MARCINKIEWICZ LAW OF LARGE NUMBERS FOR OUTER-PRODUCTS OF HEAVY-TAILED, LONG-RANGE DEPENDENT DATA

By Michael A. Kouritzin and Samira Sadeghi*

University of Alberta

The Marcinkiewicz Strong Law, $\lim_{n\to\infty} \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n}(D_k - D) = 0$ a.s. with $p \in (1,2)$, is studied for outer products $D_k = X_k \overline{X}_k^T$, where $\{X_k\}, \{\overline{X}_k\}$ are both two-sided (multivariate) linear processes (with coefficient matrices $(C_l), (\overline{C}_l)$ and i.i.d. zero-mean innovations $\{\Xi\}$, $\{\overline{\Xi}\}$). Matrix sequences C_l and $\overline{C}_l$ can decay slowly enough (as $|l| \to \infty$) that $\{X_k, \overline{X}_k\}$ have long-range dependence while $\{D_k\}$ can have heavy tails. In particular, the heavy-tail and long-range-dependence phenomena for $\{D_k\}$ are handled simultaneously and a new decoupling property is proved that shows the convergence rate is determined by the worst of the heavy-tails or the long-range dependence, but not the combination. The main result is applied to obtain Marcinkiewicz Strong Law of Large Numbers for stochastic approximation, non-linear functions forms and autocovariances.

1. Intoduction. Let $D_k = X_k \overline{X}_k^T$ be random matrices with $\{X_k\}$, $\{\overline{X}_k\}$ being $\mathbb{R}^d$-valued (possibly two-sided, multivariate) linear processes

$$(1) \qquad X_k = \sum_{l=-\infty}^{\infty} C_{k-l}\Xi_l, \quad \overline{X}_k = \sum_{l=-\infty}^{\infty} \overline{C}_{k-l}\overline{\Xi}_l.$$

defined on some probability space (Ω, F, P).

$$\left\{ \left(\Xi_l = (\xi_l^{(1)}, ..., \xi_l^{(m)}), \overline{\Xi}_l = (\overline{\xi}_l^{(1)}, ..., \overline{\xi}_l^{(m)}) \right), \ l \in \mathbb{Z} \right\}$$

are i.i.d. zero-mean random $\mathbb{R}^{m+m}$-vectors (innovations) such that $E[|\Xi_1|^2] < \infty$, $E[|\overline{\Xi}_1|^2] < \infty$ and $(C_l)_{l\in\mathbb{Z}}, (\overline{C}_l)_{l\in\mathbb{Z}}$ are $\mathbb{R}^{d\times m}$-matrix sequences satisfying $\sup_{l\in\mathbb{Z}} |l|^{\sigma}\|C_l\| < \infty, \sup_{l\in\mathbb{Z}} |l|^{\overline{\sigma}}\|\overline{C}_l\| < \infty$ for some $(\sigma, \overline{\sigma}) \in \left(\frac{1}{2}, 1\right]$. Hence, $\{D_k\}$ can have heavy tails as well as long-range dependence.

---

*Corresponding author. E-mail: ssadeghi@ualberta.ca

1

Linear process models are heavily used in finance, engineering, econometrics, and statistics. In fact, classical time-series theory mainly involves the statistical analysis of stationary linear processes. Current applications in network theory and financial mathematics leads us to study time series models where $\{D_k\}$ can have heavy tails and long memory. Heavy-tailed data exhibits frequent extremes and infinite variance, while positively-correlated long memory data displays great serial momentum or inertia. Heavy-tailed data with long-range dependence has been observed in a plethora of empirical data set over the last fifty years and so. For instance, Mandelbrot [11] observed that long memory time series often were heavy-tailed and self-similar.

The possible rates of the convergence is affected by both long-range dependence and heavy-tailed. There are two broad types of dependence for linear processes. If the coefficients (C_l) are absolutely summable and innovations have second moments, then the covariances of X_k are summable and we say that $\{X_k\}$ is short-range dependence (SRD). On the contrary, we generically say that $\{X_k\}$ is long-range dependence (LRD) if its covariances are not absolutely summable. Practically, by choosing appropriate coefficients, matrix sequence (C_l) can decay slowly enough (as $|l| \to \infty$) such that $\{X_k\}$ shows LRD. We consider $\{D_k\}$ to have LRD too in this $\{C_l\}$ non-summable case even though the second moments for D_k may not exist. There are also two general kinds of randomness. If each D_k fails to have a second moment, then we say it has heavy-tailed (HT) and is otherwise light-tailed (LT). In our setting, D_k will either have HT or LT depending upon the moments of and dependence between Ξ_1 and $\overline{\Xi}_1$.

There few general Marcinkiewicz Strong Law of Large Numbers (MSLLN) results for partial sums of X_k under both heavy-tailed and the long-range dependence and the MSLLN for partial sums of nonlinear functions of X_k is almost untouched. Our purpose here is to establish a method and a structure under which certain MSLLN for heavy-tailed and the long-range-dependent phenomena can be handled properly. Technically, our goal is to prove:

$$\lim_{n \to \infty} \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n} (D_k - D) = 0 \quad \text{a.s.} \quad \text{for} \quad p < \frac{1}{2 - \sigma - \overline{\sigma}} \wedge \alpha \wedge 2,$$

when $\max\limits_{1 \leq i,j \leq m} \sup\limits_{t \geq 0} t^{\alpha} P(|\xi_1^{(i)} \overline{\xi}_1^{(j)}| > t) < \infty$ for some $\alpha > 1$ and $\sup\limits_{l \in \mathbb{Z}} |l|^{\sigma} \|C_l\| < \infty$, $\sup\limits_{l \in \mathbb{Z}} |l|^{\overline{\sigma}} \|\overline{C}_l\| < \infty$ when $(\sigma, \overline{\sigma}) \in \left(\frac{1}{2}, 1\right]$. This format of $\{D_k\}$ is critical for our result since, it allows LRD and HT conditions decouple and convergence rate be determined by the worst of the HT requirement $p < (\alpha \wedge 2)$

and the LRD condition $p < \frac{1}{2-\sigma-\overline{\sigma}}$, but not the combination. A bifurcation happens. Consider the summation, $D_k = \sum\limits_{l,m=-\infty}^{\infty} C_{k-l}\Xi_l\overline{C}_{k-m}\overline{\Xi}_m$, broken into off-diagonal and diagonal terms. Due to the independence of $(\Xi_l, \overline{\Xi}_l)$ from $(\Xi_m, \overline{\Xi}_m)$, the off-diagonal sum $\sum\limits_{l \neq m} C_{k-l}\overline{C}_{k-m}\Xi_l\overline{\Xi}_m$ does not have heavy tails (when $\alpha > 1$). Conversely, since $\sigma + \overline{\sigma} > 1$ the diagonal sum $\sum\limits_{l=-\infty}^{\infty} C_{k-l}\overline{C}_{k-l}\Xi_l\overline{\Xi}_l$ does not experience long-range dependence. In addition, the rate of convergence depends on the worst of $(\alpha \wedge 2)$ and $\frac{1}{2-\sigma-\overline{\sigma}}$, so whenever we are in the LRD dominant case, $(\alpha > \frac{1}{2-\sigma-\overline{\sigma}})$, the off-diagonal terms dictate the rate of convergence by the LRD effect $(p < \frac{1}{2-\sigma-\overline{\sigma}})$ and in the HT dominant case, $(\alpha < \frac{1}{2-\sigma-\overline{\sigma}})$, the diagonal terms dictate the rate of convergence by HT effect $(p < \alpha)$. The bifurcation point is when $\alpha = \frac{1}{2-\sigma-\overline{\sigma}}$ and $\alpha < 2$.

2. Background. In this section we give a review of some existing literature on MSLLN or weak convergence for partial sums, sample covariance and non-linear function of partial sums with heavy-tailed and/or long-range dependence. Many existing results were only established in the scalar case. For ease of assimilation we use $\{x_k\}$, (c_l), $\{d_k\}$ and $\{\xi_k\}$ to denote these scalar versions of $\{X_k\}$, (C_l), $\{D_k\}$ and $\{\Xi_k\}$ and $\{x_{k+h}\}$ for $\{\overline{X}_k\}$ when it is a shifted version of $\{x_k\}$.

2.1. *Partial Sums.* There are many of publications that consider almost sure rates of convergence for linear processes having either LRD or HT. However, there are only a few like Louhchi and Soulier [10] that considered the combination of these two phenomena. They stated the following result for linear symmetric α-stable (SαS) processes.

Theorem 1 *Let $\{\xi_j\}_{j\in\mathbb{Z}}$ be i.i.d. sequence of SαS random variables with $1 < \alpha < 2$ and $\{c_j\}_{j\in\mathbb{Z}}$ be a bounded collection such that $\sum\limits_{j\in\mathbb{Z}} |c_j|^s < \infty$ for some $s \in [1, \alpha)$. Set $x_k = \sum\limits_{j\in\mathbb{Z}} c_{k-j}\xi_j$. Then, for $p \in (1,2)$ satisfying $\frac{1}{p} > 1 - \frac{1}{s} + \frac{1}{\alpha}$*

$$\frac{1}{n^{\frac{1}{p}}}\sum_{i=1}^{n} x_i \to 0 \ a.s.$$

The condition $s < \alpha$ ensures $\sum\limits_{j\in\mathbb{Z}} |c_j|^\alpha < \infty$ and thereby convergence of $\sum\limits_{j\in\mathbb{Z}} c_{k-j}\xi_j$. Moreover, $\{x_k\}$ not only exhibits heavy tails but also long-range

dependence if, for example, $c_j = |j|^{-\sigma}$ for $j \neq 0$ and some $\sigma \in \left(\frac{1}{2}, 1\right)$. Notice there is interactions between the heavy tail condition and the long range dependent condition. In particular for a given p, heavier tails (α becomes smaller) implies that you can not have as long range dependence (s must become smaller) and vice versa. Moreover, this result is difficult or even impossible to apply in our outer product setting due to the fact that x_k's are linear processes with SαS innovations and so x_k cannot be decomposed to product of two variables even in the scalar case.

