Electron. Commun. Probab. **22** (2017), no. 36, 1–13.
DOI: 10.1214/17-ECP67
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

A note on continuous-time stochastic approximation in infinite dimensions*

Jan Seidler† František Žák‡

Abstract

We find sufficient conditions for convergence of a continuous-time Robbins-Monro type stochastic approximation procedure in infinite dimensional Hilbert spaces in terms of Lyapunova functions, the variational approach to stochastic partial differential equations being used as the main tool.

Keywords: stochastic approximation; stochastic parabolic problems; variational solutions.
AMS MSC 2010: 60H15; 62L20.
Submitted to ECP on October 31, 2016, final version accepted on June 14, 2017.

1 Introduction

1.1

Stochastic approximation was originally introduced as a procedure for sequentially finding a zero or an extremum point of a function which can be observed only with a random measurement error; it has found many applications e.g. to recursive estimation, adaptive control or learning algorithms, see the books [BMP, BPP, Bo, Che2, KC] or [KY] for a thorough information about the stochastic approximation methods. The seminal Robbins-Monro procedure may be roughly described as follows: Let $R\colon \mathbb{R} \to \mathbb{R}$ be a function which is known to have a unique root x_0 but the observation of $R(x)$ at time $k \in \mathbb{N}$ is corrupted by a noise $e_k(x)$. Let $\alpha_n > 0$ be such that

$$\sum_{n=0}^{\infty} \alpha_n = \infty, \quad \sum_{n=0}^{\infty} \alpha_n^2 < \infty$$

and set

$$Y_{n+1} = Y_n + \alpha_n \big(R(Y_n) + e_{n+1}(Y_n) \big).$$

Then under suitable assumptions upon the function R and the random variables $e_k(x)$ it may be shown that $Y_n \to x_0$ almost surely as $n \to \infty$. M. B. Nevel'son and R. Z. Khas'minskiĭ in their book [NCh] studied a continuous-time version of stochastic approximation. In particular, they introduced a continuous-time analogue of the Robbins-Monro procedure: Consider a stochastic differential equation

$$\mathrm{d}X = \alpha(t) \big(R(X)\,\mathrm{d}t + \sigma(t,X)\,\mathrm{d}W \big), \quad X_0 = x, \tag{1.1}$$

*The research of the first author was supported in part by the GAČR grant no. 15-08819S.

†Institute of Information Theory and Automation of the Czech Academy of Sciences, 182 08 Prague, Czech Republic. E-mail: seidler@utia.cas.cz

‡Imperial College London, United Kingdom. E-mail: f.zak12@imperial.ac.uk

where W is a Wiener process and α is a strictly positive function in $L^2(\mathbb{R}_{\geq 0}) \setminus L^1(\mathbb{R}_{\geq 0})$. Sufficient conditions for X_t to converge to the zero set of R almost surely as $t \to \infty$ were found in terms of existence of a suitable Lyapunov function for (1.1). One may consult the book [Ko] or the papers [Pf, Che, La1, La2, La3, La4] for further results in this direction. Due to powerful tools from stochastic analysis, proofs in the continuous-time case may be presented in a very lucid way (cf. also [Che] for a discussion of this point).

The aim of this note is to extend the stochastic analysis approach, in the form proposed by Nevel'son and Khas'minskiĭ, to infinite-dimensional Hilbert spaces. Several results on discrete-time stochastic approximation in infinite-dimensional spaces are available, cf. e.g. [BRS, Go, KS, Ni, Yi, YZ], but the only paper using infinite-dimensional stochastic analysis to study stochastic approximation we are aware of is [BYY, § 4]. However, [BYY] treats stochastic delay equations, whilst we are interested in stochastic partial differential equations. We confine ourselves to procedures of the Robbins-Monro type in the case of a unique root, since we see our task in indicating how the ideas from [NCh] may be combined with techniques from the theory of stochastic evolution equations, not in obtaining the strongest possible results. A typical example we can cover is the following: Consider a nonlinear elliptic equation

$$\Delta u + r(u) = f \text{ in } D, \quad u = 0 \text{ on } \partial D, \tag{1.2}$$

where $D \subseteq \mathbb{R}^d$ is a bounded domain with a smooth boundary ∂D, and a stochastic parabolic equation

$$\mathrm{d}X = \alpha(t)\big(\Delta X + r(X) - f\big)\,\mathrm{d}t + \alpha(t)\sigma(t,X)\,\mathrm{d}W, \quad X_{|\mathbb{R}_{>0}\times\partial D} = 0,\ X_0 = y \tag{1.3}$$

in $L^2(D)$, driven by an infinite-dimensional Wiener process W, where $\alpha \in L^2(\mathbb{R}_{\geq 0}) \setminus L^1(\mathbb{R}_{\geq 0})$ is again a strictly positive function. Sufficient conditions on r will be found for the solution X of (1.3) to converge almost surely to the (unique) solution $u_0 \in W_0^{1,2}(D)$ of (1.2) (see Example 3.1 below).

A common approach to equations like (1.3) is to interpret them in the mild sense, as an equation

$$X_t = U(t,0)y + \int_0^t U(t,s)r(X_s)\,\mathrm{d}s + \int_0^t \alpha(s)U(t,s)\sigma(s,X_s)\,\mathrm{d}W_s,$$

where U is the evolution operator generated by $\alpha(\cdot)\Delta$. However, our proofs rely heavily on the use of Lyapunov functions, while mild solutions are not semimartingales and the Itô formula cannot be applied directly to them, approximations of a rather technical nature being needed. Thence we decided to use the theory of variational solutions, going back to [Pa] and [KR] (see e.g. the books [Cho, Chapter 6] and [LR, Chapters 4 and 5] for a more recent presentation). Moreover, this choice makes it possible to deal with quasilinear problems (see Examples 3.2, 3.3).

Before stating our main results we have to introduce some notation and recall a few basic facts about variational solutions we shall need. This is done in the next two subsections; the main results are stated and proved in Section 2, in Section 3 some illustrative examples are provided.

1.2

Let E and F be Banach spaces, we shall denote by $\mathscr{L}(E,F)$ the space of all bounded linear operators from E to F. If both spaces are Hilbert, by $\mathscr{L}_2(E,F)$ the ideal of Hilbert-Schmidt operators in $\mathscr{L}(E,F)$ will be denoted. $C^k(G)$ will stand for the space of k-times continuously differentiable real-valued functions on an open set $G \subseteq E$. If $f\colon G \to \mathbb{R}$, we shall denote by $Df(x)$ and $D^2f(x)$ the first and second Gâteaux derivative of f at the point x, respectively, provided they exist. Analogously, if $f\colon \mathbb{R}_{\geq 0} \times G \to \mathbb{R}$, $D_x f(t,x)$ and $D_x^2 f(t,x)$ will stand for the first and second Gâteaux derivative of $f(t,\cdot)$ at the point x. For spaces of (Bochner) integrable functions and Sobolev spaces we shall use standard notation; finally, by λ the Lebesgue measure on $\mathbb{R}_{\geq 0}$ will be denoted.

