A Low-rank perturbations of rotationally invariant matrices

In this appendix, we recall some known results concerning low-rank perturbations of rotationally invariant matrices [21, 22]. We present them in a form which is more convenient for our discussion, and we specialize them for rank-1 perturbations. The notations are the same as in Section 3.

The first result characterizes the asymptotic value of the largest eigenvalue of the perturbed symmetric matrix $\boldsymbol{Y} \in \mathbb{R}^{N\times N}$.

Theorem 3 (Theorem 2.1 of [21]). *Under Assumption 1, as $N \to +\infty$, the largest eigenvalue $\overline{\nu}_N$ of $\boldsymbol{Y}$ converges almost surely to*

$$\overline{\nu} := K_\rho\Big(\frac{1}{\sqrt{\lambda_*}}\Big)\mathbf{1}\big(\overline{h}\sqrt{\lambda_*} \geq 1\big) + \overline{\gamma}\,\mathbf{1}\big(\overline{h}\sqrt{\lambda_*} < 1\big).$$

This theorem thus implies that, below the phase transition, the presence of the perturbation does not modify the largest eigenvalue of the noise matrix. For the inner product between $\boldsymbol{X}$ and the eigenvector of $\boldsymbol{Y}$ corresponding to the largest eigenvalue, the following result holds.

Theorem 4 (Theorem 2.2 of [21]). *Under Assumption 1, as $N \to +\infty$, the eigenvector $\overline{\boldsymbol{v}}_N$ corresponding to the largest eigenvalue $\overline{\nu}_N$ of $\boldsymbol{Y}$ is such that, almost surely,*

$$C(\rho, \lambda_*) := \lim_{N\to+\infty} \frac{\langle \boldsymbol{X}, \overline{\boldsymbol{v}}_N\rangle^2}{N} = \Big(1 - \frac{1}{\lambda_*}R'_\rho\Big(\frac{1}{\sqrt{\lambda_*}}\Big)\Big)\mathbf{1}\big(\overline{h}\sqrt{\lambda_*} \geq 1\big).$$

Below the phase transition the spectral estimators are thus asymptotically uncorrelated with the signal while above the transition the leading eigenvector starts to align in its direction.

B Proofs for the Bayes estimator

It is useful to define a slight generalization of the model (1) that includes a "Wigner regularization" of the rotationally-invariant noise. Namely, in this section, we consider the following generalized model for the data:

$$\boldsymbol{Y}_\epsilon = \frac{\sqrt{\lambda_*}}{N}\boldsymbol{X}\boldsymbol{X}^\top + \boldsymbol{Z} + \sqrt{\epsilon}\,\boldsymbol{W}, \tag{19}$$

where, as before, $\boldsymbol{W}$ is a standard Wigner matrix and $\epsilon > 0$ is some constant. We denote by $\langle\cdot\rangle_\epsilon$ the mean w.r.t. the associated mismatched posterior obtained by replacing $\boldsymbol{Y}$ with $\boldsymbol{Y}_\epsilon$ in equation (2):

$$\langle g(\boldsymbol{x})\rangle_\epsilon := \frac{1}{Z_N(\boldsymbol{Y}_\epsilon)}\int g(\boldsymbol{x})\exp\Big(\frac{\sqrt{\lambda}}{2}\langle \boldsymbol{x}, \boldsymbol{Y}_\epsilon \boldsymbol{x}\rangle\Big)\mu_N(d\boldsymbol{x}). \tag{20}$$

At $\epsilon = 0$, the operator $\langle\cdot\rangle_0$ corresponds to the expectation w.r.t. to the original mismatched measure (2). The other notations are analogous to those for the non-regularized model, though we will add a subscript ϵ when it is useful to remark the dependence of a given quantity over ϵ. For example, we denote ρ_ϵ the weak limit of the empirical spectral density of $\boldsymbol{Y}_\epsilon$, R_ϵ its associated R-transform, $\overline{\nu}_\epsilon$ the limit of the maximum eigenvalue of $\boldsymbol{Y}_\epsilon$, $\overline{\gamma}_\epsilon$ the limit of the maximum eigenvalue of the noise matrix $\boldsymbol{Z} + \sqrt{\epsilon}\,\boldsymbol{W}$ and $\overline{h}_\epsilon := \lim_{z\downarrow\overline{\gamma}_\epsilon} H_{\rho_\epsilon}(z)$.

Our first intermediate result, Proposition 1, characterizes the log-spherical integral (which plays the role of log-moment generating function) for the regularized data matrix:

$$I_N\Big(\frac{\sqrt{\lambda}}{2}, \boldsymbol{Y}_\epsilon\Big) := \int \exp\Big(\frac{\sqrt{\lambda}}{2}\langle \boldsymbol{x}, \boldsymbol{Y}_\epsilon \boldsymbol{x}\rangle\Big)\mu_N(d\boldsymbol{x}). \tag{21}$$

Although this proposition may be interesting in itself, its main purpose is to allow us to derive the MSE of the mismatched Bayes estimator in terms of its derivatives w.r.t. ϵ and λ_*. This is in fact the reason for introducing the regularization in the first place.

Like in the analysis of mismatched inference with Gaussian noise of [71], we will make use of the following result for the asymptotic value of the rank-1 spherical integral.

Theorem 5 (Theorem 6 of [45]). *Define $\overline{h}_\epsilon := \lim_{z\downarrow\overline{\nu}_\epsilon} H_{\rho_\epsilon}(z)$. Assume that Assumption 1 holds. Recall that $\overline{\nu}_\epsilon$ is the limit of the maximum eigenvalue of $\boldsymbol{Y}_\epsilon$ and must be finite. Then,*

$$\lim_{N\to+\infty}\frac{1}{N}\ln I_N\Big(\frac{\sqrt{\lambda}}{2}, \boldsymbol{Y}_\epsilon\Big) = \frac{1}{2}\int_0^{\sqrt{\lambda}} R_{\rho_\epsilon}(t)dt\,\mathbf{1}\big(\sqrt{\lambda} \leq \overline{h}_\epsilon\big) + g_{\lambda,\epsilon}(\overline{\nu}_\epsilon)\,\mathbf{1}\big(\sqrt{\lambda} > \overline{h}_\epsilon\big).$$

We can now state our first intermediate proposition.

Proposition 1 (Spherical integral)**.** *For every $\lambda, \epsilon > 0$, define the function $g_{\lambda,\epsilon} : (\overline{\gamma}_\epsilon, +\infty) \mapsto \mathbb{R}$ as*

$$g_{\lambda,\epsilon}(x) \coloneqq \frac{\sqrt{\lambda}}{2}x - \frac{1}{2}\int d\rho_\epsilon(\gamma)\ln(x-\gamma) - \frac{1}{2} - \frac{1}{4}\ln\lambda.$$

Then, under Assumption 1, for every $\lambda, \lambda_* > 0$ we almost surely have that

$$\lim_{N\to+\infty}\frac{1}{N}\ln I_N\Big(\frac{\sqrt{\lambda}}{2}, \boldsymbol{Y}_\epsilon\Big) = f_\epsilon(\lambda,\lambda_*) \coloneqq \begin{cases} g_{\lambda,\epsilon}(K_{\rho_\epsilon}(\frac{1}{\sqrt{\lambda_*}})) & \text{if } \overline{h}_\epsilon\sqrt{\lambda_*} \geq 1 \text{ and } \lambda\lambda_* > 1, \\ g_{\lambda,\epsilon}(\overline{\gamma}_\epsilon) & \text{if } \overline{h}_\epsilon\sqrt{\lambda_*} < 1 \text{ and } \sqrt{\lambda} > \overline{h}_\epsilon, \\ \frac{1}{2}\int_0^{\sqrt{\lambda}} R_{\rho_\epsilon}(t)dt & \text{otherwise}. \end{cases}$$

Moreover we also have that

$$\lim_{N\to+\infty}\frac{1}{N}\mathbb{E}\ln I_N\Big(\frac{\sqrt{\lambda}}{2}, \boldsymbol{Y}_\epsilon\Big) = f_\epsilon(\lambda,\lambda_*).$$

Proof of Proposition 1. By Theorem 3 applied to the generalized model (19), we are almost surely under the hypothesis of Theorem 5. Combining these two results proves the first claim.

To get the result in expectation, it suffices to notice that by Assumption 1 the sequence of values of the largest eigenvalue of $\boldsymbol{Z}$ is a.s. convergent. The same is a.s. true for the Wigner matrix $\boldsymbol{W}$. There thus almost surely exists some bounded $C > 0$ such that

$$\frac{\sqrt{\lambda}}{2}\langle \boldsymbol{x}, \boldsymbol{Y}_\epsilon \boldsymbol{x}\rangle \leq \frac{\sqrt{\lambda}}{2}(C + \sqrt{\lambda_*})N.$$

Then, the last result follows by dominated convergence. □

B.1 Proof of Theorem 1

To establish this result, define the (matrix) *magnetization* and the (matrix) *overlap* according to

$$M_N \coloneqq \Big(\frac{\langle \boldsymbol{X}, \boldsymbol{x}\rangle}{N}\Big)^2, \qquad Q_N \coloneqq \Big(\frac{\langle \boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}\rangle}{N}\Big)^2.$$

respectively. Here, the supra-indices indicate two conditionally (on $\boldsymbol{Y}_\epsilon$) independent samples from the mismatched posterior $P_{\rm mis}(\cdot \mid \boldsymbol{Y}_\epsilon)$ of the generalized model (19) (with mean (20)). Theorem 1 is the consequence of the asymptotic formulas for the mean magnetization and overlap given in the lemmas below.

