Lp solution of backward stochastic differential equations driven by a marked point process

Fulvia Confortola

Abstract

We obtain existence and uniqueness in $L^p$, $p > 1$ of the solutions of a backward stochastic differential equation (BSDE for short) driven by a marked point process, on a bounded interval. We show that the solution of the BSDE can be approximated by a finite system of deterministic differential equations. As application, we address an optimal control problem for point processes of general non-Markovian type and show that BSDEs can be used to prove existence of an optimal control and to represent the value function.

Keywords Backward stochastic differential equations · Marked point processes · Stochastic optimal control

Mathematics Subject Classification 60H10 · 60G55 · 93E20

1 Introduction

In this paper, we study $L^p$ ($p > 1$) solutions of a backward stochastic differential equation (BSDE for short in the remaining) driven by a random measure, without diffusion part, on a finite time interval, of the following form:

$$Y_t + \int_t^T \int_K Z(s, x) \mu(ds, dx) = \xi + \int_t^T \int_K f(s, x, Y_s, Z_s(\cdot)) \nu(ds, dx). \quad (1.1)$$

(Here and in the following, the symbol $\int_a^b$ is to be understood as an integral over the interval $(a, b)$). The counting measure $\mu(ds, dx) = \sum_n \delta(S_n, X_n)(ds, dx)$ (where $\delta$ denotes the Dirac measure) corresponds to a marked point process $(S_n, X_n)_{n \geq 1}$, where $(S_n)$ is an increasing sequence of random times and $(X_n)$ a sequence of random vari-
ables in the state (or mark) space $K$. We denote $(A_t)$ the compensator of the counting process $(\mu([0, t] \times K))$ and by $\nu(dt, dx) = \phi_t(dx)dA_t$ the (random) compensator of $\mu$. The generator $f$ and the final condition $\xi$ are given. The unknown process is a pair $(Y_t, Z_t(\cdot))$, where $Y$ is a real adapted càdlàg process and $\{Z_t(x), t \in [0, T], x \in K\}$ is a predictable random field.

The random measure $\mu$ is fairly general, the only restriction being non-explosion (i.e., $S_n \to \infty$) and the requirement that $(A_t)$ has continuous trajectories. We allow the space $K$ to be of general type, for instance a Lusin space. Therefore, our results can also be directly applied to marked point processes with discrete state space.

The BSDEs have been introduced in their linear version by Bismut [9] in 1973, as an adjoint equation in the Pontryagin stochastic maximum principle, but the first study presenting a systematic treatment of nonlinear BSDEs is the seminal paper by Pardoux and Peng [43]. In this first paper, and in most of the subsequent ones, the driving term is a Brownian motion. Since then, there has been an increasing interest for this subject: these equations have a wide range of applications in various fields of stochastic analysis, including probabilistic techniques in partial differential equations, stochastic optimal control, mathematical finance (see, e.g., El Karoui, Peng and Quenez [25]). After the work [43], generalizations of the theory have followed several different paths: many papers have been devoted to existence and/or uniqueness results under weaker assumptions; in others the Brownian motion has been replaced with a more general process. Recently, BSDEs driven by a Brownian motion and Poisson random measure have also been considered due to their utility in the study of stochastic maximum principle, partial differential equations of non-local type, quasi-variational inequalities, impulse control and stochastic problem in stochastic finance. These so-called BSDEs with jumps have been introduced first by Tang and Li [47], followed notably by Barles, Buckdahn and Pardoux [7], Situ [46], Royer [44], Becherer [8], El Karoui, Matoussi and Ngupeyou [24], Kazi-Tani, Possamaï and Zhou [36,37], while the specific case of Lévy processes was treated by Nualart and Schoutens [42] and later Bahlali, Eddahbi and Essaky [4].

As in the original Brownian framework, in all these papers the existence of the solution of the BSDE relies upon the martingale representation theorem which holds in the case of a driving noise given by a Brownian motion and a more general compensated integer-valued random measure with deterministic compensator $\nu$ (see [9] Section 13.6.3, and Section 14.5 for general Lévy processes). This type of compensator allows to deal with explosive jump processes.

