FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

DOCTORAL THESIS

Ondřej Týbl

Stochastic Equations with Correlated Noise and Their Applications

Department of Probability and Mathematical Statistics

Supervisor of the doctoral thesis:	prof. RNDr. Bohdan Maslowski, DrSc.
Study programme:	P4M9
Study branch:	Probability and statistics, econometry and financial mathematics

Prague 2023

I declare that I carried out this doctoral thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In .... date ....

Author's signature

I would like to express my gratitude to my supervisor prof. Bohdan Maslowski and my collaborator Dr. Jan Seidler for their kindness, patience and excellent guidance during my studies.

Title: Stochastic Equations with Correlated Noise and Their Applications

Author: Ondřej Týbl

Department: Department of Probability and Mathematical Statistics

Supervisor: prof. RNDr. Bohdan Maslowski, DrSc., Department of Probability and Mathematical Statistics

Abstract: Properties of stochastic differential equations with jumps are studied. Lyapunov-type methods are derived to assess long-time behavior of solutions and general results are applied in specific cases. In the first case, conditions in terms of the geometric properties of the coefficients for stability in terms of boundedness in probability in the mean are obtained. By means of Krylov-Bogolyubov Theorem criterion for existence of invariant measures is given subsequentely. In the second case, the long-time behavior refers to existence of an almost sure single-point limit not depending on the initial condition. This result is then applied to get a continuous-time Robbins-Monro type stochastic approximation procedure for finding roots of a given function.

Keywords: Lévy processes Invariant measures Stochastic approximation procedures Lévy-driven stochastic differential equations

Contents

Introduction **2**

1 Theory review **6**
1.1 Lévy processes 6
1.2 Stochastic integration 12
1.3 Stochastic differential equations 17
1.4 Stochastic approximation procedures 20
1.5 Invariant measures for Markov processes 23

2 Invariant measures and boundedness in the mean **26**
2.1 General Lyapunov criterion 27
2.2 Stabilization by compensated jumps 31
2.3 Stabilization by uncompensated jumps 37
2.4 Invariant measure 43

3 Stochastic Approximation Procedures **45**
3.1 Main results 47
3.2 Applications 54

Bibliography **59**

List of abbreviations **63**

List of publications **64**

Introduction

Consider a dynamical system that we would like to describe by means of a stochastic differential equation. Formally, we have a physical model

$$\dot{v}(t) = f(t, v(t)) + \xi(t, v(t)),$$

where f represents the underlying physical law and ξ represents some noise. Thus, both f and ξ need to be specified and it is the modeller's task to convert their knowledge about the system into properties of both f and ξ. After doing so an obvious question arises: *how do our model assumptions play together?* Necessarily, our equation should have a solution. In the second step, we should also be interested in the solution's properties as we might reject our model if it does not encode the modeller's *a priori* assumptions.

Some of the key model properties relate to the long-time behavior of a solution. We shall be interested only in systems that are stable in some sense: either the solution does not grow unboundedly in some sense, it possess an equilibrium point represented by a stationary solution (or invariant measure) or we might even want the solution to converge to a single point.

It is in the famous work Lyapunov [1907] where methods for determining the stability properties of deterministic systems where proposed. Since then similar techniques have proved useful for stochastic systems as well and those have been referred to as Lyapunov techniques in many important works. In this Thesis we make advantage of the fact that Lyapunov methods help us to answer many questions on the growth of solution, existence of invariant measure and a convergence to a single point even when the particular shape of the solution is unknown. Therefore, also non-linear systems can be considered.

In this work we bring new results in the setting of stochastic differential equations with jumps represented by Lévy noise. The novelty lies in the fact that we are able to cover non-linear systems where no assumptions on the existence of moments of the noise are needed. We arrive at conditions fully specified in terms of the coefficients of the model and interpret them geometrically. The results can be divided into two parts.

Firstly, we investigate a stability property that may be called "boundedness of solutions in probability in the mean" which together with the Feller property verifies the existence of an invariant measure by means of the well-known Krylov–Bogolyubov Theorem. In particular, we are interested in some situations when the equation is "stabilized" in the above sense by noise. This phenomenon is well understood in the case of Gaussian noise (as discussed already in the classical Khasminskii's monograph Khasminskii [1980], for infinite-dimensional systems Leha et al. [1999]) and we focus on contribution of the stochastic jump terms. Stochastic differential equations driven by Lévy noise have been extensively studied in recent years. Important fundamental results in this field are presented, for instance, in the monograph Applebaum [2009] from which we take the basic setting of the problem.

The topic addressed is related to the problems of stability (and stabilization) of trivial solution to Lévy-driven stochastic equation which was recently studied in several papers. In Applebaum and Siakalli [2010], asymptotic a.s. stability is shown in the linear case (under conditions analogous to those in this work if restricted to such case). In Applebaum and Siakalli [2009], asymptotic stability in probability, in the mean and the moment stability is studied. The paper Nane and Ni [2017] deals with boundedness in probability and the moment boundedness, for a time-changed Lévy noise, an analogous problem in the case of Gaussian noise is also investigated in Nane and Ni [2017]. Some related results can be also found in Grigoriu [1996], Li et al. [2002] and an analogous problem for equations driven by discontinuous semimartingales are studied in Mao and Rodkina [1995]. Reference Mao [2008] presents a general treatise on stochastic stability. Existence of stationary distributions has been addressed in Bhan et al. [2012] and Jurek [1982]. In the latter paper, the existence is shown by means of Lyapunov method but the effect of stabilization by noise is not considered (the noise has to be "small enough"). In Albeverio et al. [2016], a class of invariant measures is described in general terms by means of Fokker–Planck equation. For infinite-dimensional systems, existence of invariant measure has been studied, for example, in Applebaum [2015] or Kumar and Riedle [2021]. A related problem of existence of random attractors for equations with two-sided Lévy noise has been treated in Zhang et al. [2019].

Secondly, criterions for convergence of a solution to a single point by means of Lyapunov methods are applied in the context of stochastic approximation procedures.

Stochastic approximation algorithms originally proposed as a tool for finding a root of a function (the Robbins–Monro procedure) or its minimum (the Kiefer–Wolfowitz procedure), these algorithms found various applications in optimization and machine learning. See, e.g., the books Bhatnagar et al. [2013], Borkar [2008], Browder [1963], Chen [2002], Kushner and Clark [1978], Kushner and Yin [2003] for a thorough discussion of various aspects of stochastic approximation algorithms and their use. (Let us mention also [Gwinner et al., 2022, 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

$$\mathrm{d}Y_t = \alpha(t)\,(R(Y_t) + \sigma(t, Y_t)\mathrm{d}Y_t) \tag{0.0.1}$$

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 (0.0.1) 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 Nevel'son and Khas'minskiĭ [1972] for a systematic development of these ideas and, for example, the papers Chen [1994], Komarov and Krasulina [1999], Pflug [1979] and the book Korostelev [1984] for further results on continuous-time stochastic approximation.

As discrete-time systems indicate, it is reasonable to consider more general driving noises in Eq. (0.0.1). Stochastic recursive procedures described by equations driven by semimartingales were considered by Mel'nikov [1989] and Lazrieva et al. [1997], Lazrieva et al. [2003], Lazrieva et al. [2008], Lazrieva and Toronjadze [2010]. 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 Korostelev [1984]; proofs, using Lyapunov functions techniques, are given, however, only in the discrete-time case.

The Thesis is divided into three chapters.

The first Chapter consists of 5 sections where basic definitions and standard results are stated. Exposition to the theory of Lévy processes, stochastic integration with respect to them and stochastic differential equations with jumps are mostly based on Applebaum [2009]. Sections introducing the topics of stochastic approximation procedures and invariant measures for Markov processes are also included.

The second Chapter is based on Maslowski and Týbl [2022] and is devoted to boundedness in probability in the mean and existence of invariant measures. First, the problem is posed and some preliminary standard results are recalled then results are presented in four following sections.

In Section 2.1 general Lyapunov criterion for boundedness in probability in the mean is presented. In principle, for the most general formulation of the stability theorem we only need local boundedness of the coefficients besides existence and uniqueness of solutions. However, we also present some standard conditions (Lipschitz and linear growth conditions) which verify this basic assumption and which are also helpful in more specific cases.

The general theorem from Section 2.1 is applied to the equation containing the drift, diffusion and compensated integral terms in Section 2.2. The main result (Theorem 2.2.1) is formulated for locally bounded coefficients and then specified in the linear growth case (Corollary 2.2.1). This result allows us to discuss stabilizing roles of particular terms in the equation and their mutual influence. The Section is closed by an example where such interplay of particular terms in the equation is demonstrated and also, relation to moment stability of solutions is discussed.

In Section 2.3, the general statement from Section 2.1 is applied to the equation with uncompensated integral term, drift and diffusion. Theorem 2.3.1 is analogous to Theorem 2.2.1 from the previous section. In Theorem 2.3.2, a different approach is adopted and a stability criterion is found which is expressed directly in terms of jumps. Section 2.3 is closed by two examples: In the first one, the linear equation is studied. In the second one, the influence of the parameter dividing small and big jumps is discussed.

Section 4 summarizes consequences of the previous parts for the existence of invariant measure (stationary solution) if the equation defines a Feller Markov process, which may be viewed as the main result.

In the last Chapter results on approxistochasticmation procedures are presented based on Seidler and Týbl [2023]. First, we introduce the equation we deal with precisely and we state the Itô formula in a form required in our proofs.

In Section 3.1, the main results are proved: Theorem 3.1.1 giving general sufficient conditions for convergence of solutions to a stochastic differential equation driven by a Lévy process to a singleton and its Corollary 3.1.1 concerning the Robbins–Monro procedure.

In Section 3.2, we show how to apply these results to particular systems. 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 Nevel'son and Khas'minskiĭ [1972]. It may look odd that the noise in (3.0.1) 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 (3.0.1) 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 (3.0.1) converge to the root of R under conditions weaker than those used in the diffusion case (0.0.1) 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 3.2.1, we discuss the differences between behavior of solutions to (0.0.1) and (3.0.1) in detail.

ACKNOWLEDGEMENTS *This research was partially supported by Czech Science Foundation (GAČR) Grant no. 22-12790S, the SVV Grant No. 260580 and the Grant schemes at Charles University, Project No. CZ.02.2.690.00.019 073/0016935.*

1. Theory review

1.1 Lévy processes

In this Section we briefly recall what is a Lévy process together with its relation to infinite divisibility, Lévy-Itô formula and present some examples. The exposition follows Applebaum [2009].

Definition 1.1.1. Let $n \in \mathbb{N}$ and $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis with a normal filtration $(\mathscr{F}_t)_{t\geq 0}$. An $\mathbb{R}^n$-valued $(\mathscr{F}_t)$-progressively measurable stochastic process L is called an $(\mathscr{F}_t)$-Lévy process if

- $L_0 = 0, \quad$ a.s.,
- L has stationary and independent increments,
- L is stochastically continuous, that is

  $$\mathbb{P}\left(|L_{t+h} - L_t| > \epsilon\right) \to 0, \quad h \to 0$$

  for every $t \in \mathbb{R}_{\geq 0}$ and $\epsilon \in \mathbb{R}_{>0}$ (with limit from the right only if $t = 0$),
- L has almost surely càdlàg paths.

If the normal filtration $(\mathscr{F}_t)$ is clear from the context we speak about Lévy processes for short.

The concept of Lévy processes is closely related to the notion of infinite divisibility. To describe this relation we also recall what is a weakly continuous convolution semigroup. Recall that if $x \in \mathbb{R}^n$, $n \in \mathbb{N}$ then δ_x denotes the dirac measure at x.

Definition 1.1.2. A family $(p_t, t \in \mathbb{R}_{\geq 0})$ of probability measures in $\mathbb{R}^n$, $n \in \mathbb{N}$ is said to be a weakly continuous convolution semigroup if

- $p_0 = \delta_0$,
- we have

  $$p_{s+t} = p_s * p_t, \quad s, t \in \mathbb{R}_{\geq 0},$$

  where $*$ denotes the convolution operator,
- we have

  $$p_t \to \delta_0 \quad \text{weakly}$$

  as $t \to 0+$[1].

[1] In our case this congergence simply means that $\int_{\mathbb{R}^n} f(y)\mathrm{d}p_t(y) \to f(0)$ as $t \to 0+$ for every continuous and bounded function $f : \mathbb{R}^n \to \mathbb{R}$.

Remark 1.1.1.

It is easy to see that the marginal distribution of a Lévy process at any fixed time is infinitely divisible. Indeed, if L is a Lévy process and $t \geq 0$ then we shall write

$$L_t = \sum_{j=1}^{k} \left(L_{\frac{j}{n}t} - L_{\frac{j-1}{k}t} \right)$$

for any $k \in \mathbb{N}$. So with $k \in \mathbb{N}$ given setting $Y_j = L_{\frac{j}{n}t} - L_{\frac{j-1}{k}t}, j = 1, \ldots k$ we have shown that

$$L_t = \sum_{j=1}^{k} Y_j$$

for $Y_j, j = 1, \ldots, k$ independent and identically distributed (by the independence and stationarity of the increments of L), which shows infinite divisibility of the distribution of L_t.

On the other hand, if we are given an infinitely divisible probability distribution μ on $\mathbb{R}^n$, $n \in \mathbb{N}$, then we can construct a Lévy process L with μ being the distribution of L_1 as follows: First, by the celebrated *Lévy-Khnintchine formula* (see e.g. [Sato, 1999, Thmeorem 8.1]) the characteristic function of μ is of the exponential form $u \mapsto e^{\eta(u)}$ for a suitable complex function η. By simple arguments one can show that the family of characteristic functions

$$\left(e^{t\eta}, t \in \mathbb{R}_{\geq 0} \right)$$

corresponds to some family $(p_t, t \in \mathbb{R}_{\geq 0})$ of weakly continuous convolution probability measures on $\mathbb{R}^n$. Now a canonical process of projections

$$\tilde{L}_t(\omega) := \omega(t), \quad \omega \in \Omega, t \in \mathbb{R}_{\geq 0}$$

on the set Ω of functions on $\mathbb{R}_{\geq 0}$ with values in $\mathbb{R}^n$ starting from 0 is constructed. We equip Ω with the cylinder σ-algebra $\mathscr{F}$, filtration generated by $\tilde{L}$ and a probability measure $\mathbb{P}$ given by the condition

$$\mathbb{P}\left(\{\omega \in \Omega : \omega(t_1) \in A_1, \ldots, \omega(t_n) \in A_n\}\right) := \int_{\mathbb{R}^n} \cdots \int_{\mathbb{R}^n} \mathbf{1}_{A_1}(y_1) \ldots \mathbf{1}_{A_n}(y_1 + \cdots + y_n) p_{t_1}(\mathrm{d}y_1) \ldots p_{t_n}(\mathrm{d}y_n)$$

for every $A_1, \ldots, A_n \in \mathscr{B}(\mathbb{R}^n)$, $t_1, \ldots, t_n \in \mathbb{R}_{\geq 0}$, which is justified by Kolmogorov's extension theorem. Using standard arguments one can show that $\tilde{L}$ is a process starting at 0 of stationary and independent increments and that $\tilde{L}$ is stochastically continuous. The construction is completed by taking a càdlàg modification L of $\tilde{L}$ that exists and is again a Lévy process as shown in [Applebaum, 2009, Theorem 2.1.7] and taking the augmented canonical filtration which satisfies the normality conditions by [Applebaum, 2009, Theorem 2.1.9]. For more details of this construction see Applebaum [2009].

We have seen that there is one-to-one correspondence between Lévy processes and infinitely divisible distributions. Let us now present some key examples.

Example 1.1.1. The following processes are Lévy processes:

- any linear deterministic function $t \mapsto bt$ for some $b \in \mathbb{R}^n$,
- Q-Wiener process, that is an $\mathbb{R}^n$-valued process W with continuous paths starting from zero with independent increments and such that $W_t - W_s$ has centered Normal distribution with covariance $(t-s)Q$ for some positive semi-definite and symmetric matrix $Q \in \mathbb{R}^{n\times n}$ for any $0 \le s < t < \infty$,
- Poisson process, that is a càdlàg $\mathbb{R}$-valued process N with independent increments and such that

$$\mathbb{P}\left(N_t = n\right) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n \in \mathbb{N}$$

  for any $t \in \mathbb{R}_{\ge 0}$ and some $\lambda \in \mathbb{R}_{>0}$ (which is called the intensity of N),
- compensated Poisson process, that is the process of the form

$$\tilde{N}_t = N_t - \lambda t,$$

  where N is the Poisson process with intensity $\lambda \in \mathbb{R}_{>0}$,
- compound Poisson process, that is if N is a Poisson process independent of sequence of i.i.d. $\mathbb{R}^n$-valued variables $Y_1, Y_2, \ldots$ then we consider the process

$$\sum_{j=1}^{N_t} Y_j,$$

- α-stable process, that is a Lévy process for which any marginal distribution is (strictly) α-stable with some common stability parameter $\alpha \in (0,2]$. We note that this is the only example of a self-similar Lévy process as shown e.g. in [Sato, 1999, Proposition 13.5]. We recall that a stochastic process X is self-similar if there exists so-called *Hurst index* $H \in \mathbb{R}_{>0}$ such that $(X_{at}, t \in \mathbb{R}_{\ge 0})$ and $\left(a^H X_t, t \in \mathbb{R}_{\ge 0}\right)$ have the same finite-dimensional distributions for any $a \in \mathbb{R}_{\ge 0}$. In this case we have $H = 1/\alpha$.

It is easily checked that a sum of independent Lévy processes is again a Lévy process. It is due to the famous Lévy-Itô formula that we can decompose any Lévy process into the sum of four independent processes of simple form. The above-mentioned examples: linear deterministic function, Q-Wiener process, compensated Poisson process and compound Poisson process are the key building blocks. However, before stating this result we need to introduce some notation. For more detailed exposition the reader may consult with [Jacod and Shiryaev, 2003, Chapter II.] where even more general case is treated.

First, we define the jump measure for an $(\mathscr{F}_t)$-Lévy process L as follows: we denote

$$N(t, A) := \sum_{0 \le s \le t} \mathbf{1}_A\left(\Delta L_s\right), \quad t \in \mathbb{R}_{\ge 0}, A \in \mathscr{B}(\mathbb{R}^n), \tag{1.1.1}$$

where

$$\Delta L_s := L_s - \lim_{u \to s+} L_u$$

denotes the jump of L at a given time s. One can show that N defines so-called $(\mathscr{F}_t)$-*Poisson random measure* on $\mathbb{R}_+ \times (\mathbb{R}^n \setminus \{0\})$ with some intensity measure $\mathrm{d}t\nu(\mathrm{d}y)$ (see [Jacod and Shiryaev, 2003, Chapter II.] or [Applebaum, 2009, Chapter 2.3] for the details).

As directly follows from the path regularity of a Lévy process we have that if $A \in \mathscr{B}(\mathbb{R}^n)$ is such that its closure does not contain the zero element $0 \in \mathbb{R}^n$ then there are only finitely many jumps ΔL_s with $\Delta L_s \in A$ on any finite interval $s \in [0,t]$ almost surely. Thus, the integral process

$$\int_A yN(t,\mathrm{d}y) := \int_{[0,t]\times A} yN(\mathrm{d}t,\mathrm{d}y) = \sum_{0\le s\le t} \left(\mathbf{1}_A(\Delta L_s)\Delta L_s\right), \quad t \in \mathbb{R}_{\ge 0} \tag{1.1.2}$$

which represents the sum of all jumps of L with values in A up to a given time is a piecewise constant process. In fact it is a compound Poisson process (see Example 1.1.1). On the other hand, if $A \in \mathscr{B}(\mathbb{R}^n)$ is bounded (but its closure may contain the zero element) the limit

$$\int_A y\tilde{N}(t,\mathrm{d}y) := \lim_{\epsilon\to 0+} \left(\int_{A\cap|y|\ge\epsilon} yN(t,\mathrm{d}y) - t\int_{A\cap|y|\ge\epsilon} y\nu(\mathrm{d}y) \right), \quad t \in \mathbb{R}_{\ge 0} \tag{1.1.3}$$

in $L^2(\Omega,\mathbb{R}^n)$ defines (up to taking the càdlàg modification) an $(\mathscr{F}_t)$-Lévy process.

Now we are ready to formulate the Lévy-Itô formula and provide some remarks on its applicability.

Theorem 1.1.1. Any $(\mathscr{F}_t)$-Lévy process L in $\mathbb{R}^n$ can be decomposed as

$$L_t = bt + W_t + \int_{\{|y|<c\}} y\tilde{N}(t,\mathrm{d}y) + \int_{\{|y|\ge c\}} yN(t,\mathrm{d}y), \quad t \in \mathbb{R}_{\ge 0} \tag{1.1.4}$$

almost surely, where $c \in \mathbb{R}_{>0}$, $b \in \mathbb{R}^n$, W is an $(\mathscr{F}_t)$-Wiener process with covariance Q with values in $\mathbb{R}^n$ which is independent to an $(\mathscr{F}_t)$-Poisson random measure N on $\mathbb{R}_+ \times (\mathbb{R}^n \setminus \{0\})$ with intensity measure $\mathrm{d}t\nu(\mathrm{d}y)$, where ν is a Lévy measure on $\mathbb{R}^n \setminus \{0\}$, that is

$$\int_{\mathbb{R}^n\setminus\{0\}} \left(|y|^2 \wedge 1\right) \nu(\mathrm{d}y) < \infty \tag{1.1.5}$$

Proof. See [Sato, 1999, Theorem 19.2]. □

Remark 1.1.2.

- The four terms in (1.1.4) represent the drift, diffusion, small jumps (up to size of c) and large jumps respectively. While the large jumps can be separated due to the càdlàg property into a finite random sum, the small jumps are compensated (which can be seen in the formula (1.1.3)). In more detail, both the terms

  $$t \mapsto bt \quad \text{and} \quad t \mapsto \int_{\{|y|\ge c\}} yN(t,\mathrm{d}y)$$

  are processes of finite variation while

  $$t \mapsto W_t \quad \text{and} \quad t \mapsto \int_{\{|y|<c\}} y\tilde{N}(t,\mathrm{d}y)$$

  are L^2-martingales. We see that any Lévy process is a semimartingale.

