



# Lévy Langevin Monte Carlo

David Oechsler<sup>1,2</sup>

Received: 26 February 2023 / Accepted: 17 October 2023 / Published online: 10 November 2023  
© The Author(s) 2023

## Abstract

Analogously to the well-known Langevin Monte Carlo method, in this article we provide a method to sample from a target distribution  $\pi$  by simulating a solution of a stochastic differential equation. Hereby, the stochastic differential equation is driven by a general Lévy process which—unlike the case of Langevin Monte Carlo—allows for non-smooth targets. Our method will be fully explored in the particular setting of target distributions supported on the half-line  $(0, \infty)$  and a compound Poisson driving noise. Several illustrative examples conclude the article.

**Keywords** Langevin Monte Carlo · Lévy processes · Stochastic differential equations · Invariant distributions · Limiting distributions

**Mathematics Subject Classification** Primary 60G51 · 60H10 · 60G10; Secondary 65C05

## 1 Introduction

Monte Carlo methods based on stationary Markov processes appear frequently in fields such as statistics, computer simulation and machine learning, and they have a variety of applications, for example in physics and biology, cf. (Bardenet et al. 2017; Brooks et al. 2011; Kendall et al. 2005; Şimşekli 2017; Welling and Teh 2011). These methods have in common that in order to sample from a target distribution  $\pi$  one considers sample paths of certain Markov processes to approximate  $\pi$ .

*Langevin Monte Carlo* (LMC) is one of these methods and it originates from statistical physics. It applies to absolutely continuous target distributions  $\pi(dx) = \pi(x)dx$  with smooth density functions  $\pi : \mathbb{R}^d \rightarrow \mathbb{R}_+$ , and its associated process  $(X_t)_{t \geq 0}$  is the so-called *Langevin diffusion*, that is a strong solution of the stochastic differential equation (SDE)

$$dX_t = \frac{\nabla \pi(X_t)}{\pi(X_t)} dt + \sqrt{2} dB_t, \quad (1.1)$$

where  $(B_t)_{t \geq 0}$  is a standard Brownian motion on  $\mathbb{R}^d$ , and  $\nabla \pi$  denotes the gradient of  $\pi$ . For LMC to produce samples from  $\pi$  it is required that  $(X_t)_{t \geq 0}$  is a unique strong solution for (1.1) and  $\pi$  is an invariant distribution for  $(X_t)_{t \geq 0}$ , that is

$$\int_{\mathbb{R}} \mathbb{P}^x(X_t \in B) \pi(dx) = \pi(B) \quad \text{for all } t \geq 0, B \in \mathcal{B}(\mathbb{R}). \quad (1.2)$$

However, for this to be the case it is only natural that assumptions must be made regarding  $\pi$ , e.g. that  $\nabla \pi$  exists in a suitable sense. Moreover, to sample from  $\pi$  using  $(X_t)_{t \geq 0}$  it is essential that  $(X_t)_{t \geq 0}$  converges to  $\pi$  in a suitable sense from any starting point  $x \in \text{supp } \pi$ . As solutions  $(X_t)_{t \geq 0}$  of (1.1) are almost surely continuous,  $\text{supp } \pi$  is necessarily connected for convergence to be even possible. For more on LMC see Brooks et al. (2011) or Roberts and Tweedie (1996).

Due to the constraints of LMC it is reasonable to ask whether one could construct similar methods by replacing the Brownian motion with a more general process. In this article we consider Lévy processes as driving noises. In particular, we are interested in the following question:

*Given a distribution  $\pi$  and a Lévy process  $(L_t)_{t \geq 0}$ , can we choose a drift coefficient  $\phi$  such that we can sample from  $\pi$  by simulation of a solution  $(X_t)_{t \geq 0}$  of*

$$dX_t = \phi(X_t)dt + dL_t? \quad (1.3)$$

✉ David Oechsler  
david.c.oechsler@gmail.com

<sup>1</sup> Institut für Mathematische Stochastik, Technische Universität Dresden, Helmholtzstraße 10, 01069 Dresden, Germany

<sup>2</sup> Center of Scalable Data Analytics and Artificial Intelligence (ScDS.AI), Helmholtzstraße 10, Dresden, Leipzig, Germany

To thoroughly answer this question it is essential to distinguish between the notions *infinitesimally invariant distribution*, *invariant distribution*, and *limiting distribution*. These, and some more introductory notions and well-known facts can be found in Sect. 2 of this article. After that, in Sect. 3, we investigate under which conditions a drift coefficient  $\phi$  exists such that  $\pi$  is infinitesimally invariant for  $(X_t)_{t \geq 0}$ . Clearly, there are cases for which this is not the case, think of discrete distributions, or distributions on a half-space while jumps can occur in all directions. Hence, a general answer can only exist under certain assumptions on the regularity of  $\pi$  and the compatibility of  $\pi$  and  $(L_t)_{t \geq 0}$ .

In the same section, we then find a particular set of conditions under which  $\pi$  is invariant and limiting for  $(X_t)_{t \geq 0}$ . Various examples subsequently illustrate our results. Afterwards, in Sect. 4, we present the more technical aspects of the proofs, followed in Sect. 5 by a list of possible extensions with comments on the difficulties they might pose.

Methodologically we rely on the results in Behme et al. (2022) on invariant measures of Lévy-type processes, the Foster–Lyapunov methods originating in a series of articles by Meyn and Tweedie (1992, 1993a, b), and standard techniques from the theory of ordinary differential equations.

There are various cases in the literature for which SDEs of the form (1.3) are considered with Lévy processes as driving noises. In Nguyen et al. (2019) and Şimşekli (2017) a *fractional Langevin Monte Carlo* (fLMC) method is introduced for which  $(L_t)_{t \geq 0}$  is an  $\alpha$ -stable process, and in Eliazar and Klafter (2003), several examples are produced for the case when the driving noise is a pure jump Lévy process. However, both of these studies are rather focused on practical aspects, disregarding some of the theoretical foundations, the details of which will be discussed in Remark 3.3.

Other related articles include (Priola et al. 2012; Xu and Zegarliński 2009) and (Huang et al. 2021; Zhang and Zhang 2023) in which the ergodicity of general  $\alpha$ -stable driven systems is considered. Moreover, the latter two are also concerned with *heavy-tailed sampling* by means of  $\alpha$ -stable driven SDEs, similar to fLMC in case of Zhang and Zhang (2023) and analogue to fLMC for (Huang et al. 2021). Applying their obtained general results, both articles manage to close the aforementioned gaps in the proofs of Nguyen et al. (2019) and Şimşekli (2017). Methodically, Huang et al. (2021), Priola et al. (2012), Xu and Zegarliński (2009), Zhang and Zhang (2023) are related to the present article in the sense that, among other techniques like Doeblin-coupling and Prokhorov’s theorem, Foster–Lyapunov methods are employed in order to show ergodicity of the process in question. The main difference compared to our approach, however, lies in the fact that these articles rely on proving the strong Feller property, which we specifically intend to avoid, see Sect. 3.2.

## 2 Preliminaries

Throughout this paper we denote by  $L^p(\mathbb{R})$  and  $W^{k,p}(\mathbb{R})$  the classic Lebesgue and Sobolev spaces, and by  $\mathcal{C}_0(\mathbb{R})$  the space of continuous functions vanishing at infinity, i.e. functions  $f \in \mathcal{C}(\mathbb{R})$  such that for all  $\varepsilon > 0$  there exists a compact set  $K \subset \mathbb{R}$  such that for all  $x \in \mathbb{R} \setminus K$  it holds  $|f(x)| < \varepsilon$ .

Let  $U \subseteq \mathbb{R}$  be an open set. As usual,  $\mathcal{C}_c^\infty(U)$  denotes the space of test functions, i.e. smooth functions  $f$  with compact support  $\text{supp } f \subset U$ . Linear functionals  $T : \mathcal{C}_c^\infty(U) \rightarrow \mathbb{R}$  that are continuous w.r.t. uniform convergence on compact subsets of all derivatives are called *Schwartz distributions*. The *distributional derivative* of a Schwartz distribution  $T$  is defined by  $T' : \mathcal{C}_c^\infty(U) \rightarrow \mathbb{R}, f \mapsto T(f')$ . If there exists  $N \in \mathbb{N}_0$  such that for all compact sets  $K \subset U$  there exists  $c > 0$  such that

$$|T(f)| \leq c \max\{\|f^{(n)}\|_\infty : n \leq N\}$$

for all  $f \in \mathcal{C}_c^\infty(U)$  with  $\text{supp } f \subset K$ , then the smallest such  $N$  is called the *order* of  $T$ . If no such  $N$  exists, the order of  $T$  is set to  $\infty$ .

**Markov processes and generators** Let  $(X_t)_{t \geq 0}$  be a Markov process in  $\mathbb{R}$  on the probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ . We denote

$$\mathbb{P}^x(\cdot) := \mathbb{P}(\cdot | X_0 = x) \quad \text{and} \quad \mathbb{E}^x[\cdot] := \mathbb{E}[\cdot | X_0 = x]$$

for all  $x \in \mathbb{R}$ . The *pointwise generator* of  $(X_t)_{t \geq 0}$  is the pair  $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$  defined by

$$\mathcal{A}f(x) := \lim_{t \downarrow 0} \frac{\mathbb{E}^x f(X_t) - f(x)}{t}, \quad x \in \mathbb{R}, \quad f \in \mathcal{D}(\mathcal{A})$$

where

$$\mathcal{D}(\mathcal{A}) := \left\{ f \in \mathcal{C}_0(\mathbb{R}) : \lim_{t \downarrow 0} \frac{\mathbb{E}^x f(X_t) - f(x)}{t} \text{ exists for all } x \in \mathbb{R} \right\}.$$

Further, denote by  $\mathcal{D}(\mathcal{G})$  the set of all functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  for which there exists a measurable function  $g : \mathbb{R} \rightarrow \mathbb{R}$  such that for all  $x \in \mathbb{R}$  and  $t > 0$  it holds

$$\mathbb{E}^x f(X_t) = f(x) + \mathbb{E}^x \left[ \int_0^t g(X_s) ds \right]$$

and

$$\int_0^t \mathbb{E}^x [|g(X_s)|] ds < \infty.$$

Setting  $\mathcal{G}f = g$  the pair  $(\mathcal{G}, \mathcal{D}(\mathcal{G}))$  is called the *extended generator* of  $(X_t)_{t \geq 0}$ .

These two types of generator are linked in the sense that both are extensions of the more commonly used *infinitesimal* or *strong generator* of a Markov process, cf. Kühn (2018), whose domain is given by

$$\mathcal{D} := \left\{ f \in \mathcal{C}_0(\mathbb{R}) : \lim_{t \downarrow 0} \left\| \frac{\mathbb{E} f(X_t) - f}{t} - g \right\|_\infty \right. \\ \left. \text{for some } g \in \mathcal{C}_0(\mathbb{R}) \right\}.$$

In the following, while the pointwise generator facilitates the investigation of the invariant measures of  $(X_t)_{t \geq 0}$ , we need the extended generator to show ergodicity of  $(X_t)_{t \geq 0}$ . On the one hand, it is easier to show whether a function  $f$  is contained in  $\mathcal{D}(\mathcal{A})$  than in  $\mathcal{D}$ . On the other hand, the Foster–Lyapunov methods we employ rely on the application of the extended generator to so called *norm-like* functions (see further below in Sect. 4) which are, by definition, not in  $\mathcal{C}_0(\mathbb{R})$ .

