Convergence Rate for Stochastic Consensus Algorithms with Time-Varying Noise Statistics: Asymptotic Normality

Minyi Huang

Abstract—This paper studies consensus seeking over noisy networks with time-varying noise statistics. Stochastic approximation type algorithms can ensure consensus in mean square and with probability one. For performance evaluation, we examine the long term behavior of the approximation error which consists of two naturally defined components. We show that the two components and their sum are each asymptotically normal after being normalized by the square root of time. This, in turn, characterizes the convergence rate of the algorithm. We also give the asymptotic formula for the scaled error covariances.

I. INTRODUCTION

Recent technological advances have made it possible to build large distributed systems in which different constituent components or agents may cooperatively perform complex tasks. In these systems, consensus protocols provide a basic mechanism for the agents to agree on key information concerning system operation. Consensus problems and various closely related formulations have been intensively investigated for multi-agent systems [14], [17], [20]. A comprehensive survey on recent research can be found in [18], [22]. While most existing consensus algorithms have assumed exact state averaging, which in general necessitates perfect state exchange, recently, there is an increasing attention on additional algorithms involving constant or lowered bounded state exchange, recently, there is an increasing attention on

For convenience of exposition, the two names, agent and node, will be used alternatively. The agent $A_k$ (resp., node $k$) is a neighbor of $A_i$ (resp., node $i$) if $(k, i) \in \mathcal{E}$ where $k \neq i$. Denote the neighbor set $\mathcal{N}_i = \{k | (k, i) \in \mathcal{E}\} \subset \mathcal{N}$. being a vector in span$\{1_n\}$ where $1_n = [1, \cdots, 1]^T$. We give a linear decomposition of the error into two parts where the first characterizes the oscillation within span$\{1_n\}$. We will show that after normalization, the distributions of these two error components each converge weakly to a normal distribution; this, in turn, characterizes the convergence rate of the algorithm. Some preliminary asymptotic normality analysis has been developed in [12] under i.i.d. noise assumptions. In this paper, we consider much more general noise sequences without assuming independence. We allow time-varying noise statistics by only specifying certain long term average behavior of the covariances and conditional covariances; such assumptions are applicable to models occasionally experiencing burst receiver noises while the previous i.i.d. assumptions are not. Our proof adopts the classic central limit theorem approach [23], [16], [5], [7], but some new techniques need to be developed in order to deal with the time-varying (non-convergent) noise covariances. For general asymptotic error analysis in stochastic approximation, the reader is referred to [2], [5], [15], [16].

The organization of the paper is as follows. In Section II, we formulate the stochastic consensus problem and review our previous convergence results. The main theorem on asymptotic normality is stated in Section III. Section IV contains simulations and Section V concludes the paper.

II. THE STOCHASTIC CONSENSUS PROBLEM

Consider $n$ agents distributed according to a digraph $G = (\mathcal{N}, \mathcal{E})$ consisting of a set of nodes $\mathcal{N} = \{1, \cdots, n\}$ and a set of directed edges $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$. For brevity, a directed edge will be simply called an edge. An edge from node $i$ to node $j$ is denoted as an ordered pair $(i, j)$ where $i \neq j$ (so there is no edge between a node and itself). A directed path (from $i_1$ to $i_l$) consists of a sequence of nodes $i_1, i_2, \cdots, i_l$, $l \geq 2$, such that $(i_k, i_{k+1}) \in \mathcal{E}$ for $k = 1, \cdots, l - 1$. We say node $i$ is connected to node $j$ ($\neq i$) if there exists a directed path from $i$ to $j$. The digraph $G$ is said to be strongly connected if each node $i$ is connected to any other node $j$ by a directed path. A directed tree is a digraph where each node, except the root node, has exactly one parent node. Hence, the root node is connected to any other node by a directed path. The digraph $G$ contains a spanning tree $G_t = (\mathcal{N}_t, \mathcal{E}_t)$ if $G_t$ is a directed tree such that $\mathcal{N}_t = \mathcal{N}$ and $\mathcal{E}_t \subset \mathcal{E}$. A strongly connected digraph always contains a spanning tree.
A. The Measurement Model

For agent $A_i$, denote its state at time $t$ by $x_i^t \in \mathbb{R}$, where $t \in \mathbb{Z}^+ = \{0, 1, 2, \cdots \}$. Each $A_i$ receives noisy measurements of the states of its neighbors if $\mathcal{N}_i \neq \emptyset$, where $\emptyset$ denotes the empty set. Denote the measurement by $A_i$ of $A_k$’s state by

$$y_{ik}^t = x_i^t + w_{ik}^t, \quad t \in \mathbb{Z}^+, \quad k \in \mathcal{N}_i \neq \emptyset,$$

where $w_{ik}^t \in \mathbb{R}$ is the additive noise; see Fig. 1. The underlying probability space is denoted by $(\Omega, \mathcal{F}, P)$. We call $y_{ik}^t$ the observation of the state of $A_k$ obtained by $A_i$, and assume each $A_i$ knows its own state $x_i^t$ exactly. For similar measurement modeling, see [9], [4], [24]. We introduce the assumption:

(A1) The digraph $G = (\mathcal{N}, \mathcal{E})$ contains a spanning tree. □

For each $t \in \mathbb{Z}^+$, the set of noises $\{w_{ik}^t, i \in \mathcal{N} \text{ and } k \in \mathcal{N}_i \neq \emptyset\}$ is listed into a vector $w_t$ in which the position of $w_{ik}^t$ depends only on $(i, k)$ and does not change with $t$. Define the state vector

$$x_t = [x_1^t, \cdots, x_i^t]^T, \quad t \geq 0.$$  \hspace{1cm} (2)

Denote the $\sigma$-algebras as follows: $\mathcal{F}_t = \sigma(x_0^n, w_k, k \leq t)$ (i.e., the set of all events induced by these random variables) for $t \geq 0$. Then $w_t$ is adapted to $\mathcal{F}_t$ and $\mathcal{F}_t \subset \mathcal{F}_{t+1}$. We introduce the assumption:

(A2) The sequence $\{w_t, t \in \mathbb{Z}^+\}$ constitutes a sequence of martingale differences with a uniformly bounded second order moment, i.e., $w_t$ is adapted to $\mathcal{F}_t$, $E[|w_t|^2 | \mathcal{F}_{t-1}] = 0$ for $t \geq 0$ and $\sup_{t \geq 0} E[|w_t|^2] < \infty$. In addition, $E[|x_0|^2] < \infty$. □

The following assumption with independent noises holds as a special case of (A2).

(A2') The sequence $\{w_{ik}^t, i \in \mathcal{N} \text{ and } k \in \mathcal{N}_i \neq \emptyset\}$ are independent with respect to the indices $i, k, t$ and also independent of $x_0$, and each $w_{ik}^t$ has zero mean and variance $Q_{ik}^t$. In addition, $E[|x_0|^2] < \infty$ and $\sup_{i, k, t} Q_{ik}^t < \infty$. □

B. The Stochastic Approximation Algorithm

The state of each agent is updated by the rule

$$x_{i+1}^t = (1 - a_i b_{ii})x_i^t + a_i \sum_{k \in \mathcal{N}_i} b_{ik} y_{ik}^t, \quad t \geq 0,$$

where $i \in \mathcal{N}$, $a_i > 0$ and the parameters $b_{ij}$ will be specified subsequently. Throughout our analysis, we adopt the convention: $\sum_{\emptyset} = 0$ regardless of the summation.

Case 1. If $\mathcal{N}_i \neq \emptyset$, we take:

$$b_{ii} > 0, \quad b_{ik} = 0, \quad b_{ii} = \sum_{k \in \mathcal{N}_i} b_{ik}, \quad i \in \mathcal{N}_i.$$

Case 2. If $\mathcal{N}_i = \emptyset$, we define $b_{ik} \equiv 0$ for all $k \in \mathcal{N}$ and the state of agent $i$ is fixed as its initial value: $x_{i+1}^t \equiv x_i^t$. Such a situation arises in leader following where the leader’s state is fixed as a constant at all times.

Define the matrix

$$B = \begin{pmatrix}
-b_{11} & b_{12} & \cdots & b_{1n} \\
b_{21} & -b_{22} & \cdots & b_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
b_{n1} & b_{n2} & \cdots & -b_{nn}
\end{pmatrix},$$

where $\Phi^{-1} B \Phi = \begin{pmatrix} 0 \\ \hat{B} \end{pmatrix}$, \hspace{1cm} (8)

where $\Phi = (1, \hat{\Phi})$ is nonsingular and

$$\Phi^{-1} B \Phi = \begin{pmatrix} 0 \\ \hat{B} \end{pmatrix},$$

where $\hat{B} \in \mathbb{R}^{(n-1) \times (n-1)}$ is Hurwitz. □

(ii) Letting $z_t = [z_1^t, \cdots, z_n^t]^T = \Phi^{-1} x_t$ and

$$v_t = [v_1^t, \cdots, v_n^t]^T = \Phi^{-1} w_t,$$

we have the relationship

$$z_{i+1}^t = z_i^t + a_i v_i^t, \quad t \geq 0,$$

$$\tilde{z}_{i+1}^t = (I + a_i \hat{B}) \tilde{z}_i + a_i \tilde{v}_i, \quad t \geq 0,$$

where $\tilde{v}_i$ is a feedback of $v_i$.

