On The Stability of Approximate Message Passing with Independent Measurement Ensembles

Dang Qua Nguyen and Taejoon Kim
Department of Electrical Engineering and Computer Science
The University of Kansas, Lawrence, KS 66045 USA
Email: {quand, taejoonkim}@ku.edu

***Abstract*—Approximate message passing (AMP) is a scalable, iterative approach to signal recovery. For structured random measurement ensembles, including independent and identically distributed (i.i.d.) Gaussian and rotationally-invariant matrices, the performance of AMP can be characterized by a scalar recursion called state evolution (SE). The pseudo-Lipschitz (polynomial) smoothness is conventionally assumed. In this work, we extend the SE for AMP to a new class of measurement matrices with independent (not necessarily identically distributed) entries. We also extend it to a general class of functions, called controlled functions which are not constrained by the polynomial smoothness; unlike the pseudo-Lipschitz function that has polynomial smoothness, the controlled function grows exponentially. The lack of structure in the assumed measurement ensembles is addressed by leveraging Lindeberg-Feller. The lack of smoothness of the assumed controlled function is addressed by a proposed conditioning technique leveraging the empirical statistics of the AMP instances. The resultants grant the use of the SE to a broader class of measurement ensembles and a new class of functions.**

***Index Terms*—Approximate message passing (AMP), state evolution (SE), controlled function, and random matrix theory.**

I. INTRODUCTION

The problem of signal recovery from a linear observation[1]

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{w} \tag{1}$$

appears in various fields [1]–[3], where $\mathbf{y} \in \mathbb{R}^{n\times 1}$, $\mathbf{A} \in \mathbb{R}^{n\times N}$ is a given measurement matrix, $\mathbf{x}_0 \in \mathbb{R}^{N\times 1}$ is the signal to be recovered, and $\mathbf{w} \in \mathbb{R}^{n\times 1}$ is an additive noise. Of particular interest is the case when $\mathbf{A}$ is overcomplete ($n \ll N$). However, the computational cost to solve this problem is typically prohibitive when the dimensions n and N are large.

Message passing (MP) can be applied to handle the large-dimensionality of the problem. Conventionally, sparsity is essential for MP to approach a fundamental performance limit [4]. Recently, the approximate message-passing (AMP) algorithm has received significant attention [5]–[7] because it performs surprisingly well in systems that are not sparse.

The remarkable features of AMP have inspired a wide range of applications [8]–[23]. Despite its widespread applicability, the AMP algorithm suffers from instability issues [24]–[26]. The instability is closely related to the underlying structure of the random measurement matrix $\mathbf{A}$. Understanding the dynamics of AMP for various classes of random measurement ensembles has been an outstanding open problem.

A rigorous proof of state evolution (SE) was first established by Bayati and Montanari [27]. As N tends to infinity while $\rho = \frac{n}{N}$ is kept constant, Bayati and Montanari [27] asymptotically characterized the evolution of the mean squared error (state) of AMP for the $\mathbf{A}$ with independent and identically distributed (i.i.d.) Gaussian entries. Rush *et. al* [28] showed a concentration bound of the SE in the finite n and N regime; the probability of deviation decays exponentially with n.

The validity of SE in [27] has inspired extensive research efforts on extending it to different measurement ensembles such as sub-Gaussian $\mathbf{A}$ by Bayati *et. al* [29] and Chen *et. al* [30], right-orthogonally-invariant $\mathbf{A}$ by Rangan *et. al* [31], rotationally-invariant $\mathbf{A}$ by Fan [32], unitarily-invariant $\mathbf{A}$ by Takeuchi [33], and semi-random $\mathbf{A}$ by Dudeja *et. al* [34]. A belief is that the SE for AMP might hold for an even boarder class of matrices. The key to analyzing the SE for AMP is the conditioning technique [27], [35]. This means that the current instance of the AMP algorithm is modeled as a linear combination of the previous instances that are Gaussian, plus a deviation term (non-Gaussian). A key step to the stability is leveraging the polynomial smoothness of the pseudo-Lipschitz function[2] and establishing that the contribution from the deviation term decays as n and N grow.

Recently, the controlled function has been introduced to analyze the nonlinear behavior of neural networks in machine learning [36], [37]. Unlike the pseudo-Lipschitz function, the controlled function incorporates exponential growth into its model. Hence, the pseudo-Lipschitz function can be viewed as a special case of the controlled function.

This work is motivated by the experimental observations [5], [25] that there is ample room for extending the establishment of SE for AMP to different classes of measurement ensembles

This work was supported in part by the National Science Foundation (NSF) under Grant CNS1955561, CNS2212565, CNS2225577, and the Office of Naval Research (ONR) under Grant N00014-21-1-2472.

[1] A bold lower case letter $\mathbf{a}$ is a column vector and a bold upper case letter $\mathbf{A}$ is a matrix. $\|\mathbf{a}\|_p$, $\mathbf{A}^*$, $\mathbf{A}^{-1}$, and A_{ij} denote the p-norm of $\mathbf{a}$, transpose of $\mathbf{A}$, inverse of $\mathbf{A}$, and ith row and jth column entry of $\mathbf{A}$, respectively. $\mathbf{A}(N)$ and $\mathbf{a}(N)$, respectively, are the matrix and vector indexed by N. $\mathcal{N}(\nu, \sigma^2)$ denotes the Gaussian distribution with mean ν and variance σ^2. The $\overset{d}{\Longrightarrow}$ and $\overset{a.s.}{=}$ denote the convergence in distribution and equivalence in an almost sure sense, respectively. $\langle \mathbf{u}, \mathbf{v}\rangle = \frac{1}{n}\sum_{i=1}^{n} u_i v_i$ defines the normalized inner product of $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{n\times 1}$. $\mathbb{E}_Z[\cdot]$ denotes the expectation with respect to the random variable Z. $\mathbf{0}_n$ denotes the $n \times 1$ all-zero vector.

[2] The polynomial smoothness of the pseudo-Lipschitz function is defined in Appendix A

and functions. The major contributions of this work are summarized below.

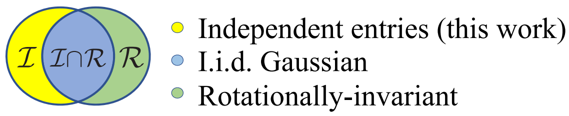

Fig. 1. Two categories of random measurement ensembles

- We extend the SE analysis for AMP to the measurement ensembles in $\mathcal{I}$ in Fig. 1, where $\mathcal{I}$ is the set of matrices with independent but not necessarily identically distributed entries. Note that the SE analysis in the prior works [27], [28], [31]–[33] focused on the set $\mathcal{R}$ in Fig. 1, where $\mathcal{R}$ is the set of rotationally-invariant matrices, including the i.i.d. Gaussian ensembles. The set $\mathcal{I}$ distinguishes it from the set $\mathcal{R}$ because it is far more incoherent than $\mathcal{R}$. This constitutes a major challenge in establishing the SE of the assumed ensembles. In our work, the lack of structure of the $\mathbf{A} \in \mathcal{I}$ is addressed by leveraging and extending the Lindeberg-Feller theorem.
- In contrast to the prior works [27]–[33] that relied on the pseudo-Lipschitz smoothness to establish SE, we generalize the SE analysis to the controlled function. The lack of smoothness of the latter is addressed by exploiting the measurability against Gaussian measure in conjunction with the conditioning technique, based on empirical statistics.

A part of our results is a theoretical justification for the conjecture made in [27] about the validity of the SE for AMP with i.i.d. non-Gaussian measurement ensembles.

II. Premilinaries

First, the concept of empirical statistics is developed. Next, the frequently used statistical lemmas are presented.

A. Empirical Statistics

The empirical law of a random vector constructed here is exploited to propose a conditioning technique in the following sections. We start by defining probability distributions.

Suppose $\mathcal{P}(\mathbb{R})$ is the collection of all probability distributions on $\mathbb{R}$ with sample space Ω. A random variable $X : \Omega \to \mathbb{R}$ has the distribution $\mu \in \mathcal{P}(\mathbb{R})$ denoted by $X \sim \mu$, if $Pr(X \in \mathcal{S}) = \mu(\mathcal{S})$, for a set $\mathcal{S} \subseteq \Omega$. For a function $f : \mathbb{R} \to \mathbb{R}$, we denote, if it exists, $\mu f = \int_{\mathbb{R}} f(x)\mu(dx)$. The first moment and second moment of the distribution μ are then given by taking $f(X) = X$ and $f(X) = X^2$, respectively, denoted as $\mathbb{E}[X] = \langle\mu\rangle$ and $\mathbb{E}[X^2] = \langle\mu^2\rangle$. The variance of the distribution μ is denoted by $\langle\mu\rangle_2$, which equals to the variance of X. Given these definitions, we define empirical distribution of deterministic vectors as follows.

1) Empirical Distribution of a Deterministic Vector: For a deterministic vector $\mathbf{v} \in \mathbb{R}^{n\times 1}$, the empirical sample mean, sample second moment, and sample variance are defined as $\langle\mathbf{v}\rangle = \frac{1}{n}\sum_{i=1}^{n} v_i, \langle\mathbf{v}^2\rangle = \frac{1}{n}\sum_{i=1}^{n} v_i^2$, and $\langle\mathbf{v}\rangle_2 = \frac{1}{n}\sum_{i=1}^{n}(v_i - \langle\mathbf{v}\rangle)^2$, respectively. We let $\delta_{\{a\}} \in \mathcal{P}(\mathbb{R})$ be the Dirac distribution with mass on $\{a\}$ and $\delta_{\{a\}} f \triangleq f(a)$ for all function $f : \mathbb{R} \to \mathbb{R}$. The empirical distribution of $\mathbf{v}$ is then defined by $\widehat{\mathbf{v}} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\{v_i\}}$. For the empirical distribution $\widehat{\mathbf{v}}$, the first moment $\langle\widehat{\mathbf{v}}\rangle = \widehat{\mathbf{v}}x = \frac{1}{n}\sum_{i=1}^{n}\delta_{\{v_i\}}x = \frac{1}{n}\sum_{i=1}^{n} v_i = \langle\mathbf{v}\rangle$, the second moment $\langle\widehat{\mathbf{v}}^2\rangle = \widehat{\mathbf{v}}x^2 = \frac{1}{n}\sum_{i=1}^{n} v_i^2 = \langle\mathbf{v}^2\rangle$, and the variance $\langle\widehat{\mathbf{v}}\rangle_2 = \frac{1}{n}\sum_{i=1}^{n}(v_i - \langle\widehat{\mathbf{v}}\rangle)^2 = \langle\mathbf{v}\rangle_2$. We are now ready to present the empirical law of a random vector.

2) Empirical Law of a Random Vector: For a random vector $\mathbf{v} : \Omega \to \mathbb{R}^{n\times 1}$, $\mathbf{v}(\omega) \in \mathbb{R}^{n\times 1}$ is a random sample of the empirical distribution $\widehat{\mathbf{v}(\omega)} \in \mathcal{P}(\mathbb{R})$. The empirical law of the random vector $\mathbf{v}$ is then defined by $\widehat{\mathbf{v}} f = \mathbb{E}[\widehat{\mathbf{v}(\omega)} f]$, for a measurable function f. Suppose that $v_i \sim \mu_i \in \mathcal{P}(\mathbb{R})$, $\forall i$. Then, we have the empirical law $\widehat{\mathbf{v}} = \frac{1}{n}\sum_{i=1}^{n}\mu_i$. Moreover, the first moment $\langle\widehat{\mathbf{v}}\rangle = \frac{1}{n}\sum_{i=1}^{n}\langle\mu_i\rangle$ and the second moment $\langle\widehat{\mathbf{v}}^2\rangle = \frac{1}{n}\sum_{i=1}^{n}\langle\mu_i^2\rangle$.

B. Frequently Used Statistics Results

***Lemma* 1.** *(Lindeberg-Feller [38]) For $n \geq 1$, we let $\{X_{n,m} : 1 \leq m \leq n\}$ be an independent triangular array with $\mathbb{E}[X_{n,m}] = 0$, $\forall m$. Suppose i) $\lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^2] = \sigma_x^2 > 0$; ii) (Lindeberg condition) for all $\epsilon > 0$, $\lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}\left[X_{n,m}^2; |X_{n,m}| > \epsilon\right] = 0$. Then $\sum_{m=1}^{n} X_{n,m} \overset{d}{\Rightarrow} \mathcal{N}(0, \sigma_x^2)$ as $n \to \infty$.*

We propose a sufficient condition for the *Lindeberg condition*.

***Proposition* 1.** *If there exists $\alpha > 0$ such that $\sup_m \mathbb{E}[X_{n,m}^{2+2\alpha}] = o(n^{-1})$, the Lindeberg condition in Lemma 1 holds.*

Proof. See Appendix B. □

Incorporating Proposition 1 into Lemma 1 leads to a variant of Lindeberg-Feller that we use for our analysis.

***Proposition* 2.** *For $n \geq 1$, we let $\{X_{n,m} : 1 \leq m \leq n\}$ be independent random triangular array with $\mathbb{E}[X_{n,m}] = 0$, $\forall m$. Suppose i) $\lim_{n\to\infty}\sum_{m=1}^{n}\mathbb{E}[X_{n,m}^2] = \sigma_x^2 < \infty$; ii) $\sup_m \mathbb{E}\left[X_{n,m}^{2+2\alpha}\right] = o(n^{-1})$, for a constant $\alpha > 0$. Then $\sum_{m=1}^{n} X_{n,m} \overset{d}{\Rightarrow} \mathcal{N}(0, \sigma_x^2)$ as $n \to \infty$.*

Proposition 2 will be further exploited to propose our key propositions in Section III-B.

In what follows, $\mathbf{I}$ denotes an identity matrix with appropriate dimensions. For a matrix $\mathbf{H} \in \mathbb{R}^{n\times t}$ $(n \geq t)$, $\mathbf{P_H} = \mathbf{H}(\mathbf{H}^*\mathbf{H})^{-1}\mathbf{H}^*$ is the orthogonal projection onto the column subspace of $\mathbf{H}$ and $\mathbf{P}_\mathbf{H}^\perp = \mathbf{I} - \mathbf{P_H}$. Two random variables X and Y are said to be equal in distribution, denoted by $X \overset{d}{=} Y$, if $\mathbb{E}[\phi(X)Z] = \mathbb{E}[\phi(Y)Z]$, for any integrable function ϕ and random variable Z. The following proposition characterizes our conditional distribution result.

***Proposition* 3.** *Given $\mathbf{A} \in \mathbb{R}^{n\times N}$, $\widetilde{\mathbf{A}} \in \mathbb{R}^{n\times N}$ such that $\mathbf{A} \overset{d}{=} \widetilde{\mathbf{A}}$, and the set $\mathcal{P}_{\mathbf{X},\mathbf{Y}} = \{\mathbf{A}|\mathbf{A}^*\mathbf{M} = \mathbf{X}, \mathbf{A}\mathbf{Q} = \mathbf{Y}\}$, where $\mathbf{M} \in \mathbb{R}^{n\times t}$, $\mathbf{X} \in \mathbb{R}^{N\times t}$, $\mathbf{Q} \in \mathbb{R}^{N\times t}$, and $\mathbf{Y} \in \mathbb{R}^{n\times t}$, the following holds*

$$\mathbf{A}|_{\mathcal{P}_{\mathbf{X},\mathbf{Y}}} \overset{d}{=} \mathbf{P}_\mathbf{M}^\perp \widetilde{\mathbf{A}} \mathbf{P}_\mathbf{Q}^\perp + \mathbf{B}, \tag{2}$$

where $\mathbf{A}|_{\mathcal{P}_{\mathbf{X},\mathbf{Y}}}$ is the orthogonal projection of $\mathbf{A}$ onto $\mathcal{P}_{\mathbf{X},\mathbf{Y}}$ and $\mathbf{B} = \mathbf{Y}(\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^* + \mathbf{M}(\mathbf{M}^*\mathbf{M})^{-1}\mathbf{X}^* - \mathbf{M}(\mathbf{M}^*\mathbf{M})^{-1}\mathbf{X}^*\mathbf{Q}(\mathbf{Q}^*\mathbf{Q})^{-1}\mathbf{Q}^*$.

