Komlós-Major-Tusnády approximation under dependence

István Berkes[1], Weidong Liu[2] and Wei Biao Wu[3]

Abstract

The celebrated results of Komlós, Major and Tusnády (1975, 1976) give optimal Wiener approximation for the partial sums of i.i.d. random variables and provide a powerful tool in probability and statistics. In this paper we extend KMT approximation for a large class of dependent stationary processes, solving a long standing open problem in probability theory. Under the framework of stationary causal processes and functional dependence measures of Wu (2005), we show that, under natural moment conditions, the partial sum processes can be approximated by Wiener process with an optimal rate. Our dependence conditions are mild and easily verifiable. The results are applied to ergodic sums, as well as to nonlinear time series and Volterra processes, an important class of nonlinear processes.

Keywords: Stationary processes, strong invariance principle, KMT approximation, weak dependence, nonlinear time series, ergodic sums

MSC 2000: 60F17, 60G10, 60G17

[1])Institute of Statistics, Graz University of Technology, Münzgrabenstrasse 11, 8010 Graz, Austria. Email: `berkes@tugraz.at` Research supported by Austrian Science Fund (FWF), Projekt P24302-N18.

[2])Department of Mathematics, Shanghai Jiao Tong University, 800 Dongchuan Road Minhang, Shanghai, China. Email: `liuweidong99@gmail.com`

[3])Department of Statistics, University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, USA. Email: `wbwu@galton.uchicago.edu`

1 Introduction

Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with $\mathsf{E}X_1 = 0$, $\mathsf{E}X_1^2 = 1$. In their seminal papers, Komlós, Major and Tusnády (1975, 1976) proved that under $\mathsf{E}|X_1|^p < \infty$, $p > 2$ there exists, after suitably enlarging the probability space, a Wiener process $\{I\!\!B(t),\, t \geq 0\}$ such that, setting $S_n = \sum_{k=1}^n X_k$, we have

$$S_n = I\!\!B(n) + o(n^{1/p}) \qquad \text{a.s.} \tag{1.1}$$

Assuming $\mathsf{E}e^{t|X_1|} < \infty$ for some $t > 0$, they obtained the approximation

$$S_n = I\!\!B(n) + O(\log n) \qquad \text{a.s.} \tag{1.2}$$

The remainder terms in (1.1) and (1.2) are optimal. These results close a long development in probability theory starting with the classical paper of Erdős and Kac (1946) introducing the method of *invariance principle.* The ideas of Erdős and Kac were developed further by Doob (1949), Donsker (1952), Prohorov (1956) and others and led to the theory of weak convergence of probability measures on metric spaces, see e.g. Billingsley (1968). In another direction, Strassen (1964) used the Skorohod representation theorem to get an almost sure approximation of partial sums of i.i.d. random variables by Wiener process. Csörgő and Révész (1974) showed that using the quantile transform instead of Skorohod embedding yields better approximation rates under higher moments and developing this idea further, Komlós et al. (1975, 1976) reached the final result in the i.i.d. case. Their results were extended to the independent, non-identically distributed case and for random variables taking values in $\mathbb{R}^d$, $d \geq 2$ by Sakhanenko, Einmahl and Zaitsev; see Götze and Zaitsev (2009) for history and references.

Due to the powerful consequences of KMT approximation (see e.g. Csörgő and Hall (1984) or the books of Csörgő and Révész (1981) and Shorack and Wellner (1986) for the scope of its applications), extending these results for dependent random variables would have a great importance, but until recently, little progress has been made in this direction. The dyadic construction of Komlós, Major and Tusnády is highly technical and utilizes conditional large deviation techniques, which makes it very difficult to extend to dependent processes. Recently a new proof of the KMT result for the simple random walk via Stein's method was given by Chatterjee (2012). The main motivation of this paper was, as stated by the author, to get "a more conceptual understanding of the problem that may allow one to go beyond sums of independent random variables". Using martingale approximation and Skorohod embedding, Wu

(2007) proved the approximation

$$S_n = \sigma \mathbb{B}(n) + o(n^{1/p}(\log n)^\gamma) \qquad \text{a.s.} \tag{1.3}$$

with some $\sigma \geq 0$, $\gamma > 0$ for a class of stationary sequences (X_k) satisfying $\mathsf{E}X_1 = 0$, $\mathsf{E}|X_1|^p < \infty$ for some $2 < p \leq 4$. Liu and Lin (2009) removed the logarithmic term from (1.3), reaching the KMT bound $o(n^{1/p})$. Recently Merlevède and Rio (2012) obtained nearly optimal strong approximation results for $p \leq 3$. Note, however, that all existing results in the dependent case concern the case $2 \leq p \leq 4$ and the applied tools (e.g. Skorohod representation) limit the accuracy of the approximation to $o(n^{1/4})$, regardless the moment assumptions on X_1.

The purpose of the present paper is to develop a new approximation technique enabling us prove the KMT approximation (1.1) for all $p > 2$ and for a large class of dependent sequences. Specifically, we will deal with stationary sequences allowing the representation

$$X_k = G(\ldots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \ldots) \qquad k \in \mathbb{Z}, \tag{1.4}$$

where ε_i, $i \in \mathbb{Z}$, are i.i.d. random variables and $G : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}$ is a measurable function. Sequences of this type have been studied intensively in weak dependence theory (see e.g. Billingsley (1968) or Ibragimov and Linnik (1971)) and many important time series models also have a representation (1.4). Processes of the type (1.4) also play an important role in ergodic theory, as sequences generated by Bernoulli shift transformations. The Bernoulli shift is a very important class of dynamical systems; see Ornstein (1974) and Shields (1973) for the deep Kolmogorov-Sinai-Ornstein isomorphism theory. There is a substantial amount of research showing that various dynamical systems are isomorphic to Bernoulli shifts. As a step further, Weiss (1974) asked

> "having shown that some physical system is Bernoullian, what does that allow one to say about the system itself? To answer such questions one must dig deeper and gain a better understanding of a Bernoulli system".

Naturally, without additional assumptions one cannot hope to prove KMT type results (or even the CLT) for Bernoulli systems; the representation (1.4) allows stationary processes that can exhibit a markedly non-i.i.d. behavior. For limit theorems under dynamic assumptions see Hofbauer and Keller (1982), Denker and Philipp (1984), Denker (1989), Volný (1999), Merlevède and Rio (2012). The classical approach to deal with systems (1.4) is to assume that G is approximable with finite dimensional functions in a certain technical sense, see Billingsley (1968) or Ibragimov and Linnik (1971). However, this approach leads to a substantial loss of accuracy and does not yield optimal results. In this paper we introduce a new, triadic decomposition scheme enabling one to deduce directly, under the dependence measure (1.5) below, the asymptotic properties of X_n in (1.4) from those of the ε_n. In particular, this allows us to carry over KMT approximation from the partial sums of the ε_n to those of X_n.

To state our weak dependence assumptions on the process in (1.4), assume $X_i \in \mathcal{L}^p$, $p > 2$, namely $\|X_i\|_p := [\mathsf{E}(|X_i|^p)]^{1/p} < \infty$. Let $(\varepsilon'_j)_{j\in\mathbb{Z}}$ be an iid copy of $(\varepsilon_j)_{j\in\mathbb{Z}}$; let $\varepsilon_{i,\{j\}} = \varepsilon_i$ if $j \neq i$ and $\varepsilon_{i,\{j\}} = \varepsilon'_i$ if $j = i$. Define the shift process $\mathcal{F}_i = (\varepsilon_{i-l},\, l \in \mathbb{Z})$, $\mathcal{F}_{i,\{j\}} = (\varepsilon_{i-l,\{j\}},\, l \in \mathbb{Z})$ and

$$\delta_{i,p} = \|X_i - X_{i,\{0\}}\|_p, \text{ where } X_{i,\{0\}} = G(\mathcal{F}_{i,\{0\}}). \tag{1.5}$$

The above quantity can be interpreted as the dependence of X_i on ε_0 and $X_{i,\{0\}}$ is a coupled version of X_i with ε_0 in the latter replaced by ε'_0. If $G(\mathcal{F}_i)$ does not functionally depend on ε_0, then $\delta_{i,p} = 0$. Throughout the paper, for a random variable $W = H(\mathcal{F}_i)$, we use the notation $W_{\{j\}} = H(\mathcal{F}_{i,\{j\}})$ for the j-coupled version of W, namely it is obtained by replacing ε_j in $\mathcal{F}_i$ by the i.i.d. copy ε'_j.

The functional dependence measure (1.5) is easy to work with and it is directly related to the underlying data-generating mechanism. In our main result Theorem 2.1, we express our dependence condition in terms of

$$\Theta_{i,p} = \sum_{j=|i|}^{\infty} \delta_{j,p}, \tag{1.6}$$

which can be interpreted as the cumulative dependence of $(X_j)_{j\geq|i|}$ on ε_0. Throughout the paper we assume that the short-range dependence condition

$$\Theta_{0,p} < \infty \tag{1.7}$$

holds. If (1.7) fails, then the process (X_i) can be long-range dependent and the partial sum processes behave no longer like Brownian motions. Our main result is introduced in Section 2, where we also include with some discussions on the conditions. The proof is given in Section 3, and some useful lemmas are provided in Section 4.

2 Main Results

We introduce some notation. For $u \in \mathbb{R}$, let $\lceil u \rceil = \min\{i \in \mathbb{Z} : i \geq u\}$ and $\lfloor u \rfloor = \max\{i \in \mathbb{Z} : i \leq u\}$. Write the $\mathcal{L}^2$ norm $\|\cdot\| = \|\cdot\|_2$. Denote by "$\Rightarrow$" the weak convergence. Before stating our main result, we first introduce a central limit theorem for S_n. Assume that X_i has mean zero, $\mathsf{E}(X_i^2) < \infty$, with covariance function $\gamma_i = \mathsf{E}(X_0 X_i)$, $i \in \mathbb{Z}$. Further assume that

$$\sum_{i=-\infty}^{\infty} \|\mathsf{E}(X_i|\mathcal{G}_0) - \mathsf{E}(X_i|\mathcal{G}_{-1})\| < \infty, \tag{2.8}$$

where $\mathcal{G}_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i)$. Then we have

$$\frac{S_n}{\sqrt{n}} \Rightarrow N(0, \sigma^2), \text{ where } \sigma^2 = \sum_{i\in\mathbb{Z}} \gamma_i. \tag{2.9}$$

Results of the above type have been known for several decades; see Hannan (1979), Woodroofe (1992), Volný (1993) and Dedecker and Merlevéde (2003) among others. Wu (2005) pointed out the inequality $\|\mathsf{E}(X_i|\mathcal{G}_0) - \mathsf{E}(X_i|\mathcal{G}_{-1})\| \le \delta_{i,2}$. Hence (2.8) follows from $\Theta_{0,2} < \infty$. With stronger moment and dependence conditions, the central limit theorem (2.9) can be improved to strong invariance principles.

There is a huge literature for central limit theorems and invariance principles for stationary processes; see for example Ibragimov and Linnik (1971), the monograph on limit theorems under dependence edited by Eberlein and Taqqu (1986), Bradley (2007), Dedecker et al (2007), Peligrad and Utev (2005) among others. To establish strong invariance principles, here we shall use the framework of stationary process (1.4) and its associated functional dependence measures (1.5). Many popular nonlinear time series processes assume this form; see Wiener (1958), Tong (1990), Priestley (1988), Shao and Wu (2007) and Wu (2011) among others.

