Stochastic Approximation Procedures for Lévy-Driven SDEs

Jan Seidler[1] · Ondřej Týbl[2]

Abstract

We consider a continuous-time Robbins–Monro-type stochastic approximation procedure for a system described by a (multidimensional) stochastic differential equation driven by a general Lévy process, and we find sufficient conditions for its convergence in terms of Lyapunov functions. While the jump part of the noise may spoil convergence to the root of the drift in some cases, we show that by a suitable choice of noise coefficients we obtain convergence under hypotheses on the drift weaker than those used in the diffusion case or convergence to a selected root in the case of multiple roots of the drift.

Keywords Stochastic approximation algorithms · Robbins–Monro procedure · Lévy-driven stochastic differential equations

Mathematics Subject Classification 60H10 · 62L20

1 Introduction

Stochastic approximation algorithms concern convergence of sequences (Y_n) of random variables defined recursively, i.e., by a stochastic difference equation $Y_{n+1} = Y_n + \alpha_n U_n$ where U_n's represent noisy observations and the step sizes $\alpha_n > 0$ satisfy suitable smallness assumptions. Originally proposed as a tool for finding a root of a function (the Robbins–Monro procedure) or its minimum (the Kiefer–Wolfowitz

✉ Ondřej Týbl
   tybl@karlin.mff.cuni.cz

   Jan Seidler
   seidler@utia.cas.cz

1 The Czech Academy of Sciences, Institute of Information Theory and Automation, Pod
   Vodárenskou Věží 4, 182 00 Prague 8, Czech Republic

2 Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics,
   Sokolovská 49/83, 186 75 Prague 8, Czech Republic

procedure), these algorithms found various applications in optimization and machine learning. See, e.g., the books [2–4, 7, 13, 14] for a thorough discussion of various aspects of stochastic approximation algorithms and their use. (Let us mention also [8, Chapter 8] for very recent applications to variational inequalities with random data.)

Nevel'son and Khas'minskii developed a continuous-time approach to stochastic approximation, which in the case of the Robbins–Monro-type procedure leads to a stochastic differential equation

$$\mathrm{d}Y_t = \alpha(t)\big(R(Y_t)\,\mathrm{d}t + \sigma(t, Y_t)\,\mathrm{d}W_t\big) \tag{1}$$

driven by a Wiener process W. Having advanced tools of stochastic analysis at their disposal—in particular the Lyapunov functions method from the stability theory of stochastic differential equations—they showed that sufficient conditions on coefficients of (1) implying convergence of its solutions almost surely as $t \to \infty$ to the (unique) root of the drift R may be found and proved in a straightforward and transparent way. See their book [21] for a systematic development of these ideas and, for example, the papers [6, 11, 22] and the book [12] for further results on continuous-time stochastic approximation.

As discrete-time systems indicate, it is reasonable to consider more general driving noises in Eq. (1). Stochastic recursive procedures described by equations driven by semimartingales were considered by Mel'nikov [20] and Lazrieva et al. [15–18]. Precise statements of their results are rather technical, but roughly speaking, the martingale part of the driving noise is a locally square integrable martingale or a random measure like a compensated Poisson random measure; proofs in these papers are based on results on convergence of semimartingales. A number of results concerning equations driven by square integrable processes with independent increments are stated in the book [12]; proofs, using Lyapunov functions techniques, are given, however, only in the discrete-time case.

In our paper, we shall study equations of the type (1) but driven by a general (multidimensional) Lévy process. Owing to the Lévy–Itô decomposition, such an equation may be written as

$$\begin{aligned}
\mathrm{d}X_t = \alpha(t)\Big(& R(X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t + \int_{\{|y|<c\}} H(X_{t-}, y)\,\tilde{N}(\mathrm{d}t, \mathrm{d}y) \\
& + \int_{\{|y|\geq c\}} K(X_{t-}, y)\,N(\mathrm{d}t, \mathrm{d}y)\Big),
\end{aligned} \tag{2}$$

where N and $\tilde{N}$ are an uncompensated and a compensated Poisson random measures, respectively, and W is a Wiener process. Compared with the available results, we admit a non-compensated Poisson process as a driving noise and essentially no hypotheses of the L^2-integrability type are needed. Employing the Lyapunov functions approach, we generalize results on convergence of the Robbins–Monro procedure from [21] to Eq. (2). It may look odd that the noise in Eq. (2) is not centered since then the last term on the right-hand side influences the drift R (e.g., if c is changed) and hence also its roots. Indeed, it may happen that solutions of (2) converge to a given point

which, however, is not a root of R. Nevertheless, a nontrivial class of coefficients H and K exists such that solutions to (2) converge to the root of R under conditions weaker than those used in the diffusion case (1) as no monotonicity-type hypotheses are needed. Moreover, in the case of a drift with multiple roots, by choosing K in a suitable way we may select a unique root of R the solutions will converge to. Again, in the diffusion case the behavior is different. In Remark 4.1, we discuss the differences between behavior of solutions to (1) and (2) in detail.

Let us note that the coefficients H and K is (2) are defined on disjoint sets $\mathbb{R}_{\geq 0} \times \mathbb{R}^m \times \{|y| < c\}$ and $\mathbb{R}_{\geq 0} \times \mathbb{R}^m \times \{|y| \geq c\}$, respectively, so we may—and will—treat them as restrictions of a single function defined on $\mathbb{R}_{\geq 0} \times \mathbb{R}^m \times \mathbb{R}^n$. This convention simplifies the form of the Itô formula.

In the next section, we introduce the equation we deal with precisely and we state the Itô formula in a form required in our proofs. In Sect. 2, the main results are proved: Theorem 3.1 giving general sufficient conditions for convergence of solutions to a stochastic differential equation driven by a Lévy process to a singleton and its Corollary 3.1 concerning the Robbins–Monro procedure, i.e., the problem (2). In Sect. 3, we show how to apply these results to particular systems.

In the rest of this section, let us introduce some notations to be used in the sequel. We set $\mathbb{R}_{\geq 0} = [0, \infty)$ and $\mathbb{R}_{>0} = (0, \infty)$. By $\mathbb{R}^{m \times n}$, we denote the space of all $m \times n$ matrices with real entries. If $A \in \mathbb{R}^{m \times n}$, then $A^T \in \mathbb{R}^{n \times m}$ is the transpose of the matrix A. Further, we denote by $\mathscr{C}_b(\mathbb{R}^m; \mathbb{R}^k)$ the set of all bounded continuous $\mathbb{R}^k$-valued functions on $\mathbb{R}^m$, and by $\| \cdot \|_\infty$ its norm, i.e., $\|u\|_\infty = \sup_{\mathbb{R}^m} |u|$. Let $\mathscr{C}^2(\mathbb{R}^m)$ be the space of all continuous real-valued functions on $\mathbb{R}^m$ having two continuous derivatives, and let the first and second Fréchet derivatives of $V \in \mathscr{C}^2(\mathbb{R}^m)$ be denoted by DV and D^2V, respectively.

