The Annals of Applied Probability
1997, Vol. 7, No. 2, 444–459

AN ALMOST SURE INVARIANCE PRINCIPLE FOR STOCHASTIC APPROXIMATION PROCEDURES IN LINEAR FILTERING THEORY[1]

BY ERICH BERGER

Universität Göttingen

In this paper we consider a class of stochastic approximation procedures that arises in linear filtering and regression theory. Our main result asserts that the stochastic approximation process satisfies an almost sure invariance principle (with a certain rate of convergence) if the partial sums of the errors do.

1. Introduction. Throughout this paper, the underlying probability space is denoted by $(\Omega, \mathscr{F}, P)$. The space $\mathbb{R}^d$ $(d \in \mathbb{N})$ will be assumed to be equipped with an arbitrary norm $\|\cdot\|$. The elements of $\mathbb{R}^d$ are considered as column vectors.

The stochastic algorithm considered in this paper arises in linear filtering and regression problems. It can be motivated in the following way.

Let $Y\colon \Omega \to \mathbb{R}$ and $Z\colon \Omega \to \mathbb{R}^d$ be two random variables such that

$$EY^2 < \infty \quad \text{and} \quad E\|Z\|^2 < \infty. \tag{1.1}$$

We are interested in determining a vector $\theta_0 \in \mathbb{R}^d$ such that $E|Y - \theta^T Z|^2$ becomes minimal by taking $\theta = \theta_0$. (Here, ^{T} denotes the transpose.) Then θ_0 is a solution of the Wiener–Hopf equation

$$\Gamma\theta - Q = 0, \tag{1.2}$$

where $Q := E(YZ)$ and $\Gamma \in \mathbb{R}^{d,d}$ (= set of all $d \times d$-matrices with real entries) is the unique symmetric (nonnegative-definite) matrix such that $\langle \Gamma t, t\rangle = E\langle Z, t\rangle^2$ for all $t \in \mathbb{R}^d$. (In case $EZ = 0$, Γ is the covariance matrix of Z.) The qualities Γ and Q are assumed to be unknown, but it is possible sequentially to observe random variables $\Gamma_k\colon \Omega \to \mathbb{R}^{d,d}$ and $Q_k\colon \Omega \to \mathbb{R}^d$ $(k \in \mathbb{N})$ having the following properties:

$$\frac{1}{n}\sum_{k=1}^{n} \Gamma_k \to \Gamma \quad \text{a.s. as } n \to \infty \tag{1.3}$$

and

$$\frac{1}{n}\sum_{k=1}^{n} Q_k \to Q \qquad \text{a.s. as } n \to \infty. \tag{1.4}$$

Received October 1995; revised October 1996.

[1]Research partially supported by Deutsche Forschungsgemeinschaft.

AMS 1991 *subject classifications*. Primary 62L20; secondary 60F17, 62M20.

Key words and phrases. Stochastic approximation, linear filtering, almost sure invariance principles.

Suppose, in addition to the above hypotheses, that

(1.5) the spectrum of Γ is contained in the set $\{z \in \mathbb{C}: \operatorname{Re} z > 0\}$

and that

$$\sup_{n\in\mathbb{N}} \frac{1}{n} \sum_{k=1}^{n} \|\Gamma_k\| < \infty \quad \text{a.s.} \tag{1.6}$$

(Starting from the given norm $\|\cdot\|$ on $\mathbb{R}^d$, the norm $\|A\|$ of a matrix $A \in \mathbb{R}^{d,d}$ is assumed to be defined by $\|A\| := \sup\{\|Ax\|: x \in \mathbb{R}^d, \ \|x\| \le 1\}$.) Then it is known that the Robbins–Monro type algorithm

$$X_{n+1} := X_n - n^{-1}(\Gamma_n X_n - Q_n), \quad n \in \mathbb{N}, \tag{1.7}$$

(with $X_1: \Omega \to \mathbb{R}^d$ being an arbitrary random variable) converges almost surely to the unique solution $\theta = \theta_0$ of equation (1.2) [cf. Walk and Zsidó (1989) and the references therein].

The natural next step is to ask for conditions ensuring that the sequence $(X_n)_{n\in\mathbb{N}}$ satisfies more refined limit theorems such as the central limit theorem, a Donsker type invariance principle, a law of the iterated logarithm or an almost sure invariance principle. To handle these problems, it is useful to rewrite (1.7) in the form

$$U_{n+1} = (I - n^{-1}\Gamma_n) U_n + n^{-1} V_n, \quad n \in \mathbb{N}, \tag{1.8}$$

where

$$U_n := X_n - \theta_0, \qquad V_n := Q_n - \Gamma_n \theta_0 \tag{1.9}$$

and I stands for the identity matrix in $\mathbb{R}^{d,d}$. Recursion formulas of the form (1.8), but with assumption (1.3) being replaced by

$$\Gamma_n \to \Gamma \quad \text{a.s. as } n \to \infty, \tag{1.10}$$

have been introduced by Fabian (1968) as an appropriate setting for proving asymptotic normality of various more classic stochastic approximation procedures, including the usual Robbins–Monro process and the Kiefer–Wolfowitz process. They have been further analyzed by (among others) Walk (1977), Mark (1982) and—in a rather systematic way—by the present author (1986). The last mentioned references contain weak invariance principles, almost sure invariance principles for the law of the iterated logarithm and also rates of convergence in the invariance principle.

In 1988, Walk established a weak invariance principle and an invariance principle for the law of the iterated logarithm for the process (1.8) under the weaker hypotheses (1.3) and (1.6) [instead of (1.10)]. More recently, Heunis (1994) and Kouritzin (1996) obtained almost sure invariance principles with rates of convergence for the process (1.8). Heunis imposed stationarity requirements and mixing conditions on the sequence $(\Gamma_n, V_n)_{n\in\mathbb{N}}$; Kouritzin's work is based on the assumption that the partial sums of the sequence $(\Gamma_n, V_n)_{n\in\mathbb{N}}$ satisfy an almost sure invariance principle.

Our objective here is to provide a general access to rates of convergence in almost sure invariance principles (and in particular to the just mentioned results of Heunis and Kouritzin) by suitably modifying the methods developed in Berger (1986). Roughly speaking, our main result (Theorem 2.1 below) asserts that rates of convergence in almost sure invariance principles for the process $(U_n)_{n\in\mathbb{N}}$ in (1.8) can be achieved if the following hold.

