CIRJE-F-993

Asymptotic Expansion for Forward-Backward SDEs with Jumps

Masaaki Fujii
University of Tokyo

Akihiko Takahashi
University of Tokyo

September 2015

Asymptotic Expansion for Forward-Backward SDEs with Jumps *

Masaaki Fujii[†] & Akihiko Takahashi[‡]

This version: 18 December, 2015

Abstract

The paper develops an asymptotic expansion method for forward-backward SDEs (FBSDEs) driven by the random Poisson measures with σ-finite compensators. The expansion is performed around the small-variance limit of the forward SDE and does not necessarily require a small size of the non-linearity in the BSDE's driver, which was actually the case for the linearization method proposed by the current authors in a Brownian setup before. A semi-analytic solution technique, which only requires a system of ODEs (one is non-linear and the others are linear) to be solved, as well as its error estimate are provided. In the case of a finite jump measure with a bounded intensity, the method can also handle sate-dependent (and hence non-Poissonian) jumps, which are quite relevant for many practical applications. Based on the stability result, we also provide a rigorous justification to use arbitrarily smooth coefficients in FBSDEs for any approximation purpose whenever rather mild conditions are satisfied.

Keywords : BSDE, jumps, random measure, asymptotic expansion, Lévy process

1 Introduction

Since it was introduced by Bismut (1973) [4] and Pardoux & Peng (1990) [36], the backward stochastic differential equations (BSDEs) have attracted many researchers. There now exist excellent mathematical reviews, such as El Karoui & Mazliak (eds.) (1997) [15], Ma & Yong (2000) [33], and Pardoux & Rascanu (2014) [38] for interested readers.

In recent years, there also appeared various applications of BSDEs to financial as well as operational problems. One can see, for example, El Karoui et al. (1997) [16], Lim (2004) [31], Jeanblanc & Hamadène (2007) [24], Cvitanić & Zhang (2013) [9], Delong (2013) [10], Touzi (2013) [46], Crépey, Bielecki & Brigo (2014) [7] and references therein. In particular, due to the financial crisis in 2008 and a bunch of new regulations that followed, various problems involving non-linearity, such as credit/funding risks, risk measures and optimal executions in illiquid markets, have arisen as central issues in the financial industry. In those practical applications, one needs concrete numerical methods which can efficiently evaluate the BSDEs.

Although Monte-Carlo simulation techniques based on the least-square regression method have been proposed and studied by many researchers, mostly in diffusion setups, (See, for

*All the contents expressed in this research are solely those of the author and do not represent any views or opinions of any institutions. The author is not responsible or liable in any manner for any losses and/or damages caused by the use of any contents in this research.

†Quantitative Finance Course, Graduate School of Economics, The University of Tokyo.

‡Quantitative Finance Course, Graduate School of Economics, The University of Tokyo.

example, Bouchard & Touzi (2004) [6], Zhang (2004) [47], Gobet et al. (2004) [23], and Bender & Denk (2007) [2].), they have not yet become the standard among practitioners due to their computational burden when applied to a big portfolio. In fact, as emphasized in Crépey, Bielecki & Brigo (2014) [7] and Crépey & Song (2015) [8], the existing regression scheme cannot be applied to realistic financial problems but only to those with rather short maturities and low dimensional underlying processes [1]. The presence of jumps, whether infrequent or infinitely active, inevitably makes the regression more time consuming. Furthermore, in certain applications such as mean-variance hedging and multiple dependent defaults, the solution of one BSDE appears in the driver of another BSDE [2]. In such a case, deriving an analytic approximation for the first BSDE seems to be the only possibility to deal with the problem in a feasible manner. From the above observation, it is clear that a simple analytic approximation method is deeply wanted.

In the current work, we develop an asymptotic expansion method for (decoupled) forward-backward SDEs driven by the Poisson random measures in addition to the standard Brownian motions. We propose an expansion around a small-variance limit of the forward SDE. The proposed scheme starts from solving a non-linear ODE which corresponds to the BSDE in which every forward component is replaced by its deterministic mean process. Every higher order approximation yields linear forward-backward SDEs which can be solved by a system of linear ODEs just like a simple affine model. The approximate solution of the BSDE including the martingale components is explicitly given as a function of the stochastic flows of the underlying forward process with non-random coefficients determined by these ODEs. Thus, one can obtain not only the current value of the solution but also its evolution by simply simulating the flows of the underlying process. In an optimal hedging problem for example, one can study the evolution of the value function as well as the associated hedging position along each path and hence their distributional information, too.

In order to justify the approximation method and to obtain its error estimate, we use the recent results of Kruse & Popier (2015) [29] regarding a priori estimates and the existence of unique $\mathbb{L}^p$-solution of a BSDE with jumps, the representation theorem based on the Malliavin's derivative for a BSDE with jumps by Delong & Imkeller (2010) [11] and Delong [10], as well as the idea of Pardoux & Peng (1992) [37] and Ma & Zhang (2002) [34] that controls the sup-norm of the martingale integrands of the BSDE. In addition to the system driven by the random Poisson measures, we also justify the expansion of a system with a state-dependent jump intensity when it is bounded. The current work also serves as a justification of a polynomial expansion method proposed in Fujii (2015) [18], at least, for a certain class of models. As a particular example, a simple closed-form expansion is presented when the underlying forward SDE belongs to (time-inhomogeneous) exponential Lévy type. Under rather mild conditions, we also provide a rigorous justification to use arbitrarily smooth coefficients in the forward-backward SDEs for any *approximation purpose*, which releases us from being bothered about the technicalities.

As for forward SDEs, the asymptotic expansion method around a small-variance limit has already been popular and been applied to a variety of real problems. It has been shown, in various numerical examples, that the first few terms of expansion are enough to achieve accurate approximation in option pricing with typical volatilities ranging from 10% to 20% and maturities up to a few years. See a recent review Takahashi (2015) [42]

[1]In Crépey & Song [8], the authors successfully applied the asymptotic expansion method we proposed in [20, 21] to a collateralized debt obligation (CDO) with 120 underlying names to evaluate credit/funding valuation adjustments.

[2]See Mania & Tevzadze (2003) [35], Pham (2010) [40] and Fujii (2015) [19] for concrete examples.

for the details and a comprehensive list of literature. For longer maturities (and hence with larger variances), one typically needs higher-order approximation. Of course, if the underlying processes are more volatile, then the expansion can even fail to converge. Here, it is important to note that the small-variance approximation can be rewritten in terms of the short-term approximation as implied by Takahashi & Yamada (2015) [43]. This naturally leads to a sub-stepping scheme of asymptotic expansion studied by Fujii (2014) [17] and Takahashi & Yamada (2015) [44], which can handle higher volatilities and longer maturities. Generalizing the proposed method to incorporate sub-stepping scheme is an important future research topic. This is expected to be done naturally by resetting the boundary condition for ODEs at each time step backwardly. Another direction of research is to establish an efficient Monte Carlo simulation scheme, in which the numerically costly regression is replaced by the semi-analytic result obtained by the current work.

The organization of the paper is as follows: Section 2 gives some preliminaries, Section 3 explains the setup of the forward-backward SDEs and their existence. Section 4 gives the representation theorem based on Malliavin's derivative, and Section 5 and 6 deal with the classical differentiability and the error estimate of the asymptotic expansion. Section 7 discusses the state-dependent intensity and Section 8 explains the implementation of the asymptotic expansion. Section 9 treats a special case of a linear forward SDE and the associated polynomial expansion. Several numerical examples are available in [18] for a certain class of models. Section 10 provides a justification to use smooth coefficients, and finally Appendix summarizes the relevant a priori estimates used in the main text.

2 Preliminaries

2.1 General Setting

$T > 0$ is some bounded time horizon. The space $(\Omega_W, \mathcal{F}_W, \mathbb{P}_W)$ is the usual canonical space for a l-dimensional Brownian motion equipped with the Wiener measure $\mathbb{P}_W$. We also denote $(\Omega_\mu, \mathcal{F}_\mu, \mathbb{P}_\mu)$ as a product of canonical spaces $\Omega_\mu := \Omega_\mu^1 \times \cdots \times \Omega_\mu^k$, $\mathcal{F}_\mu := \mathcal{F}_\mu^1 \times \cdots \times \mathcal{F}_\mu^k$ and $\mathbb{P}_\mu^1 \times \cdots \times \mathbb{P}_\mu^k$ with some constant $k \geq 1$, on which each μ^i is a Poisson measure with a compensator $\nu^i(dz)dt$. Here, $\nu^i(dz)$ is a σ-finite measure on $\mathbb{R}_0 = \mathbb{R}\backslash\{0\}$ satisfying $\int_{\mathbb{R}_0} |z|^2 \nu^i(dz) < \infty$. Throughout the paper, we work on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, where the space $(\Omega, \mathcal{F}, \mathbb{P})$ is the product of the canonical spaces $(\Omega_W \times \Omega_\mu, \mathcal{F}_W \times \mathcal{F}_\mu, \mathbb{P}_W \times \mathbb{P}_\mu)$, and that the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ is the canonical filtration completed for $\mathbb{P}$ and satisfying the usual conditions. In this construction, $(W, \mu^1, \cdots, \mu^k)$ are independent. We use a vector notation $\mu(\omega, dt, dz) := (\mu^1(\omega, dt, dz^1), \cdots, \mu^k(\omega, dt, dz^k))$ and denote the compensated Poisson measure as $\widetilde{\mu} := \mu - \nu$. We represent the $\mathbb{F}$-predictable σ-field on $\Omega \times [0, T]$ by $\mathcal{P}$.

2.2 Notation

We denote a generic constant by C_p, which may change line by line, depending on p, T and the Lipschitz constants and the bounds of the relevant functions. Let us introduce a sup-norm for a $\mathbb{R}^r$-valued function $x : [0, T] \to \mathbb{R}^r$ as

$$||x||_{[a,b]} := \sup\{|x_t|, t \in [a, b]\}$$

and write $||x||_t := ||x||_{[0,t]}$. We also use the following spaces for stochastic processes for $p \geq 2$:

• $\mathbb{S}^p_r[s,t]$ is the set of $\mathbb{R}^r$-valued adapted càdlàg processes X such that

$$||X||_{\mathbb{S}^p_r[s,t]} := \mathbb{E}\left[||X(\omega)||^p_{[s,t]}\right]^{1/p} < \infty \ .$$

• $\mathbb{H}^p_r[s,t]$ is the set of progressively measurable $\mathbb{R}^r$-valued processes Z such that

$$||Z||_{\mathbb{H}^p_r[s,t]} := \mathbb{E}\left[\Big(\int_s^t |Z_u|^2 du\Big)^{p/2}\right]^{1/p} < \infty.$$

• $\mathbb{H}^p_{r,\nu}[s,t]$ is the set of functions $\psi = \{(\psi)_{i,j}, 1 \le i \le r, 1 \le j \le k\}$, $(\psi)_{i,j} : \Omega \times [0,T] \times \mathbb{R}_0 \to \mathbb{R}$ which are $\mathcal{P} \times \mathcal{B}(\mathbb{R}_0)$-measurable and satisfy

$$||\psi||_{\mathbb{H}^p_{r,\nu}[s,t]} := \mathbb{E}\left[\Big(\sum_{i=1}^k \int_s^t \int_{\mathbb{R}_0} |\psi_u^{\cdot,i}(z)|^2 \nu^i(dz) du\Big)^{p/2}\right]^{1/p} < \infty.$$

For notational simplicity, we use $(E,\mathcal{E}) = (\mathbb{R}_0^k, \mathcal{B}(\mathbb{R}_0)^k)$ and denote the above maps $\{(\psi)_{i,j}, 1 \le i \le r, 1 \le j \le k\}$ as $\psi : \Omega \times [0,T] \times E \to \mathbb{R}^{r \times k}$ and say ψ is $\mathcal{P} \times \mathcal{E}$-measurable without referring to each component. We also use the notation such that

$$\int_s^t \int_E \psi_u(z) \widetilde{\mu}(du,dz) := \sum_{i=1}^k \int_s^t \int_{\mathbb{R}_0} \psi_u^i(z) \widetilde{\mu}^i(du,dz)$$

for simplicity. The similar abbreviation is used also for the integral with μ and ν. When we use E and $\mathcal{E}$, one should always interpret it in this way so that the integral with the k-dimensional Poisson measure does make sense. On the other hand, when we use the range $\mathbb{R}_0$ with the integrators $(\widetilde{\mu}, \mu, \nu)$, for example,

$$\int_{\mathbb{R}_0} \psi_u(z) \nu(dz) := \Big(\int_{\mathbb{R}_0} \psi_u^i(z) \nu^i(dz)\Big)_{1 \le i \le k}$$

we interpret it as a k-dimensional vector.
• $\mathcal{K}^p[s,t]$ is the set of functions (Y,Z,ψ) in the space $\mathbb{S}^p[s,t] \times \mathbb{H}^p[s,t] \times \mathbb{H}^p_\nu[s,t]$ with the norm defined by

$$||(Y,Z,\psi)||_{\mathcal{K}^p[s,t]} := \Big(||Y||^p_{\mathbb{S}^p[s,t]} + ||Z||^p_{\mathbb{H}^p[s,t]} + ||\psi||^p_{\mathbb{H}^p_\nu[s,t]}\Big)^{1/p}.$$

• $\mathbb{L}^2(E,\mathcal{E},\nu : \mathbb{R}^r)$ is the set of $\mathbb{R}^{r \times k}$-valued $\mathcal{E}$-measurable functions U satisfying

$$\begin{aligned} ||U||_{\mathbb{L}^2(E)} &:= \Big(\int_E |U(z)|^2 \nu(dz)\Big)^{1/2} \\ &:= \Big(\sum_{i=1}^k \int_{\mathbb{R}_0} |U^{\cdot,i}(z)|^2 \nu^i(dz)\Big)^{1/2} < \infty \ . \end{aligned}$$

We frequently omit the subscripts for its dimension r and the time interval $[s,t]$ when those are obvious in the context.

We use the notation of partial derivatives such that

$$\partial_\epsilon = \frac{\partial}{\partial \epsilon}, \quad \partial_x = (\partial_{x_1}, \cdots, \partial_{x_d}) = \Big(\frac{\partial}{\partial x_1}, \cdots, \frac{\partial}{\partial x_d}\Big)$$
$$\partial_x^2 = \partial_{x,x} = \Big(\frac{\partial^2}{\partial x_i \partial x_j}\Big)_{i,j=\{1,\cdots,d\}}$$

and similarly for every higher order derivative without a detailed indexing. We suppress the obvious summation of indexes throughout the paper for notational simplicity.

3 Forward and Backward SDEs

We work in the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ defined in the last section. Firstly, let us introduce the d-dimensional forward SDE of $(X_s^{t,x,\epsilon}, s \in [t,T])$ with the initial data $(t,x) \in [0,T] \times \mathbb{R}^d$ and a small constant parameter $\epsilon \in [0,1]$;

$$X_s^{t,x,\epsilon} = x + \int_t^s b(r, X_r^{t,x,\epsilon}, \epsilon) dr + \int_t^s \sigma(r, X_r^{t,x,\epsilon}, \epsilon) dW_r + \int_t^s \int_E \gamma(r, X_{r-}^{t,x,\epsilon}, z, \epsilon) \widetilde{\mu}(dr, dz) \tag{3.1}$$

where $b : [0,T] \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$, $\sigma : [0,T] \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d \times l}$ and $\gamma : [0,T] \times \mathbb{R}^d \times E \times \mathbb{R} \to \mathbb{R}^{d \times k}$. Let us also introduce the function $\eta : \mathbb{R} \to \mathbb{R}$ by $\eta(z) = 1 \wedge |z|$. Now, we make the following assumptions:

Assumption 3.1. *The functions $b(t,x,\epsilon), \sigma(t,x,\epsilon)$ and $\gamma(t,x,z,\epsilon)$ are continuous in all their arguments and continuously differentiable arbitrary many times with respect to (x,ϵ). Furthermore, there exists some positive constant K such that*
(i) for every $m \geq 0$, $|\partial_\epsilon^m b(t,0,\epsilon)| + |\partial_\epsilon^m \sigma(t,0,\epsilon)| \leq K$ uniformly in $(t,\epsilon) \in [0,T] \times [0,1]$,
(ii) for every $n \geq 1, m \geq 0$, $|\partial_x^n \partial_\epsilon^m b(t,x,\epsilon)| + |\partial_x^n \partial_\epsilon^m \sigma(t,x,\epsilon)| \leq K$ uniformly in $(t,x,\epsilon) \in [0,T] \times \mathbb{R}^d \times [0,1]$,
(iii) for every $m \geq 0$ and column $1 \leq i \leq k$, $|\partial_\epsilon^m \gamma_{\cdot,i}(t,0,z,\epsilon)|/\eta(z) \leq K$ uniformly in $(t,z,\epsilon) \in [0,T] \times \mathbb{R}_0 \times [0,1]$,
(iv) for every $n \geq 1, m \geq 0$ and column $1 \leq i \leq k$, $|\partial_x^n \partial_\epsilon^m \gamma_{\cdot,i}(t,x,z,\epsilon)|/\eta(z) \leq K$ uniformly in $(t,x,z,\epsilon) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}_0 \times [0,1]$.

We define $(\partial_x X_s^{t,x,\epsilon}, s \in [t,T])$ as the solution of the SDE (if exists) given by a formal differentiation:

$$\partial_x X_s^{t,x,\epsilon} = \int_t^s \partial_x b(r, X_r^{t,x,\epsilon}, \epsilon) \partial_x X_r^{t,x,\epsilon} dr + \int_t^s \partial_x \sigma(r, X_r^{t,x,\epsilon}, \epsilon) \partial_x X_r^{t,x,\epsilon} dW_r$$
$$+ \int_t^s \int_E \partial_x \gamma(r, X_r^{t,x,\epsilon}, z, \epsilon) \partial_x X_r^{t,x,\epsilon} \widetilde{\mu}(dr, dz) \tag{3.2}$$

and similarly for $(\partial_\epsilon X_s^{t,x,\epsilon}, s \in [t,T])$ and every higher order flow $(\partial_x^n \partial_\epsilon^m X_s^{t,x,\epsilon}, s \in [t,T])_{m,n \geq 0}$.

Proposition 3.1. *Under Assumption 3.1, the SDE (3.1) has a unique solution $X^{t,x,\epsilon} \in \mathbb{S}_d^p[t,T]$ for $\forall p \geq 2$. Furthermore, every (n,m)-times classical differentiation of $X^{t,x,\epsilon}$ with respect to (x,ϵ) is well defined and given by $(\partial_x^n \partial_\epsilon^m X_s^{t,x,\epsilon}, s \in [t,T])$, which is a unique solution of the corresponding SDE defined by the formal differentiation of the coefficients as (3.2) and belongs to $\mathbb{S}_{d^{n+1}}^p[t,T]$ for $\forall p \geq 2$.*

Proof. The existence of a unique solution $X^{t,x,\epsilon} \in \mathbb{S}^p_d[t,T]$ for $\forall p \geq 2$ is standard and can easily be proved by Lemma A.3. Since every SDE is linear, it is not difficult to recursively show that the same conclusion holds for every $\partial^n_x \partial^m_\epsilon X^{t,x,\epsilon}$. The agreement with the classical differentiation can be proved by following the same arguments in Theorem 3.1 of Ma & Zhang (2002) [34]. In particular, one can show

$$\lim_{h\to 0} \mathbb{E}||\nabla X^h - \partial_x X^{t,x,\epsilon}||^2_{[t,T]} = 0$$

where $\nabla X^h_s := \dfrac{X^{t,x+h,\epsilon}_s - X^{t,x,\epsilon}_s}{h}$, and similar relations for every higher order derivatives with respect to (x,ϵ). □

Let us now introduce the BSDE which depends on $X^{t,x,\epsilon}$ given by (3.1):

$$\begin{aligned} Y^{t,x,\epsilon}_s &= \xi(X^{t,x,\epsilon}_T) + \int_s^T f\Big(r, X^{t,x,\epsilon}_r, Y^{t,x,\epsilon}_r, Z^{t,x,\epsilon}_r, \int_{\mathbb{R}_0} \rho(z)\psi^{t,x,\epsilon}_r(z)\nu(dz)\Big)dr \\ &\quad - \int_s^T Z^{t,x,\epsilon}_r dW_r - \int_s^T \int_E \psi^{t,x,\epsilon}_r(z)\widetilde{\mu}(dr,dz), \end{aligned} \tag{3.3}$$

for $s \in [t,T]$ where $\xi : \mathbb{R}^d \to \mathbb{R}^m$, $f : [0,T] \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m\times l} \times \mathbb{R}^{m\times k} \to \mathbb{R}^m$ and $\rho : E \to \mathbb{R}^k$. We make the following assumptions:

Assumption 3.2. *There exist some positive constant $K, q \geq 0$ such that*
(i) $\xi(x)$ is continuously differentiable arbitrary many times with respect to x and satisfies $|\xi(x)| \leq K(1+|x|^q)$ for every $x \in \mathbb{R}^d$,
(ii) $|\rho_i(z)| \leq K\eta(z)$ for every $1 \leq i \leq k$ and $z \in \mathbb{R}_0$,
(iii) $f(t,x,y,z,u)$ is continuous in (t,x,y,z,u) and continuously differentiable arbitrary many times with respect to (x,y,z,u). All the partial differentials except those regarding only on x, i.e. $(\partial^n_x f(t,x,y,z,u), n \geq 1)$, are bounded by K uniformly in $(t,x,y,z,u) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m\times l} \times \mathbb{R}^m$,
(iv) $|f(t,x,0,0,0)| \leq K(1+|x|^q)$ for every $x \in \mathbb{R}^d$ uniformly in $t \in [0,T]$.