We address a class of purely discontinuous BSDEs driven by a random measure, naturally associated with a marked point process. We aim to find a solution by taking advantage of the martingale representation Theorem 5.4 in [34]. For this reason, we have to assume the process to be non-explosive (see Remark 1). We stress that the compensator $\nu(dt, dx) = \phi_t(dx)dA_t$ of a marked point process is random and not absolutely continuous in time with respect to the Lebesgue measure. It is well known that the martingale representation theorem does not hold in general when the compensator $\nu$ of the random measure $\mu$ is not deterministic, even if it is absolutely continuous in time with respect to the Lebesgue measure (see Example 18.4.5 in [15]).
We note, moreover, that the martingale representation theorem remains true if the noise has a Brownian part in addition to the compensated random measure and the paper admits a generalization in this direction.

In spite of the large literature devoted to BSDEs with driving term continuous or continuous-plus-jumps, there are relatively few results on the case of a driving term which is purely discontinuous. We cite Shen and Elliott [45] for the particularly simple “one-jump” case, or Cohen and Elliott [13,14] and Cohen and Szpruch [16] for BSDEs associated with Markov chains.

In [18] and [6] are considered BSDEs driven by more general random measures $\mu$ related to a Markov and semi-Markov jump process, respectively, and with compensator $\nu(dt, dx) = \phi_t(dx)dA_t$ such that the process $A$ is continuous. The non-Markovian case, under the same assumption on the process $A$, is studied in [17]. In [5] are obtained wellposedness results when the process $A$ is right-continuous and non-decreasing. In all these papers, existence, uniqueness and continuous dependence on the data of the solution of Eq. (1.1) have been proved in suitable weighted $L^2$ spaces, under $L^2$ summability condition on the data $\xi$ and $f$ and a Lipschitz condition on the generator $f$.

The first aim of this paper is to develop a $L^p$-theory for $p > 1$ for the class of BSDEs considered in [17].

From an abstract point of view, it seems the natural extension of the theory, but it is very important for the applications, to which the last section of the paper is devoted. In fact, in [6,8,17,18] this type of equations is used to solve optimal control problems for point processes, following a method well known in the optimal control problem for diffusions. Anyway the $L^2$ framework is too restrictive and does not allow to taking into account many optimal control problems with $p$-integrable costs which may not be square integrable.

The results of existence, uniqueness and continuous dependence on the data are presented in Sect. 3, after an introductory section devoted to notations and preliminaries. They are stated in the case of a scalar equation, but the extension to the vector-valued case is immediate.

The basic hypothesis on the generator $f$ is an uniform Lipschitz condition (see Hypothesis 1 for precise statements). In order to solve the equation, besides measurability assumptions, we require, for $p > 1$, the $L^p$ summability condition

$$\mathbb{E}\left[e^{\beta A_T} |\xi|^p \right] + \mathbb{E} \int_0^T \int_K e^{\beta A_t} |f(\omega, t, 0, 0)|^p \nu(dt, dx) < \infty,$$

to hold for a suitable $\beta$.

Note that the occurrence of exponentials of stochastic processes in the summability condition is imposed by the form of the compensator $\nu(dt, dx) = \phi_t(dx)dA_t$ of the random measure $\mu$ involving the process $A$. In the Poisson case mentioned above, we have a deterministic compensator $\nu(dt, dx) = \phi_t(dx)dA_t = \pi(dy)dt$ for some fixed measure $\pi$ on $K$, and the integrability condition reduces to a simpler form where the exponential terms disappear. For this reason, our results cannot be seen as a trivial generalization of the Poissonian case.
Some efforts have been made in relaxing the square integrability on the data. In the Brownian framework, we recall the papers of El Karoui et al. [25], Briand and Carmona [10], Briand et al. [11]. Recently, when the filtration is generated by the Brownian motion and a Poisson random measure, Yao [49] has considered the $L^p$ case, $p > 1$, Li and Wei [32] have given existence and uniqueness results for a fully coupled forward–backward SDE with $L^p$ coefficients, $p \geq 2$, Kruse and Popier [40] lately have studied a similar $L^p$ solution problem of BSDE under a right-continuous filtration which may be larger than the jump filtration. Finally, Eddahbi, Fakhouri, and Ouknine in [23] have addressed multidimensional generalized BSDEs in a more general filtration supporting a Brownian motion and an independent Poisson random measure. They have proved the existence and uniqueness of $L^p$ ($p \geq 2$)-solutions in the case of a fixed terminal time and of a random terminal time.

