Jan Gairing | Michael Högele | Tetiana Kosenkova

Transportation Distances and Noise Sensitivity of Multiplicative Lévy SDE with Applications
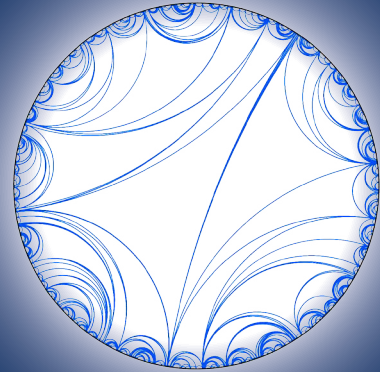
Preprints des Instituts für Mathematik der Universität Potsdam

Jan Gairing | Michael Högele | Tetiana Kosenkova

Transportation Distances and Noise Sensitivity of Multiplicative Lévy SDE with Applications

Transportation distances and noise sensitivity
of multiplicative Lévy SDE with applications

Jan Gairing [*]   Michael Högele[†]   Tetiana Kosenkova [‡]

November 23, 2015

Abstract

This article assesses the distance between the laws of stochastic differential equations with multiplicative Lévy noise on path space in terms of their characteristics. The notion of transportation distance on the set of Lévy kernels introduced by Kosenkova and Kulik yields a natural and statistically tractable upper bound on the noise sensitivity. This extends recent results for the additive case in terms of coupling distances to the multiplicative case. The strength of this notion is shown in a statistical implementation for simulations and the example of a benchmark time series in paleoclimate.

MSC 2010:   60G51; 60G52; 60J75; 62M10; 62P12;

Keywords: stochastic differential equations; multiplicative Lévy noise; Lévy type processes; heavy-tailed distributions; model selection; Wasserstein distance; time series;

1   Introduction

Many dynamical phenomena are subject to random forcing, often described by stochastic differential equations of the following type

$$dX(t) = -\nabla U(X(t))dt + d\xi(t), \qquad X_0 = x_0, \tag{1.1}$$

where $-\nabla U$ encodes the deterministic dynamics given by a potential gradient and ξ a noise signal. In general, for instance when ξ exhibits discontinuities, it is not straight-forward to describe the law of the solution X on path space in terms of the parameters which determine the distribution of ξ. Our approach allows to quantify the distance of such laws in terms of accessible quantities, both analytically and statistically.

The case where ξ is given as a discontinuous Lévy process was studied in a previous publication [7]. The authors introduced the notion of a coupling distance between Lévy measures in order to quantify the Wasserstein distance on path space. The coupling distances have been found to be sufficiently strong (in a topological sense) to quantify the convergence in functional limit theorems, yet being weak enough in order to be numerically and statistically tractable. See for instance the calibration problem of a climate time series in [8]. In many situations, however, the noise process ξ exhibits state dependence, for instance multiplicative noise. This generalization lifts Lévy diffusions to Lévy-type diffusions. To treat this class of noise processes the authors introduced in [21]

the notion of transportation distance extending the coupling distances with the help of a common reference Lévy measure. The present article establishes analogous bounds on the distance between the laws of Lévy-type diffusions on path space in terms of transportation distances.

We stress that our procedure is suitable for a large variety of phenomena modelled with jump diffusions, such as in finance (e.g. [24]) or neurosciences [3, 13, 27]. The particular application we have in mind in this article is the refinement of the analysis of the noise structure behind the paleoclimate temperature evolution studied in [8]. The climate data apparently fluctuate around two distinct metastable states with rapid transitions (see Fig. 2). Such phenomena are observed in stochastic energy balance models [1, 2, 4, 14, 16, 17].

There is a list of publications associating this time series to an underlying jump diffusion (see for instance [8, 6, 9, 11]). Using various techniques these articles aim to determine the (polynomial) jump behavior of (1.1) for different classes of heavy-tailed Lévy processes ξ. The models so far require the spacial homogeneity of the noise characteristics. The present article lifts this restriction. We may now investigate the statical behavior in the different spatial regimes, prescribed by the metastable states, and solve the corresponding model selection problem on the generic class of heavy-tailed jump diffusions. Our results of the implementation of this program applied to the mentioned climate times series are consistent with the findings in [8].

2 Preliminaries

Transportation functions and transportation distance: Consider a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \geqslant 0})$ satisfying the usual conditions in the sense of Protter [23] carrying a scalar Brownian motion $(B_t)_{t \in [0,T]}$ and an independent Cauchy Poisson random measure ν_0 on $[0, T] \times \mathbb{R}$ with intensity measure $dt \otimes \Pi_0$ given by

$$\Pi_0(dv) = \mathbf{1}_{\{v \in \mathbb{R} \setminus \{0\}\}} \frac{dv}{v^2}.$$

We define a Lévy measure to be a σ-finite Borel measure on $\mathbb{R} \setminus \{0\}$ satisfying

$$\int_{\mathbb{R}} (|v|^2 \wedge 1) \Pi(dv) < \infty. \tag{2.1}$$

In contrast to the standard definition we do not exclude point-mass in 0 which may be taken to be infinity. Nevertheless we will identify all such measures that coincide on the Borel σ-algebra $\mathcal{B}(\mathbb{R} \setminus \{0\})$ and the standard Lévy measures (without mass in 0) are the canonical representatives of those equivalence classes (for details see [21]). The key of our analysis is the observation that all standard Lévy measures with infinite mass admit a representation as a measure transform of a common reference Lévy measure. This reference measure will be chosen to be the standard symmetric Cauchy measure. In order to treat finite Lévy measures we may artificially assign an infinite point-mass to 0. The precise statement is given as follows.

Proposition 2.1. *For any standard Lévy measure Π (that is $\Pi(\{0\}) = 0$) there exists a unique Lévy measure $\tilde{\Pi}$ in the sense of (2.1) defined as $\tilde{\Pi} = \Pi$ if $\Pi(\mathbb{R}) = \infty$ and $\tilde{\Pi} = \Pi + \infty \delta_0$ else, and a unique transportation function $c : \mathbb{R} \to \mathbb{R}$ satisfying*

   1. c is non-decreasing,

   2. $c((-\infty, 0]) \subset (-\infty, 0]$ and $c([0, \infty)) \subset [0, \infty)$,

   3. c is left continuous on $(-\infty, 0]$ and right continuous on $[0, \infty)$,

such that $\tilde{\Pi} = \Pi_0 \circ c^{-1}$.

A proof is found in [21]. For the sake of readability and due to uniqueness we no longer distinguish between $\tilde{\Pi}$ and Π.

Example 2.2. 1. For the one-sided Cauchy measure $\Pi(dv) = \mathbf{1}_{\{v>0\}}\frac{dv}{v^2}$ we have the transportion function $c(v) = \mathbf{1}_{\{v>0\}}v$. Note that c maps the infinite mass of Π_0 on $(-\infty, 0]$ to the point 0. The image measure $\tilde{\Pi}(dv) = \mathbf{1}_{\{v>0\}}\frac{dv}{v^2} + \infty \cdot \delta_0$ coincides with Π outside 0.

2. The next example will be exploited extensively in the applications of Section 4. For $\alpha, \varepsilon, \lambda > 0$ consider the Pareto-type power tail $\Pi_{\alpha,\varepsilon,\lambda}(dv) = \mathbf{1}_{\{v>\varepsilon\}}\frac{\lambda dv}{v^{\alpha+1}}$. It is easy to verify that in this case the transportation function is of the form

$$c(v) = \mathbf{1}_{\{v>\frac{\alpha\varepsilon^\alpha}{\lambda}\}}\left(\frac{\lambda v}{\alpha}\right)^{1/\alpha}.$$

Note that the positive span of the family $\{\Pi_{\alpha,\varepsilon,\lambda} : \alpha, \varepsilon, \lambda \in \mathbb{Q}_+\}$ is dense in the class of heavy tailed Lévy measures on the positive half-line.

Example 2.3. Let $(\xi_i)_{i\in\mathbb{N}}$ be a sequence of real valued random variables with values in $[\varepsilon, \infty)$ for some $\varepsilon > 0$. The respective *empirical measure* is given by $\Pi_n([a,b]) = \frac{\#\{\xi_i \in (a,b], i \leqslant n\}}{n}$ on intervals $(a,b]$. We denote the n-th *order statistic* by $\varepsilon < \xi_{1:n} \leqslant \xi_{2:n} \leqslant \cdots \leqslant \xi_{n:n}$. In each interval $(\xi_{i:n}, \xi_{i+1:n}]$ we have exactly one data point with individual mass $1/n$ such that

$$\Pi_n((\xi_{i:n}, \xi_{i+1:n}]) = \frac{1}{n} = \int\limits_{c^{-1}(\xi_{i:n})}^{c^{-1}(\xi_{i+1:n})} \frac{dv}{v^2} = \frac{1}{c^{-1}(\xi_{i:n})} - \frac{1}{c^{-1}(\xi_{i+1:n})},$$

which yields

$$c^{-1}(v) = \begin{cases} 0 & v \in [0,\varepsilon) \\ 1 & v \in [\varepsilon, \xi_{1:n}) \\ \frac{n}{(n-i+1)} & v \in [\xi_{i-1:n}, \xi_{i:n}) \qquad \text{for all } i \in \{2,n\} \\ \infty & v \geqslant \xi_{n:n} . \end{cases}$$

Hence the inverse is given by

$$c(v) = \begin{cases} 0 & v \in [0,1) \\ \xi_{i:n} & v \in [\frac{n}{(n-i+1)}, \frac{n}{(n-i)}) \qquad \text{for all } i \in \{1, n-1\} \\ \xi_{n:n} & v \geqslant n. \end{cases}$$

We take advantage of the measure transform of Proposition 2.1 to compare two given Lévy measures in terms of a truncated $L^p(\Pi_0)$ distance of the respective transportation functions.

Remark 2.4. We stress that Proposition 2.1 guarantees that the previously chosen probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\geqslant 0})$ is rich enough to carry any Poisson random measures ν (in distribution) with respect to the intensity measure $dt \otimes \Pi$, where Π is a (standard) Lévy measure via

$$\nu((s,t] \times A) = \nu_0((s,t] \times c^{-1}(A)), \qquad A \in \mathcal{B}(\mathbb{R} \setminus \{0\}), \quad 0 \leqslant s \leqslant t.$$

3

Definition 2.5. Define the class

$$\mathcal{L}_\rho^p := \{\Pi \text{ (standard) Lévy measure} \mid \int_{\mathbb{R}} \rho^p(0, v)\Pi(dv) < \infty\},$$

where $p \geqslant 1$ and $\rho(x, y) := |x - y| \wedge 1$ is the truncated Euclidean distance. For Lévy measures $\Pi_i \in \mathcal{L}_\rho^p$, $i = 1, 2$ with respective representations $\Pi_i = \Pi_0 \circ c_i^{-1}$ according to Proposition 2.1 we define the *transportation distance* of order p by

$$\mathcal{T}_p(\Pi_1, \Pi_2) := \left(\int_{\mathbb{R}} \rho^p\big(c_1(u), c_2(u)\big) \Pi_0(dv) \right)^{1/p}.$$

It has been established in [21] that the transportation distance metrizes the positive cone $\mathcal{L}_\rho^p$.

