ON THE CONVERGENCE OF MOMENTS IN THE ALMOST SURE CENTRAL LIMIT THEOREM FOR STOCHASTIC APPROXIMATION ALGORITHMS

Peggy Cénac

Abstract. We study the almost sure asymptotic behaviour of stochastic approximation algorithms for the search of zero of a real function. The quadratic strong law of large numbers is extended to the powers greater than one. In other words, the convergence of moments in the almost sure central limit theorem (ASCLT) is established. As a by-product of this convergence, one gets another proof of ASCLT for stochastic approximation algorithms. The convergence result is applied to several examples as estimation of quantiles and recursive estimation of the mean.

1. Introduction

Let (X_n) be a sequence of independent and identically distributed (i.i.d.) random variables with $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[X_n^2] = \sigma^2$. For the empirical measures

$$G_n \stackrel{\text{def}}{=} \frac{1}{\log n} \sum_{k=1}^{n} \frac{1}{k} \delta_{\frac{S_k}{\sqrt{k}}} \quad \text{with} \quad S_n \stackrel{\text{def}}{=} \sum_{k=1}^{n} X_k,$$

the almost sure central limit theorem (ASCLT) states that, with probability one, $G_n \Longrightarrow G$ where G stands for the $\mathcal{N}(0, \sigma^2)$ distribution and $\Longrightarrow$ denotes the convergence in distribution. It was simultaneously established by Brosamler [5], Schatte [29] and in the present form by Lacey and Phillip [16]. Moreover, under assumptions on the moments of (X_n), the convergence of moments in the ASCLT has been established recently by Bercu [2], Bercu and Fort [3]. That is to say, the following convergences hold:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \frac{1}{k} \left(\frac{S_k}{\sqrt{k}} \right)^{2p} = \frac{\sigma^{2p}(2p)!}{2^p p!} \quad \text{a.s.,} \tag{1.1}$$

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \frac{1}{k} \left(\frac{S_k}{\sqrt{k}} \right)^{2p-1} = 0 \quad \text{a.s.} \tag{1.2}$$

In fact, this result is a simplified version for i.i.d. random variables. It is extended in a context of martingales. The case of multidimensional martingales is considered in Bercu *et al.* [4]. Despite the availability of a wide

---

[1] Institut de Mathématiques de Bourgogne, IMB UMR 5584 CNRS, 9 rue Alain Savary, BP 47870, 21078 Dijon Cedex, France. peggy.cenac@u-bourgogne.fr

literature concerning the ASCLT for independent variables, very few references can be found on the ASCLT for martingales apart from the important contribution of Chaâbane [6, 7], Chaâbane and Maâouia [8], Chaâbane *et al.* [9] and Lifshits [22, 23]. As a by-product of the convergence of moments, Bercu and Fort [3] propose a proof of the ASCLT for unidimensional martingale transforms via an original approach which uses the Carleman moment theorem. In this paper, we prove that stochastic approximation algorithms used for the search of zeros and which are known to satisfy an ASCLT (see Pelletier [27]) also fulfill the convergence of moments in the ASCLT.

More precisely, we consider a stochastic algorithm of the form

$$Z_{n+1} = Z_n + \gamma_n \left[h\left(Z_n \right) + R_{n+1} \right] + \sigma_n \varepsilon_{n+1}, \tag{1.3}$$

where the function h is defined on $\mathbb{R}$ and $\mathbb{R}$-valued. The two sequences (R_n) and (ε_n) are two real disturbances, defined on a probability space and adapted to a filtration $\mathbb{F} \stackrel{\text{def}}{=} (\mathcal{F}_n)_{n \geq 0}$. The stepsizes (γ_n) and (σ_n) are two deterministic positive sequences going to zero. This model includes the algorithms of Robbins–Monro and Kiefer–Wolfowitz as well as algorithms with Markovian disturbances. Stochastic approximation algorithms such as (1.3) used for the search of zeros of h have been widely studied under various assumptions (see Duflo [10] for an overview of these algorithms). In the context of Brownian diffusion process, Lamberton and Pagès [17, 18] establish the convergence of weighted empirical measures in the form of G_n with a view to getting an approximation of the diffusion's invariant distribution, even in the case when the diffusion may have several invariant distributions. As a corollary of their main convergence result, they get the almost sure central limit theorem in the context of Brownian diffusion processes.

Let z^* be a zero of h. Many criteria ensure the almost sure convergence of (Z_n) to z^*. In the wide literature concerning this convergence, let us refer to Benveniste *et al.* [1], Duflo [10], Dupuis and Kushner [11], Hall and Heyde [13], Kushner and Clark [15], Ljung *et al.* [24]. Moreover, in the case when (Z_n) converges almost surely to z^*, the weak convergence rate is given by

$$\sqrt{\frac{\gamma_n}{\sigma_n^2}} \left(Z_n - z^* \right) \Longrightarrow \mathcal{N}\left(0, \Sigma^2 \right), \tag{1.4}$$

where Σ^2 is a positive real number related to the second moment of (ε_n) and to the differential of the function h at the point z^* (see among many others Benveniste *et al.* [1], Duflo [10], Kushner and Clark [15], Ljung *et al.* [24], Zhu [30]).

Let us define $v_n \stackrel{\text{def}}{=} \gamma_n \sigma_n^{-2}$. The main goal of this paper is to establish an analog result of (1.1) and (1.2) in the context of stochastic approximation algorithms. More precisely, under appropriate assumptions on moments of the noise (ε_n), we prove the following result: for $p \geq 1$,

$$\lim_{n \to \infty} \frac{1}{\sum_{k=1}^n \gamma_k} \sum_{k=1}^n \gamma_k \left[\sqrt{v_k} \left(Z_k - z^* \right) \right]^p = g(p) \quad \text{a.s.}, \tag{1.5}$$

where $g(p)$ is the moment of order p of the Gaussian distribution $\mathcal{N}(0, \Sigma^2)$. The ASCLT for stochastic approximation algorithms can be viewed as a corollary of this result due to Carleman's theorem. By the way, in the scalar case, we get another proof for the ASCLT established by Pelletier [27]. Moreover, in the particular case $p = 2$, the convergence (1.5) is the quadratic strong law of large numbers established by Le Breton [19], Le Breton and Novikov [20, 21], Pelletier [26]. The main contribution of this paper is the extension (1.5) of the strong law to any power p.

The paper is organized as follows. Section 2 is devoted to the assumptions and the presentation of the main results. In Section 3, we propose some statistical applications to estimation errors for several particular cases of Robbins–Monro procedure. The proofs are established in Section 4.

2. Assumptions and main results

Let us begin with the definition of the class of positive sequences that will be used in our assumptions. This definition is introduced in Mokkadem and Pelletier [25].

Definition 2.1. Let $\alpha \in \mathbb{R}$ and (v_n) be a nonrandom positive sequence. Then (v_n) is said to be in the set $\mathcal{GS}(\alpha)$ if

$$\lim_{n \to \infty} n \left(1 - \frac{v_{n-1}}{v_n} \right) = \alpha.$$

For example, typical sequences in $\mathcal{GS}(\alpha)$ are $n^\alpha (\log n)^\beta$ or $n^\alpha (\log \log n)^\beta$, for $\alpha, \beta \in \mathbb{R}$. The assumptions we will refer to in the sequel are the following.

($H1$) Z_n converges almost surely to z^*.

($H2$) The function h is defined on $\mathbb{R}$ and z^* is a zero of h such that, on a neighborhood of z^*,

$$h(z) = H(z - z^*) + \mathcal{O}\left(|z - z^*|^2 \right), \tag{2.1}$$

where $H < 0$.