Theorem 2.1 *Assume that $X_i \in \mathcal{L}^p$ with mean 0, $p > 2$, and there exists $\alpha > p$ such that*

$$\Xi_{\alpha,p} := \sum_{j=-\infty}^{\infty} |j|^{1/2-1/\alpha} \delta_{j,p}^{p/\alpha} < \infty. \tag{2.10}$$

Further assume that there exists a positive integer sequence $(m_k)_{k=1}^{\infty}$ such that

$$M_{\alpha,p} := \sum_{k=1}^{\infty} 3^{k-k\alpha/p} m_k^{\alpha/2-1} < \infty, \tag{2.11}$$

$$\sum_{k=1}^{\infty} \frac{3^{kp/2}\Theta_{m_k,p}^p}{3^k} < \infty \tag{2.12}$$

and

$$\Theta_{m_k,p} + \min_{l\ge 0}(\Theta_{l,p} + l3^{k(2/p-1)}) = o\left(\frac{3^{k(1/p-1/2)}}{(\log k)^{1/2}}\right). \tag{2.13}$$

Then there exists a probability space $(\Omega_c, \mathcal{A}_c, P_c)$ *on which we can define random variables* X_i^c *with the partial sum process* $S_n^c = \sum_{i=1}^n X_i^c$, *and a standard Brownian motion* $\mathbb{B}_c(\cdot)$, *such that* $(X_i^c)_{i\geq 1} \stackrel{\mathcal{D}}{=} (X_i)_{i\geq 1}$ *and*

$$S_n^c - \sigma \mathbb{B}_c(n) = o_{\text{a.s.}}(n^{1/p}) \text{ in } (\Omega_c, \mathcal{A}_c, P_c). \tag{2.14}$$

Gaussian approximation results of type (2.14) have many applications in statistics. For example, Wu and Zhao (2007) dealt with simultaneous inference of trends in time series. Eubank and Speckman (1993) considered a similar problem for independent observations. Csörgő and Révész (1981) provided applications of Gaussian approximations for independent random variables. For simultaneous inference under dependence, Colin Wu, Chiang and Hoover (1998) pointed out the fundamental difficulty of not having a suitable form of Gaussian approximation with the presence of dependence. We expect (2.14) to have substantial applications in this direction.

A crucial issue in applying Theorem 2.1 is to find the sequence m_k and to verify conditions (2.10), (2.11), (2.12) and (2.13). If $\Theta_{m,p}$ decays to zero at the rate $O(m^{-\tau}(\log m)^{-A})$, where $\tau > 0$, then we have the following corollary. An explicit form of m_k can also be given. Let

$$\tau_p = \frac{p^2 - 4 + (p-2)\sqrt{p^2 + 20p + 4}}{8p}. \tag{2.15}$$

Corollary 2.1 *Assume that either one of the following holds:*

(i) $p > 4$ *and* $\Theta_{m,p} = O(m^{-\tau_p}(\log m)^{-A})$, *where* $A > \frac{2}{3}(1/p + 1 + \tau_p)$;

(ii) $p = 4$ *and* $\Theta_{m,p} = O(m^{-1}(\log m)^{-A})$ *with* $A > 3/2$;

(iii) $2 < p < 4$ *and* $\Theta_{m,p} = O(m^{-1}(\log m)^{-1/p})$.

Then there exists $\alpha > p$ *and an integer sequence* m_k *such that (2.10), (2.11), (2.12) and (2.13) are all satisfied. Hence the strong invariance principle (2.14) holds.*

Proof. If $\Theta_{m,p} = O(m^{-\tau}(\log m)^{-A})$, then

$$\begin{aligned}
\Xi_{\alpha,p} &\leq \sum_{l=1}^{\infty} 2^{l(1/2-1/\alpha)} \sum_{j=2^{l-1}}^{2^l-1} (\delta_{j,p}^{p/\alpha} + \delta_{-j,p}^{p/\alpha}) \\
&\leq \sum_{l=1}^{\infty} 2^{l(1/2-1/\alpha)} 2^{(l-1)(1-p/\alpha)} \left(\sum_{j=2^{l-1}}^{2^l-1} (\delta_{j,p} + \delta_{-j,p}) \right)^{p/\alpha} \\
&\leq \sum_{l=1}^{\infty} 2^{l(3/2-1/\alpha-p/\alpha)} \Theta_{2^{l-1},p}^{p/\alpha} \\
&= \sum_{l=1}^{\infty} 2^{l(3/2-1/\alpha-p/\alpha)} O[(2^{-l\tau} l^{-A})^{p/\alpha}],
\end{aligned}$$

which is finite if $3/2 < (1+p+p\tau)/\alpha$ or $3/2 = (1+p+p\tau)/\alpha$ and $Ap/\alpha > 1$.

(i) Write $\tau = \tau_p$. The quantity τ_p satisfies the following equation

$$\frac{\tau - (1/2 - 1/p)}{\tau/p - 1/4 + 1/(2p)} = \frac{2}{3}(1+p+p\tau). \tag{2.16}$$

Let $\alpha = \frac{2}{3}(1+p+p\tau_p)$. Then (2.10) requires that $Ap/\alpha > 1$, or $A > \alpha/p$. Let

$$m_k = \lfloor 3^{k(\alpha/p-1)/(\alpha/2-1)} k^{-1/(\alpha/2-1)} (\log k)^{-1/(p/2-1)} \rfloor, \tag{2.17}$$

which satisfies (2.11). Then $\Theta_{m_k,p} = O(m_k^{-\tau} k^{-A})$. If $A > \tau/(\alpha/2-1)$, then (2.13) holds. If $A > \tau/(\alpha/2-1)+1/p$, then (2.12) holds. Combining these three inequalities on A, we have (i), since $\alpha/p > \tau/(\alpha/2-1)+1/p$.

(ii) In this case we can choose $\alpha = 6$ and $m_k = \lfloor 3^{k/4}/k \rfloor$.

(iii) Since $2 < p < 4$, we can choose α such that $(2+p)/(3-p/2) < \alpha < (2+4p)/3$ and $m_k = \lfloor 3^{k(1/2-1/p)} \log k \rfloor$. ♢

Corollary 2.1 indicates that to establish Gaussian approximation for the stationary process (1.4), one only needs to compute the functional dependence measure $\delta_{i,p}$ in (1.5). In the following examples we shall deal with some special situations. Example 2.1 considers ergodic sums, Example 2.2 concerns some widely used nonlinear time series, and Example 2.3 deals with Volterra processes which play an important role in the study of nonlinear systems.

Example 2.1 Consider the measure-preserving transformation $Tx = 2x \bmod 1$ on $([0,1], \mathcal{B}, \mathsf{P})$, where P is the Lebesgue measure on $[0,1]$. Let $U_0 \sim uniform(0,1)$ have the dyadic expansion $U_0 = \sum_{j=0}^{\infty} \varepsilon_j / 2^{1+j}$, where ε_j are i.i.d. Bernoulli random variables with $\mathsf{P}(\varepsilon_j = 0) = \mathsf{P}(\varepsilon_j = 1) = 1/2$. Then $U_i = T^i U_0 = \sum_{j=i}^{\infty} \varepsilon_j / 2^{1+j-i}$, $i \geq 0$; see Denker and Keller (1986) for a more detailed discussion. We now compute the functional dependence measure for $X_i = g(U_i)$. Assume that $\int_0^1 g(u)du = 0$ and $\int_0^1 |g(u)|^p du < \infty$, $p > 2$. Then

$$\begin{aligned} \delta_p(i)^p &= \mathsf{E}|g(U_0) - g(U_{0,i})|^p \\ &= \frac{1}{2} \sum_{j=1}^{2^i} \int_0^1 |g(\frac{j}{2^i} + \frac{u}{2^{i+1}}) - g(\frac{j-1}{2^i} + \frac{u}{2^{i+1}})|^p du. \end{aligned} \tag{2.18}$$

If $X_i = g(U_i) = K(\sum_{j=i}^{\infty} a_{j-i} \varepsilon_j)$, where K is a Lipschitz continuous function and $\sum_{j=0}^{\infty} |a_j| < \infty$, then $\delta_p(i) = O(|a_i|)$. If g has the Haar wavelet expansion

$$g(u) = \sum_{i=0}^{\infty} \sum_{j=1}^{2^i} c_{i,j} \phi_{i,j}(u), \tag{2.19}$$

where $\phi_{i,j}(u) = 2^{i/2}\phi(2^i u - j)$ and $\phi(u) = \mathbf{1}_{0\le u<1/2} - \mathbf{1}_{1/2\le u<1}$, then

$$\delta_p(i)^p = O(2^{i(p/2-1)}) \sum_{j=1}^{2^i} |c_{i,j}|^p. \tag{2.20}$$

Example 2.2 (Nonlinear Time Series) Consider the iterated random function

$$X_i = G(X_{i-1}, \varepsilon_i), \tag{2.21}$$

where ε_i are i.i.d. and G is a measurable function (Diaconis and Freedman, 1999). Many nonlinear time series including ARCH, Threshold Autoregressive, Random coefficient Autoregressive, Bilinear Autoregressive processes, are of form (2.21). If there exists $p > 2$ and x_0 such that $G(x_0, \varepsilon_0) \in \mathcal{L}^p$ and

$$\ell_p = \sup_{x\neq x'} \frac{\|G(x,\varepsilon_0) - G(x',\varepsilon_0)\|_p}{|x-x'|} < 1, \tag{2.22}$$

then $\delta_{m,p} = O(\ell_p^m)$ and also $\Theta_{m,p} = O(\ell_p^m)$ (Wu and Shao, 2004). Hence conditions in Corollary 2.1 are trivially satisfied and thus (2.14) holds.

Example 2.3 In the study of nonlinear systems Volterra processes are of fundamental importance; see Schetzen (1980), Rugh (1981), Casti (1985), Priestley (1988) and Bendat (1990) among others. We consider the discrete-time process

$$X_n = \sum_{k=1}^{\infty} \sum_{0\le j_1<\ldots<j_k} g_k(j_1,\ldots,j_k)\varepsilon_{n-j_1}\ldots\varepsilon_{n-j_k}, \tag{2.23}$$

where ε_i are i.i.d. with mean 0, $\varepsilon_i \in \mathcal{L}^p$, $p > 2$, and g_k are called the kth order Volterra kernel. Let

$$Q_{n,k} = \sum_{n\in\{j_1,\ldots,j_k\},\, 0\le j_1<\ldots<j_k} g_k^2(j_1,\ldots,j_k). \tag{2.24}$$

Assume for simplicity that p is an even integer. Elementary calculations show that there exists a constant c_p, only depending on p, such that

$$\delta_{n,p}^2 \le c_p \sum_{k=1}^{\infty} \|\varepsilon_0\|_p^{2k} Q_{n,k}. \tag{2.25}$$

Assume that for some $\tau > 0$ and A,

$$\sum_{k=1}^{\infty} \|\varepsilon_0\|_p^{2k} \sum_{j_k\ge m,\, 0\le j_1<\ldots<j_k} g_k^2(j_1,\ldots,j_k) = O(m^{-1-2\tau}(\log m)^{-2A}) \tag{2.26}$$

as $m \to \infty$. Then

$$\sum_{n=m}^{\infty} \delta_{n,p}^2 \le c_p \sum_{k=1}^{\infty} \|\varepsilon_0\|_p^{2k} \sum_{n=m}^{\infty} Q_{n,k} = O(m^{-1-2\tau}(\log m)^{-2A}), \tag{2.27}$$

which implies $\Theta_{m,p} = O(m^{-\tau}(\log m)^{-A})$ and hence Corollary 2.1 is applicable.