2.2. *Non-linear function of partial sums.* The limit behavior of suitably normalized partial sums of stationary random variables that demonstrate either LRD or HT has been subject of study by many authors. Applications can be found in geophysics, economics, hydrology and statistics. For instance, in contexts like Whittle approximation, the asymptotic behavior of quadratic forms of stationary sequences have an important role. In addition, the efficacy of "R/S-statistic" theory that was introduced for estimating the long-run, non-periodic statistical dependence of time series by Hurst and developed by Mandelbrot [12], can be confirmed by convergence of these limit functions.

There are many results that deal with the existence and description of limit distributions of sums

$$(2) \qquad S_{n,h}(t) = \sum_{k=1}^{[nt]} (h(x_k) - E(h(x_k))), \quad t \geq 0,$$

where h is a (nonlinear) function. The limit behavior for a Gaussian LRD process $\{x_k\}$, firstly was studied by Rosenblatt [14]. Afterward, Dobrushin and Major [4] explained it in more general form. Then Taqqu [18] showed that the limit in distribution of particular normalized sums $S_{n,h}(t)$ is determined by the Hermite rank $m^* \in \{1, 2, ...\}$ of $h(x)$, which is the index of the first nonzero coefficient in the Hermite expansion. On the other hand, the behavior of nonlinear non-Gaussian LRD processes is much less commonly known. One of the most studied models of non-Gaussian LRD processes is the one-sided linear (moving average) process,

$$(3) \qquad x_k = \sum_{j=0}^{\infty} c_j \xi_{k-j},$$

in which, innovations $\xi_k, k \in \mathbb{Z}$, are independent and identically distributed (i.i.d.), have zero mean with finite variance, and coefficients c_j satisfy:

$$(4) \qquad c_j \sim c_\sigma j^{-\sigma}, \quad j \geq 1$$

for some constant $c_\sigma \neq 0$, $c_0 = 1$ and $\sigma \in (\frac{1}{2}, 1)$.

Surgailis [16] considered the limit behavior of partial sum processes $S_{n,h}(t)$ of polynomial h of linear process $\{x_k\}_{k\in\mathbb{Z}}$. Later, Giraitis and Surgailis [5][6], Avram and Taqqu [1] noticed that the only difference between this case and Gaussian case is that the Hermite rank m^* of $h(x)$ has to be replaced by the Appell rank m.

Vaiciulis [19] investigated distributional convergence for normalized partial sums of Appell polynomials $A_m(x_k)$ of linear processes x_k having both long-memory and heavy-tails in the sense $EA_m^2(x_k) = \infty$. In particular, he assumed x_k had the form (3) with innovations $\{\xi_k^m\}$ belonging to the domain of attraction of an α-stable law with $1 < \alpha < 2$ and c_j following (4). The limit was: **i)** an α-stable Levy process, **ii)** an mth order Hermite process, or **iii)** the sum of two mutually independent α-stable Levy and m^{th} order Hermite processes, depending on the value of α, m and σ where $\sigma \in (\frac{1}{2}, 1)$.

Thereafter, Surgailis [17] considered the bounded, infinitely differentiable h case where $\{x_k\}$ was LRD and had innovations with probability tail decay of $x^{-2\alpha}$ for $1 < \alpha < 2$. Suppose x_k satisfies (3) and (4). Then he showed three different limiting behaviors corresponding to three different LRD-HT setting: $n^{1-(2\sigma-1)m^*/2}S_{n,h}(t)$, $n^{\frac{1}{2\alpha\sigma}}S_{n,h}(t)$ or $n^{\frac{1}{2}}S_{n,h}(t)$ converge in distribution to respectively a Hermite process of order m^*, a $2\alpha\sigma$-stable Levy process or a Brownian motion, all at time t, for certain range of α and σ.

2.3. *Sample Covariances.* Auto-covariance functions play a substantial role in time series analysis and have diverse applications in inference problems, including hypothesis testing and parameter estimation. The natural estimator of auto-covariance is sample covariance. Hence, the convergence properties of the sample covariance is of great interest. In the case of LRD and HT, it is an area of active research.

Davis and Resnick [3] studied the distributional convergence of sample autocovariances for two-sided linear processes with innovations that were i.i.d. and had regularly varying tail probabilities of index $\alpha > 0$.

$$P(|\xi_k| > x) = x^{-2\alpha}L(x),$$

(5) $\qquad \dfrac{P(\xi_k > x)}{P(|\xi_k| > x)} \to p \quad \text{and} \quad \dfrac{P(\xi_k < -x)}{P(|\xi_k| > x)} \to q, \quad \text{as } x \to \infty,$

where $L(.)$ is a function slowly varying at infinity $\left(\lim\limits_{j\to\infty} \dfrac{L(aj)}{L(j)} = 1\right)$ and $0 \leq p \leq 1$, $q = 1 - p$. They considered the case where the innovations had finite variance (ι) but infinite fourth moment, i.e. $1 < \alpha < 2$ with absolutely summable coefficients c_j with form of (4).

Note: We choose to scale our constants, here and in the sequel, so that $\alpha < 2$ always mean HT of the object of interest, which is $x_k x_{k+h}$ or more generally $X_k \overline{X}_k$.

In case of infinite fourth moment for $\{\xi_k\}_{k \in \mathbb{Z}}$, the asymptotic distribution of normalized sample autocovariances of long-memory processes was studied by Horvath and Kokoszka [7]. Suppose we observe the realization $x_1, x_2, ..., x_{n+v}$, $n > 1, v \geq 0$, the sample autocovariances and population autocovariances are defined as

$$\hat{\gamma}_h^{(n)} = \frac{1}{n} \sum_{k=1}^{n} x_k x_{k+h}, \quad h = 0, 1, ..., v, \quad \text{and}$$

(6)
$$\gamma_h = E[x_0 x_h] = \iota \sum_{j=0}^{\infty} c_j c_{j+h},$$

respectively. Horvath and Kokoszka [7, Theorem 3.1] studied the asymptotic distribution $[\hat{\gamma}_h^{(n)} - \gamma_h]$, $h = 0, 1, ..., v$ for linear process of form (3) with coefficients and innovations satisfying (4) and (5) and a norming constant $a_n = \inf\{x : P(|\xi_1| > x) \leq n^{-1}\}$ (roughly of order $n^{\frac{1}{2\alpha}}$) satisfying

(7)
$$\lim_{n \to \infty} nP[|\xi_k| > a_n x] = x^{-2\alpha}, \ x > 0.$$

We quote this result in our notations as the following theorem.

Theorem 2 *Suppose, conditions (3), (4), (5) and (7) hold.*

  (a) *If $1 - \frac{1}{2\alpha} < \sigma < 1$ and $1 < \alpha < 2$, then*

$$na_n^{-2}[\hat{\gamma}_h^{(n)} - \gamma_h] \overset{d}{\to} \left(S - \frac{\alpha}{\alpha-1} \right) \left[\sum_{j=0}^{\infty} c_j c_{j+h} \right], \quad h = 0, 1, ..., H.$$

  where S is an α-stable random variable. For the above to hold for $\sigma = 3/4$, we must additionally assume that $a_n^{-4} n \ln n \to 0$.

  (b) *If $\frac{1}{2} < \sigma < 1 - \frac{1}{2\alpha}$ and $1 < \alpha < 2$, then*

$$n^{2\sigma-1}[\hat{\gamma}_h^{(n)} - \gamma_h] \overset{d}{\to} \iota c_\sigma^2 \left[U_\sigma(1) \right], \quad h = 0.1, ..., H.$$

  where U_σ is a Rosenblatt process. The Rosenblatt process is often defined by the iterated stochastic integral:

$U_\sigma(t) = 2 \int_{w_1 < w_2 < t} \left[\int_0^t (\tau - w_1)_+^{-\sigma} (\tau - w_2)_+^{-\sigma} d\tau \right] W(dw_1) W(dw_2),$
in which $W(.)$ is the standard Wiener process on the real line.

This theorem works for one-sided linear processes with a regularly varying tail condition and gives us weak convergence.

Notice that in Theorem 2, case **(a)** represents the HT dominant, ($\alpha < \frac{1}{2-2\sigma}$), so the diagonal terms dictate convergence to an α-stable distribution. However, case **(b)** represents the LRD dominant, ($\alpha > \frac{1}{2-2\sigma}$), hence off-diagonal terms take over and we get convergence to Rosenblatt process.

3. Main results. Our first result is in the scalar case. Later, we will extract the full vector-valued result as a second main theorem. All proofs are delayed until the next section after we have discussed the applications.