1. The convergence in (1.3) is sufficiently fast.
2. Rates of convergence in almost sure invariance principles for the partial sums of the "errors" V_n are available.

In contrast to Kouritzin's work, we do not demand that the partial sums of the sequence $(\Gamma_n)_{n\in\mathbb{N}}$ satisfy an almost sure invariance principle. In Theorem 2.1 it is assumed that the partial sums of the V_n's can be well approximated by Brownian motion. A somewhat extended version of Theorem 2.1 that has been motivated by Kouritzin's article and admits a wider class of approximating Gaussian processes (including those allowed by Kouritzin) will be described in the Appendix.

2. The result. Let $\Gamma \in \mathbb{R}^{d,d}$, and let $U_n,\ V_n\colon \Omega \to \mathbb{R}^d$ and $\Gamma_n\colon \Omega \to \mathbb{R}^{d,d}$ $(n \in \mathbb{N})$ be random variables satisfying a recursion formula of the form (1.8). We subject Γ, $(\Gamma_n)_{n\in\mathbb{N}}$ and $(V_n)_{n\in\mathbb{N}}$ to the following four conditions.

CONDITION 2.1. There is a constant $\gamma \in (\frac{1}{2}, \infty)$ such that

$$\|e^{-u\Gamma}\| < e^{-\gamma u} \quad \text{for all } u \in (0, \infty). \tag{2.1}$$

CONDITION 2.2. There is a constant $q \in (0, 1)$ such that

$$\sup_{n\in\mathbb{N}} \frac{1}{n^q} \left\| \sum_{k=1}^{n} (\Gamma_k - \Gamma) \right\| < \infty \quad \text{a.s.} \tag{2.2}$$

CONDITION 2.3.

$$\sup_{n\in\mathbb{N}} \frac{1}{n} \sum_{k=1}^{n} \|\Gamma_k\| < \infty \quad \text{a.s.} \tag{2.3}$$

CONDITION 2.4. For some $\eta \in (0, \min\{\frac{1}{2}, \gamma - \frac{1}{2}\})$, there is an $\mathbb{R}^d$-valued Brownian motion W with $W(0) = 0$ on $(\Omega, \mathscr{F}, P)$ such that

$$\left\| W(t) - \sum_{k\le t} V_k \right\| = O(t^{(1/2)-\eta}) \quad \text{a.s. for } t \ge 1. \tag{2.4}$$

REMARK 2.1. If the spectrum of Γ is contained in $\{z \in \mathbb{C}\colon \operatorname{Re} z > \frac{1}{2}\}$, then it is possible to find a constant $\gamma \in (\frac{1}{2}, \infty)$ and a norm on $\mathbb{R}^d$ such that (2.1) holds with respect to the induced norm on $\mathbb{R}^{d,d}$. [Note that all norms on a finite-dimensional real vector space are equivalent.] Conversely, (2.1) entails that the spectrum of Γ is contained in the set $\{z \in \mathbb{C}\colon \operatorname{Re} z > \gamma\}$. Compare Daleckiĭ and Kreĭn (1974), page 29ff and Walk (1977, 1980).

THEOREM 2.1. *Under Conditions* 2.1–2.4, *one has*

$$\lim_{\substack{t\to\infty\\ t>0}} t^{-(1/2)+\sigma} \|[t]\, U_{[t+1]} - Y(t)\| = 0 \quad a.s. \tag{2.5}$$

for any $\sigma \in [0, \min\{\frac{1}{3}(1-q),\ \eta,\ \gamma - \frac{1}{2} - \eta\})$, *where the process* $(Y(t))_{t\in[0,\infty)}$ *is defined by*

$$Y(t) := W(t) - \Lambda \int_0^1 s^{\Lambda - I}\, W(st)\, ds \quad \text{for } t \in [0, \infty) \tag{2.6}$$

with $\Lambda := \Gamma - I$.

The proof of Theorem 2.1 will be carried out in Sections 3 and 4.

REMARK 2.2. The above theorem can be regarded as a partial extension of Corollary 2.24 in Berger (1986). The present partly more restrictive framework has mainly been chosen to keep technicalities to a minimum. As a matter of fact, the arguments utilized in our proof can easily be carried over to the general Banach space setting underlying the just mentioned corollary.

REMARK 2.3. It is obvious that (2.5) implies a Donsker type weak invariance principle for the sequence $(U_n)_{\in\mathbb{N}}$. Combining (2.5), (2.6) and Strassen's (1964) functional version of the law of the iterated logarithm for W, it is also possible to derive both the usual and the functional version of the law of the iterated logarithm for $(U_n)_{n\in\mathbb{N}}$, thus giving an exact description of the rate of almost sure convergence of $(U_n)_{n\in\mathbb{N}}$ [cf. Mark (1982)].

REMARK 2.4. Almost sure approximations of the form (2.4) have been obtained for several classes of weakly dependent sequences $(V_k)_{k\in\mathbb{N}}$ (more precisely, for sequences whose partial sums can be well approximated by martingales or by partial sums of independent random variables). [A convenient reference is Philipp's (1986) survey. The reader may also consult the reference list in Berger (1990).] Combining these invariance principles and our Theorem 2.1, it is easy to formulate more concrete invariance principles for $(U_n)_{n\in\mathbb{N}}$. In particular, Heunis' (1994) results can immediately be extended in various directions.

3. Auxiliary results. In the following two sections, the symbol C stands for a generic positive constant. The characterizing property of these constants is that they do not depend on the indices of the sequences involved.

The aim of this section is to adapt the estimates for operator products in Berger's (1986) Lemma 3.3 to the present situation. The corresponding results are given in Lemma 3.4 and Corollary 3.1. We start with three more elementary lemmas.

LEMMA 3.1 [Noncommutative partial summation; cf. Walk and Zsidó (1989), page 173]. *Let* $(A_k)_{k\in\mathbb{N}}$, $(B_k)_{k\in\mathbb{N}}$ *and* $(D_k)_{k\in\mathbb{N}}$ *be three sequences in* $\mathbb{R}^{d,d}$. *Then*

$$\begin{aligned}\sum_{m=1}^{n} A_m\, D_m\, B_m = A_n\, S_n\, B_n + A_n \sum_{l=1}^{n-1} S_l\,(B_l - B_{l+1})\\ + \sum_{m=1}^{n-1} (A_m - A_{m+1})\, S_m\, B_m\\ + \sum_{m=1}^{n-1} (A_m - A_{m+1}) \sum_{l=1}^{m-1} S_l\,(B_l - B_{l+1}),\end{aligned} \tag{3.1}$$

where $S_l := \sum_{k=1}^{l} D_k$ *for* $l \in \{1, \dots, n\}$.