3 Proof of Theorem 2.1

The proof of Theorem 2.1 is quite intricate. To simplify notation, we assume that (X_i) is a function of one-sided Bernoulli shift:

$$X_i = G(\mathcal{F}_i), \text{ where } \mathcal{F}_i = (\cdots, \varepsilon_{i-1}, \varepsilon_i), \tag{3.28}$$

where $\varepsilon_k, k \in \mathbb{Z}$, are iid. As argued in Wu (2011), (3.28) itself defines a very large class of stationary processes and many widely used linear and nonlinear processes fall within the framework of (3.28). Our argument can be extended to the two-sided process (1.4) in a straightforward manner since our primary tool is the m-dependence approximation technique. In Section 3.1 we shall handle the pre-processing work of truncation, m-dependence approximation and blocking, and in Section 3.2 we shall apply Sakhanenko's (2006) Gaussian approximation result to the transformed processes, and establish conditional Gaussian approximations. Section 3.3 removes the conditioning, and an unconditional Gaussian approximation is obtained. In Section 3.4 we refine the unconditional Gaussian approximation in Section 3.3 by linearizing the variance function, so that one can have the readily applicable form (2.14).

3.1 Truncation, m-dependence Approximation and Blocking

For $a > 0$, define the truncation operator T_a by

$$T_a(w) = \max(\min(w, a), -a), \ w \in \mathbb{R}. \tag{3.29}$$

Then T_a is Lipschitz continuous and the Lipschitz constant is 1. For $n \geq 2$ let $h_n = \lceil (\log n)/(\log 3) \rceil$, so that $3^{h_n - 1} < n \leq 3^{h_n}$. Define

$$W_{k,l} = \sum_{i=1+3^{k-1}}^{l+3^{k-1}} [T_{3^{k/p}}(X_i) - \mathsf{E}T_{3^{k/p}}(X_i)] \tag{3.30}$$

and the m_k-dependent process

$$\tilde{X}_{k,j} = \mathsf{E}[T_{3^{k/p}}(X_j) | \varepsilon_{j-m_k}, \ldots, \varepsilon_{j-1}, \varepsilon_j] - \mathsf{E}T_{3^{k/p}}(X_j). \tag{3.31}$$

Let

$$S_n^{\dagger} = \sum_{k=1}^{h_n - 1} W_{k, 3^k - 3^{k-1}} + \sum_{i=1+3^{h_n-1}}^{n} [T_{3^{h_n/p}}(X_i) - \mathsf{E}T_{3^{h_n/p}}(X_i)] \tag{3.32}$$

and

$$\tilde{S}_n = \sum_{k=1}^{h_n-1} \tilde{W}_{k,3^k - 3^{k-1}} + \tilde{W}_{h_n, n - 3^{h_n - 1}}, \text{ where } \tilde{W}_{k,l} = \sum_{i=1+3^{k-1}}^{l+3^{k-1}} \tilde{X}_{k,i}. \tag{3.33}$$

If $n=1$, we let $S_1^\dagger = \tilde{S}_1 = 0$. Since $X_i \in \mathcal{L}^p$, we have

$$\max_{1\le i\le n} |S_i - S_i^\dagger| = a_{\text{a.s.}}(n^{1/p}). \tag{3.34}$$

Note that there exists a constant c_p such that, for all $k \ge 1$,

$$\left\| \max_{3^{k-1}<l\le 3^k} |\tilde{W}_{k,l} - W_{k,l}| \right\|_p \le c_p(3^k - 3^{k-1})^{1/2}\Theta_{1+m_k,p}. \tag{3.35}$$

Hence, by the Borel-Cantelli Lemma and condition (2.12), we have

$$\max_{1\le i\le n} |\tilde{S}_i - S_i^\dagger| = a_{\text{a.s.}}(n^{1/p}). \tag{3.36}$$

Let $q_k = \lfloor 2\times 3^{k-2}/m_k \rfloor - 2$. By (2.11), $m_k = o(3^{k(\alpha/p-1)/(\alpha/2-1)})$. Hence $\lim_{k\to\infty} q_k = \infty$. Choose $K_0 \in \mathbb{N}$ such that $q_k \ge 2$ whenever $k \ge K_0$ and let $N_0 = 3^{K_0}$. For $k \ge K_0$ define

$$B_{k,j} = \sum_{i=1+3jm_k+3^{k-1}}^{3(j+1)m_k+3^{k-1}} \tilde{X}_{k,i}, \ \ j = 1, 2, \ldots, q_k. \tag{3.37}$$

Let $B_{k,j} \equiv 0$ if $k < K_0$. In the sequel we assume throughout that $k \ge K_0$ and $n \ge N_0$. By Markov's inequality and the stationarity of the process $(\tilde{X}_{k,i})_{i\in\mathbb{Z}}$,

$$\begin{aligned} \mathsf{P}\left(\max_{1\le l\le 2\times 3^{k-1}} \left| \tilde{W}_{k,l} - \sum_{j=1}^{\lfloor l/(3m_k)\rfloor} B_{k,j} \right| \ge 3^{k/p} \right) &\le \frac{2\times 3^{k-1}}{m_k} \mathsf{P}\left(\max_{1\le l\le 3m_k} |\tilde{W}_{k,l}| \ge 3^{k/p} \right) \\ &\le \frac{3^k \mathsf{E}(\max_{1\le l\le 3m_k} |\tilde{W}_{k,l}|^\alpha)}{m_k 3^{k\alpha/p}}. \end{aligned} \tag{3.38}$$

We define the functional dependence measure for the process $(T_{3^{k/p}}(X_i))_{i\in\mathbb{Z}}$ as

$$\delta_{k,j,\iota} = \|T_{3^{k/p}}(X_i) - T_{3^{k/p}}(X_{i,\{i-j\}})\|_\iota, \tag{3.39}$$

where $\iota \ge 2$, and similarly the functional dependence measure for $(\tilde{X}_{k,i})$ as

$$\tilde{\delta}_{k,j,\iota} = \|\tilde{X}_{k,i} - \tilde{X}_{k,i,\{i-j\}})\|_\iota. \tag{3.40}$$

For those dependence measures, we can easily have the following simple relation:

$$\tilde{\delta}_{k,j,\iota} \le \delta_{k,j,\iota}, \ \ \delta_{k,j,p} \le \delta_{j,p} \text{ and } \delta_{k,j,2} \le \delta_{j,2}. \tag{3.41}$$

By the above relation, a careful check of the proof of Lemma 4.3 indicates that, under (2.10) and (2.11), there exists a constant $c = c_{\alpha,p}$ such that

$$\sum_{k=K_0}^\infty \frac{3^k}{m_k} \frac{\mathsf{E}(\max_{1\le l\le 3m_k} |\tilde{W}_{k,l}|^\alpha)}{3^{k\alpha/p}} \le c(M_{\alpha,p}\Theta_{0,2}^\alpha + \Xi_{\alpha,p}^\alpha + \|X_1\|_p^p). \tag{3.42}$$

The above inequality plays a critical role in our proof and it will be used again later. In (3.38), the largest index j is $\lfloor 2 \times 3^{k-1}/(3m_k) \rfloor = q_k + 2$. Note that B_{k,q_k} is independent of $B_{k+1,1}$. This motivates us to define the sum

$$S_n^{\diamond} = \sum_{k=K_0}^{h_n-1} \sum_{j=1}^{q_k} B_{k,j} + \sum_{j=1}^{\tau_n} B_{h_n,j}, \text{ where } \tau_n = \left\lfloor \frac{n - 3^{h_n-1}}{3m_{h_n}} \right\rfloor - 2. \tag{3.43}$$

We emphasize that the sums $\sum_{j=1}^{q_k} B_{k,j}$, $k = 1, 2, \ldots, h_n - 1$, and $\sum_{j=1}^{\tau_n} B_{h_n,j}$ are mutually independent. By (3.38), (3.42) and the Borel-Cantelli Lemma, we have

$$\max_{N_0 \le i \le n} |\tilde{S}_i - S_i^{\diamond}| = a_{\text{a.s.}}(n^{1/p}), \tag{3.44}$$

where we recall $N_0 = 3^{K_0}$. Summarizing the truncation approximation (3.34), the m-dependence approximation (3.36), and the block approximation (3.44), we have

$$\max_{N_0 \le i \le n} |S_i - S_i^{\diamond}| = a_{\text{a.s.}}(n^{1/p}) \tag{3.45}$$

and by Lemma 4.1 it remains to show that (2.14) holds with $S_n^{\diamond}$.

3.2 Conditional Gaussian Approximation

For $3^{k-1} < i \le 3^k$, $k \ge K_0$, let G_k be a measurable function such that

$$\tilde{X}_{k,i} = G_k(\varepsilon_{i-m_k}, \ldots, \varepsilon_i). \tag{3.46}$$

Recall $q_k = \lfloor 2 \times 3^{k-2}/m_k \rfloor - 2$. For $j = 1, 2, \ldots, q_k$ define

$$\mathcal{J}_{k,j} = \{3^{k-1} + (3j-1)m_k + l, \, l = 1, 2, \ldots, m_k\}. \tag{3.47}$$

Let $\mathbf{a} = (\mathbf{a}_{k,3j}, 1 \le j \le q_k)_{k=K_0}^{\infty}$ be a vector of real numbers, where $\mathbf{a}_{k,3j} = (a_l, \, l \in \mathcal{J}_{k,j})$, $j = 1, \ldots, q_k$. Define the random functions

$$\begin{aligned}
F_{k,3j}(\mathbf{a}_{k,3j}) &= \sum_{i=1+(3j-1)m_k}^{3jm_k} G_k(a_{i+3^{k-1}}, \cdots, a_{3jm_k+3^{k-1}}, \\
&\qquad \varepsilon_{3jm_k+1+3^{k-1}}, \cdots, \varepsilon_{i+m_k+3^{k-1}}); \\
F_{k,1+3j} &= \sum_{i=1+3jm_k}^{(1+3j)m_k} G_k(\varepsilon_{i+3^{k-1}}, \cdots, \varepsilon_{(1+3j)m_k+3^{k-1}}, \\
&\qquad \varepsilon_{(1+3j)m_k+1+3^{k-1}}, \cdots, \varepsilon_{i+m_k+3^{k-1}}); \\
F_{k,2+3j}(\mathbf{a}_{k,3+3j}) &= \sum_{i=1+(1+3j)m_k}^{(2+3j)m_k} G_k(\varepsilon_{i+3^{k-1}}, \cdots, \varepsilon_{(2+3j)m_k+3^{k-1}}, \\
&\qquad a_{(2+3j)m_k+1+3^{k-1}}, \cdots, a_{i+m_k+3^{k-1}}).
\end{aligned}$$