Proof. See Appendix C. □

Remark 1. An equivalence to (2) was proven in [27, Lemma 10, Eqn. (3.29)]. The underlying assumption of [27] was that A_{ij} are i.i.d. following $\mathcal{N}(0, \frac{1}{n})$, in which the conditional distribution of $\mathbf{A}$ is computed based on its unitary-invariance property (i.e., $\mathbf{A} \in \mathcal{I} \cap \mathcal{R}$). On the other hand, Proposition 3 does not rely on the unitary-invariance property. The result in Proposition 3 will be exploited in Section III–IV to characterize the conditional distribution of $\mathbf{A} \in \mathcal{I}$ in our SE analysis.

III. Main Results

In this section, we present our main results by showing the SE for AMP when $\mathbf{A} \in \mathcal{I}$ and for the controlled function.

Given the linear observation in (1), the general AMP algorithm [27] recurs the vectors $\mathbf{h}^{t+1} \in \mathbb{R}^{N\times 1}$, $\mathbf{q}^t \in \mathbb{R}^{N\times 1}$, $\mathbf{b}^t \in \mathbb{R}^{n\times 1}$, and $\mathbf{m}^t \in \mathbb{R}^{n\times 1}$, sequentially, for $t \geq 0$, through

$$\mathbf{q}^t = f_t(\mathbf{h}^t, \mathbf{x}_0), \qquad \mathbf{b}^t = \mathbf{A}\mathbf{q}^t - \lambda_t \mathbf{m}^{t-1}, \tag{3a}$$

$$\mathbf{m}^t = g_t(\mathbf{b}^t, \mathbf{w}), \qquad \mathbf{h}^{t+1} = \mathbf{A}^*\mathbf{m}^t - \xi_t \mathbf{q}^t, \tag{3b}$$

where $\mathbf{q}^0 \in \mathbb{R}^{N\times 1}$ is an initial condition, $\mathbf{h}^0 = \mathbf{0}_N$, $\mathbf{m}^{-1} = \mathbf{0}_n$, and $f_t(\cdot,\cdot)$ and $g_t(\cdot,\cdot)$ are controlled functions. The f_t and g_t are applied entry-wise when their arguments are vectors, e.g., $g_t(\mathbf{b}^t, \mathbf{w}) = (g_t(b_1^t, w_1), \ldots, g_t(b_n^t, w_n)) \in \mathbb{R}^{n\times 1}$. Suppose that f_t and g_t are differentiable with respect to their first argument. Then, the scalar λ_t and ξ_t in (3) are, respectively, $\lambda_t = \frac{1}{\rho}\langle f_t'(\mathbf{h}^t, \mathbf{x}_0)\rangle$ and $\xi_t = \langle g_t'(\mathbf{b}^t, \mathbf{w})\rangle$, where $f_t'(\mathbf{h}^t, \mathbf{x}_0) = \left(\frac{\partial f_t}{\partial h_1^t}, \ldots, \frac{\partial f_t}{\partial h_N^t}\right) \in \mathbb{R}^{N\times 1}$, $g_t'(\mathbf{b}^t, \mathbf{w}) = \left(\frac{\partial g_t}{\partial b_1^t}, \ldots, \frac{\partial g_t}{\partial b_n^t}\right) \in \mathbb{R}^{n\times 1}$, $\frac{\partial f}{\partial x}$ denotes the partial derivative of f with respect to x, and $\rho = \frac{n}{N}$ is kept being constant as n and N tend to infinity.

A. Definitions and Assumptions

1) Controlled Function: A function ϕ: $\mathbb{R}^{t\times 1} \to \mathbb{R}$ is called a controlled function [36], [37] if for $\mathbf{x} \in \mathbb{R}^{t\times 1}$,

$$|\phi(\mathbf{x})| \leq c_1 \exp(c_2 \|\mathbf{x}\|_2^\lambda), \tag{4}$$

where $c_1 > 0, c_2 > 0$, and $1 \leq \lambda < 2$ are constant. The controlled function is a general model of a nonlinear function without polynomial smoothness constraint. Although the controlled function may increase exponentially, it is integrable in $\mathcal{L}^1$ and $\mathcal{L}^2$ spaces against Gaussian measure [36], [37]. Herein, the space $\mathcal{L}^p(\mathcal{X}, \mathcal{F}, \mu)$ consists of all measurable functions $\{f\}$ on $\mathcal{X}$ such that $\int_{\mathcal{X}} |f(x)|^p d\mu(x) < \infty$, where $(\mathcal{X}, \mathcal{F}, \mu)$ is a measure space and $p \in \{1, 2\}$.

Using the inequalities $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_\lambda \leq t^{\frac{1}{\lambda}-\frac{1}{2}}\|\mathbf{x}\|_2$ for $\mathbf{x} \in \mathbb{R}^{t\times 1}$ and $1 \leq \lambda < 2$, we get from (4)

$$|\phi(\mathbf{x})| \leq c_1 \exp(c_2 \|\mathbf{x}\|_\lambda^\lambda) \leq c_1 \exp(c_2 t^{\frac{1}{\lambda}-\frac{1}{2}} \|\mathbf{x}\|_2^\lambda), \tag{5}$$

revealing that $c_1 \exp(c_2\|\mathbf{x}\|_\lambda^\lambda)$ is also a controlled function.

2) Measurement Matrix: The matrix $\mathbf{A}$ in (1) consists of independent entries $A_{ij} \sim \frac{1}{\sqrt{n}}\mu_{ij}$, with $\langle \mu_{ij}\rangle = 0$ and $\langle \mu_{ij}\rangle_2 = 1$, which are not necessarily identically distributed, $\forall i, j$, i.e., $\mathbf{A} \in \mathcal{I}$.

3) Signal $\mathbf{x}_0$ and Noise $\mathbf{w}$: The entries of signal vector $\mathbf{x}_0$ and noise $\mathbf{w}$ in (1) are i.i.d. according to the distribution μ_{X_0} and μ_W, respectively, where $\langle \mu_{X_0}\rangle = 0, \langle \mu_{X_0}\rangle_2 = \sigma_{X_0}^2$ and $\langle \mu_W\rangle = 0, \langle \mu_W\rangle_2 = \sigma_W^2$. The empirical distributions $\widehat{\mathbf{x}}_0 \overset{\text{d}}{\Rightarrow} \mu_{X_0}$ as $N \to \infty$, $\widehat{\mathbf{w}} \overset{\text{d}}{\Rightarrow} \mu_W$ as $n \to \infty$, and $\mu_{X_0} f < \infty$ and $\mu_W f < \infty$ for any controlled function f.

B. Convergence Lemmas

To find an asymptotic expression SE for the measurement ensemble $\mathcal{I}$, we establish the following lemmas.

***Proposition* 4.** *Let $A(n)$ be a random variable with $A(n) \sim \frac{1}{\sqrt{n}}\mu$, where $\langle \mu\rangle = 0$ and $\langle \mu\rangle_2 = 1$. Then $\mathbb{E}[A^{2+2\alpha}(n)] = o(n^{-2})$ for $\alpha > 1$.*

Proof. See Appendix D. □

We now present a the key proposition in our SE analysis.

***Proposition* 5.** *We let $\mathbf{A}(N) \in \mathbb{R}^{n\times N}$ be a random matrix defined as in Section III-A2. Suppose that $\mathbf{v}(N) \in \mathbb{R}^{N\times 1}$ is a deterministic vector satisfying $\lim_{N\to\infty}\langle \mathbf{v}^2(N)\rangle = s_0^2 < \infty$ and $\limsup_{N\to\infty} \frac{1}{N}\|\mathbf{v}(N)\|_{2+2\alpha}^{2+2\alpha} < \infty$ for a constant $\alpha > 1$. Then, the following empirical law converges, $\widehat{\mathbf{A}(N)\mathbf{v}(N)} \overset{d}{\Rightarrow} \mathcal{N}\left(0, s_0^2/\rho\right)$ as $N \to \infty$ while $\rho = \frac{n}{N}$ is kept being constant.*

Proof. See Appendix E. □

The next proposition is a direct consequence of Proposition 5.

***Proposition* 6.** *Suppose $\mathbf{u} \in \mathbb{R}^{N\times 1}$ and $\mathbf{v} \in \mathbb{R}^{n\times 1}$ with $\|\mathbf{u}\|_2 = \|\mathbf{v}\|_2 = 1$. Then $\mathbf{v}^*\mathbf{A}\mathbf{u} \overset{d}{\Rightarrow} \frac{Z}{\sqrt{n}}$ as $N \to \infty$, where the matrix $\mathbf{A}$ is defined as in Proposition 5 and $Z \sim \mathcal{N}(0, 1)$.*

Remark 2. It is worth noting that Proposition 6 is a generalization of Lemma 2 in [27]. In [27], the standard properties of Gaussian matrices are used to show Proposition 6. However, this can not be directly extended to the measurement matrices in $\mathcal{I}$ because its entries are not i.i.d.

C. State Evolution of General AMP Algorithm

Suppose that

$$\lim_{N\to\infty}\langle \mathbf{q}^0, \mathbf{q}^0\rangle/\rho = \sigma_0^2 < \infty, \ \limsup_{N\to\infty}\|\mathbf{q}^0\|_{2+2\alpha}^{2+2\alpha}/N < \infty, \tag{6}$$

for a constant $\alpha > 1$. Then, the SE for the general AMP in (3) is described [27] by

$$\tau_t^2 = \mathbb{E}\left[g_t^2(\sigma_t Z, W)\right], \ \ \sigma_t^2 = \mathbb{E}\left[f_t^2(\tau_{t-1}Z, X_0)\right]/\rho, \tag{7}$$

where $X_0 \sim \mu_{X_0}, W \sim \mu_W$, and $Z \sim \mathcal{N}(0,1)$ is independent of W and X_0. The SE in (7) has the same expression as the one in [27], [28] except that the matrix $\mathbf{A}$ in our setting follows the new assumptions in Section III-A2 and the controlled function.

Define a collection of vectors as a set $\mathcal{F}_{t_1,t_2} = \{\mathbf{b}^0, \ldots, \mathbf{b}^{t_1-1}, \mathbf{m}^0, \ldots, \mathbf{m}^{t_1-1}, \mathbf{h}^1, \ldots, \mathbf{h}^{t_2}, \mathbf{q}^0, \ldots, \mathbf{q}^{t_2}, \mathbf{x}_0, \mathbf{w}\}$. The recursion in (3) can then be represented by incorporating matrices as $\mathbf{Y}_t = \mathbf{A}\mathbf{Q}_t$ with $\mathbf{Y}_t = [\mathbf{b}^0|\mathbf{b}^1 + \lambda_1\mathbf{m}^0|\ldots|\mathbf{b}^{t-1} + \lambda_{t-1}\mathbf{m}^{t-2}] \in \mathbb{R}^{n\times t}$ and $\mathbf{Q}_t = [\mathbf{q}^0|\mathbf{q}^1|\ldots|\mathbf{q}^{t-1}] \in \mathbb{R}^{N\times t}$, and $\mathbf{X}_t = \mathbf{A}^*\mathbf{M}_t$ with $\mathbf{X}_t = [\mathbf{h}^1 + \xi_0\mathbf{q}^0|\ldots|\mathbf{h}^t + \xi_{t-1}\mathbf{q}^{t-1}] \in \mathbb{R}^{N\times t}$ and $\mathbf{M}_t = [\mathbf{m}^0|\ldots|\mathbf{m}^{t-1}] \in \mathbb{R}^{n\times t}$. The $\mathbf{m}_{||}^t$ and $\mathbf{q}_{||}^t$ denote, respectively, the projection of $\mathbf{m}^t$ and $\mathbf{q}^t$ onto the column spaces of $\mathbf{M}_t$ and $\mathbf{Q}_t$,

$$\mathbf{m}_{||}^t = \sum_{i=0}^{t-1}\zeta_i\mathbf{m}^i, \ \mathbf{q}_{||}^t = \sum_{i=0}^{t-1}\beta_i\mathbf{q}^i, \tag{8}$$

where ζ_i and β_i quantify, respectively, the contributions of $\mathbf{m}^i$ and $\mathbf{q}^i$ to the projected images of $\mathbf{m}^t$ and $\mathbf{q}^t$. Then their null space projections $\mathbf{q}_\perp^t$ and $\mathbf{m}_\perp^t$ are defined as

$$\mathbf{q}_\perp^t = \mathbf{q}^t - \mathbf{q}_{||}^t, \ \mathbf{m}_\perp^t = \mathbf{m}^t - \mathbf{m}_{||}^t. \tag{9}$$

The following theorem is the main result of this work.

***Theorem* 1.** *Suppose* $\{\tau_t\}_{t\geq 0}$ *and* $\{\sigma_t\}_{t\geq 0}$ *defined in* (7). *Given the AMP recursion in* (3), *the following propositions hold for* $t \geq 0$.
a)

$$\mathbf{b}^t|_{\mathcal{F}_{t,t}} \overset{d}{\Rightarrow} \sum_{j=0}^{t-1} \beta_j \mathbf{b}^j + \widetilde{\mathbf{A}}\mathbf{q}_\perp^t, \text{ as } n \to \infty, \tag{10}$$

$$\mathbf{h}^{t+1}|_{\mathcal{F}_{t+1,t}} \overset{d}{\Rightarrow} \sum_{j=0}^{t-1} \zeta_j \mathbf{h}^{j+1} + \widetilde{\mathbf{A}}^* \mathbf{m}_\perp^t, \text{ as } N \to \infty, \tag{11}$$

where $\widetilde{\mathbf{A}} \overset{d}{=} \mathbf{A}$.
b) For a constant $\alpha > 1$,

$$\limsup_{N\to\infty} \frac{1}{N} \|\mathbf{q}_\perp^t\|_{2+2\alpha}^{2+2\alpha} < \infty, \tag{12}$$

$$\limsup_{n\to\infty} \frac{1}{n} \|\mathbf{m}_\perp^t\|_{2+2\alpha}^{2+2\alpha} < \infty. \tag{13}$$

c) For all $0 \leq t_1, t_2 \leq t$,

$$\lim_{n\to\infty} \langle \mathbf{b}^{t_1}, \mathbf{b}^{t_2} \rangle \overset{a.s.}{=} \frac{1}{\rho} \lim_{N\to\infty} \langle \mathbf{q}^{t_1}, \mathbf{q}^{t_2} \rangle < \infty. \tag{14}$$

$$\lim_{N\to\infty} \langle \mathbf{h}^{t_1+1}, \mathbf{h}^{t_2+1} \rangle \overset{a.s.}{=} \lim_{n\to\infty} \langle \mathbf{m}^{t_1}, \mathbf{m}^{t_2} \rangle < \infty. \tag{15}$$

d) For all controlled functions ϕ_b *and* ϕ_h: $\mathbb{R}^{t+2} \to \mathbb{R}$,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \phi_b(\mathbf{u}_i^t) \overset{a.s.}{=} \mathbb{E}\left[\phi_b(\sigma_0 \widetilde{Z}_0, .., \sigma_t \widetilde{Z}_t, W)\right], \tag{16}$$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \phi_h(\mathbf{v}_i^t) \overset{a.s.}{=} \mathbb{E}\left[\phi_h(\tau_0 Z_0, .., \tau_t Z_t, X_0)\right], \tag{17}$$

where $\mathbf{u}_i^t = (b_i^0 \ldots, b_i^t, w_i)$ *and* $\mathbf{v}_i^t = (h_i^1, .., h_i^{t+1}, x_{0i})$. *The* $(Z_0, \ldots, Z_t)$ *and* $(\widetilde{Z}_0, \ldots, \widetilde{Z}_t)$ *are Gaussian vectors, where* Z_j *and* $\widetilde{Z}_j$ *are i.i.d. following* $\mathcal{N}(0,1)$, *for* $j = 1, 2, \ldots, t$, *and independent of* X_0 *and* W.