Remark 2.6. In the case of proper Lévy-type mesures with state dependence $\Pi(x, dv)$ Proposition 2.1 yields a family of transportation functions $(c(x, \cdot))_{x \in \mathbb{R}}$.

Stochastic differential equations: For any Poisson random measure ν on $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \geqslant 0})$ (cf. Remark 2.4) independent of W with intensity measure $dt \otimes \Pi$, Π a Lévy measure, and $a, b \in \mathbb{R}$ there is a Lévy process Z given in terms of its Lévy-Itô decomposition

$$Z(t) = at + bW(t) + \iint_{(0,t] \times [-1,1]} v\tilde{\nu}(dv, dt) + \iint_{(0,t] \times \mathbb{R} \setminus [-1,1]} v\nu(dv, dt). \tag{2.2}$$

As usual $\tilde{\nu} = \nu - dt \otimes \Pi$ denotes the compensated Poisson random measure of ν on $\mathcal{B}(\mathbb{R} \setminus \{0\})$. Recall that the law of Z is characterized by its *Lévy triplet* (a, b, Π). For further details we refer to [26]. Consider now the following formal stochastic differential equation

$$dX(t) = a(X(t))dt + b(X(t))dW(t)$$
$$+ \int_{|v| \leqslant 1} v\, \tilde{\nu}(X(t-), dt, dv) + \int_{|u| > 1} v\, \nu(X(t-), dt, dv). \tag{2.3}$$

Here ν (resp. $\tilde{\nu}$) is interpreted as a space dependent (compensated) Poisson random measure explained below. A solution X of such an equation is understood as the solution to the corresponding martingale problem for the following integro-differential operator A acting on $\varphi \in \mathcal{C}_c^2(\mathbb{R})$

$$A[\varphi](x) = a(x)\varphi'(x) + b^2(x)\varphi''(x) + \int_{\mathbb{R} \setminus \{0\}} \big(\varphi(x + v) - \varphi(x) - \varphi'(x)v\mathbf{1}_{\{|v| \leqslant 1\}}\big)\Pi(x, dv),$$

where $a, b : \mathbb{R} \to \mathbb{R}$ and $x \mapsto \Pi(x, \cdot)$ is a Lévy kernel, which associates to each $x \in \mathbb{R}$ the Lévy measure $\Pi(x, \cdot)$. We may rewrite A in terms of the Lipschitz continuous cutoff function $\tau(u) := \operatorname{sign}(u)(|u| \wedge 1)$ and $\bar{a}(x) = \Pi(x, \{u \in \mathbb{R} \mid |u| > 1\})$ as

$$A[\varphi](x) = (a(x) + \bar{a}(x))\varphi'(x) + b^2(x)\varphi''(x) + \int_{\mathbb{R} \setminus \{0\}} \big(\varphi(x + v) - \varphi(x) - \varphi'(x)\tau(v)\big)\Pi(x, dv). \tag{2.4}$$

Proposition 2.1 allows us to represent $\Pi(x, \cdot)$ as $\Pi_0 \circ c^{-1}(x, \cdot)$ in terms of a family of tansport functions $(c(x, \cdot))_{x \in \mathbb{R}}$ with respect to the Cauchy reference measure Π_0. For Lévy type processes we refer to [10] and [5].

In [21] it is shown that under the following Lipschitz and boundedness conditions on the space-dependent coefficients (a, b, Π) for any $x, y \in \mathbb{R}$

$$|a(x) - a(y)| \leqslant L_a \, |x - y|, \tag{2.5}$$

$$(b(x) - b(y))^2 \leqslant L_b \, \rho^2(x, y), \tag{2.6}$$

$$|\bar{a}(x) - \bar{a}(y)| \leqslant L_{\bar{a}} \rho(x, y), \tag{2.7}$$

$$\mathcal{T}_2^2(\Pi(x, \cdot), \Pi(y, \cdot)) \leqslant L_\Pi \, \rho^2(x, y), \tag{2.8}$$

and initial value $x_0 \in \mathbb{R}$ there exists a unique strong solution to the martingale problem associated to (2.4). On the probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \geqslant 0})$ this is given as a strong solution of the following SDE

$$dX(t) = \big(a(X(t)) + \bar{a}(X(t))\big)dt + b(X(t))dW(t)$$
$$+ \int_{\mathbb{R}} c(X(t-), v) \left[\nu_0(dt, dv) - \frac{\tau(c(X(t-), v))}{c(X(t-), v)} \Pi_0(dv)dt \right], \tag{2.9}$$

$$X(0) = x_0. \tag{2.10}$$

Under these assumptions Remark 2.4 provides a strong solution of (2.3).

The main purpose of this construction is to compare the laws of two diffusions X_i, $i = 1, 2$ on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$ given as strong solutions (2.9) with identical Brownian motion W and Cauchy Poisson random measure ν_0, whose respective coefficients (a_i, b_i, Π_i), $i = 1, 2$ satisfy (2.5), (2.6), (2.7) and (2.8). This is carried out in Theorem 3.5. In a simplified setting more suitable for the applications we have in mind this is done in Theorem 3.6.

Remark 2.7. In the case of pure jump diffusions with finite intensity, that is Lévy-type diffusions with triplet of characteristics $(a, 0, \Pi)$, where Π is a finite measure, it is obvious that we only need the Lipschitz continuity of a (2.5). In this case we consider equation

$$dX(t) = a(X(t))dt + \int_{\mathbb{R}} c(X(t-), v)\nu_0(dt, dv), \tag{2.11}$$

$$X(0) = x_0. \tag{2.12}$$

3 Main results

In this section we will first compare the laws of two diffusions in the sense of (2.9) in terms of $\mathcal{T}_2$. In a second part this is carried out in terms of $\mathcal{T}_1$ motivated by applications elaborated in Section 4.

3.1 Noise sensitivity estimates in terms of $\mathcal{T}_2$

We are interested in the sensititivity of the laws of solutions of (2.9) with respect to their parameters. The following theorem provides a quantitative estimate in terms of $\mathcal{T}_2$.

Theorem 3.1. *Let X_i, $i = 1, 2$ on $(\Omega, \mathcal{F}, \mathbb{P})$ be strong solutions of (2.9) with respective initial conditions $x_i \in \mathbb{R}$ and triplets of characteristics (a_i, b_i, Π_i) satisfying the Lipschitz conditions (2.5-2.8) for common constants $L_a, L_b, L_{\bar{a}}, L_\Pi > 0$ and $\|a_1 - a_2\|_\infty < \infty$. Then for any $T > 0$ there is constant $K > 0$ such that for $G(x) = \max\{\sqrt{x}, x\}, x \geqslant 0$ we have*

$$\mathbb{E} \sup_{t \in [0,T]} \rho^2(X_1(t), X_2(t)) \leqslant KG(\Delta), \tag{3.1}$$

where

$$\Delta = \rho(x_1, x_2) + \|a_1 - a_2\|_\infty^2 + \|\bar{a}_1 - \bar{a}_2\|_\infty^2 + \|b_1 - b_2\|_\infty^2$$

$$+ \sup_{x \in \mathbb{R}} \mathcal{T}_2(\Pi_1(x, \cdot), \Pi_2(x, \cdot)) + \sup_{x \in \mathbb{R}} \mathcal{T}_2^2(\Pi_1(x, \cdot), \Pi_2(x, \cdot)).$$

The statement of the theorem also allows to compare laws of solutions of (2.3) which are not necessarily defined on the same probability space. The natural choice of a metric between such laws is the (truncated) Wasserstein distance on path space, for details see [7] and for general reference [25].

Corollary 3.2. *Under the assumptions of Theorem 3.1 the strong solutions X_i, $i = 1, 2$ to (2.9) form a specific coupling of solutions to (2.3). Hence the bound (3.1) gives an upper bound of the Wasserstein distance of order 2 on the path space of càdlàg paths endowed with the (truncated) supremum norm. The Wasserstein distance is precisely defined as as the infimum of the left-hand side of (3.1) where the pair (X_1, X_2) ranges over all couplings of X_1 with X_2.*

The following example shows that the Lipschitz conditions (2.5-2.8) are not very restrictive in the class of Lévy-type kernels with exponential moments.

Example 3.3. An important example is given by the Gamma-type process, that is a Lévy-type process with triplet of characteristics $(0, 0, \Pi)$ and initial value $x_0 = 0$. The Lévy kernel is given as $\Pi(x, dv) = \mathbf{1}_{\{v>0\}}\lambda(x)(e^{-\gamma(x)v}/v)dv$ with bounded and Lipschitz continuous coefficients $\lambda, \gamma : [0, \infty) \to [0, \infty)$ satisfying $\sup_{x \in \mathbb{R}} \lambda(x) > 0$ and $\sup_{x \in \mathbb{R}} \gamma(x) > 0$. Indeed, Proposition 3 in [21] states the following. Let Π_j, $j = 1, 2$ be two Gamma Lévy measures $\Pi_j(dv) = \mathbf{1}_{\{v>0\}}(\gamma_j e^{-\lambda_j v}/v)dv$, $j = 1, 2$ with constants $\gamma_j, \lambda_j > 0$.

1. For two such Gamma measures Π_j, $j = 1, 2$ with the same parameter λ and different parameters $0 < \gamma_1 < \gamma_2$ there exists a constant $D = D(\lambda) > 0$ such that the following bound holds true

$$\mathcal{T}_2(\Pi_1, \Pi_2) \leqslant D(\gamma_2 - \gamma_1). \tag{3.2}$$

2. For two such Gamma measures Π_j, $j = 1, 2$ with the same parameter γ and different parameters $0 < \lambda_1 < \lambda_2$ there exists a constant $\tilde{D} = \tilde{D}(\gamma, \lambda_1, \lambda_2) > 0$ locally bounded in λ_1 around λ_2 such that

$$\mathcal{T}_2(\Pi_1, \Pi_2) \leqslant \tilde{D}(\lambda_2 - \lambda_1). \tag{3.3}$$

These results immediately yield that condition (2.8) is satisfied in either case.

3.2 Noise sensitivity estimates in terms of $\mathcal{T}_1$

In applications, however, the involved Lévy measures are heavy-tailed and do not satisfy (2.8). In [8] for instance, it was obtained that the measures that are likely to describe the noise in the paleoclimatic data are of the following shape

$$\Pi(dv) = \left(\mathbf{1}_{\{v<1\}}\frac{\lambda^-}{|v|^{\alpha^-+1}} + \mathbf{1}_{\{v>1\}}\frac{\lambda^+}{v^{\alpha^++1}}\right)dv, \quad \alpha \geqslant 2. \tag{3.4}$$

Remark 3.4. Such Lévy measures do not satisfy (2.8) due to the following reasoning. Let $\lambda^+ = 1$, $\lambda^- = 0$ and $\alpha_2 > \alpha_1 > 1$ with $|\alpha_2 - \alpha_1|$ small. Then transportation distance satisfies inequalities of the form

$$
\mathcal{T}_2^2(\Pi_1, \Pi_2) = \int_0^\infty \left(\mathbf{1}_{\{v > \alpha_1^{1/\alpha_1}\}} \left(\frac{v}{\alpha_1}\right)^{1/\alpha_1} - \mathbf{1}_{\{v > \alpha_2^{1/\alpha_2}\}} \left(\frac{v}{\alpha_2}\right)^{1/\alpha_2} \right)^2 \wedge 1) \Pi_0(dv)
$$

$$
\geqslant \int_{\alpha_1^{1/\alpha_1}}^{\alpha_2^{1/\alpha_2}} \left(\frac{v}{\alpha_1}\right)^{1/\alpha_1} \frac{1}{v^2} dv \geqslant C_{\alpha_1} |\alpha_2^{1/\alpha_2} - \alpha_1^{1/\alpha_1}| \geqslant \inf_{|x| \leqslant |\alpha_2 - \alpha_1|} g'(x) \, C_{\alpha_1} |\alpha_1 - \alpha_2|,
$$

where $g(x) = x^{1/x}$. Since $g'(\alpha_1)$ is locally bounded from below in a small neighborhood of α_1, this only implies Hölder continuity with respect to $\mathcal{T}_2$.