Let $\boldsymbol{\eta}_{k,3j} = (\varepsilon_l,\, l \in \mathcal{J}_{k,j})$, $j = 1, \ldots, q_k$, and $\boldsymbol{\eta} = (\boldsymbol{\eta}_{k,3j}, 1 \le j \le q_k)_{k=K_0}^{\infty}$. Then

$$B_{k,j} = F_{k,3j}(\boldsymbol{\eta}_{k,3j}) + F_{k,1+3j} + F_{k,2+3j}(\boldsymbol{\eta}_{k,3j+3}). \tag{3.48}$$

Note that $\mathsf{E}F_{k,1+3j} = 0$. Define the mean functions

$$\Lambda_{k,0}(\mathbf{a}_{k,3j}) = \mathsf{E}F_{k,3j}(\mathbf{a}_{k,3j}), \quad \Lambda_{k,2}(\mathbf{a}_{k,3+3j}) = \mathsf{E}F_{k,2+3j}(\mathbf{a}_{k,3+3j}).$$

Introduce the centered process

$$\begin{aligned} Y_{k,j}(\mathbf{a}_{k,3j}, \mathbf{a}_{k,3j+3}) &= [F_{k,3j}(\mathbf{a}_{k,3j}) - \Lambda_{k,0}(\mathbf{a}_{k,3j})] \\ &\quad + F_{k,3j+1} + [F_{k,3j+2}(\mathbf{a}_{k,3j+3}) - \Lambda_{k,2}(\mathbf{a}_{k,3j+3})]. \end{aligned} \tag{3.49}$$

Then $Y_{k,j}(\mathbf{a}_{k,3j}, \mathbf{a}_{k,3j+3})$, $j = 1, \ldots, q_k$, $k \ge K_0$, are mean zero independent random variables with variance function

$$\begin{aligned} V_k(\mathbf{a}_{k,3j}, \mathbf{a}_{k,3j+3}) &= \|Y_{k,j}(\mathbf{a}_{k,3j}, \mathbf{a}_{k,3j+3})\|^2 \\ &= \|F_{k,3j}(\mathbf{a}_{k,3j}) - \Lambda_{k,0}(\mathbf{a}_{k,3j})\|^2 + \|F_{k,3j+1}\|^2 \\ &\quad + 2\mathsf{E}\{F_{k,3j+1}[F_{k,3j}(\mathbf{a}_{k,3j}) - \Lambda_{k,0}(\mathbf{a}_{k,3j})]\} \\ &\quad + \|F_{k,3j+2}(\mathbf{a}_{k,3j+3}) - \Lambda_{k,2}(\mathbf{a}_{k,3j+3})\|^2 \\ &\quad + 2\mathsf{E}\{F_{k,3j+1}[F_{k,3j+2}(\mathbf{a}_{k,3j+3}) - \Lambda_{k,2}(\mathbf{a}_{k,3j+3})]\}, \end{aligned} \tag{3.50}$$

since $[F_{k,3j}(\mathbf{a}_{k,3j}) - \Lambda_{k,0}(\mathbf{a}_{k,3j})]$ and $[F_{k,3j+2}(\mathbf{a}_{k,3j+3}) - \Lambda_{k,2}(\mathbf{a}_{k,3j+3})]$ are independent. Following the definition of $S_n^{\diamond}$ in (3.43), we let

$$T_n(\mathbf{a}) = \sum_{k=K_0}^{h_n-1} \sum_{j=1}^{q_k} Y_{k,j}(\mathbf{a}_{k,3j}, \mathbf{a}_{k,3j+3}) + \sum_{j=1}^{\tau_n} Y_{h_n,j}(\mathbf{a}_{h_n,3j}, \mathbf{a}_{h_n,3j+3}). \tag{3.51}$$

Define the mean function

$$\begin{aligned} M_n(\mathbf{a}) &= \sum_{k=K_0}^{h_n-1} \sum_{j=1}^{q_k} [\Lambda_{k,0}(\mathbf{a}_{k,3j}) + \Lambda_{k,2}(\mathbf{a}_{k,3+3j})] \\ &\qquad + \sum_{j=1}^{\tau_n} [\Lambda_{h_n,0}(\mathbf{a}_{h_n,3j}) + \Lambda_{h_n,2}(\mathbf{a}_{h_n,3+3j})], \end{aligned}$$

and variance of $T_n(\mathbf{a})$:

$$Q_n(\mathbf{a}) = \sum_{k=K_0}^{h_n-1} \sum_{j=1}^{q_k} V_k(\mathbf{a}_{k,3j}, \mathbf{a}_{k,3+3j}) + \sum_{j=1}^{\tau_n} V_{h_n}(\mathbf{a}_{h_n,3j}, \mathbf{a}_{h_n,3+3j}).$$

Let

$$V_k^{\circ}(\mathbf{a}_{k,3j}) = \|[F_{k,3j}(\mathbf{a}_{k,3j}) - \Lambda_{k,0}(\mathbf{a}_{k,3j})]$$

$$
\begin{aligned}
&\quad +F_{k,3j+1} + [F_{k,3j+2}(\mathbf{a}_{k,3j}) - \Lambda_{k,2}(\mathbf{a}_{k,3j})]\|^2 \\
&= \|F_{k,3j}(\mathbf{a}_{k,3j}) - \Lambda_{k,0}(\mathbf{a}_{k,3j})\|^2 + \|F_{k,3j+1}\|^2 \\
&\quad +2\mathsf{E}\{F_{k,3j+1}[F_{k,3j}(\mathbf{a}_{k,3j}) - \Lambda_{k,0}(\mathbf{a}_{k,3j})]\} \\
&\quad +\|F_{k,3j+2}(\mathbf{a}_{k,3j}) - \Lambda_{k,2}(\mathbf{a}_{k,3j})\|^2 \\
&\quad +2\mathsf{E}\{F_{k,3j+1}[F_{k,3j+2}(\mathbf{a}_{k,3j}) - \Lambda_{k,2}(\mathbf{a}_{k,3j})]\}, \\
L_k(\mathbf{a}_{k,3j}) &= \|F_{k,3j+1} + [F_{k,3j+2}(\mathbf{a}_{k,3j}) - \Lambda_{k,2}(\mathbf{a}_{k,3j})]\|^2 \\
&= \|F_{k,3j+1}\|^2 + \|[F_{k,3j+2}(\mathbf{a}_{k,3j}) - \Lambda_{k,2}(\mathbf{a}_{k,3j})]\|^2 \\
&\quad +2\mathsf{E}\{F_{k,3j+1}[F_{k,3j+2}(\mathbf{a}_{k,3j}) - \Lambda_{k,2}(\mathbf{a}_{k,3j})]\}. \qquad (3.52)
\end{aligned}
$$

By the formulae of $V_k(\mathbf{a}_{k,3j}, \mathbf{a}_{k,3j+3})$ in (3.50) and $V_k^\circ(\mathbf{a}_{k,3j})$ and $L_k(\mathbf{a}_{k,3j})$ in (3.52), we have the following identity:

$$
L_k(\mathbf{a}_{k,3}) + \sum_{j=1}^{t} V_k(\mathbf{a}_{k,3j}, \mathbf{a}_{k,3j+3}) = \sum_{j=1}^{t} V_k^\circ(\mathbf{a}_{k,3j}) + L_k(\mathbf{a}_{k,3+3t}) \qquad (3.53)
$$

holds for all $t \ge 1$. The above identity motivates us to introduce the auxiliary process

$$
\Gamma_n(\mathbf{a}) = \sum_{k=K_0}^{h_n-1} L_k(\mathbf{a}_{k,3})^{1/2}\zeta_k + L_{h_n}(\mathbf{a}_{h_n,3})^{1/2}\zeta_{h_n}, \qquad (3.54)
$$

where $\zeta_l, l \in \mathbb{Z}$, are i.i.d. standard normal random variables which are independent of $(\varepsilon_i)_{i\in\mathbb{Z}}$. Then in view of (3.53) the variance of $T_n(\mathbf{a}) + \Gamma_n(\mathbf{a})$ is given by

$$
\begin{aligned}
Q_n^\circ(\mathbf{a}) &= \sum_{k=K_0}^{h_n-1} \left[\sum_{j=1}^{q_k} V_k^\circ(\mathbf{a}_{k,3j}) + L_k(\mathbf{a}_{k,3+3q_k})\right] \\
&\quad + \sum_{j=1}^{\tau_n} \left[V_{h_n}^\circ(\mathbf{a}_{h_n,3j}) + L_{h_n}(\mathbf{a}_{k_n,3+3\tau_n})\right]. \qquad (3.55)
\end{aligned}
$$

In studying $T_n(\mathbf{a})+\Gamma_n(\mathbf{a})$, for notational convenience, for $j = 0$ we let $Y_{k,0}(\mathbf{a}_{k,0}, \mathbf{a}_{k,3}) = L_k(\mathbf{a}_{k,3})^{1/2}\zeta_k$. We shall now apply Sakhanenko's (1991, 2006) Gaussian approximation result. To this end, for $x > 0$, we define

$$
\begin{aligned}
\Psi_h(\mathbf{a}, x, \alpha) &= \sum_{k=K_0}^{h} \sum_{j=0}^{q_k} \mathsf{E}\min\{|Y_{k,j}(\mathbf{a}_{k,3j}, \mathbf{a}_{k,3j+3})/x|^\alpha,\ |Y_{k,j}(\mathbf{a}_{k,3j}, \mathbf{a}_{k,3j+3})/x|^2\} \\
&\le \sum_{k=K_0}^{h} \sum_{j=0}^{q_k} \mathsf{E}|Y_{k,j}(\mathbf{a}_{k,3j}, \mathbf{a}_{k,3j+3})/x|^\alpha. \qquad (3.56)
\end{aligned}
$$

By Theorem 1 in Sakhanenko (2006), there exists a probability space $(\Omega_\mathbf{a}, \mathcal{A}_\mathbf{a}, \mathsf{P}_\mathbf{a})$ on which we can define a standard Brownian motion $\mathbb{B}_\mathbf{a}$ and random variables $R_{k,j}^\mathbf{a}$ such that the distributional equality

$$
(R_{k,j}^\mathbf{a})_{0\le j\le q_k,\, k\ge K_0} =_{\mathcal{D}} (Y_{k,j}(\mathbf{a}_{k,3j}, \mathbf{a}_{k,3j+3}))_{0\le j\le q_k,\, k\ge K_0} \qquad (3.57)
$$

holds, and, for the partial sum processes

$$\Upsilon_n^{\mathbf{a}} = \sum_{k=K_0}^{h-1} \sum_{j=1}^{q_k} R_{k,j}^{\mathbf{a}} + \sum_{j=1}^{\tau_n} R_{h_n,j}^{\mathbf{a}} \quad \text{and} \quad \mu_n^{\mathbf{a}} = \sum_{k=K_0}^{h-1} R_{k,0}^{\mathbf{a}} + R_{h_n,0}^{\mathbf{a}}, \tag{3.58}$$

we have for all $x > 0$ and $\alpha > p$ that

$$\mathsf{P}_{\mathbf{a}} \left[\max_{N_0 \le i \le 3^h} |(\Upsilon_i^{\mathbf{a}} + \mu_i^{\mathbf{a}}) - \mathbb{B}_{\mathbf{a}}(Q_i^{\circ}(\mathbf{a}))| \ge c_0 \alpha x \right] \le \Psi_h(\mathbf{a}, x, \alpha). \tag{3.59}$$

