Approximate Message Passing with Spectral Initialization for Generalized Linear Models – Supplementary Material

A Proof of Proposition 3.1

From assumption **(B2)** on p. 4, we recall that $h_t(u;\, y) = h_t(u;\, q(g, v))$. We write $\partial_g h_t(u;\, q(g, v))$ for the partial derivative with respect to g. We will show that $\mu_{X,t+1}$ in (3.8) can be written as:

$$\mu_{X,t+1} = \sqrt{\delta}\mathbb{E}\{\partial_g h_t(U_t;\, q(G, V))\}, \tag{A.1}$$

$$= \sqrt{\delta}\mathbb{E}\left\{h_t(U_t; Y)\left(\frac{\mathbb{E}\{G|U_t, Y\} - \mathbb{E}\{G|U_t\}}{\mathsf{Var}\{G|U_t\}}\right)\right\}. \tag{A.2}$$

From (A.2), we have that

$$\frac{\mu_{X,t+1}}{\sigma_{X,t+1}} = \frac{\sqrt{\delta}}{\sqrt{\mathbb{E}\{h_t(U_t; Y)^2\}}}\mathbb{E}\left\{h_t(U_t; Y)\left(\frac{\mathbb{E}\{G|U_t, Y\} - \mathbb{E}\{G|U_t\}}{\mathsf{Var}\{G|U_t\}}\right)\right\}. \tag{A.3}$$

The absolute value of the RHS is maximized when $h_t = c\, h_t^*$, for $c \neq 0$ and h_t^* is given in (3.18). To obtain the alternative expression in (3.19) from (3.18) we recall that U_t is Gaussian with zero mean and variance $(\mu_{U,t}^2 + \sigma_{U,t}^2)$. Furthermore, the conditional distribution of G given $U_t = u$ is Gaussian with $\mathbb{E}\{G \mid U_t = u\} = \rho_t u$ and $\mathsf{Var}(G \mid U_t = u) = (1 - \rho_t\mu_{U,t})$. Therefore, with $W \sim \mathsf{N}(0, 1)$ we have

$$\begin{aligned}\mathbb{E}\{G \mid U_t = u,\, Y = y\} &= \frac{\mathbb{E}_W\{(\rho_t u + \sqrt{1-\rho_t\,\mu_{U,t}}\, W)\, p_{Y|G}(y \mid \rho_t u + \sqrt{1-\rho_t\,\mu_{U,t}}\, W)\}}{\mathbb{E}_W\{p_{Y|G}(y \mid \rho_t u + \sqrt{1-\rho_t\,\mu_{U,t}}\, W)\}} \\ &= \rho_t u \;+\; \sqrt{1-\rho_t\,\mu_{U,t}}\,\frac{\mathbb{E}\{W p_{Y|G}(y \mid \rho_t u + \sqrt{1-\rho_t\,\mu_{U,t}}\, W)\}}{\mathbb{E}_W\{p_{Y|G}(y \mid \rho_t u + \sqrt{1-\rho_t\,\mu_{U,t}}\, W)\}}.\end{aligned} \tag{A.4}$$

Substituting (A.4) in (3.18) yields (3.19).

It remains to show (A.2), which we do by first showing (A.1). Define $e_t : \mathbb{R}^3 \to \mathbb{R}$ by

$$e_t(g, w, v) = h_t(\mu_{U,t} g + \sigma_{U,t} w;\, q(g, v)). \tag{A.5}$$

Then, using the chain rule, the partial derivative of $e_t(g, w, v)$ with respect to g is

$$\partial_g e_t(g, w, v) = \mu_{U,t} h_t'(\mu_{U,t} g + \sigma_{U,t} w;\, q(g, v)) \;+\; \partial_g h_t(u;\, q(g, v)). \tag{A.6}$$

The parameter $\mu_{X,t+1}$ in (3.8) can be written as

$$\begin{aligned}\mu_{X,t+1} &= \sqrt{\delta}\,[\mathbb{E}\{G\, e_t(G, W_{U,t}, V)\} - \mu_{U,t}\mathbb{E}\{h_t'(\mu_{U,t} G + \sigma_{U,t} W_{U,t};\, Y)\}] \\ &\overset{\text{(i)}}{=} \sqrt{\delta}\,\left[\mathbb{E}\{\partial_g e_t(G, W_{U,t}, V)\} - \mu_{U,t}\mathbb{E}\{h_t'(\mu_{U,t} G + \sigma_{U,t} W_{U,t};\, Y)\}\right] \\ &= \sqrt{\delta}\,\mathbb{E}\{\partial_g h_t(U_t;\, q(G, V))\},\end{aligned} \tag{A.7}$$

where the last equality is due to (A.6), and (i) holds due to Stein's lemma. Finally, we obtain (A.2) from (A.1) as follows:

$$\begin{aligned}\mathbb{E}\{\partial_g h_t(U_t;\, q(G, V))\} &= \mathbb{E}\left\{\mathbb{E}_{G|U_t}\left[\partial_g h_t(U_t;\, q(G, V)) \mid U_t\right]\right\} \\ &\overset{\text{(ii)}}{=} \mathbb{E}\left\{\mathbb{E}_{G|U_t}\left[h_t(U_t;\, q(G, V)) \cdot (G - \mathbb{E}\{G|U_t\})/\mathsf{Var}\{G|U_t\} \mid U_t\right]\right\} \\ &= \mathbb{E}\left\{\mathbb{E}_{G|U_t, Y}\left[h_t(U_t;\, Y) \cdot (G - \mathbb{E}\{G|U_t\})/\mathsf{Var}\{G|U_t\} \mid U_t, Y\right]\right\} \\ &= \mathbb{E}\left\{h_t(U_t; Y) \cdot ((\mathbb{E}\{G|U_t, Y\} - \mathbb{E}\{G|U_t\})/\mathsf{Var}\{G|U_t\})\right\}.\end{aligned} \tag{A.8}$$

Here step (ii) holds due to Stein's lemma. This completes the proof of the proposition. □

B Proof of the Main Result

B.1 The Artificial GAMP Algorithm

The state evolution parameters for the artificial GAMP are recursively defined as follows. Recall from (5.8) that $\tilde{X}_t = \mu_{\tilde{X},t} X + \sigma_{\tilde{X},t} W_{\tilde{X},t}$ and $\tilde{U}_t \equiv \mu_{\tilde{U},t} G + \sigma_{\tilde{U},t} W_{\tilde{U},t}$. Using (5.4), the state evolution initialization is

$$\mu_{\tilde{X},0} = \alpha, \qquad \sigma^2_{\tilde{X},0} = 1 - \alpha^2, \qquad \beta_0 = \sqrt{\mu^2_{\tilde{X},0} + \sigma^2_{\tilde{X},0}} = 1. \tag{B.1}$$

For $0 \le t \le (T-1)$, the state evolution parameters are iteratively computed by using the functions defined in (5.5) in (3.8):

$$\begin{aligned} &\mu_{\tilde{U},t} = \frac{\mu_{\tilde{X},t}}{\sqrt{\delta}\beta_t}, \qquad \sigma^2_{\tilde{U},t} = \frac{\sigma^2_{\tilde{X},t}}{\delta\, \beta_t^2}, \\ &\mu_{\tilde{X},t+1} = \frac{\mu_{\tilde{X},t}}{\sqrt{\delta}\beta_t}, \qquad \sigma^2_{\tilde{X},t+1} = \frac{1}{\beta_t^2}\, \mathbb{E}\left\{ \frac{Z_s^2 (G^2 \mu^2_{\tilde{X},t} + \sigma^2_{\tilde{X},t})}{(\lambda^*_\delta - Z_s)^2} \right\}, \\ &\beta_{t+1} = \sqrt{\mu^2_{\tilde{X},t+1} + \sigma^2_{\tilde{X},t+1}}. \end{aligned} \tag{B.2}$$

Here we recall that $G \sim \mathsf{N}(0,1)$, $Y \sim p_{Y|G}(\cdot \mid G)$, $Z_s = \mathcal{T}_s(Y)$, and the equality in (2.6) which is used to obtain the expression for $\mu_{\tilde{X},t+1}$. For $t \ge T$, the state evolution parameters are:

$$\begin{aligned} \mu_{\tilde{U},t} &= \frac{1}{\sqrt{\delta}} \mathbb{E}\{X f_{t-T}(\tilde{X}_t)\}, \\ \sigma^2_{\tilde{U},t} &= \frac{1}{\delta} \mathbb{E}\{f_{t-T}(\tilde{X}_t)^2\} - \mu^2_{\tilde{U},t}, \\ \mu_{\tilde{X},t+1} &= \sqrt{\delta}\mathbb{E}\{G h_{t-T}(\tilde{U}_t; Y)\} \,-\, \mathbb{E}\{h'_{t-T}(\tilde{U}_t; Y)\}\mathbb{E}\{X f_{t-T}(\tilde{X}_t)\}, \\ \sigma^2_{\tilde{X},t+1} &= \mathbb{E}\{h_{t-T}(\tilde{U}_t; Y)^2\}. \end{aligned} \tag{B.3}$$

Proposition B.1 (State evolution for artificial GAMP)**.** *Consider the setting of Theorem 1, the artificial GAMP iteration described in* (5.1)-(5.7)*, and the corresponding state evolution parameters defined in* (B.1)-(B.3).

For any PL*(2) function* $\psi : \mathbb{R}^2 \to \mathbb{R}$*, the following holds almost surely for* $t \ge 1$*:*

$$\lim_{d\to\infty} \frac{1}{d} \sum_{i=1}^{d} \psi(x_i, \tilde{x}_i^t) = \mathbb{E}\{\psi(X, \tilde{X}_t)\}, \tag{B.4}$$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \psi(y_i, \tilde{u}_i^t) = \mathbb{E}\{\psi(Y, \tilde{U}_t)\}. \tag{B.5}$$

Here $X \sim P_X$ *and* $Y \sim P_{Y|G}$*, with* $G \sim \mathsf{N}(0,1)$*. The random variables* $\tilde{X}_t, \tilde{U}_t$ *are defined in* (5.8).

The proposition follows directly from the state evolution result of (Javanmard and Montanari, 2013) since the initialization $\tilde{\boldsymbol{x}}^0$ of the artificial GAMP is independent of $\boldsymbol{A}$.

B.2 Analysis of the First Phase

Lemma B.2 (Fixed point of state evolution for first phase)**.** *Consider the setting of Theorem 1. Then, the state evolution recursion for the first phase, given by* (B.1)-(B.2)*, converges as* $T \to \infty$ *to the following fixed point:*

$$\mu_{\tilde{X}} \triangleq \lim_{T\to\infty} \mu_{\tilde{X},T} = \frac{a}{\sqrt{\delta}}, \qquad \sigma^2_{\tilde{X}} \triangleq \lim_{T\to\infty} \sigma^2_{\tilde{X},T} = \frac{1-a^2}{\delta}, \tag{B.6}$$

where a *is defined in* (2.7).