- The decomposition (1.1.4) is unique for a given choice of $c \in \mathbb{R}_{>0}$. In fact, for each choice of c we obtain possibly different decomposition: if b is the drift term associated with given c then

$$b + \int_{c' \leq |y| < c} y\nu(\mathrm{d}y)$$

is the drift term associated with the choice $0 < c' < c$.

- The marginal distribution of L is determined by a triplet (b, Q, ν), where $b \in \mathbb{R}^n$, Q is a covariance matrix of a $\mathbb{R}^n$-dimensional process and ν is a Lévy measure if $c \in \mathbb{R}_{>0}$ is fixed. Thus, we call this triplet the characteristics of the Lévy process L. Such definition is consistent with the definition of characteristics of a semimartingale which possibly depend on the so-called *truncation function* which is $\mathbf{1}_{\{|y|<c\}}$ in our case. The semimartingale characteristics are time-dependent and random in the general case.

- One can figure out many properties of a Lévy process L from its characteristics, for example
  - We have that $\mathbb{E}\,|L_t| < \infty$ for some (and hence all) $t \in \mathbb{R}_{>0}$ if and only if

$$\int_{\{|y| \geq c\}} |y|\, \nu(\mathrm{d}y) < \infty \tag{1.1.6}$$

    for some (and hence all) $c \in \mathbb{R}_{>0}$. In that case

$$\mathbb{E}\, L_t = t\left(b + \int_{\{|y| \geq c\}} y\nu(\mathrm{d}y)\right).$$

    In fact, finiteness of any moment of L_t is equivalent to the respective integrability of $|\cdot|$ over $\{|y| \geq c\}$ with respect to ν, that is any Lévy process with bounded jumps has all moments finite. Moreover, if if ν is a symmetric measure and $b = 0$ then L_t is centered for every $t \in \mathbb{R}_{>0}$.
  - L is of bounded variation if and only if

$$\int_{\{|y|<c\}} |y|\, \nu(dy) < \infty \quad \textit{and} \quad Q = 0$$

    for some (and hence all) $c \in \mathbb{R}_{>0}$.
  - L is continuous if and only if $\nu = 0$, that is the only continuous Lévy process is a Wiener process with drift.
  - L is a martingale if and only if (1.1.6) holds and

$$b + \int_{\{|y| \geq c\}} y\nu(\mathrm{d}y) = 0$$

    for some (and hence all) $c \in \mathbb{R}_{c>0}$.
  - L_t has rotationally invariant distribution for some (and hence all) $t \in \mathbb{R}_{>0}$ if and only if $Q = \alpha I$ for some $\alpha \in \mathbb{R}_{\geq 0}$ and ν is rotationally invariant.

- L has infinitely many jumps on any time interval almost surely if and only if ν is infinite.

  More details can be found e.g. in [Applebaum, 2009, Chapter 2.4].

- The condition (1.1.5) on the Lévy measure ν is given by the fact that even though it might be that

$$\sum_{0\le s\le t} |\Delta L_s| \mathbf{1}_{|y|\le 1}(\Delta L_s) = \infty \quad a.s.$$

  for some $t \in \mathbb{R}_{>0}$, we always have

$$\mathbb{E}\left[\sum_{0\le s\le t} |\Delta L_s|^2 \mathbf{1}_{|y|\le 1}(\Delta L_s)\right] = t\int_{|y|\le 1} |y|^2 \,\mathrm{d}\nu(y) < \infty, \quad t \in \mathbb{R}_{\ge 0}.$$

Remark 1.1.3. The decomposition (1.1.4) of a Lévy process L is obtained in several steps which we briefly comment on now. The whole procedure depends on a parameter $c \in \mathbb{R}_{>0}$ that we fix in the sequel of this remark.

- In the first step, we extract the large jump, meaning that we denote

$$\begin{aligned} L_t^{\{|y|<c\}} :=& L_t - \int_{\{|y|\ge c\}} yN(t, \mathrm{d}y) \\ =& L_t - \sum_{0\le s\le t} \left(\mathbf{1}_{\{|y|\ge c\}}(\Delta L_s)\Delta L_s\right), \quad t \in \mathbb{R}_{\ge 0} \end{aligned}$$

  (recall the definition of the integral in (1.1.2)). One can show that $L^{\{|y|<c\}}$ is again a Lévy process, moreover, it is clear, that the jumps of $L^{\{|y|<c\}}$ are bounded by the constant c. As any Lévy process with bounded jumps has all moments finite we can further define

$$\begin{aligned} \tilde{L}_t^{\{|y|<c\}} :=& L_t^{\{|y|<c\}} - \mathbb{E}\, L_t^{\{|y|<c\}} \\ =& L_t^{\{|y|<c\}} - t\mathbb{E}\, L_1^{\{|y|<c\}} \\ =& L_t^{\{|y|<c\}} - tb \end{aligned}$$

  for $t \in \mathbb{R}_{\ge 0}$ where

$$b = \mathbb{E}\left[L_1 - \int_{\{|y|\ge c\}} yN(1, \mathrm{d}y)\right].$$

- In the second step, we show that if all the jumps of L are bounded by c then

$$W_t := L_t - \int_{\{|y|<c\}} y\tilde{N}(t, \mathrm{d}y), \quad t \in \mathbb{R}_{\ge 0}$$

  (see (1.1.3) for the definition of the integral on the right-hand-side) is a Wiener process with some covariance Q. First, one can show that W is continuous local martingale. Then the result follows by Lévy characterisation theorem (see e.g. [Ikeda and Watanabe, 1981, Theorem II.6.3]).

- The previous two steps already show (1.1.4) so it remains to show that the intensity measure

$$\nu(A) := \mathbb{E}\, N(1, A)$$

is a Lévy measure which is closely related to the fact that

$$\int_{\{|y|<c\}} y\tilde{N}(1, \mathrm{d}y)$$

has a finite second moment.

The above described procedure is a particular case of a way how to obtain the semimartingale characteristics when the underlying process is Lévy, see e.g. [Jacod and Shiryaev, 2003, Chapter II].

1.2 Stochastic integration

Let L be a Lévy process with the decomposition (1.1.4). Then we formally define a stochastic integral process with respect to L via (1.1.4) as

$$\begin{aligned}\int_0^\cdot f(t)\mathrm{d}L_t &= \int_0^\cdot f(t)\mathrm{d}\left(bt + W_t + \int_{\{|y|<c\}} y\tilde{N}(t, \mathrm{d}y) + \int_{\{|y|\geq c\}} yN(t, \mathrm{d}y)\right) \\ &= \int_0^\cdot f(t)b\mathrm{d}t + \int_0^\cdot f(t)\mathrm{d}W_t \\ &\qquad + \int_0^\cdot \int_{\{|y|<c\}} f(t)y\tilde{N}(\mathrm{d}t, \mathrm{d}y) + \int_0^\cdot \int_{\{|y|\geq c\}} f(t)yN(\mathrm{d}t, \mathrm{d}y)\end{aligned} \tag{1.2.1}$$

for a suitable $\mathbb{R}^{m\times n}$-valued process f. The first integral on the right-hand side in (1.2.1) is defined pathwise, while the second one is defined using the standard Itô theory, cf. Karatzas and Shreve [1991], if $f : \mathbb{R}_{\geq 0} \times \Omega \to \mathbb{R}^{m\times n}$ is progressively measurable and

$$\int_0^t |f(s)|^2 \,\mathrm{d}s < \infty \quad \text{a.s.}$$

for every $t \in \mathbb{R}_{\geq 0}$. The aim of this section is to summarize the standard procedure how to define the last two terms in (1.2.1) following the approach in [Applebaum, 2009, Chapter 4]. In fact, we even define more generally

$$\int_0^\cdot \int_{\{|y|<c\}} H(t, y)\tilde{N}(\mathrm{d}t, \mathrm{d}y) \quad \text{and} \quad \int_0^\cdot \int_{\{|y|\geq c\}} H(t, y)N(\mathrm{d}t, \mathrm{d}y)$$

for suitable processes $H : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \times \Omega \to \mathbb{R}^m$.

First, we give two definitions.

Definition 1.2.1. Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ be a stochastic basis and $n \in \mathbb{N}$. The σ-algebra $\mathscr{P}$ which is defined as the smallest σ-algebra for which all mappings $F : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \times \Omega \to \mathbb{R}$ such that

- $(y, \omega) \mapsto F(t, y, \omega)$ is $\mathscr{B}(\mathbb{R}^n) \otimes \mathscr{F}_t$-measurable for any $t \in \mathbb{R}_{\geq 0}$,

- $t \mapsto F(t, y, \omega)$ is left-continuous for any $y \in \mathbb{R}^n$, $\omega \in \Omega$

are measurable is called *predictable σ-algebra.* If $H : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \times \Omega \to \mathbb{R}$ is $\mathscr{P}$-measurable we say that H is *predictable*, similarly we may treat also space-homogeneous mappings $H : \mathbb{R}_{\geq 0} \times \Omega \to \mathbb{R}$.

Note that any predictable process is progressively measurable, cf. [Cohen and Elliott, 2015, Remark 7.2.2.].

Definition 1.2.2. Let μ be a σ-finite measure on a measurable space $(X, \mathscr{X})$. A family of random variables $\{N(A), A \in \mathscr{X}\}$ defined on some probability space is called *Poisson random measure with intensity* μ if

- $N(A)$ follows Poisson distribution with intensity rate $\mu(A)$ for any $A \in \mathscr{X}$ (with $N(A) = \infty$ almost surely if $\mu(A) = \infty$),
- $N(A_1), \dots, N(A_n)$ are independent if $A_1, \dots, A_n \in \mathscr{X}$ are disjoint,
- almost surely we have that $A \mapsto N(A)$ defines a measure on $(X, \mathscr{X})$.

It follows from [Applebaum, 2009, Chapter 2] that the jump measure (1.1.1) defines a Poisson random measure with intensity μ that satisfies

$$\mu([0, t] \times A) = t\nu(A), \quad t \in \mathbb{R}_{\geq 0}, A \in \mathscr{B}(\mathbb{R}^n),$$

where ν it the Lévy measure from Theorem 1.1.1 (here we take $X = \mathbb{R}_{\geq 0} \times \mathbb{R}^n)^2$. Thus, we may define the (pathwise) integral

$$\int_0^t \int_A H(s, y) N(\mathrm{d}s, \mathrm{d}y), \qquad t \in \mathbb{R}_{\geq 0}, A \in \mathscr{B}(\mathbb{R}^n), 0 \notin \bar{A} \tag{1.2.2}$$

for any measurable $H : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \times \Omega \to \mathbb{R}^m$ as a finite random sum due to the fact that $\nu(A) < \infty$ if $A \in \mathscr{B}(\mathbb{R}^n), 0 \notin \bar{A}$. In the case of deterministic integrand such that $H(s, y) = h(s, y)$ almost surely for every $y \in \mathbb{R}^n$ and for some measurable $h : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \to \mathbb{R}^m$ we have

$$\begin{aligned} \mathbb{E}\left[\int_0^t \int_A h(s, y) N(\mathrm{d}s, \mathrm{d}y)\right] &= \int_0^t \int_A h(s, y) \mathrm{d}s \mathrm{d}\nu(y), \\ \mathbb{VAR}\left[\int_0^t \int_A h(s, y) N(\mathrm{d}s, \mathrm{d}y)\right] &= \int_0^t \int_A |h(s, y)|^2 \, \mathrm{d}s \mathrm{d}\nu(y) \end{aligned}$$

for $t \in \mathbb{R}_{\geq 0}, A \in \mathscr{B}(\mathbb{R}^n), 0 \notin \bar{A}$ and by the construction of N also

$$\int_0^t \int_A h(s, y) N(\mathrm{d}s, \mathrm{d}y) = \sum_{0 \leq s \leq t} \left(\mathbf{1}_A(\Delta L_s) H(s, \Delta L_s)\right), \quad a.s.$$

which is in compliance with (1.1.2).

We always assume that H in (1.2.2) is predictable to ensure nice properties of the integral.

More delicate is the case of the integral with respect to the compensated measure $\tilde{N}$ where the pathwise approach fails and an L^2-theory is used instead.

[2] The jump measure N moreover satisfies that $N(t, A) - N(s, A)$ is independent of $\mathscr{F}_s$ if it is given by an $(\mathscr{F}_t)$-Lévy process for any $0 \leq s < t < \infty$ and $A \in \mathscr{B}(\mathbb{R}^n)$. In such case we talk about an $(\mathscr{F}_t)$-Poisson process

Definition 1.2.3. Let $(X, \mathscr{X})$ be a measurable space, $\mathscr{U}$ be an algebra over X that generates the σ-algebra $\mathscr{X}$ and ρ be a σ-finite measure on the product space $(\mathbb{R}_{\geq 0} \otimes X, \mathscr{B}(\mathbb{R}_{\geq 0}) \otimes \mathscr{X})$. A family of random variables $\{M(t, A), t \in \mathbb{R}_{\geq 0}, A \in \mathscr{U}\}$ on a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$ with normal filtration $(\mathscr{F}_t)_{t\geq 0}$ is called an *$(\mathscr{F}_t)$-martingale-valued measure on $\mathbb{R}_{\geq 0} \times X$ with controlling measure ρ* if

- almost surely for any $t \in \mathbb{R}_{\geq 0}$ we have that $M(t, \cdot)$ is a pre-measure on $\mathscr{U}$,
- the process $t \mapsto M(t, A), t \in \mathbb{R}_{\geq 0}$ is a càdlàg martingale for any $A \in \mathscr{U}$,
- the increment $M(t, A) - M(s, A)$ is independent of $\mathscr{F}_s$ for any $0 \leq s < t < \infty$ and $A \in \mathscr{U}$,
- we have

$$\mathbb{E}\, M(t, A)^2 = \rho(t, A), \quad t \in \mathbb{R}_{\geq 0}, A \in \mathscr{U}.$$

An important example of a martingale-valued measure to which the developed theory will be applied comes from the jump measure of a Lévy process L. Recall that we denoted N the jump measure defined by (1.1.1). Then we define the compensated Poisson measure by

$$\tilde{N}(t, A) = N(t, A) - t\nu(A), \quad t \in \mathbb{R}_{\geq 0}, A \in \mathscr{B}(\mathbb{R}^n), 0 \notin \bar{A}, \tag{1.2.3}$$

where ν is the intensity measure of $N(1, \cdot)$. Note that $\tilde{N}$ is well defined as $\nu(A) < \infty$ for $A \in \mathscr{B}(\mathbb{R}^n), 0 \notin \bar{A}$ by the fact that there are only finitely many jumps of size larger than some fixed constant on any finite interval for the underlying Lévy process almost surely. One can show that $\tilde{N}$ is a martingale-valued measure on $\mathbb{R}_{\geq 0} \times \mathbb{R}^n$ with controlling measure $\mathrm{d}t\nu(\mathrm{d}y)$.

Now let M be as in Definition 1.2.3 and fix $T \in \mathbb{R}_{>0}$. Details of the following construction may be found in [Applebaum, 2009, Chapter 4]. We define $\mathbb{H}$ to be the Hilbert space of equivalence classes of predictable mappings $H : [0, T] \times \mathbb{R}^n \times \Omega \mapsto \mathbb{R}^m$ with respect to equality $\mathbb{P} \otimes \rho$-almost everywhere such that

$$\mathbb{E}\left[\int_0^T \int_{\mathbb{R}^n} |H(t, y)|^2 \rho(\mathrm{d}t, \mathrm{d}y)\right] < \infty$$

equipped with the scalar product

$$\langle H_1, H_2\rangle_{\mathbb{H}} := \mathbb{E}\left[\int_0^T \int_{\mathbb{R}^n} \langle H_1(t, y), H_2(t, y)\rangle \rho(\mathrm{d}t, \mathrm{d}y)\right], \quad H_1, H_2 \in \mathbb{H}.$$

Now we define the integral of a simple process against M. By a simple process we mean a linear combination

$$H = \sum_{j=1}^{m} \sum_{k=1}^{n} \alpha_k F_{t_j} \mathbf{1}_{(t_j, t_{j+1}]} \mathbf{1}_{A_k} \tag{1.2.4}$$

for some $0 \leq t_1 < t_2 < \cdots < t_{m+1} \leq T$ $\alpha_1, \ldots \alpha_k \in \mathbb{R}$, $k \in \mathbb{N}$, F_{t_j} bounded, $\left(\mathscr{F}_{t_j}\right)$-measurable and $\mathbb{R}^m$-valued random variables and $A_1, \ldots, A_n \in \mathscr{U}$, $n \in \mathbb{N}$. For a simple process of the form (1.2.4) we set

$$I_{T,M}(H) := \sum_{j=1}^{m} \sum_{k=1}^{n} \alpha_k F_{t_j} \left(M(t_{j+1}, A_k) - M(t_j, A_k)\right). \tag{1.2.5}$$

If we denote $\mathbb{S} \subset \mathbb{H}$ the linear space of simple processes of the form (1.2.4) then one can show that

$$I_{T,M} : \mathbb{S} \to L^2(\Omega, \mathbb{R}^m)$$

is a linear isometry. As $\mathbb{S}$ is dense in $\mathbb{H}$ one can uniquely extend $I_{T,H}$ to $\mathbb{H}$. Similarly to the classical case of Itô integral we may further extend $I_{T,H}$ to the linear space $\mathbb{P}$ of equivalence classes of predictable mappings $H : [0,T] \times \mathbb{R}^n \times \Omega \mapsto \mathbb{R}^m$ with respect to equality $\mathbb{P} \otimes \rho$-almost everywhere such that

$$\int_0^T \int_{\mathbb{R}^n} |H(t,y)|^2 \, \rho(\mathrm{d}t, \mathrm{d}y) < \infty, \quad a.s. \tag{1.2.6}$$

It is due to the fact that $\mathbb{S}$ is dense in $\mathbb{P}$ if $\mathbb{P}$ is equipped with a suitable (and sufficient for our considerations) topology (see [Applebaum, 2009, Chapter 4.2.2] for details) and the inequality due to [Gikhman and Skorokhod, 1972, page 20]

$$\mathbb{P}\left(|I_{T,M}(H)| > \epsilon\right) \le \frac{K}{\epsilon^2} + \mathbb{P}\left(\int_0^T \int_{\mathbb{R}^n} |H(t,y)|^2 \, \rho(\mathrm{d}t, \mathrm{d}y) > K\right)$$

for any $\epsilon, K \in \mathbb{R}_{>0}$ and $H \in \mathbb{S}$. Finally, one can show that if a predictable mapping $H : [0,T] \times \mathbb{R}^n \times \Omega \mapsto \mathbb{R}^m$ satisfies (1.2.6) for every $T \in \mathbb{R}_{>0}$ then there exists a modification of the process

$$\{I_{t,M}(H), t \in \mathbb{R}_{\ge 0}\}$$

that is a càdlàg local martingale. This modification is then denoted as

$$\int_0^{\cdot} H \mathrm{d}M.$$

In the special case of a martingale-valued measure $\tilde{N}$ with a control measure $\mathrm{d}t\mathrm{d}\nu(dy)$ associated with a Lévy process we use notation

$$\int_0^{\cdot} \int_{\{|y|<c\}} H(t,y)\tilde{N}(\mathrm{d}t, \mathrm{d}y) = \int_0^{\cdot} \mathbf{1}_{\{|y|<c\}} H \mathrm{d}\tilde{N}$$

if $\mathbf{1}_{\{|y|<c\}} H \in \mathbb{P}$, that is if

$$\int_0^T \int_{\{|y|<c\}} |H(t,y)|^2 \, \mathrm{d}t\mathrm{d}\nu(y) < \infty \quad \text{a.s.}$$

This notation is compatible with the previously adapted notation, that is we have

$$\int_0^t \int_{\{|y|<c\}} y\tilde{N}(\mathrm{d}s, \mathrm{d}y) = \int_{\{|y|<c\}} y\tilde{N}(t, \mathrm{d}y), \quad t \in \mathbb{R}_{\ge 0}, \quad \text{a.s.}$$

in the Lévy-Itô decomposition (1.1.4). Moreover, by the fact that $I_{T,H}$ is isometric between $\mathbb{H}$ and $L^2(\Omega, \mathbb{R}^m)$ we obtain the Itô-type isometry

$$\mathbb{E}\left[\int_0^T \int_{\{|y|<c\}} |H(t,y)|^2 \, \tilde{N}(\mathrm{d}t, \mathrm{d}y)\right] = \mathbb{E}\left[\int_0^T \int_{\{|y|<c\}} |H(t,y)|^2 \, \mathrm{d}t\mathrm{d}\nu(y)\right] \tag{1.2.7}$$

for $H \in \mathbb{H}$.

Remark 1.2.1. Using the above construction we can give a meaning to a process with the following decomposition

$$\int_0^\cdot F(s)\mathrm{d}s + \int_0^\cdot G(s)\mathrm{d}W_s + \int_0^\cdot \int_{\{|y|<c\}} H(s,y)\tilde{N}(\mathrm{d}s,\mathrm{d}y) \\ + \int_0^\cdot \int_{\{|y|\geq c\}} K(s,y)N(\mathrm{d}s,\mathrm{d}y) \tag{1.2.8}$$

for predictable mappings $F : \mathbb{R}_{\geq 0} \times \Omega \to \mathbb{R}^m$, $G : \mathbb{R}_{\geq 0} \times \Omega \to \mathbb{R}^{m\times n}$, $H, K : \mathbb{R}_{\geq 0} \times \mathbb{R}^n \times \Omega \to \mathbb{R}^m$ where we assume

$$\int_0^T \left(|F(s)| + |G(s)|^2 + \int_{\{|y|<c\}} |H(s,y)|^2 \nu(\mathrm{d}y) \right) \mathrm{d}t < \infty \quad \text{a.s.}$$

for every $T \in \mathbb{R}_{>0}$. Processes of the form (1.2.8) are called *Lévy-type process* and are càdlàg semimartingales. For more insight into a further generalisation of the processes of the form (1.2.8) we recommend [Ikeda and Watanabe, 1981, Chapter II] (however, a rather interesting fact is that we cannot construct the process (1.2.8) where W and N would be defined on the same filtered space and would not be independent at the same time, see [Ikeda and Watanabe, 1981, Theorem II.6.3]). On the other hand, for some Lévy processes we do not need to follow the construction leading to a Lévy-type process if we want to define a stochastic integral process with respect to them. For instance, in the case when L^α is a symmetric α-stable process with $\alpha \in (0,2)$, that is its decomposition (1.1.4) satisfies $b = 0$, $Q = 0$ and ν has a density

$$\nu(dy) = c\frac{\mathrm{d}y}{|y|^{n+\alpha}}$$

for some $c \in \mathbb{R}_{>0}$, then we may define

$$\int_0^\cdot f \mathrm{d}L^\alpha$$

directly via the usual Riemann sums as a limit in a properly chosen space if $f : \mathbb{R}_{\geq 0} \times \Omega \to \mathbb{R}^{m\times n}$ is predictable and

$$\int_0^t |f(s)|^\alpha \mathrm{d}s < \infty \quad \text{a.s.} \tag{1.2.9}$$

for every $t \in \mathbb{R}_{\geq 0}$. The condition (1.2.9) is different from the condition that one has to impose in the case of the integral process of the form (1.2.8). The details can be found in Rosinski and Woyczynski [1986].