PROOF. By partial summation,

$$\sum_{m=1}^{n} A_m\, D_m\, B_m = A_n \sum_{m=1}^{n} D_m\, B_m + \sum_{m=1}^{n-1} (A_m - A_{m+1}) \sum_{l=1}^{m} D_l\, B_l \tag{3.2}$$

and

$$\sum_{l=1}^{m} D_l\, B_l = S_m\, B_m + \sum_{k=1}^{m-1} S_k\,(B_k - B_{k+1}). \tag{3.3}$$

Combining (3.2) and (3.3) leads to (3.1). □

LEMMA 3.2. *Let* $(a_k)_{k\in\mathbb{N}}$ *be a sequence in* $[0, \infty)$ *such that*

$$\sup_{n\in\mathbb{N}} \frac{1}{n} \sum_{k=1}^{n} a_k < \infty.$$

Then, for any $\alpha \in \mathbb{R}$ *and any pair* $(k, n) \in \mathbb{N}^2$ *with* $k < n$,

$$\sum_{l=k+1}^{n} l^{\alpha}\, a_l \le \begin{cases} C\, n^{\alpha+1}, & \text{if } \alpha > -1,\\ C\left(1 + \log\left(\dfrac{n}{k}\right)\right), & \text{if } \alpha = -1,\\ C\, k^{\alpha+1}, & \text{if } \alpha < -1.\end{cases}$$

PROOF. Partial summation. □

Now let $\Gamma \in \mathbb{R}^{d,d}$, let $(\tilde{\Gamma}_j)_{j\in\mathbb{N}}$ be a sequence in $\mathbb{R}^{d,d}$, and suppose that the following three conditions are fulfilled:

(3.4) there is a constant $\gamma \in (0, \infty)$ such that (2.1) obtains;

$$\sup_{n\in\mathbb{N}} \frac{1}{n^q} \left\| \sum_{k=1}^{n} (\tilde{\Gamma}_k - \Gamma) \right\| < \infty \quad \text{for some constant } q \in (0, 1); \tag{3.5}$$

$$\sup_{n\in\mathbb{N}} \frac{1}{n} \sum_{k=1}^{n} \|\tilde{\Gamma}_k\| < \infty. \tag{3.6}$$

[The sequence $(\tilde{\Gamma}_j)_{j\in\mathbb{N}}$ should be envisaged as a typical path $(\Gamma_j(\omega))_{j\in\mathbb{N}}$ of the sequence $(\Gamma_j)_{j\in\mathbb{N}}$ introduced in Section 2.] Our purpose here is to investigate the products

$$A_{k,n} := (I - n^{-1}\tilde{\Gamma}_n) \cdot \cdots \cdot (I - k^{-1}\tilde{\Gamma}_k) \tag{3.7}$$

and

$$A^*_{k,n} := (I - n^{-1}\Gamma) \cdot \cdots \cdot (I - k^{-1}\Gamma) \tag{3.8}$$

for $(k, n) \in \mathbb{N}^2$ with $k \le n$. We also define

$$A_{k,k-1} := A^*_{k,k-1} := I \quad \text{for } k \in \mathbb{N}. \tag{3.9}$$

LEMMA 3.3. *Under the above hypotheses, one has*

$$\|A^*_{k,n}\| \le C\left(\frac{k}{n}\right)^{\gamma} \quad \textit{for all } (k, n) \in \mathbb{N}^2 \textit{ with } k \le n, \tag{3.10}$$

and there is a constant $\lambda \in \mathbb{R}$ *such that*

$$\|A_{k,n}\| \le C\left(\frac{k}{n}\right)^{\lambda} \quad \textit{for all } (k, n) \in \mathbb{N}^2 \textit{ with } k \le n. \tag{3.11}$$

PROOF. For (3.10), we refer to the proof of Lemma 3.3 in Berger (1986). To prove (3.11), we first notice that $\|A_{k,n}\| \le \exp\{\sum_{j=k}^{n} j^{-1}\|\tilde{\Gamma}_j\|\}$. Hence and from (3.6) and Lemma 3.2, we conclude that there is a constant $\lambda \in \mathbb{R}$ such that $\|A_{k,n}\| \le C(k/n)^{\lambda}$, as desired. □

LEMMA 3.4. *The notation and assumptions are as introduced before the statement of Lemma* 3.3. *Let* $\lambda \in (-\infty, \gamma)$ *be a constant with the following properties*:

$$\lambda \ne q, \qquad \gamma - \lambda \ne 1 - q \tag{3.12}$$

and

$$\|A_{k,n}\| \le C\left(\frac{k}{n}\right)^{\lambda} \quad \textit{for all } (k, n) \in \mathbb{N}^2 \textit{ with } k \le n. \tag{3.13}$$

Then

$$\|A_{k,n} - A^*_{k,n}\| \le C\left[\left(\frac{k}{n}\right)^{\lambda+1-q} + \left(\frac{k}{n}\right)^{\gamma} + \frac{k}{n}\left(1 + \log\left(\frac{n}{k}\right)\right)\right] k^{-(1-q)} \tag{3.14}$$

for any pair $(k, n) \in \mathbb{N}^2$ *with* $k \le n$.