Proof. Recall that λ_δ^* denotes the unique solution of $\zeta_\delta(\lambda) = \phi(\lambda)$ for $\lambda > \tau$ (also given by (2.6)), and define $Z = Z_s/(\lambda_\delta^* - Z_s)$, where $Z_s = \mathcal{T}_s(Y)$. Note that

$$\mathbb{E}\{Z(G^2-1)\} = \mathbb{E}\left\{\frac{Z_s(G^2-1)}{\lambda_\delta^* - Z_s}\right\} = \frac{1}{\delta}, \tag{B.7}$$

where the second equality follows from the equality in (2.6). Moreover, the inequality in (2.6) implies that

$$\frac{\mathbb{E}\{Z^2\}}{(\mathbb{E}\{Z(G^2-1)\})^2} = \delta^2 \mathbb{E}\left\{\frac{Z_s^2}{(\lambda_\delta^* - Z_s)^2}\right\} < \delta. \tag{B.8}$$

Thus, by recalling that the state evolution initialization $\mu_{\tilde{X},0} = \alpha$ is strictly positive, the result follows from Lemma 5.2 in Mondelli et al. (2020). □

Lemma B.3 (Convergence to spectral estimator)**.** *Consider the setting of Theorem 1, and consider the first phase of the artificial GAMP iteration, given by* (5.1)-(5.2) *with* $\tilde{f}_t$ *and* $\tilde{h}_t$ *defined in* (5.5)*. Then,*

$$\lim_{T\to\infty} \lim_{d\to\infty} \frac{\|\sqrt{d}\, \hat{\boldsymbol{x}}^{\rm s} - \sqrt{\delta}\, \tilde{\boldsymbol{x}}^T\|^2}{d} = 0 \quad a.s. \tag{B.9}$$

Furthermore, for any PL*(2) function* $\psi : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$*, almost surely we have:*

$$\lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d} \psi(x_i, \sqrt{d}\, \hat{x}_i^{\rm s}) = \lim_{T\to\infty}\lim_{d\to\infty} \frac{1}{d}\sum_{i=1}^{d} \psi(x_i, \sqrt{\delta}\, \tilde{x}_i^T) = \mathbb{E}\{\psi(X, \sqrt{\delta}\,(\mu_{\tilde{X}} X + \sigma_{\tilde{X}} W))\}. \tag{B.10}$$

Here $X \sim P_X$ *and* $W \sim \mathsf{N}(0,1)$ *are independent.*

Proof. As in the proof of the previous result, let $Z = Z_s/(\lambda_\delta^* - Z_s)$ and note that (B.7)-(B.8) hold. Also define

$$Z' \triangleq \frac{Z}{Z + \delta\mathbb{E}\{Z(G^2-1)\}} = \frac{Z}{Z+1} = \frac{Z}{\lambda_\delta^*}. \tag{B.11}$$

Then, the assumptions of Lemma 5.4 in (Mondelli et al., 2020) are satisfied, with the only difference of the initialization of the GAMP iteration (cf. (5.4) in this paper and (5.4) in (Mondelli et al., 2020)). However, it is straightforward to verify that the difference in the initialization does not affect the proof of Lemma 5.4 in (Mondelli et al., 2020). Thus, (B.9) follows from (5.87) of (Mondelli et al., 2020), and (B.10) follows by taking $k = 2$ in (5.31) of (Mondelli et al., 2020). □

We will also need the following result on the convergence of the GAMP iterates.

Lemma B.4 (Convergence of GAMP iterates)**.** *Consider the first phase of the artificial GAMP iteration, given by* (5.1)-(5.2) *with* $\tilde{f}_t$ *and* $\tilde{h}_t$ *defined in* (5.5)*. Then, the following limits hold almost surely:*

$$\lim_{T\to\infty}\lim_{n\to\infty} \frac{1}{n}\|\tilde{\boldsymbol{u}}^{T-1} - \tilde{\boldsymbol{u}}^{T-2}\|_2^2 = 0, \qquad \lim_{T\to\infty}\lim_{d\to\infty} \frac{1}{d}\|\tilde{\boldsymbol{x}}^{T} - \tilde{\boldsymbol{x}}^{T-1}\|_2^2 = 0. \tag{B.12}$$

Though the initialization of the GAMP in (Mondelli et al., 2020) is different from (5.4), the proof of Lemma B.4 is the same as that of Lemma 5.3 in (Mondelli et al., 2020) since it only relies on $\mu_{\tilde{X},0} = \alpha$ being strictly non-zero.

B.3 Analysis of the Second Phase

Lemma B.5. *Assume the setting of Theorem 1. Consider the artificial GAMP algorithm* (5.1)-(5.2) *with the related state evolution recursion* (B.2)-(B.3)*, and the modified version of the true GAMP algorithm* (5.13)-(5.14)*. Fix any* $\varepsilon > 0$*. Then, for* $t \geq 0$ *such that* $\sigma_{X,k}^2 > 0$ *for* $0 \leq k \leq t$*, the following statements hold:*

1.

$$\lim_{T\to\infty}\left|\mu_{\tilde{U},t+T}-\mu_{U,t}\right|=0, \qquad \lim_{T\to\infty}\left|\sigma^2_{\tilde{U},t+T}-\sigma^2_{U,t}\right|=0, \tag{B.13}$$

$$\lim_{T\to\infty}\left|\mu_{\tilde{X},T+t+1}-\mu_{X,t+1}\right|=0, \qquad \lim_{T\to\infty}\left|\sigma^2_{\tilde{X},T+t+1}-\sigma^2_{X,t+1}\right|=0. \tag{B.14}$$

2. Let $\psi:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ *be a* PL*(2) function. Then, almost surely,*

$$\lim_{T\to\infty}\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_i,\tilde{u}_i^{T+t})-\frac{1}{n}\sum_{i=1}^{n}\psi(y_i,\hat{u}_i^{t})\right|=0, \tag{B.15}$$

$$\lim_{T\to\infty}\lim_{d\to\infty}\left|\frac{1}{d}\sum_{i=1}^{d}\psi(x_i,\tilde{x}_i^{T+t+1})-\frac{1}{d}\sum_{i=1}^{d}\psi(x_i,\hat{x}_i^{t+1})\right|=0. \tag{B.16}$$

The limits in (B.14) *and* (B.16) *also hold for* $t+1=0$*.*

Proof. We will use $\kappa_t,\kappa_t',c_t,\gamma_t$ to denote generic positive constants which depend on t, but not on n,d, or ε. The values of these constants may change throughout the proof.

Proof of (B.13) **and** (B.14)**.** We prove the result by induction, starting from the base case $\left|\mu_{\tilde{X},T}-\mu_{X,0}\right|$, $|\sigma^2_{\tilde{X},T}-\sigma^2_{X,0}|$. From Lemma B.2, we have

$$\lim_{T\to\infty}\mu_{\tilde{X},T}=\mu_{\tilde{X}}=\frac{a}{\sqrt{\delta}}, \qquad \lim_{T\to\infty}\sigma^2_{\tilde{X},T}=\sigma^2_{\tilde{X}}=\frac{1-a^2}{\delta}. \tag{B.17}$$

Recalling from (3.9) that $\mu_{X,0}=\frac{a}{\sqrt{\delta}}$, $\sigma^2_{X,0}=\frac{1-a^2}{\delta}$, (B.17) implies that

$$\lim_{T\to\infty}\left|\mu_{\tilde{X},T}-\mu_{X,0}\right|=0, \qquad \lim_{T\to\infty}\left|\sigma^2_{\tilde{X},T}-\sigma^2_{X,0}\right|=0. \tag{B.18}$$

Assume towards induction that (B.14) holds with $(t+1)$ replaced by t, and that $\sigma^2_{X,k}>0$ for $0\le k\le t$. We will show that (B.13) holds, and then that (B.14) holds.

For brevity, we write $\Delta_{\mu,t},\Delta_{\sigma,t}$ for $(\mu_{X,t}-\mu_{\tilde{X},t+T})$ and $(\sigma_{X,t}-\sigma_{\tilde{X},t+T})$, respectively. By the induction hypothesis, given any $\varepsilon>0$, for T sufficiently large we have

$$|\Delta_{\mu,t}|<\kappa_t\varepsilon, \qquad |\Delta_{\sigma,t}|<\frac{\kappa_t}{\sigma_{X,t}+\sigma_{\tilde{X},t+T}}\varepsilon=\kappa_t'\varepsilon. \tag{B.19}$$

Since $\sigma_{X,t}$ is strictly positive, κ_t' is finite and bounded above.

From (3.8) we have

$$\mu_{U,t}=\frac{1}{\sqrt{\delta}}\mathbb{E}\{Xf_t(\mu_{X,t}X+\sigma_{X,t}W_{X,t})\}=\frac{1}{\sqrt{\delta}}\mathbb{E}\{Xf_t(\mu_{\tilde{X},T+t}X+\sigma_{\tilde{X},T+t}W_{X,t}\;+\;\Delta_{\mu,t}X+\Delta_{\sigma,t}W_{X,t}). \tag{B.20}$$

Recalling that f_t is Lipschitz and letting L_t denote its Lipschitz constant, we have

$$\left|f_t(\mu_{\tilde{X},T+t}X+\sigma_{\tilde{X},T+t}W_{X,t}\;+\;\Delta_{\mu,t}X+\Delta_{\sigma,t}W_{X,t})\;-\;f_t(\mu_{\tilde{X},T+t}X+\sigma_{\tilde{X},T+t}W_{X,t})\right|\le L_t\left|\Delta_{\mu,t}X+\Delta_{\sigma,t}W_{X,t}\right|. \tag{B.21}$$

Using (B.21) in (B.20), we obtain

$$\begin{aligned}\sqrt{\delta}\mu_{U,t}&\ge\mathbb{E}\{Xf_t(\mu_{\tilde{X},T+t}X+\sigma_{\tilde{X},T+t}W_{X,t})\}\;-\;L_t\mathbb{E}\{|X|\,|\Delta_{\mu,t}X+\Delta_{\sigma,t}W_{X,t}|\},\\ \sqrt{\delta}\mu_{U,t}&\le\mathbb{E}\{Xf_t(\mu_{\tilde{X},T+t}X+\sigma_{\tilde{X},T+t}W_{X,t})\}\;+\;L_t\mathbb{E}\{|X|\,|\Delta_{\mu,t}X+\Delta_{\sigma,t}W_{X,t}|\}.\end{aligned} \tag{B.22}$$

Since $W_{X,t} \stackrel{\text{d}}{=} W_{\tilde{X},t+T}$ and independent of X, we have that $\mathbb{E}\{X f_t(\mu_{\tilde{X},T+t} X + \sigma_{\tilde{X},T+t} W_{X,t})\} = \sqrt{\delta}\mu_{\tilde{U},t+T}$. Therefore, (B.22) implies

$$\sqrt{\delta}\left|\mu_{U,t} - \mu_{\tilde{U},t+T}\right| \le L_t(\Delta_{\mu,t} + \Delta_{\sigma,t}\mathbb{E}\{|W_{X,t}|\}), \tag{B.23}$$

where we have used $\mathbb{E}\{|X|^2\} < \sqrt{\mathbb{E}\{X^2\}} = 1$. Noting that $\mathbb{E}\{|W_{X,t}|\} = \sqrt{2/\pi}$, from (B.19) it follows that for sufficiently large T:

$$\left|\mu_{U,t} - \mu_{\tilde{U},t+T}\right| \le \frac{L_t}{\sqrt{\delta}}(\kappa_t + \kappa_t'\sqrt{2/\pi})\,\varepsilon < \gamma_t\,\varepsilon. \tag{B.24}$$

Next consider $\sigma^2_{U,t}$. From (3.8), we have

$$\sigma^2_{U,t} = \frac{1}{\delta}\mathbb{E}\{f_t(\mu_{X,t}X + \sigma_{X,t}W_{X,t})^2\} - \mu^2_{U,t}. \tag{B.25}$$

Furthermore, as $W_{X,t} \stackrel{\text{d}}{=} W_{\tilde{X},t+T}$ and independent of X, we also have that

$$\sigma^2_{\tilde{U},t+T} = \frac{1}{\delta}\mathbb{E}\{f_t(\mu_{\tilde{X},t+T}X + \sigma_{\tilde{X},t+T}W_{X,t})^2\} - \mu^2_{\tilde{U},t+T}. \tag{B.26}$$

Using the reverse triangle inequality, we have

$$\begin{aligned}
&\left|f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t} + \Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t})\right| \\
&\quad\ge |f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})| \\
&\qquad - \left|f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t} + \Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}) - f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})\right| \\
&\quad\ge |f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})| - L_t\,|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|,
\end{aligned} \tag{B.27}$$

where the last inequality follows from (B.21). Similarly,

$$\begin{aligned}
&\left|f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t} + \Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t})\right| \\
&\qquad\le |f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})| + L_t\,|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|.
\end{aligned} \tag{B.28}$$