2 Preliminaries

Let $m, n \in \mathbb{N}$ and suppose that Borel functions

$$f: \mathbb{R}_{\geq 0} \times \mathbb{R}^m \longrightarrow \mathbb{R}^m, \ g: \mathbb{R}_{\geq 0} \times \mathbb{R}^m \longrightarrow \mathbb{R}^{m \times n}, \ H: \mathbb{R}_{\geq 0} \times \mathbb{R}^m \times \mathbb{R}^n \longrightarrow \mathbb{R}^m,$$

and a Borel probability measure μ on $\mathbb{R}^m$ are given. We consider the equation

$$
\begin{aligned}
\mathrm{d}X_t = {} & f(t, X_t)\, \mathrm{d}t + g(t, X_t)\, \mathrm{d}W_t + \int_{\{y \in \mathbb{R}^n; \ |y| < c\}} H(t, X_{t-}, y)\, \tilde{N}(\mathrm{d}t, \mathrm{d}y) \\
& + \int_{\{y \in \mathbb{R}^n; \ |y| \geq c\}} H(t, X_{t-}, y)\, N(\mathrm{d}t, \mathrm{d}y), \qquad t \geq 0, \\
& X_0 \sim \mu,
\end{aligned}
\tag{3}
$$

for some $c \in \mathbb{R}_{>0}$ and a pair (W, N), where N is a Poisson random measure, $\tilde{N}$ is its compensated counterpart, and W is a Wiener process independent of N, see, e.g., [1, Section 2.3.1]. More precisely, recalling that a Borel measure on $\mathbb{R}^n \setminus \{0\}$ is called a Lévy measure if

$$\int_{\mathbb{R}^n \setminus \{0\}} \left(|y|^2 \wedge 1 \right) \nu(\mathrm{d}y) < \infty$$

we define a solution of (3) as follows.

Definition 2.1 A triplet $((\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geq 0}, \mathbb{P}), (W, N), X)$ is called a solution to Eq. (3) provided

(i) $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t \geq 0}, \mathbb{P})$ is a stochastic basis with a normal filtration $(\mathscr{F}_t)_{t \geq 0}$,

(ii) W is an $(\mathscr{F}_t)$-Wiener process with values in $\mathbb{R}^n$,

(iii) N is an $(\mathscr{F}_t)$-Poisson random measure N on $\mathbb{R}_{\geq 0} \times (\mathbb{R}^n \setminus \{0\})$ whose intensity is $\mathrm{d}t \, \nu(\mathrm{d}y)$ for some Lévy measure ν on $\mathbb{R}^n \setminus \{0\}$ and which is independent of W,

(iv) $\tilde{N} = N - \mathrm{d}t \, \nu(\mathrm{d}y)$, and

(v) X is an $\mathbb{R}^m$-valued $(\mathscr{F}_t)$-progressively measurable càdlàg process such that the distribution of X_0 is μ and

$$
\begin{aligned}
X_t = X_0 &+ \int_0^t f(s, X_s) \, \mathrm{d}s + \int_0^t g(s, X_s) \, \mathrm{d}W_s \\
&+ \int_0^t \int_{\{|y| < c\}} H(s, X_{s-}, y) \, \tilde{N}(\mathrm{d}s, \mathrm{d}y) \\
&+ \int_0^t \int_{\{|y| \geq c\}} H(s, X_{s-}, y) \, N(\mathrm{d}s, \mathrm{d}y) \quad \mathbb{P}\text{-a.s.}
\end{aligned}
$$

for all $t \in \mathbb{R}_{\geq 0}$.

In paragraph (v) of Definition 2.1, it is supposed implicitly that all integrals are well defined, that is,

$$\int_0^t \left\{ |f(s, X_s)| + |g(s, X_s)|^2 + \int_{\{|y| < c\}} |H(s, X_s, y)|^2 \, \nu(\mathrm{d}s) \right\} \mathrm{d}s < \infty \quad \mathbb{P}\text{-a.s.}$$

for all $t \geq 0$.

Throughout the paper, we impose the following assumption:

Assumption 2.1 We shall assume that

$$\int_{\{|y| < c\}} |H(t, x, y)|^2 \, \nu(\mathrm{d}y) < \infty \quad \text{for all } (t, x) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^m \tag{4}$$

and the function

$$(t, x) \longmapsto \int_{\{|y| \geq c\}} |H(t, x, y)| \, \nu(\mathrm{d}y) \tag{5}$$

is locally bounded on $\mathbb{R}_{\geq 0} \times \mathbb{R}^m$.

Now, let us set

$$\mathscr{V} = \left\{ V \in \mathscr{C}^2(\mathbb{R}^m); \ DV \in \mathscr{C}_b(\mathbb{R}^m; \mathbb{R}^m), \ D^2 V \in \mathscr{C}_b(\mathbb{R}^m; \mathbb{R}^{m \times m}) \right\} \tag{6}$$

and introduce an operator $\mathscr{L}$ associated with Eq. (3) that will henceforth play a crucial role. For $V \in \mathscr{V}$, we define

$$
\begin{aligned}
\mathscr{L}V : &\; \mathbb{R}_{\geq 0} \times \mathbb{R}^m \longrightarrow \mathbb{R}, \\
(t, x) \longmapsto &\; \langle f(t, x), DV(x) \rangle + \frac{1}{2} \operatorname{Tr}\left(g(t, x)^T D^2 V(x) g(t, x) \right) \\
&\; + \int_{\mathbb{R}^n \setminus \{0\}} \big[V(x + H(t, x, y)) - V(x) \\
&\; - \mathbf{1}_{\{|y| < c\}}(y) \langle H(t, x, y), DV(x) \rangle \big] \nu(\mathrm{d}y).
\end{aligned}
\tag{7}
$$

Using hypotheses (4) and (5), we can check easily that the definition of $\mathscr{L}$ is correct, see analogous considerations in the proof of Proposition 2.1.

Remark 2.1 (a) Assumption (4) can be omitted if we define $\mathscr{L}V$ as a function on the set $\{(t, x) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^m ; \text{ the right-hand side of (7) makes sense}\}$. It is a direct consequence of the integrability condition in part (v) of Definition 2.1. We only adopted (4) so that the formulation of our main results may be more straightforward.

(b) On the other hand, (5) is important and cannot be dispensed with easily. In a companion paper [19], related results on stability of solutions to (3) are obtained under a weaker hypothesis that

$$
(t, x) \longmapsto \int_{\{|y| \geq c\}} |H(t, x, y)|^p \, \nu(\mathrm{d}y) \quad \text{is locally bounded on } \mathbb{R}_{\geq 0} \times \mathbb{R}^m \tag{8}
$$

for some $p \in (0, 1)$. The same choice is possible in the present paper. Under (8), we have to restrict ourselves to a narrower class of Lyapunov functions than $\mathscr{V}$, proofs become rather complicated while the gain is not very impressive: the final criterion for convergence of the Robbins–Monro procedure remains almost the same. That is why we opted for (5).

Using the operator $\mathscr{L}$, we can state the Itô formula for smooth functions of solutions to (3) in a suitable form.

Proposition 2.1 *Assume that $V \in \mathscr{V}$ and X solves (3), then*

$$
\begin{aligned}
\mathrm{d}V(X_t) = &\; \mathscr{L}V(t, X_t)\mathrm{d}t + \langle g(t, X_t)^T DV(X_t), \cdot \rangle \mathrm{d}W_t \\
&\; + \int_{\{|y| < c\}} \big[V(X_{t-} + H(t, X_{t-}, y)) - V(X_{t-}) \big] \tilde{N}(\mathrm{d}t, \mathrm{d}y) \\
&\; + \int_{\{|y| \geq c\}} \big[V(X_{t-} + H(t, X_{t-}, y)) - V(X_{t-}) \big] N(\mathrm{d}t, \mathrm{d}y) \\
&\; - \int_{\{|y| \geq c\}} \big[V(X_t + H(t, X_t, y)) - V(X_t) \big] \nu(\mathrm{d}y)\,\mathrm{d}t.
\end{aligned}
\tag{9}
$$