Lemma 1. *In the setting considered, for all $\lambda, \lambda_* > 0$ fixed and letting $M(\cdot,\cdot)$ be given by* (9)*, we almost surely have*

$$\lim_{N\to+\infty}\langle M_N\rangle_0 = M(\lambda,\lambda_*).$$

Proof. For this proof we can set $\epsilon = 0$ all along. Note that, in the exponent of the spherical integral (21), the only term in which λ_* appears (when writing $\boldsymbol{Y}_0$ explicitly in terms of $\boldsymbol{X}\boldsymbol{X}^\intercal$) is

$$\frac{\sqrt{\lambda\lambda_*}}{2N}\langle \boldsymbol{X}, \boldsymbol{x}\rangle^2 = N\frac{\sqrt{\lambda\lambda_*}}{2}M_N.$$

Then, we have that for all $N \geq 1$

$$\frac{d}{d\sqrt{\lambda_*}}\frac{1}{N}\ln I_N\Big(\frac{\sqrt{\lambda}}{2}, \boldsymbol{Y}_0\Big) = \frac{\sqrt{\lambda}}{2}\langle M_N\rangle_0.$$

We also have

$$\frac{d^2}{(d\sqrt{\lambda_*})^2}\frac{1}{N}\ln I_N\Big(\frac{\sqrt{\lambda}}{2}, \boldsymbol{Y}_0\Big) = \frac{\lambda N}{4}\Big(\big\langle M_N^2\big\rangle_0 - \langle M_N\rangle_0^2\Big) \geq 0.$$

Thus, by the convexity of $\ln I_N(\frac{\sqrt{\lambda}}{2}, \boldsymbol{Y}_0)$ w.r.t. $\sqrt{\lambda_*}$ and Proposition 1, we get that almost surely

$$\lim_{N\to+\infty}\langle M_N\rangle_0 = \frac{2}{\sqrt{\lambda}}\frac{df_0(\lambda,\lambda_*)}{d\sqrt{\lambda_*}} = 4\sqrt{\frac{\lambda_*}{\lambda}}\frac{df_0(\lambda,\lambda_*)}{d\lambda_*}.$$

Recall that $K_\rho(x) = R_\rho(x) + 1/x$. Then, the derivative of f_0 with respect to λ_* is given by

$$\frac{df_0}{d\lambda_*}(\lambda,\lambda_*) = \frac{1}{4}\Big(\frac{1}{\lambda_*^2} - \Big(\frac{\lambda}{\lambda_*^3}\Big)^{1/2}\Big)\Big(R'_\rho\Big(\frac{1}{\sqrt{\lambda_*}}\Big) - \lambda_*\Big)\mathbf{1}\big(\bar{h}\sqrt{\lambda_*} \geq 1 \cap \lambda\lambda_* > 1\big)$$

with $\bar{h} = \bar{h}_0$. From these two last equations, the result of the lemma follows. □

Lemma 2. *In the setting considered, for all $\lambda, \lambda_* > 0$ fixed and letting $Q(\cdot,\cdot)$ be given by* (9)*, we have*

$$\lim_{N\to+\infty} \mathbb{E}\left\langle Q_N \right\rangle_0 = Q(\lambda,\lambda_*).$$

Proof. Note that in the exponent of the spherical integral (21) the only term in which ϵ appears (when writing $\boldsymbol{Y}_\epsilon$ explicitly in terms of $\boldsymbol{W}$) is $\sqrt{\lambda\epsilon}\,\boldsymbol{x}^\top \boldsymbol{W}\boldsymbol{x}/2$. Then, we have that for every $N \geq 1$

$$\frac{d}{d\sqrt{\epsilon}}\frac{1}{N}\mathbb{E}\ln I_N\Big(\frac{\sqrt{\lambda}}{2}, \boldsymbol{Y}_\epsilon\Big)\Big|_{\epsilon=0} = \frac{\sqrt{\lambda}}{2N}\mathbb{E}\left\langle \boldsymbol{x}^\top \boldsymbol{W}\boldsymbol{x}\right\rangle_0. \tag{22}$$

Furthermore,

$$\frac{d^2}{(d\sqrt{\epsilon})^2}\mathbb{E}\ln I_N\Big(\frac{\sqrt{\lambda}}{2}, \boldsymbol{Y}_\epsilon\Big) = \frac{\lambda}{4}\mathbb{E}\Big(\left\langle (\boldsymbol{x}^\top \boldsymbol{W}\boldsymbol{x})^2\right\rangle_\epsilon - \left\langle \boldsymbol{x}^\top \boldsymbol{W}\boldsymbol{x}\right\rangle_\epsilon^2\Big) \geq 0. \tag{23}$$

Since $\boldsymbol{W}$ is Wigner, its elements are distributed as independent Gaussian random variables (up to symmetry) with variance $1/N$ for the off-diagonal elements and $2/N$ on the diagonal. Using Gaussian integration by parts[3] in the r.h.s. of (22), we get that

$$\frac{d}{d\sqrt{\epsilon}}\frac{1}{N}\mathbb{E}\ln I_N\Big(\frac{\sqrt{\lambda}}{2}, \boldsymbol{Y}_\epsilon\Big) = \frac{\lambda\sqrt{\epsilon}}{2}(1 - \mathbb{E}\left\langle Q_N\right\rangle_\epsilon), \tag{24}$$

and thus

$$\frac{d}{d\epsilon}\frac{1}{N}\mathbb{E}\ln I_N\Big(\frac{\sqrt{\lambda}}{2}, \boldsymbol{Y}_\epsilon\Big)\Big|_{\epsilon=0} = \frac{\lambda}{4}(1 - \mathbb{E}\left\langle Q_N\right\rangle_0). \tag{25}$$

Hence, by the convexity of $\mathbb{E}\ln I_N(\frac{\sqrt{\lambda}}{2}, \boldsymbol{Y}_\epsilon)$ w.r.t. ϵ and Proposition 1, we obtain that

$$\lim_{N\to+\infty}\mathbb{E}\left\langle Q_N\right\rangle_0 = 1 - \frac{4}{\lambda}\frac{df_\epsilon(\lambda,\lambda_*)}{d\epsilon}\Big|_{\epsilon=0}. \tag{26}$$

To find the asymptotic value of $\mathbb{E}\left\langle Q_N\right\rangle_0$, we are thus left to compute $d_\epsilon f_\epsilon(\lambda,\lambda_*)$. We will then need to consider the three regimes of Proposition 1.

Firstly, if $\bar{h}\sqrt{\lambda_*} \geq 1$ and $\lambda\lambda_* > 1$, then

$$\begin{aligned}
\frac{df_\epsilon(\lambda,\lambda_*)}{d\epsilon}\Big|_{\epsilon=0} &= \frac{d}{d\epsilon}g_{\lambda,\epsilon}(K_{\rho_\epsilon}(\lambda_*^{-1/2}))\Big|_{\epsilon=0}\\
&= \frac{\partial g_{\lambda,\epsilon}(x)}{\partial\epsilon}\Big|_{\epsilon=0, x=K_{\rho_0}(\lambda_*^{-1/2})} + \frac{dg_{\lambda,\epsilon}(x)}{dx}\Big|_{\epsilon=0,x=K_{\rho_0}(\lambda_*^{-1/2})}\frac{dK_{\rho_\epsilon}(\lambda_*^{-1/2})}{d\epsilon}\Big|_{\epsilon=0}.
\end{aligned}$$

For computing $\partial_\epsilon g_{\lambda,\epsilon}(x)$, note that the only term of $g_{\lambda,\epsilon}$ depending on ϵ is

$$\ell(x,\epsilon) := -\frac{1}{2}\int d\rho_\epsilon(\gamma)\ln(x-\gamma),$$

for which we have $d_x\ell(x,\epsilon) = -\frac{1}{2}H_{\rho_\epsilon}(x)$. Thus, for any appropriate x_0,

$$\ell(x,\epsilon) = -\frac{1}{2}\int_{x_0}^{x} H_{\rho_\epsilon}(y)dy + C(x_0,\epsilon),$$

where $C(x_0,\epsilon)$ is a constant depending only on x_0 and ϵ. Furthermore, by the Dyson Brownian motion characterization of the eigenvalues of the matrix $\boldsymbol{Z} + \sqrt{\epsilon}\,\boldsymbol{W}$, we have that the limiting Stieltjes transform $H_{\rho_\epsilon}(x)$ is a solution of Burger's equation (see for example [4, Proposition 4.3.10]); that is,

$$\frac{dH_{\rho_\epsilon}(x)}{d\epsilon} = -H_{\rho_\epsilon}(x)\frac{dH_{\rho_\epsilon}(x)}{dx} = -\frac{1}{2}\frac{dH^2_{\rho_\epsilon}(x)}{dx}. \tag{27}$$

[3] The used formula here is $\mathbb{E}Zf(Z) = \sigma^2\mathbb{E}f'(Z)$ for $Z \sim \mathcal{N}(0,\sigma^2)$.

We then have

$$\frac{d\ell(x,\epsilon)}{d\epsilon} = -\frac{1}{2}\int_{x_0}^{x}\frac{dH_{\rho_\epsilon}(y)}{d\epsilon}dy + \frac{dC(x_0,\epsilon)}{d\epsilon} = \frac{1}{4}H^2_{\rho_\epsilon}(x) - \frac{1}{4}H^2_{\rho_\epsilon}(x_0) + \frac{dC(x_0,\epsilon)}{d\epsilon}. \tag{28}$$

As the left hand-side is independent of x_0, $-\frac{1}{4}H^2_{\rho_\epsilon}(x_0) + d_\epsilon C(x_0,\epsilon)$ is some function $c(\epsilon)$ of ϵ alone. This gives that

$$\frac{\partial g_{\lambda,\epsilon}(x)}{\partial\epsilon}\Big|_{\epsilon=0,x=K_{\rho_0}(\lambda_*^{-1/2})} = \frac{1}{4}H^2_{\rho_0}(K_{\rho_0}(\lambda_*^{-1/2})) + c(0) = \frac{1}{4\lambda_*} + c(0). \tag{29}$$

Furthermore,

$$\frac{dg_{\lambda,\epsilon}(x)}{dx} = \frac{\sqrt{\lambda}}{2} - \frac{1}{2}H_{\rho_\epsilon}(x), \tag{30}$$

and $d_\epsilon K_{\rho_\epsilon}(x) = x$ (which follows from the additivity of the R-transform and the fact that the R-transform associated with the asymptotic spectral density of the Wigner part $\sqrt{\epsilon}\boldsymbol{W}$ is ϵx). Then,

$$\frac{dg_{\lambda,\epsilon}(x)}{dx}\Big|_{\epsilon=0,x=K_{\rho_0}(\lambda_*^{-1/2})}\frac{dK_{\rho_\epsilon}(\lambda_*^{-1/2})}{d\epsilon}\Big|_{\epsilon=0} = \frac{1}{2}\Big(\sqrt{\frac{\lambda}{\lambda_*}} - \frac{1}{\lambda_*}\Big). \tag{31}$$

Combining equations (26), (29), and (31) we get that, if $\overline{h}\sqrt{\lambda_*} \geq 1$ and $\lambda\lambda_* > 1$, then

$$\lim_{N\to+\infty}\mathbb{E}\langle Q_N\rangle_0 = 1 - \frac{2}{\sqrt{\lambda\lambda_*}} + \frac{1}{\lambda\lambda_*} + c(0) = \Big(1 - \frac{1}{\sqrt{\lambda\lambda_*}}\Big)^2 + c(0).$$

We now show that $c(0) = 0$. First note that $c(0)$ does not depend on λ nor λ_*. Assume that $c(0)$ is positive and fix λ_* s.t. $\overline{h}\sqrt{\lambda_*} > 1$. Choose λ large enough so that we simultaneously have $\lambda\lambda_* > 1$ and $(1 - 1/\sqrt{\lambda\lambda_*})^2 + c(0) > 1$. This can be done because $c(0) > 0$. For these values of λ and λ_*, we thus have that $\lim_{N\to+\infty}\mathbb{E}\langle Q_N\rangle_0 > 1$. However, this is a contradiction, since for every $\lambda,\lambda_* > 0$ and $N \geq 1$ a.s. $0 \leq Q_N \leq 1$. We therefore have that $c(0)$ is non-positive. In a similar way, we can prove that $c(0)$ is non-negative. Then, we conclude that, if $\overline{h}\sqrt{\lambda_*} \geq 1$ and $\lambda\lambda_* > 1$,

$$\lim_{N\to+\infty}\mathbb{E}\langle Q_N\rangle_0 = \Big(1 - \frac{1}{\sqrt{\lambda\lambda_*}}\Big)^2.$$

