SIAM J. Control Optim.
Vol. 39, No. 2, pp. 533–570

LAW OF THE ITERATED LOGARITHM FOR A CONSTANT-GAIN LINEAR STOCHASTIC GRADIENT ALGORITHM*

J. A. JOSLIN† and A. J. HEUNIS‡

Abstract. We study almost-sure limiting properties, taken as $\varepsilon \searrow 0$, of the finite horizon sequence of random estimates $\{\theta_0^\varepsilon, \theta_1^\varepsilon, \theta_2^\varepsilon, \ldots, \theta_{\lfloor T/\varepsilon \rfloor}^\varepsilon\}$ for the linear stochastic gradient algorithm

$$\theta_{n+1}^\varepsilon = \theta_n^\varepsilon + \varepsilon \left[a_{n+1} - (\theta_n^\varepsilon)' X_{n+1}\right] X_{n+1}, \qquad \theta_0^\varepsilon \stackrel{\triangle}{=} \theta_* \quad \text{nonrandom},$$

where $T \in (0,\infty)$ is an arbitrary constant, $\varepsilon \in (0,1]$ is a (small) adaptation gain, and $\{a_n\}$ and $\{X_n\}$ are data sequences which drive the algorithm. These limiting properties are expressed in the form of a functional law of the iterated logarithm.

Key words. stochastic gradient algorithm, L-mixing processes, functional law of the iterated logarithm

AMS subject classifications. 60F15, 60F17, 93E10

PII. S0363012997331007

1. Introduction.

A commonly used stochastic gradient algorithm has the following structure:

$$\theta_{n+1}^\varepsilon = \theta_n^\varepsilon + \varepsilon \left[a_{n+1} - (\theta_n^\varepsilon)' X_{n+1}\right] X_{n+1}, \quad \theta_0^\varepsilon \stackrel{\triangle}{=} \theta_*, \quad \forall\, n = 0, 1, \ldots \tag{1.1}$$

Here $\varepsilon \in (0,1]$ is a (small) constant called the *adaptation gain*; $\theta_* \in \Re^d$ is a fixed nonrandom *initial condition*; and $\{a_n,\ n = 1,2,\ldots\}$ and $\{X_n,\ n = 1,2,\ldots\}$ are, respectively, $\Re$ and $\Re^d$-valued *data processes* which "drive" the algorithm and in turn give rise to the $\Re^d$-valued process of *estimates* $\{\theta_n^\varepsilon,\ n = 0,1,2,\ldots\}$.

Within this context one often wants to characterize asymptotic properties of the random sequence $\{\theta_0^\varepsilon, \theta_1^\varepsilon, \theta_2^\varepsilon, \ldots, \theta_{\lfloor T\varepsilon^{-1} \rfloor}^\varepsilon\}$ as $\varepsilon \to 0$, where $T \in (0,\infty)$ is a fixed but arbitrary constant called the *horizon*. Such asymptotic properties deal with the *finite horizon* characteristics of (1.1). Perhaps the most basic of these asymptotic properties is given by the so-called *ODE method* which, in the present context, essentially says the following. Under reasonably general conditions on the processes $\{a_n\}$ and $\{X_n\}$, one can relate the sequence of estimates $\{\theta_n^\varepsilon\}$ to the solution $\{\theta^0(\tau),\ \tau \in [0,\infty)\}$ of the ODE

$$\dot{\theta}^0(\tau) = \bar{b} - \bar{R}\theta^0(\tau), \quad \theta^0(0) = \theta_*, \tag{1.2}$$

(where the d-vector $\bar{b}$ and $d \times d$-matrix $\bar{R}$ in (1.2) are typically given by conditions such as (C2) in section 2) by the following *finite horizon weak law of large numbers.* For each $\delta \in (0,\infty)$ and $T \in (0,\infty)$ we have

$$\lim_{\varepsilon \to 0} P\left[\max_{0 \le \tau \le T} |\theta_{\lfloor \tau/\varepsilon \rfloor}^\varepsilon - \theta^0(\tau)| > \delta\right] = 0 \tag{1.3}$$

*Received by the editors December 2, 1997; accepted for publication (in revised form) May 11, 2000; published electronically September 15, 2000. Research for this work was supported by NSERC of Canada.

http://www.siam.org/journals/sicon/39-2/33100.html

†Siemens Canada, Kanata, Ontario K2K 2M5, Canada (jjoslin@magma.ca).

‡Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada (heunis@kingcong.uwaterloo.ca).

(see Kushner and Shwartz [19, Theorem 1, p. 178]), where $\lfloor x \rfloor$ denotes the integer part of $x \in [0, \infty)$. The ODE method is comprehensively covered in, for example, the books of Benveniste, Métivier, and Priouret [1], Kumar and Varaiya [17], and Kushner and Yin [18].

In classical probability, the law of large numbers is complemented by Donsker's functional central limit theorem (CLT), which has the following form. Suppose that $\{\xi_k,\ k = 1, 2, \ldots\}$ is an independent and identically distributed (i.i.d.) sequence of random variables with $E\xi_k = 0$, $E\xi_k^2 = 1$. Define the partial sums

$$S_n \stackrel{\triangle}{=} \sum_{k=1}^{n} \xi_k \qquad \forall\, n = 1, 2, \ldots \tag{1.4}$$

and, for each $m = 1, 2, \ldots$, let $\{\Xi^m(\tau),\ \tau \in [0, \infty)\}$ be the continuous piecewise-linear process given by

$$\Xi^m(\tau) \stackrel{\triangle}{=} \begin{cases} 0 & \text{if } \tau = 0, \\ m^{-\frac{1}{2}} S_n & \text{if } \tau = n/m, \quad \forall\, n = 1, 2, \ldots, \\ \text{linear interpolation}, & \text{otherwise}. \end{cases} \tag{1.5}$$

Then, for each $T \in (0, \infty)$, the process $\{\Xi^m(\tau),\ \tau \in [0, T]\}$ converges weakly to a standard Wiener process $\{W(\tau), \tau \in [0, T]\}$ as $m \to \infty$ (see Theorem 10.1 of Billingsley [2]). An analogous *finite horizon functional CLT* can be established for the random sequence $\{\theta_0^\varepsilon, \theta_1^\varepsilon, \theta_2^\varepsilon, \ldots, \theta_{\lfloor T\varepsilon^{-1} \rfloor}^\varepsilon\}$ obtained from (1.1). For each $\varepsilon \in (0, 1]$ define the $\Re^d$-valued continuous piecewise-linear process $\{\Theta^\varepsilon(\tau),\ \tau \in [0, \infty)\}$ by

$$\Theta^\varepsilon(\tau) \stackrel{\triangle}{=} \begin{cases} \varepsilon^{-\frac{1}{2}} \left(\theta_{\tau/\varepsilon}^\varepsilon - \theta^0(\tau) \right), & \tau = k\varepsilon, \quad \forall\, k = 0, 1, 2, \ldots, \\ \text{linear interpolation}, & \text{otherwise}. \end{cases} \tag{1.6}$$

Then, subject to certain regularity conditions on the data sequences $\{a_n\}$ and $\{X_n\}$, for each $T \in (0, \infty)$ the process $\{\Theta^\varepsilon(\tau),\ \tau \in [0, T]\}$ converges weakly in $C[0, T]$ (the space of continuous functions from $[0, T]$ into $\Re^d$) to a limiting Gauss–Markov process $\{\hat{\Theta}(\tau),\ \tau \in [0, T]\}$ as $\varepsilon \to 0$. A precise formulation of this result, pertaining to a general class of algorithms which includes (1.1) as a special case and providing a complete characterization of the Gauss–Markov limit, may be found in [1, Theorem 1, p. 107] and [3, Theorem 2, p. 969]. The Gauss–Markov limit is also discussed further in Remark 3.4.

In the context of a sum of i.i.d. random variables $\{\xi_k,\ k = 1, 2, \ldots\}$ with $E\xi_k = 0$ and $E\xi_k^2 = 1$, Donsker's functional CLT is complemented by another basic result, namely, Strassen's *functional law of the iterated logarithm* (see Theorem 3 of [25]). To see the form of this result, fix some $T \in (0, \infty)$ and put

$$K_\xi^T \stackrel{\triangle}{=} \left\{ \phi : [0, T] \to \Re \ : \phi(0) = 0,\ \phi(\cdot) \text{ abs. continuous},\ \frac{1}{2} \int_0^T |\dot{\phi}(s)|^2 \, ds \le 1 \right\}. \tag{1.7}$$

It is well known that K_ξ^T is a compact set of continuous functions (with the supremum norm of uniform convergence over $[0, T]$), and Strassen's functional law of the iterated logarithm (LIL) says the following. For P-almost all ω the sequence of continuous functions $\{\Xi^m(\tau, \omega)/\sqrt{2 \log\log m},\ \tau \in [0, T]\}$, indexed by $m = 1, 2, \ldots$, converges towards K_ξ^T as $m \to \infty$, and the set of its accumulation points coincides exactly with K_ξ^T, the sense of convergence being that of uniform convergence over $[0, T]$. This result—called by Williams [26, page 208] a "staggering generalization" of Kolmogorov's classical LIL—can be used as a tool for deriving many subtle fine-structure properties of the sample-paths of the partial-sum sequence $\{S_n\}$. For example, it can be used to show that, with probability one, the quantity

$$\frac{1}{N}\text{cardinality}\left(\left\{1 \leq n \leq N : \; S_n > \frac{1}{2}\sqrt{2n \log\log n}\right\}\right), \qquad N = 1, 2, \ldots,$$

exceeds 0.99999 for infinitely many values of N, whereas (again with probability one) the same quantity exceeds the slightly larger number 0.999999 for only finitely many N (see Strassen [25] for this as well as other examples of how one can use the functional LIL to analyze the partial sum process).

The preceding discussion suggests the problem of establishing a functional LIL for the random sequence $\{\theta_0^\varepsilon, \theta_1^\varepsilon, \theta_2^\varepsilon, \ldots, \theta_{\lfloor T\varepsilon^{-1}\rfloor}^\varepsilon\}$ arising from (1.1). This should have the same relation to the finite horizon functional CLT indicated previously as Strassen's functional LIL has to Donsker's functional CLT, and therefore should be of the following general form. For each fixed $T \in (0, \infty)$ there is a compact set $K_\Theta^T \subset C[0, T]$ (with a characterization analogous to that of K_ξ^T in (1.7)) and, for P-almost all ω, the family of continuous functions $\{\Theta^\varepsilon(\tau, \omega)/\sqrt{2\log\log\varepsilon^{-1}}, \; \tau \in [0, T]\}$, indexed by $\varepsilon \in (0, 1]$, converges towards K_Θ^T (as $\varepsilon \to 0$) and the set of its $C[0, T]$-accumulation points coincides exactly with K_Θ^T. This is a *finite horizon functional LIL* for the sequence of estimates $\{\theta_0^\varepsilon, \theta_1^\varepsilon, \theta_2^\varepsilon, \ldots, \theta_{\lfloor T\varepsilon^{-1}\rfloor}^\varepsilon\}$, and our goal is to establish a result of this kind (see Theorem 3.3 to follow) subject to certain conditions on the data sequences $\{a_n\}$ and $\{X_n\}$ which are set forth in section 2. We choose to concentrate attention on the algorithm (1.1) because there seem to be major technical obstacles to getting this result for the more general classes of fixed-gain algorithms proposed, for example, in [1] and [3], whereas the linear structure of (1.1) simplifies matters considerably.

The usual method for establishing a functional LIL, pioneered by Strassen [25], is to first prove a so-called *strong invariance principle* for the partial-sum sequence $\{S_n\}$. Essentially, this says that one can always construct a Wiener process $\{W_t\}$ on the same probability space on which the S_n of (1.4) are defined (or perhaps on some extension of this space) such that

$$S_{\lfloor t\rfloor} - W_t = o(\sqrt{t \log\log t}) \quad (t \to \infty) \quad \text{almost surely (a.s.).} \tag{1.8}$$

Then one uses (1.8), together with known sample-path properties of the Wiener process, as the basis for establishing Strassen's functional LIL. In the context of stochastic algorithms with *decreasing gain* one can follow a similar approach (see [12] and [24]) but for *constant-gain* algorithms, such as (1.1), it is not at all obvious how to formulate and prove an analogue of (1.8). Accordingly, we shall adopt a different approach, in the spirit of a method pioneered by Chover [5], who showed how to establish Strassen's functional LIL using a CLT with *rate of convergence* in place of the strong invariance principle (1.8). This general approach, which was made to work in [15] for the stochastic averaging principle, will be extended here to work for the algorithm (1.1).

The organization of the paper is as follows: in section 2 we state and discuss conditions on the sequences $\{a_n\}$ and $\{X_n\}$ which drive the algorithm (1.1). In section 3 we establish the main result, namely Theorem 3.3. The proof in section 3 relies on two key technical results, namely an auxiliary LIL (see Theorem 3.5) and an a.s. approximation theorem (see Theorem 3.6). These are established in sections 4 and 5, respectively. Another essential technical result, needed for the proofs in section 4, is a functional CLT with rate of convergence (see Theorem 4.4), and this is proved in section 6. In section 7 we develop a miscellany of subsidiary technical lemmas which are used in the previous sections. In section 8 we restate, in a form best suited to our needs, some results from probability theory which are used in section 3 to section 7. Finally, we define the notation at the beginning of each section where it is first needed.

2. Conditions.

Notation 2.1. We use the following notation: $\Re^d$, $\Re^{d\times r}$ denote the usual vector spaces of real d-dimensional column vectors and real d by r matrices, respectively, with vector norm $|x| \stackrel{\triangle}{=} (\sum_{i=1}^d x_i^2)^{1/2}$ for all $x \in \Re^d$, and matrix operator norm $|A| \stackrel{\triangle}{=} \max_{x\in\Re^r,\ |x|=1} |Ax|$ for all $A \in \Re^{d\times r}$. Write $(B)'$ for the transpose of a matrix B. For an $\Re^d$ or $\Re^{d\times r}$-valued random element X and $p \in [1,\infty)$, put $\|X\|_p \stackrel{\triangle}{=} (E[|X|^p])^{1/p}$. For $x \in [0,\infty)$, $\lfloor x \rfloor$ is the largest integer n such that $n \le x$.

The data sequences $\{a_n,\ n = 1,2,\ldots\}$ and $\{X_n,\ n = 1,2,\ldots\}$ driving the algorithm (1.1) will always be special instances of the class of L-mixing processes introduced by Gerencsér [9], and formulated in a discrete-parameter setting as follows.

DEFINITION 2.2. *Suppose that* $\{\mathcal{F}_n,\ n = 1,2,\ldots\}$ *and* $\{\mathcal{F}_n^+,\ n = 1,2,\ldots\}$ *are sequences of sub-*σ*-algebras in the probability triple* $(\Omega,\mathcal{F},P)$*, increasing and decreasing, respectively, with* $\mathcal{F}_n$ *and* $\mathcal{F}_n^+$ *independent for each* $n = 1,2,\ldots$. *An* $\Re^{d\times r}$*-valued random process* $\{z_n,\ n = 1,2,\ldots\}$ *on* $(\Omega,\mathcal{F},P)$ *is* L*-mixing with respect to the system* $(\mathcal{F}_n,\mathcal{F}_n^+)$ *when* (i) $\{z_n\}$ *is* $\{\mathcal{F}_n\}$*-adapted,* (ii) *for each* $p \in [1,\infty)$ *we have* $\sup_n \|z_n\|_p < \infty$*, and, for*

$$\gamma_p(s) \stackrel{\triangle}{=} \sup_{n>s} \|z_n - E\left[z_n \mid \mathcal{F}_{n-s}^+\right]\|_p \quad \forall\, s = 1,2,\ldots, \quad \textit{we have} \quad \sum_{1\le s<\infty} \gamma_p(s) < \infty.$$

We shall require the following strengthened notion of L-mixing.

DEFINITION 2.3. *Suppose that* $\{\mathcal{F}_n,\ n = 1,2,\ldots\}$ *and* $\{\mathcal{F}_n^+,\ n = 1,2,\ldots\}$ *are sequences of sub-*σ*-algebras in the probability triple* $(\Omega,\mathcal{F},P)$*, as in Definition* 2.2. *An* $\Re^{d\times r}$*-valued random process* $\{z_n,\ n = 1,2,\ldots\}$ *on* $(\Omega,\mathcal{F},P)$ *is geometrically* L*-mixing with respect to the system* $(\mathcal{F}_n,\mathcal{F}_n^+)$*, when* (i) $\{z_n\}$ *is* $\{\mathcal{F}_n\}$*-adapted,* (ii) $\sup_n \|z_n\|_p < \infty$ *for each* $p \in [0,\infty)$*,* (iii) *there is a constant* $\lambda \in (0,1)$ *and, for each* $p \in [1,\infty)$*, a constant* $C_p \in [0,\infty)$ *such that*

$$\sup_{n>s} \|z_n - E\left[z_n \mid \mathcal{F}_{n-s}^n\right]\|_p \le C_p\, \lambda^s \qquad \forall\, s = 1,2,\ldots,$$

where

$$\mathcal{F}_m^n \stackrel{\triangle}{=} \mathcal{F}_n \cap \mathcal{F}_m^+, \quad \textit{when} \quad 1 \le m < n. \tag{2.1}$$

(Thus, $\mathcal{F}_m^n$ *is the collection of all events which are members of both* $\mathcal{F}_n$ *and* $\mathcal{F}_m^+$*.) The constant* $\lambda \in (0,1)$ *is called a rate of the geometrically* L*-mixing process.*

Remark 2.4. Notice that conditioning is on the σ-algebra $\mathcal{F}_{n-s}^+$ in Definition 2.2, whereas it is on the smaller σ-algebra $\mathcal{F}_{n-s}^n$ in Definition 2.3 (see Remark 2.5 for more discussion on this). Using the elementary inequality $\|z - E[z \mid \mathcal{H}]\|_p \le 2\|z - E[z \mid \mathcal{G}]\|_p$, which holds for $z \in L_p(\Omega,\mathcal{F},P)$, $p \in [1,\infty)$, and sub-σ-algebras $\mathcal{G} \subset \mathcal{H} \subset \mathcal{F}$, one sees immediately that a geometrically L-mixing process with respect to a given system $(\mathcal{F}_n, \mathcal{F}_n^+)$ is also L-mixing with respect to the system $(\mathcal{F}_n, \mathcal{F}_n^+)$. Besides being motivated by L-mixing, the notion of geometric L-mixing in Definition 2.3 is also closely related to that of *exponential stability* introduced by Ljung [20] and Ljung and Caines [21] (see also Caines [4, p. 488] and Davis and Vinter [6, p. 217]). An important instance of a geometrically L-mixing process is the output of a stable finite-dimensional linear system with the form

$$x_{n+1} = Ax_n + Be_n, \qquad z_n = Cx_n + De_n, \qquad n = 1, 2, \ldots,$$

where $\{e_1, e_2, e_3, \ldots\}$ is a "driving" sequence of independent random vectors, and x_1 and $\sigma\{e_1, e_2, \ldots\}$ are independent, with $\sup_n \|e_n\|_p < \infty$ and $\|x_1\|_p < \infty$ for each $p \in [1, \infty)$. Here one defines

$$\text{(2.2)} \quad \mathcal{F}_n \stackrel{\triangle}{=} \sigma\{x_1, e_1, e_2, \ldots, e_n\}, \qquad \mathcal{F}_n^+ \stackrel{\triangle}{=} \sigma\{e_{n+1}, e_{n+2}, \ldots\} \qquad \forall\, n = 1, 2, \ldots,$$

and, by the argument in [4, pages 488-489] it is easily shown that $\{z_n\}$ is geometrically L-mixing with respect to the system $(\mathcal{F}_n, \mathcal{F}_n^+)$. It is this close link to stable linear systems that makes L-mixing a very appropriate model for signals in data communication problems. The L^p-bounds established for L-mixing in [9] and [10] also render this class of processes extremely tractable, and will be used frequently in the arguments that follow.

Remark 2.5. Suppose that $\{\mathcal{F}_n,\ n = 1, 2, \ldots\}$ and $\{\mathcal{F}_n^+,\ n = 1, 2, \ldots\}$ are sequences of sub-σ-algebras in the probability triple $(\Omega, \mathcal{F}, P)$, as in Definitions 2.2 and 2.3, and $\{z_n,\ n = 1, 2, \ldots\}$ is some $\Re^{d\times r}$-valued and geometrically L-mixing process with respect to the system $(\mathcal{F}_n, \mathcal{F}_n^+)$. For each $s = 1, 2, \ldots$, define the process $\{z_n[s], n = 1, 2, \ldots\}$ by

$$\text{(2.3)} \qquad z_n[s] \stackrel{\triangle}{=} E\left[\, z_n \mid \mathcal{F}_{n-s}^n \right] \qquad \forall\, n = 1, 2, \ldots,$$

where we put $\mathcal{F}_{n-s}^n \stackrel{\triangle}{=} \mathcal{F}_1^n$ when $n \le s$. Observe that, for each $s = 1, 2, \ldots$, the process $\{z_n[s],\ n = 1, 2, \ldots\}$ is s-dependent, since, from (2.3), one has $\sigma\{z_1[s], z_2[s], \ldots, z_m[s]\} \subset \mathcal{F}_m$, while $\sigma\{z_n[s], z_{n+1}[s], \ldots, z_{n+k}[s]\} \subset \mathcal{F}_{n-s}^+$, and the σ-algebras $\mathcal{F}_m$ and $\mathcal{F}_{n-s}^+$ are clearly independent when $n - m > s$. Thus, we see from Definition 2.3 that a geometrically L-mixing process $\{z_n\}$ can be nicely approximated by the s-dependent process $\{z_n[s]\}$. This approximation property was used by Ljung and Caines [21] to establish asymptotic normality in off-line system identification, and will likewise be necessary for proving the main results of this work.

From now on we shall always suppose that the data processes $\{a_n,\ n = 1, 2, \ldots\}$ and $\{X_n,\ n = 1, 2, \ldots\}$, which drive the recursion (1.1), are $\Re$ and $\Re^d$-valued, respectively, defined on a common probability triple $(\Omega, \mathcal{F}, P)$, and subject to the following conditions (C1) to (C4).

(C1) There are sequences $\{\mathcal{F}_n,\ n = 1, 2, \ldots\}$ and $\{\mathcal{F}_n^+,\ n = 1, 2, \ldots\}$ of sub-σ-algebras, as in Definition 2.2, such that $\{a_n\}$ and $\{X_n\}$ are geometrically L-mixing with respect to the system $(\mathcal{F}_n, \mathcal{F}_n^+)$.

(C2) There exists a d-vector $\bar{b}$ and a $d \times d$ matrix $\bar{R}$ such that the limits

$$\text{(2.4)} \quad \bar{b} \stackrel{\triangle}{=} \lim_{N\to\infty} \frac{1}{N+1} \sum_{n=n_0}^{N+n_0} E[a_n X_n]\,, \quad \bar{R} \stackrel{\triangle}{=} \lim_{N\to\infty} \frac{1}{N+1} \sum_{n=n_0}^{N+n_0} E[X_n (X_n)']$$

exist uniformly with respect to $n_0 = 1, 2, \ldots$.

To state the remaining conditions we need the following notation:

$$
(2.5)\qquad \begin{cases} b_n \stackrel{\triangle}{=} a_n X_n, \quad \bar{b}_n \stackrel{\triangle}{=} E[b_n]\,, \quad \tilde{b}_n \stackrel{\triangle}{=} b_n - \bar{b}_n, \quad \hat{b}_n \stackrel{\triangle}{=} \bar{b}_n - \bar{b}, \\ R_n \stackrel{\triangle}{=} X_n(X_n)', \quad \bar{R}_n \stackrel{\triangle}{=} E[R_n]\,, \quad \tilde{R}_n \stackrel{\triangle}{=} R_n - \bar{R}_n, \quad \hat{R}_n \stackrel{\triangle}{=} \bar{R}_n - \bar{R}. \end{cases}
$$

Also, recall that $\{\theta^0(\tau),\ \tau \in [0,\infty)\}$ is the (unique) solution of (1.2). The remaining conditions are as follows.

(C3a) For each $T \in (0,\infty)$ there exist constants $C_1(T) \in [0,\infty)$ and $\varepsilon(T) \in (0,1]$ such that

$$
\max_{0 \le k \le 1+\lfloor T\varepsilon^{-1}\rfloor} \left| \sum_{j=0}^{k-1} (\hat{b}_{j+1} - \hat{R}_{j+1}\theta^0(\varepsilon j)) \right| \le C_1(T)
$$

for each $\varepsilon \in (0, \varepsilon(T)]$.