Remark 3. The results in (16) and (17) are similar to Lemma 1b in [27] except that ϕ_b and ϕ_h in (16) and (17) are controlled functions. The results in (10) and (11) represent the convergence in the distribution of $\mathbf{b}^t|_{\mathcal{F}_{t,t}}$ and $\mathbf{h}^{t+1}|_{\mathcal{F}_{t+1,t}}$, which are obtained by applying the conditioning technique and Proposition 3 in our analysis. Compared to Lemma 1a in [27], the expressions in (10) and (11) do not include the basis-aligned deviation terms. More specifically, the $\mathbf{b}^t|_{\mathcal{F}_{t,t}}$ and $\mathbf{h}^{t+1}|_{\mathcal{F}_{t+1,t}}$ in Lemma 1a of [27] are

$$\mathbf{b}^t|_{\mathcal{F}_{t,t}} \overset{d}{=} \sum_{j=0}^{t-1} \beta_j \mathbf{b}^j + \widetilde{\mathbf{A}}\mathbf{q}_\perp^t + \widetilde{\mathbf{M}}_t \overrightarrow{\mathbf{o}}_t(1), \tag{18}$$

$$\mathbf{h}^{t+1}|_{\mathcal{F}_{t+1,t}} \overset{d}{=} \sum_{j=0}^{t-1} \zeta_j \mathbf{h}^{j+1} + \widetilde{\mathbf{A}}^* \mathbf{m}_\perp^t + \widetilde{\mathbf{Q}}_{t+1} \overrightarrow{\mathbf{o}}_t(1), \tag{19}$$

where $\widetilde{\mathbf{M}}_t \in \mathbb{R}^{n\times t}$ and $\widetilde{\mathbf{Q}}_{t+1} \in \mathbb{R}^{N\times(t+1)}$ are the orthogonal bases of the column subspaces of $\mathbf{M}_t$ and $\mathbf{Q}_{t+1}$, respectively. The $\overrightarrow{\mathbf{o}}_t(1)$ denotes a vector in $\mathbb{R}^{t\times 1}$ such that all entries converge to 0 almost surely as $N \to \infty$. The common procedure for characterizing the SE in [27], [28], [31]–[33] is to cancel the deviation terms $\widetilde{\mathbf{M}}_t \overrightarrow{\mathbf{o}}_t(1)$ and $\widetilde{\mathbf{Q}}_{t+1} \overrightarrow{\mathbf{o}}_t(1)$ in (18) and (19), respectively, by assuming the smoothness of the pseudo-Lipschitz functions. The challenge in our proof of Theorem 1 is that there is no such smoothness for ϕ_b and ϕ_h in (16) and (17), respectively, because they are controlled functions. It is shown in Section IV that these deviation terms vanish in our derivation, which leverages the concept of empirical statistics developed in Section II-A. We note that the SE in (7) is a special case of Theorem 1; its detailed derivation is relegated to Appendix G.

Remark 4. The results in Theorem 1 directly verify the conjecture about the validity of SE for i.i.d. non-Gaussian measurement ensemble in [27] because the set of such $\mathbf{A}$ is a subset of the $\mathbf{A}$ defined in Section III-A2.

IV. Proof of Theorem 1

The proof of Theorem 1 is inspired by [27]. Since a part of the proof is based on a similar technique in [27], we refer to [27] for those standard arguments, while we present the features that are unique and refined in this work. The proof is based on mathematical induction in t and the application of the conditioning technique on four sets $\mathcal{F}_{0,0}, \mathcal{F}_{1,0}, \mathcal{F}_{t,t}$, and $\mathcal{F}_{t+1,t}$ sequentially.

A. Step 1: Conditioning on $\mathcal{F}_{0,0} = \{\mathbf{q}^0, \mathbf{x}_0, \mathbf{w}\}$.

a) The convergence in (10) holds because $\mathbf{b}^0 = \mathbf{A}\mathbf{q}^0 \overset{d}{=} \widetilde{\mathbf{A}}\mathbf{q}_\perp^0$, where $\mathbf{q}_\perp^0 = \mathbf{q}^0$ because $\mathbf{Q}_0$ is an empty matrix.

b) The bound in (12) is due to (6) and $\mathbf{q}_\perp^0 = \mathbf{q}^0$.

c) Given (6), applying Proposition 5 to $\widetilde{\mathbf{A}}$ and $\mathbf{q}^0$ leads to

$$\widehat{\widetilde{\mathbf{A}}\mathbf{q}^0} \overset{\mathrm{d}}{\Rightarrow} \mathcal{N}(0, \sigma_0^2). \tag{20}$$

Hence, by Step 1a), $\lim_{n\to\infty} \langle \mathbf{b}^0, \mathbf{b}^0 \rangle \overset{a.s.}{=} \sigma_0^2$, implying $\lim_{n\to\infty} \langle \mathbf{b}^0, \mathbf{b}^0 \rangle \overset{a.s.}{=} \lim_{N\to\infty} \frac{\langle \mathbf{q}^0, \mathbf{q}^0 \rangle}{\rho} < \infty$, which completes the proof of (14).

d) We let $\widetilde{\mathbf{u}}_i^0 = (\sigma_0 \widetilde{Z}_0, w_i)$, $\forall i$, where $\widetilde{Z}_0 \sim \mathcal{N}(0,1)$. From (20), the convergence $\mathbf{u}_i^0 \overset{\mathrm{d}}{\Rightarrow} \widetilde{\mathbf{u}}_i^0$ holds. To prove (16), we first claim that $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \phi_b(\widetilde{\mathbf{u}}_i^0) \overset{a.s.}{=} \mathbb{E}[\phi_b(\sigma_0 \widetilde{Z}_0, W)]$. By the triangular inequality, we have $\left| \frac{1}{n} \sum_{i=1}^{n} \phi_b(\widetilde{\mathbf{u}}_i^0) - \mathbb{E}[\phi_b(\sigma_0 \widetilde{Z}_0, W)] \right| \leq X_1^0 + X_2^0$, where $X_1^0 = \left| \frac{1}{n} \sum_{i=1}^{n} \left(\phi_b(\widetilde{\mathbf{u}}_i^0) - \mathbb{E}_{\widetilde{Z}_0}[\phi_b(\widetilde{\mathbf{u}}_i^0)] \right) \right|$ and $X_2^0 = \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{\widetilde{Z}_0}[\phi_b(\widetilde{\mathbf{u}}_i^0)] - \mathbb{E}[\phi_b(\sigma_0 \widetilde{Z}_0, W)] \right|$. The goal is to prove that $\lim_{n\to\infty} X_1^0 \overset{a.s.}{=} 0$ and $\lim_{n\to\infty} X_2^0 \overset{a.s.}{=} 0$, respectively.

Since ϕ_b is a controlled function, we get by (5) $|\phi_b(\widetilde{\mathbf{u}}_i^0)| \leq c_1^0 \exp(c_2^0(|\sigma_0 \widetilde{Z}_0|^\lambda + |w_i|^\lambda))$, where $c_1^0 > 0$, $c_2^0 > 0$, and $1 \leq \lambda < 2$ are constant. Hence, $\mathbb{E}_{\widetilde{Z}_0}[|\phi_b(\widetilde{\mathbf{u}}_i^0)|^{2+\kappa}] \leq c_3^0 \exp(c_4^0 |w_i|^\lambda) \mathbb{E}_{\widetilde{Z}_0}[\exp(c_4^0 |\sigma_0 \widetilde{Z}_0|^\lambda)]$, where $0 < \kappa < 1$, $c_3^0 = (c_1^0)^{2+\kappa}$, and $c_4^0 = c_2^0(2+\kappa)$. Thus,

$$\mathbb{E}_{\widetilde{Z}_0}[|\phi_b(\widetilde{\mathbf{u}}_i^0)|^{2+\kappa}] \leq c_5^0 \exp(c_4^0 |w_i|^\lambda), \tag{21}$$

where $c_5^0 = c_3^0 \mathbb{E}_{\widetilde{Z}_0}[\exp(c_4^0 |\sigma_0 \widetilde{Z}_0|^\lambda)]$ is constant. Define $X_{n,i}^0 = \phi_b(\widetilde{\mathbf{u}}_i^0) - \mathbb{E}_{\widetilde{Z}_0}[\phi_b(\widetilde{\mathbf{u}}_i^0)]$, $\forall i$. To prove the thesis $\lim_{n\to\infty} X_1^0 \overset{a.s.}{=} 0$, we show that $\{X_{n,i}^0\}_{i=1}^n$ satisfy Lemma 3 in Appendix F. Indeed, applying Holder's inequality in Lemma 4 in Appendix F to $X_{n,i}^0$ gives

$$\begin{aligned} \mathbb{E}_{\widetilde{Z}_0}[|X_{n,i}^0|^{2+\kappa}] &\leq 2^{1+\kappa} \left(\mathbb{E}_{\widetilde{Z}_0}[|\phi_b(\widetilde{\mathbf{u}}_i^0)|^{2+\kappa}] + |\mathbb{E}_{\widetilde{Z}_0}[\phi_b(\widetilde{\mathbf{u}}_i^0)]|^{2+\kappa} \right), \\ &\leq c_6^0 \exp(c_4^0 |w_i|^\lambda), \end{aligned} \tag{22a}$$

where $c_6^0 = 2^{2+\kappa} c_5^0$ and (22a) follows from Lemma 5 (Lyapunov's inequality) in Appendix F and (21). Note that $\frac{1}{n} \sum_{i=1}^{n} c_6^0 \exp(c_4^0 |w_i|^\lambda) \overset{a.s.}{=} \mathbb{E}[c_6^0 \exp(c_4^0 |W|^\lambda)]$. For n sufficiently large, using (22a) gives

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_{\widetilde{Z}_0}[|X_{n,i}^0|^{2+\kappa}] \leq \mathbb{E}[c_6^0\exp(c_4^0|W|^\lambda)] < cn^{\kappa/2}, \quad (23)$$

where $c>0$ is constant and (23) follows from the facts that $\mathbb{E}[c_6^0\exp(c_4^0|W|^\lambda)] = c_7^0 < \infty$ and there exists a constant $n_0>0$ such that $c_7^0 < cn^{\kappa/2}$ for $n>n_0$. Lemma 3 in Appendix F leads to $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_{n,i}^0 \overset{a.s.}{=} 0$, implying $\lim_{n\to\infty} X_1^0 \overset{a.s.}{=} 0$.

The rest of Step 1d) is showing the convergence $\lim_{n\to\infty} X_2^0 \overset{a.s.}{=} 0$. Define $\widetilde{\phi}_b(w_i) = \mathbb{E}_{\widetilde{Z}_0}[\phi_b(\widetilde{\mathbf{u}}_i^0)]$, $\forall i$. Then $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\widetilde{\phi}_b(w_i) \overset{a.s.}{=} \mathbb{E}[\widetilde{\phi}_b(W)]$ because $\widehat{\mathbf{w}} \overset{\mathrm{d}}{\Rightarrow} \mu_W$. Hence, $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\mathbb{E}_{\widetilde{Z}_0}[\phi_b(\widetilde{\mathbf{u}}_i^0)] \overset{a.s.}{=} \mathbb{E}_W[\mathbb{E}_{\widetilde{Z}_0}[\phi_b(\sigma_0\widetilde{Z}_0,W)]] \overset{a.s.}{=} \mathbb{E}[\phi_b(\sigma_0\widetilde{Z}_0,W)]$, implying $\lim_{n\to\infty}X_2^0 \overset{a.s.}{=} 0$, concludes (16).

B. Step 2: Conditioning on $\mathcal{F}_{1,0}=\{\mathbf{b}^0,\mathbf{m}^0,\mathbf{q}^0,\mathbf{x}_0,\mathbf{w}\}$.

a) From (3), $\mathbf{h}^1|_{\mathcal{F}_{1,0}} \overset{d}{=} \mathbf{A}^*|_{\mathcal{F}_{1,0}}\mathbf{m}^0 - \xi_0\mathbf{q}^0$. Conditioning on $\mathcal{F}_{1,0}$ is equivalent to conditioning on $\{\mathbf{A}|\mathbf{A}\mathbf{q}^0=\mathbf{b}^0\}$. Hence, applying Proposition 3 yields $\mathbf{A}|_{\mathcal{F}_{1,0}} \overset{d}{=} \widetilde{\mathbf{A}}\mathbf{P}_{\mathbf{q}^0}^\perp + \frac{\mathbf{b}^0\mathbf{q}^{0*}}{\|\mathbf{q}^0\|_2^2}$. Then $\mathbf{h}^1|_{\mathcal{F}_{1,0}} \overset{d}{=} \mathbf{P}_{\mathbf{q}^0}^\perp\widetilde{\mathbf{A}}^*\mathbf{m}^0 + \frac{\mathbf{b}^{0*}\mathbf{m}^0}{\|\mathbf{q}^0\|_2^2}\mathbf{q}^0 - \xi_0\mathbf{q}^0$, resulting in

$$\mathbf{h}^1|_{\mathcal{F}_{1,0}} \overset{a.s.}{=} \widetilde{\mathbf{A}}^*\mathbf{m}^0 - \mathbf{P}_{\mathbf{q}^0}\widetilde{\mathbf{A}}^*\mathbf{m}^0 + \left(\rho\frac{\langle\mathbf{b}^0,\mathbf{m}^0\rangle}{\langle\mathbf{q}^0,\mathbf{q}^0\rangle} - \xi_0\right)\mathbf{q}^0. \quad (24)$$