($H3$) The noise (ε_n) is a martingale difference sequence such that

$$\lim_{n \to \infty} \mathbb{E}\left[\varepsilon_{n+1}^2 | \mathcal{F}_n \right] = \sigma^2 \quad \text{a.s.}$$

($H4$) The disturbance (R_n) is split into two terms such that

$$R_n = r_n + \mathcal{O}\left(|Z_{n-1} - z^*|^2 \right) \quad \text{a.s.},$$
$$|r_n| = o\left(v_n^{-1/2} (\log s_n)^{-q} \right) \quad \text{a.s.} \quad \forall q \geq 0, \tag{2.2}$$
$$\text{where} \quad s_n \stackrel{\text{def}}{=} \sum_{k=1}^n \gamma_k.$$

($H5$) The nonrandom sequences (γ_n) and (σ_n) are such that

$$(\gamma_n) \in \mathcal{GS}(-\alpha) \quad \text{with} \quad \alpha \in \left] \max\left\{ \frac{1}{2}, \frac{2}{a} \right\}, 1 \right],$$
$$(\sigma_n) \in \mathcal{GS}(-\beta) \quad \text{with} \quad \beta \in \left] \frac{\alpha}{2}, \alpha \right],$$
$$\lim_{n \to \infty} n\gamma_n > -\frac{2\beta - \alpha}{2H}.$$

Let $\xi \stackrel{\text{def}}{=} \lim_{n \to \infty} (n\gamma_n)^{-1}$. We can now define the asymptotic variance Σ^2 which appears in (1.4):

$$\Sigma^2 \stackrel{\text{def}}{=} \frac{-\sigma^2}{2H + \xi(2\beta - \alpha)}.$$

Note that ($H5$) implies that $\xi \in [0, -2H/(2\beta - \alpha)[$ and thus that Σ^2 is well defined and strictly positive.

Comments on the assumptions.

The residual term (R_n) enables the study of algorithms with small Markovian disturbances. The way these algorithms can be rewritten as equation (1.3) is detailed in Duflo [10].

Let us now comment the last assumption $(H5)$ on the gains. For the usual gains

$$\gamma_n = \frac{\gamma_0}{n^\alpha} \quad \text{and} \quad \sigma_n = \frac{\sigma_0}{\sqrt{n^{\alpha+\beta}}}, \quad \text{with} \quad \gamma_0 > 0, \ \sigma_0 > 0, \quad \text{and} \quad 0 < \beta \le \alpha,$$

if $\alpha \in]\max\{1/2, 2/a\}, 1[$ or $(\alpha = 1$ and $\beta < -2H\gamma_0)$, assumption $(H5)$ is fulfilled. In the particular case of Robbins–Monro algorithm, that is to say $R_n = 0$, $\sigma_n = \gamma_n$, $(H5)$ contains for example the gains $\gamma_n = \sigma_n = \gamma_0/n^\alpha$ with $\gamma_0 > 0$ and $\alpha \in]\max\{1/2, 2/a\}, 1[$ or $(\alpha = 1$ and $\gamma_0 > -1/2H)$. The Kiefer–Wolfowitz algorithm corresponds to the case $h = V'$ where V is observable only together with a noise. For example classical gains satisfying $(H5)$ could be

$$\gamma_n = \frac{\gamma_0}{n^\alpha}, \quad \sigma_n = n^\tau \gamma_n, \quad \frac{\alpha}{6} \le \tau < \frac{\alpha}{2}, \quad \text{and} \quad \gamma_0 > 0, \ \sigma_0 > 0,$$

$\alpha \in]\max\{1/2, 2/a\}, 1[$ or $(\alpha = 1$ and $\gamma_0 > \frac{2\tau-1}{2H})$. In these examples, the optimal rate is obtained for $\gamma_n = \gamma_0/n$ but leads to a condition on the initial γ_0. To circumvent this condition, Koval and Schwabe [14] introduced the stepsize, for any $(p, d) \in \mathbb{N}^2$,

$$\gamma_n = \frac{\gamma_0 \left(\log_p n\right)^d}{n}, \quad \text{where} \quad \log_p n = \log\left(\log_{p-1} n\right),$$

which also fulfills $(H5)$. This stepsize leads to convergence rates very close to the ones obtained with γ_0/n but without any constraints on the initial γ_0.

The main result of this paper is the following theorem, which establishes the convergence of moments in the ASCLT for stochastic algorithms.

Theorem 2.2. *Set an integer $p \ge 1$. Assume that for some real $m > 2p$, (ε_n) satisfies the moment condition*

$$\sup_{n\ge 0} \mathbb{E}\left[|\varepsilon_{n+1}|^m | \mathcal{F}_n\right] < \infty \quad a.s. \tag{2.3}$$

Under assumptions $(H1)$ to $(H5)$, one has

$$\lim_{n\to\infty} \frac{1}{s_n} \sum_{k=1}^n \gamma_k \left[\sqrt{v_k}\,(Z_k - z^*)\right]^{2p} = \frac{\Sigma^{2p}(2p)!}{2^p p!} \quad a.s. \tag{2.4}$$

$$\lim_{n\to\infty} \frac{1}{s_n} \sum_{k=1}^n \gamma_k \left[\sqrt{v_k}\,(Z_k - z^*)\right]^{2p-1} = 0 \quad a.s. \tag{2.5}$$

In the particular case $p = 1$, the convergence (2.4) gives the quadratic strong law established by Pelletier [26]. Theorem 2.2 extends this result to all powers $\ne 2$. The limiting constants in (2.4) and (2.5) correspond to the moments of the Gaussian distribution $\mathcal{N}(0, \Sigma^2)$.

The classical moment problem is related to the question whether or not a given sequence uniquely determines the associated probability distribution. The well-known Carleman theorem (see *e.g.* Feller [12]) gives a condition on the moments in order to ensure the unicity of the distribution.

Theorem 2.3 (Carleman). *A probability distribution is uniquely determined by its moments (m_n) if*

$$\sum_{n=1}^\infty m_{2n}^{-1/2n} = \infty.$$

Since the Gaussian limit distribution satisfies Carleman's moment condition, we deduce the following corollary from Theorem 2.2.

Corollary 2.4 (ASCLT)**.** *Assume that (ε_n) is a martingale difference sequence such that, for all integer $p \geq 1$,*

$$\sup_{n \geq 0} \mathbb{E}\left[|\varepsilon_{n+1}|^p | \mathcal{F}_n\right] < \infty \quad a.s.$$

Under assumptions $(H1)$ to $(H5)$, one has

$$\frac{1}{s_n} \sum_{k=1}^{n} \gamma_k \delta_{\sqrt{v_k}(Z_k - z^*)} \Longrightarrow \mathcal{N}\left(0, \Sigma^2\right) \quad a.s.$$

As a straightforward application, the following strong law holds.

Corollary 2.5. *Assume that (ε_n) is a martingale difference sequence such that, for all integer $p \geq 1$,*

$$\sup_{n \geq 0} \mathbb{E}\left[|\varepsilon_{n+1}|^p | \mathcal{F}_n\right] < \infty \quad a.s.$$

Moreover, assume that f stands for any almost everywhere continuous function, with polynomial growth at infinity. Under assumptions $(H1)$ to $(H5)$, one has

$$\lim_{n \to \infty} \frac{1}{s_n} \sum_{k=1}^{n} \gamma_k f\left(\sqrt{v_k}\left(Z_k - z^*\right)\right) = \int f \mathrm{d}G_\Sigma \quad a.s.,$$

where G_Σ denotes the $\mathcal{N}(0, \Sigma)$ distribution.

This corollary gives approximations of gaussian integral of almost everywhere continuous function with polynomial growth at infinity.

The Proof of Theorem 2.2 is given in Section 4.