In the latter paper, the Lipschitz condition of the generator on variable $y$ has been replaced by a monotonicity condition. Our work could be extended in this direction. We note, anyway, that in the optimal control problem in general the generator of the BSDE does not depend on the variable $y$.

We recall that the $L^1$ theory for the solutions of Eq. (1.1) has been developed in [19]. Our $L^p$ assumptions are not in general comparable with the $L^1$ assumptions which involve suitable doubly weighted spaces. Moreover in [19] is assumed the continuity of the process $A$, as in our paper, and

$$\mathbb{P}(S_{n+1} > T \mid \mathcal{F}_{S_n}) > 0 \quad \text{for all } n \geq 0. \quad (1.2)$$

Since we use the martingale representation theorem and fixed point arguments, we do can avoid the last technical requirement to have existence and uniqueness of the solution of BSDE (1.1).

The second purpose of the paper is presented in Sect. 5, where we illustrate an approximation scheme to solve Eq. (1.1).

We show that the solution to (1.1) can be obtained as limit of a sequence of approximating BSDEs driven by a random measure with a finite number of jumps. We use the a priori estimates to obtain an error estimate for this approximation (see Proposition 1). It involves the distribution of the last jump time, and in particular cases it can be easily computed.

It is well known from Lemma 7 in [19] that, if $A$ is continuous and the technical assumption (1.2) holds, then BSDE (1.1) can be reduced to a deterministic ordinary differential equation. We give a condition on the process $A$ which implies the (1.2) (see Lemma 4). Moreover, with the localization procedure introduced in [19] we find a suitable approximating sequence of BSDEs driven by a random measure for which condition (1.2) holds. So the original BSDE can be approximated by a sequence of deterministic ordinary differential equations.

The method to reduce the BSDE to a sequence of ODEs has been used also for a BSDE driven by a Brownian motion plus a Poisson process, see, for example, Kharroubi and Lim [34]. In recent years, there has been much interest in numerical approximation of the solution to the BSDEs, in the context of diffusion processes. Our results might be used for similar methods in the framework of pure jump processes as well.
As said before, Sect. 6 is devoted to the study of an optimal control problem for a marked process, formulated in a classical way, with the approach based on the BSDEs. For every predictable control process, the cost to be minimized is

\[
J(u(\cdot)) = \mathbb{E}_u \left( \int_0^T l(t, u_t) \, dA_t + g(X_T) \right),
\]

where \( \mathbb{E}_u \) denotes the expectation with respect to a new probability \( \mathbb{P}_u \), depending on a control process \( (u_t) \) and defined by means of an absolutely continuous change of measure: the choice of the control process modifies the compensator of the random measure under \( \mathbb{P}_u \) making it equal to \( r(t, x, u_t) \nu(dt, dx) \) for some given function \( r \).

To solve the optimal control problem, we introduce the BSDE

\[
Y_t + \int_t^T \int_K Z(s, x) \mu(ds \, dx) = g(X_T) + \int_t^T f(s, Z_s(\cdot)) \, dA_s, \quad t \in [0, T]. \tag{1.3}
\]

where the generator contains the Hamiltonian function

\[
f(\omega, t, x, \zeta(\cdot)) = \inf_{u \in U} \left\{ l(\omega, t, u) + \int_K \zeta(x) r(\omega, t, x, u) \phi_t(\omega, dy) \right\}. \tag{1.4}
\]

Assuming that the infimum is in fact a minimum, admitting a suitable selector, together with a summability condition of the form

\[
\mathbb{E} \exp(\beta A_T) + \mathbb{E}[|g(X_T^{t, x})|^p e^{\beta A_T}] < \infty
\]

for a sufficiently large value of \( \beta \), we prove that the optimal control problem has a solution and that the value function and the optimal control can be represented by means of the solution to the BSDE. In this way, we are able to treat a large class of optimal control problems with costs which are not necessarily square integrable. We stress, moreover, that to solve the optimal control problem we assume that the action space \( U \) is the union of countably many compact metrizable subsets of itself (for example, \( \mathbb{R} \) or \( \mathbb{N} \)) avoiding the stronger conditions of compactness required, for instance, in [17–19].