Theorem 3 *Let $\left\{(\xi_l, \overline{\xi}_l)\right\}_{l \in \mathbb{Z}}$ be i.i.d. zero-mean random variables such that $E[\xi_1^2] < \infty$, $E[\overline{\xi}_1^2] < \infty$ and $\sup\limits_{t \geq 0} t^\alpha P(|\xi_1 \overline{\xi}_1| > t) < \infty$ for some $\alpha > 1$. Moreover, suppose $(c_l)_{l \in \mathbb{Z}}, (\overline{c}_l)_{l \in \mathbb{Z}}$ satisfy*

$$\sup_{l \in \mathbb{Z}} |l|^\sigma |c_l| < \infty, \quad \sup_{l \in \mathbb{Z}} |l|^{\overline{\sigma}} |\overline{c}_l| < \infty \quad for \ some \quad \sigma, \overline{\sigma} \in \left(\frac{1}{2}, 1\right],$$

$d_k = \sum\limits_{l,m=-\infty}^{\infty} c_{k-l} \overline{c}_{k-m} \xi_l \overline{\xi}_m$ and $d = E[\xi_1 \overline{\xi}_1] \sum\limits_{l=-\infty}^{\infty} c_l \overline{c}_l$. Then, for p satisfying $p < \frac{1}{2-\sigma-\overline{\sigma}} \wedge \alpha \wedge 2$

$$\lim_{n \to \infty} \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n} (d_k - d) = 0 \ a.s.$$

Remark 1 *The tail probability bound ensures that $E[|\xi_1 \overline{\xi}_1|^r] < \infty$ for any $r \in (1, (\alpha \wedge 2))$ and $E[d_1]$ exists but it is possible that $E[d_1^2] = \infty$ so we are handling heavy tails for $\{d_k\}$. On the other hand, $E[|\xi_1 \overline{\xi}_1|^\alpha] < \infty$ implies our tail condition by Markov's inequality. $\sigma, \overline{\sigma}$ bound the amount of long-range dependence in $x_k = \sum\limits_{l=-\infty}^{\infty} c_{k-l} \xi_l$, $\overline{x}_k = \sum\limits_{l=-\infty}^{\infty} \overline{c}_{k-l} \overline{\xi}_l$. If σ can be taken larger than 1, then $\sum\limits_{k=1}^{\infty} E[x_0 x_k] < \infty$ and there is no long-range dependence in $\{x_k\}$. $\sigma > \frac{1}{2}$ with $E[\xi_1^2] < \infty$ ensures that $\sum\limits_{l=-\infty}^{\infty} c_{k-l} \xi_l$ converges a.s.*

Remark 2 *Notice that the constraints to handle long-range dependence, $p < \frac{1}{2-\sigma-\overline{\sigma}}$, and to handle the heavy tails, $p < (\alpha \wedge 2)$, decouple. This decoupling appears to be due to the structure of d_k. Due to the independence of $(\xi_l, \overline{\xi}_l)$ from $(\xi_m, \overline{\xi}_m)$, the off-diagonal sum $\sum\limits_{l \neq m} c_{k-l} \overline{c}_{k-m} \xi_l \overline{\xi}_m$ does not have heavy tails. Conversely, since $\sigma + \overline{\sigma} > 1$ the diagonal sum $\sum\limits_{l=-\infty}^{\infty} c_{k-l} \overline{c}_{k-l} \xi_l \overline{\xi}_l$ does not experience long-range dependence.*

We will give a simple example to verify conditions in Theorem 3. Recall, a non-negative random variable ξ obeys a power law with parameters $\beta > 1$ and $x_{\min} > 0$, written $\xi \sim PL(x_{min}, \beta)$, if it has density

$$f(x) = \frac{\beta - 1}{x_{min}} \left(\frac{x}{x_{min}}\right)^{-\beta} \quad \forall \ x \geq x_{min}$$

so $E|\xi|^r = \begin{cases} x_{min}^r \left(\frac{\beta-1}{\beta-1-r}\right) & r < \beta - 1 \\ \infty & r \geq \beta - 1 \end{cases}$.

It has a folded t distribution with parameter $\beta > 1$, written $\xi \sim Ft(\beta)$, if it has density

$$f(x) = \frac{2\Gamma(\frac{\beta}{2})}{\Gamma(\frac{\beta-1}{2})\sqrt{(\beta-1)\pi}} \left(1 + \frac{x^2}{(\beta-1)}\right)^{-\frac{\beta}{2}} \quad \forall \ x > 0$$

so $E(|\xi|^r)$ exists if and only if $r < \beta - 1$.

Example 1 *Suppose $p, q, \alpha, \beta, \overline{\beta} > 1$ are such that $\frac{1}{p} + \frac{1}{q} = 1$, $\beta > p\alpha + 1$, $\overline{\beta} > q\alpha + 1$ and $p\alpha, q\alpha \geq 2$. If ξ_1 and $\overline{\xi}_1$ have power law distribution, lets say $\xi_1 \sim Pl(x_{\min}, \beta)$, $\overline{\xi}_1 \sim Pl(\overline{x}_{\min}, \overline{\beta})$ for some $x_{\min}, \overline{x}_{\min} > 0$, then $E[\xi_1^2]$, $E[\overline{\xi}_1^2] < \infty$ and $\sup_{t \geq 0} t^\alpha P(|\xi_1\overline{\xi}_1| > t) < \infty$. If $\xi_1 \sim Ft(\beta)$, $\overline{\xi}_1 \sim Ft(\overline{\beta})$, then $E[\xi_1^2]$, $E[\overline{\xi}_1^2] < \infty$ and $\sup_{t \geq 0} t^\alpha P(|\xi_1\overline{\xi}_1| > t) < \infty$. Either way, the Theorem 3 applies with properly chosen $(c_l, \overline{c}_l)$.*

We now consider the case where X_k and $\overline{X}_k$ are (multivariate) linear processes.

Theorem 4 *Let $\{\Xi_l\}$ and $\{\overline{\Xi}_l\}$ be i.i.d. zero-mean random $\mathbb{R}^m$-vectors such that $\Xi_l = \left(\xi_l^{(1)}, ..., \xi_l^{(m)}\right)$, $\overline{\Xi}_l = \left(\overline{\xi}_l^{(1)}, ..., \overline{\xi}_l^{(m)}\right)$, $E[|\Xi_1|^2] < \infty$, $E[|\overline{\Xi}_1|^2] < \infty$ and $\max_{1 \leq i,j \leq m} \sup_{t \geq 0} t^\alpha P(|\xi_1^{(i)}\overline{\xi}_1^{(j)}| > t) < \infty$ for some $\alpha > 1$. Moreover, suppose matrix sequences $(C_l)_{l \in \mathbb{Z}}, (\overline{C}_l)_{l \in \mathbb{Z}} \in \mathbb{R}^{d \times m}$ satisfy*

$$\sup_{l \in \mathbb{Z}} |l|^\sigma \|C_l\| < \infty, \ \sup_{l \in \mathbb{Z}} |l|^{\overline{\sigma}} \|\overline{C}_l\| < \infty \quad for \ some \quad (\sigma, \overline{\sigma}) \in \left(\frac{1}{2}, 1\right],$$

X_k, $\overline{X}_k$ take form of (1), $D_k = X_k\overline{X}_k^T$ and $D = E[X_1\overline{X}_1^T]$. Then, for p satisfying $p < \frac{1}{2-\sigma-\overline{\sigma}} \wedge \alpha \wedge 2$

$$\lim_{n \to \infty} \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n} (D_k - D) = 0 \quad a.s.$$

This theorem follows by linearity of limits and Theorem 3.

3.1. *Applications.*   We give some applications of our theorems.

3.1.1. *Stochastic Approximation.*   Stochastic approximation (SA) is often used in optimization problems for linear models. Hence, the convergence properties of SA algorithms driven by linear models is of utmost interest. For illustration, we assume $\{z_k, k = 1, 2, ..\}$ and $\{y_k, k = 2, 3, ...\}$ are respectively $\mathbb{R}^d-$ and $\mathbb{R}-$valued stochastic processes, defined on some probability space (Ω, F, P), that satisfy

(8)
$$y_{k+1} = z_k^T h + \epsilon_k, \qquad \forall k = 1, 2, \ldots,$$

where h is an unknown d-dimensional parameter or weight vector of interest and $\{\epsilon_k\}$ is a noise sequence. We want to estimate the parameter vector h through the stochastic approximation algorithm:

(9)
$$h_{k+1} = h_k + \mu_k(b_k - A_k h_k),$$

where μ_k is the kth step gain of the form $\mu_k = k^{-\chi}$ for some $\chi \in \left(\frac{1}{2}, 1\right]$, $A_k = z_k z_k^T$ and $b_k = y_{k+1} z_k$.
Kouritzin and Sadeghi [9] studied the convergence and almost sure rates of convergence for the algorithm (9). Now, we can combine our main result (Theorem 4) with [9, Corollary 1] to obtain a powerful rate of convergence result for stochastic approximation.

Theorem 5 *Let $\{\Xi_l\}$ be i.i.d. zero-mean random $\mathbb{R}^m$-vectors such that*

$$\sup_{t \geq 0} t^{\alpha} P(|\Xi_1|^2 > t) < \infty \quad \text{for some } \alpha \in (1, 2)$$

$(C_l)_{l \in \mathbb{Z}}$ be $\mathbb{R}^{(d+1) \times m}$-matrices such that $\sup_{l \in \mathbb{Z}} |l|^{\sigma} \|C_l\| < \infty$ for some $\sigma \in \left(\frac{1}{2}, 1\right]$,

$$(z_k^T, y_{k+1})^T = \sum_{l=-\infty}^{\infty} C_{k-l} \Xi_l,$$

$A_k = z_k z_k^T$ and $b_k = y_{k+1} z_k$ and $A = E[z_k z_k^T]$ and $b = E[y_{k+1} z_k]$.
Then, $|h_n - h| = o(n^{-\gamma})$ as $n \to \infty$ a.s. for any $\gamma < \gamma_0^{(\chi)} \doteq (\chi - \frac{1}{\alpha}) \wedge (\chi + 2\sigma - 2)$.