We conclude this Section with the Itô formula.

Theorem 1.2.1. Let X be a process with decomposition (1.2.8). If $V \in \mathscr{C}^2(\mathbb{R}^m)$,

$DV \in \mathscr{C}_b(\mathbb{R}^m;\mathbb{R}^m)$, $D^2V \in \mathscr{C}_b(\mathbb{R}^m;\mathbb{R}^{m\times m})$ then we have

$$\begin{aligned}
V(X_T) = &\int_0^T \left(\left\langle F(t), DV(X_t)\right\rangle + \frac{1}{2}\operatorname{Tr}\left(G(t)^T D^2V(X_t)G(t)\right)\right)\mathrm{d}t\\
&+ \int_0^T \left\langle G(t)^T DV(X_t), \cdot\right\rangle \mathrm{d}W_t\\
&+ \int_0^T \int_{\{|y|<c\}} \left[V(X_{t-} + H(t,y)) - V(X_{t-})\right] \tilde{N}(\mathrm{d}t, \mathrm{d}y)\\
&+ \int_0^T \int_{\{|y|<c\}} \Big[V(X_t + H(t,y)) - V(X_t)\\
&\qquad - \left\langle DV(X_t), H(t,y)\right\rangle\Big]\,\nu(\mathrm{d}y)\,\mathrm{d}t\\
&+ \int_0^T \int_{\{|y|\geq c\}} \left[V(X_{t-} + K(t,y)) - V(X_{t-})\right] N(\mathrm{d}t, \mathrm{d}y).
\end{aligned}$$

almost surely for every $T \in \mathbb{R}_{\geq 0}$.

Proof. See Theorem 4.4.7 and Remark below in Applebaum [2009]. □

1.3 Stochastic differential equations

Similarly as in the previous Section where we exploited the Lévy-Itô decomposition (1.1.4) to define the integral with respect to a Lévy process L as in (1.2.1) we now formally consider stochastic differential equations of the form

$$\begin{aligned}
X_t &= x_0 + \int_0^t f(X_{s-})\mathrm{d}L_s\\
&= x_0 + \int_0^t f(X_{s-})\mathrm{d}\left(bs + W_s + \int_{\{|y|<c\}} y\tilde{N}(\mathrm{d}s, \mathrm{d}y) + \int_{\{|y|<c\}} yN(\mathrm{d}s,\mathrm{d}y)\right)\\
&= x_0 + \int_0^t f(X_{s-})b\mathrm{d}s + \int_0^t f(X_{s-})\mathrm{d}W_s\\
&\quad + \int_0^t \int_{\{|y|<c\}} f(X_{s-})y\tilde{N}(\mathrm{d}s,\mathrm{d}y) + \int_0^t \int_{\{|y|<c\}} f(X_{s-})yN(\mathrm{d}s,\mathrm{d}y)
\end{aligned} \tag{1.3.1}$$

for a suitable coefficient $f : \mathbb{R}_{\geq 0} \times \mathbb{R}^m \to \mathbb{R}^{m\times n}$ and some initial condition x_0. Having the general form of Lévy-type process from Remark 1.2.1 in mind we arrive at the following definition of a stochastic differential equation and its solution.

Definition 1.3.1. Let $m, n \in \mathbb{N}, c \in \mathbb{R}_{>0}$ and suppose that Borel functions

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

are given. Suppose that $\left(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in\mathbb{R}_{>0}}, \mathbb{P}\right)$ is a filtered probability space with normal filtration, W is an $(\mathscr{F}_t)$-Wiener process with values in $\mathbb{R}^n$ which is independent to an $(\mathscr{F}_t)$-Poisson random measure N on $\mathbb{R}_{\geq 0}\times(\mathbb{R}^n\setminus\{0\})$ with intensity $\mathrm{d}t\nu(\mathrm{d}y)$, where ν is a Lévy measure and $\tilde{N}$ is the compensator of N and x_0 is an $\mathscr{F}_0$-measurable $\mathbb{R}^m$-valued random variable. An $\mathbb{R}^m$-valued $(\mathscr{F}_t)$-progresively measurable càdlàg process X that satisfies

$$X_t = x_0 + \int_0^t f(s, X_s)\,\mathrm{d}s + \int_0^t g(s, X_s)\,\mathrm{d}W_s$$
$$+ \int_0^t \int_{\{|y|<c\}} H(s, X_{s-}, y)\,\tilde{N}(\mathrm{d}s, \mathrm{d}y) + \int_0^t \int_{\{|y|\geq c\}} H(s, X_{s-}, y)\,N(\mathrm{d}s, \mathrm{d}y) \quad \mathbb{P}\text{-a.s.} \tag{1.3.2}$$

for all $t \in \mathbb{R}_{\geq 0}$ is said to be a solution to the equation

$$\mathrm{d}X_t = f(t, X_t)\,\mathrm{d}t + g(t, X_t)\,\mathrm{d}W_t + \int_{\{|y|<c\}} H(t, X_{t-}, y)\,\tilde{N}(\mathrm{d}t, \mathrm{d}y)$$
$$+ \int_{\{|y|\geq c\}} H(t, X_{t-}, y)\,N(\mathrm{d}t, \mathrm{d}y), \qquad t \geq 0 \tag{1.3.3}$$

with the initial condition x_0. We say that the solution to (1.3.3) with the initial condition x_0 is unique if whenever X, Y are solutions to (1.3.3) with the initial condition x_0 then

$$\mathbb{P}\,(X(t) = Y(t), \quad t \in \mathbb{R}_{\geq 0}) = 1.$$

Denoting

$$a(t, x, z) = g(t, x)\,(g(t, y))^T\,, \quad t \in \mathbb{R}_{\geq 0}, \quad x, y \in \mathbb{R}^m$$

we formulate the standard sufficient conditions for existence of a solution to (1.3.3) (see Applebaum [2009], Section 6.2).
General Lipschitz condition: There exists measurable and locally bounded function $l_1 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ such that

$$|f(s, x) - f(s, z)|^2 \vee |a(t, x, x) - 2a(t, x, z) + a(t, z, z)|^2$$
$$\vee \int_{\{|y|<c\}} |H(t, x, y) - H(t, z, y)|^2\,\nu(\mathrm{d}y) \leq l_1(t)\,|x - z|^2\,, \tag{1.3.4}$$

for any $x, z \in \mathbb{R}^m$.
General growth condition: There exists measurable and locally bounded function $l_2 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$

$$|f(t, x)|^2 + |a(t, x, x)| + \int_{\{|y|<c\}} |H(t, x, y)|^2\,\nu(\mathrm{d}y) \leq l_2(t)(1 + |x|^2), \tag{1.3.5}$$

for any $x \in \mathbb{R}^m$.
Continuity condition: We have

$$H(\cdot, y) \in \mathscr{C}(\mathbb{R}^m; \mathbb{R}^m) \tag{1.3.6}$$

for all $y \in \{|y| \geq c\}$.

Remark 1.3.1. Based on [Applebaum, 2009, Section 6.2] several remarks shall be made.

- The existence and uniqueness of a solution to (1.3.3) under (1.3.4),(1.3.5) and (1.3.6) is shown using so-called *interlacing*. That is, first, we omit the integral with respect to N and look for a suitable process $\tilde{X}$ that would be a solution to

$$\mathrm{d}\tilde{X}_t = f(t, \tilde{X}_t)\,\mathrm{d}t + g(t, \tilde{X}_t)\,\mathrm{d}W_t + \int_{\{|y|<c\}} H(t, \tilde{X}_{t-}, y)\,\tilde{N}(\mathrm{d}t, \mathrm{d}y), \quad (1.3.7)$$

for $t \geq 0$. Such (unique) process can be found using the standard Itô isometry for the Wiener integral and the Itô-type isometry (1.2.7) together with the fact that all the stochastic integrals defining $\tilde{X}$ can be shown to be L^2-martingales and thus the classical Doob's inequality can be applied leading to the technique of Picard iterations. If $\nu(\{|y| \geq c\}) = 0$, that is there are now jumps of magnitude larger than c, the process $\tilde{X}$ is already a solution to (1.3.3). In the case $\nu(\{|y| \geq c\}) > 0$ we construct the Markov times of big jumps

$$\tau_0 = 0, \quad \tau_n = \inf\{t > \tau_{n-1} : N(t, \{|y| \geq c\}) = n\}, \quad n \in \mathbb{N}$$

and we obtain that $\tau_n > \tau_{n-1}$ for $n \in \mathbb{N}$ and $\tau_n \to \infty$ as $n \to \infty$. The solution X to (1.3.3) is then obtained as

$$\begin{aligned}
X(t) &= \tilde{X}_0(t), \quad 0 \leq t < \tau_1 \\
X(\tau_1) &= \lim_{s \to t+} \tilde{X}_0(s) + H(\lim_{s \to t+} \tilde{X}_0(s), \Delta P(\tau_1)), \quad t = \tau_1 \\
X(t) &= X(\tau_1) + \tilde{X}_1(t - \tau_1), \quad \tau_1 < t < \tau_2 \\
X(\tau_2) &= \dots
\end{aligned}$$

where $\tilde{X}_n$ is the (unique) solution to (1.3.7) with initial condition $X(\tau_n)$ for $n \in \mathbb{N} \cup \{0\}$ and

$$P_t = \int_{\{|y| \geq c\}} y N(t, \mathrm{d}y), \quad t \in \mathbb{R}_{\geq 0}$$

is the compound Poisson process composed of the large jumps. The uniqueness of the solution to (1.3.3) follows from the uniqueness for (1.3.7) and by the above interlacing procedure.

- The interlacing procedure is only one possible way how to show existence of a solution in our case. For more general semimartingale setting see [Protter, 2005, Chapter 5].

- Generally, the solution to (1.3.3) has discontinuities originating in the integrals driven by $\tilde{N}$ and N. It can be shown that no jumps occur almost surely if

$$\nu\left(y \in \mathbb{R}^n \setminus \{0\} : |H(x, y)| > 0\right) = 0, \quad (1.3.8)$$

for $x \in \mathbb{R}^m$, which corresponds to the case when (1.3.3) can be rewritten as a classical equation driven by a Wiener process. The necessary condition for path-continuity is more delicate as the solution might avoid points $x \in \mathbb{R}^m$ for which (1.3.8) fails even if the driving jump processes $\tilde{N}$ and N are non-degenerate.

- Similarly to the Wiener case it can be shown that under (1.3.4),(1.3.5) and (1.3.6) if $\nu(\{|y| \geq c\}) = 0$ then for every $t \in \mathbb{R}_{\geq 0}$ there exists a constant $l(t) \in \mathbb{R}_{>0}$ such that the second moment of the solution X to (1.3.3) can be bounded

$$\mathbb{E}\, |X(t)|^2 \leq l(t) \left(1 + \mathbb{E}\, |X(0)|^2\right).$$

  That is, whenever we have an L^2-integrable initial condition, the whole solution is L^2-integrable (with the second moment possibly growing unboundedly). This, however, may fail if the large jumps are present, i.e. when $\nu(\{|y| \geq c\}) > 0$ and the solution might not have any positive moment finite.

- If we replace (1.3.4) and (1.3.5) by its obvious counterparts that are *local in space coordinates* $x, z \in \mathbb{R}^m$ we obtain that there exists a *unique local solution* to (1.3.3). That is there exists an $(\mathscr{F}_t)$-progressively measurable càdlàg process X that satisfies (1.3.2) for all $t \in \mathbb{R}_{\geq 0}$ almost surely on the set $\{\omega \in \Omega : t \leq \tau(\omega)\}$, where τ is a suitable almost surely positive Markov time usually called *the explosion time for* (1.3.3).

1.4 Stochastic approximation procedures

Let $R : \mathbb{R}^m \to \mathbb{R}^m$ be an unknown function representing *regression equations*. Our goal is to find a root of R, a point $x_0 \in \mathbb{R}^m$ such that $R(x_0) = 0$ even though we are only able to measure the values $R(x), x \in \mathbb{R}^m$ up to some error. We aim at providing an appropriate experiment consisting of consecutive measurements of R at some points x_t for which the convergence $x_t \to x_0$ as t goes to infinity is guaranteed under only mild assumptions on the behavior of R such as, for example, continuity or monotonicity. While several methods such as Newton's tangent method work in the case when the measurement error is negligible and provide rapid convergence to a root x_0 under mild assumptions, one must be satisfied with slower convergence when the measurement error is significant. We briefly summarize basic facts from the theory of stochastic approximation procedures that tackles the case of significant measurement error which is mostly due to Robbins and Monro in Robbins and Monro [1951] (see Nevel'son and Khas'minskiĭ [1972] for a comprehensive exposition and source for this Section).

We begin with a discrete-time case, i.e. a situation when R is measured at points x_t for $t \in \mathbb{N}$. In this case it is natural to assume that the measurement of R at a point x at a time t is given as

$$R(x) + G(t, x) \tag{1.4.1}$$

for some measurable $G : \mathbb{N} \times \mathbb{R}^m \times \Omega \to \mathbb{R}^m$, where Ω represents a probability space, such that $\mathbb{E}\, G(t, x) = 0$ for any $t \in \mathbb{N}$ and $x \in \mathbb{R}^m$. The so-called *Robbins-Monro procedure* in this case is given as a sequence X_t for $t \in \mathbb{N}$ where X_0 is some given initial point and

$$X_{t+1} = X_t + \alpha(t) \left(R(X_t) + G(t, X_t)\right), \quad t \in \mathbb{N}, \tag{1.4.2}$$

for some $\{\alpha(t)\}_{t\in\mathbb{N}} \subset \mathbb{R}_{>0}$ such that

$$\sum_{t=1}^{\infty} \alpha(t) = \infty \quad \text{and} \quad \sum_{t=1}^{\infty} \alpha^2(t) < \infty. \tag{1.4.3}$$

One can show that for an L^2 convergence

$$\mathbb{E}\,|X_t - x_0|^2 \to 0, \quad t \to \infty,$$

for any initial condition x_0 in the real case, that is $m = 1$, is is enough that R is strictly decreasing continuous bounded function and $\mathbb{E}\,G^2(t,x)$ is bounded uniformly in $t \in \mathbb{N}$ and $x \in \mathbb{R}$ (cf. Robbins and Monro [1951]).

The imposed condition (1.4.3) forces the sequence of coefficients $\alpha(t)$ to decrease to zero under a specific mode, allowing us to take for instance $\alpha(t) = 1/t$. It easy to see that if the convergence is too fast, namely if $\sum_{t=1}^{\infty} \alpha(t) < \infty$ it can't be that $X_t \to x_0$ even in the case of zero measurement error, $G \equiv 0$, for any initial condition X_0 as in this case we always have

$$\sum_{t=1}^{\infty} |X_{t+1} - X_t| \le \left(\sup_{x\in\mathbb{R}} |R(x)|\right) \sum_{t=1}^{\infty} \alpha(t),$$

where the right-hand-side is finite and does not depend on X_0. Therefore, if the root x_0 is sufficiently far from our initial guess X_0 we cannot hope for the desired convergence as the sum of increments $X_{t+1} - X_t$ is not allowed to reach any value.

On the other hand, the second condition in (1.4.3) is only sufficient and not necessary. It can be weakened as in [Nevel'son and Khas'minskiĭ, 1972, Theorem 4.5] where it is only assumed that

$$\sum_{t=1}^{\infty} \alpha(t) = \infty \quad \text{and} \quad \alpha(t) \to 0, \quad t \to \infty.$$

The continuous-time case, when the measurement of R is given as in (1.4.1) but now with $t \in \mathbb{R}_{>0}$ was considered in Driml and Nedoma [1951], Sakrison [1964] and Cypkin and Nikolić [1970] where the continuous analog of (1.4.2) is naturally given as

$$\mathrm{d}X_t = \alpha(t)\left(R(X_t) + \sigma(t, X_t)\mathrm{d}W_t\right), \quad t \in \mathbb{R}_{\ge 0} \tag{1.4.4}$$

with initial condition $X_0 \in \mathbb{R}^m$ where we replaced $G(t,x)$ by a stochastic differential

$$\sigma(t, X_t)\mathrm{d}W_t$$

for some (possibly unknown) coefficient $\sigma : \mathbb{R}_{\ge 0} \times \mathbb{R}^m \to \mathbb{R}^{m\times n}$ and Wiener process W. Note that in this setting the measurement errors of R are represented by Gaussian white noise (cf. Section 3 of this work where more general noise is treated). Using the fact that the set of solutions to (1.4.4) for different (deterministic) initial conditions defines a Markov process with known generator one can show the following result in the case of one dimension and time-homogeneous diffusion coefficient:

Theorem 1.4.1. Assume that $x_0 \in \mathbb{R}^m$ and $L \in \mathbb{R}_{>0}$ are such that

$$
\begin{aligned}
&R(x)(x - x_0) < 0, \quad x \neq x_0, \\
&\sigma^2(x) \leq L(1 + x^2), \quad x \in \mathbb{R}, \\
&\int_0^\infty \alpha(t)\mathrm{d}t = \infty \quad \text{and} \quad \int_0^\infty \alpha^2(t)\mathrm{d}t < \infty,
\end{aligned}
$$

then any solution (if it exists) to (1.4.4) satisfies

$$
\lim_{t \to \infty} X_t = x_0, \quad \text{a.s.}
$$

Proof. Follows by [Nevel'son and Khas'minskiĭ, 1972, Theorem 3.8.2] □

It is easy to see that in the case when x_0 is the (unique) root of R Theorem 1.4.1 provides sufficient conditions on the noise coefficient and the unknown function R such that the stochastic approximation procedure given by (1.4.4) provides desired convergence.

In the multivariate case, $m > 1$, the same conclusion as in Theorem 1.4.1 can be obtained under conditions on R and σ that are typically expressed indirectly through existence of some type of *Lyapunov function* $V : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$. In one simple case we shall assume that V is sufficiently smooth and satisfies

$$
\begin{aligned}
V(x) = 0 \iff x = x_0 \quad \text{and} \quad V(x) \to \infty, \quad |x| \to \infty, \\
\langle R(x), DV(x) \rangle < 0, \quad x \neq x_0, \\
\mathrm{Tr}\Big(\sigma(t,x)^T D^2 V(x) \sigma(t,x)\Big) \leq L\Big(1 + V(x)\Big), \quad t \in \mathbb{R}_{\geq 0}, x \in \mathbb{R}^m
\end{aligned}
$$

for some $x_0 \in \mathbb{R}^m$ and $L \in \mathbb{R}_{>0}$, where DV and D^2V denote the first and second Fréchet derivatives of V. These conditions on V can be interpreted geometrically: under these conditions any solution X to (1.4.4) reaches eventually the surface $\{x \in \mathbb{R}^m : V(x) = c\}$ at some Markov time and afterwards remains in the set $\{x \in \mathbb{R}^m : V(x) < c\}$ which can be formally seen from the inequality

$$
\frac{\mathrm{d}V(X_t)}{\mathrm{d}t} = \alpha(t) \langle R(X_t), DV(X_t) \rangle < 0,
$$

(cf. [Nevel'son and Khas'minskiĭ, 1972, Section 4.4]). An example of a Lyapunov function is

$$
V(x) = \langle C(x - x_0), x - x_0 \rangle, \quad x \in \mathbb{R}^m
$$

for some $x_0 \in \mathbb{R}^m$ and a positive-definite matrix $C \in \mathbb{R}^m$. This choice leads to the following conditions on R and σ

$$
\begin{aligned}
&\langle R(x), C(x - x_0) \rangle < 0, \quad x \neq x_0 \\
&|\sigma(t,x)| \leq L\Big(1 + |x|\Big), \quad t \in \mathbb{R}_{\geq 0}, x \in \mathbb{R}^m
\end{aligned}
$$

for some $L \in \mathbb{R}_{>0}$.

1.5 Invariant measures for Markov processes

In this Section we remind basic notions related to the theory of Markov processes with emphasis on the relation to stochastic differential equations. We conclude this Section by stating the celebrated Krylov-Bogolyubov Theorem which is an example of how the theory of Markov processes can help us when studying properties of stochastic equations.

Definition 1.5.1. Let $m \in \mathbb{R}^m$. Any function $P : \mathbb{R}^m \times \mathscr{B}(\mathbb{R}^m) \to [0,1]$ that satisfies

- $P(x, \cdot)$ is a Borel probability measure on $\mathbb{R}^m$ for any $x \in \mathbb{R}^m$,
- $P(\cdot, A)$ is a measurable function for any $A \in \mathscr{B}(\mathbb{R}^m)$

is called a *Markov kernel.*

Naturally, any Markov kernel P can be understood as a linear operator on the space of measurable bounded functions $B_b(\mathbb{R}^m)$ by the following relation (this operator is again denoted as P):

$$Pf(x) := \int_{\mathbb{R}^m} f(y)P(x, \mathrm{d}y), \quad f \in B_b(\mathbb{R}^m).$$

One can show that in fact P is then a bounded linear operator on $B_b(\mathbb{R}^m)$ and thus it makes a good sense to talk about a composition of Markov kernels, which leads us to the following definition.