Proposition 3.2. *Under Assumption 3.2, the BSDE (3.3) has a unique solution $(Y^{t,x,\epsilon}, Z^{t,x,\epsilon}, \psi^{t,x,\epsilon})$ which belongs to $\mathbb{S}^p_m[t,T] \times \mathbb{H}^p_{m\times l}[t,T] \times \mathbb{H}^p_{m,\nu}[t,T]$ for $\forall p \geq 2$. Furthermore, it also satisfies*

$$||\hat{\Theta}^{t,x,\epsilon}||^p_{\mathcal{K}^p[t,T]} \leq C_p(1+|x|^{pq}) \tag{3.4}$$

for every $p \geq 2$.

Proof. The existence follows from Lemma B.2. In addition, one has

$$||\hat{\Theta}^{t,x,\epsilon}||^p_{\mathcal{K}^p[t,T]} \leq C_p \mathbb{E}\left[|\xi(X^{t,x,\epsilon}_T)|^p + \Big(\int_t^T |f(s, X^{t,x,\epsilon}_s, 0,0,0)|ds\Big)^p\right]$$

and hence one obtains the desired result by Lemma A.3 and the assumption of polynomial growth of $\xi(x), f(\cdot,x,0,0,0)$. □

To lighten the notation, we use the following symbol to represent the collective arguments:

$$\Theta_r^{t,x,\epsilon} := \Big(X_r^{t,x,\epsilon}, Y_r^{t,x,\epsilon}, Z_r^{t,x,\epsilon}, \int_{\mathbb{R}_0} \rho(z)\psi_r^{t,x,\epsilon}(z)\nu(dz)\Big)$$
$$\hat{\Theta}_r^{t,x,\epsilon} := \Big(Y_r^{t,x,\epsilon}, Z_r^{t,x,\epsilon}, \int_{\mathbb{R}_0} \rho(z)\psi_r^{t,x,\epsilon}(z)\nu(dz)\Big).$$

We also use $\partial_\Theta := (\partial_x, \partial_y, \partial_z, \partial_u)$ as well as $\partial_{\hat{\Theta}} := (\partial_y, \partial_z, \partial_u)$ and their higher order derivatives.

Remark

Let us remark on the practical implications of the Assumptions 3.1 and 3.2, since some readers may find that the smoothness assumption is too restrictive. Since the financial problems relevant for the BSDEs are inevitably non-linear, we are forced to consider in a portfolio level. Thus, ξ and f are likely to be given by complicated piecewise linear functions, which involve a large number of non-smooth points. The first step we can do is to approximate the overall form of these functions by smooth ones by introducing appropriate mollifiers. In the industry, this is quite common even for linear products such as a digital option to make delta hedging feasible in practice. A small additional fee arising from a mollifier is charged to a client as a hedging cost. It is also used for CVA evaluation by Henry-Labordère (2012) [25], for example. We think that making the approximation procedures more complicated by rigorously dealing with the non-smoothness fails to evaluate the relative importance of practical matters. In fact, we will provide a rigorous justification for the above smoothing arguments using the stability result in Section 10, which tells that one can safely work on $\mathcal{C}_b^\infty$ functions whenever rather mild conditions are satisfied.

4 Representation theorem for the BSDE

We define the Malliavin derivatives $D_{t,z}$ according to the conventions used in Section 3 of Delong & Imkeller (2010) [11] and Section 2.6 of Delong (2013) [10] (with $\sigma = 1$). See also Di Nunno et al (2009) [12] for details and other applications.

According to their definition, if the random variable $H(\cdot, \omega_\mu)$ is differentiable in the sense of classical Malliavin's calculus for $\mathbb{P}_\mu$-a.e. $\omega_\mu \in \Omega_\mu$, then we have the relation

$$D_{t,0}H(\omega_W, \omega_\mu) = D_t H(\cdot, \omega_\mu)(\omega_W) \ ,$$

where $D.$ is the Malliavin's derivative with respect to the Wiener direction. For the definition $D_{t,z}H$ with $z \neq 0$, the increment quotient operator is introduced

$$\mathcal{I}_{t,z}H(\omega_W, \omega_\mu) := \frac{H(\omega_W, \omega_\mu^{t,z}) - H(\omega_W, \omega_\mu)}{z}$$

where $\omega_\mu^{t,z}$ transforms a family $\omega_\mu = ((t_1, z_1), (t_2, z_2), \cdots) \in \Omega_\mu$ into a new family $\omega_\mu^{t,z}((t,z), (t_1, z_1), (t_2, z_2), \cdots) \in \Omega_\mu$. This is defined for a one-dimensional Poisson random measure. In the multi-dimensional case, $\mathcal{I}_{t,z}H$ is extended to k-dimensional vector in the obvious way. It is known that when $\mathbb{E}\Big[\int_0^T \int_E |\mathcal{I}_{t,z}H|^2 z^2 \nu(dz)dt\Big] = \mathbb{E}\Big[\sum_{i=1}^k \int_0^T \int_{\mathbb{R}_0} |\mathcal{I}_{t,z_i}H|^2 z_i^2 \nu^i(dz_i)dt\Big] < \infty$, one has $D_{t,z}H = \mathcal{I}_{t,z}H$.

Proposition 4.1. *Under Assumption 3.1, the process $X^{t,x,\epsilon}$ is Malliavin differentiable. Moreover, it satisfies*

$$\sup_{(s,z)\in[0,T]\times\mathbb{R}^k} \mathbb{E}\Big[\sup_{r\in[s,T]} |D_{s,z}X_r^{t,x,\epsilon}|^p\Big] < \infty$$

for any $\forall p \geq 2$.

Proof. This is a modification of Theorem 4.1.2 of [10] for our setting. The existence of Malliavin derivative follows from Theorem 3 in Petrou (2008) [39].

According to [39], for $z^i \neq 0$, one has

$$\begin{aligned} D_{s,z^i}X_r^{t,x,\epsilon} &= \frac{\gamma^i(s, X_{s-}^{t,x,\epsilon}, z^i, \epsilon)}{z^i} + \int_s^r D_{s,z^i}b(u, X_u^{t,x,\epsilon}, \epsilon)du \\ &+ \int_s^r D_{s,z^i}\sigma(u, X_u^{t,x,\epsilon}, \epsilon)dW_u + \int_s^r \int_E D_{s,z^i}\gamma(u, X_{u-}^{t,x,\epsilon}, z, \epsilon)\widetilde{\mu}(du, dz) \end{aligned} \quad (4.1)$$

for $s \leq r$ with $D_{s,z^i}X_r^{t,x,\epsilon} = 0$ otherwise. Here, γ^i denotes the i-th column vector and

$$D_{s,z^i}b(u, X_u^{t,x,\epsilon}, \epsilon) := \frac{1}{z^i}\big[b(u, X_u^{t,x,\epsilon} + z^i D_{s,z^i}X_u^{t,x,\epsilon}, \epsilon) - b(u, X_u^{t,x,\epsilon}, \epsilon)\big]$$

and similarly for the terms $(D_{s,z^i}\sigma(u, X_u^{t,x,\epsilon}, \epsilon), D_{s,z^i}\gamma(u, X_{u-}^{t,x,\epsilon}, z, \epsilon))$. Due to the uniformly bounded derivative of $\partial_x b, \partial_x\sigma, \partial_x\gamma/\eta$, (4.1) has the unique solution by Lemma A.3. In addition, applying the Burkholder-Davis-Gundy (BDG) and Gronwall inequalities and Lemma A.1, one obtains

$$\mathbb{E}||D_{s,z^i}X^{t,x,\epsilon}||_{[s,T]}^p \leq C_p\Big(\Big|\frac{\gamma^i(s, 0, z^i, \epsilon)}{z^i}\Big|^p + \mathbb{E}||X^{t,x,\epsilon}||_T^p\Big)$$

By Assumption 3.1 (iii), we obtain the desired result. The arguments for the Wiener direction ($z = 0$) are similar. □

Next theorem is an adaptation of Theorem 3.5.1 and Theorem 4.1.4 of [10] to our setting. We suppress the superscripts (t, x, ϵ) denoting the initial data for simplicity.

Theorem 4.1. *Under Assumptions 3.1 and 3.2,*
(a) There exists a unique solution $(Y^{s,0}, Z^{s,0}, \psi^{s,0})$ belongs to $\mathcal{K}^p$ for $\forall p \geq 2$ to the BSDE

$$Y_u^{s,0} = D_{s,0}\xi(X_T) + \int_u^T f^{s,0}(r)dr - \int_u^T Z_r^{s,0}dW_r - \int_u^T \int_E \psi_r^{s,0}(z)\widetilde{\mu}(dr, dz)$$

where

$$\begin{aligned} D_{s,0}\xi(X_T) &:= \partial_x\xi(X_T)D_{s,0}X_T \\ f^{s,0}(r) &= \partial_x f(r, \Theta_r)D_{s,0}X_r + \partial_y f(r, \Theta_r)Y_r^{s,0} + \partial_z f(r, \Theta_r)Z_r^{s,0} \\ &+ \partial_u f(r, \Theta_r)\int_{\mathbb{R}_0} \rho(z)\psi_r^{s,0}(z)\nu(dz). \end{aligned}$$

(b) For $z^i \neq 0$, there exists a unique solution $(Y^{s,z^i}, Z^{s,z^i}, \psi^{s,z^i})$ belongs to $\mathcal{K}^p$ for $\forall p \geq 2$

to the BSDE

$$Y_u^{s,z^i} = D_{s,z^i}\xi(X_T) + \int_u^T f^{s,z^i}(r)dr - \int_u^T Z_r^{s,z^i}dW_r - \int_u^T \int_E \psi_r^{s,z^i}(z)\widetilde{\mu}(dz,dr)$$

where

$$\begin{aligned} D_{s,z^i}\xi(X_T) &:= \frac{\xi(X_T + z^i D_{s,z^i}X_T) - \xi(X_T)}{z^i} \\ f^{s,z^i}(r) &:= \Big[f\Big(r, X_r + z^i D_{s,z^i}X_r, Y_r + z^i D_{s,z^i}Y_r, Z_r + z^i D_{s,z^i}Z_r \\ &\quad , \int_{\mathbb{R}_0}\rho(e)\big[\psi_r(e) + z^i D_{s,z^i}\psi_r(e)\big]\nu(de)\Big) - f\Big(r, X_r, Y_r, Z_r, \int_{\mathbb{R}_0}\rho(e)\psi_r(e)\nu(de)\Big)\Big]/z^i \end{aligned}$$

for every $1 \le i \le k$.
(c)For $u < s \le T$, set $(Y_u^{s,z}, Z_u^{s,z}, \psi_u^{s,z}) = 0$ for $z \in \mathbb{R}^k$ (i.e., including Wiener direction $z = 0$). Then, (Y, Z, ψ) is Malliavin differentiable and $(Y^{s,z}, Z^{s,z}, \psi^{s,z})$ is a version of $(D_{s,z}Y, D_{s,z}Z, D_{s,z}\psi)$.
(d)Set a deterministic function $u(t,x,\epsilon) := Y_t^{t,x,\epsilon}$ using the solution of the BSDE (3.3). If u is continuous in t and one-time continuously differentiable with respect to x, then

$$Z_s^{t,x,\epsilon} = \partial_x u(s, X_{s-}^{t,x,\epsilon}, \epsilon)\sigma(s, X_{s-}^{t,x,\epsilon}, \epsilon) \tag{4.2}$$

$$\Big(\psi_s^{t,x,\epsilon}(z)\Big)^i_{1\le i\le k} = \Big(u\big(s, X_{s-}^{t,x,\epsilon} + \gamma^i(s, X_{s-}^{t,x,\epsilon}, z^i, \epsilon), \epsilon\big) - u(s, X_{s-}^{t,x,\epsilon}, \epsilon)\Big)_{1\le i\le k} \tag{4.3}$$

for $t \le s \le T$ and $z = (z^i)_{1\le i\le k} \in \mathbb{R}^k$.

Proof. (a) and (b) can be proved by Lemma B.2, the boundedness of derivatives and the fact that $\Theta^{t,x,\epsilon} \in \mathbb{S}^p \times \mathcal{K}^p$ and $D_{s,z}X \in \mathbb{S}^p$ for $\forall p \ge 2$.
(c) can be proved as a simple modification of Theorem 3.5.1 in [10], which is a straightforward extension of Proposition 5.3 in El Karoui et.al (1997) [16] to the jump case. The conditions written for ω-dependent driver (assumptions (vii) and (viii) of [10]) can be replaced by our assumption on f, which is Lipschitz with respect to (y, z, u) and has a polynomial growth in x. Note that we already know $X^{t,x,\epsilon}, D_{s,z}X^{t,x,\epsilon} \in \mathbb{S}^p$ for $\forall p \ge 2$.
(d) follows from Theorem 4.1.4 of [10]. □

5 Classical differentiation of the BSDE with respect to x

For the analysis of our asymptotic expansion with respect to ϵ, we need to study the properties of $(\partial_\epsilon^n \hat{\Theta}^{t,x,\epsilon})$. In this section however, we investigate the properties of $(\partial_x^n \hat{\Theta}^{t,x,\epsilon})$ first, which becomes relevant to discuss the $(\partial_\epsilon^n \hat{\Theta}^{t,x,\epsilon})$ in the next section.

Lemma 5.1. *Under Assumptions 3.1 and 3.2, $\hat{\Theta}^{t,x,\epsilon}$ is classically differentiable with respect to x, and it is given by $\partial_x \hat{\Theta}^{t,x,\epsilon}$ defined as the unique solution of the BSDE with formal differentiation with respect to x:*

$$\begin{aligned} \partial_x Y_s^{t,x,\epsilon} &= \partial_x \xi(X_T^{t,x,\epsilon})\partial_x X_T^{t,x,\epsilon} + \int_s^T \partial_\Theta f(r, \Theta_r^{t,x,\epsilon})\partial_x \Theta_r^{t,x,\epsilon} dr \\ &\quad - \int_s^T \partial_x Z_r^{t,x,\epsilon} dW_r - \int_s^T \int_E \partial_x \psi_r^{t,x,\epsilon}(z)\widetilde{\mu}(dr,dz) \end{aligned} \tag{5.1}$$

and $\partial_x \hat{\Theta}^{t,x,\epsilon} \in \mathcal{K}^p[t,T]$ *satisfying*

$$||\partial_x \hat{\Theta}^{t,x,\epsilon}||^p_{\mathcal{K}^p[t,T]} \le C_p(1+|x|^{pq})$$

for any $\forall p \ge 2$.

Proof. The existence and uniqueness can be easily shown from Lemma B.2. Note that the BSDE (5.1) is linear with bounded Lipschitz constants and satisfies

$$\begin{aligned}
&||\partial_x \hat{\Theta}^{t,x,\epsilon}||^p_{\mathcal{K}^p[t,T]} \le C_p \mathbb{E}\Big[|\partial_x \xi(X_T^{t,x,\epsilon})|^p |\partial_x X_T^{t,x,\epsilon}|^p + \Big(\int_t^T |\partial_x f(r,\Theta_r^{t,x,\epsilon})||\partial_x X_r^{t,x,\epsilon}|dr\Big)^p\Big] \\
&\le C_p ||\partial_x X^{t,x,\epsilon}||^p_{\mathbb{S}^{2p}[t,T]} \Big\{ \Big(\mathbb{E}|\partial_x \xi(X_T^{t,x,\epsilon})|^{2p}\Big)^{1/2} + \Big(\mathbb{E}\Big(\int_t^T |\partial_x f(r,X_r^{t,x,\epsilon},0)|dr\Big)^{2p}\Big)^{1/2} \\
&+||\hat{\Theta}^{t,x,\epsilon}||^p_{\mathcal{K}^{2p}[t,T]}\Big\} \le C_p(1+|x|^{pq})
\end{aligned}$$

for any $\forall p \ge 2$. With a simple modification of Theorem 3.1 of [34], one can also show that

$$\lim_{h\to 0} ||\nabla^h \hat{\Theta}^{t,x,\epsilon} - \partial_x \hat{\Theta}^{t,x,\epsilon}||^2_{\mathcal{K}^2[t,T]} = 0$$

where $\nabla^h \hat{\Theta}^{t,x,\epsilon} := \dfrac{\hat{\Theta}^{t,x+h,\epsilon} - \hat{\Theta}^{t,x,\epsilon}}{h}$ with $h \neq 0$ (for each direction). This gives the agreement with the classical differentiation. □

Corollary 5.1. *Under Assumptions 3.1 and 3.2, there exists* $\partial_x u(t,x,\epsilon)$ *that has at most a polynomial growth in* x *uniformly in* $(t,\epsilon) \in [0,T] \times [0,1]$ *and continuous in* (t,x). *Furthermore,* $Z^{t,x,\epsilon}$ *and* $\int_{\mathbb{R}_0} \rho(z)\psi^{t,x,\epsilon}(z)\nu(dz)$ *belong to* $\mathbb{S}^p[t,T]$ *for every* $\forall p \ge 2$.

Proof. This is an adaptation of Corollary 3.2 of [34] to our setting. In particular, note that $\partial_x u(t,x,\epsilon) = \partial_x Y_t^{t,x,\epsilon}$ and there exists some constant $C>0$ such that

$$|\partial_x u(t,x,\epsilon)| \le ||\partial_x \hat{\Theta}^{t,x,\epsilon}||_{\mathcal{K}^p[t,T]} \le C(1+|x|^q)$$

for every $x \in \mathbb{R}^d$ uniformly in $t \in [0,T]$ by Lemma 5.1. The continuity of $\partial_x u(t,x,\epsilon)$ in (t,x) can be shown in the same way as [34] using the continuity of $X^{t,x,\epsilon}$ in (t,x), which can be seen in Lemma A.3. Then, from the representation given in (4.2), (4.3) and the above result, one sees

$$|Z_s^{t,x,\epsilon}| + \Big|\int_E \rho(z)\psi_s^{t,x,\epsilon}(z)\nu(dz)\Big| \le C(1+|X_{s-}^{t,x,\epsilon}|^{q+1})$$

which gives the desired result $\hat{\Theta}^{t,x,\epsilon} \in \mathbb{S}^p[t,T]^{\otimes 3}$ for any $p \ge 2$. □

Proposition 5.1. *Under Assumptions 3.1 and 3.2, the classical differentiation of* $\hat{\Theta}^{t,x,\epsilon}$ *with respect to* x *arbitrary many times exists. For every* $n \ge 1$, *it is given by the solution* $\partial_x^n \hat{\Theta}^{t,x,\epsilon}$ *to the BSDE*

$$\begin{aligned}
\partial_x^n Y_s^{t,x,\epsilon} &= \xi_n + \int_s^T \Big\{H_{n,r} + \partial_\Theta f(r,\Theta_r^{t,x,\epsilon})\partial_x^n \Theta_r^{t,x,\epsilon}\Big\}dr \\
&\quad - \int_s^T \partial_x^n Z_r^{t,x,\epsilon} dW_r - \int_s^T \int_E \partial_x^n \psi_r^{t,x,\epsilon}(z)\widetilde{\mu}(dr,dz) \qquad (5.2)
\end{aligned}$$

where

$$\xi_n := n!\sum_{k=1}^{n}\sum_{\beta_1+\cdots+\beta_k=n,\beta_i\geq 1}\frac{1}{k!}\partial_x^k\xi(X_T^{t,x,\epsilon})\prod_{j=1}^{k}\frac{1}{\beta_j!}\partial_x^{\beta_j}X_T^{t,x,\epsilon},$$

$$H_{n,r} := n!\sum_{k=2}^{n}\sum_{\beta_1+\cdots+\beta_k=n,\beta_i\geq 1}\sum_{i_x=0}^{k}\sum_{i_y=0}^{k-i_x}\sum_{i_z=0}^{k-i_x-i_y}\frac{\partial_x^{i_x}\partial_y^{i_y}\partial_z^{i_z}\partial_u^{k-i_x-i_y-i_z}f(r,\Theta_r^{t,x,\epsilon})}{i_x!i_y!i_z!(k-i_x-i_y-i_z)!}$$

$$\times\prod_{j_x=1}^{i_x}\frac{1}{\beta_{j_x}!}\partial_x^{\beta_{j_x}}X_r^{t,x,\epsilon}\prod_{j_y=i_x+1}^{i_x+i_y}\frac{1}{\beta_{j_y}!}\partial_x^{\beta_{j_y}}Y_r^{t,x,\epsilon}\prod_{j_z=i_x+i_y+1}^{i_x+i_y+i_z}\frac{1}{\beta_{j_z}!}\partial_x^{\beta_{j_z}}Z_r^{t,x,\epsilon}$$

$$\times\prod_{j_u=i_x+i_y+i_z+1}^{k}\frac{1}{\beta_{j_u}!}\int_{\mathbb{R}_0}\rho(z)\partial_x^{\beta_{j_u}}\psi_r^{t,x,\epsilon}(z)\nu(dz)$$

and satisfies $\partial_x^n\hat{\Theta}^{t,x,\epsilon}\in\mathbb{S}^p[t,T]^{\otimes 3}$ *for* $\forall p\geq 2$*.*

Proof. We can prove recursively by the arguments used to show Proposition 3.2, Lemma 5.1 and Corollary 5.1. We already know that $\hat{\Theta}^{t,x,\epsilon}\in\mathbb{S}^p[t,T]^{\otimes 3}$ and $\partial_x\hat{\Theta}^{t,x,\epsilon}\in\mathcal{K}^p[t,T]$ for any $p\geq 2$. The BSDE for $\partial_x^2\hat{\Theta}^{t,x,\epsilon}$ has bounded Lipschitz constants and $H_{2,r}$ is at most quadratic in $(\partial_x\hat{\Theta}_r^{t,x,\epsilon})$. Since $\xi(x), f(\cdot,x,0)$ have at most a polynomial growth in x and the fact that $\partial_x^m X^{t,x,\epsilon}$ for $m\geq 0$ and $\hat{\Theta}^{t,x,\epsilon}$ are in $\mathbb{S}^p[t,T]$ for any $p\geq 2$, one can prove the existence of the unique solution $\partial_x^2\hat{\Theta}^{t,x,\epsilon}\in\mathcal{K}^p[t,T]$ for any $p\geq 2$ by Lemma B.2. Furthermore, one can show as in Lemma 5.1 that $||\partial_x^2\hat{\Theta}^{t,x,\epsilon}||_{\mathcal{K}^p[t,T]}$ has at most polynomial growth in x. By following the arguments of Theorem 3.1 of [34], one sees this agrees with the classical differentiation in the sense of Lemma 5.1. This in turn shows the existence $\partial_x^2 u(t,x,\epsilon)=\partial_x^2 Y_t^{t,x,\epsilon}$ and also the fact that $\partial_x^2 u(t,x,\epsilon)$ has at most a polynomial growth in x. This implies that, together with Assumption 3.1 and the representation theorem (4.2) (4.3), $\partial_x Z^{t,x,\epsilon}$ and $\int_{\mathbb{R}_0}\rho(z)\partial_x\psi^{t,x,\epsilon}(z)\nu(dz)$ are in $\mathbb{S}^p[t,T]$ for $\forall p\geq 2$. Thus, we get $\partial_x\hat{\Theta}^{t,x,\epsilon}\in\mathbb{S}^p[t,T]^{\otimes 3}$.