With this result in mind we switch to the setting of a pure jump diffusions X_j with the respective triplet of characteristics $(a_j, 0, \Pi_j)$ for finite Π_j and initial conditions x_j, satisfying (2.5). For two solutions of such differential equations we can state the following result.

Theorem 3.5. *Let X_i, $i = 1, 2$ on $(\Omega, \mathcal{F}, \mathbb{P})$ be strong solutions of (2.9) with respective initial conditions $x_j \in \mathbb{R}$ and triplet of characteristics $(a_i, 0, \Pi_i)$ satisfying the Lipschitz conditions (2.5) and*

$$
\mathcal{T}_1(\Pi(x, \cdot), \Pi(y, \cdot)) \leqslant L_\Pi \rho(x, y) \qquad \text{for all } x, y \in \mathbb{R}. \tag{3.5}
$$

Then for any $T > 0$ there exists a constant $K > 0$ such that for $G(x) = \max\{\sqrt{x}, x\}, x \geqslant 0$ we have

$$
\mathbb{E} \sup_{t \in [0,T]} \rho(X_1(t), X_2(t)) \leqslant KG\Big(\rho(x_1, x_2) + \sup_{x \in \mathbb{R}} \mathcal{T}_1\big(\Pi_1(x, \cdot), \Pi_2(x, \cdot)\big)\Big). \tag{3.6}
$$

The following proposition establishes that for constants $\lambda^+ \geqslant 0$ and $\lambda^- \geqslant 0$ and a bounded and globally Lipschitz continuous function $\alpha : \mathbb{R} \to [2, \infty)$ the Lévy-type measure $\Pi(x, dv)$

$$
\Pi(x, dv) = \left(\mathbf{1}_{\{v < 1\}} \frac{\lambda^-}{|v|^{\alpha(x)+1}} + \mathbf{1}_{\{v > 1\}} \frac{\lambda^+}{v^{\alpha(x)+1}} \right) dv \tag{3.7}
$$

satisfies the $\mathcal{T}_1$-Lipschitz condition (3.5).

Proposition 3.6. *For Lévy measures Π_i, $i = 1, 2$ defined in (3.4) with $\lambda^+ = \lambda_1^+ = \lambda_2^+$ and $\lambda^- = \lambda_1^- = \lambda_2^-$ and $2 \leqslant \alpha_1^+, \alpha_2^+, \alpha_1^-, \alpha_2^-$ there exists a constant $D = D(\lambda^+, \lambda^-) > 0$ such that*

$$
\mathcal{T}_1(\Pi_1, \Pi_2) \leqslant D\Big(|\alpha_1^+ - \alpha_2^+| + |\alpha_1^- - \alpha_2^-|\Big). \tag{3.8}
$$

All proofs are found in Section 5.

4 Applications

In this section we demonstrate the benefit of the transportation distance in an empirical setting. Assume we are given a data set $\xi_1, \ldots \xi_n$ that we interpret as jumps of a Lévy process with Lévy measure Π. In Example 2.3 we have calculated the transportation kernel c_n of the empirical Lévy measure Π_n. We are particularly interested in modelling with power law tails where the corresponding transportation kernels c_α are given in Example 2.2. We are then in the position to evaluate the transportation distance between the empirical Lévy measures Π_n and such power laws. This is done in example 4.1. In the sequel we analyze the behavior of the transportation distance for simulated power law jumps. Ultimately we utilize this device to propose a simple but state dependent jump diffusion model for a paleoclimatic bench mark time series.

4.1 Analytical considerations

Example 4.1. Let us consider the transportation distance between the empirical measure $\Pi_n((a,b])$ of Example 2.3 and a power law $\Pi_{\alpha,\varepsilon}(vu) = \mathbf{1}_{\{v>\varepsilon\}}\frac{\lambda dv}{v^{\alpha+1}}$ ($\alpha \neq 1$) of Example 2.2. Recall that c_n is piecewise constant. Observe that for $0 < a < b$ and a constant $c > 0$

$$
\int_a^b |c - (\tfrac{\lambda v}{\alpha})^{1/\alpha}|v^{-2}dv = \int_{a\wedge\frac{\alpha}{\lambda}c^\alpha}^{b\wedge\frac{\alpha}{\lambda}c^\alpha} (c - (\tfrac{\lambda v}{\alpha})^{1/\alpha})v^{-2}dv + \int_{a\vee\frac{\alpha}{\lambda}c^\alpha}^{b\vee\frac{\alpha}{\lambda}c^\alpha} ((\tfrac{\lambda v}{\alpha})^{1/\alpha} - c)v^{-2}dv
$$

$$
= q_c(b \vee \tfrac{\alpha}{\lambda}c^\alpha) - q_c(b \wedge \tfrac{\alpha}{\lambda}c^\alpha) + q_c(a \wedge \tfrac{\alpha}{\lambda}c^\alpha) - q_c(a \vee \tfrac{\alpha}{\lambda}c^\alpha), \qquad (4.1)
$$

where

$$
q_c(x) = (\tfrac{\lambda}{\alpha})^{1/\alpha}\frac{x^{1/\alpha-1}}{1/\alpha - 1} + cx^{-1} .
$$

For convenience we denote

$$
\tilde{\kappa}_c^- := \tfrac{\alpha}{\lambda}(c-1)^\alpha \;\leqslant\; \tilde{\kappa}_c^0 := \tfrac{\alpha}{\lambda}(c)^\alpha \;\leqslant\; \tilde{\kappa}_c^+ := \tfrac{\alpha}{\lambda}(c+1)^\alpha .
$$

Let us restrict these quantities to the interval $(a,b]$ by setting

$$
\kappa_c^\star := \kappa_c^\star(a,b) := (a \vee \tilde{\kappa}_c^\star) \wedge b , \quad \star \in \{-,0,+\} ,
$$

which yields $a \;\leqslant\; \kappa_c^- \;\leqslant\; \kappa_c^0 \;\leqslant\; \kappa_c^+ \;\leqslant\; b$. Then formula (4.1) turns into

$$
\int_a^b (|c - (\tfrac{\lambda v}{\alpha})^{1/\alpha}| \wedge 1)v^{-2}dv
$$

$$
= \int_a^{\kappa_c^-} v^{-2}dv + \int_{\kappa_c^-}^{\kappa_c^0} (c - (\tfrac{\lambda v}{\alpha})^{1/\alpha})v^{-2}dv + \int_{\kappa_c^0}^{\kappa_c^+} ((\tfrac{\lambda v}{\alpha})^{1/\alpha} - c)v^{-2}dv + \int_{\kappa_c^+}^b v^{-2}dv
$$

$$
= q_c(\kappa_c^-) + q_c(\kappa_c^+) - 2q_c(\kappa_c^0) + \frac{1}{a} - \frac{1}{b} + \frac{1}{\kappa_c^+} - \frac{1}{\kappa_c^-}.
$$

Hence we may evaluate

$$
\mathcal{T}_1(\Pi_n, \Pi_{\alpha,\varepsilon}) = \int_0^\infty (|c_n(v) - c_\alpha(v)| \wedge 1)v^{-2}dv
$$

$$
= \int_\varepsilon^1 ((\tfrac{\lambda v}{\alpha})^{1/\alpha} \wedge 1)v^{-2}du + \sum_{i=1}^n \int_{\frac{n}{n-i+1}}^{\frac{n}{n-i}} (|\xi_{i:n} - (\tfrac{\lambda v}{\alpha})^{1/\alpha}| \wedge 1)v^{-2}dv
$$

$$
= \sum_{i=1}^n \int_{\frac{n}{n-i+1}}^{\frac{n}{n-i}} (|\xi_{i:n} - (\tfrac{\lambda v}{\alpha})^{1/\alpha}| \wedge 1)v^{-2}dv
$$

$$
= 1 + \sum_{i=1}^n q_{\xi_{i:n}}(\kappa^-(i)) + q_{\xi_{i:n}}(\kappa^+(i)) - 2q_{\xi_{i:n}}(\kappa^0(i)) + \frac{1}{\kappa^+(i)} - \frac{1}{\kappa^-(i)},
$$

8

α \ ε	0.5	0.6	0.7	0.8	0.9	1.0
1	0.1901	0.2000	0.2191	0.2429	0.2766	0.3191
2	0.0877	0.0893	0.1039	0.1017	0.1181	0.1406
3	0.0522	0.0530	0.0579	0.0619	0.0690	0.0789
4	0.0364	0.0359	0.0373	0.0400	0.0479	0.0551
5	0.0286	0.0277	0.0272	0.0294	0.0335	0.0403
6	0.0232	0.0213	0.0216	0.0238	0.0276	0.0317
7	0.0194	0.0184	0.0170	0.0182	0.0210	0.0237
8	0.0169	0.0153	0.0149	0.0149	0.0172	0.0193
9	0.0149	0.0135	0.0126	0.0137	0.0148	0.0178
10	0.0132	0.0123	0.0116	0.0114	0.0137	0.0149

Table 1: The mean out of 100 simulations of $\tilde{\mathcal{T}}_1(\Pi_n, \Pi_{\alpha,\varepsilon})$ for $n = 100$.

where we have abbreviated

$$\kappa^{\star}(i) := \kappa^{\star}_{\xi_{i:n}}\left(\tfrac{n}{n-i+1}, \tfrac{n}{n-i}\right), \quad i = 1, \ldots, n, \; \star \in \{-, 0, +\} \; .$$

Remark 4.2 (Normalization).

1. Note that the empirical measure Π_n is a probability measure and has total mass one. It is therefore reasonable to normalize $\Pi_{\alpha,\varepsilon}$ choosing $\lambda = \frac{\varepsilon^{\alpha}}{\alpha}$.