PROOF. Let $(k, n) \in \mathbb{N}^2$ be such that $k \leq n$. We have

$$A_{k,n} - A^*_{k,n} = \sum_{j=k}^{n} j^{-1} A^*_{j+1,n} (\Gamma - \tilde{\Gamma}_j) A_{k,j-1}. \tag{3.15}$$

Writing $G_{k,m} := \sum_{j=k}^{m} (\Gamma - \tilde{\Gamma}_j)$ for $m \in \mathbb{N}$ with $m \geq k$, it follows from Lemma 3.1 that

$$A_{k,n} - A^*_{k,n} = \sum_{\nu=1}^{4} J_\nu(k, n), \tag{3.16}$$

where

$$J_1(k, n) := n^{-1} A^*_{n+1,n} G_{k,n} A_{k,n-1},$$

$$J_2(k, n) := n^{-1} \sum_{l=k}^{n-1} G_{k,l} (A_{k,l-1} - A_{k,l}),$$

$$J_3(k, n) := \sum_{m=k}^{n-1} \left(\frac{1}{m} A^*_{m+1,n} - \frac{1}{m+1} A^*_{m+2,n} \right) G_{k,m} A_{k,m-1}$$

and

$$J_4(k, n) := \sum_{m=k}^{n-1} \left(\frac{1}{m} A^*_{m+1,n} - \frac{1}{m+1} A^*_{m+2,n} \right) \sum_{l=k}^{m-1} G_{k,l} \big(A_{k,l-1} - A_{k,l}\big).$$

To estimate the norms of $J_\nu(k, n)$ ($\nu \in \{1, 2, 3, 4\}$), we first notice that, for all l, $m \in \mathbb{N}$ with $l \geq k$ and $k \leq m < n$,

$$\|G_{k,l}\| \leq C\, l^q \qquad \text{[by (3.5)]}, \tag{3.17}$$

$$\|A_{k,l-1} - A_{k,l}\| \leq \frac{1}{l} \|\tilde{\Gamma}_l\| \, \|A_{k,l-1}\| \leq \frac{C}{l} \left(\frac{k}{l}\right)^\lambda \|\tilde{\Gamma}_l\| \qquad \text{[by (3.13)]} \tag{3.18}$$

and

$$\begin{aligned} \left\| \frac{1}{m} A^*_{m+1,n} - \frac{1}{m+1} A^*_{m+2,n} \right\| &\leq \frac{1}{m\,(m+1)} \|I - \Gamma\| \, \|A^*_{m+2,n}\| \\ &\leq C\, m^{-2} \left(\frac{m}{n}\right)^\gamma \qquad \text{[by (3.10)]}. \end{aligned} \tag{3.19}$$

It follows that

$$\|J_1(k, n)\| \leq C\, k^\lambda\, n^{q-1-\lambda} = C \left(\frac{k}{n}\right)^{\lambda+1-q} k^{-(1-q)} \tag{3.20}$$

[cf. (3.13)] and

$$\begin{aligned} \|J_2(k, n)\| &\leq C\, n^{-1} k^\lambda \sum_{l=k}^{n-1} l^{q-1-\lambda} \|\tilde{\Gamma}_l\| \qquad \text{[by (3.17) and (3.18)]} \\ &\leq C \left(n^{q-\lambda-1} k^\lambda + n^{-1} k^q \right) \end{aligned} \tag{3.21}$$

[by (3.6), (3.12) and Lemma 3.2]. Moreover,

$$\|J_3(k,\,n)\| \le C \sum_{m=k}^{n-1} \frac{1}{m^2}\left(\frac{m}{n}\right)^{\gamma} m^q \left(\frac{k}{m}\right)^{\lambda} \qquad \text{[by (3.13), (3.17) and (3.19)]}$$

$$\le C\, k^{\lambda}\, n^{-\gamma} \sum_{m=k}^{n-1} m^{-2+\gamma+q-\lambda} \tag{3.22}$$

$$\le C\left(k^{\lambda}\, n^{q-\lambda-1} + n^{-\gamma}\, k^{q+\gamma-1}\right) \qquad \text{[by (3.12)].}$$

Finally,

$$\|J_4(k,\,n)\| \le C \sum_{m=k}^{n-1} m^{-2}\left(\frac{m}{n}\right)^{\gamma} \sum_{l=k}^{m-1} l^{q-1}\left(\frac{k}{l}\right)^{\lambda} \|\tilde{\Gamma}_l\|$$

[by (3.17), (3.18) and (3.19)]

$$\le C\, k^{\lambda}\, n^{-\gamma} \sum_{m=k}^{n-1} m^{-2+\gamma} \sum_{l=k}^{m-1} l^{q-1-\lambda}\, \|\tilde{\Gamma}_l\|.$$

In case $\lambda < q$, it follows from (3.6) and Lemma 3.2 that

$$\begin{aligned}\|J_4(k,\,n)\| &\le C\, k^{\lambda}\, n^{-\gamma} \sum_{m=k}^{n-1} m^{-2+\gamma+q-\lambda}\\ &\le C\left(k^{\lambda}\, n^{q-\lambda-1} + n^{-\gamma}\, k^{q+\gamma-1}\right) \qquad \text{[by (3.12)].}\end{aligned} \tag{3.23}$$

Similarly, if $\lambda > q$, then

$$\begin{aligned}\|J_4(k,\,n)\| &\le C\, k^{q}\, n^{-\gamma} \sum_{m=k}^{n-1} m^{-2+\gamma}\\ &\le \begin{cases} C\, k^q\, n^{-1}, & \text{if } \gamma > 1,\\ C\, k^q\, n^{-1}\left(1+\log\left(\dfrac{n}{k}\right)\right), & \text{if } \gamma = 1,\\ C\, n^{-\gamma}\, k^{q+\gamma-1}, & \text{if } \gamma < 1.\end{cases}\end{aligned} \tag{3.24}$$

Combining (3.16), (3.20), (3.21), (3.22), (3.23), (3.24) and (3.12), we arrive at (3.14). □

COROLLARY 3.1. *The notation and assumptions are as introduced before the statement of Lemma* 3.3. *Then, for each* $r \in (0,\, \min\{1,\, \gamma\})$ *there is a constant* $C_0 \in (0,\, \infty)$ *such that*

$$\|A_{k,\,n} - A^{*}_{k,\,n}\| \le C_0 \left(\frac{k}{n}\right)^{r} k^{-(1-q)} \tag{3.25}$$

for all $(k,\, n) \in \mathbb{N}^2$ *with* $k \le n$.

PROOF. The constant λ in (3.11) can be chosen in such a way that

$$\lambda < \gamma, \qquad \frac{|q-\lambda|}{1-q} \notin \mathbb{N} \cup \{0\} \quad \text{and} \quad \frac{\gamma-\lambda}{1-q} \notin \mathbb{N}.$$

An induction argument based on (3.10) and Lemma 3.4 then completes the proof of the corollary. □

4. Proof of Theorem 2.1. In the following, it will not always be indicated that certain inequalities only hold almost surely. Moreover, the C's are allowed to depend on $\omega \in \Omega$.