1.3

Let H and K be real separable Hilbert spaces, B a reflexive Banach space embedded continuously and densely in H. Upon identifying H with its dual H^* we get a Gelfand triple $B \subseteq H \subseteq B^*$; note that – in this representation – the restriction of the dual pairing $\langle \cdot,\cdot \rangle_{B^*,B}$ to $H \times B$ coincides with the scalar product $\langle \cdot,\cdot \rangle_H$ in H. Assume that

(A) *$f\colon \mathbb{R}_{\geq 0} \times B \longrightarrow B^*$ and $\sigma\colon \mathbb{R}_{\geq 0} \times B \longrightarrow \mathscr{L}_2(K,H)$ are Borel functions, μ is a Borel probability measure on H,*

and consider a stochastic evolution equation

$$\mathrm{d}X = f(t,X)\,\mathrm{d}t + \sigma(t,X)\,\mathrm{d}W, \quad X_0 \sim \mu. \tag{1.4}$$

Definition 1.1. *$((\Omega,\mathscr{F},(\mathscr{F}_t),\boldsymbol{P}),W,X)$ is called a (variational) solution to the stochastic evolution equation (1.4) provided $(\Omega,\mathscr{F},(\mathscr{F}_t),\boldsymbol{P})$ is a filtered probability space whose filtration satisfies the usual conditions and on which a standard cylindrical $(\mathscr{F}_t)$-Wiener process W on K and a B^*-valued $(\mathscr{F}_t)$-progressively measurable process X are defined such that*

i) $X(0)$ has distribution μ,

ii) there exists an $(\mathscr{F}_t)$-progressively measurable B-valued process $\tilde{X}$ satisfying $\|\tilde{X}\|_B \in L^p_{\mathrm{loc}}(\mathbb{R}_{\geq 0})$ $\boldsymbol{P}$-almost surely for some $p \in (1,\infty)$, $X = \tilde{X}$ $\lambda \otimes \boldsymbol{P}$-almost everywhere on $\mathbb{R}_{\geq 0} \times \Omega$,

iii) $\|f(\cdot,\tilde{X}(\cdot))\|_{B^*}^{p/(p-1)} + \|\sigma(\cdot,\tilde{X}(\cdot))\|_{\mathscr{L}_2}^2 \in L^1_{\mathrm{loc}}(\mathbb{R}_{\geq 0})$ and

$$X(t) = X(0) + \int_0^t f(s,\tilde{X}(s))\,\mathrm{d}s + \int_0^t \sigma(s,\tilde{X}(s))\,\mathrm{d}W_s \quad \text{in } B^*$$

for all $t \geq 0$ $\boldsymbol{P}$-almost surely.

Since the process X solving (1.4) is in general only B^*-valued, the Itô formula cannot be used to compute $\varphi(X)$ for an arbitrary $\varphi \in C^2(H)$ and extra assumptions on φ are needed. We state here two Itô formula-type results which we shall need later.

First, let $(\Omega,\mathscr{F},(\mathscr{F}_t),\boldsymbol{P})$ be a filtered probability space satisfying the usual conditions and carrying a standard cylindrical $(\mathscr{F}_t)$-Wiener process W on K. Assume that:

1° $u_0\colon \Omega \longrightarrow H$ is an $\mathscr{F}_0$-measurable random variable,

2° $Z\colon \mathbb{R}_{\geq 0} \times \Omega \longrightarrow \mathscr{L}_2(K,H)$ is a progressively measurable process such that $\|Z\|_{\mathscr{L}_2} \in L^2_{\mathrm{loc}}(\mathbb{R}_{\geq 0})$ $\boldsymbol{P}$-almost surely,

3° $v\colon \mathbb{R}_{\geq 0} \times \Omega \longrightarrow B^*$ is a progressively measurable process with $\|v\|_{B^*} \in L^{p/(p-1)}_{\mathrm{loc}}(\mathbb{R}_{\geq 0})$ $\boldsymbol{P}$-almost surely for some $p \in (1,\infty)$,

4° if u is the B^*-valued process defined by

$$u(t) = u_0 + \int_0^t v(s)\,\mathrm{d}s + \int_0^t Z(s)\,\mathrm{d}W_s, \quad t \geq 0, \tag{1.5}$$

then there exists a B-valued process $\tilde{u}$ such that $\tilde{u} \in L^p_{\mathrm{loc}}(\mathbb{R}_{\geq 0};B)$ $\boldsymbol{P}$-almost surely and $u = \tilde{u}$ $\lambda \otimes \boldsymbol{P}$-almost everywhere on $\mathbb{R}_{\geq 0} \times \Omega$.

Then u has sample paths in $C(\mathbb{R}_{\geq 0}; H)$ $\boldsymbol{P}$-almost surely and

$$\|u(t)\|_H^2 = \|u_0\|_H^2 + \int_0^t \Big\{ 2\langle v(s), \tilde{u}(s)\rangle_{B^*,B} + \|Z_s\|_{\mathscr{L}_2}^2 \Big\} \,\mathrm{d}s \\ + 2\int_0^t \langle Z(s)^* u(s), \cdot\rangle_K \,\mathrm{d}W_s$$

for all $t \geq 0$ $\boldsymbol{P}$-almost surely, see [Pa, Théorème 3.1 on p. 57], [KR, Theorem 2.17] or [LR, Theorem 4.2.5].

Comparing this result with Definition 1.1 we see that any solution X of (1.4) has continuous sample path in H $\boldsymbol{P}$-almost surely.