In the same manner, if we assume that $\left(\partial_x^i\hat{\Theta}^{t,x,\epsilon}\right)_{i\leq n}\in\mathbb{S}^p[t,T]^{\otimes 3}$ and that $\partial_x^{n+1}\hat{\Theta}^{t,x,\epsilon}\in\mathcal{K}^p[t,T]$ for $\forall p\geq 2$ with the $\mathcal{K}^p$-norm at most a polynomial growth in x then one can show that the existence of the unique solution $\partial_x^{n+2}\hat{\Theta}^{t,x,\epsilon}\in\mathcal{K}^p[t,T]$ with the norm at most a polynomial growth in x by Lemma B.2. It then implies from the representation theorem that $\partial_x^{n+1}\hat{\Theta}^{t,x,\epsilon}\in\mathbb{S}^p[t,T]^{\otimes 3}$ for $\forall p\geq 2$. This proves the proposition. □

6 Asymptotic Expansion

We are now going to prove $\partial_\epsilon^n\hat{\Theta}^{t,x,\epsilon}\in\mathbb{S}^p[t,T]^{\otimes 3}$ for any $\forall p\geq 2$ and $n\geq 1$. Although the strategy is similar to the previous section, we actually have to study the properties of $\left(\partial_x^m\partial_\epsilon^n\hat{\Theta}^{t,x,\epsilon}\right)_{n,m\geq 0}$ since ϵ affects $u(s,X_{s-}^{t,x,\epsilon},\epsilon)$ not only from its explicit dependence but also from $X^{t,x,\epsilon}$.

Lemma 6.1. *Under Assumptions 3.1 and 3.2,* $\hat{\Theta}^{t,x,\epsilon}$ *is classically differentiable with respect to* ϵ*, and it is given by* $\partial_\epsilon\hat{\Theta}^{t,x,\epsilon}$*, which is defined as the unique solution of the BSDE* *with formal differentiation with respect to ϵ:*

$$\partial_\epsilon Y_s^{t,x,\epsilon} = \partial_x \xi(X_T^{t,x,\epsilon})\partial_\epsilon X_T^{t,x,\epsilon} + \int_s^T \partial_\Theta f(r,\Theta_r^{t,x,\epsilon})\partial_\epsilon \Theta_r^{t,x,\epsilon} dr \\ - \int_s^T \partial_\epsilon Z_r^{t,x,\epsilon} dW_r - \int_s^T \int_E \partial_\epsilon \psi_r^{t,x,\epsilon} \widetilde{\mu}(dr,dz) \ .$$

One has $\partial_\epsilon \hat{\Theta}^{t,x,\epsilon} \in \mathcal{K}^p[t,T]$ satisfying

$$||\partial_\epsilon \hat{\Theta}^{t,x,\epsilon}||_{\mathcal{K}^p[t,T]}^p \le C_p(1+|x|^{pq})$$

for any $\forall p \ge 2$.

Proof. The proof can be done similarly as in Lemma 5.1. □

We now get the following result:

Proposition 6.1. *Under Assumptions 3.1 and 3.2, the classical differentiation of $\hat{\Theta}^{t,x,\epsilon}$ with respect to ϵ arbitrary many times exists and is given by the solution $\partial_\epsilon^n \hat{\Theta}^{t,x,\epsilon}$ to the BSDE*

$$\partial_\epsilon^n Y_s^{t,x,\epsilon} = \widetilde{\xi}_n + \int_s^T \Big\{ \widetilde{H}_{n,r} + \partial_\Theta f(r,\Theta_r^{t,x,\epsilon})\partial_\epsilon^n \Theta_r^{t,x,\epsilon} \Big\} dr \\ - \int_s^T \partial_\epsilon^n Z_r^{t,x,\epsilon} dW_r - \int_s^T \int_E \partial_\epsilon^n \psi_r^{t,x,\epsilon} \widetilde{\mu}(dr,dz)$$

for every $n \ge 1$. Here, $\widetilde{\xi}_n$ and $\widetilde{H}_{n,r}$ are given by the expressions of ξ_n and $H_{n,r}$ in Proposition 5.1 with $\partial_x^{\beta_j \Theta}$ replaced by $\partial_\epsilon^{\beta_j \Theta}$. Moreover, for every $n \ge 1$, $\partial_\epsilon^n \hat{\Theta}^{t,x,\epsilon} \in \mathbb{S}^p[t,T]^{\otimes 3}$ for $\forall p \ge 2$.

Proof. We start from the result of Lemma 6.1, which implies $\partial_\epsilon u(t,x,\epsilon)$ has at most polynomial growth in x. Using the fact that $\partial_\epsilon \Theta^{t,x,\epsilon} \in \mathbb{S}^p[t,T] \times \mathcal{K}^p[t,T]$ and $\partial_x \Theta^{t,x,\epsilon} \in \mathbb{S}^p[t,T]^{\otimes 4}$, one can recursively prove as in Proposition 5.1, for every $n \ge 1$ that the classical differentiation $\partial_x^n \partial_\epsilon \hat{\Theta}^{t,x,\epsilon}$ exists and belongs to $\mathcal{K}^p[t,T]$ for $\forall p \ge 2$ with the $\mathcal{K}^p$-norm bounded by a polynomial of x. This implies $\partial_x^n \partial_\epsilon u(t,x,\epsilon)$ has at most a polynomial growth in x. Using this result and the polynomial growth property of $\partial_x^m u(t,x,\epsilon)$, the representations (4.2) and (4.3) and their derivatives, one can show that $\partial_x^{n-1}\partial_\epsilon Z^{t,x,\epsilon}$ and $\int_{\mathbb{R}_0} \rho(z)\partial_x^{n-1}\partial_\epsilon \psi^{t,x,\epsilon}(z)\nu(dz)$ are in $\mathbb{S}^p[t,T]$ for $\forall p \ge 2$. Thus, we find $\partial_x^n \partial_\epsilon \hat{\Theta}^{t,x,\epsilon} \in \mathbb{S}^p[t,T]^{\otimes 3}$ for every $n \ge 1$ by induction. Using the above result, similar procedures give that $\partial_x^n \partial_\epsilon^2 \hat{\Theta}^{t,x,\epsilon} \in \mathbb{S}^p[t,T]^{\otimes 3}$ for every $n \ge 1$ and $\forall p \ge 2$. By induction, one can finally show that, for every $n,m \ge 0$, $\partial_x^n \partial_\epsilon^m \hat{\Theta}^{t,x,\epsilon}$ exists and belongs to $\mathbb{S}^p[t,T]^{\otimes 3}$ for $\forall p \ge 2$, and hence also the claim of the proposition. □

We have shown that $\Theta^{t,x,\epsilon}$ is classically differentiable with respect to (x,ϵ) arbitrary many times and that, for every $n \ge 0$, $\partial_\epsilon^n \Theta^{t,x,\epsilon} \in \mathbb{S}^p[t,T]^{\otimes 4}$ for $\forall p \ge 2$. Let us define for $s \in [t,T]$ that

$$\Theta_s^{[n]} := \frac{1}{n!}\partial_\epsilon^n \Theta_s^{t,x,\epsilon}\Big|_{\epsilon=0}.$$

Using the differentiability and the Taylor formula, one has

$$\Theta_s^{t,x,\epsilon} = \Theta_s^{[0]} + \sum_{n=1}^{N} \epsilon^n \Theta_s^{[n]} + \frac{\epsilon^{N+1}}{N!} \int_0^1 (1-u)^N \left(\partial_\alpha^{N+1} \Theta_s^{t,x,\alpha}\right)\Big|_{\alpha=u\epsilon} du \; . \tag{6.1}$$

As we shall see later, each $\Theta^{[m]}, m \in \{1, 2, \cdots\}$ can be evaluated by solving the system of linear ODEs. Although $\Theta^{[0]}$ requires to solve a non-linear ODE as an exception, the existence of the bounded solution is guaranteed under the Assumptions 3.1 and 3.2.

The next theorem is the first main result of the paper which gives the error estimate of the approximation of $\Theta^{t,x,\epsilon}$ by the series of $\Theta^{[m]}, m \in \{0, 1, \cdots, \}$.

Theorem 6.1. *Under Assumptions 3.1 and 3.2, the asymptotic expansion of the forward-backward SDEs (3.1) and (3.3) is given by (6.1) and satisfies, with some positive constant C_p, that*

$$\left|\left|\Theta^{t,x,\epsilon} - \left(\Theta^{[0]} + \sum_{n=1}^{N} \epsilon^n \Theta^{[n]}\right)\right|\right|_{\mathbb{S}^p[t,T]}^p \le \epsilon^{p(N+1)} C_p \; . \tag{6.2}$$

Proof. This immediately follows from Propositions 3.1 and 6.1. □

7 State-dependent jump intensity

When ν is a finite measure $\nu(E) < \infty$, all the previous results hold true with slightly weaker assumptions with $\eta, \rho \equiv 1$ in Assumptions 3.1 and 3.2. In practical applications, however, there are many cases where we want to make the jump intensity state dependent. In this section, we solve this problem when the intensity is bounded.

In particular, we consider the forward-backward SDEs (3.1) and (3.3) but with the compensated random measure $\widetilde{\mu}(dr, dz)$ given by, for $1 \le i \le k$,

$$\widetilde{\mu}^i(dr, dz) = \mu^i(dr, dz) - \lambda^i(r, X_r^{t,x,\epsilon})\nu^i(dz)dr$$

where ν^i is normalized as $\nu^i(\mathbb{R}_0) = 1$ and $\lambda^i : [0, T] \times \mathbb{R}^d \to \mathbb{R}$. One can see that the random measure is not Poissonian any more and depends implicitly on ϵ through its intensity.

Assumption 7.1. *For every $1 \le i \le k$, $\nu^i(\mathbb{R}_0) = 1$ and*
(i) the function $\lambda^i(t, x)$ is continuous in (t, x), continuously differentiable arbitrary many times with respect to x with uniformly bounded derivatives,
(ii) there exist positive constants c_1, c_2 such that $0 < c_1 \le \lambda^i(t, x) \le c_2$ uniformly in $(t, x) \in [0, T] \times \mathbb{R}^d$,
(iii) for every $m \ge 0$, $|\partial_\epsilon^m \gamma_{\cdot,i}(t, x, z, \epsilon)| \le K$ uniformly in $(t, x, z, \epsilon) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}_0 \times [0, 1]$.

Lemma 7.1. *Under Assumption 7.1, one can define an equivalent probability measure $\mathbb{Q}$ by, for $s \in [t, T]$,*

$$\frac{d\mathbb{Q}}{d\mathbb{P}}\Big|_{\mathcal{F}_s} = M_s$$

where M is a strictly positive $\mathbb{P}$-martingale given by

$$M_s = 1 + \sum_{i=1}^{k} \int_t^s M_{r-} \Big(\frac{c_2}{\lambda^i(r, X_{r-}^{t,x,\epsilon})} - 1 \Big) \widetilde{\mu}^i(dr, \mathbb{R}_0) \; .$$

Under the new measure $\mathbb{Q}$, the compensated random measure becomes

$$\widetilde{\mu}^{\mathbb{Q}}(dr,dz) = \mu(dr,dz) - c_2\nu(dz)dt$$

and hence μ is Poissonian. Moreover, for $\forall s \in [t,T]$,

$$M_s \geq \exp\bigl(-(c_2-c_1)k(T-t)\bigr) \ .$$

Proof. By Kazamaki (1979) [28], it is known that if X is a BMO martingale satisfying $\Delta X_t \geq -1+\delta$ a.s. for all $t \in [0,T]$ with some strictly positive constant $\delta > 0$, then Doléans-Dade exponential $\mathcal{E}(X)$ is uniformly integrable. One can easily confirm that this condition is satisfied for a martingale

$$\int^{\cdot} \Bigl(c_2/\lambda(s, X_s^{t,x,\epsilon}) - 1\Bigr)\widetilde{\mu}(ds, \mathbb{R}_0) \ .$$

Thus the given measure change is well-defined and the first claim follows from Theorem 41 in Chapter 3 of [41]. The explicit expression

$$\begin{aligned} M_s &= \prod_{i=1}^k \left\{ \prod_{0<r\leq s} \Bigl(\frac{c_2}{\lambda^i(r, X_{r-}^{t,x,\epsilon})}\Bigr)^{\Delta\mu^i(r,\mathbb{R}_0)} \exp\Bigl(-\int_t^s (c_2 - \lambda^i(r, X_{r-}^{t,x,\epsilon}))dr\Bigr) \right\} \\ &\geq \exp\Bigl(-\int_t^s k(c_2-c_1)dr\Bigr) \end{aligned}$$

proves the second claim. ☐

In the measure $\mathbb{Q}$, we have

$$\begin{aligned} X_s^{t,x,\epsilon} &= x + \int_t^s \widetilde{b}(r, X_r^{t,x,\epsilon}, \epsilon)dr + \int_t^s \sigma(r, X_r^{t,x,\epsilon}, \epsilon)dW_r \\ &\quad + \int_t^s \int_E \gamma(r, X_{r-}^{t,x,\epsilon}, z, \epsilon)\widetilde{\mu}^{\mathbb{Q}}(dr,dz) \end{aligned} \tag{7.1}$$

$$\begin{aligned} Y_s^{t,x,\epsilon} &= \xi(X_T^{t,x,\epsilon}) + \int_s^T \widetilde{f}\Bigl(r, X_r^{t,x,\epsilon}, Y_r^{t,x,\epsilon}, Z_r^{t,x,\epsilon}, \int_{\mathbb{R}_0} \psi_r^{t,x,\epsilon}(z)\nu(dz)\Bigr)dr \\ &\quad - \int_s^T Z_r^{t,x,\epsilon}dW_r - \int_s^T \int_E \psi_r^{t,x,\epsilon}(z)\widetilde{\mu}^{\mathbb{Q}}(dr,dz) \end{aligned} \tag{7.2}$$

where

$$\widetilde{b}(s,x,\epsilon) = b(s,x,\epsilon) + \sum_{i=1}^k (c_2 - \lambda^i(s,x)) \int_{\mathbb{R}_0} \gamma^i(s,x,z^i,\epsilon)\nu(dz^i)$$

$$\widetilde{f}(s,x,y,z,u) = f(s,x,y,z,u) - \sum_{i=1}^k (c_2 - \lambda^i(s,x))u^i.$$

Theorem 7.1. *Under Assumptions 3.1, 3.2 with ρ and η replaced by 1, and Assumption 7.1, the solution $\Theta^{t,x,\epsilon}$ of the forward-backward SDEs (3.1) and (3.3) allows the asymptotic expansion with respect to ϵ and satisfies the same error estimate (6.2) in the original measure $\mathbb{P}$.*

Proof. Assumption 7.1 makes $(\widetilde{b}, \widetilde{f})$ once again satisfy Assumptions 3.1 and 3.2 with ρ, η replaced by 1. Therefore, all the results in the previous sections hold true under the measure $\mathbb{Q}$ to the equivalent FBSDEs (7.1) and (7.2). In particular this implies from Lemma 7.1 that, with some positive constant C_p,

$$\begin{aligned}
\epsilon^{p(N+1)} C_p &\geq \mathbb{E}^{\mathbb{Q}}\left[\sup_{s\in[t,T]}\left|\Theta_s^{t,x,\epsilon} - \Big(\Theta_s^{[0]} + \sum_{n=1}^{N}\epsilon^n \Theta_s^{[n]}\Big)\right|^p\right] \\
&= \mathbb{E}\left[M_T \sup_{s\in[t,T]}\left|\Theta_s^{t,x,\epsilon} - \Big(\Theta_s^{[0]} + \sum_{n=1}^{N}\epsilon^n \Theta_s^{[n]}\Big)\right|^p\right] \\
&\geq \exp\big(-k(c_2-c_1)(T-t)\big)\mathbb{E}\left[\sup_{s\in[t,T]}\left|\Theta_s^{t,x,\epsilon} - \Big(\Theta_s^{[0]} + \sum_{n=1}^{N}\epsilon^n \Theta_s^{[n]}\Big)\right|^p\right] .
\end{aligned}$$

This proves the claim. □

8 Implementation of the asymptotic expansion

In this section, we explain how to calculate $\Theta^{[n]}$, $n \in \{0, 1, 2, \cdots\}$ (semi)-analytically. As we shall see, if we introduce ϵ in a specific way to the forward SDE (3.1), then the grading structure introduced by the asymptotic expansion yields a very simple scheme requiring only a system of linear ODEs to be solved with only one exception at the zero-th order. It is also very remarkable that one can directly approximate not only $(Y^{t,x}, Z^{t,x})$ but also the $\mathbb{L}^2(E;\nu)$-valued process $\psi^{t,x}(\cdot)$. This is quite difficult, at least numerically very heavy, for regression based simulation techniques.

Let us put the initial time as $t = 0$, and take $(m = d = l = 1)$ for *notational simplicity.* The extension to higher dimensional setups is straightforward for which one only needs a proper indexing of each variable. Let us adopt a following parametrization of X with ϵ which obviously leads to *small-variance* expansion;

$$X_s^\epsilon = x + \int_0^s b(r, X_r^\epsilon, \epsilon)dr + \int_0^s \epsilon\sigma(r, X_r^\epsilon)dW_r + \int_0^s \int_{\mathbb{R}_0} \epsilon\gamma(r, X_{r-}^\epsilon, z)\widetilde{\mu}(dr, dz) ,$$

where we omit the superscript denoting the initial data $(0, x)$. Similar to the standard applications [42], this parameterization is crucial to obtain semi-analytic approximations.

As a special case, if we put $\xi(x) = e^{ikx}$ and $f \equiv 0$, the following calculation provides the estimate of X's characteristic function. Thus, its inverse Fourier transformation gives the estimate of the X's density function if exists. Note that Assumption 3.1 and hence the current scheme is not requiring the absolutely continuous density of X.

We assume Assumptions 3.1 and 3.2 (or those replaced by $\rho = \eta = 1$ and Assumption 7.1) hold throughout this section. The following result for $\Theta^{[0]}$ is obvious from the growth conditions of ξ and f.

Lemma 8.1. *The zero-th order solution* $\big(\Theta_s^{[0]}, s \in [0,T]\big)$ *is given by*

$$\begin{aligned}
X_s^{[0]} &= x + \int_0^s b(r, X_r^{[0]}, 0)dr \\
Y_s^{[0]} &= \xi(X_T^{[0]}) + \int_s^T f(r, X_r^{[0]}, Y_r^{[0]}, 0, 0)dr \\
Z^{[0]} &= \psi^{[0]}(\cdot) \equiv 0 \ .
\end{aligned} \tag{8.1}$$

which is continuous, deterministic and bounded.

Let us introduce some notations:

$$\begin{aligned}
&b^{[0]}(s) := b(s, X_s^{[0]}, 0), \quad \sigma^{[0]}(s) := \sigma(s, X_s^{[0]}), \quad \gamma^{[0]}(s,z) := \gamma(s, X_s^{[0]}, z) \\
&\xi^{[0]} := \xi(X_T^{[0]}), \quad f^{[0]}(s) := f(s, X_s^{[0]}, Y_s^{[0]}, 0, 0), \\
&\Gamma^{[0]}(s) := \int_{\mathbb{R}_0} \rho(z)\gamma^{[0]}(s,z)\nu(dz)
\end{aligned}$$

and their derivatives such that

$$\begin{aligned}
&\partial_x b^{[0]}(s) := \partial_x b(s,x,0)\Big|_{x=X_s^{[0]}}, \quad \partial_\epsilon b^{[0]}(s) = \partial_\epsilon b(s, X_s^{[0]}, \epsilon)\Big|_{\epsilon=0} \\
&\partial_x \Gamma^{[0]}(s) := \int_{\mathbb{R}_0} \rho(z)\partial_x \gamma(s,x,z)\Big|_{x=X_s^{[0]}}\nu(dz)
\end{aligned}$$

and similarly for the others.