**Lévy processes** A (one-dimensional) Lévy process  $(L_t)_{t \geq 0}$  is a Markov process in  $\mathbb{R}$  with stationary and independent increments with characteristic exponent  $\varphi(\beta) := \ln \mathbb{E}[e^{i\beta L_1}]$  given by

$$\varphi(\beta) = i\gamma\beta - \frac{1}{2}\sigma^2\beta^2 + \int_{\mathbb{R}} (e^{i\beta z} - 1) \mu(dz) \\ + \int_{\mathbb{R}^d} (e^{i\beta z} - 1 - i\beta z \mathbb{1}_{\{|z| < 1\}}) \rho(dz).$$

Here,  $\gamma \in \mathbb{R}$  is the *location parameter*,  $\sigma^2 \geq 0$  is the *Gaussian parameter*, and  $\mu$  and  $\rho$  are two measures on  $\mathbb{R}$  such that  $\mu\{0\} = \nu\{0\} = 0$  and  $\int_{\mathbb{R}} (1 \wedge |z|) \mu(dz) < \infty$  and  $\int_{\mathbb{R}} (|z| \wedge |z|^2) \rho(dz) < \infty$ , respectively. The measure  $\Pi = \mu + \rho$  is called the *jump measure*. The triplet  $(\gamma, \sigma^2, \Pi) = (\gamma, \sigma^2, \mu + \rho)$  is called the *characteristic triplet* of  $(L_t)_{t \geq 0}$ . Note that the decomposition of  $\Pi$  into  $\mu$  and  $\rho$  is not unique.

Further, denote by

$$\bar{\mu}(x) = \begin{cases} \mu(x, \infty), & x > 0, \\ \mu(-\infty, x), & x < 0, \\ 0, & x = 0, \end{cases}$$

the *integrated tail* of  $\mu$ , and by  $\bar{\mu}_s(x) := \text{sgn}(x)\bar{\mu}(x)$  the *signed integrated tail* of  $\mu$ . We similarly define the *double integrated tail* of  $\rho$  by

$$\bar{\bar{\rho}}(x) = \begin{cases} \int_{(x, \infty)} \rho(z, \infty) dz, & x > 0, \\ \int_{(-\infty, x)} \rho(-\infty, z) dz, & x < 0, \\ 0, & x = 0. \end{cases}$$

A Lévy process  $(L_t)_{t \geq 0}$  with  $\sigma^2 = 0$ ,  $\rho = 0$  and  $|\mu| < \infty$ , where  $|\cdot|$  denotes the *total variation norm*, is called a *compound Poisson process*. If, additionally,  $\text{supp } \mu \subset \mathbb{R}_+$

then  $(L_t)_{t \geq 0}$  is called a *spectrally positive compound Poisson process*.

**Invariant measures and Harris recurrence** Let  $(X_t)_{t \geq 0}$  be a Markov process on  $\mathbb{R}$  with open state space  $\mathcal{O} \subseteq \mathbb{R}$  and with pointwise generator  $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$ . As mentioned in the introduction, a measure  $\pi$  with  $\text{supp } \pi \subset \mathcal{O}$  is called *invariant* for  $(X_t)_{t \geq 0}$  if (1.2) holds. It is called *infinitesimally invariant* for  $(X_t)_{t \geq 0}$  if

$$\int_{\mathbb{R}} \mathcal{A}f(x) \pi(dx) = 0 \quad \text{for all } f \in \mathcal{C}_c^\infty(\mathcal{O}),$$

and  $\pi$  is called *limiting* for  $(X_t)_{t \geq 0}$  if  $\pi$  is a *distribution*, i.e.  $|\pi| = 1$ , and

$$\lim_{t \rightarrow \infty} \|\mathbb{P}^x(X_t \in \cdot) - \pi\|_{\text{TV}} = 0, \quad \text{for all } x \in \mathcal{O},$$

where  $\|\cdot\|_{\text{TV}}$  denotes the total variation norm.

The process  $(X_t)_{t \geq 0}$  is called *Harris recurrent* if there exists a non-trivial  $\sigma$ -finite measure  $a$  on  $\mathbb{R}$  such that for all  $B \in \mathcal{B}(\mathbb{R})$  with  $a(B) > 0$  it holds  $\mathbb{P}^x(\tau_B < \infty) = 1$  where  $\tau_B := \inf\{t \geq 0 : X_t \in B\}$ . It is well-known (cf. Meyn and Tweedie 1993b) that for any Harris recurrent Markov process  $(X_t)_{t \geq 0}$  an invariant measure  $\pi$  exists which is unique up to multiplication with a constant. If  $\pi$  is finite it can be normalized to be a distribution. In this case  $(X_t)_{t \geq 0}$  is called *positive Harris recurrent*.

### 3 Lévy Langevin Monte Carlo

Let  $\pi$  be a probability distribution on  $\mathbb{R}$ , and let  $(L_t)_{t \geq 0}$  be a Lévy process on  $\mathbb{R}$ . Further, let  $(X_t)_{t \geq 0}$  be a solution of

$$dX_t = \phi(X_t)dt + dL_t.$$

Can we choose  $\phi : \mathbb{R} \rightarrow \mathbb{R}$  in such a way that  $\pi$  is limiting for  $(X_t)_{t \geq 0}$ ? In the spirit of fLMC we call the sampling of  $(X_t)_{t \geq 0}$  in order to sample from  $\pi$  *Lévy Langevin Monte Carlo* (LLMC).

As mentioned in the introduction a general answer to this question cannot be given without certain conditions on  $\pi$  and  $(L_t)_{t \geq 0}$ . Throughout, we assume that  $\pi(dx) = \pi(x)dx$  is absolutely continuous, and that the Lévy process  $(L_t)_{t \geq 0}$  with characteristic triplet  $(\gamma, \sigma^2, \Pi)$  is not purely deterministic, i.e.  $\sigma^2 > 0$  or  $\Pi \neq 0$ . Recall that  $\Pi = \mu + \rho$  with  $\mu$  and  $\rho$  as in Sect. 2.

Define  $\mathcal{E} := \{x \in \mathbb{R} : \pi(x) > 0\}$  and assume either  $\mathcal{E} = \mathbb{R}$  or some open half-line - without loss of generality we choose in this case  $\mathcal{E} = (0, \infty)$ . This choice of  $\mathcal{E}$  is not a real restriction, as explained further in Sect. 5. Additionally we assume the following:

- (a1) If  $\mathcal{E} = (0, \infty)$ , then  $(L_t)_{t \geq 0}$  is a spectrally positive compound Poisson process.
- (a2) If  $\mathcal{E} = (0, \infty)$ , then there exists  $c > 0$  such that  $\int_0^x \pi(z) dz \leq cx\pi(x)$  for all  $x \ll 1$ .
- (a3) If  $(L_t)_{t \geq 0}$  has paths of unbounded variation, then  $\pi \in W_{loc}^{1,1}(0, \infty)$ .

### 3.1 Infinitesimally invariant distributions

**Theorem 3.1** *Let  $(L_t)_{t \geq 0}$  be a Lévy process in  $\mathbb{R}$  with characteristic triplet  $(\gamma, \sigma^2, \Pi)$  with  $\Pi = \mu + \rho$  as in Sect. 2. Let  $\pi$  be a distribution on  $\mathbb{R}$  such that (a1)–(a3) are fulfilled. Consider the SDE (1.3) with*

$$\phi(x) := \mathbb{1}_{\mathcal{E}}(x) \left( \frac{\frac{1}{2}\sigma^2\pi'(x) - \bar{\mu}_s * \pi(x) + (\bar{\rho} * \pi)'(x)}{\pi(x)} - \gamma \right). \tag{3.4}$$

Then  $\pi$  is an infinitesimally invariant distribution of any solution  $(X_t)_{t \geq 0}$  of (1.3).

The proof of Theorem 3.1 will be clearer if we point out the primary thoughts behind Assumptions (a1)–(a3) first.

**Remark 3.2** Assumptions (a1) and (a2) make sure that the process  $(X_t)_{t \geq 0}$  stays in the open half-line  $(0, \infty)$  if  $\mathcal{E} = (0, \infty)$ . The former does so by allowing only upward jumps while the latter guarantees that  $(X_t)_{t \geq 0}$  cannot drift onto 0, as we will see in the proof below.

Clearly, Assumption (a3) becomes only relevant if  $\mathcal{E} = \mathbb{R}$ , and it ensures that our choice of the drift coefficient in (3.4) is well-defined. Note that it can be weakened if  $\sigma^2 = 0$  as  $\pi \in W_{loc}^{1,1}(0, \infty)$  is sufficient but not necessary for  $(\bar{\rho} * \pi)'$  to be well-defined. However, since we discuss in this article mostly processes with paths of bounded variation we choose to omit various special cases for the sake of clarity.

**Proof of Theorem 3.1** Denote by  $\mathcal{O} \subset \mathbb{R}$  the state space of  $(X_t)_{t \geq 0}$ . In order to show that  $\mathcal{O} = \mathcal{E}$  we prove that  $X_t \in \mathcal{E}$  for all  $t \geq 0$  if  $X_0 \in \mathcal{E}$ . As this is trivially true for  $\mathcal{E} = \mathbb{R}$  we show it only for  $\mathcal{E} = (0, \infty)$ .

In this case  $(L_t)_{t \geq 0}$  is a spectrally positive compound Poisson process, by (a1). Thus,  $(X_t)_{t \geq 0}$  cannot exit  $\mathcal{E}$  via jumps. We are going to show that  $(X_t)_{t \geq 0}$  cannot exit via drift either. If no jump of  $(L_t)_{t \geq 0}$  interrupts the path of  $(X_t)_{t \geq 0}$  then  $t \mapsto X_t$  is monotone decreasing and follows the autonomous differential equation

$$\begin{cases} dX_t = \phi(X_t)dt, \\ X_0 = x > 0, \end{cases}$$

with  $\phi(x) = -\mathbb{1}_{(0, \infty)}(x) \frac{\bar{\mu}_s * \pi(x)}{\pi(x)}$ . Separation of variables yields that the time  $T$  it takes for  $(X_t)_{t \geq 0}$  to drift from  $x$  to

$x' \in [0, x]$  is given by

$$T = \int_{x'}^x \frac{\pi(z)}{\bar{\mu}_s * \pi(z)} dz.$$

By (a1) and (a2),  $\bar{\mu}_s * \pi(x) \leq cx\pi(x)$  for some constant  $c > 0$ , and thereby

$$\int_0^x \frac{\pi(z)}{\bar{\mu}_s * \pi(z)} dz = \infty \tag{3.5}$$

for all  $x > 0$ . Hence,  $(X_t)_{t \geq 0}$  cannot drift onto 0 in finite time. Therefore,  $\mathcal{O} = \mathcal{E}$  in this case as well.

