Supplementary for Approximate message passing for amplitude based optimization

June 6, 2018

Abstract

We consider an ℓ_2-regularized non-convex optimization problem for recovering signals from their noisy phaseless observations. We design and study the performance of a message passing algorithm that aims to solve this optimization problem. We consider the asymptotic setting $m, n \to \infty$, $m/n \to \delta$ and obtain sharp performance bounds, where m is the number of measurements and n is the signal dimension. We show that for complex signals the algorithm can perform accurate recovery with only $m = \frac{64}{\pi^2} - 4 \approx 2.5n$ measurements. Also, we provide sharp analysis on the sensitivity of the algorithm to noise. We highlight the following facts about our message passing algorithm: (i) Adding ℓ_2 regularization to the non-convex loss function can be beneficial. (ii) Spectral initialization has marginal impact on the performance of the algorithm. The sharp analyses in this paper, not only enable us to compare the performance of our method with other phase recovery schemes, but also shed light on designing better iterative algorithms for other non-convex optimization problems.

1 Motivation

Phase retrieval refers to the task of recovering a signal $\boldsymbol{x}_* \in \mathbb{C}^{n\times 1}$ from its m phaseless linear measurements:

$$y_a = \left| \sum_{i=1}^{n} A_{ai} x_{*,i} \right| + w_a, \quad a = 1, 2, \ldots, m, \tag{1.1}$$

where $x_{*,i}$ is the ith component of $\boldsymbol{x}_*$, $A_{ai} \overset{\text{i.i.d}}{\sim} \mathcal{CN}(0, \frac{1}{m})$ and $w_a \sim \mathcal{N}(0, \sigma_w^2)$ a Gaussian noise. The recent surge of interest has led to a better understanding of the theoretical aspects of this problem. Early theoretical results on phase retrieval, such as PhaseLift (Candès et al., 2013) and PhaseCut (Waldspurger et al., 2015), are based on semidefinite relaxations. For random Gaussian measurements, a variant of PhaseLift can recover the signal exactly (up to global phase) in the noiseless setting using $O(n)$ measurements (Candès & Li, 2014). A different convex optimization approach for phase retrieval was proposed in Goldstein & Studer (2016) and Bahmani & Romberg (2016). This method does not involve lifting and is computationally more attractive than its SDP-based counterparts. Apart from these convex relaxation approaches, non-convex optimization approaches have recently raised intensive research interests. These algorithms typically consist of a carefully designed initialization step (usually accomplished via a spectral method (Netrapalli et al., 2013)) followed by low-cost iterations such as alternating minimization algorithm (Netrapalli et al., 2013) or gradient descent variants like Wirtinger flow (Candès et al., 2015; Ma et al., 2017), truncated Wirtinger flow (Chen & Candès, 2017), amplitude flow (Wang et al., 2016; Zhang & Liang, 2016), incremental reshaped Wirtinger flow (Zhang et al., 2017) and reweighted amplitude flow (Wang et al., 2017a). Other approaches include Kaczmarz method (Wei, 2015; Chi & Lu, 2016; Tan & Vershynin, 2017; Jeong & Güntürk, 2017), trust region method (Sun et al., 2016), coordinate decent (Zeng & So, 2017), prox-linear (Duchi & Ruan, 2017), Polyak subgradient (Davis et al., 2017), block coordinate decent (Barmherzig & Sun, 2017).

Thanks to such research we now have access to several algorithms, inspired by different ideas, that are theoretically guaranteed to recover $\boldsymbol{x}_*$ exactly in the noiseless setting. Despite all these progresses, there is still a gap between the theoretical understanding of the recovery algorithms and what practitioners would like to know. For instance, for many algorithms, including Wirtinger flow and amplitude flow, the exact recovery is guaranteed with either $cn \log n$ or cn measurements, where c is often a fixed but large constant that does not depend on n. In both cases, it is often claimed that the large value of c or the existence of $\log n$ is an artifact of the proving technique and the algorithm is expected to work with cn for a reasonably small value of c. Such claims have left many users wondering

Q.1 Which algorithm should we use? The theoretical analyses may not be sharp and many factors may have impact on the simulations including the distribution of the noise, the true signal $\boldsymbol{x}_*$, and the number of measurements.

Q.2 When can we trust the performance of these algorithms in the presence of noise?

Q.3 What is the impact of initialization schemes, such as spectral initialization?

Researchers have developed certain intuition based on a combination of theoretical and empirical results, to give heuristic answers to these questions. However, as demonstrated in a series of papers in the context of compressed sensing, such folklores are sometimes inaccurate (Zheng et al., 2017). To address Question Q.1, several researchers have adopted the asymptotic framework $m, n \to \infty$, $m/n \to \delta$, and provided sharp analyses for the performance of several algorithms (Dhifallah & Lu, 2017; Dhifallah et al., 2017; Abbasi et al., 2017). This line of work studies recovery algorithms that are based on convex optimization. In this paper, we adopt the same asymptotic framework and study the following popular non-convex problem, known as amplitude-based optimization (Zhang & Liang, 2016; Wang et al., 2016):

$$\min_{\boldsymbol{x}} \quad \sum_{a=1}^{m} (y_a - |(\boldsymbol{A}\boldsymbol{x})_a|)^2 + \frac{\mu}{2}\|\boldsymbol{x}\|_2^2. \tag{1.2}$$

where $(\boldsymbol{A}\boldsymbol{x})_a$ denotes the a-th entry of $\boldsymbol{A}\boldsymbol{x}$. Note that compared to them, (1.2) has an extra ℓ_2-regularizer. Regularization is known to reduce the variance of an estimator and hence is expected to be useful when $\boldsymbol{w} \neq \boldsymbol{0}$. However, as we will clarify later in Section 2, since the loss function $\sum_{a=1}^{m} (y_a - |(\boldsymbol{A}\boldsymbol{x})_a|)^2$ is non-convex, regularization can help the iterative algorithm that aims to solve (1.2) even in the noiseless settings. To answer Q.1 to Q.3, we study a message passing algorithm that aims to solve (1.2). As a result of our studies, we present sharp characterization of the mean square error (even the constants are sharp) in both noiseless and noisy settings. Furthermore, in simulation section (Section 4.3), we compare our algorithm with other existing methods and present a quantitative characterization of the gain that spectral initialization can offer to our algorithms.

For phase retrieval, a Bayesian GAMP algorithm has been discussed in Schniter & Rangan (2015); Barbier et al. (2017). However, they did not provide rigorous performance analysis, particularly, how they handle the difficulty related to initialization, for which we will provide a solution in this paper. Further, the algorithm in Barbier et al. (2017) is based on the Bayesian framework, and performance analyses of Bayesian algorithms are often very challenging under "non-ideal" situations which the algorithms are not designed for. This paper considers an AMP algorithm referred as AMP.A for solving the popular optimization problem (1.2). Contrary to the Bayesian GAMP, the asymptotic performance of AMP.A does not depend on the signal and noise distributions except for their second moments. Further, given the fact that the most popular schemes in practice are iterative algorithms derived for solving non-convex optimization problems, the detailed analyses of AMP.A presented in our paper may also shed light on the performance of these algorithms and suggest new ideas to improve their performances.

2 AMP.A **Algorithm**

Our algorithm is based on the approximate message passing (AMP) framework (Donoho et al., 2009; Bayati & Montanari, 2011), in particular the generalized approximate message passing (GAMP) algorithm developed and analyzed in Rangan (2011) and Javanmard & Montanari (2013). Following the steps proposed in Rangan (2011), we obtain the following algorithm called, *Approximate Message Passing for Amplitude-based optimization* (AMP.A) (the derivation is shown in Appendix B in supplementary). Starting from an initial estimate $\boldsymbol{x}^0 \in \mathbb{C}^{n\times 1}$, AMP.A proceeds as follows for $t \geq 0$:

$$\boldsymbol{p}^t = \boldsymbol{A}\boldsymbol{x}^t - \frac{\lambda_{t-1}}{\delta} \cdot \frac{g(\boldsymbol{p}^{t-1}, \boldsymbol{y})}{-\mathrm{div}_p(g_{t-1})}, \tag{2.1a}$$

$$\boldsymbol{x}^{t+1} = \lambda_t \cdot \left(\boldsymbol{x}^t + \boldsymbol{A}^{\mathrm{H}} \frac{g(\boldsymbol{p}^t, \boldsymbol{y})}{-\mathrm{div}_p(g_t)} \right). \tag{2.1b}$$

In these iterations

$$g(p, y) = y \cdot \frac{p}{|p|} - p,$$

and

$$\lambda_t = \frac{-\mathrm{div}_p(g_t)}{-\mathrm{div}_p(g_t) + \mu \left(\tau_t + \frac{1}{2}\right)}, \tag{2.1c}$$

$$\tau^t = \frac{1}{\delta} \frac{\tau^{t-1} + \frac{1}{2}}{-\mathrm{div}_p(g_{t-1})} \cdot \lambda_{t-1}. \tag{2.1d}$$

In the above, $p/|p|$ at $p = 0$ can be any fixed number and does not affect the performance of AMP.A. Further, the "divergence" term $\mathrm{div}_p(g_t)$ is defined as

$$\begin{aligned} \mathrm{div}_p(g_t) &\triangleq \frac{1}{m} \sum_{a=1}^{m} \frac{1}{2} \left(\frac{\partial g(p_a^t, y_a)}{\partial p_a^R} - \mathrm{i} \frac{\partial g(p_a^t, y_a)}{\partial p_a^I} \right) \\ &= \frac{1}{m} \sum_{a=1}^{m} \frac{y_a}{2|p_a^t|} - 1, \end{aligned} \tag{2.2}$$

where p_a^R and p_a^I denote the real and imaginary parts of p_a^t respectively (i.e., $p_a^t = p_a^R + \mathrm{i}p_a^I$).

The first point that we would like to discuss here is the benefits of the regularization on AMP.A. Since the optimization problem in (1.2) is non-convex, iterative algorithms intended to solve it can get stuck at bad local minima. In this regard, regularization can still help AMP.A to escape bad local minima through continuation concept even in the noiseless setting. Continuation is popular in convex optimization for improving the convergence rate of iterative algorithms (Hale et al., 2008). In continuation we start with a value of μ for which AMP.A is capable of finding the global minimizer of (1.2). Then, once AMP.A converges we gradually change μ towards the target value of μ for which we want to solve the problem and use the previous fixed point of AMP.A as the initialization for the new AMP.A. We continue this process until we reach the value of μ we are interested in. For instance, if we would like to solve the noiseless phase retrieval problem then μ should eventually go to zero so that we do not introduce unnecessary bias.

A more general version of the continuation idea we discussed above is to let μ change at every iteration (denoted as μ^t), and set λ_t according to μ^t:

$$\lambda_t = \frac{-\mathrm{div}_p(g_t)}{-\mathrm{div}_p(g_t) + \mu^t \left(\tau_t + \frac{1}{2}\right)}, \tag{2.3}$$

This way not only we can automate the continuation process, but also let AMP.A decide which choice of μ is appropriate at a given stage of the algorithm. Our discussion so far has been heuristic. It is not clear whether and how much the generalized continuation can benefit the algorithm. To give a partial answer to this question, we focus on the following particular continuation strategy: $\mu^t = \frac{1+2\mathrm{div}_p(g_t)}{1+2\tau_t}$ and obtain the following version of AMP.A:

$$\boldsymbol{p}^t = \boldsymbol{A}\boldsymbol{x}^t - \frac{2}{\delta} g(\boldsymbol{p}^{t-1}, \boldsymbol{y}), \tag{2.4a}$$

$$\boldsymbol{x}^{t+1} = 2\left[-\mathrm{div}_p(g_t) \cdot \boldsymbol{x}^t + \boldsymbol{A}^{\mathrm{H}} g(\boldsymbol{p}^t, \boldsymbol{y})\right]. \tag{2.4b}$$

Note that this choice of μ_t removes $\mathrm{div}_p(g_t)$ from the denominator of (2.1), stabilizes the algorithm, and significantly improves the convergence behavior of AMP.A. A key property of AMP (including GAMP) is that its asymptotic behavior can be characterized exactly via the state evolution platform (Donoho et al., 2009; Bayati & Montanari, 2011; Rangan, 2011). Based on a standard asymptotic framework developed in Bayati & Montanari (2011) we can analyze the state evolution (SE), that captures the performance of AMP.A under the asymptotic framework. We assume that the sequence of instances $\{\boldsymbol{x}_*(n), \boldsymbol{A}(n), \boldsymbol{w}(n)\}$ is a converging sequence defined in Bayati & Montanari (2011). Further, without loss of generality, we assume $\frac{1}{n}\|\boldsymbol{x}_*(n)\|^2 \to \kappa = 1$. Then, roughly speaking, the estimate $\boldsymbol{x}^t$ can be modeled as $\alpha_t \boldsymbol{x}_* + \sigma_t \boldsymbol{h}$, where $\boldsymbol{h}$ behaves like an iid standard complex normal noise. Further, the scaling constant α_t and the noise standard deviation σ_t evolve according to a known deterministic rule, called the state evolution (SE), defined below.

Definition 1. *Starting from fixed* $(\alpha_0, \sigma_0^2) \in \mathbb{C}\times\mathbb{R}_+\backslash(0,0)$*, the sequences* $\{\alpha_t\}_{t\geq 1}$ *and* $\{\sigma_t^2\}_{t\geq 1}$ *are generated via the following recursion:*

$$\begin{aligned} \alpha_{t+1} &= \psi_1(\alpha_t, \sigma_t^2), \\ \sigma_{t+1}^2 &= \psi_2(\alpha_t, \sigma_t^2; \delta, \sigma_w^2), \end{aligned} \tag{2.5}$$

where $\psi_1 : \mathbb{C}\times\mathbb{R}_+ \mapsto \mathbb{C}$ *and* $\psi_2 : \mathbb{C}\times\mathbb{R}_+ \mapsto \mathbb{R}_+$ *are respectively given by (with* θ_α *being the phase of* α*):*

$$\psi_1(\alpha, \sigma^2) = e^{\mathrm{i}\theta_\alpha} \cdot \int_0^{\frac{\pi}{2}} \frac{|\alpha| \sin^2\theta}{\left(|\alpha|^2 \sin^2\theta + \sigma^2\right)^{\frac{1}{2}}} \mathrm{d}\theta, \tag{2.6a}$$

$$\begin{aligned} \psi_2(\alpha, \sigma^2; \delta, \sigma_w^2) = \frac{4}{\delta}\left(|\alpha|^2 + \sigma^2 + 1\right) \\ - \frac{4}{\delta}\int_0^{\frac{\pi}{2}} \frac{2|\alpha|^2\sin^2\theta + \sigma^2}{\left(|\alpha|^2\sin^2\theta + \sigma^2\right)^{\frac{1}{2}}}\mathrm{d}\theta + 4\sigma_w^2. \end{aligned} \tag{2.6b}$$

The state evolution framework for generalized AMP (GAMP) algorithms (Rangan, 2011) was formally proved in Javanmard & Montanari (2013). To apply the results in (Rangan, 2011; Javanmard & Montanari, 2013) to AMP.A, however, we need two generalizations. First, we need to extend the results to complex-valued models. This is straightforward by applying a complex-valued version of the conditioning lemma introduced in Rangan (2011); Javanmard & Montanari (2013). Second, existing results in Rangan (2011) and Javanmard & Montanari (2013) require the function g to be smooth. Our simulation results in Section 4 show that SE predicts the performance of AMP.A despite the fact that g is not smooth. For theoretical purpose, we use the smoothing idea discussed in Zheng et al. (2017) to prove the connection between the SE equations presented in (2.5) and the iterations of AMP.A in (2.4) rigorously. Let $\epsilon > 0$ be a small fixed number,

$$\boldsymbol{p}^t = \boldsymbol{A}\boldsymbol{x}_{\boldsymbol{\epsilon}}^t - \frac{2}{\delta} g_\epsilon(\boldsymbol{p}^{t-1}, \boldsymbol{y}), \tag{2.7a}$$

$$\boldsymbol{x}_{\boldsymbol{\epsilon}}^{t+1} = 2\left[-\mathrm{div}_p(g_{t,\epsilon}) \cdot \boldsymbol{x}_{\boldsymbol{\epsilon}}^t + \boldsymbol{A}^{\mathrm{H}} g_\epsilon(\boldsymbol{p}^t, \boldsymbol{y})\right], \tag{2.7b}$$

where $g_\epsilon(\boldsymbol{p}^{t-1}, \boldsymbol{y})$ refers to a vector produced by applying $g_\epsilon : \mathbb{C}\times\mathbb{R}_+ \mapsto \mathbb{C}$ below component-wise:

$$g_\epsilon(p, y) \triangleq y \cdot h_\epsilon(p) - p,$$

where for $p = p_1 + \mathrm{i}p_2$, $h_\epsilon(p)$ is defined as $h_\epsilon(p) \triangleq \frac{p_1 + \mathrm{i}p_2}{\sqrt{p_1^2 + p_2^2 + \epsilon}}$. Note that as $\epsilon \to 0$, $g_{t,\epsilon} \to g_t$ and hence we expect the iterations of smoothed-AMP.A converge to the iterations of AMP.A.

Theorem 1 (asymptotic characterization)**.** *Let* $\{\boldsymbol{x}_*(n), \boldsymbol{A}(n), \boldsymbol{w}(n)\}$ *be a converging sequence of instances. For each instance, let* $\boldsymbol{x}^0(n)$ *be an initial estimate independent of* $\boldsymbol{A}(n)$*. Assume that the following hold almost surely*

$$\lim_{n\to\infty} \frac{1}{n}\boldsymbol{x}_*^{\mathrm{H}}\boldsymbol{x}^0 = \alpha_0 \quad \textit{and} \quad \lim_{n\to\infty}\frac{1}{n}\|\boldsymbol{x}^0\|^2 = \sigma_0^2 + |\alpha_0|^2.$$

Let $\boldsymbol{x}_\epsilon^t(n)$ *be the estimate produced by the smoothed* AMP.A *initialized by* $\boldsymbol{x}^0(n)$ *(which is independent of* $\boldsymbol{A}(n)$*) and* $\boldsymbol{p}^{-1}(n) = \boldsymbol{0}$*. Let* $\epsilon_1, \epsilon_2, \ldots$ *denote a sequence of smoothing parameters for which* $\epsilon_i \to 0$ *as* $i \to \infty$

Then, for any iteration $t \geq 1$, the following holds almost surely

$$
\begin{aligned}
&\lim_{j\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}|x^t_{\epsilon_j,i}(n)-e^{\mathrm{i}\theta_t}\,x_{*,i}|^2 \\
&=\mathbb{E}\left[|X^t-e^{\mathrm{i}\theta_t}X_*|^2\right]=\left|1-|\alpha_t|\right|^2+\sigma_t^2,
\end{aligned}
\tag{2.8}
$$

where $\theta_t = \angle\alpha_t$, $X^t = \alpha_t X_* + \sigma_t H$ and $X_* \sim p_X$ is independent of $H \sim \mathcal{CN}(0,1)$. Further, $\{\alpha\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ are determined by (2.5) with initialization α_0 and σ_0^2.

The proof of theorem can be found in Appendix A.2 in supplementary.

3 Main results for SE mapping

3.1 Convergence of the SE for noiseless model

We now analyze the dynamical behavior of the SE. Before we proceed, we point out that in phase retrieval, one can only hope to recover the signal up to global phase ambiguity (Netrapalli et al., 2013; Candès et al., 2013, 2015), for generic signals without any structure. In light of (2.8), AMP.A is successful if $|\alpha_t| \to 1$ and $\sigma_t^2 \to 0$ as $t \to \infty$. By analyzing the SE, i.e, the update rule for (α_t, σ_t^2) in (2.6), the following two values of δ will play critical roles in the analysis:

$$
\delta_{\mathrm{AMP}} \triangleq \frac{64}{\pi^2} - 4 \approx 2.48 \quad \text{and} \quad \delta_{\mathrm{global}} \triangleq 2.
$$

The importance of δ_{AMP} and δ_{global} is revealed by the following two theorems (proofs can be found in Appendix A.3 and A.4 in supplementary file respectively):

Theorem 2 (convergence of SE)**.** *Consider the noiseless model where $\sigma_w^2 = 0$. If $\delta > \delta_{\mathrm{AMP}}$, then for any $0 < |\alpha_0| \leq 1$ and $\sigma_0^2 \leq 1$, the sequences $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ defined in (2.5) converge to*

$$
\lim_{t\to\infty} |\alpha_t| = 1 \quad \textit{and} \quad \lim_{t\to\infty} \sigma_t^2 = 0.
$$

Theorem 3 (local convergence of SE)**.** *When $\sigma_w^2 = 0$, then $(\alpha, \sigma^2) = (1, 0)$ is a fixed point of the SE in (2.6). Furthermore, if $\delta > \delta_{\mathrm{global}}$, then there exist two constants $\epsilon_1 > 0$ and $\epsilon_2 > 0$ such that the SE converges to this fixed point for any $\alpha_0 \in (1 - \epsilon_1, 1)$ and $\sigma_0^2 \in (0, \epsilon_2)$. On the other hand if $\delta < \delta_{\mathrm{global}}$, then the SE cannot converge to $(1, 0)$ except when initialized there.*

There are a couple of points that we would like to emphasize here:

1. $\alpha_0 \neq 0$ is essential for the success of AMP.A. This can be seen from the fact that $\alpha = 0$ is always a fixed point of $\psi_1(\alpha, \sigma^2)$ for any $\sigma^2 > 0$. From our definition of α_0 in Theorem 1, $\alpha_0 = 0$ is equivalent to $\lim_{n\to\infty} \frac{1}{n}\langle \boldsymbol{x}_*, \boldsymbol{x}^0\rangle = 0$. This means that the initial estimate $\boldsymbol{x}^0$ cannot be orthogonal to the true signal vector $\boldsymbol{x}_*$, otherwise there is no hope to recover the signal no matter how large δ is. This will be discussed in more details in Section 4.1.

2. Fig. 1 exhibits the basin of attraction of $(\alpha, \sigma^2) = (1, 0)$ as a function of δ. As expected, the basin of attraction shrinks as δ decreases. According to Theorem 3, if SE is initialized in the basin of attraction of $(\alpha, \sigma) = (1, 0)$, then it still converges to (α, σ^2) even if $\delta_{\mathrm{global}} < \delta < \delta_{\mathrm{AMP}}$. However, there are two points we should emphasize here: (i) we find that when $\delta < \delta_{\mathrm{AMP}}$, standard initialization techniques, such as the spectral method, do not help AMP.A much. Again details are discussed in Section 4 . Hence, the question of finding initialization in the basin of attraction of $(\alpha, \sigma^2) = (1, 0)$ (when $\delta < \delta_{\mathrm{AMP}}$) remains open for future research. (ii) As δ decreases from δ_{AMP} to δ_{global} the basin of attraction of $(\alpha, \sigma^2) = (1, 0)$ shrinks.

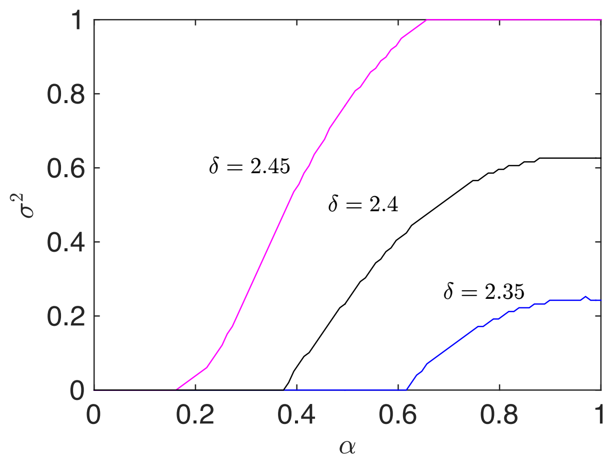

Figure 1: The regions below the curves exhibit the basin of attraction of $(\alpha, \sigma^2) = (1, 0)$ for different values of δ respectively (left to right: $\delta = 2.45, 2.4, 2.35$). The results are obtained by running the state evolution (SE) of AMP.A (complex-valued version) with α_0 and σ_0^2 chosen from 100×100 values equispaced in $[0, 1] \times [0, 1]$.

3.2 Noise sensitivity

So far we have only discussed the performance of AMP.A in the ideal setting where the noise is not present in the measurements. In general, one can use (2.5) to calculate the asymptotic MSE (AMSE) of AMP.A as a function of the variance of the noise and δ. However, as our next theorem demonstrates it is possible to obtain an explicit and informative expression for AMSE of AMP.A in the high signal-to-noise ratio (SNR) regime.

Theorem 4 (noise sensitivity)**.** *Suppose that $\delta > \delta_{\text{AMP}} = \frac{64}{\pi^2} - 4$ and $0 < |\alpha_0| \leq 1$ and $\sigma_0^2 < 1$. Then, in the high SNR regime, the asymptotic MSE defined by ($\theta_t \triangleq \angle \frac{\boldsymbol{x}_*^{\mathrm{H}} \boldsymbol{x}^t}{n}$)*

$$\text{AMSE}(\delta, \sigma_w^2) \triangleq \lim_{t \to \infty} \frac{\|\boldsymbol{x}^t - e^{\mathrm{i}\theta_t} \boldsymbol{x}_*\|_2^2}{n},$$

behaves as

$$\lim_{\sigma_w^2 \to 0} \frac{\text{AMSE}(\sigma_w^2, \delta)}{\sigma_w^2} = \frac{4}{1 - \frac{2}{\delta}}.$$

The proof of this theorem can be found in Appendix E.4 in supplementary. Note that as intuitively expected, as δ decreases the sensitivity of the algorithm to noise increases. Hence, one should set the number of measurements according to the accepted noise level in the recovered signal.

4 Initialization and Simulations

4.1 Initialization

As shown in Section 3.1, to achieve successful reconstruction, the initial estimate $\boldsymbol{x}^0$ cannot be orthogonal to the true signal $\boldsymbol{x}_*$, namely,

$$\alpha_0 = \lim_{n \to \infty} \frac{1}{n} \boldsymbol{x}_*^{\mathrm{H}} \boldsymbol{x}^0 \neq 0. \tag{4.1}$$

In many important applications (e.g., astronomic imaging and crystallography (Millane, 1990)), the signal is known to be real and nonnegative. In such cases, the following initialization of AMP.A meets the non-orthogonality requirement:

$$\boldsymbol{x}^0 = \rho \mathbf{1}, \quad \rho \neq 0.$$

(At the same time, we set $g(\boldsymbol{p}^{-1},\boldsymbol{y})=\boldsymbol{0}$.) However, finding initializations that satisfy (4.1) is not straightforward for generic complex-valued signals. Also, random initialization does not necessarily work either, since asymptotically speaking a random vector will be orthogonal to $\boldsymbol{x}_*$. One promising direction to alleviate this issue is the spectral initialization method that was introduced in (Netrapalli et al., 2013) for phase retrieval and subsequently studied in Candès et al. (2015); Chen & Candès (2017); Wang et al. (2016); Lu & Li (2017); Mondelli & Montanari (2017). Specifically, the "direction" of the signal is estimated by the principal eigenvector $\boldsymbol{v}$ ($\|\boldsymbol{v}\|^2=n$) of the following matrix:

$$\boldsymbol{D}\triangleq\boldsymbol{A}^{\mathrm{H}}\mathrm{diag}\{\mathcal{T}(y_1),\ldots,\mathcal{T}(y_m)\}\boldsymbol{A},\tag{4.2}$$

where $\mathcal{T}:\mathbb{R}_+\to\mathbb{R}$ is a nonlinear processing function, and $\mathrm{diag}\{a_1,\ldots,a_m\}$ is a diagonal matrix with diagonal entries given by $\{a_1,\ldots,a_m\}$. The exact asymptotic performance of the spectral method was characterized in Lu & Li (2017) under some regularity assumptions on $\mathcal{T}$. The analysis in Lu & Li (2017) reveals a phase transition phenomenon: the spectral estimate is not orthogonal to the signal vector $\boldsymbol{x}_*$ (i.e., (4.1) holds) if and only if δ is larger than a threshold δ_{weak}. Later, Mondelli & Montanari (2017) derived the optimal nonlinear processing function $\mathcal{T}$ (in the sense of minimizing δ_{weak}) and showed that the minimum weak threshold is $\delta_{\mathrm{weak}}=1$ for the complex-valued model.

The above discussions suggest that the spectral method can provide the required non-orthogonal initialization for AMP.A. However, the naive combination of the spectral estimate with AMP.A will not work. As shown in Figure 2, the performance of AMP.A that is initialized with the spectral method does not follow the state evolution. This is due to the fact that $\boldsymbol{x}^0$ is heavily correlated with the matrix $\boldsymbol{A}$ and violates the assumptions of SE. A trivial remedy is data splitting, i.e, we generate initialization and apply AMP.A on two separate sets of measurements (Netrapalli et al., 2013). However, this simple solution is sub-optimal in terms of sample complexity. To avoid such loss, we propose the following modification to the spectral initialization method, that we call decoupled spectral initialization:

Decoupled spectral initialization: Let $\delta>2$. Set $\boldsymbol{v}$ to be the eigenvector of $\boldsymbol{D}$ corresponding to the largest eigenvalue defined in (4.2). Let $\boldsymbol{x}^0=\rho\cdot\boldsymbol{v}$, where ρ is a fixed number which will be discussed later. Define

$$\boldsymbol{p}^0=(1-2\tau\mathcal{T}(\boldsymbol{y}))\circ\boldsymbol{A}\boldsymbol{x}^0,\tag{4.3}$$

where $\circ$ denotes entry-wise product and τ is the unique solution of[1]

$$\varphi_1(\delta,\tau)=\frac{1}{\delta},\quad \tau\in(0,\tau^\star),\tag{4.4}$$

and $\tau^\star$ is the unique solution of

$$\varphi_2(\delta,\tau^\star)=\frac{1}{\delta},\quad \tau^\star\in(0,\tau_{\max}),\tag{4.5}$$

where

$$\varphi_1(\delta,\tau)\triangleq\mathbb{E}\left[(\delta\,|Z|^2-1)\frac{2\tau\mathcal{T}(Y)}{1-2\tau\mathcal{T}(Y)}\right],\tag{4.6a}$$

$$\varphi_2(\delta,\tau)\triangleq\mathbb{E}\left[\left(\frac{2\tau\mathcal{T}(Y)}{1-2\tau\mathcal{T}(Y)}\right)^2\right].\tag{4.6b}$$

The expectations above are over $Z\sim\mathcal{CN}(0,1/\delta)$ and $Y=|Z|+W$, where $W\sim\mathcal{CN}(0,\sigma_w^2)$ is independent of Z.

Now we use $\boldsymbol{x}^0$ and $\boldsymbol{p}^0$ as the initialization for AMP.A. So far, we have not discussed how we can set ρ and $\mathcal{T}$. In this paper, we use the following $\mathcal{T}(y)$ derived by Mondelli & Montanari (2017):

$$\mathcal{T}(y)\triangleq\frac{\delta y^2-1}{\delta y^2+\sqrt{\delta}-1}.\tag{4.7}$$

[1] The uniqueness of solution in (4.4) and (4.5) is guaranteed by our choice of $\mathcal{T}(y)$ in (4.7)(Lu & Li, 2017; Mondelli & Montanari, 2017). Yet, in noisy case, (4.4) and (4.5) can only be calculated precisely if we know the variance of the noise.

Note that our initial estimate is given by $\boldsymbol{x}^{\mathbf{0}} = \rho \cdot \boldsymbol{v}$ (where $\|\boldsymbol{v}\| = \sqrt{n}$). Recall from Theorem 2 that we require $0 < |\alpha_0| < 1$ and $0 \le \sigma_0^2 < 1$ for $\delta > \delta_{\text{AMP}}$. To satisfy this condition, we can simply set $\rho = \|\boldsymbol{y}\|/\sqrt{n}$, which is an accurate estimate of $\|\boldsymbol{x}_*\|/\sqrt{n}$ in the noiseless setting (Lu & Li, 2017)[2]. Under this choice, we have $|\alpha_0|^2 + \sigma_0^2 = \rho^2 = 1$. Hence, as long as $\alpha_0 \neq 0$, we have $0 < |\alpha_0| < 1$ and $0 \le \sigma_0^2 < 1$. The choice we have picked for ρ is not necessarily optimal. We will discuss the optimal spectral initialization and what it can offer to AMP.A in Section 4.3.

In summary, our initialization in (4.3) intuitively satisfies "enough independency" requirement such that the SE for AMP.A still holds and this is supported by our numerical results in Section 4.3. We have clarified this intuition in Section 4.2. Our numerical experiments in Section 4.3 show that the estimate $\boldsymbol{x}^0$ behaves as if it is independent of the matrix $\boldsymbol{A}$. Our finding is summarized below.

Finding 1. *Let $\boldsymbol{x}^0$ and $\boldsymbol{p}^0$ be generated according to* (4.3)*, and $\{\boldsymbol{x}^t\}_{t\ge1}$ and $\{\boldsymbol{p}^t\}_{t\ge1}$ generated by the* AMP.A *algorithm as described in* (2.4)*. The AMSE converges to*

$$\lim_{n\to\infty} \frac{1}{n}\|\boldsymbol{x}^t - e^{\mathrm{i}\theta_t}\boldsymbol{x}_*\|_2^2 = (1-|\alpha_t|)^2 + \sigma_t^2,$$

where $\theta_t = \angle(\boldsymbol{x}_*^{\mathrm{H}}, \boldsymbol{x}_t)$, $\{|\alpha_t|\}_{t\ge1}$ and $\{\sigma_t^2\}_{t\ge1}$ are generated according to (2.5) *and*

$$|\alpha_0|^2 \;=\; \frac{1-\delta\varphi_2(\delta,\tau)}{1+\delta\varphi_3(\delta,\tau)} \quad \text{and} \quad \sigma_0^2 \;=\; 1-|\alpha_0|^2, \tag{4.8}$$

where τ is the solution to (4.3) *and φ_3 are defined as (φ_2 is defined in* (4.6)*)*

$$\varphi_3(\delta,\tau) \triangleq \mathbb{E}\left[\left(\delta|Z|^2-1\right)\left(\frac{2\tau\mathcal{T}(Y)}{1-2\tau\mathcal{T}(Y)}\right)^2\right], \tag{4.9}$$

where $Y = |Z| + W$.

We expect to provide a rigorous proof of this finding in a forthcoming paper.

4.2 Intuition of our initialization

Note that in conventional AMP.A, we set initial $g(\boldsymbol{p}^{-1}, \boldsymbol{y}) = \mathbf{0}$ and therefore $\boldsymbol{p}^0 = \boldsymbol{A}\boldsymbol{x}^0$. Hence, our modification in (4.3) appears to be a rescaling procedure of $\boldsymbol{p}^0$. Note that solving the principal eigenvector of $\boldsymbol{D}$ in (4.2) is equivalent to the following optimization problem:

$$\boldsymbol{v} \;=\; \underset{\|\boldsymbol{x}\|=\sqrt{n}}{\operatorname{argmin}} \; -\sum_{a=1}^{m}\mathcal{T}(y_a)\cdot\left|(\boldsymbol{A}\boldsymbol{x})_a\right|^2. \tag{4.10}$$

Following the derivations proposed in Rangan (2011), we obtain the following approximate message passing algorithm for spectral method (denoted as AMP.S):

$$\hat{\tau}^t = \frac{1}{\delta}\frac{1}{\mathrm{div}_p(h_{t-1})}\cdot\frac{\sqrt{n}}{\|\hat{\boldsymbol{r}}_{t-1}\|}, \tag{4.11a}$$

$$\hat{\boldsymbol{p}}^t = \boldsymbol{A}\hat{\boldsymbol{x}}^t - \frac{1}{\delta}\frac{h\left(\hat{\boldsymbol{p}}^{t-1}, \boldsymbol{y}, \hat{\tau}^{t-1}\right)}{\mathrm{div}_p(h_{t-1})}\cdot\frac{\sqrt{n}}{\|\hat{\boldsymbol{r}}_{t-1}\|}, \tag{4.11b}$$

$$\hat{\boldsymbol{r}}^t = \hat{\boldsymbol{x}}^t - \frac{\boldsymbol{A}^{\mathrm{H}} h\left(\hat{\boldsymbol{p}}^t, \boldsymbol{y}, \hat{\tau}^t\right)}{\mathrm{div}_p(h_t)}, \tag{4.11c}$$

$$\hat{\boldsymbol{x}}^{t+1} = -\frac{\sqrt{n}}{\|\hat{\boldsymbol{r}}_t\|}\cdot\hat{\boldsymbol{r}}^t, \tag{4.11d}$$

where we defined $h(\hat{p}, y, \hat{\tau}) \triangleq \frac{2\mathcal{T}(y)}{1-2\hat{\tau}\mathcal{T}(y)}\cdot\hat{p}$. The optimizer $\boldsymbol{v}$ of (4.10) can be regarded as the limit of the estimate $\hat{\boldsymbol{x}}^t$ under correct initialization of AMP.S. Note that AMP.S acts as a proxy and we do not intend

[2]Or one can always choose ρ to be small enough. However, this might slow down the convergence rate.

to use it for the eigenvector calculations. (There are standard numerical recipes for that purpose.) But, the correction term used in (4.3) is suggested by the Onsager correction term in AMP.S. To see that let $\hat{\boldsymbol{p}}^{\infty}$, $\hat{\boldsymbol{x}}^{\infty}$, $\hat{\tau}^{\infty}$ represent the limits of $\hat{\boldsymbol{p}}^{t}$, $\hat{\boldsymbol{x}}^{t}$, $\hat{\tau}^{t}$ respectively. Then, from (4.11a) and (4.11b), we obtain the following equation

$$\begin{aligned}\hat{\boldsymbol{p}}^{\infty} &\overset{(a)}{=} \boldsymbol{A}\hat{\boldsymbol{x}}^{\infty} - \hat{\tau}^{\infty} h\left(\hat{\boldsymbol{p}}^{\infty}, \boldsymbol{y}, \hat{\tau}^{\infty}\right), \\ &\overset{(b)}{=} \boldsymbol{A}\hat{\boldsymbol{x}}^{\infty} - \underbrace{\hat{\tau}^{\infty}\frac{2\mathcal{T}(\boldsymbol{y})}{1-2\hat{\tau}^{\infty}\mathcal{T}(\boldsymbol{y})} \circ \hat{\boldsymbol{p}}^{\infty}}_{\text{Onsager term}}\end{aligned} \tag{4.12}$$

By solving (4.12), we obtain (4.3) with rescaling of $\frac{\|\boldsymbol{y}\|}{\sqrt{n}}$ (since $\hat{\boldsymbol{x}}^{\infty} = \sqrt{n}\boldsymbol{v}$ and $\boldsymbol{x}^0 = \|\boldsymbol{y}\|\boldsymbol{v}$). Further, (4.4) and (4.5) that determine the value of $\hat{\tau}^{\infty}$ can be simplified through solving the fixed point of the following state evolution of AMP.S:

$$\hat{\alpha} = \frac{\hat{\alpha}\,\varphi_1(\delta,\hat{\tau})}{\sqrt{\hat{\alpha}^2\,\varphi_1^2(\delta,\hat{\tau}) + \frac{1}{\delta}\varphi_2(\delta,\hat{\tau}) + \frac{\hat{\alpha}^2}{\delta}\varphi_3(\delta,\hat{\tau})}}, \tag{4.13a}$$

$$1 = \frac{1}{\delta}\frac{1}{\sqrt{\hat{\alpha}^2\,\varphi_1^2(\delta,\hat{\tau}) + \frac{1}{\delta}\varphi_2(\delta,\hat{\tau}) + \frac{\hat{\alpha}^2}{\delta}\varphi_3(\delta,\hat{\tau})}}, \tag{4.13b}$$

where φ_1, φ_2 are defined in (4.6) and φ_3 is defined in (4.9).

4.3 Simulation results

We now provide simulation results to verify our analysis and compare AMP.A in (2.4) with existing algorithms. Notice that our analysis of the SE is based on a smoothing idea. Our simulation results in this section show that, for the complex-valued settting, the SE predicts the performance of AMP.A even without smoothing g.