Using (B.27), we obtain the bound

$$\begin{aligned}
\mathbb{E}\{f_t(\mu_{X,t}X + \sigma_{X,t}W_{X,t})^2\} &\ge \mathbb{E}\{f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})^2\} - L_t^2\mathbb{E}\{|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|^2\} \\
&\quad - 2L_t\sqrt{\mathbb{E}\{f_t(\mu_{X,t}X + \sigma_{X,t}W_{X,t})^2\}\cdot\mathbb{E}\{|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|^2\}}.
\end{aligned} \tag{B.29}$$

Similarly, using (B.28) we get

$$\begin{aligned}
\mathbb{E}\{f_t(\mu_{X,t}X + \sigma_{X,t}W_{X,t})^2\} &\le \mathbb{E}\{f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})^2\} + L_t^2\mathbb{E}\{|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|^2\} \\
&\quad + 2L_t\sqrt{\mathbb{E}\{f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})^2\}\cdot\mathbb{E}\{|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|^2\}}.
\end{aligned} \tag{B.30}$$

Furthermore,

$$\mathbb{E}\{|\Delta_{\mu,t}X + \Delta_{\sigma,t}W_{X,t}|^2\} \le 2\,|\Delta_{\mu,t}|^2\,\mathbb{E}\{X^2\} + 2\,|\Delta_{\sigma,t}|^2\,\mathbb{E}\{W_{X,t}\}^2 = 2(|\Delta_{\mu,t}|^2 + |\Delta_{\sigma,t}|^2).$$

From (3.8) and (B.3), we note that

$$\begin{aligned}
\mathbb{E}\{f_t(\mu_{X,t}X + \sigma_{X,t}W_{X,t})\}^2 &= \delta(\mu^2_{U,t} + \sigma^2_{U,t}), \\
\mathbb{E}\{f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{\tilde{X},T+t})\}^2 &= \delta(\mu^2_{\tilde{U},T+t} + \sigma^2_{\tilde{U},T+t}).
\end{aligned} \tag{B.31}$$

Therefore, Eqs. (B.29) and (B.30) imply that

$$\begin{aligned}
&\left|\mathbb{E}\{f_t(\mu_{X,t}X + \sigma_{X,t}W_{X,t})^2\} - \mathbb{E}\{f_t(\mu_{\tilde{X},T+t}X + \sigma_{\tilde{X},T+t}W_{X,t})^2\}\right| \\
&\le 2L_t^2(|\Delta_{\mu,t}|^2 + |\Delta_{\sigma,t}|^2) + 2L_t\sqrt{2\delta(\mu^2_{U,t} + \sigma^2_{U,t} + \mu^2_{\tilde{U},T+t} + \sigma^2_{\tilde{U},T+t})(|\Delta_{\mu,t}|^2 + |\Delta_{\sigma,t}|^2)}.
\end{aligned} \tag{B.32}$$

Using this in (B.25)-(B.26), we have

$$\begin{aligned}\left|\sigma^2_{U,t}-\sigma^2_{\tilde{U},t+T}\right| &\leq \left|\mu_{\tilde{U},T+t}-\mu_{U,t}\right|\cdot\left|\mu_{\tilde{U},T+t}+\mu_{U,t}\right| \\ &+\left(\frac{2}{\delta}L_t^2(|\Delta_{\mu,t}|^2+|\Delta_{\sigma,t}|^2)+\frac{2}{\sqrt{\delta}}L_t\sqrt{2(\mu^2_{U,t}+\sigma^2_{U,t}+\mu^2_{\tilde{U},T+t}+\sigma^2_{\tilde{U},T+t})(|\Delta_{\mu,t}|^2+|\Delta_{\sigma,t}|^2)}\right).\end{aligned} \tag{B.33}$$

From (B.19), we obtain

$$|\Delta_{\mu,t}|^2+|\Delta_{\sigma,t}|^2<\left(\kappa_t^2+(\kappa_t')^2\right)\varepsilon^2. \tag{B.34}$$

Furthermore, as f_t is Lipschitz, from (B.31) and the induction hypothesis we have

$$|\mu_{\tilde{U},T+t}|+|\mu_{U,t}|+\sigma_{U,t}+\sigma_{\tilde{U},T+t}\leq c_t, \tag{B.35}$$

for some constant c_t. Using (B.24), (B.34) and (B.35) in (B.33), we conclude that for sufficiently large T:

$$\left|\sigma^2_{U,t}-\sigma^2_{\tilde{U},T+t}\right|<\gamma_t\varepsilon. \tag{B.36}$$

Next, we show that if (B.13) holds for some $t\geq 0$ and $\sigma^2_{X,k}>0$ for $k\leq t$, then :

$$\lim_{T\to\infty}\left|\mu_{\tilde{X},T+t+1}-\mu_{X,t+1}\right|=0,\qquad \lim_{T\to\infty}\left|\sigma^2_{\tilde{X},T+t+1}-\sigma^2_{X,t+1}\right|=0. \tag{B.37}$$

We denote the Lipschitz constant of h_t by $\bar{L}_t$, and write $\bar{\Delta}_{\mu,t}$, $\bar{\Delta}_{\sigma,t}$ for $(\mu_{U,t}-\mu_{\tilde{U},t+T})$ and $(\sigma_{U,t}-\sigma_{\tilde{U},t+T})$, respectively. Using this notation, we have

$$\begin{aligned}&\left|h_t(\mu_{\tilde{U},T+t}G+\sigma_{\tilde{U},T+t}W_{U,t}+\bar{\Delta}_{\mu,t}G+\bar{\Delta}_{\sigma,t}W_{U,t};\,Y)-h_t(\mu_{\tilde{U},T+t}G+\sigma_{\tilde{U},T+t}W_{U,t};\,Y)\right| \\ &\leq \bar{L}_t\left|\bar{\Delta}_{\mu,t}G+\bar{\Delta}_{\sigma,t}W_{U,t}\right|.\end{aligned} \tag{B.38}$$

The induction hypothesis (B.13) implies that for sufficiently large T:

$$\left|\bar{\Delta}_{\mu,t}\right|<\gamma_t\varepsilon,\qquad \left|\bar{\Delta}_{\sigma,t}\right|<\frac{\gamma_t}{\sigma_{U,t}+\sigma_{\tilde{U},t+T}}\varepsilon=\gamma_t\varepsilon. \tag{B.39}$$

We note that $\sigma_{U,t}>0$ since $\sigma_{X,t}>0$. Indeed, from the discussion leading to (3.17), for a fixed $\mu_{X,t},\sigma_{X,t}$ the smallest possible ratio $\sigma^2_{U,t}/\mu^2_{U,t}$ is achieved by the Bayes-optimal choice $f_t=cf_t^*$, where $f_t^*(X_t)=E\{X|X_t\}$. Furthermore, from (3.17), in order for $\sigma_{U,t}=0$, we need $\mathbb{E}\{\mathbb{E}\{X|X_t\}^2\}=1$. From Jensen's inequality, we also have $\mathbb{E}\{\mathbb{E}\{X|X_t\}^2\}\leq\mathbb{E}\{\mathbb{E}\{X^2|X_t\}\}=1$. Therefore, $\mathbb{E}\{\mathbb{E}\{X|X_t\}^2\}=1$ only if X is a deterministic function of $X_t=\mu_{X,t}X+\sigma_{X,t}W$. But this is impossible when $\sigma_{X,t}>0$. Therefore $\sigma_{U,t}>0$, and γ_t in (B.39) is strictly positive.

From (B.38), we obtain

$$\begin{aligned}&\mathbb{E}\{Gh_t(\mu_{\tilde{U},T+t}G+\sigma_{\tilde{U},T+t}W_{U,t};\,Y)\}-\bar{L}_t\mathbb{E}\{\left|\bar{\Delta}_{\mu,t}\right|G^2+\left|\bar{\Delta}_{\sigma,t}\right|\cdot|G|\cdot|W_{U,t}|\} \\ &\quad\leq\mathbb{E}\{Gh_t(\mu_{U,t}G+\sigma_{U,t}W_{U,t};\,Y)\} \\ &\quad\leq\mathbb{E}\{Gh_t(\mu_{\tilde{U},T+t}G+\sigma_{\tilde{U},T+t}W_{U,t};\,Y)\}+\bar{L}_t\mathbb{E}\{\left|\bar{\Delta}_{\mu,t}\right|G^2+\left|\bar{\Delta}_{\sigma,t}\right|\cdot|G|\cdot|W_{U,t}|\}.\end{aligned} \tag{B.40}$$

Now, using (3.8) and (B.3), we have:

$$\begin{aligned}\frac{1}{\sqrt{\delta}}\left|\mu_{\tilde{X},T+t+1}-\mu_{X,t+1}\right|&=\Big|\mathbb{E}\{G(h_t(\tilde{U}_{T+t};\,Y)-h_t(U_t;\,Y))\} \\ &\qquad-\mu_{U,t}\big(\mathbb{E}\{h_t'(\tilde{U}_{T+t};Y)\}-\mathbb{E}\{h_t'(U_t;Y)\}\big)\;-\;\mathbb{E}\{h_t'(\tilde{U}_{T+t};Y)\}(\mu_{\tilde{U},T+t}-\mu_{U,t})\Big| \\ &\leq\bar{L}_t\big(\left|\bar{\Delta}_{\mu,t}\right|+\left|\bar{\Delta}_{\sigma,t}\right|(2/\pi)\big)\;+\;|\mu_{U,t}|\cdot|\mathbb{E}\{h_t'(\tilde{U}_{T+t};Y)\}-\mathbb{E}\{h_t'(U_t;Y)\}|\;+\;\bar{L}_t\left|\bar{\Delta}_{\mu,t}\right|.\end{aligned} \tag{B.41}$$

For the inequality above, we used (B.40) (noting that $\mathbb{E}\{|W_{U,t}|\}=\mathbb{E}\{|G|\}=\sqrt{2/\pi}$ and $\mathbb{E}\{G^2\}=1$), and the fact that $|h_t'|$ is bounded by $\bar{L}_t$, the Lipschitz constant of h_t. Now,

$$\left|\mathbb{E}\{h_t'(U_t;Y)\;-\mathbb{E}\{h_t'(\tilde{U}_{T+t};Y)\}\right|=\left|\mathbb{E}\{h_t'(\mu_{U,t}G+\,\sigma_{U,t}W_{U,t};Y)\}-\mathbb{E}\{h_t'(\mu_{\tilde{U},T+t}G+\,\sigma_{\tilde{U},T+t}W_{U,t}\,;Y)\}\right|. \tag{B.42}$$

By the induction hypothesis (B.13), we have

$$\lim_{T\to\infty} \mu_{\tilde{U},T+t} = \mu_{U,t}, \qquad \lim_{T\to\infty} \sigma_{\tilde{U},T+t} = \sigma_{U,t}. \tag{B.43}$$

Thus, as $T \to \infty$, the random variable $(\mu_{\tilde{U},T+t}G + \sigma_{\tilde{U},T+t}W_{U,t})$ converges in distribution to $\mu_{U,t}G + \sigma_{U,t}W_{U,t}$. Then, Lemma C.1 in Appendix C implies that

$$\lim_{T\to\infty} \left|\mathbb{E}\{h_t'(U_t;Y) \ - \mathbb{E}\{h_t'(\tilde{U}_{T+t};Y)\}\right| = 0. \tag{B.44}$$

Using (B.44), (B.39) and (B.35) in (B.41) proves that the first limit in (B.37) holds.

