Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT®
ELSEVIER
Stochastic Processes and their Applications 105 (2003) 299–313
stochastic processes and their applications
www.elsevier.com/locate/spa

A law of the iterated logarithm for stochastic approximation procedures in d-dimensional Euclidean space

Valery Koval[a,1], Rainer Schwabe[b,*]

[a] *Department of Mathematics, Zhytomyr Institute of Engineering and Technology, Cherniakhovsky str. 103, Zhytomyr 10005, Ukraine*

[b] *Department of Medical Biometry, University of Tübingen, Westbahnhofstr. 55, Tübingen D-72 070, Germany*

Received 4 June 2002; received in revised form 26 November 2002; accepted 27 December 2002

Abstract

In this paper, we investigate the rate of convergence for general d-dimensional stochastic approximation procedures and present an explicit expression for the asymptotic bounds in the law of the iterated logarithm.

Keywords: Stochastic approximation; Law of the iterated logarithm; Rate of convergence

1. Introduction

Let $\mathbf{R}^d$ be the d-dimensional Euclidean space with corresponding norm $\|x\| = \sqrt{x^\top x}$, where "$\top$" denotes transposition. We consider the nonlinear stochastic difference equation

$$X_n = X_{n-1} - a_n(F(X_{n-1}) + \Psi_n) + b_n V_n, \qquad n \geqslant 1, \; X_0 \in \mathbf{R}^d, \tag{1}$$

where $(a_n,\ n \geqslant 1)$ is a sequence of real numbers satisfying

$$a_n > 0, \qquad \lim_{n\to\infty} a_n = 0 \quad \text{and} \quad \sum_{n=1}^{\infty} a_n = \infty. \tag{2}$$

* Corresponding author. Institute for Mathematical Stochastics, Otto von Guericke University, PF 4120, Magdeburg D-39 016, Germany.

E-mail address: rainer.schwabe@mathematik.uni-magdeburg.de (R. Schwabe).

1 Partly supported by a research grant of the German Academic Exchange Service (DAAD).

doi:10.1016/S0304-4149(03)00016-4

Moreover, $(b_n,\ n \geqslant 1)$ is a sequence of real numbers, $(V_n,\ n \geqslant 1)$ and $(\Psi_n,\ n \geqslant 1)$ are sequences of d-dimensional random vectors such that $\|\Psi_n\| \to 0$ almost surely (a.s.) as $n \to \infty$. Finally, we assume that the equation $F(x) = 0$ has a solution θ in $\mathbf{R}^d$.

The recursive scheme (1) includes, in particular, the well-known Robbins–Monro and Kiefer–Wolfowitz procedures (see e.g. Nevelson and Khasminskii, 1976; Kushner and Yin, 1997).

In this paper, we investigate the rate of convergence of X_n towards θ as $n \to \infty$. For $d > 1$ and some special choices of the sequences $(a_n,\ n \geqslant 1)$ and $(b_n,\ n \geqslant 1)$ bounds were obtained for the rate of convergence by Ruppert (1982), Korostelev (1983, 1984), and Pelletier (1998). Koval (1998) proved that under rather general assumptions on (1) there exists a constant $L \in (0, \infty)$ such that

$$\limsup_{n \to \infty} \left(\frac{a_n}{b_n^2 \ell_n} \right)^{1/2} \|X_n - \theta\| = L \quad \text{a.s.}, \tag{3}$$

where $\ell_n = \log(\sum_{j=1}^n a_j)$. Here $\log t = \ln \max\{t, e\}$ for $t \geqslant 0$ where ln denotes the natural logarithm. Lower and upper bounds for L were given by Koval (1998), too. In the present paper we establish the exact value of L. Note that in the case $d = 1$ this problem was solved by Koval and Schwabe (1998).

Under very general assumptions on (1) the investigation of relation (3) reduces to the analogous relation for the solution $(Y_n,\ n \geqslant 1)$ of the corresponding linear stochastic difference equation

$$Y_n = (I - a_n A) Y_{n-1} + b_n V_n, \qquad n \geqslant 1, \ \ Y_0 = 0 \tag{4}$$

(see Koval, 1998, Lemma 2), where I is the identity matrix and A is a $d \times d$ matrix such that $F(x) = A(x - \theta) + \mathrm{o}(\|x - \theta\|)$ as $x \to \theta$ and the real parts $re\, \lambda_i(A)$ of its eigenvalues $\lambda_1(A), \ldots, \lambda_d(A)$ satisfy the condition

$$\lambda^* = \min_{1 \leqslant i \leqslant d} \operatorname{re} \lambda_i(A) > 0. \tag{5}$$

The remainder of the paper is organized as follows: A law of the iterated logarithm (LIL) is given for the linear recursive equation (4) in Section 2 (Theorem 1). The proof of Theorem 1 is based on a LIL for the matrix-normalized sums of independent random vectors which is proved in Section 3. Theorem 1 is then proved in Section 4. Finally, the general nonlinear recursive scheme (1) is investigated in Section 5.

In the sequel we will use the following notations: For any matrix H the supremum norm $\|\cdot\|_2$ of H is defined by $\|H\|_2 = \sqrt{\lambda_{\max}(H^\top H)}$, where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue of the corresponding symmetric matrix. In particular, if H is a positive semi-definite symmetric matrix, then this norm reduces to $\|H\|_2 = \lambda_{\max}(H)$. We write $H > 0$ if the symmetric matrix H is positive definite. By c we denote a positive constant which may vary from line to line.