Secondly, we will now establish the limit of the overlap when $\overline{h}\sqrt{\lambda_*} < 1$ and $\sqrt{\lambda} > \overline{h}$. In this region,

$$\frac{df_\epsilon(\lambda,\lambda_*)}{d\epsilon}\Big|_{\epsilon=0} = \frac{d}{d\epsilon}g_{\lambda,\epsilon}(\overline{\gamma}_\epsilon)\Big|_{\epsilon=0} = \frac{\partial g_{\lambda,\epsilon}(x)}{\partial\epsilon}\Big|_{\epsilon=0,x=\overline{\gamma}_0} + \frac{\partial g_{\lambda,\epsilon}(x)}{\partial x}\Big|_{\epsilon=0,x=\overline{\gamma}_0}\frac{\partial\overline{\gamma}_\epsilon}{\partial\epsilon}\Big|_{\epsilon=0}. \tag{32}$$

The partial derivatives $\partial_\epsilon g_{\lambda,\epsilon}(x)$ and $\partial_x g_{\lambda,\epsilon}(x)$ are obtained as before. Therefore, we only need to compute $\partial_\epsilon\overline{\gamma}|_{\epsilon=0}$. We will prove that as $\epsilon\to 0$, $\overline{\gamma}_\epsilon = \overline{\gamma}_0 + \overline{h}_0\epsilon + o(\epsilon)$. As a consequence of the large deviation principle of [44, Theorem 2.5], $\overline{\gamma}_\epsilon$ is equal to the supremum of the compact support K_ϵ of ρ_ϵ. By Assumption 2, for all $\epsilon > 0$ we have $\lim_{z\downarrow\overline{\gamma}_\epsilon}H'_{\rho_\epsilon}(z) = -\infty$. This implies that $\lim_{x\downarrow\overline{h}_\epsilon}K'_{\rho_\epsilon}(x) = 0$. Thus, $\overline{h}_\epsilon$ is a solution to the equation $K'_{\rho_0}(x) + \epsilon = 0$ (again, this can be seen by additivity of the R-transform and its link to K_{ρ_ϵ}). As K'_{ρ_0} is analytic and monotonously increasing, the solution is unique and depends smoothly on ϵ. Thus, $\overline{h}_\epsilon = \overline{h}_0 + \mathcal{O}(\epsilon)$, which means that

$$\overline{\gamma}_\epsilon = \lim_{x\downarrow\overline{h}_\epsilon}K_{\rho_\epsilon}(x) = \lim_{x\downarrow\overline{h}_0}K_{\rho_0}(x) + \epsilon\overline{h}_0 + o(\epsilon) = \overline{\gamma}_0 + \epsilon\overline{h}_0 + o(\epsilon),$$

where we used that $\lim_{x\downarrow\overline{h}_0}K'_{\rho_0}(x) = 0$. This proves that

$$\partial_\epsilon\overline{\gamma}_\epsilon|_{\epsilon=0} = \overline{h}_0.$$

Combining this with equations (26), (28), (30), and (32) gives the result in this region.

Finally, by Proposition 1 and (26), if (λ,λ_*) does not verify either of the conditions considered above, then by dominated convergence

$$\lim_{N\to+\infty}\mathbb{E}\langle Q_N\rangle_0 = 1 - \frac{2}{\lambda}\int_0^{\sqrt{\lambda}}\partial_\epsilon R_{\rho_\epsilon}(x)dx = 1 - \frac{2}{\lambda}\int_0^{\sqrt{\lambda}}x\,dx = 0.$$

This finishes the proof that $\lim_{N\to+\infty}\mathbb{E}\langle Q_N\rangle_0 = Q(\lambda,\lambda_*)$. □

Proof of Theorem 1. In order to establish the formula for the MSE, it suffices to use Lemmas 1 and 2 combined with the observation that

$$\frac{1}{2N^2}\mathbb{E}\|\langle\boldsymbol{x}\boldsymbol{x}^\top\rangle_0 - \boldsymbol{X}\boldsymbol{X}^\top\|_{\mathrm{F}}^2 = \frac{1}{2}\big(1 + \mathbb{E}\langle Q_N\rangle_0 - 2\mathbb{E}\langle M_N\rangle_0\big).$$

□

C Proofs for Approximate Message Passing

C.1 Auxiliary AMP

The iterates of the auxiliary AMP are denoted by $\tilde{\boldsymbol{z}}^t, \tilde{\boldsymbol{x}}^t \in \mathbb{R}^N$, and they are computed as follows, for $t \geq 1$:

$$\tilde{\boldsymbol{z}}^t = \boldsymbol{Z}\tilde{\boldsymbol{x}}^t - \sum_{i=1}^{t} \mathsf{b}_{t,i}\tilde{\boldsymbol{x}}^i, \quad \tilde{\boldsymbol{x}}^{t+1} = \tilde{h}_{t+1}(\tilde{\boldsymbol{z}}^1, \dots, \tilde{\boldsymbol{z}}^t, \hat{\boldsymbol{x}}^1, \boldsymbol{X}). \tag{33}$$

The iteration (33) is initialized with $\tilde{\boldsymbol{x}}^1 = \hat{\boldsymbol{x}}^1$, where $\hat{\boldsymbol{x}}^1$ is also the initialization of the true AMP (see (7)). For $t \geq 1$, the functions $\tilde{h}_{t+1} : \mathbb{R}^{t+2} \to \mathbb{R}$ are applied component-wise, and they are recursively defined as

$$\begin{aligned}\tilde{h}_{t+1}(z_1, \dots, z_t, \hat{x}_1, x) = h_{t+1}\Bigg(z_t + (\boldsymbol{B}_t)_{t,1}\hat{x}_1 + \sum_{i=2}^{t}(\boldsymbol{B}_t)_{t,i}\tilde{h}_i\left(z_1, \dots, z_{i-1}, \hat{x}_1, x\right) + \mu_t x \\ - \bar{\beta}_t \tilde{h}_{t-1}\left(z_1, \dots, z_{t-2}, \hat{x}_1, x\right)\Bigg).\end{aligned} \tag{34}$$

The idea is that the choice (34) for the denoisers $\{\tilde{h}_{t+1}\}_{t\geq 1}$ allows to "correct" for the mismatch and compensate for the wrong Onsager corrections in (8). Here, h_{t+1} is the denoiser of the true AMP (see (8)), and $(\mu_t, \bar{\beta}_t, \boldsymbol{B}_t)$ come from the state evolution recursion of the true AMP: μ_1 is given by (13), and, for $t \geq 2$, μ_t is given by (15); $\bar{\beta}_1 = 0$ and, for $t \geq 2$, $\bar{\beta}_t = \mathbb{E}[h_t'(x_{t-1})]$, where the law of x_{t-1} is given by (14); and $\boldsymbol{B}_t$ is defined via (16). We now discuss how to obtain the coefficients $\{\mathsf{b}_{t,i}\}_{i=1}^t$. Let us define the matrix $\boldsymbol{\Phi}_t \in \mathbb{R}^{t\times t}$ as

$$(\boldsymbol{\Phi}_t)_{i,j} = 0, \quad \text{for } i \leq j, \qquad (\boldsymbol{\Phi}_t)_{i,j} = \langle \partial_j \tilde{\boldsymbol{x}}^i \rangle, \quad \text{for } i > j, \tag{35}$$

where, for $j < i$, the vector $\langle \partial_j \tilde{\boldsymbol{x}}^i \rangle \in \mathbb{R}^N$ denotes the partial derivative of $\tilde{h}_i : \mathbb{R}^{i+1} \to \mathbb{R}$ with respect to the j-th input (applied component-wise). Then, the vector $(\mathsf{b}_{t,1}, \dots, \mathsf{b}_{t,t})$ is given by the last row of the matrix $\tilde{\boldsymbol{B}}_t \in \mathbb{R}^{t\times t}$ defined as

$$\tilde{\boldsymbol{B}}_t = \sum_{j=0}^{t-1} \kappa_{j+1}\boldsymbol{\Phi}_t^j, \tag{36}$$

where $\{\kappa_k\}_{k\geq 1}$ denotes the sequence of free cumulants associated to the matrix $\boldsymbol{Z}$.

C.2 State evolution of auxiliary AMP

Using Theorem 2.3 in [87], we provide a state evolution result for the auxiliary AMP (33). In particular, we show in Proposition 2 that the joint empirical distribution of $(\tilde{\boldsymbol{z}}^1, \dots, \tilde{\boldsymbol{z}}^t)$ converges to a t-dimensional Gaussian $\mathcal{N}(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}}_t)$.

The covariance matrices $\{\tilde{\boldsymbol{\Sigma}}_t\}_{t\geq 1}$ are defined recursively, starting with $\tilde{\boldsymbol{\Sigma}}_1 = \bar{\kappa}_2\mathbb{E}[\hat{x}_1^2]$, where $\hat{x}_1$ is defined in (7). Given $\tilde{\boldsymbol{\Sigma}}_t$, let

$$\begin{aligned}&(\tilde{z}_1, \dots, \tilde{z}_t) \sim \mathcal{N}(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}}_t) \text{ and independent of } (X, \hat{x}_1), \\ &\tilde{x}_s = \tilde{h}_s\left(\tilde{z}_1, \dots, \tilde{z}_{s-1}, \hat{x}_1, X\right), \quad \text{for } s \in \{2, \dots, t+1\},\end{aligned} \tag{37}$$

where $\tilde{h}_s$ is defined via (34) and we set $\tilde{x}_1 = \hat{x}_1$. Let $\tilde{\boldsymbol{\Phi}}_{t+1}, \tilde{\boldsymbol{\Delta}}_{t+1} \in \mathbb{R}^{(t+1)\times(t+1)}$ be matrices with entries given by

$$\begin{aligned}&(\tilde{\boldsymbol{\Phi}}_{t+1})_{i,j} = 0, \quad \text{for } i \leq j, \qquad (\tilde{\boldsymbol{\Phi}}_{t+1})_{i,j} = \mathbb{E}[\partial_j \tilde{x}_i], \quad \text{for } i > j, \\ &(\tilde{\boldsymbol{\Delta}}_{t+1})_{i,j} = \mathbb{E}[\tilde{x}_i\, \tilde{x}_j], \quad 1 \leq i, j \leq t+1,\end{aligned} \tag{38}$$

where $\partial_j \tilde{x}_i$ denotes the partial derivative $\partial_{z_j}\tilde{h}_i(\tilde{z}_1, \dots, \tilde{z}_{i-1}, \hat{x}_1, X)$. Then, we compute the covariance matrix $\tilde{\boldsymbol{\Sigma}}_{t+1}$ as

$$\tilde{\boldsymbol{\Sigma}}_{t+1} = \sum_{j=0}^{2t} \bar{\kappa}_{j+2}\tilde{\boldsymbol{\Theta}}_{t+1}^{(j)}, \qquad \text{with} \quad \tilde{\boldsymbol{\Theta}}_{t+1}^{(j)} = \sum_{i=0}^{j}(\tilde{\boldsymbol{\Phi}}_{t+1})^i\tilde{\boldsymbol{\Delta}}_{t+1}(\tilde{\boldsymbol{\Phi}}_{t+1}^\top)^{j-i}. \tag{39}$$

It can be verified that the $t \times t$ top left sub-matrix of $\tilde{\boldsymbol{\Sigma}}_{t+1}$ is given by $\tilde{\boldsymbol{\Sigma}}_t$.