Finally, we prove the second limit in (B.37). From (3.8), (B.3) and arguments along the same lines as (B.29)-(B.32), we obtain the bound

$$\begin{aligned}\left|\sigma^2_{X,t+1} - \sigma^2_{\tilde{X},T+t+1}\right| &= \left|\mathbb{E}\{h_t(U_t;Y)^2\} \ - \ \mathbb{E}\{h_t(\tilde{U}_{t+T};Y)^2\}\right| \\ &\le 2\bar{L}_t^2(\left|\bar{\Delta}_{\mu,t}\right|^2 + \left|\bar{\Delta}_{\sigma,t}\right|^2) + 2\bar{L}_t\sqrt{(\sigma^2_{X,t+1} + \sigma^2_{\tilde{X},T+t+1})(\left|\bar{\Delta}_{\mu,t}\right|^2 + \left|\bar{\Delta}_{\sigma,t}\right|^2)}.\end{aligned} \tag{B.45}$$

Furthermore, as h_t is Lipschitz, the formulas for $\sigma^2_{X,t+1}$ and $\sigma_{\tilde{X},T+t+1}$ (in (3.8) and (B.3)) along with the induction hypothesis (B.43) imply that

$$\sigma^2_{X,t+1} + \sigma^2_{\tilde{X},T+t+1} \le c_t, \tag{B.46}$$

for some constant c_t. By using (B.46) and (B.39), we can upper bound the RHS of (B.45) with $\kappa_{t+1}\varepsilon$, for sufficiently large T. This completes the proof of the second limit in (B.37).

Proof of (B.15) **and** (B.16)**.**

Since $\psi \in \mathrm{PL}(2)$, for $i \in [d]$ we have

$$\left|\psi(x_i, \tilde{x}_i^{T+t+1}) - \psi(x_i, \hat{x}_i^{t+1})\right| \le C\left(1 + |x_i| + |\tilde{x}_i^{T+t+1}| + |\hat{x}_i^{t+1}|\right)\left|\tilde{x}_i^{T+t+1} - \hat{x}_i^{t+1}\right|, \tag{B.47}$$

for a universal constant $C > 0$. Therefore,

$$\begin{aligned}\left|\frac{1}{d}\sum_{i=1}^d \psi(x_i, \tilde{x}_i^{T+t+1}) - \frac{1}{d}\sum_{i=1}^d \psi(x_i, \hat{x}_i^{t+1})\right| &\le \frac{C}{d}\sum_{i=1}^d \left(1 + |x_i| + |\tilde{x}_i^{T+t+1}| + |\hat{x}_i^{t+1}|\right)\left|\tilde{x}_i^{T+t+1} - \hat{x}_i^{t+1}\right| \\ \le 4C\left[1 + \frac{1}{d}\sum_{i=1}^d \Big(|x_i|^2 + |\tilde{x}_i^{T+t+1}|^2 + |\hat{x}_i^{t+1}|^2\Big)\right]^{1/2} &\frac{\|\tilde{\boldsymbol{x}}^{T+t+1} - \hat{\boldsymbol{x}}^{t+1}\|_2}{\sqrt{d}},\end{aligned} \tag{B.48}$$

where the second inequality follows from Cauchy-Schwarz. By the same argument,

$$\left|\frac{1}{n}\sum_{i=1}^n \psi(y_i, \tilde{u}_i^{T+t}) - \frac{1}{n}\sum_{i=1}^n \psi(y_i, \hat{u}_i^{t})\right| \le 4C\left[1 + \frac{1}{n}\sum_{i=1}^n \Big(|y_i|^2 + |\tilde{u}_i^{T+t}|^2 + |\hat{u}_i^{t}|^2\Big)\right]^{\frac{1}{2}} \frac{\|\tilde{\boldsymbol{u}}^{T+t} - \hat{\boldsymbol{u}}^{t}\|_2}{\sqrt{n}}. \tag{B.49}$$

We will show via induction that as $d \to \infty$: *i)* the terms inside the square brackets in (B.48) and (B.49) converge almost surely to finite deterministic values, and *ii)* as $T \to \infty$ (with the limit in T taken after the limit in d), the terms $\frac{\|\tilde{\boldsymbol{x}}^{T+t} - \hat{\boldsymbol{x}}^{t}\|_2}{\sqrt{d}}$ and $\frac{\|\tilde{\boldsymbol{u}}^{T+t+1} - \hat{\boldsymbol{u}}^{t+1}\|_2}{\sqrt{d}}$ converge to 0 almost surely.

<u>Base case $t = 0$</u>: The result (B.16) for $t + 1 = 0$ directly follows from Lemma B.3. Next, using (B.49), the LHS of (B.15) for $t = 0$ can be bounded as

$$\left|\frac{1}{n}\sum_{i=1}^n \psi(y_i, \tilde{u}_i^{T}) - \frac{1}{n}\sum_{i=1}^n \psi(y_i, \hat{u}_i^{0})\right| \le 4C\left[1 + \frac{\|\boldsymbol{y}\|_2^2}{n} + \frac{\|\tilde{\boldsymbol{u}}^T\|_2^2}{n} + \frac{\|\hat{\boldsymbol{u}}^0\|_2^2}{n}\right]^{\frac{1}{2}} \frac{\|\tilde{\boldsymbol{u}}^{T} - \hat{\boldsymbol{u}}^{0}\|_2}{\sqrt{n}}. \tag{B.50}$$

From the definition of the artificial GAMP (5.1)-(5.6), we have

$$\tilde{\boldsymbol{u}}^T = \frac{1}{\sqrt{\delta}}\boldsymbol{A}f_0(\tilde{\boldsymbol{x}}^T) - \sqrt{\delta}\,\tilde{\mathsf{b}}_T\,\boldsymbol{Z}\tilde{\boldsymbol{u}}^{T-1}, \tag{B.51}$$

where we define

$$\boldsymbol{Z} = \boldsymbol{Z}_s(\lambda_\delta^* \boldsymbol{I} - \boldsymbol{Z}_s)^{-1}, \tag{B.52}$$

with $\boldsymbol{Z}_s = \text{diag}(\mathcal{T}_s(y_1), \dots, \mathcal{T}_s(y_n))$. Similarly, defining

$$\boldsymbol{e}_1 := \tilde{\boldsymbol{u}}^{T-1} - \tilde{\boldsymbol{u}}^{T-2}, \tag{B.53}$$

we obtain $\tilde{\boldsymbol{u}}^{T-1} = \frac{1}{\sqrt{\delta}\beta_{T-1}}\left[\boldsymbol{A}\tilde{\boldsymbol{x}}^{T-1} - \boldsymbol{Z}\tilde{\boldsymbol{u}}^{T-1} + \boldsymbol{Z}\boldsymbol{e}_1\right]$, or

$$\tilde{\boldsymbol{u}}^{T-1} = \frac{1}{\sqrt{\delta}\beta_{T-1}}\left(\boldsymbol{I} + \frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\left[\boldsymbol{A}\tilde{\boldsymbol{x}}^{T-1} + \boldsymbol{Z}\boldsymbol{e}_1\right]. \tag{B.54}$$

Substituting (B.54) in (B.51), we obtain

$$\tilde{\boldsymbol{u}}^{T} = \frac{1}{\sqrt{\delta}}\boldsymbol{A}f_0(\tilde{\boldsymbol{x}}^T) - \frac{\tilde{\mathsf{b}}_T}{\beta_{T-1}}\boldsymbol{Z}\left(\boldsymbol{I} + \frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\boldsymbol{A}\tilde{\boldsymbol{x}}^{T-1} - \frac{\tilde{\mathsf{b}}_T}{\beta_{T-1}}\boldsymbol{Z}^2\left(\boldsymbol{I} + \frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\boldsymbol{e}_1. \tag{B.55}$$

Using (B.55) and the expression for $\hat{\boldsymbol{u}}^0$ from (5.12), we have

$$\begin{aligned}\frac{1}{d}\|\tilde{\boldsymbol{u}}^T - \hat{\boldsymbol{u}}^0\|_2^2 &\le 3\,\frac{\|\boldsymbol{A}f_0(\tilde{\boldsymbol{x}}^T) - \boldsymbol{A}f_0(\hat{\boldsymbol{x}}^0)\|_2^2}{\delta\, d} + 3\left\|\frac{\tilde{\mathsf{b}}_T}{\beta_{T-1}}\boldsymbol{Z}^2\left(\boldsymbol{I} + \frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\frac{\boldsymbol{e}_1}{\sqrt{d}}\right\|_2^2 \\ &\quad + \frac{3}{d}\left\|\frac{\bar{\mathsf{b}}_0\sqrt{\delta}}{\lambda_\delta^*}\boldsymbol{Z}_s\boldsymbol{A}\hat{\boldsymbol{x}}^0 - \frac{\tilde{\mathsf{b}}_T}{\beta_{T-1}}\boldsymbol{Z}\left(\boldsymbol{I} + \frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\boldsymbol{A}\tilde{\boldsymbol{x}}^{T-1}\right\|_2^2 \\ &:= 3(S_1 + S_2 + S_3).\end{aligned} \tag{B.56}$$

We now bound each of the three terms. By Cauchy-Schwarz inequality,

$$S_1 \le \|\boldsymbol{A}\|_{\text{op}}^2\,\frac{\|f_0(\tilde{\boldsymbol{x}}^T) - f_0(\hat{\boldsymbol{x}}^0)\|_2^2}{\delta\, d} \le \|\boldsymbol{A}\|_{\text{op}}^2\,\frac{L_0^2}{\delta}\cdot\frac{\|\tilde{\boldsymbol{x}}^T - \hat{\boldsymbol{x}}^0\|_2^2}{d}, \tag{B.57}$$

where L_0 is the Lipschitz constant of f_0. Since the entries of $\boldsymbol{A}$ are i.i.d. $\mathsf{N}(0, 1/d)$, almost surely the operator norm of $\boldsymbol{A}$ is bounded by a universal constant for sufficiently large d (Anderson et al., 2009). From Lemma B.3 and the definition of $\hat{\boldsymbol{x}}^0$ in (5.11), we also have

$$\lim_{T\to\infty}\lim_{d\to\infty}\frac{\|\tilde{\boldsymbol{x}}^T - \hat{\boldsymbol{x}}^0\|_2^2}{d} = \frac{1}{\delta}\cdot\frac{\|\sqrt{\delta}\tilde{\boldsymbol{x}}^T - \sqrt{d}\hat{\boldsymbol{x}}^{\text{s}}\|_2^2}{d} = 0 \quad \text{a.s.} \tag{B.58}$$

Therefore,

$$\lim_{T\to\infty}\lim_{d\to\infty} S_1 = 0 \;\; \text{a.s.} \tag{B.59}$$

Next, recalling the definition of $\boldsymbol{e}_1$ from (B.53) we bound S_2 as follows:

$$S_2 \le \frac{\tilde{\mathsf{b}}_T^2}{\beta_{T-1}^2}\|\boldsymbol{Z}^2\big(\boldsymbol{I} + \boldsymbol{Z}/(\sqrt{\delta}\beta_{T-1})\big)^{-1}\|_{\text{op}}^2\cdot\frac{\|\tilde{\boldsymbol{u}}^{T-1} - \tilde{\boldsymbol{u}}^{T-2}\|_2^2}{d}. \tag{B.60}$$

From Lemma B.4, we know that $\lim_{T\to\infty}\lim_{d\to\infty}\frac{\|\tilde{\boldsymbol{u}}^{T-1} - \tilde{\boldsymbol{u}}^{T-2}\|_2^2}{d} = 0$ almost surely. We now show that the other terms on the RHS of (B.60) are bounded almost surely. Recall from (5.7) that $\tilde{\mathsf{b}}_T = \frac{1}{n}\sum_{i=1}^d f_0'(\tilde{x}_i^T)$. Proposition B.1 guarantees that the empirical distribution of $\tilde{\boldsymbol{x}}^t$ converges to the law of $\tilde{X}_t \equiv \mu_{\tilde{X},t}X + \sigma_{\tilde{X},t}W$. Since f_0 is Lipschitz, Lemma C.1 in Appendix C therefore implies that almost surely:

$$\lim_{d\to\infty}\tilde{\mathsf{b}}_T = \frac{1}{\delta}\mathbb{E}\{f_0'(\mu_{\tilde{X},T}X + \sigma_{\tilde{X},T}W)\}. \tag{B.61}$$