2. A LIL for the linear equation (4)

In this section, we consider the linear recursive equation (4). We assume that the sequences $(a_n,\ n \geqslant 1)$ and $(b_n,\ n \geqslant 1)$ satisfy the following conditions:

$$b_n^2 / a_n \geqslant b_{n+1}^2 / a_{n+1}, \qquad n \geqslant 1, \tag{6}$$

$$\lim_{n\to\infty} \frac{1}{a_n}\left(1-\frac{b_n^2/a_n}{b_{n-1}^2/a_{n-1}}\right) = \tau < 2\lambda^*, \tag{7}$$

$$\limsup_{n\to\infty} \frac{|b_n - b_{n+1}|}{a_{n+1}|b_n|} < \infty. \tag{8}$$

Theorem 1. *Assume that conditions* (2) *and* (5)–(8) *are satisfied. If there exists a sequence of independent, identically distributed Gaussian random vectors* $(\Gamma_n,\ n \geqslant 1)$ *with zero expectation,* $\mathbf{E}(\Gamma_n) = 0$, *and covariance matrix* $\mathbf{E}(\Gamma_n \Gamma_n^\top) = D > 0,\ n \geqslant 1$, *such that*

$$\left\| \sum_{i=1}^{n} V_i - \sum_{i=1}^{n} \Gamma_i \right\| = \mathrm{o}((\ell_n/a_n)^{1/2}) \quad a.s.\ (n \to \infty), \tag{9}$$

then

$$\limsup_{n\to\infty} \left(\frac{a_n}{b_n^2 \ell_n}\right)^{1/2} \|Y_n\| = \left(2 \left\| \int_0^\infty \mathrm{e}^{((\tau/2)I - A)t} D \mathrm{e}^{((\tau/2)I - A^\top)t}\, \mathrm{d}t \right\|_2 \right)^{1/2} \quad a.s. \tag{10}$$

Consider the assumptions of Theorem 1. Conditions (6)–(8) are discussed in Koval and Schwabe (1998). The regularity condition $D > 0$ is commonly assumed (see Korostelev, 1984, p. 161; Pelletier, 1998). Condition (9) means that the sequence $(V_n,\ n \geqslant 1)$ satisfies an almost sure invariance principle with an approximation error of order $\mathrm{o}((\ell_n/a_n)^{1/2})$. This allows for a reduction of the problem to prove (10) for the Gaussian case only (cf. Lemma 3). On the other hand a.s. invariance principles are known to hold for broad classes of random sequences of various types, such as martingale difference sequences, mixing sequences, etc. (see Philipp (1986) for a survey on this topic). Therefore, assumption (9) makes it possible to consider various classes of sequences $(V_n,\ n \geqslant 1)$ simultaneously. It is worth noticing that the proofs of a couple of strong limit theorems make use of assumptions of type (9) (see, for example, Lacey and Philipp, 1990; Horváth, 1993).

In view of condition (9) stronger assumptions have to be imposed on the accuracy Δ_n in the approximation of the partial sum $\sum_{i=1}^n V_i$ by its Gaussian counterpart $\sum_{i=1}^n \Gamma_i$, if the steplengths $(a_n,\ n \geqslant 1)$ decrease at a slower rate, as $n \to \infty$, e.g.

(a) $\Delta_n = \mathrm{o}((n \log\log n)^{1/2})$ if $a_n = 1/n$,
(b) $\Delta_n = \mathrm{o}((n^a \log n)^{1/2})$ if $a_n = 1/n^a,\ 0 < a < 1$,
(c) $\Delta_n = \mathrm{o}((\log n)^{(1+\alpha)/2})$ if $a_n = 1/(\log n)^\alpha,\ \alpha > 1$.

In the above cases (a), (b), and (c), besides others, results were given, e.g. by Einmahl (1989), Berkes and Philipp (1979), Kuelbs and Philipp (1980), Morrow and Philipp (1982), and Eberlein (1986) for the approximation of $\sum_{i=1}^n V_i$ by $\sum_{i=1}^n \Gamma_i$ in $\mathbf{R}^d$ for various classes of sequences $(V_n,\ n \geqslant 1)$.

3. Auxiliary results

Let $(Z_n,\ n \geqslant 1)$ be a sequence of independent Gaussian random vectors in $\mathbf{R}^d$ with zero expectation, $\mathbf{E}(Z_n) = 0,\ n \geqslant 1$, and let $(A_n,\ n \geqslant 1)$ be an arbitrary sequence of $p \times d$ matrices ($A_n \neq 0,\ n \geqslant 1$). Denote by $S_n = \sum_{i=1}^n Z_i$ the partial sums and by $B_n = \mathbf{E}(S_n S_n^\top) = \sum_{i=1}^n \mathbf{E}(Z_i Z_i^\top)$ the corresponding covariance matrices, $n \geqslant 1$. Further, denote by $(e_n,\ n \geqslant 1)$ a sequence of normalized eigenvectors of the matrix $A_n B_n A_n^\top$ associated with the largest eigenvalue $q_n = \|A_n B_n A_n^\top\|_2$, i.e. $A_n B_n A_n^\top e_n = q_n e_n$ and $\|e_n\| = 1,\ n \geqslant 1$.

For the partial sums S_n of independent random vectors we establish next the LIL under matrix normalization which is of particular interest on its own (for more detailed information see Buldygin and Solntsev, 1997).

Proposition 1. *Let $B_{n_0} > 0$ for some $n_0 \geqslant 1$ and*

$$\|B_n\|_2 \to \infty \quad (n \to \infty). \tag{11}$$

Assume that there exists a sequence $(f_n,\ n \geqslant 1)$ of positive numbers such that

$$\limsup_{n\to\infty} \frac{f_{n+1}}{f_n} < \infty, \tag{12}$$

$$\log\log f_n \sim \log\log \|B_n\|_2 \quad (n \to \infty) \tag{13}$$

and

$$\frac{e_m^\top A_m B_m A_n^\top e_n}{(q_m q_n)^{1/2}} \leqslant c\,\frac{f_m}{f_n} \tag{14}$$

for all $n > m \geqslant n_0$. Then

$$\limsup_{n\to\infty} \frac{\|A_n S_n\|}{(2 q_n \log\log \|B_n\|_2)^{1/2}} = 1 \quad a.s. \tag{15}$$

For the proof of Proposition 1 we need the following additional lemma.