We now return to the general case. A straight-forward application of Itô's lemma and the Lévy-Itô decomposition, similar to (Schnurr 2009, Thm. 2.50), yields that for the pointwise generator  $(\mathcal{A}, \mathcal{D}(\mathcal{A}))$  of  $(X_t)_{t \geq 0}$  it holds  $\mathcal{C}_c^\infty(\mathcal{O}) \subset \mathcal{D}(\mathcal{A})$ , and

$$\begin{aligned} \mathcal{A}f(x) &= (\phi(x) + \gamma)f'(x) + \frac{1}{2}\sigma^2 f''(x) \\ &\quad + \int_{\mathbb{R}} (f(x+z) - f(x))\mu(dz) \\ &\quad + \int_{\mathbb{R}} (f(x+z) - f(x) - \nabla f(x))\rho(dz) \end{aligned}$$

for all  $f \in \mathcal{C}_c^\infty(\mathcal{O})$ . By (Behme et al. 2022, Thm. 4.2) a measure  $\eta$  is infinitesimally invariant for  $(X_t)_{t \geq 0}$  if

$$-((\phi + \gamma)\eta)' + \frac{1}{2}\sigma^2 \eta'' - (\bar{\mu}_s * \eta)' + (\bar{\rho} * \eta)'' = 0 \tag{3.6}$$

in the distributional sense w.r.t.  $\mathcal{C}_c^\infty(\mathcal{O})$ . Because of  $\mathcal{O} = \mathcal{E}$ , simply inserting  $\pi$  into (3.6) proves the claim.  $\square$

### 3.2 Invariant distributions

In general, proving that an infinitesimally invariant distribution is an invariant distribution is hard. The best-case scenario is given when  $(X_t)_{t \geq 0}$  is a Feller process and the test functions constitute a core of the pointwise generator of  $(X_t)_{t \geq 0}$ . In this case infinitesimally invariant and invariant are equivalent notions, cf. Liggett (2010).

Although there exist easily verifiable conditions on the drift coefficient  $\phi$  and  $(L_t)_{t \geq 0}$  such that a solution of (1.3) is a Feller process (cf. Kühn 2018) these have some drawbacks. The fact that typically,  $\phi$  is required to be continuous and fulfills a linear growth condition, i.e.  $|\phi(x)| \leq C(1 + |x|)$  for some  $C > 0$ , excludes many interesting cases. Moreover, even if  $(X_t)_{t \geq 0}$  is a Feller process, we are still left with the question whether the test functions form a core. The task of finding conditions for this to be the case is an open problem (cf. Böttcher et al. 2013) which has not yet been answered to the best of our knowledge.

**Remark 3.3** Both the article (Eliazar and Klafter 2003) on Lévy Langevin dynamics and the original article (Şimşekli 2017) on fLMC do not provide arguments as to why the considered target measures are invariant for the respective processes.

The chosen approach in both articles revolves around finding a stationary solution for Kolmogorov’s forward equation of the underlying SDE (1.3). This equation, which is also known as Fokker–Planck equation, is inherently connected to invariant distributions as any weak stationary solution of it can be associated to an infinitesimally invariant measure of a solution  $(X_t)_{t \geq 0}$  of (1.3). In the aforementioned articles it is suggested that the transition densities  $p(t, x, y)$  of  $(X_t)_{t \geq 0}$  defined via

$$\mathbb{P}^x(X_t \in B) = \int_B p(t, x, y) dy$$

solve the associated Kolmogorov forward equation. For many processes this is true, e.g. Feller diffusions (cf. Kallenberg 1997) just to name one. However, for SDEs (1.3) with general Lévy noises we were not able to find a reference with a rigorous proof of this claim. If it was indeed true, then any invariant measure  $\pi$  of  $(X_t)_{t \geq 0}$  would necessarily be a stationary solution of Kolmogorov’s forward equation.

Moreover, both articles are missing an argument as to why the stationary solution of Kolmogorov’s forward equation is unique. Although in Şimşekli (2017) another article (cf. Schertzer et al. 2001) is cited on this topic, in said reference uniqueness is merely argued heuristically but not proved.

However, as noted in the introduction, the articles (Huang et al. 2021; Zhang and Zhang 2023) present solutions to these issues, at least for the case of fLMC, by showing the strong Feller property of stable-driven SDE.

To ensure methodological rigor, we present in this section a different way of showing that  $\pi$  is an invariant distribution of  $(X_t)_{t \geq 0}$ , and as such even unique. We consider this approach in the special case of  $\mathcal{E} = (0, \infty)$  and  $(L_t)_{t \geq 0}$  being a compound Poisson process and discuss some of the obstacles one faces when trying to extend the results to more general frameworks further below in Sect. 5.

Let us now briefly describe our setting. Denote by

$$\mathcal{P} := \{(x_i)_{i \in \mathbb{Z}} \subset (0, \infty)^{\mathbb{Z}} : x_i < x_{i+1} \text{ for all } i \in \mathbb{Z}, \text{ and } 0 \text{ is the unique accumulation point of } (x_i)_{i \in \mathbb{Z}}\}$$

a set of partitions of the open interval  $(0, \infty)$ . We call a function  $f \in L^1_{loc}(0, \infty)$  *piecewise weakly differentiable* if there exists a partition  $(x_i)_{i \in \mathbb{Z}} \in \mathcal{P}$  such that  $f|_{(x_i, x_{i+1})} \in W^{1,1}(x_i, x_{i+1})$  for all  $i \in \mathbb{Z}$ . Analogously, we call  $f$  *piecewise Lipschitz continuous* if there exists a partition  $(x_i)_{i \in \mathbb{Z}} \in \mathcal{P}$  such that  $f|_{(x_i, x_{i+1})}$  is Lipschitz continuous for all  $i \in \mathbb{Z}$ .

Let  $(L_t)_{t \geq 0}$  be a Lévy process in  $\mathbb{R}$  and let  $\pi(dx) = \pi(x)dx$  be an absolutely continuous distribution on  $(0, \infty)$ . Our assumptions are as follows:

- (b1)  $\pi$  is a positive, piecewise weakly differentiable function, and there exist constants  $C, C', \alpha > 0$  such that  $\lim_{x \rightarrow \infty} \pi(x)e^{\alpha x} = C$ , and  $\int_0^x \pi(z)dz \leq C'\pi(x)x$  for  $x \ll 1$ .
- (b2)  $(L_t)_{t \geq 0}$  is a spectrally positive compound Poisson process, i.e. a Lévy process with characteristic triplet  $(0, 0, \mu)$  such that  $\text{supp } \mu \subset \mathbb{R}_+$  and  $\int_0^\infty (1 \vee z)\mu(dz) < \infty$ .

Note that our standing assumptions (a1) - (a3) are direct consequences of (b1) and (b2). In this setting, the drift coefficient given by (3.4) is reduced to

$$\phi(x) = -\mathbb{1}_{(0, \infty)}(x) \frac{\bar{\mu}_s * \pi(x)}{\pi(x)}, \tag{3.7}$$

and it is easy to see that  $\phi(x) \in (-\infty, 0)$  for all  $x > 0$ .

Sometimes it will be advantageous to write  $L_t = \sum_{i=1}^{N_t} \xi_i$ , where  $(N_t)_{t \geq 0}$  is a Poisson process with intensity  $|\mu|$  and  $(\xi_i)_{i \in \mathbb{N}}$  is a sequence of i.i.d. random variables distributed according to  $\mu/|\mu|$ . Note that  $\mathbb{E}\xi_1 < \infty$  by (b2).

With the following theorem we show that, under (b1) and (b2), a solution  $(X_t)_{t \geq 0}$  of (1.3) has the unique invariant distribution  $\pi$  if, additionally, one of the following two conditions is met:

- (c1) There exists  $n \in \mathbb{N}$  such that  $\text{supp } \mu \subset (1/n, n)$ , or
- (c2)  $\pi$  is piecewise Lipschitz continuous.

**Theorem 3.4** *Assume that (b1) and (b2) hold, and let  $(X_t)_{t \geq 0}$  be a solution of (1.3) with  $\phi$  as in (3.7). Then*

- (i)  $(X_t)_{t \geq 0}$  is positive Harris recurrent, and
- (ii) any invariant distribution of  $(X_t)_{t \geq 0}$  is an infinitesimally invariant distribution of  $(X_t)_{t \geq 0}$ .

*Additionally, if (c1) or (c2) are fulfilled, then*

- (iii)  $\pi$  is the unique invariant distribution of  $(X_t)_{t \geq 0}$ .

The proof of Theorem 3.4 is presented in Sect. 4, and it is divided into several steps. The first assertion is shown by using the Foster–Lyapunov method of Meyn and Tweedie (1992), Meyn and Tweedie (1993a), Meyn and Tweedie (1993b), while the second assertion is a simple application of (Behme et al. 2022, Cor. 5.4). Under (c1) we show Theorem 3.4 (iii) via techniques from the theory of ordinary differential equations. If instead (c2) is true, we approximate

$(X_t)_{t \geq 0}$  by a sequence of processes fulfilling (c1) to prove the claim.

### 3.3 Limiting distributions

The natural follow-up question of Theorem 3.4 is whether existence and uniqueness of an invariant distribution  $\eta$  for  $(X_t)_{t \geq 0}$  implies that  $\eta$  is a limiting distribution. This property of  $(X_t)_{t \geq 0}$ , i.e. the existence of a limiting distribution, is called *ergodicity*.

As before,  $\pi(dx) = \pi(x)dx$ ,  $\mathcal{E} = (0, \infty)$ , and  $(L_t)_{t \geq 0}$  is a spectrally positive compound Poisson process.

**Corollary 3.5** *Assume (b1) and (b2) hold, and let  $(X_t)_{t \geq 0}$  be a solution of (1.3) with  $\phi$  as in (3.7). Further assume that some skeleton chain of  $(X_t)_{t \geq 0}$  is irreducible, i.e. there exists  $\Delta > 0$  such that for all  $B \in \mathcal{B}((0, \infty))$  with  $\lambda^{\text{Leb}}(B) > 0$  and all  $x \in (0, \infty)$  there exists  $n \in \mathbb{N}$  such that*

$$\mathbb{P}^x(X_{n\Delta} \in B) > 0.$$

Then  $(X_t)_{t \geq 0}$  is ergodic.

**Proof** Follows directly from Theorem 3.4 (i) and (Meyn and Tweedie 1993a, Thm. 6.1).  $\square$

**Lemma 3.6 (Irreducible skeleton chain)** *Assume (b1) and (b2) hold, and let  $(X_t)_{t \geq 0}$  be a solution of (1.3) with  $\phi$  as in (3.7). Additionally assume that  $\mu = \mu_1 + \mu_2$  where  $\mu_1$  is arbitrary and  $\mu_2$  is absolutely continuous and such that  $\mu_2(I) > 0$  for all open intervals  $I \subset (0, \infty)$ . Then the 1-skeleton chain is irreducible, and  $(X_t)_{t \geq 0}$  is ergodic.*

**Proof** Without loss of generality we assume  $\mu_1 = 0$  since otherwise we may simply condition on the event that the jumps are only sampled from  $\mu_2$ .

Let  $B \in \mathcal{B}((0, \infty))$  with  $\lambda^{\text{Leb}}(B) > 0$ . Our goal is to show that for all  $x \in (0, \infty)$  there exists  $n \in \mathbb{N}$  such that

$$\mathbb{P}^x(X_n \in B) > 0. \tag{3.8}$$

It suffices to show (3.8) only for sets  $B$  for which  $\inf B > 0$  since for arbitrary  $B \in \mathcal{B}(0, \infty)$  with  $\lambda^{\text{Leb}}(B) > 0$  there exist  $0 < a < b$  such that  $\lambda^{\text{Leb}}(B \cap (a, b)) > 0$ . Moreover, it also suffices to just consider  $x < \inf B$  and  $n = 1$ . This is due to the fact that for arbitrary  $x \in (0, \infty)$  and  $m \in \mathbb{N}$  we obtain

$$\mathbb{P}^x(X_m \in B) \geq \mathbb{P}^x(X_m \in B, N_{m-1} = 0),$$

where we recall that  $(N_t)_{t \geq 0}$  is the Poisson process counting the jumps of  $(L_t)_{t \geq 0}$ , and therefore also of  $(X_t)_{t \geq 0}$ .