Substituting $\mathbf{m}^0 = g_0(\mathbf{b}^0,\mathbf{w})$ into $\langle\mathbf{b}^0,\mathbf{m}^0\rangle$ in (24) gives $\lim_{n\to\infty}\langle\mathbf{b}^0,\mathbf{m}^0\rangle \overset{a.s.}{=} \mathbb{E}[\sigma_0\widetilde{Z}_0 g_0(\sigma_0\widetilde{Z}_0,W)] \overset{a.s.}{=} \sigma_0^2\mathbb{E}[g_0'(\sigma_0\widetilde{Z},W)] \overset{a.s.}{=} \lim_{n\to\infty}\langle\mathbf{b}^0,\mathbf{b}^0\rangle\langle g_0'(\mathbf{b}^0,\mathbf{w})\rangle$, where the first equality is due to (16) applied to $\phi_b(\mathbf{u}_i^0) = b_i^0 g_0(b_i^0,w_i)$, the second equality is due to Stein's Lemma (Lemma 2) in Appendix F, the third equality is obtained by setting $\phi_b(\mathbf{u}_i^0) = g_0'(b_i^0,w_i)$ in (16). Note that $\xi_0 = \langle g_0'(\mathbf{b}^0,\mathbf{w})\rangle$. Using (14) gives $\lim_{n\to\infty}\langle\mathbf{b}^0,\mathbf{m}^0\rangle \overset{a.s.}{=} \lim_{N\to\infty}\frac{\langle\mathbf{q}^0,\mathbf{q}^0\rangle}{\rho}\xi_0$, which is plugged into (24) to result in

$$\mathbf{h}^1|_{\mathcal{F}_{1,0}} \overset{d}{=} \widetilde{\mathbf{A}}^*\mathbf{m}^0 - \mathbf{P}_{\mathbf{q}^0}\widetilde{\mathbf{A}}^*\mathbf{m}^0 + o(1)\mathbf{q}^0, \quad (25)$$

where $\lim_{n\to\infty}o(1) \overset{a.s.}{=} 0$. Using Proposition 7 in Appendix F, the second term on the right-hand side (r.h.s) of (25) converges to $\lim_{N\to\infty}\mathbf{P}_{\mathbf{q}^0}\widetilde{\mathbf{A}}^*\mathbf{m}^0 \overset{a.s.}{=} \mathbf{0}_N$. We claim that the last term converges to $\lim_{N\to\infty}o(1)\mathbf{q}^0 \overset{a.s.}{=} \mathbf{0}_N$. Indeed, this is true because (i) the empirical expectation of $o(1)\mathbf{q}^0$ satisfies $\lim_{N\to\infty}\langle\widehat{o(1)\mathbf{q}^0}\rangle \overset{a.s.}{=} 0$, which holds due to $\lim_{N\to\infty}|\langle\widehat{o(1)\mathbf{q}^0}\rangle| \leq \lim_{N\to\infty}|o(1)|\frac{1}{N}\sum_{i=1}^N|q_i^0| \overset{a.s.}{=} 0$ and (ii) the empirical variance of $o(1)\mathbf{q}^0$ converges to $\lim_{N\to\infty}\langle\widehat{o(1)\mathbf{q}^0}\rangle_2 = \lim_{N\to\infty}[o(1)]^2\langle\mathbf{q}^0,\mathbf{q}^0\rangle \overset{a.s.}{=} 0$. Thus, $\mathbf{h}^1|_{\mathcal{F}_{1,0}} \overset{\mathrm{d}}{\Rightarrow} \widetilde{\mathbf{A}}^*\mathbf{m}^0$, which concludes (11) when $t=0$ since $\mathbf{m}^0=\mathbf{m}_\perp^0$; note that $\mathbf{M}_0$ is an empty matrix.

b) The bound in (13) is equivalent to $\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^n|m_i^0|^{2+2\alpha}<\infty$. Incorporating $\phi_b(b_i^0,w_i) = |g_0(b_i^0,w_i)|^{2+2\alpha}$ for a constant $\alpha>1$ into (16) leads to $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n|m_i^0|^{2+2\alpha} \overset{a.s.}{=} \mathbb{E}[|g_0(\sigma_0\widetilde{Z}_0,W)|^{2+2\alpha}]<\infty$, which completes the proof.

c) By (16) at $t=0$, we get $\lim_{n\to\infty}\langle g_0(\mathbf{b}^0,\mathbf{w}),g_0(\mathbf{b}^0,\mathbf{w})\rangle \overset{a.s.}{=} \mathbb{E}[g_0^2(\sigma_0\widetilde{Z}_0,W)]$, leading to $\lim_{n\to\infty}\langle\mathbf{m}^0,\mathbf{m}^0\rangle \overset{a.s.}{=} \tau_0^2$ due to the definitions of $\mathbf{m}^0$ in (3) and τ_0^2 in (7). Thus, Proposition 5 (Lindeberg-Feller) holds because $\mathbf{m}_\perp^0 = \mathbf{m}^0$ and (13), resulting in $\widetilde{\mathbf{A}}^*\mathbf{m}^0 \overset{\mathrm{d}}{\Rightarrow} \mathcal{N}(0,\tau_0^2)$ and $\widehat{\mathbf{h}^1} \overset{\mathrm{d}}{\Rightarrow} \mathcal{N}(0,\tau_0^2)$, i.e., $\lim_{N\to\infty}\langle\mathbf{h}^1,\mathbf{h}^1\rangle \overset{a.s.}{=} \tau_0^2$. Hence, $\lim_{N\to\infty}\langle\mathbf{h}^1,\mathbf{h}^1\rangle \overset{a.s.}{=} \lim_{n\to\infty}\langle\mathbf{m}^0,\mathbf{m}^0\rangle<\infty$, concluding (15).

d) We let $\widetilde{\mathbf{v}}_i^0 = (\tau_0 Z_0, x_{0i})$, $\forall i$. Because $\mathbf{v}_i^0 \overset{\mathrm{d}}{\Rightarrow} \widetilde{\mathbf{v}}_i^0$ holds due to $\widehat{\mathbf{h}_i^1} \overset{\mathrm{d}}{\Rightarrow} \mathcal{N}(0,\tau_0^2)$ (Lindeberg-Feller), $\forall i$, the proof of (17) is boiled down to showing $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\phi_h(\widetilde{\mathbf{v}}_i^0) \overset{a.s.}{=} \mathbb{E}[\phi_h(\tau_0 Z_0,X_0)]$. Similar to Step 1d), by the triangular inequality $\left|\frac{1}{N}\sum_{i=1}^N\phi_h(\widetilde{\mathbf{v}}_i^0) - \mathbb{E}[\phi_h(\tau_0 Z_0,X_0)]\right| \leq Y_1^0 + Y_2^0$, where $Y_1^0 = \left|\frac{1}{N}\sum_{i=1}^N\phi_h(\widetilde{\mathbf{v}}_i^0) - \mathbb{E}_{Z_0}[\phi_h(\widetilde{\mathbf{v}}_i^0)]\right|$ and $Y_2^0 = \left|\frac{1}{N}\sum_{i=1}^N\mathbb{E}_{Z_0}[\phi_h(\widetilde{\mathbf{v}}_i^0)] - \mathbb{E}[\phi_h(\tau_0 Z_0,X_0)]\right|$. We claim that $\lim_{N\to\infty}Y_1^0 \overset{a.s.}{=} 0$ and $\lim_{N\to\infty}Y_2^0 \overset{a.s.}{=} 0$.

The convergence $\lim_{N\to\infty}Y_1^0 \overset{a.s.}{=} 0$ is treated first. Defining $Y_{N,i}^0 = \phi_h(\widetilde{\mathbf{v}}_i^0) - \mathbb{E}_{Z_0}[\phi_h(\widetilde{\mathbf{v}}_i^0)]$, the proof is equivalent to showing that $\{Y_{N,i}^0\}_{i=1}^N$ satisfy Lemma 3 in Appendix F. Since ϕ_h is a controlled function, the following holds $|\phi_h(\widetilde{\mathbf{v}}_i^0)| \leq d_1^0\exp(d_2^0(|\tau_0 Z_0|^\lambda + |x_{0i}|^\lambda))$, where $d_1^0>0$, $d_2^0>0$, and $1\leq\lambda<2$ are constant. Hence, $\mathbb{E}_{Z_0}[|\phi_h(\widetilde{\mathbf{v}}_i^0)|^{2+\kappa}] \leq d_3^0\exp(d_4^0|x_{0i}|^\lambda)\mathbb{E}_{Z_0}[\exp(d_4^0|\tau_0 Z_0|^\lambda)]$, where $0<\kappa<1$, $d_3^0 = (d_1^0)^{2+\kappa}$, and $d_4^0 = d_2^0(2+\kappa)$. Therefore,

$$\mathbb{E}_{Z_0}[|\phi_h(\widetilde{\mathbf{v}}_i^0)|^{2+\kappa}] \leq d_5^0\exp(d_4^0|x_{0i}|^\lambda), \quad (26)$$

where $d_5^0 = d_3^0\mathbb{E}_{Z_0}[\exp(d_4^0|\tau_0 Z_0|^\lambda)]$ is a constant. By Lemma 4 in Appendix F,

$$\begin{aligned}\mathbb{E}_{Z_0}[|Y_{N,i}^0|^{2+\kappa}] &\leq 2^{1+\kappa}\left(\mathbb{E}_{Z_0}[|\phi_h(\widetilde{\mathbf{v}}_i^0)|^{2+\kappa}] + |\mathbb{E}_{Z_0}[\phi_h(\widetilde{\mathbf{v}}_i^0)]|^{2+\kappa}\right),\\ &\leq d_6^0\exp(d_4^0|x_{0i}|^\lambda), \quad (27a)\end{aligned}$$

where $d_6^0 = 2^{2+\kappa}d_5^0$ and (27a) follows from Lemma 5 in Appendix F and (26). Note that $\frac{1}{N}\sum_{i=1}^N d_6^0\exp(d_4^0|x_{0i}|^\lambda) \overset{a.s.}{=} \mathbb{E}[d_6^0\exp(d_4^0|X_0|^\lambda)]$. For N is sufficiently large, we have

$$\frac{1}{N}\sum_{i=1}^N\mathbb{E}_{Z_0}[|Y_{N,i}^0|^{2+\kappa}] \leq \mathbb{E}[d_6^0\exp(d_4^0|X_0|^\lambda)] < cN^{\kappa/2}, \quad (28)$$

where $c>0$ is constant and the last estimation in (28) follows because $\mathbb{E}[d_6^0\exp(d_4^0|X_0|^\lambda)] = d_7^0 < \infty$ and there exists a constant $N_0>0$ such that $d_7^0 < cN^{\kappa/2}$ for $N>N_0$. Given (28), applying Lemma 3 in Appendix F to $\{Y_{N,i}^0\}_{i=1}^N$ gives $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N Y_{N,i}^0 \overset{a.s.}{=} 0$, implying $\lim_{N\to\infty}Y_1^0 \overset{a.s.}{=} 0$.

To show $\lim_{N\to\infty}Y_2^0 \overset{a.s.}{=} 0$, we define $\widetilde{\phi}_h(x_{0i}) = \mathbb{E}_{Z_0}[\phi_h(\widetilde{\mathbf{v}}_i^0)]$. Then the following holds, $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\widetilde{\phi}_h(x_{0i}) \overset{a.s.}{=} \mathbb{E}[\widetilde{\phi}_h(X_0)]$ because $\widehat{\mathbf{x}_0} \overset{\mathrm{d}}{\Rightarrow} \mu_{X_0}$. Thus, $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\mathbb{E}_{Z_0}[\phi_h(\widetilde{\mathbf{v}}_i^0)] \overset{a.s.}{=} \mathbb{E}_{X_0}[\mathbb{E}_{Z_0}[\phi_h(\tau_0 Z_0,X_0)]] = \mathbb{E}[\phi_h(\tau_0 Z_0,X_0)]$, resulting in $\lim_{N\to\infty}Y_2^0 \overset{a.s.}{=} 0$, concluding (17).

Suppose Theorem 1 holds up to the $(t-1)$th iteration. We prove that the thesis also holds for the tth iteration. Similar to the first two steps, we show (10), (12), (14), and (16) in Step 3 and (11), (13), (15), and (17) in Step 4, respectively, which are more complex and thus, relegated to Appendix H.

V. Conclusion

The SE analysis for AMP was extended to the class of measurement matrices with independent (not necessarily identically distributed) entries and the controlled functions. A variant of the Lindeberg-Feller was proposed to deal with the lack of the structure of the assumed measurement ensemble. An empirical statistic-based conditioning technique was proposed to cope with the lack of smoothness of the controlled functions. The results revealed a new direction to the SE analysis for boarder classes of measurement ensembles and functions.

REFERENCES

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, 1997.

[2] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] S. K. Shastri, R. Ahmad, C. A. Metzler, and P. Schniter, "Denoising generalized expectation-consistent approximation for MR image recovery," *IEEE Journal on Selected Areas in Information Theory*, pp. 1–1, 2022.

[4] T. Richardson and R. Urbanke, "The capacity of low-density parity-check codes under message-passing decoding," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 599–618, 2001.

[5] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[6] D. L. Donoho, A. Maleki, and A. Montanari, "Message passing algorithms for compressed sensing: I. motivation and construction," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, 2010, pp. 1–5.

[7] ——, "Message passing algorithms for compressed sensing: II. analysis and validation," in *2010 IEEE Information Theory Workshop on Information Theory (ITW 2010, Cairo)*, 2010, pp. 1–5.

[8] A. Montanari and R. Venkataramanan, "Estimation of low-rank matrices via approximate message passing," *Annals of Statistics*, vol. 49, 11 2017.

[9] A. Fletcher and S. Rangan, "Iterative reconstruction of rank-one matrices in noise," *Information and Inference: A Journal of the IMA*, vol. 7, pp. 531–562, 09 2018.

[10] T. Lesieur, F. Krzakala, and L. Zdeborová, "Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2017, 01 2017.

[11] P. Pandit, M. Sahraee, S. Rangan, and A. K. Fletcher, "Asymptotics of MAP inference in deep networks," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 842–846.

[12] P. Pandit, M. Sahraee-Ardakan, S. Rangan, P. Schniter, and A. Fletcher, "Inference with deep generative priors in high dimensions," *IEEE Journal on Selected Areas in Information Theory*, vol. PP, pp. 1–1, 04 2020.

[13] M. M. Emami, M. Sahraee-Ardakan, P. Pandit, S. Rangan, and A. K. Fletcher, "Generalization error of generalized linear models in high dimensions," in *International Conference on Machine Learning*, 2020.

[14] Z. Zhang, X. Cai, C. Li, C. Zhong, and H. Dai, "One-bit quantized massive MIMO detection based on variational approximate message passing," *IEEE Transactions on Signal Processing*, vol. 66, no. 9, pp. 2358–2373, 2018.

[15] C. Jeon, R. Ghods, A. Maleki, and C. Studer, "Optimality of large MIMO detection via approximate message passing," in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 1227–1231.

[16] P. Schniter, "Turbo reconstruction of structured sparse signals," in *2010 44th Annual Conference on Information Sciences and Systems (CISS)*, 2010, pp. 1–6.

[17] T. Kim and D. J. Love, "Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels," in *2015 IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2015, pp. 146–150.

[18] Q. Duan, T. Kim, H. Huang, K. Liu, and G. Wang, "Aod and aoa tracking with directional sounding beam design for millimeter wave mimo systems," in *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, 2015, pp. 2271–2276.

[19] F. Bellili, F. Sohrabi, and W. Yu, "Massive MIMO mmWave channel estimation using approximate message passing and Laplacian prior," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1–5.

[20] D. Baron, C. Rush, and Y. Yapici, "mmWave channel estimation via approximate message passing with side information," in *2020 IEEE 21st International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2020, pp. 1–5.

[21] K. Sung, B. L. Daniel, and B. A. Hargreaves, "Location constrained approximate message passing for compressed sensing MRI," *Magnetic Resonance in Medicine*, vol. 70, no. 2, pp. 370–381, 2013.

[22] J. Tan, Y. Ma, and D. Baron, "Compressive imaging via approximate message passing with image denoising," *IEEE Transactions on Signal Processing*, vol. 63, no. 8, pp. 2085–2092, 2015.

[23] C. Millard, A. T. Hess, B. Mailhe, and J. Tanner, "An approximate message passing algorithm for rapid parameter-free compressed sensing MRI," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 91–95.

[24] F. Caltagirone, L. Zdeborová, and F. Krzakala, "On convergence of approximate message passing," *2014 IEEE International Symposium on Information Theory*, pp. 1812–1816, 2014.

[25] J. Ma and L. Ping, "Orthogonal AMP," *IEEE Access*, vol. 5, pp. 2020–2033, 2017.

[26] S. Rangan, P. Schniter, A. K. Fletcher, and S. Sarkar, "On the convergence of approximate message passing with arbitrary matrices," *IEEE Trans. Inf. Theor.*, vol. 65, no. 9, p. 5339–5351, sep 2019.

[27] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, Feb 2011.

[28] C. Rush and R. Venkataramanan, "Finite sample analysis of approximate message passing algorithms," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 7264–7286, 2018.

[29] M. Bayati, M. Lelarge, and A. Montanari, "Universality in polytope phase transitions and message passing algorithms," *The Annals of Applied Probability*, vol. 25, no. 2, pp. 753 – 822, 2015.