In the first order of the expansion, we have to solve

$$X_s^{[1]} = \int_0^s \big[\partial_\epsilon b^{[0]}(r) + \partial_x b^{[0]}(r) X_r^{[1]}\big]dr + \int_0^s \sigma^{[0]}(r)dW_r + \int_0^s \int_{\mathbb{R}_0} \gamma^{[0]}(s,z)\widetilde{\mu}(dr,dz), \tag{8.2}$$

$$Y_s^{[1]} = \partial_x \xi^{[0]} X_T^{[1]} + \int_s^T \partial_\Theta f^{[0]}(r)\Theta_r^{[1]}dr - \int_s^T Z_r^{[1]}dW_r - \int_s^T \int_{\mathbb{R}_0} \psi_s^{[1]}(z)\widetilde{\mu}(dr,dz) \ . \tag{8.3}$$

Lemma 8.2. *There exits a unique solution* $\Theta^{[1]}$ *to (8.2) and (8.3) which belongs to* $\mathbb{S}^p[0,T]^{\otimes 4}$ *for* $\forall p \geq 2$*.* $\hat{\Theta}^{[1]}$ *is given by, for* $s \in [0,T]$ *and* $z \in \mathbb{R}_0$*,*

$$\begin{aligned}
Y_s^{[1]} &= y_1^{[1]}(s) X_s^{[1]} + y_0^{[1]}(s) \\
Z_s^{[1]} &= y_1^{[1]}(s)\sigma^{[0]}(s) \\
\psi_s^{[1]}(z) &= y_1^{[1]}(s)\gamma^{[0]}(s,z) \ .
\end{aligned}$$

Here, $\Big(y_1^{[1]}(s), y_0^{[1]}(s), s \in [0,T]\Big)$ *are the solutions to the following linear ODEs:*

$$\begin{aligned}
-\frac{dy_1^{[1]}(s)}{ds} &= \big(\partial_x b^{[0]}(s) + \partial_y f^{[0]}(s)\big)y_1^{[1]}(s) + \partial_x f^{[0]}(s), \\
-\frac{dy_0^{[1]}(s)}{ds} &= \partial_y f^{[0]}(s) y_0^{[1]}(s) + \Big(\partial_\epsilon b^{[0]}(s) + \partial_z f^{[0]}(s)\sigma^{[0]}(s) + \partial_u f^{[0]}(s)\Gamma^{[0]}(s)\Big)y_1^{[1]}(s)
\end{aligned}$$

with the terminal conditions $y_1^{[1]}(T) = \partial_x \xi^{[0]}$ *and* $y_0^{[1]}(T) = 0$*.*

Proof. The existence of the unique solution for $\Theta^{[1]}$ is obvious from Lemmas A.3 and B.2. The form of $Y^{[1]}$ is naturally expected from the linear structure of the BSDE and the order of ϵ. It automatically fixes the form of $Z^{[1]}$ and $\psi^{[1]}$. One can now compare the BSDE with $\hat{\Theta}^{[1]}$ substituted by the hypothesized form and what is obtained by applying Itô formula to the hypothesized $Y^{[1]}$. By comparing the coefficients of $X^{[1]}$ and the deterministic part, one obtains the given linear ODEs. The procedures are similar to those used in an Affine model when deriving its generating function. Since the hypothesized as well as the original variables satisfy the same BSDE, it provides one possible solution. But we know the solution is unique. □

In the second order of ϵ, one obtains

$$\begin{aligned} X_s^{[2]} &= \int_0^s \Big(\partial_x b^{[0]}(r) X_r^{[2]} + \frac{1}{2}\partial_x^2 b^{[0]}(r)(X_r^{[1]})^2 + \partial_x\partial_\epsilon b^{[0]}(r) X_r^{[1]} + \frac{1}{2}\partial_\epsilon^2 b^{[0]}(r)\Big) dr \\ &\quad + \int_0^s \partial_x\sigma^{[0]}(r) X_r^{[1]} dW_r + \int_{\mathbb{R}_0} \partial_x \gamma^{[0]}(r,z) X_r^{[1]} \widetilde{\mu}(dr,dz) \end{aligned} \tag{8.4}$$

and

$$\begin{aligned} Y_s^{[2]} &= \partial_x\xi^{[0]} X_T^{[2]} + \frac{1}{2}\partial_x^2\xi^{[0]}(X_T^{[1]})^2 + \int_s^T \Big(\partial_\Theta f^{[0]}(r)\Theta_r^{[2]} + \frac{1}{2}\partial_\Theta^2 f^{[0]}(r)\Theta_r^{[1]}\Theta_r^{[1]}\Big) dr \\ &\quad - \int_s^T Z_r^{[2]} dW_r - \int_s^t \psi_r^{[2]}(z)\widetilde{\mu}(dr,dz) \ . \end{aligned} \tag{8.5}$$

You can see that the dynamics of $X^{[2]}$ is linear in $X^{[2]}$ and contains $\{(X^{[1]})^j, j \le 2\}$. The BSDE for $\hat{\Theta}^{[2]}$ is linear in itself and contains $\{(\Theta^{[1]})^j, j \le 2\}$. Since we have seen $\hat{\Theta}^{[1]}$ is linear in $X^{[1]}$, the driver contains $\{(X^{[1]})^j, j \le 2\}$. Suppose that $\hat{\Theta}^{[2]}$ is linear in $X^{[2]}$ and quadratic in $X^{[1]}$. Then, one can check that this is also the case for the driver of $Y^{[2]}$ and hence consistent with the initial assumption. Although it becomes a bit more tedious, the same technique used in Lemma 8.2 gives the following result:

Lemma 8.3. *There exits a unique solution $\Theta^{[2]}$ to (8.4) and (8.5) which belongs to $\mathbb{S}^p[0,T]^{\otimes 4}$ for $\forall p \ge 2$. $\hat{\Theta}^{[2]}$ is given by, for $s \in [0,T]$ and $z \in \mathbb{R}_0$,*

$$\begin{aligned} Y_s^{[2]} &= y_2^{[2]}(s) X_s^{[2]} + y_{1,1}^{[2]}(s)(X_s^{[1]})^2 + y_1^{[2]}(s) X_s^{[1]} + y_0^{[2]}(s) \\ Z_s^{[2]} &= X_{s-}^{[1]}\Big(y_2^{[2]}(s)\partial_x\sigma^{[0]}(s) + 2y_{1,1}^{[2]}\sigma^{[0]}(s)\Big) + y_1^{[2]}(s)\sigma^{[0]}(s) \\ \psi_s^{[2]}(z) &= X_{s-}^{[1]}\Big(y_2^{[2]}(s)\partial_x\gamma^{[0]}(s,z) + 2y_{1,1}^{[2]}(s)\gamma^{[0]}(s,z)\Big) + y_{1,1}^{[2]}(s)(\gamma^{[0]}(s,z))^2 + y_1^{[2]}(s)\gamma^{[0]}(s,z) \ . \end{aligned}$$

Here, $\Big(y_2^{[2]}(s), y_{1,1}^{[2]}(s), y_1^{[2]}(s), y_0^{[2]}(s), s \in [0,T]\Big)$ are the solutions to the following linear ODEs:

$$\begin{aligned} -\frac{dy_2^{[2]}(s)}{ds} &= \Big(\partial_x b^{[0]}(s) + \partial_y f^{[0]}(s)\Big) y_2^{[2]}(s) + \partial_x f^{[0]}(s) \\ -\frac{dy_{1,1}^{[2]}(s)}{ds} &= \Big(2\partial_x b^{[0]}(s) + \partial_y f^{[0]}(s)\Big) y_{1,1}^{[2]}(s) + \frac{1}{2}\partial_x^2 f^{[0]}(s) \\ &\quad + \frac{1}{2}\partial_x^2 b^{[0]}(s) y_2^{[2]}(s) + \partial_x\partial_y f^{[0]}(s) y_1^{[1]}(s) + \frac{1}{2}\partial_y^2 f^{[0]}(s)(y_1^{[1]}(s))^2 \end{aligned}$$

$$
\begin{aligned}
-\frac{dy_1^{[2]}(s)}{ds} &= \Big(\partial_x b^{[0]}(s)+\partial_y f^{[0]}(s)\Big)y_1^{[2]}(s)+\partial_x\partial_\epsilon b^{[0]}(s)y_2^{[2]}(s)+2\partial_\epsilon b^{[0]}(s)y_{1,1}^{[2]}(s)\\
&\quad+\partial_z f^{[0]}(s)\Big(y_2^{[2]}(s)\partial_x\sigma^{[0]}(s)+2y_{1,1}^{[2]}(s)\sigma^{[0]}(s)\Big)\\
&\quad+\partial_u f^{[0]}(s)\Big(y_2^{[2]}(s)\partial_x\Gamma^{[0]}(s)+2y_{1,1}^{[2]}(s)\Gamma^{[0]}(s)\Big)\\
&\quad+\partial_y^2 f^{[0]}(s)y_1^{[1]}(s)y_0^{[1]}(s)+\partial_x\partial_y f^{[0]}(s)y_0^{[1]}(s)\\
&\quad+y_1^{[1]}(s)\Big(\partial_x\partial_z f^{[0]}(s)\sigma^{[0]}(s)+\partial_x\partial_u f^{[0]}(s)\Gamma^{[0]}(s)\Big)\\
&\quad+(y_1^{[1]}(s))^2\Big(\partial_y\partial_z f^{[0]}(s)\sigma^{[0]}(s)+\partial_y\partial_u f^{[0]}(s)\Gamma^{[0]}(s)\Big)\\
-\frac{dy_0^{[2]}(s)}{ds} &= \partial_y f^{[0]}(s)y_0^{[2]}(s)+y_{1,1}^{[2]}(s)\Big((\sigma^{[0]}(s))^2+\int_{\mathbb{R}_0}(\gamma^{[0]}(s,z))^2\nu(dz)\Big)\\
&\quad+\frac{1}{2}\partial_\epsilon^2 b^{[0]}(s)y_2^{[2]}(s)+\partial_\epsilon b^{[0]}(s)y_1^{[2]}(s)+y_1^{[2]}(s)\Big(\partial_z f^{[0]}(s)\sigma^{[0]}(s)+\partial_u f^{[0]}(s)\Gamma^{[0]}(s)\Big)\\
&\quad+y_{1,1}^{[2]}(s)\partial_u f^{[0]}(s)\int_{\mathbb{R}_0}\rho(z)(\gamma^{[0]}(s,z))^2\nu(dz)+\frac{1}{2}\partial_y^2 f^{[0]}(s)(y_0^{[1]}(s))^2\\
&\quad+(y_1^{[1]}(s))^2\Big(\frac{1}{2}\partial_z^2 f^{[0]}(s)(\sigma^{[0]}(s))^2+\frac{1}{2}\partial_u^2 f^{[0]}(s)(\Gamma^{[0]}(s))^2+\partial_z\partial_u f^{[0]}(s)\sigma^{[0]}(s)\Gamma^{[0]}(s)\Big)\\
&\quad+(y_1^{[1]}(s)y_0^{[1]}(s))\Big(\partial_y\partial_z f^{[0]}(s)\sigma^{[0]}(s)+\partial_y\partial_u f^{[0]}(s)\Gamma^{[0]}(s)\Big)
\end{aligned}
$$

with terminal conditions $y_2^{[2]}(T)=\partial_x\xi^{[0]}$, $y_{1,1}^{[2]}(T)=\frac{1}{2}\partial_x^2\xi^{[0]}$, $y_1^{[2]}(T)=y_0^{[2]}(T)=0$.

One can repeat the procedures to an arbitrary higher order. This can be checked in the following way. By a simple modification of (5.2) gives

$$
Y_s^{[n]}=G_n+\int_s^T\Big\{F_{n,r}+\partial_\Theta f^{[0]}(r)\Theta_r^{[n]}\Big\}dr-\int_s^T Z_r^{[n]}dW_r-\int_s^T\int_{\mathbb{R}_0}\psi_r^{[n]}(z)\widetilde{\mu}(dr,dz)
$$

where

$$
\begin{aligned}
G_n &:= \sum_{k=1}^n\sum_{\beta_1+\cdots+\beta_k=n,\beta_i\geq 1}\frac{1}{k!}\partial_x^k\xi(X_T^{[0]})\prod_{j=1}^k X_T^{[\beta_j]},\\
F_{n,r} &:= \sum_{k=2}^n\sum_{\beta_1+\cdots+\beta_k=n,\beta_i\geq 1}\sum_{i_x=0}^k\sum_{i_y=0}^{k-i_x}\sum_{i_z=0}^{k-i_x-i_y}\frac{\partial_x^{i_x}\partial_y^{i_y}\partial_z^{i_z}\partial_u^{k-i_x-i_y-i_z}f^{[0]}(r)}{i_x!i_y!i_z!(k-i_x-i_y-i_z)!}\\
&\times\prod_{j_x=1}^{i_x}X_r^{[\beta_{j_x}]}\prod_{j_y=i_x+1}^{i_x+i_y}Y_r^{[\beta_{j_y}]}\prod_{j_z=i_x+i_y+1}^{i_x+i_y+i_z}Z_r^{[\beta_{j_z}]}\prod_{j_u=i_x+i_y+i_z+1}^{k}\int_{\mathbb{R}_0}\rho(z)\psi_r^{[\beta_{j_u}]}(z)\nu(dz).
\end{aligned}
$$

From the shapes of $G_n, F_{n,r}$, one can confirm that $\hat{\Theta}_r^{[n]}$ is given by the polynomials

$$
\left\{\prod_{j=1}^k X_r^{[\beta_j]};\beta_1+\cdots+\beta_k=m\ (\beta_i\geq 1),\ k\leq m,\ m\leq n\right\}
$$

by induction. Since $\Theta^{[n]}$ appears only linearly both in the forward and backward SDEs the relevant ODEs become always linear.

9 A polynomial expansion

In the last section, the grading structure both for $\{X^{[n]}\}_{n\geq 0}$ and $\{\hat{\Theta}^{[n]}\}_{n\geq 0}$ played an important role. In particular, even if $\{\hat{\Theta}^{[n]}\}_{n\geq 0}$ has a grading structure, one cannot obtain the system of linear ODEs unless $\{X^{[n]}\}_{n\geq 0}$ share the same features. Suppose that the dynamics of $X^{t,x}$ is linear in itself. Then, one need not expand the forward SDE and thus can obtain the expansion of $\hat{\Theta}^{t,x,\epsilon}$ in terms of polynomials of $X^{t,x}$. If this is the case, the ODEs for the associated coefficients required in each order will be greatly simplified.

Let us consider the following forward-backward SDEs for $s\in[t,T]$:

$$
\begin{aligned}
X_s^{t,x} &= x+\int_t^s\Big(b^0(r)+b^1(r)X_r^{t,x}\Big)dr+\int_t^s\Big(\sigma^0(r)+\sigma^1(r)X_r^{t,x}\Big)dW_r\\
&\quad+\int_s^t\int_E\Big(\gamma^0(r,z)+\gamma^1(r,z)X_{r-}^{t,x}\Big)\widetilde{\mu}(dr,dz)\\
Y_s^{t,x,\epsilon} &= \xi(\epsilon X_T^{t,x})+\int_s^T f\Big(r,\epsilon X_r^{t,x},Y_r^{t,x,\epsilon},Z_r^{t,x,\epsilon},\int_{\mathbb{R}_0}\rho(z)\psi_r^{t,x,\epsilon}(z)\nu(dz)\Big)dr\\
&\quad-\int_s^T Z_r^{t,x,\epsilon}dW_r-\int_s^T\int_E\psi_r^{t,x,\epsilon}(z)\widetilde{\mu}(dr,dz)\ . \qquad (9.1)
\end{aligned}
$$

where $b^0:[0,T]\to\mathbb{R}^d$, $b^1:[0,T]\to\mathbb{R}^{d\times d}$, $\sigma^0:[0,T]\to\mathbb{R}^{d\times l}$, $\sigma^1:[0,T]\to\mathbb{R}^{d\times d\times l}$, $\gamma^0:[0,T]\times E\to\mathbb{R}^{d\times k}$, $\gamma^1:[0,T]\times E\to\mathbb{R}^{d\times d\times k}$ and ξ,f are defined as before.

Assumption 9.1. *The functions $\{b^i(t),\sigma^i(t),\gamma^i(t,z)\}$, $i\in\{0,1\}$ are continuous in t. Furthermore, there exists some positive constant K such that $\Big(|b^i(t)|+|\sigma^i(t)|+|\gamma^i(t,z)|/\eta(z)\leq K\Big)$ for $i\in\{0,1\}$ uniformly in $(t,z)\in[0,T]\times E$.*

With slight abuse of notation, let us use $\Theta_r^{t,x,\epsilon}:=\Big(\epsilon X_r^{t,x},Y_r^{t,x,\epsilon},Z_r^{t,x,\epsilon},\int_{\mathbb{R}_0}\rho(z)\psi_r^{t,x,\epsilon}(z)\nu(dz)\Big)$ in this section.

Theorem 9.1. *Under Assumptions 3.2 and 9.1, there exists a unique solution $\hat{\Theta}^{t,x,\epsilon}$ to the BSDE (9.1) and it is classically differentiable arbitrary many times with respect to ϵ. For every $n\geq 1$, $\partial_\epsilon^n\hat{\Theta}^{t,x,\epsilon}$ is given by the solution to the BSDE*

$$
\begin{aligned}
\partial_\epsilon^n Y_s^{t,x,\epsilon} &= g_n(X_T^{t,x})^n+\int_s^T\Big\{h_{n,r}+\partial_x^n f(r,\Theta_r^{t,x,\epsilon})(X_r^{t,x})^n+\partial_{\hat{\Theta}}f(r,\Theta_r^{t,x,\epsilon})\partial_\epsilon^n\hat{\Theta}_r^{t,x,\epsilon}\Big\}dr\\
&\quad-\int_s^T\partial_\epsilon^n Z_r^{t,x,\epsilon}dW_r-\int_s^T\int_E\partial_\epsilon^n\psi_r^{t,x,\epsilon}(z)\widetilde{\mu}(dr,dz)
\end{aligned}
$$

where $g_n:=\partial_x^n\xi(\epsilon X_T^{t,x})$ and

$$
\begin{aligned}
h_{n,r} &:= n!\sum_{k=2}^n\sum_{i_x=0}^{k-1}\sum_{i_y=0}^{k-i_x}\sum_{i_z=0}^{k-i_x-i_y}\sum_{\beta_{i_x+1}+\cdots+\beta_k=n-i_x,\ \beta_i\geq 1}\frac{\partial_x^{i_x}\partial_y^{i_y}\partial_z^{i_z}\partial_u^{k-i_x-i_y-i_z}f(r,\Theta_r^{t,x,\epsilon})}{i_x!i_y!i_z!(k-i_x-i_y-i_z)!}(X_r^{t,x})^{i_x}\\
&\times\prod_{j_y=i_x+1}^{i_x+i_y}\frac{1}{\beta_{j_y}!}\partial_\epsilon^{\beta_{j_y}}Y_r^{t,x,\epsilon}\prod_{j_z=i_x+i_y+1}^{i_x+i_y+i_z}\frac{1}{\beta_{j_z}!}\partial_\epsilon^{\beta_{j_z}}Z_r^{t,x,\epsilon}\prod_{j_u=i_x+i_y+i_z+1}^{k}\frac{1}{\beta_{j_u}!}\int_{\mathbb{R}_0}\rho(z)\partial_\epsilon^{\beta_{j_u}}\psi_r^{t,x,\epsilon}(z)\nu(dz)
\end{aligned}
$$

and satisfies $\partial_\epsilon^n\hat{\Theta}^{t,x,\epsilon}\in\mathbb{S}^p[t,T]^{\otimes 3}$ for $\forall p\geq 2$. Moreover, the asymptotic expansion of $\hat{\Theta}^{t,x,\epsilon}$

with respect to ϵ satisfies, with some positive constant C_p, that

$$\left|\left|\hat{\Theta}^{t,x,\epsilon} - \Big(\hat{\Theta}^{[0]} + \sum_{n=1}^{N} \epsilon^n \hat{\Theta}^{[n]}\Big)\right|\right|^p_{\mathbb{S}^p[t,T]} \le \epsilon^{p(N+1)} C_p.$$

Proof. One can follow the same arguments used to derive Proposition 6.1 and Theorem 6.1 by replacing $(X^{t,x,\epsilon})$ by $(\epsilon X^{t,x})$. Since there is no ϵ-dependence through $X^{t,x}$ in the expressions $Y_s^{t,x,\epsilon} = u(s, X_s^{t,x}, \epsilon)$ and $Z_s^{t,x,\epsilon} = \partial_x u(x, X_{s-}^{t,x}, \epsilon)\sigma(s, X_{s-}^{t,x}, \epsilon)$, one-time differentiability with respect to x and its polynomial growth property are enough to show recursively that $\partial_\epsilon^n \hat{\Theta}^{t,x,\epsilon} \in \mathbb{S}^p[t,T]$ for $\forall p \ge 2$. □

The above result also justifies the method proposed in Fujii (2015) [18] for the underlying X having a linear dynamics. For a general Affine-like process X (such as $\sigma(x) = \sqrt{x}$), it is difficult to prove within the current technique due to its non-Lipschitz nature.