Proof By [1, Theorem 4.4.7], we have

$$
\begin{aligned}
dV(X_t) = & \Big(\langle f(t, X_t), DV(X_t) \rangle + \frac{1}{2} \operatorname{Tr} \Big(g(t, X_t)^T D^2 V(X_t) g(t, X_t) \Big) \Big) dt \\
& + \langle g(t, X_t)^T DV(X_t), \cdot \rangle dW_t \\
& + \int_{\{|y|<c\}} \big[V(X_{t-} + H(t, X_{t-}, y)) - V(X_{t-}) \big] \tilde{N}(dt, dy) \\
& + \int_{\{|y|<c\}} \big[V(X_t + H(t, X_t, y)) - V(X_t) \\
& \qquad - \langle DV(X_t), H(t, X_t, y) \rangle \big] \nu(dy) \, dt \\
& + \int_{\{|y|\geq c\}} \big[V(X_{t-} + H(t, X_{t-}, y)) - V(X_{t-}) \big] N(dt, dy).
\end{aligned}
\tag{10}
$$

Now adding and substracting

$$
\int_0^t \int_{\{|y|\geq c\}} \big[V(X_s + H(s, X_s, y)) - V(X_s) \big] \nu(dy) \, ds,
\tag{11}
$$

to the right-hand side of (10) we obtain the formula (9) provided (11) is well defined for every $t \geq 0$ $\mathbb{P}$-almost surely. However, realizing that $\theta \longmapsto V(x + \theta H(s, x, y))$ is a smooth function on $[0, 1]$ and invoking boundedness of DV, we get

$$
\begin{aligned}
& \int_{\{|y|\geq c\}} \big| V(x + H(s, x, y)) - V(x) \big| \nu(dy) \\
& = \int_{\{|y|\geq c\}} \left| \int_0^1 \langle DV(x + \theta H(s, x, y)), H(s, x, y) \rangle \, d\theta \right| \nu(dy) \\
& \leq \|DV\|_\infty \int_{\{|y|\geq c\}} \big| H(s, x, y) \big| \nu(dy)
\end{aligned}
$$

for all $x \in \mathbb{R}^m$ and $s \in \mathbb{R}_{\geq 0}$. Hence,

$$
\int_0^t \int_{\{|y|\geq c\}} \big| V(X_s + H(s, X_s, y)) - V(X_s) \big| \nu(dy) \, ds < \infty \quad \mathbb{P}\text{-a.s.}
$$

follows by (5) since the paths of X are locally bounded. $\qquad\square$

3 Main Results

In this section, we first state a criterion based on Lyapunov functions for a solution to (3) to converge to a given point of the state space $\mathbb{R}^m$. The following theorem and its corollary generalize results from [21] to equations driven by Lévy processes.

Theorem 3.1 *Let Assumption [2.1] be satisfied and let there exist $x_0 \in \mathbb{R}^m$, a measurable function $\varphi : \mathbb{R}^m \longrightarrow \mathbb{R}_{\geq 0}$, a function $V \in \mathscr{V}$, and measurable functions $\alpha, \gamma : \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{>0}$ such that*

(H1) either

$$\inf_{|x-x_0|\geq\varepsilon} \varphi(x) > 0 \quad \text{for all } \varepsilon > 0 \tag{12}$$

or

$$\lim_{|x|\to\infty} V(x) = +\infty \quad \text{and} \quad \inf_{\varrho \geq |x-x_0|\geq\varepsilon} \varphi(x) > 0 \quad \text{for all } \varrho > \varepsilon > 0, \tag{13}$$

(H2) $V(x_0) = 0$, $V \in L^1(\mu)$ and

$$\inf_{|x-x_0|\geq\varepsilon} V(x) > 0 \tag{14}$$

for any $\varepsilon > 0$,
(H3) $\alpha \in L^1_{\mathrm{loc}}(\mathbb{R}_{\geq 0}) \setminus L^1(\mathbb{R}_{\geq 0})$, $\gamma \in L^1(\mathbb{R}_{\geq 0}) \cap \mathscr{C}(\mathbb{R}_{\geq 0})$ and

$$\mathscr{L}V(t,x) \leq -\alpha(t)\varphi(x) + \gamma(t)(1 + V(x)) \tag{15}$$

for all $t \geq 0$ and $x \in \mathbb{R}^m$.
Then, any solution $(\Omega, \mathscr{F}, (\mathscr{F}_t), (W, N), X)$ to [(3)] satisfies

$$\lim_{t\to\infty} X_t = x_0 \quad \mathbb{P}\text{-a.s.} \tag{16}$$

Proof Let us set

$$\xi(t) = \exp\left(\int_t^\infty \gamma(r)dr\right), \quad t \in \mathbb{R}_{\geq 0},$$

and

$$U(t,x) = \xi(t)(1 + V(x)) = \exp\left(\int_t^\infty \gamma(r)dr\right)(1 + V(x)), \quad (t,x) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^m.$$

Step 1 We establish convergence of $V(X_t)$ as $t \to \infty$. To this end, we first show that $(U(t, X_t))_{t\geq 0}$ is a supermartingale. Define

$$
\begin{aligned}
\tau_n^1 &= \inf\{t \geq 0 : |X_t| > n\}, \\
\tau_n^2 &= \inf\left\{t \geq 0 : \int_0^t |g(s, X_s)|^2 \, ds > n\right\}, \\
\tau_n^3 &= \inf\left\{t \geq 0 : \int_0^t \int_{\{|y|<c\}} |H(s, X_s, y)|^2 \, \nu(dy) \, ds > n\right\}, \\
\tau_n &= \tau_n^1 \wedge \tau_n^2 \wedge \tau_n^3
\end{aligned}
\tag{17}
$$

for $n \in \mathbb{N}$. Obviously, τ_n's are stopping times and $\tau_n \to \infty$ $\mathbb{P}$-almost surely as $n \to \infty$.

By the product rule for semimartingales, we get

$$dU(t, X_t) = (1 + V(X_t)) \, d\xi(t) + \xi(t) \, dV(X_t), \quad t \in \mathbb{R}_{\geq 0}. \tag{18}$$

Hence, combining (9) and (18), we obtain for any $n \in \mathbb{N}$ and $t \in \mathbb{R}_{\geq 0}$ (fixed but arbitrary)

$$
\begin{aligned}
&U(\tau_n \wedge t, X_{\tau_n \wedge t}) - U(0, X_0) \\
&= \int_0^{\tau_n \wedge t} \left[(1 + V(X_s))\xi'(s) + \xi(s)\mathscr{L}V(s, X_s) \right] ds \\
&\quad + \int_0^{\tau_n \wedge t} \xi(s)\langle g(s, X_s)^T DV(X_s), \cdot \rangle \, dW_s \\
&\quad + \int_0^{\tau_n \wedge t} \int_{\{|y|<c\}} \xi(s) \left[V(X_{s-} + H(s, X_{s-}, y)) - V(X_{s-}) \right] \tilde{N}(ds, dy) \\
&\quad + \int_0^{\tau_n \wedge t} \int_{\{|y|\geq c\}} \xi(s) \left[V(X_{s-} + H(s, X_{s-}, y)) - V(X_{s-}) \right] N(ds, dy) \\
&\quad - \int_0^{\tau_n \wedge t} \int_{\{|y|\geq c\}} \xi(s) \left[V(X_s + H(t, X_s, y)) - V(X_s) \right] \nu(dy) \, ds.
\end{aligned}
\tag{19}
$$

