Stochastic systems with memory and jumps

D.R. Baños,* F. Cordoni† G. Di Nunno‡ L. Di Persio§ and E.E. Røse¶

October 23rd, 2018

Abstract

Stochastic systems with memory naturally appear in life science, economy, and finance. We take the modelling point of view of stochastic functional delay equations and we study these structures when the driving noises admit jumps. Our results concern existence and uniqueness of strong solutions, estimates for the moments and the fundamental tools of calculus, such as the Itô formula. We study the robustness of the solution to the change of noises. Specifically, we consider the noises with infinite activity jumps versus an adequately corrected Gaussian noise. The study is presented in two different frameworks: we work with random variables in infinite dimensions, where the values are considered either in an appropriate L^p-type space or in the space of càdlàg paths. The choice of the value space is crucial from the modelling point of view, as the different settings allow for the treatment of different models of memory or delay. Our techniques involve tools of infinite dimensional calculus and the stochastic calculus via regularisation.

Keywords: Stochastic delay equations, memory, jump diffusions, Itô formula, moment estimates, calculus via regularisation.

AMS classification: 34K50, 60H07

1 Introduction

Delay equations are differential equations whose coefficients depend also on the past history of the solution. Besides being of mathematical interest on their own, delay equations naturally arise in many applications, ranging from mathematical biology to mathematical finance, where often the effect of the memory or delay on the evolution of the system cannot be neglected, we refer to [8, 9, 28, 29, 32, 33, 36] and references therein for applications in different areas.

When dealing with a delay differential equation (DDE), one cannot in general relay on standard existence and uniqueness theorems, but ad hoc results have to be proven. In general this is done by lifting the DDE, from having a solution with values in a finite dimensional state space, such as $\mathbb{R}^d$, to having values in an infinite dimensional path space, which has to be carefully chosen according to the specific problem. For the case of *deterministic* delay differential equations an extensive literature exists, we refer the reader to the monographs [21, 24] for details.

When considering *stochastic* delay differential equations (SDDE), that is DDE perturbed by a stochastic noise, one encounters problems that did not appear in the deterministic case or in

---

*Institute of Mathematics of the University of Barcelona, University of Barcelona, Gran Via de les Corts Catalanes 585, 08007, Barcelona, and Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, N-0316 Oslo, Norway, Email: davidru@math.uio.no

†Department of Mathematics, University of Trento, Via Sommarive, 14, 38123 Trento, Italy. Email: francesco.cordoni@unitn.it

‡Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, N-0316 Oslo, Norway, and Norwegian School of Economics and Business Administration, Helleveien 30, N-5045 Bergen, Norway. Email: giulian@math.uio.no

§Department of Computer Science, University of Verona, Strada le Grazie, 15, 37134 Verona, Italy. Email: luca.dipersio@univr.it

¶Department of Mathematics, University of Oslo, P.O. Box 1053 Blindern, 0316 Oslo, Norway. Email: elinero@math.uio.no

classical stochastic differential equations. In particular the SDDE fails to satisfy the Markov property, hence one cannot rely on the well established setting of Markov processes for the study of the solution. As in the deterministic case, however, one can apply the key idea to lift the SDDE to have values in a suitable infinite dimensional path space. In doing so, one is able to recover the Markov property, nevertheless the main drawback is that now one is dealing with an infinite dimensional stochastic partial differential equation (SPDE). Although a well established theory for SPDE's exists, some fundamental results, known in the finite dimensional case, fail to hold true in the infinite dimension. In particular when considering infinite dimensional SPDE's, the concept of *quadratic variation* is not a straightforward generalisation of the classical notion of quadratic variation. We recall that this concept is crucial in essential tools of stochastic analysis, such as the Itô formula. Some concrete results around the concept of quadratic variation in infinite dimensions have appeared only recently, see [20].

SDDE's have been first studied in the seminal works [10, 35] and then extensively studied in [26, 36, 44], though in a different, but yet related setting. Recently there has been a renewed interest in SDDE's motivated by financial applications. In [23] a path dependent stochastic calculus was first suggested and then widely developed in [12, 13].

From the stochastic calculus point of view, the technique of regularisation, recently introduced, proved to be powerful to define a stochastic integral and to prove a general Itô formula for stochastic differential equations both in finite and infinite dimensions. The first results exploiting the stochastic calculus via regularisation are found in [41, 42], where a generalisation of the Itô formula was proved. More recently in [19] a new concept of quadratic variation for Banach space-valued processes was proposed and applied to prove a suitable Itô formula for infinite dimensional stochastic processes. This triggered a stream of studies aimed at deriving a suitable Itô's formula for delay equations and at studying deterministic problems that can be tackled by a stochastic approach. Particular attention was given to the Kolmogorov equation. We refer to [14, 16, 15, 25, 26]. Eventually in [16, 25] the relationship between the *path-wise calculus* and the *Banach-space* calculus was detailed.

We remark that all the aforementioned results for delay equations are proved in the case when the driving noise is continuous, such as for a standard Brownian motion. Very few results exist when the noise allows for random jumps to happen, see, e.g. [39, 40].

As mentioned above, there are different approaches to deal with SDDE's. We work with the setting of stochastic functional delay differential equations (SFDDE), first introduced in [35] and further developed in [36, 44]. This choice is motivated by the fact that it appears to be the right compromise between a general purely infinite dimensional SPDE and a classical finite dimensional SDE. In fact, even if the stochastic delay equation is treated as an infinite dimensional equation, there is always a clear connection with its finite dimensional realisations, so that standard finite dimensional Itô calculus can be often used.

The aim of the present paper is to extend the theory of SFDDE, studied in [35] for a Brownian driving noise, to include jumps, that is to deal with noises of jump-diffusion type. Specifically we aim at settling the existence and unicity of solutions, and derive the fundamental tools of a stochastic calculus for SFDE's with jumps. We also study the robustness of the solutions of SFDDEs to changes of the driving noise. This is an important analysis in view of the future applications. From a finite dimensional point of view this was studied in e.g. [6].

We consider an $\mathbb{R}^d-$valued SDE of the form

$$
\begin{aligned}
dX(t) =&f(t, X(t), X(t + \cdot))dt + g(t, X(t), X(t + \cdot))dW(t) \\
&+ \int_{\mathbb{R}_0} h(t, X(t), X(t + \cdot))(z)\tilde{N}(dt, dz)\,, \quad t \in [0, T],
\end{aligned}
\tag{1.1}
$$

where W is a standard Brownian motion, $\tilde{N}$ is a compensated Poisson random measure where $\mathbb{R}_0 := \mathbb{R} \setminus \{0\}$, and f, g and h are some given suitable functional coefficients. With the notation $X(t + \cdot)$ we mean that the coefficient may depend also on the past values of the solution on the interval $[t - r, r]$ for some fixed delay $r > 0$. It is this dependence on the past values of the evolution that is identified as memory or, equivalently, delay. The formal introduction of the

current notation will be carried out in the next section. Notice that at this stage equation (1.1) is a finite dimensional SDE with values in $\mathbb{R}^d$.

We now *lift the process* (1.1) to have values in a suitable infinite dimensional path space. The choice of the suitable space is truly a key issue. As illustration, consider the purely diffusive case and denote the maximum delay appearing in (1.1) by $r > 0$. Then, in [44] a product space of the form $M^p := L^p([-r, 0]; \mathbb{R}^d) \times \mathbb{R}^d$, $p \in [2, \infty)$, was chosen, whereas in [36] the space of continuous functions $\mathcal{C} := \mathcal{C}([-r, 0]; \mathbb{R}^d)$ was taken as reference space. With the former choice one can rely on well-established results and techniques for L^p−spaces. Nevertheless this choice may seem artificial when dealing with a past path-dependent memory. For this reason the second choice, the space of continuous functions, is often considered the "right" space where to study delay equations, though it requires mathematically careful considerations.

When we consider the *lifting* of equation (1.1) to an infinite dimensional framework, we have to be cautious. The lift of SFDDE's of the jump-diffusion type presents two naturally possible choices of setting: the product space M^p and the space of càdlàg (right continuous with finite left limit) functions $\mathcal{D} := \mathcal{D}([-r, 0]; \mathbb{R}^d)$. We decided to carry out our study in both settings in order to give a general comprehensive and critical presentation of when and in what sense it may be more suitable to treat the study in the one or the other setting. To explain, on the one side, we have the inclusion $\mathcal{D} \subset M^p$, with the injection being continuous, so that the M^p−setting appears to be more general, on the other side we can see that the existence and uniqueness of the solution of an SFDDE cannot be established in full generality in the space M^p. This, in fact, depends on the type of delay or memory. The drawback of the M^p approach is that it does not apply, for instance, to SFDDE's with *discrete delay*, as e.g.

$$X(t) = \int_0^t X(s + \rho)ds + \int_0^t X(s + \rho)dW(s) + \int_0^t \int_{\mathbb{R}_0} X(t + \rho)z\tilde{N}(dt, dz), \qquad (1.2)$$

where $\rho \in [-r, 0)$ is a fixed parameter. We refer to the discussion in Remark 2.2 for details. This case, can be well treated in the $\mathcal{D}$-setting.

In the sequel, we thus lift equation (1.1) to have values either in M^p, with $p \in [2, \infty)$, or in $\mathcal{D}$, exploiting the notion of *segment*. We denote X_t the *segment* of the process X on an interval $[t - r, t]$, that is

$$X_t := \{X(t + \theta) : \theta \in [-r, 0]\},$$

being $r \geqslant 0$ the maximum delay. We denote the present value of the process at time t by $X(t)$. We then study an SFDDE of the form,

$$dX(t) = f(t, X_t)dt + g(t, X_t)dW(t) + \int_{\mathbb{R}_0} h(t, X_t)(z)\tilde{N}(dt, dz) \qquad (1.3)$$

$$X_0 = \eta$$

where η is a function on $[-r, 0]$.

In this paper, first we establish existence, uniqueness and moment estimates for the equation (1.3) where the *segment* X_t takes values either in $\mathcal{D}$ or it is considered as the pair $(X_t \mathbb{1}_{[-r,0)}, X(t))$ with values in M^p, depending on the framework chosen.

Then we look at the robustness of the model to changes of the noise. In particular, we study what happens if we replace the small jumps of the infinite activity Poisson random measure N, by a continuous Brownian noise B. This is done by comparing a process X with dynamics

$$dX(t) = f(t, X_t)dt + g(t, X_t)dW(t) + \int_{\mathbb{R}_0} h_0(t, X_t)\lambda(z)\tilde{N}(dt, dz) \qquad (1.4)$$

$$X_0 = \eta,$$

to the process $X^{(\epsilon)}$ defined by

$$dX^{(\epsilon)}(t) = f(t, X_t^{(\epsilon)})dt + g(t, X_t^{(\epsilon)})dW(t) + h_0(t, X_t^{(\epsilon)})\int_{|z|<\epsilon} |\lambda(z)|^2\nu(dz)dB(t)$$

$$+ \int_{|z| \geqslant \epsilon} h_0(t, X_t^{(\epsilon)}) \lambda(z) \tilde{N}(dt, dz) \tag{1.5}$$

$$X_0^{(\epsilon)} = \eta \,.$$

We remark that the choice of this approximate guarantees the same so-called total volatility, using a terminology from financial modelling.

Eventually, exploiting the *stochastic calculus via regularisation* we prove an Itô type formula for stochastic delay equations with jumps, showing that the results are in fact coherent with the results obtained in [36, 44]. We work with forward integrals in the following sense. For the stochastic processes $X = \{X_s, s \in [0, T]\}$, and $Y = \{Y_s, s \in [0, T]\}$, as mappings on $[0, T] \times \Omega$ and taking values, respectively, in $L^p([-r, 0], \mathbb{R}^d)$ and its topological dual, we define the *forward integral* of Y against X as

$$\int_0^t {}_q \langle Y_s \,, dX_s \rangle_p := \lim_{\epsilon \searrow 0} \int_0^t \left\langle Y_s \,, \frac{X_{s+\epsilon} - X_s}{\epsilon} \right\rangle_p ds \,, \tag{1.6}$$

where the limit holds in *probability*. Here we denoted the paring between $L^p([-r, 0], \mathbb{R}^d)$ and its dual by $\langle \cdot, \cdot \rangle_p$. Furthermore, if the above limit holds *uniformly in probability on compact sets* (ucp), we immediately have that the process

$$\left(\int_0^t {}_q \langle Y_s \,, dX_s \rangle_p \right)_{t \in [0, T]} \,,$$

admits a càdlàg version and we say that the *forward integral* exists. When the two processes have values in the space M^p and M^{p*}, we are able to show that the above limit holds in fact ucp, characterizing thus the *forward integral* in terms of the derivative of the process X, which coincides with the operators introduced in [44] and in [25].

The present work is structured as follows. In Section 2 we introduce the main notation used throughout the whole paper. In Section 2.3 we study existence and uniqueness results for equation (1.3) with values in $\mathcal{D}$, whereas in Section 2.4 we prove the same results in the M^p setting. Then, in Section 2.5 we prove the robustness of equation (1.3) to the change of the noise. Eventually in Section 3 we prove a suitable Itô-type formula for SFDDE's with values in M^p and in $\mathcal{D}$.

2 Stochastic functional differential equations with jumps

2.1 Notation

Let $(\Omega, \mathcal{F}, \mathbb{F} := \{\mathcal{F}_t\}_{t \in [0, T]}, P)$ be a complete, filtered probability space satisfying the usual hypotheses for some finite time horizon $T < \infty$. Let $r \geqslant 0$ be a non-negative constant denoting the maximum delay of the equations considered. We extend the filtration by letting $\mathcal{F}_s = \mathcal{F}_0$ for all $s \in [-r, 0]$. This will still be denoted by $\mathbb{F}$. Let $W = (W^1, \dots, W^m)^\mathsf{T}$ be an m-dimensional $\mathbb{F}$-adapted Brownian motion and $N = (N^1, \dots, N^n)^\mathsf{T}$ be the jump measures associated with n independent $\mathbb{F}$-adapted Lévy processes, with Lévy measures $\nu = (\nu_1, \dots, \nu_n)$ respectively. Recall $\mathbb{R}_0 := \mathbb{R} \setminus \{0\}$. We denote by $\tilde{N}$, the compensated Poisson random measure

$$\tilde{N}(dt, dz) := (N^1(dt, dz) - \nu_1(dz)dt, \dots, N^n(dt, dz) - \nu_n(dz)dt)^\mathsf{T} \,.$$

Consider the equation

$$dX(t) = f(t, X_t)dt + g(t, X_t)dW(t) + \int_{\mathbb{R}_0} h(t, X_t)(z)\tilde{N}(dt, dz) \tag{2.1}$$

$$X_0 = \eta,$$

where f, g and h are some given functionals on a space containing the segments X_t, $t \in [0, T]$ of the process X. We will give precise definitions of the segments and the coefficient functionals below.

Equations of the form (2.1) will be referred to as *stochastic functional delay differential equation* (SFDDE).

We remark that equation (2.1) is to be interpreted component-wise as a system of SFDDE of the following form:

$$dX^i(t) = f^i(t, X_t)dt + \sum_{j=1}^{m} g^{i,j}(t, X_t)dW^j(t) + \sum_{j=1}^{n} \int_{\mathbb{R}_0} h^{i,j}(t, X_t, z)\tilde{N}^j(dt, dz),$$

$$X_0^i = \eta^i, \quad i = 1, \dots, d.$$

With the component-wise interpretation in mind, it is natural to require that the images $f(t, X_t)$ and $g(t, X_t)$ of the coefficient functionals f and g are contained in the spaces $L^2(\Omega, \mathbb{R}^d)$ and $L^2(\Omega, \mathbb{R}^{d \times m})$ respectively. Similarly, we want the image $h(t, X_t)$ of h to be contained in a set of matrices with j'th column in $L^2(\Omega, L^2(\nu_j, \mathbb{R}^d))$. To express the space of all such matrices in a compact manner, we introduce the following notation. For the $\mathbb{R}^n$-valued measure $\nu = (\nu_1, \dots, \nu_n)^{\mathsf{T}}$, we will write $L^p(\nu) = L^p(\nu, \mathbb{R}^{d \times n})$ $(p \geqslant 2)$, to denote the set of measurable functions

$$H : \mathbb{R}_0 \to \mathbb{R}^{d \times n},$$

such that

$$\|H\|_{L^p(\nu)}^p := \sum_{j=1}^{n} \|H^{\cdot j}\|_{L^p(\nu_j, \mathbb{R}^d)}^p < \infty. \tag{2.2}$$

Here we have used the notation $H^{\cdot j}$ to denote the j'th column of H. Notice also that the Bochner space

$$L^q(\Omega, L^p(\nu, \mathbb{R}^{d \times n})) \quad (q \geqslant 2)$$

consists of the measurable functions $\mathcal{H} : \Omega \mapsto L^p(\nu, \mathbb{R}^{d \times n})$ such that

$$\|\mathcal{H}\|_{L^p(\Omega, L^p(\nu, \mathbb{R}^{d \times n}))}^q := E[\|\mathcal{H}\|_{L^p(\nu, \mathbb{R}^{d \times n})}^q] < \infty. \tag{2.3}$$

For convenience, we will sometimes omit to explicitly specify the spaces $\mathbb{R}^d$, $\mathbb{R}^{d \times m}$ and $\mathbb{R}^{d \times n}$, when it is clear from the context which space to consider and no confusion is possible. In this paper, when no confusion will occur, the standard Euclidean norm $\| \cdot \|_{\mathbb{R}^{k \times l}}$, will be denoted by $|\cdot|$ for any $k, l \in \mathbb{N}$.

Hereafter we introduce the relevant spaces we work with in the sequel. For $0 \leqslant u \leqslant T$, let

$$\mathcal{D}_u := \mathcal{D}([-r, u], \mathbb{R}^d), \tag{2.4}$$

denote the space of all càdlàg functions from $[-r, u]$ to $\mathbb{R}^d$, equipped with the uniform norm

$$\|\eta\|_{\mathcal{D}_u} := \sup_{-r \leqslant \theta \leqslant u} \{|\eta(\theta)|\}, \quad \eta \in \mathcal{D}_u. \tag{2.5}$$

Set $\mathcal{D} := \mathcal{D}_0$. For $2 \leqslant p < \infty$, let $L_u^p := L^p([-r, u], \mathbb{R}^d)$ and

$$M_u^p := L_u^p \times \mathbb{R}^d$$

with norm given by

$$\|(\eta, v)\|_{M_u^p}^p := \|\eta\|_{L_u^p}^p + |v|^p, \quad \eta \in M_u^p$$

Set $L^p := L_0^p$ and $M^p := M_0^p$.

We recall that the M_u^p-spaces are separable Banach spaces and M_u^2 is also a Hilbert space. On the other side $\mathcal{D}_u$ equipped with the topology given by (2.5) is a non-separable Banach space. The

space $\mathcal{D}_u$ equipped with the Skorohod topology is a separable metric space. Moreover, there exists also a topology on $\mathcal{D}_u$, equivalent to the Skorohod topology, such that $\mathcal{D}_u$ is a complete separable metric space. See e.g. [7, 37].

Observe that if $\eta \in \mathcal{D}$, then

$$\|(\eta \mathbb{1}_{[-r,0)}, \eta(0))\|_{M^p}^p = \|\eta \mathbb{1}_{[-r,0)}\|_{L^p}^p + |\eta(0)|^p \leqslant (r+1)\|\eta\|_{\mathcal{D}}^p. \tag{2.6}$$

By (2.6), and since the elements in M^p have at most one càdlàg representative, the linear functional

$$\eta \mapsto (\eta \mathbb{1}_{[-r,0)}, \eta(0))$$

is a linear continuous embedding of $\mathcal{D}$ into M^p. Note that we will write $\|\eta\|_{M^p}$ in place of $\|(\eta \mathbb{1}_{[-r,0)}, \eta(0))\|_{M^p}$.