It is not difficult to see that $(\hat{\Theta}_s^{[n]}, s \in [t,T])$ is given by the unique solution to the following BSDE:

$$\begin{aligned} Y_s^{[n]} &= \frac{1}{n!}\partial_x^n \xi(0)(X_T^{t,x})^n + \int_s^T \Big\{ \widetilde{h}_{n,r} + \frac{1}{n!}\partial_x^n f^{[0]}(r)(X_r^{t,x})^n + \partial_{\hat{\Theta}} f^{[0]}(r)\hat{\Theta}_r^{[n]} \Big\} dr \\ &\quad - \int_s^T Z_r^{[n]} dW_r - \int_s^T \int_E \psi_r^{[n]}(z)\widetilde{\mu}(dr, dz) \end{aligned} \tag{9.2}$$

where

$$\begin{aligned} \widetilde{h}_{n,r} &:= \sum_{k=2}^{n} \sum_{i_x=0}^{k-1} \sum_{i_y=0}^{k-i_x} \sum_{i_z=0}^{k-i_x-i_y} \sum_{\beta_{i_x+1}+\cdots+\beta_k = n-i_x, \beta_i \ge 1} \frac{\partial_x^{i_x}\partial_y^{i_y}\partial_z^{i_z}\partial_u^{k-i_x-i_y-i_z} f^{[0]}(r)}{i_x! i_y! i_z! (k-i_x-i_y-i_z)!} (X_r^{t,x})^{i_x} \\ &\times \prod_{j_y=i_x+1}^{i_x+i_y} Y_r^{[\beta_{j_y}]} \prod_{j_z=i_x+i_y+1}^{i_x+i_y+i_z} Z_r^{[\beta_{j_z}]} \prod_{j_u=i_x+i_y+i_z+1}^{k} \int_{\mathbb{R}_0} \rho(z)\psi_r^{[\beta_{j_u}]}(z)\nu(dz) \end{aligned}$$

and $f^{[0]}(r) := f(r, 0, Y_r^{[0]}, 0, 0)$. Since $(i_x + \sum_{j_y}\beta_{j_y} + \sum_{j_z}\beta_{j_z} + \sum_{j_u}\beta_{j_u}) = n$, one can recursively show that $\hat{\Theta}_r^{[n]}$ is given by the polynomials $\Big\{(X_r^{t,x})^j, 0 \le j \le n\Big\}$ and every coefficient is determined by the system of linear ODEs as in Section 8, which we leave as a simple exercise.

An exponential Lévy case

In the reminder of this section, let us deal with a special example of an exponential (time-inhomogeneous) Lévy dynamics for X. Let us put $m = d = l = k = 1$ and $t = 0$ for simplicity and consider $b^0 = \sigma^0 = \gamma^0 = 0$

$$X_s = x + \int_t^s X_r\Big(b(r)dr + \sigma(r)dW_r\Big) + \int_s^t \int_{\mathbb{R}_0} X_{r-}\gamma(r,z)\widetilde{\mu}(dr,dz) \tag{9.3}$$

with $b = b^1, \sigma = \sigma^1, \gamma = \gamma^1$. We omit the superscript denoting the initial data $(0, x)$.

Let us introduce the notations: $q(s,j) := \int_{\mathbb{R}_0}(\gamma(s,z))^j \nu(dz)$ for $j \ge 2$, $\Gamma(s,j) := \int_{\mathbb{R}_0}\rho(z)\big[(1+\gamma(s,z))^j - 1]\nu(dz)$ for $j \ge 1$ and $C_{n,j} := n!/(j!(n-j)!)$ for $j \le n, n \ge 2$.

Theorem 9.2. *Under Assumptions 3.1, 9.1, $m = d = l = k = 1$ and $t = 0$, the asymptotic* *expansion of the forward-backward SDEs (9.3) and (9.1) is given by, for $s \in [0,T]$,*

$$\begin{aligned} Y_s^{[0]} &= \xi(0) + \int_s^T f(r,0,Y_r^{[0]},0,0)dr \\ Z^{[0]} &= \psi^{[0]} = 0 \end{aligned} \tag{9.4}$$

and, for $n \geq 1$,

$$\begin{aligned} Y_s^{[n]} &= (X_s)^n y^{[n]}(s) \\ Z_s^{[n]} &= (X_{s-})^n y^{[n]}(s) n\sigma(s) \\ \psi_s^{[n]}(z) &= (X_{s-})^n y^{[n]}(s)\big[(1+\gamma(s,z))^n - 1\big] \end{aligned}$$

where the functions $\{y^{[j]}(s), s \in [0,T]\}_{1\leq j\leq n}$ are determined recursively by the following system of linear ODEs:

$$\begin{aligned} -\frac{dy^{[n]}(s)}{ds} \;&=\; \Big(nb(s) + \frac{1}{2}n(n-1)\sigma^2(s) + \sum_{j=2}^n C_{n,j}q(s;j) + \partial_y f^{[0]}(s) \\ &+\partial_z f^{[0]}(s)n\sigma(s) + \partial_u f^{[0]}(s)\Gamma(s;n)\Big)y^{[n]}(s) + \frac{1}{n!}\partial_x^n f^{[0]}(s) \\ &+\sum_{k=2}^n \sum_{i_x=0}^{k-1} \sum_{i_y=0}^{k-i_x-i_y} \sum_{\beta_{i_x+1}+\cdots+\beta_k = n-i_x, \beta_i\geq 1} \Bigg\{ \frac{\partial_x^{i_x}\partial_y^{i_y}\partial_z^{i_z}\partial_u^{k-i_x-i_y-i_z} f^{[0]}(s)}{i_x!i_y!i_z!(k-i_x-i_y-i_z)!} \\ &\times \prod_{j_y=i_x+1}^{i_x+i_y}\Big(y^{[\beta_{j_y}]}(s)\Big) \prod_{j_z=i_x+i_y+1}^{i_x+i_y+i_z}\Big(\beta_{j_z}\sigma(s)y^{[\beta_{j_z}]}(s)\Big) \\ &\times \prod_{j_u=i_x+i_y+i_z+1}^{k}\Big(\Gamma(s;\beta_{j_u})y^{[\beta_{j_u}]}(s)\Big)\Bigg\} \end{aligned}$$

with a terminal condition $y^{[n]}(T) = \partial_x^n\xi(0)/n!$ for every $n \geq 1$. Here, $f^{[0]}(r)$ is defined by $f(r,0,Y_r^{[0]},0,0)$ using $Y^{[0]}$ determined by (9.4).

Proof. If one supposes the form of the solution as $Y_s^{[n]} = (X_s)^n y^{[n]}(s)$, then $Z^{[n]}$ and $\psi^{[n]}$ must have the form as given. Comparing the result of Itô formula applied to $X^n y^{[n]}$ and the form of the BSDE (9.2) substituted by the hypothesized form of $\{\hat{\Theta}^{[\beta]}\}_{\beta\leq n}$, one obtains the system of ODEs given above. Since every ODE is linear, there exists a solution for every $y^{[n]}$, $n \geq 1$. Since the solution of the BSDE is unique, this must be the desired solution. □

Remark

It is interesting to observe the difference from the linearization scheme proposed in [20] for a Brownian setup. There, the BSDE is expanded around a linear driver in the first step. Then, in the second step, the resultant set of linear BSDEs are evaluated by the small-variance asymptotic expansion of the forward SDE, or by the interacting particle simulation method proposed in Fujii & Takahashi (2015) [21]. Thus, in order for the scheme of [20] works well, it requires the smallness of the non-linear terms in the driver f, although it naturally arises in many applications. Furthermore, due to the presence of large number of conditional expectations, calculating them analytically without involving the particle simulation technique [21] is unrealistic in most of the practical situations.

On the other hand, in the current scheme, the expansion of the driver is not directly performed and the significant part of non-linearity is taken into account at the zero-th order around the mean dynamics of the forward SDE as observed in (8.1). The effects of the stochasticity from the forward SDE are then taken into account perturbatively around this "mean" solution. Therefore, the current scheme is expected to be more advantageous when there exists significant non-linearity in the driver. Furthermore, the special grading structure of approximating FBSDEs makes them explicitly solvable by ODEs without any use of simulation. Since the approximate solution of $(Y, Z, \psi(\cdot))$ is explicitly given as a function of the stochastic flows of X with non-random coefficients fixed by ODEs, one can obtain not only the current value $(Y_0, Z_0, \psi_0(\cdot))$ but also its evolution by simply simulating the flows of X (or X itself for the polynomial case). Some numerical examples and error estimates are available in Fujii (2015) [18] based on this property for a certain class of models.

10 Justification to use smooth coefficients

In this section, we provide a justification to use smooth coefficients in the forward-backward SDEs for any numerical approximation purpose. Since ϵ is a perturbation parameter, we can always introduce it so that all the functions depend smoothly on ϵ. This is actually the case for the examples used in Sections 8 and 9. Thus, in this section, we concentrate on the other parameters and omit ϵ dependence from the functions. Let us first consider the forward component:

$$\widetilde{X}_s = x + \int_t^s \widetilde{b}(r, \widetilde{X}_r)dr + \int_t^s \widetilde{\sigma}(r, \widetilde{X}_r)dW_r + \int_t^s \int_E \widetilde{\gamma}(r, \widetilde{X}_r, z)\widetilde{\mu}(dr, dz) \ , \tag{10.1}$$

where $x \in \mathbb{R}^d$, $\widetilde{b} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$, $\widetilde{\sigma} : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times l}$, $\widetilde{\gamma} : [0,T] \times \mathbb{R}^d \times E \to \mathbb{R}^{d\times k}$. We omit the superscripts denoting the initial data (t, x).

Assumption 10.1. *$\widetilde{b}, \widetilde{\sigma}, \widetilde{\gamma}$ are continuous in (t, x, z). There exists some positive constant K such that, for every $x, x' \in \mathbb{R}^d$,*
(i) $|\widetilde{b}(t, x) - \widetilde{b}(t, x')| + |\widetilde{\sigma}(t, x) - \widetilde{\sigma}(t, x')| \le K|x - x'|$ uniformly in $t \in [0, T]$,
(ii) $|\widetilde{\gamma}_j(t, x, z) - \widetilde{\gamma}_j(t, x', z)| \le K\eta(z)|x - x'|$ for $1 \le j \le k$ uniformly in $(t, z) \in [0, T] \times \mathbb{R}_0$,
(iii) $||\widetilde{b}(\cdot, 0)||_T + ||\widetilde{\sigma}(\cdot, 0)||_T + ||\widetilde{\gamma}(\cdot, 0, z)||_T/\eta(z) \le K$ uniformly in $z \in E$.

The regularization technique by the convolution with appropriate mollifiers gives us the following approximating functions.

Lemma 10.1. *Under Assumption 10.1, one can choose a sequence of functions $b_n : [0,T]\times \mathbb{R}^d \to \mathbb{R}^d$, $\sigma_n : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{d\times l}$, $\gamma_n : [0,T] \times \mathbb{R}^d \times E \to \mathbb{R}^{d\times k}$ with $n \in \mathbb{N}$, which are continuous in all their arguments, continuously differentiable arbitrary many times with respect to $x \in \mathbb{R}^d$ and also satisfy, for every $n \ge 1$;*
(i) for every $m \ge 1$, $|\partial_x^m b_n(t, x)| + |\partial_x^m \sigma_n(t, x)| + |\partial_x^m \gamma_n(t, x, z)|/\eta(z)$ is uniformly bounded in $(t, x, z) \in [0, T] \times \mathbb{R}^d \times E$
(ii) for every $(t, x, z) \in [0, T] \times \mathbb{R}^d \times E$, $b_n(t, x), \sigma_n(t, x)$ and $\gamma_n(t, x, z)$ converge pointwise to $\widetilde{b}(t, x)$, $\widetilde{\sigma}(t, x)$ and $\widetilde{\gamma}(t, x, z)$, respectively,
(iii) $(b_n, \sigma_n, \gamma_n)$ satisfy the properties in Assumption 10.1 with some positive constant K' independent of n.

Proof. We consider a sequence of (symmetric) mollifiers $\varrho_n \in \mathcal{C}_0^\infty : \mathbb{R}^d \to \mathbb{R}_+$ with compact support satisfying $\int_{\mathbb{R}^d} \varrho_n(x)dx = 1$ and $\varrho_n(x) \to \delta(x)$ as $n \to \infty$ in the space of Schwartz distributions, where $\delta(\cdot)$ is a Dirac delta function. Let us define intermediate mollified functions as

$$\bar{b}_n(t,x) := \varrho_n * \widetilde{b}(t,x), \quad \bar{\sigma}_n(t,x) := \varrho_n * \widetilde{\sigma}(t,x), \quad \bar{\gamma}_n(t,x,z) := \varrho_n * \widetilde{\gamma}(t,x,z)$$

where $*$ denotes a convolution with respect to x, such as

$$\bar{b}_n(t,x) = \int_{\mathbb{R}^d} \varrho_n(x-y)\widetilde{b}(t,y)dy = \int_{\mathbb{R}^d} \widetilde{b}(t,x-y)\varrho_n(y)dy \ .$$

Since $\widetilde{b}, \widetilde{\sigma}, \widetilde{\gamma}$ are continuous, every point $x \in \mathbb{R}^d$ is a Lebesgue point. Thus, the approximated functions $\bar{b}_n, \bar{\sigma}_n, \bar{\gamma}_n$ are known to converge pointwise to $\widetilde{b}, \widetilde{\sigma}, \widetilde{\gamma}$ from the Lebesgue differentiation theorem (see, for example, Theorem 8.7 in Igari (1996) [26] or Theorem C.19 in Leoni (2009) [30]). The Lipschitz property can be shown as, for every $x, x' \in \mathbb{R}^d$,

$$\begin{aligned} |\bar{\gamma}_{n,j}(t,x,z) - \bar{\gamma}_{n,j}(t,x',z)| &\leq \int_{\mathbb{R}^d} |\widetilde{\gamma}_j(t,x-y,z) - \widetilde{\gamma}_j(t,x'-y,z)|\varrho_n(y)dy \\ &\leq K\eta(z)|x-x'| \int_{\mathbb{R}^d} \varrho_n(y)dy = K|x-x'|\eta(z) \end{aligned}$$

and similarly for the others. It is easy to see that there exists some positive constant C' satisfying

$$||\bar{b}_n(\cdot,0)||_T + ||\bar{\sigma}_n(\cdot,0)||_T + ||\bar{\gamma}_n(\cdot,0,z)||_T/\eta(z) \leq C'$$

uniformly in $z \in E$ as well as $n \in \mathbb{N}$ since ϱ_n has a compact support shrinking to the origin as $n \to \infty$. We prepare another (symmetric) mollifiers $\varsigma_n \in \mathcal{C}_0^\infty : \mathbb{R}^d \times E \to \mathbb{R}_+$ in the following way:

$$\varsigma_n(x,z) = \begin{cases} 1 & \text{for } |x| + |z| \leq n \\ 0 & \text{for } |x| + |z| \geq 2n \end{cases} \ . \tag{10.2}$$

We then construct the mollified functions as

$$b_n(t,x) = \varsigma_n(x,0)\bar{b}_n(t,x), \quad \sigma_n(t,x) = \varsigma_n(x,0)\bar{\sigma}_n(t,x), \quad \gamma_n(t,x,z) = \varsigma_n(x,z)\bar{\gamma}_n(t,x,z) \ .$$

Then, for each n, they have bounded derivatives of all orders with respect to x uniformly in (t,x,z). The pointwise convergence is clearly preserved. Lastly, one has to check that there exists a Lipschitz constant K' independent of n. By the construction in (10.2), one can arrange the mollifier in the following way: there exists a positive constant C such that

$$\sup_{(x,z) \in \mathbb{R}^d \times E} \Big| \partial_x \varsigma_n(x,z) \Big| \leq C/n$$

for every $n \in \mathbb{N}$. Then, for $\forall n \in \mathbb{N}$, one sees

$$\begin{aligned} |\partial_x \gamma_n(t,x,z)| &\leq \varsigma_n(x,z)|\partial_x \bar{\gamma}_n(t,x,z)| + |\partial_x \varsigma_n(x,z)||\bar{\gamma}_n(t,x,z)| \\ &\leq K\eta(z) + \eta(z)C/n(C' + K(2n)) \leq K'\eta(z) \end{aligned}$$

uniformly in (t,x,z). Here, we have used the fact that $\partial_x \varsigma_n(x,z)$ vanishes when $|x| \geq 2n$ and the linear growth property of $\bar{\gamma}_n$. One can similarly check $|\partial_x b_n(t,x)|, |\partial_x \sigma_n(t,x)| \leq K'$

for $\forall n \in \mathbb{N}$. The property (iii) of Assumption 10.1 is obviously preserved in the second mollification. □

This yields the following result.

Proposition 10.1. *Under Assumption 10.1, consider the process* $\widetilde{X}$ *of (10.1) and the sequence of processes* $(X^n_s, s \in [t,T])_{n\geq 1}$ *defined by*

$$X^n_s = x + \int_t^s b_n(r, X^n_r)dr + \int_t^s \sigma_n(r, X^n_r)dW_r + \int_t^s \int_E \gamma_n(r, X^n_r, z)\widetilde{\mu}(dr,dz) \qquad (10.3)$$

with b_n, σ_n *and* γ_n *given in Lemma 10.1. Then, there exist unique solutions* $\widetilde{X}, X^n$ *in* $\mathbb{S}^p$ *for* $\forall p \geq 2$*. Moreover, the following relation holds*

$$\lim_{n\to\infty} \mathbb{E}\Big[||\widetilde{X} - X^n||^p_{[t,T]}\Big] = 0$$

for $\forall p \geq 2$*.*

Proof. The existence of the unique solution for (10.1) as well as (10.3) in $\mathbb{S}^p$ for $\forall p \geq 2$ is clear from Lemma A.3. We also have, for $\forall p \geq 2$,

$$||\widetilde{X} - X^n||^p_{\mathbb{S}^p} \leq C_p \mathbb{E}\left[\Big(\int_t^T |\delta\widetilde{b}_n(r,\widetilde{X}_r)|dr\Big)^p + \Big(\int_t^T |\delta\widetilde{\sigma}_n(r,\widetilde{X}_r)|^2 dr\Big)^{p/2} + \int_t^T |\delta L^n_r|^p dr\right]$$

where $\delta\widetilde{b}_n := \widetilde{b} - b_n$, $\delta\widetilde{\sigma}_n := \widetilde{\sigma} - \sigma_n$. Furthermore δL^n is a predictable process satisfying $|\delta\widetilde{\gamma}_n|(t, \widetilde{X}_{t-}, z) \leq \delta L^n_t \eta(z)$, $d\mathbb{P}\otimes dt$-a.e. in $\Omega \times [0,T]$, where $\delta\widetilde{\gamma}_n := \widetilde{\gamma} - \gamma_n$. We can take δL^n such that $\int_t^T \mathbb{E}|\delta L^n_r|^p dr < \infty$, since we have $|\delta\widetilde{\gamma}_n|(s, \widetilde{X}_{s-}, z) \leq 2K(1 + |\widetilde{X}_{s-}|)\eta(z)$ in the current setup. See also the related discussion in Lemma A.3.

Note that C_p is independent of n thanks to Lemma 10.1 (iii). Due to the linear growth property, the inside of the expectation is dominated by $C(1 + ||\widetilde{X}||^p_{[t,T]})$ with some positive constant C independent of n. From Lemma 10.1 (ii), $(\delta\widetilde{b}_n, \delta\widetilde{\sigma}_n, \delta\widetilde{\gamma}_n)$ converge pointwise to zero. Thus, one can also take a sequence of $(\delta L^n, n \in \mathbb{N})$ converging pointwise to zero. Since $\widetilde{X} \in \mathbb{S}^p$ for $\forall p \geq 2$, the dominated convergence theorem give the desired result in the limit $n \to \infty$. [3] □

The above result implies that by choosing a large enough n one can work on X^n that is an arbitrary accurate approximation in the $\mathbb{S}^p$ sense of the original process $\widetilde{X}$, and involves only smooth coefficients $(b_n, \sigma_n, \gamma_n)$. This conclusion can be extended to the forward-backward system. Consider the BSDE driven by $\widetilde{X}$;

$$\begin{aligned}\widetilde{Y}_s = \widetilde{\xi}(\widetilde{X}_T) &+ \int_s^T \widetilde{f}\Big(r, \widetilde{X}_r, \widetilde{Y}_r, \widetilde{Z}_r, \int_{\mathbb{R}_0} \rho(z)\widetilde{\psi}_r(z)\nu(dz)\Big)dr \\ &- \int_s^T \widetilde{Z}_r dW_r - \int_s^T \int_E \widetilde{\psi}_r(z)\widetilde{\mu}(dr,dz) \qquad (10.4)\end{aligned}$$

for $s \in [t,T]$ where $\widetilde{\xi} : \mathbb{R}^d \to \mathbb{R}^m$, $\widetilde{f} : [0,T] \times \mathbb{R}^d \times \mathbb{R}^m \times \mathbb{R}^{m\times l} \times \mathbb{R}^{m\times k} \to \mathbb{R}^m$ and ρ is defined as before.