1) Accuracy of state evolution

We first consider the noiseless setting. Fig. 2 verifies the accuracy of SE predictions of AMP.A together with the proposed initialization (i.e., (4.3)). The true signal is generated as $\boldsymbol{x}_* \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$. We measure the following two quantities (averaged over 10 runs):

$$\hat{\alpha}_t = \frac{\boldsymbol{x}_*^{\mathrm{H}}\boldsymbol{x}^t}{\|\boldsymbol{x}_*\|^2} \quad \text{and} \quad \hat{\sigma}_t^2 = \frac{\|\boldsymbol{x}^t - \hat{\alpha}_t\boldsymbol{x}_*\|^2}{\|\boldsymbol{x}_*\|^2}.$$

We expect $\hat{\alpha}_t$ and $\hat{\sigma}_t^2$ to converge to their deterministic counterparts α_t and σ_t^2 (as described in Finding 1). Indeed, Fig. 2 shows that the match between the simulated $\hat{\alpha}_t$ and $\hat{\sigma}_t^2$ (solid curves) and the SE predictions (dotted curves) is precise. For reference, we also include the simulation results for the "blind approach" where the spectral initialization is incorporated into AMP.A without applying the proposed correction (i.e., we use $\boldsymbol{p}^0 = \boldsymbol{A}\boldsymbol{x}^0$ instead of (4.3)). From Fig. 2, we see that this blind approach deviates significantly from the SE predictions. Note that the blind approach still recovers the signal correctly for the current experiment, albeit $\hat{\sigma}_t^2$ deviates from theoretical predictions. However, we found that (results are not shown here) the blind approach is unstable, and can perform rather poorly for other popular choices of $\mathcal{T}$ (such as the orthogonality-promoting method proposed in (Wang et al., 2016)).

We next consider a noisy setting. In Fig. 3, we plot the simulated MSE and the corresponding SE predictions for two different cases. For the figure on the left panel, the true signal is generated as $\boldsymbol{x}_* \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$, and the decoupled spectral initialization discussed in Section 4.1 is used. For the figure on the right panel, the signal is nonnegative and we use the initialization $\boldsymbol{x}^0 = \boldsymbol{1}$ and $\boldsymbol{g}(\boldsymbol{p}^{-1}, \boldsymbol{y}) = \boldsymbol{0}$. The nonnegative signal is generated in the following way: we set 90% of the entries to be zero and remaining 10% to be constants. (Note that the signal is sparse, but the sparsity information is not exploited in the AMP.A algorithm.) The signal-to-noise ratio (SNR) is defined to be $\mathbb{E}[\|\boldsymbol{A}\boldsymbol{x}\|^2]/\mathbb{E}[\|\boldsymbol{w}\|^2]$. The figure displays the following MSE performance:

$$\text{MSE} = \inf_{\theta\in[0,2\pi)} \frac{\|\boldsymbol{x}^t - e^{\mathrm{i}\theta}\boldsymbol{x}_*\|^2}{\|\boldsymbol{x}_*\|^2}.$$

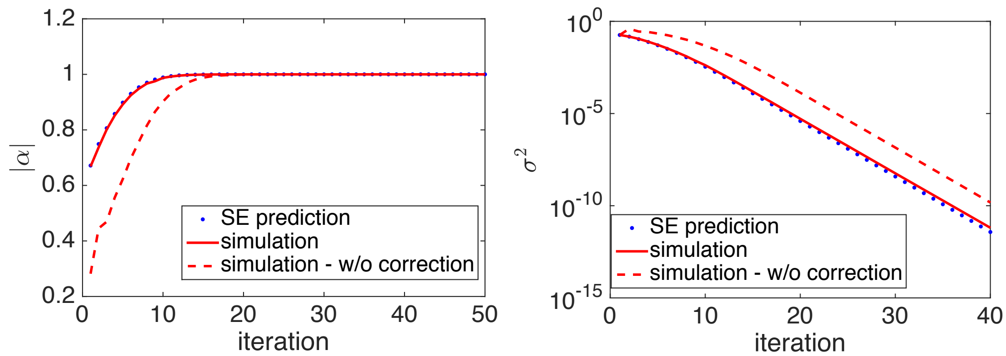

Figure 2: State evolution prediction for AMP.A with spectral initialization in the noiseless setting. **Left:** predicted and simulated results of $|\alpha|$. **Right:** predicted and simulated results of σ^2. The solid curves show the simulation results for the proposed initialization, and the dashed curves show the results for a naive approach without the proposed correction (namely, we set $\boldsymbol{p}^0 = \boldsymbol{A}\boldsymbol{x}^0$). In these experiments, $n = 5000$ and $m = 20000$. The optimal $\mathcal{T}$ in (4.7) is employed.

The SE prediction of the above MSE is given by $(1-|\alpha_t|)^2 + \sigma_t^2$. Again, we see from Fig. 3 that simulated MSE matches the SE predictions reasonably well. Further, the right figure exhibits larger fluctuations. This is mainly due to the fact that in our experiment the initialization for the right figure is less accurate than that adopted for the left figure.

2) Basin of attraction of AMP.A *and spectral initialization*

In this Section, we aim to address Q.3 we raised in the introduction. As discussed in Section 4.1, the spectral method can provide the required non-orthogonal estimate for AMP.A. Besides that, as discussed in Q.3 in Section 1, it is interesting to see if the spectral method can help AMP.A for $\delta < \delta_{\text{AMP}}$. To answer this, we need to examine whether (α_0, σ_0^2) produced by the spectral estimate can fall into the attraction basin of the good fixed point $(\alpha, \sigma^2) = (1, 0)$. Currently, the basin of attraction cannot be analytically characterized, but it can be conveniently computed via SE. Specifically, for a given (α_0, σ_0^2), we run the SE for a sufficiently large number of iterations and see if it converges to $(1, 0)$ (up to a pre-defined tolerance).

Fig. 4 plots the basin of attraction of the fixed point $(\alpha, \sigma) = (1, 0)$ for $\delta = 2.4$ or 2.41 (indicated by the blue curve). The straight line is obtained in the following way: From (Lu & Li, 2017), for a given δ and $\mathcal{T}$, the ratio σ_0/α_0 can be computed by solving a set of fixed point equations, and this ratio determines a straight line $\sigma/\alpha = \sigma_0/\alpha_0$ in the $\alpha - \sigma$ plane. The red line in Fig. 4 is obtained using $\mathcal{T}$ in (4.7). The region above the red line can be potentially achieved by certain choices of $\mathcal{T}$ together with linear scaling. On the other hand, no known $\mathcal{T}$ can achieve the region below the red line. As we see in this figure, the spectral estimate cannot fall into the basin of attraction in the current example for $\delta = 2.4$ (left subfigure). The smallest δ such that two curves intersect is numerically found to be around $\delta = 2.41$ (right subfigure) which is quite close to $\delta_{\text{AMP}} \approx 2.48$. Notice that for $\delta > \delta_{\text{AMP}}$, AMP.A works for any $\alpha_0 \neq 0$. This means that the spectral method cannot help AMP.A much besides providing an estimate not orthogonal to the true signal.

3) Comparison with existing methods

Fig. 5 displays the success recovery rate of AMP.A and the Gerchberg-Saxton algorithm (GS) (Gerchberg, 1972), truncated Wirtinger flow (TWF) (Chen & Candès, 2017), truncated amplitude flow (TAF) (Wang et al., 2017b), incremental reshaped Wirtinger flow (IRWF) (Zhang et al., 2017) and reweighted amplitude flow (RAF) (Wang et al., 2017a). Notice that the GS algorithm involves solving a least squares problem in each iteration and is thus computationally more expensive than other algorithms. For the figure in the left panel, the signal is $\boldsymbol{x}_* \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$ and the initialization is generated via the spectral method with $\mathcal{T}$ defined in (4.7). For the right panel, the signal is nonnegative (generated in the same way as that in Fig. 3) and the initial estimate is $\boldsymbol{x}^0 = \boldsymbol{1}$ for all algorithms.

We see that AMP.A outperforms all other algorithms except at $\delta = 2.7$ for the figure in the left panel. Based on simulation results not shown in this paper, we find that AMP.A outperforms IRWF consistently for a larger problem size (say $n = 2000$). However, we adopt the current setting where $n = 1000$ for ease of

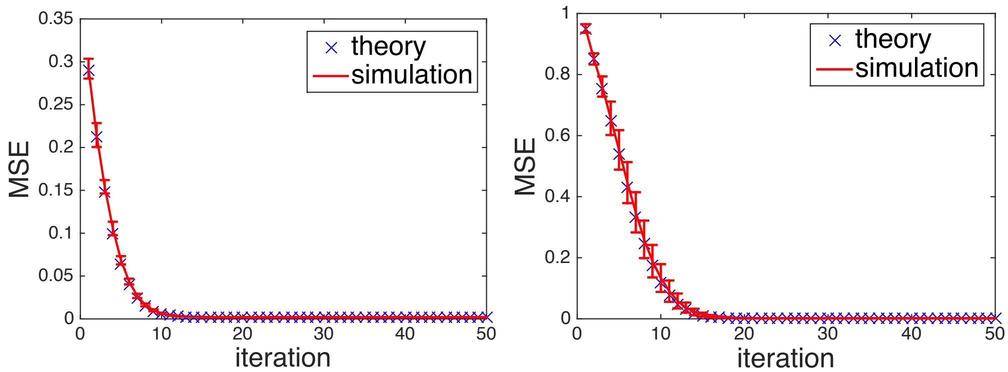

Figure 3: Simulated MSE and SE predictions in noisy settings. The solid curves show the average MSE over 10 runs. The error bars show one standard deviation.

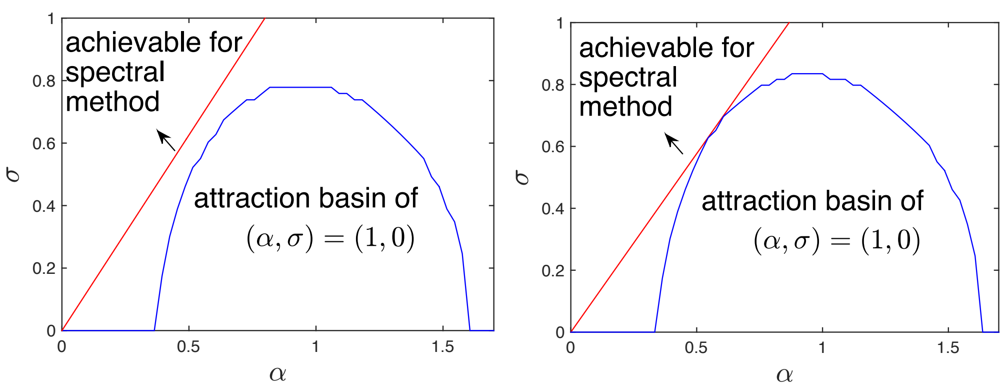

Figure 4: Plot of the attraction basin of AMP.A and the achievable region of the spectral method. **Left:** $\delta = 2.40$. **Right:** $\delta = 2.41$. In this figure, the vertical axis is σ instead of σ^2.

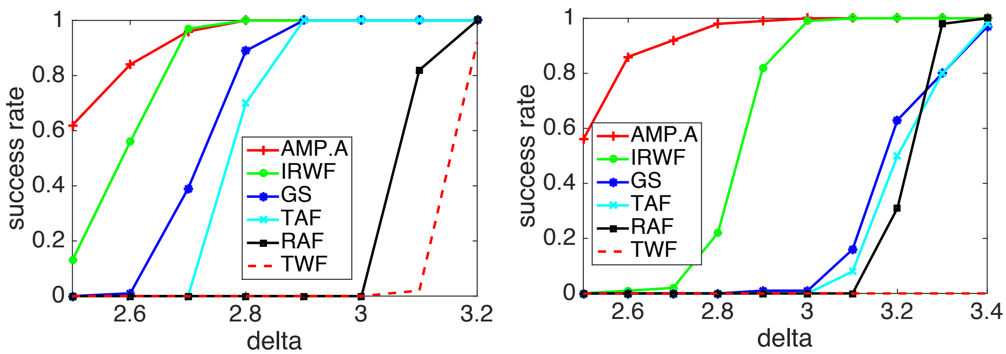

Figure 5: Recovery performance of various algorithms. We fix $n = 1000$ and vary δ. All algorithms have run 1000 iterations. Reconstruction is considered successful if the final AMSE is smaller than 10^{-10}. The success rates are measured in 100 independent realizations of $\boldsymbol{A}$ and $\boldsymbol{x}_*$. **Left:** spectral initialization with random Gaussian signal. **Right:** $\boldsymbol{x}^0 = \mathbf{1}$ and $\boldsymbol{p}^0 = \boldsymbol{A}\boldsymbol{x}^0$. The signal is nonnegative.

comparison (Chen & Candès, 2017; Wang et al., 2017b; Zhang et al., 2017; Wang et al., 2017a). Comparing the two figures in Fig. 5, we see that all algorithms are quite sensitive to the quality of initialization except for AMP.A. Notice that in the asymptotic setting where $n \to \infty$, AMP.A is able to recover the signal for all $\delta > \delta_{\mathrm{AMP}} \approx 2.48$ based on our SE analysis.

Finally, we present simulation results for the real-valued case in Fig. 6. Due to the lack of space, a thorough discussion of the real-valued AMP.A and its state evolution will be reported in a later paper. Yet, in this paper, we want to emphasize two points through Fig. 6. First, we see that AMP.A outperforms the competing algorithms with a clear phase transition between $\delta = 1.4$ and $\delta = 1.5$. This is consistent with our analysis ($\delta_{\mathrm{AMP}} = \frac{\pi^2}{4} - 1$ in real value case). Second, we notice that the IRWF algorithm (which performs best next to AMP.A in Fig. 5) is outperformed by RAF in this case. For reference, we also included the performance of the Bayesian GAMP algorithm Schniter & Rangan (2015); Barbier et al. (2017) (in conjunction with our own proposed decouple initialization to get the best performance of the Bayesian GAMP), under the assumption that the signal distribution (in this case, Gaussian) is perfectly known. As discussed in Section 1, this assumption can be unrealistic in practice. Nevertheless, the performance of Bayesian GAMP algorithm is a meaningful benchmark and hence included in Fig. 5.[3]

5 Future work

There are a couple of research directions that can be pursued in the future. First, our simulation results suggest that the AMP.A + decoupled spectral initialization can be described by a set of SE equations (see Finding 1). We hope to establish a rigorous proof for this finding. It is also interesting to investigate if the proposed decoupled spectral initialization can also work for other phase retrieval algorithms, e.g., PhaseMax. Finally, in the case of sparse signals and noisy measurements, it can be advantageous to replace the ℓ_2 regularizer by a general ℓ_p ($p \geq 0$) regularizer. How to tune the parameters in that case is largely unknown and can be a promising future direction.

[3] We also carried out simulations of Bayesian GAMP for the complex-valued case. However, we found that its performance is not competitive under the setting of Fig. 5: its recovery rate is less than 95% at $\delta = 3.5$, even when the MSE threshold is set to 10^{-6}. (Note that the MSE threshold is 10^{-10} for the curves in Fig. 5).

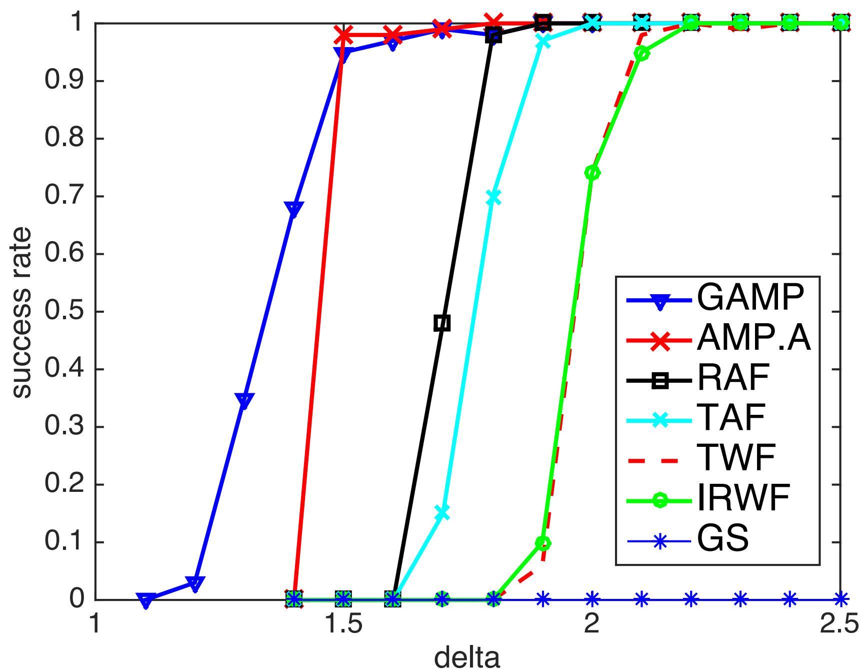

Figure 6: Recovery performance of various algorithms in real-valued case.

Acknowledgement

This work was supported in part by the U.S. National Science Foundation under Grant CIF1420328.

References

Abbasi, E., Salehi, F., and Hassibi, B. Performance of real phase retrieval. In *International Conference on Sampling Theory and Applications (SampTA)*, July 2017.

Anderson, GD and Vamanamurthy, MK. Inequalities for elliptic integrals. *Publ. Inst. Math.(Beograd)(NS)*, 37(51):61–63, 1985.

Bahmani, Sohail and Romberg, Justin. Phase retrieval meets statistical learning theory: A flexible convex relaxation. *arXiv preprint arXiv:1610.04210*, 2016.

Barbier, Jean, Krzakala, Florent, Macris, Nicolas, Miolane, Léo, and Zdeborová, Lenka. Phase transitions, optimal errors and optimality of message-passing in generalized linear models. *arXiv preprint arXiv:1708.03395*, 2017.

Barmherzig, David and Sun, Ju. A local analysis of block coordinate descent for Gaussian phase retrieval. *arXiv preprint arXiv:1712.02083*, 2017.

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

Byrd, PF and Friedman, MD. Handbook of elliptic integrals for engineers and scientists. 1971. *Berlin, Heidelberg. New York*, 1971.

Campese, Simon. Fourth moment theorems for complex Gaussian approximation. *arXiv preprint arXiv:1511.00547*, 2015.

Candès, E. J., Li, X., and Soltanolkotabi, M. Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory*, 61(4):1985–2007, April 2015.

Candès, Emmanuel J and Li, Xiaodong. Solving quadratic equations via PhaseLift when there are about as many equations as unknowns. *Foundations of Computational Mathematics*, 14(5):1017–1026, 2014.

Candès, Emmanuel J, Strohmer, Thomas, and Voroninski, Vladislav. PhaseLift: Exact and stable signal recovery from magnitude measurements via convex programming. *Communications on Pure and Applied Mathematics*, 66(8):1241–1274, 2013.

Chen, Yuxin and Candès, E. J. Solving random quadratic systems of equations is nearly as easy as solving linear systems. *Communications on Pure and Applied Mathematics*, 70:822–883, May 2017.

Chi, Yuejie and Lu, Yue M. Kaczmarz method for solving quadratic equations. *IEEE Signal Processing Letters*, 23(9):1183–1187, 2016.

Davis, Damek, Drusvyatskiy, Dmitriy, and Paquette, Courtney. The nonsmooth landscape of phase retrieval. *arXiv preprint arXiv:1711.03247*, 2017.

Dhifallah, Oussama and Lu, Yue M. Fundamental limits of PhaseMax for phase retrieval: A replica analysis. *arXiv preprint arXiv:1708.03355*, 2017.

Dhifallah, Oussama, Thrampoulidis, Christos, and Lu, Yue M. Phase retrieval via linear programming: Fundamental limits and algorithmic improvements. *arXiv preprint arXiv:1710.05234*, 2017.

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

Duchi, John C and Ruan, Feng. Solving (most) of a set of quadratic equalities: composite optimization for robust phase retrieval. *arXiv preprint arXiv:1705.02356*, 2017.

Gerchberg, Ralph W. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.

Goldstein, Tom and Studer, Christoph. PhaseMax: Convex phase retrieval via basis pursuit. *arXiv preprint arXiv:1610.07531*, 2016.

Hale, Elaine T, Yin, Wotao, and Zhang, Yin. Fixed-point continuation for ℓ_1-minimization: methodology and convergence. *SIAM Journal on Optimization*, 19(3):1107–1130, 2008.

Javanmard, Adel and Montanari, Andrea. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Information and Inference: A Journal of the IMA*, 2(2):115, 2013.

Jeong, Halyun and Güntürk, C Sinan. Convergence of the randomized Kaczmarz method for phase retrieval. *arXiv preprint arXiv:1706.10291*, 2017.

Lu, Yue M and Li, Gen. Phase transitions of spectral initialization for high-dimensional nonconvex estimation. *arXiv preprint arXiv:1702.06435*, 2017.

Ma, Cong, Wang, Kaizheng, Chi, Yuejie, and Chen, Yuxin. Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution. *arXiv preprint arXiv:1711.10467*, 2017.

Maleki, A., Anitori, L., Yang, Z., and Baraniuk, R. G. Asymptotic analysis of complex LASSO via complex approximate message passing (CAMP). *IEEE Transactions on Information Theory*, 59(7):4290–4308, July 2013.

Maleki, Arian. *Approximate message passing algorithms for compressed sensing*. Stanford University, 2010.

Millane, Rick P. Phase retrieval in crystallography and optics. *JOSA A*, 7(3):394–411, 1990.

Mondelli, Marco and Montanari, Andrea. Fundamental limits of weak recovery with applications to phase retrieval. *arXiv preprint arXiv:1708.05932*, 2017.

Netrapalli, Praneeth, Jain, Prateek, and Sanghavi, Sujay. Phase retrieval using alternating minimization. In *Advances in Neural Information Processing Systems*, pp. 2796–2804, 2013.

Rangan, S. Generalized approximate message passing for estimation with random linear mixing. In *IEEE International Symposium on Information Theory Proceedings*, pp. 2168–2172, July 2011.

Schniter, Philip and Rangan, Sundeep. Compressive phase retrieval via generalized approximate message passing. *IEEE Transactions on Signal Processing*, 63(4):1043–1055, 2015.

Sun, J., Qu, Q., and Wright, J. A geometric analysis of phase retrieval. In *IEEE International Symposium on Information Theory (ISIT)*, pp. 2379–2383, July 2016.

Tan, Yan Shuo and Vershynin, Roman. Phase retrieval via randomized Kaczmarz: Theoretical guarantees. *arXiv preprint arXiv:1706.09993*, 2017.

Waldspurger, Irène, d'spremont, Alexandre, and Mallat, Stéphane. Phase recovery, maxcut and complex semidefinite programming. *Mathematical Programming*, 149(1-2):47–81, 2015.

Wang, Gang, Giannakis, Georgios B, and Eldar, Yonina C. Solving systems of random quadratic equations via truncated amplitude flow. *arXiv preprint arXiv:1605.08285*, 2016.

Wang, Gang, Giannakis, Georgios, Saad, Yousef, and Chen, Jie. Solving most systems of random quadratic equations. In *Advances in Neural Information Processing Systems*, pp. 1865–1875, 2017a.

Wang, Miao-Kun and Chu, Yu-Ming. Asymptotical bounds for complete elliptic integrals of the second kind. *Journal of Mathematical Analysis and Applications*, 402(1):119–126, 2013.

Wang, Shuaiwen, Weng, Haolei, and Maleki, Arian. Which bridge estimator is optimal for variable selection? *arXiv preprint arXiv:1705.08617*, 2017b.

Wei, Ke. Solving systems of phaseless equations via Kaczmarz methods: A proof of concept study. *Inverse Problems*, 31(12):125008, 2015.

Zeng, Wen-Jun and So, HC. Coordinate descent algorithms for phase retrieval. *arXiv preprint arXiv:1706.03474*, 2017.

Zhang, Huishuai and Liang, Yingbin. Reshaped Wirtinger flow for solving quadratic system of equations. In *Advances in Neural Information Processing Systems*, pp. 2622–2630, 2016.

Zhang, Huishuai, Zhou, Yi, Liang, Yingbin, and Chi, Yuejie. A nonconvex approach for phase retrieval: Reshaped Wirtinger flow and incremental algorithms. *Journal of Machine Learning Research*, 18(141): 1–35, 2017.

Zheng, L., Maleki, A., Weng, H., Wang, X., and Long, T. Does ℓ_p-minimization outperform ℓ_1-minimization? *IEEE Transactions on Information Theory*, PP(99):1–1, 2017.

A Proofs of our main results

A.1 Background on Elliptic Integrals

The functions that we have in (2.5) are related to the first and second kinds of elliptic integrals. Below we review some of the properties of these functions that will be used throughout our paper. Elliptic integrals (elliptic integral of the second kind) were originally proposed for the study of the arc length of ellipsoids. Since their appearance, elliptic integrals have appeared in many problems in physics and chemistry, such as characterization of planetary orbits. Three types of elliptic integrals are of particular importance, since a large class of elliptic integrals can be reduced to these three. We introduce two of them that are of particular interest in our work.

Definition 2. *The first and second kinds of complete elliptic integrals, denoted by $K(m)$ and $E(m)$ (for $-\infty < m < 1$) respectively, are defined as (Byrd & Friedman, 1971)*

$$K(m) = \int_0^{\frac{\pi}{2}} \frac{1}{(1 - m\sin^2\theta)^{\frac{1}{2}}} \mathrm{d}\theta, \tag{A.1a}$$

$$E(m) = \int_0^{\frac{\pi}{2}} (1 - m\sin^2\theta)^{\frac{1}{2}} \mathrm{d}\theta. \tag{A.1b}$$

For convenience, we also introduce the following definition:

$$T(m) = E(m) - (1 - m)K(m). \tag{A.1c}$$

In the above definitions, we continued to use m, to follow the convention in the literature of elliptic integrals. Previously, m was defined to be the number of measurements, but such abuse of notation should not cause confusion as the exact meaning of m is usually clear from the context.

Below, we list some properties of elliptic integrals that will be used in this paper. The proofs of these properties can be found in standard references for elliptic integrals and thus omitted (e.g., (Byrd & Friedman, 1971)).

Lemma 1. *The following hold for $K(m)$ and $E(m)$ defined in* (A.1)*:*

(i) $K(0) = E(0) = \frac{\pi}{2}$. *Further, for $\epsilon \to 0$, $E(1-\epsilon)$ and $K(1-\epsilon)$ behave as*

$$\begin{aligned} E(1-\epsilon) &= 1 + \frac{\epsilon}{2}\left(\log\frac{4}{\sqrt{\epsilon}} - 0.5\right) + O(\epsilon^2 \log(1/\epsilon)) \\ K(1-\epsilon) &= \log\left(\frac{4}{\sqrt{\epsilon}}\right) + O(\epsilon\log(1/\epsilon)). \end{aligned}$$

(ii) On $m \in (0,1)$, $K(m)$ is strictly increasing, $E(m)$ is strictly decreasing, and $T(m)$ is strictly increasing.

(iii) For $m > -1$,

$$\begin{aligned} K(-m) &= \frac{1}{\sqrt{1+m}} K\left(\frac{m}{1+m}\right), \\ E(-m) &= \sqrt{1+m}\, E\left(\frac{m}{1+m}\right). \end{aligned}$$

(iv) The derivatives of $K(m)$, $E(m)$ and $T(m)$ are given by (for $m < 1$)

$$\begin{aligned} K'(m) &= \frac{E(m) - (1-m)K(m)}{2m(1-m)}, \\ E'(m) &= \frac{E(m) - K(m)}{2m}, \\ T'(m) &= \frac{1}{2}K(m). \end{aligned} \tag{A.3}$$

Furthermore, we will use a few more elliptic integrals in our work. Next lemma and its proof connects these elliptic integrals to Type I and Type II elliptic integrals.

Lemma 2. *The following equalities hold for any $m \geq 0$:*

$$\int_0^{\frac{\pi}{2}} \frac{\cos^2\theta}{\left(1+m\sin^2\theta\right)^{\frac{3}{2}}}\mathrm{d}\theta = \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(1+m\sin^2\theta\right)^{\frac{1}{2}}}\mathrm{d}\theta, \tag{A.4a}$$

$$\int_0^{\frac{\pi}{2}} \frac{3m\cos^2\theta}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta + \int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta = \int_0^{\frac{\pi}{2}} \frac{1+2m\sin^2\theta}{\left(1+m\sin^2\theta\right)^{\frac{1}{2}}}\mathrm{d}\theta. \tag{A.4b}$$

Proof. We will only prove (A.4b). (A.4a) can be proved in the same way. The idea is to express the integrals using elliptic integrals defined in (A.1), and then apply known properties of elliptic integrals (Lemma 1) to simplify the results. The same tricks in proving (A.4b) are used to derive other related integrals in this paper. Below, we will provide the full details for the proof of (A.4b), and will not repeat such calculations elsewhere. The LHS of (A.4b) can be rewritten as:

$$\int_0^{\frac{\pi}{2}} \frac{3m}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta - \int_0^{\frac{\pi}{2}} \frac{3m\sin^2\theta}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta + \int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta = \int_0^{\frac{\pi}{2}} \frac{1+2m\sin^2\theta}{\left(1+m\sin^2\theta\right)^{\frac{1}{2}}}\mathrm{d}\theta. \tag{A.5}$$

The equality in (A.5) can be proved by combining the following identities together with straightfroward manipulations:

$$\text{(i):}\quad \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta = \frac{(m+1)E\left(\frac{m}{1+m}\right) - K\left(\frac{m}{1+m}\right)}{m\sqrt{1+m}}, \tag{A.6a}$$

$$\text{(ii):}\quad \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta = \frac{K\left(\frac{m}{1+m}\right) - E\left(\frac{m}{1+m}\right)}{m\sqrt{1+m}}, \tag{A.6b}$$

$$\text{(iii):}\quad \int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta = \frac{1}{\sqrt{1+m}}E\left(\frac{m}{1+m}\right), \tag{A.6c}$$

$$\text{(iv):}\quad \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta = \frac{-(1-m)E\left(\frac{m}{1+m}\right) + K\left(\frac{m}{1+m}\right)}{3m(1+m)^{\frac{3}{2}}}, \tag{A.6d}$$

$$\text{(v):}\quad \int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta = \frac{2(m+2)E\left(\frac{m}{1+m}\right) - K\left(\frac{m}{1+m}\right)}{3(1+m)^{\frac{3}{2}}}, \tag{A.6e}$$

where $K(m)$ and $E(m)$ denote the complete elliptic integrals of the first and second kinds (see (A.1)). First, consider the identity (i) in (A.6):

$$\begin{aligned}
\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta &= \frac{1}{m}\int_0^{\frac{\pi}{2}} (1+m\sin^2\theta)^{\frac{1}{2}}\mathrm{d}\theta - \frac{1}{m}\int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta \\
&\overset{(a)}{=} \frac{1}{m}\left[E(-m) - K(-m)\right] \\
&\overset{(b)}{=} \frac{1}{m}\left[\sqrt{1+m}E\left(\frac{m}{1+m}\right) - \frac{1}{\sqrt{1+m}}K\left(\frac{m}{1+m}\right)\right],
\end{aligned}$$

where (a) is from the definition of $K(m)$ and $E(m)$ in (A.1), and (b) is from Lemma 1 (iii).

Identity (ii) can be proved as follows:

$$
\begin{aligned}
\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta &= -2\frac{\mathrm{d}}{\mathrm{d}m}\int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta \\
&= -2\frac{\mathrm{d}}{\mathrm{d}m}K(-m) \\
&\overset{(a)}{=} \frac{(1+m)K(-m)-E(-m)}{m(1+m)} \\
&\overset{(b)}{=} \frac{K\left(\frac{m}{1+m}\right)-E\left(\frac{m}{1+m}\right)}{m\sqrt{1+m}},
\end{aligned}
\tag{A.7}
$$

where (a) is due to Lemma 1 (iv) and (b) is from Lemma 1 (iii).

For identity (iii), we have

$$
\begin{aligned}
\int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta &= \int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta - m\cdot\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta \\
&\overset{(a)}{=} K(-m) - m\cdot\frac{(1+m)K(-m)-E(-m)}{m(1+m)} \\
&= \frac{E(-m)}{1+m} \\
&\overset{(b)}{=} \frac{1}{\sqrt{1+m}}E\left(\frac{m}{1+m}\right),
\end{aligned}
\tag{A.8}
$$

where step (a) follows from the third step of (A.7), and step (b) follows from Lemma 1 (iii).

Identity (iv) can be proved in a similar way:

$$
\begin{aligned}
\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta &= -\frac{2}{3}\cdot\frac{\mathrm{d}}{\mathrm{d}m}\int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta \\
&\overset{(a)}{=} -\frac{2}{3}\cdot\frac{\mathrm{d}}{\mathrm{d}m}\frac{E(-m)}{1+m} \\
&\overset{(b)}{=} \frac{(1+m)K(-m)-(1-m)E(-m)}{3m(1+m)^2} \\
&\overset{(c)}{=} \frac{-(1-m)E\left(\frac{m}{1+m}\right)-K\left(\frac{m}{1+m}\right)}{3m(1+m)^{\frac{3}{2}}},
\end{aligned}
$$

where (a) is from the third step of (A.8), step (b) is from Lemma 1 (iv) and (c) is from Lemma 1 (iii).

Lastly, identity (v) can be proved as follows:

$$
\begin{aligned}
\int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta &= \int_0^{\frac{\pi}{2}} \frac{1}{(1+m\sin^2\theta)^{\frac{3}{2}}}\mathrm{d}\theta - m\cdot\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+m\sin^2\theta)^{\frac{5}{2}}}\mathrm{d}\theta \\
&\overset{(a)}{=} \frac{E(-m)}{1+m} - m\cdot\frac{(1+m)K(-m)-(1-m)E(-m)}{3m(1+m)^2} \\
&\overset{(b)}{=} \frac{2(m+2)E\left(\frac{m}{1+m}\right)-K\left(\frac{m}{1+m}\right)}{3(1+m)^{\frac{3}{2}}},
\end{aligned}
$$

where step (a) follows from the derivations of the previous two identities and (b) is again due to Lemma 1 (iii). □

A.2 Proof of Theorem 1

Since the proof of the real-valued and complex valued signals look similar, for the sake of notational simplicity we present the proof for the real-valued signals. First note that according to Lemma 13 in Mondelli &

Montanari (2017)[4] for the smoothed AMP.A algorithm we know that almost surely

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left(x_{\epsilon_j,i}^{t+1}(n) - \text{sign}(\alpha_t) \cdot x_{*,i} \right)^2 = \mathbb{E}(X_{\epsilon_j}^{t+1} - \text{sign}(\alpha_t) \cdot X_*)^2,$$

where $X_\epsilon^t = \alpha_{\epsilon,t} X_* + \sigma_{\epsilon,t} H$ and $X_* \sim p_X$ is independent of $H \sim \mathcal{N}(0,1)$, and $\alpha_{\epsilon,t}$ and $\sigma_{\epsilon,t}$ satisfy the following iterations:

$$\begin{aligned} \alpha_{\epsilon,t+1} &= \mathbb{E}\left[\partial_z g_\epsilon(P^t, Y)\right], \\ \sigma_{\epsilon,t+1}^2 &= \mathbb{E}[g_\epsilon^2(P^t, Y)], \end{aligned}$$

where $Y = |Z| + W$, $P^t = \alpha_{\epsilon,t} Z + \sigma_{\epsilon,t} B$, where $B \sim \mathcal{N}(0, 1/\delta)$ is independent of $Z \sim \mathcal{N}(0, 1/\delta)$ and $W \sim \mathcal{N}(0, 1/\delta)$. It is also straightforward to use an induction step similar to the one presented in the proof of Theorem 1 of (Zheng et al., 2017) and show that $(\alpha_{\epsilon,t}, \sigma_{\epsilon,t}^2) \to (\alpha_t, \sigma_t^2)$ as $i \to \infty$, where (α_t, σ_t^2) satisfy

$$\begin{aligned} \alpha_{t+1} &= \mathbb{E}\left[\partial_z g(P^t, Y)\right], \\ \sigma_{t+1}^2 &= \mathbb{E}[g^2(P^t, Y)]. \end{aligned}$$

A.3 Proof of Theorem 2

The goal of this section is to prove Theorem 2. However, since the proof is very long we start with the proof sketch to help the reader navigate through the complete proof.

A.3.1 Roadmap of the proof

First note that

Lemma 3. *For any $(\alpha_0, \sigma_0^2) \in \mathbb{C} \times \mathbb{R}_+ \backslash (0,0)$, ψ_1 and ψ_2 satisfy the following properties:*

(i) $\psi_1(\alpha, \sigma^2) = \psi_1(|\alpha|, \sigma^2) \cdot e^{\mathrm{i}\theta_\alpha}$, with $e^{\mathrm{i}\theta_\alpha}$ being the phase of α;

(ii) $\psi_2(\alpha, \sigma^2) = \psi_2(|\alpha|, \sigma^2)$.

Hence, if θ_t denotes the phase of α_t, then $\theta_t = \theta_0$.

In light of this lemma, we can focus on real and nonnegative values of α_t. Then our main goal is to study the dynamics of the iterations:

$$\begin{aligned} \alpha_{t+1} &= \psi_1(\alpha_t, \sigma_t^2), \\ \sigma_{t+1}^2 &= \psi_2(\alpha_t, \sigma_t^2; \delta), \end{aligned} \tag{A.9}$$

Notice that according to the assumptions of the theorem, we assume that we initialized the dynamical system with $\alpha_0 > 0$. Our first hope is that this dynamical system will not oscillate and will converge to the solutions of the following system of nonlinear equations:

$$\begin{aligned} \alpha &= \psi_1(\alpha, \sigma^2), \\ \sigma^2 &= \psi_2(\alpha, \sigma^2; \delta), \end{aligned} \tag{A.10}$$

Hence, the first step is to characterize and understand the fixed points of the solutions of (A.10). Toward this goal we should study the properties of $\psi_1(\alpha, \sigma^2)$ and $\psi_2(\alpha, \sigma^2; \delta)$. In particular, we would like to know how the fixed points of $\psi_1(\alpha, \sigma^2)$ behave for a given σ^2 and how the fixed points of $\psi_2(\alpha, \sigma^2; \delta)$ behave for a given value of α and δ. The graphs of these functions are shown in Figure 7. We list some of the important properties of these two functions. We refer the reader to Section A.3.2 to see more accurate statement of these claims.

[4]The proof for a more general result was first presented in (Javanmard & Montanari, 2013). However, we found (Mondelli & Montanari, 2017) easier to follow. The reader may also find Claim 1 in Rangan (2011) and related discussions useful, although no formal proof was provided.

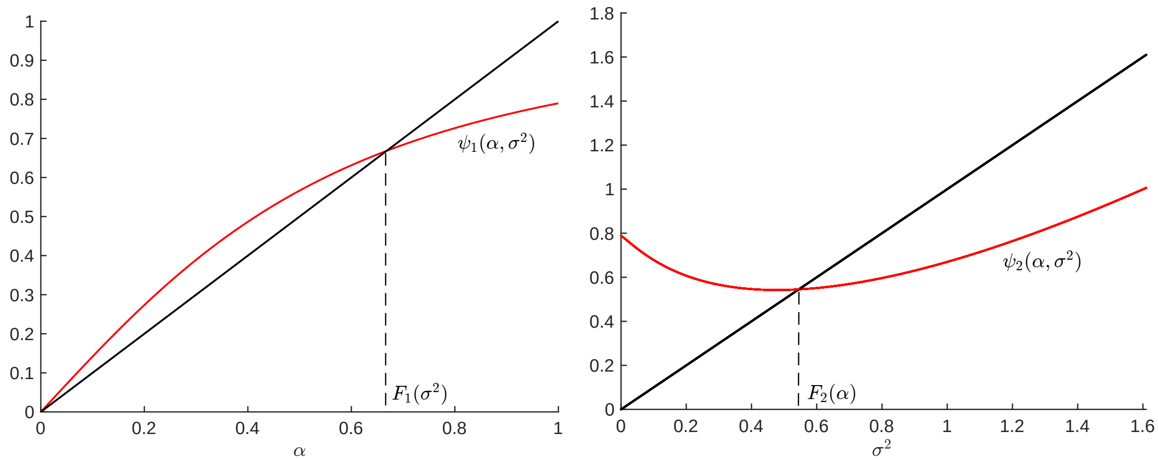

Figure 7: **Left:** plot of $\psi_1(\alpha, \sigma^2)$ against α. $\sigma^2 = 0.3$. **Right:** plot of $\psi_2(\alpha, \sigma^2; \delta)$ against σ^2. $\alpha = 0.3$ and $\delta = \delta_{\text{AMP}}$.

1. $\psi_1\left(\alpha, \sigma^2\right)$ is a concave and strictly increasing function of $\alpha > 0$, for any $\sigma^2 > 0$: This implies that $\psi_1\left(\alpha, \sigma^2\right)$ can have two fixed points: one at zero and one at $\alpha > 0$. Also, as is clear from the figure, the second fixed point is the stable one.

2. If $\delta > \delta_{\text{AMP}}$, then ψ_2 has always one stable fixed point. It may have one unstable fixed points (as a function of σ^2). See Fig. 10 for an example of this situation.