[30] W.-K. Chen and W.-K. Lam, "Universality of approximate message passing algorithms," *Electronic Journal of Probability*, vol. 26, pp. 1 – 44, 2021. [Online]. Available: https://doi.org/10.1214/21-EJP604

[31] S. Rangan, P. Schniter, and A. K. Fletcher, "Vector approximate message passing," *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6664–6684, 2019.

[32] Z. Fan, "Approximate message passing algorithms for rotationally invariant matrices," *ArXiv*, vol. abs/2008.11892, 2020.

[33] K. Takeuchi, "Rigorous dynamics of expectation-propagation-based signal recovery from unitarily invariant measurements," *IEEE Transactions on Information Theory*, vol. 66, no. 1, pp. 368–386, 2020.

[34] R. Dudeja, Y. M. Lu, and S. Sen, "Universality of approximate message passing with semi-random matrices," 2022.

[35] E. Bolthausen, "On the high-temperature phase of the Sherrington-Kirkpatrick model," *Seminar presented at EURANDOM*, September 2009.

[36] G. Yang, "Tensor programs I: Wide feedforward or recurrent neural networks of any architecture are Gaussian processes," in *Neural Information Processing Systems*, 2019.

[37] ——, "Scaling limits of wide neural networks with weight sharing: Gaussian process behavior, gradient independence, and neural tangent kernel derivation," *ArXiv*, vol. abs/1902.04760, 2019.

[38] R. Durrett, *Probability: Theory and Examples (Cambridge Series in Statistical and Probabilistic Mathematics)*. Cambridge University Press; 4 edition, 2010.

[39] C. Stein, "A bound for the error in the normal approximation to the distribution of a sum of dependent random variables," in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, Berkeley, Calif., 1972, pp. 583–602.

[40] T.-C. Hu and R. L. Taylor, "On the strong law for arrays and for the bootstrap mean and variance," *International Journal of Mathematics and Mathematical Sciences*, vol. 20, pp. 375–382, 1997.

[41] A. S. Poznyak, "4 - basic probabilistic inequalities," in *Advanced Mathematical Tools for Automatic Control Engineers: Stochastic Techniques*, A. S. Poznyak, Ed. Oxford: Elsevier, 2009, pp. 63–81.

Appendix A
Pseudo-Lipschitz Function

***Definition* 1.** *For a $k > 1$, a function $f : \mathbb{R}^{n\times 1} \to \mathbb{R}$ is said pseudo-Lipschitz of order k if there exists a constant $L > 0$ such that $|f(\mathbf{x}) - f(\mathbf{y})| \leq L(1 + \|\mathbf{x}\|^{k-1} + \|\mathbf{y}\|^{k-1})\|\mathbf{x} - \mathbf{y}\|$, for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{n\times 1}$; the first order derivative of f is bounded by a polynomial of order $(k-1)$, i.e., polynomial smoothness.*

Appendix B
Proof of Proposition 1

This follows from $\sum_{m=1}^{n} \mathbb{E}\big[X_{n,m}^2; |X_{n,m}| > \epsilon\big] \leq \sum_{m=1}^{n} \mathbb{E}\Big[\frac{X_{n,m}^{2+2\alpha}}{\epsilon^{2\alpha}}\Big]$, where the right-hand side converges to zero because $\frac{n}{\epsilon^{2\alpha}} o(n^{-1}) \to 0$ as $n \to \infty$.

Appendix C
Proof of Proposition 3

Recalling the following subspace decomposition $\mathbf{A} = \mathbf{P}_{\mathbf{M}}^{\perp}\mathbf{A}\mathbf{P}_{\mathbf{Q}}^{\perp} + \mathbf{A}\mathbf{P}_{\mathbf{Q}} + \mathbf{P}_{\mathbf{M}}\mathbf{A} - \mathbf{P}_{\mathbf{M}}\mathbf{A}\mathbf{P}_{\mathbf{Q}}$, the orthogonal projection of $\mathbf{A}$ onto $\mathcal{P}_{\mathbf{X},\mathbf{Y}}$ is given by $\mathbf{A}|_{\mathcal{P}_{\mathbf{X},\mathbf{Y}}} = \mathbf{P}_{\mathbf{M}}^{\perp}\mathbf{A}\mathbf{P}_{\mathbf{Q}}^{\perp} + \mathbf{B}$. For any integrable function ψ, $\mathbb{E}\left[\psi(\mathbf{A}|_{\mathcal{P}_{\mathbf{X},\mathbf{Y}}})\right] = \mathbb{E}\left[\psi(\mathbf{P}_{\mathbf{M}}^{\perp}\mathbf{A}\mathbf{P}_{\mathbf{Q}}^{\perp} + \mathbf{B})\right] = \mathbb{E}\left[\psi(\mathbf{P}_{\mathbf{M}}^{\perp}\widetilde{\mathbf{A}}\mathbf{P}_{\mathbf{Q}}^{\perp} + \mathbf{B})\right] = \mathbb{E}\left[\psi(\widetilde{\mathbf{A}}|_{\mathcal{P}_{\mathbf{X},\mathbf{Y}}})\right]$, where the second equality follows from the fact that $\mathbf{A} \overset{d}{=} \widetilde{\mathbf{A}}$. Hence, $\mathbf{A}|_{\mathcal{P}_{\mathbf{X},\mathbf{Y}}} \overset{d}{=} \mathbf{P}_{\mathbf{M}}^{\perp}\widetilde{\mathbf{A}}\mathbf{P}_{\mathbf{Q}}^{\perp} + \mathbf{B}$, which completes the proof.

Appendix D
Proof of Proposition 4

Denoting $B = \sqrt{n}A(n) \sim \mu$ yields $\mathbb{E}\big[A^{2+2\alpha}(n)\big] = \mathbb{E}\big[B^{2(1+\alpha)}\big]n^{-(1+\alpha)} = o(n^{-2})$, where the last step uses the facts that $\mathbb{E}\big[B^{2(1+\alpha)}\big]$ is independent of n and $\alpha > 1$.

Appendix E
Proof of Proposition 5

Denoting $X_{N,ij} = A_{ij}(N)v_j(N)$, then $\{X_{N,ij}\colon 1 \leq j \leq N\}$ is an independent zero-mean triangular array, for $i = 1, 2, \ldots, n$. We claim that $\{X_{N,ij}\colon 1 \leq j \leq N\}$ satisfies two conditions in Proposition 2, $\forall i$. First, we note that $\sum_{j=1}^{N} \mathbb{E}[X_{N,ij}^2] = \sum_{j=1}^{N} \mathbb{E}[A_{ij}^2(N)]v_j^2(N) = \frac{1}{n}\sum_{j=1}^{N} v_j^2(N) = \frac{1}{\rho}\langle \mathbf{v}^2(N)\rangle \to \frac{s_0^2}{\rho}$ as $n \to \infty$. Second, applying Proposition 4 to $A_{ij}(N)$ gives $\mathbb{E}[A_{ij}^{2+2\alpha}(N)] = o(n^{-2})$, leading to $\mathbb{E}[X_{N,ij}^{2+2\alpha}] = \mathbb{E}[A_{ij}^{2+2\alpha}(N)]|v_j(N)|^{2+2\alpha} \leq o(n^{-2})\frac{n}{\rho}\frac{1}{N}\|\mathbf{v}(N)\|_{2+2\alpha}^{2+2\alpha} = o(n^{-1})$, where the last equality holds because $\limsup_{N\to\infty} \frac{1}{N}\|\mathbf{v}(N)\|_{2+2\alpha}^{2+2\alpha} < \infty$ and ρ is a constant. Applying Proposition 2 to $\{X_{N,ij} : 1 \leq j \leq N\}$ leads to $[\mathbf{A}(N)\mathbf{v}(N)]_i = \sum_{j=1}^{N} X_{N,ij} \overset{\mathrm{d}}{\Rightarrow} \mathcal{N}\left(0, \frac{s_0^2}{\rho}\right)$ as $N \to \infty$, $\forall i$. Hence, $\mathbf{A}\widehat{(N)\mathbf{v}}(N) \overset{d}{\Rightarrow} \mathcal{N}\left(0, \frac{s_0^2}{\rho}\right)$ as $N \to \infty$.

Appendix F
Well-known Lemmas

***Lemma* 2.** *(Stein's Lemma [39]) For jointly zero-mean Gaussian random variables Z_1 and Z_2, and any function $\phi : \mathbb{R} \to \mathbb{R}$, where $\mathbb{E}[\phi'(Z_2)]$ and $\mathbb{E}[Z_1\phi(Z_2)]$ exist, the following holds $\mathbb{E}[Z_1\phi(Z_2)] = \mathrm{Cov}(Z_1, Z_2)\mathbb{E}[\phi'(Z_2)]$, where $\mathrm{Cov}(Z_1, Z_2)$ is the covariance between Z_1 and Z_2.*

***Lemma* 3.** *(Strong Law of Large Number [40]) Let $\{X_{n,m} : 1 \leq m \leq n\}$ be a triangular array of random variables with $(X_{n,1}, X_{n,2}, \ldots, X_{n,n})$ mutually independent with zero-mean for each n and $\frac{1}{n}\sum_{m=1}^{n} \mathbb{E}[|X_{n,m}|^{2+\kappa}] \leq cn^{\kappa/2}$ for some $0 < \kappa < 1$ and $c < \infty$. Then $\lim_{n\to\infty} \frac{1}{n}\sum_{m=1}^{n} X_{n,m} \overset{a.s.}{=} 0$.*

***Lemma* 4.** *(Holder's inequality [41]) For random variables X and Y, $\mathbb{E}[|X + Y|^r] \leq c_r(\mathbb{E}[|X|^r] + \mathbb{E}[|Y|^r])$, where c_r = 1 if $0 < r \leq 1$ and $c_r = 2^{r-1}$ otherwise. In particular, the inequality becomes $\mathbb{E}[|X + y|^r] \leq c_r(\mathbb{E}[|X|^r] + |y|^r])$ when $Y = y$ being a constant.*

***Lemma* 5.** *(Lyapunov's inequality [41]) Suppose a random variable X and a constant κ with $0 < \kappa < 1$, then $|\mathbb{E}[X]|^{2+\kappa} \leq \mathbb{E}[|X|^{2+\kappa}]$.*

***Proposition* 7.** *Suppose the $\mathbf{P}_{\mathbf{M}(n)} = \big(\frac{1}{\sqrt{n}}\mathbf{V}(n)\big)\big(\frac{1}{\sqrt{n}}\mathbf{V}(n)\big)^*$, where $\mathbf{M}(n) \in \mathbb{R}^{n\times t}$ $(t \leq n)$, t is a fixed constant, and $\mathbf{V}(n)$=$[\mathbf{v}_1(n), \mathbf{v}_2(n), \ldots, \mathbf{v}_t(n)] \in \mathbb{R}^{n\times t}$ is an orthogonal basis of $\mathbf{M}(n)$ such that $\mathbf{V}^*(n)\mathbf{V}(n) = n\mathbf{I}$. If we let $\mathbf{a}(n) \in \mathbb{R}^{n\times 1}$ be a random vector with independent entries, which have zero mean and finite variance σ_a^2, then $\lim_{n\to\infty} \mathbf{P}_{\mathbf{M}}(n)\mathbf{a}(n) \overset{a.s.}{=} \mathbf{0}_n$, where $\mathbf{0}_n$ is the $n \times 1$ all-zero vector.*

Proof. Denoting $\widetilde{\mathbf{a}}(n) = \frac{\mathbf{a}(n)}{\|\mathbf{a}(n)\|_2}$ yields $\mathbf{P}_{\mathbf{M}(n)}\mathbf{a}(n) = \mathbf{V}(n)\frac{\|\mathbf{a}(n)\|_2}{\sqrt{n}}\left(\frac{1}{\sqrt{n}}\mathbf{V}^*(n)\right)\widetilde{\mathbf{a}}(n)$. The proposition follows from the fact that $\frac{\|\mathbf{a}(n)\|_2}{\sqrt{n}} \overset{a.s.}{=} \sigma_a$ and $\left(\frac{1}{\sqrt{n}}\mathbf{V}^*(n)\right)\widetilde{\mathbf{a}}(n) \overset{a.s.}{=} \mathbf{0}_n$ as $n \to \infty$. ☐

Appendix G
Proof of SE in (7) Theorem 1

Substituting $\phi_b(\mathbf{u}_i^t) = (b_i^t)^2$ into (16) gives $\lim_{n\to\infty}\langle \mathbf{b}^t, \mathbf{b}^t\rangle \overset{a.s}{=} \mathbb{E}\left[\sigma_t^2 \widetilde{Z}_t^2\right] = \sigma_t^2$. Using (14) with $t_1 = t_2 = t$ yields $\lim_{N\to\infty} \frac{1}{\rho}\langle \mathbf{q}^t, \mathbf{q}^t\rangle = \sigma_t^2$. Then substituting $\phi_h(\mathbf{v}_i^{t-1}) = f_t^2(h_i^t, x_{0i}) = (q_i^t)^2$ into (17) leads to $\lim_{N\to\infty}\langle \mathbf{q}^t, \mathbf{q}^t\rangle \overset{a.s}{=} \mathbb{E}\left[f_t^2(\tau_{t-1}Z_{t-1}, X_0)\right]$, resulting in $\sigma_t^2 = \frac{1}{\rho}\mathbb{E}\left[f_t^2(\tau_{t-1}Z_{t-1}, X_0)\right]$. Showing the rest half $\tau_t^2 = \mathbb{E}\left[g_t^2(\sigma_t Z, W)\right]$ of the SE in (7) follows from the exactly same procedure as the above. Setting $\phi_h(\mathbf{v}_i^t) = (h_i^{t+1})^2$ in (17) gives $\lim_{N\to\infty}\langle \mathbf{h}^{t+1}, \mathbf{h}^{t+1}\rangle = \tau_t^2$. Using (15) with $t_1 = t_2 = t$ yields $\lim_{N\to\infty}\langle \mathbf{m}^t, \mathbf{m}^t\rangle = \tau_t^2$. Applying (16) to $\phi_b(\mathbf{u}_i^t) = g_t^2(b_i^t, w_i)$ yields $\lim_{N\to\infty}\langle \mathbf{m}^t, \mathbf{m}^t\rangle = \mathbb{E}\left[g_t^2(\sigma_t Z, W)\right]$. Therefore, $\tau_t^2 = \mathbb{E}\left[g_t^2(\sigma_t Z, W)\right]$, concluding the proof.