Definition 1.5.2. A family of Markov kernels $P = \{P_t\}_{t\in\mathbb{R}_{\geq 0}}$ is called a *(time-homogeneous) transition semigroup* if

$$P_{s+t} = P_s P_t, \quad s, t \in \mathbb{R}_{\geq 0} \tag{1.5.1}$$

and

$$P_0 \text{ is the identity operator on } B_b(\mathbb{R}^m) \tag{1.5.2}$$

Remark 1.5.1.

- The equality (1.5.1) is called *Chapman-Kolmogorov equality* and is closely connected to a Markov property,
- The condition (1.5.2) is not standardized across literature as some authors do not include it in the definition of transition semigroup and those transition semigroups that satisfy (1.5.2) are then called *normal transition semigroups.*

Now we can proceed to the definition of a Markov process which is originally due to Dynkin [1965].

Definition 1.5.3. Let $\left(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in\mathbb{R}_{\geq 0}}\right)$ be a filtered measurable space, $X = \{X_t, t \in \mathbb{R}_{\geq 0}\}$ be an $(\mathscr{F}_t)$-adapted $\mathbb{R}^m$-valued process and $(\mathbb{P}_x)_{x\in\mathbb{R}^m}$ be a family of probability measures on $(\Omega, \mathscr{F})$ and P be a transition semigroup such that

$$\mathbb{P}_x\left(\{\omega \in \Omega : X_0(\omega) = x\}\right) = 1$$

for any $x \in \mathbb{R}^m$ and

$$\mathbb{P}_x\left(\{\omega \in \Omega : X_t(\omega) \in A | \mathscr{F}_s\}\right) = P_{t-s}(X_s, A), \quad \mathbb{P}_x\text{-a.s.}$$

for any $x \in \mathbb{R}^m$, $0 \leq s \leq t < \infty$ and $A \in \mathscr{B}(\mathbb{R}^m)$.
Then the triplet $\left(\left(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in\mathbb{R}_{\geq 0}}\right), X, (\mathbb{P}_x)_{x\in\mathbb{R}^m}\right)$ is called a *Markov process* with transition semigroup P.

An important class of Markov processes are so-called *Feller Markov processes.*

Definition 1.5.4. We say that a Markov process with transition semigroup P is *Feller* if

$$P_t f \in \mathscr{C}_b(\mathbb{R}^m), \quad f \in \mathscr{C}_b(\mathbb{R}^m),$$

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

There is a confusion among authors as some of them prefer to use the space of compactly supported and continuous functions on $\mathbb{R}^m$ instead of the bigger space $\mathscr{C}_b(\mathbb{R}^m)$ leading to a stronger definition in our setting (cf. [Schilling, 1998, Theorem 3.2]). For some results (not needed in our work though) one has to impose a condition

$$P_t f \in \mathscr{C}_b(\mathbb{R}^m), \quad f \in B_b(\mathbb{R}^m),$$

which corresponds to the definition of *strong Feller Markov process.*

A particular class of Markov processes is of our interest in this work. Suppose that for any (deterministic) initial condition $x \in \mathbb{R}^m$ there exists a unique solution to the equation (1.3.3), which we denote as X^x, where we assume that all coefficients are time-homogeneous:

$$f(s,x) = \tilde{f}(x), \quad g(s,x) = \tilde{g}(x), \quad H(s,x,y) = \tilde{H}(x,y)$$

for all $x \in \mathbb{R}^m, y \in \mathbb{R}^n$ for suitable measurable functions $\tilde{f} : \mathbb{R}^m \to \mathbb{R}^m, \tilde{g} : \mathbb{R}^m \to \mathbb{R}^{m\times n}$ and $\tilde{H} : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$. It can be shown (cf. [Applebaum, 2009, Theorem 6.4.6]) that the collection $\{X^x, x \in \mathbb{R}^m\}$ gives rise to a Markov process with some transition kernel P for which we have a relation

$$P_t f(x) = \mathbb{E}\, f(X_t^x), \quad f \in B_b(\mathbb{R}^m), x \in \mathbb{R}^m,$$

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

Therefore, it is convenient to use results from theory of Markov processes to study some properties of stochastic differential equations. One example, important for this work, is the existence of invariant measure for a stochastic differential equation.

Definition 1.5.5. We say that Markov process induced by the equation (1.3.3) with time-homogeneous coefficients possesses an invariant measure if there exists a Borel probability measure μ^* on $\mathbb{R}^m$ and a solution Y to (1.3.3) such that

$$\mathbb{E}\, f(Y_t) = \int_{\mathbb{R}^m} f(x)\mu^*(dx), \quad f \in B_b(\mathbb{R}^m).$$

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

Remark 1.5.2. The notion of invariant measure of a Markov process induced by a stochastic equation is closely related to existence of a *stationary solution*, that is if μ^* is an invariant measure for (1.3.3) then there exists a solution Y to (1.3.3) such that

$$\mathbb{E}\, f(Y_0) = \int_{\mathbb{R}^m} f(x)\mu^*(dx), \quad f \in B_b(\mathbb{R}^m)$$

and the distribution of Y_t is independent of $t \in \mathbb{R}_{\geq 0}$.

Before we state Krylov-Bogolyubov Theorem we need to introduce the following definition.

Definition 1.5.6. A measurable stochastic process $Y = \{Y_t, t \in \mathbb{R}_{\geq 0}\}$ on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with values in $\mathbb{R}^d$ is *bounded in probability in the mean* if

$$\lim_{R\to\infty} \limsup_{t\to\infty} \frac{1}{t} \int_0^t \mathbb{P}\left[|Y_s| \leq R\right] \mathrm{d}s = 1 \tag{1.5.3}$$

The following statement is a particular case of Krylov-Bogolyubov Theorem.

Theorem 1.5.1 (Krylov-Bogolyubov)**.** Markov process defined by (1.3.3) with time-homogeneous coefficients possesses an invariant measure, if both

1. it is Feller,
2. there exists a (deterministic) initial condition for which the unique solution to (1.3.3) is bounded in probability in the mean.

Proof. The proof can be found e.g. in Krylov and Bogolyubov [1937]. □

2. Invariant measures and boundedness in the mean

Let $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\in\mathbb{R}_{\geq 0}}, \mathbb{P})$ be a filtered probability space with the normal filtration $(\mathscr{F}_t)_{t\in\mathbb{R}_{\geq 0}}$ and assume that W is an $(\mathscr{F}_t)$-Wiener process with values in $\mathbb{R}^n$ which is independent to an $(\mathscr{F}_t)$-Poisson random measure N on $\mathbb{R}_+ \times (\mathbb{R}^n \setminus \{0\})$. The Poisson measure N has the intensity measure $\mathrm{d}t\nu(\mathrm{d}y)$, where ν is a Lévy measure on $\mathbb{R}^n \setminus \{0\}$. Let $\tilde{N}$ denote the compensator of N.

We study an equation driven by the pair (W, N). More specifically, assume we are given Borel measurable mappings $f : \mathbb{R}^m \to \mathbb{R}^m$, $g : \mathbb{R}^m \to \mathbb{R}^{m\times n}$, $H, K : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m$ and a constant $c \in \mathbb{R}_{>0}$ and consider the equation

$$\begin{aligned} dX_t = f(X_{t-})\mathrm{d}t + g(X_{t-})dW_t + &\int_{\{|y|<c\}} H(X_{t-}, y)\tilde{N}(\mathrm{d}t, \mathrm{d}y) \\ &+ \int_{\{|y|\geq c\}} K(X_{t-}, y)N(\mathrm{d}t, \mathrm{d}y), \quad t \in \mathbb{R}_{\geq 0}. \end{aligned} \tag{2.0.1}$$

Recall the standard requirements on the coefficients in (2.0.1) that are sufficient for existence of an unique global solution for any $\mathscr{F}_0$-measurable initial condition (1.3.4), (1.3.5) and (1.3.6). In this Section we deal with time-homogeneous equation therefore we restate these conditions in our particular case[1].
Lipschitz condition: There exists $L_1 \in \mathbb{R}_{>0}$ such that

$$\begin{aligned} |f(x) - f(z)|^2 \vee |g(x) - g(z)|^2 \vee \int_{\{|y|<c\}} |H(x, y) - H(z, y)|^2 \nu(\mathrm{d}y) \\ \leq L_1 |x - z|^2, \end{aligned} \tag{LIP}$$

for any $x, z \in \mathbb{R}^m$.
Growth condition: There exists $L_2 \in \mathbb{R}_{>0}$ such that

$$\int_{\{|y|<c\}} |H(x, y)|^2 \nu(\mathrm{d}y) \leq L_2(1 + |x|^2), \tag{GRO}$$

for any $x \in \mathbb{R}^m$.
Continuity condition: We have

$$K(\cdot, y) \in \mathscr{C}(\mathbb{R}^m; \mathbb{R}^m) \tag{CON}$$

for all $y \in \{|y| \geq c\}$.

In the following, we may proceed without (LIP), (GRO), (CON), however, under these assumptions the results can be substantially simplified.

We only assume that for any $x \in \mathbb{R}^m$ the unique solution to (2.0.1) denoted as X^x is given and it defines a time-homogeneous Markov process. It is convenient to assign the Markov process in the usual manner a translation semigroup of linear operators $(S_t, t \in \mathbb{R}_{\geq 0})$ acting on $B_b(\mathbb{R}^m)$ as

$$S_t f(x) = \mathbb{E} f(X_t^x), \quad f \in B_b(\mathbb{R}^m), x \in \mathbb{R}^m, \tag{2.0.2}$$

[1] We also impose the conditions on g directly instead of slightly more general setting in (1.3.4) formulated in terms of gg^T.

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

In the sequel we consider two concepts of stability of the equation (2.0.1). The first one, existence of an invariant measure as in Definition 1.5.5, considers the Markov process induced by (2.0.1), while in the second one can be formulated for any stochastic process as in Definition 1.5.3.

2.1 General Lyapunov criterion

In this section we investigate general criterion for stability the system (2.0.1) in terms of boundedness in probability in the mean.

We will deal with a specific Lyapunov function that takes the following form. For $p \in (0,1)$ denote V_p an arbitrary (but fixed in the sequel) element of $\mathscr{C}^2(\mathbb{R}^m)$ satisfying

$$DV_p \in \mathscr{C}_b(\mathbb{R}^m; \mathbb{R}^m), \quad D^2 V_p \in \mathscr{C}_b(\mathbb{R}^m; \mathbb{R}^{m\times m}) \tag{V1}$$

$$V_p(x) = |x|^p, \quad |x| \geq 1 \tag{V2}$$

$$0 \leq V_p(x) \leq |x|^p, \quad |x| \leq 1. \tag{V3}$$

If follows that the derivatives of V_p take the following form

$$DV_p(x) = p\,|x|^{p-2}\,x \tag{2.1.1}$$

$$D^2 V_p(x) = p(p-2)\,|x|^{p-4}\,xx^T + p\,|x|^{p-2}\,I \tag{2.1.2}$$

for $x \in B_1^c$, where I is the identity in $\mathbb{R}^{m\times m}$.

Using only (V1) the Itô formula may be used to obtain the differential of $V_p(X)$, where X is a solution to (2.0.1).

Proposition 2.1.1 (Itô formula)**.** Let X be a solution to (2.0.1) and $p \in (0,1)$. Then

$$\begin{aligned} dV_p(X_t) = &\Big(\langle f(X_{t-}), DV_p(X_{t-})\rangle + \frac{1}{2}\operatorname{Tr}\left(g(X_{t-})^T D^2 V_p(X_{t-}) g(X_{t-})\right)\Big)\,\mathrm{d}t \\ &+ DV_p(X_{t-})^T g(X_{t-}) dW_t \\ &+ \int_{\{|y|<c\}} V_p(X_{t-} + H(X_{t-},y)) - V_p(X_{t-})\tilde{N}(\mathrm{d}t,\mathrm{d}y) \\ &+ \int_{\{|y|<c\}} \Big(V_p(X_{t-} + H(X_{t-},y)) - V_p(X_{t-}) \\ &\quad - \langle DV_p(X_{t-}), H(X_{t-},y)\rangle\Big)\nu(\mathrm{d}y)\mathrm{d}t \\ &+ \int_{\{|y|\geq c\}} V_p(X_{t-} + K(X_{t-},y)) - V_p(X_{t-}) N(\mathrm{d}t,\mathrm{d}y), \end{aligned} \tag{2.1.3}$$

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

Proof. Follows directly from Theorem 1.2.1. □

The form of Itô formula (2.1.3) motivates us to study the linear operator $\mathcal{L}$ that is given as follows. Denote $Dom(\mathcal{L})$ the linear subspace of $\mathscr{C}^2(\mathbb{R}^m)$ of functions $V \in \mathscr{C}^2(\mathbb{R}^m)$ such that the following prescription

$$\begin{aligned}\mathcal{L}V(x) =& \langle f(x), DV(x)\rangle \\ &+\frac{1}{2}\operatorname{Tr}\left(g(x)^T D^2V(x)g(x)\right) \\ &+\int_{\{|y|<c\}} V(x+H(x,y)) - V(x) - \langle H(x,y), DV(x)\rangle \nu(\mathrm{d}y) \\ &+\int_{\{|y|\geq c\}} V(x+K(x,y)) - V(x)\nu(\mathrm{d}y), \quad x\in\mathbb{R}^m, \end{aligned} \tag{2.1.4}$$

defines an element of $B_b(\mathbb{R}^m)$. Then define $\mathcal{L} : Dom(\mathcal{L}) \to B_b(\mathbb{R}^m)$ by (2.1.4).

In fact, our aim is to rewrite (2.1.3) as

$$\begin{aligned} dV_p(X_t) = \mathcal{L}V_p(X_{t-})\mathrm{d}t &+ DV_p(X_{t-})^T g(X_{t-})dW_t \\ &+\int_{\{|y|<c\}} V_p(X_{t-}+H(X_{t-},y)) - V_p(X_{t-})\tilde{N}(\mathrm{d}t,\mathrm{d}y) \\ &+\int_{\{|y|\geq c\}} V_p(X_{t-}+K(X_{t-},y)) - V_p(X_{t-})N(\mathrm{d}t,\mathrm{d}y) \\ &-\int_{\{|y|\geq c\}} V_p(X_{t-}+K(X_{t-},y)) - V_p(X_{t-})\nu(\mathrm{d}y)\mathrm{d}t, \end{aligned} \tag{2.1.5}$$

for $t\in\mathbb{R}_{\geq 0}$ and any $p\in(0,1)$.

We will prove (2.1.5) under some additional conditions on the coefficients.

Assumption 2.1.1. The following mappings

$$x \mapsto f(x), \tag{2.1.6}$$

$$x \mapsto g(x), \tag{2.1.7}$$

$$x \mapsto \int_{\{|y|<c\}} |H(x,y)|^2 \nu(\mathrm{d}y), \tag{2.1.8}$$

$$x \mapsto \int_{\{|y|\geq c\}} |K(x,y)|^p \nu(\mathrm{d}y) \tag{2.1.9}$$

are locally bounded on $\mathbb{R}^m$ for $p\in(0,1)$.

Note that the following would hold even under weaker assumption of local boundedness of (2.1.9) only for $p\in(0,p^*)$ for some $p^*\in\mathbb{R}_{>0}$.

Lemma 2.1.1. Fix $p\in(0,1)$. Under Assumption 2.1.1 we have that $V_p \in Dom(\mathcal{L})$ and the Itô formula (2.1.5) for V_p holds.

Proof. Let $p\in(0,1)$ be given. To show that the first two terms in (2.1.4) are well defined and locally bounded in x is straightforward. We proceed with the integral terms in more detail.
The compensated term: We use Taylor's reminder in the integral form and (V1) as follows

$$\begin{aligned}&\int_{\{|y|<c\}} |V_p(x+H(x,y)) - V_p(x) - \langle H(x,y), DV_p(x)\rangle|\,\nu(\mathrm{d}y) \\ &\qquad = \int_{\{|y|<c\}} \left|\int_0^1 H(x,y)^T D^2V_p(x+\theta H(x,y))H(x,y)(1-\theta)d\theta\right| \nu(\mathrm{d}y) \\ &\qquad \leq \left|D^2V_p\right|_\infty \int_{\{|y|<c\}} |H(x,y)|^2\,\nu(\mathrm{d}y),\end{aligned}$$

for $x \in \mathbb{R}^m$. By the Assumption Assumption 2.1.1 local boundedness of the compensated term now easily follows.
The uncompensated term: We use (V2), (V3) and estimate

$$
\begin{aligned}
&\int_{\{|y|\geq c\}} |V_p(x+K(x,y)) - V_p(x)|\,\nu(\mathrm{d}y) \\
&\quad\leq \int_{\{|y|\geq c\}} |V_p(x+K(x,y)) - |x+K(x,y)|^p|\,\nu(\mathrm{d}y) \\
&\qquad + \int_{\{|y|\geq c\}} |x+K(x,y)|^p + V_p(x)\nu(\mathrm{d}y) \\
&\quad\leq \int_{\{|y|\geq c\}} 2\nu(\mathrm{d}y) + \int_{\{|y|\geq c\}} |K(x,y)|^p + |x|^p + V_p(x)\nu(\mathrm{d}y) \\
&\quad\leq \nu(\{|y|\geq c\})\,(2+|x|^p + V_p(x)) + \int_{\{|y|\geq c\}} |K(x,y)|^p\,\nu(\mathrm{d}y) \qquad (2.1.10)
\end{aligned}
$$

for $x \in \mathbb{R}^m$ with the last term being locally bounded by the Assumption 2.1.1.

The formula (2.1.5) is valid as it is just a different form of (2.1.3) provided that

$$\int_{\{|y|\geq c\}} V_p(X_{t-} + K(X_{t-},y) - V_p(X_{t-})\nu(\mathrm{d}y)$$

is well defined for every $t \in \mathbb{R}_{\geq 0}$ almost surely, which is the case by (2.1.10) and the almost sure local boundedness of the trajectories of the solution to (2.0.1). ☐

Having Lemma 2.1.1 we will now prove the main criterion for boundedness in probability in the mean. The proof is an adaptation of work of R. Khasminskii (cf. Khasminskii [1980]) established for the special case of diffusion processes.

Theorem 2.1.1. Let the Assumption 2.1.1 hold. Then the solution to (2.0.1) with any deterministic initial condition is bounded in probability in the mean if there exists $p \in (0,1)$ such that there exists $R_0 \in \mathbb{R}_{>0}$ such that for all $R \in (R_0,\infty)$ there exists $A_R \in \mathbb{R}_{>0}$ with

$$A_R \to \infty, \quad R \to \infty \qquad (2.1.11)$$

and

$$\mathcal{L}V_p(x) \leq -A_R, \quad |x| \geq R. \qquad (2.1.12)$$

Proof. Let $x \in \mathbb{R}^m$ and write shortly $X = X^x$ for the unique solution to (2.0.1) with the initial condition x. As X is a global solution, the stopping times

$$\tau_k = \inf\{t \in \mathbb{R}_{\geq 0}, |X_{t-}| > k\},$$

for $k \in \mathbb{N}$, tend to infinity almost surely. Now fix $t \in \mathbb{R}_{\geq 0}$, $k \in \mathbb{N}$. By Lemma 2.1.1 the Itô formula (2.1.5) implies

$$
\begin{aligned}
&V_p(X_{t\wedge\tau_k}) - V(x) = \\
&\quad= \int_0^{t\wedge\tau_k} \mathcal{L}V_p(X_{s-})\mathrm{d}s + \int_0^{t\wedge\tau_k} DV_p(X_{s-})^T g(X_{s-})dW_s \\
&\qquad + \int_0^{t\wedge\tau_k}\int_{\{|y|<c\}} V_p(X_{s-} + H(X_{s-},y)) - V_p(X_{s-})\tilde{N}(\mathrm{d}s,\mathrm{d}y) \\
&\qquad + \int_0^{t\wedge\tau_k}\int_{\{|y|\geq c\}} V_p(X_{s-} + K(X_{s-},y)) - V_p(X_{s-})N(\mathrm{d}s,\mathrm{d}y) \\
&\qquad - \int_0^{t\wedge\tau_k}\int_{\{|y|\geq c\}} V_p(X_{s-} + K(X_{s-},y)) - V_p(X_{s-})\nu(\mathrm{d}y)\mathrm{d}s. \qquad (2.1.13)
\end{aligned}
$$

We compute expectations of all the stochastic integrals in (2.1.13).

