq t,\bullet}; \mathsf{A})\right] \quad \forall\ s \leq t,\ R \subset [t],$$

where $(\mathsf{X}_\star, \{\mathsf{V}_{s,R} : s \leq t, R \subset [t]\}\,; \mathsf{A})$ denote the state evolution random variables corresponding to the iterates of the intermediate OAMP algorithm that have already been constructed (by the induction hypothesis). Notice that by construction, $\bar{F}_{t+1}$ satisfies the divergence-free requirement:

$$\mathbb{E}[\partial_{s,S} \bar{F}_{t+1}(\mathsf{V}_{\leq t;\bullet}; \mathsf{A})] = 0 \quad \forall\ s \leq t,\ S \subset [t].$$

We can now extend our construction of the intermediate OAMP algorithm to the $(t+1)$th step:

$$\boldsymbol{v}_{t+1,S} = \bar{\Psi}_S(\boldsymbol{Y}) \cdot \bar{F}_{t+1}(\boldsymbol{v}_{\leq t,\bullet}; \boldsymbol{a}) \quad \forall\ S \subset \mathbb{N} \quad 1 \leq |S| < \infty. \tag{171}$$

Next, we define the postprocessing function H_{t+1}. For convenience, we introduce the vectors $(\boldsymbol{F}_\tau)_{\tau \in [t+1]}$:

$$\bar{\boldsymbol{F}}_\tau \overset{\text{def}}{=} \bar{F}_\tau(\boldsymbol{v}_{<\tau,\bullet}; \boldsymbol{a}) \quad \forall\ \tau \in [t+1]. \tag{172}$$

With these definitions, we can express the approximation in (170) as:

$$\begin{aligned}\boldsymbol{r}_{t+1} &\overset{N \to \infty}{\simeq} \Psi_{t+1}(\boldsymbol{Y}) \cdot F_{t+1}(\boldsymbol{v}_{\leq t,\bullet}; \boldsymbol{a}) + G_{t+1}(\boldsymbol{v}_{\leq t,\bullet}; \boldsymbol{a}) \\ &\overset{N \to \infty}{\simeq} \Psi_{t+1}(\boldsymbol{Y}) \cdot \boldsymbol{F}_{t+1} + \sum_{s=1}^{t} \sum_{S \subset [t]} b_{s,S} \cdot \Psi_{t+1}(\boldsymbol{Y}) \cdot \boldsymbol{v}_{s,S} + G_{t+1}(\boldsymbol{v}_{\leq t,\bullet}; \boldsymbol{a}).\end{aligned} \tag{173}$$

For any $S \subset [t+1]$ we define:

$$m_S \overset{\text{def}}{=} \mathbb{E}_{\Lambda \sim \mu}[\Psi_S(\Lambda)].$$

With this definition, we can express the approximation in (173) as:

$$\begin{aligned}
\boldsymbol{r}_{t+1} &\overset{N}{\underset{\infty}{\to}} \Psi_{t+1}(\boldsymbol{Y}) \cdot \boldsymbol{F}_{t+1} + \sum_{s=1}^{t} \sum_{S \subset [t]} b_{s,S} \cdot \Psi_{t+1}(\boldsymbol{Y}) \cdot \boldsymbol{v}_{s,S} + G_{t+1}(\boldsymbol{v}_{\leq t,\bullet}; \boldsymbol{a}) \\
&= \Psi_{t+1}(\boldsymbol{Y}) \cdot \boldsymbol{F}_{t+1} + \sum_{s=1}^{t} \sum_{S \subset [t]} b_{s,S} \cdot \Psi_{t+1}(\boldsymbol{Y}) \cdot (\Psi_S(\boldsymbol{Y}) - m_S \boldsymbol{I}_N) \cdot \boldsymbol{F}_s + G_{t+1}(\boldsymbol{v}_{\leq t,\bullet}; \boldsymbol{a}) \\
&= \Psi_{t+1}(\boldsymbol{Y}) \cdot \boldsymbol{F}_{t+1} + \sum_{s=1}^{t} \sum_{S \subset [t]} b_{s,S} \cdot (\Psi_{\{t+1\} \cup S}(\boldsymbol{Y}) - m_S \Psi_{t+1}(\boldsymbol{Y})) \cdot \boldsymbol{F}_s + G_{t+1}(\boldsymbol{v}_{\leq t,\bullet}; \boldsymbol{a}) \\
&\overset{(167),(171)}{=} \boldsymbol{v}_{t+1,\{t+1\}} + \left(\sum_{s=1}^{t} \sum_{S \subset [t]} b_{s,S} \cdot (\boldsymbol{v}_{s,S \cup \{t+1\}} - m_S \cdot \boldsymbol{v}_{s,\{t+1\}}) \right) + G_{t+1}(\boldsymbol{v}_{\leq t,\bullet}; \boldsymbol{a}) \\
&\qquad + m_{\{t+1\}} \cdot \boldsymbol{F}_{t+1} + \left(\sum_{s=1}^{t} \sum_{S \subset [t]} b_{s,S} \cdot (m_{S \cup \{t+1\}} - m_S \cdot m_{\{t+1\}}) \cdot \boldsymbol{F}_s \right).
\end{aligned} \tag{174}$$

Recalling (172),

$$\begin{aligned}
\boldsymbol{r}_{t+1} &\overset{N}{\underset{\infty}{\to}} \boldsymbol{v}_{t+1,\{t+1\}} + \left(\sum_{s=1}^{t} \sum_{S \subset [t]} b_{s,S} \cdot (\boldsymbol{v}_{s,S \cup \{t+1\}} - m_S \cdot \boldsymbol{v}_{s,\{t+1\}}) \right) + G_{t+1}(\boldsymbol{v}_{\leq t,\bullet}; \boldsymbol{a}) \\
&\qquad + m_{\{t+1\}} \cdot \bar{F}_{t+1}(\boldsymbol{v}_{\leq t,\bullet}; \boldsymbol{a}) + \left(\sum_{s=1}^{t} \sum_{S \subset [t]} b_{s,S} \cdot (m_{S \cup \{t+1\}} - m_S \cdot m_{\{t+1\}}) \cdot \bar{F}_s(\boldsymbol{v}_{\leq s,\bullet}; \boldsymbol{a}) \right).
\end{aligned}$$

In light of the above expression, we define the postprocessing function for the $t+1$th iteration as:

$$\begin{aligned}
H_{t+1}(v_{\leq t+1,\bullet}; a) &\overset{\text{def}}{=} v_{t+1,\{t+1\}} + \left(\sum_{s=1}^{t} \sum_{S \subset [t]} b_{s,S} \cdot (v_{s,S \cup \{t+1\}} - m_S \cdot v_{s,\{t+1\}}) \right) \\
&\qquad + G_{t+1}(v_{\leq t,\bullet}; \boldsymbol{a}) + m_{\{t+1\}} \cdot \breve{F}_{t+1}(v_{\leq t,\bullet}; a) + \left(\sum_{s=1}^{t} \sum_{S \subset [t]} b_{s,S} \cdot (m_{S \cup \{t+1\}} - m_S \cdot m_{\{t+1\}}) \cdot \breve{F}_s(v_{\leq s,\bullet}; a) \right).
\end{aligned}$$

Hence, we can approximate the $t+1$th iterate of the target algorithm using the iterates of the intermediate OAMP algorithm:

$$\boldsymbol{r}_{t+1} \overset{N}{\underset{\infty}{\to}} H_{t+1}(\boldsymbol{v}_{\leq t+1,\bullet}; \boldsymbol{a}).$$

This completes the construction of the intermediate OAMP algorithm with all the desired properties by induction. □

E.1.2 Polynomial Approximation (Proof of Lemma 15)

Proof of Lemma 15. Recall the intermediate OAMP algorithm takes the form:

$$\boldsymbol{v}_{t,S} = \bar{\Psi}_S(\boldsymbol{Y}) \cdot \bar{F}_t(\boldsymbol{v}_{<t,\bullet}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N},\ S \subset \mathbb{N},\ 1 \leq |S| < \infty. \tag{175}$$

To implement this algorithm using a degree-D lifted OAMP algorithm, we begin by repeating the arguments used in the proof of Lemma 5 to construct an OAMP algorithm of the form:

$$\begin{aligned}
\boldsymbol{v}_{t,S}^{(D)} = \left(\Psi_S^{(D)}(\boldsymbol{Y}) - \mathbb{E}_{\Lambda \sim \mu}[\Psi_S^{(D)}(\Lambda)] \cdot \boldsymbol{I}_N \right) \cdot \left(\bar{F}_t(\boldsymbol{v}_{<t,\bullet}^{(D)}; \boldsymbol{a}) - \sum_{s=1}^{t-1} \sum_{R \subset [t-1]} \mathbb{E}[\partial_{s,R} \bar{F}_t(\mathsf{V}_{<t,\bullet}^{(D)}; \mathsf{A})] \cdot \boldsymbol{v}_{s,R}^{(D)} \right) \\
\forall\, t \in \mathbb{N},\ S \subset \mathbb{N},\ 1 \leq |S| < \infty,
\end{aligned} \tag{176}$$

which approximates the iterates generated by the intermediate OAMP algorithm in the sense that,

$$\limsup_{D\to\infty} \operatorname*{plim\,sup}_{N\to\infty} \frac{\|\boldsymbol{v}_{t,S}^{(D)} - \boldsymbol{v}_{t,S}\|^2}{N} = 0 \quad \forall\, t \in \mathbb{N},\ S \subset \mathbb{N},\ 1 \le |S| < \infty. \tag{177}$$

In the above display:

- The matrix denoiser $\Psi_S^{(D)} : \mathbb{R} \mapsto \mathbb{R}$ is a degree-D polynomial which approximates the matrix denoiser $\Psi_S : \mathbb{R} \mapsto \mathbb{R}$ and is constructed using the Weierstrass approximation theorem.
- $(\mathsf{X}_\star, \{\mathsf{V}_{t,S}^{(D)} : t \in \mathbb{N}, S \subset \mathbb{N}\}; \mathsf{A})$ denote the state evolution random variables associated with (176).

We claim that there is a degree-D lifted AMP algorithm (Definition 6):

$$\boldsymbol{w}_{t,i}^{(D)} = \left(\boldsymbol{Y}^i - \mathbb{E}_{\Lambda\sim\mu}[\Lambda^i] \cdot \boldsymbol{I}_N\right) \cdot \bar{F}_t^{(D)}(\boldsymbol{w}_{<t,\bullet}^{(D)}; \boldsymbol{a}) \quad \forall\, t \in \mathbb{N},\ i \in [D], \tag{178}$$

along with homogeneous linear postprocessing maps

$$h_{t,S}^{(D)} : \mathbb{R}^D \mapsto \mathbb{R} \quad t \in \mathbb{N},\ S \subset \mathbb{N},\ 1 \le |S| < \infty,$$

which can exactly recover the iterates generated by the degree-D approximation to the intermediate OAMP algorithm (176):

$$\boldsymbol{v}_{t,S}^{(D)} = h_{t,S}^{(D)}(\boldsymbol{w}_{t,\bullet}^{(D)}) \quad \forall\, t \in \mathbb{N},\ S \subset \mathbb{N},\ 1 \le |S| < \infty. \tag{179}$$

Observe that (177) and (179) immediately imply Lemma 15. Indeed, for any $t \in \mathbb{N}$ and any finite subset $S \subset \mathbb{N}$:

$$\limsup_{D\to\infty} \operatorname*{plim\,sup}_{N\to\infty} \frac{\|h_{t,S}^{(D)}(\boldsymbol{w}_{t,\bullet}^{(D)}) - \boldsymbol{v}_{t,S}\|^2}{N} \overset{(179)}{=} \limsup_{D\to\infty} \operatorname*{plim\,sup}_{N\to\infty} \frac{\|\boldsymbol{v}_{t,S}^{(D)} - \boldsymbol{v}_{t,S}\|^2}{N} \overset{(177)}{=} 0,$$

as desired. Hence, the remainder of the proof consists of constructing the lifted OAMP algorithm and proving (179).

Construction of the lifted OAMP Algorithm. We will construct the desired lifted OAMP algorithm (178) by induction. As our induction hypothesis, we will assume that we have specified the functions $\bar{F}_1^{(D)}, \dots, \bar{F}_{t-1}^{(D)}$, as well as the homogenous linear postprocessing maps:

$$h_{s,S}^{(D)} : \mathbb{R}^D \mapsto \mathbb{R} \quad s < t,\ S \subset [t-1],\ S \ge 1,$$

so that the iterates resulting lifted OAMP algorithm $\{\boldsymbol{w}_{s,i}^{(D)} : s < t,\ i \in [D]\}$ can reconstruct the following iterates of the degree-D approximation to the intermediate OAMP algorithm:

$$\boldsymbol{v}_{s,S}^{(D)} = h_{s,S}^{(D)}(\boldsymbol{w}_{s,\bullet}^{(D)}) \quad \forall\, s < t,\ S \subset [t-1],\ |S| > 1. \tag{180}$$

We now extend our construction to step t. To do so, we need to construct the function $\bar{F}_t^{(D)}$ used in the tth iteration of the lifted OAMP algorithm, as well as the linear postprocessing maps:

$$h_{t,S}^{(D)} : \mathbb{R}^D \mapsto \mathbb{R} \quad S \subset [t],\ S \ge 1,$$
$$h_{s,S}^{(D)} : \mathbb{R}^D \mapsto \mathbb{R} \quad S \subset [t], S \ni t,$$

which ensure that:

$$\boldsymbol{v}_{t,S}^{(D)} = h_{t,S}^{(D)}(\boldsymbol{w}_{t,\bullet}^{(D)}) \quad \forall\, S \subset [t],\ |S| > 1, \tag{181}$$
$$\boldsymbol{v}_{s,S}^{(D)} = h_{s,S}^{(D)}(\boldsymbol{w}_{s,\bullet}^{(D)}) \quad \forall\, s < t,\ S \subset [t],\ S \ni t. \tag{182}$$

We will present the proof for (181), the proof for (182) is analogous. We begin by recalling the update rule for $\boldsymbol{v}_{t,S}^{(D)}$:

$$\boldsymbol{v}_{t,S}^{(D)} = \Big(\Psi_S^{(D)}(\boldsymbol{Y}) - \mathbb{E}_{\mathsf{\Lambda}\sim\mu}[\Psi_S^{(D)}(\mathsf{\Lambda})]\cdot \boldsymbol{I}_N\Big)\cdot\left(\bar{F}_t(\boldsymbol{v}_{<t,\bullet}^{(D)};\boldsymbol{a}) - \sum_{s=1}^{t-1}\sum_{R\subset[t-1]}\mathbb{E}[\partial_{s,R}\bar{F}_t(\mathsf{V}_{<t,\bullet}^{(D)};\mathsf{A})]\cdot \boldsymbol{v}_{s,R}^{(D)}\right)$$

$$\overset{\text{(a)}}{=} \Big(\Psi_S^{(D)}(\boldsymbol{Y}) - \mathbb{E}_{\mathsf{\Lambda}\sim\mu}[\Psi_S^{(D)}(\mathsf{\Lambda})]\cdot \boldsymbol{I}_N\Big)\cdot\left(\bar{F}_t(\{h_{s,R}^{(D)}(\boldsymbol{w}_{s,\bullet}^{(D)})\};\boldsymbol{a}) - \sum_{s=1}^{t-1}\sum_{R\subset[t-1]}\mathbb{E}[\partial_{s,R}\bar{F}_t(\mathsf{V}_{<t,\bullet}^{(D)};\mathsf{A})]\cdot h_{s,R}^{(D)}(\boldsymbol{w}_{s,\bullet}^{(D)})\right), \tag{183}$$

where step (a) follows from induction hypothesis in (180). In light of the above expression, we define $F_{t,D}$ as the composite function:

$$F_t^{(D)}(w_{<t,\bullet};a) \overset{\text{def}}{=} \bar{F}_t(\{h_{s,R}^{(D)}(w_{s,\bullet}) : s<t, R\subset[t-1], |R|>1\};a) \quad \forall\, w_{<t,\bullet}\in\mathbb{R}^{(t-1)D},\ a\in\mathbb{R}^k, \tag{184}$$

and the divergence-free iterate denoiser:

$$\bar{F}_t^{(D)}(w_{<t,\bullet};a) = F_t^{(D)}(w_{<t,\bullet};a) - \sum_{s=1}^{t-1}\sum_{i=1}^{D}\mathbb{E}[\partial_{s,i}F_t^{(D)}(\mathsf{W}_{<t,\bullet}^{(D)};\mathsf{A})]\cdot w_{s,i} \quad \forall\, w_{<t,\bullet}\in\mathbb{R}^{(t-1)D},\ a\in\mathbb{R}^k, \tag{185}$$

where $(\mathsf{X}_\star, \{\mathsf{W}_{s,i} : s<t, i\in[D]\};\mathsf{A})$ are the state evolution random variables associated with the lifted OAMP algorithm (178) that have been constructed so far (as a part of the induction hypothesis). We now extend our construction of the desired lifted OAMP algorithm to step t. At step t, the lifted OAMP algorithm generates the iterates:

$$\boldsymbol{w}_{t,i}^{(D)} = \big(\boldsymbol{Y}^i - \mathbb{E}_{\mathsf{\Lambda}\sim\mu}[\mathsf{\Lambda}^i]\cdot\boldsymbol{I}_N\big)\cdot\bar{F}_t^{(D)}(\boldsymbol{w}_{<t,\bullet}^{(D)};\boldsymbol{a}) \quad i\in[D]. \tag{186}$$