We now introduce the notion of *segment* that will play an important role in this paper.

Definition 2.1. *For any stochastic process $Y : [-r, T] \times \Omega \to \mathbb{R}^d$, and each $t \in [0, T]$, we define the* segments

$$Y_t : [-r, 0] \times \Omega \to \mathbb{R}^d, \quad \text{by} \quad Y_t(\theta, \omega) := Y(t+\theta, \omega), \quad \theta \in [-r, 0], \quad \omega \in \Omega.$$

In view of the arguments above, for each t, the segment can also be regarded as a function

$$\Omega \ni \omega \longmapsto Y_t(\cdot, \omega) \in \mathcal{D}$$

or

$$\Omega \ni \omega \longmapsto \left(Y_t(\cdot, \omega) \mathbb{1}_{[-r,0)}(\cdot), Y(t, \omega)\right) \in M^p$$

depending on the type of memory or delay in the process and provided the necessary conditions of càdlàg paths or integrability. Indeed it is this interpretation that allows for the *lifting* of equation (1.1) to an infinite dimensional framework, as anticipated in the introduction.

Remark 2.2. *We need here to stress that the two frameworks $\mathcal{D}$ or M^p are not equivalent when jumps are present and one has to choose where to set the study depending on what type of memory the model involves. As illustration, consider the case of discrete delay, as in equation (1.2) for example. There the term $X(t+\rho)$ with fix $\rho \in [-r, 0)$, corresponds to the segment $X_t = \{X(t+s), s \in [-r, 0]\}$ evaluated at the point $s = \rho$. The evaluation is not a well-defined operation in M^p. To see this, consider two elements $\eta_1, \eta_2 \in \mathcal{D}$ such that $\eta_1(s) = \eta_2(s)$ for all $s \in [-r, 0] \setminus \{\rho\}$ and $\eta_1(\rho) \neq \eta_2(\rho)$. These are clearly distinguished elements in $\mathcal{D}$. However, if we were considering them as pairs $(\eta_1, \eta_1(0)), (\eta_2, \eta_2(0))$, then they would be two representatives of the same class in M^p. Thus, for (ω, t) fixed, the evaluation $X_t(\rho) = X(t+\rho)$ is not uniquely determined in M^p.*

We recall the following definitions. Let $\mathcal{G} \subseteq \mathcal{F}$ be a P-augmented σ-algebra of events in Ω. Let $\mathcal{D}$ be equipped with the σ-algebra $\mathfrak{D}$ generated by the Skorohod topology.

Definition 2.3. *We say that a function $\eta : \Omega \to \mathcal{D}$ is a ($\mathcal{G}$-measurable) $\mathcal{D}$-valued random variable if it is $\mathcal{G}$-measurable with respect to the σ-algebra $\mathfrak{D}$ or, equivalently, we can say that η is $\mathcal{G}$-measurable if for every $\theta \in [-r, 0]$, the $\mathbb{R}^d$-valued function $\omega \mapsto \eta(\theta, \omega)$ is $\mathcal{G}$-measurable.*

Recall that $\mathfrak{D}$ coincides with the projection σ-algebra and that $\mathfrak{D} \subsetneq \mathcal{B}(\mathcal{D})$, where $\mathcal{B}(\mathcal{D})$ is the Borel σ-algebra generated by the topology given by the norm (2.5).

Definition 2.4. *We say that a function $(\eta, v) : \Omega \to M^p$ is a ($\mathcal{G}$-measurable) M^p-valued random variable if it is measurable with respect to the σ-algebras $\mathcal{G}$ and $\mathcal{B}(M^p)$, or equivalently if the function*

$$\omega \mapsto \int_{-r}^0 \eta(\theta, \omega) \phi(\theta) d\theta + v(\omega) \cdot u$$

is $\mathcal{G}$-measurable for every $(\phi, u) \in M^{p*} = M^{\frac{p}{p-1}}$.

Notice also that if η is a $\mathcal{G}$-measurable $\mathcal{D}$-valued-random variable, then it is $\mathcal{G}$-measurable as an M^p-valued random variable. Corresponding definitions apply in the cases of the D_u or M_u^p spaces above.

We are now ready to introduce the spaces of measurable $\mathcal{D}$-valued and M^p-valued random variables.

Recall that $\mathcal{D}_u$ is equipped with the σ-algebra $\mathfrak{D}_u$ generated by the Skorohod topology on $\mathcal{D}_u$. Let η be a $\mathcal{D}_u$-valued random variable. For $p \geqslant 2$, define

$$\|\eta\|_{S^p(\Omega;\mathcal{D}_u)}^p := E\left[\sup_{\theta \in [-r,u]} |\eta(\theta)|^p\right] = E[\|\eta(\theta)\|_{\mathcal{D}_u}^p],$$

and the equivalence relation $\eta_1 \sim \eta_2 \Leftrightarrow \|\eta_1 - \eta_2\|_{S^p(\Omega;\mathcal{D}_u)} = 0$. Let

$$S^p(\Omega, \mathcal{G}; \mathcal{D}_u)$$

denote the space of equivalence classes of $\mathcal{D}$-valued random variables $\omega \mapsto \eta(\omega, \cdot)$ such that $\|\eta\|_{S^p(\Omega;\mathcal{D}_u)}^p < \infty$. We have used the notation $S^p(\Omega, \mathcal{G}; \mathcal{D}_u)$ to distinguish the space above from $L^p(\Omega, \mathcal{G}; \mathcal{D}_u)$, which consists of the equivalence classes of the $(\mathcal{G}, \mathcal{B}(\mathcal{D}_u))$-measurable $\mathcal{D}_u$-valued functions.

Remark 2.5. *In the continuous setting (see e.g. [36]), the segments of an SFFDE are often considered as elements of the Bochner space $L^2(\Omega; \mathcal{C})$, where $\mathcal{C}$ denotes the set of continuous functions from $[-r, 0]$ to $\mathbb{R}^d$. We remark that the càdlàg counterpart, namely the Bochner space $L^p(\Omega, \mathcal{G}; \mathcal{D})$, turns out to be too restrictive to contain a sufficiently large class of càdlàg segments. This can bee seen from the following lemma:*

Lemma 2.6. *Suppose that X is a càdlàg Lévy-Itô process with $X \in L^p(\Omega, \mathcal{G}; \mathcal{D}[a, b])$. Then X is continuous with probability 1.*

To see why this holds, we first recall that by an equivalent definition of Bochner spaces (see [22] for more on these spaces), $L^p(\Omega, \mathcal{G}; \mathcal{D})$ consists of equivalence classes of the $(\mathcal{G}, \mathcal{B}(\mathcal{D}))$-measurable functions $X : \Omega \to \mathcal{D}$ such that the image $X(\Omega_0)$ is separable for some subset $\Omega_0 \subset \Omega$ with $P(\Omega_0) = 1$, and $E[\|X\|_{\mathcal{D}([a,b])}^p] < \infty$ holds[1]. By [27, lemma 9.12], we know that $X(\Omega_0)$ is separable if and only if there exist a countable set $\mathbb{T}_0 \in [a, b]$, such that $\Delta X(t, \omega) = 0$ whenever $t \notin \mathbb{T}_0, \omega \in \Omega_0$. In other words, except for a negligible set of sample paths of X, all the jumps of X occur at a countable number of times.

Proof of Lemma 2.6. Since $X \in L^p(\Omega, \mathcal{G}; \mathcal{D}([a, b]))$, we can choose $\Omega_0, \mathbb{T}_0$ be as above. Now since X is a càdlàg Lévy-Itô process, it also holds that $P(\omega : \Delta X(t, \omega) \neq 0) = 0$ for every t, and hence

$$\mathcal{N} := \bigcup_{t \in \mathbb{T}_0} \{\omega \in \Omega_0 : \Delta X(t, \omega) \neq 0\}$$

is a null set. But then if $\omega \in \Omega_0 \setminus \mathcal{N}$, it holds that $\Delta X(t, \omega) = 0$ for every t, that is X is continuous on $\Omega_0 \setminus \mathcal{N}$ and $P(\Omega_0 \setminus \mathcal{N}) = 1$. $\square$

We also consider the following space. For $p \geqslant 2$, let

$$L^p(\Omega, \mathcal{G}; M_u^p),$$

denote the Bochner spaces $L^p(\Omega, M_u^p)$ consisting of the M_u^p-valued random variables (η, v) such that the norm given by

$$\|(\eta, v)\|_{L^p(\Omega;M_u^p)}^p := E[\|(\eta, v)\|_{M_u^p}^p]$$

_____
[1]in fact this definition is valid for $\mathcal{D}([a, b])$ replaced by any Banach space V

is finite. We recall that both $S^p(\Omega; \mathcal{D}_u)$ and $L^p(\Omega; M_u^p)$ are Banach spaces. Observe that if $\eta \in S^p(\Omega, \mathcal{G}; \mathcal{D})$, then

$$\|(\eta, \eta(0))\|_{L^p(\Omega; M^p)}^p \leqslant (r+1)\|\eta\|_{S^p(\Omega; \mathcal{D})}^p \tag{2.7}$$

thus, it also holds that

$$S^p(\Omega, \mathcal{G}; \mathcal{D}) \subset L^p(\Omega, \mathcal{G}; M^p),$$

and the embedding is continuous. With the appropriate boundedness and integrability conditions on a càdlàg adapted process Y, then for each t, the segment Y_t can be regarded as an element in the spaces $S^p(\Omega, \mathcal{F}_t; \mathcal{D})$ or $L^p(\Omega, \mathcal{F}_t; M^p)$.

In line with the definitions given above, we also use the following notation for any $u \in [0, T]$ and $2 \leqslant p < \infty$. Let

$$S_{ad}^p(\Omega, \mathcal{F}_u; \mathcal{D}_u) \subseteq S^p(\Omega, \mathcal{F}_u; \mathcal{D}_u)$$

denote the subspace of elements in $S^p(\Omega, \mathcal{F}_u; \mathcal{D}_u)$ admitting a $\mathbb{F}$-adapted representative. We remark that if $Z \in S^p(\Omega, \mathcal{F}_T; \mathcal{D}_T)$, then we have that

$$\|Z\|_{S^p(\Omega; \mathcal{D})} \leqslant \|Z\|_{S^p(\Omega; \mathcal{D}_t)} \leqslant \|Z\|_{S^p(\Omega; \mathcal{D}_T)}. \tag{2.8}$$

Also, consider the Banach space

$$L^p(\Omega; L_u^p)$$

with the usual norm given by:

$$\|Y\|_{L^p(\Omega; L_u^p)}^p := E\left[\|Y\|_{L_u^p}^p\right] < \infty.$$

Remark 2.7. *For later use, we remark that, if the stochastic process $Y : [-r, T] \times \Omega \longrightarrow \mathbb{R}^d$ is also $Y \in L^p(\Omega; L_T^p)$, then it make sense to consider the process of the segments $\{X_t, t \in [0, T]\}$, as a mapping*

$$\omega \longmapsto \left\{ t \mapsto \left(X_t(\cdot, \omega) \mathbb{1}_{[-r,0)}(\cdot), X(t, \omega) \right) \right\}.$$

which is well-defined in $L^p([0, T], M^p)$.

Let

$$L_{ad}^p(\Omega; L_u^p) \subseteq L^p(\Omega; L_u^p)$$

denote the subspace of elements admitting $\mathbb{F}$-adapted representative.

Suppose now that $Y \in L_{ad}^p(\Omega; L_T^p)$. Since $Y(t)$ is well-defined for a.e. $t \in [-r, T]$, it makes sense to consider the segments Y_t as elements in $L^p(\Omega, \mathcal{F}_t; L_t^p)$ for a.e. t. Also,

$$\begin{aligned}
\int_0^u \|(Y_t, Y(t))\|_{L^p(\Omega; M^p)}^p dt &\leqslant \int_0^u \|Y_t\|_{L^p(\Omega; L^p)}^p dt + \int_{-r}^u \|Y(t)\|_{L^p(\Omega; \mathbb{R}^d)}^p dt \\
&= \int_0^u \int_{t-r}^t E[|Y(s)|^p] ds\, dt + \|Y\|_{L^p(\Omega; L_u^p)}^p \\
&\leqslant (u+1)\|Y\|_{L^p(\Omega; L_u^p)}^p.
\end{aligned} \tag{2.9}$$

Even though we can not consider $S_{ad}^p(\Omega; \mathcal{D}_t)$ as a subspace of $L_{ad}^p(\Omega; L_t^p)$, since the function

$$S_{ad}^p(\Omega; \mathcal{D}_t) \ni \eta \mapsto \eta \in L_{ad}^p(\Omega; L_t^p)$$

is not injective, this function is continuous, and

$$\|Y_t\|_{L^p(\Omega; L_t^p)}^p \leqslant (t+r)\|Y_t\|_{S^p(\Omega; \mathcal{D}_t)}^p. \tag{2.10}$$

2.2 Examples

To illustrate some possible ways to model memory or delay in a stochastic differential equation, we include some examples of delay terms appearing in applications.

i) **Distributed delay**: the functional

$$S_t \longmapsto \int_{-r}^{0} S(t+\theta)\alpha(d\theta)\,. \tag{2.11}$$

where α is a finite Borel measure on $[-r, 0]$, is an example of a *distributed delay*-functional. This is a general type of delay in the sense that examples ii), iii) below, can be regarded as particular cases of this one.

A general financial framework in this setting has been studied in [8, 9] where the authors considered a price evolution for the stock of the form

$$dS(t) = M(S_t)dt + N(S_t)dW(t) = \int_{-r}^{0} S(t+s)\alpha_M(ds)dt + \int_{-r}^{0} S(t+s)\alpha_N(ds)dW(t)\,,$$

α_M and α_N being suitable functions of bounded variation. See also [36, Sec. V], where α is taken as a *probability measure.*

ii) **Absolutely continuous distributed delay**: in the particular case $\alpha << \mathcal{L}$, where we have denoted by $\mathcal{L}$ the Lebesgue measure, we have that the measure α admits a density $\kappa := \frac{d\alpha}{d\mathcal{L}}$. Therefore the functional (2.11) reads as

$$S_t \longmapsto \int_{-r}^{0} S(t+\theta)\kappa(\theta)d\theta,\,.$$

A more advanced example has been provided in [29] where a functional of the form

$$(t, S_t) \longmapsto \int_{-r}^{0} \ell(t, S(t+\theta))h(\theta)d\theta,$$

for some functional ℓ, has been treated.

iii) **Discrete delay**: if we let $\alpha = \delta_\tau$, in equation (2.11), where δ_τ is the Dirac measure concentrated at $\tau \in [-r, 0]$, then we have a *discrete delay* functional, namely

$$S_t \longmapsto \int_{-r}^{0} S(t+\theta)\delta_\tau(d\theta) = S(t-\tau)\,. \tag{2.12}$$

A discrete delay model using functionals on the form (2.12), is widely used in concrete applications, spanning from mathematical biology, as in the case of the delayed Lotka-Volterra model, see, e.g. [21, 32, 36], to mathematical finance, as it happens for the delayed Black-Scholes model, see, e.g. [3, 28], or for commodities markets, see, e.g., [33]. In particular, in [3], the authors give an explicit form for the price a European call option written on an underlying evolving as

$$dS(t) = \mu S(t-a)dt + \sigma(S(t-b))dW(t)\,,$$

for $\mu \in \mathbb{R}$ and a suitable function σ.

A particular case of the discrete delay example is the no delay case, i.e. $\tau = \delta_0$. A multiple delay case, can be defined by letting $\alpha = \sum_{i=1}^{N} \delta_{\tau_i}$, $\tau_i \in [-r, 0]$, $i = 1, 2, \ldots, N$.

iv) **Brownian delay**: our setting allows also to consider delays with respect to a Brownian motion, namely

$$S_t \longmapsto \int_{t-r}^{t} S(\theta) dW(\theta).$$

Hence this permits to take *noisy memory* models into account. These cases are arising e.g. in the modelling of stochastic volatility see, e.g. [28, 43] and when dealing with stochastic control problems, see e.g. [18].

v) **Lévy delay**: similarly to the Brownian delay, we can also consider a delay with respect to a square integrable Lévy process of the form

$$S_t \longmapsto \int_{t-r}^{t} S(\theta) dL(\theta).$$

Such type of delay has been employed in [43] in order to consider some stochastic volatility models related to energy markets.

vi) **Mean field delay**: we can consider a delay of the form

$$S_t \longmapsto E\left[\int_{-r}^{0} S(t+\theta)\,\alpha(d\theta)\right],$$

where α is as in example i), see e.g. [1].

2.3 $\mathcal{D}$ framework

Fix $p \in [2, \infty)$. Consider again the equation

$$dX(t) = f(t, X_t)dt + g(t, X_t)dW(t) + \int_{\mathbb{R}_0} h(t, X_t)(z)\tilde{N}(dt, dz) \tag{2.13}$$

$$X_0 = \eta,$$

In this section, we require that $f(t, \cdot)$, $g(t, \cdot)$, $h(t, \cdot)$ are defined on $S^p(\Omega, \mathcal{F}_t; \mathcal{D})$ for each fixed t. Therefore, we introduce the space

$$\mathbf{S}_p^{\mathbb{F}} := \{(t, \psi) \in [0, T] \times S^p(\Omega, \mathcal{F}; \mathcal{D}) \text{ such that } \psi \in S^p(\Omega, \mathcal{F}_t; \mathcal{D})\}, \tag{2.14}$$

as the domain of the coefficient functionals f, g, h in the SFDDE (2.13). In particular, we will require that:

$$f : \mathbf{S}_p^{\mathbb{F}} \to L^p(\Omega, \mathbb{R}^d)$$
$$g : \mathbf{S}_p^{\mathbb{F}} \to L^p(\Omega, \mathbb{R}^{d \times m}),$$
$$h : \mathbf{S}_p^{\mathbb{F}} \to L^p(\Omega, L^2(\nu, \mathbb{R}^{d \times n})).$$

Moreover,

$$\eta \in S^p(\Omega, \mathcal{F}_0; \mathcal{D})$$

To ensure that the integrals are well-defined, the following assumptions are imposed on the coefficient functionals f, g and h.

Assumption ($\mathcal{P}$). *Whenever $Y \in S_{ad}^p(\Omega; \mathcal{D}_T)$, the process*

$$[0, T] \times \Omega \times \mathbb{R}_0 \ni (t, \omega, z) \mapsto h(t, Y_t)(\omega)(z) \in \mathbb{R}^{d \times n} \tag{2.15}$$

has a predictable version, and

$$[0, T] \times \Omega \ni (t, \omega) \mapsto f(t, Y_t)(\omega) \in \mathbb{R}^d,$$
$$[0, T] \times \Omega \ni (t, \omega) \mapsto g(t, Y_t)(\omega) \in \mathbb{R}^d$$

have progressively measurable versions.

Predictable and progressive should be interpreted in the standard sense for $\mathbb{R}^k$-valued processes (see e.g. [2]). We emphasise that the integrals in (2.13) should be interpreted with respect to the predictable and progressive versions of the respective integrands. For a range of SFDE's likely to be encountered in applications, the assumption $\mathcal{P}$ is fairly easy to verify.