By the hypothesis (H3), we may estimate

$$
\begin{aligned}
&\int_0^{\tau_n \wedge t} \left[(1 + V(X_s))\xi'(s) + \xi(s)\mathscr{L}V(s, X_s) \right] ds \\
&= \int_0^{\tau_n \wedge t} \xi(s)\left\{ -\gamma(s)(1 + V(X_s)) + \mathscr{L}(s, X_s) \right\} ds \\
&\leq - \int_0^{\tau_n \wedge t} \xi(s)\alpha(s)\varphi(X_s) \, ds \\
&\leq 0
\end{aligned}
\tag{20}
$$

as α and φ are nonnegative. Therefore, from (19) we get

$$
\begin{aligned}
&U(\tau_n \wedge t, X_{\tau_n \wedge t}) - U(0, X_0) \\
&\leq \int_0^{\tau_n \wedge t} \xi(s)\langle g(s, X_s)^T DV(X_s), \cdot \rangle \, dW_s \\
&\quad + \int_0^{\tau_n \wedge t} \int_{\{|y|<c\}} \xi(s) \left[V(X_{s-} + H(s, X_{s-}, y)) - V(X_{s-}) \right] \tilde{N}(ds, dy) \\
&\quad + \int_0^{\tau_n \wedge t} \int_{\{|y|\geq c\}} \xi(s) \left[V(X_{s-} + H(s, X_{s-}, y)) - V(X_{s-}) \right] N(ds, dy) \\
&\quad - \int_0^{\tau_n \wedge t} \int_{\{|y|\geq c\}} \xi(s) \left[V(X_s + H(t, X_s, y)) - V(X_s) \right] \nu(dy) \, ds.
\end{aligned}
\tag{21}
$$

We aim at showing that the right-hand side of (21) is a martingale for any $n \in \mathbb{N}$. This having been established we find that

$$\mathbb{E}\left[U(t \wedge \tau_n, X_{t \wedge \tau_n}) - U(0, X_0) \right] \leq 0,$$

so we may apply the Fatou lemma and arrive at

$$
\begin{aligned}
\mathbb{E}\, U(t, X_t) = \mathbb{E}\, \lim_{n \to \infty} U(t \wedge \tau_n, X_{t \wedge \tau_n}) &\leq \liminf_{n \to \infty} \mathbb{E}\, U(t \wedge \tau_n, X_{t \wedge \tau_n}) \\
&\leq \mathbb{E}\, U(0, X_0) \\
&= e^{\|\gamma\|_{L^1}} \mathbb{E}\, V(X_0) < \infty
\end{aligned}
$$

for every $t \in \mathbb{R}_{\geq 0}$, as $V \in L^1(\mu)$. Using the Fatou lemma for conditional expectations, we get in a completely analogous way that $(U(t, X_t),\ t \in \mathbb{R}_{\geq 0})$ is a supermartingale, we skip the details.

Hence, now we fix $n \in \mathbb{N}$ and we shall proceed with the terms on the right-hand side of (21) separately.

First, since $DV \in \mathscr{C}_b(\mathbb{R}^m; \mathbb{R}^m)$ by assumption we get

$$\mathbb{E} \int_0^{t \wedge \tau_n} \left| \xi(s) \langle g(s, X_s)^T DV(X_s), \cdot \rangle \right|^2 \mathrm{d}s \leq e^{2\|\gamma\|_{L^1}} \|DV\|_\infty^2 nt < \infty$$

for all $t \in \mathbb{R}_{\geq 0}$ due to the definition of τ_n^2, so the stochastic integral

$$\int_0^{\cdot \wedge \tau_n} \xi(s) \langle g(s, X_s)^T DV(X_s), \cdot \rangle \mathrm{d}W_s$$

is a martingale.

Similarly, the compensated integral

$$\int_0^{\cdot \wedge \tau_n} \int_{\{|y| < c\}} \xi(s) \big(V(X_{s-} + H(s, X_{s-}, y)) - V(X_{s-}) \big) \tilde{N}(\mathrm{d}s, \mathrm{d}y)$$

is a martingale, since proceeding as in the proof of Proposition 2.1 and invoking the definition of τ_n^3 we get

$$
\begin{aligned}
\mathbb{E} &\int_0^{t \wedge \tau_n} \int_{\{|y| < c\}} \left| \xi(s) \big(V(X_s + H(s, X_s, y)) - V(X_s) \big) \right|^2 \nu(\mathrm{d}y)\, \mathrm{d}s \\
&= \mathbb{E} \int_0^{t \wedge \tau_n} \int_{\{|y| < c\}} \left| \int_0^1 \xi(s) \langle DV(X_s + \theta H(s, X_s, y)), H(s, X_s, y) \rangle \mathrm{d}\theta \right|^2 \nu(\mathrm{d}y)\, \mathrm{d}s \\
&\leq e^{2\|\gamma\|_{L^1}} \|DV\|_\infty^2 \mathbb{E} \int_0^{t \wedge \tau_n} \int_{\{|y| < c\}} |H(s, X_s, y)|^2 \nu(\mathrm{d}y)\, \mathrm{d}s \\
&\leq e^{2\|\gamma\|_{L^1}} \|DV\|_\infty^2 nt \\
&< \infty
\end{aligned}
$$

for every $t \in \mathbb{R}_{\geq 0}$.

Finally,

$$
\mathbb{E} \int_0^{t \wedge \tau_n} \int_{\{|y| \geq c\}} \big|\xi(s)\big(V(X_s + H(s, X_s, y)) - V(X_s))\big|\, \nu(\mathrm{d}y)\, \mathrm{d}s
$$

$$
= \mathbb{E} \int_0^{t \wedge \tau_n} \int_{\{|y| \geq c\}} \left|\int_0^1 \xi(s)\langle DV(X_s + \theta H(s, X_s, y)), H(s, X_s, y)\rangle \mathrm{d}\theta\right| \nu(\mathrm{d}y)\, \mathrm{d}s
$$

$$
\leq e^{\|\gamma\|_{L^1}} \|DV\|_\infty \, \mathbb{E} \int_0^{t \wedge \tau_n} \int_{\{|y| \geq c\}} |H(s, X_s, y)|\, \nu(\mathrm{d}y)\, \mathrm{d}s
$$

$$
< \infty
$$

for all $t \in \mathbb{R}_{\geq 0}$ owing to (5). Therefore, by the same argument as in [9, Lemma II.3.1] (see the proof of formula (3.8) on page 62 therein) or by modifying slightly the definition of τ_n's and using [10, Theorem II.1.8] we have that

$$
\int_0^{\cdot \wedge \tau_n} \int_{\{|y| \geq c\}} \xi(s)\big(V(X_{s-} + H(s, X_{s-}, y)) - V(X_{s-})\big)\, N(\mathrm{d}s, \mathrm{d}y)
$$

$$
- \int_0^{\cdot \wedge \tau_n} \int_{\{|y| \geq c\}} \xi(s)\big(V(X_s + H(s, X_s, y)) - V(X_s)\big)\, \nu(\mathrm{d}y)\, \mathrm{d}s
$$

is again a martingale.

Hence, the proof that $(U(t, X_t))$ is a supermartingale is completed. Since $U(t, X_t)$ is plainly nonnegative and right-continuous, the martingale convergence theorem implies that there exists an integrable random variable $U_\infty \in L^1(\mathbb{P})$ such that $\lim_{t \to \infty} U(t, X_t) = U_\infty$ $\mathbb{P}$-a.s., whence it follows that

$$
\lim_{t \to \infty} V(X_t) = \lim_{t \to \infty} \exp\Big(-\int_t^\infty \gamma(r)\, \mathrm{d}r\Big) U(t, X_t) - 1 = U_\infty - 1 =: V_\infty \quad (22)
$$

$\mathbb{P}$-almost surely.