2 Motivation of the problem

Models of optimal portfolio liquidation have received considerable attentions in the mathematical finance and stochastic control literature in recent years, see, e.g., [1,26], [27]. These models represent financial markets where traders face the task of selling a large amount of a given asset over a short time period and need to find an optimal portfolio execution. Usually the inventory, i.e., the number of shares held at time \( t \), is modeled by an absolutely continuous process. But if we suppose to submit block trades (for possible future execution) only to a crossing network or dark pool, the inventory is best described by a pure jump process. Dark pools are alternative trading
venues that allow investors shield their orders from public view and hence to reduce or to avoid market impact and trading costs. Since orders submitted to a dark pool are not openly displayed, order execution is uncertain and often modeled by a point process (see [39]).

The literature on optimal liquidation has so far been confined to Markovian models, where the cost functions are either deterministic or driven by stochastic factors that follow a Markovian dynamics. But real-world markets call for a general mathematical framework which allows for non-Markovian factor dynamics and explicit functional dependencies of the optimal liquidation strategies on the observable factor process (see [2,29,30,41]).

**Example** We consider a trader who wants to liquidate $x$ units of a stock over the period $[0, T]$ for a given time horizon $T$. We assume that she submits block trades only to a crossing network or dark pool and that the number of shares held at time $t \in [0, T]$ is given by a real-valued marked point process $X_t$. The random measure $\mu$ associated with the process $X_t$ governs dark pool executions.

The trader submits sell orders and can act modifying the rates of jumps of different sizes. Choosing a control process $u$, she modifies the compensator of the random measure $\mu$ under the probability $\mathbb{P}_u$ (depending on $u$).

She attempts to modify the probabilities of different paths of the state process $X_t$, in a dynamic way, to minimize the expected terminal cost given, for $p > 1$, by

$$J(u(\cdot)) := \mathbb{E}_u \left( \int_0^T \lambda_t |X_t|^p dA_t + |X_T|^p \right)$$

In a portfolio liquidation framework, the coefficients $\lambda_t(\omega)$ measure the investors desire for early liquidation (risk aversion). At time $T$, not all shares have been sold and the terminal cost models the liquidation value of the remaining share position $X_T$. $J$ is given by the cost of liquidating the portfolio comprising $x$ shares during the time interval $[0, T]$ and by the end cost which penalizes the non-full liquidation.

### 3 The setting

Let $T \in (0, \infty)$ be a fixed time horizon. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a complete probability space and $(K, \mathcal{K})$ a Lusin space. Assume we have a non-explosive multivariate point process (also called marked point process) on $[0, T] \times K$: this is a sequence $(S_n, X_n)_{n \geq 1}$ of random variables with distinct times of occurrence $S_n$ and with marks $X_n$. $S_n$ taking values in $(0, T] \cup \{\infty\}$ and $X_n$ in $K$. We set $S_0 = 0$, and we assume $\mathbb{P}$-a.s., $S_1 > 0$; if $S_n \leq T$, then $S_n < S_{n+1}$ and $S_n \leq S_{n+1}$ everywhere; $\Omega = \cup \{S_n > T\}$. Note that the “mark” $X_n$ is relevant on the set $\{S_n \leq T\}$ only, but it is convenient to have it defined on the whole set $\Omega$, and without restriction we may assume that $X_n = \Delta$ when $S_n = \infty$, where $\Delta$ is a distinguished point in $K$. 

\[\square\] Springer
The multivariate point process is completely characterized by the following discrete random measure on \((0, \infty) \times K\)

\[
\mu(dt, dx) = \sum_{n \geq 1: S_n \leq T} \delta(S_n, X_n)(dt, dx),
\]

(3.1)

where \(\delta(t, x)\) denotes the Dirac measure located at point \((t, x)\).

We introduce the \(\sigma\)-algebra

\[
\mathcal{F}_t := \sigma(\mu([0, t] \times B : s \in [0, t], B \in \mathcal{K}), t \geq 0
\]

and denote by \((\mathcal{F}_t)_{t \geq 0}\) the filtration generated by the point process, which is the smallest right-continuous filtration in which \(\mu\) is optional.

As we will see, the special structure of this filtration plays a fundamental role in all what follows. We let \(\mathcal{P}\) be the predictable \(\sigma\)-field on \(\Omega \times [0, T]\), and for any auxiliary measurable space \((G, \mathcal{G})\) a function on the product \(\Omega \times [0, T] \times G\) which is measurable with respect to \(\mathcal{P} \otimes \mathcal{G}\) is called predictable.