C. Preliminary Decomposition Results

We introduce the following class of matrices in $\mathbb{R}^{n \times (n-1)}$

$$\mathcal{C}(B) = \left\{ \phi \in \{\mathbb{R}^n\} | \text{span} \{\phi\} = \text{span} \{B\} \right\},$$

where $\text{span} \{\phi\}$ denotes the subspace spanned by the columns of $\phi$. Under (A1), rank$(B) = n - 1$, and accordingly, each $\phi \in \mathcal{C}(B)$ has rank $n - 1$ (see [12]).

Lemma 4: [12] Assuming (A1), for (6) we have:

(i) For $I_n = [1, \cdots, 1]^T$ and any given $\tilde{\phi} \in \mathcal{C}(B)$, the matrix $\Phi = (I_n, \tilde{\phi})$ is nonsingular and

$$\Phi^{-1} B \Phi = \begin{pmatrix} 0 \\ \hat{B} \end{pmatrix},$$

where $\hat{B} \in \mathbb{R}^{(n-1) \times (n-1)}$ is Hurwitz. □

(ii) Letting $z_t = [z_1^t, \cdots, z_n^t]^T = \Phi^{-1} x_t$ and

$$v_t = [v_1^t, \cdots, v_n^t]^T = \Phi^{-1} w_t,$$

we have the relationship

$$z_{i+1}^t = z_i^t + a_i v_i^t, \quad t \geq 0,$$

$$\tilde{z}_{i+1}^t = (I + a_i \hat{B}) \tilde{z}_i + a_i \tilde{v}_i, \quad t \geq 0,$$

where $\tilde{v}_i$ is a feedback of $v_i$. □
where $z_t = [z_t^1, \cdots, z_t^n]^T$ and $\tilde{v}_t = [v_t^1, \cdots, v_t^n]^T$.

In fact, the first row of $\Phi^{-1}$ is given as a unique nonnegative vector $\pi = [\pi_1, \cdots, \pi_n]$ satisfying $\pi B = 0$ and $\sum_{k=1}^n \pi_k = 1$. Under (A1)-(A3), $z_t$ converges to $z_{\infty} = [z_{\infty}^1, 0, \cdots, 0]^T$ in mean square and w.p.1 (see [12]).

III. MAIN RESULTS ON ASYMPTOTIC NORMALITY

For Theorem 3, denote the limit state vector by $x_{\infty} = [x_{\infty}^1, \cdots, x_{\infty}^n]^T = x_{\infty}^1 1_n$. Since $z_t^1 = \pi x_t$, we obtain $z_{\infty}^1 = \pi (x_{\infty}^1 1_n) = x_{\infty}^1$. By the relation $x_t = x_{\infty}^1 1_n + \Phi z_t - x_{\infty}^1 1_n$, we obtain the following decomposition:

$$x_t = x_{\infty}^1 1_n + (z_t^1 - z_{\infty}^1) 1_n + \bar{\phi} \tilde{z}_t.$$ (12)

Thus, under (A1)-(A3) the approximation error for $x_t$ is decomposed into two components $x_t^{c.a} = (z_t^1 - z_{\infty}^1) 1_n$ and $x_t^{c.b} = \bar{\phi} \tilde{z}_t$ to give

$$x_t = x_{\infty}^1 1_n + x_t^{c.a} + x_t^{c.b}. \quad (13)$$

Clearly, $x_t^{c.a} \in \text{span \{1}_n \}$ and $x_t^{c.b} \in \text{span \{B \}}$. Under (A1), using the property $\pi B = 0$, we may show that

$$\text{span \{1}_n \} \cap \text{span \{B \}} = \{0 \}, \quad \mathbb{R}^n = \text{span \{1}_n \} \oplus \text{span \{B \}}$$

where $\oplus$ denotes direct sum. Hence, $x_t - x_{\infty}^1 1_n$ has a unique representation as the sum of two vectors (independent of $\tilde{\phi}$) in $\text{span \{1}_n \}$ and $\text{span \{B \}}$, respectively. Alternatively, we may use the fact that $x_t, x_{\infty}^1 1_n$ and $z_t^1 = \pi x_t, z_{\infty}^1 = \pi x_{\infty}^1 1_n$ are all independent of $\tilde{\phi}$ to check that the decomposition in (13) does not depend on the choice of $\tilde{\phi} \in \mathscr{D}(B)$. For the case of leader following, we can show that $v_t^1 \equiv 0$ and $x_t^{c.a} \equiv 0$, and the asymptotic error analysis reduces to checking $x_t^{c.b}$.

We introduce some assumption related to $w_t$, and it will be convenient to give the condition based on $v_t$, which is defined via (5) and (9).