Remark 2.6. Assumptions $(H1)$ to $(H5)$ are equivalent to those of Pelletier [27] establishing the ASCLT. Here we assume the conditional moment assumption besides to get the stronger result of convergence of moments in the ASCLT.

3. Examples of applications

Let us give here three examples of estimation for which the Theorem 2.2 apply. These examples are different forms of Robbins–Monro procedure, described as follows. Robbins–Monro algorithm (see Robbins and Monro [28]) is used for solving the equation $f(x) = \alpha$, where f is an $\mathbb{R}^d$-valued function and $\alpha \in \mathbb{R}^d$. More precisely, this algorithm is used when f can be rewritten in the form $f(x) = \mathbb{E}[F(x, X)]$. Then the solution of the equation $f(x) = \alpha$ can be recursively approximated by

$$U_{n+1} = U_n - \gamma_n \left(F\left(U_n, X_{n+1}\right) - \alpha\right),$$

where $\gamma_n \geq 0$ is a deterministic sequence going to zero and (X_n) is a sequence of independent and identically distributed random variables. Obviously, as mentioned in the introduction, Robbins–Monro algorithm is a particular case of model (1.3).

3.1. Translation parameter

Let (Y_n) be a sample from a distribution F on $\mathbb{R}$, with density f with respect to the Lebesgue measure. The sequence of random variables (Y_n) is not observable, but each Y_n is supposed to be centered. We can only observe a translation model (X_n) where $X_n = Y_n + \theta$. The real translation parameter θ is unknown. Since $\mathbb{E}[Y_1] = 0$, the parameter θ is the mean of X_n, $\theta = \mathbb{E}[X_n]$.

Without knowing the function f, we may also assume that f is even, strictly positive and continuously differentiable. Let us define the recursive estimator of θ defined by

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n - \gamma_n \left(\mathbb{1}_{\{X_{n+1} \leq \widehat{\theta}_n\}} - \frac{1}{2} \right). \tag{3.1}$$

This algorithm is a particular case of the model (1.3) with $\sigma_n = \gamma_n$, $R_{n+1} = 0$ and

$$h(z) = \frac{1}{2} - \mathbb{E}\left[\mathbb{1}_{\{X_{n+1} \leq z\}} | \mathcal{F}_n \right],$$

where $\mathcal{F}_n$ is the σ-field of events prior to n and $\mathbb{F} \stackrel{\text{def}}{=} (\mathcal{F}_n)$ is the natural filtration. Indeed, (3.1) can be rewritten as

$$\widehat{\theta}_{n+1} = \widehat{\theta}_n + \gamma_n h\left(\widehat{\theta}_n \right) + \gamma_n \varepsilon_{n+1},$$

with $\varepsilon_{n+1} = \mathbb{E}\left[\mathbb{1}_{\{X_{n+1} \leq \widehat{\theta}_n\}} | \mathcal{F}_n \right] - \mathbb{1}_{\{X_{n+1} \leq \widehat{\theta}_n\}}$. Obviously, assumptions $(H4)$ and (2.3) for all integer p are satisfied, as well as $(H3)$ with $\sigma^2 = 1/4$. It remains to trust $(H2)$. Since X_{n+1} is independent of $\mathcal{F}_n$, one has

$$h(z) = \frac{1}{2} - \mathbb{P}\left(X_{n+1} \leq z \right) = \frac{1}{2} - \mathbb{P}\left(Y_{n+1} \leq z - \theta \right).$$

The function f is continuous and strictly positive hence h is differentiable and $h'(z) = -f(z - \theta)$. Moreover θ is a zero of h. Taylor-Young formula leads to

$$h(z) - h(\theta) = -f(0)(z - \theta) + \mathcal{O}\left((z - \theta)^2 \right),$$

with $f(0) > 0$. Thus $(H2)$ holds. Consequently, for any integer p, assuming $(H1)$ and choosing the steps (γ_n) such that $(H5)$ holds, Theorem 2.2 gives the asymptotic results on the estimation errors. The sequence of estimates $(\widehat{\theta}_n)$ satisfies

$$\lim_{n \to \infty} \frac{1}{s_n} \sum_{k=1}^n \frac{1}{\gamma_k^{p-1}} \left(\widehat{\theta}_k - \theta \right)^{2p} = \frac{\Sigma^{2p}(2p)!}{2^p p!} \quad \text{a.s.}$$

$$\lim_{n \to \infty} \frac{1}{s_n} \sum_{k=1}^n \frac{1}{\gamma_k^{p-\frac{3}{2}}} \left(\widehat{\theta}_k - \theta \right)^{2p-1} = 0 \quad \text{a.s.},$$

with $\Sigma^2 = \left[4\big(2f(0) - \alpha\xi\big) \right]^{-1}$ and the ASCLT for $(\widehat{\theta}_n)$ holds:

$$\frac{1}{s_n} \sum_{k=1}^n \gamma_k \delta_{\frac{1}{\sqrt{\gamma_k}} \left(\widehat{\theta}_k - \theta \right)} \Longrightarrow \mathcal{N}\left(0, \Sigma^2 \right) \quad \text{a.s.}$$

3.2. Recursive estimation of quantiles

More generally, with a procedure of estimation analogous to the preceding one, we can give asymptotic convergence results on the error of quantile's estimation. We consider a sample (Y_n) with a strictly increasing distribution function F. Let us define q the quantile of order δ, that is to say $\delta \stackrel{\text{def}}{=} F(q)$. Without knowing the function F, we may also assume that the density $f = F'$ is continuously differentiable.

Let us define the recursive estimator

$$\widehat{q}_{n+1} = \widehat{q}_n - \gamma_n \left(\mathbb{1}_{\{Y_{n+1} \leq \widehat{q}_n\}} - \delta \right). \tag{3.2}$$

Again, the sequence $(\widehat{q}_n)$ is a particular case of (1.3) with $\sigma_n = \gamma_n$, $R_{n+1} = 0$,

$$h(z) = \delta - \mathbb{E}\left[\mathbb{1}_{\{Y_{n+1} \leq z\}} | \mathcal{F}_n \right] = \delta - F(z),$$

and $\varepsilon_{n+1} = \mathbb{E}\left[\mathbb{1}_{\{Y_{n+1}\leq\widehat{q}_n\}}|\mathcal{F}_n\right] - \mathbb{1}_{\{Y_{n+1}\leq\widehat{q}_n\}}$. Obviously, assumptions $(H4)$ and (2.3) for all integer p are satisfied, as well as $(H3)$ with $\sigma^2 = \delta(1-\delta)$. Moreover, the function h clearly satisfies $(H2)$ for the zero q of h. Consequently, for any integer p, assuming $(H1)$ and choosing the steps such that $(H5)$ holds, Theorem 2.2 leads to

$$\lim_{n\to\infty} \frac{1}{s_n} \sum_{k=1}^{n} \frac{1}{\gamma_k^{p-1}} (\widehat{q}_k - q)^{2p} = \frac{\Sigma^{2p}(2p)!}{2^p p!} \quad \text{a.s.}$$

$$\lim_{n\to\infty} \frac{1}{s_n} \sum_{k=1}^{n} \frac{1}{\gamma_k^{p-\frac{3}{2}}} (\widehat{q}_k - q)^{2p-1} = 0 \quad \text{a.s.,}$$

with $\Sigma^2 = \delta(1-\delta)/[2f(q) - \alpha\xi]$. Moreover we have

$$\frac{1}{s_n} \sum_{k=1}^{n} \gamma_k \delta_{\frac{1}{\sqrt{\gamma_k}}(\widehat{q}_k - q)} \Longrightarrow \mathcal{N}\left(0, \Sigma^2\right) \quad \text{a.s.}$$