In order to establish the Itô formula for functions more general than $\|\cdot\|_H^2$ one needs an additional hypothesis

(C) *Both B and B^* are uniformly convex.*

Let $\mathfrak{I}$ be the set of all functions $\varphi \in C^1(H)$ such that the second Gâteaux derivative $D^2\varphi(x)$ exists at all points $x \in H$, the functions φ, $D\varphi$ and $D^2\varphi$ are bounded on bounded sets in H, the mapping $x \longmapsto D^2\varphi(x)$ is continuous from H to $(\mathscr{L}(H), \text{weak}^*)$, the restriction $D\varphi_{|B}$ maps continuously $(B, \|\cdot\|)$ to (B, weak) and there exists a constant $k < \infty$ such that $\|D\varphi(x)\|_B \leq k(1 + \|x\|_B)$ for every $x \in B$. If the process u defined by (1.5) satisfies the hypotheses 1°–4° above and $\varphi \in \mathfrak{I}$ then

$$\varphi(u(t)) = \varphi(u_0) + \int_0^t \Big\{ \big\langle v(s), D\varphi(\tilde{u}(s))\big\rangle_{B^*,B} + \frac{1}{2}\operatorname{Tr}\big(D^2\varphi(u(s))Z(s)Z^*(s)\big) \Big\} \,\mathrm{d}s \\ + \int_0^t \big\langle Z(s)^* D\varphi(u(s)), \cdot\big\rangle_K \,\mathrm{d}W_s$$

for all $t \geq 0$ $\boldsymbol{P}$-almost surely, see [Pa, Théorème 4.2 on p. 65], cf. also [Kry, Theorem 3.1]. In particular, $\varphi(u)$ is a continuous real-valued semimartingale, hence the process $\psi(t, \varphi(u(t)))$ may be expressed by means of the real-valued case of the Itô formula, provided ψ belongs to the set $C^{1,2}$ of all functions $\zeta \in C^1(\mathbb{R}_{\geq 0} \times \mathbb{R})$ such that $\zeta(t, \cdot) \in C^2(\mathbb{R})$ for all $t \geq 0$ and $(t, x) \longmapsto \frac{\partial^2 \zeta}{\partial x^2}(t, x)$ is a continuous function on $\mathbb{R}_{\geq 0} \times \mathbb{R}$. We shall denote by $\mathfrak{K}$ the set of all functions ξ on $\mathbb{R}_{\geq 0} \times H$ of the form $\xi(t, x) = \psi(t, \varphi(x))$, $\psi \in C^{1,2}$, $\varphi \in \mathfrak{I}$. For $\xi \in \mathfrak{K}$ one gets the expected equality

$$\xi(t, u(t)) = \xi(0, u_0) + \int_0^t \Big\{ \frac{\partial \xi}{\partial s}(s, u(s)) + \big\langle v(s), D_x\xi(s, \tilde{u}(s))\big\rangle_{B^*,B} \\ + \frac{1}{2}\operatorname{Tr}\big(D_x^2\xi(s, u(s))Z(s)Z^*(s)\big) \Big\} \,\mathrm{d}s \\ + \int_0^t \big\langle Z(s)^* D_x\xi(s, u(s)), \cdot\big\rangle_K \,\mathrm{d}W_s.$$

Note that $D_x\xi(t, x) = \frac{\partial \psi}{\partial x}(t, \varphi(x))D\varphi(x)$, so the term $\langle v(t), D_x\xi(t, \tilde{u}(t))\rangle_{B^*,B}$ remains well defined. A special case, following from the product rule for semimartingales, is used repeatedly in the considerations below:

$$\mathrm{d}\big(g(t)\xi(t, u(t))\big) = g'(t)\xi(t, u(t)) \,\mathrm{d}t + g(t) \,\mathrm{d}\xi(t, u(t)) \tag{1.6}$$

whenever $g \in C^1(\mathbb{R}_{\geq 0})$, or, more generally,

$$\mathrm{d}\big(g(t)\xi(t, u(t))\big) = \xi(t, u(t)) \,\mathrm{d}g(t) + g(t) \,\mathrm{d}\xi(t, u(t)) \tag{1.7}$$

if $g \in C(\mathbb{R}_{\geq 0})$ is locally of finite variation.

Remark 1.2. a) The hypothesis (C) is obviously satisfied if B is a Hilbert space. Let us emphasize that (C) can be omitted if $\varphi = \|\cdot\|_H^2$ or, more generally, if processes of the form $\psi(t, \|u(t)\|_H^2)$ with $\psi \in C^{1,2}$ are considered.

b) An Itô formula for the process $\chi(t, u(t))$, where χ is a suitable smooth function on $\mathbb{R}_{\geq 0} \times H$, is proved in [Cho, Theorem 7.2.1], but under rather restrictive additional assumptions on u.

2 Main results

Following [NCh], we derive the convergence of a Robbins-Monro type procedure as an immediate corollary to a theorem providing sufficient conditions for the convergence of path of any solution of (1.4) to a singleton $\{x_0\}$, which will be established first. (In applications to stochastic approximation, x_0 will be the unique root of the drift coefficient, but on an abstract level, it may be an arbitrary point in H.) Hence, let us consider the equation (1.4), that is

$$\mathrm{d}X = f(t, X)\,\mathrm{d}t + \sigma(t, X)\,\mathrm{d}W, \quad X_0 \sim \mu,$$

and denote by L the Kolmogorov operator associated with it, namely, if $h \in \mathfrak{K}$, then we set

$$Lh(t,x) = \frac{\partial h}{\partial t}(t,x) + \big\langle f(t,x), D_x h(t,x)\big\rangle_{B^*,B} + \frac{1}{2}\operatorname{Tr}\big(D_x^2 h(t,x)(\sigma\sigma^*)(t,x)\big),$$

$t \in \mathbb{R}_{\geq 0}$, $x \in B$.

Further, let us consider the following conditions:

(H1) $\varphi\colon \mathbb{R}_{\geq 0} \times H \longrightarrow \mathbb{R}_{\geq 0}$ *is a Borel function and* $x_0 \in H$ *a point such that*

$$\inf_{t\geq 0} \inf_{\|x-x_0\|_H \geq \varepsilon} \varphi(t,x) > 0 \quad \textit{for any } \varepsilon > 0. \tag{2.1}$$

(H2) $V \in \mathfrak{K}$ *is a function satisfying*

$$\lim_{x\to x_0} \sup_{t\geq 0} V(t,x) = 0 \quad \textit{in } H, \tag{2.2}$$

$$\inf_{t\geq 0} \inf_{\|x-x_0\|_H \geq \varepsilon} V(t,x) > 0 \quad \textit{for any } \varepsilon > 0, \tag{2.3}$$

x_0 *being the point introduced in* (H1)*, and*

$$\int_H V(0,y)\,\mathrm{d}\mu(y) < \infty. \tag{2.4}$$

(H3) $\alpha, \gamma\colon \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{>0}$ *are Borel functions such that* $\alpha \in L^1_{\mathrm{loc}}(\mathbb{R}_{\geq 0}) \setminus L^1(\mathbb{R}_{\geq 0})$*,* $\gamma \in L^1(\mathbb{R}_{\geq 0})$*.*

Now we are prepared to state and prove the main theorem.