Proposition 2 (State evolution for auxiliary AMP)**.** *Consider the auxiliary AMP in* (33) *and the state evolution random variables defined in* (37)*. Let* $\tilde{\psi} : \mathbb{R}^{2t+2} \to \mathbb{R}$ *be any pseudo-Lipschitz functions of order 2. Then for each* $t \geq 1$*, we almost surely have*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \tilde{\psi}\left((\tilde{\boldsymbol{z}}^1)_i, \dots, (\tilde{\boldsymbol{z}}^t)_i, (\tilde{\boldsymbol{x}}^1)_i, \dots, (\tilde{\boldsymbol{x}}^{t+1})_i, (\boldsymbol{X})_i\right) = \mathbb{E}[\tilde{\psi}(\tilde{z}_1, \dots \tilde{z}_t, \tilde{x}_1, \dots, \tilde{x}_{t+1}, X)]. \tag{40}$$

Equivalently, as $N \to \infty$*, almost surely:*

$$(\tilde{\boldsymbol{z}}^1, \dots, \tilde{\boldsymbol{z}}^t, \tilde{\boldsymbol{x}}^1, \dots, \tilde{\boldsymbol{x}}^{t+1}, \boldsymbol{X}) \xrightarrow{W_2} (\tilde{z}_1, \dots, \tilde{z}_t, \tilde{x}_1, \dots, \tilde{x}_{t+1}, X).$$

Proof. The result follows from Theorem 2.3 in [87]. In fact, Assumption 2.1 of [87] holds because of the model assumptions on $\boldsymbol{Z}$, Assumption 2.2(a) holds because $(\boldsymbol{X}, \hat{\boldsymbol{x}}^1) \xrightarrow{W_2} (X, \hat{x}_1)$ from (7), and Assumption 2.2(b) follows from the definition of $\tilde{h}_{t+1}$ in (34) and our Assumption 3. As the auxiliary AMP in (33) is of the standard form for which the state evolution result of Theorem 2.3 in [87] holds, the proof is complete. □

C.3 Proof of Theorem 2

We start by presenting a useful technical lemma.

Lemma 3. *Let* $F : \mathbb{R}^t \to \mathbb{R}$ *be a Lipschitz function, and let* $\partial_k F$ *denote its derivative with respect to the* k*-th argument, for* $1 \leq k \leq t$*. Assume that* $\partial_k F$ *is continuous almost everywhere in the* k*-th argument, for each* k*. Let* $(V_1^{(m)}, \dots, V_t^{(m)})$ *be a sequence of random vectors in* $\mathbb{R}^t$ *converging in distribution to the random vector* $(V_1, \dots, V_t)$ *as* $m \to \infty$*. Furthermore, assume that the distribution of* $(V_1, \dots, V_t)$ *is absolutely continuous with respect to the Lebesgue measure. Then,*

$$\lim_{m\to\infty} \mathbb{E}[\partial_k F(V_1^{(m)}, \dots, V_t^{(m)})] = \mathbb{E}[\partial_k F(V_1, \dots, V_t)], \qquad 1 \leq k \leq t. \tag{41}$$

The result was proved for $t = 2$ in [18, Lemma 6]. The proof for $t > 2$ is essentially the same; see also [39, Lemma 7.14].

At this point, we are ready to give the proof of Theorem 2.

Proof of Theorem 2. The first step is to show the equivalence between the state evolution for the true AMP and the corresponding one for the auxiliary AMP. In particular, we prove that, for $t \geq 1$,

$$(\tilde{z}_1, \dots, \tilde{z}_t) \overset{\mathrm{d}}{=} (z_1, \dots, z_t), \tag{42}$$

where the random variables on the left are defined in (37), and the random variables on the right are defined in (14). As $(\tilde{z}_1, \dots, \tilde{z}_t) \sim \mathcal{N}(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}}_t)$ and $(z_1, \dots, z_t) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_t)$, (42) is readily implied by

$$\tilde{\boldsymbol{\Sigma}}_t = \boldsymbol{\Sigma}_t, \qquad \text{for all } t \geq 1. \tag{43}$$

We now show that (43) holds by induction. The base case ($t = 1$) follows from the definitions of $\tilde{\boldsymbol{\Sigma}}_1, \boldsymbol{\Sigma}_1$. Towards induction, assume that $\tilde{\boldsymbol{\Sigma}}_t = \boldsymbol{\Sigma}_t$ for some $t \geq 1$. By comparing the definition of $\hat{x}_s$ in (14) with the definition of $\tilde{x}_s$ in (37) and by using the choice of $\tilde{h}_s$ in (34) (for $s \in \{2, \dots, t+1\}$), we readily obtain that $\Delta_{t+1} = \tilde{\Delta}_{t+1}$ and $\bar{\boldsymbol{\Phi}}_{t+1} = \tilde{\boldsymbol{\Phi}}_{t+1}$. Hence, by using (17) and (39), we have that $\tilde{\boldsymbol{\Sigma}}_{t+1} = \boldsymbol{\Sigma}_{t+1}$ and the proof of (43) is complete.

The second step is to show that, for any pseudo-Lipschitz function $\psi : \mathbb{R}^{2t+2} \to \mathbb{R}$ of order 2, the following limit holds almost surely for $t \geq 1$:

$$\begin{aligned} \lim_{N\to\infty} \Bigg| \frac{1}{N} \sum_{i=1}^{N} \psi\left((\boldsymbol{x}^1)_i, \dots, (\boldsymbol{x}^t)_i, (\hat{\boldsymbol{x}}^1)_i, \dots, (\hat{\boldsymbol{x}}^{t+1})_i, (\boldsymbol{X})_i\right) \\ - \frac{1}{N} \sum_{i=1}^{N} \psi\left((\boldsymbol{u}^1)_i, \dots, (\boldsymbol{u}^t)_i, (\tilde{\boldsymbol{x}}^1)_i, \dots, (\tilde{\boldsymbol{x}}^{t+1})_i, (\boldsymbol{X})_i\right) \Bigg| = 0, \end{aligned} \tag{44}$$

where, we have defined for $s \in \{1, \dots, t\}$,

$$\boldsymbol{u}^s = \tilde{\boldsymbol{z}}^s + (\boldsymbol{B}_s)_{s,1}\tilde{\boldsymbol{x}}^1 + \sum_{i=2}^{s}(\boldsymbol{B}_s)_{s,i}\tilde{\boldsymbol{x}}^i + \mu_s \boldsymbol{X} - \bar{\beta}_s \tilde{\boldsymbol{x}}^{s-1}. \tag{45}$$

From here till the end of the argument, all the limits hold almost surely, and we use C to denote a generic positive constant, which can change from line to line and is independent of N. By using that ψ is pseudo-Lipschitz, we have that

$$\begin{aligned}
&\Bigg| \frac{1}{N}\sum_{i=1}^{N}\psi\big((\boldsymbol{x}^1)_i, \dots, (\boldsymbol{x}^t)_i, (\hat{\boldsymbol{x}}^1)_i, \dots, (\hat{\boldsymbol{x}}^{t+1})_i, (\boldsymbol{X})_i\big) \\
&\qquad - \frac{1}{N}\sum_{i=1}^{N}\psi\big((\boldsymbol{u}^1)_i, \dots, (\boldsymbol{u}^t)_i, (\tilde{\boldsymbol{x}}^1)_i, \dots, (\tilde{\boldsymbol{x}}^{t+1})_i, (\boldsymbol{X})_i\big)\Bigg| \\
&\le \frac{C}{N}\sum_{i=1}^{N}\left(1 + |(\boldsymbol{X})_i| + 2|(\hat{\boldsymbol{x}}^1)_i| + \sum_{\ell=1}^{t}\Big(|(\boldsymbol{x}^\ell)_i| + |(\boldsymbol{u}^\ell)_i| + |(\hat{\boldsymbol{x}}^{\ell+1})_i| + |(\tilde{\boldsymbol{x}}^{\ell+1})_i|\Big)\right) \\
&\qquad \cdot \left(\sum_{\ell=1}^{t}\big(|(\boldsymbol{x}^\ell)_i - (\boldsymbol{u}^\ell)_i|^2 + |(\hat{\boldsymbol{x}}^{\ell+1})_i - (\tilde{\boldsymbol{x}}^{\ell+1})_i|^2\big)\right)^{1/2} \\
&\le C(4t+3)\left[1 + \frac{\|\boldsymbol{X}\|^2}{N} + \sum_{\ell=1}^{t}\Big(\frac{\|\boldsymbol{x}^\ell\|^2}{N} + \frac{\|\boldsymbol{u}^\ell\|^2}{N} + \frac{\|\hat{\boldsymbol{x}}^{\ell+1}\|^2}{N} + \frac{\|\tilde{\boldsymbol{x}}^{\ell+1}\|^2}{N}\Big)\right]^{1/2} \\
&\qquad \cdot \left(\sum_{\ell=1}^{t}\left(\frac{\|\boldsymbol{x}^\ell - \boldsymbol{u}^\ell\|^2}{N} + \frac{\|\hat{\boldsymbol{x}}^{\ell+1} - \tilde{\boldsymbol{x}}^{\ell+1}\|^2}{N}\right)\right)^{1/2},
\end{aligned} \tag{46}$$

where the last step uses twice Cauchy-Schwarz inequality. We now inductively show that as $N \to \infty$: *(i)* each of the terms in the last line of (46) converges to zero, and *(ii)* the terms within the square brackets in (46) all converge to finite, deterministic limits.