Lemma 1. *Let $(\xi(n),\ n \geqslant 1)$ be a real-valued Gaussian sequence with zero expectation, $\mathbf{E}(\xi(n)) = 0$, non-vanishing variances $\sigma^2(n) = \mathbf{E}(\xi^2(n)) > 0$ and correlations $r(m,n) = \mathbf{E}(\xi(m)\xi(n))/(\sigma(m)\sigma(n))$. Assume that there exists a sequence $(f(n),\ n \geqslant 1)$ of positive numbers such that*

$$f(n) \to \infty \quad (n \to \infty), \tag{16}$$

$$\limsup_{n\to\infty} f(n+1)/f(n) < \infty \tag{17}$$

and for all $n > m \geqslant 1$

$$|r(m,n)| \leqslant c f(m)/f(n). \tag{18}$$

Then

$$\limsup_{n\to\infty} \frac{|\xi(n)|}{(2\sigma^2(n)\log\log f(n))^{1/2}} \geqslant 1 \quad a.s.$$

Proof. Define a subsequence $(n_i,\ i \geqslant 1)$ of indices recursively by

$$n_1 = 1, \quad n_{i+1} = \min\{j\colon\ j > n_i,\ f(n_i)/f(j) < q\}, \quad i \geqslant 1 \tag{19}$$

for some fixed $q \in (0,1)$. Note that this subsequence is infinite by (16). To establish the present lemma it is sufficient to prove the assertion for the subsequence $(n_i,\ i \geqslant 1)$, i.e.

$$\limsup_{i\to\infty} \frac{|\xi(n_i)|}{(2\sigma^2(n_i)\log\log f(n_i))^{1/2}} \geqslant 1 \quad \text{a.s.} \tag{20}$$

Let $\eta(i) = \xi(n_i)/\sigma(n_i)$ and note that $(\eta(i),\ i \geqslant 1)$ is a normalized Gaussian sequence, $\mathbf{E}(\eta(i)) = 0$, $\mathbf{E}(\eta^2(i)) = 1$, with correlation $r_\eta(i,j) = \mathbf{E}(\eta(i)\eta(j)) = r(n_i, n_j)$. By (18) and (19) we have $|r_\eta(i,j)| < cq^{j-i}$ for all $j > i \geqslant 1$. This implies that $\lim_{k\to\infty}\sup_{|j-i|\geqslant k} |r_\eta(i,j)| = 0$ and, hence, by Lai (1973, Theorem 2)

$$\limsup_{i\to\infty} \frac{|\eta(i)|}{(2\log i)^{1/2}} = 1 \quad \text{a.s.} \tag{21}$$

It follows from (19) and (17), respectively, that $f(n_i - 1) \leqslant q^{-1} f(n_{i-1})$, $i \geqslant 2$, and $f(n) \leqslant c_1 f(n-1)$, $n > n_0$, for some $c_1 > 1$. This implies $f(n_i) \leqslant c_2^i f(n_0)$, $i \geqslant 1$, where $c_2 = c_1 q^{-1} > 1$. Consequently,

$$\limsup_{i\to\infty} \frac{|\eta(i)|}{(2\log\log f(n_i))^{1/2}} \geqslant 1 \quad \text{a.s.,}$$

in view of (21) which proves assertion (20) for the subsequence $(n_i,\ i \geqslant 1)$. □

Proof of Proposition 1. Set $t_n = (2\log\log\|B_n\|_2)^{1/2}$, $n \geqslant 1$. By (11) and Theorem 2 in Koval (2002) we have

$$\limsup_{n\to\infty} \frac{\|A_n S_n\|}{q_n^{1/2} t_n} \leqslant 1 \quad \text{a.s.} \tag{22}$$

In order to establish (15) we have to prove

$$\limsup_{n\to\infty} \frac{\|A_n S_n\|}{q_n^{1/2} t_n} \geqslant 1 \quad \text{a.s.,} \tag{23}$$

in view of (22). For this set $\zeta_n = e_n^\top A_n S_n$, $n \geqslant 1$. Note that $(\zeta_n,\ n \geqslant 1)$ is a Gaussian sequence with zero expectation, $\mathbf{E}(\zeta_n) = 0$, variances $\mathbf{E}(\zeta_n^2) = q_n$ and corresponding coefficients of correlation

$$r_\zeta(m,n) = \frac{e_m^\top A_m B_m A_n^\top e_n}{(q_m q_n)^{1/2}}, \quad n > m \geqslant 1.$$

By (11), (12) and (14) it follows from Lemma 1 that

$$\limsup_{n\to\infty} \frac{|e_n^\top A_n S_n|}{(q_n 2\log\log f_n)^{1/2}} \geqslant 1 \quad \text{a.s.} \tag{24}$$

By the Cauchy–Schwarz inequality we have $\|A_n S_n\| \geqslant |e_n^\top A_n S_n|$ and assertion (23) follows from (13) and (24) which completes the proof. □

Lemma 2. *Assume that conditions* (2) *and* (5) *are satisfied. Then, for any* $\delta \in (0, \lambda^*)$, *there exist constants* c_δ *and* k_0 *such that for all* $n \geqslant k \geqslant k_0$

$$\left\| \prod_{j=k}^{n} (I - a_j A) \right\|_2 \leqslant c_\delta \prod_{j=k}^{n} (1 - a_j)^{\lambda^* - \delta}. \tag{25}$$

Proof. There exists an invertible matrix T such that $\Lambda = T^{-1}AT$ is a block diagonal matrix of r square Jordan blocks J_i of order $d_j \geqslant 1$ $(1 \leqslant j \leqslant r)$ with sum d. Let $\delta \in (0, \lambda^*)$ be fixed and consider the diagonal matrix $Q = \mathrm{diag}(1, \delta/4, (\delta/4)^2, \dots, (\delta/4)^{d-1})$. For every complex matrix $B = (b_{ij})$ of dimension $d \times d$ we define by $\|B\|_\infty = \max_{1 \leqslant i \leqslant d} \sum_{j=1}^{d} |b_{ij}|$ the norm of maximal row sums. Then, for all $n \geqslant k \geqslant 1$ we have

$$\begin{aligned} \left\| \prod_{j=k}^{n} (I - a_j A) \right\|_2 &\leqslant \sqrt{d} \|TQ\|_2 \|(TQ)^{-1}\|_2 \left\| \prod_{j=k}^{n} (I - a_j Q^{-1} T^{-1} A T Q) \right\|_\infty \\ &\leqslant c_\delta \prod_{j=k}^{n} \left\| (I - a_j Q^{-1} \Lambda Q) \right\|_\infty . \end{aligned} \tag{26}$$