Here c_0 is an absolute constant. By Jensen's inequality, for both $j = 0$ and $j > 0$, there exists a constant c_α such that

$$\mathsf{E}[|Y_{k,j}(\boldsymbol{\eta}_{k,3j}, \boldsymbol{\eta}_{k,3j+3})|^{\alpha}] \le c_\alpha \mathsf{E}(|\tilde{W}_{k,m_k}|^{\alpha}). \tag{3.60}$$

In (3.59) we let $x = 3^{h/p}$ and by Lemma 4.2 (see also (3.42)),

$$\begin{aligned} \sum_{h=K_0}^{\infty} \mathsf{E}[\Psi_h(\boldsymbol{\eta}, 3^{h/p}, \alpha)] &\le \sum_{h=K_0}^{\infty} \sum_{k=K_0}^{h} \frac{q_k + 1}{3^{\alpha h/p}} c_\alpha \mathsf{E}(|\tilde{W}_{k,m_k}|^{\alpha}) \\ &\le \sum_{k=K_0}^{\infty} \sum_{h=k}^{\infty} \frac{3^k c_\alpha}{m_k 3^{\alpha h/p}} \mathsf{E}\left(\max_{1 \le l \le 3m_k} |\tilde{W}_{k,l}|^{\alpha} \right) \\ &< \infty. \end{aligned} \tag{3.61}$$

Hence, by the Borel-Cantelli Lemma, we obtain

$$\max_{i \le n} |(\Upsilon_i^{\boldsymbol{\eta}} + \mu_i^{\boldsymbol{\eta}}) - \mathbb{B}_{\boldsymbol{\eta}}(Q_i^{\circ}(\boldsymbol{\eta}))| = o_{\text{a.s.}}(n^{1/p}). \tag{3.62}$$

The probability space for the above almost sure convergence is

$$(\Omega_*, \mathcal{A}_*, \mathsf{P}_*) = (\Omega, \mathcal{A}, \mathsf{P}) \times \prod_{\tau \in \Omega} (\Omega_{\boldsymbol{\eta}(\tau)}, \mathcal{A}_{\boldsymbol{\eta}(\tau)}, \mathsf{P}_{\boldsymbol{\eta}(\tau)}), \tag{3.63}$$

where $(\Omega, \mathcal{A}, \mathsf{P})$ is the probability space on which the random variables $(\varepsilon_i)_{i \in \mathbb{Z}}$ are defined and, for a set $A \subset \Omega_*$ with $A \in \mathcal{A}_*$, the probability measure P_* is defined as

$$\mathsf{P}_*(A) = \int_{\Omega} \mathsf{P}_{\boldsymbol{\eta}(\omega)}(A_\omega) \mathsf{P}(d\omega), \tag{3.64}$$

where A_ω is the ω-section of A. Here we recall that, for each $\mathbf{a}$, $(\Omega_{\mathbf{a}}, \mathcal{A}_{\mathbf{a}}, \mathsf{P}_{\mathbf{a}})$ is the probability space carrying $\mathbb{B}_{\mathbf{a}}$ and $R_{k,j}^{\mathbf{a}}$ given $\boldsymbol{\eta} = \mathbf{a}$. On the probability space $(\Omega_*, \mathcal{A}_*, \mathsf{P}_*)$, the random variable $R_{k,j}^{\boldsymbol{\eta}}$ is defined as $R_{k,j}^{\boldsymbol{\eta}}(\omega, \theta(\cdot)) = R_{k,j}^{\boldsymbol{\eta}(\omega)}(\theta(\omega))$, where $(\omega, \theta(\cdot)) \in \Omega_*$, $\theta(\cdot)$ is an element in $\prod_{\tau \in \Omega} \Omega_{\boldsymbol{\eta}(\tau)}$ and $\theta(\tau) \in \Omega_{\boldsymbol{\eta}(\tau)}$, $\tau \in \Omega$. The other random processes $\mu_i^{\boldsymbol{\eta}}$ and $\mathbb{B}_{\boldsymbol{\eta}}(Q_i^{\circ}(\boldsymbol{\eta}))$ can be similarly defined.

3.3 Unconditional Gaussian Approximation

In this subsection we shall work with the processes $\Upsilon_i^{\boldsymbol{\eta}}$, $\mu_i^{\boldsymbol{\eta}}$ and $I\!B_{\boldsymbol{\eta}}(Q_i^\circ(\boldsymbol{\eta}))$. Write

$$I\!B_{\mathbf{a}}(Q_n^\circ(\mathbf{a})) = \varpi_n(\mathbf{a}) + \varphi_n(\mathbf{a}), \tag{3.65}$$

where, since Brownian motions have independent increments,

$$\begin{aligned}
\varpi_n(\mathbf{a}) &= \sum_{k=K_0}^{h_n-1}\sum_{j=1}^{q_k} V_k^\circ(\mathbf{a}_{k,3j})^{1/2} Z_{k,j}^{\mathbf{a}} + \sum_{j=1}^{\tau_n} V_{h_n}^\circ(\mathbf{a}_{h_n,3j})^{1/2} Z_{h_n,j}^{\mathbf{a}}, \\
\varphi_n(\mathbf{a}) &= \sum_{k=K_0}^{h_n-1} L_k(\mathbf{a}_{k,3+3q_k})^{1/2} Z_{k,1+q_k}^{\mathbf{a}} + L_{h_n}(\mathbf{a}_{k_n,3+3\tau_n})^{1/2} Z_{h_n,1+\tau_n}^{\mathbf{a}}.
\end{aligned}$$

Here $Z_{i,l}^{\mathbf{a}}, i, l \in \mathbb{Z}$, are iid standard normal random variables. Let $Z_{i,l}^\star, i, l \in \mathbb{Z}$, independent of $(\varepsilon_j)_{j\in\mathbb{Z}}$, be also iid standard normal random variables and define

$$\begin{aligned}
\Phi_n &= \sum_{k=K_0}^{h_n-1}\sum_{j=1}^{q_k} V_k^\circ(\boldsymbol{\eta}_{k,3j})^{1/2} Z_{k,j}^\star + \sum_{j=1}^{\tau_n} V_{h_n}^\circ(\boldsymbol{\eta}_{h_n,3j})^{1/2} Z_{h_n,j}^\star, \\
\chi_n &= \sum_{k=K_0}^{h_n-1} L_k(\boldsymbol{\eta}_{k,3+3q_k})^{1/2} Z_{k,1+q_k}^\star + L_{h_n}(\boldsymbol{\eta}_{h_n,3+3\tau_n})^{1/2} Z_{h_n,1+\tau_n}^\star.
\end{aligned}$$

Since $Z_{i,l}^{\mathbf{a}}$, are iid standard normal, the conditional distribution $[\varpi_n(\boldsymbol{\eta})|\boldsymbol{\eta} = \mathbf{a}]$, namely the distribution of $\varpi_n(\mathbf{a})$, is same as that of Φ_n. Hence

$$(\Phi_i, \chi_i)_{i\ge N_0} =_{\mathcal{D}} (\varpi_i(\boldsymbol{\eta}), \varphi_i(\boldsymbol{\eta}))_{i\ge N_0}. \tag{3.66}$$

By Jensen's equality, $\mathsf{E}[|L_k(\boldsymbol{\eta}_{k,3+3j})^{1/2}|^\alpha] \le 3^\alpha \mathsf{E}(|\tilde{W}_{k,m_k}|^\alpha)$. By (3.42),

$$\begin{aligned}
\sum_{k=K_0}^{\infty} \mathsf{P}\left(\max_{1\le j\le q_k} |L_k(\boldsymbol{\eta}_{k,3+3j})^{1/2} Z_{k,1+j}^\star| \ge 3^{k/p}\right) &\le \sum_{k=K_0}^{\infty} q_k \frac{\mathsf{E}[|L_k(\boldsymbol{\eta}_{k,3})^{1/2} Z_{k,1}^\star|^\alpha]}{3^{k\alpha/p}} \\
&\le \sum_{k=K_0}^{\infty} q_k \frac{c_\alpha \mathsf{E}(|\tilde{W}_{k,m_k}|^\alpha)}{3^{k\alpha/p}} \\
&< \infty,
\end{aligned} \tag{3.67}$$

which by the Borel-Cantelli Lemma implies

$$\max_{i\le n} |\chi_i| = o_{\text{a.s.}}(n^{1/p}). \tag{3.68}$$

The same argument also implies that $\max_{i\le n} |\Gamma_i(\boldsymbol{\eta})| = o_{\text{a.s.}}(n^{1/p})$ and consequently

$$\max_{i\le n} |\mu_i^{\boldsymbol{\eta}}| = o_{\text{a.s.}}(n^{1/p}) \tag{3.69}$$

in view of (3.57) with $j = 0$. Hence, by (3.62) and (3.65), we have $\max_{i\le n} |\Upsilon_i^{\boldsymbol{\eta}} - \varpi_i(\boldsymbol{\eta})| = o_{\text{a.s.}}(n^{1/p})$. Observe that, by (3.57), (3.58), (3.48) and (3.49), we have the distributional equality,

$$(\Upsilon_i^{\boldsymbol{\eta}} + M_i(\boldsymbol{\eta}))_{i\ge N_0} =_{\mathcal{D}} (S_i^{\diamond})_{i\ge N_0}, \tag{3.70}$$

where we recall (3.43) for the definition of $S_n^{\diamond}$. Then it remains to establish a strong invariance principle for $\Phi_n + M_n(\boldsymbol{\eta})$. To this end, let

$$A_{k,j} = V_k^{\circ}(\boldsymbol{\eta}_{k,3j})^{1/2} Z_{k,j}^{\star} + \Lambda_{k,0}(\boldsymbol{\eta}_{k,3j}) + \Lambda_{k,2}(\boldsymbol{\eta}_{k,3j}), \tag{3.71}$$

which are independent random variables for $j = 1, \ldots, q_k$ and $k \ge K_0$, and let

$$S_n^{\natural} = \sum_{k=K_0}^{h_n-1} \sum_{j=1}^{q_k} A_{k,j} + \sum_{j=1}^{\tau_n} A_{h_n,j} \tag{3.72}$$

and $R_n^{\natural} = \Phi_n + M_n(\boldsymbol{\eta}) - S_n^{\natural}$. Note that

$$R_n^{\natural} = \sum_{k=K_0}^{h_n-1} [\Lambda_{k,2}(\boldsymbol{\eta}_{k,3+3q_k}) - \Lambda_{k,2}(\boldsymbol{\eta}_{k,3})] + [\Lambda_{h_n,2}(\boldsymbol{\eta}_{k_n,3+3\tau_n}) - \Lambda_{h_n,2}(\boldsymbol{\eta}_{k_n,3})].$$

Then using the same argument as in (3.67), we have

$$\max_{i\le n} |R_i^{\natural}| = \max_{i\le n} |\Phi_i + M_i(\boldsymbol{\eta}) - S_i^{\natural}| = o_{\text{a.s.}}(n^{1/p}). \tag{3.73}$$