Hence (183) can be expressed as:

$$\begin{aligned}\boldsymbol{v}_{t,S}^{(D)} &\overset{(184)}{=} \Big(\Psi_S^{(D)}(\boldsymbol{Y}) - \mathbb{E}_{\mathsf{\Lambda}\sim\mu}[\Psi_S^{(D)}(\mathsf{\Lambda})]\cdot \boldsymbol{I}_N\Big)\cdot\left(F_t^{(D)}(\boldsymbol{w}_{<t,\bullet}^{(D)};\boldsymbol{a}) - \sum_{s=1}^{t-1}\sum_{R\subset[t-1]}\mathbb{E}[\partial_{s,R}\bar{F}_t(\mathsf{V}_{<t,\bullet}^{(D)};\mathsf{A})]\cdot h_{s,R}^{(D)}(\boldsymbol{w}_{s,\bullet}^{(D)})\right)\\
&\overset{\text{(b)}}{=} \Big(\Psi_S^{(D)}(\boldsymbol{Y}) - \mathbb{E}_{\mathsf{\Lambda}\sim\mu}[\Psi_S^{(D)}(\mathsf{\Lambda})]\cdot \boldsymbol{I}_N\Big)\cdot\left(F_t^{(D)}(\boldsymbol{w}_{<t,\bullet}^{(D)};\boldsymbol{a}) - \sum_{s=1}^{t-1}\sum_{i=1}^{D}\mathbb{E}[\partial_{s,i}F_t^{(D)}(\mathsf{W}_{<t,\bullet}^{(D)};\mathsf{A})]\cdot \boldsymbol{w}_{s,i}^{(D)}\right)\\
&\overset{(185)}{=} \Big(\Psi_S^{(D)}(\boldsymbol{Y}) - \mathbb{E}_{\mathsf{\Lambda}\sim\mu}[\Psi_S^{(D)}(\mathsf{\Lambda})]\cdot \boldsymbol{I}_N\Big)\cdot\bar{F}_t^{(D)}(\boldsymbol{w}_{<t,\bullet}^{(D)};\boldsymbol{a})\end{aligned} \tag{187}$$

where the step marked (b) follows by relating the derivatives of the function $F_t^{(D)}$ and the function $\bar{F}_t$ using the chain rule and by exploiting the linearity of postprocessing maps $\{h_{s,R}\}$. Recall that $\Psi_S^{(D)}:\mathbb{R}\mapsto\mathbb{R}$ is a degree-D polynomial, and so, it can be expressed as a linear combination of the monomials $\{\lambda\mapsto\lambda^i : i\in[D]\}$:

$$\Psi_S^{(D)}(\lambda) - \mathbb{E}_{\mathsf{\Lambda}\sim\mu}[\Psi_S^{(D)}(\mathsf{\Lambda})] = \sum_{i=1}^{D}\chi_S^{(D)}[i]\cdot(\lambda^i - \mathbb{E}_{\mathsf{\Lambda}\sim\mu}[\mathsf{\Lambda}^i]) \quad i\in[D].$$

Using this representation, the formula for $\boldsymbol{v}_{t,S}^{(D)}$ obtained in (187) can be expressed as a linear combination of the newly constructed lifted OAMP iterates in (186):

$$\boldsymbol{v}_{t,S}^{(D)} = \sum_{i=1}^{D}\chi_S^{(D)}[i]\cdot\boldsymbol{w}_{t,i}^{(D)}.$$

Hence, we define the postprocessing function $h_{t,S}^{(D)}$ as:

$$h_{t,S}^{(D)}(w_{t,\bullet}) = \sum_{i=1}^{D}\chi_S^{(D)}[i]\cdot w_{t,i} \quad \forall\, w_{t,\bullet}\in\mathbb{R}^D.$$

which is linear and ensures that $\boldsymbol{v}_{t,S}^{(D)} = h_{t,S}^{(D)}(\boldsymbol{w}_{t,\bullet}^{(D)})$, as desired. This completes the construction of the desired lifted OAMP algorithm by induction and concludes the proof of this lemma. □

E.2 Proof Proposition 4

Consider any degree-D lifted OAMP algorithm (Definition 6) of the form:

$$\boldsymbol{w}_{t,i} = \left(\boldsymbol{Y}^i - \mathbb{E}_{\Lambda\sim\nu}[\Lambda^i]\cdot \boldsymbol{I}_N\right)\cdot f_t\big(\boldsymbol{w}_{1,\bullet},\ldots,\boldsymbol{w}_{t-1,\bullet};\boldsymbol{a}\big) \quad i\in[D],\ t\in\mathbb{N}, \tag{188}$$

which returns the following estimator after t iterations:

$$\widehat{\boldsymbol{w}}_t = h_t(\boldsymbol{w}_{1,\bullet},\ldots,\boldsymbol{w}_{t,\bullet};\boldsymbol{a}).$$

Let $(\mathsf{X}_\star,\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A})$ denote the state evolution random variables associated with the first t iterations of the lifted OAMP algorithm, which form a Gaussian channel in the sense of Definition 3. We observe that, by Corollary 1:

$$\begin{aligned}\operatorname*{plim}_{N\to\infty}\frac{\|\boldsymbol{x}_\star-\widehat{\boldsymbol{w}}_t\|^2}{N} &= \mathbb{E}[\{\mathsf{X}_\star - h(\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A})\}^2]\\ &\overset{\text{(a)}}{\geq} \min_{h\in L^2(\mathsf{W}_{\leq t,\bullet};\mathsf{A})}\mathbb{E}[\{\mathsf{X}_\star - h(\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A})\}^2]\\ &\overset{\text{def}}{=} \mathrm{MMSE}(\mathsf{X}_\star|\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A}).\end{aligned} \tag{189}$$

In the above display, the inequality in step (a) follows by observing that since the function h is a Lipschitz function (as required by Definition 6), and hence, $h\in L^2(\mathsf{W}_{\leq t,\bullet};\mathsf{A})$. Furthermore, if Assumption 3 holds, the lower bound above is tight since we can choose h as the MMSE estimator for the Gaussian channel $(\mathsf{X}_\star,\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A})$.

In light of (189), for a t iteration lifted OAMP algorithm which uses functions $f_{1:t}$ in its iterations, we define:

$$\mathscr{M}_t(f_1,\ldots,f_t)\overset{\text{def}}{=}\mathrm{MMSE}(\mathsf{X}_\star|\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A}),$$

where $(\mathsf{X}_\star,\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A})$ are the state evolution random variables associated with the lifted OAMP algorithm. Notice that the RHS of the above display depends on the functions $f_{1:t}$ implicitly, since the distribution of the state evolution random variables is determined by $f_{1:t}$ (recall Definition 6).

A convenient formula for $\mathscr{M}_t(f_1,\ldots,f_t)$. To determine the optimal lifted OAMP algorithm, we need to optimize the functional $\mathscr{M}_t$ with respect to $f_{1:t}$. To do so, we first derive a convenient formula for $\mathscr{M}_t(f_1,\ldots,f_t)=\mathrm{MMSE}(\mathsf{X}_\star|\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A})$. Lemma 3 in Appendix A.2 shows that the MMSE (and other relevant properties) of a multivariate Gaussian channel, such as $(\mathsf{X}_\star,\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A})$ can be expressed in terms of the MMSE of an *effective scalar Gaussian channel*:

$$(\mathsf{X}_\star,\mathsf{S}\overset{\text{def}}{=}\langle v_{\mathrm{opt}},\mathsf{W}_{\leq t,\bullet}\rangle;\mathsf{A}), \tag{190}$$

whose observation random variable S is a linear combination of the observation random variables of the original multivariate Gaussian channel. The weights of linear combination are given by $v_{\mathrm{opt}}(\mathsf{X}_\star|\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A})$ (which we will sometimes shorthand as v_{opt}, when the multivariate Gaussian channel in question is clear from the context). The effective scalar Gaussian channel operates at SNR $\omega_{\mathrm{eff}}(\mathsf{X}_\star|\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A})$, and the MMSE of the original multivariate channel and the effective scalar channel are connected via the formula:

$$\mathrm{MMSE}(\mathsf{X}_\star|\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A}) = \mathrm{MMSE}(\mathsf{X}_\star|\mathsf{S};\mathsf{A}) = \mathsf{mmse}_\pi\left(\omega_{\mathrm{eff}}(\mathsf{X}_\star|\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A})\right),$$

where for $\omega\in[0,1]$, $\mathsf{mmse}_\pi(\omega)$ is the MMSE of a scalar Gaussian channel at SNR ω (recall Definition 3). The following lemma provides a formula for the effective SNR $\omega_{\mathrm{eff}}(\mathsf{X}_\star|\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet};\mathsf{A})$. To state this formula, recall from Definition 6 that the joint distribution of the state evolution random variables $(\mathsf{X}_\star,\mathsf{W}_{\leq t,\bullet};\mathsf{A})$ is given by:

$$(\mathsf{X}_\star;\mathsf{A})\sim\pi,\quad (\mathsf{W}_{1,\bullet},\ldots,\mathsf{W}_{t,\bullet})|(\mathsf{X}_\star;\mathsf{A})\sim\mathcal{N}\left(\mathsf{X}_\star\cdot\alpha\otimes q,\alpha\alpha^\top\otimes(Q-qq^\top)+(\Sigma-\alpha\alpha^\top)\otimes\Gamma\right). \tag{191a}$$

In the above display, the entries of $q \in \mathbb{R}^D$, $Q \in \mathbb{R}^{D\times D}$, and $\Gamma \in \mathbb{R}^{D\times D}$ are given by:

$$q_i = \mathbb{E}[\Lambda_\nu^i] - \mathbb{E}[\Lambda^i],\ Q_{ij} = \mathbb{E}\left[\left(\Lambda_\nu^i - \mathbb{E}[\Lambda^i]\right)\cdot\left(\Lambda_\nu^j - \mathbb{E}[\Lambda^j]\right)\right],\ \Gamma_{ij} = \mathrm{Cov}[\Lambda^i, \Lambda^j] \quad \text{where } \Lambda \sim \mu,\ \Lambda_\nu \sim \nu, \tag{191b}$$

the entries of $\alpha \in \mathbb{R}^t$ and $\Sigma \in \mathbb{R}^{t\times t}$ are given by:

$$\alpha_s \overset{\text{def}}{=} \mathbb{E}[\mathsf{X}_\star \cdot f_s(\mathsf{W}_{<s,\bullet}; \mathsf{A})], \quad \Sigma_{s\tau} \overset{\text{def}}{=} \mathbb{E}[f_s(\mathsf{W}_{<s,\bullet}; \mathsf{A}) \cdot f_\tau(\mathsf{W}_{<\tau,\bullet}; \mathsf{A})] \quad \forall\, s, \tau \in [t], \tag{191c}$$

and $\otimes$ denotes the Kronecker (or tensor) product for matrices.

Lemma 16. *Consider a lifted OAMP algorithm which uses functions $f_1, f_2, \ldots, f_t$ with $f_i : \mathbb{R}^{(i-1)\cdot D+k} \mapsto \mathbb{R}$ in the first t iterations. Let $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$ be the state evolution random variables associated with the algorithm. Then,*

$$\omega_{\text{eff}}(\mathsf{X}_\star | \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A}) = \begin{cases} q^\mathsf{T}\left[Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right]^\dagger q & : \kappa_t \in [0,1) \\ 0 & : \kappa_t = 1. \end{cases}$$

Moreover, if $\kappa_t < 1$, then,

$$v_{\text{opt}}(\mathsf{X}_\star | \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A}) = \frac{1}{\sqrt{\omega_{\text{eff}}(\mathsf{X}_\star | \mathsf{W}_{\le t,\bullet}; \mathsf{A})}} \cdot (\Sigma^\dagger \alpha) \otimes [(1-\kappa_t)Q + \kappa_t \Gamma]^\dagger q,$$

where the scalar $\kappa_t \in [0,1]$ is defined as[2]*:*

$$\kappa_t \overset{def}{=} \min\left\{\mathbb{E}\left[\{\mathsf{X}_\star - g(\mathsf{W}_{<t,\bullet}; \mathsf{A})\}^2\right] : g \in \mathrm{Span}(f_1, \ldots, f_t)\right\}.$$

Proof. The proof of this lemma is presented in Section E.2.1. □

Combining the effective SNR formula from the lemma above with (190), we obtain the following expression for $\mathscr{M}(f_1, \ldots, f_t)$:

$$\mathscr{M}(f_1, \ldots, f_t) \overset{\text{def}}{=} \mathrm{MMSE}(\mathsf{X}_\star | \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A}) = \mathsf{mmse}_\pi(\omega_{\text{eff}}(\mathsf{X}_\star | \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})) = \mathscr{m}(\kappa_t), \tag{192a}$$

where we introduced the function $\mathscr{m} : [0,1] \mapsto [0,1]$ defined as follows:

$$\mathscr{m}(\kappa) \overset{\text{def}}{=} \begin{cases} \mathsf{mmse}_\pi\left(q^\mathsf{T}\left[Q + \frac{\kappa}{1-\kappa}\cdot\Gamma\right]^\dagger q\right) & : \kappa \in [0,1) \\ \mathsf{mmse}_\pi(0) & : \kappa = 1, \end{cases} \tag{192b}$$

where q, Q, Γ are as defined in (47).

Monotonicity of $\mathscr{m}(\cdot)$ The following monotonicity property of the function $\mathscr{m}$ introduced in (192) will play an important role in the proof of Proposition 4.

Lemma 17. *The function $\mathscr{m} : [0,1] \mapsto [0,1]$ is non-decreasing.*

Proof. See Section E.2.2. □

[2]For a collection of functions $f_1, \ldots, f_t$, $\mathrm{Span}(f_1, \ldots, f_t)$ denotes the set of all functions that can be written as a linear combination of $f_1, \ldots, f_t$.

A Greedy Approach Minimizing $\mathscr{M}_t$. To derive the optimal OAMP algorithm, our goal is to solve the variational problem:

$$\min_{f_1} \min_{f_2} \cdots \min_{f_t} \mathscr{M}_t(f_1, \ldots, f_t), \tag{193}$$

Recall that Proposition 4 claims that the iterate denoisers $(f_t^\star)_{t\in\mathbb{N}}$ that optimize (193) are defined recursively: once the optimal denoisers $f_1^\star, \ldots, f_t^\star$ have been specified, the optimal denoiser for the $t+1$ iteration is the DMMSE estimator for the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$, where $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$ denote the state evolution random variables corresponding to the first t iterations. We will show that these choices for the iterate denoisers correspond to a *greedy approach* to solving the optimization problem in (193):

$$f_1^\star \in \arg\min_{f_1} \mathscr{M}_1(f_1), \quad f_2^\star \in \arg\min_{f_2} \mathscr{M}_2(f_1^\star, f_2), \quad \ldots, \quad f_t^\star \in \arg\min_{f_t} \mathscr{M}_t(f_1^\star, f_2^\star, \ldots, f_{t-1}^\star, f_t).$$

Proving this requires us to characterize the solution of the simpler (compared to (193)) variational problem:

$$\min_{f_t} \mathscr{M}_t(f_1, \ldots, f_t),$$

which is done in the following lemma.

Lemma 18. *Let:*

1. *$f_1, \ldots, f_{t-1}$ with $f_i : \mathbb{R}^{(i-1)\times D+k} \mapsto \mathbb{R}$ be a fixed collection of iterate denoisers which satisfy the requirements imposed in the definition of lifted OAMP algorithms (Definition 6) ;*
2. *$(\mathsf{X}_\star, \mathsf{W}_{1;\bullet}, \ldots, \mathsf{W}_{t-1,\bullet}; \mathsf{A})$ be the state evolution random variables associated with the first $t-1$ iterations of a lifted OAMP algorithm which uses the functions $f_1, \ldots, f_{t-1}$ in the first $t-1$ iterations;*
3. *$f_t^\sharp : \mathbb{R}^{(t-1)\times D+k} \mapsto \mathbb{R}$ denote the DMMSE estimator for the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{1;\bullet}, \ldots, \mathsf{W}_{t-1,\bullet}; \mathsf{A})$.*

Then, for any iterate denoiser $f_t : \mathbb{R}^{(t-1)\times D+k} \mapsto \mathbb{R}$ which satisfies the requirements imposed in Definition 6, we have:

$$\mathscr{M}_t(f_1, \ldots, f_t) \geq \mathscr{M}_t(f_1, \ldots, f_t^\sharp) = m(\kappa_t^\sharp), \tag{194}$$

where $\kappa_t^\sharp \stackrel{def}{=} \mathsf{dmmse}_\pi \circ \mathsf{mmse}_\pi^{-1}(\mathscr{M}_{t-1}(f_{1:t-1})))$.