Theorem 2.1. *Suppose that* (A) *and* (C) *are satisfied and there exist functions* φ*,* V*,* α *and* γ *satisfying* (H1)–(H3) *and*

$$LV(t,x) \leq -\alpha(t)\varphi(t,x) + \gamma(t)\big[1 + V(t,x)\big] \quad \textit{for all } t \geq 0,\ x \in B. \tag{2.5}$$

Let $((\Omega, \mathscr{F}, (\mathscr{F}_t), \boldsymbol{P}), W, X)$ *be any solution to* (1.4)*, then*

$$\lim_{t\to\infty} \|X_t - x_0\|_H = 0 \quad \boldsymbol{P}\textit{-almost surely.} \tag{2.6}$$

Remark 2.2. Tracing the proof below one may check easily that – as in the finite-dimensional case – it suffices to assume instead of (2.1)–(2.3) that $V \geq 0$, there exists $\tau \geq 0$ such that

$$\lim_{x\to x_0} \sup_{t\geq\tau} V(t,x) = 0,$$

and that for any $\varepsilon > 0$ there exists $\tau' = \tau'(\varepsilon)$ such that

$$\inf_{t\geq\tau'} \inf_{\|x-x_0\|_H\geq\varepsilon} V(t,x) > 0, \quad \inf_{t\geq\tau'} \inf_{\|x-x_0\|_H\geq\varepsilon} \varphi(t,x) > 0.$$

Remark 2.3. The singleton $\{x_0\}$ may be replaced with an arbitrary closed set $\Gamma \subseteq H$. Let (2.1)–(2.3) be modified in the following way:

$$\lim_{\operatorname{dist}(x,\Gamma)\to 0} \sup_{t\geq 0} V(t,x) = 0 \quad \text{in } H,$$

$$\inf_{t\geq 0} \inf_{\operatorname{dist}(x,\Gamma)\geq\varepsilon} V(t,x) \wedge \varphi(t,x) > 0 \quad \text{for all } \varepsilon > 0$$

and let $V = 0$ on $\mathbb{R}_{\geq 0} \times \Gamma$. Then $\operatorname{dist}(X_t, \Gamma) \longrightarrow 0$ as $t \to \infty$ $\boldsymbol{P}$-almost surely. The proof requires only very straightforward changes; unfortunately, this result is usually too weak to be applied to equations with multiple roots (cf. the discussion in [NCh, Chapter 5]).

Proof. a) The first two steps of the proof are essentially known from stability theory of stochastic PDEs, but we provide them for completeness and as we shall refer to parts of the argument in the sequel. Set

$$U(t,x) = \exp\Big(\int_t^\infty \gamma(r)\,\mathrm{d}r\Big)\big[1 + V(t,x)\big], \quad (t,x) \in \mathbb{R}_{\geq 0} \times H.$$

Since $\gamma \in L^1(\mathbb{R}_{\geq 0})$, U is obviously well defined and $U \geq 0$ on $\mathbb{R}_{\geq 0} \times H$. To avoid overcomplicated formulae, we shall proceed as if γ were also continuous, i.e. $\exp(\int_\cdot^\infty \gamma(r)\,\mathrm{d}r) \in C^1(\mathbb{R}_{\geq 0})$, the general case may be handled in the same way by using (1.7) instead of (1.6). If γ is continuous, $U \in \mathfrak{K}$ and an easy calculation shows that

$$LU(t,x) \leq -\alpha(t)\varphi(t,x) \quad \text{for all } t \geq 0,\ x \in B, \tag{2.7}$$

in particular $LU \leq 0$ on $\mathbb{R}_{\geq 0} \times B$.

b) We aim at proving that $(U(t,X_t),\ t \geq 0)$ is a supermartingale. Towards this end, set

$$\tau_n = \inf\big\{t \geq 0;\ \|X_t\|_H \geq n\big\} \wedge \inf\Big\{t \geq 0;\ \int_0^t \|\sigma(r,\tilde{X}_r)\|^2_{\mathscr{L}_2} \geq n\Big\},\ n \in \mathbb{N}$$

(with the convention $\inf\emptyset = +\infty$), where $\tilde{X}$ is the process introduced in Definition 1.1. Plainly, $\tau_n \nearrow \infty$ as $n \to \infty$ $\boldsymbol{P}$-almost surely. Using the Itô formula and (2.7) we get

$$\begin{aligned} U(t\wedge\tau_n, X(t\wedge\tau_n)) - U(0,X_0) &= \int_0^{t\wedge\tau_n} LU(r,\tilde{X}_r)\,\mathrm{d}r + \int_0^{t\wedge\tau_n} \langle \sigma(r,\tilde{X}_r)^* D_x U(r,X_r), \cdot\rangle_K \,\mathrm{d}W_r \\ &\leq \int_0^{t\wedge\tau_n} \langle \sigma(r,\tilde{X}_r)^* D_x U(r,X_r), \cdot\rangle_K \,\mathrm{d}W_r. \end{aligned}$$

Note that

$$\boldsymbol{E}\int_0^{t\wedge\tau_n} \|\sigma(r,\tilde{X}_r)^* D_x U(r,X_r)\|^2_K \,\mathrm{d}r < \infty$$

for any $t \geq 0$ due to the definition of τ_n and boundedness of $D_x U$ on bounded subsets of $\mathbb{R}_{\geq 0} \times H$, as

$$\int_0^{t\wedge\tau_n} \|\sigma(r,\tilde{X}_r)^* D_x U(r,X_r)\|^2_K \,\mathrm{d}r \leq \sup_{\substack{0\leq r\leq t \\ \|z\|_H\leq n}} \|D_x U(r,z)\|^2_H \int_0^{t\wedge\tau_n} \|\sigma(r,\tilde{X}_r)\|^2_{\mathscr{L}_2} \,\mathrm{d}r.$$