Denote by $\rho = \max_{1 \leqslant i \leqslant d} |\lambda_i(A)|^2$ the squared maximal eigenvalue of the matrix A. Then

$$\begin{aligned} \left\| (I - a_j Q^{-1} \Lambda Q) \right\|_\infty &\leqslant \max_{1 \leqslant i \leqslant d} |1 - a_j \lambda_i(A)| + a_j \delta/4 \\ &\leqslant \sqrt{1 - 2 a_j \lambda^* + a_j^2 \rho} + a_j \delta/4. \end{aligned} \tag{27}$$

In view of $\lim_{j \to \infty} a_j = 0$ we have

$$\lim_{j \to \infty} \frac{1 - (1 - a_j)^{\lambda^* - \delta}}{a_j} = \lambda^* - \delta.$$

Thus, there exists k_0 such that for all $j \geqslant k_0$

$$a_j < \delta/(2\rho) \tag{28}$$

and

$$\frac{1 - (1 - a_j)^{\lambda^* - \delta}}{a_j} < \lambda^* - \frac{\delta}{2}. \tag{29}$$

By (28) and (29) it follows from (27) that for all $j \geqslant k_0$

$$\begin{aligned} \|(I - a_j Q^{-1} \Lambda Q)\|_\infty &\leqslant \sqrt{1 - 2 a_j (\lambda^* - \delta/4)} + a_j \delta/4 \\ &\leqslant 1 - a_j(\lambda^* - \delta/4) + a_j \delta/4 = 1 - a_j(\lambda^* - \delta/2) \\ &\leqslant (1 - a_j)^{\lambda^* - \delta}. \end{aligned}$$

This, together with (26), implies (25). □

4. Proof of Theorem 1

In view of conditions (6) and $\lim_{n\to\infty} a_n = 0$ we may assume without loss of generality that $a_n < \min\{1,\ 1/(2\|A\|_2)\}$ and $b_n \neq 0$ for all $n \geqslant 1$. We adopt the convention $\prod_{j=n+1}^{n}(\cdot) = 1$ or I, respectively, for matrices.

In order to simplify the presentation of the proof we start with two lemmas.

First, consider Eq. (4) when the sequence $(V_n,\ n \geqslant 1)$ is replaced by $(\Gamma_n,\ n \geqslant 1)$ defined in Theorem 1, i.e.

$$Y_n^* = (I - a_n A)Y_{n-1}^* + b_n \Gamma_n, \quad n \geqslant 1, \ Y_0^* = 0. \tag{30}$$

Lemma 3. *If conditions* (2), (5) *and* (7)–(9) *are satisfied, then*

$$\|Y_n - Y_n^*\| = o((b_n^2 \ell_n / a_n)^{1/2}) \quad a.s.\ (n \to \infty). \tag{31}$$

Proof. Denote by S_n and S_n^* the partial sums of the V_i and Γ_i, respectively, $S_n = \sum_{i=1}^n V_i$ and $S_n^* = \sum_{i=1}^n \Gamma_i$. We use Abel's identity to obtain

$$\begin{aligned}
Y_n - Y_n^* &= \sum_{k=1}^{n} b_k \prod_{j=k+1}^{n} (I - a_j A)(V_k - \Gamma_k) \\
&= \sum_{k=1}^{n-1} \left[b_k \prod_{j=k+1}^{n} (I - a_j A) - b_{k+1} \prod_{j=k+2}^{n} (I - a_j A) \right] \\
&\quad \times (S_k - S_k^*) + b_n (S_n - S_n^*) \\
&= \sum_{k=1}^{n-1} \prod_{j=k+2}^{n} (I - a_j A) b_k a_{k+1} \left(\frac{b_k - b_{k+1}}{a_{k+1} b_k} I - A \right) \\
&\quad \times (\ell_k / a_k)^{1/2} [(S_k - S_k^*)(a_k / \ell_k)^{1/2}] \\
&\quad + b_n (S_n - S_n^*).
\end{aligned}$$

Let $\delta_n = (a_n/\ell_n)^{1/2} \|S_n - S_n^*\|,\ n \geqslant 1$, and note that

$$\delta_n \to 0 \quad \text{a.s. } (n \to \infty) \tag{32}$$

in view of (9). Next we make use of condition (8) and obtain

$$\begin{aligned}
\|Y_n - Y_n^*\| &\leqslant \sum_{k=1}^{n-1} \left\| \prod_{j=k+2}^{n} (I - a_j A) \right\|_2 |b_k| a_{k+1} \\
&\quad \times \left(\left| \frac{b_k - b_{k+1}}{a_{k+1} b_k} \right| + \|A\|_2 \right) (\ell_k / a_k)^{1/2} \delta_k \\
&\quad + (b_n^2 \ell_n / a_n)^{1/2} \delta_n.
\end{aligned} \tag{33}$$

It follows from (7) (see Koval and Schwabe, 1998) that for some v, $0 < v < 2$, and for all sufficiently large n

$$\frac{b_n^2}{a_n} \geqslant \frac{b_{n-1}^2}{a_{n-1}} (1-\lambda^* a_n)^v. \tag{34}$$

Now we set

$$\sigma = (\lambda^* - \delta) - (\lambda^* + \varepsilon)v/2, \tag{35}$$

where $\delta \in (0, \lambda^*)$ and $\varepsilon > 0$ are chosen sufficiently small such that $\sigma > 0$. By Lemma 2 it follows from (2) and (5) that

$$\left\| \prod_{j=k}^{n} (I - a_j A) \right\|_2 \leqslant c \prod_{j=k}^{n} (1-a_j)^{\lambda^* - \delta} \tag{36}$$

for all $n \geqslant k \geqslant 1$. It follows from (35) and (2) that

$$\prod_{j=k}^{n} (1-a_j)^{\lambda^*-\delta} \leqslant \prod_{j=k}^{n} (1-a_j)^{\sigma} \exp\left\{ -\frac{v}{2} (\lambda^* + \varepsilon) \sum_{j=k}^{n} a_j \right\}$$

$$\leqslant c \prod_{j=k}^{n} (1-a_j)^{\sigma} \prod_{j=k}^{n} (1-\lambda^* a_j)^{v/2}, \quad 1 \leqslant k \leqslant n. \tag{37}$$