(A4) The sequence $\{v_t, t \in \mathbb{Z}^+\}$ constitutes vector random variables with zero mean and covariance $Q^\phi$ such that

$$\lim_{t \to \infty} (1/T) \sum_{k=1}^{T-k} Q^\phi_k = \bar{Q} \quad (14)$$

uniformly w.r.t. $k \geq 0$, and in addition

$$\lim_{t \to \infty} \left(1/T \right) \sum_{k=1}^{T-k} E \left[ E[v_t^1 \tilde{v}_t^1 | v_{t-1}^1, \cdots, v_{t-1-k}^1] - \bar{Q}^\phi \right] = 0. \quad \text{(15)}$$

$$\lim_{k \to \infty} \sup_{t \geq 0} |E[|v_t^1|^2 1_{|v_t^1| \geq k}] = 0. \quad \text{(16)}$$

Letting $\sigma_t^2 = E[|v_t|^2]$, then

$$\lim_{t \to \infty} \left(1/T \right) \sum_{k=1}^{T-k} \sigma_t^2 = \bar{\sigma}^2, \quad \text{(17)}$$

$$\lim_{t \to \infty} \left(1/T \right) \sum_{k=1}^{T-k} E \left[ E[|v_t|^2 | v_{t-1}^1, \cdots, v_{t-1-k}^1] - \sigma_t^2 \right] = 0, \quad \text{(18)}$$

both uniformly with respect to $k$. Finally

$$\lim_{k \to \infty} \sup_{t \geq 0} E[|v_t|^2 1_{|v_t| \geq k}] = 0. \quad \text{(19)}$$

For the special case $\{w_t, t \geq 0\}$ being an i.i.d. sequence with zero mean and finite covariance, (14)-(19) are satisfied. It must be noted that the validity of (A4) does not depend on the choice of $\tilde{\phi}$. More specifically, when a different $\tilde{\phi}$ is used, (A4) is still true as long as all the associated variances and conditional variances correspond to the new $\tilde{\phi}$.

(A3') The sequence $\{a_t, t \geq 0\}$ satisfies i) $a_t > 0$, ii) $\lim_{t \to \infty} (a_{t+1} - a_t^{-1}) = \alpha > 0$, iii) $\tilde{\phi} \in \mathcal{D} + \alpha r/2$ is Hurwitz.

It is evident that (A3') implies (A3). If a sequence of random variables $\{z_t, t \geq 0\}$ converges in distribution to a normal random variable $\xi_\infty$ with mean $\mu$ and covariance $\Sigma$, we denote $\xi_\infty \overset{d}{\rightarrow} N(\mu, \Sigma)$ and $\xi_\infty \overset{a.s.}{\rightarrow} \xi_\infty$.

Denote $D = \int_0^\infty e^{\bar{\phi} T} \tilde{D} e^{\tilde{\phi} T} dt$, and

$$D_a = \alpha^{-2} \tilde{\sigma}^2 1_n 1_n^T, \quad D_b = \alpha^{-1} \tilde{\phi} D \tilde{\phi}. \quad \text{(20)}$$

Theorem 5: Assuming (A1), (A2), (A3') and (A4) hold, we have i) $\sqrt{\tilde{\phi}} x_t^{c.a} \overset{d}{\rightarrow} N(0, D_a)$ and $\sqrt{\tilde{\phi}} x_t^{c.b} \overset{d}{\rightarrow} N(0, D_b)$ and ii) $\sqrt{\tilde{\phi}} (x_t^{c.a} + x_t^{c.b}) \overset{d}{\rightarrow} N(0, D_a + D_b).

To prove Theorem 5-i), it suffices to establish the two lemmas below. Their proofs are quite technical and are given in Appendix B. Theorem 5-ii) may be proved by first approximating $\sqrt{\tilde{\phi}} (x_t^{c.a} + x_t^{c.b})$ by the sum of finite terms of martingale differences (similar to the treatment in proving Lemma 7), and next carry out the asymptotic characteristic function estimation with $t \to \infty$.

Lemma 6: Under the assumptions of Theorem 5, $\sqrt{\tilde{\phi}} z_{t} \overset{a.s.}{\rightarrow} N(0, \alpha^{-1} D)$.

Lemma 7: Under the assumptions of Theorem 5, $\sqrt{\tilde{\phi}} (z_t - z_{\infty}^1) \overset{d}{\rightarrow} N(0, \alpha^{-2} \sigma^2)$.