From Lemma B.2, we know that $\lim_{T\to\infty}\mu_{\tilde{X},T}=\frac{a}{\sqrt{\delta}}$ and $\lim_{T\to\infty}\sigma^2_{\tilde{X},T}=\frac{1-a^2}{\delta}$. Therefore, letting $T\to\infty$ and applying Lemma C.1 again, we obtain

$$\lim_{T\to\infty}\lim_{d\to\infty}\tilde{\mathsf{b}}_T=\frac{1}{\delta}\,\mathbb{E}\left\{f_0'\Big(\frac{a}{\sqrt{\delta}}X\,+\,\frac{\sqrt{1-a^2}}{\sqrt{\delta}}W\Big)\right\}\quad\text{a.s.} \tag{B.62}$$

From Lemma B.2, we have $\beta_{T-1}\to 1/\sqrt{\delta}$ as $T\to\infty$. Also recall from assumption **(A2)** on p. 3 that τ is the supremum of the support of Z_s, and that $\lambda^*_\delta>\tau$. Therefore, $Z=\frac{Z_s}{\lambda^*_\delta-Z_s}$ has bounded support, due to which $\|\boldsymbol{Z}^2\big(\boldsymbol{I}+\boldsymbol{Z}/(\sqrt{\delta}\beta_{T-1})\big)^{-1}\|^2_{\rm op}<C$ for a universal constant $C>0$. Hence,

$$\lim_{T\to\infty}\lim_{d\to\infty}S_2=0\ \ \text{a.s.} \tag{B.63}$$

To bound S_3, we first write the term inside the norm on the second line of (B.56) as

$$\frac{\sqrt{\delta}}{\lambda^*_\delta}\boldsymbol{Z}_s\boldsymbol{A}\hat{\boldsymbol{x}}^0(\bar{\mathsf{b}}_0-\tilde{\mathsf{b}}_T)+\frac{\tilde{\mathsf{b}}_T}{\lambda^*_\delta}\boldsymbol{Z}_s\boldsymbol{A}\left(\sqrt{\delta}\hat{\boldsymbol{x}}^0-\frac{\tilde{\boldsymbol{x}}^{T-1}}{\beta_{T-1}}\right)+\frac{\tilde{\mathsf{b}}_T}{\beta_{T-1}}\left(\frac{\boldsymbol{Z}_s}{\lambda^*_\delta}-\boldsymbol{Z}\Big(\boldsymbol{I}+\frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\Big)^{-1}\right)\boldsymbol{A}\tilde{\boldsymbol{x}}^{T-1}.$$

Then, using triangle inequality and Cauchy-Schwarz, we have

$$\begin{aligned}S_3\le\ &\frac{3\delta}{(\lambda^*_\delta)^2}\frac{\|\boldsymbol{Z}_s\boldsymbol{A}\hat{\boldsymbol{x}}^0\|_2^2}{d}(\bar{\mathsf{b}}_0-\tilde{\mathsf{b}}_T)^2\,+\,\frac{3\tilde{\mathsf{b}}_T^2}{(\lambda^*_\delta)^2}\|\boldsymbol{Z}_s\boldsymbol{A}\|^2_{\rm op}\frac{\|\sqrt{\delta}\hat{\boldsymbol{x}}^0-\tilde{\boldsymbol{x}}^{T-1}/\beta_{T-1}\|_2^2}{d}\\&+\,\frac{3\tilde{\mathsf{b}}_T^2}{\beta^2_{T-1}}\,\frac{\|\boldsymbol{A}\tilde{\boldsymbol{x}}^{T-1}\|_2^2}{d}\,\left\|\frac{1}{\lambda^*_\delta}\boldsymbol{Z}_s\,-\,\boldsymbol{Z}\left(\boldsymbol{I}+\frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\right\|^2_{\rm op}:=3(S_{3a}+S_{3b}+S_{3c}).\end{aligned} \tag{B.64}$$

Using the expression for $\hat{\boldsymbol{x}}^0$ from (5.11) and applying Cauchy-Schwarz, we can bound S_{3a} as:

$$S_{3a}\le\frac{1}{(\lambda^*_\delta)^2}\|\boldsymbol{Z}_s\|^2_{\rm op}\|\boldsymbol{A}\|^2_{\rm op}\|\hat{\boldsymbol{x}}^{\rm s}\|_2^2(\bar{\mathsf{b}}_0-\tilde{\mathsf{b}}_T)^2. \tag{B.65}$$

We note that Z_s is bounded, $\|\hat{\boldsymbol{x}}^{\rm s}\|_2=1$, and $\|\boldsymbol{A}\|^2_{\rm op}$ is bounded almost surely by a universal constant for sufficiently large d. Moreover, recalling the definitions of $\bar{\mathsf{b}}_0$ and $X_0=\mu_{X,0}X+\sigma_{X,0}W_{X,0}$ from (5.15) and (3.9), we see that $\bar{\mathsf{b}}_0=\frac{1}{\delta}\mathbb{E}\{f_0'(X_0)\}$ is the limit of $\tilde{\mathsf{b}}_T$ in (B.62). Therefore $\lim_{T\to\infty}\lim_{d\to\infty}S_{3a}=0$ almost surely.

Next, we bound S_{3b}. Recalling that $\hat{\boldsymbol{x}}^0=\sqrt{d}\hat{\boldsymbol{x}}^{\rm s}/\sqrt{\delta}$, we have

$$\begin{aligned}\frac{\|\sqrt{\delta}\hat{\boldsymbol{x}}^0-\tilde{\boldsymbol{x}}^{T-1}/\beta_{T-1}\|_2^2}{d}&=\frac{\|\sqrt{d}\hat{\boldsymbol{x}}^{\rm s}-\sqrt{\delta}\tilde{\boldsymbol{x}}^T+\sqrt{\delta}\tilde{\boldsymbol{x}}^T-\sqrt{\delta}\tilde{\boldsymbol{x}}^{T-1}+\sqrt{\delta}\tilde{\boldsymbol{x}}^{T-1}-\tilde{\boldsymbol{x}}^{T-1}/\beta_{T-1}\|_2^2}{d}\\&\le\frac{3\|\sqrt{d}\hat{\boldsymbol{x}}^{\rm s}-\sqrt{\delta}\tilde{\boldsymbol{x}}^T\|_2^2}{d}+\frac{3\|\sqrt{\delta}\tilde{\boldsymbol{x}}^T-\sqrt{\delta}\tilde{\boldsymbol{x}}^{T-1}\|_2^2}{d}+\frac{3\|\tilde{\boldsymbol{x}}^{T-1}\|_2^2}{d}(\sqrt{\delta}\,-\,1/\beta_{T-1})^2.\end{aligned} \tag{B.66}$$

Lemmas B.3 and B.4 imply that the first two terms on the RHS of (B.66) tend to zero in the iterated limit $T\to\infty,d\to\infty$. Furthermore, from Lemma B.2, we have $\lim_{T\to\infty}\beta_{T-1}=1/\sqrt{\delta}$. From Proposition B.1, we also have

$$\lim_{d\to\infty}\frac{\|\tilde{\boldsymbol{x}}^{T-1}\|_2^2}{d}=\mu^2_{\tilde{X},T-1}+\sigma^2_{\tilde{X},T-1}=\beta^2_{T-1}\quad\text{a.s.} \tag{B.67}$$

Therefore, $\lim_{T\to\infty}\lim_{d\to\infty}S_{3b}=0$ almost surely.

To bound S_{3c}, recalling from (B.52) that $Z=\frac{Z_s}{\lambda^*_\delta-Z_s}$, we have

$$\frac{1}{\lambda^*_\delta}\boldsymbol{Z}_s\,-\,\boldsymbol{Z}\left(\boldsymbol{I}+\frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}=\frac{1}{\beta_{T-1}}\boldsymbol{Z}_s^2\left(\lambda^*_\delta\boldsymbol{I}+\boldsymbol{Z}_s\Big(\frac{1}{\sqrt{\delta}\beta_{T-1}}-1\Big)\right)^{-1}\frac{(\frac{1}{\sqrt{\delta}}-\beta_{T-1})}{\lambda^*_\delta}. \tag{B.68}$$

Since $\lim_{T\to\infty}\beta_{T-1}=\frac{1}{\sqrt{\delta}}$, almost surely

$$\lim_{T\to\infty}\left\|\frac{1}{\lambda^*_\delta}\boldsymbol{Z}_s\,-\,\boldsymbol{Z}\left(\boldsymbol{I}+\frac{1}{\sqrt{\delta}\beta_{T-1}}\boldsymbol{Z}\right)^{-1}\right\|^2_{\rm op}=0. \tag{B.69}$$

Thus $\lim_{T\to\infty}\lim_{d\to\infty} S_{3c} = 0$ almost surely. Using the results above in (B.64), we have shown that

$$\lim_{T\to\infty}\lim_{d\to\infty} S_3 = 0 \quad \text{a.s.} \tag{B.70}$$

Using (B.59), (B.63) and (B.70) in (B.56), and recalling that $n/d \to \delta$, we obtain

$$\lim_{T\to\infty}\lim_{n\to\infty} \frac{\|\tilde{\boldsymbol{u}}^T - \hat{\boldsymbol{u}}^0\|_2}{\sqrt{n}} = 0. \tag{B.71}$$

To complete the proof for the base case, we show that the term inside the brackets in (B.50) is finite almost surely as $n \to \infty$. First, by assumption **(B2)** on p. 4, we have $\lim_{n\to\infty} \|\boldsymbol{y}\|_2^2/n = \mathbb{E}\{Y^2\}$ almost surely. Furthermore, by Proposition B.1, we almost surely have

$$\lim_{n\to\infty} \|\tilde{\boldsymbol{u}}^T\|_2^2/n = \mu_{\tilde{U},T}^2 + \sigma_{\tilde{U},T}^2. \tag{B.72}$$

Next, using the triangle inequality, we have

$$\|\tilde{\boldsymbol{u}}^T\|_2 - \|\tilde{\boldsymbol{u}}^T - \hat{\boldsymbol{u}}^0\|_2 \le \|\hat{\boldsymbol{u}}^0\|_2 \le \|\tilde{\boldsymbol{u}}^T\|_2 + \|\tilde{\boldsymbol{u}}^T - \hat{\boldsymbol{u}}^0\|_2. \tag{B.73}$$

Combining this with (B.71), we obtain

$$\lim_{T\to\infty}\lim_{n\to\infty} \frac{\|\hat{\boldsymbol{u}}^0\|_2^2}{n} = \lim_{T\to\infty} \mu_{\tilde{U},T}^2 + \sigma_{\tilde{U},T}^2 = \mu_{U,0}^2 + \sigma_{U,0}^2 \quad \text{a.s.} \tag{B.74}$$

Therefore, using (B.50), we have shown that

$$\lim_{T\to\infty}\lim_{n\to\infty} \left| \frac{1}{n}\sum_{i=1}^n \psi(y_i, \tilde{u}_i^T) - \frac{1}{n}\sum_{i=1}^n \psi(y_i, \hat{u}_i^0) \right| = 0 \quad \text{a.s.} \tag{B.75}$$

Induction step: Assume that (B.15) holds for some t, and that (B.16) holds with $t+1$ replaced by t. Also assume towards induction that almost surely

$$\lim_{T\to\infty}\lim_{d\to\infty} \frac{\|\tilde{\boldsymbol{x}}^{T+t} - \hat{\boldsymbol{x}}^t\|_2^2}{d} = 0, \qquad \lim_{T\to\infty}\lim_{n\to\infty} \frac{\|\tilde{\boldsymbol{u}}^{T+t} - \hat{\boldsymbol{u}}^t\|_2^2}{n} = 0. \tag{B.76}$$

The limits in (B.76) hold for $t = 0$, as established in the proof of the base case (see (B.66), (B.71)).