As  $\phi(x) < 0$  for all  $x > 0$ , we may choose  $m$  large enough such that  $X_{m-1}^x < \inf B$  on  $\{N_{m-1} = 0\}$ , and consider  $X_{m-1}^x$  as a new starting point.

Thus, let  $0 < x < \inf B$ . In the following we condition on the event that exactly one jump occurs until  $t = 1$ . Denote  $Y_t := (X_t^x | N_1 = 1)$ . It holds

$$\mathbb{P}^x(X_1 \in B) \geq c\mathbb{P}^x(Y_1 \in B)$$

for some  $c > 0$ . Further, denote by  $T \in (0, 1)$  the uniformly distributed time of the jump. We show that the joint cumulative distribution function

$$\mathbb{P}(T \leq t, Y_1 \leq y)$$

is strictly monotone on  $(0, 1) \times (x, \infty)$  in both arguments. Let  $0 < t < t' < 1$  and  $y \in (x, \infty)$ . We obtain

$$\mathbb{P}(T \in (t, t'], Y_1 \leq y) \geq \mathbb{P}(T \in (t, t'], \xi_1 \leq x - Y_{t-}) > 0.$$

Indeed, if  $T \in (t, t']$  and additionally  $\xi_1 \leq x - Y_{t-}$ , then  $Y_1 \leq x < y$ , and since  $Y_{t-} < x$  we get  $\mathbb{P}(T \in (t, t'], \xi_1 \leq x - Y_{t-}) > 0$ .

Now, let  $t \in (0, 1)$  and  $x < y < y' < \infty$ . We note that for every  $t \in (0, 1)$  there exists some interval  $I \subset (0, \infty)$  such that  $Y_1 \in (y, y']$  if  $T = t$  and  $\xi_1 \in I$ . This is due to the fact that the paths of  $(X_t)_{t \geq 0}$  between two jumps are continuous and strictly decreasing. Moreover, since  $Y_1$  depends continuously on  $T$  and  $\xi_1$ , there exists  $\varepsilon > 0$  and an interval  $I' \subset (0, \infty)$  such that  $Y_1 \in (y, y']$  if  $T \in (t - \varepsilon, t]$  and  $\xi_1 \in I'$ . Thus,

$$\begin{aligned} \mathbb{P}(T \leq t, Y_1 \in (y, y']) &\geq \mathbb{P}(T \in (t - \varepsilon, t], Y_1 \in (y, y']) \\ &\geq \varepsilon\mu(I') > 0 \end{aligned}$$

by the assumption on  $\mu$ .

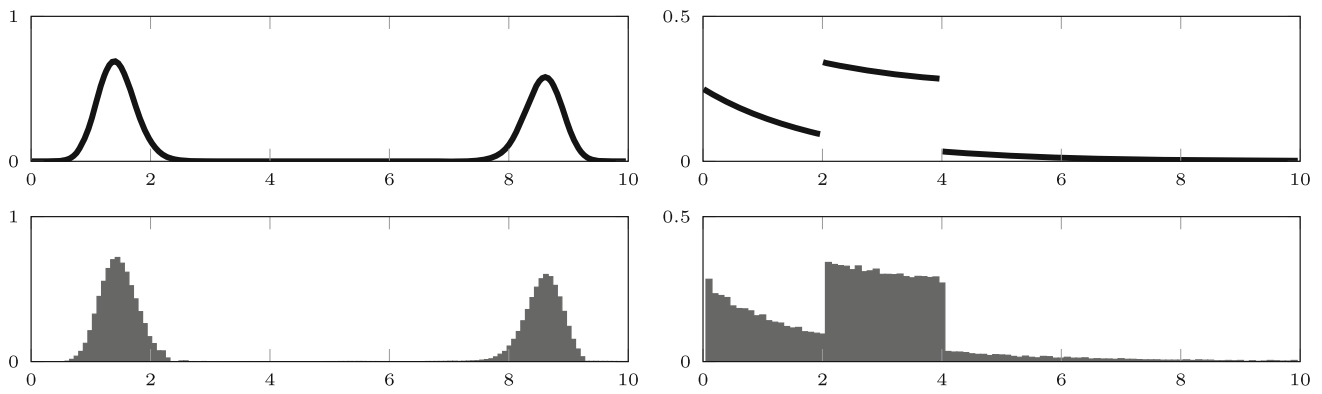
As both  $T$  and  $Y_1$  have clearly no atoms in  $(0, 1)$  and  $(x, \infty)$ , respectively, there exists a joint density function  $f_{(T, Y_1)}$  of  $(T, Y_1)$  on  $(0, 1) \times (x, \infty)$  which is strictly positive. Hence

$$\begin{aligned} \mathbb{P}^x(X_1 \in B) &\geq c\mathbb{P}^x(T \in (0, 1), Y_1 \in B) \\ &= \int_{(0,1) \times B} f_{(T, Y_1)}(t, y) dt dy > 0. \end{aligned}$$

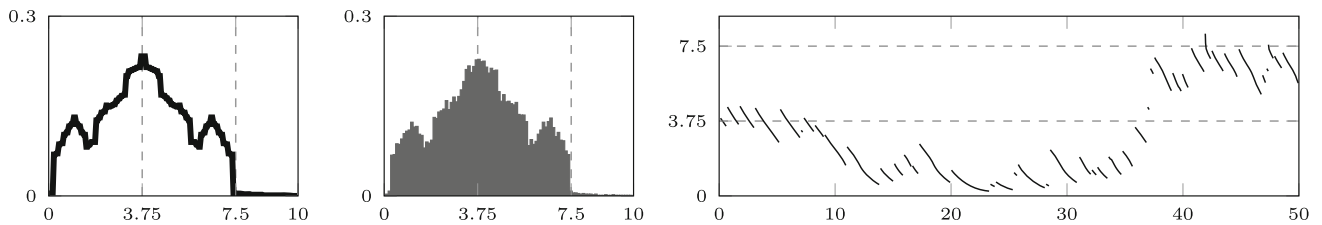
This, together with Corollary 3.5, concludes the proof.  $\square$

### 3.4 Examples

In this section we illustrate Theorem 3.4 on various examples by sampling (1.3). To this end, we first compute a realization of the path of the driving noise  $(L_t)_{t \geq 0}$ . With  $(L_t)_{t \geq 0}$  being a compound Poisson process this is straight-forward. It remains to solve a (deterministic) differential equation which is then done via the classic Euler method.



**Fig. 1** Illustration of Example 3.7 (left) and Example 3.8 (right). The target densities are displayed in the top images while the bottom images show the respective histograms with a sample size of  $N = 50,000$  each



**Fig. 2** From left to right we see the target density function  $\pi$ , a histogram of the sampled distribution with sample size  $N = 50,000$ , and an exemplary sample path of  $(X_t)_{t \geq 0}$

**Example 3.7** (double-well) In Şimşekli (2017) is pointed out that sampling from a target distribution  $\pi(dx) = \pi(x)dx$  with two separated modes is challenging for classic LMC. The lower the values of  $\pi$  are between the modes the longer it takes on average for the continuous Langevin diffusion to move from one mode to the other. This issue can be circumvented by allowing jumps. Define

$$p(x) = \exp \left\{ -\frac{1}{10}x(x - 4)(x - 6.02)(x - 10) + 0.5 \right\}, \quad x > 0,$$

which is taken from (Şimşekli 2017, Sec. 4), but shifted to the right such that both modes are contained in  $(0, 10)$ . We choose as target density

$$\pi(x) = p(x)\mathbb{1}_{(0,M)}(x) + e^{-x}\mathbb{1}_{[M,\infty)}(x)$$

for some large number  $M \gg 10$ . This ensures that the tail growth condition in (b1) is fulfilled. Note that such a technique can be helpful for various practical purposes, such as sampling, when the exact shape of the tails is not the primary focus. As the probability that the process reaches  $[M, \infty)$  tends to zero, it becomes less important how the density behaves for  $x \geq M$ .

As driving noise we consider a Lévy process  $(L_t)_{t \geq 0}$  with characteristic triplet  $(0, 0, \mu)$  where

$$\mu(dx) = e^{-x}dx + \delta_4 + 2\delta_8.$$

Clearly, conditions (b1), (b2), and (c2) are fulfilled. Thus,  $\pi$  is invariant for a solution  $(X_t)_{t \geq 0}$  of (1.3) with  $\phi$  as in (3.7) by Theorem 3.4. Further, since Lemma 3.6 applies,  $\pi$  is even limiting.

We demonstrate this example in Fig. 1.

Note that, in general, there is no closed-form expression of  $\phi$  due to the convolution term appearing in its definition. Hence, we use numerical integration to evaluate  $\phi$  for this and the following two examples.

**Example 3.8** (non-smooth density) Another important advantage of LLMC compared to LMC and also fLMC of Şimşekli (2017) is the possibility to choose non-smooth target densities. Let, for example,  $\pi(dx) = \pi(x)dx$  with

$$\pi(x) = e^{-0.5x} + \mathbb{1}_{(2,4)}(x),$$

and  $(L_t)_{t \geq 0}$  be a Lévy process with characteristic triplet  $(0, 0, \mu)$  where

$$\mu(dx) = x^2e^{-0.5x}dx + \delta_1.$$

Let  $(X_t)_{t \geq 0}$  be a solution of (1.3) with  $\phi$  as in (3.7). As for the previous example, (b1), (b2), and (c2) are met, and  $\pi$

is invariant and limiting for  $(X_t)_{t \geq 0}$  by Theorem 3.4 and Lemma 3.6.

This, too, is displayed in Fig. 1.

**Example 3.9** (DRESDEN FRAUENKIRCHE) to illustrate that our result also covers target densities with lots of detail we consider the density  $\pi$  as in Fig. 2 (left) which represents the silhouette of the DRESDEN FRAUENKIRCHE, continued by an exponential tail. We manufactured  $\pi$  in a way such that (b1) is met. As driving noise we choose a spectrally positive Lévy process with characteristic triplet  $(0, 0, \mu)$  where

$$\mu(dx) = \mathbb{1}_{\{x>0\}} e^{-\frac{x^2}{2}} dx.$$

Let  $(X_t)_{t \geq 0}$  be a solution of (1.3) with  $\phi$  as in (3.7). As for both prior examples,  $\pi$  is clearly invariant and limiting for  $(X_t)_{t \geq 0}$ . The density function, the sampled distribution and an exemplary sample path of  $(X_t)_{t \geq 0}$  can be seen in Fig. 2.

Taking a closer look we see that the process  $(X_t)_{t \geq 0}$  slows down considerably upon entering the interval  $(0, 7.5)$  on which most of the mass of  $\pi$  concentrates. In general, the drift coefficient takes on large values in areas of small mass and small values in areas of high mass. This stems from  $\pi$  appearing in the denominator of (3.7), and can be observed by inspecting the slopes of the sample path in Fig. 2.

Moreover, the process becomes slower the closer it gets to the origin. On the one hand, this is due to Assumption (b1) (and (a2), respectively), and ensures that 0 cannot be reached in finite time. On the other hand, this slowing down is caused by the convolution with the signed tail function  $\bar{\mu}_s$  in the nominator of (3.7).

This is reasonable: Because jumps go only upwards it is less likely for  $(X_t)_{t \geq 0}$  to reach the area of the left half of the silhouette (approximately the interval  $(0, 3.75)$ ) than to appear in the area of its right half. But since both sides are symmetrical the drift must compensate for that.

### 4 Proof of Theorem 3.4

In the following, whenever constants  $C, C'$  or  $\alpha$  appear in the proof below, we mean the constants of Assumption (b1).