We set \(N\) is the univariate point process

\[
N_t = \mu([0, t] \times K) = \sum_{n \geq 1} 1_{\{S_n \leq t\}}.
\]

(3.2)

It is known that there exists an increasing càdlàg predictable process \(A\) starting at \(A_0 = 0\) such that

\[
\mathbb{E} \int_0^\infty H_t \ dN_t = \mathbb{E} \int_0^\infty H_t \ dA_t
\]

for every nonnegative predictable process \(H\). The above stochastic integrals are defined for \(\mathbb{P}\)-almost every \(\omega\) as ordinary (Stieltjes) integrals. \(A\) is called the compensator, or the dual predictable projection, of \(N\).

The following assumption will hold throughout:

**Assumption A** The process \(A\) is continuous.

Since \(N\) is an adapted locally integrable increasing process, this condition is equivalent to assume the jump times \(S_n\) to be totally inaccessible ([31], Corollary 5.28), or the process \(N\) to be quasi-left continuous ([31], Definition 4.22).

An adapted càdlàg process having only totally inaccessible jumps is said to be quasi-left continuous (Def 4.22). It can be proved (see [34] Sect. 2) that, if \((K, \mathcal{K})\) is a Lusin space with its Borel \(\sigma\)-algebra, there exists a function \(\phi_t(\omega, B)\) such that

1. for every \(\omega \in \Omega, t \in [0, \infty)\), the mapping \(B \mapsto \phi_t(\omega, B)\) is a probability measure on \((K, \mathcal{K})\);
2. for every \(B \in \mathcal{K}\), the process \((\omega, t) \mapsto \phi_t(\omega, B)\) is predictable;
3. for every nonnegative \(H_t(\omega, x), \mathcal{P} \otimes \mathcal{K}\)-measurable, we have

\[
\mathbb{E} \int_0^\infty \int_K H_t(x) \mu(\{x\}) = \mathbb{E} \int_0^\infty \int_K H_t(x) \phi_t(dx) dA_t.
\]
We will use the following notation

\[ \nu(\omega, dt, dx) = dA_t(\omega) \phi_t(\omega, dx), \quad (3.3) \]

and the random measure \( \nu(\omega, dt, dx) \) will be called the compensator, or the dual predictable projection, of \( \mu \) relative to the filtration \( (\mathcal{F}_t) \).

Finally, we recall that given a fixed \( T > 0 \) and a \( \mathcal{P} \otimes \mathcal{K} \)-measurable real function \( H_t(\omega, x) \) satisfying

\[ \int_0^T \int_K |H_t(x)| \nu(dt, dx) < \infty, \quad \mathbb{P} - a.s. \quad (3.4) \]

then the following stochastic integral can be defined

\[
\int_0^t \int_K H_s(x) (\mu(ds, dx) - \nu(ds, dx)) = \int_0^t \int_K H_s(x) \mu(ds, dx) - \int_0^t \int_K H_s(x) \nu(ds, dx), \quad t \in [0, T], \quad (3.5)
\]

as the difference of ordinary integrals with respect to \( \mu \) and \( \nu \). Moreover, under (3.4), it turns out to be a finite variation martingale.

**Example 1** 1. Let \( X \) be a continuous-time homogeneous Markov chain on \( K = \{1, 2, \ldots, N\} \) described by the rates transition matrix \( Q = (\lambda(i, j))_{i, j \in K} \) \( Q \) is a real square matrix such that

1. \( \lambda(i, j) \geq 0 \) for all \( i, j \in K, i \neq j \).
2. \( \sum_{j \in K} \lambda(i, j) = 0 \) for all \( i \in K \).

The waiting time in a state \( i \in K \) has exponential distribution with rate \( \lambda(i) = -\lambda(i, i) = \sum_{j \neq i} \lambda(i, j) \) and the probability to jump in a state \( j \neq i \) is \( q(i, j) := \frac{\lambda(i, j)}{\lambda(i)} \). (If \( i \) is an absorbing state \( \lambda(i) = 0 \) and we set \( q(i, j) := \delta_{ij} \).) In this case, the random measure \( \mu \) associated with the marked point process \( X \) is specified by

\[ \mu((0, t] \times \{j\}) = \sum_{n \geq 1} 1_{\{T_n \leq t\}} 1_{\{X_{T_n} = j\}} \]

and its compensator is defined

\[ \nu(dt, \{j\}) = \lambda(X_{t-}, j)dt = q(X_{t-}, j)\lambda(X_{t-})dt. \]