We see that
$$\int_0^{\cdot\wedge\tau_n} \langle \sigma(r,\tilde{X}_r)^* D_x U(r,X_r),\cdot\rangle_K \,\mathrm{d}W_r$$
is a martingale, hence
$$\boldsymbol{E}U(t\wedge\tau_n, X(t\wedge\tau_n)) \le \boldsymbol{E}U(0,X_0) \le e^{\|\gamma\|_{L^1}}\boldsymbol{E}V(0,X_0) < \infty$$
for all $t\ge 0$ and $n\in\mathbb{N}$ by (2.4). Since $U\in C(\mathbb{R}_{\ge 0}\times H)$ and the paths of X are continuous in H we obtain
$$\begin{aligned}\boldsymbol{E}U(t,X_t) &= \boldsymbol{E}\lim_{n\to\infty} U(t\wedge\tau_n, X(t\wedge\tau_n)) \le \liminf_{n\to\infty} \boldsymbol{E}U(t\wedge\tau_n, X(t\wedge\tau_n))\\ &\le \boldsymbol{E}U(0,X_0)\end{aligned}$$
by the Fatou lemma, thus $U(t,X_t)\in L^1(\boldsymbol{P})$ for every $t\in\mathbb{R}_{\ge 0}$. Analogously, for any $0\le s\le t$ we have
$$U(t\wedge\tau_n, X(t\wedge\tau_n)) - U(s\wedge\tau_n, X(s\wedge\tau_n)) \le \int_{s\wedge\tau_n}^{t\wedge\tau_n} \langle \sigma(r,\tilde{X}_r)^* D_x U(r,X_r),\cdot\rangle_K \,\mathrm{d}W_r;$$
so
$$\boldsymbol{E}\big[U(t\wedge\tau_n, X(t\wedge\tau_n)) \,\big|\, \mathscr{F}_s\big] - U(s\wedge\tau_n, X(s\wedge\tau_n)) \le 0.$$
The Fatou lemma for conditional expectations now implies that
$$\begin{aligned}\boldsymbol{E}\big[U(t,X_t) \,\big|\, \mathscr{F}_s\big] &\le \liminf_{n\to\infty} \boldsymbol{E}\big[U(t\wedge\tau_n, X(t\wedge\tau_n)) \,\big|\, \mathscr{F}_s\big]\\ &\le \liminf_{n\to\infty} U(s\wedge\tau_n, X(s\wedge\tau_n))\\ &= U(s,X_s)\end{aligned}$$
$\boldsymbol{P}$-almost surely, which is the supermartingale property. For further use, let us note that proceeding as above we get
$$\begin{aligned}-\boldsymbol{E}\int_0^{t\wedge\tau_n} LU(r,\tilde{X}_r)\,\mathrm{d}r &= \boldsymbol{E}U(0,X_0) - \boldsymbol{E}U(t\wedge\tau_n, X(t\wedge\tau_n))\\ &\le \boldsymbol{E}U(0,X_0),\end{aligned}$$
whence
$$-\boldsymbol{E}\int_0^{\infty} LU(r,\tilde{X}_r)\,\mathrm{d}r \le \boldsymbol{E}U(0,X_0) < \infty, \tag{2.8}$$
again by the Fatou lemma.

Since $U(t,X_t)$ is a continuous nonnegative supermartingale, the martingale convergence theorem yields a random variable $U_\infty\in L^1(\boldsymbol{P})$ such that
$$\lim_{t\to\infty} U(t,X_t) = U_\infty \quad \boldsymbol{P}\text{-almost surely.}$$
From the definition of U it follows that there exists $\Omega_s\in\mathscr{F}$, $\boldsymbol{P}(\Omega_s)=1$, such that $1\le U_\infty(\omega)<\infty$ and
$$\lim_{t\to\infty} V(t,X(t,\omega)) = V_\infty(\omega) \equiv U_\infty(\omega) - 1$$
for any $\omega\in\Omega_s$.

c) Since
$$\int_0^\infty \alpha(r)\varphi(r,\tilde{X}_r)\,\mathrm{d}r \le -\int_0^\infty LU(r,\tilde{X}_r)\,\mathrm{d}r \quad \text{on } \Omega$$

by (2.7), the integral on the right-hand side is a nonnegative random variable with a finite expectation by (2.8), and $X = \tilde{X}$ λ-almost everywhere on $\mathbb{R}_{\geq 0}$ $\boldsymbol{P}$-almost surely, there exists $\Omega_i \in \mathscr{F}$, $\boldsymbol{P}(\Omega_i) = 1$, such that

$$\int_0^\infty \alpha(r)\varphi(r, X(r,\omega))\,\mathrm{d}r < \infty$$

for every $\omega \in \Omega_i$.

d) Now we check that

$$\liminf_{t\to\infty} \|X(t,\omega) - x_0\|_H = 0 \tag{2.9}$$

for all $\omega \in \Omega_i$. Striving for a contradiction assume that $\omega \in \Omega_i$ but

$$\liminf_{t\to\infty} \|X(t,\omega) - x_0\|_H > 0.$$

Then there exists $t_0 \in \mathbb{R}_{\geq 0}$ and $\varepsilon > 0$ such that $\|X(t,\omega) - x_0\|_H \geq \varepsilon$ for any $t \geq t_0$; by (H1) we may find $\delta > 0$ satisfying $\varphi(t, X(t,\omega)) \geq \delta$ for all $t \geq t_0$, therefore

$$\int_{t_0}^\infty \alpha(r)\varphi(r, X(r,\omega))\,\mathrm{d}r \geq \delta \int_{t_0}^\infty \alpha(r)\,\mathrm{d}r = +\infty$$

by (H3), however, this contradicts the definition of Ω_i.

e) It remains to prove that

$$\lim_{t\to\infty} \|X(t,\omega) - x_0\|_H = 0 \quad \text{for all } \omega \in \Omega_i \cap \Omega_s. \tag{2.10}$$

Assume that $\omega \in \Omega_i \cap \Omega_s$ but (2.10) fails. Then there exist $t_n \nearrow \infty$ and $\varepsilon > 0$ such that $\|X(t_n,\omega) - x_0\|_H \geq \varepsilon$. By (2.3), a $\eta > 0$ may be found such that $V(t_n, X(t_n,\omega)) \geq \eta$, consequently

$$\eta \leq \lim_{n\to\infty} V(t_n, X(t_n,\omega)) = V_\infty(\omega).$$

On the other hand, by (2.9) there exist $s_n \nearrow \infty$ such that $\|X(s_n,\omega) - x_0\|_H \to 0$ as $n \to \infty$, hence

$$0 \leq V_\infty(\omega) = \lim_{n\to\infty} V(s_n, X(s_n,\omega)) \leq \lim_{n\to\infty} \sup_{r\geq 0} V(r, X(s_n,\omega)) = 0$$

by (2.2). This contradiction proves (2.10) and the proof of Theorem 2.1 is completed. □

Remark 2.4. By (2.2) and Theorem 2.1,

$$\lim_{t\to\infty} V(t, X_t) = 0 \quad \boldsymbol{P}\text{-almost surely.} \tag{2.11}$$

The estimate

$$\boldsymbol{E}V(t, X_t) = e^{-\int_t^\infty \gamma\,\mathrm{d}r}\boldsymbol{E}U(t, X_t) - 1 \leq \boldsymbol{E}U(0, X_0) < \infty, \quad t \geq 0,$$

was established in the course of the proof. Therefore, if $\nu \in (0,1)$ then the set $\{V(t, X_t)^\nu,\ t \geq 0\}$ is uniformly integrable and (2.11) implies that

$$\lim_{t\to\infty} \boldsymbol{E}V(t, X_t)^\nu = 0.$$

Now we may proceed to a theorem on stochastic approximation.