#### Proof of Theorem 3.4 (i): Positive Harris recurrence

The Foster–Lyapunov method of Meyn and Tweedie (1992, 1993a, b) is tailored for processes which cover the whole real line. Hence, we consider the auxiliary process  $Y_t = s(X_t)$  where  $s : (0, \infty) \rightarrow \mathbb{R}$  is a smooth strictly monotone function such that  $s(x) = \ln(x)$  for  $x \in (0, 1 - \varepsilon)$  and  $s(x) = x$  for  $x \in (1 + \varepsilon, \infty)$  where  $0 < \varepsilon < e^{-1}$  is some constant. Clearly,  $(X_t)_{t \geq 0}$  is positive Harris recurrent if and only if  $(Y_t)_{t \geq 0}$  is positive Harris recurrent.

Central to this method are the so-called *norm-like functions* whose precise definition needs additional notation: For  $m \in \mathbb{N}$  denote  $O_m := (e^{-m}, m)$  and choose  $h_m \in C_c^\infty(\mathbb{R})$  such that  $0 \leq h_m \leq 1$ , and with  $h_m(x) = 1$  for all  $x \in O_m$ , and  $h_m(x) = 0$  for all  $x \in O_{m+1}^c$ . Let  $(\mathcal{X}_t^{(m)})_{t \geq 0}$  be the unique strong solution of

$$d\mathcal{X}_t^{(m)} = h_m(\mathcal{X}_{t-}^{(m)}) \left( \phi(\mathcal{X}_{t-}^{(m)}) dt + dL_t \right), \quad \mathcal{X}_0^{(m)} > 0.$$

We set  $\mathcal{Y}_t^{(m)} := s(\mathcal{X}_t^{(m)})$ . Clearly, this construction implies that for all  $m \in \mathbb{N}$  if  $\mathcal{X}_0^{(m)} \in O_m$  it holds  $\mathcal{Y}_t^{(m)} = Y_t$  for all  $t < T_m := \inf\{s \geq 0 : |Y_s| \geq m\}$ . For  $m \in \mathbb{N}$  denote by  $(\mathcal{G}_m, \mathcal{D}(\mathcal{G}_m))$  the extended generator of  $(\mathcal{Y}_t^{(m)})_{t \geq 0}$ . A function  $f : \mathbb{R} \rightarrow \mathbb{R}_+$  is called *norm-like* w.r.t.  $(Y_t)_{t \geq 0}$  if

- (i)  $f(x) \rightarrow \infty$  as  $x \rightarrow \pm\infty$ , and
- (ii)  $f \in \mathcal{D}(\mathcal{G}_m)$  for all  $m \in \mathbb{N}$ .

It is typical for the Foster–Lyapunov method that one only requires a single norm-like function (sometimes also called Foster–Lyapunov function) which fulfills a certain inequality. Our particular choice is presented in the lemma below.

In the following, to make notation easier, we denote  $y := s(x)$  if  $x \in (0, \infty)$  is given, and  $x := s^{-1}(y)$  if  $y \in \mathbb{R}$  is given.

**Lemma 4.1** *Let  $f \in C^1(\mathbb{R})$  with  $f(y) \in [1 + |y|, 2 + |y|]$  for all  $y \in \mathbb{R}$ , and  $f(y) = 1 + |y|$  for all  $|y| > \varepsilon$  for some  $\varepsilon > 0$ . Then  $f$  is norm-like w.r.t.  $(Y_t)_{t \geq 0}$ .*

*Moreover, for all  $m \in \mathbb{N}$  and  $y \in (-m, m)$  it holds*

$$\mathcal{G}_m f(y) = \phi(x) f_0'(x) + \int_{(0, \infty)} (f_0(x+z) - f_0(x)) \mu(dz). \tag{4.9}$$

where  $f_0(x) := f(y) = f(s(x))$ .

**Proof** Fix  $m \in \mathbb{N}$ . Itô’s formula yields

$$\begin{aligned} & \mathbb{E}^y [f(\mathcal{Y}_t^{(m)}) - f(y)] \\ &= \mathbb{E}^x [f_0(\mathcal{X}_t^{(m)}) - f_0(x)] \\ &= \mathbb{E}^x \left[ \int_0^t h_m(\mathcal{X}_{s-}^{(m)}) \phi(\mathcal{X}_{s-}^{(m)}) f_0'(\mathcal{X}_{s-}^{(m)}) ds \right. \\ & \quad \left. + \int_{z \neq 0} \int_0^t \left( f_0(\mathcal{X}_{s-}^{(m)} + h_m(\mathcal{X}_{s-}^{(m)})z) - f_0(\mathcal{X}_{s-}^{(m)}) \right) \right. \\ & \quad \left. \tilde{\mu}(\cdot, ds, dz) \right] \tag{4.10} \end{aligned}$$

where  $\tilde{\mu}$  is the jump measure of  $(L_t)_{t \geq 0}$ . To verify whether the jump measure may be replaced by the compensator under the expectation, and to subsequently swap the order of integration, we need some estimates. Clearly,  $f_0(x + h_m(x)z) - f_0(x) = 0$  for all  $z > 0$ , and  $x \notin [e^{-m-1}, m + 1]$ . On the



other hand, for all  $x \in [e^{-(m+1)}, m + 1]$  there exists  $M > 0$  such that

$$|f_0(x + h_m(x)z) - f_0(x)| \leq M \vee z$$

by the definition of  $f_0$ . Hence, by (Schnurr 2009, Thm. 2.21) and the fact that  $\int_0^\infty (1 \vee z)\mu(dz) < \infty$  by Assumption (b2),  $\tilde{\mu}(\cdot, ds, dz)$  may be replaced by  $ds\mu(dz)$  under the expectation in (4.10).

Applying Fubini's theorem and reversing the space transform, i.e. going back to  $(\mathcal{Y}_t)_{t \geq 0}$ , we obtain

$$\begin{aligned} & \mathbb{E}^y[f(\mathcal{Y}_t^{(m)}) - f(y)] \\ &= \mathbb{E}^x \left[ \int_0^t h_m(\mathcal{X}_{s-}^{(m)})\phi(\mathcal{X}_{s-}^{(m)})f_0'(\mathcal{X}_{s-}^{(m)})ds \right. \\ & \quad \left. + \int_0^t \int_{z \neq 0} (f_0(\mathcal{X}_{s-}^{(m)} + h_m(\mathcal{X}_{s-}^{(m)})z) - f_0(\mathcal{X}_{s-}^{(m)}))\mu(dz)ds \right] \\ &= \mathbb{E}^y \left[ \int_0^t g(\mathcal{Y}_{s-}^{(m)})ds \right] \end{aligned}$$

where

$$g(y) := h_m(x)\phi(x)f_0'(x) + \int_{z \neq 0} (f_0(x + h_m(x)z) - f_0(x))\mu(dz). \tag{4.11}$$

This function is clearly measurable. We observe that the integral term is continuous in  $y$  and vanishes for  $|y| \geq m + 1$ . Therefore,  $g$  is bounded and Tonelli's theorem is applicable yielding

$$\left| \mathbb{E}^y \left[ \int_0^t g(\mathcal{Y}_{s-}^{(m)}) ds \right] \right| \leq \int_0^t \mathbb{E}^y \left[ |g(\mathcal{Y}_{s-}^{(m)})| \right] ds \leq \|g\|_\infty t < \infty$$

for all  $y \in \mathbb{R}$  and all  $t \geq 0$ . Hence,  $f \in \mathcal{D}(\mathcal{G}_m)$  for all  $m \in \mathbb{N}$ . This completes the proof as we observe that (4.9) follows from the definition of the extended generator and upon realizing that the representation in (4.11) agrees with (4.9) for all  $y \in (-m, m)$ .  $\square$

The second key ingredient of the Foster–Lyapunov method is the following: A set  $K \subset \mathbb{R}$  is called *petite* for a Markov process  $(Y_t)_{t \geq 0}$  if there exists a distribution  $a$  on  $(0, \infty)$  and a non-trivial measure  $\varphi$  on  $\mathcal{B}(\mathbb{R})$  such that for all  $y \in K$  and  $B \in \mathcal{B}(\mathbb{R})$

$$\int_{(0, \infty)} \mathbb{P}^y(Y_t \in B)a(dt) \geq \varphi(B).$$

**Lemma 4.2** *All compact sets  $K \subset \mathbb{R}$  are petite for  $(Y_t)_{t \geq 0}$ .*

**Proof** We start with some helpful notation. For  $y \in \mathbb{R}$  denote by  $q_y(\cdot)$  the solution of the autonomous differential equation

$$\begin{cases} q_y' = \phi(s^{-1}(q_y))s'(s^{-1}(q_y)), \\ q_y(0) = y. \end{cases} \tag{4.12}$$

Then  $q_y(t)$  represents the (deterministic) state  $Y_t^y$  under the assumption that no jump occurs in the time interval  $[0, t]$ , that is  $N_t = 0$ . The inverse function  $q_y^{-1}(y')$  exists due to  $\phi(x) < 0$  for all  $x > 0$ . We note that it represents the time it takes to drift from  $y > 0$  to  $y' \in (0, y)$ , and is hence decreasing in  $y'$  and increasing in  $y$ .

Let  $K \subset \mathbb{R}$  be compact, without loss of generality assume  $K = [k_1, k_2]$ . Let  $a(dt) = e^{-t}dt$  and  $\varphi(dz) = c\mathbb{1}_{(k_1-1, k_1)}(z)dz$  for some  $c > 0$  which we are yet to choose. Let  $y \in K$  and  $B \in \mathcal{B}(\mathbb{R})$ . Using  $\mathbb{P}(N_t = 0) = e^{-t|\mu|}$  we compute

$$\begin{aligned} & \int_0^\infty \mathbb{P}^y(Y_t \in B)a(dt) \\ & \geq \int_0^\infty \mathbb{1}_{q_y^{-1}(B)}(t)e^{-t(1+|\mu|)}dt \\ & \geq \int_0^{q_y^{-1}(k_1-1)} \mathbb{1}_{q_y^{-1}(B)}(t)e^{-t(1+|\mu|)}\frac{q_y'(t)}{q_y'(t)}dt \\ & \geq \frac{1}{\sup\{|q_y'(t)| : t \leq q_y^{-1}(k_1-1)\}} \\ & \quad \int_{k_1-1}^y \mathbb{1}_B(z)e^{-q_y^{-1}(z)(1+|\mu|)}dz \\ & \geq \frac{1}{\sup\{|q_y'(t)| : t \leq q_y^{-1}(k_1-1)\}} \\ & \quad \int_{k_1-1}^{k_1} \mathbb{1}_B(z)e^{-q_y^{-1}(z)(1+|\mu|)}dz \\ & \geq \frac{\exp\{-q_{k_2}^{-1}(k_1-1)(1+|\mu|)\}}{\sup\{|q_y'(t)| : t \leq q_y^{-1}(k_1-1)\}} \int_{k_1-1}^{k_1} \mathbb{1}_B(z)dz. \end{aligned}$$

For the third inequality we substituted  $z := q_y(t)$  and used the fact that for all  $y, k_1 \in \mathbb{R}$  it holds  $\sup\{|q_y'(t)| : t \leq q_y^{-1}(k_1-1)\} < \infty$ . Indeed, this is implied by (4.12), and the properties of  $\phi$  and  $s$ . The fourth inequality is due to the reduction of the area of integration while the fifth inequality uses the monotonicity properties of  $q^{-1}$  described above. Lastly, choosing

$$c := \frac{\exp\{-q_{k_2}^{-1}(k_1-1)(1+|\mu|)\}}{\sup\{|q_y'(t)| : t \leq q_y^{-1}(k_1-1)\}}$$

finishes the proof.  $\square$

Finally, we are ready to prove the first claim of Theorem 3.4.