For the moment assume that the unstable fixed points do not affect the dynamics of AMP.A. Let $F_1(\sigma^2)$ denote the non-zero fixed point of ψ_1 and $F_2(\sigma^2)$ the stable fixed point of ψ_2.[5] We will prove in Lemma 10 that $F_1(\sigma^2)$ is a decreasing function and hence $F_1^{-1}(\alpha)$ is well-defined on $0 < \alpha \leq 1$. Moreover, we will show that by choosing $F_1^{-1}(0) = \frac{\pi^2}{16}$, $F_1^{-1}(\alpha)$ is continuous on [0,1]. $F_1^{-1}(\alpha)$ and $F_2(\alpha; \delta)$ are shown in Fig. 8. Note that the places these curves intersect correspond to the fixed points of (A.10). Depending on the value of δ the two curves show the following different behaviors:

1. When $\delta > \delta_{\text{AMP}}$, the dashed curve (see Fig. 8) is entirely below the solid curve except at $(\alpha, \sigma^2) = (1, 0)$. δ_{AMP} is the critical value of δ at which $F_2(0; \delta) = F_1^{-1}(0)$. Formally, we will prove the following lemma:

   Lemma 4. *If $\delta \geq \delta_{\text{AMP}} = \frac{64}{\pi^2} - 4$, then $F_1^{-1}(\alpha) > F_2(\alpha; \delta)$ holds for any $\alpha \in (0, 1)$.*

   You may find the proof of this lemma in Section A.3.4. Intuitively speaking, in this case we expect the state evolution to converge to the fixed point $(\alpha, \sigma^2) = (1, 0)$, meaning that AMP.A achieves exact recovery.

2. When $2 < \delta < \delta_{\text{AMP}}$, the two curves intersect at multiple locations, but $F_2(\alpha) < F_1^{-1}(\alpha)$ for the values of α that are close to one. This implies that AMP.A can still exactly recover $\boldsymbol{x}_*$ if the initialization is close enough to $\boldsymbol{x}_*$. However, this does not happen with spectral initialization. We will discuss this case in Theorem 3 and we do not pursue it further here.

So far, we have studied the solutions of (A.10). But the ultimate goal of analysis of AMP.A is the analysis of (A.9). In particular, it is important to show that the estimates (α_t, σ_t^2) converge to $(1, 0)$ and do not oscillate. Unfortunately, the dynamics of (α_t, σ_t^2) do not monotonically move toward the fixed point $(1, 0)$, which makes the analysis of SE complicated.

Suppose that $\delta > \delta_{\text{AMP}}$. We first show that (α_t, σ_t^2) lies within a bounded region if the initialization falls into that region.

[5] In the literature of dynamical systems, these functions are sometimes called *nullclines*. Nullclines are useful for qualitatively analyzing local dynamical behavior of two-dimensional maps (which is the case for the SE in this paper).

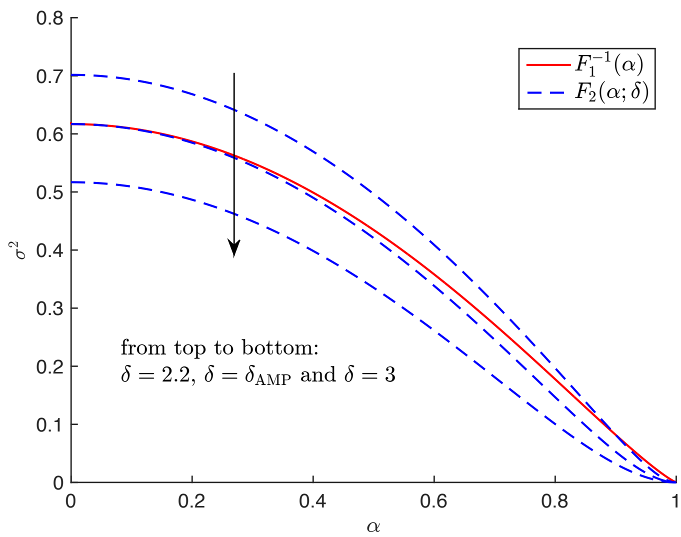

Figure 8: Plots of $F_1^{-1}(\alpha)$ and $F_2(\alpha)$ for different values of δ. When $\delta = \delta_{\text{AMP}}$, $F_1^{-1}(\alpha)$ and $F_2(\alpha;\delta)$ intersect at $\alpha = 0$.

Lemma 5. *Suppose that $\alpha_0 > 0$ and $\sigma_0^2 \leq 1$. If $\delta > \delta_{\text{AMP}} = \frac{64}{\pi^2} - 4$, then the sequences $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ generated by* (2.5) *satisfy the following:*

$$0 \leq \alpha_t \leq 1 \quad \textit{and} \quad 0 \leq \sigma_t^2 \leq \sigma_{\max}^2, \quad \forall t \geq 1,$$

where $\sigma_{\max}^2 \triangleq \max\left\{1, \frac{4}{\delta}\right\}$.

Proof. As discussed in Lemma 3, the assumption $\alpha_0 > 0$ implies that $\alpha_t > 0$, $\forall t \geq 1$. Further, from the property that $0 < \psi_1(\alpha, \sigma^2) < 1$ for $\alpha > 0$ and $\sigma^2 > 0$ (see Lemma 8 (ii)), we readily have $0 \leq \alpha_t \leq 1$. Similarly, Lemma 9 (iii) shows that if $\delta > \delta_{\text{AMP}}$, $\alpha \in [0, 1]$ and $\sigma^2 \in [0, \sigma_{\max}^2]$, then $0 \leq \psi_2(\alpha, \sigma^2; \delta) \leq \sigma_{\max}^2$. By our assumption, we have $\sigma_0^2 \leq 1 \leq \sigma_{\max}^2$, and using induction we prove $0 \leq \sigma_t^2 \leq \sigma_{\max}^2$. $\square$

From the above lemma, we see that to understand the dynamics of the SE, we only focus on the region $\mathcal{R} \triangleq \left\{(\alpha, \sigma^2) \middle| 0 < \alpha \leq 1, 0 < \sigma^2 \leq \sigma_{\max}^2\right\}$. Since the dynamic of AMP.A is complicated, we divide this region into smaller regions. See Figure 9 for an illustration.

Definition 3. *We divide $\mathcal{R} \triangleq \left\{(\alpha, \sigma^2) \middle| 0 < \alpha \leq 1, 0 < \sigma^2 \leq \sigma_{\max}^2\right\}$ into the following three sub-regions:*

$$\begin{aligned}
\mathcal{R}_0 &\triangleq \left\{(\alpha, \sigma^2) \middle| 0 < \alpha \leq 1, \frac{\pi^2}{16} < \sigma^2 \leq \sigma_{\max}^2\right\}, \\
\mathcal{R}_1 &\triangleq \left\{(\alpha, \sigma^2) \middle| 0 < \alpha \leq 1, F_1^{-1}(\alpha) < \sigma^2 \leq \frac{\pi^2}{16}\right\}, \\
\mathcal{R}_2 &\triangleq \left\{(\alpha, \sigma^2) \middle| 0 < \alpha \leq 1, 0 \leq \sigma^2 \leq F_1^{-1}(\alpha)\right\}.
\end{aligned} \tag{A.11}$$

Our next lemma shows that if (α_t, σ_t^2) is in $\mathcal{R}_1$ or $\mathcal{R}_2$ for $t \geq 1$, then (α_t, σ_t^2) converges to $(1, 0)$. The following lemma demonstrates this claim.

Lemma 6. *Suppose that $\delta > \delta_{\text{AMP}}$. If $(\alpha_{t_0}, \sigma_{t_0}^2)$ is in $\mathcal{R}_1 \cup \mathcal{R}_2$ at time t_0 (where $t_0 \geq 1$), and $\{\alpha_t\}_{t\geq t_0}$ and $\{\sigma_t^2\}_{t\geq t_0}$ are obtained via the SE in* (2.5)*, then*

(i) (α_t, σ_t^2) *remains in $\mathcal{R}_1 \cup \mathcal{R}_2$ for all $t > t_0$;*

(ii) (α_t, σ_t^2) *converges:*

$$\lim_{t\to\infty} \alpha_t = 1 \quad \textit{and} \quad \lim_{t\to\infty} \sigma_t^2 = 0.$$

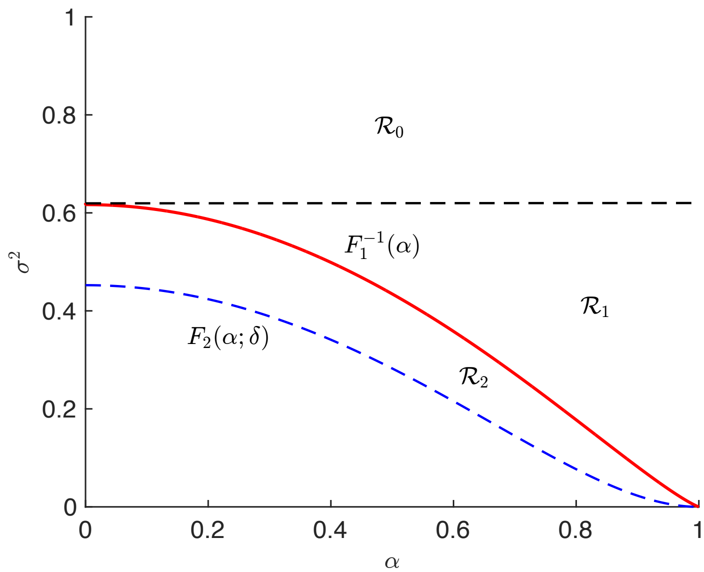

Figure 9: Illustration of the three regions in Definition 3. Note that $\mathcal{R}_2$ also includes the region below $F_2(\alpha;\delta)$.

This claim will be proved in Section A.3.5. Notice that the condition $t_0 \geq 1$ is important for part (i) to hold: if (α_0, σ_0^2) is close to the origin (and thus in $\mathcal{R}_2$), then (α_1, σ_1^2) can move to $\mathcal{R}_0$. However, this cannot happen when $t \geq 1$. In the proof given in Section A.3.5, we showed that for any $(\alpha_0, \sigma_0^2) \in \mathcal{R}$ the possible locations of (α_1, σ_1^2) are bounded from below by a curve, and once (α, σ^2) is above this curve and also in region $\mathcal{R}_1$ or $\mathcal{R}_2$, then we will prove that it cannot go to $\mathcal{R}_0$. Finally, we will prove the following Lemma that completes the proof.

Lemma 7. *Suppose that $\delta > \delta_{\text{AMP}}$. Let $\{\alpha_t\}_{t\geq 1}$ and $\{\sigma_t^2\}_{t\geq 1}$ be the sequences generated according to* (2.5) *from any $(\alpha_0, \sigma_0^2) \in \mathcal{R}_0$. Then, there exists a finite number $T \geq 1$ such that $(\alpha_T, \sigma_T^2) \in \mathcal{R}_1 \cup \mathcal{R}_2$.*

The proof of this result is in Section A.3.6. Combining the above two lemmas, it is straightforward to see that $(\alpha_t, \sigma_t^2) \to (1, 0)$, and hence the proof is complete.

Below we present the missing details.

A.3.2 Properties of ψ_1 and ψ_2

In this section we derive all the main properties of ψ_1 and ψ_2 that are used throughout the paper.

Lemma 8. *$\psi_1\left(\alpha, \sigma^2\right)$ has the following properties (for $\alpha \geq 0$):*

(i) $\psi_1\left(\alpha, \sigma^2\right)$ is a concave and strictly increasing function of $\alpha > 0$, for any given $\sigma^2 > 0$.

(ii) $0 < \psi_1(\alpha, \sigma^2) \leq 1$, for $\alpha > 0$ and $\sigma^2 > 0$.

(iii) If $0 < \sigma^2 < \pi^2/16$, then there are two nonnegative solutions to $\alpha = \psi_1(\alpha, \sigma^2)$: $\alpha = 0$ and $\alpha = F_1(\sigma^2) > 0$. Further, $F_1(\sigma^2)$ is strongly globally attracting, meaning that

$$\alpha < \psi_1(\alpha, \sigma^2) < F_1(\sigma^2), \quad \alpha \in (0, F_1(\sigma^2)), \tag{A.12a}$$

and

$$F_1(\sigma^2) < \psi_1(\alpha, \sigma^2) < \alpha, \quad \alpha \in (F_1(\sigma^2), \infty). \tag{A.12b}$$

On the other hand, if $\sigma^2 \geq \pi^2/16$ then $\alpha = 0$ is the unique nonnegative fixed point and it is strongly globally attracting.

Proof. Part (i): From (2.6), it is easy to verify that $\psi_1(\alpha, \sigma^2)$ is an increasing function of $\alpha > 0$. We now prove its concavity. To this end, we calculate its first and second partial derivatives:

$$\frac{\partial \psi_1(\alpha, \sigma^2)}{\partial \alpha} = \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta \cdot \sigma^2}{(\alpha^2 \sin^2\theta + \sigma^2)^{\frac{3}{2}}} \mathrm{d}\theta, \tag{A.13a}$$

$$\frac{\partial_1^2 \psi_1(\alpha, \sigma^2)}{\partial \alpha^2} = \int_0^{\frac{\pi}{2}} \frac{-3\sin^4\theta \cdot \sigma^2 \alpha}{(\alpha^2 \sin^2\theta + \sigma^2)^{\frac{5}{2}}} \mathrm{d}\theta < 0, \quad \forall \alpha > 0, \sigma^2 > 0. \tag{A.13b}$$

Hence, $\psi_2(\alpha, \sigma^2)$ is a concave function of α for $\alpha > 0$.

Part (ii): Positivity of ψ_1 is obvious. Also, note that

$$\psi_1(\alpha, \sigma^2) = \int_0^{\pi/2} \frac{\sin^2\theta}{(\sin^2(\theta) + \frac{\sigma^2}{\alpha^2})^{\frac{1}{2}}} d\theta \le \int_0^{\pi/2} \sin\theta d\theta = 1.$$

Proof of (iii): The claim is a consequence of the concavity of ψ_1 (with respect to α) and the following condition:

$$\frac{\partial \psi_1(\alpha, \sigma^2)}{\partial \alpha}\Big|_{\alpha=0} = 1 \Longleftrightarrow \sigma^2 = \frac{\pi^2}{16}.$$

The detailed proof is as follows. First, it is straightforward to verify that $\alpha = 0$ is always a solution to $\alpha = \psi_1(\alpha, \sigma^2)$. Define

$$\Psi_1(\alpha, \sigma^2) \triangleq \psi_1(\alpha, \sigma^2) - \alpha.$$

Since $\Psi_1(\alpha, \sigma^2)$ is a concave function of α (as $\psi_1(\alpha, \sigma^2)$ is concave), $\frac{\partial \Psi_1(\alpha, \sigma^2)}{\partial \alpha}$ is decreasing. Let's first consider $\sigma^2 > \pi^2/16$. In this case we know that

$$\frac{\partial \Psi_1(\alpha, \sigma^2)}{\partial \alpha} \le \frac{\partial \Psi_1(\alpha, \sigma^2)}{\partial \alpha}\Big|_{\alpha=0} = \frac{\partial \psi_1(\alpha, \sigma^2)}{\partial \alpha}\Big|_{\alpha=0} - 1 = \frac{\pi}{4\sigma} - 1 < 0, \tag{A.14}$$

where the second equality can be calculated from (A.13a). Since $\Psi_1(\alpha, \sigma^2)$ is a decreasing function of α and is equal to zero at zero, and it does not have any other solution. Now, consider case $\sigma^2 < \pi^2/16$. It is straightforward to confirm that

$$\frac{\partial \Psi_1(\alpha, \sigma^2)}{\partial \alpha}\Big|_{\alpha=0} = \frac{\partial \psi_1(\alpha, \sigma^2)}{\partial \alpha}\Big|_{\alpha=0} - 1 = \frac{\pi}{4\sigma} - 1 > 0.$$

Furthermore, from (A.13a) we have $\frac{\partial \psi_1(\alpha, \sigma^2)}{\partial \alpha}\Big|_{\alpha \to \infty} = 0$, and so

$$\frac{\partial \Psi_1(\alpha, \sigma^2)}{\partial \alpha}\Big|_{\alpha \to \infty} \to -1.$$

Hence, $\Psi_1(\alpha, \sigma^2) = 0$ has exactly one more solution for $\alpha > 0$. Note that since from part (ii) $\psi_1(\alpha, \sigma^2) < 1$, the solution of $\alpha = \psi_1(\alpha, \sigma^2)$ also satisfies $\alpha \le 1$.

Finally, the strong global attractiveness follows from the fact that ψ_1 is a strictly increasing function of α.

□

Lemma 9. $\psi_2\left(\alpha, \sigma^2; \delta\right)$ *has the following properties:*

(i) If $\delta < 2$, then $\sigma^2 = 0$ is a locally unstable fixed point to $\sigma^2 = \psi_2\left(\alpha, \sigma^2; \delta\right)$, meaning that

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2}\Big|_{\alpha=1, \sigma^2=0} > 1.$$

(ii) For any $\delta > 2$, $\sigma^2 = \psi_2\left(\alpha, \sigma^2; \delta\right)$ has a unique fixed point in $\sigma^2 \in [0,1]$ for any $\alpha \in [0,1]$. Further, the fixed point is (weakly) globally attracting in $\sigma^2 \in [0,1]$:

$$\sigma^2 < \psi_2(\alpha, \sigma^2; \delta), \quad \sigma^2 \in (0, F_2(\alpha)), \tag{A.15a}$$

and

$$\psi_2(\alpha, \sigma^2) < \sigma^2, \quad \sigma^2 \in (F_2(\alpha), 1). \tag{A.15b}$$

(iii) If $\delta \geq \delta_{\text{AMP}}$, then for any $\alpha \in [0,1]$, we have

$$0 \leq \psi_2(\alpha, \sigma^2; \delta) \leq \sigma^2_{\max}, \quad \sigma^2 \in [0, \sigma^2_{\max}],$$

where $\sigma^2_{\max} \triangleq \max\{1, 4/\delta\}$.

(iv) If $\delta \geq \delta_{\text{AMP}}$, then for any $\alpha \in [0,1]$, $F_2(\alpha)$ is the unique (weakly) globally attracting fixed point of $\sigma^2 = \psi_2(\alpha, \sigma^2; \delta)$ in $\sigma^2 \in [0, \sigma^2_{\max}]$. Namely,

$$\sigma^2 < \psi_2(\alpha, \sigma^2; \delta), \quad \sigma^2 \in (0, F_2(\alpha)), \tag{A.16a}$$

and

$$\psi_2(\alpha, \sigma^2) < \sigma^2, \quad \sigma^2 \in (F_2(\alpha), \sigma^2_{\max}). \tag{A.16b}$$

(v) For any $\delta > 0$, $\psi_2(\alpha, \sigma^2; \delta)$ is an increasing function of $\sigma^2 > 0$ if

$$\alpha > \alpha_* \triangleq \frac{1}{2\sqrt{1+s_*^2}} E\left(\frac{1}{1+s_*^2}\right) \approx 0.53, \tag{A.17}$$

where s_*^2 is the unique solution to

$$2E\left(\frac{1}{1+s_*^2}\right) = K\left(\frac{1}{1+s_*^2}\right).$$

Here, $K(\cdot)$ and $E(\cdot)$ denote the complete elliptic integrals introduced in (A.1). Further, when $\alpha > \alpha_*$ and $\delta > \delta_{\text{AMP}}$, then $F_2(\sigma^2)$ is strongly globally attracting in $[0, \sigma^2_{\max}]$. Specifically,

$$\sigma^2 < \psi_2(\alpha, \sigma^2; \delta) < F_2(\alpha), \quad \sigma^2 \in (0, F_2(\alpha)),$$

and

$$F_2(\alpha) < \psi_2(\alpha, \sigma^2) < \sigma^2, \quad \sigma^2 \in (F_2(\alpha), \sigma^2_{\max}).$$

Proof. First note that the partial derivative of ψ_2 w.r.t. σ^2 is given by

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} = \frac{4}{\delta}\left(1 - \frac{1}{2}\int_0^{\frac{\pi}{2}} \frac{\sigma^2}{(\alpha^2 \sin^2\theta + \sigma^2)^{\frac{3}{2}}} \mathrm{d}\theta\right). \tag{A.18}$$

Part (i): Before we proceed, we first comment on the discontinuity of the partial derivative $\frac{\partial \psi_2(\alpha,\sigma^2;\delta)}{\partial \sigma^2}$ at $\sigma^2 = 0$. Note that the formula in (A.18) was derived for non-zero values of σ^2. Naively, one may plug in $\sigma^2 = 0$ in the equation and assume that $\frac{\partial \psi_2(\alpha,\sigma^2;\delta)}{\partial \sigma^2}\Big|_{\alpha=1,\sigma^2=0} = \frac{4}{\delta}$. This is not the case since the integral $\int_0^{\pi/2} \frac{d\theta}{\sin\theta}$ is divergent. It turns out that the derivative $\frac{\partial \psi_2(\alpha,\sigma^2;\delta)}{\partial \sigma^2}$ is a continuous function of σ^2. The technical details can be found in Appendix D.

Since $\frac{\partial \psi_2(\alpha,\sigma^2;\delta)}{\partial \sigma^2}$ is continuous at $\sigma^2 = 0$, we have

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2}\Big|_{\alpha=1,\sigma^2=0} = \lim_{\sigma^2 \to 0} \frac{\partial \psi_2(1, \sigma^2; \delta)}{\partial \sigma^2}.$$

Note that if we set $m = 1/\sigma^2$, then from (A.6) we have

$$\frac{\partial \psi_2(1,\sigma^2;\delta)}{\partial \sigma^2} = \frac{4}{\delta}\left(1 - \frac{1}{2}\int_0^{\frac{\pi}{2}} \frac{\sigma^2}{(\sin^2\theta+\sigma^2)^{\frac{3}{2}}}\mathrm{d}\theta\right) = \frac{4}{\delta}\left(1-\frac{1}{2}\sqrt{\frac{m}{1+m}}E\left(\frac{m}{m+1}\right)\right).$$

It is then straightforward to use Lemma 1 to prove that

$$\lim_{m\to\infty}\frac{4}{\delta}\left(1-\frac{1}{2}\sqrt{\frac{m}{1+m}}E\left(\frac{m}{m+1}\right)\right) = \frac{2}{\delta}.$$

Hence, $\left.\frac{\partial \psi_2(\alpha,\sigma^2;\delta)}{\partial\sigma^2}\right|_{\alpha=1,\sigma^2=0} > 1$ for $\delta < 2$.

Part (ii): We first prove that the following equation has at least one solution for any $\alpha \in [0,1]$ and $\delta > 2$:

$$\sigma^2 = \psi_2(\alpha,\sigma^2;\delta), \quad \sigma^2 \in [0,1].$$

It is straightforward to verify that

$$\psi_2(\alpha,\sigma^2;\delta)|_{\sigma^2=0} = \frac{4}{\delta}(1-\alpha)^2 \geq 0. \tag{A.19}$$

We next prove our claim by proving the following:

$$\psi_2(\alpha,\sigma^2;\delta)|_{\sigma^2=1} < 1, \quad \forall \alpha\in[0,1] \text{ and } \delta > 2. \tag{A.20}$$

From (2.6b), we have

$$\psi_2(\alpha,\sigma^2;\delta)|_{\sigma^2=1} < 1 \Longleftrightarrow \underbrace{\int_0^{\frac{\pi}{2}} \frac{2\alpha^2\sin^2\theta+1}{(\alpha^2\sin^2\theta+1)^{\frac{1}{2}}}\mathrm{d}\theta - \alpha^2}_{g(\alpha^2)} > 2 - \frac{\delta}{4}. \tag{A.21}$$

We next show that $g(\alpha^2)$ in (A.21) is a concave function of α^2, and hence the minimum can only happen at either $\alpha = 0$ or $\alpha = 1$. The first two derivatives w.r.t. α^2 are given by:

$$\frac{\mathrm{d}g(\alpha^2)}{\mathrm{d}\alpha^2} = \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta\left(\alpha^2\sin^2\theta+\frac{3}{2}\right)}{(\alpha^2\sin^2\theta+1)^{\frac{3}{2}}}\mathrm{d}\theta - 1,$$

and

$$\frac{\mathrm{d}^2 g(\alpha^2)}{\mathrm{d}(\alpha^2)^2} = -\int_0^{\frac{\pi}{2}} \frac{\sin^4\theta\left(\frac{1}{2}\alpha^2\sin^2\theta+\frac{5}{4}\right)}{(\alpha^2\sin^2\theta+1)^{\frac{5}{2}}}\mathrm{d}\theta < 0.$$

The concavity of $g(\alpha^2)$ implies that its minimum happens at either $\alpha = 0$ or $\alpha = 1$. Hence, to prove (A.21), it suffices to prove that

$$g(0) = \frac{\pi}{2} > 2 - \frac{\delta}{4} \quad \text{and} \quad g(1) \approx 1.509 > 2 - \frac{\delta}{4},$$

which holds for $\delta > 2$. Hence, (A.21) holds. By combining (A.19) and (A.20) we conclude that $\psi_2(\alpha,\sigma^2;\delta)$ has at least one fixed point between $\sigma^2 = 0$ and $\sigma^2 = 1$. The next step is to prove the uniqueness of this fixed point. For the rest of the proof, we discuss two cases separately: a) $\delta > 4$ and b) $2 < \delta \leq 4$.

(a) $\delta > 4$. Define

$$\Psi_2(\alpha,\sigma^2;\delta) \triangleq \psi_2(\alpha,\sigma^2;\delta) - \sigma^2. \tag{A.22}$$

From (A.18), if $\delta > 4$, then $\frac{\partial\psi_2(\alpha,\sigma^2;\delta)}{\partial\sigma^2} < 1$, $\forall \sigma^2 > 0$. This means that $\Psi_2(\alpha,\sigma^2;\delta)$ defined in (A.22) is monotonically decreasing in $\sigma^2 > 0$. Hence, the solution to $\Psi_2(\alpha,\sigma^2;\delta) = 0$ is unique. Furthermore, the following property is a direct consequence of the monotonicity of $\Psi_2(\alpha,\sigma^2;\delta)$:

$$\Psi_2(\alpha,\sigma^2;\delta) < 0, \quad \forall 0 < \sigma^2 < F_2(\alpha), \tag{A.23a}$$

and

$$\Psi_2(\alpha,\sigma^2;\delta) > 0 > \sigma^2, \quad \forall F_2(\alpha) < \sigma^2 < 1, \tag{A.23b}$$

where $F_2(\alpha)$ denotes the solution to $\Psi_2(\alpha,\sigma^2;\delta) = 0$.

(b) $2 < \delta \le 4$. In this case, we will prove that there exists a threshold on σ^2, denoted as $\sigma_\star^2(\alpha;\delta)$ below, such that the following hold:

$$\frac{\partial \psi_2(\alpha,\sigma^2;\delta)}{\partial \sigma^2} < 1, \quad \forall \sigma^2 < \sigma_\star^2(\alpha;\delta) \quad \text{and} \quad \frac{\partial \psi_2(\alpha,\sigma^2;\delta)}{\partial \sigma^2} > 1, \quad \forall \sigma^2 \in (\sigma_\star^2(\alpha;\delta),\infty). \tag{A.24}$$

This means that $\Psi_2(\alpha,\sigma^2;\delta) = \psi_2(\alpha,\sigma^2;\delta) - \sigma^2$ is strictly decreasing on $\sigma^2 \in (0,\sigma_\star^2(\alpha;\delta))$ and increasing on $\sigma^2 \in (\sigma_\star^2(\alpha;\delta),\infty)$. Note that since we have proved that $\Psi_2(\alpha,\sigma^2;\delta)=0$ has at least one solution, we conclude that there exist exactly two solutions to $\Psi_2(\alpha,\sigma^2;\delta)=0$, one in $(0,\sigma_\star^2(\alpha;\delta))$ and the second in $(\sigma_\star^2(\alpha;\delta),\infty)$, if $\Psi_2(\alpha,\sigma^2;\delta)|_{\sigma^2=\sigma_\star^2(\alpha;\delta)} < 0$. This is the case since $\Psi_2(\alpha,\sigma^2;\delta)|_{\sigma^2=1} < 0$ (see (A.20)), and that $\Psi_2(\alpha,\sigma^2;\delta)|_{\sigma^2=1} < \Psi_2(\alpha,\sigma^2;\delta)|_{\sigma^2=\sigma_\star^2(\alpha;\delta)}$ (since the latter is the global minimum of $\Psi_2(\alpha,\sigma^2;\delta)$ in $\sigma^2 \in (0,\infty)$).

Also, it is easy to prove (A.23). In fact, the following holds:

$$\Psi_2(\alpha,\sigma^2;\delta) < 0, \quad \forall 0 < \sigma^2 < F_2(\alpha),$$

and

$$\Psi_2(\alpha,\sigma^2;\delta) > 0 > \sigma^2, \quad \forall F_2(\alpha) < \sigma^2 < \hat{F}_2(\alpha;\delta),$$

where $\hat{F}_2(\alpha;\delta) > 1$ denotes the larger solution to $\Psi_2(\alpha,\sigma^2;\delta)=0$. See Fig. 10 for an illustration.

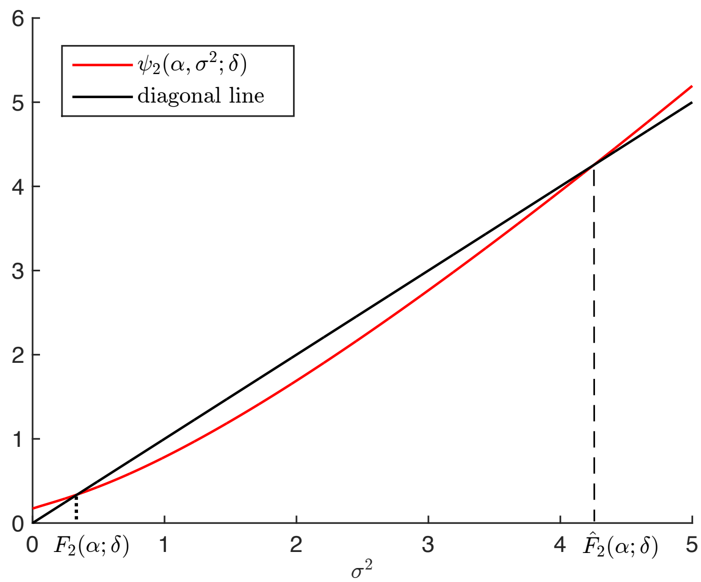

Figure 10: Plot of $\psi_2(\alpha,\sigma^2;\delta)$ for $\alpha=0.7$ and $\delta=2.1$.

From the above discussions, it remains to prove (A.24). To this end, it is more convenient to express (A.18) using elliptic integrals discussed in Section A.1:

$$\frac{\partial \psi_2(\alpha,\sigma^2;\delta)}{\partial \sigma^2} = \frac{4}{\delta}\left(1 - \frac{1}{2}\int_0^{\frac{\pi}{2}} \frac{\sigma^2}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{3}{2}}}\mathrm{d}\theta\right) \tag{A.25a}$$

$$= \frac{4}{\delta\alpha}\left(\alpha - \underbrace{\frac{1}{2\sqrt{1+s^2}}E\left(\frac{1}{1+s^2}\right)}_{f(s)}\right), \tag{A.25b}$$

where we introduced a new variable $s \triangleq \frac{\sigma}{\alpha}$ and the last step is derived using the identities in Lemma 2. Based on (A.25) we can now rewrite (A.24) as

$$f(s) > \alpha\left(1-\frac{\delta}{4}\right), \quad \forall s < \frac{\sigma_\star(\alpha;\delta)}{\alpha} \quad \text{and} \quad f(s) < \alpha\left(1-\frac{\delta}{4}\right), \quad \forall s \in \left(\frac{\sigma_\star(\alpha;\delta)}{\alpha},\infty\right). \tag{A.26}$$

To prove this, we first show that there exists s^* such that $f(s)$ is strictly increasing on $(0, s_*)$ and decreasing on (s_*, ∞), namely,

$$f'(s) > 0, \text{ for } s < s_*, \quad \text{and} \quad f'(s) < 0, \text{ for } s > s_*. \tag{A.27a}$$

s_* is in fact the unique solution to the following equation:

$$2E\left(\frac{1}{1+s_*^2}\right) = K\left(\frac{1}{1+s_*^2}\right). \tag{A.27b}$$

This can be seen from $f'(s)$ derived below:

$$\begin{aligned} f'(s) &= \frac{\mathrm{d}}{\mathrm{d}s}\frac{1}{2\sqrt{1+s^2}}E\left(\frac{1}{1+s^2}\right) \\ &= \frac{s}{2(1+s^2)^{\frac{3}{2}}}\left[K\left(\frac{1}{1+s^2}\right) - 2E\left(\frac{1}{1+s^2}\right)\right]. \end{aligned}$$

Further noting that $E(\cdot)$ is strictly decreasing in $(0,1)$ while $K(\cdot)$ is increasing, we proved (A.27).

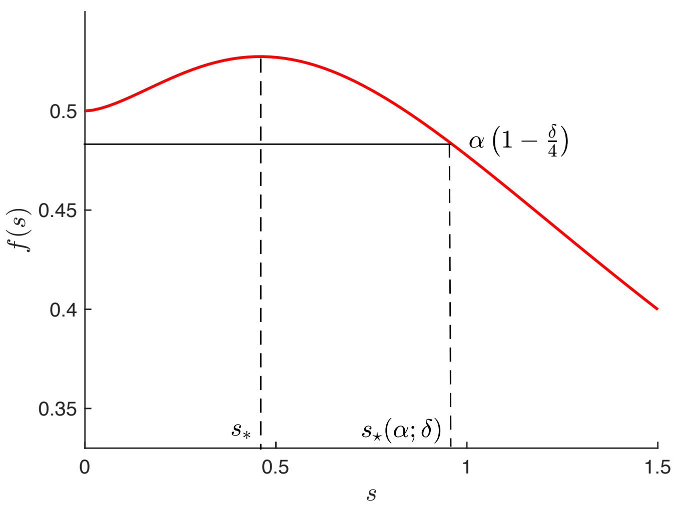

Figure 11: Illustration of $f(s)$.

Based on the above discussions, we can finally turn to the proof of (A.26). From (A.25b), it is straightforward to verify that $f(0) = \frac{1}{2}$. Therefore, when $\delta > 2$, we have

$$\alpha\left(1-\frac{\delta}{4}\right) \le 1 - \frac{\delta}{4} < \frac{1}{2} = f(0), \quad \forall \delta > 2 \text{ and } 0 \le \alpha \le 1.$$

Hence, the following equation admits a unique solution (denoted as $s_\star(\alpha;\delta)$ below):

$$f(s) = \alpha\left(1-\frac{\delta}{4}\right), \quad \forall \delta > 2 \text{ and } 0 \le \alpha \le 1.$$

See Fig. 11 for an illustration. Also, from our above discussions on the monotonicity of $f(s)$ it is straightforward to show that

$$f(s) > \alpha\left(1-\frac{\delta}{4}\right), \quad \forall s < s_\star(\alpha;\delta) \quad \text{and} \quad f(s) < \alpha\left(1-\frac{\delta}{4}\right), \forall s \in (s_\star(\alpha;\delta), \infty),$$

which proves (A.26) by setting $\sigma_\star(\alpha;\delta) \triangleq \alpha \cdot s_\star(\alpha;\delta)$. This proves (A.24), which completes the proof.

Part (iii): We will prove a stronger result: $\psi_2 \le 4/\delta$. From (2.6b), $\psi_2(\alpha, \sigma^2; \delta) \le 4/\delta$ is equivalent to

$$\alpha^2 + \sigma^2 - \int_0^{\frac{\pi}{2}} \frac{2\alpha^2 \sin^2\theta + \sigma^2}{(\alpha^2\sin^2\theta + \sigma^2)^{\frac{1}{2}}} \mathrm{d}\theta \le 0,$$

which can be further reformulated as

$$\alpha^2 \le \int_0^{\frac{\pi}{2}} \frac{2\alpha^2 \sin^2\theta}{(\alpha^2\sin^2\theta + \sigma^2)^{\frac{1}{2}}} \mathrm{d}\theta + \sigma^2 \left(\int_0^{\frac{\pi}{2}} \frac{1}{(\alpha^2\sin^2\theta + \sigma^2)^{\frac{1}{2}}} \mathrm{d}\theta - 1 \right). \tag{A.28}$$

For $0 \le \alpha \le 1$ and $\sigma^2 \le \sigma^2_{\max}$ we have

$$\begin{aligned}
\int_0^{\frac{\pi}{2}} \frac{1}{(\alpha^2\sin^2\theta + \sigma^2)^{\frac{1}{2}}} \mathrm{d}\theta &\ge \int_0^{\frac{\pi}{2}} \frac{1}{\left(\sin^2\theta + \sigma^2_{\max}\right)^{\frac{1}{2}}} \mathrm{d}\theta, \\
&\overset{(a)}{=} \int_0^{\frac{\pi}{2}} \frac{1}{\left(\sin^2\theta + \frac{4}{\delta_{\mathrm{AMP}}}\right)^{\frac{1}{2}}} \mathrm{d}\theta \\
&\approx 1.09 > 1,
\end{aligned} \tag{A.29}$$

where step (a) from $\sigma^2_{\max} = \max\{1, 4/\delta\} \ge \max\{1, 4/\delta_{\mathrm{AMP}}\} = 4/\delta_{\mathrm{AMP}} \approx 1.6$. Due to (A.29), to prove (A.28), it suffices to prove

$$\alpha^2 \le \int_0^{\frac{\pi}{2}} \frac{2\alpha^2 \sin^2\theta}{(\alpha^2\sin^2\theta + \sigma^2)^{\frac{1}{2}}} \mathrm{d}\theta,$$

or

$$1 \le \int_0^{\frac{\pi}{2}} \frac{2 \sin^2\theta}{(\alpha^2\sin^2\theta + \sigma^2)^{\frac{1}{2}}} \mathrm{d}\theta,$$

which, similar to (A.29), can be proved by

$$\int_0^{\frac{\pi}{2}} \frac{2 \sin^2\theta}{(\alpha^2\sin^2\theta + \sigma^2)^{\frac{1}{2}}} \mathrm{d}\theta \ge \int_0^{\frac{\pi}{2}} \frac{2 \sin^2\theta}{\left(\sin^2\theta + \frac{4}{\delta_{\mathrm{AMP}}}\right)^{\frac{1}{2}}} \mathrm{d}\theta \approx 1.02 > 1.$$

Part (iv): We bound the partial derivative of $\psi_2(\alpha, \sigma^2; \delta)$ for $\sigma^2 \in [0, \sigma^2_{\max}]$ as:

$$\begin{aligned}
\frac{\psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} &= \frac{4}{\delta} \left(1 - \frac{1}{2} \int_0^{\frac{\pi}{2}} \frac{\sigma^2}{(\alpha^2\sin^2\theta + \sigma^2)^{\frac{3}{2}}} \mathrm{d}\theta \right) \\
&\overset{(a)}{\le} \frac{4}{\delta} \left(1 - \frac{1}{2} \int_0^{\frac{\pi}{2}} \frac{\sigma^2}{(\theta^2 + \sigma^2)^{\frac{3}{2}}} \mathrm{d}\theta \right) \\
&\overset{(b)}{=} \frac{4}{\delta} \left(1 - \frac{1}{2} \int_0^{\frac{\pi}{2\sigma}} \frac{1}{(\tilde{\theta}^2 + 1)^{\frac{3}{2}}} \mathrm{d}\tilde{\theta} \right) \\
&\overset{(c)}{\le} \frac{4}{\delta_{\mathrm{AMP}}} \left(1 - \frac{1}{2} \int_0^{\frac{\pi}{2\sqrt{\frac{4}{\delta_{\mathrm{AMP}}}}}} \frac{1}{(\tilde{\theta}^2 + 1)^{\frac{3}{2}}} \mathrm{d}\tilde{\theta} \right) \\
&\approx 0.98 < 1,
\end{aligned} \tag{A.30}$$

where step (a) follows from the constraint $0 \le \alpha \le 1$ and the inequality $\sin\theta \le \theta$; (b) is due to the variable change $\tilde{\theta} = \theta/\sigma$; (c) is a consequence of the constraint $\sigma^2 \le \sigma^2_{\max} = \max\{1, 4/\delta\} \le \max\{1, 4/\delta_{\mathrm{AMP}}\} = 4/\delta_{\mathrm{AMP}}$. As a result of (A.30), $\Psi_2(\alpha, \sigma^2; \delta) = \psi_2(\alpha, \sigma^2; \delta) - \sigma^2$ is decreasing in $\sigma^2 \in [0, \sigma^2_{\max}]$. It is easy to verify that $\psi_2(0, \alpha; \delta) \ge 0$ for $\alpha \in [0, 1]$. Further, Lemma 9 (iii) implies that

$$\psi_2(\sigma^2_{\max}, \alpha; \delta) - \sigma^2_{\max} \le 0.$$

Hence, there exists a unique solution (which we denote as $F_2(\alpha)$) to the following equation:

$$\psi_2(\sigma,\alpha;\delta)=\sigma^2,\quad 0\le\sigma^2\le\sigma^2_{\max}.$$

Finally, the property in (A.16) is a direct consequence of the fact that $\Psi_2(\alpha,\sigma^2;\delta)=\psi_2(\alpha,\sigma^2;\delta)-\sigma^2$ is a decreasing function of $\sigma^2\le\sigma^2_{\max}$.