<u>Base case ($t = 1$).</u> We have that

$$\begin{aligned}
\boldsymbol{x}^1 - \boldsymbol{u}^1 &= \boldsymbol{Y}\hat{\boldsymbol{x}}^1 - \tilde{\boldsymbol{z}}^1 - (\boldsymbol{B}_1)_{1,1}\tilde{\boldsymbol{x}}^1 - \mu_1 \boldsymbol{X} \\
&= \boldsymbol{Z}\hat{\boldsymbol{x}}^1 + \sqrt{\lambda_*}\frac{\langle \boldsymbol{X}, \hat{\boldsymbol{x}}^1\rangle}{N}\boldsymbol{X} - \boldsymbol{Z}\tilde{\boldsymbol{x}}^1 + \mathsf{b}_{1,1}\tilde{\boldsymbol{x}}^1 - (\boldsymbol{B}_1)_{1,1}\tilde{\boldsymbol{x}}^1 - \mu_1\boldsymbol{X} \\
&= \sqrt{\lambda_*}\frac{\langle \boldsymbol{X}, \hat{\boldsymbol{x}}^1\rangle}{N}\boldsymbol{X} + \mathsf{b}_{1,1}\tilde{\boldsymbol{x}}^1 - (\boldsymbol{B}_1)_{1,1}\tilde{\boldsymbol{x}}^1 - \mu_1\boldsymbol{X},
\end{aligned} \tag{47}$$

where the first equality uses (8) and (45), the second equality uses (1) and (33), and the third equality uses that $\tilde{\boldsymbol{x}}^1 = \hat{\boldsymbol{x}}^1$. Hence, by triangle inequality,

$$\begin{aligned}
\frac{\|\boldsymbol{x}^1 - \boldsymbol{u}^1\|^2}{N} &\le 2\left(\sqrt{\lambda_*}\frac{\langle \boldsymbol{X}, \hat{\boldsymbol{x}}^1\rangle}{N} - \mu_1\right)^2 \frac{\|\boldsymbol{X}\|^2}{N} + 2\left(\mathsf{b}_{1,1} - (\boldsymbol{B}_1)_{1,1}\right)^2 \frac{\|\tilde{\boldsymbol{x}}^1\|^2}{N} \\
&\le C\left(\left(\sqrt{\lambda_*}\frac{\langle \boldsymbol{X}, \hat{\boldsymbol{x}}^1\rangle}{N} - \mu_1\right)^2 + \left(\mathsf{b}_{1,1} - (\boldsymbol{B}_1)_{1,1}\right)^2\right),
\end{aligned} \tag{48}$$

where the last inequality uses that $\tilde{\boldsymbol{x}}^1 = \hat{\boldsymbol{x}}^1$ and that $(\boldsymbol{X}, \hat{\boldsymbol{x}}^1)$ converge in W_2 to random variables with finite second moments. By using (7) and recalling that $\mu_1 = \sqrt{\lambda_*}\epsilon$ (cf. (13)), we have

$$\lim_{N\to\infty}\sqrt{\lambda_*}\frac{\langle \boldsymbol{X}, \hat{\boldsymbol{x}}^1\rangle}{N} = \sqrt{\lambda_*}\epsilon = \mu_1. \tag{49}$$

Furthermore, note that $(\boldsymbol{B}_1)_{1,1} = \bar{\kappa}_1$ (cf. (13)) and $\mathsf{b}_{1,1} = \kappa_1$ (cf. (36)). Hence, by Assumption 1, $\kappa_1 \to \bar{\kappa}_1$, as $N \to \infty$. Hence, $\mathsf{b}_{1,1} \to (\boldsymbol{B}_1)_{1,1}$ and, by combining this observation with (48) and (49), we obtain that

$$\lim_{N\to\infty}\frac{\|\boldsymbol{x}^1 - \boldsymbol{u}^1\|^2}{N} = 0. \tag{50}$$

Next, by expressing $\hat{\boldsymbol{x}}^2$ via (8) and $\tilde{\boldsymbol{x}}^2$ via (33), (34) and (45), we have that

$$\hat{\boldsymbol{x}}^2 - \tilde{\boldsymbol{x}}^2 = h_2(\boldsymbol{x}^1) - h_2(\boldsymbol{u}^1). \tag{51}$$

Thus, as h_2 is Lipschitz, (50) immediately implies that

$$\lim_{N\to\infty} \frac{\|\hat{\boldsymbol{x}}^2 - \tilde{\boldsymbol{x}}^2\|^2}{N} = 0. \tag{52}$$

An application of the triangle inequality gives that, for any $i \geq 1$,

$$\begin{aligned} \|\boldsymbol{u}^i\| - \|\boldsymbol{x}^i - \boldsymbol{u}^i\| \leq \|\boldsymbol{x}^i\| \leq \|\boldsymbol{u}^i\| + \|\boldsymbol{x}^i - \boldsymbol{u}^i\|, \\ \|\tilde{\boldsymbol{x}}^{i+1}\| - \|\hat{\boldsymbol{x}}^{i+1} - \tilde{\boldsymbol{x}}^{i+1}\| \leq \|\hat{\boldsymbol{x}}^{i+1}\| \leq \|\tilde{\boldsymbol{x}}^{i+1}\| + \|\hat{\boldsymbol{x}}^{i+1} - \tilde{\boldsymbol{x}}^{i+1}\|. \end{aligned} \tag{53}$$

Thus, by using (53) with $i = 1$ and Proposition 2, we obtain that

$$\begin{aligned} \lim_{N\to\infty} \frac{\|\boldsymbol{x}^1\|^2}{N} = \lim_{N\to\infty} \frac{\|\boldsymbol{u}^1\|^2}{N} = \mathbb{E}[(\tilde{z}_1 + (\boldsymbol{B}_1)_{1,1}\hat{x}_1 + \mu_1 X)^2], \\ \lim_{N\to\infty} \frac{\|\hat{\boldsymbol{x}}^2\|^2}{N} = \lim_{N\to\infty} \frac{\|\tilde{\boldsymbol{x}}^2\|^2}{N} = \mathbb{E}[(\tilde{x}_2)^2], \end{aligned} \tag{54}$$

which concludes the base step.

Induction step. Assume towards induction that

$$\lim_{N\to\infty} \frac{\|\boldsymbol{x}^j - \boldsymbol{u}^j\|^2}{N} = 0, \quad \text{for } j \in \{1, \dots, t\}, \tag{55}$$

$$\lim_{N\to\infty} \frac{\|\hat{\boldsymbol{x}}^j - \tilde{\boldsymbol{x}}^j\|^2}{N} = 0, \quad \text{for } j \in \{2, \dots, t+1\}, \tag{56}$$

$$\begin{aligned} \lim_{N\to\infty} \frac{\|\boldsymbol{x}^j\|^2}{N} &= \lim_{N\to\infty} \frac{\|\boldsymbol{u}^j\|^2}{N} \\ &= \mathbb{E}\left[\left(\tilde{z}_j + (\boldsymbol{B}_j)_{j,1}\hat{x}_1 + \sum_{i=2}^{j}(\boldsymbol{B}_j)_{j,i}\tilde{x}_i + \mu_j X - \bar{\beta}_j \tilde{x}_{j-1}\right)^2\right], \quad \text{for } j \in \{1, \dots, t\}, \end{aligned} \tag{57}$$

$$\lim_{N\to\infty} \frac{\|\hat{\boldsymbol{x}}^j\|^2}{N} = \lim_{N\to\infty} \frac{\|\tilde{\boldsymbol{x}}^j\|^2}{N} = \mathbb{E}[\tilde{x}_j^2], \quad \text{for } j \in \{2, \dots, t+1\}. \tag{58}$$

We now show that the following limits hold:

$$\lim_{N\to\infty} \frac{\|\boldsymbol{x}^{t+1} - \boldsymbol{u}^{t+1}\|^2}{N} = 0, \tag{59}$$

$$\lim_{N\to\infty} \frac{\|\hat{\boldsymbol{x}}^{t+2} - \tilde{\boldsymbol{x}}^{t+2}\|^2}{N} = 0, \tag{60}$$

$$\begin{aligned} \lim_{N\to\infty} \frac{\|\boldsymbol{x}^{t+1}\|^2}{N} &= \lim_{N\to\infty} \frac{\|\boldsymbol{u}^{t+1}\|^2}{N} \\ &= \mathbb{E}\left[\left(\tilde{z}_{t+1} + (\boldsymbol{B}_{t+1})_{t+1,1}\hat{x}_1 + \sum_{i=2}^{t+1}(\boldsymbol{B}_{t+1})_{t+1,i}\tilde{x}_i + \mu_{t+1} X - \bar{\beta}_{t+1}\tilde{x}_t\right)^2\right], \end{aligned} \tag{61}$$

$$\lim_{N\to\infty} \frac{\|\hat{\boldsymbol{x}}^{t+2}\|^2}{N} = \lim_{N\to\infty} \frac{\|\tilde{\boldsymbol{x}}^{t+2}\|^2}{N} = \mathbb{E}[\tilde{x}_{t+2}^2]. \tag{62}$$

By doing so, we will have proved also the induction step and consequently that (44) holds.

Using similar passages as in (47), we obtain

$$\begin{aligned} \boldsymbol{x}^{t+1} - \boldsymbol{u}^{t+1} &= \boldsymbol{Y}\hat{\boldsymbol{x}}^{t+1} - \beta_{t+1}\hat{\boldsymbol{x}}^t - \tilde{\boldsymbol{z}}^{t+1} - (\boldsymbol{B}_{t+1})_{t+1,1}\tilde{\boldsymbol{x}}^1 - \sum_{i=2}^{t+1}(\boldsymbol{B}_{t+1})_{t+1,i}\tilde{\boldsymbol{x}}^i - \mu_{t+1}\boldsymbol{X} + \bar{\beta}_{t+1}\tilde{\boldsymbol{x}}^t \\ &= \boldsymbol{Z}\hat{\boldsymbol{x}}^{t+1} + \sqrt{\lambda_*}\frac{\langle \boldsymbol{X}, \hat{\boldsymbol{x}}^{t+1}\rangle}{N}\boldsymbol{X} - \beta_{t+1}\hat{\boldsymbol{x}}^t - \boldsymbol{Z}\tilde{\boldsymbol{x}}^{t+1} + \sum_{i=1}^{t+1}\mathsf{b}_{t+1,i}\tilde{\boldsymbol{x}}^i - (\boldsymbol{B}_{t+1})_{t+1,1}\tilde{\boldsymbol{x}}^1 \\ &\quad - \sum_{i=2}^{t+1}(\boldsymbol{B}_{t+1})_{t+1,i}\tilde{\boldsymbol{x}}^i - \mu_{t+1}\boldsymbol{X} + \bar{\beta}_{t+1}\tilde{\boldsymbol{x}}^t. \end{aligned} \tag{63}$$

Hence, by triangle inequality,

$$\begin{aligned}\frac{\|\boldsymbol{x}^{t+1}-\boldsymbol{u}^{t+1}\|^2}{N} &\le C\Bigg(\frac{\|\boldsymbol{Z}\hat{\boldsymbol{x}}^{t+1}-\boldsymbol{Z}\tilde{\boldsymbol{x}}^{t+1}\|^2}{N}+\left(\sqrt{\lambda_*}\frac{\langle \boldsymbol{X},\hat{\boldsymbol{x}}^{t+1}\rangle}{N}-\mu_{t+1}\right)^2\frac{\|\boldsymbol{X}\|^2}{N}\\ &\quad+\sum_{i=1}^{t+1}(\mathsf{b}_{t+1,i}-(\boldsymbol{B}_{t+1})_{t+1,i})^2\frac{\|\tilde{\boldsymbol{x}}^i\|^2}{N}+(\bar{\beta}_{t+1}-\beta_{t+1})^2\frac{\|\hat{\boldsymbol{x}}^t\|^2}{N}+(\bar{\beta}_{t+1})^2\frac{\|\tilde{\boldsymbol{x}}^t-\hat{\boldsymbol{x}}^t\|^2}{N}\Bigg)\\ &:= C(T_1+T_2+T_3+T_4+T_5).\end{aligned} \tag{64}$$

Consider the first term. Since $\|\boldsymbol{Z}\|_{\mathrm{op}} \le C$, the induction hypothesis (56) implies that $T_1 \to 0$ as $N \to \infty$.