Example 2.8. *Most of the examples presented in Section 2.2 satisfy Assumption $\mathcal{P}$. For instance, the functional displayed in i), which is more general than ii), iii), is predictable whenever the point zero is not an atom of the measure α, i.e. the discrete delay in (2.12) is not allowed when $\tau = 0$. The mean-field delay in example vi) is deterministic and hence predictable. The Brownian delay can also be considered, since the process $t \mapsto \int_{t-r}^{t} S(\theta) dW(\theta)$ is a continuous martingale, in particular it admits a version with left-limits.*

Definition 2.9. *Suppose that the assumption $\mathcal{P}$ holds. We say that $X \in S_{ad}^p(\Omega; \mathcal{D}_T)$ is a strong solution to the equation (2.13) if for each $t \in [0, T]$*

$$X(t) = \eta(0) + \int_0^t f(s, X_s) ds + \int_0^t g(s, X_s) dW(s) + \int_0^t \int_{\mathbb{R}_0} h(s, X_s)(z) \tilde{N}(ds, dz) \qquad (2.16)$$

$$X_0 = \eta,$$

in $S_{ad}^p(\Omega; \mathcal{D}_T)$ (in particular P-a.s.). If the solution is unique, we will write $^\eta X$ to denote the solution of (2.16) with initial datum $^\eta X_0 = \eta$.

To prove existence and uniqueness of the solution of the SFDDE, we rely on the following result.

Lemma 2.10 (Kunita's inequality). *Let $q \geqslant 2$. Suppose that F, G and H are predictable processes taking values in $\mathbb{R}^d$, $\mathbb{R}^{d \times m}$ and $\mathbb{R}^{d \times n}$ respectively. If*

$$Y(t) = Y_0 + \int_0^t F(s) ds + \int_0^t G(s) dW(s) + \int_0^t \int_{\mathbb{R}_0} H(s, z) \tilde{N}(ds, dz), \quad t \in [0, T],$$

then there exists a constant $C = C(q, d, m, n, T)$, independent of the processes F, G and H and the initial value Y_0, such that whenever $t \leqslant T$ the following inequality holds

$$E[\sup_{0 \leqslant u \leqslant t} |Y(t)|^q] \leqslant C \Big\{ \|Y_0\|_{L^q(\Omega, \mathbb{R}^d)}^q + \int_0^t \Big(\|F(s)\|_{L^q(\Omega, \mathbb{R}^d)}^q + \|G(s)\|_{L^q(\Omega, \mathbb{R}^{d \times m})}^q$$

$$+ \|H(s)\|_{L^q(\Omega, L^q(\nu))}^q + \|H(s)\|_{L^q(\Omega, L^2(\nu))}^q \Big) ds \Big\} \qquad (2.17)$$

For $n = 1$ (and arbitrary m and d), this is a rewritten version of Corollary 2.12 in [34]. We have justified the extension to general n in Appendix A.1.

2.3.1 Existence, uniqueness and moment estimates

Before giving sufficient conditions for existence and uniqueness of solutions to the equation (2.13), we will establish a set of hypotheses.

Assumption. $(\mathbf{D_1})$ *There exists $L > 0$, such that whenever $t \in [0, T]$ and $\eta_1, \eta_2 \in S^p(\Omega, \mathcal{F}_t; \mathcal{D})$, then*

$$\|f(t, \eta_1) - f(t, \eta_2)\|_{L^p(\Omega; \mathbb{R}^d)}^p + \|g(t, \eta_1) - g(t, \eta_2)\|_{L^p(\Omega; \mathbb{R}^{d \times n})}^p$$

$$+ \|h(t, \eta_1) - h(t, \eta_2)\|_{L^p(\Omega, L^p(\nu))}^p + \|h(t, \eta_1) - h(t, \eta_2)\|_{L^p(\Omega, L^2(\nu))}^p$$

$$\leqslant L \|\eta_1 - \eta_2\|_{S^p(\Omega; \mathcal{D})}^p.$$

$(\mathbf{D_2})$ *There exists $K > 0$, such that whenever $t \in [0, T]$ and $\eta \in S^p(\Omega, \mathcal{F}_t; \mathcal{D})$, then*

$$\|f(t, \eta)\|_{L^p(\Omega; \mathbb{R}^d)}^p + \|g(t, \eta)\|_{L^p(\Omega; \mathbb{R}^{d \times n})}^p$$

$$+ \|h(t, \eta)\|_{L^p(\Omega, L^p(\nu))}^p + \|h(t, \eta)\|_{L^p(\Omega, L^2(\nu))}^p \qquad (2.18)$$

$$\leqslant K(1 + \|\eta\|_{S^p(\Omega; \mathcal{D})}^p).$$

Remark 2.11. *As usual, $\mathbf{D_2}$ is implied by $\mathbf{D_1}$, if we assume that whenever $\eta = 0$, the left-hand-side of inequality (2.18) is bounded by some K', uniformly in $t \in [0, T]$.*

Theorem 2.12 (**Existence and Uniqueness I**). *Consider equation (2.13) with $\mathcal{P}$ satisfied.*

(i) *Suppose that assumption $\mathbf{D_1}$ holds. If $X, Y \in S_{ad}^p(\Omega; \mathcal{D}_T)$ are strong solutions to (2.1), then $X = Y$ in $S_{ad}^p(\Omega; \mathcal{D}_T)$.*

(ii) *Suppose that assumptions $\mathbf{D_1}$ and $\mathbf{D_2}$ hold. Then there exists a strong solution $X \in S_{ad}^p(\Omega; \mathcal{D}_T)$ to the equation (2.1). Moreover, there exists $D = D(K, p, T, d, m, n) > 0$, such that*

$$\|X\|_{S_{ad}^p(\Omega;\mathcal{D}_T)}^p \leqslant e^{Dt}(Dt + \|\eta\|_{S^p(\Omega;\mathcal{D})}^p) \tag{2.19}$$

whenever $t \leqslant T$.

Proof. We will use a standard Picard iteration argument to show that a solution exists. First, we define, for each $k \geqslant 0$, a sequence of processes in $S_{ad}^p(\Omega; \mathcal{D}_T)$ inductively by

$$X^1(t) = \eta(0), \qquad\qquad\qquad\qquad t \in [0, T],$$
$$X_0^1 = \eta$$
$$X^{k+1}(t) = \eta(0) + \int_0^t f(s, X_s^k)ds + \int_0^t g(s, X_s^k)dW(s)$$
$$\qquad\qquad + \int_0^t h(s, X_s^k)(z)\tilde{N}(ds, dz), \qquad\qquad t \in [0, T]$$
$$X_0^{k+1} = \eta.$$

We immediately have that $X^1 \in S_{ad}^p(\Omega; \mathcal{D}_T)$. Also if we assume that $X^k \in S_{ad}^p(\Omega; \mathcal{D}_T)$, then by assumption $f(X^k)$, $g(X^k)$, and $h(X^k)$ admit progressive and predictable versions respectively. Thus by assumption ($\mathbf{D_2}$) it follows that

$$\int_0^T \Big(\|f(t, X_t^k)\|_{L^p(\Omega;\mathbb{R}^d)}^p + \|g(t, X_t^k)\|_{L^p(\Omega;\mathbb{R}^{d \times n})}^p$$
$$\qquad + \|h(t, X_t^k)\|_{L^p(\Omega, L^p(\nu))}^p + \|h(t, X_t^k)\|_{L^p(\Omega, L^2(\nu))}^p \Big)dt \tag{2.20}$$
$$\leqslant \int_0^T K(1 + \|X_t^k\|_{S^p(\Omega;\mathcal{D})}^p)dt \leqslant KT(1 + \|X^k\|_{S_{ad}^p(\Omega;\mathcal{D}_T)}^p) < \infty.$$

In particular, the integrands of X^{k+1} are Itô integrable, so that X^{k+1} is càdlàg and adapted, and finally by Kunita's inequality, we have that $X^{k+1} \in S_{ad}^p(\Omega; \mathcal{D}_T)$.

We now claim that for each $k \in \mathbb{N}$ the following estimate holds for every $t \in [0, T]$,

$$\|X^{k+1} - X^k\|_{S_{ad}^p(\Omega;\mathcal{D}_t)}^p \leqslant \frac{(LCt)^{k-1}}{(k-1)!}\|X^2 - X^1\|_{S_{ad}^p(\Omega;\mathcal{D}_T)}^p. \tag{2.21}$$

This trivially holds when $k = 1$. Now suppose that (2.21) holds for each $t \in [0, T]$. Using the definition of X^{k+2}, X^{k+1}, Kunita's inequality (2.17), and assumption ($\mathbf{D_2}$), we find that

$$\|X^{k+2} - X^{k+1}\|_{S_{ad}^p(\Omega;\mathcal{D}_t)}^p \leqslant C\int_0^t \Big(\|f(s, X_s^{k+1}) - f(s, X_s^k)\|_{L^p(\Omega;\mathbb{R}^d)}^p + \|g(s, X_s^{k+1}) - g(s, X_s^k)\|_{L^p(\Omega;\mathbb{R}^{d \times n})}^p$$
$$\qquad + \|h(s, X_s^{k+1}) - h(s, X_s^k)\|_{L^p(\Omega, L^p(\nu))}^p + \|h(s, X_s^{k+1}) - h(s, X_s^k)\|_{L^p(\Omega, L^2(\nu))}^p \Big)ds$$
$$\qquad \leqslant LC\int_0^t \|X_s^{k+1} - X_s^k\|_{S^p(\Omega;\mathcal{D})}^p ds \leqslant LC\int_0^t \|X^{k+1} - X^k\|_{S^p(\Omega;\mathcal{D}_s)}^p ds$$
$$\qquad \leqslant LC\int_0^t \frac{(LCs)^{k-1}}{(k-1)!}\|X^2 - X^1\|_{S_{ad}^p(\Omega;\mathcal{D}_T)}^p ds = \frac{(LCt)^k}{k!}\|X^2 - X^1\|_{S_{ad}^p(\Omega;\mathcal{D}_T)}^p.$$

12

Now, by induction, (2.21) holds for each $k \in \mathbb{N}$. In particular

$$\|X^k - X^i\|_{S_{ad}^p(\Omega;\mathcal{D}_t)}^p \leqslant \|X^2 - X^1\|_{S_{ad}^p(\Omega;\mathcal{D}_T)}^p \sum_{j=\min\{k,i\}}^{\infty} \frac{(LTC)^{j-1}}{(j-1)!} \to 0, \quad \text{as } k, i \to \infty,$$

so that $\{X^k\}_{k \geqslant 0}$ is a Cauchy sequence in $S_{ad}^p(\Omega;\mathcal{D}_T)$. Since $S_{ad}^p(\Omega;\mathcal{D}_T)$ is complete, we have that $\{X^k\}_{k \geqslant 0}$ converges to some X in $S_{ad}^p(\Omega;\mathcal{D}_T)$. Clearly $X_0 = \eta$ P-a.s.

We will now show that the limit X satisfies (2.16) by showing that

$$d := \mathbb{E}\Big[\sup_{0 \leqslant t \leqslant T}\Big|X(t) - \Big\{\eta(0) + \int_0^t f(s, X_s)ds + \int_0^t g(s, X_s)dW(s)$$
$$+ \int_0^t \int_{\mathbb{R}_0} h(s, X_s)(z)\tilde{N}(ds, dz)\Big\}\Big|^p\Big]^{1/p} = 0 \tag{2.22}$$

For arbitrary k, we subtract X^{k+1} and add its integral representation inside the supremum in (2.22). Then by the triangle inequality, Kunita's inequality, and finally the Lipschitz condition $(\mathbf{D_1})$ we find that

$$d \leqslant \|X - X^{k+1}\|_{S_{ad}^p(\Omega;\mathcal{D}_T)} + \Big\{C\int_0^T \Big(\|f(t, X_t^k) - f(t, X_t)\|_{L^p(\Omega;\mathbb{R}^d)}^p$$
$$+ \|g(t, X_t) - g(t, X_t^k)\|_{L^p(\Omega;\mathbb{R}^{d\times n})}^p + \|h(t, X_t) - h(t, X_t^k)\|_{L^p(\Omega, L^p(\nu))}^p$$
$$+ \|h(t, X_t) - h(t, X_t^k)\|_{L^p(\Omega, L^2(\nu))}^p\Big)dt\Big\}^{1/p}$$
$$\leqslant \|X - X^{k+1}\|_{S_{ad}^p(\Omega;\mathcal{D}_T)} + \Big\{CL\int_0^T \|X_t - X_t^k\|_{S^p(\Omega;\mathcal{D})}^p dt\Big\}^{1/p}$$
$$\leqslant \|X - X^{k+1}\|_{S_{ad}^p(\Omega;\mathcal{D}_T)} + (CLT)^{1/p}\|X - X^k\|_{S_{ad}^p(\Omega;\mathcal{D}_T)} \to 0.$$

Since for any $\epsilon > 0$ we have that $0 \leqslant d < \epsilon$, it follows that $d = 0$, and hence a solution exists.

Suppose now that X and Y are solutions of (2.13). We will show that $X = Y$. Exploiting the integral representation of X and Y, Kunita's inequality and the Lipschitz condition $(\mathbf{D_1})$, we have that, for all $t \in [0, T]$,

$$\|X - Y\|_{S_{ad}^p(\Omega;\mathcal{D}_t)}^p \leqslant C\int_0^t \Big(\|f(s, X_s) - f(s, Y_s)\|_{L^p(\Omega;\mathbb{R}^d)}^p + \|g(s, X_s) - g(s, Y_s)\|_{L^p(\Omega;\mathbb{R}^{d\times n})}^p$$
$$+ \|h(s, X_s) - h(s, Y_s)\|_{L^p(\Omega, L^p(\nu))}^p + \|h(s, X_s) - h(s, Y_s)\|_{L^p(\Omega, L^2(\nu))}^p\Big)ds$$
$$\leqslant CL\int_0^t \|X_s - Y_s\|_{S^p(\Omega;\mathcal{D})}^p ds \leqslant CL\int_0^t \|X - Y\|_{S^p(\Omega;\mathcal{D}_s)}^p ds.$$

and thus we have $\|X - Y\|_{S_{ad}^p(\Omega;\mathcal{D}_t)}^p = 0$ for every $t \in [0, T]$ from Grönwall's inequality.

Similarly, if X is a solution to (2.13), from the integral representations, Kunita's inequality and the linear growth condition $(\mathbf{D_2})$ we have that

$$\|X\|_{S_{ad}^p(\Omega;\mathcal{D}_t)}^p \leqslant C\Big\{\|\eta\|_{S^p(\Omega;\mathcal{D})}^p + \int_0^t \Big(\|f(s, X_s)\|_{L^p(\Omega;\mathbb{R}^d)}^p$$
$$+ \|g(s, X_s)\|_{L^p(\Omega;\mathbb{R}^{d\times n})}^p + \|h(s, X_s)\|_{L^p(\Omega, L^p(\nu))}^p + \|h(s, X_s)\|_{L^p(\Omega, L^2(\nu))}^p\Big)ds\Big\}$$
$$\leqslant C\Big(\|\eta\|_{S^p(\Omega;\mathcal{D})}^p + K\big(\int_0^t 1 + \|X_s\|_{S^p(\Omega;\mathcal{D})}^p ds\big)\Big) \leqslant C\|\eta\|_{S^p(\Omega;\mathcal{D})}^p + CKt + CK\int_0^t \|X\|_{S^p(\Omega;\mathcal{D}_s)}^p ds,$$

so applying Grönwall's inequality we obtain

$$\|X\|_{S_{ad}^p(\Omega;\mathcal{D}_t)}^p \leqslant \Big(C\|\eta\|_{S^p(\Omega;\mathcal{D})}^p + CKt\Big)e^{CKt},$$

$\square$

for all $t \in [0, T]$.

Remark 2.13 (**Path dependent SDEs**). *Suppose that for each $t \in [0,T]$ and every $\eta \in S_{ad}^p(\Omega, \mathcal{F}_0; \mathcal{D})$ it holds that,*

$$f(t, \eta)(\omega) = F(t, \eta(\omega))$$
$$g(t, \eta)(\omega) = G(t, \eta(\omega))$$
$$h(t, \eta, \omega, \zeta) = H(t, \eta(\omega), \zeta),$$

P-a.s for some deterministic functionals

$$F : [0,T] \times \mathcal{D} \to \mathbb{R}^d,$$
$$G : [0,T] \times \mathcal{D} \to \mathbb{R}^{d \times m},$$
$$H : [0,T] \times \mathcal{D} \to L^2(\nu) \cap L^p(\nu).$$

then the assumptions (**D₁**) *and* (**D₂**) *hold whenever F, G are Lipschitz continuous in the second variable, uniformly with respect to the first, and H is Lipschitz continuous in the second variable, uniformly with respect to the first, using both norms $\|\cdot\|_{L^2(\nu)}$ and $\|\cdot\|_{L^p(\nu)}$.*

2.4 M^p framework

Now, consider equation

$$dX(t) = f(t, X_t, X(t))dt + g(t, X_t, X(t))dW(t) + \int_{\mathbb{R}_0} h(t, X_t, X(t))(z)\tilde{N}(dt, dz) \tag{2.23}$$

$$(X_0, X(0)) = (\eta, x).$$

Here (2.23) we have used the notation $f(\cdot, X_t, X(t))$ to emphasize the structure of the product space of M^p. Now for each $t \in [0,T]$ we will require that $(X_t, X(t))$ belongs to the space $L^p(\Omega, \mathcal{F}_t; M^p)$ for some $p \in [2, \infty)$, that will be fixed throughout the section. Therefore, we introduce

$$\mathbf{L}_p^{\mathbb{F}} := \{(t, (\psi, v)) \in [0,T] \times L^p(\Omega, \mathcal{F}; M^p) \text{ such that } (\psi, v) \in L^p(\Omega, \mathcal{F}_t; M^p)\}, \tag{2.24}$$

In particular, we will require that:

$$f : \mathbf{L}_p^{\mathbb{F}} \to L^p(\Omega, \mathbb{R}^d))$$
$$g : \mathbf{L}_p^{\mathbb{F}} \to L^p(\Omega, \mathbb{R}^{d \times m}),$$
$$h : \mathbf{L}_p^{\mathbb{F}} \to L^p(\Omega, L^2(\nu, \mathbb{R}^{d \times n})).$$

Moreover,

$$(\eta, x) \in L^p(\Omega, \mathcal{F}_0; M^p).$$

To ensure that the integrals are well-defined, the following assumptions are imposed on the coefficient functionals f, g and h.

Assumption (**Q**). *For $Y \in L_{ad}^p(\Omega; L_T^p)$, the process*

$$[0,T] \times \Omega \times \mathbb{R}_0 \ni (t, \omega, z) \mapsto h(t, Y_t, Y(t))(\omega)(z) \in \mathbb{R}^{d \times n}$$

has a predictable version, and

$$[0,T] \times \Omega, \ni (t, \omega) \mapsto f(t, Y_t, Y(t))(\omega) \in \mathbb{R}^d,$$
$$[0,T] \times \Omega, \ni (t, \omega) \mapsto g(t, Y_t, Y(t))(\omega) \in \mathbb{R}^{d \times m}$$

have progressively measurable versions.

Definition 2.14. *We say that* $X \in L^p_{ad}(\Omega; L^p_T)$ *is a* strong solution *to* (2.23) *if for each* $t \in [0, T]$

$$X(t) = x + \int_0^t f(s, X_s, X(s))ds + \int_0^t g(s, X_s, X(s))dW(s) + \int_0^t \int_{\mathbb{R}_0} h(s, X_s, X(s))(z)\tilde{N}(ds, dz)$$

$(X_0, X(0)) = (\eta, x),$

(2.25)

in $L^p_{ad}(\Omega; L^p_T)$ *(in particular P-a.s.). If the solution is unique, we will sometimes write* $^{\eta,x}X$ *to denote the solution of* (2.25) *with initial data* $(X_0, X(0)) = (\eta, x)$.

Proposition 2.15. *Let* $Y : [0, T] \times \Omega \to \mathbb{R}^d$ *be a stochastic process with a.s. càdlàg sample paths. Then the associated* M^p*-valued* segment process

$$[0, T] \times \Omega \ni (t, \omega) \mapsto (Y_t(\omega), Y(t, \omega)) \in M^p$$

(2.26)

is a.s. càdlàg.