Proof. By Theorem 4 when $\frac{1}{p} = \chi - \gamma$, $\overline{X}_k^T = X_k^T = (z_k^T, y_{k+1})$, $\overline{\Xi}_l = \Xi_l$, $\overline{C}_l = C_l$, $\overline{\sigma} = \sigma$, and $D_k = \begin{pmatrix} z_k z_k^T & y_{k+1} z_k \\ y_{k+1} z_k^T & y_{k+1}^2 \end{pmatrix}$,

$$\frac{1}{n^{\chi - \gamma}} \sum_{k=1}^{n} (D_k - D) \to 0 \quad a.s.,$$

where $D = \begin{pmatrix} A & b \\ b^T & E[y_{k+1}^2] \end{pmatrix}$. The first d-rows of $\frac{1}{n^{\chi-\gamma}} \sum\limits_{k=1}^{n} (D_k - D) \to$ 0 a.s. then establish the MSLLN

$$\frac{1}{n^{\chi-\gamma}} \sum_{k=1}^{n}(A_k - A) \to 0 \quad \text{and} \quad \frac{1}{n^{\chi-\gamma}} \sum_{k=1}^{n}(b_k - b) \to 0 \quad a.s.$$

Now, we apply [9, Corollary 1] to complete the proof. $\square$

Remark 3 *Note that $\chi - \gamma$ satisfies the required conditions $\chi - \gamma > 2 - 2\sigma$ and $\chi - \gamma > \frac{1}{\alpha}$ in Theorem 4. Theorem 5 also appears in [9, Theorem 2].*

3.1.2. *Non-linear Function of linear processes.* As mentioned in Background, Vaiciulis [19] showed the convergence of distributions of the partial sum processes with non-linear $h(x_k)$ in terms of convergence of Appell polynomials $A_m(x_k)$ of a long-memory moving average process $\{x_k\}$ with i.i.d. innovations $\{\xi_k\}$ in the case where the variance $EA_m^2(x_k) = \infty$, and the distribution of ξ_1^m belongs to the domain of attraction of an α-stable law with $1 < \alpha < 2$.

Practically, the simplest examples of functions $h(x)$ with a given Appell rank m are Appell polynomials $h = A_m$ relative to the marginal distribution x_1 of the linear process (3). In case $m = 2$ the Appell polynomial is $A_2(x) = x^2 - \mu_2$ where $\mu_2 = Ex^2$. Viaiciulis [19, Theorems 1.1 and 1.2] proved that when $m(2\sigma - 1) < 1$, $m \geq 2$ and $\sigma \in (\frac{1}{2}, 1)$ the limit distribution of partial sums of mth Appell polynomial is either **(i)** an α-stable Levy process for $2 - 2\sigma < 1 + \frac{2}{m}(\frac{1}{\alpha} - 1)$, or **(ii)** an mth order Hermite process for $2 - 2\sigma > 1 + \frac{2}{m}(\frac{1}{\alpha} - 1)$ or **(iii)** the sum of two mutually independent processes depending on the value of α, m and σ, for $2 - 2\sigma = 1 + \frac{2}{m}(\frac{1}{\alpha} - 1)$.

Taking into account all his conditions (when $t = 1$) and transforming it to our case we write our complementary almost sure rate-of-convergence theorem.

Theorem 6 *Suppose A_2 represents the Appell polynomials with rank 2 relative to the marginal distribution x_1 of the linear process $x_k = \sum\limits_{j=0}^{\infty} c_{k-j}\xi_j$, for $p \in [1, \frac{1}{2-2\sigma} \wedge \alpha)$ when*

$$(10) \qquad \sup_{t \geq 0} t^\alpha P(\xi_1^2 > t) < \infty \quad for~some \quad \alpha \in (1, 2),$$

$$(11) \qquad \sup_{l \in \mathbb{Z}} |l|^\sigma |c_l| < \infty \quad for~some \quad \sigma \in \left(\frac{1}{2}, 1\right].$$

Then,

$$\lim_{n\to\infty} \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n} A_2(x_k) = 0 \ \ a.s.$$

One might wonder if we have obtained the best possible MSLLN. Indeed, we have. For example when $m = 2$, Viaiciulis [19] shows convergence in distribution of $\dfrac{1}{n^{(2-2\sigma)\wedge\frac{1}{\alpha}}} \sum_{k=1}^{n} A_2(x_k)$ to different non-trivial limits in cases $(2-2\sigma) > \frac{1}{\alpha}$ (LRD dominant) or $(2-2\sigma) < \frac{1}{\alpha}$ (HT dominant), respectively. Therefore, $\dfrac{1}{n^{(2-2\sigma)\wedge\frac{1}{\alpha}}} \sum_{k=1}^{n} A_2(x_k)$ cannot converge to zero almost surely. Theorem 6 gives MSLLN for Appell polynomials with rank 2 or in other word gives the convergence and almost sure rates of convergence for partial sums of second Appell polynomial when $\frac{1}{p} > (2-2\sigma) \vee \frac{1}{\alpha}$. Our result is optimal in polynomial sense and we cannot do better than that in terms of MSLLN.

3.1.3. *Autocovariances.* As mentioned in the background, autocovariance estimation under HT and LRD conditions is an active area of research. We will handle the asymptotic behavior of sample covariance function for processes with LRD, innovations of infinite 4th moment and finite variance ι. If we define the sample aurtocovariance and population autocovariance functions by $\hat{\gamma}^{(n)}(h)$ and $\gamma(h)$, as (6), we have following almost sure result.

Theorem 7 *Assume $\hat{\gamma}^{(n)}(h)$ and $\gamma(h)$, as (6) in which $x_k = \sum\limits_{j=0}^{\infty} c_{k-j}\xi_j$ and satisfies (10) and (11) with $E[\xi_1^2] = \iota$. Then for p satisfying $p < \frac{1}{2-2\sigma} \wedge \alpha \wedge 2$*

$$(12) \qquad n^{1-\frac{1}{p}} [\hat{\gamma}_h^{(n)} - \gamma_h] \to 0 \ \ a.s.$$

Proof. Note that in Theorem 3, for case $\overline{\xi}_l = \xi_l$, $E[\xi_1^2] = \iota$, $\overline{c}_l = c_{l+h}$ and $\{c_l = 0, \forall l < 0\}$ we have

$$d_k = \sum_{l=-\infty}^{k} \sum_{m=-\infty}^{k+h} c_{k-l}c_{k+h-m}\xi_l\xi_m \quad \text{and} \quad d = \iota \sum_{l=0}^{\infty} c_l c_{l+h}.$$

Hence,

$$(13) \quad \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n} (d_k - d) = \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n} \left(\sum_{l=-\infty}^{k} \sum_{m=-\infty}^{k+h} c_{k-l}c_{k+h-m}\xi_l\xi_m - \iota \sum_{l=0}^{\infty} c_l c_{l+h} \right).$$

On the other hand, (12) can be written as

$$n^{1-\frac{1}{p}}[\hat{\gamma}_h^{(n)} - \gamma_h] = \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n}(x_k x_{k+h} - Ex_0 x_h)$$

$$(14) \qquad = \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n}\left(\sum_{l=-\infty}^{k}\sum_{m=-\infty}^{k+h} c_{k-l}c_{k+h-m}\xi_l\xi_m - \iota\sum_{l=0}^{\infty}c_l c_{l+h}\right).$$

So, the result follows.□

As we saw, Theorem 2 gives the convergence to the following non trivial limits for $\frac{2\alpha-1}{2\alpha} < \sigma < 1$ and $\frac{1}{2} < \sigma < \frac{2\alpha-1}{2\alpha}$ when $1 < \alpha < 2$,

(a) $\qquad \frac{1}{a_n^2}\sum_{k=1}^{n}(x_k x_{k+h} - Ex_0 x_h) \xrightarrow{d} \left(S - \frac{\alpha}{\alpha-1}\right)\left[\sum_{l=0}^{\infty}c_l c_{l+h}\right],$

(b) $\qquad \frac{1}{n^{2-2\sigma}}\sum_{k=1}^{n}(x_k x_{k+h} - Ex_0 x_h) \xrightarrow{d} \iota c_\sigma^2\left[U_\sigma(1)\right],$

respectively, for $h = 0, 1, ..., v$.

It is clear that in the case of HT dominant, $\frac{1}{\alpha} > 2 - 2\sigma$, we have almost sure convergence (Theorem 7) when $\frac{1}{p} > \frac{1}{\alpha}$. When $\frac{1}{p} = \frac{1}{\alpha}$, we get into the case (a) and have convergence to an α-stable distribution. On the other hand, in the LRD dominant case, $\frac{1}{\alpha} < 2 - 2\sigma$, from Theorem 7) we have almost sure convergence for $\frac{1}{p} > 2 - 2\sigma$, yet for $\frac{1}{p} = (2 - 2\sigma)$ we have convergence to Rosenblatt process by (b) .

Hence, Theorem 7 shows the $a.s$ convergence for difference of sample autocovariance and population autocovariance with HT and LRD. One example can be in the case that $h = 0$. Theorem 2 and (15) give the convergence in distribution

$$\frac{1}{a_n^2}\sum_{k=1}^{n}(x_k^2 - Ex_0^2) \xrightarrow{d} (S - \frac{\alpha}{\alpha-1})\sum_{l=0}^{\infty}c_l^2$$

$$\frac{1}{n^{2-2\sigma}}\sum_{k=1}^{n}(x_k^2 - Ex_0^2) \xrightarrow{d} \iota c_\sigma^2 U_\sigma(1),$$

for $\frac{1}{p} = \frac{1}{\alpha}$ and $\frac{1}{p} = 2 - 2\sigma$, respectively.

While, Theorem 7 gives the almost sure convergence for $\frac{1}{n^{\frac{1}{p}}}\sum_{k=1}^{n}\left(x_k^2 - Ex_0^2\right)$

when $\frac{1}{p} > (2 - 2\sigma) \vee \frac{1}{\alpha}$.
When we have convergence in distribution to non-trivial limits we can not

get almost sure convergence to 0. However, by Theorem 7 we can get arbitrary close to that with polynomial rate and get optimal polynomial almost sure rate of convergence. We can not do better than that in terms of MSLLN.