(C3b) There exist constants $\varepsilon_0 \in (0,1]$, $\alpha \in [0, 3/4)$, and $C_2 \in [0,\infty)$ such that

$$
\left| \sum_{j=k+1}^{k+N} \hat{R}_{j+1}(I - \varepsilon\bar{R})^j \right| \le C_2 N^{\alpha}
$$

for all $\varepsilon \in (0,\varepsilon_0]$ and all $k, N = 1, 2, \ldots$.

(C4) There is a constant $C_3 \in [0,\infty)$ and a function $A : \Re^d \to \Re^{d\times d}$ such that $A(\theta)$ is symmetric positive definite for each $\theta \in \Re^d$, and

$$
(2.6)\qquad \left| \frac{1}{N+1} \operatorname{cov}\left(\sum_{n=n_0}^{N+n_0} \tilde{H}_n(\theta) \right) - A(\theta) \right| \le \frac{C_3[1+|\theta|^2]}{N+1}
$$

for all $\theta \in \Re^d$ and $N, n_0 = 1, 2, 3, \ldots$, where

$$
(2.7)\qquad \tilde{H}_n(\theta) \stackrel{\triangle}{=} \tilde{b}_n - \tilde{R}_n\theta \quad \forall\, \theta \in \Re^d, \quad \forall\, n = 1, 2, \ldots.
$$

Remark 2.6. Condition (C2) defines the d-vector $\bar{b}$ and the $d \times d$-matrix $\bar{R}$, which then gives the right side of the ODE (1.2). Conditions (C2), (C3a), and (C3b) control the "amount of nonstationarity" in the data processes $\{a_n\}$ and $\{X_n\}$. (C3a) is the same as the first of the conditions appearing in (3.4) of Khas'minskii [13] but is just rewritten in the context of algorithm (1.1) and plays a role similar to that of its counterpart in [13], while (C3b) is a mild condition which limits fluctuations of $\bar{R}_n \stackrel{\triangle}{=} E[X_n(X_n)']$ about $\bar{R}$ defined in (C2).

Remark 2.7. Suppose the $\Re^{d\times r}$-valued process $\{z_n,\ n = 1, 2, \ldots\}$ is geometrically L-mixing with respect to a system $(\mathcal{F}_n, \mathcal{F}_n^+)$. Then it follows at once that the centralized process $\{z_n - Ez_n,\ n = 1, 2, \ldots\}$ is geometrically L-mixing with respect to the system $(\mathcal{F}_n, \mathcal{F}_n^+)$. Moreover, if the $\Re^{r\times q}$-valued process $\{y_n,\ n = 1, 2, \ldots\}$ is also geometrically L-mixing with respect to the system $(\mathcal{F}_n, \mathcal{F}_n^+)$, then it is easily verified using Definition 2.3 that the $\Re^{d\times q}$-valued process $\{z_n y_n,\ n = 1, 2, \ldots\}$ is geometrically L-mixing with respect to system $(\mathcal{F}_n, \mathcal{F}_n^+)$. In view of these observations and condition (C1), it follows that the processes $\{\tilde{b}_n,\ n = 1, 2, \ldots\}$ and $\{\tilde{R}_n,\ n = 1, 2, \ldots\}$ given by (2.5) are zero-mean and geometrically L-mixing with respect to the system $(\mathcal{F}_n, \mathcal{F}_n^+)$, and $\{\tilde{H}_n(\theta),\ n = 1, 2, \ldots\}$ given by (2.7) is zero-mean and geometrically L-mixing for each $\theta \in \Re^d$.

Remark 2.8. Condition (C4) gives the function $A(\cdot)$ which will be used to formulate the limiting set K_Θ^T in the main result (see Theorem 3.3). In some special cases it is possible to give explicit formulae for $A(\theta)$. For example, suppose that the driving data $\{a_n, X_n\}$ is weakly stationary so that, for each θ, we have $E[\tilde{H}_1(\theta)] = E[\tilde{H}_{1+r}(\theta)]$ and $E[\tilde{H}_1(\theta)\tilde{H}_n'(\theta)] = E[\tilde{H}_{1+r}(\theta)\tilde{H}_{n+r}'(\theta)]$ for all $r, n = 1, 2, \ldots$. Then it is easily shown that (C4) follows from (C1) with $A(\theta)$ given by

$$A(\theta) \stackrel{\triangle}{=} E[\tilde{H}_1(\theta)\tilde{H}_1'(\theta)] + \sum_{n=2}^{\infty} E[\tilde{H}_1(\theta)\tilde{H}_n'(\theta)] + \sum_{n=2}^{\infty} E[\tilde{H}_n(\theta)\tilde{H}_1'(\theta)].$$

For another example, suppose that the driving data is cyclostationary. This implies that there is a positive integer constant P such that, for each θ, we have $E[\tilde{H}_n(\theta)] = E[\tilde{H}_{n+P}(\theta)]$ and, for $\Lambda(\theta, m, n) \stackrel{\triangle}{=} E[\tilde{H}_m(\theta)\tilde{H}_n'(\theta)]$, we have $\Lambda(\theta, m, n) = \Lambda(\theta, m + P, n + P)$ for all $m, n = 1, 2, \ldots$. With the help of the periodicity relation we can extend $\Lambda(\theta, m, n)$ to all integers $-\infty < m, n < \infty$. By a straightforward adaptation of the argument on page 222 of [13] or page 76 of [14] it may be shown that (C4) follows from (C1) with $A(\theta)$ given by

$$A(\theta) = \frac{1}{P} \sum_{m=1}^{P} \sum_{n=-\infty}^{\infty} \Lambda(\theta, m, n).$$

A property of the mapping $A(\theta)$ in (C4) that will soon be needed is the following lemma.

LEMMA 2.9 (proved in section 7). *Suppose conditions* (C1) *and* (C4). *Then the function $A(\cdot)$ given by* (2.6) *is locally Lipschitz continuous over $\Re^d$.*

3. The main result.

Notation 3.1. For the results of this and later sections we need the following additional notation. $C[0,T]$ indicates the space of continuous functions $f : [0,T] \to \Re^d$, for some $T \in (0, \infty)$, with norm $\|\cdot\|_C$ defined by $\|f\|_C \stackrel{\triangle}{=} \sup_{0 \le \tau \le T} |f(\tau)|$, and $AC_0[0,T] \stackrel{\triangle}{=} \{\psi \in C[0,T] :\ \psi(0) = 0,\ \psi(\cdot)\ \text{ abs. continuous}\}$. Also, for $x \in C[0,T]$ and $K \subset C[0,T]$, put $\|x - K\|_C \stackrel{\triangle}{=} \inf_{y \in K} \|x - y\|_C$. If $\{x_m,\ m = 1, 2 \ldots\}$ is a sequence in $C[0,T]$, then $\operatorname{acc}\{x_m\}$ denotes the set of its accumulation points in $C[0,T]$ (if any), and the notation $\{x_m(\tau),\ \tau \in [0,T]\} \longrightarrow\!\!\!\!\!\rightarrow K$ for some set $K \subset C[0,T]$ means that (i) $\lim_{m\to\infty} \|x_m - K\|_C = 0$ and (ii) $\operatorname{acc}\{x_m\} = K$. The $\longrightarrow\!\!\!\!\!\rightarrow$ symbol (due to Kuelbs [16]) provides a succinct language for expressing Strassen's functional LIL formulated in section 1, namely, for each $T \in (0, T)$ and processes $\{\Xi^m(\tau),\ \tau \in [0,T]\}$ given by (1.5), we have

$$\left\{ \frac{\Xi^m(\tau)}{\sqrt{2 \log\log m}},\ \tau \in [0,T] \right\} \longrightarrow\!\!\!\!\!\rightarrow K_\xi^T \qquad \text{a.s.} \tag{3.1}$$

For the main result of this paper, we must slightly extend this symbolism. If $x_\varepsilon \in C[0,T]$ for all $\varepsilon \in (0,1]$, then $\operatorname{acc}\{x_\varepsilon\}$ denotes the set of accumulation points in $C[0,T]$ (if any) as $\varepsilon \searrow 0$, and the notation $\{x_\varepsilon(\tau),\ \tau \in [0,T]\} \longrightarrow\!\!\!\!\!\rightarrow K$ means that (i) $\lim_{\varepsilon\to 0} \|x_\varepsilon - K\|_C = 0$ and (ii) $\operatorname{acc}\{x_\varepsilon\} = K$.

Remark 3.2. From now on we fix an arbitrary finite horizon $T \in (0, \infty)$. Using the $\longrightarrow\!\!\!\!\!\rightarrow$ notation of the preceding paragraph, our main result is the following LIL for algorithm (1.1), which characterizes a.s. limiting properties of $\{\Theta^\varepsilon(\tau),\ \tau \in [0,T]\}$, the restriction to $[0,T]$ of the process Θ^ε defined by (1.6).

THEOREM 3.3. *Suppose conditions* (C1)–(C4) *of section* 2*, fix some finite horizon* $T \in (0,\infty)$*, and let* $\bar{R}$*,* $A(\cdot)$*,* $\{\theta^0(\tau),\ \tau \in [0,T]\}$ *and* $\{\Theta^\varepsilon(\tau),\ \tau \in [0,T]\}$*, be defined by* (2.4), (2.6), (1.2), and (1.6)*, respectively. Define the mapping* $I_\Theta^T : AC_0[0,T] \to [0,\infty]$ *by*

$$I_\Theta^T(\phi) \triangleq \frac{1}{2}\int_0^T (\dot{\phi}(s) + \bar{R}\phi(s))' A^{-1}(\theta^0(s))(\dot{\phi}(s) + \bar{R}\phi(s))\, ds, \tag{3.2}$$

and let

$$K_\Theta^T \triangleq \{\phi \in AC_0[0,T] \colon I_\Theta^T(\phi) \le 1\}. \tag{3.3}$$

Then K_Θ^T *is a compact subset of* $C[0,T]$*, and we have*

$$\left\{ \frac{\Theta^\varepsilon(\tau)}{\sqrt{2\log\log\varepsilon^{-1}}},\ \tau \in [0,T] \right\} \longrightarrow\!\!\!\!\to K_\Theta^T \quad a.s. \tag{3.4}$$

Remark 3.4. To motivate the proof of Theorem 3.3 we briefly recall how one can establish the functional CLT giving weak convergence of $\{\Theta^\varepsilon(\tau),\ \tau \in [0,T]\}$ to some limiting Gauss–Markov process $\{\hat{\Theta}(\tau),\ \tau \in [0,T]\}$ (see section 1). For each $\varepsilon \in (0,1]$ define the piecewise-linear continuous process $\{W^\varepsilon(\tau),\ \tau \in [0,\infty)\}$ by $W^\varepsilon(0) \triangleq 0$, and

$$W^\varepsilon(\tau) \triangleq \begin{cases} \varepsilon^{\frac{1}{2}} \sum_{j=1}^k \tilde{b}_j - \tilde{R}_j \theta^0((j-1)\varepsilon) & \forall\, \tau = k\varepsilon,\ \ k = 1,2,\dots, \\ \text{linear interpolation}, & \text{otherwise}. \end{cases} \tag{3.5}$$

Fix some arbitrary $T \in (0,\infty)$ as in Remark 3.2. For $\varepsilon \in (0,1]$, define the mapping $G^\varepsilon : C[0,T] \to C[0,T]$ as follows. For every $w \in C[0,T]$, let $G^\varepsilon(w)$ be given by the solution $v \in C[0,T]$ of the recursion

$$v(\tau) = \begin{cases} w(0), & \text{if } \tau = 0, \\ w(\tau) - \varepsilon \sum_{j=0}^{k-1} \bar{R}\, v(\varepsilon j), & \text{if } \tau = k\varepsilon,\ \ k = 1,2,\dots, \\ \text{linear interpolation}, & \text{otherwise}. \end{cases} \tag{3.6}$$

By a detailed analysis of the recursion (1.1) (which is not given here) we can show

$$\Theta^\varepsilon(\cdot) \approx G^\varepsilon(W^\varepsilon)(\cdot), \tag{3.7}$$

in the sense that $\{\Theta^\varepsilon(\tau),\ \tau \in [0,T]\}$ and $\{G^\varepsilon(W^\varepsilon)(\tau),\ \tau \in [0,T]\}$ have approximately the same distributions in $C[0,T]$ for small $\varepsilon \in (0,1]$. This suggests that we can get a weak limit for $\{\Theta^\varepsilon(\tau), \tau \in [0,T]\}$ when we establish a weak limit for $\{W^\varepsilon(\tau),\ \tau \in [0,T]\}$ and show that G^ε converges suitably to some limiting mapping $G : C[0,T] \to C[0,T]$ as $\varepsilon \to 0$. Indeed, using the fact that $\{\tilde{b}_j\}$ and $\{\tilde{R}_j\}$ are geometrically L-mixing (see Remark 2.7) and trivially modifying the arguments of [1, pp. 105–106], it can be proved that $\{W^\varepsilon(\tau),\ \tau \in [0,T]\}$ converges weakly in $C[0,T]$ to the restriction to $[0,T]$ of the Gauss–Markov process $\{\hat{W}^0(\tau),\ \tau \in [0,\infty)\}$ given by

$$\hat{W}^0(\tau) \triangleq \int_0^\tau A^{\frac{1}{2}}(\theta^0(s))\, d\hat{B}(s) \qquad \forall\, \tau \in [0,\infty), \tag{3.8}$$

for some standard $\Re^d$-valued Brownian motion $\hat{B}(\cdot)$ on a probability triple $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$. In addition, it is easily seen from elementary analysis that $G^\varepsilon(\cdot)$ converges uniformly on compact subsets of $C[0,T]$ to the mapping $G : C[0,T] \to C[0,T]$ defined as follows.

For every $w \in C[0,T]$, let $G(w)$ be given by the solution $v \in C[0,T]$ of the linear integral equation

$$v(\tau) = w(\tau) - \int_0^\tau \bar{R}\, v(s)\, ds \quad \forall\, \tau \in [0,T]. \tag{3.9}$$

These facts suggest that $\{G^\varepsilon(W^\varepsilon)(\tau),\ \tau \in [0,T]\}$ converges weakly to the limit $\{G(\hat{W}^0)(\tau),\ \tau \in [0,T]\}$ and thus, in view of (3.7), that the weak limit of $\{\Theta^\varepsilon(\tau),\ \tau \in [0,T]\}$ is likewise the process $\{G(\hat{W}^0)(\tau),\ \tau \in [0,T]\}$ or, equivalently, the process given by the solution of the stochastic differential equation

$$d\hat{\Theta}(\tau) = d\hat{W}^0(\tau) - \bar{R}\hat{\Theta}(\tau)\, d\tau, \quad \hat{\Theta}(0) = 0, \quad \forall\, \tau \in [0,T]. \tag{3.10}$$

In short, one first proves a CLT giving weak convergence of the simpler process $\{W^\varepsilon(\tau),\ \tau \in [0,T]\}$ to the limit $\{\hat{W}^0(\tau),\ \tau \in [0,T]\}$ defined by (3.8), then uses the approximation (3.7) to transfer this result into a CLT giving weak convergence of $\{\Theta^\varepsilon(\tau),\ \tau \in [0,T]\}$ to the limit $\{\hat{\Theta}(\tau),\ \tau \in [0,T]\}$ defined by (3.10). Our strategy for establishing Theorem 3.3 will be based on a very analogous method. First establish a functional LIL for the process $\{W^\varepsilon(\tau),\ \tau \in [0,T]\}$ as follows.

Theorem 3.5 (proved in section 4). *Suppose conditions* (C1)–(C4) *of section* 2, *and fix some finite horizon $T \in (0,\infty)$. Let $\{\theta^0(\tau),\ \tau \in [0,T]\}$, $\{W^\varepsilon(\tau),\ \tau \in [0,T]\}$, and $A(\cdot)$ be defined by* (1.2), (3.5), and (2.6)*, respectively, and write*

$$K_W^T \triangleq \left\{ \phi \in AC_0[0,T] :\ \frac{1}{2}\int_0^T (\dot{\phi}(s))' A^{-1}(\theta^0(s))\dot{\phi}(s)\, ds \le 1 \right\}. \tag{3.11}$$

Then K_W^T is a compact subset of $C[0,T]$, and

$$\left\{ \frac{W^\varepsilon(\tau)}{\sqrt{2\log\log\varepsilon^{-1}}},\ \tau \in [0,T] \right\} \longrightarrow\!\!\!\!\!\rightarrow K_W^T \quad a.s. \tag{3.12}$$

It remains to transfer the LIL of Theorem 3.5 for the process $\{W^\varepsilon(\tau),\ \tau \in [0,T]\}$ into the one given by Theorem 3.3 for the process $\{\Theta^\varepsilon(\tau),\ \tau \in [0,T]\}$. To this end, the following almost-sure version of (3.7) is essential.

Theorem 3.6 (proved in section 5). *Suppose conditions* (C1)–(C3) *of section* 2, *and fix some finite horizon $T \in (0,\infty)$. Then, for the process $\{W^\varepsilon(\tau),\ \tau \in [0,T]\}$ and mapping $G^\varepsilon : C[0,T] \to G^\varepsilon$ defined by* (3.5) *and* (3.6)*, respectively, we have*

$$\lim_{\varepsilon \searrow 0} \|G^\varepsilon(W^\varepsilon) - \Theta^\varepsilon\|_C = 0 \ a.s. \tag{3.13}$$

With Theorems 3.5 and 3.6 available, the proof of the main result is easy.

Proof of Theorem 3.3. One sees from (3.6) and (3.9) that $G^\varepsilon(\cdot)$ and $G(\cdot)$ are linear and continuous on $C[0,T]$. Using the Arzela–Ascoli theorem, it is easily seen that $G^\varepsilon(\cdot) \to G(\cdot)$ uniformly on compact subsets of $C[0,T]$ as $\varepsilon \searrow 0$. In view of this fact, if $Y^\varepsilon \in C[0,T]$ for all $\varepsilon \in (0,1]$ is such that $\{Y^\varepsilon(\tau),\ \tau \in [0,T]\} \longrightarrow\!\!\!\!\!\rightarrow K$ for some compact $K \subset C[0,T]$, then it follows by easy analysis that $\{G^\varepsilon(Y^\varepsilon),\ \tau \in [0,T]\} \longrightarrow\!\!\!\!\!\rightarrow G(K)$. Thus, identifying $Y^\varepsilon(\tau)$ with $W^\varepsilon(\tau)/\sqrt{2\log\log\varepsilon^{-1}}$ and using Theorem 3.5, we get

$$\left\{ \frac{G^\varepsilon(W^\varepsilon)(\tau)}{\sqrt{2\log\log\varepsilon^{-1}}},\ \tau \in [0,T] \right\} \longrightarrow\!\!\!\!\!\rightarrow G(K_W^T), \qquad \text{a.s.} \tag{3.14}$$

Next, from the definitions of K_Θ^T and K_W^T (see (3.3), (3.11)), and the definition of $G(\cdot)$ given by (3.9), we see that

$$G(K_W^T) = K_\Theta^T. \tag{3.15}$$

Now K_W^T is compact (by Theorem 3.5), thus K_Θ^T is compact, and (3.4) follows from (3.14), (3.15), and Theorem 3.6. $\square$

Remark 3.7. One immediate consequence of Theorem 3.3 is that, subject to conditions (C1)–(C4), we have

$$\max_{0\le\tau\le T} |\theta^\varepsilon_{\lfloor\tau/\varepsilon\rfloor} - \theta^0(\tau)| = O(\varepsilon^{\frac{1}{2}}\sqrt{2\log\log\varepsilon^{-1}}) \quad \text{a.s.} \tag{3.16}$$

for each $T \in (0,\infty)$. We thus complement the finite horizon weak law of large numbers (1.3) with a strong law of large numbers together with an a.s. rate of convergence, which, by an argument identical to that in ([12], page 120), may be seen to be the best possible rate of convergence.

Remark 3.8. The general methodology used for establishing Theorem 3.3 is suggested by an approach developed in [15] for proving a functional LIL for random ODEs. The methods of [15] depend in an essential way on rather restrictive boundedness hypotheses which, when carried over directly into the context of (1.1), entail uniform boundedness of the driving data sequence $\{X_n\}$ with respect to $n = 1, 2, \ldots$ and $\omega \in \Omega$. Although this boundedness may be reasonably acceptable for differential equations, it is not realistic for algorithms, and it is necessary to significantly redesign the overall approach used in [15] for differential equations to suit the system (1.1). In the following sections we shall extensively use (i) nice properties of geometric L-mixing processes, in particular their approximability (see Remark 2.5) by s-dependent processes, (ii) L^p-bounds for sums of L-mixing processes over "triangular" domains (see Theorem 5.2), and (iii) the linear structure of (1.1), in order to deal with the problems caused by unboundedness of the data sequences in (1.1).

4. Proof of Theorem 3.5.

Notation 4.1. For this section we need the following additional notation: If (S,ρ) is a metric space then $\mathcal{B}(S)$ denotes its Borel σ-algebra, and if Y is a $\mathcal{F}/\mathcal{B}(S)$-measurable mapping from a triple $(\Omega,\mathcal{F},P)$ into (S,ρ), then $\mathcal{L}(Y)$ is the probability measure on $\mathcal{B}(S)$ defined by $\mathcal{L}(Y)(A) \stackrel{\triangle}{=} P\{\omega : \ Y(\omega) \in A\}$ for all $A \in \mathcal{B}(S)$. For probability measures P_1 and P_2 on the metric space $C[0,T]$ with metric given by the norm $\|\cdot\|_C$ (see Notation 3.1), let $\Pi_C(P_1,P_2)$ denote the Prohorov distance between P_1 and P_2 (see section 8 for a general definition of Prohorov distance). If $\{\alpha_n\}$ and $\{\beta_n\}$ are sequences of real numbers, then the notation $\alpha_n \ll \beta_n$ indicates the existence of a constant $C \in [0,\infty)$ such that $|\alpha_n| \le C|\beta_n|$ for all $n = 1, 2, \ldots$ If α_ε and β_ε are real numbers for each $\varepsilon \in (0,1]$ then $\alpha_\varepsilon \ll \beta_\varepsilon$ indicates the existence of numbers $\varepsilon_0 \in (0,1]$ and $C \in [0,\infty)$ such that $|\alpha_\varepsilon| \le C|\beta_\varepsilon|$ for all $\varepsilon \in (0,\varepsilon_0]$.

The proof of Theorem 3.5 relies on the following result which is an immediate consequence of combining Lemma 2.1(iv) and Theorem 4.3 of Kuelbs [16].

THEOREM 4.2. *Fix some $T \in (0,\infty)$, and suppose that $M : [0,T] \to \Re^{d\times d}$ is continuous with $M(s)$ being positive-definite symmetric for each $s \in [0,T]$. Then the set*

$$K \stackrel{\triangle}{=} \left\{ \phi \in AC_0[0,T] \colon \frac{1}{2}\int_0^T (\dot\phi(s))' M^{-1}(s)\dot\phi(s)\,ds \le 1 \right\} \tag{4.1}$$

is a compact subset of $C[0,T]$. *Define a Gaussian process* $\{\hat{Y}(\tau),\ \tau \in [0,T]\}$ *by*

$$\hat{Y}(\tau) \stackrel{\triangle}{=} \int_0^\tau M^{\frac{1}{2}}(s)\, d\hat{B}(s) \qquad \forall\, \tau \in [0,T], \tag{4.2}$$

where $\hat{B}(\cdot)$ *is a standard* $\Re^d$*-valued Brownian motion on a triple* $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{P})$, *and let* $\{Y_r,\ r = 1,2,\ldots\}$ *be a sequence of* $C[0,T]$*-valued random variables defined on a triple* $(\Omega, \mathcal{F}, P)$. *Then the following hold.* (i) *If*

$$\sum_{r=1}^{\infty} \Pi_C(\mathcal{L}(Y_r), \mathcal{L}(\hat{Y})) < \infty, \tag{4.3}$$

then

$$\lim_{r\to\infty} \left\| \frac{Y_r}{\sqrt{2\log r}} - K \right\|_C = 0, \quad a.s. \tag{4.4}$$

(ii) *If, in addition to* (4.3), *the sequence* $\{Y_r\}$ *is independent, then*

$$\operatorname{acc}\left\{ \frac{Y_r(\omega)}{\sqrt{2\log r}} \right\} = K, \quad a.s. \tag{4.5}$$

Remark 4.3. In Theorem 4.2 the set K is defined in terms of the covariance function $M(\cdot)$ of a Gaussian process $\{\hat{Y}(\tau)\}$. The theorem says that if a sequence of $C[0,T]$-valued random functions $\{Y_r\}$ converges fast enough in distribution to the Gaussian limit $\hat{Y}$ (see (4.3)), then all accumulation points of the sequence $\{Y_r(\omega)/\sqrt{2\log r},\ r = 1,2,\ldots\}$ are included *within* the set K for P-almost all ω (see (4.4)), regardless of any dependence which may exist in the sequence $\{Y_r\}$. This is true, in particular, when the Y_r are Gaussian with common distribution equal to that of $\hat{Y}$, for then (4.3) is automatically satisfied. If, in addition, the sequence $\{Y_r\}$ is *independent*, then the set of accumulation points of $\{Y_r(\omega)/\sqrt{2\log r},\ r = 1,2,\ldots\}$ is not only included within K but actually coincides *exactly* with K for P-almost all ω (see (4.5)). Taken together, (4.4) and (4.5) constitute a law of the *noniterated* logarithm. In the present section we shall use Theorem 4.4 (to follow) as a tool for verifying (4.3), and then establish Theorem 3.5 on the basis of the law of the noniterated logarithm resulting from Theorem 4.2.