Hence, we deduce that

\[ \phi_t(\{j\}) = q(X_{t-}, j) \quad \text{and} \quad dA_t = \lambda(X_{t-})dt. \]
2. Let \((S_n)_{n \geq 1}\) be an independent sequence of random variables with exponential distribution of parameter \(\lambda > 0\) and define the sequence of jump times \((T_n)_{n \geq 1}\) by

\[
T_1 = S_1; \quad T_n = S_1 + \cdots + S_n.
\]

The process \(N_t = \sum_{n \geq 1} 1\{T_n \leq t\}\) is a Poisson process with parameter \(\lambda\). It is a marked point process with \(K = 1\) and

\[
\mu(dt \times \{1\}) = dN_t.
\]

Its compensator is the predictable measure defined by

\[
\nu(dt \times \{1\}) = \lambda dt.
\]

3. Let \(\mu\) be the Poisson random measure on \(\mathbb{R}^N\). In this case, its compensator is

\[
\nu(dt, dx) = \lambda(dx)dt
\]  

(3.6)

for some \(\lambda(x)\) deterministic, nonnegative and fixed measure on \(\mathbb{R}^N\). The associated marked point process is not explosive if \(\lambda(x)\) is finite.

4. Let \(L\) a pure jump Lévy process with Lévy measure (jump intensity measure) \(\lambda(dx)\) such that

\[
\nu(\{0\}) = 0 \quad \text{and} \quad \int_{\mathbb{R}^N} 1 \wedge |x|^2 \nu(dx) < \infty.
\]

The random measure

\[
\mu(dt, dx) = \sum_{t > 0} 1_{\{\Delta X_t \neq 0\}} \delta(t, \Delta X_t)(dt, dx)
\]

is called the Poisson measure of jumps of \(L\). Its compensator is again given by (3.6).

To guarantee \(L\) to be not explosive, we have to consider only finite jump intensity measures \(\lambda\).

**Remark 1** In the sequel, we use a martingale representation theorem for the random measure \(\mu\), namely Theorem 5.4 in [34]. For this reason, we require the multivariate point process to be non-explosive on \([0, T] \times K\) and \((F_t)_{t \geq 0}\) to be the natural filtration of \(\mu\). However, the results of this paper are still valid for random measure \(\mu\) associated with a general marked point process with a finite accumulation point (for which Theorem 5.4 in [34] holds).

It is worth noticing that if the random measure \(\mu\) has a deterministic compensator \(\nu\), then the martingale representation theorem remains true (see [15] Section 13.6.3, and Section 14.5 for general Lévy processes). Hence, in this case one can consider also explosive jump processes. The martingale representation theorem does not hold.
in general when the compensator $\nu$ of the random measure $\mu$ is not deterministic, also if it is absolutely continuous in time with respect to the Lebesgue measure. (see Example 18.4.5 in [15]).

We stress that we do not require the compensator $\nu$ to be neither deterministic nor absolutely continuous in time with respect to the Lebesgue measure (see 3.3).

4 The backward equation

We denote by $B(K)$ the set of all Borel functions on $K$; if $Z$ is a measurable function on $\Omega \times [0, T] \times K$, we write $Z_{\omega,t}(x) = Z(\omega, t, x)$, so each $Z_{\omega,t}$, often abbreviated as $Z_t$ or $Z_t(\cdot)$, is an element of $B(K)$.

In the following, we will consider the backward stochastic differential equation

$$Y_t + \int_t^T \int_K Z(s, x) \mu(ds, dx) = \xi + \int_t^T \int_K f(s, x, Y_s - , Z_s(\cdot)) \nu(ds, dx),$$  \hspace{0.5cm} (4.1)

where the generator $f$ and the final condition $\xi$ are given.

**Definition 1** A solution is a pair $(Y, Z)$ consisting in an adapted càdlàg process $Y$ and a predictable function $Z$ on $\Omega \times [0, T] \times K$ satisfying

$$\int_0^T \int_K |Z(t, x)| \nu(ds, dx) < \infty \text{ a.s.},$$

such that (4.1) holds for all $t \in [0, T]$, outside a $\mathbb{P}$-null set.