The variance of $S_n^{\natural}$ equals to

$$\sigma_n^2 = \sum_{k=K_0}^{h_n-1} \sum_{j=1}^{q_k} \|A_{k,j}\|^2 + \sum_{j=1}^{\tau_n} \|A_{h_n,j}\|^2 = \sum_{k=K_0}^{h_n-1} q_k \|A_{k,1}\|^2 + \tau_n \|A_{h_n,1}\|^2. \tag{3.74}$$

Again by Theorem 1 in Sakhanenko (2006), on the same probability space that defines $(A_{k,j})_{1\le j\le q_k,\, k\ge K_0}$, by the argument in (3.59)–(3.62), there exists a standard Brownian motion $\mathbb{B}$ such that

$$\max_{i\le n} |S_i^{\natural} - \mathbb{B}(\sigma_i^2)| = o_{\text{a.s.}}(n^{1/p}). \tag{3.75}$$

3.4 Regularizing the Gaussian Approximation

In this section we shall regularize the Gaussian approximation (3.75) by replacing the variance function σ_i^2 by the asymptotic linear form ϕ_i or the linear

form $i\sigma^2$ and the latter is more easily usable. By (3.52), we obtain

$$V_k^{\circ}(\mathbf{a}_{k,3j}) \;=\; \|F_{k,3j}(\mathbf{a}_{k,3j})\|^2 - \Lambda_{k,0}(\mathbf{a}_{k,3j})^2 + \|F_{k,3j+1}\|^2$$

$$
\begin{aligned}
&+2\mathsf{E}\{F_{k,3j+1}F_{k,3j}(\mathbf{a}_{k,3j})\}\\
&+\|F_{k,3j+2}(\mathbf{a}_{k,3j})\|^2 - \Lambda_{k,2}(\mathbf{a}_{k,3j})^2\\
&+2\mathsf{E}\{F_{k,3j+1}F_{k,3j+2}(\mathbf{a}_{k,3j})\}, \qquad (3.76)
\end{aligned}
$$

which, by the expression of $A_{k,j}$, implies that

$$
\begin{aligned}
\|A_{k,j}\|^2 &= \mathsf{E}[V_k^\circ(\boldsymbol{\eta}_{k,3j})] + \mathsf{E}[\Lambda_{k,0}(\boldsymbol{\eta}_{k,3j}) + \Lambda_{k,2}(\boldsymbol{\eta}_{k,3j})]^2\\
&= 3\mathsf{E}[\tilde{W}_{k,m_k}^2 + 2\tilde{W}_{k,m_k}(\tilde{W}_{k,2m_k} - \tilde{W}_{k,m_k})]. \qquad (3.77)
\end{aligned}
$$

Let $\tilde{\gamma}_{k,i} = \mathsf{E}(\tilde{X}_{k,0}\tilde{X}_{k,i})$. Then $\nu_k := \|A_{k,j}\|^2/(3m_k)$ has the expression

$$
\begin{aligned}
\nu_k &= \frac{1}{m_k}\mathsf{E}[\tilde{W}_{k,m_k}^2 + 2\tilde{W}_{k,m_k}(\tilde{W}_{k,2m_k} - \tilde{W}_{k,m_k})]\\
&= \sum_{i=-m_k}^{m_k}\tilde{\gamma}_{k,i} + 2\sum_{i=1}^{m_k}(1 - i/m_k)\tilde{\gamma}_{k,m_k+i}. \qquad (3.78)
\end{aligned}
$$

We now prove that

$$
\nu_k - \sigma^2 = O\left[\Theta_{m_k,p} + \min_{l\ge 0}(\Theta_{l,p} + l3^{k(2/p-1)})\right], \qquad (3.79)
$$

which converges to 0 if $k \to \infty$. Let $\hat{X}_{k,i} = T_{3^{k/p}}(X_i)$ and $\hat{\gamma}_{k,i} = \mathrm{cov}(\hat{X}_{k,0}, \hat{X}_{k,i}) = \mathsf{E}(\hat{X}_{k,0}\hat{X}_{k,i}) - [\mathsf{E}(\hat{X}_{k,0})]^2$. Note that if $|X_i| \le 3^{k/p}$, then $X_i = \hat{X}_{k,i}$. Since $X_i \in \mathcal{L}^p$,

$$
\begin{aligned}
|\mathsf{E}(X_0X_i) - \mathsf{E}(\hat{X}_{k,0}\hat{X}_{k,i})| &= |\mathsf{E}(X_0X_i\mathbf{1}_{|X_0|\le 3^{k/p},\ |X_i|\le 3^{k/p}}) - \mathsf{E}(\hat{X}_{k,0}\hat{X}_{k,i})\\
&\quad + \mathsf{E}(X_0X_i\mathbf{1}_{\max(|X_0|,\ |X_i|)>3^{k/p}})|\\
&\le |\mathsf{E}(\hat{X}_{k,0}\hat{X}_{k,i}\mathbf{1}_{\max(|X_0|,\ |X_i|)>3^{k/p}})|\\
&\quad + |\mathsf{E}(X_0X_i\mathbf{1}_{\max(|X_0|,\ |X_i|)>3^{k/p}})|\\
&\le 2\mathsf{E}[(|X_0| + |X_i|)^2\mathbf{1}_{|X_0|+|X_i|>3^{k/p}}]\\
&= o(3^{k(2-p)/p}). \qquad (3.80)
\end{aligned}
$$

Clearly, we also have $\mathsf{E}(\hat{X}_{k,0}) = o(3^{k(2-p)/p})$. Hence

$$
\sup_i |\hat{\gamma}_{k,i} - \gamma_i| = o(3^{k(2-p)/p}). \qquad (3.81)
$$

For all $j \ge 1$, we have $\|W_{k,j} - \tilde{W}_{k,j}\| \le j^{1/2}\Theta_{m_k,2} \le j^{1/2}\Theta_{m_k,p}$. Then

$$
|\mathsf{E}W_{k,j}^2 - \mathsf{E}\tilde{W}_{k,j}^2| \le \|W_{k,j} - \tilde{W}_{k,j}\|\|W_{k,j} + \tilde{W}_{k,j}\| \le 2j\Theta_{m_k,p}\Theta_{0,p}. \qquad (3.82)
$$

Since $\lim_{j\to\infty} j^{-1}\mathsf{E}\tilde{W}_{k,j}^2 = \sum_{i=-m_k}^{m_k}\tilde{\gamma}_{k,i}$ and $\lim_{j\to\infty} j^{-1}\mathsf{E}W_{k,j}^2 = \sum_{i\in\mathbb{Z}}\hat{\gamma}_{k,i}$, (3.82) implies that

$$
\left|\sum_{i=-m_k}^{m_k}\tilde{\gamma}_{k,i} - \sum_{i\in\mathbb{Z}}\hat{\gamma}_{k,i}\right| \le 2\Theta_{m_k,p}\Theta_{0,p}. \qquad (3.83)
$$

Let the projection operator $\mathcal{P}_l \cdot = \mathsf{E}(\cdot|\mathcal{F}_l) - \mathsf{E}(\cdot|\mathcal{F}_{l-1})$. Then $\hat{X}_{k,i} = \sum_{l\in\mathbb{Z}} \mathcal{P}_l \hat{X}_{k,i}$. By the orthogonality of $\mathcal{P}_l, l \in \mathbb{Z}$, and inequality (3.41),

$$\begin{aligned} |\hat{\gamma}_{k,i}| &= \left| \sum_{l\in\mathbb{Z}} \sum_{l'\in\mathbb{Z}} \mathsf{E}[(\mathcal{P}_l \hat{X}_{k,0})(\mathcal{P}_{l'} \hat{X}_{k,i})] \right| \\ &\le \sum_{l\in\mathbb{Z}} \|\mathcal{P}_l \hat{X}_{k,0}\| \|\mathcal{P}_l \hat{X}_{k,i}\| \le \sum_{j=0}^{\infty} \delta_{j,p} \delta_{j+i,p}. \end{aligned} \tag{3.84}$$

The same inequality also holds for $|\gamma_i|$ and $|\tilde{\gamma}_{k,i}|$. For any $0 \le l \le m_k$, we have by (3.84) that

$$\sum_{i=l}^{\infty} (|\hat{\gamma}_{k,i}| + |\tilde{\gamma}_{k,i}| + |\gamma_i|) \le 3 \sum_{i=l}^{\infty} \sum_{j=0}^{\infty} \delta_{j,p} \delta_{j+i,p} \le 3\Theta_{0,p}\Theta_{l,p}, \tag{3.85}$$

which entails (3.79) in view of (3.81), (3.83) and (3.78).

Recall (3.74) and (3.75) for σ_n^2. Now we shall compare σ_n^2 with

$$\phi_n = \sum_{k=1}^{h_n-1} (3^k - 3^{k-1})\nu_k + (n - 3^{h_n-1})\nu_{h_n}. \tag{3.86}$$

Then ϕ_n is a piecewise linear function. Observe that, by (2.11),

$$\max_{i\le n} |\phi_i - \sigma_i^2| \le 3 \max_{k\le h_n} (m_k \nu_k) = o(n^{(\alpha/p-1)/(\alpha/2-1)}). \tag{3.87}$$

By increment properties of Brownian motions, we obtain

$$\max_{i\le n} |I\!B(\phi_i) - I\!B(\sigma_i^2)| = o_{\text{a.s.}}(n^{(\alpha/p-1)/(\alpha-2)} \log n) = o_{\text{a.s.}}(n^{1/p}). \tag{3.88}$$

Note that by (3.79), ϕ_i is asymptotically linear with slope σ^2. Here we emphasize that, under (2.10), (2.11), (2.12), a strong invariance principle with the Brownian motion $I\!B(\phi_i)$ holds in view of (3.45), (3.70), (3.73), (3.75), (3.88) and Lemma 4.1. However, the approximation $I\!B(\phi_i)$ is not convenient for use since ϕ_i is not genuinely linear.