From (B.48), we have the bound

$$\begin{aligned}
&\left| \frac{1}{d}\sum_{i=1}^d \psi(x_i, \tilde{x}_i^{T+t+1}) - \frac{1}{d}\sum_{i=1}^d \psi(x_i, \hat{x}_i^{t+1}) \right| \\
&\le 4C\left[1 + \frac{\|\boldsymbol{x}\|_2^2}{d} + \frac{\|\tilde{\boldsymbol{x}}^{T+t+1}\|_2^2}{d} + \frac{\|\hat{\boldsymbol{x}}^{t+1}\|_2^2}{d}\right]^{\frac{1}{2}} \frac{\|\tilde{\boldsymbol{x}}^{T+t+1} - \hat{\boldsymbol{x}}^{t+1}\|_2}{\sqrt{d}}.
\end{aligned} \tag{B.77}$$

Using (5.1), (5.6), (5.13) and the triangle inequality, we obtain:

$$\begin{aligned}
\frac{\|\tilde{\boldsymbol{x}}^{T+t+1} - \hat{\boldsymbol{x}}^{t+1}\|_2^2}{d} &\le \frac{2}{\delta d}\|\boldsymbol{A}^\mathsf{T} h_t(\tilde{\boldsymbol{u}}^{T+t}; \boldsymbol{y}) - \boldsymbol{A}^\mathsf{T} h_t(\hat{\boldsymbol{u}}^t; \boldsymbol{y})\|_2^2 + 2\,\frac{\|\tilde{\mathsf{c}}_{T+t} f_t(\tilde{\boldsymbol{x}}^{T+t}) - \bar{\mathsf{c}}_t f_t(\hat{\boldsymbol{x}}^t)\|_2^2}{d} \\
&\le \frac{2}{\delta d}\|\boldsymbol{A}^\mathsf{T} h_t(\tilde{\boldsymbol{u}}^{T+t}; \boldsymbol{y}) - \boldsymbol{A}^\mathsf{T} h_t(\hat{\boldsymbol{u}}^t; \boldsymbol{y})\|_2^2 + 4\,\frac{\|f_t(\tilde{\boldsymbol{x}}^{T+t})\|_2^2}{d}(\tilde{\mathsf{c}}_{T+t} - \bar{\mathsf{c}}_t)^2 + 4\,\bar{\mathsf{c}}_t^2\,\frac{\|f_t(\tilde{\boldsymbol{x}}^{T+t}) - f_t(\hat{\boldsymbol{x}}^t)\|_2^2}{d} \\
&:= 2\,S_1 + 4\,S_2 + 4\,S_3.
\end{aligned} \tag{B.78}$$

The term S_1 can be bounded as

$$S_1 \le \|\boldsymbol{A}\|_{\text{op}}^2\, \frac{\|h_t(\tilde{\boldsymbol{u}}^{T+t}; \boldsymbol{y}) - h_t(\hat{\boldsymbol{u}}^t; \boldsymbol{y})\|_2^2}{\delta d} \le \|\boldsymbol{A}\|_{\text{op}}^2\, \bar{L}_t^2\, \frac{\|\tilde{\boldsymbol{u}}^{T+t} - \hat{\boldsymbol{u}}^t\|_2^2}{\delta d}, \tag{B.79}$$

where $\bar{L}_t$ is the Lipschitz constant of the function h_t. Since the operator norm of $\boldsymbol{A}$ is bounded almost surely as $d \to \infty$, by the induction hypothesis (B.76) we have $\lim_{T\to\infty} \lim_{d\to\infty} \frac{\|\tilde{\boldsymbol{u}}^{T+t} - \hat{\boldsymbol{u}}^t\|^2}{\delta d} = 0$ almost surely. Therefore,

$$\lim_{T\to\infty} \lim_{d\to\infty} S_1 = 0 \quad \text{a.s.} \tag{B.80}$$

To bound S_2, we recall from (5.7) that $\tilde{\mathsf{c}}_{T+t} = \frac{1}{n}\sum_i h_t'(\tilde{u}_i^t; y_i)$. Proposition B.1 guarantees that the joint empirical distribution of $(\tilde{\boldsymbol{u}}^{T+t}, \boldsymbol{y})$ converges to the law of $(\tilde{U}_{T+t}, Y) \equiv (\mu_{\tilde{U},T+t} G + \sigma_{\tilde{U},T+t} W_{U,T+t},\, Y)$. Since h_t is Lipschitz, Lemma C.1 in Appendix C implies that

$$\lim_{n\to\infty} \tilde{\mathsf{c}}_{T+t} = \mathbb{E}\{h_t'(\mu_{\tilde{U},T+t} G + \sigma_{\tilde{U},T+t} W_{U,T+t},\, Y)\} \quad \text{a.s.} \tag{B.81}$$

From (B.13), we know that $\lim_{T\to\infty} \mu_{\tilde{U},T+t} = \mu_{U,t}$ and $\lim_{T\to\infty} \sigma^2_{\tilde{U},T+t} = \sigma^2_{U,t}$. Therefore applying Lemma C.1 in Appendix C again, we obtain:

$$\lim_{T\to\infty} \lim_{n\to\infty} \tilde{\mathsf{c}}_{T+t} = \mathbb{E}\{h_t'(\mu_{U,t} G + \sigma_{U,t} W_{U,t},\, Y)\} = \bar{\mathsf{c}}_t \quad \text{a.s.} \tag{B.82}$$

Next, using the result in Proposition B.1 with the test function $\psi(x, \tilde{x}) = (f_t(\tilde{x}))^2$, we almost surely have

$$\lim_{T\to\infty} \lim_{d\to\infty} \frac{\|f_t(\tilde{\boldsymbol{x}}^{T+t})\|_2^2}{d} = \lim_{T\to\infty} \mathbb{E}\{f_t(\tilde{X}_{T+t})^2\} = \mathbb{E}\{f_t(X_t)^2\}, \tag{B.83}$$

where the last equality follows from (B.13) since f_t is Lipschitz. Combining the above with (B.82), we obtain

$$\lim_{T\to\infty} \lim_{d\to\infty} S_2 = 0 \quad \text{a.s.} \tag{B.84}$$

For the third term S_3 in (B.78), since f_t is Lipschitz (with Lipschitz constant denoted by L_t), we have the bound:

$$S_3 \le \bar{\mathsf{c}}_t^2 L_t^2 \frac{\|\tilde{\boldsymbol{x}}^{T+t} - \boldsymbol{x}^t\|_2^2}{d}. \tag{B.85}$$

Thus, by the induction hypothesis (B.76), we obtain

$$\lim_{T\to\infty} \lim_{d\to\infty} S_3 = 0 \quad \text{a.s.} \tag{B.86}$$

We have therefore shown that

$$\lim_{T\to\infty} \lim_{d\to\infty} \frac{\|\tilde{\boldsymbol{x}}^{T+t+1} - \hat{\boldsymbol{x}}^{t+1}\|^2}{d} = 0 \quad \text{a.s.} \tag{B.87}$$

Next, we show that the terms inside the brackets on the RHS of (B.77) are finite almost surely as $d \to \infty$. Using the pseudo-Lipschitz test function $\psi(x, \tilde{x}) = x^2 + \tilde{x}^2$, Proposition B.1 implies that almost surely

$$\lim_{d\to\infty} \frac{1}{d} \sum_{i=1}^{d} \left(|x_i|^2 + |\tilde{x}_i^{T+t+1}|^2 \right) = \mathbb{E}\{X^2\} + \mu^2_{\tilde{X},T+t+1} + \sigma^2_{\tilde{X},T+t+1}. \tag{B.88}$$

Moreover, (B.14) implies that $\lim_{T\to\infty} \mu^2_{\tilde{X},T+t+1} + \sigma^2_{\tilde{X},T+t+1} = \mu^2_{X,t+1} + \sigma^2_{X,t+1}$. Using the triangle inequality, we have

$$\|\tilde{\boldsymbol{x}}^{T+t+1}\|_2 - \|\tilde{\boldsymbol{x}}^{T+t+1} - \hat{\boldsymbol{x}}^{t+1}\|_2 \le \|\hat{\boldsymbol{x}}^{t+1}\|_2 \le \|\tilde{\boldsymbol{x}}^{T+t+1}\|_2 + \|\hat{\boldsymbol{x}}^{t+1} - \tilde{\boldsymbol{x}}^{T+t+1}\|_2. \tag{B.89}$$

Hence, using (B.87) and Proposition B.1, we almost surely have

$$\lim_{T\to\infty} \lim_{d\to\infty} \frac{\|\hat{\boldsymbol{x}}^{t+1}\|_2^2}{d} = \lim_{T\to\infty} \lim_{d\to\infty} \frac{\|\tilde{\boldsymbol{x}}^{T+t+1}\|_2^2}{d} = \lim_{T\to\infty} \left(\mu^2_{\tilde{X},T+t+1} + \sigma^2_{\tilde{X},T+t+1} \right) = \mu^2_{X,t+1} + \sigma^2_{X,t+1}. \tag{B.90}$$

We have thus shown via (B.77) that almost surely

$$\lim_{T\to\infty} \lim_{d\to\infty} \left| \frac{1}{d} \sum_{i=1}^{d} \psi(x_i, \tilde{x}_i^{T+t+1}) - \frac{1}{d} \sum_{i=1}^{d} \psi(x_i, \hat{x}_i^{t+1}) \right| = 0. \tag{B.91}$$

To complete the proof via induction, we need to show that if (B.87) and (B.91) hold with $(t+1)$ replaced by t for some $t > 0$, then almost surely

$$\lim_{T\to\infty}\lim_{n\to\infty}\frac{\|\tilde{\boldsymbol{u}}^{T+t}-\hat{\boldsymbol{u}}^t\|_2^2}{n}=0, \qquad \lim_{T\to\infty}\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^n\psi(y_i,\tilde{u}_i^{T+t})-\frac{1}{n}\sum_{i=1}^n\psi(y_i,\hat{u}_i^t)\right|=0. \tag{B.92}$$

From (B.49), we have the bound

$$\begin{aligned}&\left|\frac{1}{n}\sum_{i=1}^n\psi(y_i,\tilde{u}_i^{T+t})-\frac{1}{n}\sum_{i=1}^n\psi(y_i,\hat{u}_i^t)\right|\\ &\leq 4C\left[1+\frac{\|\boldsymbol{y}\|_2^2}{n}+\frac{\|\tilde{\boldsymbol{u}}^{T+t}\|_2^2}{n}+\frac{\|\hat{\boldsymbol{u}}^t\|_2^2}{n}\right]^{\frac{1}{2}}\frac{\|\tilde{\boldsymbol{u}}^{T+t}-\hat{\boldsymbol{u}}^t\|_2}{\sqrt{n}}.\end{aligned} \tag{B.93}$$

Using (5.2), (5.6), (5.14) and the triangle inequality, we obtain

$$\begin{aligned}&\frac{\|\tilde{\boldsymbol{u}}^{T+t}-\hat{\boldsymbol{u}}^t\|_2^2}{n}\leq\frac{2}{\delta n}\|\boldsymbol{A}f_t(\tilde{\boldsymbol{x}}^{T+t})-\boldsymbol{A}f_t(\hat{\boldsymbol{x}}^t)\|_2^2+2\frac{\|\tilde{\mathsf{b}}_{T+t}h_{t-1}(\tilde{\boldsymbol{u}}^{T+t-1};\boldsymbol{y})-\bar{\mathsf{b}}_th_{t-1}(\hat{\boldsymbol{u}}^{t-1};\boldsymbol{y})\|_2^2}{n}\\ &\leq\frac{2}{\delta n}\|\boldsymbol{A}f_t(\tilde{\boldsymbol{x}}^{T+t})-\boldsymbol{A}f_t(\hat{\boldsymbol{x}}^t)\|_2^2+4\,\frac{\|h_{t-1}(\hat{\boldsymbol{u}}^{t-1};\boldsymbol{y})\|_2^2}{n}(\tilde{\mathsf{b}}_{T+t}-\bar{\mathsf{b}}_t)^2\\ &\quad+4\bar{\mathsf{b}}_t^2\,\frac{\|h_{t-1}(\tilde{\boldsymbol{u}}^{T+t-1};\boldsymbol{y})-h_{t-1}(\hat{\boldsymbol{u}}^{t-1};\boldsymbol{y})\|_2^2}{n}\;:=2\,S_1+4\,S_2+4\,S_3.\end{aligned} \tag{B.94}$$