Consider the second term. The following chain of equalities holds:

$$\lim_{N\to\infty}\sqrt{\lambda_*}\frac{\langle \boldsymbol{X},\hat{\boldsymbol{x}}^{t+1}\rangle}{N}=\lim_{N\to\infty}\sqrt{\lambda_*}\frac{\langle \boldsymbol{X},\tilde{\boldsymbol{x}}^{t+1}\rangle}{N}=\sqrt{\lambda_*}\mathbb{E}[X\,\tilde{x}_{t+1}]=\sqrt{\lambda_*}\mathbb{E}[X\,\hat{x}_{t+1}]=\mu_{t+1}. \tag{65}$$

Here, the first equality uses (56) together with the fact that $\|\boldsymbol{X}\|^2/N = 1$; the second equality follows from Proposition 2; the third equality uses (42) and the definitions of $\hat{x}_{t+1}$ and $\tilde{x}_{t+1}$ in (14) and (37), respectively; and the fourth equality uses the definition of μ_{t+1} in (15). Finally, using (65) and again that $\|\boldsymbol{X}\|^2/N = 1$ gives that $T_2 \to 0$ as $N \to \infty$.

Consider the third term. The following chain of equalities holds, for $1 \le j < i \le (t+1)$,

$$\lim_{N\to\infty}(\boldsymbol{\Phi}_{t+1})_{i,j}=\lim_{N\to\infty}\langle\partial_j\tilde{\boldsymbol{x}}^i\rangle=\mathbb{E}[\partial_j\tilde{x}_i]=\mathbb{E}[\partial_j\hat{x}_i]=(\bar{\boldsymbol{\Phi}}_{t+1})_{i,j} \tag{66}$$

Here, the first equality uses the definition (35); the second equality follows from Lemma 3, as $\tilde{\boldsymbol{x}}^i = \tilde{h}_i(\tilde{\boldsymbol{z}}^1,\dots,\tilde{\boldsymbol{z}}^{i-1},\hat{\boldsymbol{x}}^1,\boldsymbol{X})$ converges in W_2 (and therefore in distribution) to $\tilde{x}_i = \tilde{h}_i(\tilde{z}_1,\dots,\tilde{z}_{i-1},\hat{x}_1,X)$ and h_i satisfies Assumption 3; the third equality uses (42) and the definitions of $\hat{x}_i$ and $\tilde{x}_i$ in (14) and (37), respectively; and the fourth equality uses the definition of $(\bar{\boldsymbol{\Phi}}_{t+1})_{i,j}$ in (16). By Assumption 1, as $N \to \infty$, $\kappa_j \to \bar{\kappa}_j$ for all j. Thus, by combining (66) with the definitions of $\boldsymbol{B}_{t+1}$ and $\tilde{\boldsymbol{B}}_{t+1}$ in (16) and (36), respectively, we conclude that, as $N \to \infty$, $\mathsf{b}_{t+1,i} \to (\boldsymbol{B}_{t+1})_{t+1,i}$ for $i \in \{1,\dots,t+1\}$. By using the induction hypothesis (58), which shows that $\|\tilde{\boldsymbol{x}}^i\|^2/N$ converges to a finite limit, we conclude that $T_3 \to 0$ as $N \to \infty$.

Consider the fourth term. By using the induction hypothesis (55) and (57), together with (46), we obtain that $\boldsymbol{x}^t$ and $\boldsymbol{u}^t$ have the same W_2 limit given by Proposition 2, namely,

$$\boldsymbol{x}^t \xrightarrow{W_2} \tilde{z}_t+(\boldsymbol{B}_t)_{t,1}\hat{x}_1+\sum_{i=2}^{t}(\boldsymbol{B}_t)_{t,i}\tilde{x}_i+\mu_t X-\bar{\beta}_t\tilde{x}_{t-1}.$$

Thus, by recalling that $\beta_{t+1} = \langle h'_{t+1}(\boldsymbol{x}^t)\rangle$, an application of Lemma 3 gives that

$$\lim_{N\to\infty}\beta_{t+1}=\mathbb{E}\left[h'_{t+1}\left(\tilde{z}_t+(\boldsymbol{B}_t)_{t,1}\hat{x}_1+\sum_{i=2}^{t}(\boldsymbol{B}_t)_{t,i}\tilde{x}_i+\mu_t X-\bar{\beta}_t\tilde{x}_{t-1}\right)\right]. \tag{67}$$

Furthermore, by using (42) and recalling the definition of $\bar{\beta}_{t+1}$, we have that the RHS of (67) is equal to $\bar{\beta}_{t+1}$. Hence, by using the induction hypothesis (58), we obtain that $T_4 \to 0$ as $N \to \infty$. Finally, by using the induction hypothesis (56), we conclude that also $T_5 \to 0$ as $N \to \infty$.

As $T_i \to 0$ for $i \in \{1,\dots,5\}$, (64) implies that (59) holds. Next, as h_{t+1} is Lipschitz, (59) immediately implies (60). Then, by using (53) with $i = t+1$ and Proposition 2, we obtain that (61) and (62) hold, thus concluding the inductive proof. The result we have just proved by induction, combined with (46), gives that (44) holds.

Another application of Proposition 2, together with (44), gives that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\psi\left((\boldsymbol{x}^1)_i,\dots,(\boldsymbol{x}^t)_i,\,(\hat{\boldsymbol{x}}^1)_i,\dots,(\hat{\boldsymbol{x}}^{t+1})_i,\,(\boldsymbol{X})_i\right)=\mathbb{E}[\psi(u_1,\dots,u_t,\tilde{x}_1,\dots,\tilde{x}_{t+1},X)], \tag{68}$$

where we have defined for $s \in \{1,\dots,t\}$,

$$u_s=\tilde{z}_s+(\boldsymbol{B}_s)_{s,1}\tilde{x}_1+\sum_{i=2}^{s}(\boldsymbol{B}_s)_{s,i}\tilde{x}_i+\mu_s X-\bar{\beta}_s\tilde{x}_{s-1}. \tag{69}$$

Finally, by using (42), we have that

$$\mathbb{E}[\psi(u_1,\ldots,u_t,\tilde{x}_1,\ldots,\tilde{x}_{t+1},X)] = \mathbb{E}[\psi(x_1,\ldots,x_t,\hat{x}_1,\ldots,\hat{x}_{t+1},X)]. \tag{70}$$

By combining (68) and (70), we obtain that the desired result (18) holds, which concludes the proof. □

C.4 State evolution for Gaussian AMP with spectral initialization

In this appendix, we consider the Gaussian AMP iteration (8) with spectral initialization $\hat{\boldsymbol{x}}_1 = \overline{\boldsymbol{v}}_N$, where $\overline{\boldsymbol{v}}_N$ denotes the eigenvector of the data matrix $\boldsymbol{Y}$ associated to the largest eigenvalue $\overline{\nu}_N$. As for the Gaussian AMP previously analyzed, one can show that the joint empirical distribution of $(\boldsymbol{x}^1,\ldots,\boldsymbol{x}^t,\hat{\boldsymbol{x}}^1,\ldots,\hat{\boldsymbol{x}}^{t+1},\boldsymbol{X})$ converges (in W_2 distance) to the random vector $(x_1,\ldots,x_t,\hat{x}_1,\ldots,\hat{x}_{t+1},X)$. The law of this random vector can be captured via a state evolution recursion, which is expressed via a sequence of vectors $\boldsymbol{\mu}_t^{\mathrm{si}} = (\mu_0,\mu_1,\ldots,\mu_t)$ and matrices $\boldsymbol{\Sigma}_t^{\mathrm{si}},\boldsymbol{\Delta}_t^{\mathrm{si}},\boldsymbol{B}_t^{\mathrm{si}} \in \mathbb{R}^{(t+1)\times(t+1)}$, defined recursively as follows. Here, the superscript si highlights that the state evolution refers to an AMP with *spectral initialization*. We also note that, to ease the notation, we have shifted the initialization index from 1 to 0.

We start with the initialization

$$\mu_0^{\mathrm{si}} = \sqrt{C_{\mathrm{OS}}},\quad \boldsymbol{\Sigma}_0^{\mathrm{si}} = 1 - C_{\mathrm{OS}},\quad \boldsymbol{\Delta}_0^{\mathrm{si}} = 1/\lambda_*,\quad \boldsymbol{B}_0^{\mathrm{si}} = R_\rho\left(\frac{1}{\sqrt{\lambda_*}}\right), \tag{71}$$

where λ_* is the true SNR (see (1)), and $R_\rho(\cdot)$ denotes the R-transform of ρ. Inductively, having defined $(\boldsymbol{\mu}_{t-1}^{\mathrm{si}},\boldsymbol{\Sigma}_{t-1}^{\mathrm{si}},\boldsymbol{\Delta}_{t-1}^{\mathrm{si}},\boldsymbol{B}_{t-1}^{\mathrm{si}})$, we define the following joint law

$$\begin{aligned}
&(z_0,,\ldots,z_{t-1}) \mid X \sim \mathcal{N}(\boldsymbol{\mu}_{t-1}^{\mathrm{si}}\cdot X,\boldsymbol{\Sigma}_{t-1}^{\mathrm{si}}),\quad \hat{x}_0 = z_0/\sqrt{\lambda_*},\quad \hat{x}_1 = z_0,\\
&x_s = z_s - \bar{\beta}_s\hat{x}_{s-1} + \sum_{i=0}^{s}(\boldsymbol{B}_s^{\mathrm{si}})_{s,i}\hat{x}_i,\quad \hat{x}_{s+1} = h_{s+1}(x_s),\ \text{for } s = 1,\ldots,t-1,
\end{aligned} \tag{72}$$

where we have set $\bar{\beta}_1 = 0$, and for $t \geq 2$, $\bar{\beta}_t = \mathbb{E}[h_t'(x_{t-1})]$. Then, the parameter μ_t^{si} is