4. Proofs.

4.1. Notation List.

$|x|$ is Euclidean distance of some $\mathbb{R}^d$-vector x.
$\|C\| = \sup_{|x|=1} |Cx|$ for any $\mathbb{R}^{n \times m}$-matrix C.
$\lfloor t \rfloor \doteq \max\{i \in \mathbb{N}_0 : i \leq t\}$ and $\lceil t \rceil \doteq \min\{i \in \mathbb{N}_0 : i \geq t\}$ for any $t \geq 0$.
$a_{i,k} \overset{i}{\ll} b_{i,k}$ means that for each k there is a $c_k > 0$ that does not depend upon i such that $|a_{i,k}| \leq c_k |b_{i,k}|$ for all i,k.
$\prod_{l=p}^{q} B_l$ ($\forall B_l$ being a $R^{d \times d}$-matrix) $= B_q B_{q-1} \cdots B_p$ if $q \geq p$ or I if $p > q$.
$a \vee b = \max\{a,b\}$ and $a \wedge b = \min\{a,b\}$.

4.2. A First Light Tail Result.

We first give a result that only handles long-range dependence without heavy tails. However, our proof of Theorem 3 to follow will show that these two phenomena decouple, so we can easily build upon the Theorem 8 to handle both long-range dependence and heavy tails together.

Theorem 8 *Let $\{(\xi_l, \overline{\xi}_l),\ l \in \mathbb{Z}\}$ be i.i.d. zero-mean random variables such that $E[(1 + \xi_1^2)(1 + \overline{\xi}_1^2)] < \infty$, $(c_l, \overline{c}_l)_{l \in \mathbb{Z}}$ satisfy*

$$\sup_{l \in \mathbb{Z}} |l|^{\sigma} |c_l| < \infty, \quad \sup_{l \in \mathbb{Z}} |l|^{\overline{\sigma}} |\overline{c}_l| < \infty \quad \text{for some} \quad \sigma, \overline{\sigma} \in \left(\frac{1}{2}, 1\right],$$

$x_k = \sum_{l=-\infty}^{\infty} c_{k-l} \xi_l$, $\overline{x}_k = \sum_{l=-\infty}^{\infty} \overline{c}_{k-l} \overline{\xi}_l$, $d_k = x_k \overline{x}_k = \sum_{l,m=-\infty}^{\infty} c_{k-l} \overline{c}_{k-m} \xi_l \overline{\xi}_m$ *and*
$d = E[\xi_1 \overline{\xi}_1] \sum_{l=-\infty}^{\infty} c_{k-l} \overline{c}_{k-l} = E[\xi_1 \overline{\xi}_1] \sum_{l=-\infty}^{\infty} c_l \overline{c}_l$. *Then, for $p < \frac{1}{2-\sigma-\overline{\sigma}}$*

$$\lim_{n \to \infty} \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n} (d_k - d) = 0 \quad a.s.$$

Proof. . Insomuch as the proof of the general case only differs cosmetically from the notationally-simpler case where $\overline{\xi}_l = \xi_l$ and $\overline{c}_l = c_l = \begin{cases} 1 & l = 0 \\ |l|^{-\sigma} & l \neq 0 \end{cases}$, we only provide the proof of the later for which the constraint becomes $p < \frac{1}{2-2\sigma}$. Assume without loss of generality that $\sigma < 1$ and $E[\xi_1^2] = 1$.

Step 1: Divide partial sums into diagonal, large c, small and mixed type terms.

Let $n_r = 2^r$ and $T = T(n) = n^\nu$ for $\nu > 0$, $n \in [n_r, n_{r+1})$ and $r \in \mathbb{N}_0$, and define

$$(15) \qquad S_n^{(1)} = \sum_{k=1}^{n} \sum_{l=-\infty}^{\infty} c_{k-l}^2 \left(\xi_l^2 - 1\right)$$

$$(16) \qquad S_n^{(2)} = \sum_{k=1}^{n} \sum_{\substack{l,m=k-T \\ l \neq m}}^{k+T} c_{k-l} c_{k-m} \xi_l \xi_m$$

$$(17) \qquad S_n^{(3)} = \sum_{k=1}^{n} \sum_{\substack{(l-k)\wedge(m-k)>T \\ l \neq m}} c_{k-l} c_{k-m} \xi_l \xi_m$$

$$(18) \qquad S_n^{(4)} = \sum_{k=1}^{n} \sum_{m-k>T} \sum_{l=k-T}^{k+T} c_{k-l} c_{k-m} \xi_l \xi_m.$$

By breaking $\left\{ \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n} (d_k - d), \ n = 1, 2, ... \right\}$ into pieces and considering those pieces with different (process) distributions, we just need to show that

$$\lim_{n\to\infty} \frac{S_n^{(1)}}{n^{\frac{1}{p}}} = \lim_{n\to\infty} \frac{S_n^{(2)}}{n^{\frac{1}{p}}} = \lim_{n\to\infty} \frac{S_n^{(3)}}{n^{\frac{1}{p}}} = \lim_{n\to\infty} \frac{S_n^{(4)}}{n^{\frac{1}{p}}} = 0 \text{ a.s.,}$$

provided $p < \frac{1}{2-2\sigma}$. To handle (the diagonal terms) $S_n^{(1)}$, we let $\zeta_l = \xi_l^2 - 1$, set $K = E[\zeta_1^2]$ and use standard steps.

Step 2: Bound second moment of geometric diagonal partial sums $S_{n_r}^{(1)}$.
By symmetry and then integral approximation, we have that

$$
\begin{aligned}
&E[(S_{n_r}^{(1)})^2] \\
=\ & \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \sum_{j=1}^{n_r} \sum_{k=1}^{n_r} c_{k-l}^2 c_{j-m}^2 E[\zeta_l \zeta_m] \\
=\ & K \sum_{l=-\infty}^{\infty} \left| \sum_{k=1}^{n_r} c_{k-l}^2 \right|^2 \\
\overset{r}{\ll}\ & \sum_{k=1}^{n_r} \left(1 + 2\sum_{l=1}^{\infty} l^{-4\sigma} + 2 \sum_{j=k+1}^{n_r} \left(2(j-k)^{-2\sigma} + \sum_{l=-\infty}^{k-1} (k-l)^{-2\sigma}(j-l)^{-2\sigma} \right. \right.
\end{aligned}
$$

$$+ \sum_{l=k+1}^{j-1} (l-k)^{-2\sigma}(j-l)^{-2\sigma} + \sum_{l=j+1}^{\infty} (l-k)^{-2\sigma}(l-j)^{-2\sigma} \Bigg) \Bigg)$$

$$\overset{r}{\ll} \sum_{k=1}^{n_r} \left(1 + \sum_{j=k+1}^{n_r} ((j-k)^{-2\sigma} + (j-k)^{1-4\sigma}) \right)$$

$$(19) \quad \overset{r}{\ll} \quad n_r.$$

Note:

$$\sum_{l=k+1}^{j-1} \frac{1}{(l-k)^{2\sigma}(j-l)^{2\sigma}} \quad \leq \quad 2 \sum_{l=k+1}^{\lfloor \frac{j+k}{2} \rfloor} \frac{1}{(l-k)^{2\sigma}(j-l)^{2\sigma}}$$

$$\overset{j,k}{\ll} \quad (j-k)^{-2\sigma} \sum_{l=k+1}^{\lfloor \frac{j+k}{2} \rfloor} \frac{1}{(l-k)^{2\sigma}}$$

$$(20) \qquad\qquad\qquad\qquad\quad \overset{j,k}{\ll} \quad (j-k)^{(1-4\sigma)}.$$

Step 3: Maximal bound for geometric diagonal partial sums.
Following (19) we have for $n_r \leq n < o < n_{r+1}$

$$E[(S_o^{(1)} - S_n^{(1)})^2] \quad \leq \quad K \sum_{l=-\infty}^{\infty} \left| \sum_{k=n+1}^{o} c_{k-l}^2 \right|^2$$

$$\overset{o,n}{\ll} \quad \sum_{k=n+1}^{o} \left(1 + \sum_{j=k+1}^{o} ((j-k)^{-2\sigma} + (j-k)^{1-4\sigma}) \right)$$

$$(21) \qquad\qquad\qquad\qquad \overset{o,n}{\ll} \quad o - n.$$

Therefore, it follows by Theorem 2.4.1 of Stout [15] with $g(a,n) = Cn$ for some constant $C > 0$ that

$$E \left[\max_{n_r \leq n < o < n_{r+1}} \left(S_o^{(1)} - S_n^{(1)} \right)^2 \right] \quad \overset{r}{\ll} \quad \left(\frac{\log(2(n_{r+1}-n_r))}{\log 2} \right)^2 (n_{r+1} - n_r)$$

$$(22) \qquad\qquad\qquad\qquad\qquad\qquad \overset{r}{\ll} \quad r^2 n_r.$$

Step 4: Use previous two steps to show *normalized* diagonal sums converge.
Combining (19) and (22), one has that

$$(23) \qquad \sum_{r=0}^{\infty} E \left[\max_{n_r \leq n < n_{r+1}} \left(\frac{S_n^{(1)}}{n^{\frac{1}{p}}} \right)^2 \right] \quad \ll \quad \sum_{r=0}^{\infty} r^2 n_r^{1-\frac{2}{p}} < \infty,$$

provided $p \in (0, 2)$. It follows by Fubini's Theorem and n^{th} term divergence that

$$\lim_{n \to \infty} \frac{S_n^{(1)}}{n^{\frac{1}{p}}} = 0.$$

Step 5: Set up for off-diagonal terms.