We have $DV_p^T g \in L^\infty_{loc}(\mathbb{R}^m)$ and X_{s-} is bounded almost surely on $(0,\tau_k)$, therefore

$$\mathbb{E}\int_0^{t\wedge\tau_k} DV_p(X_{s-})^T g(X_{s-})dW_s = 0.$$

Similarly, the compensated integral is centered,

$$\mathbb{E}\int_0^{t\wedge\tau_k}\int_{\{|y|<c\}} V_p(X_{s-}+H(X_{s-},y)) - V_p(X_{s-})\tilde{N}(\mathrm{d}s,\mathrm{d}y) = 0$$

as from Taylor's expansion and local boundedness of (2.1.8) we have

$$\begin{aligned}
&\mathbb{E}\int_0^{t\wedge\tau_k}\int_{\{|y|<c\}} |V_p(X_{s-}+H(X_{s-},y)) - V_p(X_{s-})|^2\,\nu(\mathrm{d}y)\mathrm{d}s = \\
=&\mathbb{E}\int_0^{t\wedge\tau_k}\int_{\{|y|<c\}} \left|\int_0^1 \langle DV_p(X_{s-}+\theta H(X_{s-},y)), H(X_{s-},y)\rangle(1-\theta)d\theta\right|^2 \nu(\mathrm{d}y)\mathrm{d}s \\
\le&\frac{1}{3}\,|DV_p|_\infty^2\,\mathbb{E}\int_0^{t\wedge\tau_k} |H(X_{s-},y)|^2\,\nu(\mathrm{d}y)\mathrm{d}s < \infty.
\end{aligned}$$

For the uncompensated term by (2.1.10) and local boundedness of (2.1.9) we have

$$\begin{aligned}
\mathbb{E}\int_0^{t\wedge\tau_k}&\int_{\{|y|\ge c\}} |V_p(X_{s-}+K(X_{s-},y)) - V_p(X_{s-})|\,\nu(\mathrm{d}y)\mathrm{d}s \\
&\le\mathbb{E}\int_0^{t\wedge\tau_k} \nu(\{|y|\ge c\})(2+|X_{s-}|^p + V_p(X_{s-}))\mathrm{d}s \\
&\quad+\mathbb{E}\int_0^{t\wedge\tau_k}\int_{\{|y|\ge c\}} |K(X_{s-},y)|^p\,\nu(\mathrm{d}y)\mathrm{d}s < \infty.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}\int_0^{t\wedge\tau_k}&\int_{\{|y|\ge c\}} V_p(X_{s-}+K(X_{s-},y)) - V_p(X_{s-})N(\mathrm{d}s,\mathrm{d}y) \\
&=\mathbb{E}\int_0^{t\wedge\tau_k}\int_{\{|y|\ge c\}} V_p(X_{s-}+K(X_{s-},y)) - V_p(X_{s-})\nu(\mathrm{d}y)\mathrm{d}s.
\end{aligned}$$

Finally, $\mathcal{L}V_p \in L^\infty_{loc}(\mathbb{R}^m)$ so

$$\mathbb{E}\int_0^{t\wedge\tau_k} \mathcal{L}V_p(X_{s-})\mathrm{d}s.$$

is well defined. We have shown that

$$-V(x) \le \mathbb{E}\,V_p(X_{t\wedge\tau_k}) - V(x) = \mathbb{E}\int_0^{t\wedge\tau_k} \mathcal{L}V_p(X_{s-})\mathrm{d}s.$$

For $R \in (R_0,\infty)$ we have from the assumption (2.1.12) and local boundedness of $\mathcal{L}V_p$ that

$$\begin{aligned}
\mathcal{L}V_p(x) \le& - A_R 1_{\{|x|\ge R\}} + \left(\sup_{|x|<R} \mathcal{L}V_p(x)\right) 1_{\{|x|<R\}} \\
\le& - A_R 1_{\{|x|\ge R\}} + \sup_{x\in\mathbb{R}^m} \mathcal{L}V_p(x) < \infty.
\end{aligned}$$

Denote

$$\kappa = \sup_{x \in \mathbb{R}^m} \mathcal{L}V_p(x) < \infty.$$

We shall write

$$-V(x) \leq \mathbb{E} \int_0^{t \wedge \tau_k} -A_R \mathbf{1}_{\{|X_{s-} \geq R|\}} + \kappa \mathrm{d}s.$$

Now taking the limit as $k \to \infty$. We obtain

$$-V(x) \leq -A_R \int_0^t \mathbb{P}\left[|X_{s-}| \geq R\right] \mathrm{d}s + \kappa t$$

with $t \in \mathbb{R}_{\geq 0}$ arbitrary. Finally for $t \geq 1$

$$\frac{1}{t} \int_0^t P\left[|X_{s-}| \geq R\right] \mathrm{d}s \leq \frac{V(x) + \kappa}{A_R} \to 0, \quad R \to \infty$$

by (2.1.11), which already implies (1.5.3). The assertion of the theorem now follows by the equality $X_{s-} = X_s$ almost surely for any $s \in \mathbb{R}_{\geq 0}$. □

2.2 Stabilization by compensated jumps

Throughout this section we assume Assumption 2.1.1 to hold. We investigate stabilization properties of compensated jumps. Thus, for simplicity, we put K=0 in (2.0.1) and obtain the equation:

$$dX_t = f(X_{t-})\mathrm{d}t + g(X_{t-})dW_t + \int_{\{|y|<c\}} H(X_{t-}, y)\tilde{N}(\mathrm{d}t, \mathrm{d}y), \quad t \in \mathbb{R}_{\geq 0}. \quad (2.2.1)$$

The main result of this section is presented in Theorem 2.2.1 and in an important Corollary 2.2.1 where the conditions are simplified under specific growth assumptions. First, we prove some technical formulas.

For our purpose it is useful to denote

$$\mathcal{H}(x) = \{y \in \mathbb{R}^n : |y| < c, x + H(x, y) = 0\}, \quad (2.2.2)$$

for $x \in \mathbb{R}^m$.

Note that by the relations

$$|x|^2 \nu(\mathcal{H}(x)) = \int_{\{|y|<c\} \cap \mathcal{H}(x)} |H(x, y)|^2 \nu(\mathrm{d}y) \leq \int_{\{|y|<c\}} |H(x, y)|^2 \nu(\mathrm{d}y) < \infty$$

for $x \in \mathbb{R}^m$, which are due to Assumption 2.1.1, it follows that

$$\nu(\mathcal{H}(x)) < \infty, \quad x \in \mathbb{R}^m, x \neq 0. \quad (2.2.3)$$

First, we prove a technical Lemma.

Lemma 2.2.1. For $p \in (0,1)$ we have

$$\int_{\{|y|<c\}} V_p(x+H(x,y)) - V_p(x) - \langle H(x,y), DV_p(x)\rangle \nu(\mathrm{d}y)$$
$$\leq p\,|x|^p \left[\int_{\{|y|<c\}\cap\mathcal{H}(x)^c} \log\frac{|x+H(x,y)|}{|x|} - \frac{\langle H(x,y), x\rangle}{|x|^2}\nu(\mathrm{d}y)\right.$$
$$+\frac{p}{2}\int_{\{|y|<c\}\cap\mathcal{H}(x)^c}\left(\log\frac{|x+H(x,y)|}{|x|}\right)^2\nu(\mathrm{d}y)$$
$$\left.-\left(\frac{1}{p}-1\right)\nu(\mathcal{H}(x))\right] < \infty, \tag{2.2.4}$$

for $x \in B_1^c$.

Proof. Take $p \in (0,1)$. Now, fix $x \in B_1^c$ and distinguish two cases.
Step I. If $y \in \mathcal{H}(x)$ then we have

$$V_p(x+H(x,y)) - V_p(x) - \langle H(x,y), DV_p(x)\rangle = |x|^p\,(p-1)\,, \tag{2.2.5}$$

by (2.1.1).

Hence, (2.2.3) yields

$$\int_{\mathcal{H}(x)} V_p(x+H(x,y)) - V_p(x) - \langle H(x,y), DV_p(x)\rangle\nu(\mathrm{d}y) \leq$$
$$\leq p\,|x|^p\,\nu(\mathcal{H}(x)(1-\frac{1}{p}). \tag{2.2.6}$$

Step II. For the second case, when $y \in \{|y|<c\}\cap\mathcal{H}(x)^c$, we remind the particular form of Taylor's expansion

$$\frac{a^p-1}{p} = \log a + \frac{\tilde{p}}{2}(\log a)^2\,, \tag{2.2.7}$$

for $p \in \mathbb{R}_{>0}$ and some $\tilde{p} \in (0,p)$ with $a \in \mathbb{R}_{>0}$ fixed. We now apply (2.2.7) to the case

$$a = \frac{|x+H(x,y)|}{|x|}$$

and use (2.1.1) to compute

$$V_p(x+H(x,y)) - V_p(x) - \langle H(x,y), DV_p(x)\rangle =$$
$$= |x+H(x,y)| - |x|^p - p\,|x|^p\frac{\langle H(x,y), x\rangle}{|x|^2} + (V_p(x+H(x,y)) - |x+H(x,y)|^p)$$
$$= p\,|x|^p\left(\frac{\frac{|x+H(x,y)|}{|x|}-1}{p} - \frac{\langle H(x,y), x\rangle}{|x|^2}\right) + (V_p(x+H(x,y)) - |x+H(x,y)|^p)$$
$$= p\,|x|^p\left(\log\frac{|x+H(x,y)|}{|x|} + \frac{\tilde{p}}{2}\left(\log\frac{|x+H(x,y)|}{|x|}\right)^2 - \frac{\langle H(x,y), x\rangle}{|x|^2}\right) +$$
$$+ (V_p(x+H(x,y)) - |x+H(x,y)|^p)\,.$$

The term

$$(V_p(x + H(x, y)) - |x + H(x, y)|^p)$$

is not positive by the definition of V_p (cf. (V3)) and since $\tilde{p} < p$, we have

$$\begin{aligned} V_p(x + H(x, y)) - V_p(x) - &\langle H(x, y), DV_p(x)\rangle \leq \\ &\leq p\,|x|^p \left(\log \frac{|x + H(x, y)|}{|x|} - \frac{\langle H(x, y), x\rangle}{|x|^2}\right) \\ &+ \frac{p^2}{2}\,|x|^p \left(\log \frac{|x + H(x, y)|}{|x|}\right)^2. \end{aligned} \tag{2.2.8}$$

Both the terms on the right-hand side of (2.2.8) are integrable over $\{|y| < c\} \cap \mathcal{H}(x)$ as follows from the fact that $\log(a) \leq a - 1$ for any $a \in \mathbb{R}_{>0}$:

$$\begin{aligned} 0 \leq \int_{\{|y|<c\}\cap\mathcal{H}(x)} &\left(\log \frac{|x + H(x, y)|}{|x|}\right)^2 \nu(\mathrm{d}y) \leq \\ &\leq \int_{\{|y|<c\}\cap\mathcal{H}(x)} \left(\frac{|H(x, y)|}{|x|}\right)^2 \nu(\mathrm{d}y) < \infty. \end{aligned} \tag{2.2.9}$$

by local boundedness of (2.1.8). Similarly, we estimate from above

$$\begin{aligned} \int_{\{|y|<c\}\cap\mathcal{H}(x)} &\log \frac{|x + H(x, y)|}{|x|} - \frac{\langle H(x, y), x\rangle}{|x|^2} \nu(\mathrm{d}y) \\ =&\frac{1}{2} \int_{\{|y|<c\}\cap\mathcal{H}(x)} \log \frac{|x + H(x, y)|^2}{|x|^2} - 2\frac{\langle H(x, y), x\rangle}{|x|^2} \nu(\mathrm{d}y) \\ \leq&\frac{1}{2} \int_{\{|y|<c\}\cap\mathcal{H}(x)} \frac{|x + H(x, y)|^2 - |x|^2}{|x|^2} - 2\frac{\langle H(x, y), x\rangle}{|x|^2} \nu(\mathrm{d}y) \\ =& \int_{\{|y|<c\}\cap\mathcal{H}(x)} \left(\frac{|H(x, y)|}{|x|}\right)^2 \nu(\mathrm{d}y) < \infty. \end{aligned} \tag{2.2.10}$$

Estimation from below is not needed as the left-hand side of (2.2.8) is integrable. Therefore, by (2.2.8) we have

$$\begin{aligned} \int_{\{|y|<c\}\cap\mathcal{H}(x)} &V_p(x + H(x, y)) - V_p(x) - \langle H(x, y), x\rangle \nu(\mathrm{d}y) \\ \leq& p\,|x|^p \left(\int_{\{|y|<c\}\cap\mathcal{H}(x)} \log \frac{|x + H(x, y)|}{|x|} - \frac{\langle H(x, y), x\rangle}{|x|^2} \nu(\mathrm{d}y)\right. \\ &+ \left.\frac{p}{2} \int_{\{|y|<c\}\cap\mathcal{H}(x)} \left(\log \frac{|x + H(x, y)|}{|x|}\right)^2 \nu(\mathrm{d}y)\right) \end{aligned} \tag{2.2.11}$$

Now by (2.2.6) and (2.2.11) we get the desired inequality in (2.3.3). □

Now we prove the main result of this section. For this purpose, we assume that there exists $b \in \mathbb{R}$ and $\underline{\sigma}, \overline{\sigma} \in \mathbb{R}_{>0}$ and $K > 1$ such that

$$\begin{aligned} &\langle f(x), x\rangle \leq b\,|x|^2, \\ &|g(x)| \leq \overline{\sigma}\,|x|, \quad \frac{\left|g(x)^T x\right|}{|x|^2} \geq \underline{\sigma} \end{aligned} \tag{2.2.12}$$

for $x \in B_K^c$.

Remark 2.2.1. An important example of coefficient $g : \mathbb{R}^m \mapsto \mathbb{R}^{m\times n}$ that satisfies (2.2.12) is when $n = 1$ and

$$g(x) = Gx, \quad x \in \mathbb{R}^m,$$

for some $G = (g_{ij}) \in \mathbb{R}^{m\times m}$ positive-definite. Then we may take $\overline{\sigma} = |G|$ and we have

$$\left|g(x)^T x\right| = \langle Gx, x\rangle \geq \underline{\sigma}\, |x|^2\,, \quad x \in \mathbb{R}^m$$

for some $\underline{\sigma} \in \mathbb{R}_{>0}$.

Theorem 2.2.1. Assume that f, g satisfy (2.2.12). Let $R_0 \in (1,\infty)$ be such that

$$\alpha := \sup_{x\in B^c_{R_0}} \int_{\{|y|<c\}\cap\mathcal{H}(x)^c} \log\frac{|x + H(x, y)|}{|x|} - \frac{\langle H(x, y), x\rangle}{|x|^2}\nu(\mathrm{d}y) < \infty, \qquad (2.2.13)$$

and

$$\sup_{x\in B^c_{R_0}} \int_{\{|y|<c\}\cap\mathcal{H}(x)^c} \left(\log\frac{|x + H(x, y)|}{|x|}\right)^2 \nu(\mathrm{d}y) < \infty, \qquad (2.2.14)$$

where $\mathcal{H}(x)$ is defined in (2.2.2). Then the solution to (2.2.1) with any deterministic initial condition is bounded in probability in the mean if

$$b + \frac{1}{2}\overline{\sigma}^2 - \underline{\sigma}^2 + \alpha < 0. \qquad (2.2.15)$$

Moreover, the condition (2.2.15) need not to be satisfied if $\nu(\mathcal{H}(x)) \in \mathbb{R}_{>0}$ uniformly in $x \in B^c_{R_0}$.

If we assume the growth condition (GRO), then $\alpha \leq L_2 < \infty$, where L_2 is from (GRO) and α is from (2.2.13) . Moreover, the condition (2.2.14) is satisfied. We summarize this claim in the following Corollary.

Corollary 2.2.1. Assume (GRO) and (2.2.12). Then the solution to (2.2.1) with any deterministic initial condition is bounded in probability in the mean if (2.2.15) holds, where

$$\alpha := \sup_{x\in B^c_1} \int_{\{|y|<c\}\cap\mathcal{H}(x)^c} \log\frac{|x + H(x, y)|}{|x|} - \frac{\langle H(x, y), x\rangle}{|x|^2}\nu(\mathrm{d}y). \qquad (2.2.16)$$

Proof. It easily follows by (2.2.9), (2.2.10) and Theorem 2.2.1. □

Proof of Theorem 2.2.1. We have to verify that (2.1.12) holds when A_R satisfies (2.1.11).

Let $p \in (0, 1)$ be fixed. First, observe that due to (2.2.12), we have $K \in (1,\infty)$ such that

$$\begin{aligned}\langle f(x), DV_p(x)\rangle + &\frac{1}{2}\operatorname{Tr}(g(x)^T D^2V_p(x)g(x)) = \\ &= p\,|x|^{p-2}\,\langle f(x), x\rangle + \frac{1}{2}p(p-2)\,|x|^{p-4}\left|g(x)^T x\right|^2 \\ &\quad + \frac{1}{2}p\,|x|^{p-2}\,|g(x)|^2 \\ &\leq p\,|x|^p\left(b + \frac{1}{2}\overline{\sigma}^2 + \frac{1}{2}(p-1)\underline{\sigma}^2\right)\end{aligned} \qquad (2.2.17)$$

for $x \in B_K^c$.

Combining (2.2.17) with (2.3.3) we obtain

$$\begin{aligned}
\mathcal{L}V_p(x) \leq & p\,|x|^p \Bigg[b + \frac{1}{2}\overline{\sigma}^2 + (\frac{1}{2}p-1)\underline{\sigma}^2 \\
& + \int_{\{|y|<c\}\cap\mathcal{H}(x)^c} \log\frac{|x+H(x,y)|}{|x|} - \frac{\langle H(x,y),x\rangle}{|x|^2}\nu(\mathrm{d}y) \\
& + \frac{p}{2}\int_{\{|y|<c\}\cap\mathcal{H}(x)^c} \left(\log\frac{|x+H(x,y)|}{|x|}\right)^2 \nu(\mathrm{d}y) \\
& - \left(\frac{1}{p}-1\right)\nu(\mathcal{H}(x))\Bigg] \\
\leq & p\,|x|^p \Bigg[b + \frac{1}{2}\overline{\sigma}^2 + \frac{1}{2}(p-1)\underline{\sigma}^2 + \alpha \\
& + \frac{p}{2}\sup_{x\in B_{R_0}}\int_{\{|y|<c\}\cap\mathcal{H}(x)^c} \left(\log\frac{|x+H(x,y)|}{|x|}\right)^2 \nu(\mathrm{d}y)) \Bigg],
\end{aligned} \tag{2.2.18}$$

for $x \in B_R^c$, where $R = R_0 \vee K$. In (2.2.18) note that $1/p - 1 > 0$.

Now taking $p \in \mathbb{R}_{>0}$ sufficiently small we get that there exists $\kappa \in \mathbb{R}_{>0}$ such that

$$\mathcal{L}V_p(x) \leq -\kappa\,|x|^p, \tag{2.2.19}$$

for $x \in B_R^c$ if (2.2.15) holds.

Moreover, if $\nu(\mathcal{H}(x)) \in \mathbb{R}_{>0}$ uniformly in $B_{R_0}^c$, taking into account that

$$-\left(\frac{1}{p}-1\right)\nu(\mathcal{H}(x)) \to -\infty, \quad p \to 0+,$$

uniformly in $x \in B_R^c$, (2.2.18) implies (2.2.19) even if (2.2.15) does not hold.

Finally, (2.2.19) already guarantees (2.1.12) with A_R satisfying (2.1.11), which completes the proof. □

We can see that in the condition (2.2.15) the sign of α that comes from (2.2.13) determines if the compensated integral stabilizes the system. The sign of α is determined by the interaction of two terms

$$\log\frac{|x+H(x,y)|}{|x|} \quad \textit{and} \quad -\frac{\langle H(x,y),x\rangle}{|x|^2}, \tag{2.2.20}$$

for $x, y \in \mathbb{R}^n$ fixed. We can interpret the coefficient value $H(x,y)$ as the (vector) jump of the solution from state point x given that the driving noise attaines value of y. Moreover, both the terms in (2.2.20) have always opposite sign. The first one is negative if the jump $H(x,y)$ "aims towards the origin" while the second one, coming from the compensation, is negative only when $|x + H(x,y)| > |x|$. This can be seen if we rewrite it as

$$-\frac{\langle H(x,y),x\rangle}{|x|^2} = -\cos\left(\phi(x,y)\right)\frac{|H(x,y)|}{|x|},$$

where $\phi(x,y)$ is the angle between $H(x,y)$ and x. The following example quantifies this interplay in the case of simple linear equation.

Example 2.2.1. Let $\nu(\{|y|<c\})<\infty$, $m=n=2$ and consider the unique solution to the equation

$$\begin{aligned} dX_t =& \int_{\{|y|<c\}} H(X_{t-},y)\tilde{N}(\mathrm{d}t,\mathrm{d}y), \quad t\in\mathbb{R}_{\geq 0} \\ X_0 =& x_0 \in \mathbb{R}^2, \end{aligned} \tag{2.2.21}$$

where $H(x,y)=qR_\phi x$, $x,y\in\mathbb{R}^2$ for some $q\in\mathbb{R}_{>0}$, and rotation matrix R_ϕ

$$R_\phi = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix},$$

where $\phi\in[0,2\pi)$ and $x_0\neq 0$. In this case we can separate jumps and the compensating drift, so (2.2.21) shall be written as

$$\begin{aligned} dX_t =& -q\nu(\{|y|<c\})R_\phi X_{t-}\mathrm{d}t + qR_\phi X_{t-}dP_t, \quad t\in\mathbb{R}_{\geq 0} \\ X_0 =& \in \mathbb{R}^2, \end{aligned}$$

where $P=N(\cdot,\{|y|<c\})$ is a Poisson process with intensity $\nu(\{|y|<c\})$.

We can use Corollary 2.2.1 to assess boundedness of X in probability in the mean. If $q=1$ then $\mathcal{H}(x)=\{|y|<c\}$ and X is bounded in probability the mean. Rewriting (2.2.16) we see that the same holds if

$$\alpha = \int_{\{|y|<c\}} \log\frac{|x+qR_\phi x|}{|x|} - \frac{\langle qR_\phi x, x\rangle}{|x|^2}\nu(\mathrm{d}y) < 0$$

for $x\in\mathbb{R}^2, x\neq 0$. More specifically,

$$\begin{aligned} \int_{\{|y|<c\}} \log\frac{|x+qR_\phi x|}{|x|} - \frac{\langle qR_\phi x, x\rangle}{|x|^2}\nu(\mathrm{d}y) = \\ = \nu(\{|y|<c\})\left(\frac{1}{2}\log(1+2q\cos\phi+q^2) - q\cos\phi\right), \end{aligned}$$

for $x\in\mathbb{R}^2, x\neq 0$. Therefore, X is bounded in probability in the mean if

$$\log(1+2q\cos\phi+q^2) < 2q\cos\phi. \tag{2.2.22}$$

Let us inspect the condition (2.2.22) in more detail.

Denote by

$$\mathcal{S}=\{(q,\phi)\in\mathbb{R}_{>0}\times[0,2\pi) : \log(1+2q\cos\phi+q^2)<2q\cos\phi\}$$

the set of couples $(q,\phi)\in\mathbb{R}_{>0}\times[0,2\pi)$ for which (2.2.22) holds. We briefly inspect the set $\mathcal{S}$ by considering three distinct cases

Case 1: If $\phi\in[0,\frac{\pi}{2})\cup(\frac{3\pi}{2},2\pi)$ then if $q\in\mathbb{R}_{>0}$ is big enough, we have $(q,\phi)\in\mathcal{S}$. This corresponds to the case when the jumps point in the opposite direction to origin, therefore, the stabilization effect is driven by the compensating drift. The smaller $|\cos\phi|$ is, the smaller $q\in\mathbb{R}_{>0}$ is needed in order to have $(\phi,q)\in\mathcal{S}$ and if $\phi\in[0,\frac{\pi}{4}]\cup[\frac{7\pi}{4},2\pi)$ then $(\phi,q)\in\mathcal{S}$ for any $q\in\mathbb{R}_{>0}$.