3.3. Recursive estimation of the mean

Let (Y_n) be a sequence of independent and identically distributed real random variables with mean μ and variance σ^2. Set an integer $p \geq 1$. Assume that for some real $m > 2p$, (Y_n) satisfies the moment condition

$$\sup_{n\geq 0} \mathbb{E}\left[|Y_{n+1}|^m|\mathcal{F}_n\right] < \infty \quad \text{a.s.}$$

Let us consider the recursive estimator of the mean μ defined by

$$\widehat{\mu}_{n+1} = \widehat{\mu}_n + \gamma_n \left(Y_{n+1} - \widehat{\mu}_n\right).$$

This model is again a particular case of the model (1.3) with $h(z) = \mu - z$ and $\varepsilon_{n+1} = Y_{n+1} - \mu$. Assumptions $(H2)$, $(H3)$, $(H4)$ and (2.3) are satisfied. Consequently, assuming $(H1)$ and choosing the steps such that $(H5)$ holds, the cumulated estimation error satisfies

$$\lim_{n\to\infty} \frac{1}{s_n} \sum_{k=1}^{n} \frac{1}{\gamma_k^{p-1}} (\widehat{\mu}_k - \mu)^{2p} = \frac{\Sigma^{2p}(2p)!}{2^p p!} \quad \text{a.s.}$$

$$\lim_{n\to\infty} \frac{1}{s_n} \sum_{k=1}^{n} \frac{1}{\gamma_k^{p-\frac{3}{2}}} (\widehat{\mu}_k - \mu)^{2p-1} = 0 \quad \text{a.s.,}$$

with $\Sigma^2 = \sigma^2/(2 - \alpha\xi)$, as well as the ASCLT

$$\frac{1}{s_n} \sum_{k=1}^{n} \gamma_k \delta_{\frac{1}{\sqrt{\gamma_k}}(\widehat{\mu}_k - \mu)} \Longrightarrow \mathcal{N}\left(0, \Sigma^2\right) \quad \text{a.s.}$$

4. Proof of Theorem 2.2

Remark 4.1. Assumption $(H5)$ implies that $v_n \in \mathcal{GS}(-\alpha + 2\beta)$, since $v_n = \gamma_n \sigma_n^{-2}$, $\gamma_n \in \mathcal{GS}(-\alpha)$ and $\sigma_n \in \mathcal{GS}(-\beta)$. Consequently one has

$$1 - \frac{v_{n-1}}{v_n} = \frac{-\alpha + 2\beta}{n} + o\left(\frac{1}{n}\right) = \xi\gamma_n\left(-\alpha + 2\beta\right) + o\left(\gamma_n\right).$$

Finally it comes

$$\frac{v_{n-1}}{v_n} = 1 - \gamma_n\xi(2\beta - \alpha) + o\left(\gamma_n\right). \tag{4.1}$$

This remark is very helpful for the proof of Theorem 2.2 and (4.1) is widely used.

First, the convergence is established for even moments, by induction on the power p. The recursive equation (1.3) yields

$$Z_{n+1} - z^* = Y_{n+1} + \sigma_n \varepsilon_{n+1} \quad \text{where} \quad Y_{n+1} \stackrel{\text{def}}{=} Z_n - z^* + \gamma_n \left[h(Z_n) + R_{n+1} \right]. \tag{4.2}$$

In view of assumptions $(H2)$ and $(H4)$, the sequence Y_{n+1} satisfies

$$Y_{n+1} = (Z_n - z^*)\left[1 + H\gamma_n\right] + \gamma_n r_{n+1} + \mathcal{O}\left(\gamma_n |Z_n - z^*|^2\right).$$

In order to prove Theorem 2.2, we shall apply several times the following lemma, established in Mokkadem and Pelletier [25].

Lemma 4.2. *Assuming that there exists $m > 2$ such that (2.3) holds, under assumptions $(H1)$ to $(H5)$, the stepsized stochastic algorithm (Z_n) satisfies*

$$|Z_n - z^*| = \mathcal{O}\left(\sqrt{v_n^{-1} \log s_n}\right) \quad a.s. \tag{4.3}$$

In particular, Lemma 4.2 consequently leads to

$$Y_{n+1} = (Z_n - z^*)\left[1 + H\gamma_n\right] + o\left(\gamma_n v_n^{-1/2} (\log s_n)^{-q}\right) \quad \text{a.s.} \tag{4.4}$$

for all $q \geq 0$.

4.1. Even moments

Elevating (4.2) to the power $2p$ it comes

$$(Z_{n+1} - z^*)^{2p} = \sum_{\ell=0}^{2p} C_{2p}^\ell \sigma_n^\ell \varepsilon_{n+1}^\ell Y_{n+1}^{2p-\ell}.$$

Setting $V_{n+1} \stackrel{\text{def}}{=} \sqrt{v_n}(Z_{n+1} - z^*)$, we get

$$V_{n+1}^{2p} - V_n^{2p} = v_n^p Y_{n+1}^{2p} - V_n^{2p} + \gamma_n^p \varepsilon_{n+1}^{2p} + \sum_{\ell=1}^{2p-1} C_{2p}^\ell \gamma_n^{\ell/2} \left(\sqrt{v_n} Y_{n+1}\right)^{2p-\ell} \varepsilon_{n+1}^\ell. \tag{4.5}$$

From the obvious equality

$$V_{n+1}^{2p} = \sum_{k=1}^n \left(V_{k+1}^{2p} - V_k^{2p}\right) + V_1^p,$$

and from the recursive equation (4.5), one gets for any $n \geq 1$,

$$V_{n+1}^{2p} = V_1^{2p} + \mathcal{A}_{n+1}^{(p)} + \mathcal{W}_{n+1}^{(p)} + \sum_{\ell=1}^{2p-1} C_{2p}^\ell \mathcal{B}_{n+1}^{(p)}(\ell), \tag{4.6}$$

where

$$\mathcal{W}_{n+1}^{(p)} \stackrel{\text{def}}{=} \sum_{k=1}^n \gamma_k^p \varepsilon_{k+1}^{2p}, \quad \mathcal{A}_{n+1}^{(p)} \stackrel{\text{def}}{=} \sum_{k=1}^n \left(v_k^p Y_{k+1}^{2p} - V_k^{2p}\right),$$

$$\mathcal{B}_{n+1}^{(p)}(\ell) \stackrel{\text{def}}{=} \sum_{k=1}^n \gamma_k^{\ell/2} \left(\sqrt{v_k} Y_{k+1}\right)^{2p-\ell} \varepsilon_{k+1}^\ell.$$

4.1.1. Case $p = 1$

Clearly equation (4.6) can be rewritten as

$$\mathcal{A}_{n+1}^{(1)} = -\mathcal{W}_{n+1}^{(1)} + V_{n+1}^2 - V_1^2 - 2\mathcal{B}_{n+1}^{(1)}(1). \tag{4.7}$$

Since the noise (ε_n) satisfies $(H3)$ and assumption (2.3) for a real number $m > 2$, let us note that Chow's lemma (see *e.g.* Duflo [10], p. 22) yields

$$\lim_{n \to \infty} \frac{1}{s_n} \mathcal{W}_{n+1}^{(1)} = \sigma^2 \quad \text{a.s.} \tag{4.8}$$

Let us assume that

$$\frac{1}{s_n} \left[V_{n+1}^2 - V_1^2 - 2\mathcal{B}_{n+1}^{(1)}(1) \right] = o(1) + o\left(s_n^{-1} \mathcal{A}_{n+1}^{(1)} \right) \quad \text{a.s.} \tag{4.9}$$