By (36) and (37) it follows from (33) that

$$\|Y_n - Y_n^*\|$$

$$\leqslant c\,\ell_n^{1/2} \sum_{k=1}^{n-1} \left(|b_k| a_k^{-1/2} \prod_{j=k+2}^{n} (1-\lambda^* a_j)^{v/2} \right) a_{k+1} \prod_{j=k+2}^{n} (1-a_j)^{\sigma} \delta_k$$

$$+ (b_n^2 \ell_n / a_n)^{1/2} \delta_n.$$

Next, we make use of the relation $|b_k| a_k^{-1/2} \prod_{j=k+2}^{n} (1-\lambda^* a_j)^{v/2} \leqslant c |b_n| a_n^{-1/2}$, $1 \leqslant k \leqslant n-1$ which follows from (34) and obtain

$$\|Y_n - Y_n^*\| \leqslant c (b_n^2 \ell_n / a_n)^{1/2} (t_n + \delta_n), \tag{38}$$

where $t_n = \sum_{k=1}^{n-1} a_{k+1} \prod_{j=k+2}^{n} (1-a_j)^{\sigma} \delta_k$. Set $c_{nk} = a_{k+1} \prod_{j=k+2}^{n} (1-a_j)^{\sigma}$, $1 \leqslant k \leqslant n-1$. Then

$$c_{nk} \to 0 \quad (n \to \infty) \quad \text{for every } k \geqslant 1 \tag{39}$$

and

$$\sum_{k=1}^{n-1} c_{nk} = \sum_{k=1}^{n-1} \frac{a_{k+1}}{1-(1-a_{k+1})^{\sigma}} \left(\prod_{j=k+2}^{n} (1-a_j)^{\sigma} - \prod_{j=k+1}^{n} (1-a_j)^{\sigma} \right)$$

$$\to 1/\sigma \quad (n \to \infty) \tag{40}$$

by the Toeplitz lemma. Thus, again by the Toeplitz lemma, it follows from (32), (39) and (40) that $t_n = \sum_{k=1}^{n-1} c_{nk} \delta_k \to 0$ a.s. as $n \to \infty$. This, together with (32) and (38), implies relation (31) which completes the proof. □

Lemma 4. *If conditions* (2) *and* (5)–(7) *are satisfied, then*

$$\limsup_{n\to\infty}\left(\frac{a_n}{b_n^2\ell_n}\right)^{1/2}\|Y_n^*\| = \sqrt{2}M \quad a.s. \tag{41}$$

for the recursive equation (30), *where*

$$M = \left\|\int_0^\infty e^{((\tau/2)I-A)t}De^{((\tau/2)I-A^\top)t}\,dt\right\|_2^{1/2}.$$

Proof. Set $A_n=\prod_{j=1}^n(I-a_jA)$, $n\geqslant 1$. With this notation we may derive from (30) that $Y_n^*=A_n\sum_{k=1}^n b_kA_k^{-1}\Gamma_k$. Now we set $S_n=\sum_{k=1}^n b_kA_k^{-1}\Gamma_k$ and denote by $B_n=\mathbf{E}(S_nS_n^\top)=\sum_{k=1}^n b_k^2A_k^{-1}D(A_k^{-1})^\top$ and $Q_n=\mathbf{E}(Y_n^*(Y_n^*)^\top)$ the covariance matrices of S_n and Y_n^*, respectively. Note that $Y_n^*=A_nS_n$ and $Q_n=A_nB_nA_n^\top$. Moreover, let $q_n=\|Q_n\|_2$.

In order to establish relation (41) we will make use of Proposition 1. For this purpose we first prove the following asymptotic relations

$$\log\log\|B_n\|_2\sim\ell_n \quad (n\to\infty) \tag{42}$$

$$q_n\sim M^2b_n^2/a_n \quad (n\to\infty). \tag{43}$$

Consider first (42). Let $D^{1/2}$ denote the symmetric square root of D. With this notation we may obtain a lower bound

$$\begin{aligned}\|B_n\|_2 &\geqslant d^{-1}\operatorname{tr}B_n\\ &= d^{-1}\sum_{k=1}^n b_k^2\operatorname{tr}[A_k^{-1}D^{1/2}(A_k^{-1}D^{1/2})^\top]\\ &\geqslant d^{-1}\sum_{k=1}^n b_k^2\|A_k^{-1}D^{1/2}\|_2^2\\ &\geqslant \frac{\|D\|_2}{d}\sum_{k=1}^n b_k^2/\|A_k\|_2^2.\end{aligned}$$

Hence, by (36) and (6) we get

$$\begin{aligned}\|B_n\|_2 &\geqslant c\sum_{k=1}^n b_k^2\prod_{j=1}^k(1-a_j)^{2(\delta-\lambda^*)}\\ &= c\sum_{k=1}^n\frac{b_k^2}{a_k}\frac{a_k}{1-(1-a_k)^{2(\lambda^*-\delta)}}\left[\prod_{j=1}^k(1-a_j)^{2(\delta-\lambda^*)}-\prod_{j=1}^{k-1}(1-a_j)^{2(\delta-\lambda^*)}\right]\\ &\geqslant c\,\frac{b_n^2}{a_n}\prod_{j=1}^n(1-a_j)^{2(\delta-\lambda^*)}.\end{aligned}$$

It follows from (34) that $(b_n^2/a_n)\prod_{j=1}^n (1-\lambda^* a_j)^{-\nu} \geqslant c b_1^2/a_1,\ n \geqslant 1$. In view of (37) this finally implies

$$\|B_n\|_2 \geqslant c \exp\left(2\sigma \sum_{j=1}^n a_j\right), \quad n \geqslant 1. \tag{44}$$