Let $\sigma \in [0, \min\{\frac{1}{3}(1-q), \eta, \gamma-\frac{1}{2}-\eta\})$ and $r \in \left(\frac{1}{2}+\sigma, \min\{1, \gamma\}\right)$. In case $\Gamma_n \equiv \Gamma$ for all $n \in \mathbb{N}$, it follows from the upper half of the law of the iterated logarithm for Brownian motion that (2.5) is a special case of Corollary 2.24 in Berger (1986). So it remains only to show that the sequence $(U_n^*)_{n\in\mathbb{N}}$ defined by

$$U_1^* := 0 \quad \text{and} \quad U_{n+1}^* := (I - n^{-1}\Gamma)\, U_n^* + n^{-1} V_n \quad \text{for } n \in \mathbb{N} \tag{4.1}$$

approximates $(U_n)_{n\in\mathbb{N}}$ in the sense that

$$\|U_{n+1} - U_{n+1}^*\| = o(n^{-(1/2)-\sigma}) \quad \text{a.s. as } n \to \infty. \tag{4.2}$$

To prove (4.2), let $A_{k,n}$ and $A_{k,n}^*$ $[(k, n) \in \mathbb{N}^2,\ k \le n+1]$ be defined as in (3.7)–(3.9) but with $\tilde{\Gamma}_l$ being replaced by Γ_l. Then, for any $n \in \mathbb{N}$,

$$U_{n+1} - U_{n+1}^* = A_{1,n}\, U_1 + \Delta_n, \tag{4.3}$$

where

$$\Delta_1 := 0 \quad \text{and} \quad \Delta_n := \sum_{k=1}^{n-1} (A_{k+1,n} - A_{k+1,n}^*)\, k^{-1} V_k \tag{4.4}$$

for $n \in \mathbb{N} \setminus \{1\}$ (by induction). Applying Corollary 3.1, we see that

$$\|A_{1,n}\, U_1\| = O(n^{-r}) = o(n^{-(1/2)-\sigma}) \quad \text{a.s. as } n \to \infty, \tag{4.5}$$

and so the proof of Theorem 2.1 will be complete once we have shown that

$$\|\Delta_n\| = o(n^{-(1/2)-\sigma}) \quad \text{a.s. as } n \to \infty. \tag{4.6}$$

Setting $S_k := \sum_{j=1}^{k} V_j$ for $k \in \mathbb{N}$ and

$$D_{k,n} := k^{-1}(A_{k+1,n} - A_{k+1,n}^*) - (k+1)^{-1}(A_{k+2,n} - A_{k+2,n}^*) \tag{4.7}$$

for $(k, n) \in \mathbb{N}^2$ with $k < n$, partial summation leads to the identity

$$\Delta_n = \sum_{k=1}^{n-1} D_{k,n}\, S_k \tag{4.8}$$

valid for all $n \in \mathbb{N} \setminus \{1\}$. Recalling the definitions involved, we find that

$$\begin{aligned} D_{k,n} &= \frac{1}{k(k+1)}\Big[A_{k+2,n}(I-\Gamma_{k+1}) - A^*_{k+2,n}(I-\Gamma)\Big] \\ &= \frac{1}{k(k+1)}(A_{k+2,n} - A^*_{k+2,n})(I-\Gamma) \\ &\quad + \frac{1}{k(k+1)} A_{k+2,n}(\Gamma - \Gamma_{k+1}) \end{aligned} \tag{4.9}$$

for any pair $(k, n) \in \mathbb{N}^2$ with $k < n$. It follows that

$$\Delta_n = \Delta_{n,1} + \Delta_{n,2} + \Delta_{n,3}, \tag{4.10}$$

where

$$\Delta_{n,1} := \sum_{k=1}^{n-1} \frac{1}{k(k+1)}(A_{k+2,n} - A^*_{k+2,n})(I-\Gamma) S_k, \tag{4.11}$$

$$\Delta_{n,2} := \sum_{k=1}^{n-1} \frac{1}{k(k+1)} A^*_{k+2,n}(\Gamma - \Gamma_{k+1}) S_k \tag{4.12}$$

and

$$\Delta_{n,3} := \sum_{k=1}^{n-1} \frac{1}{k(k+1)}(A_{k+2,n} - A^*_{k+2,n})(\Gamma - \Gamma_{k+1}) S_k. \tag{4.13}$$

We are left to estimate $\Delta_{n,1}$, $\Delta_{n,2}$ and $\Delta_{n,3}$. We first notice that, for any $\delta \in (0, \infty)$,

$$\|S_k\| = O(k^{(1/2)+\delta}) \quad \text{a.s.} \tag{4.14}$$

[using (2.4) and the law of the iterated logarithm for Brownian motion]. In the sequel,

$$\delta \in (0, 1) \setminus \left\{\tfrac{3}{2} - r - q,\ \tfrac{5}{6} - r - \tfrac{1}{3} q,\ \tfrac{3}{2}(1-q)^{-1}\left(\tfrac{5}{6} - \gamma - \tfrac{1}{3} q\right)\right\} \tag{4.15}$$

will be assumed to be fixed. By (4.14) and Corollary 3.1,

$$\begin{aligned} \|\Delta_{n,1}\| &\le C \sum_{k=1}^{n-1} k^{-2}\, k^{-(1-q)} \left(\frac{k}{n}\right)^r k^{(1/2)+\delta} \\ &= C\, n^{-r} \sum_{k=1}^{n-1} k^{-(5/2)+q+r+\delta} \\ &\le \begin{cases} C\, n^{-r}, & \text{if } q + r + \delta < \frac{3}{2}, \\ C\, n^{-(3/2)+q+\delta}, & \text{if } q + r + \delta > \frac{3}{2} \end{cases} \end{aligned} \tag{4.16}$$

[cf. (4.15)]. We now proceed to analyze $\Delta_{n,2}$. To this end, we begin by introducing blocks

$$I_n(\ell) := [\mu_{\ell-1}, \mu_\ell) \cap \{1, \dots, n-1\} \quad \text{for } \ell,\ n \in \mathbb{N}, \tag{4.17}$$

where $\mu_0 := 0$ and

$$\mu_\ell := [\ell^{2\alpha+1}] \quad \text{for } \ell \in \mathbb{N}. \tag{4.18}$$