Letting

$$(24) \qquad a_{l,m}^{2,n} \quad = \quad 2 \sum_{k=1}^{n} 1_{m-T \le k \le l+T} c_{k-l} c_{k-m}$$

$$(25) \qquad a_{l,m}^{3,n} \quad = \quad 2 \sum_{k=1}^{n} 1_{k < l-T} c_{k-l} c_{k-m}$$

$$(26) \qquad a_{l,m}^{4,n} \quad = \quad \sum_{k=1}^{n} 1_{k < m-T} 1_{l-T \le k \le l+T} c_{k-l} c_{k-m},$$

we find that

$$
\begin{aligned}
E\left[(S_n^{(i)})^2\right] &= \sum_{l_1=-\infty}^{\infty} \sum_{m_1=l_1+1}^{\infty} a_{l_1,m_1}^{i,n} \sum_{l_2=-\infty}^{\infty} \sum_{m_2=l_2+1}^{\infty} a_{l_2,m_2}^{i,n} E\left[\xi_{l_1}\xi_{m_1}\xi_{l_2}\xi_{m_2}\right] \\
&= \sum_{l_1=-\infty}^{\infty} \sum_{m_1=l_1+1}^{\infty} a_{l_1,m_1}^{i,n} \sum_{l_2=-\infty}^{\infty} \sum_{m_2=l_2+1}^{\infty} a_{l_2,m_2}^{i,n} \delta_{l_1,l_2} \delta_{m_1,m_2} \\
(27) \qquad &= \sum_{l=-\infty}^{\infty} \sum_{m=l+1}^{\infty} \left(a_{l,m}^{i,n}\right)^2
\end{aligned}
$$

and for $n_r \le n < o < n_{r+1}$

$$(28) \qquad E\left[(S_o^{(i)} - S_n^{(i)})^2\right] = \sum_{l=-\infty}^{\infty} \sum_{m=l+1}^{\infty} \left(a_{l,m}^{i,o} - a_{l,m}^{i,n}\right)^2$$

for $i = 2, 3, 4$. Using a change of variables and the Beta distribution pdf, we have that

$$
\begin{aligned}
\sum_{l=k+1}^{j-1} c_{j-l} c_{k-l} &\stackrel{j,k}{\ll} \int_k^j (j-t)^{-\sigma} (t-k)^{-\sigma} \, dt \\
(29) \qquad &= (j-k)^{1-2\sigma} \underbrace{\int_0^1 (1-s)^{-\sigma} s^{-\sigma} ds}_{B(1-\sigma,1-\sigma)} \stackrel{j,k}{\ll} (j-k)^{1-2\sigma}.
\end{aligned}
$$

Step 6: Apply $S^{(1)}$-procedure for convergence of large c terms $\frac{S_n^{(2)}}{n^{\frac{1}{p}}}$.

Using (29) and integral approximation, one has for $n \in [n_r, n_{r+1})$

$$E\left[(S_n^{(2)})^2\right] - 4\sum_{k=1}^{n}\sum_{m>l} 1_{k-T \leq m \leq k+T} \cdot 1_{k-T \leq l \leq k+T} c_{k-l}^2 c_{k-m}^2$$

$$= 8\sum_{j>k}\sum_{m>l} 1_{j-T \leq m \leq k+T} \cdot 1_{j-T \leq l \leq k+T} c_{j-l} c_{j-m} c_{k-l} c_{k-m}$$

$$\leq 4\sum_{j>k}\left|\sum_{l=j-T}^{k+T} c_{j-l} c_{k-l}\right|^2$$

$$\leq 4\sum_{k=1}^{n}\sum_{j=k+1}^{n \wedge (k+2T)}\left|2c_{j-k} + \sum_{l=j-T}^{k-1} c_{j-l} c_{k-l} + \sum_{l=k+1}^{j-1} c_{j-l} c_{k-l} + \sum_{l=j+1}^{k+T} c_{j-l} c_{k-l}\right|^2$$

$$\overset{n}{\ll} \sum_{k=1}^{n}\sum_{j=k+1}^{k+2T}\left[(j-k)^{-2\sigma} + (j-k)^{2-4\sigma} + (j-k)^{-2\sigma}T^{2-2\sigma}\right]$$

$$\overset{n}{\ll} \quad n l(n),$$

where $l(n) = \begin{cases} T^{3-4\sigma} = n_r^{\nu(3-4\sigma)} & \sigma < \frac{3}{4} \\ \log(T) = \nu\log(n_r) & \sigma = \frac{3}{4} \\ 1 & \sigma > \frac{3}{4} \end{cases}$. Hence,

$$(30) \qquad E\left[(S_n^{(2)})^2\right] \overset{n}{\ll} n l(n) + \sum_{k=1}^{n}\left|\sum_{l=-T}^{T} c_l^2\right|^2 \overset{n}{\ll} n l(n).$$

Similarly, we have for $n_r \leq n < o < n_{r+1}$ that

$$E\left[\left(S_o^{(2)} - S_n^{(2)}\right)^2\right] \overset{o,n}{\ll} \sum_{k=n+1}^{o}\left|\sum_{l=-T}^{T} c_l^2\right|^2 + \sum_{\substack{j,k=n+1 \\ j>k}}^{o}\left|\sum_{l=j-T}^{k+T} c_{j-l} c_{k-l}\right|^2.$$

$$(31) \qquad\qquad\qquad \overset{o,n}{\ll} \quad (o-n)\, l(n).$$

Therefore, it follows by Theorem 2.4.1 of Stout that

$$E\left[\max_{n_r \leq n < o < n_{r+1}}(S_o^{(2)} - S_n^{(2)})^2\right] \overset{r}{\ll} \left(\frac{\log(2n_r)}{\log 2}\right)^2 (n_{r+1} - n_r) l(n_{r+1})$$

$$(32) \qquad\qquad\qquad\qquad \overset{r}{\ll} \quad r^2 n_r l(n_r).$$

Combining (30) with $n = n_r$ and (32), one has that

$$(33) \qquad E\left[\sum_{r=0}^{\infty} \max_{n_r \leq n < n_{r+1}} \left(\frac{S_n^{(2)}}{n^{\frac{1}{p}}}\right)^2\right] \ll \sum_{r=0}^{\infty} r^2 n_r^{1-\frac{2}{p}} l(n_r) < \infty,$$

provided $1 + \nu(3 - 4\sigma) \vee 0 < \frac{2}{p}$ (i.e. $p < \frac{2}{1+\nu(3-4\sigma)}$ when $\sigma < \frac{3}{4}$ and $p < 2$ when $\sigma \geq \frac{3}{4}$, both of which are true). It follows that $\lim_{n \to \infty} \frac{S_n^{(2)}}{n^{\frac{1}{p}}} = 0$ a.s.

Step 7: Apply $S^{(1)}$-procedure for convergence of small c terms $\frac{S_n^{(3)}}{n^{\frac{1}{p}}}$.

$$E\left[(S_n^{(3)})^2\right]$$
$$= 8 \sum_{j>k} \sum_{m>l} 1_{j+T<l} \cdot 1_{k+T<l} c_{j-l} c_{j-m} c_{k-l} c_{k-m}$$
$$+ 4 \sum_{k=1}^{n} \sum_{m>l} 1_{k+T<l} c_{k-l}^2 c_{k-m}^2$$
$$\leq 4 \sum_{j>k} \left|\sum_{l=j+T+1}^{\infty} c_{j-l} c_{k-l}\right|^2 + 2 \sum_{k=1}^{n} \left|\sum_{l=k+T+1}^{\infty} c_{k-l}^2\right|^2$$
$$\overset{n}{\ll} \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} \left|\int_{j+T}^{\infty} (t-j)^{-\sigma} (t-k)^{-\sigma} dt\right|^2 + \sum_{k=1}^{n} \left|\int_{k+T}^{\infty} (t-k)^{-2\sigma} dt\right|^2$$
$$\overset{n}{\ll} \sum_{k=1}^{n} \left(\sum_{j=k+1}^{n} \left|\int_{T}^{\infty} t^{-2\sigma} dt\right|^2 + \left|\int_{T}^{\infty} t^{-2\sigma} dt\right|^2\right)$$
$$(34) \overset{n}{\ll} n^2 T^{2-4\sigma}.$$

Similarly, we have for $n_r \leq n < o < n_{r+1}$ that

$$(35)\ E\left[\left(S_o^{(3)} - S_n^{(3)}\right)^2\right] \overset{o,n,r}{\ll} (o-n) o T^{2-4\sigma} \overset{o,n,r}{\ll} (o-n) n_{r+1}^{1+\nu(2-4\sigma)}.$$

Therefore, it follows by Theorem 2.4.1 of Stout that

$$E\left[\max_{n_r \leq n < o < n_{r+1}} \left(S_o^{(3)} - S_n^{(3)}\right)^2\right] \overset{r}{\ll} \left(\frac{\log(2n_r)}{\log 2}\right)^2 (n_{r+1} - n_r) n_{r+1}^{1+\nu(2-4\sigma)}$$
$$(36) \overset{r}{\ll} r^2 n_r^{2+\nu(2-4\sigma)}.$$

Combining (34) with $n = n_r$ and (36), one has

$$(37) \qquad E\left[\sum_{r=0}^{\infty} \max_{n_r \leq n < n_{r+1}} \left(\frac{S_n^{(3)}}{n^{\frac{1}{p}}}\right)^2\right] \ll \sum_{r=0}^{\infty} r^2 n_r^{2+\nu(2-4\sigma)-\frac{2}{p}} < \infty,$$

provided $p < \frac{1}{1+\nu(1-2\sigma)}$, which is the given condition, so $\lim_{n\to\infty} \frac{S_n^{(3)}}{n^{\frac{1}{p}}} = 0$ a.s..