Case 2: If $\phi\in[\frac{\pi}{2},\frac{3\pi}{4}]\cup[\frac{5\pi}{4},\frac{3\pi}{2}]$, then for any $q\in\mathbb{R}_{>0}$ we have $(\phi,q)\in\mathcal{S}^c$, where

$$\mathcal{S}^c=\{(q,\phi)\in\mathbb{R}_{>0}\times[0,2\pi) : \log(1+2q\cos\phi+q^2)\geq 2q\cos\phi\}$$

and we do not observe the stabilization property of the compensated integral. This case also covers both the jumps and compensation being orthogonal to the direction to the origin.
Case 3: If $\phi \in (\frac{3\pi}{4}, \frac{5\pi}{4})$, then $(\phi, q) \in \mathcal{S}$ if $q \in \mathbb{R}_{>0}$ is small enough. The larger $|\cos\phi|$ is, the larger $q \in \mathbb{R}_{>0}$ we can take to keep $(\phi, q) \in \mathcal{S}$. For $\phi = \pi$, we can even take $q > 1$; then the solution jumps over the origin and still the stabilization property is obtained. However, $q = \frac{3}{2}$ is already too large and $(\pi, \frac{3}{2}) \in \mathcal{S}^c$, which means that even though the norm of the process after jump decreases, the stabilization property is lost. This is due to the compensation term, which drives the system in the direction opposite to the jumps.

Note that the solution X can be constructed directly using the interlacing procedure. Denote τ_k, $k \in \mathbb{N}$ the arrival times for P. Using the matrix exponentiation, X_t takes the form

$$X_t = (I + R_\phi)^k e^{-\nu(\{|y|<c\})qtR_\phi} x_0,$$

if $t \in [\tau_k, \tau_{k+1})$, where $k \in \mathbb{N}_0$. Moreover, for $p \in \mathbb{R}_{>0}$, we are able to compute the moments explicitly

$$\mathbb{E}\,|X_t|^p = |x_0|^p\, e^{-\lambda t\left(1+pq\cos\phi-\left|I+qR_\phi\right|^p\right)} \tag{2.2.23}$$

for $t \in \mathbb{R}_{\geq 0}$. Using (2.2.23), it can be shown that, if $p \in \mathbb{R}_{>0}$ is sufficiently small, then

$$\mathbb{E}\,|X_t|^p \to 0, \quad t \to \infty$$

if and only if $(\phi, q) \in \mathcal{S}$ and

$$\mathbb{E}\,|X_t|^p \to \infty, \quad t \to \infty$$

if and only if $(\phi, q) \in \mathcal{S}^c$.

In Corollary 2.2.1 we have seen that (2.2.1) is a sufficient condition for boundedness in probability in the mean under some assumptions. This example shows that in the case of the linear equation, (2.2.1) provides also if and only if condition for convergence of p-th moment of the solution for sufficiently small $p \in \mathbb{R}_{>0}$ in the infinite time horizon.

2.3 Stabilization by uncompensated jumps

Throughout this section we assume the Assumption 2.1.1 to hold. We investigate stabilization properties of uncompensated jumps. Thus, we put H=0 in (2.0.1) for simplicity and obtain the following equation:

$$dX_t = f(X_{t-})\mathrm{d}t + g(X_{t-})dW_t + \int_{\{|y|\geq c\}} K(X_{t-}, y)N(\mathrm{d}t, \mathrm{d}y), \quad t \in \mathbb{R}_{\geq 0}. \tag{2.3.1}$$

The main result of this section is presented in Theorem 2.3.1. Now we proceed similarly as in the previous section. Set

$$\mathcal{K}(x) = \{y \in \mathbb{R}^n : |y| \geq c, x + K(x, y) = 0\}, \tag{2.3.2}$$

for $x \in \mathbb{R}^m$. We have the following technical Lemma.

Lemma 2.3.1. For $p \in (0, 1/2)$ we have

$$
\begin{aligned}
\int_{\{|y|<c\}} & V_p(x+K(x,y)) - V_p(x)\nu(\mathrm{d}y) \\
& \leq p\,|x|^p \Bigg[\int_{\{|y|<c\}\cap\mathcal{K}(x)^c} \log \frac{|x+K(x,y)|}{|x|}\nu(\mathrm{d}y) \\
& \qquad + \frac{p}{2} \int_{\{|y|<c\}\cap\mathcal{K}(x)^c} \left(\log \frac{|x+K(x,y)|}{|x|} \right)^2 \nu(\mathrm{d}y) \\
& \qquad - \left(\frac{1}{p} - 1 \right) \nu(\mathcal{K}(x)) \Bigg] < \infty, \qquad (2.3.3)
\end{aligned}
$$

for $x \in B_1^c$.

Proof. To prove the inequality (2.3.3) the same ideas as when proving the similar estimate (2.2.4) for the case of compensated integral can be used. This time we estimate the first integral on the right-hand side of (2.3.3) from above,

$$
\begin{aligned}
\int_{\{|y|\geq c\}\cap\mathcal{K}(x)} \log \frac{|x+K(x,y)|}{|x|}\nu(\mathrm{d}y) &= \frac{1}{p} \int_{\{|y|\geq c\}\cap\mathcal{K}(x)} \log \frac{|x+K(x,y)|^p}{|x|^p}\nu(\mathrm{d}y) \\
&\leq \frac{1}{p} \int_{\{|y|\geq c\}\cap\mathcal{K}(x)} \frac{|x+K(x,y)|^p - |x|^p}{|x|^p}\nu(\mathrm{d}y) \\
&\leq \frac{1}{p} \int_{\{|y|\geq c\}\cap\mathcal{K}(x)} \frac{|K(x,y)|^p}{|x|^p}\nu(\mathrm{d}y) < \infty,
\end{aligned}
$$

where we used the local boundedness of (2.1.9). For the second integral on the right-hand side in (2.3.3)

$$
\begin{aligned}
0 \leq \int_{\{|y|\geq c\}\cap\mathcal{K}(x)} & \left(\log \frac{|x+K(x,y)|}{|x|} \right)^2 \nu(\mathrm{d}y) = \\
&= \frac{1}{p^2} \int_{\{|y|\geq c\}\cap\mathcal{K}(x)} \left(\log \frac{|x+K(x,y)|^p}{|x|^p} \right)^2 \nu(\mathrm{d}y) \\
&\leq \frac{1}{p^2} \int_{\{|y|\geq c\}\cap\mathcal{K}(x)} \left(\frac{|x+K(x,y)|^p - |x|^p}{|x|^p} \right)^2 \nu(\mathrm{d}y) \\
&\leq \frac{1}{p^2} \int_{\{|y|\geq c\}\cap\mathcal{K}(x)} \frac{|K(x,y)|^{2p}}{|x|^{2p}}\nu(\mathrm{d}y) < \infty,
\end{aligned}
$$

again using (2.1.9) (here we need $p \in (0, 1/2)$). □

Having the estimate (2.3.3) at hand we obtain similar result for uncompensated integral as in the compensated case in Theorem 2.2.1.

Theorem 2.3.1. Assume that f, g satisfy (2.2.12). Let $R_0 \in (1, \infty)$ be such that

$$
\beta := \sup_{x\in B_{R_0}^c} \int_{\{|y|\geq c\}\cap\mathcal{K}(x)^c} \log \frac{|x+K(x,y)|}{|x|}\nu(\mathrm{d}y) < \infty \qquad (2.3.4)
$$

and

$$
\sup_{x\in B_{R_0}^c} \int_{\{|y|<c\}\cap\mathcal{K}(x)^c} \left(\log \frac{|x+K(x,y)|}{|x|} \right)^2 \nu(\mathrm{d}y) < \infty, \qquad (2.3.5)
$$

where $\mathcal{K}(x)$ is defined in (2.3.2). Then the solution to (2.3.1) with any deterministic initial condition is bounded in probability in the mean if

$$b + \frac{1}{2}\overline{\sigma}^2 - \underline{\sigma}^2 + \beta < 0. \tag{2.3.6}$$

Moreover, the condition (2.3.6) need not to be satisfied if $\nu(\mathcal{K}(x)) \in \mathbb{R}_{>0}$ uniformly in $x \in B^c_{R_0}$.

Proof. The proof is analogous to the proof of Theorem 2.2.1, we only here use the estimate (2.3.3) in place of (2.2.4). □

We see that the sign of β in (2.3.4) determines if the uncompensated integral stabilizes the system. Unlike in the case of compensated integral in previous section we are able to develop criterion which is determined directly by direction of the jumps. It turns out that in such case ad hoc computations are more efficient than using general result from Theorem 2.3.1. These computations depend heavily on the fact that the intensity of jumps $\nu(\{|y| \geq c\})$ of the uncompensated integral is finite.

Theorem 2.3.2. Assume that f, g satisfy (2.2.12). Furthermore assume that there exist $\gamma \in \mathbb{R}_{>0}$, $\alpha \in [0,1)$, $L \in \mathbb{R}_{>0}$ and $R_0 \in \mathbb{R}_{>0}$ such that

$$|x + K(x,y)| \leq \gamma\, |x|^{1-\alpha} + L \tag{2.3.7}$$

for $x \in B^c_{R_0}$ and $y \in \{|y| \geq c\}$. Then the solution to (2.3.1) is bounded in probability in the mean for any deterministic initial condition if either

- $\alpha \neq 0$,

or

- or (2.3.6) with $\beta = \log\gamma$ holds, i.e. if

$$b + \frac{1}{2}\overline{\sigma}^2 - \underline{\sigma}^2 + \nu(\{|y| \geq c\})\log\gamma < 0. \tag{2.3.8}$$

Proof. We verify that (1.5.3) holds. Taking arbitrary $p \in (0,1)$, $n \in \mathbb{N}, n \geq R_0$ and $x \in B^c_n$, (V2), (V3) and (2.3.7) yield

$$\begin{aligned}
&\int_{\{|y|\geq c\}} V_p(x+K(x,y)) - V_p(x)\nu(\mathrm{d}y) = \\
&\quad= \int_{\{|y|\geq c\}} |x+K(x,y)|^p - |x|^p + \left(V_p(x+K(x,y)) - |x+K(x,y)|^p\right)\nu(\mathrm{d}y) \\
&\quad\leq \int_{\{|y|\geq c\}} |x+K(x,y)|^p - |x|^p\,\nu(\mathrm{d}y) + 2\nu(|y|\geq c) \\
&\quad\leq \int_{\{|y|\geq c\}} \gamma^p\,|x|^{p(1-\alpha)} - |x|^p\,\nu(\mathrm{d}y) + 2\nu(|y|\geq c) \\
&\quad= p\,|x|^p\,\frac{\left(\frac{\gamma}{|x|^\alpha}\right)^p - 1}{p}\nu(|y|\geq c) + 2\nu(|y|\geq c) \\
&\quad\leq p\,|x|^p\,\frac{\left(\frac{\gamma}{n^\alpha}\right)^p - 1}{p}\nu(|y|\geq c) + 2\nu(|y|\geq c) \\
&\quad= p\,|x|^p\left(\log\frac{\gamma}{n^\alpha} + \frac{p}{2}\left(\log\frac{\gamma}{n^\alpha})\right)^2\right) + 2\nu(\{|y|\geq c\})
\end{aligned}$$

Taking into account (2.2.17) and (2.2.12) we have $K \in \mathbb{R}_{>0}$ such that

$$\mathcal{L}V_p(x) \le p\,|x|^p \left(b + \frac{1}{2}\overline{\sigma}^2 + \frac{1}{2}(p-1)\underline{\sigma}^2 + \log\frac{\gamma}{n^\alpha} + \frac{p}{2}\left(\log\frac{\gamma}{n^\alpha}\right)^2\right) + 2\nu(\{|y| \ge c\}) \quad (2.3.9)$$

for $x \in B_R^c$, where $R \ge R_0 \vee K$. For the case $\alpha = 0$ we further simplify (2.3.9) to

$$\mathcal{L}V_p(x) \le p\,|x|^p \left(b + \frac{1}{2}\overline{\sigma}^2 + \frac{1}{2}(p-1)\underline{\sigma}^2 + \log\gamma + \frac{p}{2}(\log\gamma)^2\right) + 2\nu(\{|y| < c\}).$$

If we take $p \in \mathbb{R}_{>0}$ sufficiently small, we see that if (2.3.8) holds, there exists $\kappa \in \mathbb{R}_{>0}$ and $R \in \mathbb{R}_{>0}$ such that

$$\mathcal{L}V_p(x) \le -\kappa\,|x|^p \quad (2.3.10)$$

for $x \in B_R^c$.

On the other hand, for $\alpha \in \mathbb{R}_{>0}$ the inequality (2.3.10) may be shown for some $\tilde{\kappa} \in \mathbb{R}_{>0}$ even without assumption (2.3.8) by taking $n \in \mathbb{N}$ sufficiently large. Indeed, (2.3.9) shall be rewritten as

$$\mathcal{L}V_p(x) \le p\,|x|^p\,(\omega - \alpha\log n) + \nu(\{|y| \ge c\})$$

for $x \in B_R^c$, where $R \ge R_0 \vee K$ and

$$\omega = b + \frac{1}{2}\overline{\sigma}^2 + \frac{1}{2}(p-1)\underline{\sigma}^2 + (1 + \frac{p}{2})\log\gamma.$$

□

Remark 2.3.1. Corollary 2.3.2 tells us that the uncompensated term stabilizes our system if the jumps tend towards the origin and intensity of this stabilization is proportional to the intensity of the jumps. Indeed, the system may remind stable in the sense of boundedness in probability in the mean even for jumps in the direction opposite to the origin. Also, taking $\alpha \ne 0$ in Theorem 2.3.2 we see that if the norm of the process after jump gets small enough, then the system is stabilized even with arbitrarily small intensity of the jumps.

In the case of the linear system Corollary 2.3.2

Example 2.3.1. Let $m = n = 1$, $\nu(\{|y| \ge c\}) \in \mathbb{R}_{>0}$ and in (2.3.1) we put $f(x) = bx$, $g(x) = \sigma x$, $K(x,y) = qx$, $x \in \mathbb{R}, y \in \{|y| \ge c\}$ for some $b, \sigma, q \in \mathbb{R}$, i.e. we deal with the equation

$$\begin{aligned} dX_t &= bX_{t-}\mathrm{d}t + \sigma X_{t-}dW_t + qX_{t-}dP_t, \quad t \in \mathbb{R}_{\ge 0}, \\ X_0 &= x_0 \end{aligned} \quad (2.3.11)$$

for some $x_0 \in \mathbb{R}$ where $P = \int_{\{|y|\ge c\}} N(\cdot, \{|y| \ge c\})$, is a Poisson process with intensity $\nu(\{|y| \ge c\})$. In this case we have

$$|x + K(x,y)| = |x|\,|1+q|$$

for $x \in \mathbb{R}, y \in \{|y| \geq c\}$. Therefore, if $q \neq -1$, we may put $\gamma = |1+q|, \alpha = L = 0$ in (2.3.7) and obtain boundedness in probability in the mean for (2.3.11) provided

$$b - \frac{\sigma^2}{2} + \nu(\{|y| \geq c\}) \log|1+q| < 0 \tag{2.3.12}$$

holds.

Therefore, the uncompensated term stabilizes the considered system when $q \in (-2,-1) \cup (-1,0)$. The case $q = -1$ leads to $\mathcal{K}(x) = \{|y| \geq c\}, x \in \mathbb{R}$, where $\mathcal{K}(x)$ is defined in (2.3.2). Therefore, we may use Theorem 2.3.1 directly and obtain boundedness in probability in the mean regardless the sign in (2.3.12).

Now we present an example that merges results from Sections 3 and 4. We compare the stabilization properties of compensated and uncompensated integrals occuring together and treat the constant $c \in \mathbb{R}_{>0}$ as a parameter of the problem.

Example 2.3.2. Consider the equation (2.0.1) with finite Lévy measure ν. For notational simplicity, set

$$M(x,y) = \begin{cases} H(x,y) & (x,y) \in \mathbb{R}^m \times \{|y| < c\} \\ K(x,y) & (x,y) \in \mathbb{R}^m \times \{|y| \geq c\}. \end{cases} \tag{2.3.13}$$

Therefore, (2.0.1) can be writtern as

$$\begin{aligned} dX_t = f(X_{t-})\mathrm{d}t + g(X_{t-})dW_t + &\int_{\{|y|<c\}} M(X_{t-},y)\tilde{N}(\mathrm{d}t,\mathrm{d}y) \\ &+ \int_{\{|y|\geq c\}} M(X_{t-},y)N(\mathrm{d}t,\mathrm{d}y), \quad t \in \mathbb{R}_{\geq 0}. \end{aligned} \tag{2.3.14}$$

Assume that (2.2.12) holds with f locally finite and in compliance with the assumption (2.3.7) of Theorem 2.3.2 let there exist $\gamma \in \mathbb{R}_{>0}$ such that

$$|x + M(x,y)| \leq \gamma |x|$$

for every $x \in B_1^c$ and $y \in \mathbb{R}^n \setminus \{0\}$. Furthermore, for simplicity assume that the solution X does not jump directly to the origin almost surely, i.e.

$$\nu(\{y \in \mathbb{R}^n \setminus \{0\} : x + M(x,y) = 0\}) = 0$$

for every $x \in B_1^c$.

Then Assumption 2.1.1 is satisfied and (2.3.14) can be rewritten as

$$\begin{aligned} dX_t = \left(f(X_{t-}) - \int_{\{|y|<c\}} M(X_{t-},y)\nu(\mathrm{d}y) \right)\mathrm{d}t + g(X_{t-})dW_t \\ + \int_{\mathbb{R}^n \setminus \{0\}} M(X_{t-},y)N(\mathrm{d}t,\mathrm{d}y) \end{aligned}$$

for $t \in \mathbb{R}_{\geq 0}$. Therefore, we may expect that the parameter $c \in \mathbb{R}_{>0}$ influences the stability properties of our system only through the perturbation of the drift

$$-\int_{\{|y|<c\}} M(X_{t-},y)\nu(\mathrm{d}y).$$

We now combine proofs of Theorems 2.2.1 and 2.3.1 to assess stability in terms of boundedness in probability in the mean and obtain the condition

$$\left(b - \inf_{x \in B_1^c} \int_{\{|y|<c\}} \frac{\langle M(x,y), x \rangle}{|x|^2} \nu(\mathrm{d}y)\right) + \frac{1}{2}\overline{\sigma}^2 - \underline{\sigma}^2 + \nu(\mathbb{R}^n \setminus \{0\}) \log \gamma < 0$$

We can see that stability properties of (2.3.14) may depend on $c \in \mathbb{R}_{>0}$. In the simple case

$$M(x,y) = qx, \quad x \in \mathbb{R}^m, y \in \mathbb{R}^n \setminus \{0\},$$

for some $q \in \mathbb{R}, q \neq -1$, we get condition

$$b - q\nu(\{|y| < c\}) + \frac{1}{2}\overline{\sigma}^2 - \underline{\sigma}^2 + \nu(\mathbb{R}^n \setminus \{0\}) \log |1 + q| < 0,$$

or equivalently

$$\frac{b + \frac{1}{2}\overline{\sigma}^2 - \underline{\sigma}^2}{\nu(\mathbb{R}^n \setminus \{0\})} + \log |1 + q| < qr(c), \tag{2.3.15}$$

where

$$r(c) = \frac{\nu(\{|y| < c\})}{\nu(\mathbb{R}^n \setminus \{0\})} \in [0,1], \quad c \in \mathbb{R}_{>0}.$$

The function $r : (0, \infty) \to [0,1]$ is non-decreasing and by (2.3.15) we may conclude that if $q < 0$, which corresponds to the case when the solution exhibits jumps towards origin, the chance that (2.3.15) is satisfied gets smaller with increasing c. For $q \in \mathbb{R}_{>0}$ we get the opposite behavior.

2.4 Invariant measure

In the final section we formulate the main result of this paper concerning the existence of the invariant measure for the equation (2.0.1). Due to Krylov-Bogolyubov Theorem (cf. Theorem 1.5.1), it easily follows by Theorems 2.2.1 and 2.3.1 .

We treat the compensated and uncompensated term simultaneously using the notation as in (2.3.13) and rewrite (2.0.1) as

$$\begin{aligned} dX_t = f(X_{t-})\mathrm{d}t + g(X_{t-})dW_t + &\int_{\{|y|<c\}} M(X_{t-}, y)\tilde{N}(\mathrm{d}t, \mathrm{d}y) \\ &+ \int_{\{|y|\geq c\}} M(X_{t-}, y)N(\mathrm{d}t, \mathrm{d}y), \quad t \in \mathbb{R}_{\geq 0}. \end{aligned} \tag{2.4.1}$$

To exclude the cases when the jumps in (2.4.1) aim directly into the origin, we again adopt the useful notation

$$\mathcal{M}(x) = \{y \in \mathbb{R}^n : x + M(x, y) = 0\}$$

for $x \in \mathbb{R}^m$.

Theorem 2.4.1 (Invariant measure)**.** Let Assumption 2.1.1 hold and let the equation (2.4.1) define Markov process which is Feller. Moreover, let $b, \underline{\sigma}, \overline{\sigma}, \gamma \in \mathbb{R}, \underline{\sigma}, \overline{\sigma} \in \mathbb{R}_{>0}$ be such that there exists $R_0 \in \mathbb{R}_{>0}$ such that

$$\langle f(x), x\rangle < b\,|x|^2$$

$$|g(x)| < \overline{\sigma}\,|x|\,, \qquad \frac{\left|g(x)^T x\right|}{|x|^2} > \underline{\sigma}$$

$$\int_{\mathcal{M}(x)^c} \left(\log\frac{|x + M(x, y)|}{|x|} - \mathbf{1}_{\{|y|<c\}}(y)\frac{\langle M(x, y), x\rangle}{|x|^2}\right)\nu(\mathrm{d}y) < \gamma,$$

for $x \in B^c_{R_0}$, where $\mathbf{1}_{\{|y|<c\}}$ denotes the indicator function of the set $\{|y| < c\}$, and let

$$\sup_{x\in B^c_{R_0}} \int_{\mathcal{M}(x)^c} \left(\log\frac{|x + M(x, y)|}{|x|}\right)^2 \nu(\mathrm{d}y) <\infty.$$

If

$$b + \frac{1}{2}\overline{\sigma}^2 - \underline{\sigma}^2 + \gamma < 0,$$

then the equation (2.4.1) possesses an invariant measure.