Observe that the property that the segment process is càdlàg whenever Y is càdlàg, depends on the topology of the infinite dimensional space M^p. In general, such property does not hold if we replace M^p with $\mathcal{D}$.

Proof of Proposition 2.15. It suffices to show that if $Y(\omega) : [-r, T] \to \mathbb{R}^d$ is a càdlàg path, then the function

$$[0, T] \ni t \mapsto (Y_t(\omega), Y(t, \omega)) \in M^p$$

(2.27)

is also càdlàg. The function (2.27) is right continuous. In fact, for every sequence $r_k, k \in \mathbb{N}$ with $r_k > 0$ and $r_k \to 0$ as $k \to \infty$, we have that

$$\lim_{k \to \infty} \|Y_{t+r_k}(\omega) - Y_t(\omega), Y(t + r_k, \omega) - Y(t, \omega))\|_{M^p}^p$$
$$= \lim_{k \to \infty} \int_{-r}^0 |Y(t + r_k + \beta, \omega) - Y(t + \beta, \omega)|^p d\beta + \lim_{k \to \infty} |Y(t + r_k, \omega) - Y(t, \omega)|^p = 0.$$

by the dominated convergence theorem. Now given $t \in [0, T]$, we define $(Y_t^-(\omega), Y^-(t, \omega)) \in M^p$ by

$$Y_t^-(\theta, \omega) = \begin{cases} Y_t(\theta, \omega), & \theta \in [-r, 0) \\ \lim_{u \to 0^-} Y_t(u, \omega), & \theta = 0. \end{cases}$$

Consider r_k as above, we can use the dominated convergence theorem to observe that

$$\lim_{k \to \infty} \|(Y_{t-r_k}(\omega) - Y_t^-(\omega), Y(t - r_k, \omega) - Y^-(t, \omega))\|_{M^p}^p$$
$$= \lim_{k \to \infty} \int_{-r}^0 |Y(t - r_k + \beta, \omega) - Y^-(t + \beta, \omega)|^p d\beta + \lim_{k \to \infty} |Y(t - r_k, \omega) - Y^-(t, \omega)|^p = 0,$$

and hence the function (2.27) has left limits. $\qquad\square$

2.4.1 Existence and uniqueness

The $L^p(\Omega; M^p)$-analogue of the hypotheses $(\mathbf{D_1})$ and $(\mathbf{D_2})$, are defined below.

Assumption. $(\mathbf{L_1})$ *There exists* $L > 0$, *such that whenever* $t \in [0, T]$ *and* $(\eta_1, x_1), (\eta_2, x_2) \in L^p(\Omega, \mathcal{F}_t; M^p)$, *then*

$$\|f(t, \eta_1, x_1) - f(t, \eta_2, x_2)\|_{L^p(\Omega; \mathbb{R}^d)}^p + \|g(t, \eta_1, x_1) - g(t, \eta_2, x_2)\|_{L^p(\Omega; \mathbb{R}^{d \times n})}^p$$
$$+ \|h(t, \eta_1, x_1) - h(t, \eta_2, x_2)\|_{L^p(\Omega, L^p(\nu))}^p + \|h(t, \eta_1, x_1) - h(t, \eta_2, x_2)\|_{L^p(\Omega; L^2(\nu))}^p$$
$$\leqslant L\|(\eta_1, x_1) - (\eta_2, x_2)\|_{L^p(\Omega; M^p)}^p.$$

15

($\mathbf{L_2}$) *There exists $K > 0$, such that whenever $t \in [0, T]$ and $(\eta, v) \in L^p(\Omega, \mathcal{F}_t; M^p)$, then*

$$\|f(t, \eta, x)\|^p_{L^p(\Omega; \mathbb{R}^d)} + \|g(t, \eta, x)\|^p_{L^p(\Omega; \mathbb{R}^{d \times n})}$$
$$+ \|h(t, \eta, x)\|^p_{L^p(\Omega, L^p(\nu))} + \|h(t, \eta, x)\|^p_{L^p(\Omega, L^2(\nu))}$$
$$\leqslant K(1 + \|(\eta, x)\|^p_{L^p(\Omega; M^p)}).$$

Theorem 2.16 (**Existence and Uniqueness II**). *Consider* (2.23) *with $\mathcal{Q}$ satisfied.*

(i) *Let $(\eta, x) \in L^p(\Omega, \mathcal{F}_0; M^p)$ such that η is càdlàg P-a.s. and $X \in L^p_{ad}(\Omega; L^p_T)$ be a strong solution to equation* (2.23). *Then the segment process*

$$\Omega \times [0, T] \ni (t, \omega) \mapsto (X_t(\omega), X(t, \omega)) \in M^p \tag{2.28}$$

has a càdlàg modification.

(ii) *Suppose that assumption ($\mathbf{L_1}$) holds. If $X, Y \in L^p_{ad}(\Omega; L^p_T)$ are strong solutions to* (2.23), *then $X = Y$ in $L^p_{ad}(\Omega; L^p_T)$.*

(iii) *Suppose that assumptions ($\mathbf{L_1}$) and ($\mathbf{L_2}$) hold. Then there exists a strong solution X to equation* (2.23). *Moreover, there exists $D = D(K, p, T, d, m, n) > 0$, such that*

$$\|X\|^p_{L^p_{ad}(\Omega; L^p_t)} \leqslant e^{Dt}(Dt + \|(\eta, x)\|^p_{L^p(\Omega; M^p)}), \tag{2.29}$$

whenever $t \leqslant T$.

Proof. **(i)** Recall that since X is a strong solution of (2.23), it is a semimartingale on $[0, T]$ and hence it admits a modification which is càdlàg on $[0, T]$. Since $X_0 = \eta$ is càdlàg, X is càdlàg, on $[-r, T]$. By Proposition 2.15 (i) holds.

(ii,iii) The proof is based on the same argument as for the proof of Theorem 2.12. For the sake of brevity we do not write out the details. However, we remark that if one replaces the norms $\| \cdot \|_{S^p(\Omega; \mathcal{D})}$ and $\| \cdot \|_{S^p_{ad}(\Omega; \mathcal{D}_t)}$, with the norms $\| \cdot \|_{L^p(\Omega; M^p)}$ and $\| \cdot \|_{L^p_{ad}(\Omega; L^p_t)})$ respectively, then all the inequalities hold true, except for the choice of constants. As an example, we provide the following M^p analogue of (2.20), namely

$$\int_0^T \Big(\|f(t, X^k_t, X^k(t))\|^p_{L^p(\Omega; \mathbb{R}^d)} + \|g(t, X^k_t, X^k(t))\|^p_{L^p(\Omega; \mathbb{R}^{d \times n})}$$
$$+ \|h(t, X^k_t, X^k(t))\|^p_{L^p(\Omega, L^p(\nu))} + \|h(t, X^k_t, X^k(t))\|^p_{L^p(\Omega, L^2(\nu))} \Big) dt \tag{2.30}$$
$$\leqslant \int_0^T K(1 + \|(X^k_t, X^k(t))\|^p_{L^p(\Omega; M^p)}) dt \leqslant KT + K(T+1)\|X^k\|^p_{L^p_{ad}(\Omega; L^p_T)} < \infty.$$

This follows immediately by the assumption $\mathbf{L_2}$ and inequality (2.9). $\qquad \square$

Let us stress that when the initial value is càdlàg, then the setting of Section (2.3) is more general than the one in this section. In fact, the assumptions ($\mathbf{L_1}$) and ($\mathbf{L_2}$) imply assumptions ($\mathbf{D_1}$) and ($\mathbf{D_2}$), respectively.

2.5 Robustness SFDDEs

In the present section we study robustness of SFDDE to changes of the noise. In particular, we want to approximate the solution of an SFDDE X, with an approximate processes X^ϵ, where X^ϵ are defined by substituting the integrals with respect to the small jumps with integrals with respect to scaled Brownian motions. We follow rather closely, the presentation in [6] for ordinary SDE's and remark that a related problem is also considered in [31]. In this paper we also include a

new ingredient, by giving sufficient conditions which ensure that the approximations X^ϵ converge to X in the p'th mean.

The main motivation for studying such robustness problem is that it is difficult to perform simulations of distributions corresponding to a Lévy process. Indeed, simulation of such distributions are often performed by neglecting the jumps below a certain size ϵ. However, when needed to preserve the variation of the infinite activity Lévy process, a scaled Brownian motion is typically replacing the small jumps. Under some additional assumptions, it is known that given a square integrable (1-dimensional) Lévy process with Lévy measure μ and compensated Poisson random measure $\widetilde{M}$, the expression

$$\int_{|z|<\epsilon} z^2 \mu(dz)^{-1/2} \int_0^t \int_{|z|<\epsilon} z\tilde{M}(dz, ds),$$

converges in distribution to a standard Brownian motion W, as ϵ tends to 0. We refer to [4, 11] for more details on this topic. We remark that the robustness problem in this paper, does not rely on the above mentioned additional assumptions.

2.5.1 The model

Fix $p \in [2, \infty)$. We want to consider the following dynamical systems with memory and jumps in the setting of Section 2.3:

$$X(t) = \eta(0) + \int_0^t f(s, X_s)ds + \int_0^t g(s, X_s)dW(s) + \int_0^t \int_{\mathbb{R}_0} h_0(s, X_s)\lambda(z)\tilde{N}(ds, dz), \quad t \in [0, T]$$
$$X_0 = \eta \in S^p(\Omega, \mathcal{D}). \tag{2.31}$$

Here,

$$h_0 : \mathbf{S}_p^{\mathbb{F}} \to L^p(\Omega, \mathbb{R}^{d \times k}),$$

and

$$\lambda \in L^2(\nu, \mathbb{R}^{k \times n}) \cap L^p(\nu, \mathbb{R}^{k \times n})$$

for some $k \in \mathbb{N}$. Observe that

$$h := h_0 \lambda : \mathbf{S}_p^{\mathbb{F}} \to L^p(\Omega, L^2(\nu, \mathbb{R}^{d \times n})).$$

Example 2.17. *Suppose that $n = k = d$ and that $h_0(t, \eta)$ and λ are diagonal matrices. In particular that,*

$$h_0(t, \eta)\lambda(z)$$

is a diagonal matrix with entries $h_0^{i,i}(t, \eta)\lambda_{i,i}(z)$ for $i = 1, \ldots, n$. Then the component-wise form of the jump integral in the SFDDE (2.31) is given by

$$\int_0^t \int_{\mathbb{R}_0} h_0^{i,i}(s, X_s)\lambda_{i,i}(z)\tilde{N}_i(ds, dz).$$

If we let the delay parameter r be equal to 0, then this example reduces to the problem of robustness to model choice treated in [30].

Now, let us impose the following assumptions on f, g, h_0 and λ:

Assumption. (i) *The coefficient functionals f and g are assumed to satisfy the assumptions $(\mathcal{P})$, $(\mathbf{D_1})$, $(\mathbf{D_2})$ of Section 2.3.*

(ii) *Whenever Y is a càdlàg adapted process on $[-r, T]$, then $h_0(t, Y_t)$ is predictable. Moreover, the functional h_0 satisfies the Lipschitz and linear growth conditions:*

$$\|h_0(t, \eta_1) - h_0(t, \eta_2)\|^p_{L^p(\Omega, \mathbb{R}^{d \times k})} \leqslant L\|\eta_1 - \eta_2\|^p_{S^p(\Omega; \mathcal{D})},$$

$$\|h_0(t, \eta)\|^p_{L^p(\Omega, \mathbb{R}^{d \times k})} \leqslant K(1 + \|\eta\|^p_{S^p(\Omega; \mathcal{D})}).$$

We claim now that the map $h := h_0 \lambda$ satisfies the assumptions $(\mathcal{P})$, $(\mathbf{D_1})$, $(\mathbf{D_2})$ from Section 2.3. In fact observe that

$$\|h(t, \eta_1) - h(t, \eta_2)\|^p_{L^p(\Omega, L^p(\nu))} + \|h(t, \eta_1) - h(t, \eta_2)\|^p_{L^p(\Omega, L^2(\nu))}$$

$$= \sum_{j=1}^n \|(h_0(t, \eta_1) - h_0(t, \eta_2))\lambda^{\cdot j}\|^p_{L^p(\Omega, L^p(\nu_j))} + \|(h_0(t, \eta_1) - h_0(t, \eta_2))\lambda^{\cdot j}\|^p_{L^p(\Omega, L^2(\nu_j))}$$

$$\leqslant \|h_0(t, \eta_1) - h_0(t, \eta_2)\|^p_{L^p(\Omega; \mathbb{R}^{d \times k})} \sum_{j=1}^n \left(\|\lambda^{\cdot j}\|^p_{L^p(\nu_j)} + \|\lambda^{\cdot j}\|^p_{L^2(\nu_j)} \right)$$

$$\leqslant L\|\eta_1 - \eta_2\|^p_{S^p(\Omega; \mathcal{D})} \left(\|\lambda\|^p_{L^p(\nu)} + \|\lambda\|^p_{L^2(\nu)} \right).$$

Thus h satisfies the Lipschitz assumption $(\mathbf{D}_1)$. A similar argument yields h satisfies the linear growth assumption $(\mathbf{D}_2)$. Thus, by the existence and uniqueness Theorem 2.12, the following result holds.

Corollary 2.18. *The equation* (2.31) *has a unique solution $^\eta X$ in $S^p_{ad}(\Omega; \mathcal{D}_T)$. Moreover, there exists $D = D(K, \lambda, p, T, d, m, n) > 0$, such that*

$$E\left[\sup_{-r \leqslant s \leqslant t} |^\eta X(s)|^p \right] \leqslant e^{Dt}(Dt + \|\eta\|^p_{S^p(\Omega; \mathcal{D})}), \tag{2.32}$$

for any $t \leqslant T$.

2.5.2 The approximating model

Let us first introduce some notation. For any $\varepsilon \in (0, 1)$, define $\lambda_\varepsilon(z) \in \mathbb{R}^{k \times n}$ by

$$\lambda_\varepsilon(z) = 1_{\{|z| < \varepsilon\}}(z)\lambda(z),$$

for a.e. z. Now, let B be an n-dimensional $\mathbb{F}$-adapted Brownian motion, independent of $\tilde{N}$. Independence of B and W is not required, (see e.g. [6]). We want to approximate equation (2.31) by replacing the integral with respect to the small jumps with an integral with respect to the Brownian motion B. More specifically, we will replace the integrators

$$\int_{\mathbb{R}_0} \lambda^{i,j}_\varepsilon(z) \tilde{N}^j(dt, dz), \tag{2.33}$$

with the integrators

$$\Lambda^{i,j}(\varepsilon) dB^j(t), \tag{2.34}$$

for $i = 1, \ldots, k; j = 1, \ldots, n$. Here, $\Lambda(\epsilon)$ can be any bounded deterministic function with values in $\mathbb{R}^{k \times n}$ converging to 0 as $\varepsilon \to 0$. We choose to let

$$\Lambda^{i,j}(\varepsilon) = \|\lambda^{i,j}_\varepsilon\|_{L^2(\nu_j)}.$$

This choice corresponds to what has previously been used in the literature, see e.g. [6]. A justification of this choice is considered in Remark 2.19 below. Notice now that

$$|\Lambda(\varepsilon)|_2 = \|\lambda_\varepsilon\|_{L^2(\nu_j)}.$$

Remark 2.19. *The choice $\Lambda^{i,j}(\epsilon) = \|\lambda_\varepsilon^{i,j}\|_{L^2(\nu_j)}$ above is reasonable in the sense that for a given predictable square integrable process Y, this change of integrator preserves the variance of the integrals, i.e.*

$$E\Big[\Big(\int_0^t \int_{\mathbb{R}_0} Y(s)\lambda_\varepsilon^{i,j}(z)\tilde{N}^j(ds,dz)\Big)^2\Big] = E\Big[\int_0^t Y(s)^2 ds\Big]\|\lambda_\varepsilon^{i,j}\|_{L^2(\nu_j)}^2 = E\Big[\Big(\int_0^t Y(s)\Lambda^{i,j}(\epsilon)dB^j(s)\Big)^2\Big],$$

for $i = 1, \ldots, k; j = 1, \ldots, n$. From a financial terminology perspective where these models can be applied (see e.g. [3, 6, 8, 9, 28, 29, 33]), this choice of Λ, preserves the total volatility of a process, when (2.33) is replaced by (2.34). However, this particular choice of Λ is not necessary for the analysis, as we will see in Remark 2.22 below.

Now, we are ready to exploit the dynamics of the *approximated processes* X^ϵ. Consider

$$\begin{aligned}
X^{(\epsilon)}(t) &= \eta(0) + \int_0^t f(s, X_s^\epsilon)ds + \int_0^t g(s, X_s^\varepsilon)dW(s) \\
&\quad + \int_0^t h_0(s, X_s^\epsilon)\Lambda(\epsilon)dB(s) + \int_0^t \int_{\mathbb{R}_0} h_0(s, X_s^\epsilon)(\lambda(z) - \lambda_\varepsilon(z))\tilde{N}(ds,dz) \\
X_0^{(\epsilon)} &= \eta
\end{aligned} \tag{2.35}$$

Before proceeding to the main result of this section, we make the following observations regarding the functionals in the approximated equation (2.35):

- The functionals

$$\eta \overset{g_1}{\longmapsto} h_0(t,\eta)\Lambda(\epsilon),$$
$$\eta \overset{h_1}{\longmapsto} h_0(t,\eta)(\lambda - \lambda_\varepsilon),$$

  satisfy the corresponding hypotheses from Section 2.3

- The Lipschitz and linear growth constant appearing in assumptions ($\mathbf{D_1}$) and ($\mathbf{D_2}$) can be chosen independent of ϵ. In particular, we can deduce the following linear growth estimate:

$$\begin{aligned}
&\|h_0(t,\eta)\Lambda(\epsilon)\|_{L^p(\Omega;\mathbb{R}^{d\times n})}^p + \|h_0(t,\eta)(\lambda - \lambda_\varepsilon)\|_{L^p(\Omega,L^p(\nu))}^p + \|h_0(t,\eta)(\lambda - \lambda_\varepsilon)\|_{L^p(\Omega,L^2(\nu))}^p \\
&\leqslant \|h_0(t,\eta)\|_{L(\Omega;\mathbb{R}^{d\times k})}^p \sup_{\varepsilon\in(0,1)}|\Lambda(\epsilon)|^p + \|h_0(t,\eta)\|_{L^p(\Omega;\mathbb{R}^{d\times k})}^p\|\lambda\|_{L^p(\nu)}^p + \|h_0(t,\eta)\|_{L^p(\Omega;\mathbb{R}^{d\times k})}^p\|\lambda\|_{L^2(\nu)}^p \\
&\leqslant K'(1 + \|\eta\|_{S^p(\Omega;\mathcal{D})}^p).
\end{aligned}$$

  A similar estimate holds for the Lipschitz condition ($\mathbf{D_1}$).

The following existence and uniqueness result immediately follows from Theorem 2.12.

Corollary 2.20. *For each $\epsilon > 0$, there exists a unique strong solution ${}^\eta X^\epsilon$ to the equation (2.35). Moreover, there exists a $D = D(K, \lambda, p, T, d, m, n) > 0$, independent of ϵ, such that*

$$E\left[\sup_{-r\leqslant s\leqslant t}|{}^\eta X^\epsilon(s)|^p\right] \leqslant e^{Dt}(Dt + \|\eta\|_{S^p(\Omega;\mathcal{D})}^p) \tag{2.36}$$

for any $t \leqslant T$.

Now, we are ready to state the main result of the present section. This result guarantees that, when ε tends to 0, X_ε converges to X in $S_{ad}^p(\Omega;\mathcal{D}_T)$.