It is notable that condition on p, $p < \frac{2}{1+\nu(3-4\sigma)}$, in step 6 gets more stringent when $\nu > 1$ and the same is true for condition on p, $p < \frac{1}{1+\nu(1-2\sigma)}$, in step 7 when $\nu < 1$, so the best choice that raises the same condition on p is when $\nu = 1$. Hence, we will have to satisfy $p < \frac{1}{1-2\sigma}$ in either cases.

Step 8: Apply $S^{(1)}$-procedure for convergence of mixed terms $\frac{S_n^{(4)}}{n^{\frac{1}{p}}}$. Finally, we note

$$
E\left[(S_n^{(4)})^2\right]
$$

$$
= \sum_{k=1}^{n}\sum_{m=k+T+1}^{\infty} c_{k-m}^2 \sum_{l=k-T}^{l=k+T} c_{k-l}^2 + 2\sum_{k=1}^{n}\sum_{j=k+1}^{k+2T}\sum_{m=j+T+1}^{\infty} c_{j-m}c_{k-m} \sum_{l=j-T}^{k+T} c_{j-l}c_{k-l}
$$

$$
\overset{n}{\ll} \sum_{k=1}^{n}\left\{ T^{1-2\sigma} + \sum_{j=k+1}^{k+2T} T^{1-2\sigma}\left[(j-k)^{-\sigma} + (j-k)^{1-2\sigma} + (j-k)^{-\sigma} T^{1-\sigma}\right]\right\}
$$

$$
\overset{n}{\ll} nT^{3-4\sigma}.
$$

Similarly, we have for $n_r \le n < o < n_{r+1}$ that

$$
E\left[\left(S_o^{(4)} - S_n^{(4)}\right)^2\right] \overset{o,n}{\ll} (o-n)T^{3-4\sigma}.
$$

Therefore, it follows by $\nu = 1$ and Theorem 2.4.1 of Stout that

$$
E\left[\max_{n_r \le n < o < n_{r+1}}\left(S_o^{(4)} - S_n^{(4)}\right)^2\right] \overset{r}{\ll} \left(\frac{\log(2n_r)}{\log 2}\right)^2 (n_{r+1}-n_r)n_{r+1}^{3-4\sigma} \overset{r}{\ll} r^2 n_r^{4-4\sigma}.
$$

Combining these two equations, one has

$$
(38) \qquad E\left[\sum_{r=0}^{\infty}\max_{n_r \le n < n_{r+1}}\left(\frac{S_n^{(4)}}{n^{\frac{1}{p}}}\right)^2\right] \ll \sum_{r=0}^{\infty} r^2 n_r^{(4-4\sigma)-\frac{2}{p}} < \infty,
$$

provided $p < \frac{1}{2-2\sigma}$, which is true. It follows that $\lim_{n\to\infty} \frac{S_n^{(4)}}{n^{\frac{1}{p}}} = 0$ a.s. $\square$

4.3. **Proof of Theorem 3**. Without loss of generality we assume $1 < \alpha < 2$.

Step 1: Reduce to continuous $\{(\xi_l, \overline{\xi}_l)\}$.
Let $\{(U_l)\}_{l\in\mathbb{Z}}$ be independent $[-1,1]$-uniform random variables that are independent of everything and set $\overline{U}_l = U_l$ for all l. Then, we have that

$$
\frac{1}{n^{\frac{1}{p}}}\sum_{k=1}^{n}(d_k - d) = \frac{1}{n^{\frac{1}{p}}}\sum_{k=1}^{n}\sum_{l,m=-\infty}^{\infty} c_{k-l}\overline{c}_{k-m}\left((\xi_l + U_l)(\overline{\xi}_m + \overline{U}_m) - d - \frac{2}{3}\right)
$$

$$(39) \qquad -\frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n} \sum_{l,m=-\infty}^{\infty} c_{k-l}\bar{c}_{k-m} \left(\xi_l \overline{U}_m + U_l \bar{\xi}_m + U_l \overline{U}_m - \frac{2}{3} \right).$$

However,

$$(40)\ \lim_{n\to\infty} \frac{1}{n^{\frac{1}{p}}} \sum_{k=1}^{n} \sum_{l,m=-\infty}^{\infty} c_{k-l}\bar{c}_{k-m} \left(\xi_l \overline{U}_m + U_l \bar{\xi}_m + U_l \overline{U}_m - \frac{2}{3} \right) = 0$$

by Theorem 8. Moreover, $\xi_1 + U_1, \bar{\xi}_1 + \overline{U}_1$ have the same moment and tail probability bounds as $\xi_1, \bar{\xi}_1$. Hence, without loss of generality, we can assume $\xi_l, \bar{\xi}_m$ are continuous random variables, which will be important for the truncation to follow in Step 4.

Step 2: Handle off-diagonal sum as previous proof since unaffected by heavy tails.
Suppose $S_n^{(2)}$, $S_n^{(3)}$ and $S_n^{(4)}$ are defined as in (16-18). Then, we know that

$$\lim_{n\to\infty} \frac{S_n^{(2)}}{n^{\frac{1}{p}}} = \lim_{n\to\infty} \frac{S_n^{(3)}}{n^{\frac{1}{p}}} = \lim_{n\to\infty} \frac{S_n^{(4)}}{n^{\frac{1}{p}}} = 0 \text{ a.s.,}$$

provided $p < \frac{1}{2-\sigma-\bar{\sigma}}$ by the proof of Theorem 8.

Step 3: Reduce $\xi_l \bar{\xi}_l$ (in diagonal sum) to non-negative with single atom at 0.
Noting

$$\sum_{l=-\infty}^{\infty} c_{k-l}\bar{c}_{k-l}(\xi_l\bar{\xi}_l - E[\xi_l\bar{\xi}_l])$$

$$(41)\ = \sum_{l=-\infty}^{\infty} c_{k-l}\bar{c}_{k-l}((\xi_l\bar{\xi}_l)^+ - E[(\xi_l\bar{\xi}_l)^+]) - \sum_{l=-\infty}^{\infty} c_{k-l}\bar{c}_{k-l}((\xi_l\bar{\xi}_l)^- - E[(\xi_l\bar{\xi}_l)^-]),$$

we only have to consider the case where $\xi_l\bar{\xi}_l \geq 0$ for the remainder of the proof. Moreover, insomuch as the proof of the general case only differs cosmetically from the notationally-simpler case where $\bar{\xi}_l = \xi_l$, $E[\xi_1^2] = 1$ and $\bar{c}_l = c_l = \begin{cases} 1 & l = 0 \\ |l|^{-\sigma} & l \neq 0 \end{cases}$, we only provide the proof of the later for which the long-range dependence constraint becomes $p < \frac{1}{2-2\sigma}$. We will however indicate the most significant changes that would be needed for the general case.

Step 4: Divide diagonal terms into zero-mean truncated (i.e. bounded) and remainder pieces.
Let $\kappa > 0$. Fix $u_r^+ = n_r^{\frac{\kappa}{2-\alpha}}$ to find

$$(42) \quad 2\int_0^{u_r^+} P(\xi_1^2 > s)s\,ds \overset{r}{\ll} 2\int_0^{u_r^+} s s^{-\alpha}\,ds \overset{r}{\ll} n_r^\kappa \ \ \forall\, r = 1, 2, \ldots$$

Now, by defining

$$(43) \quad \begin{cases} \overline{\zeta}_i = \overline{\zeta}_i^r = (\xi_i^2 \wedge u_r^+) - \vartheta_i, \text{ where } \vartheta_i \doteq \int_0^{u_r^+} P(\xi_i^2 > s)ds \le 1, \\ \tilde{\zeta}_i = \tilde{\zeta}_i^r = \xi_i^2 - 1 - \overline{\zeta}_i^r, \end{cases}$$

we find that

$$(44) \qquad E[\overline{\zeta}_i] = \int_0^{u_r^+} P(\xi_i^2 > t)dt - \int_0^{u_r^+} P(\xi_i^2 > t)dt = 0,$$

so both $\overline{\zeta}_i$ and $\tilde{\zeta}_i$ are zero mean, and by (42)

$$E[|\overline{\zeta}_1|^2] = E|\xi_1^2 \wedge u_r^+|^2 - \left(\int_0^{u_r^+} P(\xi_1^2 > t)dt \right)^2$$

$$(45) \qquad = 2\int_0^{u_r^+} P(\xi_1^2 > s)s\,ds - \left(\int_0^{u_r^+} P(\xi_1^2 > t)dt \right)^2 \overset{r}{\ll} n_r^\kappa \quad \forall\, r = 1, 2, \dots$$

(In the general case, we note that $\xi_1\overline{\xi}_1$ is non-negative and of continuous distribution on $(0, \infty)$ so $E[\xi_1\overline{\xi}_1 \wedge u_r^+] = \int_0^{u_r^+} P(\xi_1\overline{\xi}_1 > s)ds$ as required. We also have $\tilde{\zeta}_i^r = \xi_i\overline{\xi}_i - E[\xi_i\overline{\xi}_i] - \overline{\zeta}_i^r$.)