Appendix H
Proof of Theorem 1: Steps 3 and 4

A. Step 3: We show a), b), c), and d) of Theorem 1 conditioning on $\mathcal{F}_{t,t} = \{\mathbf{b}^0, \ldots, \mathbf{b}^{t-1}, \mathbf{m}^0, \ldots, \mathbf{m}^{t-1}, \mathbf{h}^1, \ldots, \mathbf{h}^t, \mathbf{q}^0, \ldots, \mathbf{q}^t, \mathbf{x}_0, \mathbf{w}\}$.

a) Note that conditioning on $\mathcal{F}_{t,t}$ is equivalent to conditioning on $\mathcal{P}_{\mathbf{X}_t,\mathbf{Y}_t} = \{\mathbf{A}|\mathbf{A}^*\mathbf{M}_t = \mathbf{X}_t, \mathbf{A}\mathbf{Q}_t = \mathbf{Y}_t\}$. Applying Proposition 3 to obtain the conditional distribution $\mathbf{A}|_{\mathcal{F}_{t,t}}$

and following the same procedure as in [27, Lemma 1a], the conditional distribution of $\mathbf{b}^t$ on $\mathcal{F}_{t,t}$ is expressed as

$$\mathbf{b}^t|_{\mathcal{F}_{t,t}} \overset{d}{=} \sum_{j=0}^{t-1} \beta_j \mathbf{b}^j + \widetilde{\mathbf{A}}\mathbf{q}_\perp^t - \mathbf{P}_{\mathbf{M}_t}\widetilde{\mathbf{A}}\mathbf{q}_\perp^t + \mathbf{M}_t \overrightarrow{\mathbf{o}_t}(1). \quad (29)$$

By Proposition 7 in Appendix F, the third term on the r.h.s of (29) converges to $\lim_{n\to\infty} \mathbf{P}_{\mathbf{M}_t}\widetilde{\mathbf{A}}\mathbf{q}_\perp^t \overset{a.s.}{=} \mathbf{0}_n$. Similar to Step 2a), we verify the convergence $\lim_{n\to\infty} \mathbf{M}_t \overrightarrow{\mathbf{o}}_t(1) \overset{a.s.}{=} \mathbf{0}_n$ by characterizing the expectation and variance of its empirical distribution $\widehat{\mathbf{M}_t \overrightarrow{\mathbf{o}}_t(1)}$ as $n \to \infty$. Indeed, $\lim_{n\to\infty} |\langle \mathbf{M}_t \overrightarrow{\mathbf{o}}_t(1)\rangle| \le \lim_{n\to\infty} |o(1)| \frac{1}{n}\sum_{i=1}^n \left|\sum_{j=0}^{t-1} m_i^j\right| \le \lim_{n\to\infty} |o(1)| \sum_{j=0}^{t-1} \frac{1}{n}\sum_{i=1}^n |m_i^j| \overset{a.s.}{=} 0$, where the last equality holds because applying $\phi_b(b_i^j, w_i) = g_j(b_i^j, w_i)$ to the induction hypothesis of (16), for $j < t$, leads to $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n |m_i^j| \overset{a.s.}{=} \mathbb{E}[|g_j(\sigma_j \widetilde{Z}_j, W)|] < \infty$. Hence, $\lim_{n\to\infty} \langle \mathbf{M}_t \overrightarrow{\mathbf{o}}_t(1)\rangle \overset{a.s.}{=} 0$. For the variance of $\widehat{\mathbf{M}_t \overrightarrow{\mathbf{o}}_t(1)}$, we get

$$\lim_{n\to\infty} \langle \mathbf{M}_t \overrightarrow{\mathbf{o}}_t(1)\rangle_2 = \lim_{n\to\infty} \frac{1}{n}[o(1)]^2 \sum_{i=1}^n \left(\sum_{j=0}^{t-1} m_i^j\right)^2,$$
$$\le \lim_{n\to\infty} [o(1)]^2 t \frac{1}{n}\sum_{i=1}^n \sum_{j=0}^{t-1} (m_i^j)^2, \quad (30a)$$
$$= \lim_{n\to\infty} [o(1)]^2 t \sum_{j=0}^{t-1} \langle \mathbf{m}^j, \mathbf{m}^j\rangle \overset{a.s.}{=} 0, \quad (30b)$$

where (30a) follows from the Cauchy-Schwarz inequality and (30b) holds because $\lim_{n\to\infty} \langle \mathbf{m}^j, \mathbf{m}^j\rangle \overset{a.s.}{=} \mathbb{E}[g_j^2(\sigma_j \widetilde{Z}_j, W)] < \infty$, $\forall j$, which is a direct consequence of the induction hypothesis of (16) with $\phi_b(b_i^j, w_i) = g_j^2(b_i^j, w_i)$. Hence, (30b) is equivalent to $\lim_{n\to\infty} \langle \mathbf{M}_t \overrightarrow{\mathbf{o}}_t(1)\rangle_2 \overset{a.s.}{=} 0$. Therefore, $\lim_{n\to\infty} \mathbf{M}_t \overrightarrow{\mathbf{o}}_t(1) \overset{a.s.}{=} \mathbf{0}_n$, implying

$$\mathbf{b}^t|_{\mathcal{F}_{t,t}} \overset{d}{\Rightarrow} \sum_{j=0}^{t-1} \beta_j \mathbf{b}^j + \widetilde{\mathbf{A}}\mathbf{q}_\perp^t. \quad (31)$$

b) Note that by the induction hypothesis of (17) for $\phi_h(h_i^t, x_{0i}) = |f_t(h_i^t, x_{0i})|^{2+2\alpha}$, we get $\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N |q_i^t|^{2+2\alpha} \overset{a.s.}{=} \mathbb{E}[|f_t(\tau_{t-1}Z_{t-1}, X_0)|^{2+2\alpha}] < \infty$. On the other hand, $\sum_{i=1}^N |q_{\perp i}^t|^{2+2\alpha} < \sum_{i=1}^N |q_i^t|^{2+2\alpha}$. Thus, we have $\limsup_{N\to\infty} \frac{1}{N}\sum_{i=1}^N |q_{\perp i}^t|^{2+2\alpha} < \infty$, which concludes (12).

c) For $t_1 < t$ and $t_2 = t$, we obtain

$$\lim_{n\to\infty} \langle \mathbf{b}^{t_1}, \mathbf{b}^t\rangle \overset{d}{=} \lim_{n\to\infty} \sum_{j=0}^{t-1} \beta_j \langle \mathbf{b}^{t_1}, \mathbf{b}^j\rangle + \lim_{n\to\infty} \langle \mathbf{b}^{t_1}, \widetilde{\mathbf{A}}\mathbf{q}_\perp^t\rangle, \quad (32a)$$
$$\overset{a.s.}{=} \sum_{j=0}^{t-1} \beta_j \lim_{N\to\infty} \frac{\langle \mathbf{q}^{t_1}, \mathbf{q}^j\rangle}{\rho} + \lim_{n\to\infty} \frac{\mathbf{b}^{t_1 *}\widetilde{\mathbf{A}}\mathbf{q}_\perp^t}{n}, \quad (32b)$$

where (32a) follows from (31) and (32b) results from the induction hypothesis (14) for $t_1 < t$ and $t_2 = j < t$. Now, using Proposition 6, we get $\frac{\mathbf{b}^{t_1^*}}{\|\mathbf{b}^{t_1}\|_2}\widetilde{\mathbf{A}}\frac{\mathbf{q}_\perp^t}{\|\mathbf{q}_\perp^t\|_2} \overset{d}{=} \frac{Z}{\sqrt{n}}$, where $Z \sim \mathcal{N}(0,1)$. Hence, $\frac{\mathbf{b}^{t_1^*}\widetilde{\mathbf{A}}\mathbf{q}_\perp^t}{n} \overset{d}{=} \frac{\|\mathbf{b}^{t_1}\|_2}{\sqrt{n}}\frac{\|\mathbf{q}_\perp^t\|_2}{\sqrt{N}}\frac{1}{\sqrt{\rho}}\frac{Z}{\sqrt{n}}$, i.e.,

$$\frac{\mathbf{b}^{t_1^*}\widetilde{\mathbf{A}}\mathbf{q}_\perp^t}{n} \overset{d}{=} \frac{1}{\sqrt{\rho}}\sqrt{\langle \mathbf{b}^{t_1}, \mathbf{b}^{t_1}\rangle \langle \mathbf{q}_\perp^t, \mathbf{q}_\perp^t\rangle}\frac{Z}{\sqrt{n}}. \quad (33)$$

By the induction hypothesis of (14), we have $\lim_{n\to\infty} \langle \mathbf{b}^{t_1}, \mathbf{b}^{t_1}\rangle = \lim_{n\to\infty} \frac{\langle \mathbf{q}^{t_1}, \mathbf{q}^{t_1}\rangle}{\rho} < \infty$. Moreover, the $\langle \mathbf{q}_\perp^t, \mathbf{q}_\perp^t\rangle$ converges to $\lim_{N\to\infty} \langle \mathbf{q}_\perp^t, \mathbf{q}_\perp^t\rangle < \lim_{N\to\infty} \langle \mathbf{q}^t, \mathbf{q}^t\rangle < \infty$ because using the induction hypothesis (17) we have $\langle \mathbf{q}^t, \mathbf{q}^t\rangle = \frac{1}{N}\sum_{i=1}^N f_t^2(h_i^t, x_{0i}) \overset{a.s.}{=} \mathbb{E}[f_t^2(\tau_{t-1}Z_{t-1}, X_0)] < \infty$. Thus, for $t_1 < t$,

$$\lim_{n\to\infty} \langle \widetilde{\mathbf{A}}\mathbf{q}_\perp^t, \mathbf{b}^{t_1}\rangle \overset{a.s.}{=} 0. \quad (34)$$

Substituting (34) into (32b) gives $\lim_{n\to\infty} \langle \mathbf{b}^{t_1}, \mathbf{b}^t\rangle \overset{a.s.}{=} \sum_{j=0}^{t-1} \beta_j \lim_{n\to\infty} \frac{\langle \mathbf{q}^{t_1}, \mathbf{q}^j\rangle}{\rho} = \lim_{N\to\infty} \frac{\langle \mathbf{q}^{t_1}, \mathbf{q}_{||}^t\rangle}{\rho}$ due to (8), implying

$$\lim_{n\to\infty} \langle \mathbf{b}^{t_1}, \mathbf{b}^t\rangle \overset{a.s.}{=} \lim_{N\to\infty} \frac{\langle \mathbf{q}^{t_1}, \mathbf{q}_{||}^t\rangle}{\rho} + \lim_{N\to\infty} \frac{\langle \mathbf{q}^{t_1}, \mathbf{q}_\perp^t\rangle}{\rho}, \quad (35a)$$
$$= \lim_{N\to\infty} \frac{\langle \mathbf{q}^{t_1}, \mathbf{q}^t\rangle}{\rho}, \quad (35b)$$

where (35a) follows from the fact that $\mathbf{q}^j$ is orthogonal to $\mathbf{q}_\perp^t$, for $j < t$, and (35b) holds due to (9), concluding (14) when $t_1 < t$ and $t_2 = t$.

For the case of $t_1 = t_2 = t$, it is similarly given by $\lim_{n\to\infty} \langle \mathbf{b}^t, \mathbf{b}^t\rangle \overset{d}{=} \sum_{i,j=0}^{t-1} \beta_i\beta_j \lim_{n\to\infty} \langle \mathbf{b}^i, \mathbf{b}^j\rangle + 2\sum_{i=0}^{t-1} \beta_i \lim_{n\to\infty} \langle \mathbf{b}^i, \widetilde{\mathbf{A}}\mathbf{q}_\perp^t\rangle + \lim_{n\to\infty} \langle \widetilde{\mathbf{A}}\mathbf{q}_\perp^t, \widetilde{\mathbf{A}}\mathbf{q}_\perp^t\rangle$ due to (31). Then, by (34), the following holds

$$\lim_{n\to\infty} \langle \mathbf{b}^t, \mathbf{b}^t\rangle \overset{a.s.}{=} \sum_{i,j=0}^{t-1} \beta_i\beta_j \lim_{n\to\infty} \langle \mathbf{b}^i, \mathbf{b}^j\rangle + \lim_{n\to\infty} \langle \widetilde{\mathbf{A}}\mathbf{q}_\perp^t, \widetilde{\mathbf{A}}\mathbf{q}_\perp^t\rangle. \quad (36)$$

Using Proposition 5, $\widehat{\widetilde{\mathbf{A}}\mathbf{q}_\perp^t} \overset{d}{\Rightarrow} \mathcal{N}\left(0, \lim_{N\to\infty} \frac{\langle \mathbf{q}_\perp^t, \mathbf{q}_\perp^t\rangle}{\rho}\right)$. Thus, the second moment of $\widehat{\widetilde{\mathbf{A}}\mathbf{q}_\perp^t}$ is

$$\lim_{n\to\infty} \langle \widetilde{\mathbf{A}}\mathbf{q}_\perp^t, \widetilde{\mathbf{A}}\mathbf{q}_\perp^t\rangle \overset{a.s.}{=} \lim_{N\to\infty} \frac{\langle \mathbf{q}_\perp^t, \mathbf{q}_\perp^t\rangle}{\rho}. \quad (37)$$

Now, incorporating (37) in (36), $\lim_{n\to\infty} \langle \mathbf{b}^t, \mathbf{b}^t\rangle \overset{a.s.}{=} \sum_{i,j=0}^{t-1} \beta_i\beta_j \lim_{n\to\infty} \langle \mathbf{b}^i, \mathbf{b}^j\rangle + \lim_{N\to\infty} \frac{\langle \mathbf{q}_\perp^t, \mathbf{q}_\perp^t\rangle}{\rho}$, resulting in

$$\lim_{n\to\infty} \langle \mathbf{b}^t, \mathbf{b}^t\rangle \overset{a.s.}{=} \sum_{i,j=0}^{t-1} \beta_i\beta_j \lim_{N\to\infty} \frac{\langle \mathbf{q}^i, \mathbf{q}^j\rangle}{\rho} + \lim_{N\to\infty} \frac{\langle \mathbf{q}_\perp^t, \mathbf{q}_\perp^t\rangle}{\rho}, \quad (38a)$$
$$\overset{a.s.}{=} \lim_{N\to\infty} \frac{\langle \mathbf{q}_{||}^t, \mathbf{q}_{||}^t\rangle}{\rho} + \lim_{N\to\infty} \frac{\langle \mathbf{q}_\perp^t, \mathbf{q}_\perp^t\rangle}{\rho},$$
$$\overset{a.s.}{=} \lim_{N\to\infty} \frac{\langle \mathbf{q}^t, \mathbf{q}^t\rangle}{\rho},$$

where (38a) is due to the induction hypothesis (14) for $0 \le t_1 = i, t_2 = j \le t-1$. This completes the proof of (14) at the tth iteration.

d) Defining $\lim_{N\to\infty}\frac{\langle \mathbf{q}_\perp^t,\mathbf{q}_\perp^t\rangle}{\rho}\overset{a.s.}{=}\gamma_t^2$, we can write by (37) that $\widehat{\widetilde{\mathbf{A}}\mathbf{q}_\perp^t}\overset{d}{\Longrightarrow}\mathcal{N}(0,\gamma_t^2)$. Using (31) in conjunction with the latter, we get

$$b_i^t|_{\mathcal{F}_{t,t}}\overset{d}{\Longrightarrow}\sum_{j=0}^{t-1}\beta_j b_i^j+\gamma_t Z,\ \text{for } i=1,2,\dots,n, \tag{39}$$

where $Z\sim\mathcal{N}(0,1)$. Similar to Step 1d), using (39) $\mathbf{u}_i^t\overset{d}{\Longrightarrow}\widetilde{\mathbf{u}}_i^t$, where $\mathbf{u}_i^t=(b_i^0\dots,b_i^t,w_i)$ and $\widetilde{\mathbf{u}}_i^t=(b_i^0,\dots,b_i^{t-1},\sum_{j=0}^{t-1}\beta_j b_i^j+\gamma_t Z,w_i)$, $\forall i$. To prove (16), we first claim that $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\phi_b(\widetilde{\mathbf{u}}_i^t)-\mathbb{E}\Big[\phi_b(\sigma_0\widetilde{Z}_0,\dots,\sigma_t\widetilde{Z}_t,W)\Big]\overset{a.s.}{=}0$. By the triangular inequality, $\Big|\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\phi_b(\widetilde{\mathbf{u}}_i^t)-\mathbb{E}\Big[\phi_b(\sigma_0\widetilde{Z}_0,\dots,\sigma_t\widetilde{Z}_t,W)\Big]\Big|\le X_1^t+X_2^t$, where $X_1^t=\Big|\frac{1}{n}\sum_{i=1}^n\Big(\phi_b(\widetilde{\mathbf{u}}_i^t)-\widetilde{\phi}_b(\mathbf{u}_i^{t-1})\Big)\Big|$, $X_2^t=\Big|\frac{1}{n}\sum_{i=1}^n\widetilde{\phi}_b(\mathbf{u}_i^{t-1})-\mathbb{E}[\phi_b(\sigma_0\widetilde{Z}_0,\dots,\sigma_t\widetilde{Z}_t,W)]\Big|$, and $\widetilde{\phi}_b(\mathbf{u}_i^{t-1})=\mathbb{E}_Z[\phi_b(\widetilde{\mathbf{u}}_i^t)]$. Similar to Step 1d), we verify $\lim_{n\to\infty}X_1^t\overset{a.s.}{=}0$ and $\lim_{n\to\infty}X_2^t\overset{a.s.}{=}0$.