We first show how the combination of (4.8) and (4.9) gives the convergence (2.4) for $p = 1$. From equality (4.4) one has

$$Y_{n+1}^2 = (Z_n - z^*)^2 \left(1 + 2\gamma_n H + o(\gamma_n) \right) + o\left(\gamma_n^2 v_n^{-1} (\log s_n)^{-2q} \right)$$

$$+ o\left(|Z_n - z^*| \gamma_n v_n^{-1/2} (\log s_n)^{-q} \right)$$

$$= (Z_n - z^*)^2 \left(1 + 2\gamma_n H + o(\gamma_n) \right) + o\left(v_n^{-1} \gamma_n (\log s_n)^{-q+1/2} \right),$$

the second equality resulting from Lemma 4.2. Consequently, applying this equality for $q = 1/2$ and denoting $U_n \overset{\text{def}}{=} \sqrt{v_n}(Z_n - z^*)$, the general term of $\mathcal{A}_{n+1}^{(1)}$ is of the form

$$U_n^2 \left(1 - \frac{v_{n-1}}{v_n} + 2\gamma_n H + o(\gamma_n) \right) + o(\gamma_n) \quad \text{a.s.}$$

Hence we deduce from (4.1) by applying Kronecker lemma that

$$\lim_{n \to \infty} \frac{1}{s_n} \sum_{k=1}^n \gamma_k U_k^2 = \frac{1}{2H + \xi(2\beta - \alpha)} \lim_{n \to \infty} \frac{1}{s_n} \mathcal{A}_{n+1}^{(1)} \quad \text{a.s.} \tag{4.10}$$

Then the combination of (4.7)–(4.10) leads to (2.4).

It remains to prove (4.9). Obviously since s_n increases to infinity, almost surely $V_1^{2p} = o(s_n)$. Moreover, from Lemma 4.2, we have a.s. $V_{n+1}^2 = o(s_n)$. To conclude we have to study the asymptotic behavior of $\mathcal{B}_{n+1}^{(1)}(1)$. We easily deduce from Kronecker lemma and from the decomposition (4.4) that we only have to show that

$$\sum_{k=1}^n \sqrt{\gamma_k} U_k \varepsilon_{k+1} = o\left(\mathcal{A}_{n+1}^{(1)} \right) \quad \text{a.s.} \quad \text{and} \quad \sum_{k=1}^n \gamma_k^{3/2} \varepsilon_{k+1} = o(s_n) \quad \text{a.s.}$$

The standard strong law of large numbers for regressive series (see *e.g.* Duflo [10], p. 26) yields

$$\left| \sum_{k=1}^n \sqrt{\gamma_k} U_k \varepsilon_{k+1} \right| = o\left(\sum_{k=1}^n \gamma_k U_k^2 \right) + \mathcal{O}(1) \quad \text{a.s.}$$

$$\left| \sum_{k=1}^n \gamma_k^{3/2} \varepsilon_{k+1} \right| = \mathcal{O}\left(\sum_{k=1}^n \gamma_k^3 \right) = o(s_n) \quad \text{a.s.}$$

Since from (4.10), one has

$$o\left(\sum_{k=1}^n \gamma_k v_k (Z_k - z^*)^2 \right) = o\left(\mathcal{A}_{n+1}^{(1)} \right) \quad \text{a.s.,}$$

it comes

$$\mathcal{B}_{n+1}^{(1)}(1) = o\left(\mathcal{A}_{n+1}^{(1)}\right) + o\left(s_n\right) \quad \text{a.s.,} \tag{4.11}$$

which concludes the proof of (4.9).

4.1.2. Case $p \geq 2$

Now, let $p \geq 2$ and assume that the convergence (2.4) holds for any power q with $1 \leq q \leq p - 1$. First, we prove that

$$\lim_{n \to \infty} \frac{1}{s_n} \sum_{k=1}^{n} \gamma_k U_k^{2p} = \frac{-\Sigma^2}{p\sigma^2} \lim_{n \to \infty} \frac{1}{s_n} \mathcal{A}_{n+1}^{(p)} \quad \text{a.s.} \tag{4.12}$$

From the decomposition (4.4) together with the result (4.3) of Lemma 4.2, it is easy to see that, for all $q \geq -p + 1/2$

$$Y_{n+1}^{2p} = (Z_n - z^*)^{2p} \left(1 + 2pH\gamma_n + o\left(\gamma_n\right)\right) + o\left(\gamma_n v_n^{-p} \left(\log s_n\right)^{-q}\right).$$

In particular for $q = 0$, one has

$$v_n^p Y_{n+1}^{2p} - V_n^{2p} = v_n^p (Z_n - z^*)^{2p} \left[1 - \frac{v_{n-1}^p}{v_n^p} + 2pH\gamma_n + o\left(\gamma_n\right)\right] + o\left(\gamma_n\right) \quad \text{a.s.}$$

Since $v_n \in \mathcal{GS}(2\beta - \alpha)$, the sequence v_n^p belongs to $\mathcal{GS}\left(p(2\beta - \alpha)\right)$ and hence

$$1 - \frac{v_{n-1}^p}{v_n^p} = p(2\beta - \alpha)\xi\gamma_n + o\left(\gamma_n\right).$$

Finally, we get

$$v_n^p Y_{n+1}^{2p} - V_n^{2p} = p\gamma_n U_n^{2p} \left[\xi(2\beta - \alpha) + 2H\right] \left(1 + o(1)\right) + o\left(\gamma_n\right),$$

and (4.12) is a straightforward application of Kronecker lemma. Thus, we have to show that

$$\lim_{n \to \infty} \frac{1}{s_n} \mathcal{A}_{n+1}^{(p)} = -p\sigma^2(2p - 1)g\left[2(p - 1)\right] \quad \text{a.s.} \tag{4.13}$$

The proof of (4.13) is constructed in the following way. Let us note that Lemma 4.2 implies that $V_{n+1}^{2p} = o(s_n)$. First we establish the convergence of $s_n^{-1}\mathcal{W}_{n+1}^{(p)}$ to zero. Then we show that

$$\forall \ell \in \{1, \ldots, 2p - 1\} \backslash \{2\} \quad \mathcal{B}_{n+1}^{(p)}(\ell) = o\left(\mathcal{A}_{n+1}^{(p)}\right) + o(s_n) \quad \text{a.s.}$$

$$\lim_{n \to \infty} \mathcal{B}_{n+1}^{(p)}(2) = g\left[2(p - 1)\right]\sigma^2 \quad \text{a.s.}$$

The decomposition (4.6) will lead to convergence (4.13).

Since the noise (ε_n) satisfies (2.3), by applying Chow's lemma it comes

$$\mathcal{W}_{n+1}^{(p)} = \mathcal{O}\left(\sum_{k=1}^{n} \gamma_k^p\right) \quad \text{a.s.}$$

In addition, γ_n converges almost surely to zero, hence $\gamma_n^p = o(\gamma_n)$ a.s. Consequently, one gets from Kronecker's Lemma

$$\mathcal{W}_{n+1}^{(p)} = o\left(s_n\right) \quad \text{a.s.} \tag{4.14}$$

For establishing the asymptotic behavior of $\mathcal{B}_{n+1}^{(p)}$, let us split it into two terms, $\mathcal{B}_{n+1}^{(p)}(\ell) = \mathcal{C}_{n+1}^{(p)}(\ell) + \mathcal{D}_{n+1}^{(p)}(\ell)$ with

$$\mathcal{C}_{n+1}^{(p)}(\ell) \stackrel{\text{def}}{=} \sum_{k=1}^{n} \gamma_k^{\ell/2} \left(\sqrt{v_k} Y_{k+1}\right)^{2p-\ell} e_{k+1}(\ell),$$

$$\mathcal{D}_{n+1}^{(p)}(\ell) \stackrel{\text{def}}{=} \sum_{k=1}^{n} \gamma_k^{\ell/2} \left(\sqrt{v_k} Y_{k+1}\right)^{2p-\ell} \mathbb{E}\left[\varepsilon_{k+1}^{\ell} | \mathcal{F}_k\right],$$

and $e_{n+1}(\ell) \stackrel{\text{def}}{=} \varepsilon_{n+1}^{\ell} - \mathbb{E}\left[\varepsilon_{n+1}^{\ell} | \mathcal{F}_n\right]$. First, we claim that for any $3 \le \ell \le 2p - 1$,

$$\left| \mathcal{D}_{n+1}^{(p)}(\ell) \right| = o(s_n) \quad \text{a.s.} \tag{4.15}$$