Now we calculate an upper bound for $\|B_n\|_2$. It follows from (36) and (37) that

$$\begin{aligned} q_n &\leqslant \|D\|_2 \sum_{k=1}^n b_k^2 \left\| \prod_{j=k+1}^n (I - a_j A) \right\|_2^2 \\ &\leqslant c\|D\|_2 \sum_{k=1}^n \frac{b_k^2}{a_k} \prod_{j=k+1}^n (1-\lambda^* a_j)^{\nu} a_k \prod_{j=k+1}^n (1-a_j)^{2\sigma}. \end{aligned}$$

We make use of the inequality

$$\frac{b_k^2}{a_k} \prod_{j=k+1}^n (1-\lambda^* a_j)^{\nu} \leqslant c\,\frac{b_n^2}{a_n}, \quad 1 \leqslant k \leqslant n, \tag{45}$$

which follows from (34) and obtain

$$q_n \leqslant c(b_n^2/a_n) \sum_{k=1}^n a_k \prod_{j=k+1}^n (1-a_j)^{2\sigma} \leqslant c(b_n^2/a_n), \quad n \geqslant 1 \tag{46}$$

(see also (40)). By (46) and (6) we have

$$\begin{aligned} \|B_n\|_2 &\leqslant \|Q_n\|_2\, \|A_n^{-1}\|_2^2 \\ &\leqslant c(b_n^2/a_n) \left\| \prod_{j=1}^n (I - a_j A)^{-1} \right\|_2 \\ &\leqslant c(b_1^2/a_1) \prod_{j=1}^n \|(I - a_j A)^{-1}\|_2 \\ &\leqslant c \prod_{j=1}^n \frac{1}{1 - a_j \|A\|_2} \\ &\leqslant c \exp\left\{ (\|A\|_2 + \tilde{\varepsilon}) \sum_{j=1}^n a_j \right\} \end{aligned} \tag{47}$$

for some $\tilde{\varepsilon} > 0$ and all sufficiently large n. The asymptotic relation (42) then follows from (44) and (47).

Next, we consider the second asymptotic relation (43). Set $G_n = (a_n/b_n^2)Q_n,\ n \geqslant 1$. Obviously the sequence $(G_n,\ n \geqslant 1)$ satisfies the recursive relation

$$G_n = \alpha_n (I - a_n A) G_{n-1} (I - a_n A^{\top}) + a_n D, \quad n \geqslant 1,$$

where $\alpha_n=(a_n/b_n^2)(b_{n-1}^2/a_{n-1})$, $n\geqslant 2$, $\alpha_1=a_1/b_1^2$, $G_0=0$. We can rewrite this recursion as

$$G_n=\left[I-a_n\left(A-\frac{\tau}{2}I\right)\right]G_{n-1}\left[I-a_n\left(A^\top-\frac{\tau}{2}I\right)\right]+a_nD+a_nU_n, \qquad (48)$$

where

$$U_n=\left(\frac{\alpha_n-1}{a_n}-\tau\right)G_{n-1}-\frac{\tau^2}{4}a_nG_{n-1}+\frac{\tau}{2}a_n(AG_{n-1}+G_{n-1}A^\top)$$
$$+(\alpha_n-1)[a_nAG_{n-1}A^\top-AG_{n-1}-G_{n-1}A^\top].$$

It follows from (7) that $\alpha_n\to 1$ and $(\alpha_n-1)/a_n\to\tau$ as n tends to infinity. Observe also that $\sup_{n\geqslant 1}\|G_n\|_2<\infty$ in view of (46). This and $\lim_{n\to\infty}a_n=0$ imply

$$\|U_n\|_2\to 0 \quad (n\to\infty). \qquad (49)$$

Set $\bar{A}=A-(\tau/2)I$ and note that by (7) the eigenvalues $\lambda_1(\bar{A}),\dots,\lambda_d(\bar{A})$ of the matrix $\bar{A}$ satisfy the condition

$$\bar{\lambda}=\min_{1\leqslant i\leqslant d}\operatorname{re}\lambda_i(\bar{A})=\lambda^*-\frac{\tau}{2}>0. \qquad (50)$$

Consider the slightly simpler recursive equation

$$G_n^*=(I-a_n\bar{A})G_{n-1}^*(I-a_n\bar{A}^\top)+a_nD, \quad n\geqslant 1, \qquad (51)$$

where $G_0^*=0$. From (48) and (51) we can derive that

$$\|G_n-G_n^*\|_2\leqslant\sum_{k=1}^n a_k\left\|\prod_{j=k+1}^n(I-a_j\bar{A})\right\|_2^2\cdot\|U_k\|_2.$$

By Lemma 2 it follows from (2) and (50) that $\|\prod_{j=k}^n(I-a_j\bar{A})\|_2^2\leqslant c\prod_{j=k}^n(1-a_j)^{2(\bar{\lambda}-\delta)}$, for all $n\geqslant k\geqslant 1$, where $0<\delta<\bar{\lambda}$. Hence, $\|G_n-G_n^*\|_2\leqslant c\sum_{k=1}^n a_k\prod_{j=k+1}^n(1-a_j)^{2(\bar{\lambda}-\delta)}$ $\|U_k\|_2$. This, together with (49) and the Toeplitz lemma implies

$$\|G_n-G_n^*\|_2\to 0 \quad (n\to\infty). \qquad (52)$$

Moreover, we have $G_n^*=\sum_{k=1}^n a_k\prod_{j=k+1}^n(I-a_j\bar{A})D\prod_{j=k+1}^n(I-a_j\bar{A}^\top)$, $n\geqslant 1$, from definition (51). By Zhu (1996, Lemmas 9 and 10) it follows from (2) and (50) that $\|G_n^*\|_2\to M^2$ and, hence, $\|G_n\|_2\to M^2$ in view of (52), as n tends to infinity, which completes the proof of the second asymptotic relation (43).

In view of relations (42) and (43) it is sufficient to check the conditions of Proposition 1 for Eq. (30) in order to establish the limiting behaviour (41). Obviously, $B_1=b_1^2A_1^{-1}D(A_1^{-1})^\top>0$ and condition (11) is fulfilled by (42) and (2). Set $f_n=q_n^{1/2}/\prod_{j=1}^n(1-a_j)^\gamma$, $n\geqslant 1$, where $\gamma=\lambda^*-\delta$ (see (35)). Then $\lim_{n\to\infty}f_{n+1}/f_n=1$ by (2), (7) and (43) which establishes condition (12).