Corollary 2.5. *Let* (C) *be satisfied, let* $R\colon B \longrightarrow B^*$ *and* $\sigma\colon \mathbb{R}_{\geq 0} \times B \longrightarrow \mathscr{L}_2(K, H)$ *be Borel function and* μ *a Borel probability measure on* H. *Let* $x_0 \in B$ *be such that*

$R(x_0) = 0$. *Suppose that there exist* $V \in \mathfrak{I} \cap L^1(\mu)$ *and a Borel function* $\varphi\colon H \longrightarrow \mathbb{R}_{\geq 0}$ *satisfying*

$$V(x_0) = 0, \qquad \inf_{\|x-x_0\|_H \geq \varepsilon} \bigl\{V(x) \wedge \varphi(x)\bigr\} > 0 \quad \textit{for any } \varepsilon > 0,$$
$$\bigl\langle R(x), DV(x)\bigr\rangle_{B^*,B} \leq -\varphi(x) \quad \textit{for all } x \in B, \tag{2.12}$$

and

$$\operatorname{Tr}\bigl(D^2V(x)(\sigma\sigma^*)(t,x)\bigr) \leq K\bigl(1 + V(x)\bigr) \quad \textit{for some } K < \infty \textit{ and all } (t,x) \in \mathbb{R}_{\geq 0} \times B.$$

Let $\alpha\colon \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{>0}$ *be a Borel function such that*

$$\int_0^\infty \alpha(r)\,\mathrm{d}r = \infty, \quad \int_0^\infty \alpha^2(r)\,\mathrm{d}r < \infty.$$

Then any solution $(\Omega, \mathscr{F}, (\mathscr{F}_t), \boldsymbol{P}), W, X)$ *of the equation*

$$\mathrm{d}X(t) = \alpha(t)R(X(t))\,\mathrm{d}t + \alpha(t)\sigma(t, X(t))\,\mathrm{d}W(t), \quad X(0) \sim \mu, \tag{2.13}$$

satisfies

$$\lim_{t\to\infty} \|X(t) - x_0\|_H = 0 \quad \boldsymbol{P}\textit{-almost surely.}$$

If, moreover, $V(x) \geq \varpi\|x - x_0\|_H^2$ *for some* $\varpi \in \mathbb{R}_{>0}$ *and all* $x \in H$*, then*

$$\lim_{t\to\infty} \boldsymbol{E}\|X(t) - x_0\|^\nu = 0$$

for any $\nu \in (0,2)$.

Remark 2.6. a) Note that (2.12) may be satisfied only if x_0 is the unique root of R.

b) As in Theorem 2.1, we do not assume that there exists a unique solution of (2.13), we only claim that if a solution exists, then it converges to the root of R. Of course, in examples the problem of existence and uniqueness of solutions gains prominence.

3 Examples

Example 3.1. Let $\Lambda \subseteq \mathbb{R}^d$ be a bounded open set with a sufficiently smooth boundary $\partial\Lambda$, $g\colon \mathbb{R} \longrightarrow \mathbb{R}$ a Borel function and f a (generalized) function on Λ. Let us consider a nonlinear elliptic equation

$$\Delta u + g(u) = f \text{ in } \Lambda, \quad u = 0 \text{ on } \partial\Lambda. \tag{3.1}$$

Set $H = L^2(\Lambda)$, $B = W_0^{1,2}(\Lambda)$ and denote by G the superposition operator defined by g. Assume that G is a continuous mapping from B to H and that there exists $\varrho \in \mathbb{R}$ such that

$$\bigl\langle G(u) - G(v), u - v\bigr\rangle_H \leq \varrho\|u - v\|_H^2, \tag{3.2}$$
$$\bigl\langle G(u), u\bigr\rangle_H \leq \varrho\bigl(1 + \|u\|_H^2\bigr), \quad \|G(u)\|_H \leq \varrho\bigl(1 + \|u\|_B\bigr)$$

for all $u, v \in B$. Note that (3.2) is surely satisfied if g is either Lipschitz continuous or nonincreasing. Let $\sigma\colon \mathbb{R}_{\geq 0} \times B \longrightarrow \mathscr{L}_2(K,H)$ be a Borel function such that

$$\sup_{0\leq t\leq T} \sup_{x\in B} \frac{\|\sigma(t,x)\|_{\mathscr{L}_2}}{1 + \|x\|_H} + \sup_{0\leq t\leq T} \sup_{\substack{x,y\in B\\ x\neq y}} \frac{\|\sigma(t,x) - \sigma(t,y)\|_{\mathscr{L}_2}}{\|x - y\|_H} < \infty \tag{3.3}$$

for any $T \in \mathbb{R}_{\geq 0}$. Finally, let $\alpha\colon \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{>0}$ satisfy

$$\alpha \in L^2(\mathbb{R}_{\geq 0}) \setminus L^1(\mathbb{R}_{\geq 0}) \quad \text{and} \quad 0 < \inf_{[0,T]} \alpha \leq \sup_{[0,T]} \alpha < \infty \tag{3.4}$$

for any $T \in \mathbb{R}_{\geq 0}$, let $f \in B^*$ and let μ be a Borel probability measure on H with a finite second moment, i.e. $\|\cdot\|_H \in L^2(\mu)$. Then it may be checked easily that all hypotheses of Theorem 4.2.4 in [LR] are satisfied and hence there exists a unique solution $((\Omega, \mathscr{F}, (\mathscr{F}_t), \boldsymbol{P}), W, X)$ to the stochastic parabolic equation

$$\mathrm{d}X = \alpha(t)\big(\Delta X + G(X) - f\big)\,\mathrm{d}t + \alpha(t)\sigma(t,X)\,\mathrm{d}W(t), \quad X(0) \sim \mu,$$

the Dirichlet Laplacian Δ being interpreted as an operator in $\mathscr{L}(B, B^*)$ in a natural way. Assume that there exists a weak solution $u_0 \in B$ of (3.1); one may consult e.g. [BS], [Pr, Chapter 9] or references therein for results in this direction. We want to apply Corollary 2.5 with $V = \|\cdot - u_0\|_H^2$. Since u_0 solves (3.1),

$$\big\langle \Delta u + G(u) - f, DV(u)\big\rangle_{B^*,B} = 2\big\langle \Delta(u-u_0) + G(u) - G(u_0), u - u_0\big\rangle_{B^*,B}$$

and it is known that

$$\big\langle \Delta(u-u_0), u-u_0\big\rangle_{B^*,B} \leq -\varkappa\|u-u_0\|_H^2$$

for some $\varkappa > 0$ and all $u \in B$, so Corollary 2.5 implies that

$$\lim_{t\to\infty} \|X(t) - u_0\|_H = 0 \ \boldsymbol{P}\text{-almost surely} \quad \text{and} \quad \lim_{t\to\infty} \boldsymbol{E}\|X(t)-u_0\|^{2-\varepsilon} = 0 \tag{3.5}$$

for all $\varepsilon \in (0,2)$, provided

$$\big\langle G(u) - G(u_0), u-u_0\big\rangle_H \leq (\varkappa - \eta)\|u-u_0\|_H^2 \tag{3.6}$$

for some $\eta > 0$ and all $u \in B$, and

$$\sup_{t\geq 0}\sup_{x\in B} \frac{\|\sigma(t,x)\|_{\mathscr{L}_2}}{1+\|x\|_H} < \infty. \tag{3.7}$$