Proof. The proof of this lemma is presented in Section E.2.3. □

Using the above intermediate lemmas, we can now present the proof of Proposition 4.

Proof of Proposition 4. To prove the claim of Proposition 4, we need to show that for any $t \in \mathbb{N}$ and for any choice of iterate denoisers $f_{1:t}$ which satisfy the requirements imposed in Definition 6,

$$\mathscr{M}_t(f_1, \ldots, f_t) \geq \mathscr{M}_t(f_1^\star, \ldots, f_t^\star), \tag{195}$$

where $f_{1:t}^\star$ denote the functions used in the iterations of the optimal lifted OAMP algorithm in (49), which are specified recursively as follows: after the functions $f_1^\star, \ldots, f_{t-1}^\star$ have been specified, $f_t^\star$ was defined as the DMMSE estimator for the Gaussian channel corresponding to the first $t-1$ iterations of the resulting lifted OAMP algorithm. By the Lemma 18, $f_{1:t}^\star$ are precisely the functions derived from the greedy heuristic:

$$f_1^\star \in \arg\min_{f_1} \mathscr{M}_1(f_1), \quad f_2^\star \in \arg\min_{f_2} \mathscr{M}_2(f_1^\star, f_2), \quad \ldots, \quad f_t^\star \in \arg\min_{f_t} \mathscr{M}_t(f_1^\star, f_2^\star, \ldots, f_{t-1}^\star, f_t),$$

where the minimization is understood to be over iterate denoisers which satisfy the requirements imposed in Definition 6. We will obtain the desired conclusion in (195) by induction. As our induction hypothesis, we assume that for any choice of the iterate denoisers $f_1, \ldots, f_{t-1}$, we have:

$$\mathscr{M}_{t-1}(f_1, \ldots, f_{t-1}) \geq \mathscr{M}_{t-1}(f_1^\star, \ldots, f_{t-1}^\star). \tag{196}$$

In order to obtain the desired conclusion (195), we observe that:

$$\begin{aligned}\mathscr{M}_t(f_1,\dots,f_t) &\overset{(a)}{\geq} \mathscr{m}(\kappa_t^\sharp), \text{ where } \kappa_t^\sharp \overset{\text{def}}{=} \mathsf{dmmse}_\pi \circ \mathsf{mmse}_\pi^{-1}(\mathscr{M}_{t-1}(f_{1:t-1})))\\ &\overset{(b)}{\geq} \mathscr{m}(\kappa_t^\star), \text{ where } \kappa_t^\star \overset{\text{def}}{=} \mathsf{dmmse}_\pi \circ \mathsf{mmse}_\pi^{-1}(\mathscr{M}_{t-1}(f^\star_{1:t-1})))\\ &\overset{(c)}{=} \mathscr{M}_t(f_1^\star,\dots,f_t^\star).\end{aligned}$$

In the above display:

- Step (a) used the lower bound for $\mathscr{M}_t(f_1,\dots,f_t)$ provided in Lemma 18.
- In step (b), we observed by the induction hypothesis that $\mathscr{M}_{t-1}(f^\star_{1:t-1}) \leq \mathscr{M}_{t-1}(f_{1:t-1})$. Since mmse_π and dmmse_π are non-increasing functions (Lemma 1 and Lemma 4), we conclude that $\kappa_t^\star \leq \kappa_t^\sharp$. Finally, since $\mathscr{m}$ is non-increasing (Lemma 17), we concluded that $\mathscr{m}(\kappa_t^\sharp) \geq \mathscr{m}(\kappa_t)$.
- In step (c), we again appealed to Lemma 18:

$$\mathscr{M}_t(f_1^\star,\dots,f_t^\star) = \min_{f_t} \mathscr{M}_t(f_1^\star,\dots,f_t) = \mathscr{m}(\kappa_t^\star), \text{ where } \kappa_t^\star \overset{\text{def}}{=} \mathsf{dmmse}_\pi \circ \mathsf{mmse}_\pi^{-1}(\mathscr{M}_{t-1}(f^\star_{1:t-1}))).$$

This concludes the proof of Proposition 4. □

E.2.1 Proof of Lemma 16

Proof of Lemma 16. Recall from Definition 6 that the joint distribution of the state evolution random variables is given by:

$$(\mathsf{X}_\star;\mathsf{A}) \sim \pi, \quad (\mathsf{W}_{1,\bullet},\dots,\mathsf{W}_{t,\bullet})|(\mathsf{X}_\star;\mathsf{A}) \sim \mathcal{N}\left(\mathsf{X}_\star \cdot \alpha \otimes q, \alpha\alpha^\top \otimes (Q - qq^\top) + (\Sigma - \alpha\alpha^\top) \otimes \Gamma\right). \tag{197}$$

In the above display, the entries of $q \in \mathbb{R}^D$, $Q \in \mathbb{R}^{D\times D}$, and $\Gamma \in \mathbb{R}^{D\times D}$ are given by:

$$q_i = \mathbb{E}[\mathsf{\Lambda}_\nu^i] - \mathbb{E}[\mathsf{\Lambda}^i],\ Q_{ij} = \mathbb{E}\left[\left(\mathsf{\Lambda}_\nu^i - \mathbb{E}[\mathsf{\Lambda}^i]\right)\cdot\left(\mathsf{\Lambda}_\nu^j - \mathbb{E}[\mathsf{\Lambda}^j]\right)\right],\ \Gamma_{ij} = \mathrm{Cov}[\mathsf{\Lambda}^i,\mathsf{\Lambda}^j] \quad \text{where } \mathsf{\Lambda}\sim\mu,\ \mathsf{\Lambda}_\nu\sim\nu, \tag{198}$$

and the entries of $\alpha \in \mathbb{R}^t$ and $\Sigma \in \mathbb{R}^{t\times t}$ are given by:

$$\alpha_s \overset{\text{def}}{=} \mathbb{E}[\mathsf{X}_\star \cdot f_s(\mathsf{W}_{<s,\bullet};\mathsf{A})], \quad \Sigma_{s\tau} \overset{\text{def}}{=} \mathbb{E}[f_s(\mathsf{W}_{<s,\bullet};\mathsf{A})\cdot f_\tau(\mathsf{W}_{<\tau,\bullet};\mathsf{A})] \quad \forall\, s,\tau \in [t]. \tag{199}$$

The proof of the claim is obtained by applying the effective SNR formula for general multivariate Gaussian channels stated as Lemma 3 in Appendix A.2 to the Gaussian channel $(\mathsf{X}_\star;\mathsf{W}_{\leq t,\bullet};\mathsf{A})$. According to this formula:

$$\omega_{\text{eff}}(\mathsf{X}_\star|\mathsf{W}_{\leq t,\bullet};\mathsf{A}) = \begin{cases}\frac{\vartheta}{1+\vartheta} & \text{if } \alpha\otimes q \in \mathrm{Range}(\alpha\alpha^\top\otimes(Q-qq^\top)+(\Sigma-\alpha\alpha^\top)\otimes\Gamma),\\ 1 & \text{if } \alpha\otimes q \notin \mathrm{Range}(\alpha\alpha^\top\otimes(Q-qq^\top)+(\Sigma-\alpha\alpha^\top)\otimes\Gamma),\end{cases} \tag{200a}$$

where:

$$\vartheta \overset{\text{def}}{=} \alpha\otimes q^\top(\alpha\alpha^\top\otimes(Q-qq^\top)+(\Sigma-\alpha\alpha^\top)\otimes\Gamma)^\dagger\alpha\otimes q. \tag{200b}$$

The claim of the lemma is obtained by simplifying the above formula.

Orthogonalization. To obtain the simplified formula stated in the lemma, it will be convenient to construct an orthonormal basis for $\mathrm{Span}(f_{1:t})$. Let $r = \mathrm{Rank}(\Sigma)$ and $\widehat{f}_1,\widehat{f}_2,\dots,\widehat{f}_r : \mathbb{R}^{(t-1)D} \mapsto \mathbb{R}$ denote the basis functions constructed by orthogonalizing the functions $f_{1:t}$ with respect to the state evolution random variables $(\mathsf{W}_{<t,\bullet};\mathsf{A})$, via the Gram-Schmidt procedure. These basis functions are orthonormal in the sense that:

$$\mathbb{E}[\widehat{f}_s(\mathsf{W}_{<t,\bullet};\mathsf{A})\widehat{f}_\tau(\mathsf{W}_{<t,\bullet};\mathsf{A})] = \begin{cases}1 & : s=\tau\\ 0 & : \text{otherwise.}\end{cases}$$

We can express the functions $f_{1:t}$ as a linear combination of the basis functions:

$$f_\tau = \sum_{s=1}^{r} L_{\tau s} \widehat{f}_s \quad \forall\, \tau \;\in\; [t], \quad L_{\tau s} \stackrel{\text{def}}{=} \mathbb{E}[f_\tau(\mathsf{W}_{<\tau,\bullet};\mathsf{A})\widehat{f}_s(\mathsf{W}_{<t,\bullet};\mathsf{A})] \quad \forall\, s \;\in [r],\; \tau \in [t]. \tag{201}$$

Let L denote the $t \times r$ matrix whose entries are given by the coefficients $(L_{\tau s})_{\tau\in[t],s\in[r]}$ defined above. Using the decomposition in (201), along with the definitions of α, Σ in (199), we obtain the following alternative expressions for α and Σ:

$$\Sigma = LL^\top, \quad \alpha = L\widehat{\alpha}, \quad \widehat{\alpha} \stackrel{\text{def}}{=} (\mathbb{E}[\mathsf{X}_\star \widehat{f}_1(\mathsf{W}_{<t,\bullet};\mathsf{A})], \mathbb{E}[\mathsf{X}_\star \widehat{f}_2(\mathsf{W}_{<t,\bullet};\mathsf{A})] \cdots, \mathbb{E}[\mathsf{X}_\star \widehat{f}_t(\mathsf{W}_{<t,\bullet};\mathsf{A})])^\top. \tag{202}$$

Similarly, by exploiting the fact that $\widehat{f}_{1:t}$ form an orthonormal basis for $\mathrm{Span}(f_{1:t})$, we have the following alternate formula for κ_t:

$$\begin{aligned}\kappa_t &\stackrel{\text{def}}{=} \min\left\{\mathbb{E}\left[\{\mathsf{X}_\star - g(\mathsf{W}_{<t,\bullet};\mathsf{A})\}^2\right] : g \in \mathrm{Span}(f_1,\ldots,f_t)\right\} \\ &= 1 - \|\widehat{\alpha}\|^2 \in [0,1].\end{aligned} \tag{203}$$

These alternate expressions for α, Σ, κ_t will be useful in our subsequent calculations.

Useful Linear Algebraic Results. The proof relies on the following intermediate linear algebraic claims.

$$\text{For any } \lambda \geq 0, \quad q \in \mathrm{Range}(Q + \lambda\Gamma), \tag{204}$$

$$\text{If } \kappa_t < 1, \quad q^\top \left(Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right)^\dagger q \neq 1 \Leftrightarrow \alpha \otimes q \in \mathrm{Range}(\alpha\alpha^\top \otimes (Q - qq^\top) + (\Sigma - \alpha\alpha^\top)\otimes\Gamma). \tag{205}$$

We defer the justification of these claims to the end of the proof.

Corner Cases. We begin by verifying the claim of the lemma for two corner cases. Consider the situation where $\kappa_t = 1$. In this case, (202) and (203) imply that $\alpha = 0$. As a result, using the formula for the effective SNR for a general multivariate Gaussian channel given in Lemma 3 (Appendix A.2), the effective SNR of the Gaussian channel in (197) is $\omega_{\text{eff}} = 0$, as claimed. Next, we consider the case when:

$$\kappa_t < 1, \quad q^\top \left(Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right)^\dagger q = 1.$$

From (205), we conclude that:

$$\alpha \otimes q \notin \mathrm{Range}(\alpha\alpha^\top \otimes (Q - qq^\top) + (\Sigma - \alpha\alpha^\top)\otimes\Gamma).$$

Hence, from (200), we conclude that the effective SNR of the Gaussian channel in (197) is $\omega_{\text{eff}} = 1$, as claimed.

Derivation of the claimed formulae. The remaining task is to prove the claim of the lemma when:

$$\kappa_t < 1, \quad q^\top \left(Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right)^\dagger q \neq 1,$$

and we work under the above assumption for the remainder of the proof. From (205), we conclude that:

$$\alpha \otimes q \in \mathrm{Range}(\alpha\alpha^\top \otimes (Q - qq^\top) + (\Sigma - \alpha\alpha^\top)\otimes\Gamma).$$

From (200), we have that:

$$\omega_{\text{eff}}(\mathsf{X}_\star|\mathsf{W}_{\leq t,\bullet};\mathsf{A}) = \frac{\vartheta}{1+\vartheta}, \tag{206}$$

$$v_{\text{opt}}(\mathsf{X}_\star|\mathsf{W}_{\leq t,\bullet};\mathsf{A}) = \frac{1}{\sqrt{\vartheta+\vartheta^2}} \cdot (\alpha\alpha^\top \otimes (Q - qq^\top) + (\Sigma - \alpha\alpha^\top)\otimes\Gamma)^\dagger \cdot \alpha \otimes q \tag{207}$$

where:

$$\vartheta \stackrel{\text{def}}{=} \alpha \otimes q^\top (\alpha\alpha^\top \otimes (Q - qq^\top) + (\Sigma - \alpha\alpha^\top) \otimes \Gamma)^\dagger \alpha \otimes q. \tag{208}$$

To compute the pseudoinverse $(\alpha\alpha^\top \otimes (Q - qq^\top) + (\Sigma - \alpha\alpha^\top) \otimes \Gamma)^\dagger$, we define $\overline{\alpha}$ as the unit vector along $\widehat{\alpha}$ and P as the projector orthogonal to $\widehat{\alpha}$:

$$\overline{\alpha} \stackrel{\text{def}}{=} \frac{\widehat{\alpha}}{\|\widehat{\alpha}\|}, \quad P \stackrel{\text{def}}{=} I - \overline{\alpha}\overline{\alpha}^\top. \tag{209}$$

We have,

$$\begin{aligned}
(\alpha\alpha^\top \otimes (Q - qq^\top) + (\Sigma - \alpha\alpha^\top) \otimes \Gamma)^\dagger &\overset{(202)}{=} (L^\dagger \otimes I)^\top \cdot [\widehat{\alpha}\widehat{\alpha}^\top \otimes (Q - qq^\top - \Gamma) + I \otimes \Gamma]^\dagger \cdot (L^\dagger \otimes I) \\
&\overset{(203),(209)}{=} (L^\dagger \otimes I)^\top \cdot [\overline{\alpha}\overline{\alpha}^\top \otimes \{(1-\kappa_t)\cdot(Q - qq^\top) + \kappa_t \cdot \Gamma\} + P \otimes \Gamma]^\dagger \cdot (L^\dagger \otimes I) \\
&= (L^\dagger \otimes I)^\top \cdot [\overline{\alpha}\overline{\alpha}^\top \otimes \{(1-\kappa_t)\cdot(Q - qq^\top) + \kappa_t \cdot \Gamma\}^\dagger + P \otimes \Gamma^\dagger] \cdot (L^\dagger \otimes I).
\end{aligned}$$

The above formula for the pseudo-inverse yields:

$$\begin{aligned}
&(\alpha\alpha^\top \otimes (Q - qq^\top) + (\Sigma - \alpha\alpha^\top) \otimes \Gamma)^\dagger \cdot \alpha \otimes q = (L^{\dagger\top}\widehat{\alpha}) \otimes [\{(1-\kappa_t)\cdot(Q - qq^\top) + \kappa_t \cdot \Gamma\}^\dagger \cdot q] \\
&\overset{(a)}{=} (1-\kappa_t)^{-1} \cdot \left\{1 - q^\top \left(Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right)^\dagger q\right\}^{-1} \cdot (L^{\dagger\top}\widehat{\alpha}) \otimes \left(Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right)^\dagger q \\
&\overset{(202)}{=} (1-\kappa_t)^{-1} \cdot \left\{1 - q^\top \left(Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right)^\dagger q\right\}^{-1} \cdot (\Sigma^\dagger\alpha) \otimes \left(Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right)^\dagger q,
\end{aligned} \tag{210}$$

where we used the Sherman-Morrison formula in (a). Combining the computation above with the definition of ϑ in (208) yields the following formula for ϑ:

$$\begin{aligned}
\vartheta &\overset{(203)}{=} (\alpha^\top\Sigma^\dagger\alpha)\cdot(1-\kappa_t)^{-1} \cdot \left\{1 - q^\top \left(Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right)^\dagger q\right\}^{-1} \cdot q^\top \left(Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right)^\dagger q \\
&\overset{(a)}{=} \left\{1 - q^\top \left(Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right)^\dagger q\right\}^{-1} \cdot q^\top \left(Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right)^\dagger q
\end{aligned}$$

where the equality marked (a) follows by observing that $\alpha^\top\Sigma^\dagger\alpha = \|\widehat{\alpha}\|^2 = 1 - \kappa_t$. Plugging the expression of the previous display in (206) yields:

$$\begin{aligned}
\omega_{\text{eff}}(\mathsf{X}_\star|\mathsf{W}_{\le t,\bullet};\mathsf{A}) &= q^\top \left(Q + \frac{\kappa_t}{1-\kappa_t}\cdot\Gamma\right)^\dagger q, \\
v_{\text{opt}}(\mathsf{X}_\star|\mathsf{W}_{\le t,\bullet};\mathsf{A}) &= \frac{1}{\sqrt{\omega_{\text{eff}}(\mathsf{X}_\star|\mathsf{W}_{\le t,\bullet};\mathsf{A})}} \cdot (\Sigma^\dagger\alpha) \otimes ((1-\kappa_t)\cdot Q + \kappa_t\cdot\Gamma)^\dagger q
\end{aligned}$$

as claimed. To finish the proof, we need to prove the linear algebraic claims in (204) and (205).