By (43) we have $f_n\geqslant c(b_n/a_n^{1/2})\prod_{j=1}^n(1-a_j)^{-\gamma}$. Analogously to (44) we obtain $f_n\geqslant c\exp(\sigma\sum_{j=1}^n a_j)$, $n\geqslant 1$. Conversely, by (43) and (6) we have $f_n\leqslant c\prod_{j=1}^n(1-a_j)^{-\gamma}\leqslant c\exp\{\gamma(1+\tilde{\varepsilon})\sum_{j=1}^n a_j\}$ for some $\tilde{\varepsilon}>0$ and all sufficiently large n. Consequently, these two inequalities imply $\log\log f_n\sim\ell_n$ as n tends to infinity which, together with (42) establishes condition (13).

Finally, we get condition (14) by invoking inequality (36):

$$\frac{e_m^\top A_m B_m A_n^\top e_n}{(q_m q_n)^{1/2}} \leqslant \frac{q_m \|A_n A_m^{-1}\|_2}{(q_m q_n)^{1/2}}$$

$$\leqslant c\left(\frac{q_m}{q_n}\right)^{1/2} \prod_{j=m+1}^{n} (1-a_j)^\gamma$$

$$= c\,\frac{f_m}{f_n},$$

for all $n > m \geqslant 1$. □

Proof of Theorem 1. The result is immediate from Lemmas 3 and 4. □

5. A LIL for the nonlinear equation (1)

Theorem 2. *Assume that*

$$X_n \to \theta \quad a.s.\ (n \to \infty), \tag{53}$$

and either

$$\|\Psi_n\| = \mathrm{O}([b_n^2/a_n]^{1/2}) \quad a.s.\ (n \to \infty) \tag{54}$$

or

$$\|\Psi_n\| = \mathrm{o}(\|X_{n-1} - \theta\|) \quad a.s.\ (n \to \infty) \tag{55}$$

are satisfied. Assume further that

$$F(x) = A(x-\theta) + \mathrm{o}(\|x-\theta\|) \quad as\ x \to \theta \tag{56}$$

and that conditions (2) *and* (5)–(9) *are fulfilled. Then*

$$\limsup_{n\to\infty} \left(\frac{a_n}{b_n^2 \ell_n}\right)^{1/2} \|X_n - \theta\|$$

$$= \left(2\left\|\int_0^\infty \mathrm{e}^{((\tau/2)I - A)t} D \mathrm{e}^{((\tau/2)I - A^\top)t}\,\mathrm{d}t\right\|_2\right)^{1/2} \quad a.s. \tag{57}$$

Proof. By Lemma 2 in Koval (1998) it follows from (2), (7), (10) and (53) to (56) that $\|X_n - Y_n\| = \mathrm{o}((b_n^2 \ell_n / a_n)^{1/2})$ a.s., for $n \to \infty$, which, together with (10) implies the result (57). □

Remark. Note that the Lemma 2 in Koval (1998) is formulated for $\Psi \equiv 0$. However, its proof goes through if either condition (54) or (55) is satisfied. In fact, in the case of (54) we need additionally (45) which is ensured by condition (7).

We will give a short discussion of the assumptions imposed in Theorem 2. The consistency condition (53) is commonly assumed for investigations of the rate of convergence in the situation of a nonlinear response function F (see e.g. Mark, 1982, pp. 36–37; Korostelev, 1984, p. 161; Pelletier, 1998). Problem (53) has been investigated in great detail by Nevel'son and Khasminskii (1976), Korostelev (1984) and Kushner and Yin (1997), besides others. Concerning condition (56) it has been used by Korostelev (1984, pp. 160–161). Pelletier (1998) imposes a more stringent assumption. An analogous condition to (54) has been used, e.g. by Zhu (1996) in the central limit theorem (see also Nevel'son and Khasminskii, 1976, Chapter 4, Section 2). Conditions similar to (55) appear, e.g. in Korostelev (1984) and Pelletier (1998).

Next, we derive two corollaries of Theorem 2 for the Robbins–Monro and Kiefer–Wolfowitz procedures under special conditions on the sequences $(a_n,\ n \geqslant 1)$ and $(b_n,\ n \geqslant 1)$ (cf. Korostelev, 1984, pp. 160–162, and Pelletier, 1998).

Corollary 1. *Assume conditions* (53), (56) *and* (5). *Let*

$$a_n = 1/n, \qquad b_n = 1/n^{(1+\beta)/2}, \tag{58}$$

where $0 < \beta \leqslant 1$. *Further assume that* $\lambda^* > \beta/2$ *and either*

$$\|\Psi_n\| = \mathrm{O}(1/n^{\beta/2}) \quad a.s.\ (n \to \infty)$$

or (55) *is satisfied. If, additionally,* (9) *holds, i.e.*

$$\left\| \sum_{i=1}^{n} V_i - \sum_{i=1}^{n} \Gamma_i \right\| = \mathrm{o}((n \log\log n)^{1/2}) \quad a.s.\ (n \to \infty),$$

then

$$\limsup_{n\to\infty} \left(\frac{n^{\beta}}{\log\log n} \right)^{1/2} \|X_n - \theta\|$$

$$= \left(2 \left\| \int_0^{\infty} \mathrm{e}^{((\beta/2)I - A)t} D \mathrm{e}^{((\beta/2)I - A^{\top})t}\, \mathrm{d}t \right\|_2 \right)^{1/2} \quad a.s.$$

Corollary 2. *Assume conditions* (53), (56) *and* (5). *Let*

$$a_n = 1/n^{a}, \quad 0 < a < 1, \quad \text{and} \quad b_n = 1/n^{b}, \quad \text{where } a/2 < b \leqslant a. \tag{59}$$

Further assume that either

$$\|\Psi_n\| = \mathrm{O}(n^{(a/2)-b}) \quad a.s.\ (n \to \infty)$$

or (55) *is satisfied. If, additionally,* (9) *holds, i.e.*

$$\left\| \sum_{i=1}^{n} V_i - \sum_{i=1}^{n} \Gamma_i \right\| = \mathrm{o}((n^{a} \log n)^{1/2}) \quad a.s.\ (n \to \infty),$$

then

$$\limsup_{n\to\infty} \left(\frac{n^{2b-a}}{\log n} \right)^{1/2} \|X_n - \theta\| = \left(2(1-a) \left\| \int_0^{\infty} \mathrm{e}^{-At} D \mathrm{e}^{-A^{\top}t}\, \mathrm{d}t \right\|_2 \right)^{1/2} \quad a.s.$$

The proofs of the corollaries are immediate as each of (58) and (59) implies conditions (2) and (6)–(8) with $\tau = \beta$ and 0, respectively.