Corollary 8: Under the assumptions of Theorem 5,

$$\lim_{t \to \infty} tE\{x_t^{c.a} (x_t^{c.a})^T\} = D_a = (\alpha^{-2} \tilde{\sigma}^2) 1_n 1_n^T, \quad \text{(21)}$$

$$\lim_{t \to \infty} tE\{x_t^{c.b} (x_t^{c.b})^T\} = D_b = \alpha^{-1} \tilde{\phi} D \tilde{\phi}^T, \quad \text{(22)}$$

$$\lim_{t \to \infty} tE\{(x_t^{c.a} + x_t^{c.b})(x_t^{c.a} + x_t^{c.b})^T\} = D_a + D_b. \quad \text{(23)}$$

Proof: We obtain (21) from a direct calculation of $\lim_{t \to \infty} t \sum_{k=1}^{\infty} \sigma_k^2$. By (B.3) and (B.7), we obtain (22). To show (23), we first take expectation to eliminate the two cross terms in the expansion by using the martingale difference property of the terms in the series representation of $x_t^{c.a}$ and $x_t^{c.b}$, and the right hand side follows from (21) and (22).

Again, we remark that both $(\alpha^{-2} \tilde{\sigma}^2) 1_n 1_n^T$ and $\alpha^{-1} \tilde{\phi} D \tilde{\phi}^T$ are independent of the particular choice of $\tilde{\phi} \in \mathcal{D}(B)$.

IV. SIMULATIONS

We consider a digraph shown in Fig. 2. The noises $\{w_t^1, w_t^{21}, w_t^{23}, w_t^{11}, t \geq 0\}$ are independent and satisfy: for $k = 0, 1, \cdots$ (i) if $20k \leq t \leq 20k + 17$, each has a uniform distribution on $[-0.2, 0.2]$ (with $\sigma_k^2 = 0.0133$); (ii) if $20k + 18 \leq t < 20(k + 1)$, each has a uniform distribution on $[-0.7, 0.7]$ (with $\sigma_k^2 = 0.1633$). Thus, the noise variances periodically reach a much higher level, which models burst receiver noises. The initial state vector is $x_{[t=0]} = [5, 4, 2]^T$. 


For algorithm (6), we take
\[
B = \begin{bmatrix}
-1 & 1 & 0 \\
0.5 & -1 & 0.5 \\
1 & 0 & -1
\end{bmatrix}, \quad \Phi = \begin{bmatrix}
1 & -1 & 1 \\
1 & 0.5 & -1 \\
1 & 1 & 0
\end{bmatrix}
\] (24)
and \(a_0 = 0.5, \ a_t = \frac{a_0}{t} \) for \( t \geq 1 \). The 3 eigenvalues of \( B \) are 0, \(-1.5 \pm 0.5i\). The first two columns in \( B \) are used to construct \( \Phi \). We express \( x_t^{e,b} \) by \( \tilde{z}_t \) and consequently by \( x_t \).

The asymptotic normality conclusion of Theorem 5 holds for this example since (A4) is satisfied and \( \bar{B} + \alpha I/2 \) is Hurwitz with eigenvalues \(-0.5 \pm 0.5i\) when the associated \( B \) is given by (24) and \( \alpha = 2 \). The convergence of \( x_t \) is shown in Fig. 3 which displays the first 200 iterates, and \( \{\sqrt{\tilde{x}_t^{e,b}}, t \geq 0\} \) is displayed in Fig. 4.

V. CONCLUSIONS

We have presented asymptotic normality results for the scaled error terms in stochastic consensus algorithms. Our analysis is applicable to average consensus based algorithms with additive noises, which amounts to imposing additional conditions (i.e. \( B \) has zero row and column sums, corresponding to averaging with balanced graphs [17]).

APPENDIX A: PRELIMINARY LEMMAS

We need some preliminary lemmas before proving Lemmas 6 and 7 in Appendix B.

Lemma 9: [23], [16], [5] Suppose \( \{\xi_{tk}, \ t \geq 1, 1 \leq k \leq t\} \) forms an array of martingale differences, i.e.,
\[
E[\xi_{tk}|\xi_1, \ldots, \xi_{t(k-1)}] = 0.
\]

Denote \( S_{tk} = E\xi_{tk}\xi_{tk}^T, \ R_{tk} = E(\xi_{tk}\xi_{tk}^T|\xi_{t1}, \ldots, \xi_{t(k-1)}), \ S_t = \sum_{k=1}^t S_{tk}, \ \text{and} \ \zeta_t = \sum_{k=1}^t \xi_{tk}. \) Assume
\[
\sup_{t \geq 1} \sum_{k=1}^t E|\xi_{tk}|^2 < \infty,
\]
\[
\lim_{t \to \infty} \sum_{k=1}^t E|S_{tk} - R_{tk}| = 0,
\]
\[
\lim_{t \to \infty} S_t = S,
\]
\[
\lim_{t \to \infty} \sum_{k=1}^t E|\xi_{tk}|^2 (|\xi_{tk}| > \varepsilon), \ \forall \varepsilon > 0.
\]

Then \( \zeta_t \sim N(0, S) \).