2. In general the transportation distance is unbounded. In this example the support of the Lévy measures it bounded away from zero by ε. Hence the transportation distance is bounded from above by $\int_{\varepsilon}^{\infty} u^{-2} du = \varepsilon^{-1}$. For the sake comparability we therefore normalize $\mathcal{T}_1(\Pi_n, \Pi_{\alpha,\varepsilon})$ and set
$$\tilde{\mathcal{T}}_1(\Pi_n, \Pi_{\alpha,\varepsilon}) := \varepsilon \cdot \mathcal{T}_1(\Pi_n, \Pi_{\alpha,\varepsilon}) \; . \tag{4.2}$$

4.2 Simulated data

Let us assess the statistical behavior of the transportation distance between power laws and empirical counterparts in a simulation study. To be precise we consider a family of Lévy measures $(\Pi_{\alpha,\varepsilon})_{\alpha,\varepsilon}$ defined in Example 2.2.2, where the intensity parameter $\lambda = \frac{\varepsilon^{\alpha}}{\alpha}$ is chosen such that it is a probability measure ($\Pi_{\alpha,\varepsilon}(\mathbb{R}_+) = 1$). For α, ε fixed, we generate an i.i.d. sample $\xi_1, \ldots, \xi_n$ (for $n = 100$) distributed according to $\Pi_{\alpha,\varepsilon}$, and evaluate the transportation distance $\tilde{\mathcal{T}}_1$ between its empirical measure Π_n and $\Pi_{\alpha,\varepsilon}$. This experiment is repeated 100 times and the mean value is given in Table 1, the corresponding standard deviation in Table 2. It is apparent that the magnitude of the renormalized distance $\tilde{\mathcal{T}}_1(\Pi_n, \Pi_{\alpha,\varepsilon})$ changes considerably over the parameter range. However the standard deviation is always smaller than mean value by an order of magnitude. Increasing α from 1 to 10 results in a decrease of the mean distance by more than a digit. However increasing ε tends to increase $\mathcal{T}_1$ grosso modo.

This behavior is explained by the Cauchy weighting with v^{-2}. Increasing α shifts mass away from zero to larger values which are damped stronger.

α $\backslash$ ε	0.5	0.6	0.7	0.8	0.9	1.0
1	0.0322	0.0367	0.0425	0.0545	0.0743	0.0744
2	0.0147	0.0171	0.0327	0.0275	0.0402	0.0408
3	0.0075	0.0098	0.0189	0.0183	0.0231	0.0263
4	0.0041	0.0062	0.0094	0.0107	0.0174	0.0185
5	0.0031	0.0061	0.0064	0.0089	0.0126	0.0136
6	0.0030	0.0038	0.0057	0.0077	0.0099	0.0115
7	0.0020	0.0035	0.0036	0.0053	0.0070	0.0089
8	0.0016	0.0026	0.0037	0.0040	0.0058	0.0064
9	0.0016	0.0026	0.0040	0.0036	0.0052	0.0057
10	0.0012	0.0025	0.0025	0.0034	0.0041	0.0056

Table 2: The standard deviation out of 100 simulations of $\tilde{\mathcal{T}}_1(\Pi_n, \Pi_{\alpha,\varepsilon})$ for $n = 100$.

4.3 Climatic time series

Let us turn to the model selection problem for the paleoclimatic time series. We may extend the preceeding considerations to the dynamical set up of a jump diffusion model. Figure 3(a) shows the histogram of the data and indicates the presence of two well defined regimes, separated by a threshold $s^* \approx -0.8$ centered around the values 0.5, which corresponds to a *warm* regime, and -2, a *cold* regime. It is therefore reasonable to discretize the state space into these two regimes, separated by a threshold. We assume the characteristics of the noise to be constant in each of the regimes. Figure 3(b) shows the histograms of the increments in each regime. In both regimes the distributions seem nearly symmetric and may exhibit polynomial tails.

We will consider a finite intensity version of the model (2.3) given as (2.11). In particular we drop the Gaussian component ($b = 0$). The random measure ν has a Lévy kernel $\Pi(x, dv)$ given by a space dependent version of $\Pi_{\alpha,\varepsilon}$ of Example 2.2, namely

$$\Pi'(x, dv) = \mathbf{1}(v > \varepsilon^+)\Pi_{\alpha^+(x),\varepsilon^+}(dv) + \mathbf{1}(v < -\varepsilon^-)\Pi_{\alpha^-(x),\varepsilon^-}(dv) \ . \tag{4.3}$$

The parameters $\alpha^\pm$ are assumed to be constant in each (slightly reduced) regime, $x > s^* + \delta, x < s^* - \delta$. The separation $\delta > 0$ is introduced in view of Proposition 3.6 and Theorem 3.5 to allow for a Lipschitz interpolation of the parameters $\alpha^\pm$,

$$\alpha^\pm(x) = \begin{cases} \alpha^{1,\pm}, & x > s^* + \delta, \\ \alpha^{-1,\pm}, & x < s^* - \delta, \\ \text{Lipschitz interpolation}, & \text{otherwise.} \end{cases} \tag{4.4}$$

The class of models is then given by solutions to the SDE

$$dX(t) = a(X(t))dt + \int_{\mathbb{R}} c(X(t-), v)\nu_0(dt, dv) \ , \tag{4.5}$$

where the Cauchy kernel c is constant in each of the regimes (4.4) and determined by Example 2.2.2. Between the regimes it inherits the continuous interpolation of the parameters $\alpha^\pm$. We will now interpret big increments (larger than the thresholds ε^+ or smaller than $-\varepsilon^-$) of the time series as jumps distributed according to $\Pi(x, \cdot)$. We extract from the series four sub-samples containing big positive or negative increments in each of the two regimes. The remaining increments are

considered to be continuous and will be neglected. Theses four sub-samples are now processed in the same way as our simulation study in the previous subsection. Their respective empirical measures are compared to a family of Lévy measures $\Pi_{\alpha^\pm, \varepsilon^\pm}$ as before.

The thresholds $\varepsilon^\pm$ are chosen in accordance to the polynomial decay of the empirical distribution and taken from the preceding study [8] being $\varepsilon^+ = 0.36, \varepsilon^- = 0.34$ for comparability. The result is presented in Fig. 1. As can be seen, in each of the cases there is a clear minimizing exponent for the transportation distances varying in α. We find the minimizers

$$\alpha^{1,+} = 2.8 \ , \qquad\qquad \alpha^{1,-} = 2.9 \ ,$$
$$\alpha^{-1,+} = 3.6 \ , \qquad\qquad \alpha^{-1,-} = 4.3 \ .$$

We stress that the minimum distances we find are small and of order 0.05 (note that we consider the normalized distance $\tilde{\mathcal{T}}_1$ with values in $[0,1]$). These findings allow to select a simple jump diffusion model of type (4.5) with Lévy kernel (4.3).

Theorem 3.5 allows to control the behavior of the law of this simple model (on path space). We assume that the drift a in equation (4.5) is Lipschitz and recall that the interpolation in (4.4) is Lipschitz, too. The Lipschitz condition (3.5) in Theorem 3.5 for Π' follows from Proposition 3.6. Theorem 3.5 guarantees that the laws of such jump diffusions on a reasonable time horizon are controlled by the transportation distance of the Lévy measures. Hence the jump diffusions selected by the values obtained by the minimization procedure provide a good estimate on the space of such models.

It is possible to obtain the rate of convergence of the empirical Lévy measure Π_n under the transportation distance similarly to the case of coupling distances (c.f. [8]). In this study we content ourselves to the convincing simulation results of Section 4.2. As already mentioned a similar fitting approach for a family of polynomial Lévy measures to this time-series is adopted in [8]. The frame work there is restricted to models with additive noise and does not allow for spacial dependence of the jump kernel and cannot distinguish different regimes. However, the polynomial decay of the left and right tail of the jump measure is studied. As a result the minimizing exponents for the coupling-distance there are found to be $\alpha^+ = 3.6$ (positive jumps) and $\alpha^- = 3.55$ (negative jumps).

Our findings seem to be supportive with this result given that we consider different spacial regimes. Combining the negative jumps in the different regimes should lead to an average of the exponents $\alpha^{1,-}, \alpha^{-1,-}$. If we average the exponents weighted by the number of negative jumps in the corresponding regime we obtain

$$\frac{2.9 \times 302 + 4.3 \times 228}{302 + 228} = 3.5 \approx 3.55 \ .$$

For the positive jumps averaging behaves slightly worse and we obtain roughly

$$\frac{2.8 \times 301 + 3.6 \times 593}{301 + 593} = 3.33 \approx 3.6 \ .$$

In any case this study confirms earlier findings in [8] of a robust polynomial tail behavior with exponents α well beyond 2. Originally a jump diffusion model with α-stable noise component was proposed in [6] which restricts α to values in $(0, 2)$. A waiting time analysis of 42 transition between warm and cold temperature regimes pointed to $\alpha \approx 1.75$. Later a similar model was investigated by refined methods in [9] that supported an exponent of $\alpha \approx 1.75$ within the framework of α-stable Lévy noise. We stress that the class we are considering here are generic in the class of heavy tailed Lévy measures and approximate α-stable ones. The findings of α being close to 2 within the class of α-stable distributions may also indicate a possible exponent beyond 2, since their methods perform poorly for α approaching 2.
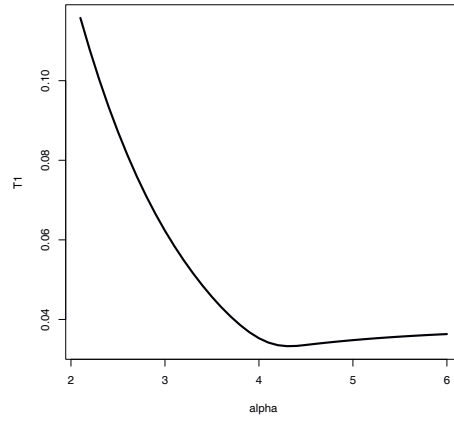
11
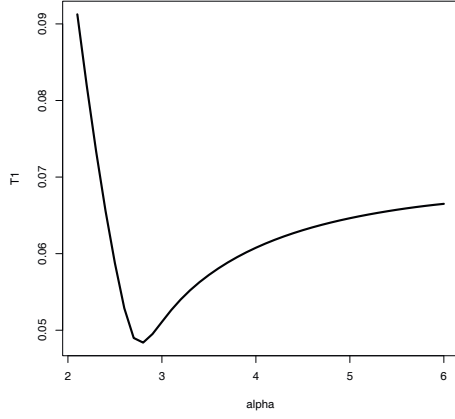
(a) Positive jumps in the warm regime:
$\alpha_{\min} = 2.8,\ \tilde{\mathcal{T}}_1(\Pi_n, \Pi_{\alpha_{\min}}) = 0.048,\ n = 301$
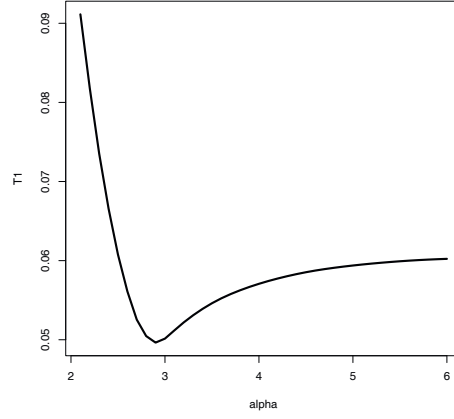
(b) Negative jumps in the warm regime:
$\alpha_{\min} = 2.9,\ \tilde{\mathcal{T}}_1(\Pi_n, \Pi_{\alpha_{\min}}) = 0.050,\ n = 302$
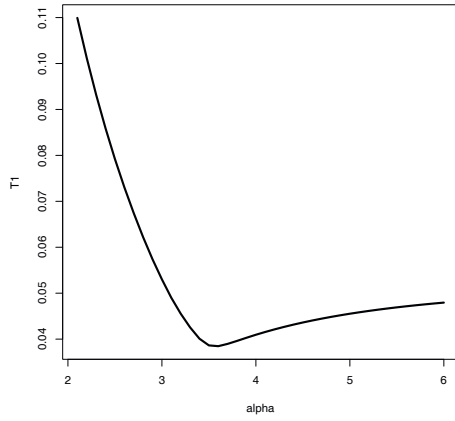
(c) Posititve jumps in the cold regime:
$\alpha_{\min} = 3.6,\ \tilde{\mathcal{T}}_1(\Pi_n, \Pi_{\alpha_{\min}}) = 0.039,\ n = 593$