First showing $\lim_{n\to\infty}X_1^t\overset{a.s.}{=}0$ is of interest. By (5), $|\phi_b(\widetilde{\mathbf{u}}_i^t)|\le c_1^t\exp\Big(c_2^t\Big(\sum_{j=0}^{t-1}|b_i^t|^\lambda+\Big|\sum_{j=0}^{t-1}\beta_j b_i^j+\gamma_t Z\Big|^\lambda+|w_i|^\lambda\Big)\Big)$, where $c_1^t>0$, $c_2^t>0$, and $1\le\lambda<2$ are constants. Using the inequality $\|\mathbf{x}\|_1^\lambda\le(t+1)^{\lambda-1}\|\mathbf{x}\|_\lambda^\lambda$ for $\mathbf{x}\in\mathbb{R}^{(t+1)\times 1}$, we get $|\phi_b(\widetilde{\mathbf{u}}_i^t)|\le c_1^t\exp\Big(c_2^t\Big(\sum_{j=0}^{t-1}(1+(t+1)^{\lambda-1}|\beta_j|^\lambda)|b_i^t|^\lambda+(t+1)^{\lambda-1}|\gamma_t|^\lambda|Z|^\lambda+|w_i|^\lambda\Big)\Big)$. Hence, $\mathbb{E}_Z[|\phi_b(\widetilde{\mathbf{u}}_i^t)|^{2+\kappa}]\le c_3^t\exp\Big(c_4^t\Big(\sum_{j=0}^{t-1}|b_i^j|^\lambda+|w_i|^\lambda\Big)\Big)\mathbb{E}_Z\left[\exp(c_4^t|Z|^\lambda)\right]$, where $0<\kappa<1$, $c_3^t=(c_1^t)^{2+\kappa}$, and $c_4^t=(2+\kappa)c_2^t\max\Big\{1+(t+1)^{\lambda-1}|\beta_0|^\lambda,\dots,1+(t+1)^{\lambda-1}|\beta_{t-1}|^\lambda,(t+1)^{\lambda-1}|\gamma_t|^\lambda\Big\}$, resulting in

$$\mathbb{E}_Z[|\phi_b(\widetilde{\mathbf{u}}_i^t)|^{2+\kappa}]\le c_5^t\exp\Big(c_4^t\Big(\sum_{j=0}^{t-1}|\beta_j b_i^j|^\lambda+|w_i|^\lambda\Big)\Big), \tag{40}$$

and $c_5^t=c_3^t\mathbb{E}_Z\left[\exp(c_4^t\gamma_t^\lambda|Z|^\lambda)\right]$ is constant. We define $X_{n,i}^t=\phi_b(\widetilde{\mathbf{u}}_i^t)-\widetilde{\phi}_b(\mathbf{u}_i^{t-1})=\phi_b(\widetilde{\mathbf{u}}_i^t)-\mathbb{E}_Z[\phi_b(\widetilde{\mathbf{u}}_i^t)]$ such that $X_1^t=|\frac{1}{n}\sum_{i=1}^n X_{n,i}^t|$. To prove $\lim_{n\to\infty}X_1^t\overset{a.s.}{=}0$, we show that $\{X_{n,i}^t\}_{i=1}^n$ satisfy Lemma 3 in Appendix F. Indeed, $\mathbb{E}_Z[|X_{n,i}^t|^{2+\kappa}]$ is upper bouned as follows,

$$\begin{aligned}\mathbb{E}_Z[|X_{n,i}^t|^{2+\kappa}]&\le 2^{1+\kappa}\Big(\mathbb{E}_Z\left[|\phi_b(\widetilde{\mathbf{u}}_i^t)|^{2+\kappa}\right]+\big|\mathbb{E}_Z[\phi_b(\widetilde{\mathbf{u}}_i^t)]\big|^{2+\kappa}\Big), &&\text{(41a)}\\ &\le 2^{2+\kappa}\mathbb{E}_Z\left[|\phi_b(\widetilde{\mathbf{u}}_i^t)|^{2+\kappa}\right], &&\text{(41b)}\\ &\le c_6^t\exp\Big(c_4^t\Big(\sum_{j=0}^{t-1}|\beta_j b_i^j|^\lambda+|w_i|^\lambda\Big)\Big), &&\text{(41c)}\end{aligned}$$

where (41a) follows from Lemma 4 (Holder's inequality) in Appendix F, (41b) follows from Lemma 5 (Lyapunov's inequality) in Appendix F, and (41c) follows from (40) with $c_6^t=2^{2+\kappa}c_5^t$. We denote the last term of (41c) as $\psi_b(\mathbf{u}_i^{t-1})=c_6^t\exp\Big(c_4^t\Big(\sum_{j=0}^{t-1}|\beta_j b_i^j|^\lambda+|w_i|^\lambda\Big)\Big)$. Then, $\psi_b(\mathbf{u}_i^{t-1})$ is a controlled function. From (41c), we get, for n is sufficiently large,

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n\mathbb{E}_Z[|X_{n,i}^t|^{2+\kappa}]&\le\frac{1}{n}\sum_{i=1}^n\psi_b(\mathbf{u}_i^{t-1}),\\ &\overset{a.s.}{=}\mathbb{E}[\psi_b(\sigma_0\widetilde{Z}_0,\dots,\sigma_{t-1}\widetilde{Z}_{t-1},W)], &&\text{(42a)}\\ &< cn^{\kappa/2}, &&\text{(42b)}\end{aligned}$$

where c is a positive constant, (42a) is due to the induction hypothesis (16), and (42b) holds because $\mathbb{E}[\psi_b(\sigma_0\widetilde{Z}_0,\dots,\sigma_{t-1}\widetilde{Z}_{t-1},W)]=c_7^t<\infty$ and there exists n_t, a positive constant, such that $c_7^t<cn^{\kappa/2}$ for $n>n_t$. By Lemma 3 in Appendix F, we get $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n X_{n,i}^t\overset{a.s.}{=}0$, implying

$$\frac{1}{n}\sum_{i=1}^n\Big(\phi_b(\widetilde{\mathbf{u}}_i^t)-\widetilde{\phi}_b(\mathbf{u}_i^{t-1})\Big)\overset{a.s.}{=}0, \tag{43}$$

which proving $\lim_{n\to\infty}X_1^t\overset{a.s.}{=}0$.

Now, showing $\lim_{n\to\infty}X_2^t\overset{a.s.}{=}0$ is of interest. By the induction hypothesis (16), $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\widetilde{\phi}_b(\mathbf{u}_i^{t-1})\overset{a.s.}{=}\mathbb{E}[\widetilde{\phi}_b(\sigma_0\widetilde{Z}_0,\dots,\sigma_{t-1}\widetilde{Z}_{t-1},W)]$, resulting in

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\widetilde{\phi}_b(\mathbf{u}_i^{t-1})\\ &=\mathbb{E}\left[\mathbb{E}_Z\left[\phi_b(\sigma_0\widetilde{Z}_0,\dots,\sigma_{t-1}\widetilde{Z}_{t-1},\sum_{j=0}^{t-1}\beta_j\sigma_j\widetilde{Z}_j+\gamma_t Z,W)\right]\right],\\ &=\mathbb{E}\left[\phi_b(\sigma_0\widetilde{Z}_0,\dots,\sigma_{t-1}\widetilde{Z}_{t-1},\sum_{j=0}^{t-1}\beta_j\sigma_j\widetilde{Z}_j+\gamma_t Z,W)\right],\end{aligned} \tag{44}$$

where (44) follows from the substitution $\widetilde{\phi}_b(\mathbf{u}_i^{t-1})=\mathbb{E}_Z[\phi_b(\widetilde{\mathbf{u}}_i^t)]$. Therefore, showing $\lim_{n\to\infty}X_2^t\overset{a.s.}{=}0$ is equivalent to proving $\sum_{j=0}^{t-1}\beta_j\sigma_j\widetilde{Z}_j+\gamma_t Z=\sigma_t\widetilde{Z}_t$, where $\widetilde{Z}_t\sim\mathcal{N}(0,1)$ and σ_t is defined in (7).

In particular, for $\phi_b(\mathbf{u}_i^t)=(b_i^t)^2$, we get $\phi_b(\widetilde{\mathbf{u}}_i^t)=\Big(\sum_{j=0}^{t-1}\beta_j b_i^j+\gamma_t Z\Big)^2$ because $\widetilde{\mathbf{u}}_i^t=(b_i^0,\dots,b_i^{t-1},\sum_{j=0}^{t-1}\beta_j b_i^j+\gamma_t Z,w_i)$. Combining (43) and (44),

$$\begin{aligned}\lim_{n\to\infty}\langle\mathbf{b}^t,\mathbf{b}^t\rangle&=\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\phi_b(\mathbf{u}_i^t)\\ &\overset{d}{=}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\phi_b(\widetilde{\mathbf{u}}_i^t)\overset{a.s.}{=}\mathbb{E}\Big[\Big(\sum_{j=0}^{t-1}\beta_j\sigma_j\widetilde{Z}_j+\gamma_t Z\Big)^2\Big].\end{aligned} \tag{45}$$

Using (14), $\lim_{n\to\infty}\langle\mathbf{b}^t,\mathbf{b}^t\rangle\overset{a.s.}{=}\lim_{N\to\infty}\frac{\langle\mathbf{q}^t,\mathbf{q}^t\rangle}{\rho}=\sigma_t^2$, where the last equality holds because by the induction hypothesis (17) for $\phi_h(\mathbf{v}_i^{t-1})=f_t^2(h_i^t,x_{0i})$ in (17), $\frac{1}{\rho}\lim_{N\to\infty}\langle\mathbf{q}^t,\mathbf{q}^t\rangle\overset{a.s.}{=}\frac{1}{\rho}\mathbb{E}[f_t^2(\tau_{t-1}Z,X_0)]=\sigma_t^2$. Hence, $\mathbb{E}\left[\Big(\sum_{j=0}^{t-1}\beta_j\sigma_j\widetilde{Z}_j+\gamma_t Z\Big)^2\right]\overset{a.s.}{=}\sigma_t^2$, implying $\sum_{j=0}^{t-1}\beta_j\sigma_j\widetilde{Z}_j+\gamma_t Z=\sigma_t\widetilde{Z}_t$ due to (45), verifying that $\lim_{n\to\infty}X_2^t\overset{a.s.}{=}0$, which completes the proof of (16).

B. Step 4: We show a), b), c), and d) of Theorem 1 conditioning on $\mathcal{F}_{t+1,t} = \{\mathbf{b}^0,\ldots,\mathbf{b}^t,\mathbf{m}^0,\ldots,\mathbf{m}^t,\mathbf{h}^1,\ldots,\mathbf{h}^t,\mathbf{q}^0,\ldots,\mathbf{q}^t,\mathbf{x}_0,\mathbf{w}\}$.

The proof of Step 4 is similar to the proof of Step 3. Thus, we only present the features that are unique in Step 4.

a) Similar to Step 3a), using Proposition 3 to characterize $\mathbf{A}|_{\mathcal{F}_{t+1,t}}$ and following the same procedure as in [27, Lemma 1a], the $\mathbf{h}^{t+1}|_{\mathcal{F}_{t+1,t}}$ is

$$\mathbf{h}^{t+1}|_{\mathcal{F}_{t+1,t}} \overset{d}{=} \sum_{j=0}^{t-1}\zeta_j\mathbf{h}^{j+1} + \widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t - \mathbf{P}_{\mathbf{Q}_{t+1}}\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t + \mathbf{Q}_t\overrightarrow{\mathbf{o}}_t(1). \tag{46}$$

By Proposition 7 in Appendix F, $\lim_{N\to\infty}\mathbf{P}_{\mathbf{Q}_{t+1}}\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t \overset{a.s.}{=} \mathbf{0}_N$. Similar to Step 3a), we verify that $\lim_{N\to\infty}\mathbf{Q}_t\overrightarrow{\mathbf{o}}_t(1) \overset{a.s.}{=} \mathbf{0}_N$ by characterizing (i) the expectation of the empirical distribution $\widehat{\mathbf{Q}_t\overrightarrow{\mathbf{o}}_t(1)}$ is bounded as $\lim_{N\to\infty}|\langle\mathbf{Q}_t\overrightarrow{\mathbf{o}}_t(1)\rangle| \leq \lim_{N\to\infty}|o(1)|\sum_{j=0}^{t-1}\frac{1}{N}\sum_{i=1}^N|q_i^j| \overset{a.s.}{=} 0$ and (ii) the empirical variance of $\widehat{\mathbf{Q}_t\overrightarrow{\mathbf{o}}_t(1)}$ is bounded an converges to $\lim_{N\to\infty}\langle\mathbf{Q}_t\overrightarrow{\mathbf{o}}_t(1)\rangle_2 \leq \lim_{N\to\infty}[o(1)]^2 t\sum_{j=0}^{t-1}\langle\mathbf{q}^j,\mathbf{q}^j\rangle \overset{a.s.}{=} 0$. Therefore, using $\lim_{N\to\infty}\mathbf{Q}_t\overrightarrow{\mathbf{o}}_t(1) \overset{a.s.}{=} \mathbf{0}_N$, we get

$$\mathbf{h}^{t+1}|_{\mathcal{F}_{t+1,t}} \overset{d}{\Longrightarrow} \sum_{j=0}^{t-1}\zeta_j\mathbf{h}^{j+1} + \widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t, \tag{47}$$

which completes the proof of (11).

b) Using (16) for $\phi_b(b_i^t,w_i) = |g_t(b_i^t,w_i)|^{2+2\alpha}$, we get $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n|m_i^t|^{2+2\alpha} \overset{a.s.}{=} \mathbb{E}[|g_t(\sigma_t Z_t,W)|^{2+2\alpha}] < \infty$. Because $\sum_{i=1}^n|m_{\perp i}^t|^{2+2\alpha} < \sum_{i=1}^n|m_i^t|^{2+2\alpha}$, the following holds $\limsup_{n\to\infty}\sum_{i=1}^n|m_{\perp i}^t|^{2+2\alpha} < \infty$, which concludes (13).

c) For $t_1 < t$ and $t_2 = t$, we have $\lim_{N\to\infty}\langle\mathbf{h}^{t_1+1},\mathbf{h}^{t+1}\rangle \overset{d}{=} \lim_{N\to\infty}\sum_{j=0}^{t-1}\zeta_j\langle\mathbf{h}^{t_1+1},\mathbf{h}^j\rangle + \lim_{N\to\infty}\langle\mathbf{h}^{t_1+1},\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t\rangle$ due to (47), resulting in

$$\lim_{N\to\infty}\langle\mathbf{h}^{t_1+1},\mathbf{h}^{t+1}\rangle \overset{a.s.}{=} \sum_{j=0}^{t-1}\zeta_j\lim_{n\to\infty}\langle\mathbf{m}^{t_1},\mathbf{m}^j\rangle + \lim_{N\to\infty}\frac{\mathbf{m}_\perp^{t^*}\widetilde{\mathbf{A}}\mathbf{h}^{t_1+1}}{N}, \tag{48}$$

where (48) is by the induction hypothesis (15). Note that $\frac{\mathbf{m}_\perp^{t^*}}{\|\mathbf{m}_\perp^t\|_2}\widetilde{\mathbf{A}}\frac{\mathbf{h}^{t_1+1}}{\|\mathbf{h}^{t_1+1}\|_2} \overset{d}{=} \frac{Z}{\sqrt{n}}$ due to Proposition 6. The second term in (48) is represented as

$$\lim_{N\to\infty}\frac{\mathbf{m}_\perp^{t^*}\widetilde{\mathbf{A}}\mathbf{h}^{t_1+1}}{N} \overset{d}{=} \lim_{N\to\infty}\frac{\|\mathbf{m}_\perp^t\|_2}{\sqrt{n}}\frac{\|\mathbf{h}^{t_1+1}\|_2}{\sqrt{N}}\frac{\sqrt{n}}{\sqrt{N}}\frac{Z}{\sqrt{n}},$$
$$= \sqrt{\rho}\lim_{N\to\infty}\sqrt{\langle\mathbf{m}_\perp^t,\mathbf{m}_\perp^t\rangle\langle\mathbf{h}^{t_1+1},\mathbf{h}^{t_1+1}\rangle}\frac{Z}{\sqrt{n}} \overset{a.s.}{=} 0, \tag{49}$$

Substituting (49) into (48) yields

$$\lim_{N\to\infty}\langle\mathbf{h}^{t_1+1},\mathbf{h}^{t+1}\rangle = \lim_{n\to\infty}\langle\mathbf{m}^{t_1},\sum_{j=0}^{t-1}\zeta_j\mathbf{m}^j\rangle = \lim_{n\to\infty}\langle\mathbf{m}^{t_1},\mathbf{m}_{||}^t\rangle,$$
$$= \lim_{n\to\infty}\langle\mathbf{m}^{t_1},\mathbf{m}_{||}^t\rangle + \lim_{n\to\infty}\langle\mathbf{m}^{t_1},\mathbf{m}_\perp^t\rangle = \lim_{n\to\infty}\langle\mathbf{m}^{t_1},\mathbf{m}^t\rangle,$$

concluding (15) when $t_1 < t$ and $t_2 = t$.