THEOREM 4.4 (proved in section 6). *Suppose conditions* (C1)–(C4) *of section* 2 *hold and fix some finite horizon* $T \in (0,\infty)$. *Let* $\{\hat{W}^0(\tau),\ \tau \in [0,T]\}$ *be defined by* (3.8), *and let* $\{W^\varepsilon(\tau),\ \tau \in [0,T]\}$ *be defined by* (3.5). *Then there is a constant* $\eta \in (0,\infty)$ *such that*

$$\Pi_C(\mathcal{L}(W^\varepsilon), \mathcal{L}(\hat{W}^0)) \ll \varepsilon^\eta.$$

Remark 4.5. Fix an arbitrary finite horizon $T \in (0,\infty)$. For the proofs in this and in later sections we shall need the $\Re^d$-valued process $\{\xi_n^\varepsilon,\ n = 1,2,\ldots\}$ defined for each $\varepsilon \in (0,1]$ by

$$\xi_n^\varepsilon \stackrel{\triangle}{=} \tilde{H}_n(\theta^0(((n-1)\varepsilon) \wedge T)) \equiv \tilde{b}_n - \tilde{R}_n\theta^0(((n-1)\varepsilon) \wedge T) \quad \forall\, n = 1,2,\ldots \tag{4.6}$$

(recall (2.7) and (2.5)). To simplify the notation, we do not indicate dependence of ξ_n^ε upon the horizon T, which is fixed throughout this work. It follows at once from

(3.5) that we have

$$W^\varepsilon(\tau) = \varepsilon^{\frac{1}{2}} \left(\sum_{\nu=1}^{\lfloor \tau/\varepsilon \rfloor} \xi_\nu^\varepsilon + (\tau/\varepsilon - \lfloor \tau/\varepsilon \rfloor) \xi^\varepsilon_{\lfloor \tau/\varepsilon \rfloor + 1} \right) \qquad \forall\, \tau \in [0, T]. \tag{4.7}$$

Since $\{\tilde{b}_n\}$ and $\{\tilde{R}_n\}$ are zero-mean and geometrically L-mixing (see Remark 2.7), and $\theta^0(((n-1)\varepsilon) \wedge T)$ is uniformly bounded with respect to $\varepsilon \in (0, 1]$, $n = 1, 2, \ldots$, and nonrandom, it follows at once that $\{\xi_n^\varepsilon,\ n = 1, 2, \ldots\}$ is zero-mean and geometrically L-mixing. In fact, if

$$\xi_n^\varepsilon[s] \stackrel{\triangle}{=} E\left[\xi_n^\varepsilon \mid \mathcal{F}_{n-s}^n \right] = \tilde{b}_n[s] - \tilde{R}_n[s] \theta^0((n-1)\varepsilon \wedge T) \qquad \forall\, s, n = 1, 2, \ldots, \tag{4.8}$$

then there is a constant $\lambda \in (0, 1)$, and, for each $p \in [1, \infty)$, a constant $C_p \in [0, \infty)$, such that

$$\sup_{n > s} \| \xi_n^\varepsilon - \xi_n^\varepsilon[s] \|_p \le C_p\, \lambda^s \qquad \forall\, s = 1, 2, \ldots, \quad \forall\, \varepsilon \in (0, 1], \tag{4.9}$$

and it follows from Remark 2.5 that $\{\xi_n^\varepsilon[s],\ n = 1, 2, 3, \ldots\}$ is a zero-mean s-dependent process for each $s = 1, 2, \ldots$ and $\varepsilon \in (0, 1]$. Notice that λ and C_p in (4.9) are *uniform* with respect to $\varepsilon \in (0, 1]$.

Remark 4.6. We can apply Theorem 8.5 separately to the geometrically L-mixing processes $\{\tilde{b}_n\}$ and $\{\tilde{R}_n\}$ in (4.6), and use the uniform boundedness in (ε, n) of $\theta^0(((n-1)\varepsilon) \wedge T)$ to get the following. For each $p \in [2, \infty)$ there is a constant $C_p^1 \in [0, \infty)$ such that for any nonrandom sequence $\{A_n\}$ of $d \times d$-matrices we have

$$\left\| \sum_{n=1}^N A_n \xi_n^\varepsilon \right\|_p \le C_p^1 \left(\sum_{n=1}^N |A_n|^2 \right)^{\frac{1}{2}} \qquad \forall\, N = 1, 2, \ldots, \quad \forall\, \varepsilon \in (0, 1].$$

Notice that the dependencies of the constants in Theorem 8.5, and uniformity in $\varepsilon \in (0, 1]$ of C_p and λ in (4.9), entail that C_p^1 does not depend on ε, N, or the sequence $\{A_n\}$.

Proof of Theorem 3.5. From Lemma 2.9 we know that $A(\theta^0(s))$ is continuous in $s \in [0, T]$. Upon comparing (3.11) and (4.1), we see from Theorem 4.2 that K_W^T is compact. We now use Theorem 4.2 to show that a.s.

$$\text{(i)}\ \ \lim_{\varepsilon \searrow 0} \left\| \frac{W^\varepsilon}{\sqrt{2 \log\log \varepsilon^{-1}}} - K_W^T \right\|_C = 0, \quad \text{(ii)}\ \ K_W^T = \operatorname{acc} \left\{ \frac{W^\varepsilon}{\sqrt{2 \log\log \varepsilon^{-1}}} \right\}, \tag{4.10}$$

which gives Theorem 3.5.

Proof of (4.10)(i). Without loss of generality take $T = 1$. Fix $\sigma \in [\frac{9}{10}, 1)$ and put

$$\varepsilon_r \stackrel{\triangle}{=} \exp(-r^\sigma) \qquad \forall\, r = 1, 2, 3, \ldots. \tag{4.11}$$

Then, for each $\varepsilon \in [\varepsilon_{r+1}, \varepsilon_r]$, we have

$$\left\|\frac{W^\varepsilon(\cdot)}{\sqrt{2\log\log\varepsilon^{-1}}}-\frac{K_W^T}{\sqrt{\sigma}}\right\|_C \le \left\|\frac{W^\varepsilon(\cdot)}{\sqrt{2\log\log\varepsilon^{-1}}}-\frac{W^{\varepsilon_r}(\cdot)}{\sqrt{2\log\log\varepsilon^{-1}}}\right\|_C$$
$$+\left\|\frac{W^{\varepsilon_r}(\cdot)}{\sqrt{2\log\log\varepsilon^{-1}}}-\frac{W^{\varepsilon_r}(\cdot)}{\sqrt{2\log\log\varepsilon_r^{-1}}}\right\|_C+\left\|\frac{W^{\varepsilon_r}(\cdot)}{\sqrt{2\log\log\varepsilon_r^{-1}}}-\frac{K_W^T}{\sqrt{\sigma}}\right\|_C,$$

and thus

$$\sup_{\varepsilon_{r+1}\le\varepsilon\le\varepsilon_r}\left\|\frac{W^\varepsilon}{\sqrt{2\log\log\varepsilon^{-1}}}-\frac{K_W^T}{\sqrt{\sigma}}\right\|_C \le \sup_{\varepsilon_{r+1}\le\varepsilon\le\varepsilon_r}\frac{\|W^\varepsilon-W^{\varepsilon_r}\|_C}{\sqrt{2\log\log\varepsilon_r^{-1}}}$$
$$+\left\|\frac{W^{\varepsilon_r}}{\sqrt{2\log\log\varepsilon_r^{-1}}}\right\|_C\left(1-\sqrt{\frac{\log\log\varepsilon_r^{-1}}{\log\log\varepsilon_{r+1}^{-1}}}\right)+\left\|\frac{W^{\varepsilon_r}}{\sqrt{2\log\log\varepsilon_r^{-1}}}-\frac{K_W^T}{\sqrt{\sigma}}\right\|_C. \tag{4.12}$$

Remark 4.7. We will show that the three terms on the right of (4.12) go to zero a.s. when $r\to\infty$ for each $\sigma\in[\frac{9}{10},1)$. Since we can choose σ arbitrarily close to 1, this gives (4.10)(i). Our choice of the sequence $\{\varepsilon_r\}$ is determined by the following considerations. From (4.7) we see that the number of terms in the sum for $W^\varepsilon(\tau)$ increases reciprocally with decreasing ε. Thus, to control the supremum appearing in the first term on the right side of (4.12), we want the difference $(\varepsilon_{r+1}^{-1}-\varepsilon_r^{-1})$ to not be too large, which means that $\{\varepsilon_r\}$ must go to zero quite slowly. On the other hand, in the course of the following proof, we shall need Theorem 4.2(i) to deal with the last term on the right of (4.12), using Theorem 4.4 to verify a bound of the form (4.3) (with $Y_r\triangleq W^{\varepsilon_r}$, $\hat{Y}\triangleq\hat{W}^0$), and for this it is important that $\{\varepsilon_r\}$ not go to zero *too slowly.* Our choice of the sequence $\{\varepsilon_r\}$ turns out to be the right compromise, meeting both of these requirements.

We now deal with the first term on the right of (4.12). For each $\gamma\in[0,\infty)$, define piecewise-constant process $\{S_\gamma(\tau),\ \tau\in[0,1]\}$ by

$$S_\gamma(\tau)\triangleq\begin{cases}\sum_{j=1}^{\lfloor\tau\gamma\rfloor}\left[\tilde{b}_j-\tilde{R}_j\theta^0((j-1)/\gamma)\right], & \text{when } \gamma>0,\\ 0, & \text{when } \gamma=0,\end{cases} \tag{4.13}$$

and observe, from (4.7) and (4.13), that

$$W^\varepsilon(\tau)=\varepsilon^{\frac12}\{S_{\varepsilon^{-1}}(\tau)+(\tau/\varepsilon-\lfloor\tau/\varepsilon\rfloor)\,\xi^\varepsilon_{\lfloor\tau/\varepsilon\rfloor+1}\}\qquad\forall\,\tau\in[0,1]. \tag{4.14}$$

Put $N_\varepsilon\triangleq\lfloor\varepsilon^{-1}\rfloor$ for all $\epsilon\in(0,1]$, and observe from (4.14) that

$$\|W^\varepsilon-W^{\varepsilon_r}\|_C\le\varepsilon^{\frac12}\max_{k=1,\dots,N_\varepsilon}|\xi_k^\varepsilon|+\|\varepsilon^{\frac12}S_{\varepsilon^{-1}}-\varepsilon_r^{\frac12}S_{\varepsilon_r^{-1}}\|_C+\varepsilon_r^{\frac12}\max_{k=1,\dots,N_{\varepsilon_r}}|\xi_k^{\varepsilon_r}| \tag{4.15}$$

for all $\epsilon\in[\varepsilon_{r+1},\varepsilon_r]$. Now we need the following result.

LEMMA 4.8 (proved in section 7). *Suppose that $\{\gamma_n\}$ is a sequence of random variables (either $\Re^d$- or $\Re^{d\times r}$-valued for all n) on a probability space $(\Omega,\mathcal{F},P)$, such that $\sup_n\|\gamma_n\|_8<\infty$. Then, for almost always (a.a.) ω, there exists some constant $C(\omega)\in[0,\infty)$ such that*

$$\max_{n=1,2,\dots,N}|\gamma_i(\omega)|\le C(\omega)N^{\frac17}\qquad\forall\ N=1,2,\dots.$$

By (4.6) and Lemma 4.8, for P-a.a. ω there are constants $c_1(\omega), c_2(\omega) \in [0,\infty)$ such that

$$(4.16)\quad \varepsilon^{\frac{1}{2}} \max_{k=1,\dots,N_\varepsilon} |\xi_k^\varepsilon(\omega)| \le \varepsilon^{\frac{1}{2}} \left\{ \max_{k=1,2,\dots,N_\varepsilon} |\tilde{b}_k(\omega)| + \max_{k=1,\dots,N_\varepsilon} |\tilde{R}_k(\omega)||\theta^0((k-1)\varepsilon)| \right\}$$

$$\le \varepsilon^{\frac{1}{2}} c_1(\omega) \left(N_\varepsilon^{\frac{1}{7}} + N_\varepsilon^{\frac{1}{7}} \max_{0\le\tau\le1} |\theta^0(\tau)| \right) \le c_2(\omega) \varepsilon_r^{\frac{5}{14}} \quad \forall\, \varepsilon \in [\varepsilon_{r+1}, \varepsilon_r], \quad \forall\, r = 1,2,3,\dots.$$

For the second term on the right of (4.15), fix some integer $p \ge 3$ and observe that

$$(4.17)\qquad E\left[\sup_{\varepsilon_{r+1}\le\varepsilon\le\varepsilon_r} \|\varepsilon^{\frac{1}{2}} S_{\varepsilon^{-1}} - \varepsilon_r^{\frac{1}{2}} S_{\varepsilon_r^{-1}}\|_C^{2p} \right]$$

$$\le 2^{2p} \left\{ E\left[\sup_{\varepsilon_{r+1}\le\varepsilon\le\varepsilon_r} (|\varepsilon_r^{\frac{1}{2}} - \varepsilon^{\frac{1}{2}}| \|S_{\varepsilon^{-1}}\|_C)^{2p} \right] + E\left[\sup_{\varepsilon_{r+1}\le\varepsilon\le\varepsilon_r} (\varepsilon_r^{\frac{1}{2}} \|S_{\varepsilon^{-1}} - S_{\varepsilon_r^{-1}}\|_C)^{2p} \right] \right\}.$$

To deal with the expectations on the right side of (4.17) we need the next result, which is a slight variant of Lemma 3.7 in [15], and is proved in exactly the same way.

LEMMA 4.9. *Suppose conditions* (C1)–(C2) *of section* 2 *hold, and define* $S_\gamma(\tau)$, $\tau \in [0,1]$, *as in* (4.13) *for each* $\gamma \in [0,\infty)$. *Then, corresponding to each integer* $p \ge 3$, *there exists a constant* $\alpha_p \in (0,\infty)$ *such that*

$$E\left[\sup_{\gamma\le u\le\eta} \|S_u - S_\gamma\|_C^{2p} \right] \le \alpha_p \eta (\eta - \gamma)^{p-1} \qquad \forall\, 0 \le \gamma < \eta < \infty.$$

Since $\sigma < 1$ in (4.11), we see from the mean-value theorem applied to $r \to r^\sigma$ that $\varepsilon_r/\varepsilon_{r+1} \le e^\sigma$. Thus, from Lemma 4.9,

$$(4.18)\quad E\left[\sup_{\varepsilon_{r+1}\le\varepsilon\le\varepsilon_r} (\varepsilon_r^{\frac{1}{2}} \|S_{\varepsilon^{-1}} - S_{\varepsilon_r^{-1}}\|_C)^{2p} \right] = \varepsilon_r^p E\left[\sup_{\varepsilon_r^{-1}\le u\le\varepsilon_{r+1}^{-1}} \|S_u - S_{\varepsilon_r^{-1}}\|_C^{2p} \right]$$

$$\le \alpha_p \varepsilon_r^p \varepsilon_{r+1}^{-1} \left(\varepsilon_{r+1}^{-1} - \varepsilon_r^{-1} \right)^{p-1} \le e^\sigma \alpha_p \{ (\varepsilon_r - \varepsilon_{r+1})(\varepsilon_{r+1}^{-1}) \}^{p-1} \stackrel{\triangle}{=} e^\sigma \alpha_p B_r^p \qquad \forall\, r = 1,2,\dots.$$

In the same way, one shows that the first expectation on the right side of (4.17) has an identical upper bound, and thus the quantity on the left side of (4.17) is $O(B_r^p)$, where the constant implied by O depends only on p. By applying the mean value theorem to $r \to \exp(-r^\sigma)$, we easily see that $B_r^p \le \{\sigma e^\sigma r^{(\sigma-1)}\}^{(p-1)}$. Then, fixing integer $p > \frac{2-\sigma}{1-\sigma}$, it follows that $r^{(\sigma-1)(p-1)} = O(r^{-\beta-1})$ for some $\beta > 0$, and thus the sequence $\{B_r^p,\ r = 1,2,\dots\}$ is summable. Hence, from (4.17) and the monotone convergence theorem,

$$E\left[\sum_{r\ge1} \sup_{\varepsilon_{r+1}\le\varepsilon\le\varepsilon_r} \|\varepsilon^{\frac{1}{2}} S_{\varepsilon^{-1}} - \varepsilon_r^{\frac{1}{2}} S_{\varepsilon_r^{-1}}\|_C^{2p} \right] < \infty,$$

which implies

$$(4.19)\qquad \lim_{r\to\infty} \left[\sup_{\varepsilon_{r+1}\le\varepsilon\le\varepsilon_r} \|\varepsilon^{\frac{1}{2}} S_{\varepsilon^{-1}} - \varepsilon_r^{\frac{1}{2}} S_{\varepsilon_r^{-1}}\|_C \right] = 0 \text{ a.s.}$$

Similarly to (4.16), one sees that the third term on the right side of (4.15) is $O(\varepsilon_r^{\frac{5}{14}})$ a.s., and hence using (4.15), (4.16), and (4.19), we get

$$\lim_{r\to\infty}\left\{\sup_{\varepsilon_{r+1}\le\varepsilon\le\varepsilon_r}\|W^\varepsilon - W^{\varepsilon_r}\|_C\right\} = 0 \text{ a.s.} \tag{4.20}$$

This controls the first term on the right of (4.12). As for the third term, use Theorem 4.4 to find some constants $c_1>0$, $\eta>0$ and $r_0>0$ such that $\Pi_C(\mathcal{L}(W^{\varepsilon_r}),\mathcal{L}(\hat{W}^0)) \le c_1\exp(-\eta r^\sigma)$ for all $r \ge r_0$, thus $\sum_r \Pi_C(\mathcal{L}(W^{\varepsilon_r}),\mathcal{L}(\hat{W}^0)) < \infty$. Hence Theorem 4.2(i) (with $M(\tau)\triangleq A(\theta^0(\tau))$, $Y_r(\cdot)\triangleq W^{\varepsilon_r}(\cdot)$, $\hat{Y}(\cdot)\triangleq\hat{W}^0(\cdot)$, and $K\triangleq K_W^T$) establishes the convergence $\lim_{r\to\infty}\|\frac{W^{\varepsilon_r}(\cdot)}{\sqrt{2\log r}} - K_W^T\|_C = 0$ *a.s.* Now $\sigma\log r = \log\log\varepsilon_r^{-1}$ (see (4.11)); therefore,

$$\lim_{r\to\infty}\left\|\frac{W^{\varepsilon_r}}{\sqrt{2\log\log\varepsilon_r^{-1}}} - \frac{K_W^T}{\sqrt{\sigma}}\right\|_C = 0 \text{ a.s.} \tag{4.21}$$

As for the second term on the right side of (4.12), clearly $\lim_r \frac{\log\log\varepsilon_r^{-1}}{\log\log\varepsilon_{r+1}^{-1}} = 1$ (by L'Hôpital formula), and thus (4.21) and boundedness of K_W^T in $C[0,1]$ shows that this term goes a.s. to zero with $r\to\infty$. It follows that the quantity on the left side of (4.12) tends a.s. to zero as $r\to\infty$.

Proof of (4.10)(ii). Fix $\sigma\in(1,3/2)$, and define ε_r as in (4.11).

Remark 4.10. In contrast to the situation of Remark 4.7, where $\sigma<1$ so that ε_r^{-1} increases quite slowly with r, here we take $\sigma>1$ in (4.11) so that ε_r^{-1} increases *extremely rapidly* with r. In view of (4.7), this ensures that the sum for $W^{\varepsilon_{r+1}}(\tau)$ involves many more terms than does the sum for $W^{\varepsilon_r}(\tau)$, or, equivalently, the sums for $W^{\varepsilon_{r+1}}(\tau)$ and $W^{\varepsilon_r}(\tau)$ have few terms in common. The geometric L-mixing of $\{\xi_\nu^\varepsilon\}$ then suggests that the sequence $\{W^{\varepsilon_r}(\cdot)\}$ of $C[0,1]$-valued random variables should be approximately independent, and we can then expect to use Theorem 4.2(ii) to obtain (4.10)(ii). Indeed, in the course of the following proof we shall make this intuition rigorous, using the geometric L-mixing of $\{\xi_\nu^\varepsilon\}$ and rapid increase of $\{\varepsilon_r\}$ to construct a sequence $\{W_2^r(\cdot)\}$ of independent $C[0,1]$-valued random variables to which we can apply Theorem 4.2(ii), and which approximates the sequence $\{W^{\varepsilon_r}(\cdot)\}$ in a strong sense (see (4.31)). From this it is easy to deduce (4.10)(ii).

Recalling $\lambda\in(0,1)$ in (4.9), define

$$q(r)\triangleq\left\lfloor\frac{r^2}{2\ln\lambda^{-1}}\right\rfloor,\quad \tau_r\triangleq\varepsilon_r\left(1+q(r)+\frac{1}{\varepsilon_{r-1}}\right)\ \forall\, r=1,2,\ldots, \tag{4.22}$$

$$\zeta_u^\varepsilon\triangleq\xi_{\lfloor u\rfloor+1}^\varepsilon,\quad \zeta_u^\varepsilon[q]\triangleq\xi_{\lfloor u\rfloor+1}^\varepsilon[q]\quad\forall\varepsilon\in(0,1],\quad\forall u\in[0,\infty),\quad\forall q=1,2\ldots\lfloor u\rfloor. \tag{4.23}$$

For each $r=1,2,\ldots$, define $\{W_1^r(\tau),\ \tau\in[0,1]\}$ and $\{W_2^r(\tau),\ \tau\in[0,1]\}$ by

$$W_1^r(\tau)\triangleq\begin{cases}0, & \text{if } 0\le\tau\le\tau_r,\\ \varepsilon_r^{\frac{1}{2}}\displaystyle\int_{\tau_r/\varepsilon_r}^{\tau/\varepsilon_r}\zeta_u^{\varepsilon_r}\,du, & \text{if } \tau_r<\tau\le 1,\end{cases} \tag{4.24}$$

$$W_2^r(\tau)\triangleq\begin{cases}0, & \text{if } 0\le\tau\le\tau_r,\\ \varepsilon_r^{\frac{1}{2}}\displaystyle\int_{\tau_r/\varepsilon_r}^{\tau/\varepsilon_r}\zeta_u^{\varepsilon_r}[q(r)]\,du, & \text{if } \tau_r<\tau\le 1.\end{cases} \tag{4.25}$$

We first show that $\{W_2^r(\cdot)\}$ is a sequence of $C[0,1]$-valued *independent* random variables. To this end, observe from (4.22), (4.23), and (4.25) that

$$\mathcal{A}_r \stackrel{\triangle}{=} \sigma\{W_2^r(\tau),\ 0\le\tau\le 1\} \subset \sigma\{\xi^{\varepsilon_r}_{2+q(r)+\lfloor\varepsilon_{r-1}^{-1}\rfloor}[q(r)],\dots,\xi^{\varepsilon_r}_{1+\lfloor\varepsilon_r^{-1}\rfloor}[q(r)]\}. \tag{4.26}$$

Now $\xi^{\varepsilon_r}_j[q]$ is $\mathcal{F}^j_{j-q}$-measurable for all $q, j = 1, 2, \dots$ (see (4.8)), and thus from (4.26) we find

$$\mathcal{A}_r \subset \mathcal{F}^{1+\lfloor\varepsilon_r^{-1}\rfloor}_{2+\lfloor\varepsilon_{r-1}^{-1}\rfloor}, \tag{4.27}$$

where $\mathcal{F}^n_m$ is given by (2.1). Now one sees from Definition 2.3 that the finite collection of σ-algebras $\{\mathcal{F}^{n_1}_{m_1}, \mathcal{F}^{n_2}_{m_2}, \dots, \mathcal{F}^{n_r}_{m_r}\}$ is independent when $m_1 < n_1 < m_2 < n_2 < \cdots < m_r < n_r$; hence (4.27) shows that $\{W_2^r(\cdot),\ r=1,2,\dots\}$ is a sequence of $C[0,1]$-valued independent random variables. In order to use Theorem 4.2(ii) on this sequence, we next show that

$$\sum_{r=1}^{\infty} \Pi_C(\mathcal{L}(W_2^r), \mathcal{L}(\hat{W}^0)) < \infty, \tag{4.28}$$

where $\{\hat{W}^0(\tau),\ \tau\in[0,1]\}$ is defined by (3.8). By the triangle inequality we have

$$\begin{aligned}\Pi_C(\mathcal{L}(W_2^r), \mathcal{L}(\hat{W}^0)) \le \Pi_C(\mathcal{L}(W_2^r), \mathcal{L}(W_1^r)) &+ \Pi_C(\mathcal{L}(W_1^r), \mathcal{L}(W^{\varepsilon_r})) \\ &+ \Pi_C(\mathcal{L}(W^{\varepsilon_r}), \mathcal{L}(\hat{W}^0)),\end{aligned} \tag{4.29}$$

and it remains to upper-bound the three terms on the right side of (4.29). To this end we require the following lemma.