Proof of claim (204). Fix any $\lambda \ge 0$. To show that $q \in \text{Range}(Q + \lambda\Gamma)$, it is sufficient to verify that q is orthogonal to $\text{Null}(Q + \lambda\Gamma)$. Towards this goal, we consider any $v \in \text{Null}(Q + \lambda\Gamma)$ and aim to show that $\langle v, q\rangle = 0$. Since v lies in the null space of $Q + \lambda\Gamma$,

$$v^\top(Q + \lambda\Gamma)v = 0 \implies v^\top Q v = 0,$$

where the implication is obtained by observing that $Q \succeq 0$ and $\Gamma \succeq 0$ (cf. (198)). Recalling the definition of q from (198), we obtain:

$$|\langle q, v\rangle|^2 = \left|\mathbb{E}\left[\sum_{i=1}^D v_i \cdot (\Lambda_\nu^i - \mathbb{E}[\Lambda^i])\right]\right|^2 \overset{(a)}{\le} \mathbb{E}\left[\left|\sum_{i=1}^D v_i \cdot (\Lambda_\nu^i - \mathbb{E}[\Lambda^i])\right|^2\right] \overset{(198)}{=} v^\top Q v = 0.$$

In the above display, $\Lambda \sim \mu, \Lambda_\nu \sim \nu$ and the inequality in (a) follows from the Cauchy–Schwarz inequality. Hence, $\langle q, v\rangle = 0$ for any $v \in \text{Null}(Q + \lambda\Gamma)$, which implies the claim in (204).

Proof of claim (205). For convenience, we introduce the definitions:

$$S \stackrel{\text{def}}{=} \alpha\alpha^\top \otimes (Q - qq^\top) + (\Sigma - \alpha\alpha^\top) \otimes \Gamma, \quad T \stackrel{\text{def}}{=} Q + \frac{\kappa_t}{1-\kappa_t} \cdot \Gamma.$$

Assuming $\kappa_t < 1$, our goal is to show the equivalence:

$$q^\top T^\dagger q \neq 1 \Leftrightarrow \alpha \otimes q \in \text{Range}(S).$$

Consider the vector:

$$v \stackrel{\text{def}}{=} (L^{\dagger\top}\widehat{\alpha}) \otimes (T^\dagger q).$$

A short computation reveals that:

$$Sv = (1-\kappa_t) \cdot (1 - q^\top T^\dagger q) \cdot \alpha \otimes q, \quad \langle q, v\rangle = (1-\kappa_t) \cdot q^\top T^\dagger q.$$

If $q^\top T^\dagger q \neq 1$, we have found a vector $u = (1-\kappa_t)^{-1} \cdot (1 - q^\top T^\dagger q)^{-1} \cdot v$ such that $Su = \alpha \otimes q$, which certifies that $\alpha \otimes q \in \text{Range}(S)$. On the other hand, if $q^\top T^\dagger q = 1$, then $Sv = 0$ and we have found a vector $v \in \text{Null}(S)$ such that $\langle v, q\rangle = (1-\kappa_t) \neq 0$, which certifies that $\alpha \otimes q \notin \text{Range}(S)$. This concludes the proof of this lemma. □

E.2.2 Proof of Lemma 17

Proof of Lemma 17. Recall from (192) that:

$$\mathscr{m}(\kappa) \stackrel{\text{def}}{=} \begin{cases} \mathsf{mmse}_\pi\left(q^\mathsf{T}\left[Q + \frac{\kappa}{1-\kappa}\cdot\Gamma\right]^\dagger q\right) & : \kappa \in [0,1) \\ \mathsf{mmse}_\pi(0) & : \kappa = 1. \end{cases} \tag{211}$$

The expressions for the vector $q \in \mathbb{R}^D$ and the *positive semi-definite* matrices $Q, \Gamma \in \mathbb{R}^{D\times D}$ appear in (47), but will not be needed in the proof. To show that $\mathscr{m}$ is a non-decreasing function, it suffices to verify that for any $0 \le \kappa \le \kappa' < 1$, we have $\mathscr{m}(\kappa) \le \mathscr{m}(\kappa') \le \mathscr{m}(1)$. Indeed we have,

$$\begin{aligned} 0 \le \kappa \le \kappa' < 1 &\implies 0 \le \frac{\kappa}{1-\kappa} \le \frac{\kappa'}{1-\kappa'} \stackrel{\text{(a)}}{\implies} 0 \preceq Q + \frac{\kappa}{1-\kappa}\cdot\Gamma \preceq Q + \frac{\kappa'}{1-\kappa'}\cdot\Gamma \\ &\stackrel{\text{(b)}}{\implies} q^\top\left(Q + \frac{\kappa}{1-\kappa}\cdot\Gamma\right)^\dagger q \ge q^\top\left(Q + \frac{\kappa'}{1-\kappa'}\cdot\Gamma\right)^\dagger \ge 0 \\ &\stackrel{\text{(c)}}{\implies} \mathsf{mmse}_\pi\left(q^\mathsf{T}\left[Q + \frac{\kappa}{1-\kappa}\cdot\Gamma\right]^\dagger q\right) \le \mathsf{mmse}_\pi\left(q^\mathsf{T}\left[Q + \frac{\kappa'}{1-\kappa'}\cdot\Gamma\right]^\dagger q\right) \le \mathsf{mmse}_\pi(0), \\ &\stackrel{\text{(d)}}{\implies} \mathscr{m}(\kappa) \le \mathscr{m}(\kappa') \le \mathscr{m}(1). \end{aligned} \tag{212}$$

In the above display, step (a) follows by observing that Q, Γ are positive semidefinite ($\preceq$ denotes the standard Loewner partial ordering of symmetric matrices defined by the positive semidefinite cone), step (c) relies on the monotonicity of $\mathsf{mmse}_\pi(\cdot)$ (Fact 1). Finally, step (b) follows from the Fenchel conjugate formula for convex quadratic forms (recall from (204) that $q \in \text{Range}(Q + \lambda\Gamma)$ for any $\lambda \ge 0$):

$$\begin{aligned} q^\top\left(Q + \frac{\kappa}{1-\kappa}\cdot\Gamma\right)^\dagger q &= \sup_{\xi\in\mathbb{R}^D} 2\langle q, \xi\rangle - \xi^\top\left(Q + \frac{\kappa}{1-\kappa}\cdot\Gamma\right)\xi \\ &\stackrel{(212)}{\ge} \sup_{\xi\in\mathbb{R}^D} 2\langle q, \xi\rangle - \xi^\top\left(Q + \frac{\kappa'}{1-\kappa'}\cdot\Gamma\right)\xi = q^\top\left(Q + \frac{\kappa'}{1-\kappa'}\cdot\Gamma\right)^\dagger q. \end{aligned}$$

This verifies that $\mathscr{m}$ is non-decreasing. □

E.2.3 Proof of Lemma 18

Proof of Lemma 18. Given a collection of iterate denoisers $f_1, \ldots, f_t$, which satisfy the requirements of Definition 6, our goal is to show that:

$$\mathscr{M}_t(f_1, f_2, \ldots, f_{t-1}, f_t) \geq \mathscr{m}(\kappa_t^\sharp) \text{ and,} \tag{213}$$

$$\mathscr{M}_t(f_1, f_2, \ldots, f_{t-1}, f_t^\sharp) = \mathscr{m}(\kappa_t^\sharp), \tag{214}$$

where $\kappa_t^\sharp \stackrel{\text{def}}{=} \mathsf{dmmse}_\pi \circ \mathsf{mmse}_\pi^{-1}(\mathscr{M}_{t-1}(f_{1:t-1})))$ and $f_t^\sharp$ is the DMMSE estimator for the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t-1,\bullet}; \mathsf{A})$ generated by the state evolution random variables associated with the lifted OAMP algorithm which uses the denoisers $f_1, \ldots, f_{t-1}$ in the first $t-1$ iterations. From Lemma 16, we know that:

$$\mathscr{M}_t(f_1, f_2, \ldots, f_t) = \mathscr{m}(\kappa_t) \tag{215a}$$

where the scalar $\kappa_t \in [0, 1]$ was defined as:

$$\kappa_t \stackrel{\text{def}}{=} \min \left\{ \mathbb{E}\left[\{\mathsf{X}_\star - g(\mathsf{W}_{<t,\bullet}; \mathsf{A})\}^2\right] : g \in \mathrm{Span}(f_1, \ldots, f_t) \right\}. \tag{215b}$$

We consider the two claims of the lemma (213) and (214) one by one.

Proof of Claim (213). In light of the monotonicity of $\mathscr{m}$ (Lemma 17), it suffices to show that

$$\kappa_t \stackrel{\text{def}}{=} \min \left\{ \mathbb{E}\left[\{\mathsf{X}_\star - g(\mathsf{W}_{<t,\bullet}; \mathsf{A})\}^2\right] : g \in \mathrm{Span}(f_1, \ldots, f_t) \right\} \geq \kappa_t^\sharp \stackrel{\text{def}}{=} \mathsf{dmmse}_\pi \circ \mathsf{mmse}_\pi^{-1}(\mathscr{M}_{t-1}(f_{1:t-1}))). \tag{216}$$

We can lower bound κ_t as follows:

$$\begin{aligned}
\kappa_t &\stackrel{\text{def}}{=} \min \left\{ \mathbb{E}\left[\{\mathsf{X}_\star - g(\mathsf{W}_{<t,\bullet}; \mathsf{A})\}^2\right] : g \in \mathrm{Span}(f_1, \ldots, f_t) \right\} \\
&\stackrel{(a)}{\geq} \min \left\{ \mathbb{E}\left[\{\mathsf{X}_\star - g(\mathsf{W}_{<t,\bullet}; \mathsf{A})\}^2\right] : g \in L^2(\mathsf{W}_{<t,\bullet}; \mathsf{A}),\ \mathbb{E}[\mathsf{Z}_{s,i} g(\mathsf{W}_{<t,\bullet}; \mathsf{A})] = 0\ \forall\ s \in [t-1],\ i \in [D] \right\} \\
&\stackrel{\text{def}}{=} \mathrm{DMMSE}(\mathsf{X}_\star | \mathsf{W}_{<t,\bullet}; \mathsf{A}).
\end{aligned} \tag{217}$$

In the above display, the random variables $\mathsf{Z}_{<t,\bullet}$ represent the Gaussian noise random variables for the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{<t,\bullet}; \mathsf{A})$. To obtain the inequality in step (a), we used the fact that any $g \in \mathrm{Span}(f_1, \ldots, f_t)$ satisfies the divergence-free constraints:

$$\mathbb{E}[\mathsf{Z}_{s,i} \cdot g(\mathsf{W}_{<t,\bullet}; \mathsf{A})] = 0\ \forall\ s \in [t-1],\ i \in [D].$$

Indeed, by Gaussian integration by parts, for any $i < t$, $s < t$ and $j \in [D]$,

$$\mathbb{E}[\mathsf{Z}_{s,j} \cdot f_i(\mathsf{W}_{<i,\bullet}; \mathsf{A})]$$

can be expressed as a linear combination of the expected derivatives

$$\{\mathbb{E}[\partial_{\tau,\ell} f_i(\mathsf{W}_{<i,\bullet}; \mathsf{A})] : \tau < i, \ell \in [D]\},$$

which are all identically zero thanks to the divergence-free requirement imposed on the iterate denoisers $f_1, \ldots, f_{t-1}$ in Definition 6. We now cast the lower bound on κ_t obtained in (217) in the desired form (216). Using the DMMSE formula for general multivariate Gaussian channels (stated as Lemma 3 in Appendix Appendix A.2) we obtain:

$$\kappa_t \geq \mathrm{DMMSE}(\mathsf{X}_\star | \mathsf{W}_{<t,\bullet}; \mathsf{A}) = \mathsf{dmmse}_\pi(\omega_{\text{eff}}),$$

where ω_{eff} is the effective SNR of the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{<t,\bullet}; \mathsf{A})$. On the other hand, the MMSE formula for general multivariate Gaussian channels (stated as Lemma 3 in Appendix Appendix A.2) yields:

$$\mathscr{M}_{t-1}(f_1, \ldots, f_{t-1}) \stackrel{\text{def}}{=} \mathrm{MMSE}(\mathsf{X}_\star | \mathsf{W}_{<t,\bullet}; \mathsf{A}) = \mathsf{mmse}_\pi(\omega_{\text{eff}}).$$

Since $\mathsf{mmse}_\pi(\cdot)$ is a strictly decreasing function (Fact 1), it has a well-defined inverse[3] mmse_π^{-1}, which can be used to relate the previous two displays:

$$\kappa_t \geq \mathsf{dmmse}_\pi(\omega_{\mathrm{eff}}) = \mathsf{dmmse}_\pi \circ \mathsf{mmse}_\pi^{-1}(\mathscr{M}_{t-1}(f_1, \ldots, f_{t-1})) \stackrel{\mathrm{def}}{=} \kappa_t^\sharp, \tag{218}$$

which proves the claim made in (216) and concludes the proof of the first part of the lemma.