**Remark 2** Equation (4.1) can be rewritten as follows:

$$Y_t + \sum_{n \geq 1} Z(S_n, X_n) 1_{\{t < S_n \leq T\}} = \xi + \int_t^T \int_K f(s, x, Y_s - , Z_s(\cdot)) \nu(ds, dx).$$  \hspace{0.5cm} (4.2)

Since $A$ is continuous, (4.2) yields, outside a $\mathbb{P}$-null set:

$$\Delta Y_{S_n} = Z(S_n, X_n) \text{ if } S_n \leq T \text{ and } n \geq 1,$$

$Y$ is continuous outside $\{S_1, \ldots, S_n, \ldots\}$.  \hspace{0.5cm} (4.3)

In other words, $Y$ completely determines the predictable function $Z$ outside a null set with respect to the measure $\mathbb{P}(d\omega)\mu(\omega, dt, dx)$ and hence also outside a $\mathbb{P}(d\omega)\nu(\omega, dt, dx)$-null set. Equivalently, if $(Y, Z)$ is a solution and $Z'$ is another predictable function, then $(Y, Z')$ being another solution is the same as having $Z' = Z$ outside a $\mathbb{P}(d\omega)\mu(\omega, dt, dx)$-null set, and the same as having $Z' = Z$ outside a $\mathbb{P}(d\omega)\nu(\omega, dt, dx)$-null set.
Hence, it is possible to define a solution to (4.1) an adapted càdlàg process $Y$ for which there exists a predictable function $Z$ satisfying

$$
\int_0^T \int_K |Z(s, x)| \nu(ds, dx) < \infty \text{ a.s.,}
$$

such that the pair $(Y, Z)$ satisfies (4.1) for all $t \in [0, T]$, outside a $\mathbb{P}$-null set. Then, uniqueness of the solution means that, for any two solutions $Y$ and $Y'$, we have $Y_t = Y'_t$ for all $t \in [0, T]$, outside a $\mathbb{P}$-null set.

### 4.1 The $L^p$ theory, $p > 1$

We introduce the Banach space $L^p_\beta$, depending on a parameter $\beta > 0$, of equivalence classes of pairs of processes $(Y, Z)$ on $[0, T]$ such that $Y$ is progressive, $Z$ is predictable and the norm

$$
\|(Y, Z)\| = \mathbb{E} \left[ \int_0^T \int_K (|Y_t|^p + |Z(t, x)|^p) e^{\beta A t} \nu(dt, dx) \right]
$$

is finite. Elements of $L^p_\beta$ are identified up to almost sure equality with respect to the measure $\mathbb{P}(d\omega) \nu(dt, dx)$, i.e., when the norm of their difference is zero. We sometimes identify processes $(Y, Z)$ with their equivalence classes in the usual way.

Let us consider the following assumptions on the data $\xi$ and $f$:

**Hypothesis 1**

1. The final condition $\xi : \Omega \to \mathbb{R}$ is $\mathcal{F}_T$-measurable and $\mathbb{E} e^{\beta A T} |\xi|^p < \infty$.
2. $f$ is a real-valued function on $\Omega \times [0, T] \times K \times \mathbb{R} \times \mathcal{B}(K)$, such that
   
   (i) for any predictable function $Z$ on $\Omega \times [0, T] \times K$ the mapping
   
   $$(\omega, t, x, y) \mapsto f(\omega, t, x, y, Z_{\omega, t}(\cdot))$$

   is predictable;

   (ii) there exist $L \geq 0, L' \geq 0$ such that for every $\omega \in \Omega, t \in [0, T], x \in K, y, y' \in \mathbb{R}, \zeta, \zeta' \in \mathcal{B}(K)$ we have

   $$
   |f(\omega, t, x, y', \zeta) - f(\omega, t, x, y, \zeta)| 
   \leq L'|y' - y| \int_K |f(\omega, t, x, y, \zeta) - f(\omega, t, x, y, \zeta')| \phi_{\omega, t}(dx)
   \leq L \left( \int_K |\zeta'(v) - \zeta(v)|^p \phi_{\omega, t}(dv) \right)^{1/p} = L|\zeta - \zeta'|_{L^p(\phi_{\omega, t})}
   $$

   where $\phi_{\omega, t}$ are the measures occurring in (3.3);

   (iii) We have

   $$
   \mathbb{E} \int_0^T \int_K e^{\beta A t} |f(t, x, 0, 0)|^p \nu(dt, dx) < \infty \text{ a.s.}
   $$
The measurability condition (i) imposed on the generator is somehow awkward; however, it seems to be unavoidable. Indeed, we notice that the same condition is imposed in [19], assumption (2.8), and a similar condition is required in [17], assumption (3.2).