[3] In $p=2$, one can see more directly $||\widetilde{X} - X^n||^2_{\mathbb{S}^2} \to 0$ since the integral of δL^n can be replaced by that of $\delta\widetilde{\gamma}_n$ (See a remark below Lemma A.3.). Taking an appropriate subsequence if necessary, one can also show that $(X^n_s, s \in [t,T])_{n\geq 1}$ is almost surely uniformly convergent to $(\widetilde{X}_s, s \in [t,T])$ by the Borel-Cantelli lemma.

Assumption 10.2. *The functions $\widetilde{\xi}$ and $\widetilde{f}$ are continuous in all their arguments. There exists some positive constants $K, q \geq 0$ such that*
(i)$|\widetilde{\xi}(x)| + |\widetilde{f}(t,x,0,0,0)| \leq K(1+|x|^q)$ for every $x \in \mathbb{R}^d$ uniformly in $t \in [0,T]$.
(ii) $|\widetilde{f}(t,x,y,z,u) - \widetilde{f}(t,x,y',z',u')| \leq K(|y-y'|+|z-z'|+|u-u'|)$ for every $(y,z,u),(y',z',u') \in \mathbb{R}^m \times \mathbb{R}^{m\times l} \times \mathbb{R}^{m\times k}$ uniformly in $(t,x) \in [0,T]\times\mathbb{R}^d$.

Lemma 10.2. *Under Assumption 10.2, one can choose a sequence of functions $\xi_n : \mathbb{R}^d \to \mathbb{R}^m$, $f_n : [0,T]\times\mathbb{R}^d\times\mathbb{R}^m\times\mathbb{R}^{m\times l}\times\mathbb{R}^{m\times k} \to \mathbb{R}^m$ with $n \in \mathbb{N}$, which are continuous in all their arguments and satisfy, for every $n \geq 1$;*
(i) (ξ_n, f_n) are continuously differentiable arbitrary many times with respect to (x,y,z,u) with derivatives bounded uniformly in $(t,x,y,z,u) \in [0,T]\times\mathbb{R}^d\times\mathbb{R}^m\times\mathbb{R}^{m\times l}\times\mathbb{R}^{m\times k}$,
(ii) for every $(t,x,y,z,u) \in [0,T]\times\mathbb{R}^d\times\mathbb{R}^m\times\mathbb{R}^{m\times l}\times\mathbb{R}^{m\times k}$, ξ_n and f_n converge pointwise to $\widetilde{\xi}$ and $\widetilde{f}$, respectively,
(iii) (ξ_n, f_n) satisfy Assumption 10.2 with some positive constant K'' independent of n.

Proof. The first step of the mollification can be done exactly the same way as in Lemma 10.1, which gives us $\bar{\xi}_n(x)$ and $\bar{f}_n(t,x,\hat{\Theta}) := \bar{f}_n(t,x,y,z,u)$. In order to achieve the property (iii), one has to take care of the polynomial growth property of the driver with respect to x. One can take the second sequence of mollifiers as

$$\varsigma_n(x,\hat{\Theta}) = \begin{cases} 1 & \text{for } |x|^q + |\hat{\Theta}| \leq n \\ 0 & \text{for } |x|^q + |\hat{\Theta}| \geq 2n \end{cases}$$

and then control their first derivatives, with some positive constant C, by

$$\sup_{(x,\hat{\Theta})\in\mathbb{R}^d\times\mathbb{R}^{m(1+l+k)}} |\partial_{\hat{\Theta}}\varsigma_n(x,\hat{\Theta})| \leq C/n$$

for $\forall n \in \mathbb{N}$. Then, one can check that

$$\xi_n(x) := \varsigma_n(x,0)\bar{\xi}_n(x), \quad f_n(t,x,\hat{\Theta}) := \varsigma_n(x,\hat{\Theta})\bar{f}_n(t,x,\hat{\Theta})$$

satisfy the desired property similarly as in Lemma 10.1. □

Finally, we get the main result of this section:

Proposition 10.2. *Under Assumption 10.2, consider the process $(\widetilde{Y}, \widetilde{Z}, \widetilde{\psi})$ of (10.4) and the sequence of processes $(Y_s^{n,m}, Z_s^{n,m}, \psi_s^{n,m})_{s\in[t,T]}, (n,m) \in \{1,2,\cdots\}$ defined as the solution to following BSDE*

$$\begin{aligned} Y_s^{n,m} &= \xi_m(X_T^n) + \int_s^T f_m\Big(r, X_r^n, Y_r^{n,m}, Z_r^{n,m}, \int_{\mathbb{R}_0} \rho(z)\psi_r^{n,m}(z)\nu(dz)\Big)dr \\ &\quad - \int_s^T Z_r^{n,m} dW_r - \int_s^T \int_E \psi_r^{n,m}(z)\widetilde{\mu}(dr,dz) \end{aligned} \tag{10.5}$$

where X^n is the solution of (10.3) and (ξ_m, f_m) are the mollified functions given in Lemma 10.2. Then, there exist unique solutions $(\widetilde{Y}, \widetilde{Z}, \widetilde{\psi})$, $(Y^{n,m}, Z^{n,m}, \psi^{n,m})_{n,m\geq 1}$ in $\mathcal{K}^p[t,T]$ for $\forall p \geq 2$. Moreover the following relation holds

$$\lim_{m\to\infty}\lim_{n\to\infty}\Big|\Big|(\delta Y^{n,m}, \delta Z^{n,m}, \delta\psi^{n,m})\Big|\Big|^p_{\mathcal{K}^p[t,T]} = 0 \ ,$$

for $\forall p \geq 2$, where $\delta Y^{n,m} := \widetilde{Y} - Y^{n,m}$, $\delta Z^{n,m} := \widetilde{Z} - Z^{n,m}$ and $\delta\psi^{n,m} := \widetilde{\psi} - \psi^{n,m}$.

Proof. The existence of the unique solution $(\widetilde{Y},\widetilde{Z},\widetilde{\psi})$ and $(Y^{n,m},Z^{n,m},\psi^{n,m})$ in $\mathcal{K}^p$ for $\forall p\geq 2$ is clear from Lemma B.2. We have, for $\forall p\geq 2$,

$$\Big|\Big|(\delta Y^{n,m},\delta Z^{n,m},\delta\psi^{n,m})\Big|\Big|^p_{\mathcal{K}^p[t,T]}\leq C_p\mathbb{E}\left[|\delta\xi^{n,m}|^p+\Big(\int_t^T|\delta f^{n,m}(r)|dr\Big)^p\right]$$

by the stability result, where $\delta\xi^{n,m}:=\widetilde{\xi}(\widetilde{X}_T)-\xi_m(X_T^n)$ and

$$\begin{aligned}\delta f^{n,m}(r) &:= \widetilde{f}\Big(r,\widetilde{X}_r,\widetilde{Y}_r,\widetilde{Z}_r,\int_{\mathbb{R}_0}\rho(z)\widetilde{\psi}_r(z)\nu(dz)\Big)\\ &\quad -f_m\Big(r,X_r^n,\widetilde{Y}_r,\widetilde{Z}_r,\int_{\mathbb{R}_0}\rho(z)\widetilde{\psi}_r(z)\nu(dz)\Big)\ .\end{aligned}$$

Firstly, let us fix m. Since $\partial_x\xi_m$ and $\partial_x f_m$ are bounded, the result of Proposition 10.1 yields

$$\lim_{n\to\infty}\Big|\Big|(\delta Y^{n,m},\delta Z^{n,m},\delta\psi^{n,m})\Big|\Big|^p_{\mathcal{K}^p[t,T]}\leq C_p\mathbb{E}\left[|\delta\xi^m|^p+\Big(\int_t^T|\delta f^m(r,\widetilde{\Theta}_r)|dr\Big)^p\right]$$

with $\delta\xi^m:=\widetilde{\xi}(\widetilde{X}_T)-\xi_m(\widetilde{X}_T)$ and $\delta f^m(r,\widetilde{\Theta}_r):=(\widetilde{f}-f_m)\Big(r,\widetilde{X}_r,\widetilde{Y}_r,\widetilde{Z}_r,\int_{\mathbb{R}_0}\rho(z)\widetilde{\psi}_r(z)\nu(dz)\Big)$. Since $\widetilde{\Theta}\in\mathbb{S}^p\times\mathcal{K}^p$ for $\forall p\geq 2$ and $(\widetilde{f},f_m)$ have the linear growth in (y,z,u) and the polynomial growth in x with proportional coefficients independent of m, passing to the limit $m\to\infty$ yields the desired result from the pointwise convergence of the mollified functions and the dominated convergence theorem. Notice also that one can achieve the same convergence with the flipped order of limits $\lim_{n\to\infty}\lim_{m\to\infty}$ by using the fact that $(X_s^n, s\in[t,T])_{n\in\mathbb{N}}$ is almost surely uniformly convergent to $(\widetilde{X}_s,s\in[t,T])$ by taking an appropriate subsequence if necessary. $\square$

Propositions 10.1 and 10.2 imply that one can work on the process Θ^n defined by the smooth coefficients $(b_n,\sigma_n,\gamma_n,\xi_n,f_n)$ as an arbitrary accurate approximation in the $\mathbb{S}^p\times\mathcal{K}^p$ sense of the original one $\widetilde{\Theta}$, which only satisfies Assumptions 10.1 and 10.2. In fact, we can weaken the assumptions further. There is no difficulty to add discontinuities to $\widetilde{\xi}$ and $\widetilde{f}$ with respect to x as long as they are all Lebesgue points. Furthermore, by using additional mollifiers such as (6.18) in [27], one can introduce finite number of bounded jumps to these functions in x and yet still obtain the approximated functions ξ_n and f_n satisfying the same properties.

A Useful a priori estimates: FSDE

In this Appendix, we summarize the useful a priori estimates for the (B)SDEs with jumps. The following result taken from Lemma 5-1 of Bichteler, Gravereaux and Jacod (1987) [3] is essential for analysis of a σ-finite random measure:

Lemma A.1. *Let $\eta:\mathbb{R}\to\mathbb{R}$ be defined by $\eta(z)=1\wedge|z|$. Then, for $\forall p\geq 2$, there exists a constant δ_p depending on p,T,m,k such that*

$$\mathbb{E}\left[\sup_{t\in[0,T]}\Big|\int_0^t\int_E U(s,z)\widetilde{\mu}(ds,dz)\Big|^p\right]\leq\delta_p\int_0^T\mathbb{E}|L_s|^p ds \qquad \text{(A.1)}$$

if U is an $\mathbb{R}^{m\times k}$-valued $\mathcal{P}\otimes\mathcal{E}$-measurable function on $\Omega\times[0,T]\times E$ and L is a predictable process satisfying $|U_{\cdot,i}(\omega,s,z)|\leq L_s(\omega)\eta(z)$ for each column $1\leq i\leq k$.

Since $\int_E \eta(z)^p\nu(dz)<\infty$ for $\forall p\geq 2$, the above lemma tells that one can use a BDG-like inequality with a compensator ν whenever the integrand of the random measure divided by η is dominated by some integrable random variable. The following result from Chapter 1 Section 9 Lemma 6 of Liptser & Shiryayev (1989) [32] or Lemma 2.1 of Dzhaparidze & Valkeila (1990) [13] is also important:

Lemma A.2. *Let ψ belong to $\mathbb{H}^2_\nu[0,T]$. Then, for $p\geq 2$, there exists some constant $C_p>0$ depending only on p such that*

$$\mathbb{E}\Bigl(\int_0^T\int_E|\psi_s(z)|^2\nu(dz)ds\Bigr)^{p/2}\leq C_p\mathbb{E}\Bigl(\int_0^T\int_E|\psi_s(z)|^2\mu(ds,dz)\Bigr)^{p/2}.$$

For $t_1\leq t_2\leq T$ and $\mathbb{R}^d$-valued $\mathcal{F}_{t_i}$-measurable random variable x^i, let us consider $\{X^i_t,t\in[t_i,T]\}_{1\leq i\leq 2}$ as a solution of the following SDE:

$$X^i_t=x^i+\int_{t_i}^t\widetilde{b}^i(s,X^i_s)ds+\int_{t_i}^t\widetilde{\sigma}^i(s,X^i_s)dW_s+\int_{t_i}^t\int_E\widetilde{\gamma}^i(s,X^i_{s-},z)\widetilde{\mu}(ds,dz)\qquad\text{(A.2)}$$

where $\widetilde{b}^i:\Omega\times[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$, $\widetilde{\sigma}^i:\Omega\times[0,T]\times\mathbb{R}^d\to\mathbb{R}^{d\times l}$, and $\widetilde{\gamma}^i:\Omega\times[0,T]\times\mathbb{R}^d\times E\to\mathbb{R}^{d\times k}$.

Assumption A.1. *For $i\in\{1,2\}$, the map $(\omega,t)\mapsto\widetilde{b}^i(\omega,t,\cdot)$ is $\mathbb{F}$-progressively measurable, $(\omega,t)\mapsto\widetilde{\sigma}^i(\omega,t,\cdot),\widetilde{\gamma}^i(\omega,t,\cdot)$ are $\mathbb{F}$-predictable, and there exists some constant $K>0$ such that, for every $x,x'\in\mathbb{R}^d$ and $z\in E$,*

$$\begin{aligned}&|\widetilde{b}^i(\omega,t,x)-\widetilde{b}^i(\omega,t,x')|+|\widetilde{\sigma}^i(\omega,t,x)-\widetilde{\sigma}^i(\omega,t,x')|\leq K|x-x'|\\&|\widetilde{\gamma}^i_{\cdot,j}(\omega,t,x,z)-\widetilde{\gamma}^i_{\cdot,j}(\omega,t,x',z)|\leq K\eta(z)|x-x'|,\quad 1\leq j\leq k\end{aligned}$$

$d\mathbb{P}\otimes dt$-a.e. in $\Omega\times[0,T]$. Furthermore, for some $p\geq 2$,

$$\mathbb{E}\left[|x^i|^p+\Bigl(\int_{t_i}^T|\widetilde{b}^i(s,0)|ds\Bigr)^p+\Bigl(\int_{t_i}^T|\widetilde{\sigma}^i(s,0)|^2ds\Bigr)^{p/2}+\int_{t_i}^T|L^i_s|^pds\right]<\infty$$

where L^i is some $\mathbb{F}$-predictable process satisfying $|\widetilde{\gamma}^i_{\cdot,j}(\omega,t,0,z)|\leq L^i_t(\omega)\eta(z)$ for every column vector $\{\widetilde{\gamma}^i_{\cdot,j},\ 1\leq j\leq k\}$.

The following lemma is an extension of Lemma A.1 given in [5] to a σ-finite measure by using (A.1).

Lemma A.3. *Under Assumption A.1, the SDE (A.2) has a unique solution and there exists some constant $C_p>0$ such that,*

$$\begin{aligned}||X^i||^p_{\mathbb{S}^p_d[t_i,T]}\leq C_p\mathbb{E}\Bigl[&|x^i|^p+\Bigl(\int_{t_i}^T|\widetilde{b}^i(s,0)|ds\Bigr)^p\\&+\Bigl(\int_{t_i}^T|\widetilde{\sigma}^i(s,0)|^2ds\Bigr)^{p/2}+\int_{t_i}^T|L^i_s|^pds\Bigr]\end{aligned}\qquad\text{(A.3)}$$

and, for all $t_i \le s \le t \le T$,

$$\mathbb{E}\left[\sup_{s\le u\le t}|X^i_u - X^i_s|^p\right] \le C_p A^i_p|t-s| \tag{A.4}$$

where

$$A^i_p := \mathbb{E}\left[|x^i|^p + ||\widetilde{b}^i(\cdot,0)||^p_{[t_i,T]} + ||\widetilde{\sigma}^i(\cdot,0)||^p_{[t_i,T]} + ||L^i||^p_{[t_i,T]}\right] .$$

Moreover, for $t_2 \le t \le T$,

$$\begin{aligned}||\delta X||^p_{\mathbb{S}^p_d[t_2,T]} &\le C_p\Big(\mathbb{E}|x^1-x^2|^p + A^1_p|t_2-t_1|\Big)\\ &+C_p\mathbb{E}\left[\Big(\int_{t_2}^T|\delta\widetilde{b}_t|dt\Big)^p + \Big(\int_{t_2}^T|\delta\widetilde{\sigma}_t|^2dt\Big)^{p/2} + \int_{t_2}^T|\delta L_t|^pdt\right]\end{aligned} \tag{A.5}$$

where $\delta X := X^1 - X^2$, $\delta\widetilde{b}_\cdot := (\widetilde{b}^1-\widetilde{b}^2)(\cdot,X^1_\cdot)$, $\delta\widetilde{\sigma}_\cdot := (\widetilde{\sigma}^1-\widetilde{\sigma}^2)(\cdot,X^1_\cdot)$ and δL is a predictable process satisfying $|\delta\widetilde{\gamma}|(\omega,t,z) \le \delta L_t(\omega)\ \eta(z)$, $d\mathbb{P}\otimes dt$-a.e. in $\Omega\times[0,T]$, where $\delta\widetilde{\gamma}(\omega,t,z) := (\widetilde{\gamma}^1-\widetilde{\gamma}^2)(\omega,t,X^1_{t-}(\omega),z)$.

Proof. The existence of a unique solution is given in pp.237 of Gikhman & Skorohod (1972) [22] or Section 6.2 of Applebaum (2009) [1], for example. For the sake of completeness, let us give a sketch of proof for the other estimates.

Set a sequence of stopping times $\Big(\tau_n := \inf\{t\ge t_i; |X^i_s|\ge n\}\wedge T,\ n\in\mathbb{N}\Big)$. Then, using the fact that $|\widetilde{\gamma}^i(s,X^i_{s-},z)| \le (L^i_s + K|X^i_{s-}|)\eta(z)$, Lemma A.1 and the Burkholder-Davis-Gundy (BDG) inequality, one obtains

$$\begin{aligned}\mathbb{E}|X^i_{\tau_n}|^p &\le C_p\int_{t_i}^{\tau_n}\mathbb{E}|X^i_s|^pds\\ &+ C_p\mathbb{E}\left[|x^i|^p + \Big(\int_{t_i}^{\tau_n}|\widetilde{b}^i(s,0)|ds\Big)^p + \Big(\int_{t_i}^{\tau_n}|\widetilde{\sigma}^i(s,0)|^2ds\Big)^{p/2} + \int_{t_i}^{\tau_n}|L^i_s|^pds\right] .\end{aligned}$$

Using the Gronwall inequality and passing to the limit $n\to\infty$, one obtains the estimate for $\Big(\sup_{t\in[t_i,T]}\mathbb{E}|X^i_t|^p\Big)$. Using the BDG inequality and Lemma A.1 once again, one obtains the first estimate (A.3). A similar analysis yields

$$\begin{aligned}\mathbb{E}\sup_{u\in[s,t]}|X^i_u - X^i_s|^p &\le C_p\mathbb{E}\left[\Big(\int_s^t|\widetilde{b}^i(r,0)|dr\Big)^p + \Big(\int_s^t|\widetilde{\sigma}^i(r,0)|^2dr\Big)^{p/2} + \int_s^t|L^i_r|^pdr\right]\\ &+C_p(t-s)\mathbb{E}||X^i||^p_{[t_i,T]},\end{aligned}$$

which gives second estimate (A.4).

As for the last estimate (A.5), notice first that

$$|\widetilde{\gamma}^1-\widetilde{\gamma}^2|(s,X^1_{s-},z) \le (L^1_s + L^2_s + 2K|X^1_{s-}|)\eta(z) .$$

Since $X^1\in\mathbb{S}^p$, there exists a predictable process δL satisfying $|\widetilde{\gamma}^1-\widetilde{\gamma}^2|(s,X^1_{s-},z) \le \delta L_s\eta(z)$, $d\mathbb{P}\otimes ds$-a.e. and $\int_{t_2}^T\mathbb{E}|\delta L_r|^pdr < \infty$ as desired. Separating the integration range, applying the BDG inequality and Lemma A.1, one obtains

$$\begin{aligned}\mathbb{E}||\delta X||^p_{[t_2,t]} &\leq C_p\mathbb{E}\Big[|x^1-x^2|^p+\Big(\int_{t_1}^{t_2}|\widetilde{b}^1(s,0)|ds\Big)^p+\Big(\int_{t_1}^{t_2}|\widetilde{\sigma}^1(s,0)|^2ds\Big)^{p/2}\\ &\quad+\int_{t_1}^{t_2}|L^1_s|^pds+(t_2-t_1)||X^1||^p_{[t_1,t_2]}\Big]+C_p\mathbb{E}\Big[\int_{t_2}^t|\delta X_s|^pds\\ &\quad+\Big(\int_{t_2}^t|\delta\widetilde{b}_s|ds\Big)^p+\Big(\int_{t_2}^t|\delta\widetilde{\sigma}_s|^2ds\Big)^{p/2}+\int_{t_2}^t|\delta L_s|^pds\Big]\ .\end{aligned}$$

Using the first two results and the Gronwall inequality, one obtains (A.5). □

Remark

Note that when $p=2$, one can replace $\int^\cdot |L^i_s|^2ds$ (resp. $\int^\cdot |\delta L_s|^2ds$) by $\int^\cdot\int_E|\widetilde{\gamma}^i(s,0,z)|^2\nu(dz)ds$ (resp. $\int^\cdot\int_E|\delta\widetilde{\gamma}(s,z)|^2\nu(dz)ds$) by simply applying the BDG inequality. Furthermore, when the compensator is finite $\nu(E)<\infty$, the above replacement is possible for any $\forall p\geq 2$ thanks to Lemma B.3 (see below).