Proof of Claim (214). In light of the formula in (215), it suffices to show that:

$$\min\left\{\mathbb{E}\left[\{\mathsf{X}_\star - g(\mathsf{W}_{<t,\bullet}; \mathsf{A})\}^2\right] : g \in \mathrm{Span}(f_1, \ldots, f_{t-1}, f_t^\sharp)\right\} = \kappa_t^\sharp.$$

From (218), we already know that:

$$\min\left\{\mathbb{E}\left[\{\mathsf{X}_\star - g(\mathsf{W}_{<t,\bullet}; \mathsf{A})\}^2\right] : g \in \mathrm{Span}(f_1, \ldots, f_{t-1}, f_t^\sharp)\right\} \geq \kappa_t^\sharp$$

Hence, we only need to show the upper bound:

$$\min\left\{\mathbb{E}\left[\{\mathsf{X}_\star - g(\mathsf{W}_{<t,\bullet}; \mathsf{A})\}^2\right] : g \in \mathrm{Span}(f_1, \ldots, f_{t-1}, f_t^\sharp)\right\} \leq \kappa_t^\sharp. \tag{219}$$

We upper bound the LHS in (219) by taking $g = f_t^\sharp$ and use the DMMSE formula for a general multivariate Gaussian channel (see Lemma 3 in Appendix A.2) to obtain:

$$\begin{aligned}\min\left\{\mathbb{E}\left[\{\mathsf{X}_\star - g(\mathsf{W}_{<t,\bullet}; \mathsf{A})\}^2\right] : g \in \mathrm{Span}(f_1, \ldots, f_{t-1}, f_t^\sharp)\right\} &\leq \mathbb{E}[(\mathsf{X}_\star - f_t^\sharp(\mathsf{W}_{<t,\bullet}; \mathsf{A}))^2] \\ &\stackrel{(a)}{=} \mathrm{DMMSE}(\mathsf{X}_\star | \mathsf{W}_{<t,\bullet}; \mathsf{A}) \\ &\stackrel{\text{Lem. 3}}{=} \mathsf{dmmse}_\pi(\omega_{\mathrm{eff}}) \stackrel{(218)}{=} \kappa_t^\sharp,\end{aligned}$$

where step (a) follows by recalling that $f_t^\sharp$ was the DMMSE estimator for $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t-1,\bullet}; \mathsf{A})$. This concludes the proof of the lemma. □

E.3 Proof of Proposition 5

We begin by reminding the reader that the optimal degree-D lifted OAMP algorithm was given by:

$$\boldsymbol{w}_{t,i} = \left(\boldsymbol{Y}^i - \mathbb{E}_{\Lambda \sim \mu}[\Lambda^i] \cdot \boldsymbol{I}_N\right) \cdot f_t^\star\left(\boldsymbol{w}_{1,\bullet}, \ldots, \boldsymbol{w}_{t-1,\bullet}; \boldsymbol{a}\right) \quad \forall\, i \in [D],\ t \in \mathbb{N}, \tag{220}$$

At the end of t iterations, the optimal lifted OAMP algorithm estimates $\boldsymbol{x}_\star$ by:

$$\widehat{\boldsymbol{w}}_t^{(D)} = h_t^\star(\boldsymbol{w}_{1,\bullet}, \ldots, \boldsymbol{w}_{t,\bullet}; \boldsymbol{a}).$$

Description of the iterate denoisers and post-processing functions. The iterate denoisers functions $f_1^\star, f_2^\star, \ldots$ and the post-processing functions $h_1^\star, h_2^\star, \cdots$ were defined recursively: once $f_1^\star, \ldots, f_t^\star$ have been specified, let $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$ denote the state evolution random variables corresponding to the first t iterations of the resulting lifted OAMP algorithm. Then, the iterate denoiser for step $t+1$ is the DMMSE estimator for the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$. Lemma 3 in Appendix A.2 provides a general formula for the DMMSE estimator of a multivariate Gaussian channel. An application of this formula yields:

$$f_{t+1}^\star(w; a) \stackrel{\mathrm{def}}{=} \bar{\varphi}\left(\left\langle v_t^{(D)}, w\right\rangle; a \;\middle|\; \omega_t^{(D)}\right) \quad \forall\, w \in \mathbb{R}^{tD},\ a \in \mathbb{R}^k \tag{221a}$$

[3] Strictly speaking, Fact 1 guarantees that $\mathsf{mmse}_\pi : [0,1] \mapsto [0,1]$ is strictly increasing only when $\mathsf{mmse}_\pi(0) = \mathbb{E}[\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}]] > 0$. In the corner case when $\mathsf{mmse}_\pi(0) = \mathbb{E}[\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}]] = 0$, notice that $\mathsf{dmmse}_\pi(0) = \mathbb{E}[\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}] = 0$. Since mmse_π and dmmse_π are non-negative, non-increasing functions (cf. Fact 1 and Lemma 4), this means that $\mathsf{mmse}_\pi(\omega) = \mathsf{dmmse}_\pi(\omega) = 0$ for any $\omega \in [0,1]$. In this situation, we can define mmse_π^{-1} arbitrarily. Irrespective of our convention, $\mathsf{dmmse}_\pi \circ \mathsf{mmse}_\pi^{-1}(\omega) = 0$, since dmmse_π is a constant function which takes the value 0.

where:

$$v_t^{(D)} \stackrel{\text{def}}{=} v_{\text{opt}}(\mathsf{X}_\star|\mathsf{W}_{1,\bullet},\dots,\mathsf{W}_{t,\bullet};\mathsf{A}), \quad \omega_t^{(D)} \stackrel{\text{def}}{=} \omega_{\text{eff}}(\mathsf{X}_\star|\mathsf{W}_{1,\bullet},\dots,\mathsf{W}_{t,\bullet};\mathsf{A}) \tag{221b}$$

denote the optimal linear combination and effective SNR of the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet},\dots,\mathsf{W}_{t,\bullet};\mathsf{A})$ (as defined in Lemma 3). Moreover, the estimator at the end of t iterations is obtained by applying the MMSE estimator for the channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet},\dots,\mathsf{W}_{t,\bullet};\mathsf{A})$ to the iterates:

$$\widehat{\boldsymbol{w}}_t^{(D)} = h_t^\star(\boldsymbol{w}_{\le t,\bullet};a), \tag{221c}$$

$$h_t^\star(w;a) \stackrel{\text{def}}{=} \varphi\left(\left\langle v_t^{(D)}, w\right\rangle; a \,\middle|\, \omega_t^{(D)}\right) \quad \forall\, w \in \mathbb{R}^{tD},\ a \in \mathbb{R}^k \tag{221d}$$

Dynamics of the Optimal Lifted OAMP Algorithm. By Corollary 1, the joint distribution of $(\mathsf{X}_\star, \mathsf{W}_{\le t,\bullet};\mathsf{A})$ is given by:

$$(\mathsf{X}_\star;\mathsf{A}) \sim \pi, \tag{222a}$$

$$(\mathsf{W}_{1,\bullet},\dots,\mathsf{W}_{t,\bullet})|(\mathsf{X}_\star;\mathsf{A}) \sim \mathcal{N}\left(\mathsf{X}_\star \cdot \alpha_t \otimes q^{(D)}, \alpha_t\alpha_t^\top \otimes (Q^{(D)} - q^{(D)}q^{(D)\top}) + (\Sigma_t - \alpha_t\alpha_t^\top)\otimes \Gamma^{(D)}\right). \tag{222b}$$

In the above display, the entries of $q^{(D)} \in \mathbb{R}^D$, $Q^{(D)} \in \mathbb{R}^{D\times D}$, and $\Gamma^{(D)} \in \mathbb{R}^{D\times D}$ are given by:

$$q_i^{(D)} = \mathbb{E}[\Lambda_\nu^i] - \mathbb{E}[\Lambda^i],\ Q_{ij}^{(D)} = \mathbb{E}\left[\left(\Lambda_\nu^i - \mathbb{E}[\Lambda^i]\right)\cdot\left(\Lambda_\nu^j - \mathbb{E}[\Lambda^j]\right)\right],\ \Gamma_{ij}^{(D)} = \mathrm{Cov}[\Lambda^i,\Lambda^j] \quad \text{where } \Lambda\sim\mu,\ \Lambda_\nu\sim\nu, \tag{222c}$$

the entries of $\alpha_t \in \mathbb{R}^t$ and $\Sigma_t \in \mathbb{R}^{t\times t}$ are given by:

$$(\alpha_t)_s \stackrel{\text{def}}{=} \mathbb{E}[\mathsf{X}_\star \cdot f_s^\star(\mathsf{W}_{<s,\bullet};\mathsf{A})], \quad (\Sigma_t)_{s\tau} \stackrel{\text{def}}{=} \mathbb{E}[f_s^\star(\mathsf{W}_{<s,\bullet};\mathsf{A})\cdot f_\tau^\star(\mathsf{W}_{<\tau,\bullet};\mathsf{A})] \quad \forall\, s,\tau \in [t], \tag{222d}$$

and $\otimes$ denotes the Kronecker (or tensor) product for matrices. Finally, the asymptotic MSE of the estimator $\widehat{\boldsymbol{w}}_t^{(D)}$ returned after t iterations is given by:

$$\operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|^2}{N} = \mathbb{E}\left[(\mathsf{X}_\star - h_t^\star(\mathsf{W}_{\le t,\bullet};\mathsf{A}))^2\right] = \mathrm{MMSE}(\mathsf{X}_\star|\mathsf{W}_{\le t,\bullet};\mathsf{A}) \stackrel{\text{Lem. 3}}{=} \mathsf{mmse}_\pi(\omega_t^{(D)}). \tag{222e}$$

The following lemma provides a recursive formula for the effective SNR $\omega_t^{(D)} \stackrel{\text{def}}{=} \omega_{\text{eff}}(\mathsf{X}_\star|\mathsf{W}_{1,\bullet},\dots,\mathsf{W}_{t,\bullet};\mathsf{A})$ and the optimal linear combination $v_t^{(D)} \stackrel{\text{def}}{=} v_{\text{opt}}(\mathsf{X}_\star|\mathsf{W}_{1,\bullet},\dots,\mathsf{W}_{t,\bullet};\mathsf{A})$ for the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet},\dots,\mathsf{W}_{t,\bullet};\mathsf{A})$ at step t.

Lemma 19. *Let* $\omega_0^{(D)} \stackrel{def}{=} 0$ *and* $\mathsf{d}_0^{(D)} \stackrel{def}{=} \mathsf{mmse}_\pi(0) = \mathsf{dmmse}_\pi(0) = \mathbb{E}[\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}]]$.

1. *If* $\mathsf{d}_0^{(D)} < 1$, *then* $\omega_t^{(D)}$ *and* $v_t^{(D)}$ *admit the following recursive characterization:*

$$\begin{aligned}
\mathsf{d}_t^{(D)} &= \mathsf{dmmse}_\pi(\omega_{t-1}^{(D)}),\\
\omega_t^{(D)} &= q^{(D)\top}\left[Q^{(D)} + \frac{\mathsf{d}_t^{(D)}}{1-\mathsf{d}_t^{(D)}}\cdot\Gamma^{(D)}\right]^\dagger q^{(D)},\\
v_t^{(D)} &= (\omega_t^{(D)})^{-\frac{1}{2}}\cdot(1-\mathsf{d}_t^{(D)})^{-1}\cdot e_t \otimes \left[Q^{(D)} + \frac{\mathsf{d}_t^{(D)}}{1-\mathsf{d}_t^{(D)}}\cdot\Gamma^{(D)}\right]^\dagger q^{(D)},
\end{aligned}$$

   where $e_t = (0,0,\dots,1)^\top \in \mathbb{R}^t$ *denotes the last standard basis vector. Moreover, for each* $t\in\mathbb{N}$,

$$\omega_t^{(D)} > 0, \quad \mathsf{d}_t^{(D)} < 1 \quad \forall\, t \in \mathbb{N}.$$

2. *If* $\mathsf{d}_0^{(D)} = 1$, *then* $\omega_t^{(D)} = 0$ *for all* $t\in\mathbb{N}$.

Proof. See Appendix E.3.1. □

Eliminating Corner Cases. Next, we observe that the proof of Proposition 5 is quite simple in the corner cases $\mathsf{mmse}_\pi(0) = \mathbb{E}[\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}]] = 1$ and $\mathsf{mmse}_\pi(0) = \mathbb{E}[\mathrm{Var}[\mathsf{X}_\star|\mathsf{A}]] = 0$.

Case 1: $\mathsf{mmse}_\pi(0) = 1$**.** In this case, combining (222e) and Lemma 19 (Claim 2), we conclude that for any $t \in \mathbb{N}$,

$$\lim_{N\to\infty} \frac{\|\widehat{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|^2}{N} = \mathsf{mmse}_\pi(\omega_t^{(D)}) = \mathsf{mmse}_\pi(0) = 1.$$

On the other hand, the optimal OAMP algorithm (17) returns the estimator $\widehat{\boldsymbol{x}}_t = \mathbf{0}$ for any $t \in \mathbb{N}$. Hence,

$$\lim_{N\to\infty} \frac{\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_\star\|^2}{N} = \lim_{N\to\infty} \frac{\|\boldsymbol{x}_\star\|^2}{N} = \mathbb{E}[\mathsf{X}_\star^2] = 1 = \lim_{N\to\infty} \frac{\|\widehat{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|^2}{N},$$

as claimed.

Case 2: $\mathsf{mmse}_\pi(0) = 0$**.** In this situation, we observe that for any $t \in \mathbb{N}$,

$$0 \le \lim_{N\to\infty} \frac{\|\widehat{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|^2}{N} = \mathsf{mmse}_\pi(\omega_t^{(D)}) \overset{\text{Fact 1}}{\le} \mathsf{mmse}_\pi(0) = 0 \implies \lim_{N\to\infty} \frac{\|\widehat{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|^2}{N} = 0.$$

On the other hand, the simplified OAMP algorithm returns the estimator $\widehat{\boldsymbol{x}}_t = \varphi(\boldsymbol{a}|0)$ for any $t \in \mathbb{N}$. Hence,

$$\lim_{N\to\infty} \frac{\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_\star\|^2}{N} = \mathbb{E}[(\mathsf{X}_\star - \varphi(\mathsf{A}|0))^2] = \mathsf{mmse}_\pi(0) = 0 = \lim_{N\to\infty} \frac{\|\widehat{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|^2}{N},$$

as claimed.

Hence, in the remainder of the proof, we will assume that:

$$\mathsf{mmse}_\pi(0) = \mathsf{dmmse}_\pi(0) \in (0, 1). \tag{223}$$

Simplifying the Optimal Lifted OAMP Algorithm. Recall the estimator computed at the end of t iterations of the optimal lifted OAMP algorithm is:

$$\widehat{\boldsymbol{w}}_t = h_t^\star(\boldsymbol{w}_{1,\bullet}, \ldots, \boldsymbol{w}_{t,\bullet}),$$

where $h_t^\star$ is the MMSE estimator for the channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet})$:

$$h_t^\star(w; a) \overset{\text{def}}{=} \varphi\left(\left\langle v_t^{(D)}, w \right\rangle ; a|\omega_t^{(D)} \right) \quad \forall\, w \in \mathbb{R}^{tD},\ a \in \mathbb{R}^k.$$

To compute this estimator, we only need to track the following vector $\boldsymbol{x}_t^{(D)}$, which is a linear combination of the lifted OAMP iterates $\boldsymbol{w}_{\le t,\bullet}$ with weights given by $v_t^{(D)}$:

$$\boldsymbol{x}_t^{(D)} \overset{\text{def}}{=} \sum_{s=1}^{t} \sum_{i=1}^{D} (v_t^{(D)})_{s,i} \cdot \boldsymbol{w}_{s,i}. \tag{224}$$

Indeed, the estimator returned by the optimal lifted OAMP algorithm can be computed using $\boldsymbol{x}_t^{(D)}$:

$$\widehat{\boldsymbol{w}}_t^{(D)} = \varphi(\boldsymbol{x}_t^{(D)}; \boldsymbol{a}|\omega_t^{(D)}).$$

In fact, $\boldsymbol{x}_t^{(D)}$ can be tracked using a simpler iterative algorithm. To see this, we observe using the formula for $v_t^{(D)}$ from Lemma 19 (Claim 1) that we have:

$$\boldsymbol{x}_t^{(D)} \overset{\text{def}}{=} \sum_{i=1}^{D} (v_t^{(D)})_{s,i} \cdot \boldsymbol{w}_{s,i} = \frac{1}{(1-\mathsf{d}_t^{(D)})\sqrt{\omega_t^{(D)}}} \sum_{i=1}^{D} \left\{ \left(Q^{(D)} + \frac{\mathsf{d}_t^{(D)}}{1-\mathsf{d}_t^{(D)}} \cdot \Gamma^{(D)} \right)^{\dagger} q^{(D)} \right\}_i \cdot \boldsymbol{w}_{t,i} \tag{225a}$$

$$\overset{(220)}{=} \frac{1}{(1-\mathsf{d}_t^{(D)})\sqrt{\omega_t^{(D)}}} \left[\sum_{i=1}^{D} \left\{ \left(Q^{(D)} + \frac{\mathsf{d}_t^{(D)}}{1-\mathsf{d}_t^{(D)}} \cdot \Gamma^{(D)} \right)^{\dagger} q^{(D)} \right\}_i \cdot \left\{ \boldsymbol{Y}^i - \mathbb{E}_{\Lambda\sim\mu}[\Lambda^i] \cdot \boldsymbol{I}_N \right\} \right] \cdot f_t^{\star}(\boldsymbol{w}_{<t;\bullet}; \boldsymbol{a}). \tag{225b}$$

In light of the above formula, we introduce the degree-D polynomial matrix denoising function:

$$\Psi_{\star}^{(D)}(\lambda;\rho) \overset{\text{def}}{=} \sum_{i=1}^{D} \left\{ \left(Q^{(D)} + \frac{1}{\rho} \cdot \Gamma^{(D)} \right)^{\dagger} q^{(D)} \right\}_i \cdot (\lambda^i - \mathbb{E}_{\Lambda\sim\mu}[\Lambda^i]) \quad \forall\, \lambda \in \mathbb{R},\ \rho \in (0,\infty], \tag{226}$$

as well as the sequence $(\rho_t^{(D)})_{t\in\mathbb{N}}$:

$$\rho_t^{(D)} \overset{\text{def}}{=} \frac{1}{\mathsf{d}_t^{(D)}} - 1, \quad \mathsf{d}_t^{(D)} = \frac{1}{\mathsf{dmmse}_{\pi}(\omega_t^{(D)})} - 1 \quad \forall\, t \in \mathbb{N}. \tag{227}$$