As we have already mentioned, (3.6) is satisfied if g is either nonincreasing, or Lipschitz continuous with a sufficiently small Lipschitz constant.

Example 3.2. Let $\varLambda \subseteq \mathbb{R}^d$ be a bounded domain with a sufficiently smooth boundary and $p \in (2,\infty)$. Set $B = W_0^{1,p}(\varLambda)$ and $H = L^2(\varLambda)$, we shall consider the p-Laplacian

$$\Delta_p u = \operatorname{div}\big(|\nabla u|^{p-2}\nabla u\big),$$

that is, rigorously, an operator $\Delta_p\colon B \longrightarrow B^*$ defined by

$$\langle \Delta_p u, v\rangle_{B^*,B} = -\int_D |\nabla u(r)|^{p-2}\langle \nabla u(r), \nabla v(r)\rangle\,\mathrm{d}r, \quad u,v \in B.$$

Let $f \in H$. It follows from [BS, Theorem 2.6.8] that the quasilinear elliptic equation

$$\Delta_p u = f \text{ in } \varLambda, \quad u = 0 \text{ on } \partial\varLambda \tag{3.8}$$

has a unique weak solution $u_0 \in B$. Likewise, the stochastic equation

$$\mathrm{d}X = \alpha(t)\big(\Delta_p X - f\big)\,\mathrm{d}t + \alpha(t)\sigma(t,X)\,\mathrm{d}W(t), \quad X(0) \sim \mu$$

has a unique variational solution $((\Omega, \mathscr{F}, (\mathscr{F}_t), \boldsymbol{P}), W, X)$ if α and σ satisfy (3.4) and (3.3), respectively, and μ is a Borel probability measure on H with a finite second moment, see the discussion in [LR, Example 4.1.9]. Again we shall use Corollary 2.5 with $V = \|\cdot - u_0\|_H^2$. Due to the inequality

$$\big\langle \|t\|^{p-2}t - \|s\|^{p-2}s, t-s\big\rangle \geq c_p\|t-s\|^p \quad \text{for a } c_p > 0 \text{ and all } s,t \in \mathbb{R}^d \tag{3.9}$$

(see e.g. [Si, p. 210]) the operator $-\Delta_p$ is strongly monotone,

$$\begin{aligned}\langle \Delta_p u - \Delta_p v, u - v\rangle_{B^*,B} &= -\int_G \big\langle \|\nabla u(r)\|^{p-2}\nabla u(r) - \|\nabla v(r)\|^{p-2}\nabla v(r), u(r) - v(r)\big\rangle \,\mathrm{d}r \\ &\leq -c_p\|u-v\|_B^p\end{aligned}$$

for all $u, v \in B$, whence we have

$$\begin{aligned}\langle \Delta_p u - f, DV(u)\rangle_{B^*,B} &= 2\langle \Delta_p u - \Delta_p u_0, u - u_0\rangle_{B^*,B} \\ &\leq -2c_p\|u-u_0\|_B^p \\ &\leq -\tilde{c}\,\|u-u_0\|_H^p\end{aligned}$$

for some constant $\tilde{c} > 0$ and all $u \in B$. Therefore, if σ satisfies (3.7) then (3.5) holds true for all $\varepsilon \in (0,2)$.

Example 3.3. In this example equations involving the porous medium operator will be considered. Let $\Lambda \subseteq \mathbb{R}^d$ be a bounded domain with a sufficiently smooth boundary and $p \in (2,\infty)$, set $B = L^p(\Lambda)$, $H = (W_0^{1,2}(\Lambda))^*$ and $\Psi(s) = s|s|^{p-2}$ for $s \in \mathbb{R}$, and define

$$A\colon B \longrightarrow B^*,\ u \longmapsto \Delta\Psi(u).$$

Details may be found e.g. in [LR, Example 4.1.11]; note that the dualities appearing in this example must be handled with some care, in particular, $\langle \Delta u, v\rangle_{B^*,B} = -\langle u, v\rangle_{L^{p/(p-1)},L^p}$ for all $u \in L^{p/(p-1)}(D)$, $v \in L^p(D)$. Let $\sigma\colon \mathbb{R}_{\geq 0} \times B \longrightarrow \mathscr{L}_2(K,H)$ and $\alpha\colon \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{>0}$ satisfy (3.3) and (3.4), respectively, let $f \in B^*$ and let μ be a Borel probability measure on H with a finite second moment. Then there exists a unique solution X of

$$\mathrm{d}X = \alpha(t)\big(A(X) - f\big)\,\mathrm{d}t + \alpha(t)\sigma(t,X)\,\mathrm{d}W(t), \quad X(0) \sim \mu.$$

Using the inequality (3.9) with $d = 1$ one may check that $-A$ is strongly monotone:

$$\begin{aligned}\langle A(u) - A(v), u - v\rangle_{B^*,B} &= \langle \Delta(\Psi(u) - \Psi(v)), u - v\rangle_{B^*,B} \\ &= -\int_D (\Psi(u(r)) - \Psi(v(r)))(u(r) - v(r))\,\mathrm{d}r \\ &\leq -c_p\|u-v\|_B^p.\end{aligned}$$

It follows, first, that the problem $Au = f$ has a unique solution $u_0 \in B$ and, secondly, choosing $V = \|\cdot - u_0\|_H^2$ we get

$$\langle A(u) - f, DV(u)\rangle_{B^*,B} \leq -\hat{c}\,\|u-u_0\|_H^p$$

for some $\hat{c} > 0$ and any $u \in B$. Therefore, (3.5) holds provided (3.7) is satisfied.