(d) Negative jumps in the cold regime:
$\alpha_{\min} = 4.3,\ \tilde{\mathcal{T}}_1(\Pi_n, \Pi_{\alpha_{\min}}) = 0.033,\ n = 228$

Figure 1: Comparison of the empirical Lévy measure in the different regimes and different tails to the family $(\Pi_\alpha)_{\alpha \in (2,6)}$.
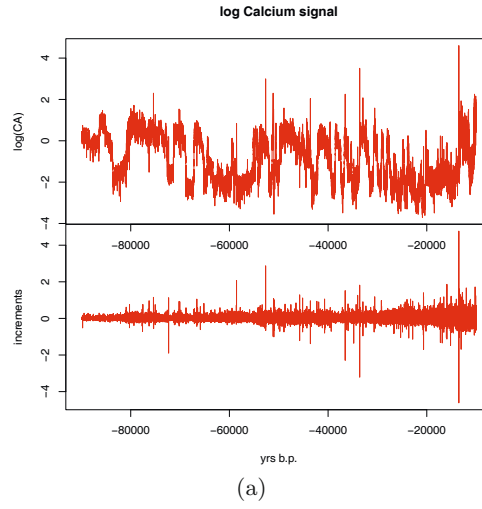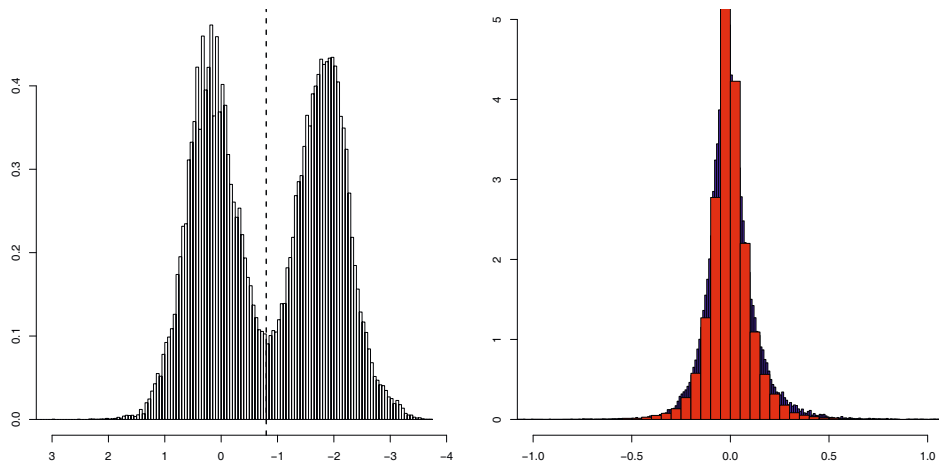
log Calcium signal

(a)

Figure 2: Timeseries

(a) The distribution of the data

(b) Superposition of increments in the cold and warm regime

Figure 3: Histograms

13

5 Proofs

Proof. of Theorem 3.1:

The strategy of the proof is to bound first $\mathbb{E}\big[(X_1(t) - X_2(t))^2 \wedge 1\big]$ with the help of the cut-off technique developed in [7]. Using the elementary inequality

$$y \wedge 1 \leqslant \frac{4}{\pi} \arctan y, \quad \text{for } y \geqslant 0 \tag{5.1}$$

we obtain for $F(y) = \arctan(y^2)$

$$\rho^2(x,y) \leqslant \frac{4}{\pi} F(|x-y|), \qquad \forall x, y \in \mathbb{R}.$$

Applying Itô's formula and Doob's maximal inequality to $F(|X_1 - X_2|)$ we establish Gronwall estimates for the process $F(|X_1 - X_2|)$. As a preparation for the two dimensional Itô formula for $F(|X_1 - X_2|)$ we rewrite the process $Y(t) = X_1(t) - X_2(t)$ and $a_i^*(x) = a_i(x) + \bar{a}_i(x)$ with common Lipschitz constant $L_{a^*} = L_a + L_{\bar{a}}$ as follows

$$
\begin{aligned}
d(X_1 - X_2)(t) =& (a_1^*(X_1) - a_2^*(X_2))dt + (b_1(X_1) - b_2(X_2))dW(t) \\
&+ \int\limits_{|v|>1} (c_1(X_1(t-),v) - c_2(X_2(t-),v)\nu_0(dv,dt) \\
&- \int\limits_{|v|>1} \Big(\tau(c_1(X_1(t-),v)) - \tau(c_2(X_2(t-),v))\Big)\Pi_0(dv)dt \\
&- \int\limits_{|v|\leqslant 1} \Big(\tau(c_1(X_1(t-),v)) - c_1(X_1(t-),v)\Big)\Pi_0(dv)dt \\
&+ \int\limits_{|v|\leqslant 1} \Big(\tau(c_2(X_2(t-),v)) - c_2(X_2(t-),v)\Big)\Pi_0(dv)dt \\
&- \int\limits_{|v|\leqslant 1} c_1(X_1(t-),v)\tilde{\nu}_0(dv,dt) + \int\limits_{|v|\leqslant 1} c_2(X_2(t-),v)\tilde{\nu}_0(dv,dt),
\end{aligned}
$$

where

$$\tau(v) = (|v| \wedge 1)\operatorname{sign}(v), \ v \in \mathbb{R}.$$

Applying Itô's formula given for instance in Chapter 2 in [20] we get

$$
\begin{aligned}
dF(Y(t)) =& F'(Y(t-))(a_1^*(X_1) - a_2^*(X_2))dt + F'(Y(t))(b_1(X_1) - b_2(X_2))dW(t) \\
&+ \frac{1}{2}F''(Y(t))(b_1(X_1) - b_2(X_2))^2 dt \\
&- \int\limits_{|v|\leqslant 1} F'(Y(t-))\Big(\tau(c_1(X_1(t-),v)) - c_1(X_1(t-),v)\Big)\Pi_0(dv)dt \\
&+ \int\limits_{|v|\leqslant 1} F'(Y(t-))\Big(\tau(c_2(X_2(t-),v)) - c_2(X_2(t-),v)\Big)\Pi_0(dv)dt \\
&- \int\limits_{|v|>1} \Big(F'(Y(t-))(\tau(c_1(X_1(t-),v)) - \tau(c_2(X_2(t-),v)))\Big)\Pi_0(dv)dt
\end{aligned}
$$

14

$$+ \int\limits_{|v|>1} \Big(F(Y(t-) + (c_1(X_1(t),v) - c_2(X_2(t),v))) - F(Y(t-)) \Big) \nu_0(dv,dt)$$

$$+ \int\limits_{|v|\leqslant 1} \Big(F(Y(t-) + (c_1(X_1(t),v) - c_2(X_2(t),v))) - F(Y(t-)) \Big) \tilde{\nu}_0(dv,dt)$$

$$+ \int\limits_{|v|\leqslant 1} \Big(F(Y(t-) + (c_1(X_1(t),v) - c_2(X_2(t),v))) - F(Y(t-)) \Big)$$

$$- F'(Y(t-))(c_1(X_1(t),v) - c_2(X_2(t),v))\Pi_0(dv)dt.$$

After rearrangements, we finally obtain the representation

$$F(Y(t)) = F(|x_1 - x_2|) + M(t) + \int\limits_0^t \int\limits_{\mathbb{R}} g(Y(s), X_1(s), X_2(s))\Pi_0(dv)ds, \tag{5.2}$$

where

$$M(t) = \int\limits_0^t F'(Y(s))(b_1(X_1(s)) - b_2(X_2(s)))dW(s)$$

$$+ \int\limits_0^t \int\limits_{\mathbb{R}} \Big(F(Y(s-) + (c_1(X_1(s),v) - c_2(X_2(s),v))) - F(Y(s-)) \Big)[\nu_0(dv,ds) - \Pi_0(dv)ds]$$

$$\tag{5.3}$$

and

$$g(y, u_1, u_2) = F'(y)(a_1^*(u_1) - a_2^*(u_2)) + \frac{1}{2}F''(y)(b_1(u_1) - b_2(u_2))^2$$

$$+ \int\limits_{\mathbb{R}} [F(y + (c_1(u_1,v) - c_2(u_2,v))) - F(y) - F'(y)(\tau(c_1(u_1,v)) - \tau(c_2(u_2,v)))]\Pi_0(dv)$$

$$=: g_1(y, u_1, u_2) + g_2(y, u_1, u_2) + g_3(y, u_1, u_2), \quad (y, u_1, u_2) \in \mathbb{R}^3, \quad y = u_1 - u_2. \tag{5.4}$$

For the function F and its derivatives, we have the following bounds taken from [7]

$$F'(y) = \frac{2y}{1 + y^4}, \quad \text{implying} \quad |F'(y)| \leqslant \frac{3^{3/4}}{2} \text{ for all } y \in \mathbb{R}, \tag{5.5}$$

$$F'(y)y = \frac{2y^2}{1 + y^4} \leqslant (2y^2) \wedge 1 = 2(y^2 \wedge \frac{1}{2}) \leqslant \frac{F(y)}{\arctan(1/2)}, \tag{5.6}$$

$$F''(y) = 2\frac{1 - 3y^4}{(1 + y^4)^2}, \quad \text{implying} \quad |F''(y)| \leqslant 2, \tag{5.7}$$

$$|F(y + \delta) - F(y) - F'(y)\delta| \leqslant \frac{\delta^2}{2} \sup_v |F''(v)| \leqslant \delta^2. \tag{5.8}$$

Now we can bound each summand of (5.4). Using (2.5) and (5.6) we obtain

$$g_1(y, u_1, u_2) \leqslant \frac{L_{a^*}}{\arctan(1/2)}F(y) + \frac{3^{3/4}}{2}\|a_1^* - a_2^*\|_\infty \quad \forall u_1, u_2, y = u_1 - u_2 \in \mathbb{R}. \tag{5.9}$$

15

Then by (2.6) g_2 is not greater than

$$g_2(y, u_1, u_2) \leqslant (b_1(u_1) - b_2(u_2))^2 \qquad \forall u_1, u_2, y = u_1 - u_2 \in \mathbb{R}. \tag{5.10}$$