Part (v): In (A.25), we have derived the following:

$$\frac{\psi_2(\alpha,\sigma^2;\delta)}{\partial\sigma^2}=\frac{4}{\delta\alpha}\left(\alpha-f(s)\right),$$

where $s\triangleq\frac{\sigma}{\alpha}$. From (A.25b), we see that $\psi_2(\alpha,\sigma^2;\delta)$ is an increasing function of σ^2 if the following holds:

$$\alpha>f(s).$$

Further, (A.27) implies that the maximum of $f(s)$ happens at s_*, i.e.,

$$\max_{s>0}\ f(s)=\frac{1}{2\sqrt{1+s_*^2}}E\left(\frac{1}{1+s_*^2}\right)\triangleq\alpha_*, \tag{A.31}$$

where s_*^2 is the unique solution to

$$2E\left(\frac{1}{1+s_*^2}\right)=K\left(\frac{1}{1+s_*^2}\right). \tag{A.32}$$

Clearly, $\alpha>\alpha_*$ immediately implies $\alpha>f(s)$, which further guarantees that $\psi_2(\alpha,\sigma^2;\delta)$ is monotonically increasing on $\sigma^2>0$. Finally, the strong global attractiveness of $F_2(\alpha)$ is a direct consequence of part (iv) of this lemma together with the monotonicity of ψ_2. $\square$

A.3.3 Properties of F_1 and F_2

In this section we derive the main properties of the functions F_1 and F_2 introduced in Section A.3.1. These properties play major roles in the results of the paper.

Lemma 10. *The following hold for $F_1(\sigma^2)$ and $F_2(\alpha;\delta)$ (for $\delta>2$):*

(i) $F_1(0)=1$ and $\lim_{\sigma^2\to\frac{\pi^2}{16}^-}F_1(\sigma^2)=0$. Further, by choosing $F_1(\frac{\pi^2}{16})=0$, we have $F_1(\sigma^2)$ is continuous on $\left[0,\frac{\pi^2}{16}\right]$ and strictly decreasing in $\left(0,\frac{\pi^2}{16}\right)$;

(ii) F_2 is a continuous function of $\alpha\in[0,1]$ and $\delta\in(2,\infty)$. $F_2(1;\delta)=0$, and $F_2(0;\delta)=\left(\frac{-\pi+\sqrt{\pi^2+4(\delta-4)}}{\delta-4}\right)^2$ for $\delta\neq4$ and $F_2(0;4)=4/\pi^2$.

Proof. Part (i): We first verify $F_1(0)=1$ and $\lim_{\sigma^2\to\frac{\pi^2}{16}^-}F_1(\sigma^2)=0$. First, $F_1(0)=1$ can be seen from the following facts: (a) $\psi_1(\alpha,0)=1$ for $\alpha>0$, see (2.6a); and (b) By definition, $F_1(0)$ is the non-zero solution to $\alpha=\psi_1(\alpha,0)$. Then, by Lemma 8 (iii) and continunity of ψ_1, we know F_1 is continuous on $[0,\frac{\pi^2}{16})$, and further $\lim_{\sigma^2\to\frac{\pi^2}{16}^-}F_1(\sigma^2)=0$ since $\sigma^2=\frac{\pi^2}{16}$ corresponds to a case where the non-negative solution to $\psi_1(\alpha,\sigma^2)=\alpha$ decreases to zero. Next, we prove the monotonicity of F_1. Note that

$$F_1(\sigma^2)=\psi_1(F_1(\sigma^2),\sigma^2),$$

Differentiation w.r.t. σ^2 yields

$$F_1'(\sigma^2)=\partial_2\psi_1(F_1(\sigma^2),\sigma^2)+\partial_1\psi_1(F_1(\sigma^2),\sigma^2)\cdot F_1'(\sigma^2),$$

where $\partial_2\psi_1(F_1(\sigma^2),\sigma^2)\triangleq\frac{\partial\psi_1(\alpha,\sigma^2)}{\partial\sigma^2}\Big|_{\alpha=F_1(\sigma^2)}$ and $\partial_1\psi_1(F_1(\sigma^2),\sigma^2)\triangleq\frac{\partial\psi_1(\alpha,\sigma^2)}{\partial\alpha}\Big|_{\alpha=F_1(\sigma^2)}$. Hence,

$$\left[1-\partial_1\psi_1(F_1(\sigma^2),\sigma^2)\right]\cdot F_1'(\sigma^2)=\partial_2\psi_1(F_1(\sigma^2),\sigma^2). \tag{A.33}$$

We have proved in (A.14) that $\frac{\partial\psi_1(\alpha,\sigma^2)}{\partial\alpha}\Big|_{\alpha=0} < 1$ when $\sigma^2 < \frac{\pi^2}{16}$. Together with the concavity of ψ_1 w.r.t. α (cf. Lemma 8 (i)), we have

$$\frac{\partial\psi_1(\alpha,\sigma^2)}{\partial\alpha}\Big|_{\alpha=F_1(\sigma^2)} < \frac{\partial\psi_1(\alpha,\sigma^2)}{\partial\alpha}\Big|_{\alpha=0} < 1, \quad \forall\sigma^2 < \frac{\pi^2}{16}. \tag{A.34}$$

Further, from (2.6a), it is straightforward to see that ψ_1 is a strictly decreasing function of σ^2, and thus

$$\partial_2\psi_1(F_1(\sigma^2),\sigma^2) = \frac{\partial\psi_1(\alpha,\sigma^2)}{\partial\alpha}\Big|_{\alpha=F_1(\sigma^2)} < 0. \tag{A.35}$$

Substituting (A.34) and (A.35) into (A.33), we obtain

$$F_1'(\sigma^2) < 0, \quad \forall\sigma^2 < \frac{\pi^2}{16}.$$

Proof of (ii): By Lemma 9 (ii) and continuity of ψ_2, it is straightforward to check that F_2 is continuous. Moreover, we have proved that $\sigma^2 = F_2(\alpha;\delta)$ is the unique solution to the following equation (for $\delta > 2$):

$$\sigma^2 = \frac{4}{\delta}\left(\alpha^2+\sigma^2+1-\int_0^{\frac{\pi}{2}}\frac{2\alpha^2\sin^2\theta+\sigma^2}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{1}{2}}}\mathrm{d}\theta\right), \quad \sigma^2\in[0,1]. \tag{A.36}$$

When $\alpha = 0$, (A.36) reduces

$$\sigma^2 = \frac{4}{\delta}\left(\sigma^2+1-\frac{\pi}{2}\sigma\right), \quad \sigma^2\in[0,1],$$

which has two possible solutions (for $\delta\neq 4$):

$$\sigma_1 = \frac{-\pi+\sqrt{\pi^2+4(\delta-4)}}{\delta-4} \quad \text{and} \quad \sigma_2 = \frac{-\pi-\sqrt{\pi^2+4(\delta-4)}}{\delta-4}.$$

(For the special case $\delta = 4$, $\sigma_1 = 2/\pi$.) However, σ_2 is invalid due to our constraint $0<\sigma^2<1$. This can be seen as follows. First, $\sigma_2 < 0$ for $\delta > 4$ and hence invalid. When $2<\delta<4$, we have

$$\sigma_2 = \frac{\pi+\sqrt{\pi^2-4(4-\delta)}}{4-\delta} > \frac{\pi}{4-\delta} > 1.$$

Hence, $F_2(0;\delta) = \sigma_1$. When $\alpha = 1$, (A.36) becomes:

$$\sigma^2 = \frac{4}{\delta}\left(2+\sigma^2-\int_0^{\frac{\pi}{2}}\frac{2\sin^2\theta+\sigma^2}{(\sin^2\theta+\sigma^2)^{\frac{1}{2}}}\mathrm{d}\theta\right), \quad \sigma^2\in[0,1].$$

It is straightforward to verify that $\sigma^2 = 0$ is a solution. Also, from Lemma 9 (ii), $\sigma^2 = 0$ is a also the unique solution. Hence, $F_2(1;\delta) = 0$.

□

A.3.4 Proof of Lemma 4

In Lemma 9, we have proved that $F_2(\alpha;\delta)$ is the unique globally attracting fixed point of ψ_2 in $\sigma^2\in[0,1]$ (for $\delta > 2$), and from (A.15) we have

$$\sigma^2 > F_2(\alpha;\delta) \Longleftrightarrow \psi_2(\alpha,\sigma^2;\delta) < \sigma^2, \quad \sigma^2\in[0,1]. \tag{A.37}$$

Here, our objective is to prove that $F_1^{-1}(\alpha) < F_2(\alpha;\delta)$ holds for any $\alpha\in(0,1)$ when $\delta\geq\delta_{\mathrm{AMP}}$. From (A.37) and noting that $F_1^{-1}(\alpha)\leq\pi^2/16<1$ (from Lemma 10), our problem can be reformulated as proving the following inequality (for $\delta>\delta_{\mathrm{AMP}}$):

$$\psi_2(\alpha,F_1^{-1}(\alpha);\delta) < F_1^{-1}(\alpha), \quad \forall\alpha\in(0,1). \tag{A.38}$$

Since $\psi_2(\alpha, F_1^{-1}(\alpha); \delta)$ is a strictly decreasing function of δ (see (2.6b)), it suffices to prove that (A.38) holds for $\delta = \delta_{\text{AMP}}$:

$$\psi_2(\alpha, F_1^{-1}(\alpha); \delta_{\text{AMP}}) < F_1^{-1}(\alpha), \quad \forall \alpha \in (0,1). \tag{A.39}$$

We now make some variable changes for (A.39). From (2.6a), ψ_1 in can be rewritten as the following for $\alpha > 0$:

$$\psi_1(\alpha, \sigma^2) = \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(\sin^2\theta + \frac{\sigma^2}{\alpha^2}\right)^{\frac{1}{2}}} \mathrm{d}\theta.$$

By definition, $F_1(\sigma^2)$ is the solution to $\alpha = \psi_1(\alpha, \sigma^2)$, and hence the following holds:

$$\alpha = \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(\sin^2\theta + \frac{F_1^{-1}(\alpha)}{\alpha^2}\right)^{\frac{1}{2}}} \mathrm{d}\theta.$$

At this point, it is more convenient to make the following variable change:

$$s \triangleq \frac{\sqrt{F_1^{-1}(\alpha)}}{\alpha}, \tag{A.40}$$

from which we get

$$\alpha = \phi_1(s) \triangleq \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(\sin^2\theta + s^2\right)^{\frac{1}{2}}} \mathrm{d}\theta. \tag{A.41}$$

Notice that $\phi_1 : \mathbb{R}_+ \mapsto [0,1]$ is a monotonically decreasing function, and it defines a one-to-one map between α and s. From the above definitions, we have

$$F_1^{-1}(\alpha) = s^2\alpha^2 = s^2\phi_1^2(s), \tag{A.42}$$

where the first equality is from (A.40) and the second step from (A.41). Using the relationship in (A.42), we can reformulate the inequality in (A.39) into the following equivalent form:

$$\psi_2(\phi_1(s), s^2\phi_1^2(s); \delta_{\text{AMP}}) < s^2\phi_1^2(s), \quad \forall s > 0. \tag{A.43}$$

Substituting (A.41) and (2.6b) into (A.43) and after some manipulations, we can finally write our objective as:

$$\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(\sin^2\theta + s^2)^{\frac{1}{2}}} d\theta \cdot \int_0^{\frac{\pi}{2}} \frac{(1-\gamma s^2)\sin^2\theta + s^2}{(\sin^2\theta + s^2)^{\frac{1}{2}}} d\theta > 1, \quad \forall s > 0. \tag{A.44}$$

where we defined

$$\gamma \triangleq 1 - \frac{\delta_{\text{AMP}}}{4} = 2 - \frac{16}{\pi^2}. \tag{A.45}$$

In the next two subsections, we prove (A.44) for $s^2 > 0.07$ and $s^2 \leq 0.07$ using different techniques.

(i) Case I: We make another variable change:

$$t \triangleq \frac{1}{s^2}.$$

Using the variable t, we can rewrite (A.44) into the following:

$$G(t) \triangleq \frac{g_1(t)}{g_2(t)} - \frac{1}{g_2^2(t)} \geq \gamma, \quad \forall t \in [0, 14.3). \tag{A.46a}$$

where γ is defined in (A.45), and

$$g_1(t) \triangleq \int_0^{\frac{\pi}{2}} (1 + t\sin^2\theta)^{\frac{1}{2}} d\theta, \tag{A.46b}$$

$$g_2(t) \triangleq \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1 + t\sin^2\theta)^{\frac{1}{2}}} d\theta. \tag{A.46c}$$

Notice that if we could prove (A.46a) for $t < 14.3$, we would have proved (A.44) for $s^2 > 0.07$, since $14.3 > 1/0.07 \approx 14.28$. For the ease of later discussions, we define

$$
\begin{aligned}
g_3(t) &\triangleq \int_0^{\frac{\pi}{2}} \frac{\sin^4\theta}{(1+t\sin^2\theta)^{\frac{3}{2}}} d\theta, \\
g_4(t) &\triangleq \int_0^{\frac{\pi}{2}} \frac{\sin^6\theta}{(1+t\sin^2\theta)^{\frac{5}{2}}} d\theta.
\end{aligned}
$$

The following identities related to $\{g_1(t), g_2(t), g_3(t), g_4(t)\}$ will be used in our proof:

$$
\begin{aligned}
g_1'(t) &= \frac{1}{2} g_2(t), \\
g_2'(t) &= -\frac{1}{2} g_3(t), \\
g_3'(t) &= -\frac{3}{2} g_4(t).
\end{aligned}
\tag{A.47}
$$

We now prove (A.46a). First, it is straightforward to verify that equality holds for (A.46a) at $t = 0$, i.e.,

$$
G(0) = \gamma. \tag{A.48}
$$

Hence, to prove that $G(t) \geq \gamma$ for $t \in [0, 14.3)$, it is sufficient to prove that $G(t)$ is an increasing function of t on $t \in [0, 14.3)$. To this end, we calculate the derivative of $G(t)$:

$$
\begin{aligned}
G'(t) &= \frac{g_1'(t)g_2(t) - g_1(t)g_2'(t)}{g_2^2(t)} - \left(\frac{-2g_2'(t)}{g_2^3(t)}\right) \\
&\overset{(a)}{=} \frac{\frac{1}{2}g_2^2(t) + \frac{1}{2}g_1(t)g_3(t)}{g_2^2(t)} - \frac{g_3(t)}{g_2^3(t)} \\
&= 1 + \frac{1}{2}\frac{g_1(t)g_3(t)}{g_2^2(t)} - \frac{g_3(t)}{g_2^3(t)} \\
&= \frac{1}{2}\frac{g_3(t)}{g_2^3(t)}\Bigg(\underbrace{\frac{g_2^3(t)}{g_3(t)}}_{G_1(t)} + \underbrace{g_1(t)g_2(t)}_{G_2(t)} - 2\Bigg),
\end{aligned}
$$

where step (a) follows from the identities listed in (A.47). Since $g_3(t) > 0$, we have

$$
G'(t) > 0 \Longleftrightarrow G_1(t) + G_2(t) - 2 > 0.
$$

It remains to prove that $G_1(t) + G_2(t) - 2 > 0$ for $t < 14.3$. Our numerical results suggest that $G_1(t) + G_2(t)$ is a monotonically decreasing function for $t > 0$, and $G_1(t) + G_2(t) \to 2$ as $t \to \infty$. However, directly proving the monotonicity of $G_1(t) + G_2(t)$ seems to be quite complicated. We use a different strategy here. We will prove that (at the end of this section)

- $G_1(t)$ is monotonically increasing;
- $G_2(t)$ is monotonically decreasing.

As a consequence, the following hold true for any $c_2 > c_1 > 0$:

$$
G_1(t) + G_2(t) - 2 \geq G_1(c_1) + G_2(c_2) - 2, \quad \forall t \in [c_1, c_2].
$$

Hence, if we verify that $G_1(c_1) + G_2(c_2) - 2 > 0$, we will be proving the following:

$$
G_1(t) + G_2(t) - 2 > 0, \quad \forall t \in [c_1, c_2].
$$

To this end, we verify that $G_1(c_1) + G_2(c_2) - 2 > 0$ hold for a sequence of c_1 and c_2: $[c_1, c_2] = [0, 0.49]$, $[c_1, c_2] = [0.49, 1.08]$, $[c_1, c_2] = [1.08, 1.78]$, $[c_1, c_2] = [1.78, 2.56]$, $[c_1, c_2] = [2.56, 3.47]$, $[c_1, c_2] =$

$[3.47, 4.47]$, $[c_1, c_2] = [4.47, 5.56]$, $[c_1, c_2] = [5.56, 6.77]$, $[c_1, c_2] = [6.67, 8.08]$, $[c_1, c_2] = [8.08, 9.5]$, $[c_1, c_2] = [9.5, 11]$, $[c_1, c_2] = [11, 12.6]$, $[c_1, c_2] = [12.6, 14.3]$. Combining all the above results proves

$$G_1(t) + G_2(t) - 2 > 0, \quad \forall t \in [0, 14.3].$$

From the above discussions, it only remains to prove the monotonicity of $G_1(t)$ and $G_2(t)$. Consider $G_1(t)$ first:

$$\begin{aligned} G_1'(t) &= \left(\frac{g_2^3(t)}{g_3(t)}\right)' \\ &= \frac{3g_2^2(t)g_2'(t)g_3(t) - g_2^3(t)g_3'(t)}{g_3^2(t)} \\ &= \frac{-\frac{3}{2}g_2^2(t)g_3^2(t) + \frac{3}{2}g_2^3(t)g_4(t)}{g_3^2(t)} \\ &= -\frac{3}{2}g_2^2(t) + \frac{3}{2}\frac{g_2^3(t)g_4(t)}{g_3^2(t)} \\ &= \frac{3}{2}\frac{g_2^2(t)}{g_3^2(t)} \cdot [-g_3^2(t) + g_2(t)g_4(t)]. \end{aligned} \tag{A.49}$$

Applying the Cauchy-Schwarz inequality yields:

$$\begin{aligned} g_2(t)g_4(t) &= \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+t\sin^2\theta)^{\frac{1}{2}}}d\theta \cdot \int_0^{\frac{\pi}{2}} \frac{\sin^6\theta}{(1+t\sin^2\theta)^{\frac{5}{2}}}d\theta \\ &\geq \left(\int_0^{\frac{\pi}{2}} \frac{\sin^4\theta}{(1+t\sin^2\theta)^{\frac{3}{2}}}d\theta\right)^2 \\ &= g_3^2(t). \end{aligned} \tag{A.50}$$

Combining (A.49) and (A.50), we proved that $G_1'(t) \geq 0$, and therefore $G_1(t)$ is monotonically increasing. For $G_2(t)$, we have

$$\begin{aligned} G_2'(t) &= g_1'(t)g_2(t) + g_1(t)g_2'(t) \\ &= \frac{1}{2}g_2^2(t) + g_1(t)\left(-\frac{1}{2}g_3(t)\right) \\ &= \frac{1}{2}[g_2^2(t) - g_1(t)g_3(t)]. \end{aligned}$$

Again, using Cauchy-Schwarz we have

$$\begin{aligned} g_1(t)g_3(t) &= \int_0^{\frac{\pi}{2}} (1+t\sin^2\theta)^{\frac{1}{2}}d\theta \cdot \int_0^{\frac{\pi}{2}} \frac{\sin^4\theta}{(1+t\sin^2\theta)^{\frac{3}{2}}}d\theta \\ &\geq \left(\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(1+t\sin^2\theta)^{\frac{1}{2}}}d\theta\right)^2 \\ &= g_2^2(t). \end{aligned}$$

Combining the previous two equations leads to $G_2'(t) \geq 0$, which completes our proof.

(ii) Case II: We next prove (A.44) for $s^2 \leq 0.07$, which is based on a different strategy. Some manipulations of the RHS of (A.44) yields:

$$\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(\sin^2\theta + s^2)^{\frac{1}{2}}}d\theta \cdot \int_0^{\frac{\pi}{2}} \frac{\left(1-\gamma s^2\right)\sin^2\theta + s^2}{(\sin^2\theta + s^2)^{\frac{1}{2}}}d\theta = \frac{E(x)T(x)}{x} - \frac{\gamma(1-x)T^2(x)}{x^2}, \tag{A.51a}$$

where $E(\cdot)$, $K(\cdot)$ and $T(\cdot)$ are elliptic integrals defined in (A.1), γ is a constant defined in (A.45), and x is a new variable:

$$x \triangleq \frac{1}{1+s^2}. \tag{A.51b}$$

From our reformulation in (A.51), the inequality in (A.44) for $s^2 < 0.07$ becomes

$$\frac{E(x)T(x)}{x} - \gamma\frac{(1-x)T^2(x)}{x^2} > 1, \quad x \in [0.93, 1). \tag{A.52}$$

Note that $0.93 < 1/(1+0.07)$ and thus proving the above inequality for $x \in [0.93, 1)$ is sufficient to prove the original inequality for $s^2 \leq 0.07$ (note that $x \triangleq 1/(1+s^2)$, see (A.51b)).

With some further calculations, (A.52) can be reformulated as

$$\frac{x}{T^2(x)}\frac{E(x)T(x)-x}{(1-x)} > \gamma, \quad x \in [0.93, 1). \tag{A.53}$$

The following inequality is due to Eqn. (1) in Anderson & Vamanamurthy (1985)

$$T(x) < x < 1, \quad \forall x \in (0, 1).$$

Hence,

$$\frac{x}{T^2(x)}\frac{E(x)T(x)-x}{(1-x)} > \frac{E(x)T(x)-x}{1-x}, \quad \forall x \in (0, 1),$$

and to prove (A.53) it suffices to prove the following

$$\frac{E(x)T(x)-x}{1-x} > \gamma, \quad \forall x \in [0.93, 1). \tag{A.54}$$

To this end, we will prove that the LHS of (A.54) is a strictly increasing function of $x \in [0.93, 1)$. If this is true, we would have

$$\frac{E(x)T(x)-x}{1-x} > \frac{E(x)T(x)-x}{1-x}|_{x=0.93} \approx 0.385 > \gamma = 2 - \frac{16}{\pi^2} \approx 0.3789, \quad \forall x \in [0.93, 1).$$

We next prove the monotonicity of $\frac{E(x)T(x)-x}{1-x}$. From the identities in Lemma 1, we derive the following

$$[E(x)T(x)-x]' = \frac{E^2(x) - 2(1-x)E(x)K(x) + (1-x)K^2(x)}{2x} - 1.$$

Hence, to prove that $\frac{E(x)T(x)-x}{1-x}$ is monotonically increasing, it is sufficient to prove the following inequality:

$$\left(\frac{E^2(x) - 2(1-x)E(x)K(x) + (1-x)K^2(x)}{2x} - 1\right)(1-x) - [E(x)T(x)-x](-1) > 0. \tag{A.55}$$

Now, substituting $T(x) = E(x) - (1-x)K(x)$ into (A.55) and after some manipulations, we finally reformulate the inequality to be proved into the following form:

$$T(x)^2 > 2x - xE^2(x).$$

It can be verified that equality holds at $x = 1$. We next prove that $T(x)^2 + xE(x)^2 - 2x$ is monotonically decreasing on $[0.93, 1)$. We differentiate once more:

$$(T(x)^2 + xE(x)^2 - 2x)' = 2E(x)^2 - (1-x)K(x)^2 - 2.$$

Our problem boils down to proving $2E(x)^2-(1-x)K(x)^2-2<0$ for $x\in[0.93,1)$. We can verify that $2E(x)^2-(1-x)K(x)^2-2=0$ holds at $x=1$. We finish by showing that $2E(x)^2-(1-x)K(x)^2-2$ is monotonically increasing in $x\in[0.93,1)$. To this end, we differentiate again:

$$\begin{aligned}
[2E(x)^2-(1-x)K(x)^2-2]' &= \frac{K(x)^2-3E(x)K(x)+2E(x)^2}{x} \\
&= \frac{\left[K(x)-\frac{3}{2}E(x)\right]^2-\frac{1}{2}E(x)^2}{x}.
\end{aligned} \tag{A.56}$$

We note that $K(x)-\left(\frac{3}{2}+\frac{1}{\sqrt{2}}\right)E(x)$ is a monotonically increasing function in (0,1) since $K(x)$ is monotonically increasing and $E(x)$ is monotonically decreasing. We verify that $K(x)-\left(\frac{3}{2}+\frac{1}{\sqrt{2}}\right)E(x)>0$ when $x\geq 0.93$. Hence,

$$K(x)-\left(\frac{3}{2}+\frac{1}{\sqrt{2}}\right)E(x)>0,\quad \forall x\in[0.93,1),$$

and therefore

$$\left(K(x)-\frac{3}{2}E(x)\right)^2>\frac{1}{2}E(x)^2,\quad \forall x\in[0.93,1). \tag{A.57}$$

Substituting (A.57) into (A.56), we prove that $[2E(x)^2-(1-x)K(x)^2-2]'>0$ for $x\in[0.93,1)$, which completes the proof.

A.3.5 Proof of Lemma 6

First, we introduce a function that will be crucial for our proof.

Definition 4. *Define*

$$L(\alpha;\delta)\triangleq\frac{4}{\delta}\left(1-\frac{\phi_2^2(\phi_1^{-1}(\alpha))}{4\left[1+(\phi_1^{-1}(\alpha))^2\right]}\right),\quad \alpha\in(0,1), \tag{A.58}$$

where $\phi_1:\mathbb{R}_+\mapsto[0,1]$ and $\phi_2:\mathbb{R}_+\mapsto\mathbb{R}_+$ below:

$$\phi_1(s)\triangleq\int_0^{\frac{\pi}{2}}\frac{\sin^2\theta}{\left(\sin^2\theta+s^2\right)^{\frac{1}{2}}}\mathrm{d}\theta, \tag{A.59a}$$

$$\phi_2(s)\triangleq\int_0^{\frac{\pi}{2}}\frac{2\sin^2\theta+s^2}{\left(\sin^2\theta+s^2\right)^{\frac{1}{2}}}\mathrm{d}\theta, \tag{A.59b}$$

where ϕ_1^{-1} is the inverse functions of ϕ_1. The existence of ϕ_1^{-1} follows from its monotonicity, which can be seen from its definition.

In the following, we list some preliminary properties of $L(\alpha;\delta)$. The main proof for Lemma 6 comes afterwards.

- **Preliminaries:**

  The following lemma helps us clarify the importance of L in the analysis of the dynamics of SE:

  Lemma 11. *For any $\alpha>0$, $\sigma^2>0$ and $\delta>0$, the following holds:*

  $$L\left[\psi_1(\alpha,\sigma^2);\delta\right]\leq\psi_2(\alpha,\sigma^2;\delta), \tag{A.60}$$

  where ψ_1 and ψ_2 are the SE maps defined in (2.6)*, and $L(\alpha;\delta)$ is defined in* (A.58)*.*

  Proof. Define $\mathcal{X}\triangleq\{(\alpha,\sigma^2)|\alpha>0,\sigma^2>0\}$. Let $\mathcal{Y}$ be the image of $\mathcal{X}$ under the SE map in (2.6). We will prove that the following holds for an arbitrary $C\in[0,1]$:

  $$L\left(C;\delta\right)=\min_{(\hat{\alpha},\hat{\sigma}^2)\in\mathcal{X}}\psi_2(\hat{\alpha},\hat{\sigma}^2;\delta), \tag{A.61}$$

where $(\hat{\alpha}, \hat{\sigma}^2)$ satisfies the constraint

$$\psi_1(\hat{\alpha}, \hat{\sigma}^2) = C.$$

If (A.61) holds, we would have proved (A.60). To see this, consider arbitrary (α, σ^2) such that $\psi_1(\alpha, \sigma^2) = C$. Then, we have

$$L\left[\psi_1(\alpha, \sigma^2); \delta\right] \overset{(a)}{=} \min_{(\hat{\alpha}, \hat{\sigma}^2)} \ \psi_2(\hat{\alpha}, \hat{\sigma}^2; \delta) \overset{(b)}{\leq} \psi_2(\alpha, \sigma^2; \delta),$$

where step (a) follows from (A.61) and $\psi_1(\alpha, \sigma^2) = C$, and step (b) holds since the choice $\hat{\alpha} = \alpha$ and $\hat{\sigma}^2 = \sigma^2$ is feasible for the constraint $\psi_1(\hat{\alpha}, \hat{\sigma}^2) = \psi_1(\alpha, \sigma^2)$. This is precisely (A.60).

We now prove (A.61). From (2.6a) we have

$$\psi_1(\alpha, \sigma^2) = \int_0^{\pi/2} \frac{\alpha \sin^2\theta}{(\alpha^2 \sin^2\theta + \sigma^2)^{1/2}} d\theta.$$

Furthermore, from the definition of ϕ_1 in (A.59a) we have

$$\psi_1(\hat{\alpha}, \hat{\sigma}^2) = \phi_1\left(\frac{\hat{\sigma}}{\hat{\alpha}}\right) = C \Longrightarrow s \triangleq \frac{\hat{\sigma}}{\hat{\alpha}} = \phi_1^{-1}(C). \tag{A.62}$$

Similarly, from (2.6b), i.e. the definition of ψ_2, and the definition of ϕ_2 in (A.59b), we can express $\psi_2(\hat{\alpha}, \hat{\sigma}^2; \delta)$ as

$$\begin{aligned}
\psi_2(\hat{\alpha}, \hat{\sigma}^2; \delta) &= \frac{4}{\delta}\left[\hat{\alpha}^2 + \hat{\sigma}^2 + 1 - \hat{\alpha} \cdot \phi_2\left(\frac{\hat{\sigma}}{\hat{\alpha}}\right)\right] \\
&= \frac{4}{\delta}\left[(1+s^2)\hat{\alpha}^2 + 1 - \hat{\alpha} \cdot \phi_2(s)\right].
\end{aligned}$$

From (A.62), we see that fixing $\psi_1(\hat{\alpha}, \hat{\sigma}^2) = C$ is equivalent to fixing $s = \phi_1^{-1}(C)$. Further, for a fixed s, $\psi_2(\hat{\alpha}, \hat{\sigma}^2)$ is a quadratic function of $\hat{\alpha}$, and the minimum happens at

$$\hat{\alpha}_{\min} = \frac{\phi_2(s)}{2(1+s^2)} = \frac{\phi_2(\phi_1^{-1}(C))}{2\left[1 + \left(\phi_1^{-1}(C)\right)^2\right]},$$

and $\psi_2(\hat{\alpha}_{\min}, \hat{\sigma}^2; \delta)$ is

$$\psi_2(\hat{\alpha}_{\min}, \hat{\sigma}^2; \delta) = \frac{4}{\delta}\left(1 - \frac{\phi_2^2(s)}{4(1+s^2)}\right) = \frac{4}{\delta}\left(1 - \frac{\phi_2^2(\phi_1^{-1}(C))}{4\left(1 + \left[\phi_1^{-1}(C)\right]^2\right)}\right) = L\left(C; \delta\right),$$

where the last step is from the definition of L is (A.58). This completes the proof. $\square$

To understand the implication of this lemma, let us consider the t^{th} iteration of the SE:

$$\begin{aligned}
\alpha_{t+1} &= \psi_1(\alpha_t, \sigma_t^2), \\
\sigma_{t+1}^2 &= \psi_2(\alpha_t, \sigma_t^2; \delta),
\end{aligned}$$

Note that according to Lemma 11, no matter where (α_t, σ_t^2) is, $(\alpha_{t+1}, \sigma_{t+1}^2)$ will fall above the $\sigma^2 = L(\alpha; \delta)$ curve. This function is a key component in the dynamics of AMP.A. Before we proceed further we discuss two main properties of the function $L(\alpha; \delta)$.

Lemma 12. *$L(\alpha; \delta)$ is a strictly decreasing function of $\alpha \in (0, 1)$.*

Proof. Recall from (A.58) that $L(\alpha;\delta)$ is defined as

$$L(\alpha;\delta) \triangleq \frac{4}{\delta}\left(1-\frac{\phi_2^2(\phi_1^{-1}(\alpha))}{4(1+(\phi_1^{-1}(\alpha))^2)}\right)$$
$$= \frac{4}{\delta}\left(1-I_2[\phi_1^{-1}(\alpha)]\right),$$

where $I_2:\mathbb{R}_+\mapsto\mathbb{R}_+$ is defined as

$$I_2(s) \triangleq \frac{\phi_2^2(s)}{4(1+s^2)}. \tag{A.63}$$

From (A.59a), it is easy to see that $\phi_1(s)$ is a decreasing function. Hence, to prove that $L(\alpha;\delta)$ is a decreasing function of α, it suffices to prove that $I_2(s)$ is strictly decreasing.

Substituting (A.59b) into (A.63) yields:

$$\begin{aligned}
I_2(s) &= \frac{\phi_2^2(s)}{4(1+s^2)}\\
&= \frac{1}{4(1+s^2)}\left(\int_0^{\frac{\pi}{2}}\frac{2\sin^2\theta+s^2}{\left(\sin^2\theta+s^2\right)^{\frac{1}{2}}}\right)^2\\
&\overset{(a)}{=} \frac{1}{4}\left[2E\left(\frac{1}{1+s^2}\right)-\frac{s^2}{1+s^2}K\left(\frac{1}{1+s^2}\right)\right]^2\\
&= \frac{1}{4}\left[2E(x)-(1-x)K(x)\right]^2,
\end{aligned}$$

where step (a) is obtained through similar calculations as those in (A.6), and in the last step we defined $x=\frac{1}{1+s^2}$. Hence, to prove that $I_2(s)$ is a decreasing function of s, it suffices to prove that $[2E(x)-(1-x)K(x)]^2$ is an increasing function of x. Further, $2E(x)-(1-x)K(x)=T(x)+E(x)>0$ (form the definition of $T(x)$ in (A.1)), our problem reduces to proving that $2E(x)-(1-x)K(x)$ is increasing. To this end, differentiation yields

$$[2E(x)-(1-x)K(x)]' \overset{(a)}{=} \frac{E(x)-(1-x)K(x)}{2x} \overset{(b)}{=} \frac{1}{2}T(x) \overset{(c)}{>} 0,$$

where (a) is from the differentiation identities in Lemma 1, (b) is from (A.1), and $T(x)>0$ follows from Lemma 1 (ii) together with the fact that $T(0)=0$. ☐

The next lemma compares the function $L(\alpha;\delta)$ with $F_1^{-1}(\alpha)$.

Lemma 13. *If $\delta>\delta_{\mathrm{AMP}}$, then*

$$F_1^{-1}(\alpha) > L(\alpha;\delta), \quad \forall\alpha\in(0,1).$$

Proof. We prove by contradiction. Suppose that $L(\hat{\alpha};\delta)\geq F_1^{-1}(\hat{\alpha})$ at some $\hat{\alpha}\in(0,1)$. If this is the case, then there exists a $\hat{\sigma}^2$ such that

$$F_1^{-1}(\hat{\alpha}) \leq \hat{\sigma}^2 \leq L(\hat{\alpha};\delta). \tag{A.64}$$

Since F_1 is a decreasing function (see Lemma 10), the first inequality implies that $\hat{\alpha}\geq F_1(\hat{\sigma}^2)$. Then, based on the global attractiveness property in Lemma 8 (iii), we have

$$\psi_1(\hat{\alpha},\hat{\sigma}^2) \leq \hat{\alpha}. \tag{A.65}$$

Further, Lemma 4 shows that $F_1^{-1}(\hat{\alpha}) > F_2(\hat{\alpha};\delta)$ for $\delta>\delta_{\mathrm{AMP}}$, and using (A.64) we also have $\hat{\sigma}^2\geq F_1^{-1}(\hat{\alpha})>F_2(\hat{\alpha};\delta)$. Also, from (A.64),

$$\hat{\sigma}^2 \leq L(\hat{\alpha};\delta) \overset{(a)}{<} L(0;\delta) = \frac{4}{\delta}\left(1-\frac{\pi^2}{16}\right) < \frac{4}{\delta} \leq \sigma^2_{\max},$$

where (a) is due to the monotonicity of $L(\alpha;\delta)$ (see Lemma 12). From the above discussions, $F_2(\hat{\alpha};\delta) < \hat{\sigma}^2 < \sigma^2_{\max}$. We then have (for $\delta > \delta_{\text{AMP}}$):

$$\psi_2(\hat{\alpha},\hat{\sigma}^2;\delta) \overset{(a)}{<} \hat{\sigma}^2 \overset{(b)}{\leq} L(\hat{\alpha};\delta) \overset{(c)}{\leq} L\left[\psi_1(\hat{\alpha},\hat{\sigma}^2);\delta\right], \tag{A.66}$$

where step (a) follows from the global attractiveness property in Lemma 9 (iv), step (b) is due to the hypothesis in (A.64), step (c) is from (A.65) together with the monotonicity of $L(\alpha;\delta)$ (see Lemma 12). Note that (A.66) shows that $\psi_2(\hat{\alpha},\hat{\sigma}^2;\delta) < L\left[\psi_1(\hat{\alpha},\hat{\sigma}^2);\delta\right]$, which contradicts Lemma 11, where we proved that $\psi_2(\alpha,\sigma^2;\delta) \geq L\left[\psi_1(\alpha,\sigma^2);\delta\right]$ for any $\alpha > 0$, $\sigma^2 > 0$ and $\delta > 0$. Hence, we must have that $L(\alpha;\delta) < F_1^{-1}(\alpha)$ for any $\alpha \in (0,1)$. ☐

Lemma 14. *The following holds for any $\alpha \in (0,1)$ and $\delta > 0$,*

$$L(\alpha;\delta) > \frac{4}{\delta}\left(1 - \frac{\pi^2}{16} - \frac{1}{2}\alpha^2\right), \tag{A.67}$$

where $L(\alpha,\delta)$ is defined in (A.58).

Proof. From (A.58), proving (A.67) is equivalent to proving:

$$1 - \frac{\phi_2^2(\phi_1^{-1}(\alpha))}{4\left[1 + (\phi_1^{-1}(\alpha))^2\right]} > 1 - \frac{\pi^2}{16} - \frac{1}{2}\alpha^2, \quad \forall \alpha \in (0,1), \tag{A.68}$$

where $\phi_1 : [0,\infty) \mapsto [0,1]$ and $\phi_2 : [0,\infty) \mapsto [0,\infty)$ are defined as (see (A.59a) and (A.59b)):

$$\phi_1(s) = \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(\sin^2\theta + s^2\right)^{\frac{1}{2}}} \mathrm{d}\theta, \tag{A.69a}$$

$$\phi_2(s) = \int_0^{\frac{\pi}{2}} \frac{2\sin^2\theta + s^2}{\left(\sin^2\theta + s^2\right)^{\frac{1}{2}}} \mathrm{d}\theta. \tag{A.69b}$$

We make a variable change:

$$\alpha = \phi_1(s).$$

Simple calculations show that (A.68) can be reformulated as the following

$$\frac{1}{1+s^2}\phi_2^2(s) < \frac{\pi^2}{4} + 2\phi_1^2(s), \quad s \in (0,\infty). \tag{A.70}$$

Let us further define

$$\phi_3(s) \equiv \int_0^{\frac{\pi}{2}} (\sin^2\theta + s^2)^{\frac{1}{2}} \mathrm{d}\theta. \tag{A.71}$$

From (A.69) and (A.71), we have

$$\phi_2(s) = \phi_1(s) + \phi_3(s),$$

and (A.70) can be reformulated as

$$\left[\phi_1(s) + \phi_3(s)\right]^2 - (1+s^2)\left[\frac{\pi^2}{4} + 2\phi_1^2(s)\right] < 0. \tag{A.72}$$

To this end, we can write the LHS of (A.72) into a quadratic form of $\phi_1(s)$:

$$\begin{aligned}
&\left[\phi_1(s) + \phi_3(s)\right]^2 - (1+s^2)\left[\frac{\pi^2}{4} + 2\phi_1^2(s)\right] \\
&= \phi_1^2(s) + \phi_3^2(s) + 2\phi_1(s)\phi_3(s) - (1+s^2)\left[\frac{\pi^2}{4} + 2\phi_1^2(s)\right] \\
&= -(1+2s^2)\phi_1^2(s) + 2\phi_1(s)\phi_3(s) - \frac{\pi^2}{4}(1+s^2) + \phi_3^2(s).
\end{aligned}$$

Hence, to prove that this quadratic form is negative everywhere, it suffices to prove that the discriminant is negative, i.e.,

$$4\phi_3^2(s) + 4(1+2s^2)\left[-\frac{\pi^2}{4}(1+s^2) + \phi_3^2(s)\right] < 0,$$

or

$$\phi_3^2(s) < \frac{\pi^2}{8}(1+2s^2).$$

Finally, by Cauchy-Schwarz we have

$$\begin{aligned}
\phi_3^2(s) &= \left[\int_0^{\frac{\pi}{2}} (\sin^2\theta + s^2)^{\frac{1}{2}} \mathrm{d}\theta\right]^2 \\
&\le \int_0^{\frac{\pi}{2}} 1\mathrm{d}\theta \cdot \int_0^{\frac{\pi}{2}} \left(\sqrt{\sin^2\theta + s^2}\right)^2 \mathrm{d}\theta \\
&= \frac{\pi}{2}\left(\frac{\pi}{4} + \frac{\pi}{2}s^2\right) = \frac{\pi^2}{8}(1+2s^2),
\end{aligned}$$

which completes our proof. □

Lemma 15. *For any $\alpha \in [0,1]$, $\psi_2(\alpha, \sigma^2; \delta_{\mathrm{AMP}})$ is an increasing function of σ^2 on $\sigma^2 \in [L(\alpha;\delta_{\mathrm{AMP}}), \infty)$, where the function $L(\alpha;\delta)$ is defined in (4).*

Proof. From Lemma 9 (v), the case $\alpha > \alpha_* \approx 0.53$ is trivial since then $\psi_2(\sigma^2, \alpha; \delta_{\mathrm{AMP}})$ is strictly increasing in $\sigma^2 \in \mathbb{R}_+$. In the rest of this proof, we assume that $\alpha < \alpha_*$. We have derived in (A.18) that

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} > 0 \Longleftrightarrow \alpha > \frac{1}{2\sqrt{1+s^2}} E\left(\frac{1}{1+s^2}\right) = f(s), \tag{A.73}$$

where

$$s \triangleq \frac{\sigma}{\alpha}.$$

Hence, the result of Lemma 15 can be reformulated as proving the following:

$$\alpha > f(s), \quad \forall s \ge \frac{\sqrt{L(\alpha;\delta_{\mathrm{AMP}})}}{\alpha}, \qquad \alpha \in [0, \alpha_*).$$

We proceed in three steps:

(i) In Lemma 14, we proved that the following holds for any $\alpha \in [0,1]$:

$$L(\alpha;\delta_{\mathrm{AMP}}) \ge \hat{L}(\alpha, \delta_{\mathrm{AMP}}) \triangleq \frac{4}{\delta_{\mathrm{AMP}}}\left(1 - \frac{\pi^2}{16} - \frac{1}{2}\alpha^2\right). \tag{A.74}$$

For convenience, define

$$\hat{s}(\alpha) \triangleq \frac{\sqrt{\hat{L}(\alpha;\delta_{\mathrm{AMP}})}}{\alpha}. \tag{A.75}$$

(ii) We prove that $f(s)$ is monotonically decreasing on $s \in [\hat{s}(\alpha), \infty)$ for $\alpha < \alpha_*$.