Theorem 2.21 (Robustness)**.** *Suppose that X satisfies equation (2.31) and X^ϵ satisfies equation (2.35). Then there exist a constant $A := A(p, T, \eta, K, L, \lambda) > 0$, independent of ϵ, such that*

$$E\left[\sup_{-r\leqslant s\leqslant t}|{}^\eta X(s) - {}^\eta X^\epsilon(s)|^p\right] \leqslant Ae^{At}(\|\lambda_\varepsilon\|_{L^2(\nu)}^p + \|\lambda_\varepsilon\|_{L^p(\nu)}^p). \tag{2.37}$$

19

Proof. Writing out the integral representation of $X(s)$ and $X^\epsilon(s)$, we have that

$$X(s) - X^\epsilon(s) = \int_0^s f(u, X_u) - f(u, X_u^\epsilon)du + \int_0^s g(u, X_u) - g(u, X_u^\epsilon)dW(u)$$
$$+ \int_0^s \int_{\mathbb{R}_0^d} (h_0(u, X_u) - h_0(u, X_u^\epsilon))\lambda(z) + h_0(u, X_u^\epsilon)\lambda_\varepsilon(z)\tilde{N}(du, dz)$$
$$- \int_0^s h_0(u, X_u^\epsilon)\Lambda_p(\epsilon)dB(u),$$
$$X_0 - X_0^\epsilon = 0.$$

Let us first consider some estimates for the integrands of $\tilde{N}$ and B. Observe that

$$\|(h_0(u, X_u) - h_0(u, X_u^\epsilon))\lambda + h_0(u, X_u^\epsilon)\lambda_\varepsilon\|_{L^p(\Omega;L^p(\nu))}$$
$$\leqslant \|(h_0(u, X_u) - h_0(u, X_u^\varepsilon))\|_{L^p(\Omega;\mathbb{R}^{d\times k})}\|\lambda\|_{L^p(\nu)} + \|h_0(u, X_u^\epsilon)\|_{L^p(\Omega;\mathbb{R}^{d\times k})}\|\lambda_\varepsilon\|_{L^p(\nu)}$$
$$\leqslant L^{1/p}\|X_u - X_u^\epsilon\|_{S^p(\Omega;\mathcal{D})}\|\lambda\|_{L^p(\nu)} + K^{1/p}(1 + \|X_u^\epsilon\|_{S^p(\Omega;\mathcal{D})})^{1/p}\|\lambda_\varepsilon\|_{L^p(\nu)},$$

and hence

$$\|(h_0(u, X_u) - h_0(u, X_u^\epsilon))\lambda + h_0(u, X_u^\epsilon)\lambda_\varepsilon\|_{L^p(\Omega, L^p(\nu))}^p$$
$$\leqslant 2^{p-1}\Big(L\|X_u - X_u^\epsilon\|_{S^p(\Omega;\mathcal{D})}^p\|\lambda\|_{L^p(\nu)}^p + K(1 + \|X_u^\epsilon\|_{S^p(\Omega;\mathcal{D})}^p)\|_{L^p(\Omega)}\|\lambda_\varepsilon\|_{L^p(\nu)}^p\Big). \tag{2.38}$$

In an analogous manner we have that

$$\|(h_0(u, X_u) - h_0(u, X_u^\epsilon))\lambda + h_0(u, X_u^\epsilon)\lambda_\varepsilon\|_{L^p(\Omega, L^2(\nu))}^p$$
$$\leqslant 2^{p-1}\Big(L\|X_u - X_u^\epsilon\|_{S^p(\Omega;\mathcal{D})}^p\|\lambda\|_{L^2(\nu)}^p + K(1 + \|X_u^\epsilon\|_{S^p(\Omega;\mathcal{D})}^p)\|\lambda_\varepsilon\|_{L^2(\nu)}^p\Big), \tag{2.39}$$

and that

$$\|h_0(u, X_u^\epsilon)\Lambda(\varepsilon)\|_{L^p(\Omega;\mathbb{R}^{d\times n})}^p \leqslant K(1 + \|X_u^\epsilon\|_{S^p(\Omega;\mathcal{D})}^p)|\Lambda(\varepsilon)|^p$$

Using Lemma 2.10, the Lipschitz condition $(\mathbf{D_1})$, estimates (2.38), (2.39), and Corollary 2.20 we have that there exist a constant $D' := D'(p, K, L, \lambda)$, independent of ε such that

$$\alpha_\varepsilon(t) := E\Big[\sup_{-r\leqslant s\leqslant t} |^\eta X(s) - {}^\eta X^\epsilon(s)|^p\Big]$$
$$\leqslant \int_0^t \Big(\|f(u, X_u) - f(u, X_u^\epsilon)\|_{L^p(\Omega;\mathbb{R}^d)}^p + \|g(u, X_u) - g(u, X_u^\epsilon)\|_{L^p(\Omega;\mathbb{R}^{d\times n})}^p$$
$$+ \|(h_0(u, X_u) - h_0(u, X_u^\epsilon))\lambda + h_0(u, X_u^\epsilon)\lambda_\varepsilon\|_{L^p(\Omega;L^p(\nu))}^p$$
$$+ \|(h_0(u, X_u) - h_0(u, X_u^\epsilon))\lambda + h_0(u, X_u^\epsilon)\lambda_\varepsilon\|_{L^p(\Omega, L^2(\nu))}^p$$
$$+ \|h_0(u, X_u^\epsilon)\Lambda(\varepsilon)\|_{L^p(\Omega;\mathbb{R}^{d\times n})}^p\Big)du$$
$$\leqslant D'\int_0^t \Big(\|X_u - X_u^\epsilon\|_{S^p(\Omega;\mathcal{D})}^p + (1 + \|X_u^\epsilon\|_{S^p(\Omega;\mathcal{D})}^p)(\|\lambda_\varepsilon\|_{L^2(\nu)}^p + \|\lambda_\varepsilon\|_{L^p(\nu)}^p + |\Lambda(\varepsilon)|^p)\Big)du$$
$$\leqslant D'\int_0^t \alpha_\varepsilon(u)du + tD'\big(1 + e^{Dt}(Dt + \|\eta\|_{S^p(\Omega;\mathcal{D})}^p)\big)(\|\lambda_\varepsilon\|_{L^2(\nu)}^p + \|\lambda_\varepsilon\|_{L^p(\nu)}^p + |\Lambda(\varepsilon)|^p).$$

Now, set $B_t := tD'\big(1 + e^{Dt}(Dt + \|\eta\|_{S^p(\Omega;\mathcal{D})}^p)\big)$ which is a non-decreasing function in t and hence by Grönwall's inequality, it follows that

$$\alpha_\varepsilon(t) \leqslant B_t e^{D't}\big(\|\lambda_\varepsilon\|_{L^2(\nu)}^p + \|\lambda_\varepsilon\|_{L^p(\nu)}^p + |\Lambda(\varepsilon)|^p\big).$$

Since $|\Lambda(\varepsilon)|^p = \|\lambda_\varepsilon\|_{L^2(\nu)}^p$, the result holds with $A := \max\{2B_T, D'\}$. $\qquad\square$

Remark 2.22. *We have chosen to scale the Brownian motions B^j in the equation* (2.35) *for X^ϵ with $\Lambda^{i,j}(\epsilon) := \|\lambda^{i,j}_\varepsilon\|_{L^2(\nu_j)}$. However, if we return to* (2.33)-(2.34), *we could let Λ_ϵ be any $\mathbb{R}^{k \times n}$- valued function $\Lambda(\varepsilon) \geqslant 0$, $\varepsilon \geqslant 0$, bounded from above and converging to 0 as $\varepsilon \to 0$. Corollary 2.20 and Theorem 2.21 still hold, with the inequality* (2.37) *replaced by*

$$E\left[\sup_{-r \leqslant s \leqslant t} |{}^\eta X(s) - {}^\eta X^\epsilon(s)|^p\right] \leqslant A'e^{A't}(\|\lambda_\varepsilon\|^p_{L^2(\nu)} + \|\lambda_\varepsilon\|^p_{L^p(\nu)} + |\Lambda(\varepsilon)|^p). \qquad (2.40)$$

This can be easily seen by reexamining the proofs of Corollary 2.20 and Theorem 2.21.

3 Itô's formula

In this section we aim at deriving Itô's formula for the SFDDE's studied in Section 2, which we recall, have the form (2.1),

$$dX(t) = f(t, X_t)dt + g(t, X_t)dW(t) + \int_{\mathbb{R}_0} h(t, X_t, z)\tilde{N}(dt, dz)$$
$$X_0 = \eta\,,$$

where X_t is the segment of the process X in $[t - r, t]$, with $r > 0$ a finite delay, taking values in a suitable path space, and $X(t) \in \mathbb{R}^d$ the present value of the process X. For the whole section we work in the M^p-framework. See Section 2.4. Moreover, we assume that f, g and h are *deterministic* functionals, see Remark 2.13.

The main problem, when dealing with the SFDDE (2.1) is that the infinite dimensional process $(X_t)_{t \in [0,T]}$ fails, in general, to be a semimartingale and standard Itô calculus does not apply. In order to overcome this problem several approaches have been used in the literature.

The first attempts go back to [36, 44] where an Itô-type formula for continuous SFDDE was proved via Malliavin calculus. More recently, exploiting the concepts of *horizontal derivative* and *vertical derivative*, a path-wise Itô calculus for non-anticipative stochastic differential equation was derived in [12, 13].

In [20], an Itô formula for Banach-valued continuous processes was proved exploiting the *calculus via regularisation*, where an application to window processes (see [20, Definition 1.4]) is also provided. Several works have followed studying Itô-type formulae for delay equations exploiting the *calculus via regularisation* and showing that the Banach-valued setting and the path-dependent setting can be in fact connected, see, e.g. [14, 16]. Eventually, the connection between the Banach space stochastic calculus and the path-wise calculus was made definitely clear in [15, 25].

We remark that the literature on Itô formulae by the calculus via regularisation deals with equations driven by continuous noises. In this paper, we focus on the SFDDE's with jumps, thus extending the existing literature in this respect. We have chosen to consider the approach of the *calculus via regularisation*, first introduced in [41, 42], which was proved to be well-suited when dealing with infinite-dimensional processes or in the non-semimartingale case, see e.g. [14, 15, 17, 19, 20]. In particular, we prove an Itô formula for the SFDDE (2.1) with values in M^p and we show that our result is coherent with those of [36, 44]. In the Appendix we provide a connection with the path-dependent calculus developed in [12, 13].

Recall that, for a finite delay $r > 0$, $L^p := L^p([-r, 0]; \mathbb{R}^d)$ endowed with the standard norm $\| \cdot \|_{L^p}$, $p \in [2, \infty)$. In what follows we implicitly identify the topological dual of L^p, i.e. $(L^p)^*$, with L^q being $\frac{1}{p} + \frac{1}{q} = 1$, via the natural isomorphism given by

$$J : L^q \to (L^p)^*$$
$$g \mapsto J(g) = {}_q\langle g, \cdot \rangle_p,$$

where $J(g)$ acts on L^p as follows

$$J(g)(f) = {}_q\langle g, f \rangle_p = \int_{-r}^0 g(s) \cdot f(s)ds, \quad g \in L^q, \quad f \in L^p,$$

being $\cdot$ in the integral the usual scalar product in $\mathbb{R}^d$. It is well-known that J is a continuous linear isomorphism and hence, with a slight abuse of notation, we just write $h \in (L^p)^*$ when we actually mean $J^{-1}(h) \in L^q$, i.e. $(L^p)^* \cong L^q$.

Moreover, we denote by $C^1(L^p)$ the space of once Fréchet differentiable, not necessarily linear, functionals $F : L^p \to \mathbb{R}$ with continuous derivative, that is $DF : L^p \to L(L^p, \mathbb{R})$ where $L(L^p, \mathbb{R})$ denotes the space of continuous linear operators from L^p to $\mathbb{R}$. Now, since F is $\mathbb{R}$-valued, we actually have that $L(L^p, \mathbb{R}) = (L^p)^*$. Hence we can regard $DF(f)$, $f \in L^p$ as an element in L^q via J^{-1}. In a summary, we identify $DF(f)$ with $J^{-1}(DF(f))$ and simply write $DF : L^p \to L^q$ so that

$$DF(f)(g) = {}_q\langle DF(f), g\rangle_p = \int_{-r}^0 DF(f)(s) \cdot g(s) ds, \quad f \in L^p, \quad g \in L^q,$$

where the first equality is, by an abuse of notation, meant as an identification.

Also, recall that $|\cdot|$ denotes the Euclidean norm in $\mathbb{R}^d$. Finally, recall that $M^p = L^p \times \mathbb{R}^d$ endowed with the standard product norm.

Let the SFDDE

$$\begin{cases} dX(t) = f(t, X_t, X(t))dt + g(t, X_t, X(t))dW(t) + \int_{\mathbb{R}_0} h(t, X_t, X(t))(z)\tilde{N}(dt, dz), \\ (X_0, X(0)) = (\eta, x) \in M^p, \end{cases} \quad (3.1)$$

for $t \in [0, T]$, $T < \infty$. We assume that $f : [0, T] \times M^p \to \mathbb{R}^d$, $g : [0, T] \times M^p \to \mathbb{R}^{d \times m}$ and $h : [0, T] \times M^p \times \mathbb{R}_0 \to \mathbb{R}^{d \times n}$ satisfy Assumptions $(\mathbf{L_1})$ and $(\mathbf{L_2})$ so that Theorem 2.16 holds and equation (3.1) admits a unique strong solution.

In the sequel every process is indexed by the time $t \in [0, T]$ and following [20], if necessary, we extend the process $X = (X(t))_{t \in [0,T]}$ to the positive real line as follows

$$X(t) := \begin{cases} X(t) & \text{for } t \in [0, T], \\ X(T) & \text{for } t > T, \end{cases} .$$

Next, we consider the definition of forward integral.

Definition 3.1. *Let* $X = \{X(s), s \in [0, T]\}$ *and* $Y = \{Y(s), s \in [0, T]\}$ *two* $\mathbb{R}^d-$*valued process. For every* $t \in [0, T]$ *we define the forward integral of* Y *w.r.t.* X *by* $\int_0^t Y(s) \cdot dX(s)$ *as the following limit,*

$$\int_0^t Y(s) \cdot dX(s) := \lim_{\epsilon \searrow 0} \int_0^t Y(s) \cdot \frac{X(s + \epsilon) - X(s)}{\epsilon} ds, \quad (3.2)$$

where the convergence is uniformly on compacts in probability (ucp).

Similarly, let $X = \{X_s, s \in [0, T]\}$, *and* $Y = \{Y_s, s \in [0, T]\}$, *be* L^p-*valued and* L^q-*valued processes, respectively. For every* $t \in [0, T]$ *we define the* L^p-*forward integral of* Y *w.r.t.* X *as the following limit,*

$$\int_0^t \int_{-r}^0 Y_s(\theta) \cdot dX_s(\theta) := \lim_{\epsilon \searrow 0} \int_0^t \int_{-r}^0 Y_s(\theta) \cdot \frac{X_{s+\epsilon}(\theta) - X_s(\theta)}{\epsilon} d\theta ds, \quad (3.3)$$

where the convergence is uniformly on compacts in probability. We introduce the short-hand notation:

$$\int_0^t {}_q\langle Y_s, dX_s\rangle_p := \int_0^t \int_{-r}^0 Y_s(\theta) \cdot dX_s(\theta).$$

Recall that a sequence of real-valued processes $\{X^n\}_{n \geqslant 1}$ converges to a process X *uniformly on compacts in probability* (ucp), if for each $t \in [0, T]$ we have that

$$\sup_{0 \leqslant s \leqslant t} |X_s^n - X_s| \to 0,$$

in probability. See e.g. [38, p.57]. In this section, if not otherwise stated, any limit will be taken in the ucp sense.

Remark 3.2. *Following [20], in Definition 3.1 we have considered the ucp limit. In fact the space of càdlàg functions is a metrizable space with the metric induced by the ucp topology, see, e.g. [38, p.57]. This implies that, being the approximating sequence in the right-hand-side of equation (3.2) càdlàg, the ucp convergence ensures that the limiting process, that is the forward integral, is also càdlàg.*

Let us now introduce the notation we will use in the present work.

Definition 3.3. *Let $F : [0, T] \times L^p \times \mathbb{R}^d \to \mathbb{R}$ a given function, we say that*

$$F \in C^{1,1,2} \left([0, T] \times L^p \times \mathbb{R}^d \right) ,$$

if F is continuously differentiable w.r.t. the first variable, Fréchet differentiable with continuous derivative w.r.t. the second variable, and twice continuously differentiable w.r.t. the third variable.

We thus denote by ∂_t the derivative w.r.t. to time, $D_i F$ the Fréchet derivative w.r.t. the i-th component of the segment X_t and ∂_i the derivative w.r.t. the i-th component of the present state $X(t)$ and finally, $\partial_{i,j}$ the second order derivative w.r.t. the i, j-th component of $X(t)$.

We will then define the Fréchet gradient w.r.t. the segment as

$$D := (D_1, \ldots, D_d) ,$$

the gradient w.r.t. the present state

$$\nabla_x := (\partial_1, \ldots, \partial_d) ,$$

and the Hessian matrix w.r.t. the present state

$$\nabla_x^2 := (\partial_{i,j})_{i,j=1,\ldots,d} .$$

Definition 3.4. *Let $\eta \in W^{1,p}([-r, 0]; \mathbb{R}^d) =: W^{1,p}$, then we define by $\partial_{\theta,i}\eta$ and $\partial_{\theta,i}^+\eta$ the weak derivative and the right weak derivative, respectively, of the i-th component of η. Accordingly we define the gradient as*

$$\nabla_\theta := (\partial_{\theta,1}, \ldots, \partial_{\theta,d}) , \quad resp. \ \nabla_\theta^+ := \left(\partial_{\theta,1}^+, \ldots, \partial_{\theta,d}^+ \right) .$$

Eventually, in proving Itô's formula, we will need the notion of *modulus of continuity* of operators between infinite-dimensional normed spaces.

Definition 3.5 (Modulus of continuity). *Let $(Y_1, \| \cdot \|_{Y_1})$ and $(Y_2, \| \cdot \|_{Y_2})$ be two normed spaces and $F : Y_1 \to Y_2$ a uniformly continuous function. We define the modulus of continuity of F as*

$$\varpi(\epsilon) := \sup_{\|y-y'\|_{Y_1} \leqslant \epsilon} \|F(y) - F(y')\|_{Y_2}, \quad \epsilon > 0.$$

We thus have the following Itô's formula for SFDDE (3.1).