Next, under condition (2.13), we shall linearize the variance function ϕ_i, so that one can have the readily applicable form (2.14). Based on the form of ϕ_i, we write

$$I\!B(\phi_n) = \sum_{k=1}^{h_n-1} \sum_{j=1}^{3^k-3^{k-1}} \nu_k^{1/2} Z_{k,j} + \sum_{j=1}^{n-3^{h_n-1}} \nu_{h_n}^{1/2} Z_{h_n,j}, \tag{3.89}$$

where $Z_{k,j}$ are i.i.d. standard normal random variables. Define

$$I\!B^\ddagger(n) = \sum_{k=1}^{h_n-1} \sum_{j=1}^{3^k-3^{k-1}} Z_{k,j} + \sum_{j=1}^{n-3^{h_n-1}} Z_{h_n,j}, \tag{3.90}$$

which is a standard Brownian motion with values at positive integers. Then we can write

$$\mathbb{B}(\phi_n) - \sigma \mathbb{B}^{\ddagger}(n) = \sum_{i=2}^{n} b_i Z_i, \tag{3.91}$$

where $(Z_2, Z_3, Z_4, \ldots) = (Z_{1,1}, Z_{1,2}, Z_{2,1}, Z_{2,2}, \ldots, Z_{2,6}, \ldots, Z_{k,1}, \ldots, Z_{k,3^k-3^{k-1}}, \ldots)$ is a lexicographic re-arrangement of $Z_{k,j}$,

and the coefficients $b_n = \nu_{h_n}^{1/2} - \sigma$. Then

$$\begin{aligned} \varsigma_n^2 &= \|\mathbb{B}(\phi_n) - \sigma \mathbb{B}^{\ddagger}(n)\|^2 = \sum_{i=2}^{n} b_i^2 \\ &= \sum_{k=1}^{h_n-1} (3^k - 3^{k-1})(\nu_k^{1/2} - \sigma)^2 + (n - 3^{h_n-1})(\nu_{h_n}^{1/2} - \sigma)^2 \end{aligned} \tag{3.92}$$

and ς_n^2 is nondecreasing. If $\lim_{n\to\infty} \varsigma_n^2 < \infty$, then trivially we have

$$\mathbb{B}(\phi_n) - \sigma \mathbb{B}^{\ddagger}(n) = o_{\text{a.s.}}(n^{1/p}). \tag{3.93}$$

We shall now prove (3.93) under the assumption that $\lim_{n\to\infty} \varsigma_n^2 = \infty$. Under the latter condition, note that we can represent $\mathbb{B}(\phi_n) - \sigma \mathbb{B}^{\ddagger}(n)$ as another Brownian motion $\mathbb{B}_0(\varsigma_n^2)$, by the Law of the Iterated Logarithm for Brownian motion, we have

$$\overline{\underline{\lim}}_{n\to\infty} \frac{\mathbb{B}(\phi_n) - \sigma \mathbb{B}^{\ddagger}(n)}{\sqrt{2\varsigma_n^2 \log\log \varsigma_n^2}} = \pm 1 \text{ almost surely.} \tag{3.94}$$

Then (3.93) follows if we can show that

$$\varsigma_n^2 \log\log n = o(n^{2/p}). \tag{3.95}$$

Note that (3.79) and (2.13) imply that $3^k(\nu_k^{1/2} - \sigma)^2 = o(3^{2k/p}/\log k)$, which entails (3.95) in view of (3.92). ♢

4 Some Useful Lemmas

In this section we shall provide some lemmas that are used in Section 3. Lemma 4.1 is a "gluing" lemma and it concerns how to combine almost sure convergences in different probability spaces. Lemma 4.2 relates truncated and original moments, and Lemma 4.3 gives an inequality for moments of maximum sums.

Lemma 4.1 *Let $(T_{1,n})_{n\geq 1}$ and $(U_{1,n})_{n\geq 1}$ be two sequences of random variables defined on the probability space $(\Omega_1, \mathcal{A}_1, P_1)$ such that $T_{1,n} - U_{1,n} \to 0$ almost surely; let*

$(T_{2,n})_{n\geq 1}$ *and* $(U_{2,n})_{n\geq 1}$ *be another two sequences of random variables defined on the probability space* $(\Omega_2, \mathcal{A}_2, P_2)$ *such that* $T_{2,n} - U_{2,n} \to 0$ *almost surely. Assume that the distributional equality* $(U_{1,n})_{n\geq 1} \overset{\mathcal{D}}{=} (T_{2,n})_{n\geq 1}$ *holds. Then we can construct a probability space* $(\Omega^\dagger, \mathcal{A}^\dagger, P^\dagger)$ *on which we can define* $(T'_{1,n})_{n\geq 1}$ *and* $(U'_{2,n})_{n\geq 1}$ *such that* $(T'_{1,n})_{n\geq 1} \overset{\mathcal{D}}{=} (T_{1,n})_{n\geq 1}$*,* $(U'_{2,n})_{n\geq 1} \overset{\mathcal{D}}{=} (U_{2,n})_{n\geq 1}$ *and* $T'_{1,n} - U'_{2,n} \to 0$ *almost surely in* $(\Omega^\dagger, \mathcal{A}^\dagger, P^\dagger)$*.*

Proof. Let $\mathbf{T_1} = (T_{1,n})_{n\geq 1}$, $\mathbf{U_1} = (U_{1,n})_{n\geq 1}$, $\mathbf{T_2} = (T_{2,n})_{n\geq 1}$, $\mathbf{U_2} = (U_{2,n})_{n\geq 1}$; let $\mu_{\mathbf{T_1}|\mathbf{U_1}}$ and $\mu_{\mathbf{U_2}|\mathbf{T_2}}$ denote the conditional distribution of $\mathbf{T_1}$ given $\mathbf{U_1}$, resp. the conditional distribution of $\mathbf{U_2}$ given $\mathbf{T_2}$. Let $(\Omega^\dagger, \mathcal{F}^\dagger, P^\dagger)$ be a probability space on which there exists a vector $\mathbf{U_1'}$ distributed as $\mathbf{U_1}$. By enlarging $(\Omega^\dagger, \mathcal{F}^\dagger, P^\dagger)$ if necessary, there exist random vectors $\mathbf{T_1'}$ and $\mathbf{U_2'}$ on this probability space such that the conditional distribution of $\mathbf{T_1'}$ given $\mathbf{U_1'}$ equals $\mu_{\mathbf{T_1}|\mathbf{U_1}}$ and the conditional distribution of $\mathbf{U_2'}$ given $\mathbf{U_1'}$ equals $\mu_{\mathbf{U_2}|\mathbf{T_2}}$. Then by $\mathbf{U_1} \overset{\mathcal{D}}{=} \mathbf{T_2}$ we have $(\mathbf{T_1'}, \mathbf{U_1'}) \overset{\mathcal{D}}{=} (\mathbf{T_1}, \mathbf{U_1})$ and $(\mathbf{U_1'}, \mathbf{U_2'}) \overset{\mathcal{D}}{=} (\mathbf{T_2}, \mathbf{U_2})$, so that for the components we have $T'_{1,n} - U'_{1,n} \to 0$ a.s. and $U'_{1,n} - U'_{2,n} \to 0$ a.s., so that $T'_{1,n} - U'_{2,n} \to 0$ a.s. $\diamondsuit$

Lemma 4.2 *Let* $X \in \mathcal{L}^p$*,* $2 < p < \alpha$*. Then there exists a constant* $c = c_{\alpha,p}$ *such that*

$$\sum_{i=1}^{\infty} 3^i P(|X| \geq 3^{i/p}) + \sum_{i=1}^{\infty} 3^i E \min(|X/3^{i/p}|^\alpha, |X/3^{i/p}|^2) \leq cE(|X|^p). \tag{4.96}$$

Proof. That the first sum is finite follows from

$$\sum_{i=1}^{\infty} 3^i \mathsf{P}(|X| \geq 3^{i/p}) \leq 3 \sum_{i=1}^{\infty} \int_{3^{i-1}}^{3^i} \mathsf{P}(|X|^p > u) du \leq 3\mathsf{E}(|X|^p). \tag{4.97}$$

For the second one, let $q_i = \mathsf{P}(3^{i-1} \leq |X|^p < 3^i)$. Then

$$\begin{aligned}
\sum_{i=1}^{\infty} 3^i \mathsf{E}(|X/3^{i/p}|^2 \mathbf{1}_{|X|^p \geq 3^i}) &\leq \sum_{i=1}^{\infty} 3^i \sum_{j=1+i}^{\infty} 3^{(j-i)2/p} q_j \\
&= \sum_{j=2}^{\infty} \sum_{i=1}^{j-1} 3^i 3^{(j-i)2/p} q_j \\
&= c_1 \sum_{j=2}^{\infty} 3^j q_j \leq c_1 \mathsf{E}(|X|^p).
\end{aligned} \tag{4.98}$$

for some constant c_1 only depending on p and α. Similarly, there exists c_2 such that

$$\begin{aligned}
\sum_{i=1}^{\infty} 3^i \mathsf{E}(|X/3^{i/p}|^\alpha \mathbf{1}_{|X|^p < 3^i}) &\leq \sum_{i=1}^{\infty} 3^i \sum_{j=-\infty}^{i} 3^{(j-i)\alpha/p} q_j \\
&= \sum_{j=-\infty}^{\infty} \sum_{i=\max(1,j)}^{\infty} 3^{i(1-\alpha/p)} 3^{j\alpha/p} q_j \leq c_2 \mathsf{E}(|X|^p).
\end{aligned}$$

For the last relation, we consider two cases $\sum_{j=-\infty}^{0}$ and $\sum_{j=1}^{\infty}$ separately. The lemma then follows from (4.97) and (4.98). It is easily seen that (4.96) also holds with the factor 3 therein replaced by any $\theta > 1$. In this case the constant c depends on p, α and θ. $\diamondsuit$

Lemma 4.3 *Recall (2.10) and (2.11) for $\Xi_{\alpha,p}$ and $M_{\alpha,p}$, respectively, and (3.30) for $W_{k,l}$. Then there exists a constant c, only depending of α and p, such that*

$$\sum_{k=1}^{\infty} \frac{3^k}{m_k} \frac{\mathsf{E}(\max_{1\le l\le m_k} |W_{k,l}|^\alpha)}{3^{k\alpha/p}} \le cM_{\alpha,p}\Theta_{0,2}^\alpha + c\Xi_{\alpha,p}^\alpha + c\|X_1\|_p^p. \tag{4.99}$$

Proof of Lemma 4.3. Recall (3.39) for the functional dependence measure $\delta_{k,j,\iota}$. Since T_a has Lipschitz constant 1, we have

$$\begin{aligned}\delta_{k,j,\iota}^\iota &\le \mathsf{E}[\min(2\times 3^{k/p},\, |X_i - X_{i,\{i-j\}}|)^\iota] \\ &\le 2^\iota \mathsf{E}[\min(3^{k/p},\, |X_j - X_{j,\{0\}}|)^\iota].\end{aligned} \tag{4.100}$$

We shall apply the Rosenthal-type inequality in Liu, Xiao and Wu (2011): there exists a constant c, only depending on α, such that

$$\begin{aligned}\left\| \max_{1\le l\le m_k} |W_{k,l}| \right\|_\alpha &\le cm_k^{1/2}\left[\sum_{j=1}^{m_k}\delta_{k,j,2} + \sum_{j=1+m_k}^{\infty}\delta_{k,j,\alpha} + \|T_{3^{k/p}}(X_1)\|_2\right] \\ &\quad + cm_k^{1/\alpha}\left[\sum_{j=1}^{m_k} j^{1/2-1/\alpha}\delta_{k,j,\alpha} + \|T_{3^{k/p}}(X_1)\|_\alpha\right] \\ &\le c(I_k + II_k + III_k),\end{aligned} \tag{4.101}$$

where

$$\begin{aligned}I_k &= m_k^{1/2}\sum_{j=1}^{\infty}\delta_{j,2} + m_k^{1/2}\|X_1\|_2, \\ II_k &= m_k^{1/\alpha}\sum_{j=1}^{\infty} j^{1/2-1/\alpha}\delta_{k,j,\alpha}, \\ III_k &= m_k^{1/\alpha}\|T_{3^{k/p}}(X_1)\|_\alpha.\end{aligned} \tag{4.102}$$

Here we have applied the inequality $\delta_{k,j,2} \le \delta_{j,2}$, since T_a has Lipschitz constant 1. Since $\sum_{j=1}^{\infty}\delta_{j,2} + \|X_1\|_2 \le 2\Theta_{0,2}$, by (2.11), we obtain the upper bound $cM_{\alpha,p}\Theta_{0,2}^\alpha$ in (4.99), which corresponds to the first term I_k in (4.101). For the third term III_k, we obtain the bound $c\|X_1\|_p^p$ in (4.101) in view of Lemma 4.2 by noting that $|T_{3^{k/p}}(X_1)| \le \min(3^{k/p},\, |X_1|)$ and $\min(|v|^\alpha, v^2) \ge \min(|v|^\alpha, 1)$.