$$\mu_t^{\mathrm{si}} = \sqrt{\lambda_*}\mathbb{E}[X\,\hat{x}_t]. \tag{73}$$

Next, we decompose the second moment matrix $\boldsymbol{\Delta}_t^{\mathrm{si}}$ of $(\hat{x}_0,\ldots,\hat{x}_t)$ into four parts:

$$\boldsymbol{\Delta}_t^{\mathrm{si}} = \bar{\boldsymbol{\Delta}}_t^{\mathrm{si}} + \tilde{\boldsymbol{\Delta}}_t^{\mathrm{si}} + (\tilde{\boldsymbol{\Delta}}_t^{\mathrm{si}})^\intercal + \hat{\boldsymbol{\Delta}}_t^{\mathrm{si}}, \tag{74}$$

where $\bar{\boldsymbol{\Delta}}_t^{\mathrm{si}},\tilde{\boldsymbol{\Delta}}_t^{\mathrm{si}},\hat{\boldsymbol{\Delta}}_t^{\mathrm{si}} \in \mathbb{R}^{(t+1)\times(t+1)}$ are defined as follows:

$$\begin{aligned}
&(\bar{\boldsymbol{\Delta}}_t^{\mathrm{si}})_{i,j} = \mathbb{E}[\hat{x}_{i-1}\,\hat{x}_{j-1}]\ \text{ for }\ i \in \{2,\ldots,t+1\}, j \in \{2,\ldots,t+1\},\ \text{ and }\ (\bar{\boldsymbol{\Delta}}_t^{\mathrm{si}})_{i,j} = 0\ \text{ otherwise},\\
&(\tilde{\boldsymbol{\Delta}}_t^{\mathrm{si}})_{i,j} = \mathbb{E}[\hat{x}_{i-1}\,\hat{x}_{j-1}]\ \text{ for }\ i = 1, j \in \{2,\ldots,t+1\}\ \text{ and }\ (\tilde{\boldsymbol{\Delta}}_t^{\mathrm{si}})_{i,j} = 0\ \text{ otherwise},\\
&(\hat{\boldsymbol{\Delta}}_t^{\mathrm{si}})_{i,j} = \mathbb{E}[\hat{x}_{i-1}\,\hat{x}_{j-1}]\ \text{ for }\ i = j = 1,\ \text{ and }\ (\hat{\boldsymbol{\Delta}}_t^{\mathrm{si}})_{i,j} = 0\ \text{ otherwise}.
\end{aligned} \tag{75}$$

Then, let us define the matrix $\bar{\boldsymbol{\Phi}}_t^{\mathrm{si}} \in \mathbb{R}^{(t+1)\times(t+1)}$ as

$$(\bar{\boldsymbol{\Phi}}_t^{\mathrm{si}})_{i,j} = 0\ \text{ if }\ i \leq j,\ \text{ and }\ (\bar{\boldsymbol{\Phi}}_t)_{i,j} = \mathbb{E}[\partial_j\hat{x}_{i-1}]\ \text{ if }\ i > j. \tag{76}$$

At this point, we define the two matrices $\boldsymbol{B}_t^{\mathrm{aux}},\tilde{\boldsymbol{B}}_t^{\mathrm{aux}} \in \mathbb{R}^{(t+1)\times(t+1)}$ as

$$\boldsymbol{B}_t^{\mathrm{aux}} = \sum_{j=0}^{\infty}\bar{\kappa}_{j+1}\bar{\boldsymbol{\Phi}}_t^j,\qquad \tilde{\boldsymbol{B}}_t^{\mathrm{aux}} = \sum_{j=0}^{\infty}\tilde{\kappa}_{j+1}\bar{\boldsymbol{\Phi}}_t^j, \tag{77}$$

where

$$\tilde{\kappa}_s = \sum_{j=0}^{\infty}\bar{\kappa}_{j+s}\left(\frac{1}{\sqrt{\lambda_*}}\right)^j,\qquad \text{for } s \geq 1. \tag{78}$$

Then, the first column of $\boldsymbol{B}_t^{\mathrm{si}}$ consists in the first column of $\tilde{\boldsymbol{B}}_t^{\mathrm{aux}}$, and the remaining columns of $\boldsymbol{B}_t^{\mathrm{si}}$ are equal to the corresponding columns of $\boldsymbol{B}_t^{\mathrm{aux}}$.

Finally, we define the covariance matrix $\mathbf{\Sigma}_t^{\mathrm{si}}$ as

$$\mathbf{\Sigma}_t^{\mathrm{si}} = \sum_{j=0}^{\infty}\sum_{i=0}^{j}(\bar{\mathbf{\Phi}}_t^{\mathrm{si}})^i(\bar{\kappa}_{j+2}\bar{\mathbf{\Delta}}_t^{\mathrm{si}} + \tilde{\kappa}_{j+2}\tilde{\mathbf{\Delta}}_t^{\mathrm{si}} + \tilde{\kappa}_{j+2}(\tilde{\mathbf{\Delta}}_t^{\mathrm{si}})^\top + \hat{\kappa}_{j+2}\hat{\mathbf{\Delta}}_t^{\mathrm{si}})((\bar{\mathbf{\Phi}}_t^{\mathrm{si}})^\top)^{j-i}, \tag{79}$$

where we have set

$$\hat{\kappa}_s = \sum_{j=0}^{\infty}(j+1)\bar{\kappa}_{j+s}\left(\frac{1}{\sqrt{\lambda_*}}\right)^j, \qquad \text{for } s \geq 1. \tag{80}$$

Having defined the state evolution recursion above, we can prove that (18) holds, where we recall that the AMP algorithm is initialized with the spectral estimate $\hat{\boldsymbol{x}}_1 = \overline{\boldsymbol{v}}_N$. The proof follows steps similar to those detailed in Appendices C.1-C.3 to show Theorem 2. In this case, the auxiliary AMP is given by the AMP with spectral initialization described in Section 3.2 of [87], and the denoisers[4] u_t are taken to be the following:

$$\begin{aligned} u_1(z_0) &= z_0, \\ u_{t+1}(z_0,\dots,z_t) &= h_{t+1}\left(z_t + \sum_{i=0}^{t}(\boldsymbol{B}_t^{\mathrm{si}})_{t,i}u_i\left(z_0,\dots,z_{i-1}\right) - \bar{\beta}_t u_{t-1}\left(z_0,\dots,z_{t-2}\right)\right), \end{aligned} \tag{81}$$

where the matrix $\boldsymbol{B}_t^{\mathrm{si}}$ is obtained via the state evolution recursion described above.

D Implementation details and additional numerical results

D.1 Correct AMP: algorithm and corresponding state evolution

In our experiments, for both the correct and Gaussian AMP, we assume to have access to an initialization $\hat{\boldsymbol{x}}^1 \in \mathbb{R}^N$ s.t. (7) holds. Then, for $t \geq 1$, the *correct AMP* iteration reads

$$\boldsymbol{x}_{\mathrm{c}}^t = \boldsymbol{Y}\hat{\boldsymbol{x}}_{\mathrm{c}}^t - \sum_{i=1}^{t}\mathsf{b}_{t,i}^{\mathrm{c}}\hat{\boldsymbol{x}}_{\mathrm{c}}^i, \quad \hat{\boldsymbol{x}}_{\mathrm{c}}^{t+1} = h_{t+1}(\boldsymbol{x}_{\mathrm{c}}^t,\dots,\boldsymbol{x}_{\mathrm{c}}^1). \tag{82}$$

To obtain the coefficients $\{\mathsf{b}_{t,i}^{\mathrm{c}}\}_{i=1}^t$, we define the matrix $\mathbf{\Phi}_t^{\mathrm{c}} \in \mathbb{R}^{t\times t}$ as

$$(\mathbf{\Phi}_t^{\mathrm{c}})_{i,j} = 0, \quad \text{for } i \leq j, \qquad (\mathbf{\Phi}_t^{\mathrm{c}})_{i,j} = \langle \partial_j \hat{\boldsymbol{x}}_{\mathrm{c}}^i\rangle, \quad \text{for } i > j, \tag{83}$$

where, for $j < i$, the vector $\langle \partial_j \hat{\boldsymbol{x}}_{\mathrm{c}}^i\rangle \in \mathbb{R}^N$ denotes the partial derivative of $h_i : \mathbb{R}^{i-1} \to \mathbb{R}$ with respect to the j-th input (applied component-wise). Then, the vector $(\mathsf{b}_{t,1}^{\mathrm{c}},\dots,\mathsf{b}_{t,t}^{\mathrm{c}})$ is given by the last row of the matrix $\boldsymbol{B}_t^{\mathrm{c}} \in \mathbb{R}^{t\times t}$ defined as

$$\boldsymbol{B}_t^{\mathrm{c}} = \sum_{j=0}^{t-1}\kappa_{j+1}(\mathbf{\Phi}_t^{\mathrm{c}})^j, \tag{84}$$

where $\{\kappa_k\}_{k\geq 1}$ denotes the sequence of free cumulants associated to the matrix $\boldsymbol{Y}$. By using the results of [37, 87] (e.g., Theorem 2.3 in [87]), one can obtain a state evolution result for the correct AMP (82). More specifically, we have that

$$(\boldsymbol{x}_{\mathrm{c}}^1,\dots,\boldsymbol{x}_{\mathrm{c}}^t,\, \hat{\boldsymbol{x}}_{\mathrm{c}}^1,\dots,\hat{\boldsymbol{x}}_{\mathrm{c}}^{t+1},\, \boldsymbol{X}) \xrightarrow{W_2} (x_1^{\mathrm{c}},\dots,x_t^{\mathrm{c}},\, \hat{x}_1^{\mathrm{c}},\dots,\hat{x}_{t+1}^{\mathrm{c}},\, X). \tag{85}$$

The law of the random vector $(x_1^{\mathrm{c}},\dots,x_t^{\mathrm{c}},\, \hat{x}_1^{\mathrm{c}},\dots,\hat{x}_{t+1}^{\mathrm{c}})$ is expressed via a sequence of vectors $\boldsymbol{\mu}_t^{\mathrm{c}} = (\mu_1^{\mathrm{c}},\dots,\mu_t^{\mathrm{c}})$ and matrices $\mathbf{\Sigma}_t^{\mathrm{c}}, \mathbf{\Delta}_t^{\mathrm{c}} \in \mathbb{R}^{t\times t}$ defined recursively as follows. We start with the initialization

$$\mu_1^{\mathrm{c}} = \sqrt{\lambda_*}\epsilon, \quad \mathbf{\Sigma}_1^{\mathrm{c}} = \bar{\kappa}_2\mathbb{E}[\hat{x}_1^2], \quad \mathbf{\Delta}_1^{\mathrm{c}} = \mathbb{E}[\hat{x}_1^2], \tag{86}$$

where λ_* is the SNR (see (1)), ϵ is given in (7), and $\{\bar{\kappa}_k\}_{k\geq 1}$ are the free cumulants associated to the asymptotic spectral measure of the noise ρ. For $t \geq 1$, given $\boldsymbol{\mu}_t^{\mathrm{c}}, \mathbf{\Sigma}_t^{\mathrm{c}}, \mathbf{\Delta}_t^{\mathrm{c}}$, let

$$\begin{aligned} &(x_1^{\mathrm{c}},\dots,x_t^{\mathrm{c}}) = (\mu_1^{\mathrm{c}},\dots,\mu_t^{\mathrm{c}})X + (w_1^{\mathrm{c}},\dots,w_t^{\mathrm{c}}), \\ &(w_1^{\mathrm{c}},\dots,w_t^{\mathrm{c}}) \sim \mathcal{N}(\mathbf{0},\mathbf{\Sigma}_t^{\mathrm{c}}) \text{ and independent of } (X,\hat{x}_1), \\ &\hat{x}_s^{\mathrm{c}} = h_s\left(x_1^{\mathrm{c}},\dots,x_{s-1}^{\mathrm{c}}\right), \quad \text{for } s \in \{2,\dots,t+1\}, \end{aligned} \tag{87}$$

[4]For the denoisers, we use the notation u_t for consistency with (3.9)-(3.10) in [87].