Using arguments along the same lines as (B.80)-(B.86) (omitted for brevity), we can show that almost surely

$$\lim_{T\to\infty}\lim_{n\to\infty}S_1=\lim_{T\to\infty}\lim_{n\to\infty}S_2=\lim_{T\to\infty}\lim_{n\to\infty}S_3=0.$$

Hence $\lim_{T\to\infty}\lim_{n\to\infty}\frac{\|\tilde{\boldsymbol{u}}^{T+t}-\hat{\boldsymbol{u}}^t\|_2}{\sqrt{n}}=0$ almost surely. Furthermore, using a triangle inequality argument as in (B.89), we obtain $\lim_{T\to\infty}\lim_{n\to\infty}\frac{\|\tilde{\boldsymbol{u}}^{T+t}\|_2^2}{n}=\lim_{T\to\infty}\lim_{n\to\infty}\frac{\|\hat{\boldsymbol{u}}^t\|_2^2}{n}$ almost surely. By Proposition B.1 and (B.13), the latter limit equals $\mu_{U,t}^2+\sigma_{U,t}^2$. Using these limits in (B.93) yields the result (B.92), and completes the proof of the lemma. □

B.4 Putting Everything Together: Proof of Theorem 1

We will first use Lemma B.5 to show that the result of the theorem holds for the GAMP iteration $(\hat{\boldsymbol{x}}^t,\hat{\boldsymbol{u}}^t)$, i.e., under the assumptions of Theorem 1, we almost surely have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\psi(y_i,\hat{u}_i^t)=\mathbb{E}\left\{\psi(Y,\,\mu_{U,t}G+\sigma_{U,t}W_{U,t})\right\},\quad t\geq 0, \tag{B.95}$$

$$\lim_{d\to\infty}\frac{1}{d}\sum_{i=1}^d\psi(x_i,\hat{x}_i^{t+1})=\mathbb{E}\left\{\psi(X,\,\mu_{X,t+1}X+\sigma_{X,t+1}\,W_{X,t+1}\right\},\quad t+1\geq 0. \tag{B.96}$$

Consider the LHS of (B.96). Using the triangle inequality, for any $T>0$, we have

$$\begin{aligned}&\left|\frac{1}{d}\sum_{i=1}^d\psi(x_i,\hat{x}_i^{t+1})-\mathbb{E}\left\{\psi(X,\,\mu_{X,t+1}X+\sigma_{X,t+1}\,W_{X,t+1}\right\}\right|\\ &\leq\left|\frac{1}{d}\sum_{i=1}^d\psi(x_i,\hat{x}_i^{t+1})-\frac{1}{d}\sum_{i=1}^d\psi(x_i,\tilde{x}_i^{T+t+1})\right|+\left|\frac{1}{d}\sum_{i=1}^d\psi(x_i,\tilde{x}_i^{T+t+1})-\mathbb{E}\{\psi(X,\mu_{\tilde{X},T+t+1}X+\sigma_{\tilde{X},T+t+1}W_{\tilde{X},T+t+1})\}\right|\\ &+\left|\mathbb{E}\{\psi(X,\mu_{\tilde{X},T+t+1}X+\sigma_{\tilde{X},T+t+1}W_{\tilde{X},T+t+1})\}-\mathbb{E}\{\psi(X,\mu_{X,t+1}X+\sigma_{X,t+1}W_{X,t+1})\}\right|\;:=T_1+T_2+T_3.\end{aligned} \tag{B.97}$$

We first bound T_3 using the pseudo-Lipschitz property of ψ, noting that $W_{\tilde{X},T+t}$ and $W_{X,t}$ are both $\sim \mathsf{N}(0,1)$:

$$\begin{aligned}
T_3 &\leq \mathbb{E}\left\{\left|\psi(X, \mu_{\tilde{X},T+t+1}X + \sigma_{\tilde{X},T+t+1}W) - \psi(X, \mu_{X,t+1}X + \sigma_{X,t+1}W)\right|\right\}, \qquad W \sim \mathsf{N}(0,1) \\
&\leq C\mathbb{E}\Bigg\{\left(1 + \left[X^2 + \mu^2_{\tilde{X},T+t+1}X^2 + \sigma^2_{\tilde{X},T+t+1}W^2\right]^{1/2} + \left[X^2 + \mu^2_{X,t+1}X^2 + \sigma^2_{X,t+1}W^2\right]^{1/2}\right) \\
&\qquad\qquad \cdot \left(X^2(\mu_{\tilde{X},T+t+1} - \mu_{X,t+1})^2 + W^2(\sigma_{\tilde{X},T+t+1} - \sigma_{X,t+1})^2\right)^{1/2}\Bigg\} \\
&\leq 3C\left(3 + \mu^2_{\tilde{X},T+t+1} + \sigma^2_{\tilde{X},T+t+1} + \mu^2_{X,t+1} + \sigma^2_{X,t+1}\right)^{1/2}\left((\mu_{\tilde{X},T+t+1} - \mu_{X,t+1})^2 + (\sigma_{\tilde{X},T+t+1} - \sigma_{X,t+1})^2\right)^{1/2},
\end{aligned} \tag{B.98}$$

where we have used Cauchy-Schwarz inequality in the last line. From Lemma B.5 (Eq. (B.14)), we know that $\lim_{T\to\infty}\left|\mu_{\tilde{X},T+t+1} - \mu_{X,t+1}\right| = 0$ and $\lim_{T\to\infty}\left|\sigma_{\tilde{X},T+t+1} - \sigma_{X,t+1}\right| = 0$. Therefore, $\lim_{T\to\infty} T_3 = 0$. Next, from (B.16) we have that $\lim_{T\to\infty}\lim_{d\to\infty} T_1 = 0$ almost surely. Furthermore, by Proposition B.1, for any $T > 0$ we almost surely have $\lim_{d\to\infty} T_2 = 0$. Letting $T, d \to \infty$ (with the limit in d taken first) and noting that the LHS of (B.97) does not depend on T, we obtain that (B.96) holds.

The proof of (B.95) uses a bound similar to (B.97) and arguments along the same lines. It is omitted for brevity.

Next, we prove the main result by showing that under the assumptions of the theorem, almost surely

$$\lim_{n\to\infty}\left|\frac{1}{n}\sum_{i=1}^{n}\psi(y_i, u_i^t) - \frac{1}{n}\sum_{i=1}^{n}\psi(y_i, \hat{u}_i^t)\right| = 0, \qquad \lim_{n\to\infty}\frac{\|\boldsymbol{u}^t - \hat{\boldsymbol{u}}^t\|_2^2}{n} = 0, \quad t \geq 0 \tag{B.99}$$

$$\lim_{d\to\infty}\left|\frac{1}{d}\sum_{i=1}^{d}\psi(x_i, x_i^{t+1}) - \frac{1}{d}\sum_{i=1}^{d}\psi(x_i, \hat{x}_i^{t+1})\right| = 0, \qquad \lim_{d\to\infty}\frac{\|\boldsymbol{x}^{t+1} - \hat{\boldsymbol{x}}^{t+1}\|_2^2}{d} = 0, \quad t+1 \geq 0. \tag{B.100}$$

Combining (B.100)-(B.99) with (B.96)-(B.95) yields the results in (3.11) and (3.12).

The proof of (B.100) and (B.99) is via induction and uses arguments very similar to those to prove (B.15)-(B.16). To avoid repetition we only provide a few steps. Noting that $\boldsymbol{x}^0 = \hat{\boldsymbol{x}}^0$, we now show (B.100), under the induction hypothesis that (B.99) holds and also that (B.100) holds with $t+1$ replaced by t.

Since $\psi \in \text{PL}(2)$, we have

$$\left|\frac{1}{d}\sum_{i=1}^{d}\psi(x_i, x_i^{t+1}) - \frac{1}{d}\sum_{i=1}^{d}\psi(x_i, \hat{x}_i^{t+1})\right| \leq 4C\left[1 + \frac{\|\boldsymbol{x}\|_2^2}{d} + \frac{\|\boldsymbol{x}^{t+1}\|_2^2}{d} + \frac{\|\hat{\boldsymbol{x}}^{t+1}\|_2^2}{d}\right]^{\frac{1}{2}}\frac{\|\boldsymbol{x}^{t+1} - \hat{\boldsymbol{x}}^{t+1}\|_2}{\sqrt{d}}. \tag{B.101}$$

Furthermore, using the definitions of $\boldsymbol{x}^{t+1}$ and $\hat{\boldsymbol{x}}^{t+1}$, and the triangle inequality we have

$$\begin{aligned}
\frac{\|\boldsymbol{x}^{t+1} - \hat{\boldsymbol{x}}^{t+1}\|_2^2}{d} &\leq \frac{2}{\delta d}\|\boldsymbol{A}^{\mathsf{T}}h_t(\boldsymbol{u}^t;\boldsymbol{y}) - \boldsymbol{A}^{\mathsf{T}}h_t(\hat{\boldsymbol{u}}^t;\boldsymbol{y})\|_2^2 + 4\,\frac{\|f_t(\boldsymbol{x}^t)\|_2^2}{d}(\mathsf{c}_t - \bar{\mathsf{c}}_t)^2 + 4\bar{\mathsf{c}}_t^2\frac{\|f_t(\boldsymbol{x}^t) - f_t(\hat{\boldsymbol{x}}^t)\|_2^2}{d} \\
&\leq \frac{2\bar{L}_t^2}{\delta}\|\boldsymbol{A}\|_{\text{op}}^2\frac{\|\boldsymbol{u}^t - \hat{\boldsymbol{u}}^t\|_2^2}{d} + 4\,\frac{\|f_t(\boldsymbol{x}^t)\|_2^2}{d}(\mathsf{c}_t - \bar{\mathsf{c}}_t)^2 + 4\bar{\mathsf{c}}_t^2L_t^2\frac{\|\boldsymbol{x}^t - \hat{\boldsymbol{x}}^t\|_2^2}{d},
\end{aligned} \tag{B.102}$$

where $L_t, \bar{L}_t$ are the Lipschitz constants of f_t, h_t, respectively. By the induction hypothesis and Lemma C.1, the terms $\frac{\|\boldsymbol{u}^t - \hat{\boldsymbol{u}}^t\|_2^2}{d}$, $\frac{\|\boldsymbol{x}^t - \hat{\boldsymbol{x}}^t\|_2^2}{d}$, and $(\mathsf{c}_t - \bar{\mathsf{c}}_t)^2$ tend to zero. Furthermore, by the induction hypothesis, we almost surely have $\frac{\|f_t(\boldsymbol{x}^t)\|_2^2}{d} \to \mathbb{E}\{f_t(X_t)^2\}$, and by (B.96), $\frac{\|\hat{\boldsymbol{x}}^{t+1}\|_2^2}{d} \to (\mu^2_{X,t+1} + \sigma^2_{X,t+1})$ as $d \to \infty$. Finally, by a triangle inequality argument analogous to (B.89), we also have

$$\lim_{d\to\infty}\frac{\|\boldsymbol{x}^{t+1}\|_2^2}{d} = \lim_{d\to\infty}\frac{\|\hat{\boldsymbol{x}}^{t+1}\|_2^2}{d} = (\mu^2_{X,t+1} + \sigma^2_{X,t+1}) \quad \text{a.s.}$$

Using these limits in (B.101) proves (B.100). The proof of (B.99) (under the induction hypothesis that (B.100) holds with $(t+1)$ replaced by t) is along the same lines: we use a bound similar to (B.101) and a decomposition of $\frac{\|\boldsymbol{u}^t - \hat{\boldsymbol{u}}^t\|_2^2}{n}$ similar to (B.102). This completes the proof of the theorem. □

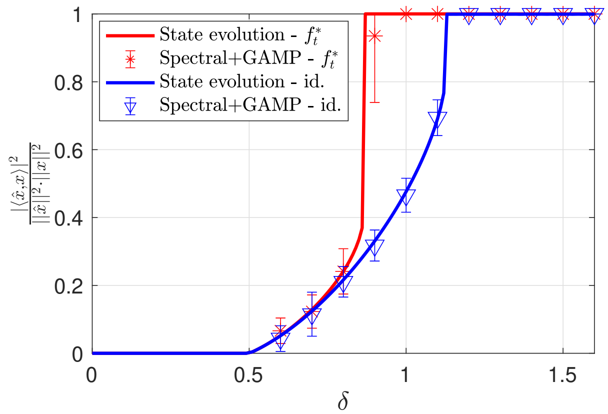

Figure 3: Performance comparison between two different choices of f_t for a binary prior $P_X(1) = P_X(-1) = \frac{1}{2}$. The Bayes-optimal choice $f_t = f_t^*$ (in red) has a lower threshold compared to f_t equal to identity (in blue).