(iii) We prove that the following holds for $\alpha < \alpha_*$:

$$\alpha > f(\hat{s}(\alpha)).$$

Clearly, (A.73) follows from the above claims. Here, we introduce the function $\hat{L}$ since $\hat{L}$ has a simple closed-form formula and is easier to manipulate than $L(\alpha)$. We next prove step (ii). From (A.27), it suffices to prove that

$$\hat{s}(\alpha) > s_*, \quad \forall \alpha < \alpha_*,$$

where s_* and α_* are defined in (A.32) and (A.31) respectively. To this end, we note that the following holds for $\alpha < \alpha_*$:

$$\hat{s}(\alpha) = \frac{\sqrt{\hat{L}(\alpha;\delta_{\text{AMP}})}}{\alpha} > \frac{\sqrt{\hat{L}(\alpha_*;\delta_{\text{AMP}})}}{\alpha_*} \approx 1.18,$$

where the inequality follows from the fact that $\hat{L}$ in (A.74) is strictly decreasing in α, and the last step is calculated from (A.74) and $\alpha_* \approx 0.527$. Finally, numerical evaluation of (A.32) shows that $s_* \approx 0.458$. Hence, $\hat{s}(\alpha) > s_*$, which completes the proof.

We next prove step (iii). First, simple manipulations yields

$$\hat{s}^2(\alpha) \overset{(a)}{=} \frac{\hat{L}(\alpha)}{\alpha^2} \overset{(b)}{=} \frac{4}{\delta_{\text{AMP}}}\left[\left(1-\frac{\pi^2}{16}\right)\cdot\frac{1}{\alpha^2}-\frac{1}{2}\right], \tag{A.76}$$

where (a) is from the definition of $\hat{s}(\alpha)$ in (A.75) and (b) is due to (A.74). Using (A.76), we further obtain

$$\alpha = \sqrt{\frac{16-\pi^2}{4\delta_{\text{AMP}}\hat{s}^2(\alpha)+8}}. \tag{A.77}$$

Now, from (A.77) and (A.25b), we have

$$\alpha - f(\hat{s}(\alpha)) > 0 \Longleftrightarrow \sqrt{\frac{16-\pi^2}{4\delta_{\text{AMP}}\hat{s}^2(\alpha)+8}} - \frac{1}{2\sqrt{1+\hat{s}^2(\alpha)}}E\left(\frac{1}{1+\hat{s}^2(\alpha)}\right) > 0. \tag{A.78}$$

We prove (A.78) by showing that the following stronger result holds:

$$\sqrt{\frac{16-\pi^2}{4\delta_{\text{AMP}}t^2+8}} - \frac{1}{2\sqrt{1+t^2}}E\left(\frac{1}{1+t^2}\right) > 0, \quad \forall t \in \mathbb{R}_+. \tag{A.79}$$

For convenience, we make a variable change:

$$x \triangleq \frac{1}{1+t^2}.$$

With some straightforward calculations, we can rewrite (A.79) as

$$E(x) < \sqrt{\frac{16-\pi^2}{\delta_{\text{AMP}}(1-x)+2x}}$$

The following upper bound on $E(x)$ is due to Eqn. (1.2) in Wang & Chu (2013):

$$E(x) < \frac{\pi}{2}\sqrt{1-\frac{x}{2}}, \quad \forall x \in (0,1].$$

Hence, it is sufficient to prove that

$$\frac{\pi}{2}\sqrt{1-\frac{x}{2}} < \sqrt{\frac{16-\pi^2}{\delta_{\text{AMP}}(1-x)+2x}},$$

which can be reformulated as

$$\left(1-\frac{x}{2}\right)(\delta_{\text{AMP}}-(\delta_{\text{AMP}}-2)x) < \frac{4}{\pi^2}(16-\pi^2) = \delta_{\text{AMP}}$$

where the second equality follows from the definition $\delta_{\text{AMP}} = \frac{64}{\pi^2} - 4$. The above inequality holds since $0 < 1 - \frac{x}{2} < 1$ and $0 < \delta_{\text{AMP}} - (\delta_{\text{AMP}} - 2)x < \delta_{\text{AMP}}$. This completes the proof. □

Lemma 16. *For any $\alpha \in [0,1]$, $\psi_2\left(\alpha, L(\alpha;\delta);\delta\right)$ is a strictly decreasing function of $\delta > 0$, where $L(\alpha;\delta)$ is defined in* (A.58).

Proof. From the definition of $L(\alpha;\delta)$ in (A.58), we can write

$$\psi_2\left(\alpha, L(\alpha;\delta);\delta\right) = \psi_2\left(\alpha, \frac{1}{\delta}\bar{\sigma}^2;\delta\right),$$

where (note that $\bar{\sigma}$ is not the conjugate of σ)

$$\bar{\sigma}^2 \triangleq 4\left(1 - \frac{\phi_2^2(\phi_1^{-1}(\alpha))}{4\left[1+(\phi_1^{-1}(\alpha))^2\right]}\right).$$

A key observation here is that $\bar{\sigma}^2$ does not depend on δ. Clearly, Lemma 16 is implied by the following stronger result:

$$\frac{\partial\psi_2\left(\alpha, \frac{1}{\delta}\bar{\sigma}^2;\delta\right)}{\partial\delta} < 0, \quad \forall\bar{\sigma}^2 > 0, \alpha > 0, \delta > 0,$$

which we will prove in the sequel. For convenience, we define

$$\bar{s} \triangleq \frac{\bar{\sigma}}{\alpha},\ \gamma \triangleq \frac{1}{\delta} \text{ and } s = \sqrt{\gamma}\bar{s}. \tag{A.80}$$

Using these new variables, we have

$$\begin{aligned}\psi_2\left(\alpha, \frac{1}{\delta}\bar{\sigma}^2;\delta\right) &= \psi_2\left(\alpha, \gamma\bar{\sigma}^2;\gamma^{-1}\right)\\ &= 4\gamma\left((1+\gamma\bar{s}^2)\alpha^2 + 1 - \alpha\int_0^{\frac{\pi}{2}} \frac{2\sin^2\theta + \gamma\bar{s}^2}{\left(\sin^2\theta+\gamma\bar{s}^2\right)^{\frac{1}{2}}}\mathrm{d}\theta\right),\end{aligned}$$

where the last equality is from the definition of ψ_2 in (2.6b). It remains to prove that $\psi_2\left(\alpha, \gamma\bar{\sigma}^2;\gamma^{-1}\right)$ is an increasing function of γ. The partial derivative of $\psi_2(\alpha,\sigma^2;\delta)$ w.r.t. γ is given by

$$\begin{aligned}\frac{\partial\psi_2\left(\alpha, \gamma\bar{\sigma}^2;\gamma^{-1}\right)}{\partial\gamma} &= 4(1+2\gamma\bar{s}^2)\alpha^2 - 4\alpha\left(\int_0^{\frac{\pi}{2}}\frac{2\sin^2\theta+\gamma\bar{s}^2}{(\sin^2\theta+\gamma\bar{s}^2)^{\frac{1}{2}}}\mathrm{d}\theta + \frac{1}{2}\int_0^{\frac{\pi}{2}}\frac{\gamma^2\bar{s}^4}{(\sin^2\theta+\gamma\bar{s}^2)^{\frac{3}{2}}}\mathrm{d}\theta\right) + 4\\ &\overset{(a)}{=} (1+2s^2)\alpha^2 - 4\alpha\left(\int_0^{\frac{\pi}{2}}\frac{2\sin^2\theta+s^2}{\left(\sin^2\theta+s^2\right)^{\frac{1}{2}}}\mathrm{d}\theta + \frac{1}{2}\int_0^{\frac{\pi}{2}}\frac{s^4}{(\sin^2\theta+s^2)^{\frac{3}{2}}}\mathrm{d}\theta\right) + 4\\ &\overset{(b)}{=} 4(1+2s^2)\alpha^2 - 4\alpha\left(\frac{(5s^2+4)E\left(\frac{1}{1+s^2}\right) - 2s^2K\left(\frac{1}{1+s^2}\right)}{2\sqrt{1+s^2}}\right) + 4,\end{aligned} \tag{A.81}$$

where in step (a) we used the relationship $s^2 = \gamma\bar{s}^2$ (see (A.80)), and step (b) is from the identities in (A.6). From (A.81), we see that $\frac{\partial\psi_2\left(\alpha,\gamma\bar{\sigma}^2;\gamma^{-1}\right)}{\partial\gamma}$ is a quadratic function of α. Therefore, to prove $\frac{\partial\psi_2\left(\alpha,\gamma\bar{\sigma}^2;\gamma^{-1}\right)}{\partial\gamma} > 0$, it suffices to show that the discriminant is negative:

$$\left(\frac{(5s^2+4)E\left(\frac{1}{1+s^2}\right) - 2s^2K\left(\frac{1}{1+s^2}\right)}{2\sqrt{1+s^2}}\right)^2 - 4(1+2s^2) < 0. \tag{A.82}$$

Further, to prove (A.82), it is sufficient to prove that the following two inequalities hold:

$$(5s^2+4)E\left(\frac{1}{1+s^2}\right) - 2s^2K\left(\frac{1}{1+s^2}\right) > 0, \tag{A.83a}$$

and

$$(5s^2+4)E\left(\frac{1}{1+s^2}\right)-2s^2K\left(\frac{1}{1+s^2}\right)<4\sqrt{1+s^2}\sqrt{1+2s^2}. \tag{A.83b}$$

We first prove (A.83a). It is sufficient to prove the following

$$(4s^2+4)E\left(\frac{1}{1+s^2}\right)-2s^2K\left(\frac{1}{1+s^2}\right)>0. \tag{A.84}$$

Applying a variable change $x=\frac{1}{1+s^2}$, we can rewrite (A.84) as

$$\frac{4E(x)-2(1-x)K(x)}{x}>0.$$

The above inequality holds since

$$4E(x)-2(1-x)K(x)>2E(x)-2(1-x)K(x)=2T(x)>0,$$

where the last equality is from the definition of $T(x)$ in (A.1).

We next prove (A.83b). Again, applying the variable change $x=\frac{1}{1+s^2}$ and after some straightforward manipulations, we can rewrite (A.83b) as

$$h(x)/x<0, \quad x\in(0,1),$$

where

$$h(x)\triangleq(5-x)E(x)-2(1-x)K(x)-4\sqrt{2-x}<0.$$

Hence, we only need to prove $h(x)<0$ for $0<x<1$. First, we note that $\lim_{x\to 1^-}h(x)=0$, from the fact that $E(1)=1$ and $\lim_{x\to 1^-}(1-x)K(x)=0$ (see Lemma 1 (i)). We finish the proof by showing that $h(x)$ is strictly increasing in $x\in(0,1)$. Using the identities in (A.3), we can obtain

$$h'(x)=\frac{3}{2}\frac{(1-x)(E(x)-K(x))}{x}+\frac{2}{\sqrt{2-x}}.$$

To prove $h'(x)>0$, it is equivalent to prove

$$\begin{aligned}\frac{4x}{3(1-x)\sqrt{2-x}}&>K(x)-E(x)\\&=\int_0^{\frac{\pi}{2}}\frac{1}{(1-x\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta-\int_0^{\frac{\pi}{2}}(1-x\sin^2\theta)^{\frac{1}{2}}\mathrm{d}\theta\\&=\int_0^{\frac{\pi}{2}}\frac{x\sin^2\theta}{(1-x\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta.\end{aligned} \tag{A.85}$$

Noting $0<x<1$, we can get the following

$$\int_0^{\frac{\pi}{2}}\frac{x\sin^2\theta}{(1-x\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta<\int_0^{\frac{\pi}{2}}\frac{x\sin^2\theta}{1-x\sin^2\theta}\mathrm{d}\theta=\frac{\pi}{2}\left(\frac{1}{\sqrt{1-x}}-1\right).$$

Hence, to prove (A.85), it suffices to prove

$$\frac{4x}{3(1-x)\sqrt{2-x}}>\frac{\pi}{2}\left(\frac{1}{\sqrt{1-x}}-1\right),$$

which can be reformulated as

$$\frac{8}{3\pi}\frac{1}{\sqrt{2-x}}>\frac{\sqrt{1-x}}{1+\sqrt{1-x}}.$$

The inequality holds since

$$\frac{8}{3\pi}\frac{1}{\sqrt{2-x}} > \frac{8}{3\pi}\frac{1}{\sqrt{2}} > \frac{1}{2}, \quad \forall x \in (0,1),$$

and

$$\frac{\sqrt{1-x}}{1+\sqrt{1-x}} < \frac{1}{2}, \quad \forall x \in (0,1).$$

$\square$

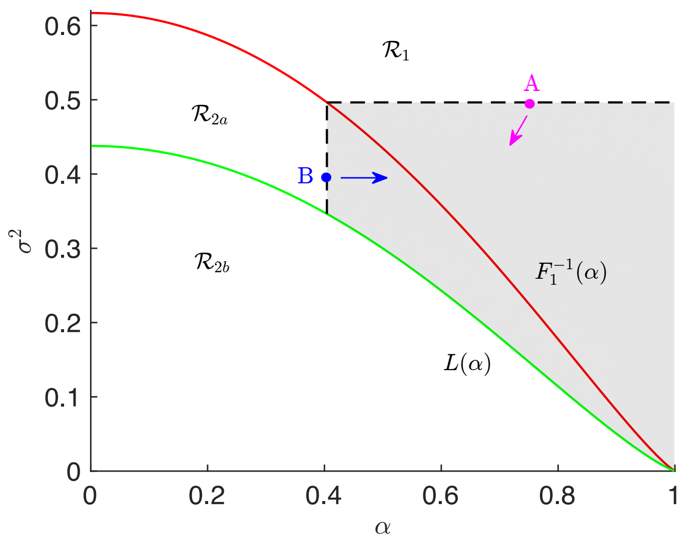

Figure 12: Illustration of the convergence behavior. $\mathcal{R}_1$ and $\mathcal{R}_2$ are defined in Definition 3. For both point A and point B, $B_1(\alpha, \sigma^2)$ and $B_2(\alpha, \sigma^2)$ are given by the two dashed lines. After one iteration, $\mathcal{R}_{2b}$ will not be achievable and we can focus on $\mathcal{R}_{2a}$.

- **Main proof**

We now return to the main proof for Lemma 6. Notice that by Lemma 11, $(\alpha_{t_0}, \sigma^2_{t_0})$ cannot fall below the curve $L(\alpha;\delta)$ for $t_0 \geq 1$. Hence, for $\mathcal{R}_2$, we can focus on the region above $L(\alpha;\delta)$ (including $L(\alpha;\delta)$), which we denote as $\mathcal{R}_{2a}$. See Fig. 12 for illustration.

We will first prove that if $(\alpha, \sigma^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$, then the next iterates $\psi_1(\alpha, \sigma^2)$ and $\psi_2(\alpha, \sigma^2)$ satisfy the following:

$$\psi_1(\alpha, \sigma^2) \geq B_1(\alpha, \sigma^2), \tag{A.87a}$$

and

$$\psi_2(\alpha, \sigma^2) \leq B_2(\alpha, \sigma^2), \tag{A.87b}$$

where $B_1(\alpha, \sigma^2)$ and $B_2(\alpha, \sigma^2)$ are defined as

$$\begin{aligned} B_1(\alpha, \sigma^2) &\triangleq \min\left\{\alpha, F_1(\sigma^2)\right\}, \\ B_2(\alpha, \sigma^2) &\triangleq \max\left\{\sigma^2, F_1^{-1}(\alpha)\right\}. \end{aligned} \tag{A.88}$$

Note that when (α, σ^2) is on F_1^{-1} (i.e., $\sigma^2 = F_1^{-1}(\alpha)$), equalities in (A.87a) and (A.87b) can be achieved. Further, this is the only case when either of the equality is achieved. Also, it is easy to see that if (α, σ^2) is on F_1^{-1}, then $(\psi_1(\alpha, \sigma^2), \psi_2(\alpha, \sigma^2))$ cannot be on F_1^{-1}.

Since F_1^{-1} separates $\mathcal{R}_1$ and $\mathcal{R}_{2a}$, (A.88) can also be written as

$$\left[B_1(\alpha,\sigma^2), B_2(\alpha,\sigma^2)\right] = \begin{cases} [F_1(\sigma^2), \sigma^2] & \text{if } (\alpha,\sigma^2)\in\mathcal{R}_1, \\ [\alpha, F_1^{-1}(\alpha)] & \text{if } (\alpha,\sigma^2)\in\mathcal{R}_{2a}. \end{cases} \tag{A.89}$$

As a concrete example, consider the situation shown in Fig. 12. In this case, for both point A and point B, $B_1(\alpha,\sigma^2)$ and $B_2(\alpha,\sigma^2)$ are given by the two dashed lines. This directly follows from (A.89) by noting that point A is in region $\mathcal{R}_1$ and point B is in region $\mathcal{R}_{2a}$. Let $\mathcal{R}_{2a}\backslash F_1^{-1}(\alpha)$ be a shorhand for $\{(\alpha,\sigma^2)|(\alpha,\sigma^2)\in\mathcal{R}_{2a}, \alpha\neq F_1(\sigma^2)\}$. To prove the strict inequality in (A.87), we deal with $(\alpha,\sigma^2)\in\mathcal{R}_1$ and $(\alpha,\sigma^2)\in\mathcal{R}_{2a}\backslash F_1^{-1}(\alpha)$ separately.

1. Assume that $(\alpha,\sigma^2)\in\mathcal{R}_1$. Using (A.89), the inequality in (A.87) can be rewritten as

$$\psi_1(\alpha,\sigma^2) > F_1(\sigma^2) \quad \text{and} \quad \psi_2(\alpha,\sigma^2) < \sigma^2. \tag{A.90}$$

Since $(\alpha,\sigma^2)\in\mathcal{R}_1$, we have $\sigma^2 > F_1^{-1}(\alpha)$. Then, applying (A.12) proves $\psi_1(\alpha,\sigma^2) > F_1(\sigma^2)$. Further, using Lemma 4, we have $\sigma^2 > F_1^{-1}(\alpha) > F_2(\alpha)$. Also, Lemma 5 guarantees that $\sigma^2 < \sigma^2_{\max}$. Hence, $F_1^{-1}(\alpha) < \sigma^2 < \sigma^2_{\max}$ and applying Lemma 9 (iv) yields $\psi_2(\alpha,\sigma^2) < \sigma^2$.

2. We now consider the case where $(\alpha,\sigma^2)\in\mathcal{R}_{2a}\backslash F_1^{-1}(\alpha)$. Similar to (A.90), we need to prove

$$\psi_1(\alpha,\sigma^2) > \alpha \quad \text{and} \quad \psi_2(\alpha,\sigma^2) < F_1^{-1}(\alpha). \tag{A.91}$$

The inequality $\psi_1(\alpha,\sigma^2) > \alpha$ can be proved by the global attractiveness in Lemma 8 (iii) and the fact that $\sigma^2 < F_1^{-1}(\alpha)$ when $(\alpha,\sigma^2)\in\mathcal{R}_{2a}\backslash F_1^{-1}(\alpha)$. The proof for $\psi_2(\alpha,\sigma^2) < F_1^{-1}(\alpha)$ is considerably more complicated and is detailed in Lemma 17 below.

Lemma 17. *For any $(\alpha,\sigma^2)\in\mathcal{R}_{2a}$ (see Definition 3) and $\delta\geq\delta_{\mathrm{AMP}}$, the following holds:*

$$\psi_2(\alpha,\sigma^2;\delta) < F_1^{-1}(\alpha), \tag{A.92}$$

where ψ_2 is the SE map in (2.6b) *and F_1^{-1} is the inverse of F_1 defined in Lemma 8.*

Proof. The following holds when $(\alpha,\sigma^2)\in\mathcal{R}_{2a}$:

$$\psi_2(\alpha,\sigma^2;\delta) \leq \max_{\hat{\sigma}^2\in\mathcal{D}_\alpha} \psi_2(\alpha,\hat{\sigma}^2;\delta),$$

where

$$\mathcal{D}_\alpha \triangleq \left\{\hat{\sigma}^2 \middle| L(\alpha;\delta)\leq\sigma^2\leq F_1^{-1}(\alpha)\right\}. \tag{A.93}$$

Hence, to prove (A.92), it suffices to prove that the following holds for any $\delta\geq\delta_{\mathrm{AMP}}$ and $\alpha\in[0,1]$:

$$\max_{\hat{\sigma}^2\in\mathcal{D}_\alpha} \psi_2(\alpha,\hat{\sigma}^2;\delta) < F_1^{-1}(\alpha). \tag{A.94}$$

We next prove (A.94). We consider the three different cases:

(i) $\alpha\in[\alpha_*,1]$ and all $\delta\in[\delta_{\mathrm{AMP}},\infty)$, where α_* is defined in (A.17).

(ii) $\alpha\in[0,\alpha_*)$ and $\delta\in[\delta_{\mathrm{AMP}},17]$.

(iii) $\alpha\in[0,\alpha_*)$ and $\delta\in(17,\infty)$.

Case (i): Lemma 9 (v) shows that ψ_2 is an increasing function of σ^2 in $\mathbb{R}_+$. Hence, by noting (A.93), we have

$$\max_{\hat{\sigma}^2\in\mathcal{D}_\alpha} \psi_2(\alpha,\hat{\sigma}^2;\delta) = \psi_2(\alpha,F_1^{-1}(\alpha);\delta).$$

Therefore, proving (A.98) reduces to proving

$$\psi_2(\alpha,F_1^{-1}(\alpha);\delta) \leq F_1^{-1}(\alpha). \tag{A.95}$$

Finally, (A.95) follows from the global attractiveness property in Lemma 9 (iv) and the inequality $F_1^{-1}(\alpha) > F_2(\alpha;\delta)$ in Lemma 4.

Case (ii): We will prove that the following holds for $\alpha \in [0, \alpha_*)$ and $\delta \in [\delta_{\mathrm{AMP}}, 17]$ (at the end of this proof)

$$\max_{\hat{\sigma}^2 \in \mathcal{D}_\alpha} \psi_2(\alpha, \sigma^2; \delta) = \max\left\{\psi_2(\alpha, L(\alpha;\delta);\delta),\ \psi_2(\alpha, F_1^{-1}(\alpha);\delta)\right\}. \tag{A.96}$$

Namely, the maximum of ψ_2 over σ^2 is achieved at either $\sigma^2 = L(\alpha;\delta)$ or $\sigma^2 = F_1^{-1}(\alpha)$. Hence, we only need to prove that the following holds for any $\alpha \in [0, \alpha_*)$ and $\delta \geq \delta_{\mathrm{AMP}}$:

$$\max\left\{\psi_2(\alpha, L(\alpha;\delta);\delta),\ \psi_2(\alpha, F_1^{-1}(\alpha);\delta)\right\} \leq F_1^{-1}(\alpha). \tag{A.97}$$

In the sequel, we first use (A.96) to prove (A.94), and the proof for (A.96) will come at the end of this proof.

Firstly, it is easy to see that $\psi_2(\alpha, F_1^{-1}(\alpha);\delta)$ is a decreasing function of δ, since $\psi_2(\alpha, \sigma^2; \delta)$ is a decreasing function of δ and $F_1^{-1}(\alpha)$ does not depend on δ. Further, Lemma 16 shows that $\psi_2(\alpha, L(\alpha;\delta);\delta)$ is also a decreasing function of δ. (Notice that unlike $F_1^{-1}(\alpha)$, $L(\alpha;\delta)$ depends on δ, and thus Lemma 16 is nontrivial.) Hence, to prove (A.97) for $\delta \geq \delta_{\mathrm{AMP}}$, it suffices to prove (A.97) for $\delta = \delta_{\mathrm{AMP}}$, namely,

$$\max\left\{\psi_2(\alpha, L(\alpha;\delta);\delta_{\mathrm{AMP}}),\ \psi_2(\alpha, F_1^{-1}(\alpha);\delta_{\mathrm{AMP}})\right\} \leq F_1^{-1}(\alpha). \tag{A.98}$$

When $\delta = \delta_{\mathrm{AMP}}$, we prove in Lemma 15 that ψ_2 is an increasing function of σ^2 in $\sigma^2 \in [L(\alpha;\delta_{\mathrm{AMP}}), \infty)$. (Such monotonicity generally does not hold if δ is too large.) Further, Lemma 13 shows that $F_1^{-1}(\alpha) > L(\alpha;\delta_{\mathrm{AMP}})$. Hence,

$$\psi_2(\alpha, L(\alpha;\delta);\delta_{\mathrm{AMP}}) \leq \psi_2(\alpha, F_1^{-1}(\alpha);\delta_{\mathrm{AMP}}),$$

and thus proving (A.98) reduces to proving

$$\psi_2(\alpha, F_1^{-1}(\alpha);\delta_{\mathrm{AMP}}) \leq F_1^{-1}(\alpha),$$

which follows from the same argument as that for (A.95).

Case (iii): Lemma 9 (iii) shows that $\psi_2(\alpha;\sigma^2;\delta) \leq \frac{4}{\delta}$ for any $\sigma^2 \in [0, \sigma^2_{\max}]$. It is easy to see that $\mathcal{D}_\alpha \subset [0, \sigma^2_{\max}]$, and thus

$$\max_{\sigma^2 \in \mathcal{D}_\alpha} \psi_2(\alpha, \sigma^2; \delta) \leq \frac{4}{\delta} \leq \frac{4}{17} \approx 0.235. \tag{A.99}$$

Further, Lemma 10 shows that $F_1^{-1} : [0,1] \mapsto [0, \pi^2/16]$ is monotonically decreasing. Hence,

$$F_1^{-1}(\alpha) > F_1^{-1}(\alpha_*) \approx 0.415, \tag{A.100}$$

where the numerical constant is calculated from the closed form formula $F_1^{-1}(\alpha) = \alpha^2 \cdot \left[\phi_1^{-1}(\alpha)\right]^2$ (see (A.42)) and $\alpha_* \approx 0.5274$ (from (A.17)). Comparing (A.99) and (A.100) shows that (A.94) holds in this case.

It only remains to prove (A.96). We have shown in (A.25) that

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} = \frac{4}{\delta \alpha}\left(\alpha - \underbrace{\frac{1}{2\sqrt{1+s^2}} E\left(\frac{1}{1+s^2}\right)}_{f(s)}\right), \tag{A.101}$$

where $s \triangleq \sigma/\alpha$. Further, we have proved in (A.27) that $f(s)$ is strictly increasing on $[0, s_*)$ and strictly decreasing on (s_*, ∞), where s_* is defined in (A.32). Hence, when $f(0) = 0.5 < \alpha < f(s_*) = \alpha_*$, there exist two solutions to

$$\alpha = f(s),$$

denoted as $s_1(\alpha)$ and $s_2(\alpha)$, respectively. Also, from (A.101) and noting the definition $s = \sigma/\alpha$, we have

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} > 0 \Longleftrightarrow \sigma^2 \in \left[0, \sigma_1^2(\alpha)\right) \cup \left(\sigma_2^2(\alpha), \infty\right),$$

$$\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} \leq 0 \Longleftrightarrow \sigma^2 \in \left[\sigma_1^2(\alpha), \sigma_2^2(\alpha)\right],$$

where $\sigma_1^2(\alpha) \triangleq \alpha^2 s_1^2(\alpha)$ and $\sigma_2^2(\alpha) \triangleq \alpha^2 s_2^2(\alpha)$. Hence, for fixed α where $\alpha \in (f(0), f(s_*))$, $\sigma_1^2(\alpha)$ is a local maximum of ψ_2 and $\sigma_2^2(\alpha)$ is a local minimum. Clearly, if

$$L(\alpha;\delta) \geq \sigma_1^2(\alpha), \tag{A.102}$$

then the maximum of ψ_2 over $\sigma^2 \in [L(\alpha;\delta), F_1^{-1}(\alpha)]$ can only happen at either $L(\alpha;\delta)$ or $F_1^{-1}(\alpha)$, which will prove (A.96). Further, for the degenerate case $\alpha \in (0, f(0))$, ψ_2 only has a local minimum, and it is easy to see that (A.96) also holds. Thus, we only need to prove that (A.102) holds when $\delta < 17$. This can be proved as follows:

$$\sigma_1^2(\alpha) \overset{(a)}{\leq} s_*^2 \cdot \alpha^2 \overset{(b)}{\leq} s_*^2 \cdot \alpha_*^2, \tag{A.103}$$

where (a) is from the fact that $s_1(\alpha) \leq s_*$ and (b) is from our assumption $\alpha \leq \alpha_*$. On the other hand, since $L(\alpha)$ is a decreasing function of α (see Lemma 12), and thus for $\alpha \leq \alpha_*$ we have

$$\begin{aligned} L(\alpha;\delta) &\geq L(\alpha_*;\delta) \\ &= \frac{4}{\delta}\left(1 - \frac{\phi_2^2(\phi_1^{-1}(\alpha_*))}{4\left[1 + (\phi_1^{-1}(\alpha_*))^2\right]}\right), \end{aligned} \tag{A.104}$$

where the last step is from Definition A.58. Based on (A.103) and (A.104), we see that $L(\alpha;\delta) > \sigma_1^2(\alpha)$ for $\alpha \leq \alpha_*$ if

$$\delta \leq \frac{4}{s_*^2 \cdot \alpha_*^2}\left(1 - \frac{\phi_2^2(\phi_1^{-1}(\alpha_*))}{4\left[1 + (\phi_1^{-1}(\alpha_*))^2\right]}\right) \approx 17.04,$$

where the numerical constant is calculated based on the definition of α_* in (A.31), the definition of s_* in (A.32), and that of ϕ_1 and ϕ_2 in Definition A.58. Hence, the condition $\delta < 17$ is enough for our purpose. This concludes our proof. $\square$

Now we turn our attention to the proof of part (i) of Lemma 6. Suppose that $(\alpha, \sigma^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$. Then, using (A.87) and based on the fact that $F_1(\alpha)$ is a strictly decreasing function, we know that $(\psi_1(\alpha,\sigma^2), \psi_2(\alpha,\sigma^2)) \in \mathcal{R}_1 \cup \mathcal{R}_2$. (See Definition 3.) Further, Lemma 7 shows that $(\psi_1(\alpha,\sigma^2), \psi_2(\alpha,\sigma^2)) \notin \mathcal{R}_{2b}$. Hence, $(\psi_1(\alpha,\sigma^2), \psi_2(\alpha,\sigma^2)) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$. Applying this argument recursively shows that if $(\alpha_{t_0}, \sigma_{t_0}^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$, then $(\alpha_t, \sigma_t^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$ for all $t > t_0$. An illustration of the situation is shown in Fig. 12.

Now we can discuss the proof of part (ii) of Lemma 6. To proceed, we introduce two auxiliary sequences $\{\tilde{\alpha}_{t+1}\}_{t\geq t_0}$ and $\{\tilde{\sigma}_{t+1}^2\}_{t\geq t_0}$, defined as:

$$\tilde{\alpha}_{t+1} = B_1(\alpha_t, \sigma_t^2) \quad \text{and} \quad \tilde{\sigma}_{t+1}^2 = B_2(\alpha_t, \sigma_t^2), \tag{A.105}$$

where B_1 and B_2 are defined in (A.88). Note that the definitions of $B_1(\alpha,\sigma^2)$ and $B_2(\alpha,\sigma^2)$ require $(\alpha,\sigma^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a}$, and such requirement is satisfied here due to part (i) of this lemma. Noting the SE update $\alpha_{t+1} = \psi_1(\alpha_t, \sigma_t^2)$ and $\sigma_{t+1}^2 = \psi_2(\alpha_t, \sigma_t^2)$, and recall the inequalities in (A.87), we obtain the following:

$$\alpha_{t+1} \geq \tilde{\alpha}_{t+1} \quad \text{and} \quad \sigma_{t+1}^2 \leq \tilde{\sigma}_{t+1}^2, \quad \forall t \geq t_0. \tag{A.106}$$

Namely, $\{\tilde{\alpha}_{t+1}\}_{t\geq t_0}$ and $\{\tilde{\sigma}_{t+1}^2\}_{t\geq t_0}$ are "worse" than $\{\alpha_{t+1}\}_{t\geq t_0}$ and $\{\sigma_{t+1}^2\}_{t\geq t_0}$, respectively, at each iteration. We next prove that

$$\lim_{t\to\infty} \tilde{\alpha}_{t+1} = 1 \quad \text{and} \quad \lim_{t\to\infty} \tilde{\sigma}_{t+1}^2 = 0, \tag{A.107}$$

which together with (A.106), and the fact that $\alpha_{t+1} \leq 1$ and $\sigma_{t+1} > 0$ (since $(\alpha_t, \sigma_t^2) \in \mathcal{R}_{2a}$), leads to the results we want to prove:

$$\lim_{t\to\infty} \alpha_{t+1} = 1 \quad \text{and} \quad \lim_{t\to\infty} \sigma_{t+1}^2 = 0.$$

It remains to prove (A.107). First, notice that $\tilde{\alpha}_{t+1} \leq 1$ and $\tilde{\sigma}_{t+1}^2 \geq 0$ ($\forall t \geq t_0$), from the definition in (A.88). We then show that the sequence $\{\tilde{\alpha}_{t+1}\}_{t \geq t_0}$ is monotonically non-decreasing and $\{\tilde{\sigma}_{t+1}^2\}_{t \geq t_0}$ is monotonically non-increasing, namely,

$$\tilde{\alpha}_{t+2} \geq \tilde{\alpha}_{t+1} \quad \text{and} \quad \tilde{\sigma}_{t+2}^2 \leq \tilde{\sigma}_{t+1}^2, \quad \forall t \geq t_0, \tag{A.108}$$

and equalities of (A.108) hold only when the equalities in (A.87) hold. Then we can finish the proof by the fact that $\tilde{\alpha}$ and $\tilde{\sigma}^2$ will improve strictly in at most two consecutive iterations and the ratios $\frac{\tilde{\alpha}_{t+2}}{\tilde{\alpha}_t}$, $\frac{\tilde{\sigma}_{t+2}^2}{\tilde{\sigma}_t^2}$ are continuous functions of (α_t, σ_t^2) on $[\tilde{\alpha}_{t_0}, 1] \times [0, \sigma_{\max}^2]$. (This is essentially due to the fact that equalities in (A.87) can be achieved when $\sigma^2 = F_1^{-1}(\alpha)$, but this cannot happen in two consecutive iterations. See the discussions below (A.88).)

To prove (A.108), we only need to prove the following (based on the definition in (A.105))

$$B_1\left[\psi_1, \psi_2\right] \geq B_1(\alpha, \sigma^2) \quad \text{and} \quad B_2\left[\psi_1, \psi_2\right] \leq B_2(\alpha, \sigma^2), \quad \forall (\alpha, \sigma^2) \in \mathcal{R}_1 \cup \mathcal{R}_{2a},$$

where ψ_1 and ψ_2 are shorthands for $\psi_1(\alpha, \sigma^2)$ and $\psi_2(\alpha, \sigma^2; \delta)$. From (A.88), the above inequalities are equivalent to

$$\min\left\{\psi_1, F_1(\psi_2)\right\} \geq B_1(\alpha, \sigma^2), \tag{A.109}$$

and

$$\max\left\{\psi_2, F_1^{-1}(\psi_1)\right\} \leq B_2(\alpha, \sigma^2). \tag{A.110}$$

Note that (A.87) already proves the following

$$\psi_1 \geq B_1(\alpha, \sigma^2) \quad \text{and} \quad \psi_2 \leq B_2(\alpha, \sigma^2).$$

Hence, to prove (A.109) and (A.110), we only need to prove

$$F_1(\psi_2) \geq B_1(\alpha, \sigma^2) \quad \text{and} \quad F_1^{-1}(\psi_1) \leq B_2(\alpha, \sigma^2).$$

To prove $F_1(\psi_2) \geq B_1(\alpha, \sigma^2)$, we note that

$$\begin{aligned}
\psi_2 &\overset{(a)}{\leq} B_2(\alpha, \sigma^2) \\
&\overset{(b)}{=} \max\left\{\sigma^2, F_1^{-1}(\alpha)\right\} \\
&\overset{(c)}{=} F_1^{-1}\left(\min\left\{F_1(\sigma^2), \alpha\right\}\right) \\
&\overset{(d)}{=} F_1^{-1}\left(B_1(\alpha, \sigma^2)\right),
\end{aligned}$$

where (a) is from (A.87b), (b) is from (A.88), and (c) is due to the fact that F_1^{-1} is strictly decreasing, and (d) from (A.87). Hence, since F_1 is strictly decreasing, we have

$$F_1(\psi_2) \geq F_1\left[F_1^{-1}\left(B_1(\alpha, \sigma^2)\right)\right] = B_1(\alpha, \sigma^2).$$

Further, it is straightforward to see that if both inequalities are strict in (A.87) then

$$\min\left\{\psi_1, F_1(\psi_2)\right\} > B_1(\alpha, \sigma^2).$$

This shows that equalities of (A.108) hold only when the equalities in (A.87) hold.

The proof for $F_1^{-1}(\psi_1) \leq B_2(\alpha, \sigma^2)$ is similar and omitted.

A.3.6 Proof of Lemma 7

Suppose that $(\alpha, \sigma^2) \in \mathcal{R}_0$. From Definition 3, we have

$$\frac{\pi^2}{16} < \sigma^2 \leq \sigma_{\max}^2. \tag{A.111}$$

Further, F_1^{-1} is monotonically decreasing and hence (for $\delta > \delta_{\text{AMP}}$)

$$\frac{\pi^2}{16} = F_1^{-1}(0) > F_1^{-1}(\alpha) \geq F_2(\alpha;\delta), \tag{A.112}$$

where the last inequality is due to Lemma 4. Combining (A.111) and (A.112) yields

$$F_2(\alpha;\delta) < \sigma^2 \leq \sigma^2_{\max}. \tag{A.113}$$

By the global attractiveness property in Lemma 9 (iv), (A.113) implies

$$\psi_2(\alpha;\sigma^2;\delta) < \sigma^2.$$

From the above analysis, we see that as long as $\frac{\pi^2}{16} < \sigma_t^2 \leq \sigma^2_{\max}$ (and also $0 < \alpha_t < 1$), σ^2_{t+1} will be strictly smaller than σ_t^2:

$$\sigma^2_{t+1} = \psi_2(\alpha_t;\sigma_t^2;\delta) < \sigma_t^2.$$

Hence, there exists a finite number $T \geq 1$ such that

$$\sigma^2_{T-1} > \frac{\pi^2}{16} \quad \text{and} \quad \sigma^2_T \leq \frac{\pi^2}{16}.$$

Otherwise, σ_t^2 will converge to a $\bar{\sigma}^2$ in $\mathcal{R}_0$. This implies that $\bar{\sigma}^2$ is a fixed point of ψ_2 for certain value of $0 < \alpha \leq 1$. However, we know from part (i) of Lemma 10 and Lemma 4 that this cannot happen.