Remark. The restrictions $\beta \leqslant 1$ and $b \leqslant a$ in Corollaries 1 and 2, respectively, are associated with assumption (53) for the Robbins–Monro and Kiefer–Wolfowitz procedures (see e.g. Kushner and Yin, 1997).

We, finally, consider an application of assumption (9) to the classical Robbins–Monro procedure

$$X_n = X_{n-1} - a_n \Phi(X_{n-1}, \varepsilon_n), \quad n \geqslant 1,$$

for finding a solution θ of the equation $F(x)=0$, where $F(x)=E\Phi(x,\varepsilon_1)$ and $(\varepsilon_n, n \geqslant 1)$ is a sequence of independent, identically distributed random variables. Then V_n can be chosen as $V_n = \Phi(X_{n-1}, \varepsilon_n) - F(X_{n-1})$ in (1) and $(V_n, \mathscr{F}_n, n \geqslant 1)$ is a martingale difference sequence, where $\mathscr{F}_n$ is the σ-field generated by the $X_k, k \leqslant n$. Let $Z(x) = \Phi(x, \varepsilon_1) - F(x)$ and assume that

$$\lim_{x \to \theta} EZ^2(x) = \sigma^2 \tag{60}$$

and

$$\sup_x E|Z(x)|^{2+\delta} < \infty \quad \text{for some } \delta > 0. \tag{61}$$

From (53) and (60) it follows that

$$\lim_{n \to \infty} EV_n^2 = \lim_{n \to \infty} E(V_n^2 | \mathscr{F}_{n-1}) = \sigma^2 \quad \text{a.s.} \tag{62}$$

From Corollary 4.2 in Hall and Heyde (1980) (cf. formula (4.45) there) by standard arguments we obtain

$$\left| \sum_{i=1}^n V_i - \sum_{i=1}^n \Gamma_i \right| = o((n \log\log n)^{1/2}) \quad \text{a.s. } (n \to \infty),$$

by using (61) and (62) where $(\Gamma_n, n \geqslant 1)$ is a sequence of independent, identically distributed Gaussian random variables with $E\Gamma_n = 0, E\Gamma_n^2 = \sigma^2$. Then, assumption (9) is satisfied for $a_n = a/n$ with $a > 0$.

References

Berkes, I., Philipp, W., 1979. Approximation theorems for independent and weakly dependent random vectors. Ann. Probab. 7, 29–54.

Buldygin, V., Solntsev, S., 1997. Asymptotic Behaviour of Linearly Transformed Sums of Random Variables. Kluwer, Dordrecht.

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

Einmahl, U., 1989. Extensions of results of Komlós, Major and Tusnády to the multivariate case. J. Multivariate Anal. 28, 20–68.

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and its Application. Academic Press, New York.
Horváth, L., 1993. A note on the law of the iterated logarithm for Abel sums. C. R. Math. Rep. Acad. Sci. Canada 15, 213–217.
Korostelev, A.P., 1983. On upper functions for stochastic approximation procedures. Theory Probab. Appl. 28, 806–811.
Korostelev, A.P., 1984. Stochastic Recursive Procedures. Local Properties. Nauka, Moscow (in Russian).
Koval, V.A., 1998. Rate of convergence of stochastic approximation procedures in a Banach space. Cybernet. Systems Anal. 34, 386–394.
Koval, V., 2002. A new law of the iterated logarithm in R^d with application to matrix-normalized sums of random vectors. J. Theoret. Probab. 15, 249–257.
Koval, V., Schwabe, R., 1998. Exact bounds for the rate of convergence in general stochastic approximation procedures. Stochastic Anal. Appl. 16, 501–515.
Kuelbs, J., Philipp, W., 1980. Almost sure invariance principles for partial sums of mixing B-valued random variables. Ann. Probab. 8, 1003–1036.
Kushner, H.J., Yin, G.G., 1997. Stochastic Approximation Algorithms and Applications. Springer, New York.
Lacey, M.T., Philipp, W., 1990. A note on the almost sure central limit theorem. Statist. Probab. Lett. 9, 201–205.
Lai, T.L., 1973. Gaussian processes, moving averages and quick detection problems. Ann. Probab. 1, 825–837.
Mark, G., 1982. Loglog-invarianzprinzipien für Prozesse der stochastischen Approximation. Mitteil. Math. Sem. Giessen 153.
Morrow, G., Philipp, W., 1982. An almost sure invariance principle for Hilbert space valued matingales. Trans. Amer. Math. Soc 273, 231–251.
Nevel'son, M.B., Khasminskii, R.Z., 1976. Stochastic Approximation and Recursive Estimation. Amer. Math. Soc., Providence, RI.
Pelletier, M., 1998. On the almost sure asymptotic behaviour of stochastic algorithms. Stochastic Process. Appl. 78, 217–244.
Philipp, W., 1986. Invariance principles for independent and weakly dependent random variables. In: Eberlein, E., Taqqu, M.S. (Eds.), Dependence in Probability and Statistics, Vol. 11. Birkhäuser, Basel, pp. 225–268.
Ruppert, D., 1982. Almost sure approximations to the Robbins–Monro and Kiefer–Wolfowitz processes with dependent noise. Ann. Probab. 10, 178–187.
Zhu, Y.M., 1996. Asymptotic normality for a vector stochastic difference equation with applications in stochastic approximation. J. Multivar. Anal. 57, 101–118.