The estimate of g_3 is slightly more involved and for convenience we introduce for fixed u_i the notation

$$\zeta_i(v) := c_i(u_i, v), \qquad v \in \mathbb{R}.$$

Let us first consider the integral g_3 on the set $\{v \in \mathbb{R} \mid |\zeta_1(v) - \zeta_2(v)| \leqslant 1\}$ and apply

$$F(y + (z_1 - z_2)) - F(y) \leqslant F'(y)(z_1 - z_2) + (z_1 - z_2)^2$$

to the first summand of

$$\int\limits_{\{v \in \mathbb{R} \mid |\zeta_1(v) - \zeta_2(v)| \leqslant 1\}} \Big([F(y + (\zeta_1(v) - \zeta_2(v))) - F(y)] - F'(y)[\tau(\zeta_1(v)) - \tau(\zeta_2(v))] \Big) \Pi_0(dv)$$

$$\leqslant \int\limits_{\{v \in \mathbb{R} \mid |\zeta_1(v) - \zeta_2(v)| \leqslant 1\}} \Big(F'(y) \Big[\zeta_1(v) - \zeta_2(v) - \tau(\zeta_1(v)) + \tau(\zeta_2(v)) \Big] + (\zeta_1(v) - \zeta_2(v))^2 \Big) \Pi_0(dv).$$

Observe that the term in square brackets vanishes if both, $|\zeta_i(v)| \leqslant 1$ for $i = 1, 2$. Hence the first integral reads

$$\int\limits_{\{v \in \mathbb{R} \mid |\zeta_1(v) - \zeta_2(v)| \leqslant 1\}} F'(y) \Big[\zeta_1(v) - \zeta_2(v) - \tau(\zeta_1(v)) + \tau(\zeta_2(v)) \Big] \Pi_0(dv)$$

$$= \int\limits_{\substack{\{v \in \mathbb{R} \mid |\zeta_1(v) - \zeta_2(v)| \leqslant 1 \\ \text{and } |\zeta_1(v)| > 1 \text{ or } |\zeta_2(v)| > 1\}}} F'(y) \Big[\zeta_1(v) - \zeta_2(v) - \tau(\zeta_1(v)) + \tau(\zeta_2(v)) \Big] \Pi_0(dv)$$

$$\leqslant 2 F'(y) \int\limits_{\substack{\{v \in \mathbb{R} \mid |\zeta_1(v) - \zeta_2(v)| \leqslant 1 \\ \text{and } |\zeta_1(v)| > 1 \text{ or } |\zeta_2(v)| > 1\}}} \Big[|\zeta_1(v) - \zeta_2(v)| \wedge 1 \Big] \Pi_0(dv)$$

$$\leqslant 2 F'(y) \Big(\int\limits_{\mathbb{R}} \Big[|\zeta_1(v) - \zeta_2(v)| \wedge 1 \Big]^2 \Pi_0(dv) \Big)^{\frac{1}{2}} \Pi_0 \Big(\{v \in \mathbb{R} \mid |\zeta_1(v)| > 1 \text{ or } |\zeta_2(v)| > 1\} \Big)^{\frac{1}{2}}$$

$$\leqslant 2 F'(y) \mathcal{T}_2(\Pi_1(u_1, \cdot), \Pi_2(u_2, \cdot)) \Big(\Pi_0(\{v \in \mathbb{R} \mid |\zeta_1(v)| > 1\}) + \Pi_0(\{v \in \mathbb{R} \mid |\zeta_2(v)| > 1\}) \Big)^{\frac{1}{2}},$$

where we have used that $|\zeta_1(v) - \zeta_2(v)| \leqslant 1$ implies

$$|\zeta_1(v) - \zeta_2(v)| + |\tau(\zeta_1(v)) - \tau(\zeta_2(v))| \leqslant 2(|\zeta_1(v) - \zeta_2(v)| \wedge 1).$$

We consider now the integral g_3 on the remainder set $\{v \in \mathbb{R} \mid |\zeta_1(v) - \zeta_2(v)| > 1\}$. By its definition in (5.4) g_3 is uniformly bounded by $\pi + \frac{3^{\frac{3}{4}}}{2}$ due to the boundedness of F and (5.5). Therefore the function g_3 in (5.4) is bounded by the sum

$$g_3(y, u_1, u_2) \leqslant 2 F'(y) \mathcal{T}_2(\Pi_1(u_1, \cdot), \Pi_2(u_2, \cdot)) \Big(\Pi_0(\{v : |\zeta_1(v)| > 1\} + \Pi_0(\{v : |\zeta_2(v)| > 1\}) \Big)^{\frac{1}{2}}$$

16

$$+ (\pi + \frac{3^{\frac{3}{4}}}{2}) \int\limits_{\mathbb{R}} ((\zeta_1(v) - \zeta_2(v))^2 \wedge 1) \Pi_0(dv)$$

$$\leqslant 2F'(y)\mathcal{T}_2(\Pi_1(u_1, \cdot), \Pi_2(u_2, \cdot))\Big(\Pi_0(\{v : |\zeta_1(v)| > 1\} + \Pi_0(\{v : |\zeta_2(v)| > 1\})\Big)^{\frac{1}{2}}$$

$$+ (\pi + \frac{3^{\frac{3}{4}}}{2})\mathcal{T}_2(\Pi_1(u_1, \cdot), \Pi_2(u_2, \cdot))^2. \tag{5.11}$$

Summarising (5.9), (5.10) and (5.11) we get for the process $F(Y(t))$ the following estimate almost surely

$$F(Y(t))$$

$$\leqslant \frac{\pi}{2}\rho^2(x_1, x_2) + T\frac{3^{3/4}}{2}\|a_1^* - a_2^*\|_\infty - \frac{L_{a^*}}{\arctan(1/2)} \int\limits_0^t F(Y(s))ds + \int\limits_0^t \Big((b_1(X_1(s)) - b_2(X_2(s)))^2 ds$$

$$+ 2\int\limits_0^t F'(Y(s-))\mathcal{T}_2(\Pi_1(x_1, \cdot), \Pi_2(x_2, \cdot))\Big(\Pi_0(\{v : |\zeta_1(v)| > 1\})^{\frac{1}{2}} + \Pi_0(\{v : |\zeta_2(v)| > 1\})^{\frac{1}{2}}\Big)ds$$

$$+ \int\limits_0^t \Big(\pi + \frac{3^{\frac{3}{4}}}{2}\Big)\mathcal{T}_2^2(\Pi_1(X_1(s-), \cdot), \Pi_2(X_2(s-), \cdot)\Big)ds + M(t).$$

In the sequel we apply the Lipschitz continuity of the coefficients b_i, $i = 1, 2$ and the kernels Π_i, $i = 1, 2$, however taken at different points. By (2.6) and (2.8) we obtain ω-wise

$$\mathcal{T}_2(\Pi_1(X_1(s-), \cdot), \Pi_2(X_2(s-), \cdot)) \leqslant L_\Pi(|Y(s-)| \wedge 1) + \sup_{x \in \mathbb{R}} \mathcal{T}_2(\Pi_1(x, \cdot), \Pi_2(x, \cdot))$$

and by (5.1)

$$\mathcal{T}_2^2(\Pi_1(X_1(s-), \cdot), \Pi_2(X_2(s-), \cdot)) \leqslant \frac{8L_\Pi}{\pi}F(Y(s-)) + 2\sup_{x \in \mathbb{R}} \mathcal{T}_2^2(\Pi_1(x, \cdot), \Pi_2(x, \cdot)),$$

$$(b_1(X_1) - b_2(X_2))^2 \leqslant \frac{8L_b}{\pi}F(Y(s)) + 2\sup_{x \in \mathbb{R}}(b_1(x) - b_2(x))^2. \tag{5.12}$$

Setting

$$K_1 := \sup_{u_1, u_2 \in \mathbb{R}} \Big(\Pi_0(\{v : |c_1(u_1, v)| > 1\}) + \Pi_0(\{v : |c_2(u_2, v)| > 1\})\Big)^{\frac{1}{2}}$$

and the help of (5.6) we obtain almost surely

$$F(Y(t)) \leqslant \frac{\pi}{2}\rho^2(x_1, x_2) + T\frac{3^{3/4}}{2}\|a_1^* - a_2^*\|_\infty + \frac{L_{a^*}}{\arctan(1/2)} \int\limits_0^t F(Y(s))ds + \frac{8L_b}{\pi} \int\limits_0^t F(Y(s))ds$$

$$+ 2t\sup_{x \in \mathbb{R}}(b_1(x) - b_2(x))^2$$

$$+ M(t) + 2K_1 \int\limits_0^t |F'(Y(s))||Y(s)|ds + 4K_1 t \sup_{x \in \mathbb{R}} \mathcal{T}_2(\Pi_1(x, \cdot), \Pi_2(x, \cdot))$$

$$+ (8 + \frac{4 \cdot 3^{\frac{3}{4}}}{\pi})L_\Pi \int\limits_0^t F(Y(s))ds + (2\pi + 3^{\frac{3}{4}})t \sup_{x \in \mathbb{R}} \mathcal{T}_2^2(\Pi_1(x, \cdot), \Pi_2(x, \cdot))$$

17

$$\leqslant \frac{\pi}{2}\rho^2(x_1,x_2) + \Big(\frac{L_{a^*}+2K_1}{\arctan(1/2)} + \frac{8L_b}{\pi} + (8+\frac{4\cdot 3^{\frac{3}{4}}}{\pi})L_\Pi\Big)\int_0^t F(Y(s))ds + M(t)$$
$$+ 2t\sup_{x\in\mathbb{R}}(b_1(x)-b_2(x))^2 + 4tK_1\sup_{x\in\mathbb{R}}\mathcal{T}_2(\Pi_1(x,\cdot),\Pi_2(x,\cdot))$$
$$+ T\frac{3^{3/4}}{2}\|a_1^*-a_2^*\|_\infty + t(2\pi+3^{\frac{3}{4}})\sup_{x\in\mathbb{R}}\mathcal{T}_2^2(\Pi_1(x,\cdot),\Pi_2(x,\cdot)). \tag{5.13}$$

We define
$$Q := \frac{L_{a^*}+2K_1}{\arctan(1/2)} + \frac{8L_b}{\pi} + (8+\frac{4\cdot 3^{\frac{3}{4}}}{\pi})L_\Pi,$$

take the expectation and apply Gronwall's lemma for $t \in [0,T]$

$$\mathbb{E}[F(Y(t))] \leqslant \Big(\frac{\pi}{2}\rho^2(x_1,x_2) + T\frac{3^{3/4}}{2}\|a_1^*-a_2^*\|_\infty + 2T\sup_{x\in\mathbb{R}}(b_1(x)-b_2(x))^2 + 4TK_1\sup_{x\in\mathbb{R}}\mathcal{T}_2(\Pi_1(x,\cdot),\Pi_2(x,\cdot))$$
$$+ T(2\pi+3^{\frac{3}{4}})\sup_{x\in\mathbb{R}}\mathcal{T}_2^2(\Pi_1(x,\cdot),\Pi_2(x,\cdot))\Big)e^{QT}. \tag{5.14}$$