We give two lemmas without proof for reasons of space.

Lemma 10: Let \( \delta \in (0, 1) \) and suppose the sequence of nonnegative numbers \( b_1, t \geq 1 \) satisfies \( s_t = \sum_{i=1}^t b_i = o(t) \). Then \( \lim_{t \to \infty} t^{-\delta} \sum_{k=1}^t t^{-(1-\delta)} b_t = 0 \).

Lemma 11: Suppose the sequence of nonnegative numbers \( b_i, i \geq 1 \) satisfies \( \lim_{T \to \infty} (1/T) \sum_{i=k}^{k+T-1} b_i = 0 \) uniformly w.r.t. \( k \). Then \( \lim_{k \to \infty} k \sum_{j=k}^{\infty} j^{-2} b_j = 0 \).

APPENDIX B: PROOF OF ASYMPTOTIC NORMALITY

We shall use \( C > 0 \) to denote a generic constant which may vary from place to place.

Proof of Lemma 6: We have the recursion
\[
\tilde{z}_{t+1}/\sqrt{\bar{a}_{t+1}} = \left(\sqrt{\bar{a}_t}/\sqrt{\bar{a}_{t+1}}\right)(I + a_t\bar{B})\left(\tilde{z}_t/\sqrt{\bar{a}_t}\right)
+ \left(a_t/\sqrt{\bar{a}_t}\right)\tilde{v}_t.
\] (B.1)

It is straightforward to show [5]
\[
(\sqrt{\bar{a}_t}/\sqrt{\bar{a}_{t+1}})(I + a_t\bar{B}) = I + a_t\bar{B} + o(\bar{a}_t),
\]
\[
\delta I + a_t\bar{B}.
\] (B.2)

Denote \( \Pi_{l,i} = \Pi_{k=i+1}^l (I + a_k\bar{B}_k), \ l \geq i, \) where \( (I + a_{i+1}\bar{B}_{i+1}) \) appears as the most right term in the successive matrix product. Denote \( \Pi_{i,i} \equiv I \). By elementary estimates (see, e.g. [5]), it can be shown that \( |\Pi_{l,i} - \exp\left\{\sum_{k=i+1}^l a_k\right\}(I + \alpha I/2)| \to 0 \) uniformly with respect to \( l \) when \( i \to \infty \), and \( |\Pi_{l,i}| \leq C_{\Pi} \exp\left(-\eta\sum_{k=i+1}^l a_k\right) \), where \( C_{\Pi} \) and \( \eta \) are fixed constants.

By (B.1)-(B.2), we have \( \tilde{z}_{t+1}/\sqrt{\bar{a}_{t+1}} = (I + a_t\bar{B}_t)(\tilde{z}_t/\sqrt{\bar{a}_t}) + (a_t/\sqrt{\bar{a}_t})\tilde{v}_t \). Denote \( \tilde{z}_t = \tilde{z}_t/\sqrt{\bar{a}_t} \). Then
\[
\tilde{z}_{t+1} = \prod_{i=0}^{l} (1 + a_t\bar{B}_i)\tilde{z}_0 + \sum_{i=0}^{l} \Pi_{i,l}(a_i/\sqrt{\bar{a}_{i+1}})\tilde{v}_i.
\] (B.3)
Since $\lim_{t \to \infty} \prod_{k=0}^{t-1}(1 + a_k \bar{B}_k) = 0$, it suffices to show asymptotic normality of $\sum_{i=0}^{t} \Pi_{j,i}(\hat{\alpha}_i / \sqrt{\bar{a}_{i+1}})\bar{v}_i$. Define $\xi_k = \sum_{i=0}^{k-1} \Pi_{j,i}(\hat{\alpha}_i / \sqrt{\bar{a}_{i+1}})\bar{v}_i$. We verify conditions (A.2)-(A.3) in Lemma 9 with $0 \leq k \leq t$; then (A.1) follows from (A.3).

Denote the $\sigma$-algebra $\mathcal{F}_t = \mathcal{F}(\bar{v}_0, \ldots, \bar{v}_t)$. In fact,

$$\Delta_t \triangleq \sum_{k=0}^{t} E[S_{tk} - R_{tk}] \leq C \sum_{i=0}^{t} a_i \Pi_{j,i} E[|\bar{v}_i|^{\gamma} |\mathcal{F}_{i-1}] - E|\bar{v}_i|^{\gamma}].$$

Since there exist $0 < \kappa \leq \beta$ such that $\kappa(i+1)^{-1} \leq a_i \leq \beta(i+1)^{-1}, i \geq 0$, we have $a_i \Pi_{j,i} \leq C(i+1)^{-1} \exp(-2\kappa(i+1)^{k-1}) \leq C(i+1)^{-1+\delta}(1+1)^{\delta}$, where $\delta = 2\kappa$. We take a small $\kappa$ so that $\delta < 1$. Then the relation

$$\Delta_t \leq C(t+1)^{-\delta} \sum_{k=0}^{t} (k+1)^{-1+\delta} E|\bar{v}_i|^{\gamma} |\mathcal{F}_{k-1}] - E|\bar{v}_i|^{\gamma} |$$

combined with Lemma 10 gives $\lim_{t \to \infty} \sum_{k=0}^{t} E[S_{tk} - R_{tk}] = 0$.