Here

$$\alpha := \frac{1+2q}{4(1-q)} \quad \text{so that } \frac{1}{2\alpha+1} = \frac{2}{3}(1-q). \tag{4.19}$$

We also define $\tau(n) := \min\{\ell \in \mathbb{N}: \mu(\ell) \geq n\}$ for $n \in \mathbb{N}$. Our point of departure is the decomposition

$$\Delta_{n,2} = \sum_{j=1}^{3} \Delta_{n,2,j}, \tag{4.20}$$

where

$$\Delta_{n,2,1} := \sum_{k=1}^{n-1} \frac{1}{k(k+1)} A^*_{k+2,n} (\Gamma - \Gamma_{k+1})(S_k - W(k)), \tag{4.21}$$

$$\Delta_{n,2,2} := \sum_{\ell=1}^{\infty} \sum_{k \in I_n(\ell)} \frac{1}{k(k+1)} A^*_{k+2,n} (\Gamma - \Gamma_{k+1}) W(\mu_{\ell-1}) \tag{4.22}$$

and

$$\Delta_{n,2,3} := \sum_{\ell=1}^{\infty} \sum_{k \in I_n(\ell)} \frac{1}{k(k+1)} A^*_{k+2,n} (\Gamma - \Gamma_{k+1}) \Big(W(k) - W(\mu_{\ell-1})\Big). \tag{4.23}$$

Here,

$$\begin{aligned} \|\Delta_{n,2,1}\| &\leq C \sum_{k=1}^{n-1} \frac{1}{k^2} \left(\frac{k}{n}\right)^{\gamma} \|\Gamma - \Gamma_{k+1}\| \, \|S_k - W(k)\| \qquad \text{[by (3.10)]} \\ &\leq C\, n^{-\gamma} \sum_{k=1}^{n-1} k^{-(3/2)+\gamma-\eta} \|\Gamma - \Gamma_{k+1}\| \qquad \text{[by (2.4)]} \\ &\leq C\, n^{-(1/2)-\eta} \end{aligned} \tag{4.24}$$

[by (2.3) and Lemma 3.2; note that $\gamma - \eta > \frac{1}{2}$]. It is easy to show that

$$\limsup_{\ell\to\infty} \ell^{-\alpha-\delta} \sup_{\mu_{\ell-1} \leq s < \mu_\ell} \|W(s) - W(\mu_{\ell-1})\| = 0 \quad \text{a.s.} \tag{4.25}$$

Using this result in (4.23), we get

$$
\begin{aligned}
\|\Delta_{n,2,3}\| &\le C \sum_{\ell=1}^{\infty} \sum_{k\in I_n(\ell)} \frac{1}{k\,(k+1)}\, \|A^*_{k+2,n}\|\, \|\Gamma-\Gamma_{k+1}\|\, \ell^{\alpha+\delta} \\
&\le C \sum_{\ell=1}^{\infty} \sum_{k\in I_n(\ell)} \frac{1}{k^2} \left(\frac{k}{n}\right)^{\gamma} \|\Gamma-\Gamma_{k+1}\|\, \ell^{\alpha+\delta} \qquad \text{[by (3.10)]} \\
&\le C \sum_{\ell=1}^{\infty} \sum_{k\in I_n(\ell)} \frac{1}{k^2} \left(\frac{k}{n}\right)^{\gamma} \|\Gamma-\Gamma_{k+1}\|\, k^{(\alpha+\delta)/(2\alpha+1)} \\
&= C\, n^{-\gamma} \sum_{k=1}^{n-1} k^{-2+\gamma}\, k^{(\alpha+\delta)/(2\alpha+1)}\, \|\Gamma-\Gamma_{k+1}\|.
\end{aligned} \tag{4.26}
$$

In case $\delta < \frac{3}{2}(1-q)^{-1}(\frac{5}{6}-\gamma-\frac{1}{3}q)$, that is, if $\gamma+(\alpha+\delta)/(2\alpha+1) < 1$ [cf. (4.19)], this entails that

$$
\|\Delta_{n,2,3}\| \le C\, n^{-\gamma} \tag{4.27}
$$

[by (2.3) and Lemma 3.2]. Similarly, if $\delta > \frac{3}{2}(1-q)^{-1}(\frac{5}{6}-\gamma-\frac{1}{3}q)$, then

$$
\begin{aligned}
\|\Delta_{n,2,3}\| &\le C\, n^{-1}\, n^{(\alpha+\delta)/(2\alpha+1)} \\
&\le C\, n^{-1/2}\, n^{-1/(2(2\alpha+1))}\, n^{\delta/(2\alpha+1)} \\
&= C\, n^{-1/2}\, n^{-(1/3)(1-q)}\, n^{\delta/(2\alpha+1)}.
\end{aligned} \tag{4.28}
$$

To estimate $\Delta_{n,2,2}$, we first consider the sums

$$
H_{m,p} := \sum_{k=m+1}^{p} \frac{1}{k\,(k+1)}\, A^*_{k+2,n}\, (\Gamma-\Gamma_{k+1}) \tag{4.29}
$$

for $(m,\ p) \in \mathbb{N}^2$ with $m < p$. Setting $G_{m+2,\,l+1} := \sum_{j=m+1}^{l} (\Gamma-\Gamma_{j+1})$ for $l \in \mathbb{N}$ with $l > m$, partial summation yields

$$
\begin{aligned}
H_{m,p} = {} & \frac{1}{p\,(p+1)}\, A^*_{p+2,n}\, G_{m+2,\,p+1} \\
& + \sum_{k=m+1}^{p-1} \left[\frac{1}{k\,(k+1)}\, A^*_{k+2,n} - \frac{1}{(k+1)\,(k+2)}\, A^*_{k+3,n} \right] G_{m+2,\,k+1}.
\end{aligned} \tag{4.30}
$$

Now

$$
\begin{aligned}
& \left\| \frac{1}{k\,(k+1)}\, A^*_{k+2,n} - \frac{1}{(k+1)\,(k+2)}\, A^*_{k+3,n} \right\| \\
& \quad \le \frac{1}{k\,(k+1)\,(k+2)}\, \|2I-\Gamma\|\, \|A^*_{k+3,n}\| \\
& \quad \le C\, k^{-3} \left(\frac{k}{n}\right)^{\gamma} \qquad \text{[by (3.10)]}.
\end{aligned} \tag{4.31}
$$