This leads to the following formula for $\boldsymbol{x}_t^{(D)}$:

$$\begin{aligned} \boldsymbol{x}_t^{(D)} &= \frac{1}{\sqrt{\omega_t^{(D)}}} \left(\frac{1}{\rho_t^{(D)}} + 1 \right) \cdot \Psi_{\star}^{(D)}(\boldsymbol{Y}; \rho_t^{(D)}) \cdot f_t^{\star}(\boldsymbol{w}_{1;\bullet}, \ldots, \boldsymbol{w}_{t-1;\bullet}; \boldsymbol{a}) \\ &\overset{(221),(224)}{=} \frac{1}{\sqrt{\omega_t^{(D)}}} \left(\frac{1}{\rho_t^{(D)}} + 1 \right) \cdot \Psi_{\star}^{(D)}(\boldsymbol{Y}; \rho_t^{(D)}) \cdot \bar{\varphi}(\boldsymbol{x}_{t-1}^{(D)}; \boldsymbol{a} | \omega_{t-1}^{(D)}) \end{aligned}$$

In summary, we have obtained an OAMP algorithm (in the sense of Definition 4):

$$\boldsymbol{x}_t^{(D)} = \frac{1}{\sqrt{\omega_t^{(D)}}} \left(\frac{1}{\rho_t^{(D)}} + 1 \right) \cdot \Psi_{\star}^{(D)}(\boldsymbol{Y}; \rho_t^{(D)}) \cdot \bar{\varphi}(\boldsymbol{x}_{t-1}^{(D)}; \boldsymbol{a} | \omega_{t-1}^{(D)}) \quad \forall\, t \in \mathbb{N}, \tag{228}$$

which can reconstruct the estimator $\widehat{\boldsymbol{w}}_t^{(D)}$ computed by the optimal degree-D lifted OAMP algorithm in (220):

$$\widehat{\boldsymbol{w}}_t^{(D)} = \varphi(\boldsymbol{x}_t^{(D)}; \boldsymbol{a} | \omega_t^{(D)}) \quad \forall\, t \in \mathbb{N}.$$

Large-degree limit. Notice the similarity between the simplified version of the optimal degree-D lifted OAMP algorithm in (228) and the optimal OAMP algorithm introduced in (17). Our goal will be to show that the performance of the simplified OAMP algorithm in (228) converges to the performance of the optimal OAMP algorithm in (17) as $D \to \infty$. To do so, we will rely on the following lemma which shows that the matrix denoiser $\Psi_{\star}^{(D)}$ used in the simplified OAMP algorithm (228) can be viewed as the minimizer of a variational problem over degree-D polynomials. By studying the same variational problem over $L^2(\mu+\nu)$, we obtain a candidate for the optimal matrix denoiser in the large degree limit, which turns out to be precisely the matrix denoiser (17e) used by the optimal OAMP algorithm in (17).

Lemma 20. *Let $\mathscr{L}_D$ and $\mathscr{L}_\infty$ denote the following functions on the domain $(0,\infty)$:*

$$\mathscr{L}_D(\rho) \overset{def}{=} \inf_{\Psi\in\mathcal{P}_D} \mathbb{E}[|\Psi(\Lambda_\nu) - 1|^2] + \frac{1}{\rho} \cdot \mathbb{E}[\Psi^2(\Lambda)] \quad \textit{subject to} \quad \mathbb{E}[\Psi(\Lambda)] = 0, \tag{229}$$

$$\mathscr{L}_\infty(\rho) \overset{def}{=} \inf_{\Psi\in L^2(\mu+\nu)} \mathbb{E}[|\Psi(\Lambda_\nu) - 1|^2] + \frac{1}{\rho} \cdot \mathbb{E}[\Psi^2(\Lambda)] \quad \textit{subject to} \quad \mathbb{E}[\Psi(\Lambda)] = 0, \tag{230}$$

where $\Lambda \sim \mu, \Lambda_\nu \sim \nu$, $\mathcal{P}_D$ denotes the set of all polynomial functions on $\mathbb{R}$ with degree at most D, and $L^2(\mu+\nu)$ denotes the set of all real valued functions on $\mathbb{R}$ which are square integrable with respect to $\mu+\nu$. Then, we have:

1. *For any $\rho \in (0,\infty)$, the function $\Psi_\star^{(D)}(\cdot;\rho)$ defined in (226) is the minimizer of the variational problem in (229) and,*

$$\mathscr{L}_D(\rho) = 1 - q_D^\top \cdot \left(Q^{(D)} + \frac{1}{\rho}\cdot \Gamma^{(D)}\right)^\dagger \cdot q^{(D)} = 1 - \mathbb{E}[\Psi_\star^{(D)}(\Lambda_\nu;\rho)].$$

2. *For any $\rho \in (0,\infty)$, the function:*

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

   is a minimizer of the variational problem in (230). In the above display, for any $\lambda \in \mathbb{R}$, $\phi(\lambda) \stackrel{def}{=} (1-\pi\theta\mathscr{H}_\mu(\lambda))^2 + \pi^2\theta^2\mu^2(\lambda)$, where $\mathscr{H}_\mu$ is the Hilbert transform of μ and θ is the SNR of the spiked matrix model (1).

3. *For any $D \in \mathbb{N}\cup\{\infty\}$, the function $\mathscr{L}_D$ maps the interval $(0,\infty)$ to $(0,1]$.*

4. *Let $(\rho_D)_{D\in\mathbb{N}}$ be any non-decreasing sequence in $(0,\infty)$ which converges to $\rho \in (0,\infty)$ as $D\to\infty$. Then, $\mathscr{L}_D(\rho_D) \downarrow \mathscr{L}_\infty(\rho)$ as $D\to\infty$. Moreover, $\Psi_\star^{(D)}(\cdot\,;\rho_D)$ converges to $\Psi_\star(\cdot\,;\rho)$ in $L^2(\mu+\nu)$ as $D\to\infty$.*

5. *For any $\rho \in (0,\infty)$, the function $\Psi_\star(\cdot;\rho)$ satisfies $\mathbb{E}[\Psi_\star(\Lambda_\nu;\rho)] = 1 - \mathscr{L}_\infty(\rho)$.*

Proof. The proof of this lemma is provided in Appendix E.3.2. □

Proof of Proposition 5. We now have all the ingredients to complete the proof of Proposition 5. Recall that the optimal estimator that can be computed using t iterations of a degree-D lifted OAMP algorithm is:

$$\widehat{\boldsymbol{w}}_t^{(D)} = \varphi(\boldsymbol{x}_t^{(D)};\boldsymbol{a}|\omega_t^{(D)}),$$

where the iterates $\boldsymbol{x}_t$ are computed using the update rule:

$$\boldsymbol{x}_t^{(D)} = \frac{1}{\sqrt{\omega_t^{(D)}}}\left(\frac{1}{\rho_t^{(D)}}+1\right)\cdot \Psi_\star^{(D)}(\boldsymbol{Y};\rho_t^{(D)}) \cdot \bar{\varphi}(\boldsymbol{x}_{t-1}^{(D)};\boldsymbol{a}|\omega_{t-1}^{(D)}) \quad \forall\, t \in \mathbb{N}. \tag{231}$$

Furthermore, the limiting mean-squared error of the estimator $\widehat{\boldsymbol{w}}_t^{(D)}$ is given by:

$$\operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|^2}{N} = \mathsf{mmse}_\pi(\omega_t^{(D)}). \tag{232}$$

In the above equations, the parameters $\{\omega_t^{(D)}\}_{t\in\mathbb{N}}, \{\mathsf{d}_t^{(D)}\}_{t\in\mathbb{N}}, \{\rho_t^{(D)}\}_{t\in\mathbb{N}}$ are updated according to the following recursion from Lemma 19 (also recall the definition of $\rho_t^{(D)}$ from (227)) initialized with $\omega_0^{(D)} = 0$:

$$\mathsf{d}_t^{(D)} = \mathsf{dmmse}_\pi(\omega_{t-1}^{(D)}), \quad \rho_t^{(D)} = \frac{1}{\mathsf{d}_t^{(D)}} - 1, \quad \omega_t^{(D)} = (q^{(D)})^\top \left[Q^{(D)} + \frac{1}{\rho_t^{(D)}}\cdot\Gamma^{(D)}\right]^\dagger q^{(D)} \stackrel{\text{(a)}}{=} 1 - \mathscr{L}_D(\rho_t^{(D)}), \tag{233}$$

where the equality marked (a) follows from Lemma 20 (Claim (2)). To prove Proposition 5, we need to show that in the large degree limit ($D\to\infty$), the performance of the estimator $\widehat{\boldsymbol{w}}_t^{(D)}$ is identical to the performance of the estimator $\widehat{\boldsymbol{x}}_t$:

$$\widehat{\boldsymbol{x}}_t \stackrel{\text{def}}{=} \varphi(\boldsymbol{x}_t;\boldsymbol{a}|\omega_t) \quad \forall\, t \in \mathbb{N},$$

returned by the optimal OAMP algorithm from (17):

$$\boldsymbol{x}_t = \frac{1}{\sqrt{\omega_t}}\left(1+\frac{1}{\rho_t}\right)\cdot \Psi_\star(\boldsymbol{Y};\rho_t)\cdot \bar{\varphi}(\boldsymbol{x}_{t-1};\boldsymbol{a}|\omega_{t-1}) \quad \forall\, t \in \mathbb{N}. \tag{234}$$

In the above equation, the parameters $\{\omega_t\}_{t\in\mathbb{N}}, \{\mathsf{d}_t\}_{t\in\mathbb{N}}, \{\rho_t\}_{t\in\mathbb{N}}$ are updated according to the following recursion initialized with $\omega_0 = 0$:

$$\mathsf{d}_t = \mathsf{dmmse}_\pi(\omega_{t-1}), \quad \rho_t = \frac{1}{\mathsf{d}_t} - 1, \quad \omega_t = 1 - \mathscr{L}_\infty(\rho_t). \tag{235}$$

Let $(\mathsf{X}_\star, (\mathsf{X}_t)_{t\in\mathbb{N}}; \mathsf{A})$ denote the state evolution random variables corresponding to the above algorithm. We will show by induction that for each $t \in \mathbb{N}$,

1. for any $D \in \mathbb{N}$, the sequences the sequences $\{\omega_t^{(D)}\}_{t\in\mathbb{N}}, \{\mathsf{d}_t^{(D)}\}_{t\in\mathbb{N}}, \{\rho_t^{(D)}\}_{t\in\mathbb{N}}$ generated by (233) satisfy:

$$\omega_t^{(D)} \in [0,1), \quad \mathsf{d}_t^{(D)} \in (0,1), \quad \rho_t^{(D)} \in (0,\infty) \quad \forall\, t \in \mathbb{N}. \tag{236}$$

   In particular, the update equation for $\rho_t^{(D)}$ in (233) is well-defined (avoids division by zero).

2. $\omega_t^{(D)} \uparrow \omega_t$, $\mathsf{d}_t^{(D)} \downarrow \mathsf{d}_t$, and $\rho_t^{(D)} \uparrow \rho_t$ as $D \to \infty$.

These claims immediately imply Proposition 5:

$$\operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{x}}_t - \boldsymbol{x}_\star\|^2}{N} \overset{\text{Prop. 1}}{=} \mathsf{mmse}_\pi(\omega_t) \overset{\text{(a)}}{=} \lim_{D\to\infty} \mathsf{mmse}_\pi(\omega_t^{(D)}) \overset{(232)}{=} \lim_{D\to\infty} \operatorname*{plim}_{N\to\infty} \frac{\|\widehat{\boldsymbol{w}}_t^{(D)} - \boldsymbol{x}_\star\|^2}{N},$$

as claimed. In the above display step (a) follows from the claim $\omega_t^{(D)} \uparrow \omega_t$ as $D \to \infty$. We now prove the three claims made above. As the induction hypothesis, we assume the claims hold for some $t \in \mathbb{N}$ and verify that they continue to hold at step $t+1$.

Proof of Claim (1). We focus on proving (236) holds at iteration $t+1$, assuming it holds at iteration t. The induction hypothesis $\omega_t^{(D)} \geq 0$ and the monotonicity of $\mathsf{dmmse}_\pi(\cdot)$ (see Lemma 4 in Appendix A.2) imply that $\mathsf{d}_{t+1}^{(D)} = \mathsf{dmmse}_\pi(\omega_t^{(D)}) \leq \mathsf{dmmse}_\pi(0) = \mathsf{mmse}_\pi(0) < 1$ (recall (223)). On the other hand, since $\omega_t^{(D)} < 1$, by the strict monotonicity of mmse_π (see Fact 1 in Appendix A.2), we conclude that $\mathsf{d}_{t+1}^{(D)} = \mathsf{dmmse}_\pi(\omega_t^{(D)}) \geq \mathsf{mmse}_\pi(\omega_t^{(D)}) > \mathsf{mmse}_\pi(1) = 0$. Hence, $\mathsf{d}_{t+1}^{(D)} \in (0,1)$. Since $\rho_{t+1}^{(D)} = (\mathsf{d}_{t+1}^{(D)})^{-1} - 1$, we conclude that $\rho_{t+1}^{(D)} \in (0,\infty)$. Finally, since $\omega_{t+1} = 1 - \mathscr{L}_D(\rho_{t+1}^{(D)})$ and Lemma 20 guarantees that $\mathscr{L}_D : (0,\infty) \mapsto (0,1]$, we conclude that $\omega_t^{(D)} \in [0,1)$, as claimed in (236). This proves the first claim.

Proof of Claim (2). We recall from (233) that $\mathsf{d}_{t+1}^{(D)} = \mathsf{dmmse}_\pi(\omega_t^{(D)})$. By the induction hypothesis $\omega_t \uparrow \omega_t \in (0,1)$. Since $\mathsf{dmmse}_\pi(\cdot)$ is a non-increasing continuous function (see Lemma 4 in Appendix A.2), we conclude that $\mathsf{d}_{t+1} \downarrow \mathsf{dmmse}_\pi(\omega_t) \overset{(235)}{=} \mathsf{d}_{t+1} \in (0,1)$ and $\rho_{t+1}^{(D)} \overset{(233)}{=} (\mathsf{d}_{t+1}^{(D)})^{-1} - 1 \uparrow (\mathsf{d}_{t+1})^{-1} - 1 \overset{(235)}{=} \rho_t$. Finally, recalling the update rule for $\omega_{t+1}^{(D)}$ from (233) and using Lemma 20 (Claim 4), we conclude that: $\omega_{t+1}^{(D)} = 1 - \mathscr{L}_D(\rho_{t+1}^{(D)}) \uparrow 1 - \mathscr{L}_\infty(\rho_{t+1}) \overset{(235)}{=} \omega_{t+1}$. This proves the second claim. This concludes the proof of Proposition 5.

E.3.1 Proof of Lemma 19

The proof of Lemma 19 relies on the following intermediate result.

Claim 3. For any $s, t \in \mathbb{N}$ with $s \leq t$:

$$\mathbb{E}[f_s^\star(\mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{s-1,\bullet}; \mathsf{A}) \cdot f_t^\star(\mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t-1,\bullet}; \mathsf{A})] = \mathbb{E}[\mathsf{X}_\star \cdot f_s^\star(\mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{s-1,\bullet}; \mathsf{A})].$$

We will first prove Lemma 19 assuming this claim. The proof of Claim 3 is provided at the end of this section.

Proof of Lemma 19. Recall that for each $t \in \mathbb{N}$,

$$v_t^{(D)} \stackrel{\text{def}}{=} v_{\text{opt}}(\mathsf{X}_\star | \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A}), \quad \omega_t^{(D)} \stackrel{\text{def}}{=} \omega_{\text{eff}}(\mathsf{X}_\star | \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A}),$$

where $(\mathsf{X}_\star, (\mathsf{W}_{t,\bullet})_{t\in\mathbb{N}}; \mathsf{A})$ are the state evolution random variables associated with the optimal lifted OAMP algorithm. In addition, we will find it helpful to introduce the quantity:

$$\mathsf{d}_t^{(D)} \stackrel{\text{def}}{=} \min \left\{ \mathbb{E}\left[\{\mathsf{X}_\star - g(\mathsf{W}_{<t,\bullet}; \mathsf{A})\}^2 \right] : g \in \text{Span}(f_1^\star, \ldots, f_t^\star) \right\}.$$

Since $\text{Span}(f_1^\star) \subset \text{Span}(f_1^\star, f_2^\star) \subset \cdots \text{Span}(f_1^\star, \ldots, f_t^\star) \subset \cdots$, $(\mathsf{d}_t^{(D)})_{t\in\mathbb{N}}$ is a non-increasing sequence.