LEMMA 4.11 (proved in section 7). *Suppose the hypotheses of Theorem* 3.5. *Then we have*

$$E\left[\max_{0\le\tau\le 1} |W_1^r(\tau) - W_2^r(\tau)|^4\right] \ll \varepsilon_r.$$

Then, from Lemma 8.6(ii) (with $c \stackrel{\triangle}{=} 4$) we get $\Pi_C(\mathcal{L}(W_1^r), \mathcal{L}(W_2^r)) \ll (\varepsilon_r^{\frac{1}{4}})^{\frac{4}{5}} = \varepsilon_r^{\frac{1}{5}}$. Next, we need the following lemma.

LEMMA 4.12 (proved in section 7). *Suppose the hypotheses of Theorem* 3.5. *Then we have*

$$E\left[\max_{0\le\tau\le 1} |W^{\varepsilon_r}(\tau) - W_1^r(\tau)|^4\right] \ll r^{-10}.$$

By Lemmas 4.12 and 8.6(ii) (with $c \stackrel{\triangle}{=} 4$) we get $\Pi_C(\mathcal{L}(W_1^r), \mathcal{L}(W^{\varepsilon_r})) \ll (r^{-\frac{5}{2}})^{\frac{4}{5}} = r^{-2}$. By Theorem 4.4 there exists $r_0, \eta \in (0,\infty)$ such that $\Pi_C(\mathcal{L}(W^{\varepsilon_r}), \mathcal{L}(\hat{W}^0)) \ll \varepsilon_r^{\eta}$ for all $r \ge r_0$, and hence by (4.29), we get $\Pi_C(\mathcal{L}(W_2^r), \mathcal{L}(\hat{W}^0)) \ll \varepsilon_r^{\frac{1}{5}} + \varepsilon_r^{\eta} + r^{-2}$, and (4.28) follows. Thus, by independence of the sequence $\{W_2^r(\cdot),\ r = 1, 2, \dots,\}$ and Theorem 4.2(ii) (with $M(\tau) \stackrel{\triangle}{=} A(\theta^0(\tau))$, $Y_r(\cdot) \stackrel{\triangle}{=} W_2^r(\cdot)$, $\hat{Y}(\cdot) \stackrel{\triangle}{=} \hat{W}^0(\cdot)$, and $K \stackrel{\triangle}{=} K_W^T$), we get $\mathrm{acc}\{\frac{W_2^r}{\sqrt{2\log r}}\} = K_W^T$ a.s. Now $\sigma \log r = \log\log \varepsilon_r^{-1}$ (see (4.11)), and hence

$$\mathrm{acc}\left\{\frac{W_2^r}{\sqrt{2\log\log\varepsilon_r^{-1}}}\right\} = \frac{K_W^T}{\sqrt{\sigma}}. \tag{4.30}$$

From Lemma 4.11 and Borel–Cantelli, we have $\lim_{r\to\infty} \|W_1^r - W_2^r\|_C = 0$ a.s. Similarly, by Lemma 4.12, $\lim_{r\to\infty} \|W_1^r - W^{\varepsilon_r}\|_C = 0$ a.s. Thus, by the triangle inequality,

$$\lim_{r\to\infty} \|W_2^r - W^{\varepsilon_r}\|_C = 0 \text{ a.s.} \tag{4.31}$$

From (4.30), (4.31), and the fact that $\{W^{\varepsilon_r}\}$ is a subnet of $\{W^\varepsilon\}$,

$$\frac{K_W^T}{\sqrt{\sigma}} = \operatorname{acc}\left\{\frac{W^{\varepsilon_r}}{\sqrt{2\log\log\varepsilon_r^{-1}}}\right\} \subset \operatorname{acc}\left\{\frac{W^{\varepsilon}}{\sqrt{2\log\log\varepsilon^{-1}}}\right\} \quad \text{a.s.} \tag{4.32}$$

Since (4.32) holds for σ arbitrarily near 1 with $\sigma > 1$, we have

$$K_W^T \subset \operatorname{acc}\left\{\frac{W^{\varepsilon}}{\sqrt{2\log\log\varepsilon^{-1}}}\right\} \quad \text{a.s.} \tag{4.33}$$

Now (4.10)(i) and (4.33) yield (4.10)(ii) as required. $\square$

5. Proof of Theorem 3.6.

Remark 5.1. For ease of notation put

$$Z^\varepsilon(\tau) \stackrel{\triangle}{=} G^\varepsilon(W^\varepsilon)(\tau) \qquad \forall\, \tau \in [0,T]. \tag{5.1}$$

From (3.6) and (4.7) we see that, for each $\varepsilon \in (0,1]$, the sequence $\{Z^\varepsilon(\varepsilon k)\}$ is given by

$$Z^\varepsilon(\varepsilon(k+1)) = (I - \varepsilon\bar{R})Z^\varepsilon(\varepsilon k) + \varepsilon^{\frac{1}{2}}\xi_{k+1}^\varepsilon, \qquad Z^\varepsilon(0) = 0, \tag{5.2}$$

which is a linear system driven by the geometrically L-mixing process $\{\xi_k^\varepsilon\}$. Theorem 3.6 effectively relates $\{\Theta^\varepsilon(\varepsilon k)\}$ to the output of this system.

Without loss of generality we shall take $T = 1$. Put $N_\varepsilon \stackrel{\triangle}{=} \lfloor \varepsilon^{-1} \rfloor$ for all $\varepsilon \in (0,1]$. From (1.1), (1.2), and (2.5),

$$\theta_k^\varepsilon = \theta_* + \varepsilon\sum_{j=0}^{k-1}\left(b_{j+1} - R_{j+1}\theta_j^\varepsilon\right), \qquad \theta^0(\varepsilon k) = \theta_* + \int_0^{\varepsilon k}\left(\bar{b} - \bar{R}\theta^0(s)\right)\, ds. \tag{5.3}$$

In view of (1.6) we have $\Theta^\varepsilon(\varepsilon k) \stackrel{\triangle}{=} \varepsilon^{-\frac{1}{2}}(\theta_k^\varepsilon - \theta^0(\varepsilon k))$, and thus, using (5.3), (3.5), and (2.5),

$$\begin{aligned}
\Theta^\varepsilon(\varepsilon k) &= \varepsilon^{\frac{1}{2}}\sum_{j=0}^{k-1} b_{j+1} - \varepsilon^{\frac{1}{2}}\sum_{j=0}^{k-1} R_{j+1}\theta_j^\varepsilon + \varepsilon^{\frac{1}{2}}\sum_{j=0}^{k-1}(\tilde{b}_{j+1} - \tilde{R}_{j+1}\theta^0(\varepsilon j)) \\
&\quad - \varepsilon^{\frac{1}{2}}\sum_{j=0}^{k-1}((b_{j+1} - \bar{b}_{j+1}) - (R_{j+1} - \bar{R}_{j+1})\theta^0(\varepsilon j)) - k\varepsilon^{\frac{1}{2}}\bar{b} + \varepsilon^{-\frac{1}{2}}\int_0^{\varepsilon k}\bar{R}\theta^0(s)\, ds \\
&= W^\varepsilon(\varepsilon k) - \varepsilon^{\frac{1}{2}}\sum_{j=0}^{k-1} R_{j+1}(\theta_j^\varepsilon - \theta^0(\varepsilon j)) + \varepsilon^{\frac{1}{2}}\sum_{j=0}^{k-1}\big((\bar{b}_{j+1} - \bar{b}) \\
&\quad - (\bar{R}_{j+1} - \bar{R})\theta^0(\varepsilon j)) + \varepsilon^{-\frac{1}{2}}\left(\int_0^{\varepsilon k}\bar{R}\theta^0(s)\, ds - \varepsilon\sum_{j=0}^{k-1}\bar{R}\theta^0(\varepsilon j)\right),
\end{aligned}$$

and hence, again using (2.5),

$$\Theta^\varepsilon(\varepsilon k) = W^\varepsilon(\varepsilon k) - \varepsilon \sum_{j=0}^{k-1} R_{j+1}\Theta^\varepsilon(\varepsilon j) + \varepsilon^{\frac{1}{2}} \sum_{j=0}^{k-1} (\hat{b}_{j+1} - \hat{R}_{j+1}\theta^0(\varepsilon j)) + \varepsilon^{-\frac{1}{2}}\left(\int_0^{\varepsilon k} \bar{R}\theta^0(s)\,ds - \varepsilon \sum_{j=0}^{k-1} \bar{R}\theta^0(\varepsilon j)\right) \tag{5.4}$$

for all $\varepsilon \in (0,1]$ and $k = 1,2,\ldots,N_\varepsilon$. Define

$$\Delta^\varepsilon(\varepsilon k) \stackrel{\triangle}{=} \Theta^\varepsilon(\varepsilon k) - Z^\varepsilon(\varepsilon k) \qquad \forall\, k = 0,1,2,\ldots,N_\varepsilon. \tag{5.5}$$

From (2.5) we know that $\bar{R} = R_{j+1} - (\tilde{R}_{j+1} + \hat{R}_{j+1})$, and thus, using (5.1) and (3.6),

$$Z^\varepsilon(\varepsilon k) = W^\varepsilon(\varepsilon k) - \varepsilon \sum_{j=0}^{k-1} R_{j+1} Z^\varepsilon(\varepsilon j) + \varepsilon \sum_{j=0}^{k-1} (\tilde{R}_{j+1} + \hat{R}_{j+1}) Z^\varepsilon(\varepsilon j). \tag{5.6}$$

The quantity in brackets in the fourth term on the right side of (5.4) is clearly $O(\varepsilon)$ uniformly with respect to $k = 1,2,\ldots,N_\varepsilon$, while condition (C3a) ensures that the third term on the right side of (5.4) is $O(\varepsilon^{\frac{1}{2}})$ uniformly in $k = 1,2,\ldots,N_\varepsilon$. Subtracting (5.6) from (5.4) and taking magnitudes gives

$$|\Delta^\varepsilon(\varepsilon k)| \le \varepsilon \sum_{j=0}^{k-1} |R_{j+1}|\,|\Delta^\varepsilon(\varepsilon j)| + (|I_k^\varepsilon| + |J_k^\varepsilon| + O(\varepsilon^{\frac{1}{2}})), \tag{5.7}$$

where the constant implied by $O(\varepsilon^{\frac{1}{2}})$ is uniform with respect to $k = 1,2,\ldots,N_\varepsilon$, and

$$I_k^\varepsilon \stackrel{\triangle}{=} \varepsilon \sum_{j=0}^{k-1} \tilde{R}_{j+1} Z^\varepsilon(\varepsilon j), \quad J_k^\varepsilon \stackrel{\triangle}{=} \varepsilon \sum_{j=0}^{k-1} \hat{R}_{j+1} Z^\varepsilon(\varepsilon j) \quad \forall\, k = 1,2,\ldots,1+N_\varepsilon. \tag{5.8}$$

By (5.7), the fact that $\Delta^\varepsilon(0) = 0$, and the discrete-parameter version of Gronwall–Bellman's inequality (obtained by taking $\mu(\cdot)$ to be counting measure with weights $\varepsilon|R_{j+1}|$ over the nonnegative integers $j = 0,1,2,\ldots$ in Theorem 5.1 on page 498 of [8]), we get

$$\max_{0\le k\le N_\varepsilon+1} |\Delta^\varepsilon(\varepsilon k)| \le \exp\left(\varepsilon \sum_{j=1}^{N_\varepsilon} |R_j|\right) \max_{1\le k\le N_\varepsilon+1} (|I_k^\varepsilon| + |J_k^\varepsilon| + O(\varepsilon^{\frac{1}{2}})) \tag{5.9}$$

for all $\varepsilon \in (0,1]$. The proof proceeds by showing that, as $\varepsilon \searrow 0$, the right-hand side of (5.9) converges to zero a.s. Since $Z^\varepsilon(\cdot)$ and $\Theta^\varepsilon(\cdot)$ are both linear on intervals of the form $[(k-1)\varepsilon, k\varepsilon)$, $k = 1,2,\ldots,N_\varepsilon$ and continuous, this gives Theorem 3.6. Now products of geometrically L-mixing processes are geometrically L-mixing (see Remark 2.7), and thus one easily sees from condition (C1) and (2.5) that $\{|R_n|^2\}$ is geometrically L-mixing; in view of Theorem 8.5 and a standard use of the Borel–Cantelli lemma, it then follows that

$$\limsup_{\varepsilon\searrow 0} \varepsilon \sum_{j=1}^{N_\varepsilon} |R_j| < \infty, \quad \text{a.s.} \tag{5.10}$$

In light of (5.9) and (5.10), Theorem 3.6 will be established when we show that

$$\limsup_{\varepsilon\searrow 0}\max_{0\le k\le N_\varepsilon+1}|I_k^\varepsilon|=0 \text{ a.s.} \quad\text{and}\quad \limsup_{\varepsilon\searrow 0}\max_{0\le k\le N_\varepsilon+1}|J_k^\varepsilon|=0 \text{ a.s.} \tag{5.11}$$

To establish the first limit of (5.11), note from (5.2) that $Z^\varepsilon(\varepsilon j)=\varepsilon^{\frac12}(I-\varepsilon\bar R)^j\sum_{i=0}^{j-1}(I-\varepsilon\bar R)^{-i-1}\xi_{i+1}^\varepsilon$ for all $j=0,1,\dots,N_\varepsilon+1$, from which we see that I_k^ε in (5.8) can be written as

$$I_k^\varepsilon=\varepsilon^{\frac32}\sum_{j=0}^{k-1}\tilde R_{j+1}(I-\varepsilon\bar R)^j\Gamma_j^\varepsilon, \quad\text{where}\quad \Gamma_j^\varepsilon\triangleq\sum_{i=1}^{j}(I-\varepsilon\bar R)^{-i}\xi_i^\varepsilon,\quad \Gamma_0^\varepsilon\triangleq 0. \tag{5.12}$$

The following very special case of Theorem 1.1 of Gerencsér [10] is essential for establishing (5.11).

THEOREM 5.2. *Let $\{f_1(i)\}_{i=0}^\infty$ and $\{f_2(i)\}_{i=0}^\infty$ be real-valued nonrandom sequences, and let $\{u_1(i)\}_{i=0}^\infty$ and $\{u_2(i)\}_{i=0}^\infty$ be zero-mean geometrically L-mixing processes with rate $\lambda\in(0,1)$. Then there exists a constant $c\in(0,\infty)$ such that, for $l,k=1,2\dots$, with $l>k$, we have*

$$\left\|\sum_{j=k}^{l-1}\sum_{i=0}^{j-1}f_1(i)u_1(i)f_2(j)u_2(j)\right\|_4\le c[\psi_1(l,k)+\psi_2(l,k)],$$

where (taking $f_1(i)\triangleq f_2(i)\triangleq 0$ when $i<0$)

$$\begin{cases}\psi_1(l,k)\triangleq\left(\sum_{i=k}^{l}h(i,l,k)(|f_2(i-1)|+|f_2(i-2)|)\varphi_1^2(i)\right)^{\frac12},\\ \psi_2(l,k)\triangleq\sum_{i=1}^{l}h(i,l,k)|f_1(i-2)|,\end{cases} \tag{5.13}$$

$$\varphi_1(i)\triangleq\left(\sum_{j=0}^{i-1}f_1^2(j)\right)^{\frac12},\qquad h(i,l,k)\triangleq\sum_{j=i\vee k}^{l}|f_2(j-1)|\lambda^{j-i},\quad i=1,\dots,l. \tag{5.14}$$

Remark 5.3. The constant c is invariant with respect to $l,k=1,2,\dots$, and the sequences $\{f_1(i)\}$ and $\{f_2(i)\}$, but may depend on the quantities λ, $\sup_{i\ge0}\|u_1(i)\|_8$ and $\sup_{j\ge0}\|u_2(j)\|_8$. Theorem 1.1 of [10] is established in a continuous-parameter setting and gives general L^p-bounds on arbitrarily many multiple integrals of L-mixing processes. Theorem 5.2 stated here is a specialization of this result to the discrete-parameter case for an L^4-bound on two summations of geometrically L-mixing processes.

Since we can diagonalize $\bar R$ (by condition (C2)) there is no loss of generality in supposing that $d=1$. For each $\varepsilon\in(0,1)$ and $i,j=0,1,2,\dots$, define

$$u_1(\varepsilon;i)\triangleq\xi_{i+1}^\varepsilon\qquad\text{and}\qquad u_2(j)\triangleq\tilde R_{j+1}, \tag{5.15}$$

$$f_1(\varepsilon;i)\triangleq\begin{cases}(1-\varepsilon\bar R)^{-i-1}, & \text{if } i\ge0,\\ 0, & \text{if } i<0,\end{cases}\qquad f_2(\varepsilon;j)\triangleq\begin{cases}(1-\varepsilon\bar R)^{j}, & \text{if } j\ge0,\\ 0, & \text{if } j<0,\end{cases} \tag{5.16}$$

and note by Remark 4.5 that $u_1(\varepsilon;\cdot)$ and $u_2(\cdot)$ are zero-mean geometrically L-mixing with some rate $\lambda\in(0,1)$. From (5.12), (5.15), and (5.16) we can write

$$E[|I_l^\varepsilon - I_k^\varepsilon|^4] = \varepsilon^6 E\left[\left(\sum_{j=k}^{l-1}\sum_{i=0}^{j-1} f_1(\varepsilon;i)u_1(\varepsilon;i)f_2(\varepsilon;j)u_2(j)\right)^4\right] \tag{5.17}$$

for all $1 \le k < l \le 1 + N_\varepsilon$. Now define $\psi_1(l,k;\varepsilon)$, $\psi_2(l,k;\varepsilon)$, $\varphi_1(i;\varepsilon)$, and $h(i,l,k;\varepsilon)$ exactly as in (5.13) to (5.14), but allowing for the parametrization by ε in (5.15) and (5.16). Put $\varepsilon_0 \stackrel{\triangle}{=} 1/2$ when $\bar{R} = 0$, and put $\varepsilon_0 \stackrel{\triangle}{=} 1/(2\bar{R})$ when $\bar{R} > 0$. In view of (5.14) and (5.16), we have $h(i,l,k;\varepsilon) \le \sum_{j=i\vee k}^{l} \lambda^{j-i}$, and clearly $\max_{1\le i\le N_\varepsilon}(1-\varepsilon\bar{R})^{-i} = O(1)$ for all $\varepsilon \in (0,\varepsilon_0]$. Now $\sum_{i=1}^{k}\{\sum_{j=k}^{l}\lambda^{j-i}\} = \{\sum_{i=1}^{k}\lambda^{k-i}\}(1-\lambda^{l-k+1})/(1-\lambda) = O(1-\lambda^{l-k+1}) = O(l-k)$. Thus, from (5.13) and $\lambda \in (0,1)$, there are constants $c_1, c_2 \in (0,\infty)$ such that

$$\psi_2(l,k;\varepsilon) \le c_1\left(\sum_{i=1}^{k}\sum_{j=k}^{l}\lambda^{j-i} + \sum_{i=k+1}^{l}\sum_{j=i}^{l}\lambda^{j-i}\right) \le c_2(l-k) \tag{5.18}$$

for all $\varepsilon \in (0,\varepsilon_0]$, and $1 \le k < l \le 1+N_\varepsilon$. Similarly, from the fact that $\max_{1\le i\le N_\varepsilon}(1-\varepsilon\bar{R})^{-2i} = O(1)$ for all $\varepsilon \in (0,\varepsilon_0]$, and (5.14), we get $\varphi_1^2(i;\varepsilon) \le c_3 i$ for all $\varepsilon \in (0,\varepsilon_0]$, $i = 1,\ldots,N_\varepsilon$. Then, from (5.13), and $h(i,l,k;\varepsilon) \le \sum_{j=i}^{l}\lambda^{j-i}$ (when $i \ge k$),

$$\psi_1^2(l,k;\varepsilon) \le c_4\sum_{i=k}^{l}\left(\sum_{j=i}^{l}\lambda^{j-i}\right)\varphi_1^2(i;\varepsilon) \le c_5\sum_{i=k}^{l} i \le c_6(l^2-k^2) \tag{5.19}$$

for all $\varepsilon \in (0,\varepsilon_0]$, and $1 \le k < l \le 1+N_\varepsilon$, where $c_3, c_4, c_5, c_6 \in (0,\infty)$ are constants. Using Theorem 5.2, (5.17), (5.18), and (5.19), we find a constant $c_7 \in (0,\infty)$ such that

$$E[|I_l^\varepsilon - I_k^\varepsilon|^4] \le c_7\varepsilon^6[(l^2-k^2)^2 + (l-k)^4] \le 2c_7\varepsilon^6[l^2-k^2]^2 \tag{5.20}$$

for all $\varepsilon \in (0,1]$, and $1 \le k < l \le 1+N_\varepsilon$. Since $I_1^\varepsilon = 0$ (see (5.2), (5.8)), from (5.20) and Theorem 8.1(ii) (with $\gamma \stackrel{\triangle}{=} 2$, $\nu \stackrel{\triangle}{=} 4$, $h(i,j) \stackrel{\triangle}{=} j^2 - i^2$, $1 \le i \le j \le 1+N_\varepsilon$), there are constants $c_8, c_9 \in (0,\infty)$ such that

$$E\left[\max_{1\le k\le N_\varepsilon+1}|I_k^\varepsilon|^4\right] = E\left[\max_{1\le k\le N_\varepsilon+1}|I_k^\varepsilon - I_1^\varepsilon|^4\right] \le c_8\varepsilon^6[h(1,1+N_\varepsilon)]^2 \le c_9\varepsilon^2 \tag{5.21}$$

for all $\varepsilon \in (0,\varepsilon_0]$. From (5.21) (with $\varepsilon \stackrel{\triangle}{=} 1/n$) and the Borel–Cantelli theorem,

$$\lim_{n\to\infty}\max_{1\le k\le n+1}\left|I_k^{1/n}\right| = 0 \quad \text{a.s.},$$

and thus

$$\lim_{\varepsilon\searrow 0}\max_{1\le k\le N_\varepsilon+1}\left|I_k^{1/N_\varepsilon}\right| = 0 \quad \text{a.s.} \tag{5.22}$$

To get the first limit of (5.11) from (5.22) we must fill the gaps between successive $1/N_\varepsilon$. For this we observe

$$\max_{1\le k\le N_\varepsilon+1}|I_k^\varepsilon| \le \max_{1\le k\le N_\varepsilon+1}\left|I_k^{1/N_\varepsilon}\right| + B_\varepsilon^1 + B_\varepsilon^2, \tag{5.23}$$

which results from the triangle inequality, (5.12), $N_\varepsilon^{\frac{3}{2}}\varepsilon^{\frac{3}{2}} \le 1$, and the definition of B_ε^1 and B_ε^2 in (5.24) and (5.25). Now we need the following lemma.