C An Auxiliary Lemma

The following result is proved in (Bayati and Montanari, 2011, Lemma 6).

Lemma C.1. *Let $F\colon \mathbb{R}^2 \to \mathbb{R}$ be a Lipschitz function, and let $F'(u, v)$ denote its derivative with respect to the first argument at $(u, v) \in \mathbb{R}^2$. Assume that $F'(\cdot, v)$ is continuous almost everywhere in the first argument, for each $v \in \mathbb{R}$. Let (U_m, V_m) be a sequence of random vectors in $\mathbb{R}^2$ converging in distribution to the random vector (U, V) as $m \to \infty$. Furthermore, assume that the distribution of U is absolutely continuous with respect to the Lebesgue measure. Then,*

$$\lim_{m\to\infty} \mathbb{E}\{F'(U_m, V_m)\} = \mathbb{E}\{F'(U, V)\}.$$

D Bayes-optimal GAMP for a Binary-valued Prior

Consider the noiseless phase retrieval model, where each entry of the signal $\boldsymbol{x}$ takes value in $\{-1, 1\}$, with $P_X(1) = 1 - p_X(-1) = \mathsf{p}$. In Figure 3, we take $\mathsf{p} = \frac{1}{2}$, and compare the performance of the GAMP algorithm with spectral initialization for two different choices of the function f_t: f_t equal to identity (in blue) and $f_t = f_t^*$ (in red), where f_t^* is the Bayes-optimal choice (3.15). By computing the conditional expectation, we have

$$\begin{aligned} f_t^*(s) &= 2\mathbb{P}(X = 1 \mid \mu_{X,t} X + \sigma_{X,t} W = s) - 1 \\ &= \frac{2}{1 + \frac{1-\mathsf{p}}{\mathsf{p}} \exp\left(\frac{-2s\mu_{X,t}}{\sigma_{X,t}^2}\right)} - 1. \end{aligned} \tag{D.1}$$

The rest of the setting is analogous to that of Figure 1. There is a significant performance gap between the Bayes-optimal choice $f_t = f_t^*$ and the choice $f_t(x) = x$. As in the previous experiment, we observe very good agreement between the GAMP algorithm and the state evolution prediction of Theorem 1. We remark that for this setting, the information-theoretically optimal overlap (computed using the formula in (Barbier et al., 2019)) is 1 for all $\delta > 0$. Since the components of $\boldsymbol{x}$ are in $\{-1, 1\}$, there are 2^d choices for $\boldsymbol{x}$. The information-theoretically optimal estimator picks the choice that is consistent with $y_i = \langle \boldsymbol{x}, \boldsymbol{a}_i \rangle$, $i \in [n]$. (Since $\boldsymbol{A}$ is Gaussian, with high probability this solution is unique.)

E Complex-valued GAMP

Consider a complex sensing matrix $\boldsymbol{A}$ with rows distributed as $(\boldsymbol{a}_i) \sim_{\text{i.i.d.}} \mathsf{CN}(0, \boldsymbol{I}_d/d))$, for $i \in [n]$. The output of the GLM $\boldsymbol{y} \in \mathbb{C}^n$ is generated as $p_{Y|G}(\boldsymbol{y} \mid \boldsymbol{g})$, where $\boldsymbol{g} = \boldsymbol{A}\boldsymbol{x}$. The GAMP algorithm for the complex setting has been studied in the context of phase retrieval by (Schniter and Rangan, 2014; Ma et al., 2019). Here, we briefly review the complex GAMP and present some numerical results for complex GAMP with spectral initialization.

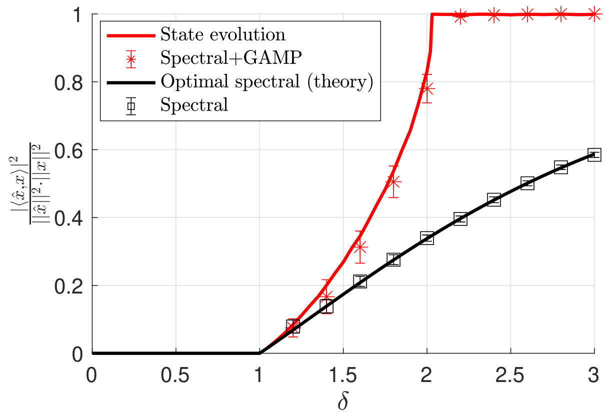

Figure 4: Performance comparison between complex GAMP with spectral initialization (in red) and the spectral method alone (in black) for a Gaussian prior $P_X \sim \mathsf{CN}(0, 1)$. On the x-axis, we have the sampling ratio $\delta = n/d$; on the y-axis, we have the normalized squared scalar product between the signal and the estimate. The experimental results ($*$ and $\Box$ markers) are in excellent agreement with the theoretical predictions (solid lines) given by state evolution for GAMP and Lemma 2.1 for the spectral method. Error bars indicate one standard deviation around the empirical mean.

As in Section 4, we take f_t to be the identity function, and $h_t = \sqrt{\delta} h_t^*$, where h_t^* is given in (3.18). To obtain a compact state evolution recursion, we initialize with a scaled version of the spectral estimator $\hat{\boldsymbol{x}}^{\rm s}$:

$$\boldsymbol{x}^0 = \sqrt{d}\,\frac{a}{1-a^2}\,\hat{\boldsymbol{x}}^{\rm s}, \qquad \boldsymbol{u}^0 = \frac{1}{\sqrt{\delta}}\boldsymbol{A}\boldsymbol{x}^0 - \frac{1}{\sqrt{\delta}\lambda_\delta^*}\,\boldsymbol{Z}_s\boldsymbol{A}\boldsymbol{x}^0. \tag{E.1}$$

The iterates are then computed as:

$$\boldsymbol{x}^{t+1} = \boldsymbol{A}^{\mathsf{H}} h_t^*(\boldsymbol{u}^t; \boldsymbol{y}) - \mathsf{c}_t f_t(\boldsymbol{x}^t), \tag{E.2}$$

$$\boldsymbol{u}^{t+1} = \frac{1}{\sqrt{\delta}}\boldsymbol{A}\boldsymbol{x}^{t+1} - \frac{1}{\sqrt{\delta}} h_t^*(\boldsymbol{u}^t; \boldsymbol{y}). \tag{E.3}$$

Here, the Onsager coefficient c_t is given by (Schniter and Rangan, 2014):

$$\mathsf{c}_t = \frac{\sqrt{\delta}}{\mathsf{Var}(G \mid U_t = u)}\left(\frac{\mathsf{Var}\{G \mid U_t = u,\, Y = y\}}{\mathsf{Var}(G \mid U_t = u)} - 1\right). \tag{E.4}$$

For this choice of f_t, h_t, the state evolution iteration can be written in terms of a single parameter $\mu_t \equiv \mu_{X,t}$. For $t \geq 0$:

$$\begin{aligned} &\mu_{U,t} = \frac{1}{\sqrt{\delta}}\mu_t, \qquad \sigma_{U,t}^2 = \frac{\mu_t}{\delta}, \qquad \sigma_{X,t}^2 = \mu_{X,t} = \mu_t, \\ &\mu_{t+1} = \sqrt{\delta}\,\mathbb{E}\left\{|h_t^*(U_t;\, Y)|^2\right\}. \end{aligned} \tag{E.5}$$

The recursion is initialized with $\mu_0 = \frac{a^2}{1-a^2}$. Moreover, the parameter μ_{t+1} can be consistently estimated from the iterate $\boldsymbol{u}^t$ as $\hat{\mu}_{t+1} = \sqrt{\delta}\|h^*(\boldsymbol{u}^t; \boldsymbol{y})\|_2^2/n$. It can also be estimated as the positive solution of the quadratic equation $\hat{\mu}_{t+1}^2 + \hat{\mu}_{t+1} = \|\boldsymbol{x}^{t+1}\|_2^2/d$.

We now discuss some numerical results for noiseless (complex) phase retrieval, where $y_i = |(\boldsymbol{A}\boldsymbol{x})_i|^2$, for $i \in [n]$. For a given measurement matrix $\boldsymbol{A}$, note that replacing $\boldsymbol{x}$ by $e^{i\theta}\boldsymbol{x}$ leaves the measurement $\boldsymbol{y}$ unchanged. Therefore the performance of any estimator is measured up to a constant phase rotation:

$$\min_{\theta \in [0, 2\pi)} \frac{\left|\langle \hat{\boldsymbol{x}},\, e^{i\theta}\boldsymbol{x}\rangle\right|^2}{\|\boldsymbol{x}\|_2^2\,\|\hat{\boldsymbol{x}}\|_2^2}. \tag{E.6}$$

Figure 4 shows the performance of GAMP with spectral initialization when the signal $\boldsymbol{x}$ is uniform on the d-dimensional *complex* sphere with radius $\sqrt{d}$, and the sensing vectors $(\boldsymbol{a}_i) \sim_{\text{i.i.d.}} \mathsf{CN}(0, \boldsymbol{I}_d/d)$.

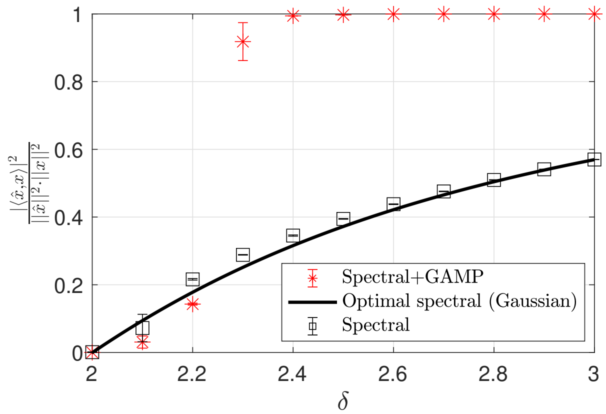

Figure 5: Performance comparison between complex GAMP with spectral initialization (in red) and the spectral method alone (in black) for a model of coded diffraction patterns.

Figure 5 shows the performance with coded diffraction pattern sensing vectors, given by (4.1). The signal $\boldsymbol{x}$ is the image in Figure 2a, which is a $d_1 \times d_2 \times 3$ array with $d_1 = 820$ and $d_2 = 1280$. The three components $\boldsymbol{x}_j \in \mathbb{R}^d$ ($j \in \{1, 2, 3\}$ and $d = d_1 \cdot d_2$) are treated separately, and the performance is measured via the average squared normalized scalar product $\frac{1}{3}\sum_{j=1}^{3} \frac{|\langle \hat{\boldsymbol{x}}_j, \boldsymbol{x}_j \rangle|^2}{\|\hat{\boldsymbol{x}}_j\|_2^2 \|\boldsymbol{x}_j\|_2^2}$.

The red points in Figure 5 are obtained by running the complex GAMP algorithm with spectral initialization, as given in (E.1)-(E.4). We perform $n_{\text{sample}} = 5$ independent trials and show error bars at 1 standard deviation. For comparison, the black points correspond to the empirical performance of the spectral method alone, and the black curve gives the theoretical prediction for the optimal squared correlation for Gaussian sensing vectors (see Theorem 1 of (Luo et al., 2019)).