Step 5: Moment Bound for truncated using the proof of Theorem 8. Noting $\{\overline{\zeta}_i\}$ are i.i.d. with $E[\overline{\zeta}_1] = 0$ and $E[\overline{\zeta}_1^2] < \infty$ and defining

$$(46) \qquad\qquad S_n^{(1)} = \sum_{k=1}^n \sum_{l=-\infty}^\infty c_{k-l}^2 \overline{\zeta}_l,$$

one finds from (23) in the proof of Theorem 8 that

$$(47) \qquad\qquad E\left[\max_{n_r \le n < n_{r+1}} \left(S_n^{(1)} \right)^2 \right] \le E|\overline{\zeta}_1|^2 r^2 n_r.$$

Hence, it follows by (45) that

$$(48) \qquad\qquad E\left[\max_{n_r \le n < n_{r+1}} \left(S_n^{(1)} \right)^2 \right] \overset{r}{\ll} r^2 n_r^{1+\kappa}.$$

Step 6: Moment Bound for remainder using Doob's inequality. Turning to the $\tilde{\zeta}_i^r$ and using the formula

$$(49) \quad E[g(X)] = \int_0^\infty g'(t)P(X > t)dt - \int_{-\infty}^0 g'(t)P(X < t)dt,$$

one has by our tail probability bounds that the non-negative part of $\tilde{\xi}_1$ satisfies

$$
\begin{aligned}
E|\tilde{\zeta}_1^+|^\tau &= \tau \int_0^\infty s^{\tau-1} P(\xi_1^2 > u_r^+ + s + 1 - \vartheta_1) ds \\
&\leq \tau \int_0^\infty s^{\tau-1} P(\xi_1^2 > u_r^+ + s) ds \text{ since } \vartheta_1 \leq 1 \\
&\overset{r}{\ll} \int_{u_r^+}^\infty (s - u_r^+)^{\tau-1} s^{-\alpha} ds \\
&\leq \int_{u_r^+}^{2u_r^+} (s - u_r^+)^{\tau-1} ds (u_r^+)^{-\alpha} + \int_{2u_r^+}^\infty (s - u_r^+)^{\tau-\alpha-1} ds
\end{aligned}
$$

$$
(50) \qquad \overset{r}{\ll} (u_r^+)^{\tau-\alpha} \overset{r}{\ll} n_r^{\frac{\kappa(\tau-\alpha)}{2-\alpha}},
$$

for $1 < \tau < \alpha$. Therefore, it follows by Jensen's inequality and Doob's L_p inequality that

$$
\begin{aligned}
E^{\frac{1}{\tau}} \left[\sup_{n_r \leq n < n_{r+1}} \left| \sum_{k=1}^n \sum_{l=-\infty}^\infty c_l^2 \tilde{\zeta}_{k-l}^r \right|^\tau \right] &\leq E^{\frac{1}{\tau}} \left[\left| \sum_{l=-\infty}^\infty c_l^2 \sup_{n_r \leq n < n_{r+1}} \left| \sum_{k=1}^n \tilde{\zeta}_{k-l}^r \right| \right|^\tau \right] \\
&\overset{r}{\ll} \sum_{l=-\infty}^\infty c_l^2 E^{\frac{1}{\tau}} \left[\sup_{n_r \leq n < n_{r+1}} \left| \sum_{k=1}^n \tilde{\zeta}_{k-l}^r \right|^\tau \right] \\
&\overset{r}{\ll} \sum_{l=-\infty}^\infty c_l^2 E^{\frac{1}{\tau}} \left[\left| \sum_{k=1}^{n_{r+1}-1} \tilde{\zeta}_{k-l}^r \right|^\tau \right]
\end{aligned}
$$

$$
(51) \qquad \overset{r}{\ll} n_r \|\tilde{\zeta}_1^r\|_\tau,
$$

so by (50,51)

$$
(52) \qquad E \left[\sup_{n_r \leq n < n_{r+1}} \left| \sum_{k=1}^n \sum_{l=-\infty}^\infty c_l^2 \tilde{\zeta}_{k-l}^r \right|^\tau \right] \overset{r}{\ll} n_r^{\tau - \frac{\kappa(\alpha-\tau)}{2-\alpha}}.
$$

Step 7: Use Truncation and Error Term bounds with Borel-Cantelli for convergence.

Combining (48) and (52), one has that

$$
P \left(\sup_{n_r \leq n < n_{r+1}} \left| \sum_{k=1}^n \sum_{l=-\infty}^\infty c_l^2 \zeta_{k-l} \right| > 2\epsilon n_r^{\frac{1}{p}} \right)
$$

$$
\leq \frac{E \left[\sup_{n_r \leq n < n_{r+1}} \left| \sum_{k=1}^n \sum_{l=-\infty}^\infty c_l^2 \overline{\zeta}_{k-l}^r \right|^2 \right]}{\epsilon^2 n_r^{\frac{2}{p}}} + \frac{E \left[\sup_{n_r \leq n < n_{r+1}} \left| \sum_{k=1}^n \sum_{l=-\infty}^\infty c_l^2 \tilde{\zeta}_{k-l}^r \right|^\tau \right]}{\epsilon^\tau n_r^{\frac{\tau}{p}}}
$$

$$\overset{r}{\ll} \quad r^2 n_r^{1+\kappa-\frac{2}{p}} + n_r^{\tau - \frac{\kappa(\alpha-\tau)}{2-\alpha} - \frac{\tau}{p}}$$

$$(53) \quad \overset{r}{\ll} \quad r^2 n_r^{1-\frac{\alpha}{p}} + n_r^{\tau - \frac{\alpha}{p}},$$

by letting $\kappa = \frac{2-\alpha}{p}$. Hence, if $\tau \in \left(1, \frac{\alpha}{p}\right)$, then

$$(54) \qquad \sum_{r=1}^{\infty} P\left(\sup_{n_r \leq n < n_{r+1}} \left|\sum_{k=1}^{n} \sum_{l=-\infty}^{\infty} c_l^2 \zeta_{k-l}\right| > 2\epsilon n_r^{\frac{1}{p}}\right) < \infty,$$

under our heavy-tail condition $p < \alpha$ and

$$(55) \qquad n^{-\frac{1}{p}} \sum_{k=1}^{n} \sum_{l=-\infty}^{\infty} c_l^2 \zeta_{k-l} \to 0 \quad \text{a.s.},$$

by Borel-Cantelli. The proof is complete. $\square$

REFERENCES

[1] Avram, F. and Taqqu, M. S. (1987). *Generalized powers of strongly dependent random variables.* Ann. Probab., **vol**. 15, pp. 767-775.

[2] Berbee, H. (1987). *Convergence rates in the strong law for a bounded mixing sequence.* Probability Theory and Related Fields, **vol**. 74, pp. 253-270.

[3] Davis, R.A. and Resnick, S.I. (1986). *Limit theory for the sample covariance and correlation functions of moving averages.* Ann. Statist., **vol**. 14, pp. 533-558.

[4] Dobrushin, R.L. and Major, P. (1979). *Non-central limit theorems for non-linear functions of Gaussian fields.* Z. Wahrscheinlichkeitstheorie Verw. Geb., **vol**. 50, pp. 27-52.

[5] Giraitis L. and Surgailis, D. (1986). *Multivariate Appell polynomials and the central limit theorem.* Dependence in Probability and Statistics, E. Eberlein and M. S. Taqqu (Eds), Birkhauser, Boston, pp. 21-71.

[6] Giraitis L. and Surgailis, D. (1989). *Limit theorem for polynomials of linear process with long-range dependence.* Lith. Math. J., **vol**. 29(2), pp. 128-145.

[7] Horvath, L. and Kokoszka, P. (2008). *Sample autocovariances of long-memory time series.* Bernoulli, **vol**. 14(2), pp. 405-418.

[8] Karagiannis, T., Molle, M. and Faloutsos, M. (2004). *Long-Range Dependence Ten Years of Internet Traffic Modeling.* IEEE Computer Society.

[9] Kouritzin, M. A. and Sadeghi, S. *Convergence Rates and Decoupling in Linear Stochastic Approximation Algorithms.* submitted to Annals of Statistics.

[10] Louhchi, S. and Soulier, P. (2000). *Marcinkiewicz-Zegmond Strong Laws for Infinite Variance Time Series.* Statistical Inference for Stochastic Processes, **vol**. 3, pp. 31-40.

[11] Mandelbrot, B. and Wallis J. (1968). *Noah, Joseph and operational hydrology.* Water Resources Research, **vol**. 4, pp. 909-918.

[12] Mandelbrot, B. (1972). *Statistical methodology for non-periodic cycles: from the covariance to R/S analysis.* Ann. Econ. and Social Measurement, **vol**. 1, pp. 259-290.

[13] Rio, E. (1995). *A Maximal Inequality and Dependent Marcinkiewicz-Zegmond Strong Laws.* The Annals of Probability, **vol.**23 No. 2, pp. 918-937.

[14] Rosenblatt, M. (1961). *Independence and dependence.* Proc. 4th Berkeley Symp. Math. Statist. Probab., pp. 411-443.

[15] Stout, W. F. (1974). *Almost Sure Convergence.* Academic Press Inc., pp. 126.

[16] Surgailis,D. (1982). *Zones of attraction of self-similar multiple integrals.* Lithuanian Math. J., **vol**. 22, pp. 327-340.

[17] Surgailis,D. (2004). *Stable limits of sums of bounded functions of long-memory moving averages with finite variance.* Bernoulli, **vol**. 10(2), pp. 327-355.

[18] Taqqu, M.S. (1979). *Convergence of integrated processes of arbitrary Hermite rank.* Z. Wahrscheinlichkeitstheorie Verw. Geb., **vol**. 50, pp. 53-83.

[19] Vaiciulis, M. (2003). *Convergence of sums of Appell polynomials with infinite variance.* Lithuanian Math. J., **vol**. 43, pp. 80-98.

[20] Varotsos, C. and Kirk-Davidoff, D. (2006). *Long-memory processes in global ozone and temperature variations at the region 60^0 S-60^0 N.* Atmospheric Chemistry and Physics, **vol.**6, pp. 4093-4100.

[21] Wu, W. B. and Min, W. (2005). *On linear processes with dependent innovations.* Stochastic Processes and their Applications, **vol.** 115, pp. 939-958.

Department of Mathematical and Statistical Sciences
University of Alberta, Edmonton, AB T6G 2G1 Canada