LEMMA 5.4 (proved in section 7). *Suppose* (C1)–(C3) *of section* 2. *Let* Γ_j^ε *be defined by* (5.12) *and* $N_\varepsilon \stackrel{\triangle}{=} \lfloor \varepsilon^{-1} \rfloor$. *Then* (a) *we have*

$$(5.24) \quad B_\varepsilon^1 \stackrel{\triangle}{=} \varepsilon^{\frac{3}{2}} \max_{1\le k\le N_\varepsilon+1} \left| \sum_{j=0}^{k-1} \tilde{R}_{j+1}((I-\varepsilon\bar{R})^j - (I - N_\varepsilon^{-1}\bar{R})^j)\Gamma_j^{N_\varepsilon^{-1}} \right| \ll \varepsilon^{\frac{1}{2}},$$

$$(5.25) \quad B_\varepsilon^2 \stackrel{\triangle}{=} \varepsilon^{\frac{3}{2}} \max_{1\le k\le N_\varepsilon+1} \left| \sum_{j=0}^{k-1} \tilde{R}_{j+1}(I-\varepsilon\bar{R})^j(\Gamma_j^\varepsilon - \Gamma_j^{N_\varepsilon^{-1}}) \right| \ll \varepsilon^{\frac{3}{14}},$$

and (b) *we have bounds identical to* (5.24) *and* (5.25), *but with* $\hat{R}_{j+1}$ *in place of* $\tilde{R}_{j+1}$.

The first limit of (5.11) follows from (5.22), (5.23), and Lemma 5.4 (a). In the same way, using condition (C3b) and Lemma 5.4(b), we can establish the second limit in (5.11) (the proof is similar to, but easier than that of the first limit in (5.11) since the matrices $\hat{R}_{j+1}$ in the definition of J_k^ε are nonrandom, whereas the $\tilde{R}_{j+1}$ in the definition of I_k^ε are random—see (5.8)). ☐

6. Proof of Theorem 4.4.

Notation 6.1. In this section we shall require the following additional notation: $\mathcal{Z} \stackrel{\triangle}{=} \{\dots,-2,-1,0,1,2,\dots\}$, $\mathcal{Z}_+ \stackrel{\triangle}{=} \{1,2,\dots\}$. Also, $\#A$ indicates the cardinality of a finite set A. For positive integer m, let $\Pi_2^m(P_1,P_2)$ denote the Prohorov distance between probability measures P_1 and P_2 on the metric space $\Re^m$ with metric given by the norm $|\cdot|$ (see Notation 2.1), and let $\mathcal{N}_m(b,Q)$ denote the normal distribution in $\Re^m$ with m-dimensional mean vector b and covariance Q.

Remark 6.2. Theorem 4.4 is a functional CLT with rate of convergence for the process $\{W^\varepsilon(\tau),\ \tau \in [0,T]\}$, which is derived from summing the geometrically L-mixing random vectors $\{\xi_n^\varepsilon\}$ (see Remark 4.5), and may be regarded as a generalization to a function-space setting of [11, Lemma A.2.1]. This latter result gives rates of convergence in a classical (i.e., nonfunctional) CLT for a sum of geometrically L-mixing random vectors in terms of bounds on characteristic functions over finite-dimensional Euclidean space. The function-space result is considerably more difficult to establish because the Prohorov distance for probability measures on $C[0,T]$ does not relate nicely to characteristic functions. Theorem 4.4 also extends the result of Yurinskii [27, section 2], which is a functional CLT with rate of convergence for sums of independent random vectors, and bears clear similarities to [14, Lemma A6.1], which is a functional CLT with rate of convergence for a sum of strong mixing random vectors subject to quite stringent boundedness conditions that do not generally apply to algorithms. We emphasize that Theorem 4.4 involves a combination of a *function-space* rate of convergence (as contrasted with the classical rate in [11]) for a sum of *dependent* random vectors (compared with the independent case in [27]), and subject to only weak boundedness (as contrasted with uniform boundedness in [14]). The combination of all these elements presents technical challenges not found in [11], [14], or [27], and is the main reason for the somewhat lengthy proofs of this section.

Without loss of generality we shall prove Theorem 4.4 with $T = 1$. For all $k=1,2,\dots$ and $\varepsilon \in (0,1]$ define the kd-dimensional vectors Ξ_k^ε and $\hat{\Xi}_k^0$ by

$$\Xi_k^\varepsilon \triangleq \begin{bmatrix} W^\varepsilon(\frac{1}{k}) \\ W^\varepsilon(\frac{2}{k}) \\ \vdots \\ W^\varepsilon(1) \end{bmatrix} \quad \text{and} \quad \hat{\Xi}_k^0 \triangleq \begin{bmatrix} \hat{W}^0(\frac{1}{k}) \\ \hat{W}^0(\frac{2}{k}) \\ \vdots \\ \hat{W}^0(1) \end{bmatrix}. \tag{6.1}$$

In order to establish Theorem 4.4 we need the following CLT giving weak convergence of Ξ_k^ε to $\hat{\Xi}_k^0$ as $\varepsilon \to 0$ and $k \to \infty$, where k increases slowly enough that it is much less than ε^{-1} (the slowly increasing k is needed when we establish weak convergence in a function space setting).

THEOREM 6.3. *Let* Ξ_k^ε *and* $\hat{\Xi}_k^0$ *be as defined in* (6.1)*. Under conditions* (C1)–(C4) *of section* 2*, there are constants* $c \in (0,\infty)$ *and* $\varepsilon_0 \in (0,1)$ *such that, for all* $\varepsilon \in (0,\varepsilon_0]$ *and* $k=1,2,\ldots,\lfloor \varepsilon^{-\frac{1}{36}} \rfloor$*, we have*

$$\Pi_2^{kd}(\mathcal{L}(\Xi_k^\varepsilon), \mathcal{L}(\hat{\Xi}_k^0)) \leq c\varepsilon^{\frac{1}{48}}. \tag{6.2}$$

Proof of Theorem 6.3. Fix arbitrary $\varepsilon \in (0,1]$ and $k=1,2,\ldots$. Recalling that we take $T=1$ in the proof, we can use (4.7) to write the lth element $W^\varepsilon(l/k)$ of Ξ_k^ε in the form

$$W^\varepsilon(l/k) = \varepsilon^{\frac{1}{2}} \left(\sum_{j=1}^{l} \left[\sum_{\nu \in J_j^{\varepsilon,k}} \xi_\nu^\varepsilon \right] + V_l^{\varepsilon,k} \right) \qquad \forall\, l = 1,2,\ldots,k, \tag{6.3}$$

where $V_l^{\varepsilon,k}$ is a small "interpolation term" given by

$$V_l^{\varepsilon,k} \triangleq \left(\frac{l}{\varepsilon k} - \left\lfloor \frac{l}{\varepsilon k} \right\rfloor \right) \xi^\varepsilon_{\lfloor l/(\varepsilon k) \rfloor + 1} \qquad \forall\, l = 1,2,\ldots,k, \tag{6.4}$$

and $J_j^{\varepsilon,k}$ is the block of consecutive integers given by

$$J_j^{\varepsilon,k} \triangleq \left\{ 1 + \left\lfloor \frac{j-1}{\varepsilon k} \right\rfloor, \ldots, \left\lfloor \frac{j}{\varepsilon k} \right\rfloor \right\} \qquad \forall\, j = 1,2,\ldots,k. \tag{6.5}$$

We will call the $J_j^{\varepsilon,k}$ the "basic blocks." We see from (6.3) that the lth row of Ξ_k^ε involves summation over the basic blocks $J_1^{\varepsilon,k}, \ldots J_l^{\varepsilon,k}$.

Remark 6.4. The proof is based on the Markov–Bernstein technique of long blocks alternated with short blocks. The basic idea is as follows. If $\sum_{\nu=1}^m \rho_\nu$ is a sum of weakly-dependent random vectors $\{\rho_\nu\}$ (e.g., the elements of a geometric L-mixing sequence), then we can study its asymptotic properties by partitioning the interval of summation $1,2,\ldots,m$ into long blocks G_i^m of integers, all of equal length p_m, separated by short blocks H_i^m of integers, likewise of equal length q_m, giving a pattern of adjacent blocks $G_1^m, H_1^m, G_2^m, H_2^m, \ldots, G_r^m, H_r^m, H_{r+1}^m$. Here the integer $r \equiv r_m$ is equal to the *maximum* number of long block/short block pairs (of total length $p_m + q_m$) which one can fit completely into the interval of summation $1,2,\ldots,m$, and H_{r+1}^m is a "remainder block" whose cardinality is less than that of a long block/short block pair, namely $p_m + q_m$. On the one hand, if the short blocks H_i^m are long enough, then terms of the form $\sum_{\nu \in G_i^m} \rho_\nu$ and $\sum_{\nu \in G_j^m} \rho_\nu$ for $i \neq j$ involve widely separated (hence approximately independent) summands, and are thus themselves almost independent. On the other hand, if the long blocks are much longer than the short blocks then the contribution to the total sum of the ρ_ν for ν belonging to the short blocks H_i^m is negligible, and thus $\sum_{\nu=1}^m \rho_\nu \approx \sum_{i=1}^{r_m} \{\sum_{\nu\in G_i^m} \rho_\nu\}$. We thus have a sum (over $i = 1, 2, \ldots, r_m$) of almost independent terms on the right-hand side, and we can use this independence to study the limiting properties of the original sum of random vectors.

In our problem matters are slightly more complicated because we do not have a fixed interval of summation. Instead, different entries $W^\varepsilon(l/k)$ of the vector Ξ_k^ε involve different intervals of summation, as may be seen from (6.3). We therefore partition not the overall intervals of summation for $W^\varepsilon(l/k)$, but rather the basic blocks $J_j^{\varepsilon,k}$ in (6.3), into adjacent long and short blocks $G_{i,j}^{\varepsilon,k}$ and $H_{i,j}^{\varepsilon,k}$ of consecutive integers ordered according to the pattern

$$G_{1,j}^{\varepsilon,k},\, H_{1,j}^{\varepsilon,k},\, G_{2,j}^{\varepsilon,k},\, H_{2,j}^{\varepsilon,k},\, G_{3,j}^{\varepsilon,k}, H_{3,j}^{\varepsilon,k}, \ldots, G_{r,j}^{\varepsilon,k},\, H_{r,j}^{\varepsilon,k},\, H_{r+1,j}^{\varepsilon,k}. \tag{6.6}$$

The blocks $G_{i,j}^{\varepsilon,k}$, $i = 1, 2, \ldots, r$ have common cardinality $p_{\varepsilon,k}$ and the alternate blocks $H_{i,j}^{\varepsilon,k}$, $i = 1, 2, \ldots, r$ have common cardinality $q_{\varepsilon,k}$ given by

$$p_{\varepsilon,k} \triangleq \lfloor k^{-1}\varepsilon^{-\frac{1}{3}} \rfloor, \qquad q_{\varepsilon,k} \triangleq \lfloor k^{-1}\varepsilon^{-\frac{1}{9}} \rfloor, \tag{6.7}$$

while r is chosen to completely fit the largest number of consecutive pairs of blocks $G_{i,j}^{\varepsilon,k}, H_{i,j}^{\varepsilon,k}$ (of total length $p_{\varepsilon,k} + q_{\varepsilon,k}$) into $J_j^{\varepsilon,k}$, namely

$$r \equiv r_{\varepsilon,k} \triangleq \lfloor (\# J_j^{\varepsilon,k}) / (p_{\varepsilon,k} + q_{\varepsilon,k}) \rfloor. \tag{6.8}$$

The $G_{i,j}^{\varepsilon,k}$ are long blocks and the $H_{i,j}^{\varepsilon,k}$ are short blocks for all $i = 1, 2, \ldots, r$, while the last block $H_{r+1,j}^{\varepsilon,k}$ is a "remainder block" whose cardinality is less than $p_{\varepsilon,k} + q_{\varepsilon,k}$. Thus,

$$\#\left(G_{i,j}^{\varepsilon,k}\right) = p, \quad \#\left(H_{i,j}^{\varepsilon,k}\right) = q, \quad \text{and} \quad 0 \le \#\left(H_{r+1,j}^{\varepsilon,k}\right) < p + q, \tag{6.9}$$

where, for brevity, p and q are now used for $p_{\varepsilon,k}$ and $q_{\varepsilon,k}$, respectively. Denoting the first integer of the block $G_{i,j}^{\varepsilon,k}$ by $b_{i,j}^{\varepsilon,k}$, from (6.6) we see that

$$b_{i,j}^{\varepsilon,k} \triangleq 1 + \left\lfloor \frac{j-1}{k\varepsilon} \right\rfloor + (i-1)(p+q) \qquad \forall\, i = 1, \ldots, r+1, \tag{6.10}$$

and, for each $j = 1, 2, \ldots, k$, we clearly have $G_{i,j}^{\varepsilon,k} = [b_{i,j}^{\varepsilon,k}\ ,\ b_{i,j}^{\varepsilon,k} + p)$ and $H_{i,j}^{\varepsilon,k} = [b_{i,j}^{\varepsilon,k} + p\ ,\ b_{i,j}^{\varepsilon,k} + p + q)$, for all $i = 1, \ldots, r$, while $H_{r+1,j}^{\varepsilon,k} = [b_{r,j}^{\varepsilon,k} + p + q\ ,\ j/(k\varepsilon)]$.

Remark 6.5. To summarize, the lth entry $W^\varepsilon(l/k)$ of Ξ_k^ε is partitioned into a sum of ξ_ν^ε over the basic blocks $J_j^{\varepsilon,k}$ indexed by $j = 1, 2, \ldots, l$ (see (6.3)), and each basic block $J_j^{\varepsilon,k}$ is itself partitioned into a sequence of long block/short block pairs $G_{i,j}^{\varepsilon,k}, H_{i,j}^{\varepsilon,k}$, indexed by $i = 1, 2, \ldots, r_{\varepsilon,k}$, and a remainder block $H_{r+1,j}^{\varepsilon,k}$ (see (6.6)).

For each $j=1, \ldots, k$ and $i=1, \ldots, r$, define the d-vectors

$$Y_{i,j}^{\varepsilon,k} \triangleq \sum_{\nu\in G_{i,j}^{\varepsilon,k}} \xi_\nu^\varepsilon, \qquad Z_{i,j}^{\varepsilon,k} \triangleq \sum_{\nu\in H_{i,j}^{\varepsilon,k}} \xi_\nu^\varepsilon, \qquad Z_{r+1,j}^{\varepsilon,k} \triangleq \sum_{\nu\in H_{r+1,j}^{\varepsilon,k}} \xi_\nu^\varepsilon, \tag{6.11}$$

and observe from (6.3) that

$$W^\varepsilon(l/k) = \varepsilon^{\frac{1}{2}} \sum_{j=1}^{l} \left\{ \sum_{i=1}^{r} (Y_{i,j}^{\varepsilon,k} + Z_{i,j}^{\varepsilon,k}) + Z_{r+1,j}^{\varepsilon,k} \right\} + \varepsilon^{\frac{1}{2}} V_l^{\varepsilon,k} \tag{6.12}$$

for all $l = 1, 2, \ldots, k$. Next, define the kd-vector $\tilde{Y}_{i,j}^{\varepsilon,k}$ by concatenating the d-dimensional zero vector $(j-1)$ times, and then by joining to this the $(k-j+1)$-fold concatenation of the d-vector $Y_{i,j}^{\varepsilon,k}$, namely, for each $j = 1, 2, \ldots, k$ and $i = 1, 2, \ldots, r$,

$$\tilde{Y}_{i,j}^{\varepsilon,k} \triangleq (\underbrace{0, 0, \ldots, 0}_{d(j-1)\ 0\text{'s}}, \underbrace{(Y_{i,j}^{\varepsilon,k})', \ldots, (Y_{i,j}^{\varepsilon,k})'}_{(k-j+1)\ Y_{i,j}^{\varepsilon,k}\text{'s}})'. \tag{6.13}$$

Likewise, for each $j = 1, 2, \ldots, k$ and $i = 1, 2, \ldots, r+1$, put

$$\tilde{Z}_{i,j}^{\varepsilon,k} \triangleq (\underbrace{0, 0, \ldots, 0}_{d(j-1)\ 0\text{'s}}, \underbrace{(Z_{i,j}^{\varepsilon,k})', \ldots, (Z_{i,j}^{\varepsilon,k})'}_{(k-j+1)\ Z_{i,j}^{\varepsilon,k}\text{'s}})', \tag{6.14}$$

and let $\tilde{V}^{\varepsilon,k}$ be the kd-vector formed by concatenating the d-vectors $V_l^{\varepsilon,k}$, $l = 1, 2, \ldots, k$:

$$\tilde{V}^{\varepsilon,k} \triangleq ((V_1^{\varepsilon,k})', (V_2^{\varepsilon,k})', \ldots, (V_k^{\varepsilon,k})')'. \tag{6.15}$$

From (6.1), (6.12), and (6.13) to (6.15), we find

$$\Xi_k^\varepsilon = \varepsilon^{\frac{1}{2}} \sum_{j=1}^{k} \left\{ \sum_{i=1}^{r} \tilde{Y}_{i,j}^{\varepsilon,k} + \sum_{i=1}^{r+1} \tilde{Z}_{i,j}^{\varepsilon,k} \right\} + \varepsilon^{\frac{1}{2}} \tilde{V}^{\varepsilon,k}. \tag{6.16}$$

Remark 6.6. In the context of summing independent random vectors, Yurinskii [27] introduced the trick of adding a concatenation of zero-vectors to get sums of vectors of common length kd. This motivates the definitions of (6.13) and (6.14). Observe, from (6.11), (6.13), and (6.14), that for all $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, k$, the kd-vectors $\tilde{Y}_{i,j}^{\varepsilon,k}$ and $\tilde{Z}_{i,j}^{\varepsilon,k}$ are derived by effectively summing ξ_ν^ε over the long blocks $G_{i,j}^{\varepsilon,k}$ and short blocks $H_{i,j}^{\varepsilon,k}$, respectively, while $\tilde{Z}_{r+1,j}^{\varepsilon,k}$ is obtained by summing ξ_ν^ε over the (infrequent) remainder blocks $H_{r+1,j}^{\varepsilon,k}$. In the light of (6.16), this suggests

$$\Xi_k^\varepsilon \approx \varepsilon^{\frac{1}{2}} \sum_{j=1}^{k} \sum_{i=1}^{r} \tilde{Y}_{i,j}^{\varepsilon,k},$$

and, since $G_{i,j}^{\varepsilon,k}$ is separated from $G_{i_1,j_1}^{\varepsilon,k}$ by blocks of length q or more when $(i,j) \neq (i_1, j_1)$, it seems plausible that the $\tilde{Y}_{i,j}^{\varepsilon,k}$ for different (i,j) are approximately independent. For this intuition to help in establishing Theorem 6.3 we must use the fact that $\{\xi_\nu^\varepsilon\}$ is geometrically L-mixing (see Remark 4.5). To this end, with q given by (6.7), and recalling (4.8), for each $j = 1, \ldots, k$ and $i = 1, \ldots, r$, define the d-vectors

$$Y_{i,j}^{\varepsilon,k}[q] \triangleq \sum_{\nu \in G_{i,j}^{\varepsilon,k}} \xi_\nu^\varepsilon[q], \qquad Z_{i,j}^{\varepsilon,k}[q] \triangleq \sum_{\nu \in H_{i,j}^{\varepsilon,k}} \xi_\nu^\varepsilon[q], \qquad Z_{r+1,j}^{\varepsilon,k}[q] \triangleq \sum_{\nu \in H_{r+1,j}^{\varepsilon,k}} \xi_\nu^\varepsilon[q]. \tag{6.17}$$

Motivated by (6.13) to (6.15), for each $j = 1, 2, \ldots, k$, put

$$\tilde{Y}_{i,j}^{\varepsilon,k}[q] \triangleq (\underbrace{0, 0, \ldots, 0}_{d(j-1)\ 0\text{'s}}, \underbrace{(Y_{i,j}^{\varepsilon,k}[q])', \ldots, (Y_{i,j}^{\varepsilon,k}[q])'}_{(k-j+1) Y_{i,j}^{\varepsilon,k}[q]\text{'s}})' \quad \forall\, i = 1, \ldots, r, \tag{6.18}$$

$$\tilde{Z}^{\varepsilon,k}_{i,j}[q] \stackrel{\triangle}{=} (\underbrace{0,0,\ldots,0}_{d(j-1)\ 0\text{'s}}, \underbrace{(Z^{\varepsilon,k}_{i,j}[q])',\ldots,(Z^{\varepsilon,k}_{i,j}[q])'}_{(k-j+1) Z^{\varepsilon,k}_{i,j}[q]\text{'s}})' \quad \forall\, i=1,\ldots,r+1, \tag{6.19}$$

$$\tilde{V}^{\varepsilon,k}[q] \stackrel{\triangle}{=} ((V^{\varepsilon,k}_1[q])', (V^{\varepsilon,k}_2[q])', \ldots, (V^{\varepsilon,k}_k[q])')', \tag{6.20}$$

where (motivated by (6.4)) $V^{\varepsilon,k}_l[q] \stackrel{\triangle}{=} (l/(k\varepsilon) - \lfloor l/(\varepsilon k)\rfloor)\xi^{\varepsilon}_{\lfloor l/(\varepsilon k)\rfloor+1}[q]$ for all $l=1,2,\ldots,k$. Also put

$$\Xi^{\varepsilon}_k[q] \stackrel{\triangle}{=} \varepsilon^{\frac{1}{2}} \sum_{j=1}^{k} \left\{ \sum_{i=1}^{r} \tilde{Y}^{\varepsilon,k}_{i,j}[q] + \sum_{i=1}^{r+1} \tilde{Z}^{\varepsilon,k}_{i,j}[q] \right\} + \varepsilon^{\frac{1}{2}} \tilde{V}^{\varepsilon,k}[q]. \tag{6.21}$$

Using the triangle inequality for the Prohorov metric, we can write

$$\begin{aligned} \Pi^{kd}_2(\mathcal{L}(\Xi^{\varepsilon}_k),\mathcal{L}(\hat{\Xi}^0_k)) &\le \Pi^{kd}_2(\mathcal{L}(\Xi^{\varepsilon}_k), \mathcal{L}(\Xi^{\varepsilon}_k[q])) \\ &\quad + \Pi^{kd}_2\left(\mathcal{L}(\Xi^{\varepsilon}_k[q]),\ \mathcal{L}\left(\varepsilon^{\frac{1}{2}} \sum_{j=1}^{k}\sum_{i=1}^{r} \tilde{Y}^{\varepsilon,k}_{i,j}[q]\right)\right) \\ &\quad + \Pi^{kd}_2\left(\mathcal{L}\left(\varepsilon^{\frac{1}{2}} \sum_{j=1}^{k}\sum_{i=1}^{r} \tilde{Y}^{\varepsilon,k}_{i,j}[q]\right),\ \mathcal{N}_{kd}\left(0, \varepsilon \sum_{j=1}^{k}\sum_{i=1}^{r} \operatorname{cov}(\tilde{Y}^{\varepsilon,k}_{i,j}[q])\right)\right) \\ &\quad + \Pi^{kd}_2\left(\mathcal{N}_{kd}\left(0, \varepsilon \sum_{j=1}^{k}\sum_{i=1}^{r} \operatorname{cov}(\tilde{Y}^{\varepsilon,k}_{i,j}[q])\right),\ \mathcal{L}(\hat{\Xi}^0_k)\right). \end{aligned} \tag{6.22}$$

To get Theorem 6.3 we must establish upper bounds for each term on the right-hand side of (6.22). We will write c, c_1, etc. for nonnegative finite constants that may vary from one use to the next.