Step 2 Now we show that

$$
\liminf_{t \to \infty} |X_t - x_0| = 0 \quad \mathbb{P}\text{-a.s.} \tag{23}
$$

Let $\omega \in \Omega$ be such that

$$
|X_t(\omega) - x_0| \geq \varepsilon
$$

for some $t_0 \in \mathbb{R}_{\geq 0}$ and $\varepsilon > 0$ and all $t \geq t_0$. If (12) is satisfied, then clearly a $\delta > 0$ may be found such that

$$
\varphi(X_t(\omega)) \geq \delta \quad \text{for all } t \geq t_0. \tag{24}
$$

If (13) is satisfied, then note that by (22) we may assume that $V(X_t(\omega))$ converges to a finite limit as $t \to \infty$, so by the first part of (13) there exists a constant $\zeta = \zeta(\omega)$ such that

$$\sup_{t \geq 0} |X_t(\omega)| \leq \zeta.$$

Hence, the second part of (13) implies that

$$\varphi(X_t(\omega)) \geq \inf_{\zeta \geq |x| \geq \varepsilon} \varphi(x) \geq \delta$$

for some $\delta > 0$ and all $t \geq t_0$, that is, (24) again holds. Thus, we have

$$\int_{t_0}^{\infty} \alpha(s) \varphi(X_s(\omega)) \, \mathrm{d}s = \infty,$$

because $\alpha \in L^1_{\mathrm{loc}}(\mathbb{R}_{\geq 0}) \backslash L^1(\mathbb{R}_{\geq 0})$. Therefore, (23) is established provided we show that

$$\int_0^{\infty} \alpha(s) \varphi(X_s) \, \mathrm{d}s < \infty \quad \mathbb{P}\text{-a.s.} \tag{25}$$

As $\xi \geq 1$, we have

$$\int_0^{t \wedge \tau_n} \alpha(s) \varphi(X_s) \, \mathrm{d}s \leq - \int_0^{t \wedge \tau_n} \left[(1 + V(X_s)) \xi'(s) + \xi(s) \mathscr{L} V(s, X_s) \right] \mathrm{d}s$$

for all $t \in \mathbb{R}_{\geq 0}$ and $n \in \mathbb{N}$ by (20). Using (19) together with the fact that the stochastic integrals in (19) are centered and $U \geq 0$, we obtain

$$\begin{aligned}
\mathbb{E} \int_0^{t \wedge \tau_n} \alpha(s) \varphi(X_s) \, \mathrm{d}s &\leq - \mathbb{E} \int_0^{t \wedge \tau_n} \left[(1 + V(X_s)) \xi'(s) + \xi(s) \mathscr{L} V(s, X_s) \right] \mathrm{d}s \\
&= \mathbb{E} \left\{ U(0, X_0) - U(t \wedge \tau_n, X_{t \wedge \tau_n}) \right\} \\
&\leq \mathbb{E} \, U(0, X_0)
\end{aligned}$$

for all $t \in \mathbb{R}_{\geq 0}$ and $n \in \mathbb{N}$, thus passing first $n \to \infty$ and then $t \to \infty$ and applying the monotone convergence theorem twice, we find the estimate

$$\mathbb{E} \int_0^{\infty} \alpha(s) \varphi(X_s) \, \mathrm{d}s \leq \mathbb{E} \, U(0, X_0) = e^{\|\gamma\|_{L^1}} \mathbb{E} \, V(X_0)$$

the right-hand side of which is finite by (H2). We see that (25) holds true.

Step 3 It remains to show that

$$\lim_{t \to \infty} X_t = x_0 \quad \mathbb{P}\text{-a.s.} \tag{26}$$

Suppose that $\omega \in \Omega$ is such that

$$\left| X_{t_n}(\omega) - x_0 \right| \geq \varepsilon$$

for some $\varepsilon > 0$ and a sequence $t_n \nearrow \infty$. By the hypothesis (H2) of Theorem 3.1, an $\eta > 0$ may be found for which

$$V(X_{t_n}(\omega)) \geq \eta \tag{27}$$

for every $n \in \mathbb{N}$. We shall show that then either

$$\lim_{t \to \infty} V(X_t(\omega)) = V_\infty(\omega) \tag{28}$$

or

$$\liminf_{t \to \infty} \left| X_t(\omega) - x_0 \right| = 0 \tag{29}$$

does not hold, where V_∞ is defined by (22). Indeed, (27) together with (28) imply that $V_\infty(\omega) \geq \eta$. On the other hand, if (29) is satisfied, then there exists a sequence $r_n \nearrow \infty$ such that

$$\lim_{n \to \infty} X_{r_n}(\omega) = x_0,$$

hence, again by (28) and (H2),

$$V_\infty(\omega) = \lim_{n \to \infty} V(X_{r_n}(\omega)) = V(x_0) = 0,$$

which is a contradiction. However, we have already shown that both (28) and (29) hold for $\mathbb{P}$-almost all $\omega \in \Omega$, which concludes the proof of Theorem 3.1. $\qquad\square$

Now we focus on a particular case of Eq. (3) corresponding to the continuous-time stochastic approximation procedure of Robbins–Monro type with a general Lévy noise. Recall that in this setting we are looking for a stochastic differential equation such that its solutions converge to a root of the drift R for a class of noise coefficients as wide as possible. Namely, we consider the equation

$$\begin{aligned}
\mathrm{d}X_t = \alpha(t) \Big(R(X_t)\, \mathrm{d}t + \sigma(t, X_t)\, \mathrm{d}W_t + \int_{\{|y| < c\}} K(X_{t-}, y)\, \tilde{N}(\mathrm{d}t, \mathrm{d}y) \\
+ \int_{\{|y| \geq c\}} K(X_{t-}, y)\, N(\mathrm{d}t, \mathrm{d}y) \Big), \quad t \geq 0 \\
X_0 \sim \mu,
\end{aligned} \tag{30}$$

with Borel coefficients

$$\alpha \colon \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{> 0}, \ \ R \colon \mathbb{R}^m \longrightarrow \mathbb{R}^m, \ \ \sigma \colon \mathbb{R}_{\geq 0} \times \mathbb{R}^m \longrightarrow \mathbb{R}^{m \times n}, \ \ K \colon \mathbb{R}^m \times \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

and a Borel probability measure μ on $\mathbb{R}^m$. The driving noise (W, N) is the same as in (3). Since the function K is independent of time now, Assumption 2.1 takes the following form:

Assumption 3.1 We shall assume that

$$\int_{\{|y| < c\}} |K(x, y)|^2 \, \nu(dy) < \infty \quad \text{for all } x \in \mathbb{R}^m$$

and the function

$$\int_{\{|y| \geq c\}} \big| K(\cdot, y) \big| \nu(dy)$$

is locally bounded on $\mathbb{R}^m$.

Let us state a result which one obtains applying Theorem 3.1 to (30).

Corollary 3.1 *Let Assumption 3.1 be satisfied. Let there exist $x_0 \in \mathbb{R}^m$, a function $V \in \mathcal{V} \cap L^1(\mu)$ with $V(x_0) = 0$ and a measurable function $\varphi \colon \mathbb{R}^m \longrightarrow \mathbb{R}_{\geq 0}$ such that*

$$\inf_{\varrho \geq |x - x_0| \geq \varepsilon} \varphi(x) > 0 \quad \text{for all } \varrho > \varepsilon > 0 \tag{31}$$

and

$$\lim_{|x| \to \infty} V(x) = +\infty, \quad \inf_{|x - x_0| \geq \varepsilon} V(x) > 0 \quad \text{for all } \varepsilon > 0. \tag{32}$$

Assume further that $\alpha \in \mathscr{C}(\mathbb{R}_{\geq 0}, \mathbb{R}_{> 0})$ satisfies

$$\int_0^\infty \alpha(r) \, dr = \infty, \quad \int_0^\infty \alpha^2(r) \, dr < \infty. \tag{33}$$

Let there exist a constant $K_\sigma \in \mathbb{R}_{\geq 0}$ and a function $\beta \in \mathscr{C}(\mathbb{R}_{\geq 0}) \cap L^1(\mathbb{R}_{\geq 0})$ such that

$$\Big\langle R(x) + \int_{\{|y| \geq c\}} K(x, y) \, \nu(dy), \, DV(x) \Big\rangle \leq -\varphi(x), \tag{34}$$

$$\mathrm{Tr}\big(\sigma(t, x)^T D^2 V(x) \sigma(t, x)\big) \leq K_\sigma \big(1 + V(x)\big) \tag{35}$$

and

$$\int_{\mathbb{R}^n \setminus \{0\}} \big[V(x + \alpha(t) K(x, y)) - V(x) - \alpha(t) \big\langle K(x, y), \, DV(x) \big\rangle \big] \, \nu(dy)$$
$$\leq \beta(t) \big(1 + V(x)\big) \tag{36}$$

for all $x \in \mathbb{R}^m$ and $t \in \mathbb{R}_{\geq 0}$.