Now establish the estimate for $\mathbb{E}[\sup_{t\in[0,T]} F(Y(t))]$. For that purpose we take the supremum over both sides of (5.13) and observe that

$$\mathbb{E}[\sup_{t\in[0,T]} F(Y(t))] \leqslant \frac{\pi}{2}\rho^2(x_1,x_2) + Q\int_0^T \mathbb{E}[F(Y(s))]ds + \mathbb{E}[\sup_{t\in[0,T]}|M(t)|]+$$
$$T\Big(\frac{3^{3/4}}{2}\|a_1^*-a_2^*\|_\infty + 2\sup_{x\in\mathbb{R}}(b_1(x)-b_2(x))^2 + 4K_1\sup_{x\in\mathbb{R}}\mathcal{T}_2(\Pi_1(x,\cdot),\Pi_2(x,\cdot))$$
$$+ (2\pi+3^{\frac{3}{4}})\sup_{x\in\mathbb{R}}\mathcal{T}_2^2(\Pi_1(x,\cdot),\Pi_2(x,\cdot))\Big). \tag{5.15}$$

For the integrand of the first integral we apply (5.14). In order to bound the martingal term we use Doob's maximum inequality

$$\mathbb{E}[\sup_{t\in[0,T]}|M(t)|] \leqslant \mathbb{E}[\sup_{t\in[0,T]}|M(t)|^2]^{\frac{1}{2}} \leqslant 4\mathbb{E}[|M(T)|^2]^{\frac{1}{2}}$$

and rewrite $\mathbb{E}[|M(T)|^2]$ as follows

$$\mathbb{E}[|M(T)|^2] = \mathbb{E}\Big[\Big(\int_0^T\int_{\mathbb{R}}\Big[F\big(Y(t-)+(c_1(X_1(t),v)-c_2(X_2(t),v))\big)-F(Y(t-))\Big]\nu_0(dv,dt)$$
$$- \int_0^T\int_{\mathbb{R}}[F\big(Y(t-)+(c_1(X_1(t),v)-c_2(X_2(t),v))\big)-F(Y(t-))]\Pi_0(dv)dt\Big)^2\Big]$$
$$+ \int_0^T\mathbb{E}\Big[\Big(F'(Y(t))\big(b_1(X_1(t))-b_2(X_2(t))\big)\Big)^2\Big]dt$$
$$= \int_0^T\int_{\mathbb{R}}\mathbb{E}\Big[\Big(F\big(Y(t-)+(c_1(X_1(t),v)-c_2(X_2(t),v))\big)-F(Y(t-))\Big)^2\Big]\Pi_0(dv)dt$$

$$+ \int_0^T \mathbb{E}\left[\left(F'(Y(t))\Big(b_1(X_1(t)) - b_2(X_2(t))\Big)\right)^2\right]dt.$$

The analogous separation argument as for the term g_3 of the cases $\{v \in \mathbb{R} \mid |c_1(x,v) - c_2(x,v)| \leqslant 1\}$ and its complement and using $F(y+\delta) - F(y) \leqslant 2\delta$ for $\delta \in (0,1)$ and $y \in \mathbb{R}$ we obtain with the help of (2.6), (2.8) and (5.1) the estimate

$$\mathbb{E}[|M(T)|^2] \leqslant 2 \int_0^T \int_{\mathbb{R}} \mathbb{E}\left[(c_1(X_1(t),v) - c_2(X_2(t),v))^2 \wedge 1\right]\Pi_0(dv)dt$$

$$+ 2 \int_0^T \mathbb{E}\left[\Big(b_1(X_1(t)) - b_2(X_2(t))\Big)^2\right]dt$$

$$= 2 \int_0^T \mathbb{E}\left[\mathcal{T}_2^2(\Pi_1(X_1(t),\cdot), \Pi_2(X_2(t),\cdot)\right]dt + 2\int_0^T \mathbb{E}\left[\Big(b_1(X_1(t)) - b_2(X_2(t))\Big)^2\right]dt$$

$$\leqslant \frac{16 L_\Pi}{\pi} \int_0^T \mathbb{E}\left[F(Y(t))\right]dt + 4T \sup_{x \in \mathbb{R}} \mathcal{T}_2^2(\Pi_1(x,\cdot), \Pi_2(x,\cdot))$$

$$+ \frac{16 L_b}{\pi} \int_0^T \mathbb{E}\left[F(Y(t))\right]dt + 4T \sup_{x \in \mathbb{R}}(b_1(x) - b_2(x))^2.$$

The last inequality comes from the triangle inequality (5.12). Finally inserting (5.14) we obtain

$$\mathbb{E}\left[\sup_{t \in [0,T]} F(Y(t))\right]$$

$$\leqslant (T^2 Q e^{QT} + T)\left(\frac{\pi}{2}\rho^2(x_1, x_2) + \frac{3^{3/4}}{2}\|a_1^* - a_2^*\|_\infty + 2\sup_{x \in \mathbb{R}}(b_1(x) - b_2(x))^2\right.$$

$$\left. + 4K_1 \sup_{x \in \mathbb{R}} \mathcal{T}_2(\Pi_1(x,\cdot), \Pi_2(x,\cdot)) + (2\pi + 3^{\frac{3}{4}}) \sup_{x \in \mathbb{R}} \mathcal{T}_2^2(\Pi_1(x,\cdot), \Pi_2(x,\cdot))\right) +$$

$$\left(32(L_\Pi + L_b)\rho^2(x_1,x_2) + \frac{64(L_\Pi + L_b)}{2\pi}T\|a_1^* - a_2^*\|_\infty^2 + \left(\frac{32}{\pi}(L_\Pi + L_b)T^2 e^{QT} + 4T\right)\sup_{x \in \mathbb{R}}(b_1(x) - b_2(x))^2\right.$$

$$+ \frac{64}{\pi}(L_\Pi + L_b)K_1 e^{QT} \sup_{x \in \mathbb{R}} \mathcal{T}_2(\Pi_1(x,\cdot), \Pi_2(x,\cdot)) +$$

$$\left. \left(\frac{16}{\pi}(L_\Pi + L_b)(2\pi + 3^{\frac{3}{4}})T^2 e^{QT} + 4T\right)\sup_{x \in \mathbb{R}} \mathcal{T}_2^2(\Pi_1(x,\cdot), \Pi_2(x,\cdot))\right)^{1/2}.$$

Taking the maximum of all the constants which we denote by $D = D(K_1, L_{a^*}, L_b, L_\Pi, T)$, using $x + \sqrt{x} \leqslant 2G(x)$ for any $x \in \mathbb{R}$, and rearranging the terms we obtain the desired estimate (3.1). This finishes the proof. $\qquad \square$

Proof. of the Theorem 3.5
In order to get estimate (3.6) we may use the same scheme as in the proof of the Theorem (3.1) but for a $\mathcal{C}^1$-function $H^\delta : \mathbb{R} \to \mathbb{R}_+$ instead of F satisfying

$$\|H^\delta\|_\infty \leqslant 2, \qquad \|(H^\delta)'\|_\infty \leqslant 2, \qquad \lim_{\delta \to 0+} \|H^\delta(y) - (1 \wedge |y|)\|_\infty = 0, \tag{5.16}$$

since there is no Brownian part and the intensity of the jump part is finite. We take the process $Y(t) = X_1(t) - X_2(t)$, which is the difference of the processes defined in (2.11), and apply Itô's formula for $H^\delta(Y)$. Analogously to (5.2) we obtain for $H^\delta(Y)$ the simpler formula

$$H^\delta(Y(t)) = \int\limits_0^t \Big(a_1(X_1(s)) - a_2(X_2(s))\Big)ds + M(t)$$
$$+ \int\limits_0^t \int\limits_{\mathbb{R}} \big[H^\delta(Y(s) + (c_1(X_1(s), v) - c_2(X_2(s), v))) - H^\delta(Y(s))\big]\Pi_0(dv)ds + H^\delta(Y(0)), \tag{5.17}$$

where the martingale term $M^\delta(t)$ is given as

$$M^\delta(t) = \int\limits_0^t \int\limits_{\mathbb{R}} \Big[H^\delta(Y(s-) + (c_1(X_1(s), v) - c_2(X_2(s), v))) - H^\delta(Y(s-))\Big][\nu_0(dv, ds) - \Pi_0(dv)ds].$$

We obtain the estimate for the inner integral in the second summand of (5.17) on two separate sets

$$\{v : |c_1(X_1(s), v) - c_2(X_2(s), v)| \leqslant 1\} \text{ and } \{v : |c_1(X_1(s), v) - c_2(X_2(s), v)| > 1\} \tag{5.18}$$

for fixed $s \in [0, t]$. Using the fact that $|(H^\delta)'(y)| \leqslant 2$ and a Taylor expansion of H^δ in y yields on the set $\{v \in \mathbb{R} : |c_1(X_1(s), v) - c_2(X_2(s), v)| \leqslant 1\}$ almost surely

$$|H^\delta(y + (c_1(x_1, v) - c_2(x_v, v))) - H^\delta(y)| \leqslant 2|c_1(x_1, v) - c_2(x_2, v))|. \tag{5.19}$$

Hence we obtain the following ω-wise estimates

$$H^\delta(Y(t)) \leqslant M^\delta(t) + 2L_a \int\limits_0^t (|Y(s)| \wedge 1)ds + 2t\|a_1 - a_2\|_\infty \tag{5.20}$$
$$+ 2\int\limits_0^t \int\limits_{\{v:|c_1(X_1(s),v)-c_2(X_2(s),v)|\leqslant 1\}} |c_1(X_1(s), v) - c_2(X_2(s), v))|\, \Pi_0(dv)ds$$
$$+ 4\int\limits_0^t \int\limits_{\{v:|c_1(X_1(s),v)-c_2(X_2(s),v)|>1\}} \Pi_0(dv)ds + H^\delta(Y(0))$$
$$\leqslant M^\delta(t) + 2L_a \int\limits_0^t (|Y(s)| \wedge 1)ds + 2t\|a_1 - a_2\|_\infty$$
$$+ 4\int\limits_0^t \int\limits_{\mathbb{R}} \big(|c_1(X_1(s), v) - c_2(X_2(s), v))| \wedge 1\big)\, \Pi_0(dv)ds + H^\delta(Y(0))$$

20

$$= M^{\delta}(t) + 4L_a \int\limits_0^t (|Y(s)| \wedge 1)ds + 2t\|a_1 - a_2\|_{\infty}$$

$$+ 4 \int\limits_0^t \mathcal{T}_1(\Pi_1(X_1(s), \cdot), \Pi_2(X_2(s), \cdot))ds + H^{\delta}(Y(0)), \tag{5.21}$$

where we have used the boundedness of the function H^{δ} and the fact that $2|y| \wedge 4 \leqslant 4(|y| \wedge 1)$ for $y \in \mathbb{R}$. The inner integral in the last expression is by definition $\mathcal{T}_1(\Pi_1(X_1(s), \cdot), \Pi_2(X_2(s), \cdot))$. For the second term on the right-hand side of (5.21) we apply the triangular inequality for $\mathcal{T}_1$ and the Lipschitz condition (3.5). We take the expectation on (5.21), send $\delta \to 0$ and obtain

$$\mathbb{E}\big[|Y(t)| \wedge 1\big] \leqslant (2L_a + 4L_{\Pi}) \int\limits_0^t \mathbb{E}\big[|Y(s)| \wedge 1\big]ds + 2t\|a_1 - a_2\|_{\infty} + 4T \sup_{x \in \mathbb{R}} \mathcal{T}_1(\Pi_1(x, \cdot), \Pi_2(x, \cdot))$$

$$+ \rho(x_1, x_2). \tag{5.22}$$

Gronwall's lemma for $t \in [0, T]$ and the monotonicity on the right-hand side imply

$$\mathbb{E}\left[|Y(t)| \wedge 1\right] \leqslant e^{(2L_a + 4L_{\Pi})T}\big(2T\|a_1 - a_2\|_{\infty} + 4T \sup_{x \in \mathbb{R}} \mathcal{T}_1(\Pi_1(x, \cdot), \Pi_2(x, \cdot)) + \rho(x_1, x_2)\big). \tag{5.23}$$