Theorem 3.6 (Itô's formula). *Let X be the solution to equation (3.1). Let $F : [0, T] \times L^p \times \mathbb{R}^d \to \mathbb{R}$ be such that $F \in C^{1,1,2} \left([0, T] \times L^p \times \mathbb{R}^d \right)$; for any $t \in [0, T]$, $\eta \in L^p$ and $x \in \mathbb{R}^d$, $DF(t, \eta, x) \in W^{1,q}$, (q such that $\frac{1}{p} + \frac{1}{q} = 1$) and $\nabla_\theta DF(t, \cdot, x) : L^p \to L^q$ is uniformly continuous. Then the following limit exists in the ucp sense,*

$$\lim_{\epsilon \searrow 0} \frac{1}{\epsilon} \int_0^t {}_q\langle DF(s, X_s, X(s)), X_{s+\epsilon} - X_s\rangle_p \, ds =: \int_0^t {}_q\langle DF(s, X_s, X(s)), dX_s\rangle_p. \quad (3.4)$$

Moreover, for $t \in [0, T]$, we have that

$$F(t, X_t, X(t)) = F(0, X_0, X(0)) + \int_0^t \partial_t F(s, X_s, X(s)) ds + \int_0^t {}_q \langle DF(s, X_s, X(s)), dX_s \rangle_p$$

$$+ \int_0^t \nabla_x F(s, X_s, X(s)) \cdot dX(s) + \frac{1}{2} \int_0^t Tr \left[g^*(s, X_s, X(s)) \nabla_x^2 F(s, X_s, X(s)) g(s, X_s, X(s)) \right] ds$$

$$+ \int_0^t \int_{\mathbb{R}_0} \left(F(s, X_s, X(s) + h(s, X_s, X(s))(z)) - F(s, X_s, X(s)) \right) N(ds, dz)$$

$$- \int_0^t \int_{\mathbb{R}_0} \nabla_x F(s, X_s, X(s)) h(s, X_s, X(s))(z) N(ds, dz),$$

$$(3.5)$$

holds, P-a.s., where we have denoted by Tr the trace and by g^* the adjoint of g and the one but last term in equation (3.5) *has to be intended component-wise, that is*

$$\int_0^t \int_{\mathbb{R}_0} \left(F(s, X_s, X(s) + h(s, X_s, X(s))(z)) - F(s, X_s, X(s)) \right) N(ds, dz)$$

$$:= \sum_{i=1}^n \int_0^t \int_{\mathbb{R}_0} \left(F(s, X_s, X(s) + h^{\cdot, i}(s, X_s, X(s))(z)) - F(s, X_s, X(s)) \right) N^i(ds, dz).$$

Proof. Let $t \in [0, T]$. First, observe that for $\varepsilon > 0$ small enough, we have

$$\frac{1}{\epsilon} \int_0^t F(s + \epsilon, X_{s+\epsilon}, X(s + \epsilon)) - F(s, X_s, X(s)) ds =$$

$$\frac{1}{\epsilon} \int_\epsilon^{t+\epsilon} F(s, X_s, X(s)) ds - \frac{1}{\varepsilon} \int_0^t F(s, X_s, X(s)) ds = \qquad (3.6)$$

$$= \frac{1}{\epsilon} \int_t^{t+\epsilon} F(s, X_s, X(s)) ds - \frac{1}{\epsilon} \int_0^\epsilon F(s, X_s, X(s)) ds,$$

which, by the continuity of F and X_s and the right-continuity of $X(s)$, $s \in [0, T]$, recalling remark 3.2 and arguing as in [5, Theorem 3.10], converges ucp to

$$F(t, X_t, X(t)) - F(0, X_0, X(0)).$$

The first part of (3.6) can be rewritten as

$$\frac{1}{\epsilon} \int_0^t F(s + \epsilon, X_{s+\epsilon}, X(s + \epsilon)) - F(s, X_s, X(s)) ds$$

$$= \underbrace{\frac{1}{\epsilon} \int_0^t F(s + \epsilon, X_{s+\epsilon}, X(s + \epsilon)) - F(s, X_{s+\epsilon}, X(s + \epsilon)) ds}_{J_\epsilon^1}$$

$$+ \underbrace{\frac{1}{\epsilon} \int_0^t F(s, X_{s+\epsilon}, X(s + \epsilon)) - F(s, X_{s+\epsilon}, X(s)) ds}_{J_\epsilon^2}$$

$$+ \underbrace{\frac{1}{\epsilon} \int_0^t F(s, X_{s+\epsilon}, X(s)) - F(s, X_s, X(s)) ds}_{J_\epsilon^3}.$$

Following the same arguments as in the proof of [20, Theorem 5.2] we can show that

$$\lim_{\epsilon \searrow 0} J_\epsilon^1 = \int_0^t \partial_t F(s, X_s, X(s)) ds, \quad \text{ucp.}$$

24

Let us now consider J_ϵ^2. A straightforward application of [5, Theorem 3.10] implies that

$$
J_\epsilon^2 \to \int_0^t \nabla_x F(s, X_s, X(s)) \cdot dX(s)
$$
$$
+ \frac{1}{2} \int_0^t Tr\left[g^*(s, X_s, X(s)) \nabla_x^2 F(s, X_s, X(s)) g(s, X_s, X(s)) \right] ds
$$
$$
+ \int_0^t \int_{\mathbb{R}_0} \left(F(s, X_s, X(s) + h(s, X_s, X(s), z)) - F(s, X_s, X(s)) \right) N(ds, dz)
$$
$$
- \int_0^t \int_{\mathbb{R}_0} \nabla_x F(s, X_s, X(s)) h(s, X_s, X(s), z) N(ds, dz), \quad \text{as } \epsilon \searrow 0.
$$

Let us now show that

$$
\lim_{\epsilon \searrow 0} J_\epsilon^3 = \int_0^t {}_q\langle DF(s, X_s, X(s)), dX_s\rangle_p ds.
$$

By an application of the infinite-dimensional version of Taylor's theorem of order one (see e.g. [45, Ch. 4, Theorem 4.C]), we obtain

$$
\frac{1}{\epsilon} \int_0^t F(s, X_{s+\epsilon}, X(s)) - F(s, X_s, X(s)) \, ds
$$
$$
= \frac{1}{\epsilon} \int_0^t \int_0^1 {}_q\langle DF(s, X_s + \tau(X_{s+\epsilon} - X_s), X(s)), X_{s+\epsilon} - X_s\rangle_p \, d\tau ds
$$
$$
= \frac{1}{\epsilon} \int_0^t \int_0^1 \int_{-r}^0 DF(s, X_s + \tau(X_{s+\epsilon} - X_s), X(s))(\alpha) \cdot (X(s+\epsilon+\alpha) - X(s+\alpha)) d\alpha d\tau ds
$$
$$
= -\int_0^t \underbrace{\frac{1}{\epsilon} \int_0^1 \int_{-r}^0 \left(DF(s, X_{s,s+\epsilon}^\tau, X(s))(\alpha+\epsilon) - DF(s, X_{s,s+\epsilon}^\tau, X(s))(\alpha) \right) \cdot X(s+\epsilon+\alpha) d\alpha d\tau}_{J_{\epsilon,s}^{3,1}} \, ds
$$
$$
+ \int_0^t \underbrace{\frac{1}{\epsilon} \int_0^1 \int_{-r}^0 DF(s, X_{s,s+\epsilon}^\tau, X(s))(\alpha+\epsilon) \cdot X(s+\alpha+\epsilon) d\alpha d\tau}_{J_{\epsilon,s}^{3,2}} \, ds
$$
$$
- \int_0^t \underbrace{\frac{1}{\epsilon} \int_0^1 \int_{-r}^0 DF(s, X_{s,s+\epsilon}^\tau, X(s))(\alpha) \cdot X(s+\alpha) d\alpha d\tau}_{J_{\epsilon,s}^{3,2}} \, ds,
$$

$$(3.7)$$

where we have denoted by $X_{s,s+\epsilon}^\tau := X_s + \tau(X_{s+\epsilon} - X_s)$. We apply the change of variables $g(\alpha) = \alpha + \epsilon$ to the first term of $J_\epsilon^{3,2}$ in Equation (3.7) in order to obtain

$$
J_{\epsilon,s}^{3,2} = \frac{1}{\epsilon} \int_0^1 \int_{-r+\epsilon}^\epsilon DF(s, X_{s,s+\epsilon}^\tau, X(s))(\alpha) \cdot X(s+\alpha) d\alpha d\tau
$$
$$
- \frac{1}{\epsilon} \int_0^1 \int_{-r}^0 DF(s, X_{s,s+\epsilon}^\tau, X(s))(\alpha) \cdot X(s+\alpha) d\alpha d\tau
$$
$$
= \underbrace{\frac{1}{\epsilon} \int_0^1 \int_0^\epsilon DF(s, X_{s,s+\epsilon}^\tau, X(s))(\alpha) \cdot X(s+\alpha) d\alpha d\tau}_{J_{\epsilon,s}^{3,2,1}}
$$
$$
- \underbrace{\frac{1}{\epsilon} \int_0^1 \int_{-r}^{-r+\epsilon} DF(s, X_{s,s+\epsilon}^\tau, X(s))(\alpha) \cdot X(s+\alpha) d\alpha d\tau}_{J_{\epsilon,s}^{3,2,2}}.
$$

We thus have, from the continuity of DF and X_s, that s-a.e.,

$$\lim_{\epsilon \searrow 0} J_{\epsilon,s}^{3,2,1} = DF(s, X_s, X(s))(0) \cdot X(s), \quad \lim_{\epsilon \searrow 0} J_{\epsilon}^{3,2,2} = DF(s, X_s, X(s))(-r) \cdot X(s-r).$$

Let $\nabla_\theta DF(s, X_{s,s+\epsilon}^\tau, X(s))$ denote a version of the weak derivative of $DF(s, X_{s,s+\epsilon}^\tau, X(s)) \in W^{1,q}$. Consider $J_{\epsilon,s}^{3,1}$. Using the mean value-theorem and interchanging the order of integration by Fubini's theorem we have

$$
\begin{aligned}
J_{\epsilon,s}^{3,1} &= \frac{1}{\epsilon} \int_0^1 \int_{-r}^0 \int_\alpha^{\alpha+\epsilon} \nabla_\theta DF(s, X_{s,s+\epsilon}^\tau, X(s))(\beta) \, d\beta \cdot X(s+\epsilon+\alpha) d\alpha d\tau \\
&= \frac{1}{\epsilon} \int_0^1 \int_{-r}^{-r+\epsilon} \int_{-r}^\beta \nabla_\theta DF(s, X_{s,s+\epsilon}^\tau, X(s))(\beta) \cdot X(s+\epsilon+\alpha) \, d\alpha d\beta d\tau \\
&\quad + \frac{1}{\epsilon} \int_0^1 \int_{-r+\epsilon}^0 \int_{\beta-\epsilon}^\beta \nabla_\theta DF(s, X_{s,s+\epsilon}^\tau, X(s))(\beta) \cdot X(s+\epsilon+\alpha) \, d\alpha d\beta d\tau \\
&\quad + \frac{1}{\epsilon} \int_0^1 \int_0^\epsilon \int_{\beta-\epsilon}^0 \nabla_\theta DF(s, X_{s,s+\epsilon}^\tau, X(s))(\beta) \cdot X(s+\epsilon+\alpha) \, d\alpha d\beta d\tau.
\end{aligned}
\tag{3.8}
$$

Now, we add and subtract integral terms so that the second integral on the right-hand side of (3.8) goes from $-r$ to 0, that is

$$
\begin{aligned}
J_{\epsilon,s}^{3,1} &= \frac{1}{\epsilon} \int_0^1 \int_{-r}^0 \int_{\beta-\epsilon}^\beta \nabla_\theta DF(s, X_{s,s+\epsilon}^\tau, X(s))(\beta) \cdot X(s+\epsilon+\alpha) \, d\alpha d\beta d\tau \\
&\quad + \frac{1}{\epsilon} \int_0^1 \int_0^\epsilon \int_{\beta-\epsilon}^0 \nabla_\theta DF(s, X_{s,s+\epsilon}^\tau, X(s))(\beta) \cdot X(s+\epsilon+\alpha) \, d\alpha d\beta d\tau \\
&\quad - \frac{1}{\epsilon} \int_0^1 \int_{-r}^{-r+\epsilon} \int_{\beta-\epsilon}^{-r} \nabla_\theta DF(s, X_{s,s+\epsilon}^\tau, X(s))(\beta) \cdot X(s+\epsilon+\alpha) \, d\alpha d\beta d\tau.
\end{aligned}
\tag{3.9}
$$

Then, Lebesgue's differentiation theorem implies that the second and third integral on the right-hand side of (3.9) converge to 0 as $\epsilon \to 0$ for a.e. s. Concerning the first integral on the right-hand side of (3.9), we add and subtract the corresponding terms in order to have

$$
\begin{aligned}
&\frac{1}{\epsilon} \int_0^1 \int_{-r}^0 \int_{\beta-\epsilon}^\beta \nabla_\theta DF(s, X_{s,s+\epsilon}^\tau, X(s))(\beta) \cdot X(s+\epsilon+\alpha) \, d\alpha d\beta d\tau \\
&= \underbrace{\frac{1}{\epsilon} \int_0^1 \int_{-r}^0 \left(\nabla_\theta DF(s, X_{s,s+\epsilon}^\tau, X(s))(\beta) - \nabla_\theta DF(s, X_s, X(s))(\beta)\right) \cdot \left(\int_{\beta-\epsilon}^\beta X(s+\epsilon+\alpha)d\alpha\right) d\beta d\tau}_{J_{\epsilon,s}^{3,1,1}} \\
&\quad + \underbrace{\frac{1}{\epsilon} \int_0^1 \int_{-r}^0 \nabla_\theta DF(s, X_s, X(s))(\beta) \cdot \left(\int_{\beta-\epsilon}^\beta X(s+\epsilon+\alpha)d\alpha\right) d\beta d\tau}_{J_{\epsilon,s}^{3,1,2}}.
\end{aligned}
\tag{3.10}
$$

Using Hölder's inequality with exponents, $p, q \geqslant 2$, $\frac{1}{p} + \frac{1}{q} = 1$ on $J_{\epsilon,s}^{3,1,1}$ we have

$$
\begin{aligned}
&|J_{\epsilon,s}^{3,1,1}| \\
&\leqslant \int_0^1 \|\nabla_\theta DF(s, X_{s,s+\epsilon}^\tau, X(s)) - \nabla_\theta DF(s, X_s, X(s))\|_{L^q} d\tau \left(\int_{-r}^0 \left|\frac{1}{\epsilon} \int_\beta^{\beta+\epsilon} X_s(\alpha)d\alpha\right|^p d\beta\right)^{1/p} \\
&\leqslant \int_0^1 \|\nabla_\theta DF(s, X_{s,s+\epsilon}^\tau, X(s)) - \nabla_\theta DF(s, X_s, X(s))\|_{L^q} d\tau \, \|X_s^*\|_{L^p},
\end{aligned}
$$

where, for a.e. s,

$$X_s^*(\beta) := \sup_{\epsilon > 0} \frac{1}{\epsilon} \int_\beta^{\beta + \epsilon} |X_s(\alpha)| \, d\alpha, \quad \beta \in [-r, 0]$$

is the Hardy-Littlewood maximal function evaluated at β. For a.e. s, the Hardy-Littlewood operator $X_s \mapsto X_s^*$ is a (non-linear) bounded operator from L^p to L^p, $p > 1$. Hence, we can apply Lebesgue's dominated convergence theorem and the fact that by Lebesgue's differentiation theorem we have

$$\lim_{\epsilon \searrow 0} \left(\int_{-r}^0 \frac{1}{\epsilon} \int_{\beta - \epsilon}^\beta |X(s + \epsilon + \alpha)|^p \, d\alpha d\beta \right)^{1/p} = \|X_s\|_{L^p}.$$

The above arguments imply the following estimate and convergence s-a.e.,

$$|J_{\epsilon,s}^{3,1,1}| \leqslant \varpi_s(\epsilon) \|X_s\|_{L^p} \to 0, \quad \epsilon \to 0 \,,$$

where $\varpi_s(\epsilon)$ denotes the modulus of continuity of $\nabla_\theta DF(s, \cdot, X(s))$ from Definition 3.5. Further, we can formally apply integration by parts to $J_{\epsilon,s}^{3,1,2}$ in order to obtain

$$J_{\epsilon,s}^{3,1,2} = \frac{1}{\epsilon} DF(s, X_s, X(s))(\beta) \cdot \int_{\beta - \epsilon}^\beta X_{s+\epsilon}(\alpha) d\alpha \bigg|_{-r}^0 - \int_{-r}^0 DF(s, X_s, X(s))(\beta) \cdot \frac{X_{s+\epsilon}(\beta) - X_s(\beta)}{\epsilon} d\beta.$$

(3.11)

Then it follows that

$$J_{\epsilon,s}^{3,1,2} \to DF(s, X_s, X(s))(-r) \cdot X(s-r) - DF(s, X_s, X(s))(0) \cdot X(s) - {}_q\langle DF(s, X_s, X(s)), dX_s\rangle_p,$$

s-a.e. as $\epsilon \searrow 0$. Altogether, we have finally shown that

$$\lim_{\epsilon \searrow 0} J_\epsilon^3 = \int_0^t {}_q\langle DF(s, X_s, X(s)), dX_s\rangle_p \,.$$

This corresponds to (3.4). Hence, we conclude the proof. $\qquad \square$

In Appendix **??** it is shown, exploiting the results obtained in [25], that the Itô formula (3.5) is coherent with the Itô formula for path-dependent processes with jumps proved in [12], as well as the results obtained in [25].

Let us consider the *forward integral*

$$\int_0^t {}_q\langle DF(s, X_s, X(s)), dX_s\rangle_p \,,$$

introduced in Theorem 3.6, we want now to relate the *forward integral* to the operator introduced in [44]. In fact, since the right-derivative operator introduced in Definition 3.4 is the infinitesimal generator of the left-shift semigroup introduced in [44], it can be shown that the *forward integral* does coincide, under some suitable regularity conditions, with the operator $\mathcal{S}F(s, X_s, X(s))$ introduced in [44].

Proposition 3.7. *Let X be the solution to equation (3.1) and let $F : [0, T] \times L^p \times \mathbb{R}^d \to \mathbb{R}$ be such that $F \in C^{1,1,2}([0,T] \times L^p \times \mathbb{R}^d)$ and such that the forward integral defined in Equation (3.4) is well-defined as a limit in probability uniformly on compacts. Additionally, let us assume that $X_s \in W^{1,p}$ for every $s \in [0, T]$. Then*

$$\int_0^t {}_q\langle DF(s, X_s, X(s)), dX_s\rangle_p = \int_0^t {}_q\langle DF(s, X_s, X(s)), \nabla_\theta^+ X_s\rangle_p ds = \int_0^t \mathcal{S}F(s, X_s, X(s)) ds$$

(3.12)

holds P-a.s., where $\mathcal{S}F(s, X_s, X(s))$ is the operator introduced in [44, Section. 9].

Proof. Let $X_s \in W^{1,p}$. From the fundamental theorem of calculus for absolutely continuous functions we have, P-a.s.,

$$\lim_{\epsilon \searrow 0} {}_q\langle DF(s, X_s, X(s)), \frac{X_{s+\epsilon} - X_s}{\epsilon} - \nabla_\theta^+ X_s \rangle_p$$

$$\lim_{\epsilon \searrow 0} \int_{-r}^0 DF(s, X_s, X(s))(\beta) \cdot \left(\frac{(X(s+\epsilon+\beta) - X(s+\beta))}{\epsilon} - \nabla_\theta^+ X(s+\beta) \right) d\beta$$

$$= \lim_{\epsilon \searrow 0} \int_{-r}^0 DF(s, X_s, X(s))(\beta) \cdot \left(\frac{1}{\epsilon} \int_{s+\beta}^{s+\beta+\epsilon} \nabla_\theta X(r) dr - \nabla_\theta^+ X(s+\beta) \right) d\beta$$

Now, by Lebesgue's dominated convergence theorem, which is justified by analogous arguments as for the convergence of (3.10), we finally get

$$\lim_{\epsilon \searrow 0} \int_{-r}^0 DF(s, X_s, X(s))(\beta) \cdot \left(\frac{1}{\epsilon} \int_{s+\beta}^{s+\beta+\epsilon} \nabla_\theta X(r) dr - \nabla_\theta^+ X(s+\beta) \right) d\beta = 0.$$

$\square$

Exploiting the standard definition of the Poisson random measure, we can now give another formulation of the Itô formula (3.6).