Combining (4.30), (3.10), (2.2) and (4.31), we find that

$$\|H_{m,\,p}\| \le C\,p^{-2}\left(\frac{p}{n}\right)^r p^q + C\sum_{k=m+1}^{p-1} k^{-3}\left(\frac{k}{n}\right)^r k^q \qquad (4.32)$$
$$\le C\,n^{-r}\,m^{-2+r+q} \qquad (\text{since } r+q<2).$$

Hence and from (4.22) and the law of the iterated logarithm for Brownian motion, we infer that

$$\|\Delta_{n,\,2,\,2}\| \le C\,n^{-r}\sum_{\ell=2}^{\tau(n)} \mu_{\ell-1}^{-2+r+q}\,\mu_{\ell-1}^{(1/2)+\delta} \qquad (4.33)$$
$$\le C\,n^{-r}\sum_{\ell=1}^{\tau(n)} \ell^{-(2\alpha+1)((3/2)-r-q-\delta)}.$$

In case $\delta > \frac{5}{6} - r - \frac{1}{3}\,q$, we have $(2\alpha+1)(\frac{3}{2}-r-q-\delta) < 1$, and so

$$\|\Delta_{n,\,2,\,2}\| \le C\,n^{-r}\,\tau(n)^{-(2\alpha+1)((3/2)-r-q-\delta)+1} \qquad (4.34)$$
$$\le C\,n^{-(3/2)+q+\delta}\,n^{1/(2\alpha+1)} \le C\,n^{-(1/2)-(1/3)(1-q)+\delta}.$$

Similarly, if $\delta < \frac{5}{6} - r - \frac{1}{3}\,q$, then

$$\|\Delta_{n,\,2,\,2}\| \le C\,n^{-r}. \qquad (4.35)$$

Finally,

$$\|\Delta_{n,\,3}\| \le C\sum_{k=1}^{n-1} k^{-2}\left(\frac{k}{n}\right)^r k^{-(1-q)}\,\|\Gamma-\Gamma_{k+1}\|\,\|S_k\| \qquad (\text{by Corollary 3.1})$$
$$\le C\,n^{-r}\sum_{k=1}^{n-1} k^{-(5/2)+q+r+\delta}\,\|\Gamma-\Gamma_{k+1}\| \qquad [\text{by } (4.14)] \qquad (4.36)$$
$$\le \begin{cases} C\,n^{-r}, & \text{if } q+r+\delta<\frac{3}{2},\\ C\,n^{-(3/2)+q+\delta}, & \text{if } q+r+\delta>\frac{3}{2}\end{cases}$$

[by (2.3) and Lemma 3.2].

Combining (4.10), (4.16), (4.20), (4.24), (4.27), (4.28), (4.34), (4.35), (4.36) and (4.15), we arrive at (4.6) (note that δ can be taken arbitrarily small). □

APPENDIX

An extension. This Appendix is mainly addressed to readers who are interested in comparing the present approach with Kouritzin's (1996) work. An aspect of his article that is not covered by Theorem 2.1 is that he permits the approximating Gaussian process [in the analogue of (2.4)] to be more general than Brownian motion. The generalization of Theorem 2.1 stated as Theorem 2.1′ eliminates this disadvantage and, at the same time, provides a possible answer to the problem of giving a more manageable extension of the class of approximating Gaussian processes admitted by Kouritzin.

We retain the notation introduced in Section 2. Instead of using Condition 2.4, we now employ the following condition.

CONDITION 2.4′. For some $\eta \in \left(0, \min\{\frac{1}{2}, \gamma - \frac{1}{2}\}\right)$, there is an $\mathbb{R}^d$-valued centered Gaussian process $(X(t))_{t\in[0,\infty)}$ [defined on $(\Omega, \mathscr{F}, P)$] with continuous sample paths satisfying the following assumptions:

$$\lim_{n\to\infty} n^{-1-\varepsilon} E\|X(n)\|^2 = 0 \quad \text{for any } \varepsilon \in (0, \infty), \tag{A.1}$$

$$\lim_{n\to\infty} n^{-\rho-\varepsilon} \max_{k\in\{1,\dots,[n^\rho]\}} E\|X(n+k) - X(n)\|^2 = 0 \quad \text{for any } \rho \in \left(\tfrac{1}{3}, 1\right) \text{ and any } \varepsilon \in (0, \infty), \tag{A.2}$$

$$\left\| X(t) - \sum_{k\le t} V_k \right\| = O(t^{(1/2)-\eta}) \quad \text{a.s. for } t \ge 1. \tag{A.3}$$

THEOREM 2.1′. *Under Conditions* 2.1–2.3 *and* 2.4′ *one has*

$$\lim_{\substack{t\to\infty\\ t>0}} t^{-(1/2)+\sigma} \|[t]U_{[t+1]} - \hat{Y}(t)\| = 0 \quad a.s. \tag{A.4}$$

for any $\sigma \in [0, \min\{\frac{1}{3}(1-q), \eta, \gamma - \frac{1}{2} - \eta\})$, *where the process* $(\hat{Y}(t))_{t\in[0,\infty)}$ *is defined by*

$$\hat{Y}(t) := X(t) - (1+t)^{-1}\Lambda \int_0^t \left(\frac{1+s}{1+t}\right)^{\Lambda - I} X(s)\,ds \quad \text{for } t \in [0, \infty) \tag{A.5}$$

with $\Lambda := \Gamma - I$.

REMARK A.1. Under the assumption that X satisfies a suitable growth condition in the interval $[0, 1]$, we could just as well have defined $\hat{Y}$ by analogy with (2.6). The definition (A.5) essentially coincides with the one utilized by Kouritzin (1996).