References

[BS] M. Badiale, E. Serra: *Semilinear elliptic equations for beginners*, Springer, London 2011. MR-2722059

[BYY] J. Bao, G. Yin, C. Yuan: Ergodicity for functional stochastic differential equations and applications, *Nonlinear Anal. 98*(2014), 66–82. MR-3158446

[BRS] K. Barty, J.-S. Roy, C. Strugarek: Hilbert-valued perturbed subgradient algorithms, *Math. Oper. Res. 32* (2007), 551–562. MR-2348233

[BMP] A. Benveniste, M. Métivier, P. Priouret: *Adaptive algorithms and stochastic approximations*, Springer, Berlin 1990. MR-1082341

[BPP] S. Bhatnagar, H. L. Prasad, L. A. Prashanth: *Stochastic recursive algorithms for optimization*, Lecture Notes in Control Inform. Sci. 434, Springer, London 2013. MR-2963530

[Bo] V. S. Borkar: *Stochastic approximation: A dynamical systems viewpoint*, Cambridge University Press, Cambridge 2008. MR-2442439

[Che] H.-F. Chen: Continuous-time stochastic approximation: Convergence and asymptotic efficiency, *Stochastics Stochastic Rep. 51* (1994), 217–239. MR-1380765

[Che2] H.-F. Chen: *Stochastic approximation and its applications*, Kluwer, Dordrecht 2002. MR-1942427

[Cho] P.-L. Chow: *Stochastic partial differential equations*, 2nd ed., CRC Press, Boca Raton 2015. MR-3288853

[Go] L. Goldstein: Minimizing noisy functionals in Hilbert space: An extension of the Kiefer-Wolfowitz procedure, *J. Theoret. Probab. 1* (1988), 189–204. MR-0938258

[Ko] A. P. Korostelev: *Stochastic recurrent procedures. Local properties* (Russian), Nauka, Moscow 1984. MR-0756811

[Kry] N. V. Krylov: A relatively short proof of Itô's formula for SPDEs and its applications, *Stoch. Partial Differ. Equ. Anal. Comput. 1* (2013), 152–174. MR-3327504

[KR] N. V. Krylov, B. L. Rozovskiĭ: Stochastic evolution equations (Russian), *Current problems in mathematics, Vol. 14*, pp. 71–147, VINITI, Moscow 1979 (English translation: *J. Soviet Math. 16* (1981), 1233–1277). MR-0570795

[KC] H. J. Kushner, D. S. Clark: *Stochastic approximation methods for constrained and unconstrained systems*, Springer, New York 1978. MR-0499560

[KS] H. J. Kushner, A. Shwartz: Stochastic approximation in Hilbert space: Identification and optimization of linear continuous parameter systems, *SIAM J. Control Optim. 23* (1985), 774–793. MR-0798059

[KY] H. J. Kushner, G. G. Yin: *Stochastic approximation and recursive algorithms and applications*, 2nd ed., Springer, New York 2003. MR-1993642

[La1] N. Lazrieva, T. Sharia, T. Toronjadze: The Robbins-Monro type stochastic differential equations. I. Convergence of solutions, *Stochastics Stochastics Rep. 61* (1997), 67–87. MR-1473914

[La2] N. Lazrieva, T. Sharia, T. Toronjadze: The Robbins-Monro type stochastic differential equations. II. Asymptotic behaviour of solutions, *Stoch. Stoch. Rep. 75* (2003), 153–180. MR-1978897

[La3] N. Lazrieva, T. Sharia, T. Toronjadze: Semimartingale stochastic approximation procedure and recursive estimation, *J. Math. Sci. (N.Y.) 153* (2008), 211–261. MR-2731941

[La4] N. Lazrieva, T. Toronjadze: The Robbins-Monro type stochastic differential equations. III. Polyak's averaging, *Stochastics 82* (2010), 165–188. MR-2677544

[LR] W. Liu, M. Röckner: *Stochastic partial differential equations: An introduction*, Springer, Cham 2015. MR-3410409

[NCh] M. B. Nevel'son, R. Z. Khas'minskiĭ: *Stochastic approximation and recurrent estimation* (Russian), Nauka, Moscow 1972 (English translation: M. B. Nevel'son, R. Z. Has'minskiĭ: *Stochastic approximation and recursive estimation*, American Mathematical Society, Providence 1973). MR-0391442

[Ni] R. Nixdorf: An invariance principle for a finite dimensional stochastic approximation method in a Hilbert space, *J. Multivariate Anal. 15* (1984), 252–260. MR-0763600

[Pa] E. Pardoux: *Équations aux dérivées partielles stochastiques non linéaires monotones. Etude de solutions fortes de type Ito*, Thèse, Université Paris Sud, 1975, `https://www.i2m.univ-amu.fr/~pardoux/Pardoux_these.pdf`

[Pf] G. Pflug: Stetige stochastische Approximation, *Metrika 26* (1979), 139–150. MR-0548412

[Pr] R. Precup: *Linear and semilinear partial differential equations*, Walter de Gruyter & Co., Berlin 2013. MR-2986215

[Si] J. Simon: Régularité de la solution d'une équation non linéaire dans $\mathbb{R}^N$, *Journées d'Analyse Non Linéaire (Besançon, 1977)*, pp. 205–227, Lecture Notes in Math. 665, Springer, Berlin 1978. MR-0519432

[Yi] G. Yin: On H-valued stochastic approximation: Finite dimensional projections, *Stochastic Anal. Appl. 10*(1992), 363–377. MR-1166794

[YZ] G. Yin, Y. M. Zhu: On H-valued Robbins-Monro processes, *J. Multivariate Anal. 34*(1990), 116–140. MR-1062551

Acknowledgments. We are indebted to M. Ondreját for very useful discussions concerning this paper.

Electronic Journal of Probability
Electronic Communications in Probability

Advantages of publishing in EJP-ECP

- Very high standards
- Free for authors, free for readers
- Quick publication (no backlog)
- Secure publication (LOCKSS[1])
- Easy interface (EJMS[2])

Economical model of EJP-ECP

- Non profit, sponsored by IMS[3], BS[4] , ProjectEuclid[5]
- Purely electronic

Help keep the journal free and vigorous

- Donate to the IMS open access fund[6] (click here to donate!)
- Submit your best articles to EJP-ECP
- Choose EJP-ECP over for-profit journals

[1]LOCKSS: Lots of Copies Keep Stuff Safe http://www.lockss.org/
[2]EJMS: Electronic Journal Management System http://www.vtex.lt/en/ejms.html
[3]IMS: Institute of Mathematical Statistics http://www.imstat.org/
[4]BS: Bernoulli Society http://www.bernoulli-society.org/
[5]Project Euclid: https://projecteuclid.org/
[6]IMS Open Access Fund: http://www.imstat.org/publications/open.htm