If $(\Omega, \mathscr{F}, (\mathscr{F}_t), (W, N), X)$ *is a solution to (30), then*

$$\lim_{t \to \infty} X_t = x_0 \quad \mathbb{P}\text{-a.s.} \tag{37}$$

Proof To see that Corollary 3.1 follows immediately from Theorem 3.1, it suffices to check that the hypothesis (H3) is satisfied. However, the operator $\mathscr{L}$ associated with (30) takes the form

$$
\begin{aligned}
\mathscr{L}V(t, x) &= \alpha(t)\langle R(x), DV(x)\rangle + \frac{\alpha^2(t)}{2} \operatorname{Tr}\big(\sigma(t, x)^T D^2 V(x)\sigma(t, x)\big) \\
&\quad + \int_{\mathbb{R}^n \setminus \{0\}} \big[V(x + \alpha(t)K(x, y)) - V(x) \\
&\qquad - \alpha(t)\mathbf{1}_{\{|y| < c\}}(y)\langle K(x, y), DV(x)\rangle\big] \,{}^\circ(dy) \\
&= \alpha(t)\Big\langle R(x) + \int_{\{|y| \geq c\}} K(x, y)\nu(dy), DV(x)\Big\rangle \\
&\quad + \frac{\alpha^2(t)}{2} \operatorname{Tr}\big(\sigma(t, x)^T D^2 V(x)\sigma(t, x)\big) \\
&\quad + \int_{\mathbb{R}^n \setminus \{0\}} \big[V(x + \alpha(t)K(x, y)) - V(x) - \alpha(t)\langle K(x, y), DV(x)\rangle\big] \nu(dy)
\end{aligned}
$$

for any $x \in \mathbb{R}^m$ and $t \in \mathbb{R}_{>0}$; the last term on the right-hand side is well defined owing to Assumption 3.1. The assumptions of Corollary 3.1 thus imply that

$$\mathscr{L}V(t, x) \leq -\alpha(t)\varphi(x) + \frac{1}{2}\big(K_\sigma \alpha^2(t) + 2\beta(t)\big)\big(1 + V(x)\big).$$

Since $(K_\sigma \alpha^2 + 2\beta) \in L^1(\mathbb{R}_{\geq 0}) \cap \mathscr{C}(\mathbb{R}_{\geq 0})$, the proof is completed. □

Remark 3.1 (a) As in Theorem 3.1, we may replace (31) and (32) with

$$\inf_{|x - x_0| \geq \varepsilon} \big(V(x) \wedge \varphi(x)\big) > 0 \quad \text{for any } \varepsilon > 0. \tag{38}$$

(b) If the function

$$x \longmapsto \Big\langle R(x) + \int_{\{|y| \geq c\}} K(x, y)\,\nu(dy), DV(x)\Big\rangle$$

is continuous on $\mathbb{R}^m$ and

$$\Big\langle R(x) + \int_{\{|y| \geq c\}} K(x, y)\,\nu(dy), DV(x)\Big\rangle < 0 \quad \text{for } x \neq x_0$$

we may set

$$\varphi(x) = -\left\langle R(x) + \int_{\{|y| \geq c\}} K(x, y)\, \nu(\mathrm{d}y), DV(x) \right\rangle, \quad x \in \mathbb{R}^m,$$

then both (31) and (34) are satisfied.

If $H = 0$ and $K = 0$, then Theorem 3.1 and Corollary 3.1 correspond essentially to [21], Theorems 3.8.1 and 4.4.1, respectively.

4 Applications

Sufficient conditions for convergence of a solution X of (30) to a point are given in Corollary 3.1 in terms of a Lyapunov function V. Choosing a particular Lyapunov function, we get more applicable criteria in terms of the coefficients of (30). If $K = 0$, then $V = |\cdot - x_0|^2$ is a standard choice; however, in the general case, we must proceed in a different way since we need a Lyapunov function belonging to the system $\mathscr{V}$.

Example 4.1 Let $x_0 \in \mathbb{R}^m$ and let us set

$$V : \mathbb{R}^m \longrightarrow \mathbb{R}_{\geq 0}, \; x \longmapsto \log\left(1 + |x - x_0|^2\right).$$

Obviously, the Fréchet derivatives of V are given by

$$DV(x) = 2\,\frac{x - x_0}{1 + |x - x_0|^2},$$

$$D^2 V(x) = \frac{2}{1 + |x - x_0|^2}\, I - \frac{4}{\left(1 + |x - x_0|^2\right)^2}\,(x - x_0)(x - x_0)^T,$$

for all $x \in \mathbb{R}^m$ and thus $V \in \mathscr{V}$, furthermore, $V(x) \to +\infty$ as $|x| \to \infty$.

Let Assumption 3.1 be satisfied and suppose that the coefficients σ and K of (30) satisfy the linear growth condition: there exists a constant $L \in \mathbb{R}_{\geq 0}$ such that

$$|\sigma(t, x)|^2 + \int_{\mathbb{R}^n \setminus \{0\}} |K(x, y)|^2\, \nu(\mathrm{d}y) \leq L\left(1 + |x|^2\right) \tag{39}$$

for all $x \in \mathbb{R}^m$ and $t \geq 0$. Denote by $\mathfrak{k}$ the function

$$\mathfrak{k} : \mathbb{R}^m \longrightarrow \mathbb{R}, \; x \longmapsto \left\langle R(x) + \int_{\{|y| \geq c\}} K(x, y)\, \nu(\mathrm{d}y), x - x_0 \right\rangle.$$

Since

$$\left\langle R(x) + \int_{\{|y| \geq c\}} K(x, y)\, \nu(\mathrm{d}y), DV(x) \right\rangle = \frac{2}{1 + |x - x_0|^2}\, \mathfrak{k}(x) \tag{40}$$

for all $x \in \mathbb{R}^m$, (34) is satisfied with the choice

$$\varphi : x \longmapsto -\frac{2\mathfrak{k}(x)}{1 + |x - x_0|^2}. \tag{41}$$

The function φ defined by (41) surely satisfies (31) if $\mathfrak{k}$ is continuous and

$$\mathfrak{k}(x) < 0 \quad \text{for all } x \neq x_0. \tag{42}$$

If $\mathfrak{k}$ is not continuous, it may be difficult to check (31) and a more feasible way may be to strengthen (42) assuming that there exists $\eta > 0$ such that

$$\mathfrak{k}(x) \leq -\eta |x - x_0|^2 \quad \text{for all } x \in \mathbb{R}^m. \tag{43}$$

In this case, we may set

$$\varphi : x \longmapsto \frac{2\eta |x - x_0|^2}{1 + |x - x_0|^2}$$

obtaining a function that clearly satisfies (31). We claim that the other hypotheses of Corollary 3.1 (in the version of Remark 3.1) are also satisfied.