**Proof of Theorem 3.4 (i)** We show that there exist some positive constants  $c, d > 0$  and a closed petite set  $K \subset \mathbb{R}$  such that

$$\mathcal{G}_m f(y) \leq -c + d\mathbb{1}_K(y) \tag{4.13}$$

for all  $m \in \mathbb{N}$  and  $y \in (-m, m)$ . Then (Meyn and Tweedie 1993b, Thm. 4.2) implies that  $(Y_t)_{t \geq 0}$  is positive Harris recurrent, and therefore,  $(X_t)_{t \geq 0}$  is positive Harris recurrent as well.

Clearly, the function

$$\mathcal{G}_m f : y \mapsto \phi(x)f'_0(x) + \int_{(0,\infty)} (f_0(x+z) - f_0(x))\mu(dz) \tag{4.14}$$

is continuous, and bounded on  $(-m, m)$ .

Hence, (4.13) follows if we can show that  $\limsup_{y \rightarrow \pm\infty} \mathcal{G}_m f(y) \leq -c$  for some  $c > 0$ . We start with  $y \rightarrow +\infty$ . Note that for  $y \gg 1$  we have  $x = s^{-1}(y) = y$ , and, on the one hand  $f'_0(y) = 1$ , and, on the other hand  $f_0(y+z) - f_0(y) = z$ .

With Assumption (b1) we obtain

$$\begin{aligned} \limsup_{y \rightarrow +\infty} \phi(y) &= -\liminf_{y \rightarrow +\infty} (\bar{\mu}_s * \pi(y)) \frac{e^{\alpha y}}{C} \\ &\leq -\liminf_{y \rightarrow +\infty} \frac{1}{C} \int_{(0,M)} \bar{\mu}_s(z)\pi(y-z)e^{\alpha y} dz \end{aligned}$$

for all arbitrary, but fixed  $M > 0$ . Thus, also with Assumption (b1),

$$\begin{aligned} \limsup_{y \rightarrow +\infty} \phi(y)f'_0(y) &\leq -\liminf_{y \rightarrow +\infty} \int_{(0,M)} \bar{\mu}_s(z)e^{\alpha z} dz \\ &= -\int_{(0,M)} \bar{\mu}_s(z)e^{\alpha z} dz \end{aligned}$$

Since  $M > 0$  was arbitrary this yields  $\limsup_{y \rightarrow +\infty} \phi(y)f'_0(y) < \int_{(0,\infty)} \bar{\mu}_s(z)dz = -\mathbb{E}\xi_1$ . Now, for the second term of (4.14) we observe that for  $x = y \gg 1$

$$\int_{(0,\infty)} (f_0(x+z) - f_0(x))\mu(dz) = \int_{(0,\infty)} z\mu(dz) = \mathbb{E}\xi_1.$$

Consequently, there exists  $c > 0$  such that  $\mathcal{G}_m f(y) < -c < 0$  for  $y \gg 1$ .

Next, consider the behavior for  $y \rightarrow -\infty$ , and start with the observation that for  $y \ll -1$  one has  $x = s^{-1}(y) = e^y$ , and  $f'_0(x) = e^{-y}$ . With the definition of  $\phi$  and Assumption (b1) we obtain

$$|\phi(e^y)| \leq C'|\mu|e^y$$

for  $y \ll -1$ . Therefore,  $\phi(e^y)f'_0(e^y)$  is bounded for  $y \ll -1$ .

Finally, to find a suitable estimate for the second term of (4.14) for  $y \ll -1$  we fix  $M > 0$  such that  $\mu([M, \infty)) > 0$ . Observe that for  $y \ll -1$  it holds  $f_0(e^y + z) - f_0(e^y) < 0$  for all  $z \in (0, M)$ . Further, there exists  $K > 0$  such that  $f_0(e^y + z) < M' + z$  for all  $z \in [M, \infty)$ . We then compute

$$\begin{aligned} &\int_{(0,\infty)} (f_0(e^y + z) - f_0(e^y))\mu(dz) \\ &= \int_{(0,M)} (f_0(e^y + z) - f_0(e^y))\mu(dz) \\ &\quad + \int_{[M,\infty)} (f_0(e^y + z) - f_0(e^y))\mu(dz) \\ &\leq \int_{[M,\infty)} (K + z)\mu(dz) - f_0(e^y)\mu[M, \infty). \end{aligned}$$

As  $\int_{(0,\infty)} z\mu(dz) < \infty$  this implies

$$\lim_{y \rightarrow -\infty} \int_{(0,\infty)} (f_0(e^y + z) - f_0(e^y))\mu(dz) = -\infty,$$

and therefore,  $\mathcal{G}_m f(y) < -c$  for  $y \ll -1$  with the same  $c$  as above. This completes the proof.  $\square$

**Proof of Theorem 3.4 (ii): Invariant distributions are infinitely invariant**

**Proof of Theorem 3.4 (ii)** The claim follows from (Behme et al. 2022, Cor. 5.4) if we can show that  $\frac{1}{t} |\mathbb{E}^x f(X_t) - f(x)| < \infty$  for all  $f \in \mathcal{C}_c^\infty(0, \infty)$  and all  $t \geq 0$ .

Analogously to the proof of Lemma 4.1, a straight-forward application of Itô's formula yields

$$\frac{1}{t} |\mathbb{E}^x f(X_t) - f(x)| = \frac{1}{t} \left| \mathbb{E}^x \int_0^t g(X_{s-}) ds \right|$$

for  $f \in \mathcal{C}_c^\infty(0, \infty)$ , where

$$g(x) = \phi(x)f'(x) + \int_{(0,\infty)} (f(x+z) - f(x))\mu(dz).$$

Clearly, if  $f \in \mathcal{C}_c^\infty(0, \infty)$ , then  $g$  is bounded and it follows

$$\frac{1}{t} |\mathbb{E}^x f(X_t) - f(x)| \leq \|g\|_\infty < \infty.$$

$\square$

**Proof of Theorem 3.4 (iii): Uniqueness of the invariant distribution**

For the third assertion we require one of the additional assumptions. As described above we start with Assumption (c1), i.e. there exists  $n \in \mathbb{N}$  such that  $\text{supp } \mu \subset (1/n, n)$ .

*Proof of Theorem 3.4 (iii) under (c1)* It has been shown in (Behme et al. 2022, Thm. 4.2) that any infinitesimally invariant measure  $\eta$  of a solution  $(X_t)_{t \geq 0}$  of (1.3) necessarily solves the distributional equation

$$-(\phi\eta)' + \mu * \eta - |\mu|\eta = 0 \tag{4.15}$$

on  $(0, \infty)$ . To show that there exists only one probability distribution solving (4.15) we first need some regularity properties for  $\phi$ . A straight-forward calculation yields that for all  $x \geq 0$  the representation

$$\phi(x) = \frac{\int_0^x (\mu * \pi(z) - |\mu|\pi(z))dz}{\pi(x)} \tag{4.16}$$

holds. As  $\pi$  is piecewise weakly differentiable and  $\pi(x) > 0$  for  $x > 0$  it follows that  $1/\pi$  is piecewise weakly differentiable as well. Thus,  $\phi$  is piecewise weakly differentiable w.r.t. the same partition as  $\pi$ , since the numerator of the right-hand side of (4.16) is the primitive of a locally integrable function, and as such contained in  $W_{loc}^{1,1}(0, \infty)$ . Further, as  $\phi(x) < 0$  for all  $x > 0$  we infer that at least  $1/\phi \in L_{loc}^1(0, \infty)$ .

This property of  $\phi$  allows us to transform (4.15) into

$$\eta' = \frac{\mu * \eta - \phi'\eta - |\mu|\eta}{\phi}. \tag{4.17}$$

We note that the right-hand side of (4.17) defines a Schwartz distribution (cf. (Behme et al. 2022, Lem. 2.2)) if  $\eta$  is a real-valued Radon measure which we can assume as we are only looking for solutions which are probability distributions. More importantly, in this case the right-hand side of (4.17) is even a Schwartz distribution of order 0, i.e. it can be identified with some real-valued Radon measure on  $(0, \infty)$ .

In summary, the distributional derivative of  $\eta$  can be identified with a real-valued Radon measure which implies that  $\eta$  itself can be identified with a locally integrable function. But now, if we insert a locally integrable function  $\eta$  into the right-hand side of (4.17) we obtain a locally integrable function plus a discrete measure with atoms at the discontinuities of  $\phi$ . Integrating on both sides of (4.17) tells us that any solution of (4.15) is piecewise weakly differentiable w.r.t. the same partition as  $\pi$ .

Let  $n \in \mathbb{N}$  such that  $\text{supp } \mu \subset (1/n, n)$ . To solve (4.15) we integrate both sides and obtain

$$-\phi\eta + \bar{\mu}_s * \eta = c_1 \tag{4.18}$$

for some  $c_1 \in \mathbb{R}$ . Observe that for  $x \gg 1$  we have

$$|\phi(x)| \leq |\mu| \int_{1/n}^n \frac{\pi(x-z)}{\pi(x)} dz \leq |\mu|e^{\alpha n},$$

by Assumption (b1). Thus,  $\limsup_{x \rightarrow \infty} |\phi(x)| < \infty$ . From Young's convolution inequality it follows that  $\|\bar{\mu}_s * \eta\|_1 \leq \|\bar{\mu}_s\|_1 \|\eta\|_1 < \infty$ . Thus, for any absolutely continuous measure  $\eta$  the left-hand side of (4.18) can be identified with an element of  $L^1(\mathbb{R}_+)$ . Consequently,  $c_1 = 0$ .

Above we have seen that any probability distribution  $\eta$  solving (4.15) is absolutely continuous. Denoting by  $H(x) := \eta((0, x])$  the cumulative distribution function of  $\eta$  we thus know that the density function  $H'$  of  $\eta$  is integrable. If we once again use the fact that  $\int_0^x (\mu * H'(z) - |\mu|H'(z))dz = \bar{\mu}_s * (H')(x)$ , and that  $\mu * (H')(x) = 0$  for  $x \in (0, 1/n]$ , we obtain from (4.18) the equation

$$H'(x) = -\frac{|\mu|H(x)}{\phi(x)}, \quad x \in (0, 1/n]. \tag{4.19}$$

Caratheodory's theorem (cf. (Hale 1980, Thm. 5.3)) implies that (4.19) has for each initial value  $H(\varepsilon) = c_2 \in \mathbb{R}$  a unique solution. Clearly, for arbitrary but fixed  $c_2 \in \mathbb{R}$  the solution of (4.19) is given by  $H(x) = \frac{c_2}{F(\varepsilon)}F(x)$  where  $F(x) := \pi(0, x]$  is the cumulative distribution function of our target distribution  $\pi$ . This is easily seen with the definition of  $\phi$ .