Based on a similar argument, we also have $\psi_1(\alpha;\sigma^2) < \alpha$ and so $\alpha_{t+1} < \alpha_t$ for $t \leq T-1$. Further, we can show that $\alpha_t > 0$ (i.e., $\alpha_t \neq 0$) for all $0 \leq t \leq T$. First, $\alpha_0 > 0$ follows from our assumption. Further, from (2.6a) we see that $\alpha_{t+1} > 0$ if $\alpha_t > 0$. Then, using a simple induction argument we prove that $\alpha_t > 0$ for all $0 \leq t \leq T$. Putting things together, we showed that there exists a finite number $T \geq 1$ such that

$$0 < \alpha_T \leq 1 \quad \text{and} \quad \sigma_T^2 \leq \frac{\pi^2}{16}.$$

(Recall that we have proved in Lemma 5 that $\alpha_T \leq 1$.) From Definition 3, $(\alpha_T, \sigma_T^2) \in \mathcal{R}_1 \cup \mathcal{R}_2$.

A.4 Proof of Theorem 3

We consider the two different cases separately: (1) $\delta > \delta_{\text{global}}$ and (2) $\delta < \delta_{\text{global}}$.

A.4.1 Case $\delta > \delta_{\text{global}}$

In this section, we will prove that when $\delta > \delta_{\text{global}}$ the state evolution converges to the fixed point $(\alpha, \sigma^2) = (1, 0)$ if initialized close enough to the fixed point. We first prove the following lemma, which shows that F_1^{-1} is larger than $F_2(\alpha;\delta)$ for α close to one.

Lemma 18. *Suppose that $\delta > \delta_{\text{global}} = 2$. Then, there exists an $\epsilon > 0$ such that the following holds:*

$$F_1^{-1}(\alpha) > F_2(\alpha;\delta), \quad \forall \alpha \in (1-\epsilon, 1). \tag{A.114}$$

Proof. In Lemma 4, we proved that $F_1^{-1}(\alpha) > F_2(\alpha;\delta)$ holds for all $\alpha \in (0,1)$ when $\delta > \delta_{\text{AMP}} \approx 2.5$. Here, we will prove that $F_1^{-1}(\alpha) > F_2(\alpha;\delta)$ holds for α close to 1 when $\delta > \delta_{\text{global}} = 2$. Similar to the manipulations given in Section A.3.4, the inequality (A.114) can be re-parameterized into the following:

$$\int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(\sin^2\theta + s^2)^{\frac{1}{2}}} d\theta \cdot \int_0^{\frac{\pi}{2}} \frac{(1-\gamma s^2)\sin^2\theta + s^2}{(\sin^2\theta + s^2)^{\frac{1}{2}}} d\theta > 1, \quad \forall s \in (0, \xi), \tag{A.115}$$

where $\gamma \triangleq 1 - \delta/4$ and $\xi = \phi_1^{-1}(\epsilon)$ (see (A.41) for the definition of ϕ_1). Again, it is more convenient to express (A.115) using elliptic integrals (cf. (A.52))

$$\frac{E(x)T(x)}{x} - \frac{\gamma(1-x)T^2(x)}{x^2} > 1, \quad \forall x \in \left(\frac{1}{1+\xi}, 1\right), \tag{A.116}$$

where we made a variable change $x \triangleq 1/(1+s^2)$. To this end, we can verify that

$$\lim_{x\to 1} \frac{E(x)T(x)}{x} - \frac{\gamma(1-x)T^2(x)}{x^2} = 1.$$

To complete the proof, we only need to show that the derivative of the LHS of (A.116) in a small neighborhood of $x=1$ is strictly negative when $\delta > \delta_{\text{global}} = 2$. Using the formulas listed in Section A.1, we can derive the following:

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{E(x)T(x)}{x} - \frac{\gamma(1-x)T^2(x)}{x^2}\right)\Big|_{x\to 1} \\
&= \frac{2\gamma(x-4)E(x)\cdot(1-x)K(x) + [4\gamma(1-x)+x]\cdot(1-x)K^2(x) + [2\gamma(2-x)-x]E^2(x)}{2x^3}\Big|_{x\to 1} \\
&= \gamma - \frac{1}{2},
\end{aligned}$$

where the last step is due to the facts that $E(x)=1$ and $\lim_{x\to 1}(1-x)K(x)=0$. See Section A.1 for more details. Hence, the above derivative is negative if $\gamma < \frac{1}{2}$ or $\delta > 2$ by noting the definition $\gamma = 1-\delta/4$. $\square$

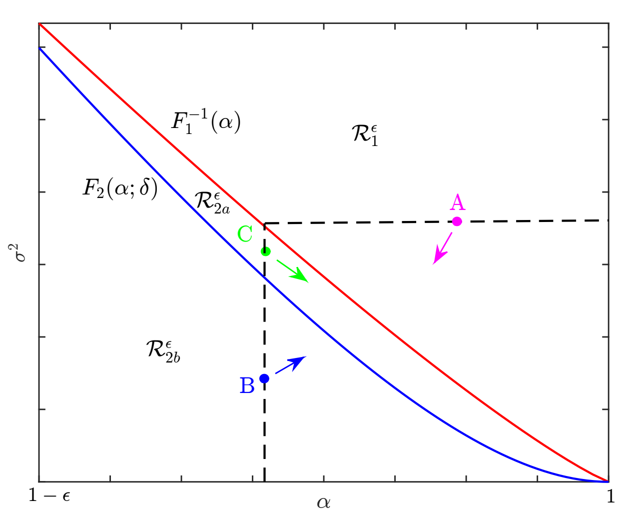

Figure 13: Illustration of the local convergence behavior when $\delta > \delta_{\text{global}}$. For all the three points shown in the figure, B_1 and B_2 are given by the dashed lines.

We now turn to the proof of Lemma 3. The idea of the proof is similar to that of Theorem 2. There are some differences though, since now δ can be smaller than δ_{AMP} and some results in the proof of Theorem 2 do not hold for the case considered here. On the other hand, as we focus on the range $\alpha \in (1-\epsilon, 1) > \alpha_*$, and under this condition we know that $F_2(\sigma^2;\delta)$ is strongly globally attracting (see Lemma 9-(v)), which means that $\psi_2(\alpha,\sigma^2)$ moves towards the fixed point $F_2(\alpha;\delta)$, but cannot move to the other side of $F_2(\alpha;\delta)$.

We continue to prove the local convergence of the state evolution. We divide the region $\mathcal{R}^\epsilon \triangleq \{(\alpha,\sigma^2)|1-\epsilon \le \alpha \le 1, 0 \le \sigma^2 \le F_1^{-1}(1-\epsilon)\}$ into the following sub-regions:

$$\begin{aligned}
\mathcal{R}_1^\epsilon &\triangleq \left\{(\alpha,\sigma^2)\middle|1-\epsilon \le \alpha \le 1, F_1^{-1}(\alpha) < \sigma^2 \le F_1^{-1}(1-\epsilon)\right\}, \\
\mathcal{R}_{2a}^\epsilon &\triangleq \left\{(\alpha,\sigma^2)\middle|1-\epsilon \le \alpha \le 1, F_2(\alpha;\delta) < \sigma^2 \le F_1^{-1}(\alpha)\right\} \\
\mathcal{R}_{2b}^\epsilon &\triangleq \left\{(\alpha,\sigma^2)\middle|1-\epsilon \le \alpha \le 1, 0 \le \sigma^2 \le F_2(\alpha;\delta)\right\}.
\end{aligned} \tag{A.117}$$

Similar to the proof of Lemma 6 discussed in Section A.3.5, we will show that if $(\alpha,\sigma^2) \in \mathcal{R}^\epsilon$ then the new states (ψ_1,ψ_2) can be bounded as follows:

$$\psi_1(\alpha,\sigma^2) \ge B_1(\alpha,\sigma^2) \quad \text{and} \quad \psi_2(\alpha,\sigma^2) \le B_2(\alpha,\sigma^2), \quad \forall(\alpha,\sigma^2) \in \mathcal{R}^\epsilon, \tag{A.118}$$

where

$$B_1(\alpha,\sigma^2)=\min\left\{\alpha,F_1(\sigma^2)\right\}\quad\text{and}\quad B_2(\alpha,\sigma^2)=\max\left\{\sigma^2,F_1^{-1}(\alpha)\right\}.$$

Based on the strong global attractiveness of ψ_1 (Lemma 8-iii) and ψ_2 (Lemma 9-v) and the additional result (A.15), it is straightforward to show the following:

$$\begin{aligned}\psi_1(\alpha,\sigma^2)\geq F_1(\sigma^2)\quad&\text{and}\quad\psi_2(\alpha,\sigma^2)\leq\sigma^2,\quad\forall(\alpha,\sigma^2)\in\mathcal{R}_1^{\epsilon},\\ \psi_1(\alpha,\sigma^2)\geq\alpha\quad&\text{and}\quad\psi_2(\alpha,\sigma^2)\leq\sigma^2,\quad\forall(\alpha,\sigma^2)\in\mathcal{R}_{2a}^{\epsilon},\\ \psi_1(\alpha,\sigma^2)\geq\alpha\quad&\text{and}\quad\psi_2(\alpha,\sigma^2)\leq F_2(\alpha;\delta),\quad\forall(\alpha,\sigma^2)\in\mathcal{R}_{2b}^{\epsilon},\end{aligned}$$

which, together with the definitions given in (A.117) and the fact that $F_2(\alpha;\delta)<F_1^{-1}(\alpha)$ (cf. Lemma 18), proves (A.118). The rest of the proof follows that in Section A.3.5. Namely, we construct two auxiliary sequences $\{\tilde{\alpha}_{t+1}\}$ and $\{\tilde{\sigma}_{t+1}^2\}$ where

$$\tilde{\alpha}_{t+1}=B_1(\alpha_t,\sigma_t^2)\quad\text{and}\quad\tilde{\sigma}_{t+1}^2=B_2(\alpha_t,\sigma_t^2),$$

and show that $\{\tilde{\alpha}_{t+1}\}$ and $\{\tilde{\sigma}_{t+1}^2\}$ monotonically converge to 1 and 0 respectively. The detailed arguments can be found in Section A.3.5 and will not be repeated here.

A.4.2 Case $\delta<\delta_{\text{global}}$

We proved in (A.25) that

$$\frac{\partial\psi_2(\alpha,\sigma^2;\delta)}{\partial\sigma^2}=\frac{4}{\delta\alpha}\left(\alpha-\underbrace{\frac{1}{2\sqrt{1+s^2}}E\left(\frac{1}{1+s^2}\right)}_{f(s)}\right),$$

where $s=\frac{\sigma}{\alpha}$. Hence, we have (note that $E(1)=1$)

$$\partial_2\psi_2(\alpha,0)\triangleq\frac{\partial\psi_2(\alpha,\sigma^2)}{\partial\sigma^2}\Big|_{\sigma^2=0}=\frac{4}{\delta}\left(1-\frac{1}{2\alpha}\right),\quad\forall\alpha>0.\tag{A.119}$$

Therefore,

$$\partial_2\psi_2(\alpha,0)>1,\quad\forall\alpha>\frac{2}{4-\delta}.$$

When $\delta<\delta_{\text{global}}=2$, we have $\frac{2}{4-\delta}<1$ and therefore there exists a constant α^* that satisfies the following:

$$\frac{2}{4-\delta}<\alpha^*<1,$$

which together with (A.119) yields

$$\partial_2\psi_2(\alpha^*,0)>1.$$

Further, as discussed in the proof of Lemma 9-(i), $\partial_2\psi_2(\alpha^*,\sigma^2)$ is a continuous function of σ^2. Hence, there exists $\xi^*>0$ such that

$$\partial_2\psi_2(\alpha^*,\sigma^2)>1,\quad\forall\sigma^2\in[0,\xi^*].\tag{A.120}$$

Further, we have shown in (A.18) that

$$\frac{\partial\psi_2(\alpha,\sigma^2;\delta)}{\partial\sigma^2}=\frac{4}{\delta}\left(1-\frac{1}{2}\int_0^{\frac{\pi}{2}}\frac{\sigma^2}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{3}{2}}}\mathrm{d}\theta\right),$$

and it is easy to see that $\partial_2\psi_2(\alpha,\sigma^2;\delta)$ is an increasing function of $\alpha\in(0,\infty)$. Hence, together with (A.120) we get the following

$$\partial_2\psi_2(\alpha,\sigma^2;\delta)>1,\quad\forall(\alpha,\sigma^2)\in[\alpha^*,1]\times[0,\xi^*],$$

which means that $\psi_2(\alpha,\sigma^2)-\sigma^2$ is a strictly increasing function of σ^2 for $(\alpha,\sigma^2)\in[\alpha^*,1]\times[0,\xi^*]$. Hence,

$$\psi_2(\alpha,\sigma^2)-\sigma^2>\psi_2(\alpha,0)=\frac{4}{\delta}(1-\alpha)^2\geq0,\quad\forall(\alpha,\sigma^2)\in[\alpha^*,1]\times[0,\xi^*].$$

This implies that σ^2 moves away from 0 in a neighborhood of the fixed point $(1,0)$.

B Derivations of AMP.A

For the convenience of the readers (especially those who are not familiar with AMP), we provide a sketch of the derivations of the AMP.A algorithm in this appendix. Our derivations follow the approach proposed in (Rangan, 2011). However, there are some differences specially in the last steps of our derivation.

For simplicity, we focus on the real-valued case. Consider the following optimization problem:

$$\min_{\boldsymbol{x}} \quad \sum_{a=1}^{m} (y_a - |(\boldsymbol{Ax})_a|)^2 + \frac{\mu}{2}\|\boldsymbol{x}\|_2^2, \tag{B.1}$$

where μ is a penalization parameter. We now sketch the derivations of the AMP.A algorithm intended for solving (B.1). First, we construct the following joint pdf for (B.1):

$$\ell(\boldsymbol{x}) = \frac{1}{Z}\prod_{a=1}^{m} \exp\left[-\beta\left(y_a - |(\boldsymbol{Ax})_a|\right)^2\right] \cdot \prod_{i=1}^{n} \exp\left(-\beta \cdot \frac{\mu}{2}x_i^2\right), \tag{B.2}$$

where Z is a normalizing constant, $(\boldsymbol{Ax})_a$ and y_a denote the a-th entries of $\boldsymbol{Ax}$ and $\boldsymbol{y}$, and $\beta > 0$ is parameter (the inverse temperature). Define

$$f(y,z) = \exp\left(-\beta\left(y - |z|\right)^2\right). \tag{B.3}$$

Following Chapter 5 in Maleki (2010), we proceed in three steps:

- Derive the sum-product belief propagation (BP) algorithm for (B.2).
- Approximate the BP update rules.
- Find the message update rules in the limit of $\beta \to \infty$.

The above procedure is slightly different from the original derivations in (Rangan, 2011) (which is derived directly from the max-sum belief propagation algorithm) but equivalent. The sum-product BP algorithm reads

$$\hat{m}^t_{a\to i}(x_i) \simeq \int_{\boldsymbol{x}\backslash i} f(y_a, (\boldsymbol{Ax})_a) \prod_{j\neq i} \mathrm{d}m^t_{j\to a}(x_j), \tag{B.4a}$$

$$m^{t+1}_{i\to a}(x_i) \simeq \prod_{b\neq a} \hat{m}^t_{b\to i}(x_i) \cdot \exp\left(-\beta \cdot \frac{\mu}{2}x_i^2\right). \tag{B.4b}$$

We next simplify the above BP update rules.

B.1 Messages from factor nodes to variable nodes

Let $x^t_{j\to a}$ and $v^t_{j\to a}/\beta$ be the mean and variance of the incoming message $m^t_{j\to a}$ (here $v^t_{j\to a}$ is $O(1)$ and the variance of $m^t_{j\to a}$ is $O(1/\beta)$ as $\beta \to \infty$ (Maleki, 2010)). Note that the calculation of the message $\hat{m}^t_{a\to i}(x_i)$ in (B.4a) can be interpreted as the expectation of $f(y_a, (\boldsymbol{Ax})_a)$ with respect to random vector $\boldsymbol{x}\backslash i$ that has product measure $\prod_{j\neq i} \mathrm{d}m^t_{j\to a}(x_j)$. Since in this interpretation the elements of $\boldsymbol{x}\backslash i$ are independent, based on a heuristic central limit theorem argument, we assume that $Z_a \triangleq (\boldsymbol{Ax})_a$ is Gaussian distributed, with mean and variance respectively given by Chapter 5.2 in Maleki (2010)

$$\begin{aligned} s^t_a &\triangleq \sum_{j\neq i} A_{aj}x^t_{j\to a} + A_{ai}x_i, \\ &= A_{ai}(x_i - x^t_{i\to a}) + \underbrace{\sum_{j=1}^{n} A_{aj}x^t_{j\to a}}_{p^t_a}, \end{aligned} \tag{B.5}$$

$$\frac{\tau^t_a}{\beta} \triangleq \frac{1}{\beta}\sum_{j\neq i} A^2_{aj}v^t_{j\to a} \approx \frac{1}{\beta}\sum_{j=1} A^2_{aj}v^t_{j\to a}.$$

Based on this approximation, the message $\hat{m}^t_{a\to i}(x_i)$ in (B.4a) can be expressed as follows

$$\begin{aligned}\hat{m}^t_{a\to i}(x_i) &= \mathbb{E}\left\{\exp\left[-\beta(y_a-|Z_a|)^2\right]\right\}\\ &= \int \exp\left[-\beta(y_a-|z|)^2\right]\cdot\mathcal{N}\left(z; A_{ai}(x_i-x^t_{i\to a})+p^t_a, \tau^t_a/\beta\right)\mathrm{d}z,\end{aligned} \tag{B.6}$$

where the expectation in step (a) is over $Z_a=(\boldsymbol{A}\boldsymbol{x})_a$ (with respect to the product distribution $\prod_{j\neq i}\mathrm{d}m^t_{j\to a}(x_j)$). Following (Rangan, 2011), we define

$$H\left(p,y,v/\beta\right) \triangleq \log\left[\int \exp\left(-\beta(y-|z|)^2\right)\cdot\mathcal{N}\left(z;p,v/\beta\right)\mathrm{d}z\right]. \tag{B.7}$$

Using this definition, we can write $\log\left[\hat{m}^t_{a\to i}\left(x_i\right)\right]$ in (B.6) as

$$\log\left[\hat{m}^t_{a\to i}\left(x_i\right)\right] = H\left(A_{ai}\left(x_i-x^t_{i\to a}\right)+p^t_a, \tau^t_a/\beta\right).$$

Noting $A_{ai}=O_p\left(\frac{1}{\sqrt{n}}\right)$, following (Rangan, 2011) we apply a second order Taylor expansion to $\log\left[\hat{m}^t_{a\to i}(x_i)\right]$ (amounts to a Gaussian approximation of $\hat{m}^t_{a\to i}(x_i)$) :

$$H\left(A_{ai}\left(x_i-x^t_{i\to a}\right)+p^t_a, y_a, \tau^t_a/\beta\right) \approx H_a(t)+A_{ai}\left(x_i-x^t_{i\to a}\right)H'_a(t)+\frac{1}{2}A^2_{ai}\left(x_i-x^t_{i\to a}\right)^2H''_a(t) \tag{B.8a}$$

$$= \frac{1}{2}A^2_{ai}H''_a(t)x_i^2+\left[A_{ai}H'_a(t)-A^2_{ai}x^t_{i\to a}H''_a(t)\right]x_i+\text{const}, \tag{B.8b}$$

where we have omitted constant terms (relative to x_i), and $H_a(t)$, $H'_a(t)$ and $H''_a(t)$ are short-hands for

$$\begin{aligned}H_a(t) &= H(p^t_a, y_a, \tau^t_a/\beta),\\ H'_a(t) &= \frac{\partial H(p,y,\tau/\beta)}{\partial p}\Big|_{p=p^t_a, y=y_a, \tau=\tau^t_a}\\ H''_a(t) &= \frac{\partial^2 H(p,y,\tau/\beta)}{\partial p^2}\Big|_{p=p^t_a, y=y_a, \tau=\tau^t_a}.\end{aligned}$$

B.2 Messages from variable nodes to factor nodes

The message from x_i to F_a is

$$m^{t+1}_{i\to a}\left(x_i\right) \simeq \prod_{b\neq a}\hat{m}^t_{b\to i}\left(x_i\right)\cdot\exp\left(-\beta\cdot\frac{\mu}{2}x_i^2\right). \tag{B.9}$$

From the Gaussian approximation in (B.8), $m^{t+1}_{i\to a}\left(x_i\right)$ is also Gaussian. Consider the following term:

$$\begin{aligned}\log\left[m^{t+1}_{i\to a}\right] &\simeq \sum_{b\neq a}\log\left[\hat{m}^t_{b\to i}(x_i)\right]-\frac{\mu\beta}{2}x_i^2\\ &\approx \frac{1}{2}\left(\sum_{b\neq a}A^2_{bi}H''_b(t)-\beta\mu\right)x_i^2+\left(\sum_{b\neq a}A_{bi}H'_b(t)-\sum_{b\neq a}A^2_{bi}H''_b(t)x^t_{i\to b}\right)x_i,\end{aligned} \tag{B.10}$$

where the second approximation comes from (B.8). Comparing (B.10) with the exponent of a Gaussian pdf, we find that its variance (which we denote by $v^{t+1}_{i\to a}/\beta$) and mean are respectively given by

$$\frac{v^{t+1}_{i\to a}}{\beta} = \frac{1}{-\sum\limits_{b\neq a}A^2_{bi}H''_b(t)+\beta\mu} = \underbrace{\frac{1}{-\sum\limits_{b=1}^{m}A^2_{bi}\cdot H''_b(t)+\beta\mu}}_{v^{t+1}_i/\beta} + O_p\left(\frac{1}{n}\right), \tag{B.11}$$

and

$$x_{i\to a}^{t+1} = \frac{v_i^{t+1}}{\beta} \cdot \left(\sum_{b\neq a} A_{bi} \cdot H_b'(t) - \sum_{b\neq a} A_{bi}^2 \cdot H_b''(t) \cdot x_{i\to b}^t \right). \tag{B.12}$$

The approximation in (B.11) is due to our assumption $\mathbb{E}[A_{bi}^2] = 1/m$. In (B.12), we have approximated $v_{i\to a}^{t+1}$ by v_i^{t+1} and omit the $O_p(1/n)$ error term.

B.3 From BP to AMP

We assume that the message $x_{i\to a}^{t+1}$ has the following structure Chapter 5.2.4 in Maleki (2010):

$$x_{i\to a}^{t+1} = x_i^{t+1} + \delta x_{i\to a}^{t+1} + O_p\left(\frac{1}{n}\right),$$

where $x_i^{t+1} = O_p(1)$ and $\delta x_{i\to a}^{t+1} \sim O_p\left(1/\sqrt{n}\right)$. From (B.12), we can identify x_i^{t+1} and $\delta x_{i\to a}^{t+1}$ (which is the term that depends on the index a) to be the following

$$x_{i\to a}^{t+1} = \underbrace{\frac{v_i^{t+1}}{\beta} \cdot \left(\sum_{b=1}^m A_{bi} \cdot H_b'(t) - \sum_{b=1}^m A_{bi}^2 \cdot H_b''(t) \cdot x_{i\to b}^t \right)}_{x_i^{t+1}} \tag{B.13a}$$

$$\underbrace{-\frac{v_i^{t+1}}{\beta} \cdot A_{ai} \cdot H_a'(t)}_{\delta x_{i\to a}^{t+1}} + \underbrace{\frac{v_i^{t+1}}{\beta} \cdot A_{ai}^2 \cdot H_a''(t) \cdot x_{i\to a}^t}_{O_p(1/n)}. \tag{B.13b}$$

We further simplify x_i^{t+1} (i.e., the first term in the above equation) as follows

$$x_i^{t+1} = \frac{v_i^{t+1}}{\beta} \cdot \left[\sum_{b=1}^m A_{bi} \cdot H_b'(t) - \sum_{b=1}^m A_{bi}^2 \cdot H_b''(t) \cdot x_{i\to b}^t \right] \tag{B.14a}$$

$$= \frac{v_i^{t+1}}{\beta} \cdot \left[\sum_{b=1}^m A_{bi} \cdot H_b'(t) - \sum_{b=1}^m A_{bi}^2 \cdot H_b''(t) \cdot x_i^t \right] + O_p\left(\frac{1}{n}\right). \tag{B.14b}$$

The approximation error in the above is $O_p(1/n)$ since

$$\sum_{b=1}^m A_{bi}^2 \cdot H_b''(t) \cdot \delta x_{i\to b}^t = -\frac{v_i^t}{\beta} \sum_{b=1}^m A_{bi}^3 \cdot H_b''(t) \cdot H_b'(t) = O_p\left(\frac{1}{n}\right),$$

where we used $\delta x_{i\to b}^t = -v_i^{t+1}/\beta \cdot A_{bi} H_b'(t)$ in the previous equation. Ignoring the $O_p(1/n)$ term, the update in (B.14) becomes

$$\begin{aligned} x_i^{t+1} &= \frac{v_i^{t+1}}{\beta} \cdot \sum_{b=1}^m A_{bi} H_b'(t) + \frac{v_i^{t+1}}{\beta} \left(-\sum_{b=1}^m A_{bi}^2 H_b''(t) \right) \cdot x_i^t \\ &= \frac{v_i^{t+1}}{\beta} \cdot \left(-\sum_{b=1}^m A_{bi}^2 H_b''(t) \right) \cdot \left(\frac{\sum_{b=1}^m A_{bi} H_b'(t)}{-\sum_{b=1}^m A_{bi}^2 H_b''(t)} + x_i^t \right). \end{aligned} \tag{B.15}$$

We now return to the update of p_a^{t+1} defined in (B.5):

$$
\begin{aligned}
p_a^{t+1} &\triangleq \sum_{j=1}^{n} A_{aj} x_{j\to a}^{t+1} \\
&\overset{(a)}{\approx} \sum_{j=1}^{n} A_{aj} \left(x_j^{t+1} - \frac{v_j^{t+1}}{\beta} \cdot A_{aj} \cdot H_a'(t) \right) \\
&= \left(\sum_{j=1}^{n} A_{aj} x_j^{t+1} \right) - \frac{\left(\sum_{j=1} A_{aj}^2 v_j^{t+1} \right)}{\beta} \cdot H_a'(t) \\
&\overset{(b)}{=} \left(\sum_{j=1}^{n} A_{aj} x_j^{t+1} \right) - \frac{\tau_a^{t+1}}{\beta} \cdot H_a'(t)
\end{aligned}
\tag{B.16}
$$

where step (a) is due to (B.13) and step (b) is from the definition in (B.5).

B.4 Large β Limit

Putting (B.5), (B.16), (B.11), (B.15), we obtain the following simplified BP update rules ($\forall a = 1, \ldots, m$ and $\forall i = 1, \ldots, n$):

$$
\tau_a^t = \sum_{j=1}^{n} A_{aj}^2 v_j^t, \tag{B.17a}
$$

$$
p_a^t = \sum_{j=1}^{n} A_{aj} x_j^t - \frac{\tau_a^t}{\beta} \cdot H_a'(t-1), \tag{B.17b}
$$

$$
v_i^{t+1} = \frac{\beta}{-\sum\limits_{b=1}^{m} A_{bi}^2 \cdot H_b''(t) + \beta\mu}, \tag{B.17c}
$$

$$
x_i^{t+1} = \frac{v_i^{t+1}}{\beta} \cdot \left(-\sum_{b=1}^{m} A_{bi}^2 H_b''(t) \right) \cdot \left(x_i^t + \frac{\sum_{b=1}^{m} A_{bi} H_b'(t)}{-\sum_{b=1}^{m} A_{bi}^2 H_b''(t)} \right), \tag{B.17d}
$$

where $H_b'(t)$ and $H_b''(t)$ are shorthands for $H'(p_b^t, y_b, \tau_b^t/\beta)$ and $H''(p_b^t, y_b, \tau_b^t/\beta)$ respectively. The algorithm summarized above is a special form the generalized AMP (GAMP) algorithm derived in (Rangan, 2011) (see Algorithm 1).

We further approximate the variance updates in (B.17a) and (B.17c) by averaging over $\boldsymbol{A}$ (based on some heuristic concentration arguments). After this approximation, τ_a^t becomes invariant to the index a (denoted as τ^t below). We can then write (B.17) into the following vector form:

$$
\begin{aligned}
\tau^t &= \frac{1}{\delta} \frac{1}{-\mathrm{div}_p(\hat{g}_{t-1})} \cdot \lambda_{t-1}, \\
\boldsymbol{p}^t &= \boldsymbol{A}\boldsymbol{x}^t - \frac{1}{\delta} \frac{\hat{g}\left(\boldsymbol{p}^{t-1}, \boldsymbol{y}, \tau^{t-1}/\beta\right)}{-\mathrm{div}_p(\hat{g}_{t-1})} \cdot \lambda_{t-1}, \\
\boldsymbol{x}^{t+1} &= \lambda_t \cdot \left(\boldsymbol{x}^t + \frac{\boldsymbol{A}^{\mathrm{T}} \hat{g}\left(\boldsymbol{p}^t, \boldsymbol{y}, \tau^t/\beta\right)}{-\mathrm{div}_p(\hat{g}_t)} \right),
\end{aligned}
\tag{B.18}
$$

where we defined:

$$
\begin{aligned}
\hat{g}(p, y, \tau/\beta) &\triangleq \frac{H'(p, y, \tau/\beta)}{\beta}, \\
\mathrm{div}_p(\hat{g}_t) &\triangleq \frac{1}{m} \sum_{a=1}^{m} \partial_p \hat{g}(p_a^t, y_a, \tau^t/\beta), \\
\lambda_t &\triangleq \frac{-\mathrm{div}_p(\hat{g}_t)}{-\mathrm{div}_p(\hat{g}_t) + \mu}.
\end{aligned}
$$

We next consider the zero-temperature limit, i.e., $\beta \to \infty$. From the definition of H in (B.7), it can be verified that (Rangan, 2011):

$$\hat{g}(p, y, \tau/\beta) = \frac{\mathbb{E}[z, p, y, \tau/\beta] - p}{\tau},$$

where $\mathbb{E}[z, p, y, \tau/\beta]$ denotes the posterior mean estimator of z w.r.t. the distribution $p(z|p, y, \tau/\beta) \propto \exp\left[-\beta(y - |z|)^2 - \beta\frac{1}{2\tau}(z-p)^2\right]$. As $\beta \to \infty$, the posterior mean concentrates around the minimum of the posterior probability, i.e., $\mathbb{E}[z, p, y, \tau/\beta] \to \text{prox}(p, y, \tau)$ where

$$\text{prox}(p, y, \tau) \triangleq \underset{z}{\text{argmin}}\ (y - |z|)^2 + \frac{(z-p)^2}{2\tau}, \tag{B.19}$$

which has the following closed-form expression (for $\tau > 0$):

$$\text{prox}(p, y, \tau) = \frac{2\tau y + |p|}{1 + 2\tau} \cdot \text{sign}(p).$$

Here, sign(0) can be arbitrarily defined to be $+1$ or -1. The function $\hat{g}$ becomes:

$$\hat{g}(p, y, \tau) = \frac{\text{prox}(p, y, \tau) - p}{\tau} = \frac{2}{1 + 2\tau} \cdot \underbrace{(y \cdot \text{sign}(p) - p)}_{g(p,y)}. \tag{B.20}$$

B.5 Summary of AMP.A

After some algebra, we can finally express (B.18) using g (instead of $\hat{g}$, see (B.20)) as the following:

$$\begin{aligned}
\tau^t &= \frac{1}{\delta} \frac{\tau^{t-1} + \frac{1}{2}}{-\text{div}_p(g_{t-1})} \cdot \lambda_{t-1}, \\
\boldsymbol{p}^t &= \boldsymbol{A}\boldsymbol{x}^t - \frac{1}{\delta} \frac{g\left(\boldsymbol{p}^{t-1}, \boldsymbol{y}\right)}{-\text{div}_p(g_{t-1})} \cdot \lambda_{t-1}, \\
\boldsymbol{x}^{t+1} &= \lambda_t \cdot \left(\boldsymbol{x}^t + \frac{\boldsymbol{A}^{\mathrm{T}} g\left(\boldsymbol{p}^t, \boldsymbol{y}\right)}{-\text{div}_p(g_t)}\right),
\end{aligned} \tag{B.21}$$

where

$$\lambda_t = \frac{-\text{div}_p(g_{t-1})}{-\text{div}_p(g_{t-1}) + \mu\left(\tau_t + \frac{1}{2}\right)}$$

There are a couple of points we want to emphasize:

- When $\mu = 0$, the update of $\boldsymbol{p}^t$ and $\boldsymbol{x}^{t+1}$ are independent of the parameter τ. This is why we prefer to use $g(p, y)$ instead of $\hat{g}(p, y, \tau)$, see (B.20).
- Calculating the divergence term $\text{div}_p(g)$ is tricky due to the discontinuity of $g(p, y)$ at $p = 0$. Unlike the complex-valued case, a simple empirical average does not work well. We postpone our discussions on this issue to a future paper.

B.6 Heuristic derivations of the state evolution

According to (2.4), the complex-valued version of AMP.A proceeds as follows

$$x_i^{t+1} = -2\text{div}_p(g_t) \cdot x_i^t + 2\underbrace{\sum_{a=1}^m \bar{A}_{ai} g(p_a^t, y_a)}_{T}, \tag{B.22a}$$

where

$$\text{div}_p(g_t) \triangleq \frac{1}{m}\sum_{a=1}^m \frac{1}{2}\left(\frac{\partial g(p_a^t, y_a)}{\partial p_a^R} - \mathrm{i}\frac{\partial g(p_a^t, y_a)}{\partial p_a^I}\right). \tag{B.22b}$$

Suppose that at each iteration the elements of $\boldsymbol{x}^t$ are distributed as

$$x_i^t \overset{d}{=} \alpha_t x_{*,i} + \sigma_t h_i, \quad \forall i = 1, \dots, n, \tag{B.23}$$

where $x_{*,i}$ represents the ith entry of the true signal vector $\boldsymbol{x}_*$ and $h_i \sim \mathcal{CN}(0,1)$ is independent of x_i^t. Rigorous proof of the state evolution framework is based on the conditioning technique developed in (Bayati & Montanari, 2011; Rangan, 2011; Javanmard & Montanari, 2013). Here, our goal is show the reader how to heuristically derive the state evolution (SE) recursion, namely, *given α_t and σ_t, how to derive α_{t+1} and σ_{t+1}.* Following (Donoho et al., 2009; Bayati & Montanari, 2011), we make the following heuristic assumptions to derive the SE:

(i) We ignore the Onsager correction term, i.e., we assume that $\boldsymbol{p}^t$ is generated as (cf. (2.4)):

$$p_a^t = \sum_j^n A_{aj} x_j^t, \quad \forall a = 1, \dots, m.$$

(ii) We assume that $\boldsymbol{x}^t$ is independent of $\boldsymbol{A}$.

We derive α_{t+1} and σ_{t+1} separately in the following two subsections.

B.6.1 Derivations of α_{t+1}

To derive α_{t+1}, we will calculate the expectation of the term T in (B.22a) by treating $\boldsymbol{x}_*$ and $\boldsymbol{x}^t$ as constants. In other words, the expectations in this section are conditioned on $\boldsymbol{x}_*$ and $\boldsymbol{x}^t$. We now consider the expectation of a single entry in T:

$$\begin{aligned}
\mathbb{E}\Big[\bar{A}_{ai} g\left(p_a^t, y_a\right)\Big] &= \mathbb{E}\left[\bar{A}_{ai} \cdot g\bigg(\sum_{j=1}^n A_{aj} x_j^t,\ \Big|\sum_{j=1}^n A_{aj} x_{*,j}\Big| + w_a\bigg)\right] \\
&= \mathbb{E}\left[\bar{A}_{ai} \sum_{j=1}^n A_{aj} x_j^t\right] \cdot \mathbb{E}\left[\partial_p g(p_a^t, y_a)\right] + \mathbb{E}\left[\bar{A}_{ai} \sum_{j=1}^n A_{aj} x_{*,j}\right] \cdot \mathbb{E}\left[\partial_z g(p_a^t, y_a)\right] \\
&= \frac{1}{m} x_i^t \cdot \mathbb{E}\left[\partial_p g(p_a^t, y_a)\right] + \frac{1}{m} x_{*,i} \cdot \mathbb{E}\left[\partial_z g(p_a^t, y_a)\right],
\end{aligned} \tag{B.24}$$

where the last step is from Stein's lemma (for complex Gaussian random variables) Lemma 2.3 in Campese (2015), and $\partial_p g(p_a^t, y_a)$ and $\partial_z g(p_a^t, |z_a| + w_a)$ are defined as

$$\begin{aligned}
\partial_p g(p, y) &\triangleq \frac{1}{2}\left(\frac{\partial}{\partial p_R} g(p, y) - \mathrm{i}\frac{\partial}{\partial p_I} g(p, y)\right), \\
\partial_z g(p, |z| + w) &\triangleq \frac{1}{2}\left(\frac{\partial}{\partial z_R} g(p, |z| + w) - \mathrm{i}\frac{\partial}{\partial z_I} g(p, |z| + w)\right),
\end{aligned}$$

where p_R and p_I are the real and imaginary parts of p (i.e., $p = p_R + \mathrm{i}p_I$) and z_R and z_I are the real and imaginary parts of z. Similar expressions also appeared in the complex AMP algorithm (CAMP) developed for solving the LASSO problem (Maleki et al., 2013). In AMP.A, $g(p, y) = y \cdot p/|p| - p$ and based on the above definitions we can derive that

$$\begin{aligned}
\partial_p g(p, y) &= \frac{y}{2|p|} - 1, \\
\partial_z g(p, |z| + w) &= \frac{\bar{z}p}{2|z|\,|p|} = \frac{1}{2} e^{\mathrm{i}(\theta_p - \theta_z)},
\end{aligned}$$

where θ_p and θ_z are the phases of p and z respectively. Note that in rigorous calculations we should be careful about the discontinuity of g. In this heuristic calculations we have ignored this issue. We will discuss this issue in our future paper. Substituting (B.24) into (B.22a) yields

$$\begin{aligned}\mathbb{E}[T] &= \frac{1}{m}\sum_{a=1}^{m}\mathbb{E}\left[\partial_p g(p_a^t, y_a)\right]\cdot x_i^t + \frac{1}{m}\sum_{a=1}^{m}\mathbb{E}\left[\partial_z g(p_a^t, y_a|)\right]\cdot x_{*,i} \\ &\approx \mathrm{div}_p(g_t)\cdot x_i^t + \mathrm{div}_z(g_t)\cdot x_{*,i},\end{aligned}\tag{B.25}$$

where in the last step we assumed that the empirical averages of the partial derivatives $\mathrm{div}_p(g_t) = \frac{1}{m}\sum_{a=1}^{m}\partial_p g(p_a^t, y_a)$ and $\mathrm{div}_z(g_t) = \frac{1}{m}\sum_{a=1}^{m}\partial_z g(p_a^t, |z_a| + w_a)$ converge to their expectations. Substituting (B.25) into (B.22a) yields

$$\begin{aligned}\mathbb{E}[x_i^{t+1}] &= -2\mathrm{div}_p(g_t)\cdot x_i^t + 2\mathbb{E}[T] \\ &= 2\mathrm{div}_z(g_t)\cdot x_{*,i}.\end{aligned}$$

From our assumption in (B.23), we have $\mathbb{E}[x_i^{t+1}] = \alpha_{t+1}\cdot x_{*,i}$. This result combined with (B.25) leads to

$$\alpha_{t+1} = 2\mathrm{div}_z(g_t).\tag{B.26}$$

Finally, when $\boldsymbol{x}$ and $\boldsymbol{x}^t$ are independent of $\boldsymbol{A}$, and by central limit theorem we can assume that both $p_a^t = \sum_{i=1}^{n} A_{ai}x_i^t$ and $z_a = \sum_{i=1}^{n} A_{ai}x_{*,i}$ are Gaussian, and their joint distribution is specified by the relationship $p_a^t \overset{d}{=} \alpha_t z_a + \sigma_t b_a$ where $z_a \sim \mathcal{CN}(0, 1/\delta)$ and $b_i \sim \mathcal{CN}(0, 1/\delta)$ are independent.