In what follows we denote $\Delta := 2T\|a_1 - a_2\|_{\infty} + 4T \sup_{x \in \mathbb{R}} \mathcal{T}_1(\Pi_1(x, \cdot), \Pi_2(x, \cdot)) + \rho(x_1, x_2)$. Further, taking the supremum in $t \in [0, T]$ on both sides of (5.21) we use the monotonicity on the right-hand side and take the expectation

$$\mathbb{E}\big[\sup_{t \in [0,T]} H^{\delta}(Y(s))\big] \leqslant \mathbb{E}\big[\sup_{t \in [0,T]} |M_t^{\delta}|\big] + (2L_a + 4L_{\Pi}) \int\limits_0^T \mathbb{E}\big[(|Y(s)| \wedge 1)\big]ds + \Delta. \tag{5.24}$$

For the martingale term in (5.24) we apply Doob's maximal moment inequality with $p = 2$, which yields

$$\mathbb{E} \sup_{t \in [0,T]} |M_t^{\delta}|^2 \leqslant 4\mathbb{E}|M_T^{\delta}|^2$$

and hence

$$\mathbb{E} \sup_{t \in [0,T]} |M_t^{\delta}|^2 \leqslant 4\mathbb{E} \int\limits_0^T \int\limits_{\mathbb{R}} \Big[H^{\delta}(Y(s-) + (c_1(X_1(s), v) - c_2(X_2(s), v))) - H^{\delta}(Y(s-))\Big]^2 \Pi_0(dv)ds. \tag{5.25}$$

Using the same separation argument for (5.18) as in the estimate (5.21) we obtain

$$\mathbb{E}\big[\sup_{t \in [0,T]} |M_t^0|^2\big] \leqslant 4\mathbb{E} \int\limits_0^T \int\limits_{\mathbb{R}} \Big(|c_1(X_1(s), v) - c_2(X_2(s), v)|^2 \wedge 1\Big) \Pi_0(dv)ds. \tag{5.26}$$

Using the fact that $|z|^2 \wedge 1 \leqslant |z| \wedge 1$ for $z \in \mathbb{R}$ we get analogously to (5.22)

$$\mathbb{E}\big[\sup_{t \in [0,T]} |M_t^0|^2\big] \leqslant 16L_{\Pi} \int\limits_0^T \mathbb{E}[(|Y(s)| \wedge 1)]ds + 4\Delta. \tag{5.27}$$

Inserting (5.23) in (5.26) and finally both of them in (5.24) we conclude

$$
\mathbb{E}\big[\sup_{t\in[0,T]} (|Y(s)| \wedge 1) \big]
$$

$$
\leqslant \sqrt{\big(16 L_\Pi T e^{(2L_a+4L_\Pi)T} + 4\big)\big(2T\|a_1-a_2\|_\infty + 4T \sup_{x\in\mathbb{R}} \mathcal{T}_1(\Pi_1(x,\cdot),\Pi_2(x,\cdot)) + 2\rho(x_1,x_2)\big)}
$$

$$
+ \big((2L_a+4L_\Pi)T e^{(2L_a+4L_\Pi)T} + 1\big)\big(2T\|a_1-a_2\|_\infty + 4T \sup_{x\in\mathbb{R}} \mathcal{T}_1(\Pi_1(x,\cdot),\Pi_2(x,\cdot)) + 2\rho(x_1,x_2)\big).
$$

$$(5.28)$$

Taking K as the maximum of all appearing constants and $x + \sqrt{x} \leqslant 2G(x)$ we have obtained the required estimate. This finishes the proof. $\qquad\square$

Proof. of the Proposition 3.6: Analogously to (2.2) we get the functions c_i

$$
c_i(y) = -\,(|y|\lambda_-)^{1/\alpha_i^-}\,\mathbf{1}_{\{y\leqslant -\frac{1}{\lambda_-}\}} + (y\lambda_+)^{1/\alpha_i^+}\,\mathbf{1}_{\{y\geqslant \frac{1}{\lambda_+}\}}. \tag{5.29}
$$

Thus, the distance $\mathcal{T}_1$ is calculated as

$$
\mathcal{T}_1(\Pi_1,\Pi_2) = \int_0^\infty \left(\mathbf{1}_{\{y\geqslant \frac{1}{\lambda_+}\}} \left|(y\lambda_+)^{1/\alpha_1^+} - (y\lambda_+)^{1/\alpha_2^+}\right| \wedge 1 \right) \frac{dy}{y^2}
$$

$$
+ \int_{-\infty}^0 \left(\mathbf{1}_{\{y\leqslant -\frac{1}{\lambda_-}\}} \left||y\lambda_-|^{1/\alpha_1^-} - |y\lambda_-|^{1/\alpha_2^-}\right| \wedge 1 \right) \frac{dy}{y^2}. \tag{5.30}
$$

The integrands are integrable with the absolute value since $\alpha_i^\pm > 2$. The first summand equals

$$
\int_{\frac{1}{\lambda_+}}^\infty \left(\left|(y\lambda_+)^{1/\alpha_1} - (y\lambda_+)^{1/\alpha_2}\right| \wedge 1 \right) \frac{dy}{y^2} = \lambda_+ \int_0^1 \left(t^{-1/\alpha_1} - t^{-1/\alpha_2} \right) dt
$$

$$
\leqslant \frac{\alpha_1}{\alpha_1-1} - \frac{\alpha_2}{\alpha_2-1} \leqslant \alpha_2 - \alpha_1, \tag{5.31}
$$

with the help of the change of variables $t = 1/(y\lambda_+)$ and assuming that $2 < \alpha_1 < \alpha_2$, where we have dropped the $+$ superscript for convenience. The analogous computation for the second integral in (5.30) yields the estimate in Proposition (3.6). $\qquad\square$

References

[1] L. Arnold. Hasselmann's program revisited: The analysis of stochasticity in deterministic climate models. *in Imkeller, Peter (ed.) et al., Stochastic climate models. Basel: Birkhäuser. Prog. Probab.* 49:141-157, 2001.

[2] R. Benzi, G. Parisi, A. Sutera, A. Vulpiani. Stochastic resonance in climate change. *Tellus*, 34:10-16, 1982.

[3] N. Berglund, D. Landon. Mixed-mode oscillations and interspike interval statistics in the stochastic FitzHugh-Nagumo model. *Nonlinearity*, 25:2303-2335, 2012.

[4] N. Berglund, B. Gentz. Metastability in simple climate models: Pathwise analysis of slowly driven Langevin equations. *Stoch. Dyn.*, 2:327-356, 2002.

[5] Björn Böttcher, René Schilling, and Jian Wang. *Lévy Matters III, Lévy-Type Processes: Construction, Approximation and Sample Path Properties*, volume 2099 of *Lecture Notes in Mathematics*. Springer-Verlag, 2013.

[6] P.D. Ditlevsen. Observation of a stable noise induced millennial climate changes from an ice-core record. *Geophysical Research Letters*, 26 (10):1441–1444, 1999.

[7] J. Gairing, M. Högele, T. Kosenkova, A. Kulik: Coupling distances between Lévy measures and applications to noise sensitivity of SDE. Stochastics and Dynamics, 15 (2), 2015.

[8] J. Gairing, M. Högele, T. Kosenkova, A. Kulik: On the calibration of Lévy driven time series with coupling distances with an application in paleoclimate. Springer INdAM volume "Mathematical Paradigms of Climate Sciences", ISSN: 2281-518X

[9] J. Gairing, P. Imkeller. Stable CLTs and rates for power variation of α-stable Lévy processes. *Methodology and Computing in Applied Probability*, 1-18, 2013.

[10] I. I. Gikhman, A. V. Skorokhod, Stochastic diffential equations and their applications, 2nd edition Naukova dumka, Kyiv, 1982 (in Russian).

[11] C. Hein, P. Imkeller, and I. Pavlyukevich. Limit theorems for p-variations of solutions of SDEs driven by additive stable Lévy noise and model selection for paleo-climatic data. *Interdisciplinary Math. Sciences*, 8:137-150, 2009.

[12] A. Debussche, M. Högele, P. Imkeller. The dynamics of non-linear reaction-diffusion equations with small Lévy noise. *Springer Lecture Notes in Mathematics, Vol. 2085*, 2013.

[13] C. Doss, M. Thieullen. Oscillations and random perturbations of a FitzHugh-Nagumo system. *Preprint hal-00395284*, 2009.

[14] K. Hasselmann. Stochastic climate models: Part I. Theory. *Tellus*, 28:473–485, 1976.

[15] M. Högele, I. Pavlyukevich. The exit problem from the neighborhood of a global attractor for heavy-tailed Lévy diffusions. *Stochastic Analysis and Applications*, 32 (1), 163–190 (2014).

[16] P. Imkeller, A. Monahan. Conceptual stochastic climate models. *Stochastics and Dynamics*, 2:311–326, 2002.

[17] P. Imkeller. Energy balance models: Viewed from stochastic dynamics. *Prog. Prob.*, 49:213–240, 2001.

[18] P. Imkeller and I. Pavlyukevich. First exit times of SDEs driven by stable Lévy processes. *Stochastic Processes and their Applications*, 116(4):611–642, 2006.

[19] P. Imkeller and I. Pavlyukevich. Metastable behaviour of small noise Lévy-driven diffusions. *ESAIM: Probability and Statistics*, 12:412–437, 2008.

[20] N. Ikeda and S. Watanabe. Stochastic differential equations and diffusion processes. North-Holland, Kodansha ltd., Tokyo, 1981.

[21] T. Kosenkova, A. Kulik: Transportation distance between the Lévy measures and stochastic equations for Lévy-type processes. Modern Stochastics: Theory and Applications, 1, 2014, p.49-64.

[22] I. Pavlyukevich. First exit times of solutions of stochastic differential equations with heavy tails. *Stochastics and Dynamics*, 11, Exp.No.2/3:1–25, 2011.

[23] Philip E. Protter. *Stochastic integration and differential equations*. Springer-Verlag Berlin Heidelberg, Berlin, 2004.

[24] S. T. Rachev (ed.). Handbook of heavy-tailed distributions in finance. *Elsevier/North-Holland*. Amsterdam, 2003.

[25] S. T. Rachev and L. Rüschendorf. Mass Transportation Problems. Vol.I: Theory, Vol.II.: Applications. *Probability and its Applications*. Springer-Verlag, New York 1998.

[26] K. Sato, *Lévy processes and infinitely divisible distributions*, vol. 68 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1999.

[27] H.C. Tuckwell, R. Rodriguez, F.Y.M. Wan, Determination of firing times for the stochastic FitzHugh-Nagumo neuronal model. *Neural Computation*, 15:143 – 159, 2003.