We shall now deal with II_k. Let $\beta = \alpha/(\alpha-1)$, so that $\beta^{-1} + \alpha^{-1} = 1$; let $\lambda_j = (j^{1/2-1/\alpha}\delta_{j,p}^{p/\alpha})^{-1/\beta}$. Recall (2.10) for $\Xi_{\alpha,p}$. By Hölder's inequality,

$$\left(\sum_{j=1}^{\infty} j^{1/2-1/\alpha}\delta_{k,j,\alpha}\right)^\alpha \le \Xi_{\alpha,p}^{\alpha/\beta}\sum_{j=1}^{\infty}\lambda_j^\alpha (j^{1/2-1/\alpha}\delta_{k,j,\alpha})^\alpha. \tag{4.103}$$

Hence, by (4.100) and Lemma 4.2, we complete the proof of (4.99) in view of

$$\begin{aligned} \sum_{k=1}^{\infty} \frac{3^k}{m_k} \frac{II_k^\alpha}{3^{\alpha k/p}} &\le \sum_{k=1}^{\infty} 3^{k-k\alpha/p} \Xi_{\alpha,p}^{\alpha/\beta} \sum_{j=1}^{\infty} \lambda_j^\alpha (j^{1/2-1/\alpha}\delta_{k,j,\alpha})^\alpha \\ &= \Xi_{\alpha,p}^{\alpha/\beta} \sum_{j=1}^{\infty} \lambda_j^\alpha j^{\alpha/2-1} \sum_{k=1}^{\infty} 3^{k-k\alpha/p} \delta_{k,j,\alpha}^\alpha \\ &\le \Xi_{\alpha,p}^{\alpha/\beta} \sum_{j=1}^{\infty} \lambda_j^\alpha j^{\alpha/2-1} c_{\alpha,p} \delta_{j,p}^p = c_{\alpha,p} \Xi_{\alpha,p}^\alpha. \end{aligned} \tag{4.104}$$

◇

References

Bendat, J. S. (1990) *Nonlinear System Analysis and Identification from Random Data,* Wiley, New York.

Berkes, I. and Philipp, W. (1979) Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7**, no. 1, 29–54.

Billingsley, P. (1968) *Convergence of probability measures.* Wiley, New York.

Burton, R. and Denker, M. (1987) On the central limit theorem for dynamical systems, *Transactions Amer. Math. Soc.* **302** 715–726.

Bradley, R. C. (2007) *Introduction to strong mixing conditions.* Kendrick Press, Heber City, UT.

Casti, J. L. (1985) *Nonlinear System Theory* Academic, Orlando, FL.

Chatterjee, S. (2012) A new approach to strong embeddings. *Probability Theory Related Fields* **152**, 231–264.

Chow, Y. S. and Teicher, H. (1988) *Probability Theory*, 2nd ed. Springer, New York.

Csörgő, M. and Révész, P. (1974/75) A new method to prove Strassen type laws of invariance principle. I, II. *Z. Wahrsch. Verw. Gebiete* **31**, 255–259; ibid. **31** (1974/75), 261–269.

Csörgő, M. and Révész, P. (1981) *Strong approximations in probability and statistics.* Academic Press, New York.

Csörgő, S. and Hall, P. (1984) The Komlós-Major-Tusnády approximations and their applications. *Austral. J. Statist.* **26**, 189–218.

Dedecker, J. and Merlevède, F. (2003). The conditional central limit theorem in Hilbert spaces. *Stochastic Process. Appl.* **108** 229–262.

Dedecker, J., Doukhan, P., Lang, G., León R., José R., Louhichi, S. and Prieur, C. (2007) *Weak Dependence: With Examples and Applications*, Springer, New York.

Denker, M. (1989) The central limit theorem for dynamical systems. In *Dynamical systems and ergodic theory (Warsaw, 1986)*, 33–62, Banach Center Publ., 23, PWN, Warsaw, 1989.

Denker, M. and Keller, G. (1986) Rigorous statistical procedures for data from dynamical systems. *J. Statist. Phys.* **44** 67–93.

Denker, M. and Philipp, W. (1984) Approximation by Brownian motion for Gibbs measures and flows under a function. *Ergodic Theory Dynam. Systems* **4** 541–552.

Diaconis, P. and Freedman. D. (1999) Iterated random functions. *SIAM Rev.* **41** 45–76.

Donsker, M. D. (1952) Justification and extension of Doob's heuristic approach to the Komogorov-Smirnov theorems. *Ann. Math. Statistics* **23**, 277–281.

Doob, J. L. (1949) Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Statistics* **20**, 393–403.

Doukhan, P. (2003). Models, inequalities, and limit theorems for stationary sequences. In *Theory and applications of long-range dependence* (P. Doukhan, G. Oppenheim and M. S. Taqqu, eds.) 43?-100. Birkhäuser, Boston, MA.

Eberlein, E. (1986). On strong invariance principles under dependence assumptions. *Ann. Probab.* **14** 260–270.

Eberlein, E. and Taqqu, M. S. (eds.) (1986). *Dependence in probability and statistics: A survey of recent results.* Birkhauser, Boston.

Erdős, P. and Kac, M. (1946) On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.* **52** 292–302.

Eubank, R. L. and Speckman, P. L. (1993) Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88** 1287–1301.

Gordin, M. I. (1969). The central limit theorem for stationary processes. *Dokl. Akad. Nauk SSSR*, **188** 739-741.

Götze F. and Zaitsev, A. Yu. (2009) Bounds for the rate of strong approximation in the multidimensional invariance principle. *Theory Probab. Appl.* **53** 59–80

Hannan, E. J. (1979). The central limit theorem for time series regression. *Stochastic Process. Appl.* **9** 281–289.

Hofbauer, F. and Keller, G. (1982) Ergodic properties of invariant measures for piecewise monotonic transformations, *Math. Z.* **180** 119–140

Ibragimov, I. A. and Linnik, Yu. V. (1971) Independent and stationary sequences of random variables. Groningen, Wolters-Noordhoff.

KOMLÓS, J., MAJOR, P. and TUSNÁDY, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131; II. (1976) **34** 33–58.

LIU, WEIDONG and LIN, ZHENGYAN (2009) Strong approximation for a class of stationary processes. *Stochastic Process. Appl.* **119** 249–280.

LIU, WEIDONG, HAN XIAO and WU, WEI BIAO (2011) Probability Inequalities Under Dependence. *Draft.*

MERLEVÈDE, F. and RIO, E. (2012) Strong approximation of partial sums under dependence conditions with application to dynamical systems. *Stochastic Process. Appl.* **122** 386–417

ORNSTEIN, D.S. (1974) Ergodic theory, randomness, and dynamical systems" , Yale Univ. Press.

PELIGRAD, M. and UTEV, S. (2005). A new maximal inequality and invariance principle for stationary sequences. *Ann. Probab.* **33** 798–815.

PHILIPP, W. (1986) Invariance principles for independent and weakly dependent random variables. In *Dependence in probability and statistics: A Survey of Recent Results.* (E. Eberlein and M. S. Taqqu, eds.), 225–268, Birkhäuser Boston, MA.

PRIESTLEY, M. B. (1988). *Nonlinear and Nonstationary Time Series Analysis.* Academic Press, London.

PROHOROV, YU. V. (1956) Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.* **1** 157–214

RIO, E. (1995). The functional law of the iterated logarithm for stationary strongly mixing sequences. *Ann. Probab.* **23** 1188–1203.

RUGH, W. J. (1981) *Nonlinear System Theory: The Volterra-Wiener Approach* Johns Hopkins Univ. Press, Baltimore.

SAKHANENKO, A. I. (1991) On the accuracy of normal approximation in the invariance principle. *Siberian Advances in Mathematics.* **1** 58–91.

SAKHANENKO, A. I. (2006) Estimates in the invariance principle in terms of truncated power moments. *Siberian Math. J.* **47** 1113–1127.

SCHETZEN, M. (1980) *The Volterra and Wiener Theories of Nonlinear Systems* Wiley, New York.

SHAO, Q. M. (1993). Almost sure invariance principles for mixing sequences of random variables. *Stochastic Process. Appl.* **48** 319–334.

SHAO, Q. M. (1995). Strong approximation theorems for independent random variables and their applications, *J. Multivariate Anal.,* **52** 107–130.

Shao, X. and Wu, W. B. (2007) Asymptotic spectral theory for nonlinear time series. Annals of Statistics, Vol. 35, No. 4, 1773–1801.

Shields, P. (1973) *The theory of Bernoulli shifts,* Univ. Chicago Press.

Shorack, G. R. and Wellner, J. A. (1986) *Empirical processes with applications to statistics.* Wiley, New York.

Stout, W. F. (1974) *Almost sure convergence.* New York: Academic Press.

Strassen, V. (1964) An invariance principle for the law of the iterated logarithm. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **3**, 211–226.

Strassen, V. (1967) Almost sure behaviour of sums of independent random variables and martingales. *Proceedings of the 5th Berkeley Symposium of Mathematical Statistics and Probability*, **2**, 315-343. Univ. of California Press, Berkeley.

Tong, H. (1990) *Nonlinear time series. A dynamical system approach.* Oxford University Press.

Volný, D. (1993) Approximating martingales and the central limit theorem for strictly stationary processes. *Stoc. Proc. Appl.* **44** 41–74.

Volný, D. (1999) Invariance principles and Gaussian approximation for strictly stationary processes. *Trans. Amer. Math. Soc.* **351** 3351–3371.

Weiss, B. (1974) The Structure of Bernoulli Systems. *Proceedings of the International Congress of Mathematicians.* Vancouver.

Wiener, N. (1958) *Nonlinear Problems in Random Theory.* MIT Press, Cambridge, MA.

Woodroofe, M. (1992) A central limit theorem for functions of a Markov chain with applications to shifts. *Stoc. Proc. Appl.*, **41** 33–44.

Wu, C. O., Chiang, C. T. and Hoover, D. R. (1998) Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.* **93** 1388–1402.

Wu, W. B. (2005) Nonlinear system theory: another look at dependence. *Proc. Nat. Acad. Sci. USA* **102** no. 40, 14150–14154.

Wu, W. B. and Shao, X. (2004) Limit theorems for iterated random functions. *J. Appl. Probab.* **41** 425–436.

Wu, W. B. (2007) Strong invariance principles for dependent random variables. *Annals of Probability*, 35, 2294–2320.

Wu, W. B. (2011) Asymptotic theory for stationary processes. *Statistics and Its Interface* **4** 207–226.

Wu, W. B. and Zhao, Z. (2007) Inference of Trends in Time Series. *Journal of the Royal Statistical Society, Series B,* **69**, 391–410.