B.6.2 Derivations of σ_{t+1}^2

From (B.23), σ_{t+1}^2 can be derived as

$$\sigma_{t+1}^2 = \mathrm{var}[x_i^{t+1}] = \mathrm{var}[-2\mathrm{div}_p(g_t)\cdot x_i^t + 2T] = 4\cdot\mathrm{var}[T].\tag{B.27}$$

Further,

$$\begin{aligned}\mathbb{E}[|T|^2] &= \mathbb{E}\left[\left|\sum_{a=1}^{m} A_{ai}g(p_a^t, |z_a|)\right|^2\right] \\ &= \sum_{a=1}^{m}\mathbb{E}\Big[|A_{ai}|^2\cdot|g_a|^2\Big] + \sum_a\sum_{b\neq a}\mathbb{E}\left[\bar{A}_{ia}\bar{g}_a A_{ib}g_b\right] \\ &\overset{(a)}{\approx} \frac{1}{m}\sum_{a=1}^{m}\mathbb{E}\Big[|g_a|^2\Big] + \sum_a\sum_{b\neq a}\mathbb{E}\left[\bar{A}_{ia}\bar{g}_a\right]\cdot\mathbb{E}[A_{ib}g_b] \\ &\approx \frac{1}{m}\sum_{a=1}^{m}\mathbb{E}\Big[|g_a|^2\Big] + \frac{m(m-1)}{m^2}\cdot|\mathbb{E}[T]|^2 \\ &\approx \frac{1}{m}\sum_{a=1}^{m}\mathbb{E}\Big[|g_a|^2\Big] + |\mathbb{E}[T]|^2,\end{aligned}\tag{B.28}$$

where g_a and g_b are shorthands for $g(p_a^t, y_a)$ and $g(p_b^t, y_b)$ respectively, and step (a) follows from the heuristic assumption that the correlation between $|A_{ai}|^2$ and $|g_a|^2$, and the correlation between $A_{ia}g_a$ and $A_{ib}g_b$ can be ignored. Hence, combining (B.27) and (B.28) we obtain

$$\sigma_{t+1}^2 = 4\left(\mathbb{E}\left[|T|^2\right] - |\mathbb{E}\left[T\right]|^2\right) \approx \frac{4}{m}\sum_{a=1}^{m}\mathbb{E}\Big[|g_a(p_a^t, y_a)^2\Big],$$

where as argued below (B.26) the joint distribution of p_a^t and z_a are specified by $p_a^t \overset{d}{=} \alpha_t z_a + \sigma_t b_a$ where $z_a \sim \mathcal{CN}(0, 1/\delta)$ and $b_a \sim \mathcal{CN}(0, 1/\delta)$ are independent.

C Simplifications of SE maps

C.1 Auxiliary Results

Here we collect some auxiliary results that will be used in the simplification of the state evolution equation.

Lemma 19. *The following identities hold for any $a \in \mathbb{R}$ and $b \in \mathbb{R}_+$:*

$$\int_0^{2\pi}\int_0^{\infty} r\cos\theta \exp\left(-\frac{r^2-2ar\cos\theta}{b}\right)\mathrm{d}r\mathrm{d}\theta = 2a\sqrt{b}\sqrt{\pi}\int_0^{\frac{\pi}{2}}\cos^2\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta, \tag{C.1a}$$

$$\int_0^{2\pi}\int_0^{\infty} r\sin\theta \exp\left(-\frac{r^2-2ar\cos\theta}{b}\right)\mathrm{d}r\mathrm{d}\theta = 0. \tag{C.1b}$$

Proof. We first consider (C.1a):

$$\begin{aligned}
&\int_0^{2\pi}\int_0^{\infty} r\cos\theta\exp\left(-\frac{r^2-2a\cdot r\cos\theta}{b}\right)\mathrm{d}\theta\mathrm{d}r\\
&=\int_0^{2\pi}\cos\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta\int_0^{\infty} r\exp\left(-\frac{(r-a\cos\theta)^2}{b}\right)\mathrm{d}r\\
&\overset{(a)}{=}\int_0^{2\pi}\cos\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\left[\frac{1}{2}b\exp\left(\frac{-a^2\cos^2\theta}{b}\right)+a\cos\theta\sqrt{b\pi}\Phi\left(\frac{\sqrt{2}a\cos\theta}{\sqrt{b}}\right)\right]\mathrm{d}\theta\\
&=\int_0^{2\pi}\frac{1}{2}b\cos\theta\mathrm{d}\theta+\int_0^{2\pi}a\cos^2\theta\sqrt{b\pi}\exp\left(\frac{a^2\cos^2\theta}{b}\right)\Phi\left(\frac{\sqrt{2}a\cos\theta}{\sqrt{b}}\right)\mathrm{d}\theta\\
&\overset{(b)}{=}\int_0^{\pi}a\cos^2\theta\sqrt{b\pi}\exp\left(\frac{a^2\cos^2\theta}{b}\right)\Phi\left(\frac{\sqrt{2}a\cos\theta}{\sqrt{b}}\right)\mathrm{d}\theta+\int_0^{\pi}a\cos^2\hat{\theta}\sqrt{b\pi}\exp\left(\frac{a^2\cos^2\hat{\theta}}{b}\right)\Phi\left(-\frac{\sqrt{2}a\cos\hat{\theta}}{\sqrt{b}}\right)\mathrm{d}\hat{\theta}\\
&=\int_0^{\pi}a\cos^2\theta\sqrt{b\pi}\exp\left(\frac{a^2\cos^2\theta}{b}\right)\left[\Phi\left(\frac{\sqrt{2}a\cos\theta}{\sqrt{b}}\right)+\Phi\left(-\frac{\sqrt{2}a\cos\theta}{\sqrt{b}}\right)\right]\mathrm{d}\theta\\
&\overset{(c)}{=}a\sqrt{b\pi}\int_0^{\pi}\cos^2\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta\\
&\overset{(d)}{=}2a\sqrt{b\pi}\int_0^{\frac{\pi}{2}}\cos^2\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta,
\end{aligned} \tag{C.2}$$

where step (a) is from the integral ($\Phi(x)$ denotes the CDF of the standard Gaussian distribution):

$$\int_0^{\infty} r\exp\left(-\frac{(r-m)^2}{v}\right)\mathrm{d}r=\frac{1}{2}b\exp\left(\frac{-m^2}{v}\right)+m\sqrt{v\pi}\Phi\left(\frac{\sqrt{2}m}{\sqrt{v}}\right),\quad \forall m\in\mathbb{R}, v\in\mathbb{R}_+,$$

step (b) is from the variable change $\hat{\theta}=\theta-\pi$, step (c) is from the fact that $\Phi(x)+\Phi(-x)=1$, and step (d) is from

$$\begin{aligned}
&\int_0^{\pi}\cos^2\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta\\
&=\int_0^{\frac{\pi}{2}}\cos^2\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta+\int_{\frac{\pi}{2}}^{\pi}\cos^2\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta\\
&=\int_0^{\frac{\pi}{2}}\cos^2\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta+\int_{\frac{\pi}{2}}^{0}\cos^2\hat{\theta}\exp\left(\frac{a^2\hat{\cos}^2\theta}{b}\right)(-\mathrm{d}\hat{\theta})\quad(\hat{\theta}=\pi-\theta)\\
&=2\int_0^{\frac{\pi}{2}}\cos^2\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta.
\end{aligned}$$

The identity in (C.1b) can be derived based on similar calculations:

$$\begin{aligned}\int_0^{2\pi}\int_0^{\infty} r\sin\theta\exp\left(-\frac{r^2-2b\cdot r\cos\theta}{b}\right)\mathrm{d}\theta\mathrm{d}r &= a\sqrt{b\pi}\int_0^{\pi}\frac{1}{2}\sin 2\theta\exp\left(\frac{a^2\cos^2\theta}{b}\right)\mathrm{d}\theta\\ &= 0.\end{aligned}$$

□

Lemma 20. *Let $\tilde{Z}\sim\mathcal{N}(0,1)$ be a standard Gaussian random variable. Then, for any $x\in\mathbb{R}$, the following identities hold:*

$$\begin{aligned}\mathbb{E}\left[|\tilde{Z}|\cdot\phi\left(x|\tilde{Z}|\right)\right] &= \frac{1}{\pi}\frac{1}{1+x^2},\\ \mathbb{E}\left[\Phi\left(x|\tilde{Z}|\right)\right] &= \frac{1}{\pi}\arctan(x)+\frac{1}{2},\\ \mathbb{E}\left[\tilde{Z}^2\cdot\Phi\left(x|\tilde{Z}|\right)\right] &= \frac{1}{\pi}\arctan(x)+\frac{1}{2}+\frac{1}{\pi}\frac{x}{1+x^2},\end{aligned}\tag{C.3}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, PDF and CDF functions of the standard Gaussian distribution.

Proof. Consider the first identity:

$$\begin{aligned}\mathbb{E}\left[|\tilde{Z}|\cdot\phi\left(x|\tilde{Z}|\right)\right] &= \int_{-\infty}^{\infty}|z|\phi(x\,|z|)\phi(z)\mathrm{d}z\\ &\overset{(a)}{=} 2\int_0^{\infty} z\phi(x\,z)\phi(z)\mathrm{d}z\\ &\overset{(b)}{=} \frac{1}{\pi}\int_0^{\infty} z\exp\left[-(1+x^2)\frac{z^2}{2}\right]\mathrm{d}z\\ &= \frac{1}{\pi}\frac{1}{1+x^2},\end{aligned}\tag{C.4}$$

where (a) is from the symmetry of ϕ and (b) from the definition $\phi(x)=1/\sqrt{2\pi}e^{-x^2/2}$. Further,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}x}\mathbb{E}\left[\Phi\left(x|\tilde{Z}|\right)\right] &= \frac{\mathrm{d}}{\mathrm{d}x}\int_{-\infty}^{\infty}\Phi(x\,|z|)\phi(z)\mathrm{d}z = \frac{\mathrm{d}}{\mathrm{d}x}\int_0^{\infty}2\Phi(x\,z)\phi(z)\mathrm{d}z\\ &= \int_0^{\infty}2\frac{\mathrm{d}}{\mathrm{d}x}\Phi(x\,z)\phi(z)\mathrm{d}z = \int_0^{\infty}2z\phi(x\,z)\phi(z)\mathrm{d}z\\ &= \frac{1}{\pi}\frac{1}{1+x^2},\end{aligned}\tag{C.5}$$

where the last equality is from (C.4). Hence,

$$\mathbb{E}\left[\Phi\left(x|\tilde{Z}|\right)\right] = \int_{-\infty}^{x}\frac{1}{\pi}\frac{1}{1+t^2}\mathrm{d}t = \frac{1}{\pi}\arctan(x)+\frac{1}{2}.\tag{C.6}$$

Finally, the third identity in (C.3) can be derived as follows:

$$
\begin{aligned}
\mathbb{E}\left[\tilde{Z}^2\cdot\Phi\left(x|\tilde{Z}|\right)\right] &= \int_{-\infty}^{\infty} z^2\Phi(x\,|z|)\phi(z)\mathrm{d}z \\
&= \int_{0}^{\infty} z^2\Phi(xz)\phi(z)\mathrm{d}z \\
&\overset{(a)}{=} -2\int_{0}^{\infty} z\Phi(xz)\mathrm{d}\phi(z) \\
&= -2\left\{z\Phi(xz)\phi(z)\big|_0^\infty - \int_0^\infty \phi(z)\left[\Phi(xz)+xz\phi(xz)\mathrm{d}z\right]\right\} \\
&= 2\int_0^\infty \phi(z)\Phi(xz)\mathrm{d}z + x\cdot 2\int_0^\infty z\phi(xz)\phi(z)\mathrm{d}z \\
&\overset{(b)}{=} \frac{1}{\pi}\arctan(x)+\frac{1}{2}+\frac{1}{\pi}\frac{x}{1+x^2},
\end{aligned}
\tag{C.7}
$$

where (a) is from the identity $\phi'(z) = z\phi(z)$ and (b) from our previously derived identities in (C.4) and (C.6). □

C.2 Complex-valued AMP.A

From Definition 1, the SE equations are given by

$$
\begin{aligned}
\psi_1(\alpha,\sigma^2) &= 2\cdot\mathbb{E}\left[\partial_z g(p,Y)\right] \\
&= \mathbb{E}\left[\frac{\bar{Z}P}{|Z|\,|P|}\right], \\
\psi_2(\alpha,\sigma^2;\delta,\sigma_w^2) &= 4\cdot\mathbb{E}\left[|g(P,Y)|^2\right] \\
&= 4\cdot\mathbb{E}\left[(|Z|-|P|+W)^2\right] \\
&= \underbrace{4\cdot\mathbb{E}\left[(|Z|-|P|)^2\right]}_{\psi_2(\alpha,\sigma^2;\delta)} + 4\sigma_w^2.
\end{aligned}
\tag{C.8}
$$

In the above, $Z\sim\mathcal{CN}(0,1/\delta)$, $P=\alpha Z+\sigma B$ where $B\sim\mathcal{CN}(0,1/\delta)$ is independent of Z, and $Y=|Z|+W$ where $W\sim\mathcal{CN}(0,\sigma_w^2)$ independent of both Z and B. We first consider a special case $\sigma^2=0$ ($\alpha\neq 0$). When $\sigma=0$, we have $P=\alpha Z+\sigma B=\alpha Z$, and therefore

$$
\begin{aligned}
\psi_1(\alpha,0) &= \mathbb{E}\left[\frac{\alpha\bar{Z}Z}{\alpha|Z|\,|Z|}\right] = 1, \\
\psi_2(\alpha,0;\delta,\sigma_w^2) &= 4\cdot\mathbb{E}\left[(|Z|-|\alpha Z|)^2\right]+4\sigma_w^2 = \frac{4}{\delta}\left(1-|\alpha|\right)^2+4\sigma_w^2.
\end{aligned}
$$

We next turn to the general case where $\sigma^2\neq 0$. Later, we will see that our formulas derived for positive σ^2 covers the special case $\sigma^2=0$ as well. Lemma 21 can simplify our derivations.

Lemma 21. *ψ_1 and ψ_2 in* (C.8) *have the following properties (for any $\alpha\in\mathbb{C}\backslash 0$ and $\sigma^2\geq 0$):*

(i) $\psi_1(\alpha,\sigma^2)=\psi_1(|\alpha|,\sigma^2)\cdot e^{\mathrm{i}\theta_\alpha}$, with $e^{\mathrm{i}\theta_\alpha}$ being the phase of α;

(ii) $\psi_2(\alpha,\sigma^2;\delta)=\psi_2(|\alpha|,\sigma^2;\delta)$.

Proof. Note that for ψ_1 and ψ_2 defined in (C.8), we have $P|Z\sim\mathcal{CN}(\alpha Z,\sigma^2/\delta)$. Consider the random variable $\tilde{P}\triangleq P\cdot e^{-\mathrm{i}\theta_\alpha}$. Based on the rotational invariance of circularly-symmetric Gaussian, we have $\tilde{P}|Z\sim\mathcal{CN}(|\alpha|Z;\sigma^2/\delta)$. Hence,

$$
\psi_1(\alpha,\sigma^2)=\mathbb{E}\left[\frac{\bar{Z}P}{|Z|\,|P|}\right]=e^{\mathrm{i}\theta_\alpha}\cdot\mathbb{E}\left[\frac{\bar{Z}\tilde{P}}{|Z|\,|\tilde{P}|}\right]=e^{\mathrm{i}\theta_\alpha}\cdot\psi_1(|\alpha|,\sigma^2).
$$

The proof of $\psi_2(\alpha,\sigma^2;\delta)=\psi_2(|\alpha|,\sigma^2;\delta)$ follows from a similar argument: the joint distribution of $|Z|$ and $|P|$ does not depend on θ_α, and thus $\psi_2(\alpha,\sigma^2)=4\mathbb{E}\left[(|Z|-|P|)^2\right]$ does not depend on θ_α. $\square$

Note that Lemma 21 also holds for $\alpha=0$ if we define $\angle 0=0$.

Remark 1. *In the following, we will derive ψ_1 and ψ_2 for the case where α is real and nonnegative. The results for complex-valued α can be easily derived from those for nonnegative α, based on Lemma 21.*

We can also write ψ_1 as

$$\psi_1(\alpha,\sigma^2)=\mathbb{E}\left[\frac{\bar{Z}P}{|Z|\,|P|}\right]=\mathbb{E}[e^{\mathrm{i}(\theta_p-\theta_z)}].$$

Note that $\theta_p-\theta_z$ is the phase of an auxiliary variable $\hat{P}\triangleq e^{-\mathrm{i}\theta_z}P=\alpha|Z|+\sigma e^{-\mathrm{i}\theta_z}B$. Further, from the rotational invariance, conditioned on $|Z|$, $\hat{P}$ is distributed as $\hat{P}\sim\mathcal{CN}(\alpha|Z|,\sigma^2/\delta)$. Hence, the expectation of its phase can be calculated as

$$\begin{aligned}
\mathbb{E}\left[e^{\mathrm{i}(\theta_p-\theta_z)}\big|\,|Z|\right]&=\int_0^{2\pi}\int_0^\infty e^{\mathrm{i}\theta}\cdot\frac{1}{\pi\sigma^2/\delta}\exp\left(-\frac{\left|re^{\mathrm{i}\theta}-\alpha|Z|\right|^2}{\sigma^2/\delta}\right)\cdot r\mathrm{d}r\mathrm{d}\theta\\
&=\frac{1}{\pi\sigma^2/\delta}\exp\left(-\frac{\alpha^2|Z|^2}{\sigma^2/\delta}\right)\cdot\int_0^{2\pi}\int_0^\infty re^{\mathrm{i}\theta}\cdot\frac{1}{\pi\sigma^2/\delta}\exp\left(-\frac{r^2-2\alpha|Z|\cos\theta r}{\sigma^2/\delta}\right)\mathrm{d}r\mathrm{d}\theta\\
&=\frac{1}{\pi\sigma^2/\delta}\exp\left(-\frac{\alpha^2|Z|^2}{\sigma^2/\delta}\right)\cdot\int_0^{2\pi}\int_0^\infty r\cos\theta\cdot\frac{1}{\pi\sigma^2/\delta}\exp\left(-\frac{r^2-2\alpha|Z|\cos\theta r}{\sigma^2/\delta}\right)\mathrm{d}r\mathrm{d}\theta\\
&\quad+\mathrm{i}\frac{1}{\pi\sigma^2/\delta}\exp\left(-\frac{\alpha^2|Z|^2}{\sigma^2/\delta}\right)\cdot\int_0^{2\pi}\int_0^\infty r\sin\theta\cdot\frac{1}{\pi\sigma^2/\delta}\exp\left(-\frac{r^2-2\alpha|Z|\cos\theta r}{\sigma^2/\delta}\right)\mathrm{d}r\mathrm{d}\theta\\
&=2\int_0^{\frac{\pi}{2}}\frac{\alpha|Z|}{\sqrt{\pi}\sqrt{\sigma^2/\delta}}\cos^2\theta\exp\left(-\frac{\alpha^2|Z|^2\sin^2\theta}{\sigma^2/\delta}\right)\mathrm{d}\theta,
\end{aligned}\tag{C.9}$$

where the last step follow the following two identities together with some straightforward manipulations:

$$\int_0^{2\pi}\int_0^\infty r\cos\theta\exp\left(-\frac{r^2-2\alpha|Z|\cos\theta r}{\sigma^2/\delta}\right)\mathrm{d}r\mathrm{d}\theta=\frac{2\alpha\sigma\sqrt{\pi}}{\sqrt{\delta}}\int_0^{\frac{\pi}{2}}\cos^2\theta\exp\left(\frac{\alpha^2|Z|^2\cos^2\theta}{\sigma^2/\delta}\right)\mathrm{d}\theta,\tag{C.10a}$$

$$\int_0^{2\pi}\int_0^\infty r\sin\theta\exp\left(-\frac{r^2-2\alpha|Z|\cos\theta r}{\sigma^2/\delta}\right)\mathrm{d}r\mathrm{d}\theta=0.\tag{C.10b}$$

The above identities are proved in Lemma 19 in Appendix C.1. Using (C.9) and noting that $Z\sim\mathcal{CN}(0,1/\delta)$, we further average our result over $|Z|$:

$$\begin{aligned}
\mathbb{E}\left[e^{\mathrm{i}(\theta_p-\theta_z)}\right]&=\mathbb{E}\left\{2\int_0^{\frac{\pi}{2}}\frac{\alpha|Z|}{\sqrt{\pi}\sqrt{\sigma^2/\delta}}\cos^2\theta\exp\left(-\frac{\alpha^2|Z|^2\sin^2\theta}{\sigma^2/\delta}\right)\mathrm{d}\theta\right\}\\
&\overset{(a)}{=}\int_0^\infty 2\delta r\exp\left(-\delta r^2\right)\cdot\left(2\int_0^{\frac{\pi}{2}}\frac{\alpha r}{\sqrt{\pi}\sqrt{\sigma^2/\delta}}\cos^2\theta\exp\left(-\frac{\alpha^2r^2\sin^2\theta}{\sigma^2/\delta}\right)\mathrm{d}\theta\right)\mathrm{d}r\\
&=\frac{4\alpha\delta^{3/2}}{\sqrt{\pi}\sigma}\int_0^{\frac{\pi}{2}}\cos^2\theta\mathrm{d}\theta\int_0^\infty r^2\exp\left(-\delta\left(1+\frac{\alpha^2\sin^2\theta}{\sigma^2}\right)r^2\right)\mathrm{d}r\\
&\overset{(b)}{=}\frac{\alpha}{\sigma}\int_0^{\frac{\pi}{2}}\cos^2\theta\left(1+\frac{\alpha^2\sin^2\theta}{\sigma^2}\right)^{-\frac{3}{2}}\mathrm{d}\theta\\
&\overset{(c)}{=}\frac{\alpha}{\sigma}\int_0^{\frac{\pi}{2}}\frac{\sin^2\theta}{\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{\frac{1}{2}}}\mathrm{d}\theta\\
&=\int_0^{\frac{\pi}{2}}\frac{\alpha\sin^2\theta}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{1}{2}}}\mathrm{d}\theta,
\end{aligned}\tag{C.11}$$

where step (a) follows since the density of $|Z|$ is $f_{|Z|}(r)=\int_0^{2\pi}\delta/\pi\exp(-\delta r^2)r\mathrm{d}\theta=2\delta r\exp(-\delta r^2)$, and step (b) follows from the identity $\int_0^\infty r^2\exp(-ar^2)\mathrm{d}r=\sqrt{\pi}/4\cdot a^{-3/2}$, and (c) is derived in (A.5).

We next derive $\psi_2(\alpha,\sigma^2;\delta)$. From (C.8), we have

$$\begin{aligned}\psi_2(\alpha,\sigma^2;\delta)&=4\mathbb{E}\left[(|Z|-|P|)^2\right]\\&=4\left(\frac{1+\alpha^2+\sigma^2}{\delta}-2\cdot\mathbb{E}\left\{|ZP|\right\}\right),\end{aligned}$$

where the last step is from $Z\sim\mathcal{CN}(0,1/\delta)$ and $P\sim\mathcal{CN}(0,(\alpha^2+\sigma^2)/\delta)$. We next calculate $\mathbb{E}[|ZP|]$. Again, conditioned on $|Z|$, P is distributed as $P\sim\mathcal{CN}(\alpha|Z|,\sigma^2/\delta)$. We first calculate $\mathbb{E}[|P|\,|\,|Z|]$:

$$\begin{aligned}\mathbb{E}\left[|P|\,\big|\,|Z|\right]&=\int_{\mathbb{C}}|P|\frac{1}{\pi\sigma^2/\delta}\exp\left(-\frac{\big|P-\alpha|Z|\big|^2}{\sigma^2/\delta}\right)\mathrm{d}P\\&=\int_0^{2\pi}\int_0^\infty r\frac{1}{\pi\sigma^2/\delta}\exp\left(-\frac{|re^{\mathrm{i}\theta}-\alpha|Z||^2}{\sigma^2/\delta}\right)\cdot r\mathrm{d}r\mathrm{d}\theta\\&=\frac{1}{\pi\sigma^2/\delta}\int_0^{2\pi}\exp\left(-\frac{\alpha^2|Z|^2\sin^2\theta}{\sigma^2/\delta}\right)\mathrm{d}\theta\int_0^\infty r^2\exp\left(-\frac{(r-\alpha|Z|\cos\theta)^2}{\sigma^2/\delta}\right)\mathrm{d}r\\&=\frac{2}{\sqrt{\pi\sigma^2/\delta}}\int_0^{\frac{\pi}{2}}\left(\alpha^2|Z|^2\cos^2\theta+\frac{\sigma^2}{2\delta}\right)\exp\left(-\frac{\alpha^2|Z|^2\sin^2\theta}{\sigma^2/\delta}\right)\mathrm{d}\theta,\end{aligned}\tag{C.12}$$

where in the last step we used the following indentity

$$\int_0^\infty r^2\exp\left(-\frac{(r-m)^2}{v}\right)\mathrm{d}r=\frac{mv}{2}\exp\left(-\frac{m^2}{v}\right)+\sqrt{v\pi}\left(m^2+\frac{v}{2}\right)\Phi\left(\sqrt{\frac{2}{v}}\cdot m\right),\quad\forall m\in\mathbb{R},v\in\mathbb{R}_+$$

and some manipulations similar to those in (C.2). Following the same procedure as that in (C.11), we further calculate $\mathbb{E}[|ZP|]$ as:

$$\begin{aligned}\mathbb{E}[|ZP|]&=\int_0^\infty r\cdot 2r\delta\exp\left(-\delta r^2\right)\cdot\left(\frac{2}{\sqrt{\pi\sigma^2/\delta}}\int_0^{\frac{\pi}{2}}\left(\alpha^2r^2\cos^2\theta+\frac{\sigma^2}{2\delta}\right)\exp\left(-\frac{\alpha^2r^2\sin^2\theta}{\sigma^2/\delta}\right)\mathrm{d}\theta\right)\mathrm{d}r\\&=\int_0^{\frac{\pi}{2}}\int_0^\infty\frac{4\delta^{3/2}}{\sqrt{\pi}\sigma}\left(\alpha^2\cos^2\theta\cdot r^4+\frac{\sigma^2}{2\delta}\cdot r^2\right)\exp\left(-\delta\left(1+\frac{\alpha^2\sin^2\theta}{\sigma^2}\right)r^2\right)\mathrm{d}r\mathrm{d}\theta\\&=\frac{3\alpha^2}{2\sigma\delta}\int_0^{\frac{\pi}{2}}\cos^2\theta\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{-\frac{5}{2}}\mathrm{d}\theta+\frac{\sigma}{2\delta}\int_0^{\frac{\pi}{2}}\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{-\frac{3}{2}}\mathrm{d}\theta,\end{aligned}\tag{C.13}$$

where in the last step we used the following identities: $\int_0^\infty r^4\exp(-ar^2)\mathrm{d}r=3\sqrt{\pi}/8\cdot a^{-5/2}$ and $\int_0^\infty r^2\exp(-ar^2)\mathrm{d}r=\sqrt{\pi}/4\cdot a^{-3/2}$. Finally, using (C.13) we have

$$\begin{aligned}\psi_2(\alpha,\sigma^2;\delta)&=4\left(\frac{1+\alpha^2+\sigma^2}{\delta}-2\cdot\mathbb{E}\left\{|Z||P|\right\}\right)\\&\overset{(a)}{=}4\left\{\frac{1+\alpha^2+\sigma^2}{\delta}-2\left[\frac{3\alpha^2}{2\sigma\delta}\int_0^{\frac{\pi}{2}}\cos^2\theta\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{-\frac{5}{2}}\mathrm{d}\theta+\frac{\sigma}{2\delta}\int_0^{\frac{\pi}{2}}\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{-\frac{3}{2}}\mathrm{d}\theta\right]\right\}\\&=\frac{4}{\delta}\left\{1+\alpha^2+\sigma^2-\frac{\sigma}{2}\left[\frac{3\alpha^2}{\sigma^2}\int_0^{\frac{\pi}{2}}\cos^2\theta\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{-\frac{5}{2}}\mathrm{d}\theta+\int_0^{\frac{\pi}{2}}\left(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta\right)^{-\frac{3}{2}}\mathrm{d}\theta\right]\right\}\\&\overset{(b)}{=}\frac{4}{\delta}\left(1+\alpha^2+\sigma^2-\sigma\int_0^{\frac{\pi}{2}}\frac{1+2\frac{\alpha^2}{\sigma^2}\sin^2\theta}{(1+\frac{\alpha^2}{\sigma^2}\sin^2\theta)^{\frac{1}{2}}}\mathrm{d}\theta\right)\\&=\frac{4}{\delta}\left(1+\alpha^2+\sigma^2-\int_0^{\frac{\pi}{2}}\frac{2\alpha^2\sin^2\theta+\sigma^2}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{1}{2}}}\mathrm{d}\theta\right),\end{aligned}$$

where (a) is from (C.13), and the derivations of step (b) is more involved and are given in Lemma 2.

D Continuity of the partial derivative $\frac{\partial \psi_2(\alpha,\sigma^2)}{\partial \sigma^2}$ at $(\alpha, \sigma^2) = (1, 0)$

Note that in the proof of Lemma 9-(i) we showed that the $\lim_{(\alpha,\sigma^2)\to(1,0)} \frac{\partial \psi_2(\alpha,\sigma^2)}{\partial \sigma^2} = \frac{2}{\delta}$. Our goal here is to show that the derivative exists at $(\alpha, \sigma^2) = (1, 0)$ and it is equal to $\frac{2}{\delta}$.

D.1 Proof of the main claim

Our goal in this section is to show that $\left.\frac{\partial \psi_2(\alpha,\sigma^2)}{\partial \sigma^2}\right|_{(1,0)} = \frac{2}{\delta}$. From the definition of the partial derivative, we have

$$
\begin{aligned}
\left.\frac{\partial \psi_2(\alpha,\sigma^2)}{\partial \sigma^2}\right|_{(1,0)} &= \lim_{\sigma^2\to 0} \frac{1}{\sigma^2}(\psi_2(1,\sigma^2) - \psi_2(1,0)) \\
&= \lim_{\sigma^2\to 0} \frac{4}{\delta\sigma^2}(1+\sigma^2+1-\int_0^{\pi/2} \frac{2\sin^2\theta+\sigma^2}{(\sin^2\theta+\sigma^2)^{\frac{1}{2}}}d\theta - 2 + \int_0^{\pi/2} 2\sin\theta d\theta) \\
&= \lim_{\sigma^2\to 0} \frac{4}{\delta\sigma^2}(\sigma^2 - \int_0^{\pi/2} \frac{2\sin^2\theta+\sigma^2}{(\sin^2\theta+\sigma^2)^{\frac{1}{2}}}d\theta + 2) \qquad \text{(D.1)}
\end{aligned}
$$

Define $m \triangleq 1/\sigma^2$. Then,

$$
\begin{aligned}
\left.\frac{\partial \psi_2(\alpha,\sigma^2)}{\partial \sigma^2}\right|_{(1,0)} &= \lim_{m\to\infty} \frac{4m}{\delta}(\frac{1}{m} - \int_0^{\pi/2} \frac{2\sqrt{m}\sin^2\theta + 1/\sqrt{m}}{(m\sin^2\theta+1)^{\frac{1}{2}}}d\theta + 2) \\
&\overset{(a)}{=} \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)E(\frac{m}{m+1}) - K(\frac{m}{m+1})}{\sqrt{m(m+1)}} - \frac{1}{\sqrt{m(m+1)}}K\left(\frac{m}{m+1}\right) + 2\right) \\
&= \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)E(\frac{m}{m+1})}{\sqrt{m(m+1)}} + \frac{1}{\sqrt{m(m+1)}}K\left(\frac{m}{m+1}\right) + 2\right). \qquad \text{(D.2)}
\end{aligned}
$$

To obtain Equality (a) we have used (A.6). By employing Lemma 1 (i) we have

$$
\begin{aligned}
&\lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)E(\frac{m}{m+1})}{\sqrt{m(m+1)}} + \frac{1}{\sqrt{m(m+1)}}K\left(\frac{m}{m+1}\right) + 2\right) \\
&= \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)}{\sqrt{m(m+1)}}\left(1+\frac{1}{2}\frac{\log 4\sqrt{m+1}}{m+1} - \frac{1}{4(m+1)}\right) + \frac{1}{\sqrt{m(m+1)}}\log 4\sqrt{m+1} + 2\right) \\
&= \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)}{\sqrt{m(m+1)}}\left(1 - \frac{1}{4(m+1)}\right) + 2\right) \\
&= \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)}{\sqrt{m(m+1)}} + 2 + \frac{1}{2\sqrt{m(m+1)}}\right) \\
&= \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{m} - 2\frac{(m+1)}{\sqrt{m(m+1)}} + 2\right) + \lim_{m\to\infty} \frac{4m}{\delta}\left(\frac{1}{2\sqrt{m(m+1)}}\right) = 0 + \frac{2}{\delta}. \qquad \text{(D.3)}
\end{aligned}
$$

Again we emphasize that we have also shown in the proof of Lemma 9 that $\lim_{(\alpha,\sigma^2)\to(1,0)} \frac{\partial \psi_2(\alpha,\sigma^2)}{\partial \sigma^2} = \frac{2}{\delta}$. Hence, $\frac{\partial \psi_2(\alpha,\sigma^2)}{\partial \sigma^2}$ is continuous at $(\alpha, \sigma^2) = (1, 0)$.

E Proofs of Theorems 4

In light of Lemma 3, we assume that $\alpha_0 \geq 0$ throughout this Appendix.

E.1 Discussion

The goal of this section is to prove Theorems 4. The strategy is similar to the proof of Theorem 2. We first construct the functions F_1^{-1} and F_2. Then, we show that these two functions will intersect at exactly one point when $\delta > \delta_{\text{AMP}}$. Finally, we discuss the dynamics of the state evolution and show that (α_t, σ_t^2) converge to the intersection of F_1^{-1} and F_2. However, there are a few differences that make the proof of the noisy case more challenging:

1. Recall that in the noiseless case, the curve F_1^{-1} is entirely above F_2 (except for the fixed point $(1, 0)$) if $\delta > \delta_{\text{AMP}}$. See the plot in Fig. 9. On the other hand, when there is some noise, the curve F_2 will move up a little bit (while F_1^{-1} is unchanged) and will cross F_1 at a certain $\alpha_\star \in (0, 1)$. As shown in Fig. 14, F_1^{-1} is above F_2 for $\alpha < \alpha_\star$ and is below F_2 when $\alpha > \alpha_\star$.
2. In the noisy setting the dynamic of SE becomes more challenging. In fact (α_t, σ_t^2) can move in any direction around the fixed point. That makes the proof of convergence of (α_t, σ_t^2) more complicated.
3. In the noiseless setting the location of the fixed point of SE was $(\alpha, \sigma^2) = (1, 0)$. This is not the case for the noisy settings where the location of the fixed point depends on the noise variance.

In the sections below we go over the entire proof, but will skip the parts that are similar to the proof of the noiseless setting which was discussed in Section A.3.

E.2 Preliminaries

In the noisy setting, $\psi_1(\alpha; \sigma^2)$ remains unchanged, and $\psi_2(\alpha, \sigma^2; \delta)$ is replaced by $\psi_2(\alpha, \sigma^2; \delta, \sigma_w^2)$ below:

$$\begin{aligned}\psi_2(\alpha, \sigma^2; \delta, \sigma_w^2) &= \psi_2(\alpha, \sigma^2; \delta) + 4\sigma_w^2 &\text{(E.1a)}\\ &= \frac{4}{\delta}\left\{\alpha^2 + \sigma^2 + 1 - \alpha\left[\phi_1\left(\frac{\sigma}{\alpha}\right) + \phi_3\left(\frac{\sigma}{\alpha}\right)\right]\right\} + 4\sigma_w^2, &\text{(E.1b)}\end{aligned}$$

where

$$\begin{aligned}\phi_1(s) &\triangleq \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{(\sin^2\theta + s^2)^{\frac{1}{2}}}\mathrm{d}\theta,\\ \phi_3(s) &\triangleq \int_0^{\frac{\pi}{2}} (\sin^2\theta + s^2)^{\frac{1}{2}}\,\mathrm{d}\theta.\end{aligned} \tag{E.2}$$

Before we proceed to the analysis of ψ_1, ψ_2, F_1, and F_2, we list a few identities for ϕ_1 and ϕ_3 which will be used in our proofs later.

Lemma 22. *ϕ_1 and ϕ_3 satisfy the following properties:*

$$\begin{aligned}\phi_1(s) &= \frac{(1+s^2)E\left(\frac{1}{1+s^2}\right) - s^2 K\left(\frac{1}{1+s^2}\right)}{\sqrt{1+s^2}},\\ \phi_3(s) &= \sqrt{1+s^2}\,E\left(\frac{1}{1+s^2}\right),\\ \phi_1(0) &= 1,\\ \left.\frac{\mathrm{d}\phi_1(s)}{\mathrm{d}s^2}s^2\right|_{s=0} &= \left.\frac{s^2(E-K)}{2\sqrt{1+s^2}}\right|_{s=0} = 0,\\ \left.\frac{\mathrm{d}\phi_1(s)\phi_3(s)}{\mathrm{d}s^2}\right|_{s=0} &= \left.\frac{1}{2}\left(\frac{(1+s^2)E^2 - s^2K^2}{1+s^2}\right)^2\right|_{s=0} = \frac{1}{2},\end{aligned} \tag{E.3}$$

where E and K are shorthands for $E\left(\frac{1}{1+s^2}\right)$ and $K\left(\frac{1}{1+s^2}\right)$ respectively in the last two identities.

The proof of this lemma is a simple application of the identities we derived in Section A.1, and is hence skipped.

Our next lemma summarizes the main properties of ψ_1, ψ_2, F_1 and F_2 in the noisy phase retrieval problem.

Lemma 23. *Let $\tilde{\sigma}^2_{\max} \triangleq \sigma^2_{\max} + 4\sigma^2_w$, where $\sigma^2_{\max} = \max\{1, 4/\delta\}$. For any $\delta > \delta_{\mathrm{AMP}}$, there exists $\epsilon > 0$ such that when $0 < \sigma^2_w < \epsilon$ the following statements hold simultaneously:*

(a) For $0 \le \alpha \le 1$, we have $\psi_2(\alpha, \sigma^2; \delta, \sigma^2_w) \le \tilde{\sigma}^2_{\max}$, $\forall \sigma^2 \in [0, \tilde{\sigma}^2_{\max}]$.

(b) For $0 \le \alpha \le 1$, $\sigma^2 = \psi_2(\alpha, \sigma^2; \delta) + 4\sigma^2_w$ admits a unique globally attracting fixed point, denoted as $F_2(\alpha; \delta, \sigma^2_w)$, in $\sigma^2 \in [0, \tilde{\sigma}^2_{\max}]$. Further, if $\alpha \ge \alpha_*$ (note that $\alpha_* \approx 0.53$ is defined in (A.17)*), then $F_2(\alpha; \delta, \sigma^2_w)$ is strongly globally attractive. Finally, $F_2(\alpha; \delta, \sigma^2_w)$ is a continuous function of σ^2_w.*

(c) The equation $F_1^{-1}(\alpha) = F_2(\alpha; \delta, \sigma^2_w)$ has a unique nonzero solution in $\alpha \in [0, 1]$. Let $\alpha_\star(\delta, \sigma^2_w)$ be that unique solution. Then, $F_1^{-1}(\alpha) > F_2(\alpha; \delta, \sigma^2_w)$ for $0 \le \alpha < \alpha_\star(\delta, \sigma^2_w)$ and $F_1^{-1}(\alpha) < F_2(\alpha; \delta, \sigma^2_w)$ for $\alpha_\star(\delta, \sigma^2_w) < \alpha \le 1$.

(d) There exists $\hat{\alpha}(\delta, \sigma^2_w)$, such that $F_2(\alpha; \delta, \sigma^2_w)$ is strictly decreasing on $\alpha \in (0, \hat{\alpha}(\delta, \sigma^2_w))$ and strictly increasing on $(\hat{\alpha}(\delta, \sigma^2_w), 1)$. Further, $\alpha_\star(\delta, \sigma^2_w) < \hat{\alpha}(\delta, \sigma^2_w) < 1$.

(e) Define $L(\alpha; \delta, \sigma^2_w) \triangleq L(\alpha; \delta) + 4\sigma^2_w$, where $L(\alpha; \delta)$ is defined in (A.58)*. Then, $L(\alpha; \delta, \sigma^2_w) < F_1^{-1}(\alpha)$ for all $\alpha \in (0, \alpha_*]$, where $\alpha_* \approx 0.53$ is defined in* (A.17)*.*

(f) For any $\alpha \in (0, \alpha_*]$ and $\sigma^2 \in [L(\alpha; \delta, \sigma^2_w), F_1^{-1}(\alpha)]$, we have $\psi_2(\alpha, \sigma^2; \delta, \sigma^2_w) \triangleq \psi_2(\alpha, \sigma^2; \delta) + 4\sigma^2_w < F_1^{-1}(\alpha)$.