PROOF OF THEOREM 2.1′. Let $\sigma \in [0, \min\{\frac{1}{3}(1-q), \eta, \gamma - \frac{1}{2} - \eta\})$. The sequence $(X(n))_{n\in\mathbb{N}}$ being Gaussian, it follows from (A.1) and from the Borel–Cantelli lemma that, for any $\varepsilon \in (0, \infty)$,

$$n^{-1-\varepsilon}\|X(n)\|^2 \to 0 \quad \text{a.s. as } n \to \infty. \tag{A.6}$$

Similarly, (A.2) entails that the following analogue of (4.25) holds:

$$\limsup_{\ell\to\infty} \ell^{-\alpha-\delta} \sup_{k\in\{\mu_{\ell-1},\dots,\mu_\ell\}} \|X(k) - X(\mu_{\ell-1})\| = 0 \quad \text{a.s.} \tag{A.7}$$

for any $\delta \in (0, \infty)$ [cf. (4.18) and (4.19)]. Taking (A.6) and (A.7) into account and arguing as in the proof of Theorem 2.1 with the role of Corollary 2.24 in

Berger (1986) being taken over by Theorem 2.23 in the same work, we find that

$$\lim_{n\to\infty} n^{-(1/2)+\sigma} \| n\, U_{n+1} - Y_n^* \| = 0 \quad \text{a.s.,} \tag{A.8}$$

where $Y_n^* := \sum_{k=1}^{n} (k/n)^{\Lambda} (X(k) - X(k-1))$ for $n \in \mathbb{N}$. Now let

$$\tilde{Y}(t) := X(t+1) - (1+t)^{-1} \Lambda \int_0^t \left(\frac{1+v}{1+t} \right)^{\Lambda - I} X(v+1)\, dv \quad \text{for } t \in [0, \infty). \tag{A.9}$$

In view of (A.8), the desired relation (A.4) will be established once we have shown that

$$\begin{aligned} &\text{(a) } \|Y_{[t]}^* - \tilde{Y}(t)\| = O(t^{(1/2)-\eta}) \quad \text{a.s.} \quad \text{and} \\ &\text{(b) } \|\tilde{Y}(t) - \hat{Y}(t)\| = O(t^{(1/2)-\eta}) \quad \text{a.s.} \end{aligned} \tag{A.10}$$

for $t \geq 1$. To verify (A.10)(a), we first notice that

$$\tilde{Y}(t) = X(t+1) - \Lambda \int_{1/(1+t)}^{1} s^{\Lambda - I} X(s(t+1))\, ds \quad \text{for } t \in [0, \infty). \tag{A.11}$$

Arguing as in Berger (1986) [formulas (2.30) and (2.31)], we conclude that, for any $n \in \mathbb{N} \setminus \{1\}$,

$$\begin{aligned} \|Y_n^* - \tilde{Y}(n-1)\| &= \left\| \Lambda \sum_{k=1}^{n-1} \int_{k/n}^{(k+1)/n} s^{\Lambda - I} (X(k) - X(sn))\, ds \right\| \\ &= O(n^{(1/2)-\eta}) \|\Lambda\| \sum_{k=1}^{n-1} \int_{k/n}^{(k+1)/n} s^{\gamma - (3/2) - \eta}\, ds \\ &\qquad\qquad \text{[using (2.1) and (A.3)]} \\ &= O(n^{(1/2)-\eta}) \quad \text{a.s.} \qquad (\text{since } \gamma - \tfrac{1}{2} - \eta > 0) \end{aligned} \tag{A.12}$$

and

$$\begin{aligned} &\max_{t\in[0,1]} \|\tilde{Y}(n-1+t) - \tilde{Y}(n-1)\| \\ &\quad \leq \max_{t\in[0,1]} \|X(n+t) - X(n)\| \\ &\qquad + \|\Lambda\| \int_{1/n}^{1} s^{\gamma-2} \max_{t\in[0,1]} \|X((n+t)s) - X(sn)\|\, ds \\ &\qquad + \|\Lambda\| \int_{1/(n+1)}^{1/n} s^{\gamma-2} \max_{t\in[0,1]} \|X((n+t)s)\|\, ds \qquad \text{[cf. (2.1)]} \\ &\quad = O(n^{(1/2)-\eta}) \quad \text{a.s.} \quad \text{[by (A.3)].} \end{aligned} \tag{A.13}$$

Combining (A.12) and (A.13) gives (A.10)(a). Finally, (A.10)(b) follows by an easy modification of the arguments just employed in (A.13), and so the proof of Theorem 2.1′ is complete. □

REFERENCES

BERGER, E. (1986). Asymptotic behaviour of a class of stochastic approximation procedures. *Probab. Theory Related Fields* **71** 517–552.

BERGER, E. (1990). An almost sure invariance principle for stationary ergodic sequences of Banach space valued random variables. *Probab. Theory Related Fields* **84** 161–201.

DALECKIĬ, JU. L. and KREĬN, M. G. (1974). *Stability of Solutions of Differential Equations in Banach Space.* Amer. Math. Soc., Providence, RI.

FABIAN, V. (1968). On asymptotic normality in stochastic approximation. *Ann. Math. Statist.* **39** 1327–1332.

HEUNIS, A. (1994). Rates of convergence for an adaptive filtering algorithm driven by stationary dependent data. *SIAM J. Control Optim.* **32** 116–139.

KOURITZIN, M. A. (1996). On the interrelation of almost sure invariance principles for certain stochastic adaptive algorithms and for partial sums of random variables. *J. Theoret. Probab.* **9** 811–840.

MARK, G. (1982). Log-log-Invarianzprinzipien für Prozesse der stochastischen Approximation. *Mitt. Math. Sem. Giessen* **153** 1–87.

PHILIPP, W. (1986). Invariance principles for independent and weakly dependent random variables. In *Dependence in Probability and Statistics* (E. Eberlein and M. S. Taqqu, eds.) 225–268. Birkhäuser, Boston.

STRASSEN, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete* **3** 211–226.

WALK, H. (1977). An invariance principle for the Robbins–Monro process in a Hilbert space. *Z. Wahrsch. Verw. Gebiete* **39** 135–150.

WALK, H. (1980). A functional central limit theorem for martingales in $C(K)$ and its application to sequential estimates. *J. Reine Angew. Math.* **314** 117–135.

WALK, H. (1988). Limit behaviour of stochastic approximation processes. *Statist. Decisions* **6** 109–128.

WALK, H. and ZSIDÓ, L. (1989). Convergence of the Robbins–Monro method for linear problems in a Banach space. *J. Math. Anal. Appl.* **139** 152–177.

INSTITUT FÜR MATHEMATISCHE STOCHASTIK
UNIVERSITÄT GÖTTINGEN
LOTZESTR. 13
D-37083 GÖTTINGEN
GERMANY
E-MAIL: berger@math.uni-goettingen.de