First term on RHS of (6.22). Define the d-vectors $W^{\varepsilon}_q(l/k) \stackrel{\triangle}{=} \varepsilon^{\frac{1}{2}} \sum_{\nu=1}^{\lfloor l/(k\varepsilon)\rfloor} \xi^{\varepsilon}_{\nu}[q] + V^{\varepsilon,k}_l[q]$ for all $l=1,2,\ldots,k$, (compare (4.7)). Using Cauchy–Schwarz for discrete sums and (4.9), we get constant c such that

$$E[|W^{\varepsilon}(l/k) - W^{\varepsilon}_q(l/k)|^2] \le \varepsilon E\left[\left(1+\frac{l}{\varepsilon k}\right) \sum_{\nu=1}^{1+\lfloor l/(\varepsilon k)\rfloor} |\xi^{\varepsilon}_{\nu} - \xi^{\varepsilon}_{\nu}[q]|^2\right] \le c\varepsilon^{-1}\lambda^q \tag{6.23}$$

for each $\varepsilon \in (0,1]$, $k=1,2,\ldots$, and $1 \le l \le k$. From (6.21) it follows that $\Xi^{\varepsilon}_k[q]$ is given by the concatenation of the d-vectors $W^{\varepsilon}_q(l/k)$, $l=1,2,\ldots,k$ (in just the same way that Ξ^{ε}_k is given in (6.1) by the concatenation of the d-vectors $W^{\varepsilon}(l/k)$, $l=1,\ldots,k$), and so, from (6.23):

$$\|\Xi^{\varepsilon}_k - \Xi^{\varepsilon}_k[q]\|_2 = \left(\sum_{l=1}^{k} E[|W^{\varepsilon}(l/k) - W^{\varepsilon}_q(l/k)|^2]\right)^{\frac{1}{2}} \le c^{\frac{1}{2}} k^{\frac{1}{2}} \varepsilon^{-\frac{1}{2}} \lambda^{\frac{q}{2}}. \tag{6.24}$$

By (6.24) and Lemma 8.6(ii), for all $\varepsilon \in (0,1]$ and $k=1,2,\ldots$, we get

$$\Pi^{kd}_2(\mathcal{L}(\Xi^{\varepsilon}_k), \mathcal{L}(\Xi^{\varepsilon}_k[q])) \le (c^{\frac{1}{2}} k^{\frac{1}{2}} \varepsilon^{-\frac{1}{2}} \lambda^{\frac{q}{2}})^{\frac{2}{3}} = c^{\frac{1}{3}} k^{\frac{1}{3}} \varepsilon^{-\frac{1}{3}} \lambda^{\frac{q}{3}}. \tag{6.25}$$

Second term on RHS of (6.22). From (6.21), (6.19), and (6.20), one easily verifies that the kd-dimensional vector $(\Xi^{\varepsilon}_k[q] - \varepsilon^{\frac{1}{2}} \sum_{j=1}^{k}\sum_{i=1}^{r} \tilde{Y}^{\varepsilon,k}_{i,j}[q])$ is the concatenation of the sequence of d-dimensional vectors $\varepsilon^{\frac{1}{2}}(\sum_{j=1}^{m}\sum_{i=1}^{r+1} Z^{\varepsilon,k}_{i,j}[q] + V^{\varepsilon,k}_m[q])$ for all $m = 1, 2, \ldots, k$, and thus

$$\left\| \Xi_k^\varepsilon[q] - \varepsilon^{\frac{1}{2}} \sum_{j=1}^{k} \sum_{i=1}^{r} \tilde{Y}_{i,j}^{\varepsilon,k}[q] \right\|_2 = \varepsilon^{\frac{1}{2}} \left(\sum_{m=1}^{k} \left\| \sum_{i=1}^{r+1} \sum_{j=1}^{m} Z_{i,j}^{\varepsilon,k}[q] + V_m^{\varepsilon,k}[q] \right\|_2^2 \right)^{\frac{1}{2}} \tag{6.26}$$

$$\leq \varepsilon^{\frac{1}{2}} \sum_{m=1}^{k} \left(\left\| \sum_{i=1}^{r} \sum_{j=1}^{m} Z_{i,j}^{\varepsilon,k}[q] \right\|_2 + \left\| \sum_{j=1}^{m} Z_{r+1,j}^{\varepsilon,k}[q] \right\|_2 + \left\| V_m^{\varepsilon,k}[q] \right\|_2 \right),$$

where we have used the inequality $(\sum_{m=1}^k a_i^2)^{\frac{1}{2}} \leq \sum_{m=1}^k |a_i|$ and then Minkowski's inequality to get the inequality in (6.26). To bound the terms in brackets on the right of (6.26) we shall assume that $d = 1$; the general multivariate case only involves more cumbersome notation. Since $\{\xi_\nu^\varepsilon[q]\}$ is a q-dependent zero-mean process (see Remark 4.5) and distinct blocks $H_{i,j}^{\varepsilon,k}$ are separated by long blocks of cardinality greater than q (recall (6.6) and (6.9)), we see from (6.17) that $Z_{i,j}^{\varepsilon,k}[q]$ and $Z_{i_1,j_1}^{\varepsilon,k}[q]$ are independent zero-mean random variables when $(i, j) \neq (i_1, j_1)$, and thus

$$\left\| \sum_{i=1}^{r} \sum_{j=1}^{m} Z_{i,j}^{\varepsilon,k}[q] \right\|_2 = \left(\sum_{i=1}^{r} \sum_{j=1}^{m} E|Z_{i,j}^{\varepsilon,k}[q]|^2 \right)^{1/2}. \tag{6.27}$$

Now $\{\xi_\nu^\varepsilon\}$ is a geometrically L-mixing zero-mean process; hence Remark 4.6, (6.9), and (6.11) give a constant c such that $\|Z_{i,j}^{\varepsilon,k}\|_2 \leq cq^{\frac{1}{2}}$ for all $\varepsilon \in (0, 1]$, for all $k = 1, 2, \ldots$, for all $i = 1, \ldots, r$, and for all $j = 1, \ldots, k$. Also, from (4.9), (6.11), (6.17), and Minkowski's inequality, we find $\|Z_{i,j}^{\varepsilon,k} - Z_{i,j}^{\varepsilon,k}[q]\|_2 \leq c_1 q \lambda^q$, and thus

$$\|Z_{i,j}^{\varepsilon,k}[q]\|_2 \leq \|Z_{i,j}^{\varepsilon,k}\|_2 + \|Z_{i,j}^{\varepsilon,k}[q] - Z_{i,j}^{\varepsilon,k}\|_2 \leq cq^{\frac{1}{2}} + c_1 q \lambda^q \leq c_2 q^{\frac{1}{2}} \tag{6.28}$$

for some constants c_1, c_2. Hence, by (6.27), we get $\| \sum_{i=1}^r \sum_{j=1}^m Z_{i,j}^{\varepsilon,k}[q]\|_2 \leq c_2 (rmq)^{\frac{1}{2}}$ for all $\varepsilon \in (0, 1]$, for all $k = 1, 2, \ldots$, and for all $m = 1, \ldots, k$. Similarly, we have $\| \sum_{j=1}^m Z_{r+1,j}^{\varepsilon,k}[q]\|_2 \leq c_2(m(p+q))^{\frac{1}{2}}$. Now, from (6.8), (6.7), and $\#J_j^{\varepsilon,k} = \lfloor k^{-1}\varepsilon^{-1} \rfloor$ (see (6.5)), we can find some $\varepsilon_2 \in (0, 1]$ such that

$$r \equiv r_{\varepsilon,k} \leq 2\varepsilon^{-\frac{2}{3}} \qquad \forall\, k = 1, 2, \ldots, \left\lfloor \varepsilon^{-\frac{1}{36}} \right\rfloor, \quad \forall\, \varepsilon \in (0, \varepsilon_2] \tag{6.29}$$

(e.g., $\varepsilon_2 \stackrel{\triangle}{=} 1/16$). But clearly $\|V_m^{\varepsilon,k}[q]\|_2 = O(1)$ (uniformly with respect to ε, k, m); thus from (6.26), (6.7),

$$\left\| \Xi_k^\varepsilon[q] - \varepsilon^{\frac{1}{2}} \sum_{j=1}^{k} \sum_{i=1}^{r} \tilde{Y}_{i,j}^{\varepsilon,k}[q] \right\|_2 \leq c_3 \varepsilon^{\frac{1}{2}} \sum_{m=1}^{k} \{ (rmq)^{\frac{1}{2}} + (m(p+q))^{\frac{1}{2}} + O(1) \} \tag{6.30}$$

$$\leq c_4 \varepsilon^{\frac{1}{2}} k \{ (\varepsilon^{-\frac{2}{3}} k k^{-1} \varepsilon^{-\frac{1}{9}})^{\frac{1}{2}} + (k k^{-1} \varepsilon^{-\frac{1}{3}})^{\frac{1}{2}} + O(1) \} \leq c_5 k \varepsilon^{\frac{1}{9}}$$

for some constants c_3, c_4, c_5; hence, by Lemma 8.6(ii), for each $\varepsilon \in (0, \varepsilon_2]$ we have

$$\Pi_2^{kd} \left(\mathcal{L}(\Xi_k^\varepsilon[q]), \mathcal{L}\left(\varepsilon^{\frac{1}{2}} \sum_{j=1}^{k} \sum_{i=1}^{r} \tilde{Y}_{i,j}^{\varepsilon,k}[q] \right) \right) \leq (c_5 k \varepsilon^{\frac{1}{9}})^{\frac{2}{3}} = c_5^{\frac{2}{3}} k^{\frac{2}{3}} \varepsilon^{\frac{2}{27}} \tag{6.31}$$

for all $k = 1, 2, \ldots, \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$.

Third term on RHS of (6.22). Here we shall need the following special case of a finite-dimensional CLT for sums of independent random vectors due to Yurinskii [27, Theorem 1]. Note that the upper-bound we state here is somewhat less precise than the bound given in [27, Theorem 1] but is easier to use.

THEOREM 6.7. *There is a constant $\gamma \in (0, \infty)$ with the following property. For all positive integers m and finite sequences $\{u_1, u_2, \ldots, u_n\}$ of zero-mean independent $\Re^m$-valued random vectors we have*

$$\Pi_2^m\left(\mathcal{L}\left(\sum_{\alpha=1}^{n} u_\alpha\right), \mathcal{N}_m\left(0, \sum_{\alpha=1}^{n} \operatorname{cov}(u_\alpha)\right)\right) \leq \gamma m^{\frac{1}{3}}(\mu_n)^{\frac{2}{9}} \quad \textit{when} \quad \mu_n \stackrel{\triangle}{=} \sum_{\alpha=1}^{n} E|u_\alpha|^3 < 1.$$

The $\tilde{Y}_{i,j}^{\varepsilon,k}[q]$ given by (6.18) are zero-mean and independent as i and j vary (since $\{\xi_\nu^\varepsilon[q]\}$ is a q-dependent process, and the long blocks $G_{i,j}^{\varepsilon,k}$ are separated by either short blocks $H_{i,j}^{\varepsilon,k}$ or short block/remainder block pairs $H_{r,j}^{\varepsilon,k}, H_{r+1,j}^{\varepsilon,k}$, all of which have length at least q). In Theorem 6.7 we will identify m with kd, n with kr, the summation index α with (i,j) (for $j = 1, \ldots, k$ and $i = 1, \ldots, r$), u_α with $\varepsilon^{\frac{1}{2}}\tilde{Y}_{i,j}^{\varepsilon,k}[q]$, and μ_n with $\mu_{kr} \stackrel{\triangle}{=} \sum_{j=1}^{k}\sum_{i=1}^{r} \varepsilon^{\frac{3}{2}} E|\tilde{Y}_{i,j}^{\varepsilon,k}[q]|^3$. From (6.18) we have $|\tilde{Y}_{i,j}^{\varepsilon,k}[q]|^2 = (k - j+1)|Y_{i,j}^{\varepsilon,k}[q]|^2 \leq k|Y_{i,j}^{\varepsilon,k}[q]|^2$, and, exactly as for (6.28), we can use Remark 4.6 to find constant c_1 such that $\|Y_{i,j}^{\varepsilon,k}[q]\|_3 \leq c_1 p^{\frac{1}{2}}$ for all $\varepsilon \in (0,1]$, for all $k = 1, 2, \ldots$, for all $i = 1, \ldots, r$, and for all $j = 1, \ldots, k$. Thus, from (6.7), we have $E|\tilde{Y}_{i,j}^{\varepsilon,k}[q]|^3 \leq c_1^3 k^{\frac{3}{2}} p^{\frac{3}{2}} = c_1^3 \varepsilon^{-\frac{1}{2}}$; hence (6.29) gives $\mu_{kr} = \sum_{j=1}^{k}\sum_{i=1}^{r} \varepsilon^{\frac{3}{2}} E|\tilde{Y}_{i,j}^{\varepsilon,k}[q]|^3 \leq 2c_1^3 k \varepsilon^{\frac{1}{3}}$. Define

$$\tilde{T}^{\varepsilon,k} \stackrel{\triangle}{=} \varepsilon \sum_{j=1}^{k}\sum_{i=1}^{r} \operatorname{cov}(\tilde{Y}_{i,j}^{\varepsilon,k}[q]) \equiv \sum_{\alpha} \operatorname{cov}(u_\alpha), \tag{6.32}$$

and fix some $\varepsilon_3 \in (0, \varepsilon_2]$ such that $2c_1^3 k \varepsilon^{\frac{1}{3}} < 1$ for all $\varepsilon \in (0, \varepsilon_3]$, $k = 1, 2, \ldots, \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$. Then Theorem 6.7 gives a constant c_2 such that, for all $\varepsilon \in (0, \varepsilon_3]$, $k = 1, 2, \ldots, \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$, we have

$$\Pi_2^{kd}\left(\mathcal{L}\left(\varepsilon^{\frac{1}{2}} \sum_{j=1}^{k}\sum_{i=1}^{r} \tilde{Y}_{i,j}^{\varepsilon,k}[q]\right), \mathcal{N}_{kd}(0, \tilde{T}^{\varepsilon,k})\right) \leq c_1 (kd)^{\frac{1}{3}} (k\varepsilon^{\frac{1}{3}})^{\frac{2}{9}} = c_2 k^{\frac{5}{9}} \varepsilon^{\frac{2}{27}}. \tag{6.33}$$

Fourth term on RHS of (6.22). To simplify the notation, we use the following convention. If B is a $kd \times kd$ matrix, then $B_{m,n}$ denotes the (m,n)th of the $d \times d$ submatrices into which B can be partitioned for all $m, n = 1, 2, \ldots, k$. If v is a kd-vector, then v_m denotes the d-vector consisting of the $(d(m-1)+1)$th to (dm)th elements of v. Fix arbitrary $\varepsilon \in (0,1]$ and $k = 1, 2, \ldots$. Define the $kd \times kd$ matrices $\hat{T}^{\varepsilon,k}$ and T^k such that their (m,n)th d-dimensional submatrices are respectively given by

$$\hat{T}_{m,n}^{\varepsilon,k} \stackrel{\triangle}{=} \varepsilon \sum_{j=1}^{m\wedge n}\sum_{i=1}^{r} \sum_{\nu \in G_{i,j}^{\varepsilon,k}} A(\theta^0((\nu-1)\varepsilon)) \quad \text{and} \quad T_{m,n}^k \stackrel{\triangle}{=} (\operatorname{cov}\hat{\Xi}_k^0)_{m,n} \tag{6.34}$$

for all $m, n = 1, 2, \ldots, k$, where $A(\cdot)$ is given by condition (C4). Then, from (6.1) and (3.8), we see that $\mathcal{L}(\hat{\Xi}_k^0) = \mathcal{N}_{kd}(0, T^k)$, so that the fourth term on the right side of (6.22) is bounded as follows:

$$\begin{aligned}\Pi_2^{kd}(\mathcal{N}_{kd}(0, \tilde{T}^{\varepsilon,k}), \mathcal{L}(\hat{\Xi}_k^0)) &\leq \Pi_2^{kd}(\mathcal{N}_{kd}(0, \tilde{T}^{\varepsilon,k}), \mathcal{N}_{kd}(0, \hat{T}^{\varepsilon,k})) \\ &\quad + \Pi_2^{kd}(\mathcal{N}_{kd}(0, \hat{T}^{\varepsilon,k}), \mathcal{N}_{kd}(0, T^k)).\end{aligned} \tag{6.35}$$

We will now use Theorem 8.7 to bound each of the terms on the right side of (6.35). In order to deal with the first term we find an expression for $|\tilde{T}^{\varepsilon,k}_{m,n} - \hat{T}^{\varepsilon,k}_{m,n}|$. From (6.18),

$$[\mathrm{cov}(\tilde{Y}^{\varepsilon,k}_{i,j}[q])]_{m,n} = \begin{cases} \mathrm{cov}(Y^{\varepsilon,k}_{i,j}[q]), & \text{when } j \le (m \wedge n), \\ 0, & \text{when } (m \wedge n) < j, \end{cases}$$

and thus, using (6.32),

$$\tilde{T}^{\varepsilon,k}_{m,n} = \varepsilon \sum_{j=1}^{m\wedge n} \sum_{i=1}^{r} \mathrm{cov}(Y^{\varepsilon,k}_{i,j}[q]). \tag{6.36}$$

On combining (6.36), (6.34), and (6.17), we get, for all $m, n = 1, 2, \ldots, k$,

$$|\tilde{T}^{\varepsilon,k}_{m,n} - \hat{T}^{\varepsilon,k}_{m,n}| = \varepsilon \left| \sum_{j=1}^{m\wedge n} \sum_{i=1}^{r} \left[\mathrm{cov}\left(\sum_{\nu \in G^{\varepsilon,k}_{i,j}} \xi^{\varepsilon}_{\nu}[q] \right) - \sum_{\nu \in G^{\varepsilon,k}_{i,j}} A(\theta^0((\nu-1)\varepsilon)) \right] \right|. \tag{6.37}$$

To upper-bound the term in square braces on the right-hand side of (6.37), partition the long block $G^{\varepsilon,k}_{i,j}$ into *adjacent* blocks of integers $Q^{\varepsilon,k}_{\alpha,i,j}$ ordered according to the pattern

$$Q^{\varepsilon,k}_{1,i,j}, Q^{\varepsilon,k}_{2,i,j}, \ldots, Q^{\varepsilon,k}_{l,i,j}, Q^{\varepsilon,k}_{l+1,i,j}, \tag{6.38}$$

where $l \stackrel{\triangle}{=} \lfloor p/q^2 \rfloor$. The blocks $Q^{\varepsilon,k}_{\alpha,i,j}$, $\alpha = 1, \ldots, l$ have common cardinality q^2, and $Q^{\varepsilon,k}_{l+1,i,j}$ is a "remainder block" whose cardinality is less than q^2. Recalling (6.7), fix some $\varepsilon_{41} \in (0, \varepsilon_3]$ such that

$$l \equiv l_{\varepsilon,k} \stackrel{\triangle}{=} \lfloor p/q^2 \rfloor \le 2k\varepsilon^{-\frac{1}{9}} \qquad \forall\, k = 1, 2, \ldots, \lfloor \varepsilon^{-\frac{1}{36}} \rfloor, \quad \forall\, \varepsilon \in (0, \varepsilon_{41}]. \tag{6.39}$$

If $s^{\varepsilon,k}_{\alpha,i,j}$ denotes the first integer of the block $Q^{\varepsilon,k}_{\alpha,i,j}$ then clearly

$$s^{\varepsilon,k}_{\alpha,i,j} \stackrel{\triangle}{=} b^{\varepsilon,k}_{i,j} + (\alpha - 1)q^2 \qquad \forall\, \alpha = 1, \ldots, l+1 \tag{6.40}$$

(recall that $b^{\varepsilon,k}_{i,j}$ is the first integer in $G^{\varepsilon,k}_{i,j}$), so that $Q^{\varepsilon,k}_{\alpha,i,j} \stackrel{\triangle}{=} [s^{\varepsilon,k}_{\alpha,i,j}, s^{\varepsilon,k}_{\alpha,i,j} + q^2)$ for all $\alpha = 1, \ldots, l$, while $Q^{\varepsilon,k}_{l+1,i,j} \stackrel{\triangle}{=} [s^{\varepsilon,k}_{l+1,i,j}, b^{\varepsilon,k}_{i,j} + p)$. Henceforth, to lighten the notation, we write Q_α for $Q^{\varepsilon,k}_{\alpha,i,j}$ and s_α for $s^{\varepsilon,k}_{\alpha,i,j}$ when convenient. Also, recalling (2.7) and (2.3), put

$$\tilde{H}_n(\theta)[s] \stackrel{\triangle}{=} E[\tilde{H}_n(\theta) | \mathcal{F}^n_{n-s}] = \tilde{b}_n[s] - \tilde{R}_n[s]\theta \qquad \forall\, s, n = 1, 2, \ldots, \quad \theta \in \Re^d, \tag{6.41}$$

and observe from (4.8) that $\xi^{\varepsilon}_{\nu}[q] = \tilde{H}_{\nu}(\theta^0((\nu-1)\varepsilon))[q]$ for all $\nu \in G^{\varepsilon,k}_{i,j}$. One then sees from (6.37) that, for each $m, n = 1, \ldots, k$,

$$|\tilde{T}^{\varepsilon,k}_{m,n} - \hat{T}^{\varepsilon,k}_{m,n}| \le \varepsilon \sum_{j=1}^{m\wedge n} \sum_{i=1}^{r} \{ \mathrm{I}^{\varepsilon,k}_{i,j} + \mathrm{II}^{\varepsilon,k}_{i,j} + \mathrm{III}^{\varepsilon,k}_{i,j} + \mathrm{IV}^{\varepsilon,k}_{i,j} \}, \tag{6.42}$$

where

$$\mathrm{I}_{i,j}^{\varepsilon,k} \triangleq \left| \mathrm{cov}\left(\sum_{\nu \in G_{i,j}^{\varepsilon,k}} \xi_\nu^\varepsilon[q] \right) - \sum_{\alpha=1}^{l+1} \mathrm{cov}\left(\sum_{\nu \in Q_\alpha} \xi_\nu^\varepsilon[q] \right) \right|, \tag{6.43}$$

$$\mathrm{II}_{i,j}^{\varepsilon,k} \triangleq \left| \sum_{\alpha=1}^{l+1} \left\{ \mathrm{cov}\left(\sum_{\nu \in Q_\alpha} \xi_\nu^\varepsilon[q] \right) - \mathrm{cov}\left(\sum_{\nu \in Q_\alpha} \tilde{H}_\nu(\theta^0(\varepsilon s_\alpha))[q] \right) \right\} \right|, \tag{6.44}$$

$$\mathrm{III}_{i,j}^{\varepsilon,k} \triangleq \left| \sum_{\alpha=1}^{l+1} \left\{ \mathrm{cov}\left(\sum_{\nu \in Q_\alpha} \tilde{H}_\nu(\theta^0(\varepsilon s_\alpha))[q] \right) - \left(\#Q_\alpha \right) A(\theta^0(\varepsilon s_\alpha)) \right\} \right|, \tag{6.45}$$

$$\mathrm{IV}_{i,j}^{\varepsilon,k} \triangleq \left| \sum_{\alpha=1}^{l+1} (\#Q_\alpha) A(\theta^0(\varepsilon s_\alpha)) - \sum_{\nu \in G_{i,j}^{\varepsilon,k}} A(\theta^0((\nu-1)\varepsilon)) \right|. \tag{6.46}$$

Remark 6.8. Our calculation of an upper-bound for $|\tilde{T}_{m,n}^{\varepsilon,k} - \hat{T}_{m,n}^{\varepsilon,k}|$ is closely based on the proof of Lemma 3.1 of Khas'minskii [13], which motivates introduction of the blocks Q_α and the upper-bound (6.42) in terms of the quantities (6.43) to (6.46). The idea is that over the blocks Q_α we "freeze" the time-varying quantity $\theta^0(\varepsilon(\nu-1))$, $\nu \in G_{i,j}^{\varepsilon,k}$, at the value $\theta^0(\varepsilon s_\alpha)$ corresponding to the first member s_α of Q_α. The block Q_α is long enough for the averaging postulated in condition (C4) to come into play (where we identify $\{n_0, \ldots, N+n_0\}$ and θ in (2.6) with Q_α and $\theta^0(\varepsilon s_\alpha)$, respectively) and provide an upper bound for (6.45). At the same time, Q_α is short enough to ensure that $\theta^0((\nu-1)\varepsilon) \approx \theta^0(\varepsilon s_\alpha)$, and this will give us upper-bounds for (6.44) and (6.46). Finally, we shall use L-mixing of $\{\xi_\nu^\varepsilon\}$ to bound (6.43). We now proceed to make this intuition precise. For the term in (6.43), put

$$B_{i,j}^{\varepsilon,k} \triangleq (G_{i,j}^{\varepsilon,k} \times G_{i,j}^{\varepsilon,k}) - \bigcup_{\alpha=1}^{l+1} (Q_\alpha \times Q_\alpha) \tag{6.47}$$

(recall that Q_α is short for $Q_{\alpha,i,j}^{\varepsilon,k}$) and let $B_+ \triangleq \{(\nu,\mu) \in B_{i,j}^{\varepsilon,k} : \mu > \nu\}$, $D_\alpha \triangleq \{(\nu,\mu) \in \mathcal{Z}^2 : -\infty < \nu < s_\alpha \le \mu < \infty\}$ for all $\alpha = 2, \ldots, l+1$. It follows from (6.47) that

$$B_+ \subset \bigcup_{\alpha=2}^{l+1} D_\alpha. \tag{6.48}$$

By (6.39), (6.48), and Remark 8.4, we have constants $\lambda \in (0,1)$ and c, c_1 with

$$\mathrm{I}_{i,j}^{\varepsilon,k} = \left| \sum_{(\nu,\mu) \in B_{i,j}^{\varepsilon,k}} E\big[(\xi_\nu^\varepsilon[q])(\xi_\mu^\varepsilon[q])'\big] \right| \le c \sum_{(\nu,\mu) \in B_+} \lambda^{\mu-\nu} \tag{6.49}$$

$$\le c \sum_{\alpha=2}^{l+1} \left\{ \sum_{(\nu,\mu) \in D_\alpha} \lambda^{\mu-\nu} \right\} \le c \sum_{\alpha=2}^{l+1} \left\{ \sum_{n=1}^{\infty} n\lambda^n \right\} \le c_1 l = c_1 l_{\varepsilon,k} \le c_1 k \varepsilon^{-\frac{1}{9}}$$

for all $\varepsilon \in (0, \varepsilon_{41}]$, $k = 1, \ldots, \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$, $i = 1, \ldots, r$, $j = 1, \ldots, k$. Next, we upper-bound the term in (6.44): put $\eta_\nu^1 \triangleq \xi_\nu^\varepsilon[q] \equiv \tilde{H}_\nu(\theta^0((\nu-1)\varepsilon))[q]$ and $\eta_\nu^2 \triangleq \tilde{H}_\nu(\theta^0(\varepsilon s_\alpha))[q]$ for all $\nu \in Q_\alpha$. Then

$$\operatorname{cov}\left(\sum_{\nu \in Q_\alpha} \eta_\nu^1\right) - \operatorname{cov}\left(\sum_{\nu \in Q_\alpha} \eta_\nu^2\right) = \sum_{\nu,\mu \in Q_\alpha} E\left[(\eta_\nu^1 - \eta_\nu^2)(\eta_\mu^1)' + (\eta_\nu^2)(\eta_\mu^1 - \eta_\mu^2)'\right]. \tag{6.50}$$

Since $\{\theta^0(\tau),\ \tau \in [0,1]\}$ is Lipschitz continuous and $\{\tilde{R}_n[s]\}$, $\{\eta_n^1\}$ are clearly L_2-bounded (uniformly in $s, n = 1,2,\ldots$), we find constant c such that

$$\left|E\left[(\eta_\nu^1 - \eta_\nu^2)(\eta_\mu^1)'\right]\right| \le \|\tilde{R}_\nu[q]\|_2 \, \|\eta_\mu^1\|_2 |\theta^0(\varepsilon s_\alpha) - \theta^0(\varepsilon(\nu-1))| \le c\varepsilon(\#Q_\alpha) \tag{6.51}$$

for all $\mu,\ \nu \in Q_\alpha$, with an identical bound for $|E[(\eta_\nu^2)(\eta_\mu^1 - \eta_\mu^2)']|$. From (6.39), (6.50), and (6.51), we find

$$\mathrm{II}_{i,j}^{\varepsilon,k} \le c\varepsilon l(\#Q_\alpha)^3 \le c\varepsilon(2k\varepsilon^{-\frac{1}{9}})(k^{-2}\varepsilon^{-\frac{2}{9}})^3 \le c_1\varepsilon^{\frac{2}{9}} \tag{6.52}$$

for all $\varepsilon \in (0, \varepsilon_{41}]$, $k = 1,2,\ldots,\lfloor \varepsilon^{-\frac{1}{36}} \rfloor$, $i = 1,\ldots,r$, $j = 1,\ldots,k$. For the term given by (6.45), we need the following lemma.