For any $x \in \mathbb{R}^m$, we may compute using (39)

$$\begin{aligned}
&\text{Tr}\big(\sigma(t, x)^T D^2 V(x) \sigma(t, x)\big) \\
&= \frac{2}{1 + |x - x_0|^2} |\sigma(t, x)|^2 - \frac{4}{\big(1 + |x - x_0|^2\big)^2} \big|\sigma(t, x)^T (x - x_0)\big|^2 \\
&\leq \frac{2}{1 + |x - x_0|^2} |\sigma(t, x)|^2 \\
&\leq 2L \frac{1 + |x|^2}{1 + |x - x_0|^2} \\
&= 4L\Big(1 + \frac{|x_0|^2}{1 + |x - x_0|^2}\Big) \\
&\leq 4L\big(1 + |x_0|^2\big)\big(1 + V(x)\big)
\end{aligned} \tag{44}$$

and (35) follows. Finally, we verify that (36) holds with the choice $\beta = 2\alpha^2 L(1 + |x_0|^2)$. Using that $\log(y) \leq y - 1$ for all $y > 0$ plainly and the definition of V, we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^n \setminus \{0\}} \big[V(x + \alpha(t) K(x, y)) - V(x) - \alpha(t)\langle K(x, y), DV(x)\rangle\big] \nu(\mathrm{d}y) \\
&= \int_{\mathbb{R}^n \setminus \{0\}} \Big[\log\Big(\frac{1 + |x + \alpha(t) K(x, y) - x_0|^2}{1 + |x - x_0|^2}\Big)
\end{aligned}$$

$$
\begin{aligned}
&\left. -\frac{2\alpha(t)}{1+|x-x_0|^2}\langle K(x,y), x-x_0\rangle\right] \nu(dy) \\
&\leq \frac{1}{1+|x-x_0|^2} \int_{\mathbb{R}^n\setminus\{0\}} \left[|x-x_0+\alpha(t)K(x,y)|^2 - |x-x_0|^2\right. \\
&\quad \left. -2\alpha(t)\langle K(x,y), x-x_0\rangle\right] \nu(dy) \\
&= \frac{\alpha^2(t)}{1+|x-x_0|^2} \int_{\mathbb{R}^n\setminus\{0\}} |K(x,y)|^2\, \nu(dy) \\
&\leq \alpha^2(t)L\, \frac{1+|x|^2}{1+|x-x_0|^2} \\
&\leq 2\alpha^2(t)L\big(1+|x_0|^2\big)\big(1+V(x)\big)
\end{aligned}
\tag{45}
$$

for all $t \in \mathbb{R}_{\geq 0}$ and $x \in \mathbb{R}^m$. Note also that Assumption 3.1 clearly follows from (39).

Therefore, whenever $\alpha \in \mathscr{C}(\mathbb{R}_{\geq 0}, \mathbb{R}_{>0})$ obeys (33) and $((W, N), X)$ is a solution to (30), then X converges almost surely to x_0 as $t \to \infty$.

Remark 4.1 It should be stressed that under the hypotheses of Example 4.1 the point $x_0 \in \mathbb{R}^m$ the solution of (30) converges to need not be a root of the drift R; therefore, *a priori* it might be misleading to speak about a Robbins–Monro stochastic approximation procedure. Let us discuss this problem more carefully: Our main positive results are illustrated in paragraphs (d) and (f), while (c) contains a counterexample. In (a), (b) and (e), particular cases related to hitherto available results are treated.

(a) Assume that $K = 0$. Then, (42) reduces to

$$
\langle R(x), x-x_0\rangle < 0 \quad \text{for all } x \neq x_0.
\tag{46}
$$

Hence, if R is continuous (which is a rather natural assumption) we have $R(x_0) = 0$ (as it is well known from the theory of monotone mappings, see, e.g., [5, Lemma 1] for a much more general result) and plainly x_0 is the unique root of R. If σ satisfies the linear growth condition and R is a continuous function such that (46) holds, then

$$
\lim_{t\to\infty} X_t = x_0 \quad \mathbb{P}\text{-almost surely}
\tag{47}
$$

for any solution of the equation

$$
dX_t = \alpha(t)\big(R(X_t)\,dt + \sigma(t, X_t)\,dW_t\big), \quad X_0 \sim \mu.
\tag{48}
$$

This is a classical result going back to [21].

(b) If the driving Lévy noise has a purely discontinuous component, but there are no large jumps, that is, $\nu\{|x| \geq a\} = 0$ for some $a \in (0, \infty)$, then the results are virtually the same as in the diffusion case. Indeed, if R is continuous, obeys (46), and σ and K have at most linear growth, then (47) holds for any solution of

$$
dX_t = \alpha(t)\big(R(X_t)\,dt + \sigma(t, X_t)\,dW_t
$$

$$+ \int_{\{|y|<a\}} K(X_{t-}, y) \tilde{N}(\mathrm{d}t, \mathrm{d}y) \Big), \quad X_0 \sim \mu. \tag{49}$$

Again, x_0 is the unique root of R. Related results, obtained by different methods, may be found in [15, 20].

(c) In the general case $K \neq 0$ and $\nu\{|y| \geq c\} > 0$, the situation changes considerably. This should not be surprising: the last term on the right-hand side of (30), that is, the process

$$\int_0^{\cdot} \int_{\{|y|\geq c\}} K(X_{t-}, y) \, N(\mathrm{d}t, \mathrm{d}y) \tag{50}$$

is not centered in general. Moreover, if we would like to keep the driving Lévy noise in (3) but to use a representation with a different c it results in a change of the drift (and, *a fortiori*, of the roots of the drift). Hence, Corollary 3.1 need not be applicable to the Robbins–Monro procedure, as it implies convergence to a point x_0 such that $R(x_0) \neq 0$. Indeed, if in the setting of Example 4.1 the function $\mathfrak{k}$ is continuous and satisfies (42), then we only know that

$$R(x_0) + \int_{\{|y|\geq c\}} K(x_0, y) \, \nu(\mathrm{d}y) = 0$$

The following simple example illustrates this phenomenon. Define the coefficients R and K by

$$R \colon x \longmapsto A(x - a), \quad K \colon (x, y) \longmapsto B(x - b)$$

for some $a, b \in \mathbb{R}^m$ and matrices $A, B \in \mathbb{R}^{m \times m}$ such that $A + B$ is invertible and negative definite, and $A(x_0 - a) \neq 0$ where we set $x_0 = (A + B)^{-1}(Aa + Bb)$. We can assume for simplicity that $\nu\{|y| \geq c\} = 1$. Then,

$$\begin{aligned}
\mathfrak{k}(x) &= \Big\langle A(x - a) + \int_{\{|y|\geq c\}} B(x - b) \, \nu(\mathrm{d}y), x - x_0 \Big\rangle \\
&= \big\langle (A + B)x - (Aa + Bb), x - x_0 \big\rangle \\
&= \big\langle (A + B)(x - x_0), x - x_0 \big\rangle \\
&\leq -\eta |x - x_0|^2
\end{aligned}$$

for some $\eta > 0$ and all $x \neq x_0$, however, $R(x_0) \neq 0$.