Then, $\bar{\mathbf{\Phi}}^{\mathrm{c}}_{t+1}, \mathbf{\Delta}^{\mathrm{c}}_{t+1} \in \mathbb{R}^{(t+1)\times(t+1)}$ are matrices with entries given by

$$\begin{aligned} &(\bar{\mathbf{\Phi}}^{\mathrm{c}}_{t+1})_{i,j} = 0, \quad \text{for } i \le j, \qquad (\bar{\mathbf{\Phi}}^{\mathrm{c}}_{t+1})_{i,j} = \mathbb{E}[\partial_j \hat{x}^{\mathrm{c}}_i], \quad \text{for } i > j, \\ &(\mathbf{\Delta}^{\mathrm{c}}_{t+1})_{i,j} = \mathbb{E}[\hat{x}^{\mathrm{c}}_i \, \hat{x}^{\mathrm{c}}_j], \quad 1 \le i, j \le t+1, \end{aligned} \tag{88}$$

where $\partial_j \hat{x}^{\mathrm{c}}_i$ denotes the partial derivative $\partial_{x^{\mathrm{c}}_j} h_i(x^{\mathrm{c}}_1, \dots, x^{\mathrm{c}}_{i-1})$. Finally, we compute μ^{c}_{t+1} and $\mathbf{\Sigma}^{\mathrm{c}}_{t+1}$ as

$$\begin{aligned} \mu^{\mathrm{c}}_{t+1} &= \mathbb{E}[X \hat{x}^{\mathrm{c}}_{t+1}], \\ \mathbf{\Sigma}^{\mathrm{c}}_{t+1} &= \sum_{j=0}^{2t} \bar{\kappa}_{j+2} \sum_{i=0}^{j} (\bar{\mathbf{\Phi}}^{\mathrm{c}}_{t+1})^i \mathbf{\Delta}^{\mathrm{c}}_{t+1} ((\bar{\mathbf{\Phi}}^{\mathrm{c}}_{t+1})^\top)^{j-i}. \end{aligned} \tag{89}$$

As usual, the $t \times t$ top left sub-matrix of $\mathbf{\Sigma}^{\mathrm{c}}_{t+1}$ is given by $\mathbf{\Sigma}^{\mathrm{c}}_t$.

D.2 Choice of non-linearities

For the denoiser $h_{t+1}(x_1, \dots, x_t)$ of the correct AMP iteration (82), we use the posterior-mean that takes into account all the past iterates, namely,

$$h_{t+1}(x_1, \dots, x_t) = \mathbb{E}[X \mid (x^{\mathrm{c}}_1, \dots, x^{\mathrm{c}}_t) = (x_1, \dots, x_t)], \tag{90}$$

where $X, x^{\mathrm{c}}_1, \dots, x^{\mathrm{c}}_t$ are the state evolution random variables defined above. If the distribution of $\boldsymbol{X}$ is uniform on the sphere, then $X \sim \mathcal{N}(0,1)$ and the conditional expectation (90) can be simplified as

$$h_{t+1}(x_1, \dots, x_t) = \frac{(\boldsymbol{\mu}^{\mathrm{c}}_t)^\top (\mathbf{\Sigma}^{\mathrm{c}}_t)^{-1} \boldsymbol{x}_t}{1 + (\boldsymbol{\mu}^{\mathrm{c}}_t)^\top (\mathbf{\Sigma}^{\mathrm{c}}_t)^{-1} \boldsymbol{\mu}^{\mathrm{c}}_t}, \tag{91}$$

where we use the short-hand $\boldsymbol{x}_t = (x_1, \dots, x_t)$. The state evolution parameters $\boldsymbol{\mu}^{\mathrm{c}}_t$ and $\mathbf{\Sigma}^{\mathrm{c}}_t$ needed to implement the denoiser h_{t+1} are estimated consistently from the data.

For the denoiser $h_{t+1}(x_t)$ of the Gaussian AMP iteration (8), we use the posterior-mean denoiser that takes into account a single iterate, namely,

$$h_{t+1}(x_t) = \frac{\mu^{\mathrm{G}}_t}{(\mu^{\mathrm{G}}_t)^2 + (\mathbf{\Sigma}^{\mathrm{G}}_t)_{t,t}} x_t. \tag{92}$$

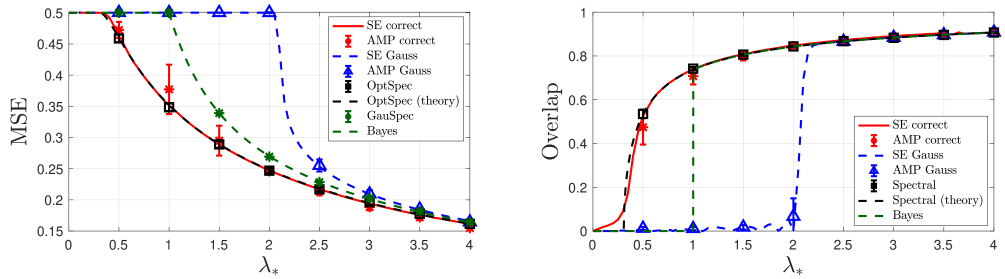
In (92), we use the supra-index G (as opposed to c) to indicate that these quantities correspond to the Gaussian AMP (8) (as opposed to the correct AMP (82)). As usual, the parameters μ^{G}_t and $(\mathbf{\Sigma}^{\mathrm{G}}_t)_{t,t}$ are obtained from the data by using the recursion of the correct AMP specialized to the case of Gaussian noise. Let us highlight that this recursion can be implemented also in the mismatched setting, as it depends only on data. However, it does *not* lead to consistent estimates of the state evolution parameters as derived in Theorem 2, because of the mismatch.

D.3 Additional numerical results

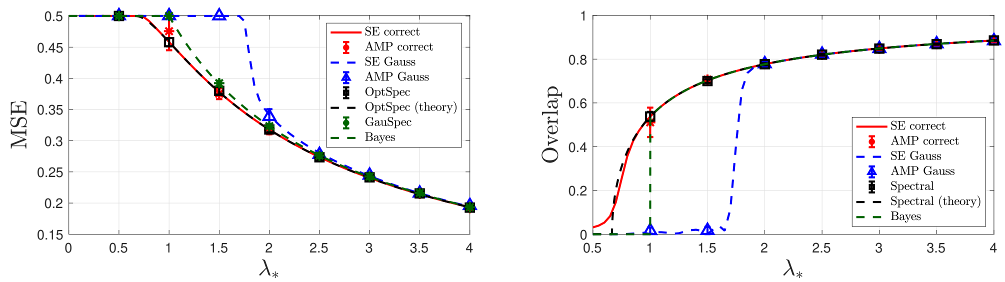
Let the matrix $\boldsymbol{A} \in \mathbb{R}^{N \times N}$ be an orthogonal matrix with "Rademacher spectrum" $\rho = \frac{1}{2}(\delta_1 + \delta_{-1})$ (the eigenvalues are i.i.d. uniform ± 1) and $\boldsymbol{W}$ a standard Wigner matrix. Then, the noise is $\boldsymbol{Z}_t := \sqrt{t}\,\boldsymbol{W} + \sqrt{1-t}\,\boldsymbol{A}$ for $t \in [0,1]$. For $t = 1$, this coincides with the pure Wigner case: here, the Bayes estimator is optimal, and the Gaussian AMP is also optimal unless a statistical-to-computational gap is present [28, 12, 54]. In contrast, for $t \in [0,1)$, there is a mismatch and our results give a sharp asymptotic characterization of the Bayes and AMP estimators, cf. Theorem 1 and 2, respectively. By additivity of the R-transform for (asymptotically) free random matrices [70], denoting $R_t(x)$ the R-transform of $\boldsymbol{Z}_t$, we obtain

$$R_t(x) = tx + \frac{\sqrt{4(1-t)x^2 + 1} - 1}{2x}.$$

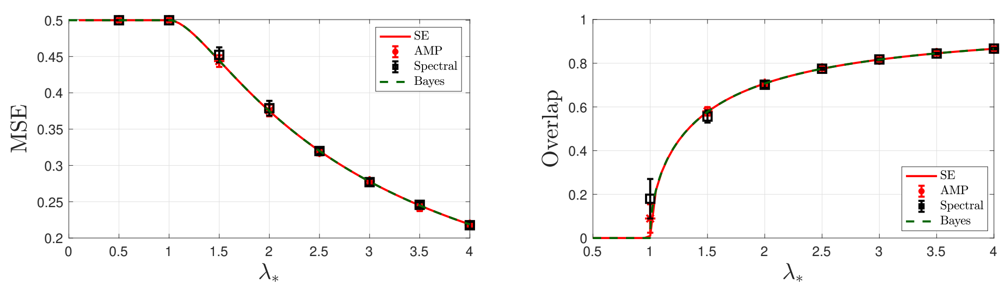
In Fig. 2, we take $t \in \{0.2, 0.5, 1\}$. We note that, for this model, Assumption 1 clearly holds, and we have verified numerically that Assumption 2 holds too. As t goes from 0 to 1, we get closer to a model without mismatch and, therefore, the performance gap between mismatched algorithms (Gaussian AMP, GauSpec, Bayes) and optimal ones (correct AMP, OptSpec) shrinks. As expected, all curves collapse at $t = 1$. The phenomenology described at the end of Section 4 can also be observed in this setting.



(a) $t = 0.2$.



(b) $t = 0.5$.



(c) $t = 1$.

Figure 2: MSE (on the left) and overlap (on the right), as a function of the true SNR λ_*, when the noise spectrum is the free convolution of Rademacher and semicircle spectra.