One can easily see from the Hölder inequality and from the moment assumption (2.3) that each moment $\left| \mathbb{E}\left[\varepsilon_{n+1}^{\ell} | \mathcal{F}_n\right] \right|$ is almost surely bounded and consequently

$$\left| \mathcal{D}_{n+1}^{(p)}(\ell) \right| = \mathcal{O}\left(\sum_{k=1}^{n} \gamma_k^{\ell/2} \left(\sqrt{v_k} Y_{k+1}\right)^{2p-\ell} \right) \quad \text{a.s.}$$

If ℓ is even, for all $q \ge 0$, (4.4) leads to,

$$\begin{aligned}
\gamma_n^{\ell/2} \left(\sqrt{v_n} Y_{n+1}\right)^{2p-\ell} &= \gamma_n^{\ell/2} \left[U_n(1 + o(1)) + o\left(\gamma_n (\log s_n)^{-q}\right) \right]^{2p-\ell} \\
&= \gamma_n^{\ell/2} \left[U_n^{2p-\ell}(1 + o(1)) + o\left(\gamma_n U_n^{2p-\ell-1} (\log s_n)^{-q}\right) \right] \\
&= \gamma_n^{\ell/2} U_n^{2p-\ell} [1 + o(1)] + o\left(\gamma_n^{\ell/2+1} (\log s_n)^{-q+p-\ell/2-1/2}\right).
\end{aligned}$$

Then the induction assumption together with Kronecker lemma yield

$$\sum_{k=1}^{n} \gamma_k^{\ell/2} \left(\sqrt{v_k} Y_{k+1}\right)^{2p-\ell} = o(s_n) \quad \text{a.s.}$$

Consequently the convergence (4.15) holds for ℓ even such that $3 \le \ell \le 2p - 1$. Otherwise in the case when ℓ is odd and $\ell \ge 3$ (since $\mathcal{D}_{n+1}(1) = 0$ is a trivial case), applying the Cauchy–Schwarz inequality it comes

$$\left| \mathcal{D}_{n+1}^{(p)}(\ell) \right|^2 = \mathcal{O}\left(\left(\sum_{k=1}^{n} \gamma_k \left(\sqrt{v_k} Y_{k+1}\right)^{2p-2} \right) \left(\sum_{k=1}^{n} \gamma_k^{\ell-1} \left(\sqrt{v_k} Y_{k+1}\right)^{2(p-\ell+1)} \right) \right) \quad \text{a.s.}$$

Again the induction assumption and Kronecker's lemma lead to (4.15). For the last term $\mathcal{D}_{n+1}^{(2)}(\ell)$, the convergence

$$\lim_{n \to \infty} \frac{1}{s_n} \mathcal{D}_{n+1}^{(p)}(2) = g[2(p-1)] \sigma^2 \tag{4.16}$$

is a consequence of the induction assumption and Kronecker lemma too.

It remains to study the asymptotic behavior of $\mathcal{C}_{n+1}^{(p)}(\ell)$ for any $1 \le \ell \le 2p - 1$. We claim that

$$\left| \mathcal{C}_{n+1}^{(p)}(\ell) \right| = o(s_n) + o\left(\mathcal{A}_{n+1}^{(p)}\right) \quad \text{a.s.} \tag{4.17}$$

We easily deduce from (4.4) and from Kronecker lemma that we only have to show that

$$\sum_{k=1}^{n} \gamma_k^{\ell/2} U_k^{2p-\ell} e_{k+1}(\ell) = o\left(s_n\right) + o\left(\mathcal{A}_{n+1}^{(p)}\right) \quad \text{a.s.}$$

$$\sum_{k=1}^{n} \gamma_k^{1+\ell/2} e_{k+1}(\ell) = o\left(s_n\right) \quad \text{a.s.}$$

According to the standard strong law of large numbers for regressive series one has, for any $\delta > 0$,

$$\left| \sum_{k=1}^{n} \gamma_k^{\ell/2} U_k^{2p-\ell} e_{k+1}(\ell) \right|^2 = o\left(\sum_{k=1}^{n} \gamma_k^{\ell} U_k^{2(2p-\ell)} \left[\log \sum_{k=1}^{n} \gamma_k^{\ell} U_k^{2(2p-\ell)} \right]^{1+\delta} \right) \quad \text{a.s.}$$

$$\left| \sum_{k=1}^{n} \gamma_k^{1+\ell/2} e_{k+1}(\ell) \right| = o\left(\sum_{k=1}^{n} \gamma_k^{2+\ell} \right) \quad \text{a.s.}$$

Since γ_n decreases to zero, it comes in particular

$$\left| \sum_{k=1}^{n} \gamma_k^{1+\ell/2} e_{k+1}(\ell) \right| = o\left(s_n\right) \quad \text{a.s.}$$

First, assume that $1 \leq \ell \leq p-1$. Since

$$\sum_{k=1}^{n} \gamma_k^{\ell} U_k^{2(2p-\ell)} = \mathcal{O}\left(\sup_{1 \leq k \leq n} U_k^{2p} \sum_{k=1}^{n} \gamma_k^{\ell} U_k^{2(p-\ell)} \right) \quad \text{a.s.,}$$

the conjonction of Lemma 4.2 and the induction assumption leads to

$$\left| \sum_{k=1}^{n} \gamma_k^{\ell/2} U_k^{2p-\ell} e_{k+1}(\ell) \right| = o(s_n) \quad \text{a.s.}$$

The particular case $\ell = p$ is also a consequence of both the strong law of large numbers and Kronecker lemma and one has

$$\sum_{k=1}^{n} \gamma_k^{p} U_k^{2p} = o\left(\mathcal{A}_{n+1}^{(p)} \right) \quad \text{a.s.}$$

Consequently Kronecker lemma leads to (4.17).

Now assume that $p < \ell \leq 2p-1$. According to Chow's lemma, one has

$$\left| \sum_{k=1}^{n} \gamma_k^{\ell/2} U_k^{2p-\ell} e_{k+1}(\ell) \right| = o\left(\nu_n(\ell)\right) \quad \text{a.s.} \quad \text{with} \quad \nu_n(\ell) \overset{\text{def}}{=} \sum_{k=1}^{n} \left| \gamma_k^{\ell/2} U_k^{2p-\ell} \right|^{2p/\ell}.$$

Applying Hölder inequality with the exponents ℓ/p and $\ell/(\ell-p)$, the induction assumption implies $\nu_n(\ell) = o(s_n)$, which concludes the proof of (4.17).

Finally the conjonction of the convergences (4.14)–(4.17), leads to

$$\lim_{n \to \infty} \frac{1}{s_n} \left[\mathcal{A}_{n+1}^{(p)} + \mathcal{W}_{n+1}^{(p)} + \sum_{\ell=1}^{2p-1} C_{2p}^{\ell} \mathcal{B}_{n+1}^{(p)}(\ell) \right] = \lim_{n \to \infty} \frac{1}{s_n} \mathcal{A}_{n+1}^{(p)} + C_{2p}^2 \sigma^2 g\left[2(p-1)\right] \quad \text{a.s.}$$

So, from the decomposition (4.6) we get

$$\lim_{n \to \infty} \frac{1}{s_n} \mathcal{A}_{n+1}^{(p)} = -C_{2p}^2 \sigma^2 g\left[2(p-1)\right] \quad \text{a.s.,}$$

which concludes the proof of (2.4).