Lemma 6.9 (proved in section 7). *Suppose conditions* (C1)–(C4) *of section* 2, *and let* $\{\tilde{H}_n(\theta)[s]\}$ *be defined by* (6.41). *Then there are constants* $C \in [0,\infty)$ *and* $\lambda \in (0,1)$ *such that*

$$\left|\operatorname{cov}\left(\sum_{n=n_0}^{N+n_0} \tilde{H}_n(\theta)[s]\right) - A(\theta)(N+1)\right| \le C[1+|\theta|^2][1+(N+1)^2\lambda^s]$$

for all $\theta \in \Re^d$, $s, N, n_0 = 1,2,3,\ldots$.

By Lemma 6.9 (identifying θ, s, and the interval $\{n_0,\ldots,n_0+N\}$ in Lemma 6.9 with $\theta^0(\varepsilon s_\alpha)$, q, and Q_α, respectively), uniform boundedness of $\theta^0(((n-1)\varepsilon) \wedge T)$ (in (ε, n)), and recalling $\#(Q_\alpha) = q^2$, we find constants c, c_1, such that

$$\mathrm{III}_{i,j}^{\varepsilon,k} \le c\sum_{\alpha=1}^{l+1}[1+q^4\lambda^q] \le c_1 l \le c_1 k\varepsilon^{-\frac{1}{9}} \tag{6.53}$$

for all $\varepsilon \in (0,\varepsilon_{41}]$ and $k = 1,2,\ldots,\lfloor \varepsilon^{-\frac{1}{36}} \rfloor$, $i = 1,\ldots,r$, $j = 1,\ldots,k$. Now we bound the term given by (6.46). From Lemma 2.9 we see that $\tau \to A(\theta^0(\tau))$ is globally Lipschitz continuous over the interval $\tau \in [0,1]$, and hence there are constants c, c_1 such that, for all $\varepsilon \in (0,\varepsilon_{41}]$ and $k = 1,2,\ldots,\lfloor \varepsilon^{-\frac{1}{36}} \rfloor$,

$$\begin{aligned}\mathrm{IV}_{i,j}^{\varepsilon,k} &\le c\sum_{\alpha=1}^{l+1}\sum_{\nu \in Q_\alpha} |\varepsilon s_\alpha - (\nu-1)\varepsilon| \le c_1\varepsilon l\,(\#Q_\alpha)^2 \\ &\le c_1\varepsilon(k\varepsilon^{-\frac{1}{9}})(k^{-2}\varepsilon^{-\frac{2}{9}})^2 \le c_1\varepsilon^{\frac{4}{9}}.\end{aligned} \tag{6.54}$$

Combining (6.42), and (6.49) to (6.54), and using (6.29) gives constants c, c_1, such that for all $\varepsilon \in (0,\varepsilon_{41}]$ and all $k = 1,2,\ldots,\lfloor \varepsilon^{-\frac{1}{36}} \rfloor$,

$$\begin{aligned}\max_{1 \le m,n \le k} |\tilde{T}_{m,n}^{\varepsilon,k} - \hat{T}_{m,n}^{\varepsilon,k}| &\le c\varepsilon\sum_{j=1}^{k}\sum_{i=1}^{r}\{k\varepsilon^{-\frac{1}{9}} + \varepsilon^{\frac{2}{9}} + k\varepsilon^{-\frac{1}{9}} + \varepsilon^{\frac{4}{9}}\} \\ &\le c_1(\varepsilon kr)(k\varepsilon^{-\frac{1}{9}}) \le c_1\varepsilon^{\frac{1}{6}}.\end{aligned} \tag{6.55}$$

Thus, there must exist some $\varepsilon_{42} \in (0, \varepsilon_{41}]$ such that $kd^{\frac{1}{2}} \max_{1 \le m,n \le k} |\tilde{T}^{\varepsilon,k}_{m,n} - \hat{T}^{\varepsilon,k}_{m,n}| \le c_1 \varepsilon^{\frac{1}{6}} kd^{\frac{1}{2}} < 1$ for all $\varepsilon \in (0, \varepsilon_{42}]$ and $k = 1, 2 \ldots \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$, and we can use Theorem 8.7 with (6.55) to find a constant c such that

$$\Pi_2^{kd}(\mathcal{N}_{kd}(0, \tilde{T}^{\varepsilon,k}), \mathcal{N}_{kd}(0, \hat{T}^{\varepsilon,k})) \le ck^{\frac{3}{4}} (\varepsilon^{\frac{1}{6}})^{\frac{1}{4}} \le c\varepsilon^{\frac{1}{48}} \tag{6.56}$$

for all $\varepsilon \in (0, \varepsilon_{42}]$ and $k = 1, 2, \ldots, \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$. To bound the second term on the right side of (6.35), fix some $1 \le m, n \le k$. One sees from (3.8) that $\{\hat{W}^0(\tau),\ 0 \le \tau \le 1\}$ is a zero-mean Gaussian process with independent increments and covariance $\int_0^\tau A(\theta^0(s))\, ds$. Thus, from (6.1) and (6.34), we have $T^k_{m,n} = \varepsilon \int_0^{m \wedge n/(k\varepsilon)} A(\theta^0(u\varepsilon))\, du$, and hence

$$\begin{aligned} T^k_{m,n} - \hat{T}^{\varepsilon,k}_{m,n} &= \varepsilon \sum_{j=1}^{m \wedge n} \sum_{i=1}^{r} \sum_{\nu \in H^{\varepsilon,k}_{i,j}} A(\theta^0((\nu - 1)\varepsilon)) \\ &\quad + \varepsilon \left(\int_0^{m \wedge n/(k\varepsilon)} [A(\theta^0(u\varepsilon)) - A(\theta^0(\lfloor u \rfloor \varepsilon))]\, du \right). \end{aligned} \tag{6.57}$$

Now, the last term on the right of (6.57) is clearly $O(\varepsilon)$ uniformly with respect to $1 \le m, n \le k \le \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$ (since Lemma 2.9 ensures that $\tau \to A(\theta^0(\tau))$ is globally Lipschitz continuous over $\tau \in [0,1]$). Hence, using (6.57), there must be constants c, c_1 such that for $\varepsilon \in (0, \varepsilon_{42}]$ and $k = 1, 2, \ldots, \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$, we have

$$\begin{aligned} \max_{1 \le m,n \le k} |\hat{T}^{\varepsilon,k}_{m,n} - T^k_{m,n}| &\le c\varepsilon kr(\# H^{\varepsilon,k}_{i,j}) + O(\varepsilon) \\ &\le c\varepsilon k(2\varepsilon^{-\frac{2}{3}}) k^{-1} \varepsilon^{-\frac{1}{9}} + O(\varepsilon) \le c_1 \varepsilon^{\frac{2}{9}}. \end{aligned} \tag{6.58}$$

By (6.58) there is clearly some $\varepsilon_{43} \in (0, \varepsilon_{42}]$ such that $kd^{\frac{1}{2}} \max_{1 \le m,n \le k} |\hat{T}^{\varepsilon,k}_{m,n} - T^k_{m,n}| < 1$ for all $\varepsilon \in (0, \varepsilon_{43}]$ and $k = 1, 2 \ldots \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$, and hence, by Theorem 8.7 and (6.58), we get

$$\Pi_2^{kd}(\mathcal{N}_{kd}(0, \hat{T}^{\varepsilon,k}), \mathcal{N}_{kd}(0, T^k)) \le c_1 k^{\frac{3}{4}} (\varepsilon^{\frac{2}{9}})^{\frac{1}{4}} \le c_1 \varepsilon^{\frac{5}{144}} \tag{6.59}$$

for all $\varepsilon \in (0, \varepsilon_{43}]$ and $k = 1, 2 \ldots \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$. Combining (6.32), (6.35), (6.56), and (6.59), we find a constant c such that

$$\Pi_2^{kd}\left(\mathcal{N}_{kd}\left(0, \varepsilon \sum_{j=1}^{k} \sum_{i=1}^{r} \operatorname{cov} \tilde{Y}^{\varepsilon,k}_{i,j}[q] \right), \mathcal{L}(\hat{\Xi}^0_k) \right) \le c\varepsilon^{\frac{1}{48}} \tag{6.60}$$

for all $\varepsilon \in (0, \varepsilon_{43}]$, $k = 1, 2, \ldots, \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$. Finally, we see that the four terms on the right of (6.22) are upper-bounded by (6.25), (6.31), (6.33), and (6.60), respectively. Since $\lambda \in (0,1)$ we have $\lambda^{\varepsilon^{-\beta}} = O(\varepsilon^\alpha)$ for constants $\alpha, \beta \in (0, \infty)$. Thus, there must be constants c, c_1, such that

$$\Pi_2^{kd}(\mathcal{L}(\Xi^\varepsilon_k), \mathcal{L}(\hat{\Xi}^0_k)) \le c(k^{\frac{1}{3}} \varepsilon^{-\frac{1}{3}} \lambda^{\frac{q}{3}} + k^{\frac{2}{3}} \varepsilon^{\frac{2}{27}} + k^{\frac{5}{9}} \varepsilon^{\frac{2}{27}} + \varepsilon^{\frac{1}{48}}) \le c_1 \varepsilon^{\frac{1}{48}}$$

for all $\varepsilon \in (0, \varepsilon_{43}]$ and $k = 1, 2, \ldots, \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$. Now Theorem 6.3 follows with $\varepsilon_0 \triangleq \varepsilon_{43}$. □

Proof of Theorem 4.4. For each $k = 1, 2, 3, \ldots$, $\varepsilon \in (0,1]$ define continuous process $\{W^\varepsilon_k(\tau),\ \tau \in [0,1]\}$ by

$$W^\varepsilon_k(\tau) \triangleq \begin{cases} W^\varepsilon(\tau) & \text{for } \tau = i/k,\ i = 0, 1, \ldots, k, \\ \text{linear interpolation,} & \text{otherwise.} \end{cases} \tag{6.61}$$

In the same way, define $\{\hat{W}_k^0(\tau),\ 0\le\tau\le 1\}$ in terms of $\{\hat{W}^0(\tau),\ 0\le\tau\le 1\}$. Also, put $\gamma \stackrel{\triangle}{=} 72$ and $k(\varepsilon) \stackrel{\triangle}{=} \lfloor \varepsilon^{-\frac{1}{\gamma}} \rfloor$ for all $\varepsilon \in (0,1]$. By the triangle inequality we can write

$$\text{(6.62)}\quad \Pi_C(\mathcal{L}(W^\varepsilon),\mathcal{L}(\hat{W}^0)) \le \Pi_C(\mathcal{L}(W^\varepsilon),\mathcal{L}(W^\varepsilon_{k(\varepsilon)})) + \Pi_C(\mathcal{L}(W^\varepsilon_{k(\varepsilon)}),\mathcal{L}(\hat{W}^0_{k(\varepsilon)})) + \Pi_C(\mathcal{L}(\hat{W}^0_{k(\varepsilon)}),\mathcal{L}(\hat{W}^0)).$$

Now upper-bound each term on the right side of (6.62) as follows.

Third term on RHS of (6.62). By the definition of $\{\hat{W}_k^0(\tau),\ 0\le\tau\le 1\}$ (following (6.61)), we have

$$\text{(6.63)}\quad \|\hat{W}^0 - \hat{W}^0_{k(\varepsilon)}\|_C \le 2 \max_{i=0,1,\dots,k(\varepsilon)-1} \left\{ \max_{\frac{i}{k(\varepsilon)}\le\tau\le\frac{i+1}{k(\varepsilon)}} \left| \hat{W}^0(\tau) - \hat{W}^0\left(\frac{i}{k(\varepsilon)}\right)\right| \right\},$$

and thus, from the Chebyshev inequality,

$$\text{(6.64)}\quad \begin{aligned} &\hat{P}[\|\hat{W}^0 - \hat{W}^0_{k(\varepsilon)}\|_C \ge \varepsilon^{\frac{1}{8\gamma}}] \\ &\le \sum_{i=0}^{k(\varepsilon)-1} \hat{P}\left\{ \max_{\frac{i}{k(\varepsilon)}\le\tau\le\frac{(i+1)}{k(\varepsilon)}} \left| \hat{W}^0(\tau) - \hat{W}^0\left(\frac{i}{k(\varepsilon)}\right)\right| \ge \frac{1}{2}\varepsilon^{\frac{1}{8\gamma}} \right\} \\ &\le 16\varepsilon^{-\frac{1}{2\gamma}} \sum_{i=0}^{k(\varepsilon)-1} \hat{E}\left\{ \max_{\frac{i}{k(\varepsilon)}\le\tau\le\frac{(i+1)}{k(\varepsilon)}} \left| \hat{W}^0(\tau) - \hat{W}^0\left(\frac{i}{k(\varepsilon)}\right)\right|^4 \right\}. \end{aligned}$$

By Doob's maximal L_p inequality [8, Proposition 2.16, p. 63] applied to the nonnegative submartingale $\{|\hat{W}^0(\tau) - \hat{W}^0(i/k(\varepsilon))|,\ \tau \in [i/k(\varepsilon), 1]\}$, we get

$$\text{(6.65)}\quad \begin{aligned} &\hat{E}\left\{ \max_{\frac{i}{k(\varepsilon)}\le\tau\le\frac{(i+1)}{k(\varepsilon)}} \left| \hat{W}^0(\tau) - \hat{W}^0\left(\frac{i}{k(\varepsilon)}\right)\right|^4 \right\} \\ &\le \left(\frac{4}{3}\right)^4 \hat{E}\left| \hat{W}^0\left(\frac{i+1}{k(\varepsilon)}\right) - \hat{W}^0\left(\frac{i}{k(\varepsilon)}\right)\right|^4 \ll k^{-2}(\varepsilon). \end{aligned}$$

Combining (6.64) and (6.65) gives

$$\hat{P}[\|\hat{W}^0 - \hat{W}^0_{k(\varepsilon)}\|_C \ge \varepsilon^{\frac{1}{8\gamma}}] \ll k^{-1}(\varepsilon)\varepsilon^{-\frac{1}{2\gamma}} \le \varepsilon^{\frac{1}{2\gamma}} \le \varepsilon^{\frac{1}{8\gamma}},$$

and thus, by Lemma 8.6(i), we get

$$\text{(6.66)}\quad \Pi_C(\mathcal{L}(\hat{W}^0),\mathcal{L}(\hat{W}^0_{k(\varepsilon)})) \ll \varepsilon^{\frac{1}{8\gamma}}.$$

Second term on RHS of (6.62). We need the next result, which is suggested by the arguments on page 246 of Yurinskii [27].

LEMMA 6.10 (proved in section 7). *For Ξ_k^ε and $\hat{\Xi}_k^0$ given by* (6.1)*, and W^ε and $\hat{W}^0$ given by* (6.61)*, we have*

$$\Pi_C(\mathcal{L}(W_k^\varepsilon),\mathcal{L}(\hat{W}_k^0)) \le \Pi_2^{kd}(\mathcal{L}(\Xi_k^\varepsilon),\mathcal{L}(\hat{\Xi}_k^0)) \quad \forall\, k = 1,2,\dots, \quad \varepsilon \in (0,1].$$

From Lemma 6.10, Theorem 6.3, and the fact that $k(\varepsilon) < \lfloor \varepsilon^{-\frac{1}{36}} \rfloor$,

$$\text{(6.67)}\quad \Pi_C(\mathcal{L}(W^\varepsilon_{k(\varepsilon)}),\mathcal{L}(\hat{W}^0_{k(\varepsilon)})) \le \Pi_2^{kd}(\mathcal{L}(\Xi^\varepsilon_{k(\varepsilon)}),\mathcal{L}(\hat{\Xi}^0_{k(\varepsilon)})) \ll \varepsilon^{\frac{1}{48}}.$$

First term on RHS of (6.62). For continuous $\varphi : [0,1] \to \Re$ and $0 \le \tau_1 \le \mu_1 < \mu_2 \le \tau_2 \le 1$ we have $\max_{\mu_1 \le \tau \le \mu_2} |\varphi(\tau) - \varphi(\mu_1)| \le 2 \max_{\tau_1 \le \tau \le \tau_2} |\varphi(\tau) - \varphi(\tau_1)|$. Thus, from (3.5) and (4.6), we get

$$\begin{aligned}
(6.68) \qquad & \max_{\frac{i}{k(\varepsilon)} \le \tau \le \frac{(i+1)}{k(\varepsilon)}} \left| W^\varepsilon(\tau) - W^\varepsilon\left(\frac{i}{k(\varepsilon)}\right)\right| \\
&\le 2 \max_{\lfloor \frac{i}{\varepsilon k(\varepsilon)} \rfloor \le k \le 1 + \lfloor \frac{(i+1)}{\varepsilon k(\varepsilon)} \rfloor} \left| W^\varepsilon(\varepsilon k) - W^\varepsilon\left(\varepsilon \left\lfloor \frac{i}{\varepsilon k(\varepsilon)} \right\rfloor\right)\right| \\
&\le 2 \max_{1+\lfloor \frac{i}{\varepsilon k(\varepsilon)} \rfloor \le k \le 1 + \lfloor \frac{i+1}{\varepsilon k(\varepsilon)} \rfloor} \varepsilon^{\frac{1}{2}} \left| \sum_{j=1+\lfloor i/(\varepsilon k(\varepsilon)) \rfloor}^{k} \xi_j^\varepsilon \right|
\end{aligned}$$

for all $i = 0, 1, \ldots, k(\varepsilon) - 1$. Now fix integers M, N with $1 \le M < N$. We see from Remark 4.6 that there is a constant $C_1 \in [0, \infty)$, not depending on the sequence $\{\xi_j^\varepsilon\}$, or M, N, such that $E[|\sum_{j=i}^k \xi_j^\varepsilon|^4] \le C_1[k - i + 1]^2$ for all $M \le i \le k \le N$ and each $\varepsilon \in (0, 1]$. Hence, by Theorem 8.1(i) (with $\gamma \stackrel{\triangle}{=} 2$, $\nu \stackrel{\triangle}{=} 4$, $g(i,k) \stackrel{\triangle}{=} k - i + 1$, $M \le i \le k \le N$), there is some constant $A(4,2) \in [0, \infty)$ such that $E[\max_{M \le k \le N} |\sum_{j=M}^k \xi_j^\varepsilon|^4] \le A(4,2)[g(M,N)]^2$ for all $1 \le M < N$ and $\varepsilon \in (0,1]$. Taking $M \stackrel{\triangle}{=} 1 + \lfloor i/(\varepsilon k(\varepsilon)) \rfloor$, $N \stackrel{\triangle}{=} 1 + \lfloor (i+1)/(\varepsilon k(\varepsilon)) \rfloor$, we see from (6.68) that there are constants $C_2, C_3 \in [0, \infty)$ such that

$$(6.69) \quad E\left[\max_{\frac{i}{k(\varepsilon)} \le \tau \le \frac{(i+1)}{k(\varepsilon)}} \left| W^\varepsilon(\tau) - W^\varepsilon\left(\frac{i}{k(\varepsilon)}\right)\right|^4\right] \le \varepsilon^2 C_2 [g(M,N)]^2 \le C_3 k(\varepsilon)^{-2}$$

for all $\varepsilon \in (0,1]$. Now (6.63) and (6.64) continue to hold with $\hat{P}$, $\hat{E}$, and $\hat{W}^0$ replaced by P, E, and W^ε, respectively. Thus, we can repeat the argument which gave (6.66), but using (6.69) in place of (6.65), to see

$$(6.70) \qquad \Pi_C(\mathcal{L}(W^\varepsilon), \mathcal{L}(W^\varepsilon_{k(\varepsilon)})) \ll \varepsilon^{\frac{1}{8\gamma}}.$$

The conclusion follows from (6.62) and the upper-bounds given by (6.66), (6.67), and (6.70). ∎

7. Proofs of technical lemmas.

Proof of Lemma 2.9. It is enough to take $d = 1$. From (C4) we have $A(\theta) = \lim_{N\to\infty} \operatorname{cov}(\sum_{\nu=1}^N \tilde{H}_\nu(\theta))$, and thus (see (2.7)),

$$\begin{aligned}
|A(\theta_1) - A(\theta_2)| &\le 2 \limsup_{N\to\infty} \frac{1}{N} \sum_{\nu=1}^N \sum_{\nu=1}^N |E[\tilde{b}_\nu \tilde{R}_\mu]| \, |\theta_1 - \theta_2| \\
&\quad + \limsup_{N\to\infty} \frac{1}{N} \sum_{\nu=1}^N \sum_{\nu=1}^N |E[\tilde{R}_\nu \tilde{R}_\mu]| \, |\theta_1^2 - \theta_2^2|.
\end{aligned}$$

From Lemma 8.3(i) it easily follows that the double sums on the right-hand side are $O(N)$, as required for local Lipschitz continuity. ∎

Proof of Lemma 4.8. Since $P(|\gamma_n| \ge n^{\frac{1}{7}}) \le E\big[|\gamma_n|^8\big]/n^{\frac{8}{7}}$, we must have that $\sum_{n=1}^\infty P(|\gamma_n| \ge n^{\frac{1}{7}}) < \infty$. By Borel–Cantelli, for a.a. ω there exists an integer $L(\omega)$ such that $|\gamma_n(\omega)| \le n^{\frac{1}{7}}$ a.s. for all $n > L(\omega)$. Hence we can find $C(\omega)$ such that $|\gamma_n| \leq C(\omega)n^{\frac{1}{7}}$ for all $n \geq 1$, and the result follows from the monotonicity of $n \to n^{\frac{1}{7}}$. $\square$