(d) Therefore, in the general case of (30) we must add the assumption $R(x_0) = 0$ if Corollary 3.1 is to be applied to stochastic approximation; for Eqs. (48) and (49) this is redundant. On the other hand, by choosing K in an appropriate way we may obtain (47) under rather mild hypotheses on R. Let us assume that $R(x_0) = 0$ and R is Lipschitz continuous, denote by $\mathrm{Lip}(R)$ its Lipschitz constant. If K satisfies, still in the setting of Example 4.1,

$$\Big\langle \int_{\{|y|\geq c\}} K(x, y) \, \nu(\mathrm{d}y), x - x_0 \Big\rangle \leq -(\mathrm{Lip}(R) + 1)|x - x_0|^2 \quad \text{for all } x \in \mathbb{R}^m,$$

then Corollary 3.1 is applicable. In the diffusion case (48), the mere Lipschitz continuity of R need not be sufficient for the convergence of the stochastic approximation procedure. (Indeed, consider (48) with the choice $m = n = 1$, $R(x) = \sigma(t, x) = x$ for $(t, x) \in \mathbb{R}_{\geq 0} \times \mathbb{R}$, $V = |\cdot|^2$, and $\alpha(t) = (1 + t)^{-1}$ for $t \geq 0$, then all assumptions of Corollary 3.1 are satisfied except the hypothesis (34), R is plainly globally Lipschitz continuous having 0 as its only root, nevertheless, a simple direct calculation shows that $X_t \to \infty$ $\mathbb{P}$-a.s. as $t \to \infty$.)

(e) If

$$\int_{\{|y| \geq c\}} K(x, y)\, \nu(\mathrm{d}y) = 0 \quad \text{for all } x \in \mathbb{R}^m$$

then the process (50) is centered and we see that any solution X to (30) converges to the unique root of R under the hypothesis that R is a continuous function satisfying (46) (and σ and K has at most linear growth). This result may be compared with theorems stated in [12] where equations driven by centered square integrable processes with independent increments are dealt with. We do not need L^2-integrability, on the other hand sharper asymptotic results than mere convergence almost surely are established in [12] at the price of more restrictive assumptions on noise coefficients and the cumulant process of the driving Lévy process.

(f) Finally, note that the hypotheses of Example 4.1 may be satisfied even if R has multiple roots. The coefficient K then "selects" a root of R which a solution to (30) converges to. This may happen only if a noncentered non-compensated Poisson process is allowed as a driving noise. As we have already indicated above, large jumps of the Lévy process virtually change the drift and, consequently, it is possible that a solution to (30) no longer converges to some (or all) of its roots. Again, in the diffusion case or for Eq. (49) the situation is completely different, see, e.g., [21, Chapter 5]. For example, let $m = 1$ and let σ and K satisfy (39) and

$$x \cdot \int_{\{|y| \geq c\}} K(x, y)\, \nu(\mathrm{d}y) \leq -2|x|^2 \quad \text{for all } x \in \mathbb{R}.$$

Then, any solution to

$$\mathrm{d}X_t = \alpha(t)\Big(\sin X_t\, \mathrm{d}t + \sigma(t, X_t)\, \mathrm{d}W_t + \int_{\{|y| < c\}} K(X_{t-}, y)\, \tilde{N}(\mathrm{d}t, \mathrm{d}y)$$
$$+ \int_{\{|y| \geq c\}} K(X_{t-}, y)\, N(\mathrm{d}t, \mathrm{d}y)\Big), \quad t \geq 0$$
$$X_0 \sim \mu,$$

satisfies

$$\lim_{t \to \infty} X_t = 0 \quad \mathbb{P}\text{-a.s.}$$

(g) It is possible to allow coefficients K depending on time, i.e., defined on $\mathbb{R}_{\geq 0} \times \mathbb{R}^m \times \mathbb{R}^n$. If Eq. (49) is considered, that is, there are no large jumps, this change

results in a trivial modification of the assumptions. In the general case, however, the hypotheses become cumbersome and thus we content ourselves with time-independent K's.

5 Conclusions

We extended a Lyapunov-functions-based approach to convergence of a continuous-time Robbins–Monro procedure of stochastic approximation from diffusion processes to systems defined by a stochastic differential equation driven by a general Lévy process. While for a driving noise with small jumps only our results are essentially comparable with available results (albeit our proofs are different), if large jumps are allowed we showed that new phenomena may occur: the large jumps may force the procedure to converge to a "fake" root of the drift, on the other hand, if the noise coefficient is properly chosen, we obtain convergence under hypothesis weaker than those of the standard theory.

References

1. Applebaum, D.: Lévy Processes and Stochastic Calculus, 2nd edn. Cambridge University Press, Cambridge (2009)
2. Benveniste, A., Métivier, M., Priouret, P.: Adaptive Algorithms and Stochastic Approximations. Springer, Berlin (1990)
3. Bhatnagar, S., Prasad, H. L., Prashanth, L. A.: Lecture notes in control information science. In: Stochastic Recursive Algorithms for Optimization, vol. 434. Springer, London (2013)
4. Borkar, V.S.: Stochastic Approximation: A Dynamical Systems Viewpoint. Cambridge University Press, Cambridge (2008)
5. Browder, F.E.: Nonlinear elliptic boundary value problems. Bull. Am. Math. Soc. **69**, 862–874 (1963)
6. Chen, H.-F.: Continuous-time stochastic approximation: convergence and asymptotic efficiency. Stoch. Stoch. Rep. **51**, 217–239 (1994)
7. Chen, H.-F.: Stochastic Approximation and its Applications. Kluwer, Dordrecht (2002)
8. Gwinner, J., Jadamba, B., Khan, A.A., Raciti, F.: Uncertainty Quantification in Variational Inequalities. Theory, Numerics, and Applications. CRC Press, Boca Raton (2022)
9. Ikeda, N., Watanabe, S.: Stochastic Differential Equations and Diffusion Processes. North-Holland, Amsterdam (1981)
10. Jacod, J., Shiryaev, A.N.: Limit Theorems for Stochastic Processes, 2nd edn. Springer, Berlin (2003)

11. Komarov, S.V., Krasulina, T.P.: One-sided convergence of continuous processes of stochastic approximation. J. Math. Sci. **93**, 379–384 (1999)
12. Korostelev, A.P.: Stochastic Recurrent Procedures. Local properties (Russian). Nauka, Moscow (1984)
13. Kushner, H.J., Clark, D.S.: Stochastic Approximation Methods for Constrained and Unconstrained systems. Springer, New York (1978)
14. Kushner, H.J., Yin, G.G.: Stochastic Approximation and Recursive Algorithms and Applications, 2nd edn. Springer, New York (2003)
15. Lazrieva, N., Sharia, T., Toronjadze, T.: The Robbins-Monro type stochastic differential equations. I. Convergence of solutions. Stoch. Stoch. Rep. **61**, 67–87 (1997)
16. Lazrieva, N., Sharia, T., Toronjadze, T.: The Robbins-Monro type stochastic differential equations. II. Asymptotic behaviour of solutions. Stoch. Stoch. Rep. **75**, 153–180 (2003)
17. Lazrieva, N., Sharia, T., Toronjadze, T.: Semimartingale stochastic approximation procedure and recursive estimation. J. Math. Sci. **153**, 211–261 (2008)
18. Lazrieva, N., Toronjadze, T.: The Robbins-Monro type stochastic differential equations. III. Polyak's averaging. Stochastics **82**, 165–188 (2010)
19. Maslowski, B., Týbl, O.: Invariant measures and boundedness in the mean for stochastic equations driven by Lévy noise. Stoch. Dyn. **22**, 2240019 (2022)
20. Mel'nikov, A. V.: Stochastic approximation procedures for semimartingales (Russian). In: Statistics and Control of Random Processes, pp. 147–256. Nauka (Russian) (1989)
21. Nevel'son, M.B., Khas'minskiĭ, R.Z.: Stochastic approximation and recurrent estimation (Russian). Nauka, Moscow (1972) (English translation: Nevel'son, M.B., Has'minskiĭ, R.Z.: Stochastic approximation and recursive estimation. American Mathematical Society, Providence (1973))
22. Pflug, G.: Stetige stochastische approximation. Metrika **26**, 139–150 (1979)