Case 1: $\mathsf{d}_0^{(D)} < 1$. Notice that since $(\mathsf{d}_t^{(D)})_{t\in\mathbb{N}}$ is a non-increasing sequence and $\mathsf{d}_0^{(D)} < 1$, we conclude that $\mathsf{d}_t^{(D)} < 1$ for all $t \in \mathbb{N}$, which is one of the claims pertaining to this case made in the lemma. To prove the other claims, we will show by induction on t that:

$$\mathsf{d}_t^{(D)} = \mathsf{dmmse}_\pi(\omega_{t-1}^{(D)}), \tag{237a}$$

$$\omega_t^{(D)} = q_D^\top \left[Q^{(D)} + \frac{\mathsf{d}_t^{(D)}}{1 - \mathsf{d}_t^{(D)}} \cdot \Gamma^{(D)} \right]^\dagger q^{(D)}, \tag{237b}$$

$$\omega_t^{(D)} > 0, \tag{237c}$$

$$v_t^{(D)} = (\omega_t^{(D)})^{-\frac{1}{2}} \cdot (1 - \mathsf{d}_t^{(D)})^{-1} \cdot e_1 \otimes \left[\cdot Q^{(D)} + \frac{\mathsf{d}_t^{(D)}}{1 - \mathsf{d}_t^{(D)}} \cdot \Gamma^{(D)} \right]^\dagger q^{(D)}. \tag{237d}$$

As our induction hypothesis we assume that (237) holds, and show that (237) also holds for $t+1$. Indeed, since $f_{t+1}^\star$ is the DMMSE estimator for the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{\le t,\bullet}; \mathsf{A})$, by Lemma 18,

$$\mathsf{d}_{t+1}^{(D)} \stackrel{\text{def}}{=} \min \left\{ \mathbb{E}\left[\{\mathsf{X}_\star - g(\mathsf{W}_{\le t,\bullet}; \mathsf{A})\}^2 \right] : g \in \text{Span}(f_1^\star, \ldots, f_{t+1}^\star) \right\} = \mathsf{dmmse}_\pi(\omega_t^{(D)}).$$

By appealing to Lemma 16, we conclude that:

$$\omega_{t+1}^{(D)} = q_D^\top \left[Q^{(D)} + \frac{\mathsf{d}_{t+1}^{(D)}}{1 - \mathsf{d}_{t+1}^{(D)}} \cdot \Gamma^{(D)} \right]^\dagger q^{(D)} > 0$$

$$v_{t+1}^{(D)} = (\omega_{t+1}^{(D)})^{-\frac{1}{2}} \cdot (1 - \mathsf{d}_{t+1}^{(D)})^{-1} \cdot (\Sigma_{t+1}^\dagger \alpha_{t+1}) \otimes \left[Q^{(D)} + \frac{\mathsf{d}_{t+1}^{(D)}}{1 - \mathsf{d}_{t+1}^{(D)}} \cdot \Gamma^{(D)} \right]^\dagger q^{(D)}.$$

From Claim 3, we know that:

$$(\alpha_{t+1})_s = (\Sigma_{t+1})_{s,t+1} \quad \forall\, s \in [t+1] \implies \alpha_{t+1} = \Sigma_{t+1} e_{t+1} \implies \Sigma_{t+1}^\dagger \alpha_{t+1} = e_{t+1}.$$

Hence,

$$v_{t+1}^{(D)} = (\omega_{t+1}^{(D)})^{-\frac{1}{2}} \cdot (1 - \mathsf{d}_{t+1}^{(D)})^{-1} \cdot e_{t+1} \otimes \left[Q^{(D)} + \frac{\mathsf{d}_{t+1}^{(D)}}{1 - \mathsf{d}_{t+1}^{(D)}} \cdot \Gamma^{(D)} \right]^\dagger q^{(D)},$$

as desired. This proves the claim of the lemma when $\mathsf{d}_0^{(D)} < 1$.

Case 2: $\mathsf{d}_0^{(D)} = 1$. In this situation, we will show by induction that for any $t \in \mathbb{N}$,

$$\mathsf{d}_t^{(D)} = 1, \quad \omega_t^{(D)} = 0.$$

We assume that the above claim holds at step t as the induction hypothesis. As in the previous case, by Lemma 18, we have:

$$\mathsf{d}_{t+1}^{(D)} = \mathsf{dmmse}_\pi(\omega_t^{(D)}) = \mathsf{dmmse}_\pi(0) = \mathsf{d}_0^{(D)} = 1.$$

Appealing to Lemma 16, we conclude that $\omega_{t+1}^{(D)} = 0$, as desired.This proves the claim of the lemma. □

We now present the proof of Claim 3.

Proof of Claim 3. By the Tower property, we have:

$$\mathbb{E}[\mathsf{X}_\star \cdot f_s^\star(\mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{s-1,\bullet}; \mathsf{A})] = \mathbb{E}[\mathbb{E}[\mathsf{X}_\star | \mathsf{W}_{<t,\bullet}, \mathsf{A}] f_s^\star(\mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{s-1,\bullet}; \mathsf{A})]. \tag{238}$$

Since $\mathbb{E}[\mathsf{X}_\star | \mathsf{W}_{<t,\bullet}, \mathsf{A}]$ is the MMSE estimator for the channel $(\mathsf{X}_\star, \mathsf{W}_{<t,\bullet}; \mathsf{A})$. Using the formula for the MMSE estimator for a general multivariate Gaussian channel (see Lemma 3 in Appendix A.2), we obtain:

$$\mathbb{E}[\mathsf{X}_\star | \mathsf{W}_{<t,\bullet}, \mathsf{A}] = \varphi\left(\langle \mathsf{W}_{<t,\bullet}, v_{\text{opt}} \rangle ; \mathsf{A} | \omega_{\text{eff}}\right), \tag{239}$$

where we use the shorthand notations ω_{eff} and v_{opt} denote the effective SNR and optimal linear combination[4] for the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t-1,\bullet}; \mathsf{A})$. Similarly, Lemma 3 gives the following formula for $f_t^\star$, the DMMSE estimator for this channel:

$$f_t^\star(\mathsf{W}_{<t,\bullet}; \mathsf{A}) = \bar{\varphi}\left(\langle \mathsf{W}_{<t,\bullet}, v_{\text{opt}} \rangle ; \mathsf{A} | \omega_{\text{eff}}\right). \tag{240}$$

Recall the formula for $\bar{\varphi}$ (the DMMSE estimator in a scalar Gaussian channel) provided in Definition 3 (cf. (11)):

$$\bar{\varphi}(x; a|\omega) \stackrel{\text{def}}{=} \begin{cases} \left(1 - \sqrt{\omega} \cdot \beta(\omega)\right)^{-1} \cdot \left(\varphi(x; a|\omega) - \beta(\omega) \cdot x\right) & : \omega < 1 \\ \varphi(x; a|\omega) & : \omega = 1 \end{cases}, \tag{241}$$

where:

$$\beta(\omega) \stackrel{\text{def}}{=} \frac{1}{\sqrt{1-\omega}} \cdot \mathbb{E}[\mathsf{Z} \varphi(\sqrt{\omega}\mathsf{X}_\star + \sqrt{1-\omega}\mathsf{Z}; \mathsf{A}|\omega)] \quad \text{where } (\mathsf{X}_\star; \mathsf{A}) \sim \pi,\ \mathsf{Z}|\mathsf{X}_\star; \mathsf{A} \sim \mathcal{N}(0,1).$$

Combining (239), (240), and (241) yields the following formula relating the MMSE and DMMSE estimators for the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$:

$$\mathbb{E}[\mathsf{X}_\star | \mathsf{W}_{<t,\bullet}, \mathsf{A}] = \begin{cases} \left(1 - \sqrt{\omega_{\text{eff}}} \cdot \beta\left(\omega_{\text{eff}}\right)\right) \cdot f_t^\star(\mathsf{W}_{<t,\bullet}; \mathsf{A}) + \beta\left(\omega_{\text{eff}}\right) \cdot \langle v_{\text{opt}}, \mathsf{W}_{<t,\bullet} \rangle & : \omega_{\text{eff}} < 1 \\ f_t^\star(\mathsf{W}_{<t,\bullet}; \mathsf{A}) & : \omega_{\text{eff}} = 1. \end{cases}$$

We substitute the above expression in (238). If $\omega_{\text{eff}} = 1$, the claim of the lemma is immediate. If $\omega_{\text{eff}} < 1$, we have:

$$\begin{aligned} &\mathbb{E}[\mathsf{X}_\star \cdot f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A})] \\ &\quad = \left(1 - \sqrt{\omega_{\text{eff}}} \cdot \beta\left(\omega_{\text{eff}}\right)\right) \cdot \mathbb{E}[f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A}) \cdot f_t^\star(\mathsf{W}_{<t,\bullet}; \mathsf{A})] + \beta\left(\omega_{\text{eff}}\right) \cdot \mathbb{E}[\langle v_{\text{opt}}, \mathsf{W}_{<t,\bullet} \rangle \cdot f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A})]. \end{aligned}$$

Let $\mathsf{Z}_{1,\bullet}, \ldots, \mathsf{Z}_{t-1,\bullet}$ denote the Gaussian noise random variables in the Gaussian channel $(\mathsf{X}_\star, \mathsf{W}_{1,\bullet}, \ldots, \mathsf{W}_{t,\bullet}; \mathsf{A})$. Since $(\mathsf{X}_\star, \langle v_{\text{opt}}, \mathsf{W}_{<t,\bullet} \rangle ; \mathsf{A})$ forms a scalar Gaussian channel with SNR ω_{eff}:

$$\langle v_{\text{opt}}, \mathsf{W}_{<t,\bullet} \rangle = \sqrt{\omega_{\text{eff}}} \cdot \mathsf{X}_\star + \sqrt{1 - \omega_{\text{eff}}} \cdot \mathsf{Z}_{\text{eff}}, \quad \mathsf{Z}_{\text{eff}} \stackrel{\text{def}}{=} \frac{\langle v_{\text{opt}}, \mathsf{Z}_{<t,\bullet} \rangle}{\sqrt{1 - \omega_{\text{eff}}}}.$$

Hence,

$$\begin{aligned} \mathbb{E}[\mathsf{X}_\star \cdot f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A})] &= (1 - \sqrt{\omega_{\text{eff}}} \cdot \beta(\omega_{\text{eff}})) \cdot \mathbb{E}[f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A}) \cdot f_t^\star(\mathsf{W}_{<t,\bullet}; \mathsf{A})] \\ &\qquad + \beta(\omega_{\text{eff}}) \cdot \mathbb{E}[(\sqrt{\omega_{\text{eff}}} \cdot \mathsf{X}_\star + \sqrt{1 - \omega_{\text{eff}}} \cdot \mathsf{Z}_{\text{eff}}) \cdot f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A})] \\ &= (1 - \sqrt{\omega_{\text{eff}}} \cdot \beta(\omega_{\text{eff}})) \cdot \mathbb{E}[f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A}) \cdot f_t^\star(\mathsf{W}_{<t,\bullet}; \mathsf{A})] + \sqrt{\omega_{\text{eff}}} \cdot \beta(\omega_{\text{eff}}) \cdot \mathbb{E}[\mathsf{X}_\star \cdot f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A})]. \end{aligned} \tag{242}$$

The final equality in the previous display follows by observing that:

$$\mathbb{E}[\mathsf{Z}_{\text{eff}} \cdot f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A})] = \mathbb{E}[\mathbb{E}[\mathsf{Z}_{\text{eff}} | \mathsf{X}_\star, \mathsf{Z}_{<s,\bullet}, \mathsf{A}] \cdot f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A})] = \mathbb{E}[\mathbb{E}[\mathsf{Z}_{\text{eff}} | \mathsf{Z}_{<s,\bullet}] \cdot f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A})] = 0.$$

This is because $\mathbb{E}[\mathbb{E}[\mathsf{Z}_{\text{eff}} | \mathsf{Z}_{<s,\bullet}]]$ is a linear combination of $\mathsf{Z}_{<s,\bullet}$ (since $(\mathsf{Z}_{<s,\bullet}, \mathsf{Z}_{\text{eff}})$ are jointly Gaussian) and the DMMSE estimator $f_s^\star$ is uncorrelated with the channel noise $\mathsf{Z}_{<s,\bullet}$. Rearranging (242) gives us:

$$\mathbb{E}[\mathsf{X}_\star \cdot f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A})] = \mathbb{E}[f_s^\star(\mathsf{W}_{<s,\bullet}; \mathsf{A}) \cdot f_t^\star(\mathsf{W}_{<t,\bullet}; \mathsf{A})],$$

which is precisely the claim set out to prove. □

[4] We suppress the dependence of these quantities on the iteration number t and the degree D of the lifted OAMP for notational convenience since it does not play a role in the proof.

E.3.2 Proof of Lemma 20

Proof of Lemma 20. We consider each of the claims made in the lemma.

Proof of Claim (1). To solve the variational problem:

$$\mathscr{L}_D(\rho) \stackrel{\text{def}}{=} \inf_{\Psi \in \mathcal{P}_D} \mathbb{E}[|\Psi(\Lambda_\nu) - 1|^2] + \frac{1}{\rho} \cdot \mathbb{E}[\Psi^2(\Lambda)] \quad \text{subject to} \quad \mathbb{E}[\Psi(\Lambda)] = 0,$$

we parameterize any polynomial f of degree at most D which satisfies $\mathbb{E}[\Psi(\Lambda)] = 0$ as:

$$\Psi(\lambda) = \sum_{i=0}^{D} v_i \cdot (\lambda^i - \mathbb{E}_{\Lambda \sim \mu}[\Lambda^i])) \quad \forall\, \lambda \in \mathbb{R}.$$

Hence,

$$\mathscr{L}_D(\rho) = \min_{v \in \mathbb{R}^D} \mathbb{E}\left| \sum_{i=1}^{D} v_i \cdot (\Lambda_\nu^i - \mathbb{E}_{\Lambda \sim \mu}[\Lambda^i])) - 1 \right|^2 + \frac{1}{\rho} \cdot \mathbb{E}\left| \sum_{i=1}^{D} v_i \cdot (\Lambda^i - \mathbb{E}_{\Lambda \sim \mu}[\Lambda^i])) \right|^2. \tag{243}$$

Expanding the square norms and recalling that the entries of $q^{(D)} \in \mathbb{R}^D$, $Q^{(D)} \in \mathbb{R}^{D \times D}$, and $\Gamma^{(D)} \in \mathbb{R}^{D \times D}$ are given by:

$$q_i^{(D)} = \mathbb{E}[\Lambda_\nu^i] - \mathbb{E}[\Lambda^i],\ Q_{ij}^{(D)} = \mathbb{E}\left[\left(\Lambda_\nu^i - \mathbb{E}[\Lambda^i]\right) \cdot \left(\Lambda_\nu^j - \mathbb{E}[\Lambda^j]\right)\right],\ \Gamma_{ij}^{(D)} = \mathrm{Cov}[\Lambda^i, \Lambda^j] \quad \text{where } \Lambda \sim \mu,\ \Lambda_\nu \sim \nu, \tag{244}$$

we obtain the following formula for $\mathscr{L}_D(\rho)$:

$$\mathscr{L}_D(\rho) = \min_{v \in \mathbb{R}^D} 1 + v^\top Q^{(D)} v - 2\left\langle q^{(D)}, v \right\rangle + \frac{v^\top \Gamma^{(D)} v}{\rho}.$$

The optimizer of the above quadratic form is:

$$v = \left(Q^{(D)} + \frac{1}{\rho} \cdot \Gamma^{(D)}\right)^\dagger \cdot q^{(D)}.$$

Hence, a minimizer of (243) is:

$$\Psi(\lambda) = \sum_{i=1}^{D} \left\{ \left(Q^{(D)} + \frac{1}{\rho} \cdot \Gamma^{(D)}\right)^\dagger \cdot q^{(D)} \right\}_i \cdot (\lambda^i - \mathbb{E}_{\Lambda \sim \mu}[\Lambda^i]) \stackrel{(226)}{=} \Psi_\star^{(D)}(\lambda; \rho) \quad \forall\, \lambda \in \mathbb{R},$$

and:

$$\mathscr{L}_D(\rho) = 1 - {q^{(D)}}^\top \left(Q^{(D)} + \frac{1}{\rho} \cdot \Gamma_D\right)^\dagger q^{(D)} = 1 - \left\langle q^{(D)}, v \right\rangle = 1 - \mathbb{E}[\Psi_\star^{(D)}(\Lambda_\nu; \rho)] \quad \text{where } \Lambda_\nu \sim \nu,$$

as claimed.