For $t_1 = t_2 = t$, by (47),

$$\lim_{N\to\infty}\langle\mathbf{h}^{t+1},\mathbf{h}^{t+1}\rangle \overset{d}{=} \sum_{i,j=0}^{t-1}\zeta_i\zeta_j\lim_{N\to\infty}\langle\mathbf{h}^{i+1},\mathbf{h}^{j+1}\rangle$$
$$+ 2\sum_{i=0}^{t-1}\zeta_i\lim_{N\to\infty}\langle\mathbf{h}^{i+1},\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t\rangle + \lim_{N\to\infty}\langle\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t,\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t\rangle. \tag{50}$$

Then, by (49), the following holds

$$\lim_{N\to\infty}\langle\mathbf{h}^{t+1},\mathbf{h}^{t+1}\rangle \overset{a.s.}{=} \sum_{i,j=0}^{t-1}\zeta_i\zeta_j\lim_{N\to\infty}\langle\mathbf{h}^{i+1},\mathbf{h}^{j+1}\rangle + \lim_{N\to\infty}\langle\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t,\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t\rangle. \tag{51}$$

By Proposition 5, the empirical distribution of $\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t$ converges to $\widehat{\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t} \overset{d}{\Longrightarrow} \mathcal{N}(0,\lim_{n\to\infty}\langle\mathbf{m}_\perp^t,\mathbf{m}_\perp^t\rangle)$. Hence, the second moment of $\widehat{\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t}$ converges to

$$\lim_{N\to\infty}\langle\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t,\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t\rangle \overset{a.s.}{=} \lim_{n\to\infty}\langle\mathbf{m}_\perp^t,\mathbf{m}_\perp^t\rangle. \tag{52}$$

Substituting (52) into (51) leads to $\lim_{N\to\infty}\langle\mathbf{h}^{t+1},\mathbf{h}^{t+1}\rangle \overset{a.s.}{=} \sum_{i,j=0}^{t-1}\zeta_i\zeta_j\lim_{N\to\infty}\langle\mathbf{h}^{i+1},\mathbf{h}^{j+1}\rangle + \lim_{n\to\infty}\langle\mathbf{m}_\perp^t,\mathbf{m}_\perp^t\rangle$, implying

$$\lim_{N\to\infty}\langle\mathbf{h}^{t+1},\mathbf{h}^{t+1}\rangle \overset{a.s.}{=} \sum_{i,j=0}^{t-1}\zeta_i\zeta_j\lim_{n\to\infty}\langle\mathbf{m}^i,\mathbf{m}^j\rangle + \lim_{n\to\infty}\langle\mathbf{m}_\perp^t,\mathbf{m}_\perp^t\rangle,$$
$$= \lim_{n\to\infty}\langle\mathbf{m}_{||}^t,\mathbf{m}_{||}^t\rangle + \lim_{n\to\infty}\langle\mathbf{m}_\perp^t,\mathbf{m}_\perp^t\rangle = \lim_{n\to\infty}\langle\mathbf{m}^t,\mathbf{m}^t\rangle.$$

Therefore, (15) also holds for $t_1 = t_2 = t$, which completes the proof.

d) Defining $\lim_{n\to\infty}\langle\mathbf{m}_\perp^t,\mathbf{m}_\perp^t\rangle \overset{a.s.}{=} \Gamma_t^2$, we can write

$$\widehat{\widetilde{\mathbf{A}}^*\mathbf{m}_\perp^t} \overset{d}{\Longrightarrow} \mathcal{N}(0,\Gamma_t^2). \tag{53}$$

Using (53) and (47), the following convergence holds

$$h_i^{t+1}|_{\mathcal{F}_{t+1,t}} \overset{d}{\Longrightarrow} \sum_{j=0}^{t-1}\zeta_j h_i^{j+1} + \Gamma_t Z, \tag{54}$$

where $Z \sim \mathcal{N}(0,1)$. Similar to Step 3d), we can write, using (54), $\mathbf{v}_i^t \overset{d}{\Longrightarrow} \widetilde{\mathbf{v}}_i^t$, where $\mathbf{v}_i^t = (h_i^1,..,h_i^{t+1},x_{0i})$ and $\widetilde{\mathbf{v}}_i^t = (h_i^1,\ldots,h_i^t,\sum_{j=0}^{t-1}\zeta_j h_i^{j+1} + \Gamma_t Z, x_{0i})$. Hence, to prove (17) we first claim that $\left|\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\phi_h(\widetilde{\mathbf{v}}_i^t) - \mathbb{E}\left[\phi_h(\tau_0 Z_0,\ldots,\tau_t Z_t,X_0)\right]\right| \overset{a.s.}{=} 0$. Similar to Step 2d), using the triangular inequality, we verify that $\lim_{N\to\infty}Y_1^t \overset{a.s.}{=} 0$ and $\lim_{N\to\infty}Y_2^t \overset{a.s.}{=} 0$, where $Y_1^t = \left|\frac{1}{N}\sum_{i=1}^N\left(\phi_h(\widetilde{\mathbf{v}}_i^t) - \widetilde{\phi}_h(\mathbf{v}_i^{t-1})]\right)\right|$ and $Y_2^t = \left|\frac{1}{N}\sum_{i=1}^N\widetilde{\phi}_h(\mathbf{v}_i^{t-1}) - \mathbb{E}\left[\phi_h(\tau_0 Z_0,\ldots,\tau_t Z_t,X_0)\right]\right|$, and $\widetilde{\phi}_h(\mathbf{v}_i^{t-1}) = \mathbb{E}_Z[\phi_h(\widetilde{\mathbf{v}}_i^t)]$, $\forall i$.

First, showing $\lim_{N\to\infty}Y_1^t \overset{a.s.}{=} 0$ is of interest. By (5), $|\phi_h(\widetilde{\mathbf{v}}_i^t)| \leq d_1^t\exp\left(d_2^t\left(\sum_{j=0}^{t-1}|h_i^{j+1}|^\lambda + \left|\sum_{j=0}^{t-1}\zeta_j h_i^{j+1} + \Gamma_t Z_i\right|^\lambda + |x_{0i}|^\lambda\right)\right)$, where $d_1^t > 0$, $d_2^t > 0$, and $1 \leq \lambda < 2$ are constants. Using the inequality $\|\mathbf{x}\|_1^\lambda \leq (t+1)^{\lambda-1}\|\mathbf{x}\|_\lambda^\lambda$ for

$\mathbf{x} \in \mathbb{R}^{(t+1)\times 1}$, we get $|\phi_h(\widetilde{\mathbf{v}}_i^t)| \le d_1^t \exp\Big(d_2^t\Big(\sum_{j=0}^{t-1}(1+(t+1)^{\lambda-1}|\zeta_j|^\lambda)|h_i^{j+1}|^\lambda + (t+1)^{\lambda-1}|\Gamma_t|^\lambda|Z_i|^\lambda + |x_{0i}|^\lambda\Big)\Big)$. Hence, $\mathbb{E}_Z\Big[|\phi_b(\widetilde{\mathbf{v}}_i^t)|^{2+\kappa}\Big] \le d_5^t \exp\Big[d_4^t\Big(\sum_{j=0}^{t-1}|h_i^{j+1}|^\lambda + |x_{0i}|^\lambda\Big)\Big]$, where $0 < \kappa < 1$, $d_4^t = d_2^t(2+\kappa)\max\Big\{1+(t+1)^{\lambda-1}|\alpha_0|^\lambda, \ldots, 1+(t+1)^{\lambda-1}|\alpha_{t-1}|^\lambda, (t+1)^{\lambda-1}|\Gamma_t|^\lambda\Big\}$, and $d_5^t = (d_1^t)^{2+\kappa}\mathbb{E}_Z\Big[\exp(d_4^t|Z_i|^\lambda)\Big]$ are constants. Define $Y_{N,i}^t = \phi_h(\widetilde{\mathbf{v}}_i^t) - \mathbb{E}_Z[\phi_h(\widetilde{\mathbf{v}}_i^t)]$, $\forall i$. To prove the convergence $\lim_{n\to\infty} Y_1^t \overset{a.s.}{=} 0$, we will show that $\{Y_{N,i}^t\}_{i=1}^N$ satisfy the condition in Lemma 3 in Appendix F. Indeed, the $\mathbb{E}_Z[|Y_{N,i}^t|^{2+\kappa}]$ is upper bounded as follows.

$$\begin{aligned}
\mathbb{E}_Z[|Y_{N,i}^t|^{2+\kappa}] &\le 2^{1+\kappa}\Big(\mathbb{E}_Z[|\phi_h(\widetilde{\mathbf{v}}_i^t)|^{2+\kappa}] + |\mathbb{E}_Z[\phi_h(\widetilde{\mathbf{v}}_i^t)]|^{2+\kappa}\Big),\\
&\le 2^{2+\kappa}\mathbb{E}_Z[|\phi_h(\widetilde{\mathbf{v}}_i^t)|^{2+\kappa}]\\
&\le d_6^t \exp\Big(d_4^t\Big(\sum_{j=0}^{t-1}|\zeta_j h_i^{j+1}|^\lambda + |x_{0i}|^\lambda\Big)\Big),\\
&\triangleq \quad \psi_h(\mathbf{v}_i^{t-1}).
\end{aligned} \tag{55a}$$

Then, $\psi_h(\mathbf{v}_i^{t-1})$ is a controlled function. From (55a), we get, for N is sufficiently large,

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{n}\mathbb{E}_Z[|Y_{N,i}^t|^{2+\kappa}] \quad &\le \quad \frac{1}{N}\sum_{i=1}^{n}\psi_h(\mathbf{v}_i^{t-1}),\\
&\overset{a.s.}{=} \quad \mathbb{E}[\psi_h(\tau_0 Z_0,\ldots,\tau_{t-1}Z_{t-1},X_0)],\\
&< \quad cN^{\kappa/2},
\end{aligned} \tag{56a}$$

where c is a positive constant and (56a) holds because $\mathbb{E}[\psi_b(\sigma_0\widetilde{Z}_0,\ldots,\sigma_{t-1}\widetilde{Z}_{t-1},W)] = d_7^t < \infty$ and there exists N_t, a positive constant, such that $d_7^t < cN^{\kappa/2}$ for $N > N_t$. Using Lemma 3 in Appendix F, $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N Y_{N,i}^t \overset{a.s.}{=} 0$, implying $\lim_{N\to\infty} Y_1^t \overset{a.s.}{=} 0$.

We are now ready to verify the convergence $\lim_{N\to\infty} Y_2^t \overset{a.s.}{=} 0$. Applying the induction hypothesis (17) for $\widetilde{\phi}_b(\mathbf{v}_i^{t-1})$ gives

$$\begin{aligned}
&\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\widetilde{\phi}_h(\mathbf{v}_i^{t-1}) \overset{a.s.}{=} \mathbb{E}[\widetilde{\phi}_h(\tau_0 Z_0,\ldots,\tau_{t-1}Z_{t-1},X_0)],\\
&= \mathbb{E}\Big[\mathbb{E}_Z[\phi_h(\tau_0 Z_0,\ldots,\tau_{t-1}Z_{t-1},\sum_{j=0}^{t-1}\zeta_j\tau_j Z_j + \Gamma_t Z, X_0)]\Big],\\
&= \mathbb{E}\Big[\phi_h(\tau_0 Z_0,\ldots,\tau_{t-1}Z_{t-1},\sum_{j=0}^{t-1}\zeta_j\tau_j Z_j + \Gamma_t Z, X_0)\Big].
\end{aligned}$$

Therefore, showing $\lim_{N\to\infty} Y_2^t \overset{a.s.}{=} 0$ is equivalent to proving $\sum_{j=0}^{t-1}\zeta_j\tau_j Z_j + \Gamma_t Z = \tau_t Z_t$, where $Z_t \sim \mathcal{N}(0,1)$ and τ_t is defined in (7). Similar to the proof of $\lim_{n\to\infty} X_2^t \overset{a.s.}{=} 0$ in Step 3d), setting $\phi_h(\mathbf{v}_i^t) = (h_i^t)^2$, i.e., $\phi_h(\widetilde{\mathbf{v}}_i^t) = \Big(\sum_{j=0}^{t-1}\zeta_j h_i^{j+1} + \Gamma_t Z\Big)^2$, we get

$$\begin{aligned}
&\lim_{N\to\infty}\langle\mathbf{h}^{t+1},\mathbf{h}^{t+1}\rangle = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\phi_h(\mathbf{v}_i^t)\\
&\overset{d}{=} \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\phi_h(\widetilde{\mathbf{v}}_i^t) \overset{a.s.}{=} \mathbb{E}\Big[\Big(\sum_{j=0}^{t-1}\zeta_j\tau_j Z_j + \Gamma_t Z\Big)^2\Big].
\end{aligned}$$

Using (15), we get $\lim_{N\to\infty}\langle\mathbf{h}^{t+1},\mathbf{h}^{t+1}\rangle \overset{a.s.}{=} \lim_{n\to\infty}\langle\mathbf{m}^t,\mathbf{m}^t\rangle = \tau_t^2$, where the last equality holds by the induction hypothesis (16) for $\phi_b(\mathbf{u}_i^t) = g_t^2(b_i^t, w_i)$, resulting in $\lim_{n\to\infty}\langle\mathbf{m}^t,\mathbf{m}^t\rangle \overset{a.s.}{=} \mathbb{E}[g_t^2(\sigma_t\widetilde{Z}_t, W)] = \tau_t^2$. Hence, $\mathbb{E}\Big[\Big(\sum_{j=0}^{t-1}\zeta_j\tau_j Z_j + \Gamma_t Z\Big)^2\Big] \overset{a.s.}{=} \tau_t^2$, which implies $\sum_{j=0}^{t-1}\zeta_j\tau_j Z_j + \Gamma_t Z = \tau_t Z_t$. Thus, it is verified that $\lim_{N\to\infty} Y_2^t \overset{a.s.}{=} 0$, which completes the proof of (17).