Proof of Lemma 4.11. Recall the notation defined in (4.22) to (4.25). Clearly,

$$\max_{0\leq\tau\leq 1} |W_1^r(\tau) - W_2^r(\tau)| \leq \varepsilon_r^{-\frac{1}{2}} \int_{\tau_r}^1 |\zeta_{t/\varepsilon_r}^{\varepsilon_r} - \zeta_{t/\varepsilon_r}^{\varepsilon_r}[q(r)]|\, dt.$$

Hence, by Jensen's inequality,

$$\max_{0\leq\tau\leq 1} |W_1^r(\tau) - W_2^r(\tau)|^4 \leq \varepsilon_r^{-1} \int_0^{1/\varepsilon_r} I[\tau_r/\varepsilon_r, 1/\varepsilon_r](u)\, |\zeta_u^{\varepsilon_r} - \zeta_u^{\varepsilon_r}[q(r)]|^4\, du.$$

Then, by (4.23), (4.22), (4.9), the fact that $(\tau_r/\varepsilon_r) \geq 1 + q(r)$, and Fubini,

$$E\left[\max_{0\leq\tau\leq 1} |W_1^r(\tau) - W_2^r(\tau)|^4\right] \ll \varepsilon_r^{-1} \int_0^{1/\varepsilon_r} \lambda^{4q(r)}\, du \ll \exp(2(r^\sigma - r^2)).$$

By the mean value theorem for the mapping $\alpha \to r^\alpha$ (r constant), there exists $\gamma \in [\sigma, 2]$ such that $r^\sigma - r^2 = (\sigma - 2)(\ln r)r^\gamma \leq (\sigma - 2)r^\sigma$, so $\exp\left(2(r^\sigma - r^2)\right) \leq \varepsilon_r^{2(2-\sigma)} \leq \varepsilon_r$ (since $\sigma < 3/2$). $\square$

Proof of Lemma 4.12. Recall the notation defined in (4.22)–(4.25). From (3.5),

$$W^{\varepsilon_r}(\tau) = \varepsilon_r^{\frac{1}{2}} \int_0^{\tau/\varepsilon_r} \zeta_u^{\varepsilon_r}\, du, \quad 0 \leq \tau \leq 1. \tag{7.1}$$

Thus, from (7.1) and (4.24),

$$W^{\varepsilon_r}(\tau) - W_1^r(\tau) = \begin{cases} \varepsilon_r^{\frac{1}{2}} \displaystyle\int_0^{\tau/\varepsilon_r} \zeta_u^{\varepsilon_r}\, du, & \text{if } 0 \leq \tau \leq \tau_r, \\ \varepsilon_r^{\frac{1}{2}} \displaystyle\int_0^{\tau_r/\varepsilon_r} \zeta_u^{\varepsilon_r}\, du, & \text{if } \tau_r \leq \tau \leq 1. \end{cases} \tag{7.2}$$

Therefore, from (7.2) and (4.23) we have

$$\max_{0\leq\tau\leq 1} |W^{\varepsilon_r}(\tau) - W_1^r(\tau)| = \varepsilon_r^{\frac{1}{2}} \max_{1\leq k\leq 1+\lfloor\tau_r/\varepsilon_r\rfloor} \left|\sum_{j=1}^k \xi_j^{\varepsilon_r}\right|. \tag{7.3}$$

By Remark 4.6 and Theorem 8.1(i) (with $\nu = 4$, $\gamma = 2$, and $g(i,j) \triangleq j - i + 1$), we easily see there is a constant $c \in [0, \infty)$, not depending on r, such that

$$E\left[\max_{1\leq k\leq 1+\lfloor\tau_r/\varepsilon_r\rfloor} \left|\sum_{j=1}^k \xi_j^{\varepsilon_r}\right|^4\right] \leq c\left(\frac{\tau_r}{\varepsilon_r}\right)^2. \tag{7.4}$$

Combining (7.4) with (7.3) and using the fact that $(\tau_r/\varepsilon_r) \ll \varepsilon_{r-1}^{-1}$ gives

$$\begin{aligned} E\left[\max_{0\leq\tau\leq 1} |W^{\varepsilon_r}(\tau) - W_1^r(\tau)|^4\right] &\ll \varepsilon_r^2 \left(\frac{\tau_r}{\varepsilon_r}\right)^2 \\ &\ll \left(\frac{\varepsilon_r}{\varepsilon_{r-1}}\right)^2 = \exp\left(2(r-1)^\sigma - 2r^\sigma\right). \end{aligned} \tag{7.5}$$

Now $r^\sigma - (r-1)^\sigma = \sigma s^{\sigma-1}$ for some $s \in [r-1, r]$, and hence $\exp(2((r-1)^\sigma - r^\sigma)) \leq \exp(-2\sigma(r-1)^{\sigma-1})$. Observe that, since $\sigma > 1$, we have $5\ln r \ll \sigma(r-1)^{\sigma-1}$, or, equivalently, $\exp\left(-\sigma(r-1)^{\sigma-1}\right) \ll r^{-5}$ for all $r \geq 1$. Now the result follows from (7.5). ☐

Proof of Lemma 5.4. We use the notation defined in the proof of Lemma 3.6. For (5.24) we have

$$\varepsilon^{\frac{3}{2}} \max_{1\leq k\leq N_\varepsilon+1}\left|\sum_{j=0}^{k-1} \tilde{R}_{j+1}\left((I-\varepsilon\bar{R})^j - (I-N_\varepsilon^{-1}\bar{R})^j\right)\Gamma_j^{N_\varepsilon^{-1}}\right| \tag{7.6}$$

$$\leq \varepsilon^{\frac{3}{2}} N_\varepsilon \cdot \max_{0\leq j\leq N_\varepsilon} |\tilde{R}_{j+1}| \cdot \max_{0\leq j\leq N_\varepsilon} \left|(I-\varepsilon\bar{R})^j - (I-N_\varepsilon^{-1}\bar{R})^j\right| \cdot \max_{0\leq j\leq N_\varepsilon} |\Gamma_j^{N_\varepsilon^{-1}}|.$$

We bound the $\max_j |\cdot|$ factors on the right of (7.6). For the third factor, we see from $\max_{1\leq i\leq n} |I - n^{-1}\bar{R}|^{-2i} = O(1)$ (uniformly in n) and Remark 4.6, that there is a constant $c_1 \in (0, \infty)$ such that

$$E\left[\left|\sum_{k=i}^{j}(I-n^{-1}\bar{R})^{-k}\xi_k^{n^{-1}}\right|^8\right] \leq c_1(j-i+1)^4 \quad \forall\, 1\leq j\leq k\leq n. \tag{7.7}$$

From (5.12), (7.7), and Theorem 8.1(i) (with $\nu \triangleq 8$, $\gamma \triangleq 4$, $g(i,j) \triangleq j-i+1$, $1 \leq i \leq j \leq n$), we have

$$E\left[\max_{0\leq j\leq n}\left|\Gamma_j^{n^{-1}}\right|^8\right] \ll n^4, \tag{7.8}$$

and thus, from Borel–Cantelli,

$$\max_{0\leq j\leq n} |\Gamma_j^{n^{-1}}| = O(n^{\frac{6}{7}}) \text{ a.s.} \tag{7.9}$$

so that the third $\max_j |\cdot|$ factor on the right of (7.6) is $O(\varepsilon^{-\frac{6}{7}})$ a.s. It follows from Lemma 4.8 that the first $\max_j |\cdot|$ factor on the right of (7.6) is $O(\varepsilon^{-\frac{1}{7}})$ a.s. while the second $\max_j |\cdot|$ factor is easily shown by Taylor's theorem to be $O(\varepsilon)$. Thus, from (7.6), the quantity in (5.24) is a.s. $O(\varepsilon^{\frac{1}{2}}\varepsilon^{-\frac{1}{7}}\varepsilon\varepsilon^{-\frac{6}{7}}) = O(\varepsilon^{\frac{1}{2}})$. The proofs for (5.25) and part (b) are similar and are omitted. ☐

Proof of Lemma 6.9. From Remark 2.7 we know that $\{\tilde{b}_n\}$ and $\{\tilde{R}_n\}$ are geometrically L-mixing and zero-mean. Thus (recalling (2.7) and (6.41)), there are constants $\lambda \in (0,1)$ and $C_1 \in [0,\infty)$ such that $\|\tilde{H}_n(\theta)\|_2 \leq C_1[1+|\theta|]$, $\|\tilde{H}_n(\theta)[s]\|_2 \leq C_1[1+|\theta|]$, and $\|\tilde{H}_n(\theta) - \tilde{H}_n(\theta)[s]\|_2 \leq C_1[1+|\theta|]\lambda^s$ for all $\theta \in \Re^d$, for all $s, n = 1, 2, \ldots$. Put

$$D^1(n,m,s,\theta) \triangleq (\tilde{H}_n(\theta) - \tilde{H}_n(\theta)[s])(\tilde{H}_m(\theta))'.$$

By Cauchy–Schwarz and the preceding bounds,

$$|ED^1(n,m,s,\theta)| \leq \|\tilde{H}_n(\theta) - \tilde{H}_n(\theta)[s]\|_2 \|\tilde{H}_m(\theta)\|_2 \leq C_2[1+|\theta|^2]\lambda^s,$$

and the same upper-bound clearly holds for the expectation of

$$D^2(n,m,s,\theta) \triangleq \tilde{H}_n(\theta)[s](\tilde{H}_m(\theta) - \tilde{H}_m(\theta)[s])'.$$

Then

$$\begin{aligned}(7.10)\qquad &\left|\operatorname{cov}\left(\sum_{n=n_0}^{N+n_0}\tilde{H}_n(\theta)\right)-\operatorname{cov}\left(\sum_{n=n_0}^{N+n_0}\tilde{H}_n(\theta)[s]\right)\right|\\ &=\left|\sum_{n=n_0}^{N+n_0}\sum_{m=n_0}^{N+n_0}E[D^1(n,m,s,\theta)+D^2(n,m,s,\theta)]\right|\\ &\leq 2C_2[1+|\theta|^2](N+1)^2\lambda^s.\end{aligned}$$

The result now follows from condition (C4) and (7.10). □

Proof of Lemma 6.10. Fix some $k=1,2,\ldots$, and let $|\cdot|_\infty$ denote the maximum norm on $\Re^{kd}$ (i.e., $|x|_\infty \triangleq \max_{1\leq i\leq kd}|x^i|$). Let $\Pi_\infty^{kd}(P_1,P_2)$ denote Prohorov distance between probability measures P_1 and P_2 on $\Re^{kd}$ with norm $|\cdot|_\infty$, and let $C_k[0,1]$ be the subset of $C[0,1]$ comprising all continuous functions $f:[0,1]\to\Re^d$, with $f(0)=0$, which are piecewise linear with break-points at i/k, $i=1,\ldots,k-1$ (see right-hand side of (6.61)). Then the metric spaces $(\Re^{kd},|\cdot|_\infty)$ and $(C_k[0,1],\|\cdot\|_C)$ are homeomorphic. Since the paths of $\{W_k^\varepsilon(\tau),\ \tau\in[0,1]\}$ and $\{\hat{W}_k^0(\tau),\ \tau\in[0,1]\}$ are in $C_k[0,1]$, it easily follows from (6.1) and (6.61) that $\Pi_C(\mathcal{L}(W_k^\varepsilon),\mathcal{L}(\hat{W}_k^0))=\Pi_\infty^{kd}(\mathcal{L}(\Xi_k^\varepsilon),\mathcal{L}(\hat{\Xi}_k^0))$. Now, for $\eta>0$ and closed $A\subset\Re^{kd}$, put $A^\eta\triangleq\{x\in\Re^{kd}:|x-a|<\eta \text{ for some } a\in A\}$ and $A_\infty^\eta\triangleq\{x\in\Re^{kd}:|x-a|_\infty<\eta \text{ for some } a\in A\}$. Since $|x|_\infty\leq|x|$, we have $A^\eta\subset A_\infty^\eta$, and from this, together with (8.1), we get $\Pi_\infty^{kd}(\mathcal{L}(\Xi_k^\varepsilon),\mathcal{L}(\hat{\Xi}_k^0))\leq\Pi_2^{kd}(\mathcal{L}(\Xi_k^\varepsilon),\mathcal{L}(\hat{\Xi}_k^0))$, as required. □

8. Useful results. In this section we collect for easy reference some simple adaptations of results from probability which are used in the previous sections. The following maximal inequalities, due to Móricz [23, Theorem 1] and Longnecker and Serfling [22], are needed for lines (5.21), (6.69), (7.4), and (7.8).

THEOREM 8.1. *Suppose that M and N are integers, $1\leq M<N<\infty$, and $\mathcal{Y}$ is a normed vector space with norm $\|\cdot\|$.*

(i) *Let $\{z_k,\ k=M,M+1,\ldots,N\}$ be arbitrary $\mathcal{Y}$-valued random variables. Suppose there are constants $c,\nu\in(0,\infty)$, $\gamma\in(1,\infty)$, and an $\Re$-valued mapping $g(i,j)$ defined for $M\leq i\leq j\leq N$, such that $E[\|\sum_{k=i}^j z_k\|^\nu]\leq c[g(i,j)]^\gamma$ for all $M\leq i\leq j\leq N$, and $g(i,j)+g(j+1,k)\leq g(i,k)$ for all $M\leq i\leq j<k\leq N$. Then there is a constant $A(\nu,\gamma)\in[0,\infty)$ such that $E[\max_{M\leq n\leq N}\|\sum_{k=M}^n z_k\|^\nu]\leq cA(\nu,\gamma)[g(M,N)]^\gamma$. The constant $A(\nu,\gamma)$ depends on ν and γ only.*

(ii) *Let $\{Q_k,\ k=M,M+1,\ldots,N\}$ be arbitrary $\mathcal{Y}$-valued random variables. Suppose there are constants $c,\nu\in(0,\infty)$, $\gamma\in(1,\infty)$, and an $\Re$-valued mapping $h(i,j)$ defined for $M\leq i\leq j\leq N$, such that $E[\|Q_j-Q_i\|^\nu]\leq c[h(i,j)]^\gamma$ for all $M\leq i\leq j\leq N$, and $h(i,j)+h(j,k)\leq h(i,k)$ for all $M\leq i\leq j\leq k\leq N$. Then, for the constant $A(\nu,\gamma)$ of (i), we have $E[\max_{M\leq n\leq N}\|Q_n-Q_M\|^\nu]\leq cA(\nu,\gamma)[h(M,N)]^\gamma$.*

Remark 8.2. The constant $A(\nu,\gamma)$ has nothing to do with M,N, the random variables $\{z_k\}$, $\{Q_k\}$, or the functions $g(\cdot,\cdot)$, $h(\cdot,\cdot)$ in (i) and (ii). Indeed, (2.1)–(2.3) of [22] give $A(\nu,\gamma)$ explicitly as a function of $\nu>0$ and $\gamma>1$ only. Notice that (ii) follows upon applying (i) of Theorem 8.1 to the sequence $\{z_k,\ k=M,M+1,\ldots,N-1\}$ defined by $z_k\triangleq Q_{k+1}-Q_k$ for all $k=M,M+1,\ldots,N-1$, with $g(i,j)\triangleq h(i,j+1)$ for all $M\leq i\leq j\leq N-1$.

By trivially adapting the arguments for Lemma A.1.2(a)(b) of [11], we get the following lemma.

LEMMA 8.3. *Suppose that* $\{z_n^1,\ n=1,2,\ldots\}$ *and* $\{z_n^2,\ n=1,2,\ldots\}$ *are* $\Re$*-valued zero-mean geometrically* L*-mixing processes with respect to the system* $(\mathcal{F}_n,\mathcal{F}_n^+)$. *Then there are constants* $C_1,C_2\in[0,\infty)$ *and* $\lambda\in(0,1)$ *such that* (i) $|E\{z_n^1 z_m^2\}|\leq C_1\lambda^{|n-m|}$ *for all* $m,n=1,2,\ldots,$ *and* (ii) $|E\{z_n^1[s]z_m^2[s]\}|\leq C_2\lambda^{|n-m|}$ *for all* $m,n,s=1,2,\ldots$ *(recall* (2.3)*).*

Remark 8.4. Since $\theta^0((n-1)\varepsilon\wedge T)$ is uniformly bounded (in $\varepsilon\in(0,1]$, $n=1,2,\ldots$) and $\{\tilde{b}_n\}$ and $\{\tilde{R}_n\}$ are zero-mean geometrically L-mixing (by Remark 2.7), we see from (4.8) and Lemma 8.3(ii) that there are constants $C\in[0,\infty)$ and $\lambda\in(0,1)$ such that $|E[(\xi_\nu^\varepsilon[s])(\xi_\mu^\varepsilon[s])']|\leq C\,\lambda^{|\mu-\nu|}$ for all $\varepsilon\in(0,1]$, for all $s,\mu,\nu=1,2,\ldots$.

The following result, which is a special case of Theorem 1.1 of Gerencsér [9], is repeatedly used.

THEOREM 8.5. *Suppose that* $\{u_j\}$ *is a zero-mean* $\Re^{l\times r}$*-valued geometrically* L*-mixing process, and* $\{A_j\}$ *is a nonrandom sequence of* $d\times l$ *matrices. Then, for each* $p\in[2,\infty)$, $\|\sum_{j=1}^N A_j u_j\|_p \ll (\sum_{j=1}^N |A_j|^2)^{\frac{1}{2}}$. *The constant implied by* $\ll$ *depends only upon* p, $\sup_{j\geq 1}\|u_j\|_p$, *and the rate* λ *of the geometrically* L*-mixing process* $\{u_j\}$, *in particular, does NOT depend on* N *or the sequence* $\{A_j\}$.

In the remainder of this section we summarize some relevant facts about the *Prohorov metric* for probability measures on a metric space. Suppose that (S,ρ) is a metric space, and let P_1 and P_2 be two probability measures on $(S,\mathcal{B}(S))$. Define the number

$$\Pi(P_1,P_2)\stackrel{\triangle}{=}\inf\{\eta\in(0,\infty):\ P_1(A)\leq P_2(A^\eta)+\eta\ \text{ for all closed } A\subset S\}, \tag{8.1}$$

where $A^\eta\stackrel{\triangle}{=}\{s\in S:\ \rho(s,a)\leq\eta\ \text{ for some } a\in A\}$ for $A\subset S$. All we need to know is that the mapping $\Pi(\cdot,\cdot)$ defined by (8.1) is indeed a metric in the set of all probability measures on $(S,\mathcal{B}(S))$, called the *Prohorov metric*, and when (S,ρ) is a *separable* metric space then $\Pi(\cdot,\cdot)$ is a metric for the topology of weak convergence of probability measures on $(S,\mathcal{B}(S))$. See [8, Theorem 3.1, p. 108]. Thus, one can use the Prohorov metric to quantify *rates* of weak convergence of probability measures. In fact, this is the significance of condition (4.3) in Theorem 4.2. We also make repeated use of the following simple result which is an immediate consequence of (8.1) and the Chebyshev inequality.

LEMMA 8.6. *Suppose that* X *and* Y *are random variables defined on* $(\Omega,\mathcal{F},P)$ *with values in a metric space* (S,ρ). (i) *If, for some* $\beta\in(0,\infty)$, *we have* $P\{\rho(X,Y)\geq\beta\}\leq\beta$, *then* $\Pi(\mathcal{L}(X),\mathcal{L}(Y))\leq\beta$. (ii) *If, for some* $\beta\in(0,\infty)$ *and* $c\in[1,\infty)$, *we have* $\|\rho(X,Y)\|_c\leq\beta$, *then* $\Pi(\mathcal{L}(X),\mathcal{L}(Y))\leq\beta^{\frac{c}{c+1}}$.

The next result (used for (6.56) and (6.59)) is a special case of Dehling [7, Theorem 7, p. 400], and upper-bounds the Prohorov distance between Gaussian distributions on Euclidean space.

THEOREM 8.7. *There exists a constant* $\beta\in[0,\infty)$ *such that*

$$\Pi_2^{kd}\left(\mathcal{N}_{kd}(0,T),\mathcal{N}_{kd}(0,S)\right)\leq\beta k^{\frac{3}{4}}d^{\frac{5}{8}}\left(\max_{1\leq m,n\leq k}|T_{m,n}-S_{m,n}|\right)^{\frac{1}{4}}$$

for all $k,d=1,2,\ldots,$ *and all* $kd\times kd$ *symmetric positive semidefinite matrices* T *and* S *such that* $kd^{\frac{1}{2}}\max_{1\leq m,n\leq k}|T_{m,n}-S_{m,n}|<1$.

Acknowledgment. Many thanks indeed to the referee for an exceptionally careful and detailed review which led to numerous improvements in this work.

REFERENCES

[1] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Springer-Verlag, New York, 1990.
[2] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1968.
[3] J. A. Bucklew, T. G. Kurtz, and W. A. Sethares, *Weak convergence and local stability properties of fixed step size recursive algorithms*, IEEE Trans. Inform. Theory, 30 (1993), pp. 966–978.
[4] P. E. Caines, *Linear Stochastic Systems*, Wiley, New York, 1988.
[5] J. Chover, *On Strassen's version of the log log law*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 8 (1967), pp. 83–90.
[6] M. H. A. Davis and R. B. Vinter, *Stochastic Modelling and Control*, Chapman and Hall, London, UK, 1985.
[7] H. Dehling, *Limit theorems for sums of weakly dependent Banach space valued random variables*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 63 (1983), pp. 393–432.
[8] S. N. Ethier and T. G. Kurtz, *Markov Processes Characterization and Convergence*, Wiley, New York, 1986.
[9] L. Gerencsér, *On a class of mixing processes*, Stochastics Stochastics Rep., 26 (1989), pp. 165–191.
[10] L. Gerencsér, *Multiple integrals with respect to L-mixing processes*, Statist. Probab. Lett., 17 (1993), pp. 73–85.
[11] A. J. Heunis, *Asymptotic properties of prediction error estimators in approximate system identification*, Stochastics, 24 (1988), pp. 1–43.
[12] A. J. Heunis, *Rates of convergence for an adaptive filtering algorithm driven by stationary dependent data*, SIAM J. Control Optim., 32 (1994), pp. 116–139.
[13] R. Z. Khas'minskii, *On stochastic processes defined by differential equations with a small parameter*, Theory Probab. Appl., 11 (1966), pp. 211–228.
[14] M. A. Kouritzin and A. J. Heunis, *Rates of convergence in a central limit theorem for stochastic processes defined by differential equations with a small parameter*, J. Multivariate Anal., 43 (1992), pp. 58–109.
[15] M. A. Kouritzin and A. J. Heunis, *A law of the iterated logarithm for stochastic processes defined by differential equations with a small parameter*, Ann. Probab., 22 (1994), pp. 659–679.
[16] J. Kuelbs, *A strong convergence theorem for Banach space valued random variables*, Ann. Probab., 4 (1976), pp. 744–771.
[17] P. R. Kumar and P.P. Varaiya, *Stochastic Systems: Estimation, Identification, and Adaptive Control*, Prentice-Hall, Englewood Cliffs, NJ, 1986.
[18] H. J. Kushner and G. G. Yin, *Stochastic Approximation Algorithms and Applications*, Springer-Verlag, New York, 1997.
[19] H. J. Kushner and A. Shwartz, *Weak convergence and asymptotic properties of adaptive filters with constant gains*, IEEE Trans. Inform. Theory, 30 (1984), pp. 177–182.
[20] L. Ljung, *Convergence analysis of parametric identification methods*, IEEE Trans. Automat. Control, 23 (1978), pp. 770–783.
[21] L. Ljung and P. E. Caines, *Asymptotic normality of prediction error estimators for approximate system models*, Stochastics, 3 (1979), pp. 29–46.
[22] M. Longnecker and R. J. Serfling, *General moment and probability inequalities for the maximum partial sum*, Studia Sci. Math. Hungar., 30 (1977), pp. 129–133.
[23] F. A. Móricz, *Moment inequalities and the strong law of large numbers*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 35 (1976), pp. 299–314.
[24] H. Pezezhki-Esfahani and A.J. Heunis, *Strong diffusion approximations for recursive stochastic algorithms*, IEEE Trans. Inform. Theory, 43 (1997), pp. 512–523.
[25] V. Strassen, *An invariance principle for the law of the iterated logarithm*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 3 (1964), pp. 211–226.
[26] D. Williams, *Probability with Martingales*, Cambridge University Press, Cambridge, UK, 1991.
[27] V. V. Yurinskii, *On the error of the Gaussian approximation for convolutions*, Theory Probab. Appl., 22 (1977), pp. 236–247.