Below we simply write $Q_t = \tilde{Q}_t$. Next, we verify (A.3) in Lemma 9. We have

$$\begin{aligned}
S_t &= \sum_{k=0}^{t} S_{tk} = \sum_{i=0}^{t} a_i^2 \Pi_{j,i}Q_i^T \\
&= \sum_{i=0}^{t} a_i \Pi_{j,i}Q_i^T + \sum_{i=0}^{t} (a_i^2 + o(a_i^2)) \Pi_{j,i}Q_i^T \\
&\leq S^{(1)} + S^{(2)}. 
\end{aligned}$$

Since $Q$ is a bounded sequence implied by (14), we have

$$|S^{(2)}| \leq C \sum_{i=0}^{t} a_i^2 \Pi_{j,i}^2 + C \sum_{i=0}^{t} a_i^2 \Pi_{j,i}^2 \leq S^{(2,a)} + S^{(2,b)}.$$ 

Since $\Pi_{j,i}$ is bounded, for any given $\xi > 0$, we can take a sufficiently large $T_1$ such that $\sup_{t \geq T_1} S^{(2,a)} \leq \xi$. On the other hand, $\lim_{t \to \infty} \Pi_{j,i} = 0$. Hence, $\lim_{t \to \infty} S^{(2)} = 0$.

We continue to determine the limit for $S^{(1)}$. By extending the treatment in [5] (pp. 125) to a sequence of time-varying covariances, we may use (B.4) to show that

$$\lim_{t \to \infty} |S^{(1)}| = \sum_{i=0}^{t} a_i \exp(B \sum_{k=1}^{t} a_k) \exp(B^T \sum_{k=1}^{t} a_k) = 0,$$

where $B = \tilde{B} + aT/2$. Subsequently, we need to show that the second term in the difference has a limit as $t \to \infty$.

Fix any $\xi > 0$. First, by (A4) we may take a large $T_0$ such that $\sup_{t \geq T_0} |(1/T_0) \sum_{i=0}^{t} \hat{Q}_i - \tilde{Q}| \leq \xi$. For the proof below, once $T_0$ is selected, it suffices to consider $t = kT_0$. Denote $M_{i,j} = \exp(B \sum_{k=i+1}^{j} a_k)$. The estimates below appear a bit technical. However, the basic idea is relatively simple. Intuitively, when $j$ is large, on the time window $[jT_0 + 1, (j+1)T_0]$, since $a_i$ varies slowly, the pair $(a_i, M_{i,j})$ appearing in $a_i M_{i,j} Q_i M_i^T$ may be frozen as at its value at the starting time $jT_0 + 1$, incurring only a small error. Let $K_0$ be fixed, we may verify that for all $K_0T_0 < t < j \leq i + T_0, |a_i M_{i,j} - a_{j,i} M_{j,i} Q_i M_i^T| \leq C \xi a_i^2$, where $C$ does not depend on $(T_0, K_0)$.

For $k \geq K_0 + 1$, we have

$$\begin{aligned}
\sum_{i=0}^{kT_0} a_i M_{i,j} Q_i M_i^T &= \sum_{i=0}^{kT_0} a_i M_{i,j} Q_i M_i^T + \sum_{j=K_0T_0 + 1}^{k-1} a_i M_{i,j} Q_i M_i^T \\
&\leq \sum_{j=K_0T_0 + 1}^{k-1} a_i^2 + C \xi 
\end{aligned}$$

where $t = kT_0$. Then we set $\hat{S}_t = \sum_{j=K_0T_0 + 1}^{kT_0} a_i M_{i,j} Q_i M_i^T$ and it is easy to show that

$$\begin{aligned}
|S_t - \sum_{j=K_0T_0 + 1}^{kT_0} a_i M_{i,j} Q_i M_i^T| &\leq C T_0 \sum_{i=K_0T_0 + 1}^{kT_0} a_i^2.
\end{aligned}$$

Hence

$$\begin{aligned}
\hat{S}_t &= \left\{ \sum_{j=K_0T_0 + 1}^{kT_0} a_i M_{i,j} Q_i M_i^T \right\} \leq C T_0 \sum_{i=K_0T_0 + 1}^{kT_0} a_i^2 + C \xi,
\end{aligned}$$

where $C$ does not depend on $T_0$ and we obtain the term $C \xi$ by using (14) and the fact $\sup_{k \geq 0} \sum_{i=0}^{k} a_i |M_{i,j} Q_i M_i^T| < \infty$.