4.2. Odd moments

Now let us deal with the odd moments with the proof of (2.5). Elevating equation (4.2) to power $2p-1$ and proceeding exactly as in the case of even moments, it comes

$$V_{n+1}^{2p-1} = V_1^{2p-1} + A_{n+1}^{(p)} + W_{n+1}^{(p)} + \sum_{\ell=1}^{2p-2} C_{2p-1}^{\ell} B_{n+1}^{(p)}(\ell), \qquad (4.18)$$

where

$$W_{n+1}^{(p)} \stackrel{\text{def}}{=} \sum_{k=1}^{n} \gamma_k^{p-1/2} \varepsilon_{k+1}^{2p-1}, \quad A_{n+1}^{(p)} \stackrel{\text{def}}{=} \sum_{k=1}^{n} \left(v_k^{p-1/2} Y_{k+1}^{2p-1} - V_k^{2p-1} \right),$$

$$B_{n+1}^{(p)}(\ell) \stackrel{\text{def}}{=} \sum_{k=1}^{n} \gamma_k^{\ell/2} \left(\sqrt{v_k} Y_{k+1} \right)^{2p-\ell-1} \varepsilon_{k+1}^{\ell}.$$

4.2.1. Case $p = 1$

For $p = 1$, the decomposition (4.18) can be rewritten as

$$A_{n+1}^{(1)} = V_{n+1} - V_1 - W_{n+1}^{(1)}. \qquad (4.19)$$

An application of the standard strong law of large numbers for regressive series leads to

$$\left| W_{n+1}^{(1)} \right| = o\left(s_n \right) \quad \text{a.s.}$$

Clearly $V_1 = o(s_n)$ since s_n increases to infinity and again from Lemma 4.2, one has $V_{n+1} = o(s_n)$ a.s. Hence we deduce from (4.19) that $A_{n+1} = o(s_n)$ a.s.

Moreover, due to the decomposition (4.4) the general term of A_{n+1} satisfies for all $q \geq 0$,

$$\sqrt{v_n} Y_{n+1} - V_n = U_n \left[1 + H\gamma_n - \sqrt{\frac{v_{n-1}}{v_n}} \right] + o\left(\gamma_n \left(\log s_n \right)^{-q} \right) \quad \text{a.s.} \qquad (4.20)$$

Since $v_n \in \mathcal{GS}(2\beta - \alpha)$, $\sqrt{v_n} \in \mathcal{GS}(\beta - \alpha/2)$ and consequently

$$1 - \sqrt{\frac{v_{n-1}}{v_n}} = \xi \gamma_n (\beta - \alpha/2) + o\left(\gamma_n \right).$$

In particular for $q = 0$, the equation (4.20) yields

$$\sqrt{v_n} Y_{n+1} - V_n = \gamma_n U_n \left[H + \xi(\beta - \alpha/2) \right] (1 + o(1)) + o\left(\gamma_n \right) \quad \text{a.s.},$$

which clearly leads to (2.5) from Kronecker lemma, since

$$\lim_{n \to \infty} \frac{1}{s_n} \sum_{k=1}^{n} \gamma_k U_k = \frac{1}{H + \xi(\beta - \alpha/2)} \lim_{n \to \infty} \frac{1}{s_n} A_{n+1}^{(1)} = 0 \quad \text{a.s.}$$

4.2.2. Case $p \geq 2$

Now let $p \geq 2$ and assume that convergence (2.5) holds for any power q with $1 \leq q \leq p-1$. The structure of the proof is the same as in the case of even moments. It consists in establishing the convergence $A_{n+1}^{(p)} = o(s_n)$ from the decomposition (4.18).

First, Chow's Lemma implies

$$W_{n+1}^{(p)} = \mathcal{O}\left(\sum_{k=1}^{n} \gamma_k^{p-1/2}\right) \quad \text{a.s.}$$

Since $p \geq 2$ and γ_n goes to zero, one gets

$$W_{n+1}^{(p)} = o\left(s_n\right) \quad \text{a.s.} \tag{4.21}$$

Moreover, clearly we have $V_1^p = o(s_n)$ a.s. and $V_{n+1}^p = o(s_n)$ a.s. from Lemma 4.2. Let us split $B_{n+1}^{(p)}(\ell)$ into two terms, $B_{n+1}^{(p)}(\ell) = C_{n+1}^{(p)}(\ell) + D_{n+1}^{(p)}(\ell)$ with

$$C_{n+1}^{(p)}(\ell) \stackrel{\text{def}}{=} \sum_{k=1}^{n} \gamma_k^{\ell/2} \left(\sqrt{v_k}Y_{k+1}\right)^{2p-\ell-1} e_{k+1}(\ell),$$

$$D_{n+1}^{(p)}(\ell) \stackrel{\text{def}}{=} \sum_{k=1}^{n} \gamma_k^{\ell/2} \left(\sqrt{v_k}Y_{k+1}\right)^{2p-\ell-1} \mathbb{E}\left[\varepsilon_{k+1}^\ell | \mathcal{F}_k\right].$$

For the first term $C_{n+1}^{(p)}$, we deduce from (4.4) and from Kronecker lemma that it suffices to show that

$$\sum_{k=1}^{n} \gamma_k^{\ell/2} U_k^{2p-\ell-1} e_{k+1}(\ell) = o\left(s_n\right) \quad \text{a.s.} \tag{4.22}$$

$$\sum_{k=1}^{n} \gamma_k^{1+\ell/2} e_{k+1}(\ell) = o\left(s_n\right) \quad \text{a.s.}$$

Once again, one gets from the strong law of large numbers that for all $\delta > 0$,

$$\left|\sum_{k=1}^{n} \gamma_k^{\ell/2} U_k^{2p-\ell-1} e_{k+1}(\ell)\right|^2 = o\left(\sum_{k=1}^{n} \gamma_k^\ell U_k^{2(2p-\ell-1)} \left[\log \sum_{k=1}^{n} \gamma_k^\ell U_k^{2(2p-\ell-1)}\right]^{1+\delta}\right) \quad \text{a.s.}$$

$$\left|\sum_{k=1}^{n} \gamma_k^{1+\ell/2} e_{k+1}(\ell)\right| = o\left(\sum_{k=1}^{n} \gamma_k^{2+\ell}\right) = o\left(s_n\right) \quad \text{a.s.}$$

First, assume that $\ell \geq p - 1$. The convergence (2.4) implies

$$\sum_{k=1}^{n} \gamma_k U_k^{2(2p-\ell-1)} = \mathcal{O}\left(s_n\right) \quad \text{a.s.}$$

Hence, (4.22) is proven since γ_n goes to zero. Now let $1 \leq \ell < p - 1$. Since

$$\sum_{k=1}^{n} \gamma_k^\ell U_k^{2(2p-\ell-1)} = \mathcal{O}\left(\sup_{1 \leq k \leq n} U_k^{2(p-\ell-1)} \sum_{k=1}^{n} \gamma_k^\ell U_k^{2p}\right) \quad \text{a.s.,}$$

the conjonction of Lemma 4.2 and convergence (2.4) leads to (4.22). Finally, we have established that

$$C_{n+1}^{(p)}(\ell) = o\left(s_n\right) \quad \text{a.s.} \tag{4.23}$$

Let us now study the asymptotic behavior of $D_{n+1}(\ell)$ in the following three cases.