B Useful a priori estimates: BSDE

Consider the following BSDE:

$$Y_t=\widetilde{\xi}+\int_t^T\widetilde{f}(s,Y_s,Z_s,\psi_s)ds-\int_t^T Z_sdW_s-\int_t^T\int_E\psi_s(z)\widetilde{\mu}(ds,dz)\ , \tag{B.1}$$

where $\xi:\Omega\to\mathbb{R}^m$, $\widetilde{f}:\Omega\times[0,T]\times\mathbb{R}^m\times\mathbb{R}^{m\times l}\times\mathbb{L}^2(E,\mathcal{E},\nu;\mathbb{R}^m)\to\mathbb{R}^m$. In this section, we use $\langle\cdot,\cdot\rangle$ to denote an inner product of m-dimensional vectors for clarity.

Assumption B.1. *(i) $\widetilde{\xi}$ is $\mathcal{F}_T$-measurable and the map $(\omega,t)\mapsto\widetilde{f}(\omega,t,\cdot)$ is $\mathbb{F}$-progressively measurable. There exists a solution (Y,Z,ψ) to the BSDE (B.1).*
(ii) For $\forall\lambda\in(0,1)$, there exist an $\mathbb{F}$-progressively measurable continuous process with bounded variation $(V^\lambda_s,s\in[0,T])$ with $V^\lambda_0=0$ and an $\mathbb{F}$-progressively measurable increasing process $(N^\lambda_s,s\in[0,T])$ with $N_0=0$ such that, as a signed measure on $\mathbb{R}_+$,

$$\langle Y_s,\widetilde{f}(s,Y_s,Z_s,\psi_s)\rangle ds\leq|Y_s|^2dV^\lambda_s+|Y_s|dN^\lambda_s+\lambda(|Z_s|^2+||\psi_s||^2_{\mathbb{L}^2(E)})ds\ .$$

(iv) There exists some $p\geq 2$ such that $\mathbb{E}\Big[||e^{V^\lambda}Y||^p_T+\Big(\int_0^Te^{V^\lambda_s}dN^\lambda_s\Big)^p\Big]<\infty$ is satisfied for every $\forall\lambda\in(0,1)$.

Lemma B.1. *Suppose Assumption B.1 hold true. Then, there exists some $\exists\lambda\in(0,1)$ such that the following inequality is satisfied;*

$$\begin{aligned}&\mathbb{E}||e^{V^\lambda}Y||^p_T+\mathbb{E}\Big(\int_0^Te^{2V^\lambda_s}|Z_s|^2ds\Big)^{\frac{p}{2}}+\mathbb{E}\Big(\int_0^T\int_Ee^{2V^\lambda_s}|\psi_s(z)|^2\mu(ds,dz)\Big)^{\frac{p}{2}}\\ &+\mathbb{E}\Big(\int_0^T\int_Ee^{2V^\lambda_s}|\psi_s(z)|^2\nu(dz)ds\Big)^{\frac{p}{2}}\leq C_{p,\lambda}\mathbb{E}\Big[e^{pV^\lambda_T}|\widetilde{\xi}|^p+\Big(\int_0^Te^{V^\lambda_s}dN^\lambda_s\Big)^p\Big]\ ,\end{aligned}$$

where $C_{p,\lambda}$ is a positive constant depending only on p,λ.

Proof. The following proof is an improvement of Proposition 2 of Kruse & Popier (2015) [29] by following the idea of Proposition 6.80 of Pardoux & Rascanu (2014) [38], which yields a slightly sharper a priori estimate for $p \geq 2$.

First step: Introduce a sequence of stopping times with $n \in \mathbb{N}$,

$$\begin{aligned}\tau_n \quad := \quad & \inf\Big\{t \geq 0; \int_0^t e^{2V_s^\lambda}|Z_s|^2 ds + \int_0^t \int_E e^{2V_s^\lambda}|\psi_s(z)|^2\big(\mu(ds,dz) + \nu(dz)ds\big) \\ & +||e^{V^\lambda}Y||_t + \int_0^t e^{V_s^\lambda} dN_s^\lambda \geq n\Big\} \wedge T.\end{aligned}$$

One obtains by applying Itô formula

$$\begin{aligned}& |Y_0|^2 + \int_0^{\tau_n} e^{2V_s^\lambda}|Z_s|^2 ds + \int_0^{\tau_n}\int_E e^{2V_s^\lambda}|\psi_s(z)|^2\mu(ds,dz) \\ &= e^{2V_{\tau_n}^\lambda}|Y_{\tau_n}|^2 + \int_0^{\tau_n} e^{2V_s^\lambda} 2\Big(\langle Y_s, \widetilde{f}(s,Y_s,Z_s,\psi_s)\rangle ds - |Y_s|^2 dV_s^\lambda\Big) \\ &\quad - \int_0^{\tau_n} e^{2V_s^\lambda} 2\langle Y_s, Z_s dW_s\rangle - \int_0^{\tau_n} e^{2V_s^\lambda} 2\langle Y_{s-}, \psi_s(z)\rangle \widetilde{\mu}(ds,dz) \\ &\leq e^{2V_{\tau_n}^\lambda}|Y_{\tau_n}|^2 + \int_0^{\tau_n} e^{2V_s^\lambda} 2\Big(|Y_s| dN_s^\lambda + \lambda(|Z_s|^2 + ||\psi_s||^2_{\mathbb{L}^2(E)})ds\Big) \\ &\quad - \int_0^{\tau_n} e^{2V_s^\lambda} 2\langle Y_s, Z_s dW_s\rangle - \int_0^{\tau_n}\int_E e^{2V_s^\lambda} 2\langle Y_{s-}, \psi_s(z)\rangle \widetilde{\mu}(ds,dz)\ .\end{aligned}$$

The BDG (or Davis when $p = 2$) inequality yields, with some positive constant C_p depending only on p,

$$\begin{aligned}& \mathbb{E}\left[\Big(\int_0^{\tau_n} e^{2V_s^\lambda}|Z_s|^2 ds\Big)^{\frac{p}{2}} + \Big(\int_0^{\tau_n}\int_E e^{2V_s^\lambda}|\psi_s(z)|^2\mu(ds,dz)\Big)^{\frac{p}{2}}\right] \\ &\leq C_p\mathbb{E}\left[||e^{V^\lambda}Y||^p_{\tau_n} + \Big(\int_0^{\tau_n} e^{V_s^\lambda} dN_s^\lambda\Big)^p\right] \\ &\quad +\lambda^{\frac{p}{2}} C_p\mathbb{E}\left[\Big(\int_0^{\tau_n} e^{2V_s^\lambda}|Z_s|^2 ds\Big)^{\frac{p}{2}} + \Big(\int_0^{\tau_n} e^{2V_s^\lambda}||\psi_s||_{\mathbb{L}^2(E)} ds\Big)^{\frac{p}{2}}\right] \\ &\quad +C_p\mathbb{E}\left[\Big(\int_0^{\tau_n} e^{4V_s^\lambda}|Y_s|^2|Z_s|^2 ds\Big)^{\frac{p}{4}} + \Big(\int_0^{\tau_n}\int_E e^{4V_s^\lambda}|Y_s|^2|\psi_s(z)|^2\mu(ds,dz)\Big)^{\frac{p}{4}}\right]\ .\end{aligned}$$

With an arbitrary constant $\epsilon > 0$, one has

$$\begin{aligned}C_p\mathbb{E}\left[\Big(\int_0^{\tau_n} e^{4V_s^\lambda}|Y_s|^2|Z_s|^2 ds\Big)^{\frac{p}{4}}\right] &\leq C_p\mathbb{E}\left[||e^{V^\lambda}Y||^{\frac{p}{2}}_{\tau_n}\Big(\int_0^{\tau_n} e^{2V_s^\lambda}|Z_s|^2 ds\Big)^{\frac{p}{4}}\right] \\ &\leq \frac{C_p^2}{4\epsilon}\mathbb{E}\Big[||e^{V^\lambda}Y||^p_{\tau_n}\Big] + \epsilon\mathbb{E}\left[\Big(\int_0^{\tau_n} e^{2V_s^\lambda}|Z_s|^2 ds\Big)^{\frac{p}{2}}\right]\end{aligned}$$

and similarly

$$\begin{aligned}& C_p\mathbb{E}\left[\Big(\int_0^{\tau_n}\int_E e^{4V_s^\lambda}|Y_s|^2|\psi_s(z)|^2\mu(ds,dz)\Big)^{\frac{p}{4}}\right] \\ &\quad \leq \frac{C_p^2}{4\epsilon}\mathbb{E}\Big[||e^{V^\lambda}Y||^p_{\tau_n}\Big] + \epsilon\mathbb{E}\left[\Big(\int_0^{\tau_n}\int_E e^{2V_s^\lambda}|\psi_s(z)|^2\mu(ds,dz)\Big)^{\frac{p}{2}}\right].\end{aligned}$$

Thus, one obtains

$$(1-\epsilon-\lambda^{\frac{p}{2}}C_p)\mathbb{E}\Big(\int_0^{\tau_n} e^{2V_s^\lambda}|Z_s|^2ds\Big)^{\frac{p}{2}}+(1-\epsilon)\mathbb{E}\Big(\int_0^{\tau_n}\int_E e^{2V_s^\lambda}|\psi_s(z)|^2\mu(ds,dz)\Big)^{\frac{p}{2}}$$
$$-\lambda^{\frac{p}{2}}C_p\mathbb{E}\Big(\int_0^{\tau_n}\int_E e^{2V_s^\lambda}|\psi_s(z)|^2\nu(dz)ds\Big)^{\frac{p}{2}}\leq C_p'\mathbb{E}\left[||e^{V^\lambda}Y||_{\tau_n}^p+\Big(\int_0^{\tau_n}e^{V_s^\lambda}dN_s^\lambda\Big)^p\right].$$

Firstly, choose some $\epsilon\in(0,1)$. Then, by Lemma A.2, there exists a $\lambda\in(0,1)$ depending only on p so that the 3rd term is absolutely smaller than the 2nd term. Redefining the coefficients and passing to the limit $\tau_n\to T$ yields

$$\mathbb{E}\left[\Big(\int_0^T e^{2V_s^\lambda}|Z_s|^2ds\Big)^{\frac{p}{2}}+\Big(\int_0^T\int_E e^{2V_s^\lambda}|\psi_s(z)|^2\mu(ds,dz)\Big)^{\frac{p}{2}}\right]$$
$$+\mathbb{E}\left[\Big(\int_0^T\int_E e^{2V_s^\lambda}|\psi_s(z)|^2\nu(dz)ds\Big)^{\frac{p}{2}}\right]\leq C_{p,\lambda}\mathbb{E}\left[||e^{V^\lambda}Y||_T^p+\Big(\int_0^T e^{V_s^\lambda}dN_s^\lambda\Big)^p\right].\quad\text{(B.2)}$$

Second step: Put $\theta(y):=|y|^p$. Then, Itô formula yields

$$d(e^{pV_s^\lambda}|Y_s|^p)=e^{pV_s^\lambda}\Big(p|Y_s|^pdV_s^\lambda+p|Y_{s-}|^{p-2}\langle Y_{s-},dY_s\rangle+\frac{1}{2}\mathrm{Tr}(\partial_y^2\theta(Y_s)Z_sZ_s^\top)ds\Big)$$
$$+\int_E e^{pV_s^\lambda}\Big(|Y_{s-}+\psi_s(z)|^p-|Y_{s-}|^p-p|Y_{s-}|^{p-2}\langle Y_{s-},\psi_s(z)\rangle\Big)\mu(ds,dz).$$

Using the same sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$,

$$e^{pV_t^\lambda}|Y_t|^p=e^{pV_{\tau_n}^\lambda}|Y_{\tau_n}|^p+\int_t^{\tau_n}e^{pV_s^\lambda}p|Y_s|^{p-2}\Big(\langle Y_s,\widetilde{f}(s,Y_s,Z_s,\psi_s)\rangle ds-|Y_s|^2dV_s^\lambda\Big)$$
$$-\int_t^{\tau_n}e^{pV_s^\lambda}\frac{1}{2}\mathrm{Tr}(\partial_y^2\theta(Y_s)Z_sZ_s^\top)ds$$
$$-\int_t^{\tau_n}\int_E e^{pV_s^\lambda}\Big(|Y_{s-}+\psi_s(z)|^p-|Y_{s-}|^p-p|Y_{s-}|^{p-2}\langle Y_{s-},\psi_s(z)\rangle\Big)\mu(ds,dz)$$
$$-\int_t^{\tau_n}e^{pV_s^\lambda}p|Y_s|^{p-2}\langle Y_s,Z_sdW_s\rangle-\int_t^{\tau_n}\int_E e^{pV_s^\lambda}p|Y_{s-}|^{p-2}\langle Y_{s-},\psi_s(z)\rangle\widetilde{\mu}(ds,dz).$$

Let us mention the fact that

$$\mathrm{Tr}(\partial_y^2\theta(Y_s)Z_sZ_s^\top)\geq p|Y_s|^{p-2}|Z_s|^2,$$
$$|Y_{s-}+\psi_s^i(z)|^p-|Y_{s-}|^p-p|Y_{s-}|^{p-2}\langle Y_{s-},\psi_s^i(z)\rangle\geq p(p-1)3^{1-p}|Y_{s-}|^{p-2}|\psi_s^i(z)|^2,$$

for every $i\in\{1,\cdots,k\}$. The latter is obtained by evaluating the residual of Taylor formula [29]. Setting $\kappa_p:=\min\Big(\frac{p}{2},p(p-1)3^{1-p}\Big)$, one obtains

$$e^{pV_t^\lambda}|Y_t|^p+\kappa_p\int_t^{\tau_n}e^{pV_s^\lambda}|Y_s|^{p-2}|Z_s|^2ds+\kappa_p\int_t^{\tau_n}\int_E e^{pV_s^\lambda}|Y_{s-}|^{p-2}|\psi_s(z)|^2\mu(ds,dz)$$
$$\leq e^{pV_{\tau_n}^\lambda}|Y_{\tau_n}|^p+\int_t^{\tau_n}e^{pV_s^\lambda}p|Y_s|^{p-2}\Big(|Y_s|dN_s^\lambda+\lambda(|Z_s|^2+||\psi_s||^2_{\mathbb{L}^2(E)})ds\Big)$$
$$-\int_t^{\tau_n}e^{pV_s^\lambda}p|Y_s|^{p-2}\langle Y_s,Z_sdW_s\rangle-\int_t^{\tau_n}\int_E e^{pV_s^\lambda}p|Y_{s-}|^{p-2}\langle Y_{s-},\psi_s(z)\rangle\widetilde{\mu}(ds,dz).\quad\text{(B.3)}$$

Putting $t=0$ and taking expectation give

$$\mathbb{E}\left[\kappa_p\int_0^{\tau_n} e^{pV_s^\lambda}|Y_s|^{p-2}|Z_s|^2ds+\kappa_p\int_0^{\tau_n}\int_E e^{pV_s^\lambda}|Y_{s-}|^{p-2}|\psi_s(z)|^2\mu(ds,dz)\right]$$
$$\leq \mathbb{E}\left[e^{pV_{\tau_n}^\lambda}|Y_{\tau_n}|^p+\int_0^{\tau_n} e^{pV_s^\lambda}p|Y_s|^{p-1}dN_s^\lambda\right]+\lambda\mathbb{E}\left[\int_0^{\tau_n} e^{pV_s^\lambda}p|Y_s|^{p-2}\big(|Z_s|^2+||\psi_s||^2_{\mathbb{L}^2(E)}\big)ds\right] .$$

By Lemma A.2, one obtains

$$\mathbb{E}\left[\int_0^{\tau_n} e^{pV_s^\lambda}|Y_s|^{p-2}|Z_s|^2ds+\int_0^{\tau_n}\int_E e^{pV_s^\lambda}|Y_{s-}|^{p-2}|\psi_s(z)|^2\mu(ds,dz)\right]$$
$$\leq C_{p,\lambda}\mathbb{E}\left[e^{pV_{\tau_n}^\lambda}|Y_{\tau_n}|^p+\int_0^{\tau_n} e^{pV_s^\lambda}|Y_s|^{p-1}dN_s^\lambda\right] \tag{B.4}$$

by choosing a small $\lambda\in(0,1)$.

Now, applying the Davis inequality (See Chap.I, Sec. 9, Theorem 6 in [32]) to (B.3),

$$\mathbb{E}\Big[||e^{V^\lambda}Y||^p_{\tau_n}\Big]+\mathbb{E}\left[\kappa_p\int_0^{\tau_n} e^{pV_s^\lambda}|Y_s|^{p-2}|Z_s|^2ds+\kappa_p\int_0^{\tau_n}\int_E e^{pV_s^\lambda}|Y_{s-}|^{p-2}|\psi_s(z)|^2\mu(ds,dz)\right]$$
$$\leq \mathbb{E}\left[e^{pV_{\tau_n}^\lambda}|Y_{\tau_n}|^p+\int_0^{\tau_n} e^{pV_s^\lambda}p|Y_s|^{p-1}dN_s^\lambda\right]+\lambda\mathbb{E}\left[\int_0^{\tau_n} e^{pV_s^\lambda}p|Y_s|^{p-2}\big(|Z_s|^2+||\psi_s||^2_{\mathbb{L}^2(E)}\big)ds\right]$$
$$+C\mathbb{E}\Big(\int_0^{\tau_n} e^{2pV_s^\lambda}|Y_s|^{2p-2}|Z_s|^2ds\Big)^{\frac{1}{2}}+C\mathbb{E}\Big(\int_0^{\tau_n}\int_E e^{2pV_s^\lambda}|Y_{s-}|^{2p-2}|\psi_s(z)|^2\mu(ds,dz)\Big)^{\frac{1}{2}} ,$$

where C is some positive constant. By Lemma A.2, one can choose $\lambda\in(0,1)$ small enough (depending only on p) so that

$$\mathbb{E}\Big[||e^{V^\lambda}Y||^p_{\tau_n}\Big]\leq \mathbb{E}\left[e^{pV_{\tau_n}^\lambda}|Y_{\tau_n}|^p+\int_0^{\tau_n} e^{pV_s^\lambda}p|Y_s|^{p-1}dN_s^\lambda\right]$$
$$+C\mathbb{E}\Big(\int_0^{\tau_n} e^{2pV_s^\lambda}|Y_s|^{2p-2}|Z_s|^2ds\Big)^{\frac{1}{2}}+C\mathbb{E}\Big(\int_0^{\tau_n}\int_E e^{2pV_s^\lambda}|Y_{s-}|^{2p-2}|\psi_s(z)|^2\mu(ds,dz)\Big)^{\frac{1}{2}} .$$

By retaking a smaller λ in the first step if necessary, one can use a common $\lambda\in(0,1)$ both in the first and second steps.

Note that

$$C\mathbb{E}\Big(\int_0^{\tau_n} e^{2pV_s^\lambda}|Y_s|^{2p-2}|Z_s|^2ds\Big)^{\frac{1}{2}}\leq C\mathbb{E}\left[||e^{V^\lambda}Y||^{\frac{p}{2}}_{\tau_n}\Big(\int_0^{\tau_n} e^{pV_s^\lambda}|Y_s|^{p-2}|Z_s|^2ds\Big)^{\frac{1}{2}}\right]$$
$$\leq \epsilon\mathbb{E}\Big[||e^{V^\lambda}Y||^p_{\tau_n}\Big]+\frac{C^2}{4\epsilon}\mathbb{E}\left[\int_0^{\tau_n} e^{pV_s^\lambda}|Y_s|^{p-2}|Z_s|^2ds\right] ,$$

and similarly

$$C\mathbb{E}\Big(\int_0^{\tau_n}\int_E e^{2pV_s^\lambda}|Y_{s-}|^{2p-2}|\psi_s(z)|^2\mu(ds,dz)\Big)^{\frac{1}{2}}$$
$$\leq \epsilon\mathbb{E}\Big[||e^{V^\lambda}Y||^p_{\tau_n}\Big]+\frac{C^2}{4\epsilon}\mathbb{E}\left[\int_0^{\tau_n}\int_E e^{pV_s^\lambda}|Y_{s-}|^{p-2}|\psi_s(z)|^2\mu(ds,dz)\right] .$$

Thus, taking $\epsilon = 1/4$, one obtains

$$\mathbb{E}\Big[||e^{V^\lambda}Y||^p_{\tau_n}\Big] \le C_p\mathbb{E}\left[e^{pV^\lambda_{\tau_n}}|Y_{\tau_n}|^p + \int_0^{\tau_n} e^{pV^\lambda_s}|Y_s|^{p-1}dN^\lambda_s\right]$$
$$+C_p\mathbb{E}\left[\int_0^{\tau_n} e^{pV^\lambda_s}|Y_s|^{p-2}|Z_s|^2ds + \int_0^{\tau_n}\int_E e^{pV^\lambda_s}|Y_{s-}|^{p-2}|\psi_s(z)|^2\mu(ds,dz)\right] .$$

Then the inequality (B.4) implies

$$\mathbb{E}\Big[||e^{V^\lambda}Y||^p_{\tau_n}\Big] \le C_{p,\lambda}\mathbb{E}\left[e^{pV^\lambda_{\tau_n}}|Y_{\tau_n}|^p + \int_0^{\tau_n} e^{pV^\lambda_s}|Y_s|^{p-1}dN^\lambda_s\right] .$$

Passing to the limit $\tau_n \to T$, the monotone convergence in the left and the dominated convergence in the right-hand side give

$$\mathbb{E}\Big[||e^{V^\lambda}Y||^p_{T}\Big] \le C_{p,\lambda}\mathbb{E}\left[e^{pV^\lambda_{T}}|\widetilde{\xi}|^p + \int_0^{T} e^{pV^\lambda_s}|Y_s|^{p-1}dN^\lambda_s\right] .$$

By Young's inequality, for an arbitrary $\epsilon > 0$, one has that

$$\mathbb{E}\left[\int_0^T e^{pV^\lambda_s}|Y_s|^{p-1}dN^\lambda_s\right] \le \mathbb{E}\left[||e^{V^\lambda}Y||^{p-1}_T\int_0^T e^{V^\lambda_s}dN^\lambda_s\right]$$
$$\le \frac{p-1}{p}\epsilon^{\frac{p}{p-1}}\mathbb{E}\Big[||e^{V^\lambda}Y||^p_T\Big] + \frac{1}{p\epsilon^p}\mathbb{E}\left[\Big(\int_0^T e^{V^\lambda_s}dN^\lambda_s\Big)^p\right] .$$

Hence, by taking ϵ small, one obtains

$$\mathbb{E}\Big[||e^{V^\lambda}Y||^p_{T}\Big] \le C_{p,\lambda}\mathbb{E}\left[e^{pV^\lambda_{T}}|\widetilde{\xi}|^p + \Big(\int_0^{T} e^{V^\lambda_s}dN^\lambda_s\Big)^p\right] ,$$

Combining with the result (B.2) in *First step*, one obtains the desired result. □

Now, let us introduce the maps $\widetilde{\xi}^i : \Omega \to \mathbb{R}^m$ and $\widetilde{f}^i : \Omega \times [0,T] \times \mathbb{R}^m \times \mathbb{R}^{m\times l} \times \mathbb{L}^2(E,\mathcal{E},\nu;\mathbb{R}^m) \to \mathbb{R}^m$ with $i \in \{1,2\}$.