Proof of Claim (2). Our goal is to solve the optimization problem:

$$\mathscr{L}_\infty(\rho) \stackrel{\text{def}}{=} \inf_{\Psi \in L^2(\mu+\nu)} \mathbb{E}[|\Psi(\Lambda_\nu) - 1|^2] + \frac{1}{\rho} \cdot \mathbb{E}[\Psi^2(\Lambda)] \quad \text{subject to} \quad \mathbb{E}[\Psi(\Lambda)] = 0. \tag{245}$$

This is exactly the problem considered in (35c), where we showed that the minimizer of (245) is:

$$\Psi(\lambda) = 1 - \left(\mathbb{E}_{\Lambda \sim \mu}\left[\frac{\phi(\Lambda)}{\phi(\Lambda) + \rho}\right]\right)^{-1} \cdot \frac{\phi(\lambda)}{\phi(\lambda) + \rho} \stackrel{\text{def}}{=} \Psi_\star(\lambda; \rho) \quad \forall\, \lambda \in \mathbb{R}, \tag{246}$$

as claimed.

Proof of Claim (3). We need to show that for any $D \in \mathbb{N} \cup \{\infty\}$, the function $\mathscr{L}_D$ maps the interval $(0,\infty)$ to $(0,1]$. We will show this claim for $\mathscr{L}_\infty$ and the exact same argument works for $\mathscr{L}_D$ for any $D \in \mathbb{N}$. The definition of $\mathscr{L}_\infty(\rho)$ (cf. (245)) implies that $\mathscr{L}_\infty(\rho) \geq 0$. Taking $\Psi = 0$ in (245) shows that $\mathscr{L}_\infty(\rho) \leq 1$. To prove Claim (3), we need to verify that if $\rho \in (0,\infty)$ then, $\mathscr{L}_\infty(\rho) > 0$. We prove this by contradiction. If $\mathscr{L}_\infty(\rho) = 0$, then the function $\Psi_\star(\cdot;\rho)$, which minimizes the objective in (245) satisfies $\mathbb{E}[|\Psi_\star(\Lambda_\nu;\rho) - 1|^2] + \rho^{-1} \cdot \mathbb{E}[|\Psi_\star(\Lambda;\rho)|^2] = 0$. This means that:

$$\nu(\{\lambda \in \mathbb{R} : \Psi_\star(\lambda;\rho) = 1\}) = 1, \quad \mu(\{\lambda \in \mathbb{R} : \Psi_\star(\lambda;\rho) = 0\}) = 1.$$

In particular μ, ν are mutually singular. However, this contradicts Lemma 1 which tells us that $\nu_{\|}$, the absolutely continuous part of ν with respect to the Lebesgue measure has density $\frac{\mathrm{d}\nu_{\|}}{\mathrm{d}\lambda}(\lambda) = \frac{\mu(\lambda)}{\phi(\lambda)}$, and hence $\nu_{\|} \ll \mu$.

Proof of Claim (4). Consider a non-decreasing sequence $(\rho_D)_{D\in\mathbb{N}}$ which converges to $\rho \in (0,\infty)$ as $D \to \infty$. We begin by observing that $\mathscr{L}_D(\rho_D)$ is a non-increasing sequence:

$$\begin{aligned}
\mathscr{L}_{D+1}(\rho_{D+1}) &\overset{(243)}{=} \inf_{\Psi\in\mathcal{P}_{D+1}} \mathbb{E}[|\Psi(\Lambda_\nu) - 1|^2] + \frac{1}{\rho_{D+1}} \cdot \mathbb{E}[\Psi^2(\Lambda)] \quad \text{subject to} \quad \mathbb{E}[\Psi(\Lambda)] = 0 \\
&\overset{(a)}{\leq} \inf_{\Psi\in\mathcal{P}_{D+1}} \mathbb{E}[|\Psi(\Lambda_\nu) - 1|^2] + \frac{1}{\rho_{D}} \cdot \mathbb{E}[\Psi^2(\Lambda)] \quad \text{subject to} \quad \mathbb{E}[\Psi(\Lambda)] = 0 \\
&\overset{(b)}{\leq} \inf_{\Psi\in\mathcal{P}_{D}} \mathbb{E}[|\Psi(\Lambda_\nu) - 1|^2] + \frac{1}{\rho_{D}} \cdot \mathbb{E}[\Psi^2(\Lambda)] \quad \text{subject to} \quad \mathbb{E}[\Psi(\Lambda)] = 0 \overset{(243)}{=} \mathscr{L}_D(\rho_D).
\end{aligned}$$

In the above display, step (a) follows by observing that $\rho_D \leq \rho_{D+1}$. Step (b) relies on the observation that $\mathcal{P}_D \subset \mathcal{P}_{D+1}$ and the infimum over larger sets is smaller. Hence $(\mathscr{L}_D(\rho_D))_{D\in\mathbb{N}}$ is a non-increasing sequence. To show that $\mathscr{L}_D(\rho_D) \downarrow \mathscr{L}_\infty(\rho)$, we begin by observing that by repeating the argument used in the previous display, we can also obtain:

$$\mathscr{L}_D(\rho_D) \geq \mathscr{L}_\infty(\rho).$$

Hence, we have obtained a lower bound for $\mathscr{L}_D(\rho_D)$ in terms of $\mathscr{L}_\infty(\rho)$. To show that in fact $\mathscr{L}_D(\rho_D) \downarrow \mathscr{L}_\infty(\rho)$ we will also need an upper bound. Towards this goal, we consider a sequence of functions $\{\widehat{\Psi}_D\}_{D\in\mathbb{N}}$ indexed by $D \in \mathbb{N}$ such that $\widehat{\Psi}_D : \mathbb{R} \mapsto \mathbb{R}$ is a degree-D polynomial and:

$$\lim_{D\to\infty} \|\widehat{\Psi}_D - \Psi_\star(\cdot;\rho)\|^2_{\mu+\nu} = 0 \quad \Leftrightarrow \quad \lim_{D\to\infty} \left(\mathbb{E}_{\Lambda\sim\mu}|\widehat{\Psi}_D(\Lambda) - \Psi_\star(\Lambda;\rho)|^2 + \mathbb{E}_{\Lambda_\nu\sim\nu}|\widehat{\Psi}_D(\Lambda_\nu) - \Psi_\star(\Lambda_\nu);\rho)|^2\right) = 0. \tag{247}$$

The existence of these polynomial approximations follows by the fact that polynomials form a dense subset of $L^2(\mu+\nu)$ since $\mu+\nu$ is a compactly supported measure [70, Corollary 14.24, Definition 14.1]. We can use the function $\lambda \mapsto \widehat{\Psi}_D(\lambda) - \mathbb{E}[\widehat{\Psi}_D(\Lambda)]$ as a candidate minimizer to upper bound $\mathscr{L}_D(\rho_D)$:

$$\mathscr{L}_\infty(\rho) \leq \mathscr{L}_D(\rho_D) \leq \mathbb{E}\left[\left|\widehat{\Psi}_D(\Lambda_\nu) - 1 - \mathbb{E}[\widehat{\Psi}_D(\Lambda)]\right|^2\right] + \frac{1}{\rho_D} \cdot \mathbb{E}\left[\left|\widehat{\Psi}_D(\Lambda) - \mathbb{E}[\widehat{\Psi}_D(\Lambda)]\right|^2\right]$$

We let $D \to \infty$ in the above display and exploit (247) and $\rho_D \to \rho \in (0,\infty)$ to conclude that:

$$\lim_{D\to\infty} \mathscr{L}_D(\rho_D) = \mathbb{E}[|\Psi_\star(\Lambda_\nu;\rho) - 1|^2] + \frac{1}{\rho} \cdot \mathbb{E}[\Psi^2_\star(\Lambda;\rho)] = \mathscr{L}_\infty(\rho),$$

as claimed. Finally, we show that:

$$\lim_{D\to\infty} \mathbb{E}_{\Lambda\sim\mu}|\Psi^{(D)}_\star(\Lambda;\rho_D) - \Psi_\star(\Lambda;\rho)|^2 = 0, \quad \lim_{D\to\infty} \mathbb{E}_{\Lambda_\nu\sim\nu}|\Psi^{(D)}_\star(\Lambda_\nu) - \Psi_\star(\Lambda_\nu);\rho_D)|^2 = 0$$

To do so, we will exploit the strong convexity of the objective function:

$$\mathscr{O}(\Psi) \overset{\text{def}}{=} \mathbb{E}[|\Psi(\Lambda_\nu) - 1|^2] + \frac{1}{\rho} \cdot \mathbb{E}[\Psi^2(\Lambda)]. \tag{248}$$

Notice that $\mathscr{O}(\Psi) - \mathbb{E}[|\Psi(\Lambda_\nu) - 1|^2]$ is convex and hence for any $\lambda \in (0,1)$ and any $\Psi, \widehat{\Psi} \in L^2(\mu+\nu)$,

$$\begin{aligned}\mathscr{O}(\lambda\Psi + (1-\lambda)\widehat{\Psi}) - \mathbb{E}[|\lambda\Psi(\Lambda_\nu) + (1-\lambda)\widehat{\Psi}(\Lambda_\nu) - 1|^2] \\ \leq \lambda\mathscr{O}(\Psi) + (1-\lambda)\mathscr{O}(\widehat{\Psi}) - \lambda\mathbb{E}[|\Psi(\Lambda_\nu) - 1|^2] - (1-\lambda)\mathbb{E}[|\widehat{\Psi}(\Lambda_\nu) - 1|^2].\end{aligned}$$

Rearranging yields:

$$\mathbb{E}[|\Psi(\Lambda_\nu) - \widehat{\Psi}(\Lambda_\nu)|^2] \leq \frac{\lambda\mathscr{O}(\Psi) + (1-\lambda)\mathscr{O}(\widehat{\Psi}) - \mathscr{O}(\lambda\Psi + (1-\lambda)\Psi)}{\lambda(1-\lambda)}.$$

We instantiate the above inequality with $\lambda = 1/2$, $\Psi = \Psi_\star(\cdot\,;\rho)$ and $\widehat{\Psi} = \Psi_\star^{(D)}(\cdot\,;\rho_D)$. For notational convenience, we define the "midpoint" between these two matrix denoisers as $\Phi^{(D)} \stackrel{\text{def}}{=} (\Psi_\star(\cdot\,;\rho) + \Psi_\star^{(D)}(\cdot\,;\rho_D))/2$:

$$\mathbb{E}[|\Psi(\Lambda_\nu) - \widehat{\Psi}(\Lambda_\nu)|^2] \leq 2 \cdot \left[\mathscr{O}(\Psi_\star(\cdot\,;\rho)) + \mathscr{O}(\Psi_\star^{(D)}(\cdot\,;\rho_D)) - 2\mathscr{O}(\Phi^{(D)})\right] \tag{249}$$

Recalling that $\Psi_\star(\cdot\,;\rho)$ is the minimizer of the variational problem that defines $\mathscr{L}_\infty(\rho)$ (claim 2 of the lemma):

$$\mathscr{O}(\Psi_\star(\cdot\,;\rho)) \stackrel{(248)}{=} \mathbb{E}[|\Psi_\star(\Lambda_\nu) - 1|^2] + \frac{1}{\rho} \cdot \mathbb{E}[\Psi_\star^2(\Lambda)] = \mathscr{L}_\infty(\rho). \tag{250}$$

Moreover, since $\Phi^{(D)}$ is a feasible point for the optimization problem that defines $\mathscr{L}_\infty(\rho)$ (recall (245)), we conclude that:

$$\mathscr{O}(\Phi^{(D)}) \stackrel{(248)}{=} \mathbb{E}[|\Phi^{(D)}(\Lambda_\nu) - 1|^2] + \frac{1}{\rho} \cdot \mathbb{E}[\Phi^{(D)}(\Lambda)^2] \stackrel{(245)}{\geq} \mathscr{L}_\infty(\rho). \tag{251}$$

Recall that $\rho_D \uparrow \rho$ and that $\Psi_\star^{(D)}(\cdot\,;\rho_D)$ is the minimizer of the variational problem that defines $\mathscr{L}_D(\rho_D)$. Hence,

$$\begin{aligned}\mathscr{O}(\Psi_\star^{(D)}(\cdot\,;\rho_D)) &\stackrel{(248)}{=} \mathbb{E}[|\Psi_\star^{(D)}(\Lambda_\nu;\rho_D) - 1|^2] + \frac{1}{\rho} \cdot \mathbb{E}[\Psi_\star^{(D)}(\Lambda;\rho_D)^2] \\ &\leq \mathbb{E}[|\Psi_\star^{(D)}(\Lambda_\nu;\rho_D) - 1|^2] + \frac{1}{\rho_D} \cdot \mathbb{E}[\Psi_\star^{(D)}(\Lambda;\rho_D)^2] = \mathscr{L}_D(\rho_D).\end{aligned} \tag{252}$$

Plugging in (250), (251), and (252) into (249), we obtain:

$$\mathbb{E}[|\Psi(\Lambda_\nu) - \widehat{\Psi}(\Lambda_\nu)|^2] \leq 2 \cdot (\mathscr{L}_D(\rho_D) - \mathscr{L}_\infty(\rho)).$$

We have already shown that $\mathscr{L}_D(\rho_D) \to \mathscr{L}_\infty(\rho)$ as $D \to \infty$. Hence,

$$\lim_{D\to\infty} \mathbb{E}[|\Psi(\Lambda_\nu) - \widehat{\Psi}(\Lambda_\nu)|^2] = 0. \tag{253}$$

An analogous argument yields $\lim_{D\to\infty} \mathbb{E}[|\Psi(\Lambda) - \widehat{\Psi}(\Lambda)|^2] = 0$, which completes the proof of Claim (4) in the statement of the lemma.

Proof of Claim (5). Recall that from the first claim of the lemma:

$$1 - \mathbb{E}_{\Lambda\sim\nu}[\Psi_\star^{(D)}(\Lambda;\rho)] = \mathscr{L}_D(\rho).$$

Taking $D \to \infty$ and using the fact that $\mathscr{L}_D(\rho) \to \mathscr{L}_\infty(\rho)$ and (253) yields the desired conclusion. This concludes the proof of the lemma. □

F Some Miscellaneous Results

Fact 2 (Benaych-Georges and Nadakuditi 12, Proposition 9.3). Let $\boldsymbol{W}$ be a $N \times N$ random noise matrix which satisfies Assumption 2 and let $\boldsymbol{u}$ be a N-dimensional random vector which is independent of $\boldsymbol{W}$ and satisfies: $\|\boldsymbol{u}\|^2/N \xrightarrow{\mathbb{P}} r$ as $N \to \infty$. Then, for any continuous function $f : \mathbb{R} \mapsto \mathbb{R}$,

$$\frac{\boldsymbol{u}^\top f(\boldsymbol{W})\boldsymbol{u}}{N} \xrightarrow{\mathbb{P}} r \cdot \mathbb{E}[f(\mathsf{A})] \quad \mathsf{\Lambda} \sim \mu,$$

where μ is the limiting spectral distribution of $\boldsymbol{W}$.

Proof. Recall from Assumption 2 that the eigen-decomposition of $\boldsymbol{W}$ is given by:

$$\boldsymbol{W} = \boldsymbol{U} \cdot \mathrm{diag}(\lambda_1(\boldsymbol{W}), \ldots, \lambda_N(\boldsymbol{W})) \cdot \boldsymbol{U}^\top,$$

where the matrix of eigenvectors $\boldsymbol{U} \sim \mathsf{Unif}(\mathbb{O}(N))$ is a Haar-distributed random orthogonal matrix independent of the eigenvalues $\lambda_1(\boldsymbol{W}), \ldots, \lambda_N(\boldsymbol{W})$. Defining $\boldsymbol{v} \stackrel{\text{def}}{=} \boldsymbol{U}^\mathsf{T}\boldsymbol{u}/\|\boldsymbol{u}\| \sim \mathrm{Unif}(\mathbb{S}^{N-1})$, we find that:

$$\frac{\boldsymbol{u}^\top f(\boldsymbol{W})\boldsymbol{u}}{N} = \frac{\|\boldsymbol{u}\|^2}{N} \cdot \sum_{i=1}^{N} v_i^2 \cdot f(\lambda_i(\boldsymbol{W})).$$

Since $\boldsymbol{v} \sim \mathrm{Unif}(\mathbb{S}^{N-1})$, by [12, Proposition 9.3] we have that:

$$\sum_{i=1}^{N} v_i^2 \cdot f(\lambda_i(\boldsymbol{W})) \xrightarrow{\mathbb{P}} \mathbb{E}[f(\mathsf{\Lambda})], \quad \mathsf{\Lambda} \sim \mu.$$

The claim follows by combining the above conclusion with the hypothesis $\|\boldsymbol{u}\|^2/N \xrightarrow{\mathbb{P}} r$ using Slutsky's theorem. □