For  $x > 1/n$  Eq. (4.18) reads

$$H'(x) = \frac{\int_0^x \mu * (H')(z)dz - |\mu|H(x)}{\phi(x)}. \tag{4.20}$$

From Assumption (c1) it follows that for all  $0 < x < b$  and any function  $f \in L^1(\mathbb{R}_+)$  holds

$$\mu * f(x) = \mu * (f|_{[0, b-\frac{1}{n}]})(x).$$

Consider Eq.(4.20) on  $[m/n, (m+1)/n]$  for some  $m \in \mathbb{N}$ . As initial condition we assume  $H(x) = c_3 F(x)$  for all  $x \in (0, m/n]$  and some  $c_3 \in \mathbb{R}$ . This results in the equation

$$H'(x) = \frac{c_3 \int_0^x \mu * \pi(z)dz - |\mu|H(x)}{\phi(x)}, \quad x \in [m/n, (m+1)/n]$$

for which Caratheodory's theorem again ensures a unique solution. Hence, by induction over  $m$  and subsequent normalization it follows that  $\pi$  is the unique infinitesimally invariant distribution of  $(X_t)_{t \geq 0}$ .

Finally, Theorem 3.4 (i), that is positive Harris recurrence, implies existence and uniqueness of an invariant distribution (cf. (Meyn and Tweedie 1993b, Sec. 4)). But this unique distribution must be  $\pi$  due to Theorem 3.4 (ii). This proves the claim.  $\square$

*Proof of Theorem 3.4 (iii) under (c2)* Assume  $\phi$  is piecewise locally Lipschitz continuous. In this case we show the assertion by approximating  $(X_t)_{t \geq 0}$  with a sequence of processes meeting Assumption (c1).

Recall that  $L_t := \sum_{i=0}^{N_t} \xi_i$ , and define

$$L_t^{(n)} := \sum_{i=0}^{N_t} \xi_i \mathbb{1}_{\{\frac{1}{n} < \xi_i < n\}}$$

for all  $n \in \mathbb{N}$ . Evidently,  $(L_t^{(n)})_{t \geq 0}$  is a Lévy process with characteristic triplet  $(0, 0, \mu^{(n)})$  where  $\mu^{(n)}(B) = \mu(B \cap (\frac{1}{n}, n))$  for all  $B \in \mathcal{B}(0, \infty)$ . To avoid the trivial case we consider only  $n \in \mathbb{N}$  large enough such that  $\text{supp } \mu \cap (\frac{1}{n}, n) \neq \emptyset$ .

Further denote

$$\phi_n(x) = -\frac{\bar{\mu}_s^{(n)} * \pi(x)}{\pi(x)}.$$

Observe that for  $n \in \mathbb{N}$  large enough and all  $x \in (0, \infty)$  it holds

$$\begin{aligned} |\phi(x) - \phi_n(x)| &= \frac{(\bar{\mu}_s - \bar{\mu}_s^{(n)}) * \pi(x)}{\pi(x)} \\ &\leq \mu((0, 1/n] \cup [n, \infty)) \frac{\int_0^x \pi(z) dz}{\pi(x)}. \end{aligned}$$

With  $\pi$  being a probability density the numerator is bounded. Using Assumption **(b1)** and the fact that  $\pi$  is bounded away from zero on compact intervals in  $(0, \infty)$  reveals that for all compact sets  $K \subset [0, \infty)$  there exists  $c > 0$  such that

$$|\phi(x) - \phi_n(x)| \leq c\mu((0, 1/n] \cup [n, \infty)) \tag{4.21}$$

for all  $x \in K$ .

For  $n \in \mathbb{N}$  large enough we consider now solutions  $(X_t^{(n)})_{t \geq 0}$  of the stochastic differential equations

$$dX_t^{(n)} = \phi_n(X_{t-}^{(n)})dt + L_t^{(n)}, \quad X_0^{(n)} \sim \pi. \tag{4.22}$$

Note that, by construction, the processes  $(X_t)_{t \geq 0}$  and  $(X_t^{(n)})_{t \geq 0}$ ,  $n \in \mathbb{N}$ , are defined on the same probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ , and that for all  $\omega \in \Omega$  the set of jump times of  $(X_t^{(n)}(\omega))_{t \geq 0}$  is a subset of the jump times of  $(X_t(\omega))_{t \geq 0}$ .

The proof of Theorem 3.4 (iii) under **(c1)** implies that  $(X_t^{(n)})_{t \geq 0}$  is a stationary process with invariant distribution  $\pi$ . Assume  $X_0 = X_0^{(n)}$ , and let us show that for all  $t \geq 0$  it holds  $X_t^{(n)} \rightarrow X_t$  in law for  $n \rightarrow \infty$ . This type of continuous dependence on the coefficients is well-known for the case when  $\phi$  is locally Lipschitz continuous and satisfies a linear growth condition, cf. (Jacod and Shiryaev 2013, Thm. IX.6.9). Unfortunately, these conditions are not necessarily fulfilled in our case, as we require merely piecewise Lipschitz continuity for  $\phi$ . In the following,  $\mathbf{p} \in \mathcal{P}$  denotes a partition w.r.t. which  $\phi$  is piecewise Lipschitz continuous.

Fix  $\omega \in \Omega$  and  $T > 0$ , denote  $a := X_0(\omega)$ , and set

$$t_1 := \inf\{t \geq 0 : \Delta X_t(\omega) \neq 0 \text{ or } X_t(\omega) \in \mathbf{p}\}.$$

Further, for  $t \geq 0$  denote by  $q(t) := X_t(\omega)$  and  $q_n(t) := X_t^{(n)}(\omega)$  the paths of the respective processes. On the interval  $[0, t_1)$ ,  $q$  and  $q_n$  are governed by the autonomous integral equations

$$q(t) = a + \int_0^t \phi(q(s))ds$$

and

$$q_n(t) = a + \int_0^t \phi_n(q_n(s))ds,$$

respectively. Hence, for all  $t \in [0, t_1)$

$$\begin{aligned} |q(t) - q_n(t)| &\leq \int_0^t |\phi(q(s)) - \phi_n(q_n(s))|ds \\ &\leq \int_0^t |\phi(q(s)) - \phi(q_n(s))| + |\phi(q_n(s)) - \phi_n(q_n(s))|ds \\ &\leq \int_0^t \ell |q(s) - q_n(s)|ds + c\mu((0, 1/n] \cup [n, \infty)) \end{aligned}$$

for some constants  $\ell, c > 0$ . This is due to the estimate in (4.21), the fact that  $q$  is strictly decreasing, and to the piecewise Lipschitz continuity of  $\phi$ . For the latter we note that  $\phi \leq \phi_n$  which implies  $q(t) \leq q_n(t)$  for all  $t \in [0, t_1)$ . In other words, if  $q$  reaches a discontinuity of  $\phi$  at  $t_1$ , i.e.  $q(t_1) \in \mathbf{p}$ , then it reaches it ahead of  $q_n$  which allows the estimate above.

Grönwall’s inequality (cf. (Hale 1980, Cor. I.6.6)) then yields for all  $t \in [0, t_1)$

$$|q(t) - q_n(t)| \leq c\mu((0, 1/n] \cup [n, \infty)) \left(1 + \int_0^t e^{\ell(t-s)} ds\right) \tag{4.23}$$

which vanishes for  $n \rightarrow \infty$ .

Our strategy is now to iterate this step until we surpass the time  $T$ . By design there are two cases:  $q$  either jumps at  $t_1$  or hits a discontinuity of  $\phi$ . Note that it can be ruled out that both events occur at the same time as the probability of this happening is zero. In the same way we exclude  $q$  jumping onto a discontinuity of  $\phi$  because  $q$  is strictly decreasing, jumps are space homogeneous, and the set of discontinuities of  $\phi$  has no accumulation points in  $(0, \infty)$ .

The first case, i.e.  $q$  jumps at  $t_1$ , is simple. Clearly, there exists  $N \in \mathbb{N}$  such that for all  $n > N$  it holds  $\Delta L_{t_1} = \Delta L_{t_1}^{(n)}$ . Consequently, for all  $\varepsilon > 0$  there exists  $N' \in \mathbb{N}$  such that for all  $n > N'$  it holds  $|q(t_1) - q_n(t_1)| < \varepsilon$ . Choosing  $\varepsilon$  small enough ensures that  $q$  and  $q_n$  both jump into the same

interval  $(x_i, x_{i+1})$  of the partition w.r.t. which  $\phi$  is piecewise Lipschitz continuous. Let

$$t_2 := \inf\{t \geq t_1 : \Delta X_t(\omega) \neq 0 \text{ or } X_t(\omega) \in \mathbf{p}\}. \tag{4.24}$$

For large enough  $n \in \mathbb{N}$  we obtain

$$|q(t) - q_n(t)| \leq \varepsilon + \int_{t_1}^{t_2} \ell |q(s) - q_n(s)| ds + c\mu((0, 1/n] \cup [n, \infty)).$$

for all  $t \in [t_1, t_2)$ , and some (possibly different) constants  $\ell, c > 0$ . Applying Grönwall’s inequality again concludes this iteration step.

For the second case, i.e. if  $q$  hits a discontinuity of  $\phi$  at  $t_1$ , we argue differently. We denote by

$$t_2(n) := \inf\{t \geq t_1 : q_n(t) = q(t_1)\}$$

the time at which  $q_n$  also reaches the discontinuity of  $\phi$  at  $q(t_1)$ . We require some observations: First,  $t_2(n) < \infty$  for all  $n \in \mathbb{N}$  large enough since  $\phi_n < 0$  is bounded away from zero on compact sets  $K \subset (0, \infty)$ . Second,  $q_n(t_1) \rightarrow q(t_1)$  for  $n \rightarrow \infty$  due to (4.23) and the fact that in this case  $q$  and  $q_n$  are continuous on  $[0, t_1]$ . Third, from  $\phi_n \leq \phi_{n-1} < 0$  for all  $n$  large enough it follows that  $q'_n \leq q'_{n-1} < 0$  on  $[t_1, t_2(n)]$ .

Thus,  $t_2(n) \rightarrow t_1$  for  $n \rightarrow \infty$ . Further, also by the continuity of  $q$ , it holds  $q(t_2(n)) \rightarrow q(t_1)$  for  $n \rightarrow \infty$ . Hence, for every  $\varepsilon > 0$  we can choose  $N \in \mathbb{N}$  such that for all  $n > N$

$$|q(t_2(n)) - q_n(t_2(n))| = |q(t_2(n)) - q(t_1)| < \varepsilon.$$

Using the above definition (4.24) of  $t_2$  we obtain for all  $t \in [t_2(n), t_2)$

$$|q(t) - q_n(t)| \leq \varepsilon' + \int_{t_2(n)}^{t_2} \ell |q(s) - q_n(s)| ds + c\mu((0, 1/n] \cup [n, \infty)).$$

Note that on  $[t_2(n), t_2)$  both  $q$  and  $q_n$  act on the same interval  $(x_i, x_{i+1})$  of the partition  $\mathbf{p}$  w.r.t. which  $\phi$  is piecewise Lipschitz continuous. Grönwall’s inequality then shows that for all  $t \in [t_2(n), t_2)$  it holds  $q_n(t) \rightarrow q(t)$  for  $n \rightarrow \infty$ . But because  $q_n$  is continuous on  $[0, t_2)$  and  $t_2(n) \rightarrow t_1$  for  $n \rightarrow \infty$ , this property extends to  $[t_1, t_2)$ .

Finally, iteration and the fact that  $T$  and  $\omega$  have been chosen arbitrarily yields  $X_t^{(n)} \rightarrow X_t$  almost surely for  $n \rightarrow \infty$  and all  $t \geq 0$ . This implies weak convergence, and therefore,  $X_t \sim \pi$  for all  $t \geq 0$ , i.e.  $\pi$  is invariant for  $(X_t)_{t \geq 0}$ . By Theorem 3.4 (i),  $(X_t)_{t \geq 0}$  is positive Harris recurrent, and hence,  $\pi$  is the unique invariant distribution of  $(X_t)_{t \geq 0}$ .  $\square$

## 5 Outlook

It is only natural to try extending Theorem 3.1 and Theorem 3.4 to more general settings, e.g. higher dimensions, more complicated driving noises, or target measures with disconnected supports, heavy tails or atoms. We believe that for many of these cases similar methods to the ones we employed here yield similar results. Generally speaking, one only has to show that the process in question is positive Harris recurrent, and that there exists a unique infinitesimally invariant distribution. The remaining steps are in most instances trivial or at least easy to prove under mild conditions.