Proof. The statement follows directly from Krylov-Bogolyubov Theorem by Theorem 2.2.1 and Theorem 2.3.1. □

Now recall the standard assumptions (LIP), (GRO) and (CON) (which translate into assumptions on M in an obvious way) under which the existence of Markov process defined by (2.4.1) is guaranteed. It is known (cf. [Applebaum, 2009, Section 6.6 and 6.7]) that this process is Feller if in place of (GRO) the following stronger condition is assumed.

Growth condition II: There exist $H_1 : \mathbb{R}^m \mapsto \mathbb{R}_+, H_2 : \mathbb{R}^n \mapsto \mathbb{R}_+$ such that

$$|H(x,y)| \leq H_1(x)H_2(y), \tag{GRO II}$$

$x \in \mathbb{R}^m, y \in \{|y| < c\}$, H_1 is Lipschitz continuous and $\int_{\{|y|<c\}} H_2(y)^2 \nu(\mathrm{d}y) < \infty$.

Indeed, in (cf. [Applebaum, 2009, Note after Theorem 6.6.3]) it is shown that under the assumptions (LIP), (GRO II) and (CON) the equation (2.0.1) defines Markov process such that the mapping

$$x \mapsto X_t^x, \quad x \in \mathbb{R}^m$$

has an almost surely continuous modification for $t \in \mathbb{R}_{>0}$. This already implies Feller property as in Definition 1.5.4 by the Dominated Convergence Theorem.

Moreover, Theorem 2.4.1 may be simplified as follows.

Theorem 2.4.2 (Invariant measure II)**.** Let Assumption (LIP), (GRO II) and (CON) hold. Moreover, let $b, \underline{\sigma}, \overline{\sigma}, \gamma \in \mathbb{R}$, $\underline{\sigma}, \overline{\sigma} \in \mathbb{R}_{>0}$ be such that there exists $R_0 \in \mathbb{R}_{>0}$ such that

$$\langle f(x), x\rangle < b\,|x|^2$$
$$|g(x)| < \overline{\sigma}\,|x|\,, \qquad \frac{\left|g(x)^T x\right|}{|x|^2} > \underline{\sigma}$$
$$\int_{\mathcal{M}(x)^c} \left(\log \frac{|x + M(x,y)|}{|x|} - \mathbf{1}_{\{|y|<c\}}(y) \frac{\langle M(x,y), x\rangle}{|x|^2} \right) \nu(\mathrm{d}y) < \gamma,$$

for $x \in B^c_{R_0}$.
If

$$b + \frac{1}{2}\overline{\sigma}^2 - \underline{\sigma}^2 + \gamma < 0,$$

then the equation (2.4.1) possesses an invariant measure.

Proof. It follows directly from Krylov-Bogolyubov Theorem by Corollary 2.2.1 and Corollary 2.3.2 since (2.4.1) defines Feller Markov process. □

3. Stochastic Approximation Procedures

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

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

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

$$\begin{aligned} \mathrm{d}X_t = f(t, X_t)\,\mathrm{d}t + g(t, X_t)\,\mathrm{d}W_t + &\int\limits_{\{y\in\mathbb{R}^n;\ |y|<c\}} H(t, X_{t-}, y)\,\tilde{N}(\mathrm{d}t, \mathrm{d}y) \\ + &\int\limits_{\{y\in\mathbb{R}^n;\ |y|\geq c\}} H(t, X_{t-}, y)\,N(\mathrm{d}t, \mathrm{d}y), \qquad t \geq 0, \\ X_0 \sim \mu, & \end{aligned} \tag{3.0.1}$$

for some $c \in \mathbb{R}_{>0}$ and a pair (W, N), where N is a Poisson random measure, $\tilde{N}$ its compensated counterpart, and W is a Wiener process independent of N, see e.g. [Applebaum, 2009, Section 2.3.1]. As (in contrast to the previous sections) we deal with a weak solution to (3.0.1) we recall its definition.

Definition 3.0.1. A triplet $((\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P}), (W, N), X)$ is called a solution to the equation (3.0.1) 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 3.0.1 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 3.0.1. We shall assume that

$$\int_{\{|y|<c\}} |H(t,x,y)|^2 \; \nu(\mathrm{d}y) < \infty \quad \text{for all } (t,x) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^m \tag{3.0.2}$$

and the function

$$(t,x) \longmapsto \int_{\{|y|\geq c\}} |H(t,x,y)| \; \nu(\mathrm{d}y) \tag{3.0.3}$$

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

Now, let us set

$$\mathscr{V} = \Big\{ V \in \mathscr{C}^2(\mathbb{R}^m);\; DV \in \mathscr{C}_b(\mathbb{R}^m;\mathbb{R}^m),\; D^2V \in \mathscr{C}_b(\mathbb{R}^m;\mathbb{R}^{m\times m}) \Big\} \tag{3.0.4}$$

and introduce an operator $\mathscr{L}$ associated with the equation (3.0.1) that will henceforth play a crucial role. For $V \in \mathscr{V}$ we define

$$\begin{aligned} &\mathscr{L}V \colon \mathbb{R}_{\geq 0} \times \mathbb{R}^m \longrightarrow \mathbb{R}, \\ &\quad (t,x) \longmapsto \Big\langle f(t,x), DV(x) \Big\rangle + \frac{1}{2} \operatorname{Tr}\Big(g(t,x)^T D^2V(x) g(t,x) \Big) \\ &\qquad + \int_{\mathbb{R}^n \setminus \{0\}} \Big[V(x + H(t,x,y)) - V(x) - \mathbf{1}_{\{|y|<c\}}(y) \Big\langle H(t,x,y), DV(x) \Big\rangle \Big] \, \nu(\mathrm{d}y). \end{aligned} \tag{3.0.5}$$

Using hypotheses (3.0.2) and (3.0.3) we can check easily that the definition of $\mathscr{L}$ is correct, see analogous considerations in the proof of Proposition 3.0.1.

Remark 3.0.1. a) The assumption (3.0.2) can be omitted if we define $\mathscr{L}V$ on the set $\{(t,x) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^m;$ the right-hand side of (3.0.5) makes sense$\}$ only. It is a direct consequence of the integrability condition in part (v) of Definition 3.0.1. We only adopted (3.0.2) so that the formulation of our main results may be more straightforward.

b) On the other hand, (3.0.3) is important and cannot be dispensed with easily. In a companion paper Maslowski and Týbl [2022] related results on stability of solutions to (3.0.1) are obtained under a weaker hypothesis that

$$(t,x) \longmapsto \int_{\{|y|\geq c\}} |H(t,x,y)|^p \, \nu(\mathrm{d}y) \quad \text{is locally bounded on } \mathbb{R}_{\geq 0} \times \mathbb{R}^m \tag{3.0.6}$$

for some $p \in (0,1)$. The same choice is possible in the present paper. Under (3.0.6) 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 (3.0.3).

Using the operator $\mathscr{L}$ we can state the Itô formula for smooth functions of solutions to (3.0.1) in a suitable form.

Proposition 3.0.1. Assume that $V \in \mathscr{V}$ and X solves (3.0.1), then

$$\begin{aligned} \mathrm{d}V(X_t) = \mathscr{L}V(t,X_t)\mathrm{d}t &+ \Big\langle g(t,X_t)^T DV(X_t), \cdot \Big\rangle \, \mathrm{d}W_t \\ &+ \int_{\{|y|<c\}} \Big[V(X_{t-} + H(t,X_{t-},y)) - V(X_{t-}) \Big] \, \tilde{N}(\mathrm{d}t,\mathrm{d}y) \\ &+ \int_{\{|y|\geq c\}} \Big[V(X_{t-} + H(t,X_{t-},y)) - V(X_{t-}) \Big] \, N(\mathrm{d}t,\mathrm{d}y) \\ &- \int_{\{|y|\geq c\}} \Big[V(X_t + H(t,X_t,y)) - V(X_t) \Big] \, \nu(\mathrm{d}y) \, \mathrm{d}t. \end{aligned} \tag{3.0.7}$$

Proof. By Theorem 1.2.1 we have

$$
\begin{aligned}
dV(X_t) = &\Big(\Big\langle f(t,X_t), DV(X_t)\Big\rangle + \frac{1}{2}\operatorname{Tr}\Big(g(t,X_t)^T D^2V(X_t)g(t,X_t)\Big)\Big)\,\mathrm{d}t \\
&+ \Big\langle g(t,X_t)^T DV(X_t), \cdot\Big\rangle\,\mathrm{d}W_t \\
&+ \int_{\{|y|<c\}} \Big[V(X_{t-} + H(t,X_{t-},y)) - V(X_{t-})\Big]\,\tilde{N}(\mathrm{d}t,\mathrm{d}y) \\
&+ \int_{\{|y|<c\}} \Big[V(X_t + H(t,X_t,y)) - V(X_t) \\
&\qquad - \Big\langle DV(X_t), H(t,X_t,y)\Big\rangle\Big]\,\nu(\mathrm{d}y)\,\mathrm{d}t \\
&+ \int_{\{|y|\geq c\}} \Big[V(X_{t-} + H(t,X_{t-},y)) - V(X_{t-})\Big]\,N(\mathrm{d}t,\mathrm{d}y).
\end{aligned}
\tag{3.0.8}
$$

Now adding and subtracting

$$
\int_0^t \int_{\{|y|\geq c\}} \Big[V(X_s + H(s,X_s,y)) - V(X_s)\Big]\,\nu(\mathrm{d}y)\,\mathrm{d}s, \tag{3.0.9}
$$

to the right-hand side of (3.0.8) we obtain the formula (3.0.7) provided (3.0.9) 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

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

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

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

follows by (3.0.3) since the paths of X are locally bounded. □

3.1 Main results

In this Section, we first state a criterion based on Lyapunov functions for a solution to (3.0.1) to converge to a given point of the state space $\mathbb{R}^m$. The following theorem and its corollary generalize results from Nevel'son and Khas'minskiĭ [1972] to equations driven by Lévy processes.

Theorem 3.1.1. Let the Assumption 3.0.1 be satisfied and let there exist $x_0 \in \mathbb{R}^m$, a measurable function $\varphi\colon \mathbb{R}^m \longrightarrow \mathbb{R}_{\geq 0}$, a function $V \in \mathscr{V}$, and measurable functions $\alpha, \gamma\colon \mathbb{R}_{\geq 0} \longrightarrow \mathbb{R}_{>0}$ such that

(H1) either

$$
\inf_{|x-x_0|\geq\varepsilon} \varphi(x) > 0 \quad \text{for all } \varepsilon > 0 \tag{3.1.1}
$$

or

$$
\lim_{|x|\to\infty} V(x) = +\infty \quad \text{and} \quad \inf_{\varrho\geq|x-x_0|\geq\varepsilon} \varphi(x) > 0 \quad \textit{for all } \varrho > \varepsilon > 0, \tag{3.1.2}
$$

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

$$\inf_{|x-x_0|\geq\varepsilon} V(x) > 0 \tag{3.1.3}$$

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

$$\mathscr{L}V(t,x) \leq -\alpha(t)\varphi(x) + \gamma(t)(1+V(x)) \tag{3.1.4}$$

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

Then any solution $(\Omega, \mathscr{F}, (\mathscr{F}_t), (W,N), X)$ to (3.0.1) satisfies

$$\lim_{t\to\infty} X_t = x_0 \quad \mathbb{P}\text{-a.s.} \tag{3.1.5}$$

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

$$\begin{aligned}
\tau_n^1 &= \inf\{t \geq 0 : |X_t| > n\}, \\
\tau_n^2 &= \inf\Big\{t \geq 0 : \int_0^t |g(s,X_s)|^2 \,\mathrm{d}s > n\Big\}, \\
\tau_n^3 &= \inf\Big\{t \geq 0 : \int_0^t \int_{\{|y|<c\}} |H(s,X_s,y)|^2 \,\nu(\mathrm{d}y)\,\mathrm{d}s > n\Big\}, \\
\tau_n &= \tau_n^1 \wedge \tau_n^2 \wedge \tau_n^3
\end{aligned} \tag{3.1.6}$$

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

$$\mathrm{d}U(t,X_t) = (1+V(X_t))\,\mathrm{d}\xi(t) + \xi(t)\,\mathrm{d}V(X_t), \quad t \in \mathbb{R}_{\geq 0}. \tag{3.1.7}$$

Hence combining (3.0.7) and (3.1.7) we obtain for any $n \in \mathbb{N}$ and $t \in \mathbb{R}_{\geq 0}$ (fixed but arbitrary)

$$\begin{aligned}
U(\tau_n \wedge t, &X_{\tau_n\wedge t}) - U(0,X_0) \\
&= \int_0^{\tau_n\wedge t} \Big[(1+V(X_s))\xi'(s) + \xi(s)\mathscr{L}V(s,X_s)\Big]\,\mathrm{d}s \\
&\quad + \int_0^{\tau_n\wedge t} \xi(s)\Big\langle g(s,X_s)^T DV(X_s), \cdot\Big\rangle\,\mathrm{d}W_s \\
&\quad + \int_0^{\tau_n\wedge t}\int_{\{|y|<c\}} \xi(s)\Big[V(X_{s-} + H(s,X_{s-},y)) - V(X_{s-})\Big]\,\tilde{N}(\mathrm{d}s,\mathrm{d}y) \\
&\quad + \int_0^{\tau_n\wedge t}\int_{\{|y|\geq c\}} \xi(s)\Big[V(X_{s-} + H(s,X_{s-},y)) - V(X_{s-})\Big]\,N(\mathrm{d}s,\mathrm{d}y) \\
&\quad - \int_0^{\tau_n\wedge t}\int_{\{|y|\geq c\}} \xi(s)\Big[V(X_s + H(t,X_s,y)) - V(X_s)\Big]\,\nu(\mathrm{d}y)\,\mathrm{d}s.
\end{aligned} \tag{3.1.8}$$

By the hypothesis (H3) we may estimate

$$\begin{aligned}\int_0^{\tau_n\wedge t}&\Big[(1+V(X_s))\xi'(s)+\xi(s)\mathscr{L}V(s,X_s)\Big]\,\mathrm{d}s\\&=\int_0^{\tau_n\wedge t}\xi(s)\Big\{-\gamma(s)(1+V(X_s))+\mathscr{L}(s,X_s)\Big\}\,\mathrm{d}s\\&\le-\int_0^{\tau_n\wedge t}\xi(s)\alpha(s)\varphi(X_s)\,\mathrm{d}s\\&\le 0\end{aligned}\tag{3.1.9}$$

as α and φ are non-negative. Therefore, from (3.1.8) we get

$$\begin{aligned}U(\tau_n\wedge t,&X_{\tau_n\wedge t})-U(0,X_0)\\&\le\int_0^{\tau_n\wedge t}\xi(s)\Big\langle g(s,X_s)^TDV(X_s),\cdot\Big\rangle\,\mathrm{d}W_s\\&\quad+\int_0^{\tau_n\wedge t}\int_{\{|y|<c\}}\xi(s)\Big[V(X_{s-}+H(s,X_{s-},y))-V(X_{s-})\Big]\,\tilde N(\mathrm{d}s,\mathrm{d}y)\\&\quad+\int_0^{\tau_n\wedge t}\int_{\{|y|\ge c\}}\xi(s)\Big[V(X_{s-}+H(s,X_{s-},y))-V(X_{s-})\Big]\,N(\mathrm{d}s,\mathrm{d}y)\\&\quad-\int_0^{\tau_n\wedge t}\int_{\{|y|\ge c\}}\xi(s)\Big[V(X_s+H(t,X_s,y))-V(X_s)\Big]\,\nu(\mathrm{d}y)\,\mathrm{d}s.\end{aligned}\tag{3.1.10}$$

We aim at showing that the right-hand side of (3.1.10) is a martingale for any $n\in\mathbb{N}$. This having been established we find that

$$\mathbb{E}\Big[U(t\wedge\tau_n,X_{t\wedge\tau_n})-U(0,X_0)\Big]\le 0,$$

so we may apply the Fatou lemma and arrive at

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

for every $t\in\mathbb{R}_{\ge0}$, as $V\in L^1(\mu)$. Using the Fatou lemma for conditional expectations we get in a completely analogous way that $(U(t,X_t),\,t\in\mathbb{R}_{\ge0})$ 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 (3.1.10) separately.

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

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

for all $t\in\mathbb{R}_{\ge0}$ due to the definition of τ_n^2 , so the stochastic integral

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

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 3.0.1 and invoking the definition of τ_n^3 we get

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

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

Finally,

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

for all $t \in \mathbb{R}_{\ge 0}$ owing to (3.0.3). Therefore, by the same argument as in [Ikeda and Watanabe, 1981, 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 [Jacod and Shiryaev, 2003, Theorem II.1.8] we have that

$$
\begin{aligned}
\int_0^{\cdot\wedge\tau_n} & \int_{\{|y|\ge c\}} \xi(s)\Big(V(X_{s-}+H(s,X_{s-},y))-V(X_{s-})\Big)\, N(\mathrm{d}s,\mathrm{d}y) \\
&- \int_0^{\cdot\wedge\tau_n} \int_{\{|y|\ge c\}} \xi(s)\Big(V(X_s+H(s,X_s,y))-V(X_s)\Big)\, \nu(\mathrm{d}y)\,\mathrm{d}s
\end{aligned}
$$

is again a martingale.

Hence the proof that $(U(t,X_t))$ is a supermartingale is completed. Since $U(t,X_t)$ it plainly non-negative 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

$$
\lim_{t\to\infty} V(X_t) = \lim_{t\to\infty} \exp\Big(-\int_t^\infty \gamma(r)\,\mathrm{d}r\Big) U(t,X_t) - 1 = U_\infty - 1 =: V_\infty \tag{3.1.11}
$$

$\mathbb{P}$-almost surely.

Step 2: Now we show that

$$
\liminf_{t\to\infty} \Big|X_t - x_0\Big| = 0 \quad \mathbb{P}\text{-a.s.} \tag{3.1.12}
$$

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

$$
\Big|X_t(\omega) - x_0\Big| \ge \varepsilon
$$

for some $t_0 \in \mathbb{R}_{\geq 0}$ and $\varepsilon > 0$ and all $t \geq t_0$. If (3.1.1) is satisfied then clearly a $\delta > 0$ may be found such that

$$\varphi(X_t(\omega)) \geq \delta \quad \text{for all } t \geq t_0. \tag{3.1.13}$$

If (3.1.2) is satisfied then note that by (3.1.11) we may assume that $V(X_t(\omega))$ converges to a finite limit as $t \to \infty$, so by the first part of (3.1.2) there exists a constant $\zeta = \zeta(\omega)$ such that

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

Hence the second part of (3.1.2) 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, (3.1.13) again holds. Thus we have

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

because $\alpha \in L^1_{\text{loc}}(\mathbb{R}_{\geq 0}) \setminus L^1(\mathbb{R}_{\geq 0})$. Therefore, (3.1.12) is established provided we show that

$$\int_0^{\infty} \alpha(s)\varphi(X_s)\,\mathrm{d}s < \infty \quad \mathbb{P}\text{-a.s.} \tag{3.1.14}$$

As $\xi \geq 1$ we have

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

for all $t \in \mathbb{R}_{\geq 0}$ and $n \in \mathbb{N}$ by (3.1.9). Using (3.1.8) together with the fact that the stochastic integrals in (3.1.8) are centered and $U \geq 0$ we obtain

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

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 (3.1.14) holds true.

Step 3: It remains to show that

$$\lim_{t\to\infty} X_t = x_0 \quad \mathbb{P}\text{-a.s.} \tag{3.1.15}$$

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

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

for some $\varepsilon > 0$ and a sequence $t_n \nearrow \infty$. By the hypothesis (H2) of Theorem 3.1.1, an $\eta > 0$ may be found for which

$$V(X_{t_n}(\omega)) \geq \eta \tag{3.1.16}$$

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

$$\lim_{t \to \infty} V(X_t(\omega)) = V_\infty(\omega) \tag{3.1.17}$$

or

$$\liminf_{t \to \infty} \Big| X_t(\omega) - x_0 \Big| = 0 \tag{3.1.18}$$

does not hold, where V_∞ is defined by (3.1.11). Indeed, (3.1.16) together with (3.1.17) imply that $V_\infty(\omega) \geq \eta$. On the other hand, if (3.1.18) 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 (3.1.17) 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 (3.1.17) and (3.1.18) hold for $\mathbb{P}$-almost all $\omega \in \Omega$, which concludes the proof of Theorem 3.1.1. □

Now we focus on a particular case of the equation (3.0.1) 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 coeffients as wide as possible. Namely, we consider the equation

$$\begin{aligned} \mathrm{d}X_t = \alpha(t)\Big(R(X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t + \int_{\{|y|<c\}} K(X_{t-}, y)\,\tilde{N}(\mathrm{d}t, \mathrm{d}y) \\ + \int_{\{|y|\geq c\}} K(X_{t-}, y)\,N(\mathrm{d}t, \mathrm{d}y)\Big), \quad t \geq 0 \\ X_0 \sim \mu, \end{aligned} \tag{3.1.19}$$

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 (3.0.1). Since the function K is independent of time now, Assumption 3.0.1 takes the following form:

Assumption 3.1.1. We shall assume that

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

and the function

$$\int_{\{|y|\geq c\}} \Big| K(\cdot, y) \Big|\, \nu(\mathrm{d}y)$$

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

Let us state a result which one obtains applying Theorem 3.1.1 to (3.1.19).