Case 1. $\ell = 2$. It is easy to see from the decomposition (4.4) and assumption ($H3$) that the general term of $D_{n+1}^{(p)}(2)$ is of the form

$$\sigma^2 \gamma_k U_k^{2p-3} [1 + o(1)] + o(s_k) \quad \text{a.s.}$$

Then, the induction assumption together with Kronecker lemma lead to

$$\lim_{n \to \infty} \frac{1}{s_n} D_{n+1}^{(p)}(2) = 0 \quad \text{a.s.}$$

Case 2. ℓ is even and such that $4 \leq \ell \leq 2(p-1)$. Let us remind that for each ℓ, the conditional moment $|\mathbb{E}[\varepsilon_{n+1}^\ell | \mathcal{F}_n]|$ is almost surely bounded. Hence, one has

$$D_{n+1}^{(p)}(\ell) = \mathcal{O} \left(\sum_{k=1}^{n} \gamma_k^{\ell/2} \left(\sqrt{v_k} Y_{k+1} \right)^{2p-\ell-1} \right) \quad \text{a.s.}$$

Then the decomposition (4.4) together with the induction assumption imply that $D_{n+1}^{(p)}(\ell) = o(s_n)$ a.s.

Case 3. ℓ is odd and such that $3 \leq \ell \leq 2p - 3$. Again since for each ℓ, the conditional moment $|\mathbb{E}[\varepsilon_{n+1}^\ell | \mathcal{F}_n]|$ is almost surely bounded, one has

$$D_{n+1}^{(p)}(\ell) = \mathcal{O} \left(\sum_{k=1}^{n} \gamma_k^{\ell/2} \left(\sqrt{v_k} Y_{k+1} \right)^{2p-\ell-1} \right) \quad \text{a.s.}$$

Once again, (4.4) and (2.4) imply that $D_{n+1}^{(p)}(\ell) = o(s_n)$ a.s.

Finally, we have proven that

$$\lim_{n \to \infty} s_n^{-1} A_n^{(p)} = 0 \quad \text{a.s.}$$

Moreover, due to (4.4), the general term of $A_n^{(p)}$ satisfies

$$v_n^{p-1/2} Y_{n+1}^{2p-1} - V_n^{2p-1} = U_n^{2p-1} \left[1 - \frac{v_{n-1}^{p-1/2}}{v_n^{p-1/2}} + (2p-1) H \gamma_n + o(\gamma_n) \right] \quad \text{a.s.}$$

Since $v_n \in \mathcal{GS}(2\beta - \alpha)$, one has $v_n^{p-1/2} \in \mathcal{GS}\big((2\beta - \alpha)(p - 1/2)\big)$ and consequently

$$1 - \frac{v_{n-1}^{p-1/2}}{v_n^{p-1/2}} = \xi \gamma_n \left(p - \frac{1}{2} \right) (2\beta - \alpha) + o(\gamma_n) \quad \text{a.s.}$$

Then it comes

$$v_n^{p-1/2} Y_{n+1}^{2p-1} - V_n^{2p-1} = \left(p - \frac{1}{2} \right) U_n^{2p-1} \gamma_n \left[2H + \xi(2\beta - \alpha) \right] (1 + o(1)) + o(\gamma_n) \quad \text{a.s.,}$$

which leads to (2.5) by applying Kronecker Lemma.

References

[1] A. Benveniste, M. Métivier and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, New York, *Appl. Math.* **22** (1990).

[2] B. Bercu, On the convergence of moments in the almost sure central limit theorem for martingales with statistical applications. *Stoc. Proc. Appl.* **111** (2004) 157–173.

[3] B. Bercu and J.-C. Fort, A moment approach for the almost sure central limit theorem for martingales. *Stud. Sci. Math. Hung.* (2006).

[4] B. Bercu, P. Cènac and G. Fayolle, On the almost sure central limit theorem for vector martingales: Convergence of moments and statistical applications. *J. Appl. Probab.* **46** (2009) 151–169.

[5] G.A. Brosamler, An almost everywhere central limit theorem. *Math. Proc. Cambridge Philos. Soc.* **104** (1988) 213–246.

[6] F. Chaâbane, Version forte du théorème de la limite centrale fonctionnel pour les martingales. *C. R. Acad. Sci. Paris Sér. I Math.* **323** (1996) 195–198.

[7] F. Chaâbane, Invariance principles with logarithmic averaging for martingales. *Stud. Sci. Math. Hung.* **37** (2001) 21–52.

[8] F. Chaâbane and F. Maâouia, Théorèmes limites avec poids pour les martingales vectorielles. *ESAIM: PS* **4** (2000) 137–189 (electronic).

[9] F. Chaâbane, F. Maâouia and A. Touati, Génèralisation du théorème de la limite centrale presque-sûr pour les martingales vectorielles. *C. R. Acad. Sci. Paris Sér. I Math.* **326** (1998) 229–232.

[10] M. Duflo, *Random Iterative Methods*. Springer-Verlag (1997).

[11] P. Dupuis and H.J. Kushner, Stochastic approximation and large deviations: Upper bounds and w.p.l convergence. *SIAM J. Control Optim.* **27** (1989) 1108–1135.

[12] W. Feller, *An introduction to probability theory and its applications* II. John Wiley, New York (1966).

[13] P. Hall and C.C. Heyde, *Martingale Limit Theory and Its Application*. Academic Press, New York, NY (1980).

[14] V. Koval and R. Schwabe, Exact bounds for the rate of convergence of stochastic approximation procédures. *Stoc. Anal. Appl.* **16** (1998) 501–515.

[15] H.J. Kushner and D.S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer-Verlag, Berlin (1978).

[16] M. Lacey and W. Phillip, A note on the almost sure central limit theorem. *Stat. Probab. Lett.* **9** (1990) 201–205.

[17] D. Lamberton and G. Pagès, Recursive computation of the invariant distribution of a diffusion. *Bernoulli* **8** (2002) 367–405.

[18] D. Lamberton and G. Pagès, Recursive computation of the invariant distribution of a diffusion: the case of a weakly mean reverting drift. *Stoch. Dyn.* **3** (2003) 435–451.

[19] A. Le Breton, About the averaging approach schemes for stochastic approximations. *Math. Methods Stat.* **2** (1993) 295–315.

[20] A. Le Breton and A. Novikov, Averaging for estimating covariances in stochastic approximation. *Math. Methods Stat.* **3** (1994) 244–266.

[21] A. Le Breton and A. Novikov, Some results about averaging in stochastic approximation. *Metrika* **42** (1995) 153–171.

[22] M.A. Lifshits, Lecture notes on almost sure limit theorems. *Publications IRMA* **54** (2001) 1–25.

[23] M.A. Lifshits, Almost sure limit theorem for martingales, in *Limit theorems in probability and statistics* II (Balatonlelle, 1999). János Bolyai Math. Soc., Budapest (2002) 367–390.

[24] L. Ljung, G. Pflug and H. Walk, *Stochastic Approximation and Optimization of Random Systems*. Birkhäuser, Boston (1992).

[25] A. Mokkadem and M. Pelletier, A companion for the Kiefer–Wolfowitz–Blum stochastic approximation algorithm. *Ann. Stat.* (2007).

[26] M. Pelletier, On the almost sure asymptotic behaviour of stochastic algorithms. *Stoch. Proc. Appl.* **78** (1998) 217–244.

[27] M. Pelletier, An almost sure central limit theorem for stochastic approximation algorithms. *J. Multivar. Anal.* **71** (1999) 76–93.

[28] H. Robbins and S. Monro, A stochastic approximation method. *Ann. Math. Stat.* **22** (1951) 400–407.

[29] P. Schatte, On strong versions of central limit theorem. *Math. Nachr.* **137** (1988) 249–256.

[30] Y. Zhu, Asymptotic normality for a vector stochastic difference equation with applications in stochastic approximation. *J. Multivar. Anal.* **57** (1996) 101–118.