Theorem 3.8 (Itô's formula). *Let the hypothesis of Theorem 3.6 hold, then*

$$F(t, X_t, X(t)) = F(0, X_0, X(0)) + \int_0^t {}_q\langle DF(s, X_s, X(s)), dX_s \rangle_p ds +$$

$$+ \int_0^t \partial_t F(s, X_s, X(s)) ds + \int_0^t \nabla_x F(s, X_s, X(s)) \cdot dX(s) ds +$$

$$+ \frac{1}{2} \int_0^t \int_0^t Tr \left[g^*(s, X_s, X(s)) \nabla_x^2 F(s, X_s, X(s)) g(s, X_s, X(s)) \right] ds +$$

$$+ \sum_{s \leqslant t} F(s, X_s, X(s)) - F(s, X_s, X(s-)) - \Delta X(s) \cdot \nabla_x F(s, X_s, X(s)),$$

(3.13)

holds P-a.s., where the notation is as in Theorem 3.6 and $\Delta X(s)$ is the jump of the process X at time s, namely

$$\Delta X(s) := X(s) - X(s-).$$

Proof. It immediately follows from Theorem 3.6 and [2, Theorem 4.4.10]. $\square$

Hereinafter, we state a crucial probabilistic property of the solution of SFDDE (2.1) which is needed for the derivation of Feynman-Kac's formula also stated below. As it is perceptible, the finite-dimensional process ${}^{(\eta,x)}X(t)$, $t \in [0, T]$, $(\eta, x) \in M^p$ is not Markov, since the value of ${}^{(\eta,x)}X(t)$ depends on the *near* past. Nevertheless, if we enlarge the state space and regard the process X as a process of the segment, i.e. with infinite-dimensional state space, in the present case M^p, then such process is indeed Markovian.

The proof follows almost immediately given the fact that the Markov property of the solution is fully known for the case without jumps, i.e. $h = 0$, see [35, Theorem (III. 1.1)], which actually follows from measure theoretical properties of the driving noise and not path or distributional properties of it. More concretely, one would expect the Markov property of the solution to hold if, for instance, the driving noises have independent increments which is the case in our setting.

In order to state the Markov property one is compelled to look at solutions starting at time $t_1 \geqslant 0$, that is we hereby consider ${}^{(t_1,\eta,x)}X_t$, $t \geqslant t_1$, $(\eta, x) \in M^p$, the segment of the solution starting in (η, x) at times $t_1 \geqslant 0$, i.e.

$${}^{(t_1,\eta,x)}X(t) = \eta(0) + \int_{t_1}^t f(s, {}^{(t_1,\eta,x)}X_s, {}^{(t_1,\eta,x)}X(s)) ds + \int_{t_1}^t g(s, {}^{(t_1,\eta,x)}X_s, {}^{(t_1,\eta,x)}X(s)) dW(s)$$

$$+ \int_{t_1}^t \int_{\mathbb{R}_0} h(s, {}^{(t_1,\eta,x)}X_s, {}^{(t_1,\eta,x)}X(s), z) \widetilde{N}(ds, dz),$$

for every $t \in [t_1, T]$ and $^{(t_1, \eta, x)}X(t) = \eta(t - t_1)$ for every $t \in [t_1 - t, t_1)$. Define further the family of operators

$$T_t^{t_1} : L^2(\Omega, \mathcal{F}_{t_1}; M^p) \longrightarrow L^2(\Omega, \mathcal{F}_t; M^p)$$

$$(\eta, x) \longmapsto \left({}^{(t_1, \eta, x)}X_t, \, {}^{(t_1, \eta, x)}X(t) \right).$$

We denote $T_t = T_t^0$. It turns out that, under hypotheses $(\mathbf{L_1})$ and $(\mathbf{L_2})$, $T_t^{t_1}$ is Lipschitz continuous and the family of operators $\{T_t^{t_1}\}_{0 \leqslant t_1 \leqslant t \leqslant T}$ defines a semigroup on $L^2(\Omega, M^p)$, i.e.

$$T_{t_2}(\eta, x) = T_{t_2}^{t_1} \circ T_{t_1}(\eta, x),$$

for every $0 \leqslant t_1 \leqslant t_2 \leqslant T$ and $(\eta, x) \in L^2(\Omega, \mathcal{F}_0; M^p)$. The latter property can easily be obtained by showing that both sides of the identity solve the same SFDDE and the fact that solutions are unique, see [35, Theorem (II. 2.2)] for the case $h = 0$.

Theorem 3.9 (The Markov property). *Assume hypotheses* $(\mathbf{L_1})$ *and* $(\mathbf{L_2})$ *hold and* $^{(\eta, x)}X(t)$, $t \in [0, T]$, $(\eta, x) \in M^p$ *denotes the unique strong solution of the SFDDE* (2.1). *Then the random field*

$$\left\{ \left({}^{(\eta, x)}X_t, \, {}^{(\eta, x)}X(t) \right) : t \in [0, T], (\eta, x) \in M^p \right\}$$

describes a Markov process on M^p *with transition probabilities given by*

$$p(t_1, (\eta, x), t_2, B) = P(\omega \in \Omega, \, T_{t_2}^{t_1}(\eta, x)(\omega) \in B),$$

for $0 \leqslant t_1 \leqslant t_2 \leqslant T$, $(\eta, x) \in M^p$ *and Borel set* $B \in \mathcal{B}(M^p)$. *In other words, for any* $(\eta, x) \in L^2(\Omega, \mathcal{F}_0; M^p)$ *and Borel set* $B \in \mathcal{B}(M^p)$, *the Markov property*

$$P(T_{t_2}(\eta, x) \in B | \mathcal{F}_{t_1}) = p(t_1, T_{t_1}(\eta, x), t_2, B) = P(T_{t_2}(\eta, x) \in B | T_{t_1}(\eta, x))$$

holds a.s. on Ω.

Proof. we can see that for every $0 \leqslant t_1 \leqslant t_2 \leqslant T$ and every $(\eta, x) \in M^p$, the mapping $B \mapsto p(t_1, (\eta, x), t_2, B) = P \circ (T_{t_2}^{t_1}(\eta, x))^{-1}(B)$, $B \in \mathcal{B}(M^p)$ defines a probability measure on M^p, since the random variable $T_{t_2}^{t_1}(\eta, x) : \Omega \to M^p$ is $(\mathcal{F}, \mathcal{B}(M^p))$-measurable. We would then like to show that, if $0 \leqslant t_1 \leqslant t_2 \leqslant T$, then

$$P\left(\omega \in \Omega : T_{t_2}(\eta, x)(\omega) \in B | \mathcal{F}_{t_1} \right)(\omega') = p(t_1, T_{t_1}(\eta, x)(\omega'), t_2, B) \tag{3.14}$$

for almost all $\omega' \in \Omega$, every Borel set $B \in \mathcal{B}(M^p)$ and any $(\eta, x) \in L^2(\Omega, \mathcal{F}_0; M^p)$. The right-hand side of (3.14) equals

$$\int_\Omega 1_B \left(\left(T_{t_2}^{t_1}(T_{t_1}(\eta)(\omega')) \right)(\omega) \right) P(d\omega)$$

for almost all $\omega' \in \Omega$. Hence, by the definition of conditional expectation, identity (3.14) is synonymous with

$$\int_A 1_B \left(T_{t_2}(\eta, x)(\omega) \right) P(d\omega) = \int_A \int_\Omega 1_B \left(\left(T_{t_2}^{t_1}(T_{t_1}(\eta, x)(\omega')) \right)(\omega) \right) P(d\omega) P(d\omega') \tag{3.15}$$

for all $A \in \mathcal{F}_{t_1}$ and all $B \in \mathcal{B}(M^p)$. In a summary, the main challenge is to verify relation (3.15) which is stated in quite general terms.

Let $\mathcal{G}_t$, $t \in [0, T]$ be the σ-algebra generated by $\{N((s, u], B), t < s \leqslant u \leqslant T, B \in \mathcal{B}(\mathbb{R}_0)\}$. The key steps in proving (3.15) according to [35, Theorem (III. 1.1)] are the following. First, the family of operators $\{T_t\}_{t \in [0, T]}$ defines a semigroup on $L^2(\Omega, M^p)$. Secondly, the σ-algebras $\mathcal{F}_t$ and $\mathcal{G}_t$ are independent for every $t \in [0, T]$, and $\{T_t^{t_1}(\eta, x)\}_{t \geqslant t_1}$ is adapted to $\{\mathcal{F}_t \cap \mathcal{G}_{t_1}\}_{t \geqslant t_1}$, being each $\mathcal{F}_t \cap \mathcal{G}_{t_1}$ independent of $\mathcal{F}_{t_1}$. Finally, a key point to prove (3.15) is that one can first prove

$$\int_A f \left(T_{t_2}(\eta, x)(\omega) \right) P(d\omega) = \int_A \int_\Omega f \left(\left(T_{t_2}^{t_1}(T_{t_1}(\eta, x)(\omega')) \right)(\omega) \right) P(d\omega) P(d\omega') \tag{3.16}$$

for any bounded continuous function $f : M^p \to \mathbb{R}$. Then one can use a limit argument to show the relation (3.15) for the case $f = 1_B$ being B just an open set of M^p and eventually for any general indicator function on Borel sets. The argument which transfers (3.16) into the case $f = 1_B$ for any open set B in M^p requires that the state space in consideration is separable so that 1_B, being B an open set of M^p, can be approximated by uniformly continuous partitions of unity $\{f_m\}_{m \in \mathbb{N}}$, $f_m : M^p \to \mathbb{R}$ such that $\lim_{m \to \infty} f_m = 1_B$. All these properties above mentioned are indeed satisfied in our framework. $\qquad \square$

Exploiting Itô's formula from Theorem 3.6 together with the Markov property from Theorem 3.9, we can now prove a Feynman-Kac theorem for M^p−valued SFDDE's with jumps.

Theorem 3.10 (Feynman-Kac theorem)**.** *Let $(X_t, X(t))$ be the solution to the SFDDE (3.1). Assume that a function $F : [0, T] \times L^p \times \mathbb{R}^d \to \mathbb{R}$ satisfies the hypothesis of Theorem 3.6 and that F solves the following path-dependent partial integro-differential equation (PPIDE)*

$$\begin{cases} \partial_t F(t, \eta, x) & = {}_q\langle DF(s, \eta, x), d\eta \rangle_p + \nabla_x F(t, \eta, x) \cdot f(t, \eta) \\ & \quad + \frac{1}{2} Tr\left[g(t, \eta, x) g^*(t, \eta, x) \nabla_x^2 F(t, \eta, x) \right] \\ & \quad + \int_{\mathbb{R}_0} \left(F(t, \eta, x + h(t, \eta, x)(z)) - F(t, \eta, x) - \nabla_x F(t, \eta, x) h(t, \eta, x)(z) \right) \nu(dz), \\ F(T, \eta, x) & = \Phi(\eta, x), \end{cases}$$

(3.17)

with Φ a given function defined on $M^p = L^p \times \mathbb{R}^d$, then we have the following representation

$$F(t, \eta, x) = \mathbb{E}\left[\Phi(X_T, X(T)) | X_t = \eta, X(t) = x \right], \quad t \in [0, T]. \tag{3.18}$$

Conversely, consider the function F given by (3.18) where $\Phi \in C^{1,2}(L^p \times \mathbb{R}^d)$ and for any $\eta \in L^p$ and $x \in \mathbb{R}^d$, $D\Phi(\eta, x) \in W^{1,q}$, (q such that $\frac{1}{p} + \frac{1}{q} = 1$) and $\nabla_\theta D\Phi(\cdot, x) : L^p \to L^q$ is uniformly continuous, then F satisfies (3.17).

Proof. We have to show that if a function $F : [0, T] \times L^p \times \mathbb{R}^d \to \mathbb{R}$ satisfies the PIDE (3.17), then we have that the representation (3.18) holds. Let us assume X is the unique solution to equation (3.1), as in Section 2.4 we will use the notation ${}^{(\tau, \eta, x)}X$ to denote the process with initial time $\tau \in [0, T]$ and initial value $(\eta, x) \in M^p$.

If F satisfies equation (3.17) by Itô's formula (3.5) we have

$$\begin{aligned} & F(T, {}^{(\tau,\eta,x)} X_T, {}^{(\tau,\eta,x)} X(T)) - F(\tau, \eta, x) \\ & = \int_\tau^T \nabla_x F(s, {}^{(\tau,\eta,x)} X_s, {}^{(\tau,\eta,x)} X(s)) \cdot g(s, {}^{(\tau,\eta,x)} X_s, {}^{(\tau,\eta,x)} X(s)) dW(s) \\ & \quad + \int_\tau^T \int_{\mathbb{R}_0} \left(F(s, {}^{(\tau,\eta,x)} X_s, {}^{(\tau,\eta,x)} X(s) + h(s, {}^{(\tau,\eta,x)} X_s, {}^{(\tau,\eta,x)} X(s))(z) \tilde{N}(ds, dz) \right. \\ & \quad \left. - \int_\tau^T \int_{\mathbb{R}_0} F(s, {}^{(\tau,\eta,x)} X_s, {}^{(\tau,\eta,x)} X(s)) \right) \tilde{N}(ds, dz). \end{aligned} \tag{3.19}$$

Taking now the expectation, exploiting the fact that the right-hand side of equation (3.19) has null conditional expectation and using the terminal condition we have for any $0 \leqslant \tau < t \leqslant T$,

$$F(t, \eta, x) = E\left[\Phi(X_T, X(T)) | X_t = \eta, X(t) = x \right],$$

and the result is proved.

Conversely, let us now suppose that F is given by (3.18), then from the Markov property from Theorem 3.9 of the M^p-valued process and the tower rule for the conditional expectation, we have that for $0 \leqslant \tau < t \leqslant T$,

$$\begin{aligned} & E\left[F(t, X_t, X(t)) - F(\tau, X_\tau, X(\tau)) | X_\tau = \eta, X(\tau) = x \right] = \\ & = E\left[E\left[\Phi(X_T, X(T)) | X_t, X(t) \right] - E\left[\Phi(X_T, X(T)) | X_\tau, X(\tau) \right] | X_\tau, X(\tau) \right] = \\ & = E\left[\Phi(X_T, X(T)) | X_\tau, X(\tau) \right] - E\left[\Phi(X_T, X(T)) | X_\tau, X(\tau) \right] = 0. \end{aligned}$$

On the other side, the assumptions on Φ imply that F fulfils the hypothesis of Theorem 3.6. In particular, applying the Itô formula to Φ and taking the conditional expectation, we can see that F is time differentiable. Then we can apply the Itô's formula to the function F. Hence, for $0 \leqslant \tau < t \leqslant T$,

$$F(t, X_t, X(t)) = F(\tau, X_\tau, X(\tau)) + \int_\tau^t {}_q\langle DF(s, X_s, X(s)), dX_s \rangle_p + \int_\tau^t \partial_t F(s, X_s, X(s)) ds$$

$$+ \int_\tau^t \nabla_x F(s, X_s, X(s)) \cdot dX(s) + \frac{1}{2} \int_\tau^t Tr\left[g^*(s, X_s, X(s)) \nabla_x^2 F(s, X_s, X(s)) g(s, X_s, X(s)) \right] ds$$

$$+ \int_\tau^t \int_{\mathbb{R}_0} \left(F(s, X_s, X(s) + h(s, X_s, X(s))(z)) - F(s, X_s, X(s)) \right) \nu(dz) ds$$

$$- \int_\tau^t \int_{\mathbb{R}_0} \nabla_x F(s, X_s, X(s)) h(s, X_s, X(s))(z) \nu(dz) ds$$

$$- \int_\tau^t \int_{\mathbb{R}_0} \nabla_x F(s, X_s, X(s)) h(s, X_s, X(s))(z) \tilde{N}(ds, dz)$$

$$+ \int_\tau^t \int_{\mathbb{R}_0} \left(F(s, X_s, X(s) + h(s, X_s, X(s))(z)) - F(s, X_s, X(s)) \right) \tilde{N}(ds, dz).$$

Then taking conditional expectation the PPIDE (3.17) holds true. □

A   Appendix

A.1   Kunita's inequality

In Section 2.3, we introduced a general version of *Kunita's inequality*, (Corollary 2.12 in [34]). For $n = 1$, this is a rewritten version of Corrolary 2.12 in [34]). Below, we explain how to extend the result to general n.

Proof of Lemma 2.10. Notice that since norms on $\mathbb{R}^n$ are equivalent, it holds that

$$\sum_{j=1}^n |a_j|^q \leqslant C_0 \Big(\sum_{j=1}^n |a_j| \Big)^q, \text{ and}$$

$$\Big(\sum_{j=1}^n |a_j| \Big)^q \leqslant C_1 \sum_{j=1}^n |a_j|^q,$$

for some constants C_0, C_1 depending only on n and q. We may assume that $C_0 > 1$

$$\sum_{j=1}^n \|H^{\cdot j}(s)\|_{L^2(\nu_j, \mathbb{R}^d)}^q = \sum_{j=1}^n \Big(\int_{\mathbb{R}_0} |H^{\cdot j}(s, z)|^2 \nu_j(dz) \Big)^{\frac{q}{2}}$$

$$\leqslant C_0 \Big(\sum_{j=1}^n \int_{\mathbb{R}_0} |H^{\cdot j}(s, z)|^2 \nu_j(dz)^{\frac{1}{2}} \Big)^q = C_0 \|H(s)\|_{L^2(\nu, \mathbb{R}^d)}^q.$$

Then, if we write out "the columns wise" form of the $\tilde{N}$-integral, we obtain

$$\sup_{0 \leqslant u \leqslant t} \Big| \sum_{j=1}^n \int_0^u \int_{\mathbb{R}_0} H^{\cdot j}(s, z) \tilde{N}(ds, dz) \Big|^q \leqslant n^{q-1} \sup_{0 \leqslant u \leqslant t} \sum_{j=1}^n \Big| \int_0^u \int_{\mathbb{R}_0} H^{\cdot j}(s, z) \tilde{N}(ds, dz) \Big|^q$$

$$\leqslant n^{q-1} \sum_{j=1}^n \int_0^t \|H^{\cdot j}(s)\|_{L^q(\Omega, L^q(\nu_j, \mathbb{R}^d))} + \|H^{\cdot j}(s)\|_{L^q(\Omega, L^2(\nu_j, \mathbb{R}^d))} ds$$

$$= n^{q-1} \int_0^t E\Big[\sum_{j=1}^n \|H^{\cdot,j}(s)\|_{L^q(\nu_j, \mathbb{R}^d)}^q + \sum_{j=1}^n \|H^{\cdot,j}(s)\|_{L^2(\nu_j, \mathbb{R}^d)}^q \Big] ds$$

$$\leqslant n^{q-1} \int_0^t E\Big[\|H(s)\|_{L^q(\nu, \mathbb{R}^{d\times n})}^q + C_0 \|H(s)\|_{L^2(\nu, \mathbb{R}^{d\times n})}^q \Big] ds$$

$$\leqslant n^{q-1} C_0 \int_0^t \|H(s)\|_{L^q(\Omega, L^q(\nu, \mathbb{R}^{d\times n}))}^q + \|H(s)\|_{L^q(\Omega, L^2(\nu, \mathbb{R}^{d\times n}))}^q ds.$$

$\square$

In what follows we provide a connection between Itô's formula (3.6) and the path-dependent Itô's formula given in [12, 13] which relies on the concepts of vertical and horizontal derivative, there introduced. Let us first set the notation we use in the current section.

Let $(\Omega, \mathcal{F}, P)$ be the probability space with $\Omega = \mathcal{D}([0, T], \mathbb{R}^d)$ endowed with the P-augmented (right-continuous) filtration $\{\mathcal{F}_t\}_{t\in[0,T]}$ generated by the canonical process $Y : [0, T] \times \Omega \to \mathbb{R}^d$, $Y(t, \omega) = \omega(t)$ and here $\mathcal{F} := \mathcal{F}_T$. In this setting we define, for every $\omega \in \Omega$ and $t \in [0, T]$, $\omega_t := \{\omega(s), 0 \leqslant s \leqslant t\} \in \mathcal{D}([0, t])$, the trajectory up to time t. A stochastic process is a function $\varphi : [0, T] \times \Omega \to \mathbb{R}^d$, $(t, \omega) \mapsto \varphi(t, \omega)$. In addition, we say φ is non-anticipative if it is defined on $\mathcal{D}([0, t]; \mathbb{R}^d)$, i.e. $\varphi(t, \omega) = \varphi(t, \omega_t) := \varphi_t(\omega_t)$.