(g) $F_2(1; \delta, \sigma^2_w) < F_1^{-1}(\alpha_*)$.

Proof. In the following, we will prove that each part of the lemma holds when σ^2_w is smaller than a constant. Hence, the statements hold simultaneously when σ^2_w is smaller than the minimum of those constants.

Part (a): In Lemma 9-(iii) we proved that, for the noiseless setting, $\psi_2(\alpha; \sigma^2; \delta) \le \sigma^2_{\max}$ for $\sigma^2 \in [0, \sigma^2_{\max}]$. If fact, it is easy to verify that our proof can be strengthened to $\psi_2(\alpha; \sigma^2; \delta) \le \sigma^2_{\max}$ for $\sigma^2 \in [0, 2]$, see (A.29). Note that $\sigma^2_{\max} = \max\{1, 4/\delta\} \le 4/\delta_{\mathrm{AMP}} \approx 1.6$. Hence, $\psi_2(\alpha; \sigma^2; \delta) \le \sigma^2_{\max}$ for $\sigma^2 \in [0, \tilde{\sigma}^2_{\max}] = \sigma^2_{\max} + 4\sigma^2_w$ when σ^2_w is small. Further, $\psi_2(\alpha; \sigma^2; \delta, \sigma^2_w) = \psi_2(\alpha; \sigma^2; \delta) + 4\sigma^2_w$, and hence $\psi_2(\alpha; \sigma^2; \delta, \sigma^2_w) \le \tilde{\sigma}^2_{\max}$ for $\sigma^2 \in [0, \tilde{\sigma}^2_{\max}]$.

Part (b): The claim is a consequence of three facts: (i) $\psi_2(\alpha, \sigma^2; \delta, \sigma^2_w) \le \sigma^2$ at $\sigma^2 = \tilde{\sigma}^2_{\max}$; (ii) $\frac{\partial \psi_2(\alpha, \sigma^2; \delta, \sigma^2_w)}{\partial \sigma^2} < 1$ when $\sigma^2 \in [0, \tilde{\sigma}^2_{\max}]$, and (iii) if $\alpha \ge \alpha_*$, then $\frac{\partial \psi_2(\alpha, \sigma^2; \delta, \sigma^2_w)}{\partial \sigma^2} > 0$ for any $\sigma^2 \ge 0$. Fact (i) has been proved in part (a) of this lemma. For Fact (ii), recall that in (A.30) we have proved $\frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2} < 1$ when $\sigma^2 \in [0, \sigma^2_{\max}]$. Again, similar to part (a) of this lemma, we can argue that the result actually holds for $\sigma^2 \in [0, \tilde{\sigma}^2_{\max}]$. We prove Fact (ii) by further noting $\psi_2(\alpha, \sigma^2; \delta, \sigma^2_w) = \psi_2(\alpha, \sigma^2; \delta) + 4\sigma^2_w$ and hence $\frac{\partial \psi_2(\alpha, \sigma^2; \delta, \sigma^2_w)}{\partial \sigma^2} = \frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2}$. Fact (iii) follows from Lemma 9-(v) and the fact that $\frac{\partial \psi_2(\alpha, \sigma^2; \delta, \sigma^2_w)}{\partial \sigma^2} = \frac{\partial \psi_2(\alpha, \sigma^2; \delta)}{\partial \sigma^2}$.

We now show that $F_2(\alpha; \delta, \sigma^2_w)$ is a continuous function of σ^2_w. Let x be an arbitrary constant in $(0, \epsilon)$. Suppose that $\lim_{\sigma^2_w \to x^-} F_2(\alpha; \delta, \sigma^2_w) = y_1$ and $\lim_{\sigma^2_w \to x^+} F_2(\alpha; \delta, \sigma^2_w) = y_2$, where $y_1, y_2 \in [0, \tilde{\sigma}^2_{\max}]$ and $y_1 \ne y_2$. Since F_2 is the fixed point of ψ_2, we then have $y_1 = \psi_2(\alpha, y_1; \delta) + 4x$ and $y_2 = \psi_2(\alpha, y_2; \delta) + 4x$, which leads to $y_1 - \psi_2(\alpha, y_1; \delta) = y_2 - \psi_2(\alpha, y_2; \delta)$. However, we have shown in Lemma 9 that $\tilde{\Psi}_2(\alpha, \sigma^2; \delta) \triangleq \sigma^2 - \psi_2(\alpha, \sigma^2; \delta) - C$ is a strictly increasing function of σ^2 in $[0, \tilde{\sigma}^2_{\max}]$, and hence for any $C \in \mathbb{R}$ there cannot be two solutions to $\tilde{\Psi}_2(\alpha, \sigma^2; \delta) = 0$. This leads to contradiction.

Part (c): It is more convenient to introduce a variable change:

$$s \triangleq \phi_1^{-1}(\alpha) \quad \text{and} \quad s_\star(\delta, \sigma^2_w) = \phi_1^{-1}\left(\alpha_\star(\delta, \sigma^2_w)\right).$$

As have been argued in Section A.3.4, $F_1^{-1}(\alpha) \le F_1^{-1}(0) = \pi^2/16 < \tilde{\sigma}^2_{\max}$. Then, by the global attractiveness of $F_2(\alpha; \delta, \sigma^2_w)$ (part (b) of this lemma) and noting that $\phi_1 : [0, \infty] \mapsto [0, 1]$ is a decreasing function, our claim can be equivalently refomulated as

$$\psi_2\left(\phi_1(s), s^2\phi_1^2(s); \delta\right) + 4\sigma^2_w > s^2\phi_1^2, \quad \forall s \in [0, s_\star(\delta, \sigma^2_w)), \tag{E.4}$$

and

$$\psi_2\left(\phi_1(s), s^2\phi_1^2(s);\delta\right) + 4\sigma_w^2 < s^2\phi_1^2, \quad \forall s > s_\star(\delta,\sigma_w^2).$$

From the definition of ψ_2 in (E.1) and after straightforward manipulations, we can write (E.4) into

$$T(s^2,\delta,\sigma_w^2) < 0, \quad \forall s \in [0, s_\star(\delta,\sigma_w^2)) \quad \text{and} \quad T(s^2,\delta,\sigma_w^2) > 0, \quad \forall s > s_\star(\delta,\sigma_w^2), \tag{E.5}$$

where

$$T(s^2,\delta,\sigma_w^2) \triangleq \left(1-\frac{4}{\delta}\right)\phi_1^2(s)s^2 + \frac{4}{\delta}\phi_1(s)\phi_3(s) - \left(\frac{4}{\delta}+4\sigma_w^2\right). \tag{E.6}$$

From (E.5), we have

$$\frac{\partial T(s^2,\sigma_w^2)}{\partial s^2} = \left(1-\frac{4}{\delta}\right)\left(\phi_1^2(s) + 2\phi_1(s)\frac{\mathrm{d}\phi_1(s)}{\mathrm{d}s^2}s^2\right) + \frac{4}{\delta}\frac{\mathrm{d}\phi_1(s)\phi_3(s)}{\mathrm{d}s^2}. \tag{E.7}$$

Applying the identities listed in (E.3), we obtain

$$\left.\frac{\partial T(s^2,\sigma_w^2)}{\partial s^2}\right|_{s=0} = 1-\frac{2}{\delta} > 0.$$

Further, $\frac{\partial T(s^2,\sigma_w^2)}{\partial s^2}$ is a continuous function at $s^2=0$, and thus there exists $\epsilon>0$ such that

$$\frac{\partial T(s^2,\sigma_w^2)}{\partial s^2} > 0, \quad \forall s^2 \in [0,\epsilon].$$

The above result shows that $T(s^2,\sigma_w^2)$ is monotonically increasing in $s^2\in[0,\epsilon]$. Further, from (E.6) we have

$$T(s^2,\delta,\sigma_w^2) = T(s^2,\delta,0) - 4\sigma_w^2.$$

It is straightforward to show that $T(0,\delta,\sigma_w^2) = -\sigma_w^2 < 0$. Hence, $T(s^2,\delta,\sigma_w^2)=0$ has a unique solution if the following holds:

$$\inf_{s^2\geq\epsilon} T(s^2,\delta,\sigma_w^2) > 0,$$

or equivalently

$$4\sigma_w^2 < \inf_{s^2\geq\epsilon} T(s^2,\delta,0). \tag{E.8}$$

Lemma 4 proves that $F_1^{-1}(\alpha) > F_2(\alpha;\delta)$ for $\alpha\in(0,1)$ for any $\delta>\delta_{\mathrm{AMP}}$, which, after re-parameterization implies that $T(s^2,\delta,0)>0$ for $s>0$ if $\delta>\delta_{\mathrm{AMP}}$. Hence, $\inf_{s^2\geq\epsilon} T(s^2,\delta,0)$ is strictly positive, and there exists sufficiently small σ_w^2 such that (E.8) holds.

Part (d): From the fixed point equation $F_2 = \psi_2(\alpha,F_2;\delta,\sigma_w^2)$ where (F_2 denotes $F_2(\alpha;\delta,\sigma_w^2)$), we can derive the following (cf. (A.33))

$$\left(1-\partial_2\psi_2(\alpha,F_2;\delta,\sigma_w^2)\right)\cdot\frac{\mathrm{d}F_2(\alpha;\delta,\sigma_w^2)}{\mathrm{d}\alpha} = \partial_1\psi_2(\alpha,F_2;\delta,\sigma_w^2).$$

Similar to the proof of part (b), $1-\partial_2\psi_2(\alpha,F_2;\delta,\sigma_w^2)>0$ when σ_w^2 is sufficiently small. Hence, proving $\partial_1\psi_2(\alpha,F_2;\delta,\sigma_w^2)<0$ is simplified to proving that there exists $\hat{\alpha}(\delta,\sigma_w^2)$ such that

$$\partial_1\psi_2(\alpha,F_2;\delta,\sigma_w^2) < 0, \quad \forall\alpha\in\left(0,\hat{\alpha}(\delta,\sigma_w^2)\right), \tag{E.9a}$$

and

$$\partial_1\psi_2(\alpha,F_2;\delta,\sigma_w^2) > 0, \quad \forall\alpha\in\left(\hat{\alpha}(\delta,\sigma_w^2),\, 1\right). \tag{E.9b}$$

From (2.6) and after some calculations, we obtain the following

$$\begin{aligned}\frac{\partial\psi_2(\alpha,\sigma^2;\delta,\sigma_w^2)}{\partial\alpha} &= \frac{4}{\delta}\left(2\alpha - \int_0^{\frac{\pi}{2}}\frac{2\alpha^3\sin^4\theta + 3\alpha\sigma^2\sin^2\theta}{(\alpha^2\sin^2\theta+\sigma^2)^{\frac{3}{2}}}\mathrm{d}\theta\right)\\ &= \frac{4}{\delta}\left(2\alpha - 2\underbrace{\int_0^{\frac{\pi}{2}}\frac{\sin^4\theta+\frac{3}{2}s^2\sin^2\theta}{(\sin^2\theta+s^2)^{\frac{3}{2}}}\mathrm{d}\theta}_{h(s)}\right),\end{aligned} \tag{E.10}$$

where $s \triangleq \sigma/\alpha$. Then, we can reformulate (E.9) as

$$\alpha < h\left(\frac{\sqrt{F_2(\alpha;\delta,\sigma_w^2)}}{\alpha}\right), \quad \forall \alpha \in \left(0, \hat{\alpha}(\delta,\sigma_w^2)\right),$$

and

$$\alpha > h\left(\frac{\sqrt{F_2(\alpha;\delta,\sigma_w^2)}}{\alpha}\right), \quad \forall \alpha \in \left(\hat{\alpha}(\delta,\sigma_w^2),\, 1\right).$$

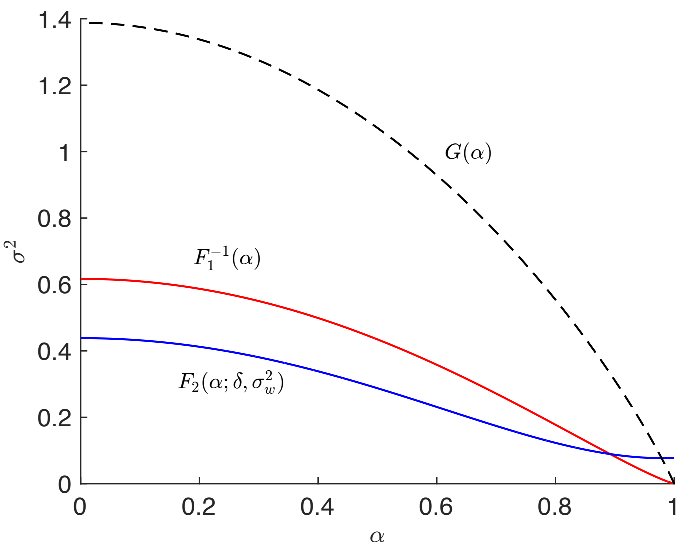

Figure 14: Depiction of $F_1^{-1}(\alpha)$, $F_2(\alpha;\delta,\sigma_w^2)$ and $G(\alpha)$. $\alpha_\star(\delta,\sigma_w^2)$: solution to $F_1^{-1}(\alpha) = F_2(\alpha;\delta,\sigma_w^2)$. $\hat{\alpha}(\delta,\sigma_w^2)$: solution to $G^{-1}(\alpha) = F_2(\alpha;\delta,\sigma_w^2)$.

From the definition given in (E.10), it is easy to show that $h : \mathbb{R}_+ \mapsto [0,1]$ is a decreasing function. Then, the above inequality can be further simplified to

$$F_2(\alpha;\delta,\sigma_w^2) < \left[\alpha \cdot h^{-1}(\alpha)\right]^2 \triangleq G(\alpha), \quad \forall \alpha \in \left(0, \hat{\alpha}(\delta,\sigma_w^2)\right), \tag{E.11a}$$

and

$$F_2(\alpha;\delta,\sigma_w^2) > \left[\alpha \cdot h^{-1}(\alpha)\right]^2 = G(\alpha), \quad \forall \alpha \in \left(\hat{\alpha}(\delta,\sigma_w^2),\, 1\right). \tag{E.11b}$$

Similar to (E.4) and (E.5), (E.11) can be re-parameterized as

$$\psi_2\left(h(s), s^2\phi_1^2(s);\delta\right) + 4\sigma_w^2 > s^2h^2, \quad \forall s < \hat{s}(\delta,\sigma_w^2), \tag{E.12}$$

and

$$\psi_2\left(\phi_1(s), s^2\phi_1^2(s);\delta\right) + 4\sigma_w^2 < s^2h^2, \quad \forall s > \hat{s}(\delta,\sigma_w^2), \tag{E.13}$$

where $\hat{s}(\delta,\sigma_w^2) \triangleq h^{-1}\left(\hat{\alpha}(\delta,\sigma_w^2)\right)$. We skip the proof for (E.12) since it is very similar to the proof of part (c) of this lemma. (Note that to apply the above re-parameterization (which is based on the global attractiveness of F_2, i.e., part (b) of this lemma), we need to ensure $G(\alpha) < \tilde{\sigma}^2_{\max}$. This can be seen from the fact that $G(\alpha) \le G(0) = (3\pi/8)^2 \approx 1.38$ while $\tilde{\sigma}^2_{\max} + 4\sigma_w^2$ and $\sigma^2_{\max} = \max\{1, 4/\delta\} > \max\{1, 4/\delta_{\mathrm{AMP}}\} \approx 1.6$.)

Finally, to show $\hat{\alpha}(\delta,\sigma_w^2) > \alpha_\star(\delta,\sigma_w^2)$, we will prove that $G(\alpha) > F_1^{-1}(\alpha)$ for $\alpha \in [0,1)$. See the plot in Fig. 14. Since $G(\alpha) = [\alpha \cdot h^{-1}(\alpha)]^2$ and $F_1^{-1}(\alpha) = [\alpha \cdot \phi_1^{-1}(\alpha)]^2$, we only need to prove $h^{-1}(\alpha) > \phi_1^{-1}(\alpha)$. Noting that both ϕ_1 and h are monotonically decreasing functions, it suffices to prove $h(s) > \phi_1(s)$ for $s > 0$, which directly follows from their definitions (cf. (E.10) and (A.59a)):

$$\begin{aligned}
h(s)-\phi_1(s) &= \int_0^{\frac{\pi}{2}} \frac{\sin^4\theta+\frac{3}{2}s^2\sin^2\theta}{(\sin^2\theta+s^2)^{\frac{3}{2}}}\mathrm{d}\theta - \int_0^{\frac{\pi}{2}} \frac{\sin^2\theta}{\left(\sin^2\theta+s^2\right)^{\frac{1}{2}}}\mathrm{d}\theta \\
&= \int_0^{\frac{\pi}{2}} \frac{\frac{1}{2}s^2\sin^2\theta}{(\sin^2\theta+s^2)^{\frac{3}{2}}} > 0, \quad \forall s>0.
\end{aligned}$$

Part (e): First note that $L(\alpha;\delta,\sigma_w^2)=L(\alpha;\delta)+4\sigma_w^2$. Hence, the proof for the claim is straightforward if the inequality $L(\alpha;\delta)<F_1^{-1}(\alpha)$ is strict for $\alpha\leq\alpha_*$. This is the case since Lemma 13 shows that $L(\alpha;\delta)\leq F_1^{-1}(\alpha)$ for $\alpha\leq 1$, but equality only happends at $\alpha=1$.

Part (f): In Lemma 17, we have proved the following result in the case of $\sigma_w^2=0$:

$$\psi_2(\alpha,\sigma^2;\delta)<F_1^{-1}(\alpha), \quad \forall 0\leq\alpha\leq\alpha_*,\ L(\alpha;\delta)<\sigma^2<F_1^{-1}(\alpha).$$

(In fact, the above inequality holds for α up to one.) In the noisy case, ψ_2 increases a little bit: $\psi_2(\alpha,\sigma^2;\delta,\sigma_w^2)=\psi_2(\alpha,\sigma^2;\delta)+4\sigma_w^2$. Hence, when σ_w^2 is sufficiently small, we still have

$$\psi_2(\alpha,\sigma^2;\delta,\sigma_w^2)<F_1^{-1}(\alpha), \quad \forall 0\leq\alpha\leq\alpha_*,\ L(\alpha;\delta)<\sigma^2<F_1^{-1}(\alpha). \tag{E.14}$$

Clearly, the inequality in (E.14) also holds for $L(\alpha;\delta,\sigma_w^2)<\sigma^2<F_1^{-1}(\alpha)$, since $L(\alpha;\delta,\sigma_w^2)=L(\alpha;\delta)+4\sigma_w^2>L(\alpha;\delta)$.

Part (g): Note that $F_1^{-1}(\alpha_*)\approx F_1^{-1}(0.53)>0$ does not depend on σ_w^2. Further, $F_2(1;\delta,0)=0$ and $F_2(1;\delta,\sigma_w^2)$ is a continuous function of σ_w^2. Hence, $F_2(1;\delta,\sigma_w^2)<F_1^{-1}(\alpha_*)$ for small enough σ_w^2. □

E.3 Convergence of the SE

Our next lemma proves that the state evolution still converges to the desired fixed point for $0<\alpha_0\leq 1$ and $\sigma_0^2\leq 1$ if $\delta>\delta_{\text{AMP}}$.

Lemma 24. *Let $\{\alpha_t\}_{t\geq1}$ and $\{\sigma_t^2\}_{t\geq1}$ be two state sequences generated according to* (2.5) *from α_0 and σ_0^2. Let ϵ be the constant required in Lemma 23. Then, the following holds for any $\delta>\delta_{\text{AMP}}$, $0<\sigma_w^2<\epsilon$, and $0<\alpha_0\leq1$ and $\sigma_0^2\leq1$:*

$$\lim_{t\to\infty}\alpha_t=\alpha_\star(\delta,\sigma_w^2) \quad \textit{and} \quad \lim_{t\to\infty}\sigma_t^2=\sigma_\star^2(\delta,\sigma_w^2),$$

where $\alpha_\star(\delta,\sigma_w^2)$ is the unique positive solution to $F_1^{-1}(\alpha)=F_2(\alpha;\delta,\sigma_w^2)$ and $\sigma_\star^2(\delta,\sigma_w^2)=F_1^{-1}(\alpha_\star(\delta,\sigma_w^2))$.

Proof. From Lemma 23-(a), when σ_w^2 is small enough, $(\alpha_t,\sigma_t^2)\in\mathcal{R}$ for all $t\geq1$, where $\mathcal{R}\triangleq\{(\alpha,\sigma^2)|0<\alpha\leq1,\ 0\leq\sigma^2\leq\tilde{\sigma}_{\max}^2\}$, where $\tilde{\sigma}_{\max}^2=\max\{1,4/\delta\}+4\sigma_w^2$. We divide $\mathcal{R}$ into several regions and discuss the dynamical behaviors of the state evolution for different regions separately. Specifically, we define

$$\begin{aligned}
\mathcal{R}_0 &\triangleq \left\{(\alpha,\sigma^2)|0<\alpha\leq1,\ \pi^2/16<\sigma^2\leq\tilde{\sigma}_{\max}^2\right\}, \\
\mathcal{R}_1 &\triangleq \left\{(\alpha,\sigma^2)|F_1^{-1}(\alpha_*)\leq\sigma^2\leq\pi^2/16,\ F_1(\sigma^2)\leq\alpha\leq1\right\}, \\
\mathcal{R}_2 &\triangleq \left\{(\alpha,\sigma^2)|0<\alpha\leq\alpha_*,\ 0\leq\sigma^2<F_1^{-1}(\alpha)\right\}, \\
\mathcal{R}_3 &\triangleq \left\{(\alpha,\sigma^2)|\alpha_*\leq\alpha\leq1,\ 0\leq\sigma^2<F_1^{-1}(\alpha_*)\right\},
\end{aligned} \tag{E.15}$$

where $\alpha_*\approx0.53$ was defined in (A.17). Notice that $\alpha_\star(\delta,0)=1$, and therefore it is guaranteed that $\alpha_\star(\delta,\sigma_w^2)>\alpha_*$ for small enough σ_w^2. See Fig. 15 for illustration. To prove the lemma, we will prove the following arguments:

(i) If $(\alpha_{t_0},\sigma_{t_0}^2)\in\mathcal{R}_0$, then there exists a finite $T_1\geq1$ such that $(\alpha_{t_0+T_1},\sigma_{t_0+T_1}^2)\in\mathcal{R}\backslash\mathcal{R}_0$.

(ii) If $(\alpha_{t_0},\sigma_{t_0}^2)\in\mathcal{R}_1\cup\mathcal{R}_2$ for $t_0\geq1$ (i.e., after one iteration), then there exists a finite $T_2\geq1$ such that $(\alpha_{t_0+T_2},\sigma_{t_0+T_2}^2)\in\mathcal{R}_3$.

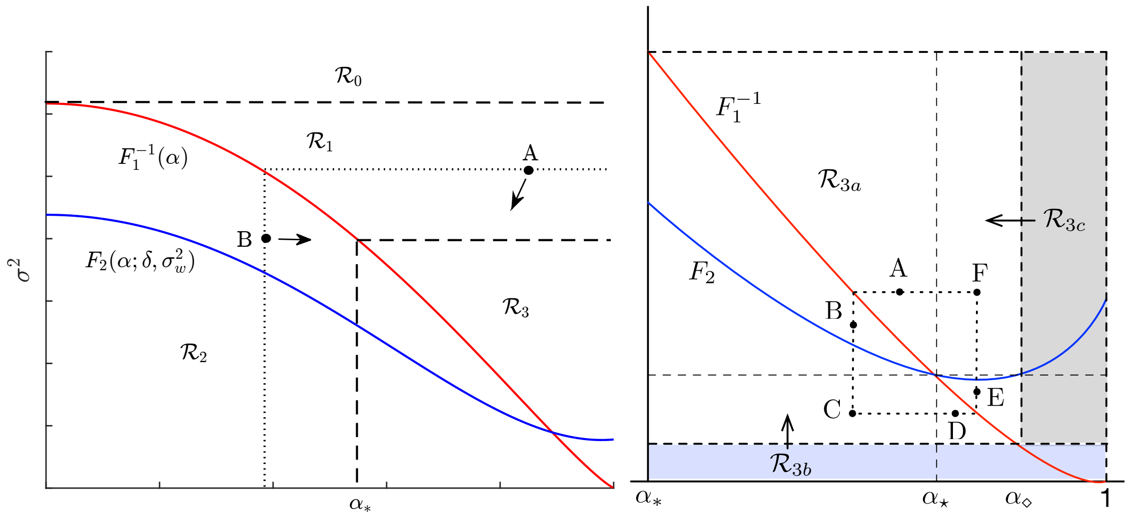

Figure 15: Dynamical behavior the state evolution in the low noise regime. **left:** points in $\mathcal{R}_1$ and $\mathcal{R}_2$ will eventually move to $\mathcal{R}_3$. Here, $\alpha_* \approx 0.53$. **Right:** Illustration of $\mathcal{R}_3$. Points in $\mathcal{R}_{3b}$ and $\mathcal{R}_{3c}$ will eventually move to $\mathcal{R}_{3a}$. For points in $\mathcal{R}_{3a}$ (marked A, B, C, D, E, F), we can form a small rectangular region that bounds the remaining trajectory. Note that the lower and right bounds for A and B (and also the upper and left bounds for D and E) are given by $\sigma_\star^2$ and $\alpha_\star$ respectively.

(iii) We show that if $(\alpha_{t_0}, \sigma_{t_0}^2) \in \mathcal{R}_3$ for $t_0 \geq 0$, then $(\alpha_t, \sigma_t^2) \in \mathcal{R}_3$ for all $t > t_0$, and (α_t, σ_t^2) converges to $(\alpha_\star, \sigma_\star^2)$.

The proof of (i) is similar to that of Lemma 7 and therefore omitted here.

Proof of (ii): Following the proof of Lemma 6, we argue that if $(\alpha_t, \sigma_t^2) \in \mathcal{R}_1 \cup \mathcal{R}_2$ then the following holds

$$\alpha_{t+1} \geq B_1(\alpha_t, \sigma_t^2) \quad \text{and} \quad \sigma_{t+1}^2 \geq B_2(\alpha_t, \sigma_t^2), \tag{E.16}$$

where $B_1(\alpha_t, \sigma_t^2) = \min\left\{\alpha_t, F_1(\sigma_t^2)\right\}$ and $B_2(\alpha_t, \sigma_t^2) = \max\left\{\sigma_t^2, F_1^{-1}(\alpha_t)\right\}$. Then, it is easy to show that $(\alpha_{t+1}, \sigma_{t+1}^2) \in \mathcal{R}_1 \cup \mathcal{R}_2 \cup \mathcal{R}_3$. Applying this recursively, we see that (α, σ^2) either moves to $\mathcal{R}_3$ at a certain time or stays in $\mathcal{R}_1 \cup \mathcal{R}_2$. We next prove that the latter case cannot happen. Suppose that $(\alpha_t, \sigma_t^2) \in \mathcal{R}_1 \cup \mathcal{R}_2$ for $t \geq t_0$. If this is the case, then it can be shown that

$$B_1(\alpha_t, \sigma_t^2) \leq B_1(\alpha_{t+1}, \sigma_{t+1}^2) \quad \text{and} \quad B_2(\alpha_t, \sigma_t^2) \geq B_2(\alpha_{t+1}, \sigma_{t+1}^2), \quad \forall t > t_0. \tag{E.17}$$

On the other hand, since we assume $(\alpha_t, \sigma_t^2) \in \mathcal{R}_1 \cup \mathcal{R}_2$ for $t \geq t_0$, B_1 is upper bounded by α_* and B_2 lower bounded by $F_1^{-1}(\alpha_*)$. Hence, this means the sequences B_1 and B_2 converges to α_* and $F_1^{-1}(\alpha_*)$, respectively. This cannot happen since there is no fixed point in $\mathcal{R}_1 \cup \mathcal{R}_2$.

The proof for (E.16) and (E.17) are basically the same as those for the noiseless counterparts and hence skipped here. Please refer to the proof of Lemma 6. We only need to show that some of the key inequalities used in the proof of Lemma 6 still hold in the noisy case, which have been listed in Lemma 23 (e) and (f).

Proof of (iii): Lemma 23-(c), (d) and (g) imply that $F_2 < F_1^{-1}(\alpha_*)$ for all $\alpha \in [\alpha_*, 1]$. Then, based on the strong global attractiveness of F_1 and F_2, it is easy to show that if $(\alpha_{t_0}, \sigma_{t_0}^2) \in \mathcal{R}_3$ then $(\alpha_t, \sigma_t^2) \in \mathcal{R}_3$ for all $t \geq t_0$. We have proved in Lemma 23-(d) that F_2 is a decreasing function of α on $[0, \hat{\alpha}]$ and increasing on $[\hat{\alpha}, 1]$, where $\alpha_\star < \hat{\alpha} < 1$. Then, the maximum of F_2 on $[\alpha_\star, 1]$ can only happen at either $\alpha_\star$ or 1. We assume that the latter case happens; it will be clear that our proof for the former case is a special case of the proof for the latter one. See the right panel of Fig. 15.

As discussed above, we assume that $F_2(1; \delta, \sigma_w^2) > F_2(\alpha_\star; \delta, \sigma_w^2)$. Hence, by Lemma 23-(d), there exists a unique number $\alpha_\diamond \in (\alpha_\star, 1)$ such that $F_2(\alpha_\diamond; \delta, \sigma_w^2) = F_2(\alpha_\star; \delta, \sigma_w^2)$. See the plot in the right panel of

Fig. 15. We further divide $\mathcal{R}_3$ into four regions:

$$
\begin{aligned}
\mathcal{R}_{3a} &\triangleq \left\{(\alpha,\sigma^2)|\alpha_* \le \alpha \le \alpha_\diamond,\ F_1^{-1}(\alpha_\diamond) < \sigma^2 \le F_1^{-1}(\alpha_*)\right\},\\
\mathcal{R}_{3b} &\triangleq \left\{(\alpha,\sigma^2)|\alpha_* \le \alpha \le 1,\ 0 \le \sigma^2 < F_1^{-1}(\alpha_\diamond)\right\},\\
\mathcal{R}_{3c} &\triangleq \left\{(\alpha,\sigma^2)|\alpha_\diamond < \alpha \le 1,\ F_1^{-1}(\alpha_\diamond) \le \sigma^2 < F_1^{-1}(\alpha_*)\right\}.
\end{aligned}
$$

Based on the strong global attractiveness of F_1 and F_2 (and similar to the proof of part (i) of this lemma), we can show the following:

- if $(\alpha_{t_0},\sigma_{t_0}^2) \in \mathcal{R}_{3a}$, then $(\alpha_{t_0+1},\sigma_{t_0+1}^2)$ can only be in $\mathcal{R}_{3a}$;
- if $(\alpha_{t_0},\sigma_{t_0}^2) \in \mathcal{R}_{3b}$, then $(\alpha_{t_0+1},\sigma_{t_0+1}^2)$ can be in $\mathcal{R}_{3a}$, $\mathcal{R}_{3b}$ or $\mathcal{R}_{3c}$;
- if $(\alpha_{t_0},\sigma_{t_0}^2) \in \mathcal{R}_{3c}$, then $(\alpha_{t_0+1},\sigma_{t_0+1}^2)$ can be in $\mathcal{R}_{3c}$ or $\mathcal{R}_{3a}$.

Putting things together, and similar to the treatment of $\mathcal{R}_0$, it can be shown that there exists a finite T_3 such that $(\alpha_t,\sigma_t^2) \in \mathcal{R}_{3a}$ for all $t \ge t_0 + T_3$.

It only remains to prove that if $(\alpha_{t'},\sigma_{t'}^2) \in \mathcal{R}_{3a}$ at a certain $t' \ge 0$, then $\{(\alpha_t,\sigma_t^2)\}_{t\ge t'}$ converges to $(\alpha_\star,\sigma_\star^2)$. To this end, define

$$
\begin{aligned}
B_1^{\text{low}}(\alpha,\sigma^2) &\triangleq \min\left\{\alpha_\star,\alpha,F_1(\sigma^2)\right\},\\
B_1^{\text{up}}(\alpha,\sigma^2) &\triangleq \max\left\{\alpha_\star,\alpha,F_1(\sigma^2)\right\},\\
B_2^{\text{low}}(\alpha,\sigma^2) &\triangleq \min\left\{\sigma_\star^2,\sigma^2,F_1^{-1}(\alpha)\right\} = F_1^{-1}\left(B_1^{\text{up}}(\alpha,\sigma^2)\right),\\
B_2^{\text{up}}(\alpha,\sigma^2) &\triangleq \max\left\{\sigma_\star^2,\sigma^2,F_1^{-1}(\alpha)\right\} = F_1^{-1}\left(B_1^{\text{low}}(\alpha,\sigma^2)\right).
\end{aligned}
$$

See examples depicted in Fig. 15. Using the strong global attractiveness of F_1 and F_2 and noting that $F_1^{-1}(\alpha) > F_2(\alpha) > \sigma_\star^2$ for $\alpha \in [\alpha_*,\alpha_\star)$ and $F_1^{-1}(\alpha) < F_2(\alpha) < \sigma_\star^2$ for $\alpha \in (\alpha_\star,\alpha_\diamond)$, it can be proved that

$$
\begin{aligned}
B_1^{\text{low}}(\alpha_t,\sigma_t^2) &\le \alpha_{t+1} \le B_1^{\text{up}}(\alpha_t,\sigma_t^2),\\
B_2^{\text{low}}(\alpha_t,\sigma_t^2) &\le \sigma_{t+1}^2 \le B_2^{\text{up}}(\alpha_t,\sigma_t^2).
\end{aligned}
$$

Further, the sequences $\{B_1^{\text{low}}(\alpha_t,\sigma_t^2)\}_{t\ge t'}$ and $\{B_2^{\text{low}}(\alpha_t,\sigma_t^2)\}_{t\ge t'}$ are monotonically non-decreasing and $\{B_1^{\text{up}}(\alpha_t,\sigma_t^2)\}_{t\ge t'}$ and $\{B_2^{\text{up}}(\alpha_t,\sigma_t^2)\}_{t\ge t'}$ are monotonically non-increasing. Also, B_1^{low} and B_2^{low} are upper bounded by $\alpha_\star$ and $\sigma_\star^2$, and B_1^{up} and B_2^{up} are lowered bounded by $\alpha_\star$ and $\sigma_\star^2$. Together with some arguments about the strict monotonicity of $\{B_1^{\text{low}}(\alpha_t,\sigma_t^2)\}_{t\ge t'}$ and $\{B_2^{\text{low}}(\alpha_t,\sigma_t^2)\}_{t\ge t'}$ (see discussions below (A.108)), we have

$$
\begin{aligned}
\lim_{t\to\infty} B_1^{\text{low}}(\alpha_t,\sigma_t^2) &= \lim_{t\to\infty} B_1^{\text{up}}(\alpha_t,\sigma_t^2) = \alpha_\star,\\
\lim_{t\to\infty} B_2^{\text{low}}(\alpha_t,\sigma_t^2) &= \lim_{t\to\infty} B_2^{\text{up}}(\alpha_t,\sigma_t^2) = \sigma_\star^2,
\end{aligned}
$$

which implies that $\lim_{t\to\infty}\alpha_{t+1} = \alpha_\star$ and $\lim_{t\to\infty}\sigma_{t+1}^2 = \sigma_\star^2$. We skip the proofs for the above statements since similar arguments have been repeatedly used in this paper. $\square$

E.4 Proof of Theorem 4

According to Lemma 24, we know that (α_t,σ_t^2) converges to the unique fixed point of the state evolution equation. We now analyze the location of this fixed point and further derive the noise sensitivity. Applying a variable change $s \triangleq \sigma/\alpha$, we obtain the following equations for this unique fixed point:

$$
\alpha = \phi_1(s), \tag{E.18a}
$$

$$
\sigma^2 = \frac{4}{\delta}\left\{\alpha^2 + \sigma^2 + 1 - \alpha\left[\phi_1(s) + \phi_3(s)\right]\right\} + 4\sigma_w^2, \tag{E.18b}
$$

where ϕ_1 and ϕ_3 are defined in (E.2). Using (E.18a) and $\sigma^2 = \alpha^2 s^2 = \phi_1^2(s) s^2$, and after some algebra, we can write (E.18b) as

$$T(s^2, \sigma_w^2) \triangleq \left(1 - \frac{4}{\delta}\right) \phi_1^2(s) s^2 + \frac{4}{\delta} \phi_1(s)\phi_3(s) - \left(\frac{4}{\delta} + 4\sigma_w^2\right) = 0. \tag{E.19}$$

Differentiating with respect to s^2 yields

$$\frac{\partial T(s^2, \sigma_w^2)}{\partial s^2} = \left(1 - \frac{4}{\delta}\right)\left(\phi_1^2(s) + 2\phi_1(s)\frac{\mathrm{d}\phi_1(s)}{\mathrm{d}s^2} s^2\right) + \frac{4}{\delta}\frac{\mathrm{d}\phi_1(s)\phi_3(s)}{\mathrm{d}s^2}. \tag{E.20}$$

Using the identities listed in (E.3), we have

$$\left.\frac{\partial T(s^2, \sigma_w^2)}{\partial s^2}\right|_{s=0} = 1 - \frac{2}{\delta}.$$

Also, it is straightforward to see that $\frac{\partial T(s^2, \sigma_w^2)}{\partial \sigma_w^2} = -4$. Note that we have an implicit relation between s^2 and σ_w^2, and by the implicit function theorem we have

$$\lim_{\sigma_w^2 \to 0} \frac{\mathrm{d}s^2}{\mathrm{d}\sigma_w^2} = -\lim_{s^2 \to 0}\left(\frac{\partial T(s^2, \sigma_w^2)}{\partial s^2}\right)^{-1} \frac{\partial T(s^2, \sigma_w^2)}{\partial \sigma_w^2} = \frac{4}{1 - \frac{2}{\delta}}.$$

Further, s is a continuously differentiable function of σ_w^2. Hence, by the mean value theorem we know that

$$\frac{s^2}{\sigma_w^2} = \left.\frac{\mathrm{d}s^2}{\mathrm{d}\sigma_w^2}\right|_{\tilde{\sigma}_w^2},$$

where $0 \leq \tilde{\sigma}_w \leq \sigma_w$. By taking $\lim_{\sigma_w^2 \to 0}$ from both sides of the above equality we have

$$\lim_{\sigma_w^2 \to 0} \frac{s^2}{\sigma_w^2} = \lim_{\tilde{\sigma}_w \to 0} \left.\frac{\mathrm{d}s^2}{\mathrm{d}\sigma_w^2}\right|_{\tilde{\sigma}_w^2} = -\lim_{s^2 \to 0}\left(\frac{\partial T(s^2, \sigma_w^2)}{\partial s^2}\right)^{-1} \frac{\partial T(s^2, \sigma_w^2)}{\partial \sigma_w^2} = \frac{4}{1 - \frac{2}{\delta}}.$$

To derive the noise sensitivity, we notice that

$$\begin{aligned} \mathrm{AMSE}(\sigma_w^2, \delta) &= (\alpha - 1)^2 + \sigma^2 \\ &= [\phi_1(s) - 1]^2 + s^2 \phi_1^2(s). \end{aligned}$$

As shown in (E.3), $\phi_1(s)$ can be expressed using elliptic integrals as:

$$\phi_1(s) = \sqrt{1 + s^2} E\left(\frac{1}{1 + s^2}\right) - \frac{s^2}{\sqrt{1 + s^2}} K\left(\frac{1}{1 + s^2}\right).$$

From Lemma 1-(i), $E(1 - \epsilon) = 1 + O(\epsilon \log \epsilon^{-1})$, hence $\sqrt{1 + s^2} E\left(\frac{1}{1+s^2}\right) = 1 + O(s^2 \log s^{-1})$. Further, since $K(1 - \epsilon) = O(\log \epsilon^{-1})$, we have $\frac{s^2}{\sqrt{1+s^2}} K\left(\frac{1}{1+s^2}\right) = O(s^2 \log s^{-1})$. Therefore, $\phi_1(s) - 1 = O(s^2 \log s^{-1})$. Hence, $\lim_{s^2 \to 0} \frac{[\phi_1(s) - 1]^2}{s^2} = 0$ and so

$$\lim_{s^2 \to 0} \frac{\mathrm{AMSE}(\sigma_w^2, \delta)}{s^2} = \lim_{s^2 \to 0} \frac{[\phi_1(s) - 1]^2}{s^2} + \phi_1^2(s) = 1.$$

Finally,

$$\begin{aligned} \lim_{\sigma_w^2 \to 0} \frac{\mathrm{AMSE}(\sigma_w^2, \delta)}{\sigma_w^2} &= \lim_{s^2 \to 0} \frac{\mathrm{AMSE}(\sigma_w^2, \delta)}{s^2} \cdot \lim_{\sigma_w^2 \to 0} \frac{s^2}{\sigma_w^2} \\ &= \frac{4}{1 - \frac{2}{\delta}}. \end{aligned}$$