In the following we shortly comment on the problems one faces when trying this approach on some of the more general cases.

**Jump measures with heavy tails** The restriction to light tailed jumps, that is when  $\mathbb{E}\xi_1 < \infty$ , is only needed in the proof of Theorem 3.4 (i). It is not entirely clear whether additional conditions are necessary to show positive Harris recurrence in the case of heavy tailed jumps, i.e.  $\mathbb{E}\xi_1 = \infty$ . Yet, as heavy tailed jumps imply larger (negative) values of the drift coefficient  $\phi$  by the definition in (3.7) searching a more sophisticated norm-like function is the most promising approach.

**Subordinators as driving noise** In Assumption (a1) we required  $(L_t)_{t \geq 0}$  to be a spectrally positive compound Poisson process if  $\mathcal{E} = (0, \infty)$ . In Theorem 3.1 one might also wish to allow (pure jump) subordinators for  $(L_t)_{t \geq 0}$ , i.e. Lévy processes with characteristic triplet  $(0, 0, \mu)$  for which  $\text{supp } \mu \subset \mathbb{R}_+$  and  $\int_{(0, \infty)} (1 \wedge z)\mu(dz) < \infty$ .

However, in this case (3.5) in the proof of Theorem 3.1 does in general not hold. Thus, one needs to find a different way of showing that  $0 \notin \mathcal{O}$ , e.g. by proving that  $(\ln(X_t))_{t \geq 0}$  does not explode. Moreover, uniqueness of the infinitesimally invariant distribution, that is Theorem 3.4 (iii), has to be shown differently. This is because there exists no subinterval of  $(0, \infty)$  that cannot be reached by jumps, and because a non-zero amount of jumps does occur almost surely during every time interval. Thus, the proofs of Theorem 3.4 (iii) under (c1) and (c2), respectively, do not apply in this case.

**Target distributions with full support** One might wish to extend the results of Theorem 3.4 for the case when  $\mathcal{E} = \mathbb{R}$ . However, just like in the previous paragraph, the approach used for the proof of Theorem 3.4 (iii) fails due to the fact that there exists no subinterval of  $\mathbb{R}$  that cannot be reached by jumps. Therefore, the proof of the uniqueness of the solution of (3.6) requires different arguments.

**Target measures with disconnected supports** Allowing only  $\mathcal{E} = \mathbb{R}$  or  $\mathcal{E} = (0, \infty)$  seems restrictive as the ability to cross gaps is one of the main advantages of the presence of jumps. Intending to allow disconnected supports of the target measure  $\pi$  one has to assume three things:

- (i)  $\mathcal{E}$  is an open set,
- (ii) jumps in both directions are possible, i.e.  $\Pi((-\infty, 0)) > 0$  and  $\Pi((0, \infty)) > 0$ ,
- (iii) jumps can only land in  $\mathcal{E}$ , i.e.  $\mathcal{E} + \text{supp } \Pi \subseteq \mathcal{E}$ .

With those three assumptions one can show that, apart from  $\mathcal{E} = \mathbb{R}$  and  $\mathcal{E}$  being some half-line, the only option is that  $\mathcal{E}$  is periodic, that is there exists  $p > 0$  such that  $\mathcal{E} + p = \mathcal{E}$ .

However, if  $\mathcal{E} \neq \mathbb{R}$  is periodic and  $(X_t)_{t \geq 0}$  with state space  $\mathcal{O} = \mathcal{E}$  solves (1.3), then  $(X_t)_{t \geq 0}$  cannot be positive Harris recurrent—regardless of the drift coefficient  $\phi$ . The reason for this is simple: The jumps of  $(X_t)_{t \geq 0}$  are space-homogeneous, and  $\mathcal{E}$  consists of countably many intervals of the same length that can only be connected by jumps. Thus, the mass of an invariant measure concentrated on each of these segments is the same. Therefore, no invariant measure can be finite (apart from the trivial measure).

**Target measures with atoms** An invariant measure  $\pi$  with  $\pi(\{x_0\}) > 0$  for one or more  $x_0 \in \mathbb{R}$  can only be achieved by a solution  $(X_t)_{t \geq 0}$  of (1.3) if  $(X_t)_{t \geq 0}$  comes to a halt at  $x_0$ . One possible solution might be to set  $\phi(x_0) = 0$ . At least heuristically this makes sense considering that the denominator in the original definition (3.4) of  $\phi$  is the density function of  $\pi$ .

However, extending Theorem 3.4 to this case needs a new idea since the current proof relies on the fact that any solution of (3.6) can be associated to a locally integrable function.

**Target measures with arbitrary tails** By Assumption (b1), we require  $\pi$  to have an exponential tail. This is mostly needed in the proof of Theorem 3.4 (i). As with heavy tailed jumps, using a more sophisticated norm-like function will most likely enable us to consider target measures  $\pi$  for which only  $|\pi(x)| \leq ce^{-\alpha x}$  for all  $x \gg 1$  and some constants  $c, \alpha > 0$ .

In case  $\pi$  has a heavy tail, that is when  $|\pi(x)| \geq cx^{-(1+\alpha)}$  for all  $x \gg 1$  and some constants  $c, \alpha > 0$ , it is not clear whether  $(X_t)_{t \geq 0}$  is positive Harris recurrent or not.

**Higher dimensions** Theorem 3.1 can be extended easily to target measures  $\pi$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  and  $d$ -dimensional driving noises  $(L_t)_{t \geq 0}$  with  $d \geq 2$ . Simply use the multi-dimensional counterparts (cf. Behme et al. 2022) to all occurring terms in the definition (3.4) of the drift coefficient, but note that it is essential to make sure that jumps can only land in  $\mathcal{E}$ , and that  $(X_t)_{t \geq 0}$  cannot drift onto  $\partial\mathcal{E}$ .

However, Eq. (3.6) becomes a partial differential equation in the multi-dimensional case. Thus, Theorem 3.4 cannot be extended with the same approach.

**Lévy-type driving noise** A solution to some of the problems mentioned above, e.g. disconnected supports or atoms, might be to select space dependent driving noises. For the case when  $(L_t)_{t \geq 0}$  is a Lévy-type process (for details see Böttcher et al. 2013) (Behme et al. 2022) provides the required framework for defining the drift coefficient. Just

like with higher dimensions extending Theorem 3.1 to this setting is feasible while the extension of Theorem 3.4 might require a new approach.

**Acknowledgements** I would like to thank Anita Behme for her advice and I am very grateful for her helpful comments and suggestions. Further, I wish to express my gratitude towards the anonymous referees for their valuable recommendations.

**Author Contributions** D.O. wrote the manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Bardenet, R., Doucet, A., Holmes, C.C.: On Markov chain Monte Carlo methods for tall data. *J. Mach. Learn. Res.*, **18**(47), (2017)
- Behme, A., Oechsler, D.: Invariant measures of Lévy-type operators and their associated Markov processes. Preprint on (2022). [arXiv:2208.07668](https://arxiv.org/abs/2208.07668)
- Böttcher, B., Schilling, R., Wang, J.: Lévy-Type Processes: Construction, Approximation and Sample Path Properties. In: Barndorff-Nielsen, O.E., Bertoin, J., Jacod, J., Klüppelberg, C. (eds.) *Lévy Matters III. Lecture Notes in Mathematics*, vol. 2099. Springer, New York (2013)
- Brooks, S., Gelman, A., Jones, G., Meng, X.-L.: *Handbook of Markov Chain Monte Carlo*. CRC Press, Boca Raton (2011)
- Eliazar, I., Klafter, J.: Lévy-driven Langevin systems: targeted stochasticity. *J. Stat. Phys.* **111**(3), 739–768 (2003)
- Hale, J.K.: *Ordinary Differential Equations*, 2nd edn. Robert E. Krieger Publishing Company (1980)
- Huang, L.-J., Majka, M.B., Wang, J.: Approximation of heavy-tailed distributions via stable-driven SDEs. *Bernoulli* **27**(3), 2040–2068 (2021)
- Jacod, J., Shiryaev, A.: *Limit Theorems for Stochastic Processes*, vol. 288. Springer, New York (2013)
- Kallenberg, O.: *Foundations of Modern Probability*. Springer, New York (1997)
- Kendall, W.S., Liang, F., Wang, J.-S.: *Markov Chain Monte Carlo: Innovations and Applications*, vol. 7. World Scientific, Singapore (2005)

- Kühn, F.: Solutions of Lévy-driven SDEs with unbounded coefficients as Feller processes. *Proc. Am. Math. Soc.* **146**(8), 3591–3604 (2018)
- Liggett, T.M.: *Continuous Time Markov Processes: An Introduction*. American Mathematical Soc, New York (2010)
- Meyn, S.P., Tweedie, R.L.: Stability of Markovian processes I: Criteria for discrete-time chains. *Adv. Appl. Prob.* **24**(3), 542–574 (1992)
- Meyn, S.P., Tweedie, R.L.: Stability of Markovian processes II: Continuous-time processes and sampled chains. *Adv. Appl. Prob.* **25**(3), 487–517 (1993)
- Meyn, S.P., Tweedie, R.L.: Stability of Markovian processes III: Foster–Lyapunov criteria for continuous time processes. *Adv. Appl. Prob.* **25**(3), 518–548 (1993)
- Nguyen, T.H., Simsekli, U., Richard, G.: Non-asymptotic analysis of Fractional Langevin Monte Carlo for non-convex optimization. In: *International Conference on Machine Learning*, pp. 4810–4819. PMLR (2019)
- Priola, E., Shirikyan, A., Xu, L., Zabczyk, J.: Exponential ergodicity and regularity for equations with Lévy noise. *Stoch. Process. Their Appl.* **122**(1), 106–133 (2012)
- Protter, P.E.: *Stochastic Integration and Differential Equations*, 2nd edn. Springer, New York (2004)
- Roberts, G.O., Tweedie, R.L.: Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2**(4), 341–363 (1996)
- Schertzer, D., Larchevêque, M., Duan, J., Yanovsky, V., Lovejoy, S.: Fractional Fokker-Planck equation for nonlinear stochastic differential equations driven by non-Gaussian Lévy stable noises. *J. Math. Phys.* **42**(1), 200–212 (2001)
- Schnurr, A.: *The symbol of a Markov semimartingale*. PhD thesis, TU Dresden (2009)
- Şimşekli, U.: *Fractional Langevin Monte Carlo: Exploring Lévy driven stochastic differential equations for Markov Chain Monte Carlo*. In: *International Conference on Machine Learning*, pp. 3200–3209. PMLR (2017)
- Welling, M., Teh, Y. W.: Bayesian learning via stochastic gradient Langevin dynamics. In: *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 681–688 (2011)
- Xu, L., Zegarliński, B.: Ergodicity of the finite and infinite dimensional  $\alpha$ -stable systems. *Stoch. Anal. Appl.* **27**(4), 797–824 (2009)
- Zhang, Xialong, Zhang, Xicheng: Ergodicity of supercritical SDEs driven by  $\alpha$ -stable processes and heavy-tailed sampling. *Bernoulli* **29**(3), 1933–1958 (2023)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.