Let $\varphi = \{\varphi_t, t \in [0, T]\}$ be a non-anticipative stochastic process and $\{e_i\}_{i=1}^d \subset \mathbb{R}^d$ the canonical basis, we define the so-called *vertical derivative* as the following (path-wise) limit

$$\mathcal{D}^{V,i}\varphi_t(\omega_t) = \lim_{h\to 0} \frac{\varphi_t(\omega_t^{he_i}) - \varphi_t(\omega_t)}{h},$$

where $\omega_t^{he_i}(s) := \omega_t(s) + he_i 1_{\{t\}}(s)$, for every $s \in [0, t]$. Here, $\omega_t^{he_i}$ means adding a jump of size h at time t on the direction of e_i and hence the name. We then define the *vertical gradient* of φ_t as

$$\mathcal{D}^V \varphi_t = \left(\mathcal{D}^{V,1}\varphi_t, \ldots, \mathcal{D}^{V,d}\varphi_t \right).$$

Furthermore, we define the *horizontal derivative* as the following (path-wise) limit

$$\mathcal{D}^H \varphi_t(\omega_t) = \lim_{h\searrow 0} \frac{\varphi_{t+h}(\omega_{t,h}) - \varphi_t(\omega_t)}{h},$$

where $\omega_{t,h}(s) := \omega_t(s) 1_{[0,t]}(s) + \omega_t(t) 1_{(t,t+h]}(s)$, for every $s \in [0, t+h]$. Here, $\omega_{t,h}$ is the extension of the trajectory ω_t on $[0, t]$ to $[0, t+h]$ by an horizontal line of length h at $\omega_t(t)$ and hence the name.

We consider a functional $F : [0, T] \times \mathcal{D}([0, T]; \mathbb{R}^d) \to \mathbb{R}$ which will act on processes φ_t. We say F is non-anticipative if

$$F(t, \psi) = F(t, \psi_t) =: F_t(\psi_t),$$

for every non-anticipative stochastic process ψ_t. Next, we state an Itô formula for $F_t(\psi_t)$ where F_t is a non-anticipative functional which is once horizontally and twice vertically differentiable. This result is taken from [12, Proposition 6].

Theorem A.1 (Functional Itô's formula). *Consider an $\mathbb{R}^d$-valued non-anticipative stochastic process φ_t which admits the following càdlàg semimartingale representation*

$$\varphi_t = \varphi_0 + \int_0^t \mu(s) ds + \int_0^t \sigma(s) dW(s) + \int_0^t \int_{\mathbb{R}_0} \gamma(s-, z) \widetilde{N}(ds, dz)$$

for processes $\mu : [0, T] \to \mathbb{R}^d$, $\sigma : [0, T] \to \mathbb{R}^{d\times m}$ and $\gamma : [0, T] \times \mathbb{R}_0 \to \mathbb{R}^{d\times n}$ such that $\int_0^T E\left[|\mu(s)| + \|\sigma(s)\|^2 + \int_{\mathbb{R}_0} \|\gamma(s, z)\|^2 \nu(dz) \right] ds < \infty$ being $\|\cdot\|$ a matrix norm.

Let F be a once horizontally and twice vertically differentiable non-anticipative functional satisfying some technical continuity conditions on F (see [12, Proposition 6]), $\mathcal{D}^V F_t$, $\mathcal{D}^V \mathcal{D}^V F_t$ and $\mathcal{D}^H F_t$. Then for any t the following functional Itô formula holds P-a.s.

$$
\begin{aligned}
F_t(\varphi_t) = {} & F_0(\varphi_0) + \int_{(0,t]} \mathcal{D}^H F_s(\varphi_{s_-}) ds + \int_{(0,t]} \mathcal{D}^V F_s(\varphi_{s_-}) \, dX(s) \\
& + \int_{(0,t]} \frac{1}{2} Tr \left[\sigma^*(s) \mathcal{D}^V \mathcal{D}^V F_s(\varphi_{s_-}) \sigma(s) \right] ds \\
& + \int_{(0,t]} \int_{\mathbb{R}_0} \mathcal{D}^V F_s(\varphi_{s_-}) \left(F_s(\varphi_{s_-} + \gamma(s-,z)1_{\{s\}}) - F_s(\varphi_{s_-}) - \gamma(s-,z) \right) N(ds, dz) .
\end{aligned}
\tag{A.1}
$$

To be able to show that the path-dependent Itô's formula (A.1) and the Itô's formula from Theorem 3.6 do coincide we need first to connect the two settings. Such a link can be established following [25], where the following operators are considered:

- the *restriction* operator, for every $t \in [0, r]$

$$
\begin{aligned}
& \mathbf{R}_t : \mathcal{D}([-r, 0], \mathbb{R}^d) \to \mathcal{D}([0, t], \mathbb{R}^d) , \\
& \mathbf{R}_t(f)(s) = f(s - t) , \quad s \in [0, t) ,
\end{aligned}
$$

- the *backward extension* operator, for every $t \in (0, r)$

$$
\begin{aligned}
& \mathbf{B}_t : \mathcal{D}([0, t], \mathbb{R}^d) \to \mathcal{D}([-r, 0], \mathbb{R}^d) , \\
& \mathbf{B}_t(f)(s) = f(0) \mathbb{1}_{[-r,-t)}(s) + f(t + s) \mathbb{1}_{[-t,0)}(s) , \quad s \in [-r, 0) ,
\end{aligned}
$$

Let us consider a non-anticipative functional $b : [0, T] \times \mathcal{D}([0, T]; \mathbb{R}^d) \to \mathbb{R}$, $b(t, \psi) = b(t, \psi_t) =: b_t(\psi_t)$ for any non-anticipative stochastic process ψ, then one can define a different functional $\widehat{b}$ on $[0, T] \times \mathcal{D}([-r, 0]; \mathbb{R}^d) \times \mathbb{R}^d$ as

$$
\widehat{b}(t, X_t, X(t)) := b_t \left(\tilde{\mathbf{R}}_t X_t \right) , \quad (X_t, X(t)) \in \mathcal{D}([-r, 0]; \mathbb{R}^d) \times \mathbb{R}^d ,
$$

with

$$
\tilde{\mathbf{R}}_t X_t(s) := \begin{cases} \mathbf{R}_t(X_t)(s) & \text{if } s \in [0, t) \\ X_t(s) & \text{if } s = t \end{cases} .
$$

The converse holds true as well, in fact let us consider a given functional $\widehat{b}$ on $[0, T] \times \mathcal{D}([-r, 0]; \mathbb{R}^d)$, then we can obtain a corresponding functional b_t on $\mathcal{D}([0, t]; \mathbb{R}^d)$ as

$$
b_t(\varphi_t) := \widehat{b}(t, \mathbf{B}_t \varphi_t, \varphi_t(t)) , \quad \varphi_t \in \mathcal{D}([0, t]; \mathbb{R}^d) ,
\tag{A.2}
$$

see [25] for details.

We can now show how the vertical and horizontal derivatives can be written in terms of the Fréchet derivative D and the derivative with respect to the present state. Part of the next theorem was already established in [25, Theorem 7.1].

Proposition A.2. *Consider a function $F : [0, T] \times \mathcal{D}([-r, 0]; \mathbb{R}^d) \to \mathbb{R}$ and let us define $u_t : \mathcal{D}([0, t]; \mathbb{R}^d) \to \mathbb{R}$ as above in (A.2) $u_t(X_t) := F(t, \mathbf{B}_t X_t, X(t))$. Then the i-th vertical derivative $\mathcal{D}^{V,i}$ of u_t coincides with the derivative with respect to the present state $X^i(t)$ of F, namely*

$$
\mathcal{D}^{V,i} u_t(X_t) = \partial_{x_i} F(t, \mathbf{B}_t X_t, X(t)) .
\tag{A.3}
$$

Furthermore, we have

$$
u_t(X_t^{h^i}) - u_t(X_t) = F(t, \mathbf{B}_t X_t, X(t) + h^i(t, \mathbf{B}_t X_t, X(t)) - F(t, \mathbf{B}_t X_t, X(t)) .
\tag{A.4}
$$

If we assume that $X_t \in W^{1,p}$, then

$$
\mathcal{D}^H u_t(X_t) = \partial_t F(t, \mathbf{B}_t X_t, X(t)) + \left\langle DF(t, \mathbf{B}_t X_t, X(t)), \nabla_\theta^+ \mathbf{B}_t X_t \right\rangle_{\mathcal{D}} ,
\tag{A.5}
$$

holds, where the notation is given in Section 3.

33

Proof. Concerning (A.3) we have

$$\mathcal{D}^{V,i}u_t(X_t) = \lim_{h\to 0}\frac{1}{h}\left(u_t(X_t^h) - u_t(X_t)\right) = \lim_{h\to 0}\frac{1}{h}\left(F(t,\mathbf{B}_tX_t^h, X^h(t)) - F(t,\mathbf{B}_tX_t, X(t))\right)$$
$$= \lim_{h\to 0}\frac{1}{h}\left(F(t, X(t) + h, \mathbf{B}_tX_t^h) - F(t, X(t), \mathbf{B}_tX_t)\right) = \partial_i F(t, X(t), \mathbf{B}_tX_t)\,. \tag{A.6}$$

For what concerns (A.4), proceeding as in (A.6), we immediately have

$$u_t(X_t^{h^i}) - u_t(X_t) = F(t, \mathbf{B}_tX_t^{h^i}, X^{h^i}(t)) - F(t, \mathbf{B}_tX_t, X(t)) =$$
$$= F(t, X(t) + h^i, \mathbf{B}_tX_t^{h^i}) - F(t, X(t), \mathbf{B}_tX_t)\,.$$

We refer to [25, Theorem 7.1] for a proof of equation (A.5) $\qquad\qquad\square$

In the framework of this section, exploiting the previous proposition we have that, for suitable regular coefficients, Itô's formula from Theorem 3.6 and the path-dependent Itô's formula in Theorem A.1 coincide. In particular let us consider a process X evolving according to

$$\begin{cases} dX_t = f(t, X_t)dt + g(t, X_t)dW(t) + \int_{\mathbb{R}_0} h(t, X_t, z)\tilde{N}(dt, dz)\,, \\ X_0 = \eta\,, \end{cases} \tag{A.7}$$

for some suitably regular enough coefficients f, g and h. Then proceeding as above we have that Equation (A.7) can be written as a path dependent process

$$\begin{cases} dX_t = \hat{f}_t(X_t)dt + \hat{g}_t(X_t)dW(t) + \int_{\mathbb{R}_0} \hat{h}_t(X_t)\tilde{N}(dt, dz)\,, \\ X_0 = \eta\,, \end{cases}$$

with $\hat{f}_t$, $\hat{g}_t$ and $\hat{h}_t$ defined as in (A.2). Then we have the following result.

Theorem A.3. *Let* $F : [0,T] \times M^p \to \mathbb{R}$, $F \in C^{1,1,2}([0,T] \times \mathcal{D} \times \mathbb{R}^d)$ *and let us define* $u_t : \mathcal{D}([0,t];\mathbb{R}^d) \to \mathbb{R}$ *as in* (A.2) $u_t(X_t) := F(t, \mathbf{B}_tX_t, X(t))$. *Then Itô's formula from Theorem 3.6 and the path dependent Itô's formula from Theorem A.1 coincide.*

Proof. It is straightforward from Proposition A.2 exploiting the backward extension operator $\mathbf{B}_t$ and eventually using Itô's formula from Theorem 3.6 and the *path-dependent* Itô formula from Theorem A.1. $\qquad\qquad\square$

References

[1] N. Agram and E. E. Røse. Optimal control of forward–backward mean-field stochastic delayed systems. *Afrika Matematika*, 29(1-2):149–174, 2018.

[2] D. Applebaum. *Lévy Processes and Stochastic Calculus*, volume 116 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, second edition, 2009.

[3] M. Arriojas, Y. Hu, S.-E. Mohammed, and G. Pap. A delayed Black and Scholes formula. *Stoch. Anal. Appl.*, 25(2):471–492, 2007.

[4] S. Asmussen and J. Rosiński. Approximations of small jumps of Lévy processes with a view towards simulation. *J. Appl. Probab.*, 38(2):482–493, 2001.

[5] E. Bandini and F. Russo. Weak dirichlet processes with jumps. *Stochastic Processes and their Applications*, 127(12):4139–4189, 2017.

[6] F. E. Benth, G. Di Nunno, and A. Khedher. Robustness of option prices and their deltas in markets modelled by jump-diffusions. *Commun. Stoch. Anal.*, 5(2):285–307, 2011.

[7] P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons Inc., New York, 1968.

[8] M.-H. Chang and R. K. Youree. The European option with hereditary price structures: basic theory. *Appl. Math. Comput.*, 102(2-3):279–296, 1999.

[9] M.-H. Chang and R. K. Youree. Infinite-dimensional Black-Scholes equation with hereditary structure. *Appl. Math. Optim.*, 56(3):395–424, 2007.

[10] A. Chojnowska-Michalik. Representation theorem for general stochastic delay equations. *Acad. Pol. Sci., Ser. Sci. Math. Astron. Phys*, 26:635–642, 1978.

[11] S. Cohen and J. Rosiński. Gaussian approximation of multivariate Lévy processes with applications to simulation of tempered stable processes. *Bernoulli*, 13(1):195–210, 2007.

[12] R. Cont and D.-A. Fournié. Change of variable formulas for non-anticipative functionals on path space. *Journal of Functional Analysis*, 259(4):1043–1072, 2010.

[13] R. Cont and D.-A. Fournié. Functional Itô calculus and stochastic integral representation of martingales. *The Annals of Probability*, 41(1):109–133, 2013.

[14] A. Cosso, C. Di Girolami, and F. Russo. Calculus via regularizations in Banach spaces and Kolmogorov-type path-dependent equations. In G. Budzban, H. R. Hughes, and H. Schurz, editors, *Probability on Algebraic and Geometric Structures*, volume 668. American Mathematical Soc., 2016.

[15] A. Cosso and F. Russo. A regularization approach to functional Itô calculus and strong-viscosity solutions to path-dependent PDEs. *arXiv:1401.5034*.

[16] A. Cosso and F. Russo. Functional itô versus banach space stochastic calculus and strict solutions of semilinear path-dependent equations. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 19(04):1650024, 2016.

[17] R. Coviello, C. Di Girolami, and F. Russo. On stochastic calculus related to financial assets without semimartingales. *Bulletin des Sciences mathématiques*, 135(6):733–774, 2011.

[18] K. Dahl, S.-E. Mohammed, B. Øksendal, and E. E. Røse. Optimal control of systems with noisy memory and bsdes with malliavin derivatives. *Journal of Functional Analysis*, 271(2):289–329, 2016.

[19] C. Di Girolami and F. Russo. Generalized covariation and extended Fukushima decomposition for Banach space-valued processes: applications to windows of Dirichlet processes. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 15(02), 2012.

[20] C. Di Girolami and F. Russo. Generalized covariation for Banach space valued processes, Itô formula and applications. *Osaka J. Math.*, 51(3):729–783, 2014.

[21] O. Diekmann, S. A. Van Gils, S. V. Lunel, and H.-O. Walther. *Delay equations: functional-, complex-, and nonlinear analysis*. Springer-Verlag, 1995.

[22] N. Dunford and J. T. Schwartz. *Linear operators. Part I.* Wiley Classics Library. John Wiley & Sons, Inc., New York, 1988. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication.

[23] B. Dupire. Functional Itô calculus. *Bloomberg Portfolio Research paper*, 2009.

[24] K.-J. Engel and R. Nagel. *One-parameter semigroups for linear evolution equations*, volume 194. Springer Science & Business Media, 2000.

[25] F. Flandoli, G. Zanco, et al. An infinite-dimensional approach to path-dependent kolmogorov equations. *The Annals of Probability*, 44(4):2643–2693, 2016.

[26] M. Fuhrman, F. Masiero, and G. Tessitore. Stochastic equations with delay: optimal control via BSDEs and regular solutions of Hamilton-Jacobi-Bellman equations. *SIAM Journal on Control and Optimization*, 48(7):4624–4651, 2010.

[27] S. Janson and S. Kaijser. Higher moments of Banach space valued random variables. *Mem. Amer. Math. Soc.*, 238(1127):vii+110, 2015.

[28] Y. Kazmerchuk, A. Swishchuk, and J. Wu. A continuous-time Garch model for stochastic volatility with delay. *Can. Appl. Math. Q.*, 13(2):123–149, 2005.

[29] Y. Kazmerchuk, A. Swishchuk, and J. Wu. The pricing of options for securities markets with delayed response. *Math. Comput. Simulation*, 75(3-4):69–79, 2007.

[30] A. Khedher. Computation of the delta in multidimensional jump-diffusion setting with applications to stochastic volatility models. *Stoch. Anal. Appl.*, 30(3):403–425, 2012.

[31] J. Kiessling and R. Tempone. Diffusion approximation of Lévy processes with a view towards finance. *Monte Carlo Methods Appl.*, 17(1):11–45, 2011.

[32] Y. Kuang. *Delay differential equations with applications in population dynamics*, volume 191 of *Mathematics in Science and Engineering*. Academic Press, Inc., Boston, MA, 1993.

[33] U. Kuchler and E. Platen. *Time delay and noise explaining cyclical fluctuations in prices of commodities*. University of Technology, Sydney, 2007.

[34] H. Kunita. Stochastic differential equations based on Lévy processes and stochastic flows of diffeomorphisms. In *Real and stochastic analysis*, Trends Math., pages 305–373. Birkhäuser Boston, Boston, MA, 2004.

[35] S. E. A. Mohammed. *Stochastic Functional Differential Equations*, volume 99 of *Research Notes in Mathematics*. Pitman (Advanced Publishing Program), Boston, MA, 1984.

[36] S.-E. A. Mohammed. Stochastic differential systems with memory: theory, examples and applications. In *Stochastic analysis and related topics, VI (Geilo, 1996)*, volume 42 of *Progr. Probab.*, pages 1–77. Birkhäuser Boston, Boston, MA, 1998.

[37] K. R. Parthasarathy. *Probability measures on metric spaces*. Probability and Mathematical Statistics, No. 3. Academic Press, Inc., New York-London, 1967.

[38] P. E. Protter. *Stochastic Integration and Differential Equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2005.

[39] M. Reiß. *Nonparametric estimation for stochastic delay differential equations*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, 2002.

[40] M. Reiß, M. Riedle, and O. van Gaans. Delay differential equations driven by Lévy processes: stationarity and Feller properties. *Stochastic Process. Appl.*, 116(10):1409–1432, 2006.

[41] F. Russo and P. Vallois. The generalized covariation process and Itô formula. *Stochastic Process. Appl.*, 59(1):81–104, 1995.

[42] F. Russo and P. Vallois. Itô formula for C^1-functions of semimartingales. *Probab. Theory Related Fields*, 104(1):27–41, 1996.

[43] A. Swishchuk. *Modeling and Pricing of Swaps for Financial and Energy Markets with Stochastic Volatilities.* World Scientific. World Scientific Publishing Co. Pte. Ltd. , Singapore., 2013.

[44] F. Yan and S. Mohammed. A stochastic calculus for systems with memory. *Stoch. Anal. Appl.*, 23(3):613–657, 2005.

[45] E. Zeidler. *Applied Functional Analysis. Main Principles and Their Applications.* Springer-Verlag 1995. New York, third edition, 1995.