Assumption B.2. *(i) For $i \in \{1,2\}$, $\widetilde{\xi}^i$ is $\mathcal{F}_T$-measurable and the map $(\omega,t) \mapsto \widetilde{f}^i(\omega,t,\cdot)$ is $\mathbb{F}$-progressively measurable.*
(ii) For every (y,z,ψ), $(y',z',\psi') \in \mathbb{R}^m \times \mathbb{R}^{m\times l} \times \mathbb{L}^2(E,\mathcal{E},\nu;\mathbb{R}^m)$, there exists a positive constant $K > 0$ such that

$$|\widetilde{f}^i(\omega,t,y,z,\psi) - \widetilde{f}^i(\omega,t,y',z',\psi')| \le K\Big(|y-y'| + |z-z'| + ||\psi-\psi'||_{\mathbb{L}^2(E)}\Big)$$

$d\mathbb{P}\otimes dt$-a.e. in $\Omega \times [0,T]$.
(iii) For both $i \in \{1,2\}$, there exists some $p \ge 2$ such that

$$\mathbb{E}\left[|\widetilde{\xi}|^p + \Big(\int_0^T |\widetilde{f}(s,0,0,0)|ds\Big)^p\right] < \infty .$$

Lemma B.2. *(a)Under Assumption B.2, the BSDE*

$$Y^i_t = \widetilde{\xi}^i + \int_t^T \widetilde{f}^i(s,Y^i_s,Z^i_s,\psi^i_s)ds - \int_t^T Z^i_sdW_s - \int_t^T\int_E \psi^i_s(z)\widetilde{\mu}(ds,dz) \tag{B.5}$$

has a unique solution (Y^i, Z^i, ψ^i) which belongs to $\mathbb{S}^p_m[0,T] \times \mathbb{H}^p_{m\times l}[0,T] \times \mathbb{H}^p_{m,\nu}[0,T]$ satisfying the inequality

$$||(Y^i, Z^i, \psi^i)||^p_{\mathcal{K}^p[0,T]} \leq C_p \mathbb{E}\left[|\widetilde{\xi}|^p + \Big(\int_0^T |\widetilde{f}^i(s,0,0,0)|ds\Big)^p\right] \tag{B.6}$$

where C_p is some positive constant depending only on (p, K, T). Moreover, if $A^i_2 := \mathbb{E}\Big[|\widetilde{\xi}^i|^2 + ||\widetilde{f}^i(\cdot, 0)||^2_T\Big] < \infty$, then

$$\mathbb{E}\Big[\sup_{s\leq u\leq t} |Y^i_u - Y^i_s|^2\Big] \leq C_2\left[A^i_2|t-s|^2 + \Big(\int_s^t |Z^i_u|^2 du\Big) + \int_s^t\int_E |\psi^i_u(z)|^2\nu(dz)du\right]. \tag{B.7}$$

(b) Fix $\widetilde{\xi}^1, \widetilde{\xi}^2 \in \mathbb{L}^p(\Omega, \mathcal{F}_T, \mathbb{P}; \mathbb{R}^m)$ and let (Y^i, Z^i, ψ^i) be the solution of (B.5) for $i \in \{1,2\}$. Then, for all $t \in [0,T]$,

$$\begin{aligned}&\mathbb{E}\left[||\delta Y||^p_{[t,T]} + \Big(\int_t^T |\delta Z_s|^2 ds\Big)^{p/2} + \Big(\int_t^T\int_E |\delta\psi_s(z)|^2\mu(ds,dz)\Big)^{p/2}\right]\\ &+\mathbb{E}\left[\Big(\int_t^T\int_E |\delta\psi_s(z)|^2\nu(dz)ds\Big)^{p/2}\right] \leq C_p\mathbb{E}\left[|\delta\xi|^p + \Big(\int_t^T |\delta\widetilde{f}_s|ds\Big)^p\right]\end{aligned} \tag{B.8}$$

where $\delta\xi := \widetilde{\xi}^1 - \widetilde{\xi}^2$, $\delta Y := Y^1 - Y^2$, $\delta Z := Z^1 - Z^2$, $\delta\psi := \psi^1 - \psi^2$ and $\delta\widetilde{f}_\cdot := (\widetilde{f}^1 - \widetilde{f}^2)(\cdot, Y^1_\cdot, Z^1_\cdot, \psi^1_\cdot)$.

Remark

Note that in [29], the estimates (B.6) and (B.8) are slightly weaker, where the right hand side is given by $\Big(\int_0^T |\widetilde{f}(s,0,0,0)|^p ds\Big)$ instead of $\Big(\int_0^T |\widetilde{f}(s,0,0,0)|ds\Big)^p$. This stems from Lemma B.1 and can be crucial if one needs to apply a fixed-point theorem for a short maturity T.

Proof. Firstly, assume the existence of a solution to (B.5) such that $(Y^i, Z^i, \psi^i) \in \mathcal{K}^p[0,T]$ for both $i \in \{1,2\}$. One has

$$\begin{aligned}\langle Y^i_s, \widetilde{f}^i(s, Y^i_s, Z^i_s, \psi^i_s)\rangle ds &\leq |Y^i_s|\Big(|\widetilde{f}^i(s,0)| + K\big(|Y^i_s| + |Z^i_s| + ||\psi^i_s||_{\mathbb{L}^2(E)}\big)\Big)ds\\ &\leq |Y^i_s|^2\Big(K + \frac{K^2}{2\lambda}\Big)ds + |Y^i_s||\widetilde{f}^i(s,0)|ds + \lambda(|Z^i_s|^2 + ||\psi^i_s||^2_{\mathbb{L}^2(E)})ds\end{aligned}$$

for $\forall \lambda > 0$. One can easily check that Assumption B.1 is satisfied by choosing

$$V^\lambda_t := \Big(K + \frac{K^2}{2\lambda}\Big)t, \quad N^\lambda_t := \int_0^t |\widetilde{f}^i(s,0)|ds,$$

for $t \in [0,T]$. Thus Lemma B.1 proves the inequality (B.6).

The BDG inequality yields

$$\mathbb{E}\left[\sup_{u\in[s,t]} |Y^i_u - Y^i_s|^2\right] \leq C_2\mathbb{E}\left[\Big(\int_s^t |\widetilde{f}^i(r, Y^i_r, Z^i_r, \psi^i_r)|dr\Big)^2 + \int_s^t \Big(|Z^i_r|^2 + ||\psi^i_r||^2_{\mathbb{L}^2(E)}\Big)dr\right]$$

which, together with the estimate (B.6), proves (B.7). For (b), it is easy to check

$$|f^1(s,Y_s^1,Z_s^1,\psi_s^1)-f^2(s,Y_s^2,Z_s^2,\psi_s^2)|\le|\delta f_s|+K\Big(|\delta Y_s|+|\delta Z_s|+||\delta\psi_s||_{\mathbb{L}^2(E)}\Big).$$

Thus, Assumption B.1 is satisfied once again by choosing

$$V_t^\lambda:=\Big(K+\frac{K^2}{2\lambda}\Big)t,\quad N_t^\lambda:=\int_0^t|\delta f(s)|ds\ .$$

Therefore, the estimate (B.8) immediately follows from Lemma B.1.

Now, let us prove the existence in (a). The uniqueness is already proven by (b). The following is a simple modification of Theorem 5.17 [38] given for a diffusion setup. Consider a sequence of BSDEs (the superscript $i\in\{1,2\}$ is omitted), for $n\in\mathbb{N}$,

$$Y_t^{n+1}=\widetilde{\xi}+\int_t^T\widetilde{f}(s,Y_s^n,Z_s^n,\psi_s^n)ds-\int_t^T Z_s^{n+1}dW_s-\int_t^T\int_E\psi_s^{n+1}(z)\widetilde{\mu}(ds,dz)\ .$$

Suppose that $(Y^n,Z^n,\psi^n)\in\mathcal{K}^p[0,T]$. Then, from the linear growth property, it is obvious that

$$\widetilde{\xi}+\int_t^T\widetilde{f}(s,Y_s^n,Z_s^n,\psi_s^n)ds\in\mathbb{L}^p(\Omega,\mathcal{F}_T,\mathbb{P};\mathbb{R}^m)\ .$$

Thus the martingale representation theorem (see, for example, Theorem 5.3.6 in [1]) implies that there exists a unique solution $(Y^{n+1},Z^{n+1},\psi^{n+1})\in\mathcal{K}^p[0,T]$. Let us define this map as $(Y^{n+1},Z^{n+1},\psi^{n+1})=\Phi(Y^n,Z^n,\psi^n)$. Denote $(\delta Y^n,\delta Z^n,\delta\psi^n):=(Y^n-Y^{n-1},Z^n-Z^{n-1},\psi^n-\psi^{n-1})$. Then, (B.8) (with a zero Lipschitz constant) implies

$$\begin{aligned}
&||(\delta Y^{n+1},\delta Z^{n+1},\delta\psi^{n+1})||^p_{\mathcal{K}^p[0,T]}\\
&\le C_p\mathbb{E}\left[\Big(\int_0^T|\widetilde{f}(s,Y_s^n,Z_s^n,\psi^n)-\widetilde{f}(s,Y_s^{n-1},Z_s^{n-1},\psi_s^{n-1})|ds\Big)^p\right]\\
&\le C_p'\mathbb{E}\left[\Big(\int_0^T\big[|\delta Y_s^n|+|\delta Z_s^n|+||\delta\psi_s^n||_{\mathbb{L}^2(E)}\big]ds\Big)^p\right]\\
&\le C_p'\max(T^p,T^{\frac{p}{2}})||(\delta Y^n,\delta Z^n,\delta\psi^n)||^p_{\mathcal{K}^p[0,T]}\ .
\end{aligned}\tag{B.9}$$

Thus, if the terminal time T is small enough so that $\alpha:=C_p'\max(T^p,T^{\frac{p}{2}})<1$, then the map Φ is strictly contracting. In this case, by the fixed point theorem in the Banach space, there exists a solution $(Y,Z,\psi)\in\mathcal{K}^p[0,T]$ to the BSDE (B.5). For general T, one can consider a time partition $0=T_0<T_1<\cdots<T_N=T$. By taking $[T_{N-1},T]$ small enough, the above arguments guarantee that there exists a solution $(Y,Z,\psi)\in\mathcal{K}^p[T_{N-1},T]$. By the uniqueness of the solution, one can repeat the same procedures for the interval $[T_{N-2},T_{N-1}]$ with the new terminal value $Y_{T_{N-1}}$. Repeating N times, one proves the desired result. □

The following lemma is useful when one deals with the jumps of finite measure.

Lemma B.3. *Suppose $\nu^i(\mathbb{R}_0)<\infty$ for every $1\le i\le k$. Given $\psi\in\mathbb{H}^2_\nu[0,T]$, let M be defined by $M_t:=\int_0^t\int_E\psi_s(z)\widetilde{\mu}(ds,dz)$ on $[0,T]$. Then, for $\forall p\ge2$, $k_p||\psi||^p_{\mathbb{H}^p_\nu[0,T]}\le||M||^p_{\mathbb{S}^p[0,T]}\le K_p||\psi||^p_{\mathbb{H}^p_\nu[0,T]}$, where k_p,K_p are positive constant depend only on $p,\nu(E)$ and*

T.

Proof. See pp.125 of [14], for example. □

Acknowledgement

The research is partially supported by Center for Advanced Research in Finance (CARF) and JSPS KAKENHI (Grant Number 25380389). The authors are grateful to professor Giulia Di Nunno for her excellent lecture on Malliavin's derivative on Lévy process at the University of Tokyo in September 2015.

References

[1] Applebaum, D., 2009, *Lévy Processes and Stochastic Calculus (2nd edition)*, Cambridge studies in advanced mathematics, Cambridge University Press, LN.

[2] Bender, C. and Denk, R., 2007, *A forward scheme for backward SDE*, Stochastic Processes and their Applications, 117, 12, 1793-1832.

[3] Bichteler, K., Gravereaux, J. and Jacod, J., 1987, *Malliavin calculus for processes with jumps*, Stochastics Monographs, Gordon and Breach science publishers, LN.

[4] Bismut, J.M., 1973, *Conjugate convex functions in optimal stochastic control*, J. Math. Anal. Apl. 44, 384-404.

[5] Bouchard, B. and Elie, R., 2008, *Discrete-time approximation of decoupled Forward-Backward SDE with Jumps*, Stochastic Processes and their Applications, 118, 53-75.

[6] Bouchard, B. and Touzi, N., 2004, *Discrete-time approximation and Monte-Carlo simulation of backward stochastic differential equations*, Stochastic Processes and their Applications, 111, 2, 175-206.

[7] Crépey, S. and Bielecki, T.R. with an introductory dialogue by Damiano Brigo, 2014, *Counterparty Risk and Funding*, CRC press, Taylor & Francis Group, NY.

[8] Crépey, S. and Song, S., 2015, *Counterparty Risk and Funding: Immersion and Beyond*, Working paper.

[9] Cvitanić, J. and Zhang, J., 2013, *Contract theory in continuous-time methods*, Springer,Berlin.

[10] Delong, L., 2013, *Backward Stochastic Differential Equations with Jumps and Their Actuarial and Financial Applications*, Springer-Verlag, LN.

[11] Delong, L. and Imkeller, P., 2010, *On Malliavin's differentiability of BSDEs with time delayed generators driven by Brownian motions and Poisson random measures*, Stochastic Processes and their Applications, 120, 9, 1748-1775.

[12] Di Nunno, G., Oksendal, B. and Proske, F., 2009, *Malliavin Calculus for Lévy Processes with Applications to Finance*, Springer, NY.

[13] Dzhaparidze, K. and Valkeila, E., 1990, *On the Hellinger type distances for filtered experiments*, Probab. Theory Related Fields, 85, 1, 105-117.

[14] Elie, R., 2006, *Contrôle stochastique et méthodes numériques en finance mathématique*, Ph.D Thesis, University Paris-Dauphine.

[15] El Karoui, N. and Mazliak, L. (eds.), 1997, *Backward stochastic differential equations*, Addison Wesley Longman Limited, U.S..

[16] El Karoui, N., Peng, S. and Quenez, M.-C., 1997, *Backward stochastic differential equations in finance*, Mathematical Finance, 7(1), 1-71.

[17] Fujii, M., 2014, *Momentum-space approach to asymptotic expansion for stochastic filtering*, Annals of the Institute of Statistical Mathematics, Vol. 66, 93-120.

[18] Fujii, M., 2015, *A polynomial scheme of asymptotic expansion for backward SDEs and option pricing*, Quantitative Finance, in press.

[19] Fujii, M., 2015, *Optimal position management for a market-maker with stochastic price impacts*, Working paper, CARF-F-360.

[20] Fujii, M. and Takahashi, A., 2012, *Analytical approximation for non-linear FBSDEs with perturbation scheme*, International Journal of Theoretical and Applied Finance, 15, 5, 1250034 (24).

[21] Fujii, M. and Takahashi, A., 2015, *Perturbative Expansion Technique for Non-linear FBSDEs with Interacting Particle Method*, Asia-Pacific Financial Markets, 22, 3, 283-304.

[22] Gikhman, I.I. and Skorohod, A.V., 1972, *Stochastic differential equations*, Springer, Berlin.

[23] Gobet, E., Lemor, J.P. and Warin, X., 2005, *A regression-based Monte Carlo method to solve backward stochastic differential equations*, The Annals of Applied Probability, 15, 3, 2172-2202.

[24] Hamadène, S. and Jeanblanc, M., 2007, *On the stating and stopping problem: Application in reversible investments*, Mathematics of Operations Research, Vol. 32, No. 1, pp. 182-192.

[25] Henry-Labordère, P., 2012, *Cutting CVA's complexity*, Risk Magazine, July Issue, 67-73.

[26] Igari, S., 1996, *Introduction of Real Analysis*, Iwanami publisher, in Japanese.

[27] Karatzas, I. and Shreve, S.E., 1998, *Brownian motion and stochastic calculus: Second edition*, Springer, NY.

[28] Kazamaki, N., 1979, *A sufficient condition for the uniform integrability of exponential martingales*, Math. Rep. Toyama Univ. 2, 1-11. MR-0542374.

[29] Kruse, T. and Popier, T., 2015, *BSDEs with monotone generator driven by Brownian and Poisson noises in a general filtration*, Stochastics, in press.

[30] Leoni, G., 2009, *A First Course in Sobolev Spaces*, Graduate Studies in Mathematics, Vol. 105, American Mathematical Society, US.

[31] Lim, A.E., *Quadratic Hedging and Mean-Variance portfolio selection with random parameters in an incomplete market*, Mathematics of Operations Research, Vol. 29, No. 1, 132-161.

[32] Liptser, R.Sh. and Shiryayev, A.N., 1989, *Theory of Martingales*, Mathematics and Its Applications, Kluwer Academic Publishers, LN.

[33] Ma, J. and Yong, J., 2000, *Forward-backward stochastic differential equations and their applications*,Springer, Berlin.

[34] Ma, J. and Zhang, J., 2002, *Representation theorems for backward stochastic differential equations*, The annals of applied probability, 12, 4, 1390-1418.

[35] Mania, M. and Tevzadze, R., 2003, *Backward stochastic PDE and imperfect hedging*, International Journal of Theoretical and Applied Finance, Vol. 6, 7, 663-692.

[36] Pardoux, E. and Peng, S., 1990, *Adapted solution of a backward stochastic differential equations*, Systems Control Lett., 14, 55-61.

[37] Pardoux, E. and Peng, S., 1992, *Backward stochastic differential equations and quasilinear parabolic partial differential equations*, Lecture Notes in Control and Inform. Sci. 176, 200-217, Springer, New York.

[38] Pardoux, E. and Rascanu, A., 2014, *Stochastic Differential Equations, Backward SDEs, Partial Differential Equations*, Springer International Publishing, Switzerland.

[39] Petrou, E., 2008, *Malliavin calculus in Lévy spaces and applications to finance*, Electric Journal of Probability, 13, 27, 852-879.

[40] Pham, H., 2010, *Stochastic control under progressive enlargement of filtrations and applications to multiple defaults risk management*, Stochastic processes and their Applications, 120, 1795-1820.

[41] Protter, P.E., 2005, *Stochastic Integration and Differential Equations*, Springer, NY.

[42] Takahashi, A., 2015, *Asymptotic Expansion Approach in Finance*, in P Fritz, J Gatheral, A Gulisashvili, A Jacquier and J Teichmann (eds.), *Large Deviations and Asymptotic Methods in Finance*, pp. 345-411, Springer.

[43] Takahashi, A. and Yamada, T., 2015, *On error estimates for asymptotic expansions with Malliavin weights: Application to stochastic volatility model*, Mathematics of Operations Research, Vol. 40, No. 3, pp. 513-541.

[44] Takahashi, A. and Yamada, T., 2015, *A weak approximation with asymptotic expansion and multidimensional Malliavin weights*, Annals of Applied Probability, in press.

[45] Takahashi, A. and Yamada, T., 2015, *An asymptotic expansion of forward-backward SDEs with a perturbed driver*, Forthcoming in International Journal of Financial Engineering.

[46] Touzi, N., 2013, *Optimal stochastic control, stochastic target problems, and backward SDEs*, Springer, New York.

[47] Zhang, J., 2004, *A numerical scheme for bsdes*, The annals of applied probability, 14, 1, 459-488.