Corollary 3.1.1. Let Assumption 3.1.1 be satisfied. Let there exist $x_0 \in \mathbb{R}^m$, a function $V \in \mathscr{V} \cap L^1(\mu)$ with $V(x_0) = 0$ and a measurable function $\varphi \colon \mathbb{R}^m \longrightarrow \mathbb{R}_{\geq 0}$ such that

$$\inf_{\varrho \geq |x-x_0| \geq \varepsilon} \varphi(x) > 0 \quad \textit{for all } \varrho > \varepsilon > 0 \tag{3.1.20}$$

and

$$\lim_{|x| \to \infty} V(x) = +\infty, \qquad \inf_{|x-x_0| \geq \varepsilon} V(x) > 0 \quad \textit{for all } \varepsilon > 0. \tag{3.1.21}$$

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

$$\int_0^\infty \alpha(r)\,\mathrm{d}r = \infty, \qquad \int_0^\infty \alpha^2(r)\,\mathrm{d}r < \infty. \tag{3.1.22}$$

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

$$\Big\langle R(x) + \int_{\{|y| \geq c\}} K(x,y)\,\nu(\mathrm{d}y), DV(x) \Big\rangle \leq -\varphi(x), \tag{3.1.23}$$

$$\operatorname{Tr}\Big(\sigma(t,x)^T D^2 V(x) \sigma(t,x)\Big) \leq K_\sigma \Big(1 + V(x)\Big) \tag{3.1.24}$$

and

$$\int_{\mathbb{R}^n \setminus \{0\}} \Big[V(x + \alpha(t) K(x,y)) - V(x) - \alpha(t) \Big\langle K(x,y), DV(x) \Big\rangle \Big]\,\nu(\mathrm{d}y) \\ \leq \beta(t) \Big(1 + V(x)\Big) \tag{3.1.25}$$

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 (3.1.19) then

$$\lim_{t \to \infty} X_t = x_0 \quad \mathbb{P}\text{-a.s.} \tag{3.1.26}$$

Proof. To see that Corollary 3.1.1 follows immediately from Theorem 3.1.1 it suffices to check that the hypothesis (H3) is satisfied. However, the operator $\mathscr{L}$ associated with (3.1.19) takes the form

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

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 3.1.1. The assumptions of Corollary 3.1.1 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 3.1.1. (a) As in Theorem 3.1.1 we may replace (3.1.20) and (3.1.21) with

$$\inf_{|x-x_0|\geq\varepsilon} \Big(V(x) \wedge \varphi(x)\Big) > 0 \quad \text{for any } \varepsilon > 0. \tag{3.1.27}$$

(b) If the function

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

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

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

we may set

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

then both (3.1.20) and (3.1.23) are satisfied.

If $H = 0$ and $K = 0$ then Theorem 3.1.1 and Corollary 3.1.1 correspond essentially to Nevel'son and Khas'minskiĭ [1972], Theorems 3.8.1 and 4.4.1, respectively.

3.2 Applications

Sufficient conditions for convergence of a solution X of (3.1.19) to a point are given in Corollary 3.1.1 in terms of a Lyapunov function V. Choosing a particular Lyapunov function we get more applicable criteria in terms of the coefficients of (3.1.19). 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 3.2.1. Let $x_0 \in \mathbb{R}^m$ and let us set

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

Obviously, the Fréchet derivatives of V are given by

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

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

Let Assumption 3.1.1 be satisfied and suppose that the coefficients σ and K of (3.1.19) satisfy the linear growth condition: there exists a constant $L \in \mathbb{R}_{\geq 0}$ such that

$$|\sigma(t,x)|^2 + \int_{\mathbb{R}^n \setminus \{0\}} |K(x,y)|^2 \, \nu(\mathrm{d}y) \leq L\Big(1 + |x|^2\Big) \tag{3.2.1}$$

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

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

Since

$$\Big\langle R(x) + \int_{\{|y| \geq c\}} K(x,y) \, \nu(\mathrm{d}y), DV(x) \Big\rangle = \frac{2}{1 + |x - x_0|^2} \mathfrak{k}(x) \tag{3.2.2}$$

for all $x \in \mathbb{R}^m$, (3.1.23) is satisfied with the choice

$$\varphi \colon x \longmapsto -\frac{2\mathfrak{k}(x)}{1 + |x - x_0|^2}. \tag{3.2.3}$$

The function φ defined by (3.2.3) surely satisfies (3.1.20) if $\mathfrak{k}$ is continuous and

$$\mathfrak{k}(x) < 0 \quad \text{for all } x \neq x_0. \tag{3.2.4}$$

If $\mathfrak{k}$ is not continuous, it may be dificult to check (3.1.20) and a more feasible way may be to strengthen (3.2.4) assuming that there exists $\eta > 0$ such that

$$\mathfrak{k}(x) \leq -\eta |x - x_0|^2 \quad \text{for all } x \in \mathbb{R}^m. \tag{3.2.5}$$

In this case we may set

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

obtaining a function that clearly satisfies (3.1.20). We claim that the other hypotheses of Corollary 3.1.1 (in the version of Remark 3.1.1) are also satisfied.

For any $x \in \mathbb{R}^m$ we may compute using (3.2.1)

$$\begin{aligned}
&\operatorname{Tr}\Big(\sigma(t,x)^T D^2 V(x) \sigma(t,x)\Big) \\
&\qquad = \frac{2}{1 + |x - x_0|^2} |\sigma(t,x)|^2 - \frac{4}{\Big(1 + |x - x_0|^2\Big)^2} \Big|\sigma(t,x)^T (x - x_0)\Big|^2 \\
&\qquad \leq \frac{2}{1 + |x - x_0|^2} |\sigma(t,x)|^2 \\
&\qquad \leq 2L \frac{1 + |x|^2}{1 + |x - x_0|^2} \\
&\qquad = 4L\Big(1 + \frac{|x_0|^2}{1 + |x - x_0|^2}\Big) \\
&\qquad \leq 4L\Big(1 + |x_0|^2\Big)\Big(1 + V(x)\Big)
\end{aligned} \tag{3.2.6}$$

and (3.1.24) follows. Finally, we verify that (3.1.25) holds with the choice $\beta = 2\alpha^2 L(1+|x_0|^2)$. Using that $\log(y) \le y-1$ for all $y>0$ plainly and the definition of V we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^n\setminus\{0\}} \Big[V(x+\alpha(t)K(x,y)) - V(x) - \alpha(t)\Big\langle K(x,y), DV(x)\Big\rangle\Big]\,\nu(\mathrm{d}y)\\
&\qquad = \int_{\mathbb{R}^n\setminus\{0\}} \Big[\log\Big(\frac{1+|x+\alpha(t)K(x,y)-x_0|^2}{1+|x-x_0|^2}\Big)\\
&\qquad\qquad - \frac{2\alpha(t)}{1+|x-x_0|^2}\Big\langle K(x,y), x-x_0\Big\rangle\Big]\,\nu(\mathrm{d}y)\\
&\qquad \le \frac{1}{1+|x-x_0|^2}\int_{\mathbb{R}^n\setminus\{0\}} \Big[|x-x_0+\alpha(t)K(x,y)|^2 - |x-x_0|^2\\
&\qquad\qquad - 2\alpha(t)\Big\langle K(x,y), x-x_0\Big\rangle\Big]\,\nu(\mathrm{d}y)\\
&\qquad = \frac{\alpha^2(t)}{1+|x-x_0|^2}\int_{\mathbb{R}^n\setminus\{0\}} |K(x,y)|^2\,\nu(\mathrm{d}y)\\
&\qquad \le \alpha^2(t)L\,\frac{1+|x|^2}{1+|x-x_0|^2}\\
&\qquad \le 2\alpha^2(t)L\Big(1+|x_0|^2\Big)\Big(1+V(x)\Big)
\end{aligned} \tag{3.2.7}$$

for all $t\in\mathbb{R}_{\ge 0}$ and $x\in\mathbb{R}^m$. Note also that Assumption 3.1.1 clearly follows from (3.2.1).

Therefore, whenever $\alpha \in \mathscr{C}(\mathbb{R}_{\ge 0}, \mathbb{R}_{>0})$ obeys (3.1.22) and $((W,N),X)$ is a solution to (3.1.19) then X converges almost surely to x_0 as $t\to\infty$.

Remark 3.2.1. It should be stressed that under the hypotheses of Example 3.2.1 the point $x_0\in\mathbb{R}^m$ the solution of (3.1.19) 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 (3.2.4) reduces to

$$\langle R(x), x-x_0\rangle < 0 \quad \text{for all } x\ne x_0. \tag{3.2.8}$$

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. [Browder, 1963, 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 (3.2.8) holds, then

$$\lim_{t\to\infty} X_t = x_0 \quad \mathbb{P}\text{-almost surely} \tag{3.2.9}$$

for any solution of the equation

$$\mathrm{d}X_t = \alpha(t)\Big(R(X_t)\,\mathrm{d}t + \sigma(t,X_t)\,\mathrm{d}W_t\Big), \quad X_0\sim\mu. \tag{3.2.10}$$

This is a classical result going back to Nevel'son and Khas'minskiĭ [1972].

(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 (3.2.8), and σ and K have at most linear growth then (3.2.9) holds for any solution of

$$\mathrm{d}X_t = \alpha(t)\Big(R(X_t)\,\mathrm{d}t + \sigma(t, X_t)\,\mathrm{d}W_t + \int_{\{|y|<a\}} K(X_{t-}, y)\tilde{N}(\mathrm{d}t, \mathrm{d}y)\Big), \quad X_0 \sim \mu. \tag{3.2.11}$$

Again, x_0 is the unique root of R. Related results, obtained by different methods, may be found in Mel'nikov [1989], Lazrieva et al. [1997].

(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 (3.1.19), that is, the process

$$\int_0^{\cdot} \int_{\{|y|\geq c\}} K(X_{t-}, y)\, N(\mathrm{d}t, \mathrm{d}y) \tag{3.2.12}$$

is not centered in general. Moreover, if we would like to keep the driving Lévy noise in (3.0.1) 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 3.1.1 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 3.2.1 the function $\mathfrak{k}$ is continuous and satisfies (3.2.4) 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 (3.1.19) we must add the assumption $R(x_0) = 0$ if Corollary 3.1.1 is to be applied to stochastic approximation; for equations (3.2.10) and (3.2.11) this is redundant. On the other hand, by choosing K in an approprite way we may obtain (3.2.9) 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 3.2.1,

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

then Corollary 3.1.1 is applicable. In the diffusion case (3.2.10) the mere Lipschitz continuity of R need not be sufficient for the convergence of the stochastic approximation procedure. (Indeed, consider (3.2.10) 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 3.1.1 are satisfied except the hypothesis (3.1.23), R is plainly globally Lipschitz continuous having 0 as its only root, nevertheless, a simple direct calculation shows that $X_t \to \infty$ almost surely 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 (3.2.12) is centered and we see that any solution X to (3.1.19) converges to the unique root of R under the hypothesis that R is a continuous function satisfying (3.2.8) (and σ and K has at most linear growth). This result may be compared with theorems stated in Korostelev [1984] 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 extablished in Korostelev [1984] 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 3.2.1 may be satisfied even if R has multiple roots. The coefficient K then "selects" a root of R which a solution to (3.1.19) converges to. This may happen only if a noncentered uncompensated 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 (3.1.19) no longer converges to some (or all) of its roots. Again, in the diffusion case or for the equation (3.2.11) the situation is completely different, see e.g. [Nevel'son and Khas'minskiĭ, 1972, Chapter 5]. For example, let $m = 1$ and let σ and K satisfy (3.2.1) 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

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

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 equation (3.2.11) 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.

Bibliography

S. Albeverio, L. Di Persio, E. Mastrogiacomo, and B. Smii. A class of Lévy driven SDEs and their explicit invariant measures. *Potential Analysis*, 45:229–259, 2016.

D. Applebaum. *Lévy Processes and Stochastic Calculus.* Cambridge University Press, 2nd edition, 2009.

D. Applebaum. Infinite dimensional Ornstein-Uhlenbeck processes driven by Lévy processes. *Probability Surveys*, 12:33–54, 2015.

D. Applebaum and M. Siakalli. Asymptotic stability of stochastic differential equations driven by Lévy noise. *Journal of Applied Probability*, 46:1116–1129, 2009.

D. Applebaum and M. Siakalli. Stochastic stabilization of dynamical systems using Lévy noise. *Stochastics and Dynamics*, 10:509–527, 2010.

C. Bhan, P. Chakraborty, and V. S. Mandrekar. Invariant measure and stability of the solution of a stochastic differential equation driven by a jump Lévy process. *International Journal of Contemporary Mathematical Sciences*, 7:33–52, 2012.

S. Bhatnagar, H. L. Prasad, and L. A. Prashanth. Stochastic recursive algorithms for optimization. In *Lecture Notes in Control Inform. Sci.*, volume 434. Springer, London, 2013.

V. S. Borkar. *Stochastic Approximation: A Dynamical Systems Viewpoint.* Cambridge University Press, Cambridge, 2008.

F. E. Browder. Nonlinear elliptic boundary value problems. *Bulletin of the American Mathematical Society*, 69:862–874, 1963.

H.-F. Chen. Continuous-time stochastic approximation: Convergence and asymptotic efficiency. *Stochastics and Stochastic Reports*, 51:217–239, 1994.

H.-F. Chen. *Stochastic Approximation and its Applications.* Kluwer, Dordrecht, 2002.

S.N. Cohen and R.J. Elliott. *Stochastic Calculus and Applications*, volume 2. Springer, 2015.

J.Z. Cypkin and Z.J. Nikolić. *Foundations of the Theory of Learning Systems (Russian).* Nauka, Moscow, 1970. English translation: Cypkin, J.Z. and Nikolić, Z.J. (1973). Foundations of the Theory of Learning Systems. Academic Press, New York.

M. Driml and J. Nedoma. On the Kiefer-Wolfowitz approximation method. *Časopis Pro Pěstování Matematiky*, pages 400–407, 1951.

E.B. Dynkin. *Markov Processes.* Springer, Berlin, 1965.

I.I. Gikhman and A.V. Skorokhod. *Stochastic Differential Equations.* Springer, Berlin, 1972.

M. Grigoriu. Lyapunov exponents for nonlinear systems with Poisson white noise. *Physics Letters A*, 217:258–262, 1996.

J. Gwinner, B. Jadamba, A. A. Khan, and F. Raciti. *Uncertainty Quantification in Variational Inequalities: Theory, Numerics, and Applications.* CRC Press, Boca Raton, 2022.

N. Ikeda and S. Watanabe. *Stochastic Differential Equations and Diffusion Processes.* North-Holland, Amsterdam, 1981.

J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes.* Springer, Berlin, 2nd edition, 2003.

Z. J. Jurek. An integral representation of operator-self decomposable random variables. *Bulletion de l'Académie Polonaise des Sciences*, 30, 1982.

I Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*, volume 113. Springer Science & Business Media, 1991.

R. Z. Khasminskii. *Stochastic Stability of Differential Equations.* Sojtjoff and Noordhoff, 1980.

S. V. Komarov and T. P. Krasulina. One-sided convergence of continuous processes of stochastic approximation. *Journal of Mathematical Sciences (N.Y.)*, 93:379–384, 1999.

A. P. Korostelev. *Stochastic Recurrent Procedures: Local Properties (Russian).* Nauka, Moscow, 1984.

N. Krylov and N. Bogolyubov. La théorie générale de la mesure dans son application à l'étude des systèmes dynamiques de la mécanique non linéaire. *Annals of Mathematics*, pages 65–113, 1937.

U. Kumar and M. Riedle. Invariant measure for the stochastic Cauchy problem driven by a cylindrical Lévy process. *Journal of Mathematical Analysis and Applications*, 493, 2021.

H. J. Kushner and D. S. Clark. *Stochastic Approximation Methods for Constrained and Unconstrained Systems.* Springer, New York, 1978.

H. J. Kushner and G. G. Yin. *Stochastic Approximation and Recursive Algorithms and Applications.* Springer, New York, 2nd edition, 2003.

N. Lazrieva and T. Toronjadze. The Robbins-Monro type stochastic differential equations. III. Polyak's averaging. *Stochastics*, 82:165–188, 2010.

N. Lazrieva, T. Sharia, and T. Toronjadze. The Robbins-Monro type stochastic differential equations. I. convergence of solutions. *Stochastics and Stochastic Reports*, 61:67–87, 1997.

N. Lazrieva, T. Sharia, and T. Toronjadze. The Robbins-Monro type stochastic differential equations. II. asymptotic behaviour of solutions. *Stochastics and Stochastic Reports*, 75:153–180, 2003.

N. Lazrieva, T. Sharia, and T. Toronjadze. Semimartingale stochastic approximation procedure and recursive estimation. *Journal of Mathematical Sciences (N.Y.)*, 153:211–261, 2008.

G. Leha, B. Maslowski, and G. Ritter. Stability of solutions to semilinear stochastic evolution equations. *Stochastic Analysis and Applications*, 17:1009–1051, 1999.

C. W. Li, Z. Dong, and R. Situ. Almost sure stability of linear stochastic differential equations with jumps. *Probability Theory and Related Fields*, 123:121–155, 2002.

A. Lyapunov. Problème général de la stabilité du mouvement (French). In *Annales de la Faculté des Sciences de Toulouse: Mathématiques*, volume 9, pages 203–474, 1907.

X. Mao. *Stochastic Differential Equations and Applications*. Elsevier, 2nd edition, 2008.

X. Mao and A. E. Rodkina. Exponential stability of stochastic differential equations driven by discontinuous semimartingales. *Stochastics: An International Journal of Probability and Stochastic Processes*, 55:207–224, 1995.

B. Maslowski and O. Týbl. Invariant measures and boundedness in the mean for stochastic equations driven by lévy noise. *Stochastics and Dynamics*, 22:Paper No. 2240019, 25 pp., 2022.

A. V. Mel'nikov. Stochastic approximation procedures for semimartingales (russian). *Statistics and Control of Random Processes*, pages 147–256, 1989.

E. Nane and Y. Ni. Stability of the solution of stochastic differential equation driven by time-changed Lévy noise. *Proceedings of the American Mathematical Society*, 145:3085–3104, 2017.

M. B. Nevel'son and R. Z. Khas'minskiĭ. *Stochastic Approximation and Recurrent Estimation (Russian)*. Nauka, Moscow, 1972. English translation: Nevel'son, M. B., Has'minskiĭ, R. Z. (1973). Stochastic Approximation and Recursive Estimation. American Mathematical Society, Providence.

G. Pflug. Stetige stochastische Approximation. *Metrika*, 26:139–150, 1979.

P.E. Protter. *Stochastic Differential Equations*. Springer, 2005.

H. Robbins and S. Monro. A stochastic approximation method. *The Annals of Mathematical Statistics*, pages 400–407, 1951.

J. Rosinski and W.A. Woyczynski. On Itô stochastic integration with respect to p-stable motion: inner clock, integrability of sample paths, double and multiple integrals. *The Annals of Probability*, pages 271–286, 1986.

D.J. Sakrison. A continuous Kiefer-Wolfowitz procedure for random processes. *The Annals of Mathematical Statistics*, 35(2):590–599, 1964.

K.-I. Sato. *Lévy Processes and Infinitely Divisible Distributions.* Cambridge university press, 1st edition, 1999.

R.L. Schilling. Conservativeness and extensions of Feller semigroups. *Positivity*, 2(3):239–256, 1998.

J. Seidler and O. Týbl. Stochastic approximation procedures for Lévy-driven SDEs. *Journal of Optimization Theory and Applications*, 197(2):817–837, 2023.

X. Zhang, Y. Xu, B. Schmalfuß, and B. Pei. Random attractors for stochastic differential equations driven by two-sided Lévy processes. *Stochastic Analysis and Applications*, 37:1028–1041, 2019.

List of abbreviations

Measure theory	
A^c	complement of the set A
$\mathscr{B}(X)$	the Borel σ-algebra over a topological space X
$\mathscr{A} \otimes \mathscr{B}$	product of σ-algebras $\mathscr{A}, \mathscr{B}$
$\mu * \nu$	convolution of probability measures μ, ν
$\mathbf{1}_A$	indicator function of a set A
$\{\lvert y\rvert < c\}$	$\{y \in \mathbb{R}^n : \lvert y\rvert \leq c\}$ for $c > 0$
$\{\lvert y\rvert \geq c\}$	$\{y \in \mathbb{R}^n : \lvert y\rvert > c\}$ for $c > 0$
Euclidean and other spaces	
A^T	transposition of a matrix A
$\mathrm{Tr}\, A$	trace of a matrix X
$\lvert\cdot\rvert$	the Euclidean norm (or the Frobenius norm in the case of matrices)
$\langle \cdot, \cdot \rangle$	the scalar product in the Euclidean space
$\mathbb{R}_{\geq 0}$	$\{t \in \mathbb{R} : t \geq 0\}$
$\mathbb{R}_{>0}$	$\{t \in \mathbb{R} : t > 0\}$
$\mathscr{C}(X, Y)$	the space of continuous functions $g : X \mapsto Y$ with the supremal norm
$\lvert f\rvert_\infty$	the supremal norm of a bounded function f
$L^p(\Omega), L^p(\mu)$	the Lebesgue spaces of equivalence classes of integrable functions
$L^p_{loc}(\Omega), L^p_{loc}(\mu)$	the Lebesgue spaces of equivalence classes of locally integrable functions
$B_b(X)$	space of bounded measurable functions on X
DV, D^2V	the first (resp. the second) Fréchet derivative
Probability theory	
$X \sim \mu$	the random variable X has a distribution μ
$\mathbb{E}$	expectation of a random variable
$\mathbb{VAR}$	variance of a random variable
a.s.	equality $\mathbb{P}$-almost surely
$\mathcal{L}$	generator of a Markov semigroup associated with an SDE

List of publications

Maslowski, B. and Ondřej Týbl. *Invariant measures and boundedness in the mean for stochastic equations driven by Lévy noise.* Stochastics and Dynamics, 22.03:2240019, 2022.

Existence of invariant measures and average stability in the mean are studied for stochastic differential equations driven by Lévy process. In particular, some natural conditions are found that verify stabilization of the equation (in the sense of the existence of invariant measures) by jump noise terms. These conditions are verified in several examples.

Seidler, J. and Ondřej Týbl. *Stochastic Approximation Procedures for Lévy-Driven SDEs.* Journal of Optimization Theory and Applications, 197.2:817-837, 2023.

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.