By switching the index $jT_0 + 1$ back to $i$ in (B.5) and taking into account the error incurred, we obtain the estimate

$$\begin{aligned}
|S_t - \sum_{i=K_0T_0 + 1}^{kT_0} a_i M_{i,j} Q_i M_i^T| \leq C T_0 \sum_{i=K_0T_0 + 1}^{kT_0} a_i^2 + C \xi + o(1),
\end{aligned}$$

where $t = kT_0$ and $C$ does not depend on $(T_0, K_0)$. This gives

$$\lim_{t \to \infty} |S_t - \sum_{i=K_0T_0 + 1}^{kT_0} a_i M_{i,j} Q_i M_i^T| = 0,$$

where the second term in the difference has a limit with a standard integral representation $\int_{0}^{\infty} e^{B^T \tilde{Q}} \tilde{B}^T \tilde{Q} dt$ (see [16], [5]). Finally, we may verify (A.4) by elementary estimates. \qedhere

**Proof of Lemma 7:** We write $z_k^2 - z_k^\ast = - \sum_{j=k+1}^{\infty} a_j \nu_j$. Then

$$E|\hat{\delta}_k|^2 = O(1/k).$$

as $k \to \infty$. It suffices to show that $\sqrt{\sum_{k=1}^{k+1} a_j \nu_j}$ converges in distribution to a normal random variable.

Denote $\Sigma_k = \sqrt{\sum_{j=1}^{j} a_j \nu_j}$ and $s_k = \sqrt{\sum_{j=1}^{j} a_j^2 \nu_j^2}$ for $j \geq 0$. Let $\hat{B}$ be the imaginary unit, and for all the estimates below, $t$ is interpreted as a real number. We can show that

$$\Delta_t \triangleq |E\exp[i\pi \Sigma_k N] - \exp(-s_k N^2/2)|$$

$$\leq \exp(-s_k N^2/2) \sum_{j=0}^{\infty} E|\exp[i\pi \Sigma_k j + s_k j^2/2] - \exp[i\pi \Sigma_k j + s_k j^2/2]|$$

$$= O(1/k).$$
for any $N > k$. Now we have
\[
E \exp \left[ i \theta \sum_{j<k} + s_k r_j^2 / 2 \right] = E \exp \left[ i \theta \sum_{j=k} + s_k r_j^2 / 2 \right] \leq \exp \left( s_k r^2 / 2 \right) E \exp \left[ i \theta \sqrt{f} \right] - \exp \left( - \left( r^2 k_a^2 \sigma_j^2 / 2 \right) \right) \triangleq \exp \left( s_k r^2 / 2 \right) D_k.
\]
For random variable $\xi$, denote $H(t, \xi) = e^{t \xi} - 1 - i r \xi + r^2 \xi^2 / 2$. Let $H_k^{(1)} = H(t, \xi)$. Then
\[
D_k = |EH_k^{(1)} - H_k^{(2)}| & \leq H_k^{(1)} + |H_k^{(2)}| + (t^2 k_a^2 / 2) E \left[ |v_j|^2 \right] \sum_{j<k} \sigma_j^2.
\]
By the elementary inequality $|H(t, \xi)| \leq \sin |t^2 \xi^2 / 3| \leq \sin |t^2 \xi^2 / 3| / 6 \leq \sin |t^2 \xi^2 / 3|$ (see 6), we have
\[
|H_k^{(1)}| & \leq t^3 k_a^2 \sigma_j^2 \left\{ 1 \left( |v_j|^2 > \epsilon k^3 / 3 \right) \right\}.
\]
for all $\epsilon > 0$, and furthermore, $|H_k^{(2)}| \leq t^3 k_a^2 \sigma_j^4$. Let $\varepsilon_k = \sup_{k,j,k < k} (s_k / 2) / \epsilon$. We have
\[
\Delta_{k,N} \leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} |v_j|^2 \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(1)} \right) + \sum_{j=k}^{k-N} t^3 k_a^2 \sigma_j^4 \left( \Delta_{k,N}^{(2)} \right) + \sum_{j=k}^{k-N} t^3 k_a^2 \sigma_j^4 \left( \Delta_{k,N}^{(3)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} |v_j|^2 \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]
\[
\leq C \epsilon k^{1/3} \sum_{j=k}^{k-N} \left( |v_j|^2 > \epsilon k^3 / 3 \right) \left( \Delta_{k,N}^{(4)} \right)
\]