However, it is satisfied when we deal with a BSDE in order to solve an optimal control problem driven by a multivariate point processes. In this framework, the suitable formulation is the following one

\[ Y_t + \int_t^T \int_K Z(s, x) \mu(ds, dx) = \xi + \int_t^T \tilde{f}(s, Y_{s-}, \eta_s Z_s) dA_s, \quad (4.7) \]

where \( \eta_{\omega, t} \) is a real-valued map on \( B(K) \), with

\[ |\eta_{\omega, t} \xi - \eta_{\omega, t} \xi'| \leq \int_K |\xi'(v) - \xi(v)| \phi_{\omega, t}(dv). \quad (4.8) \]

\( Z \) predictable on \( \Omega \times [0, T] \times K \) \( \Rightarrow \) the process \( (\omega, t) \mapsto \eta_{\omega, t} Z_{\omega, t} \) is predictable,

\( \tilde{f} \) is a predictable function on \( \Omega \times [0, T] \times \mathbb{R} \times \mathbb{R} \) satisfying:

\[ |\tilde{f}(t, y', z') - \tilde{f}(t, y, z)| \leq L'|y' - y| + L|z' - z| \]

\[ \mathbb{E} \int_0^T \int_K |\tilde{f}(t, x, 0, 0)|^p e^{\beta A_t} \nu(dt, dx) < \infty \quad (4.9) \]

Equation (4.7) reduces to (4.1) upon taking

\[ f(\omega, s, x, y, \xi) = \tilde{f}(\omega, s, y, \eta_{\omega, s} \xi), \quad (4.10) \]

and (4.9) for \( \tilde{f} \) plus (4.8) for \( \eta_{\omega, t} \) yield, with the Hölder inequality, (4.5) for \( f \).

4.1.1 A priori estimate

In this section, we provide some a priori estimates for the solutions of Eq. 4.1.

We start with a Lemma of Ito type.

**Lemma 1** Let \( \beta \in \mathbb{R} \). If \((Y, Z) \in L^p_\beta\) is a solution of (4.1), we have almost surely

\[
|Y_t|^{p} e^{\beta A_t} + \int_t^T \int_K (|Y_{s-}| + Z_s(y))^{p} - |Y_{s-}|^p \mu(ds, dy) + \beta \int_t^T |Y_s|^p e^{\beta A_s} dA_s = |\xi|^p e^{\beta A_T} + \int_t^T \int_K p|Y_{s-}|^{p-1} \text{sign}(Y_{s-}) f(s, y, Y_s, Z_s(y)) e^{\beta A_s} \nu(ds, dy).
\]

(4.11)
Lemma 3 For every $s_n$'s, and $U_T = V_T$, it suffices to check that outside a null set we have $\Delta U_{s_n} = \Delta V_{s_n}$ and also $U_t - U_s = V_t - V_s$ if $s_n \leq t < s < s_{n+1} \wedge T$, for all $n \geq 0$. The first property is obvious because $\Delta Y_{s_n} = Z(S_{s_n}, X_{s_n})$ a.s. and $A$ is continuous. The second property follows from $Y_t - Y_s = \int_t^s \int K f(v, y, Y_r, Z_v(\cdot)) v(\text{d}v, \text{d}y)$, implying $|Y_t|^p - |Y_s|^p = \int_t^s \int K p|Y_r|^p-1 \text{sign}(Y_{r-}) f(r, y, Y_r, Z_r(\cdot)) v(\text{d}r, \text{d}x)$ plus a standard change of variables formula. \hfill \square

Lemma 2 Let $p > 1$ and $a, b \in \mathbb{R}$. Then for every $\epsilon > 0$

$$|a + b|^p \leq (1 + \epsilon)|a|^p + c_\epsilon |b|^p \tag{4.12}$$

where $c_\epsilon = \left(1 - \left(\frac{1}{1+\epsilon}\right)^{1-p}\right)^{1-p}$.

Proof By using the convexity of the function $x \mapsto |x|^p$, we deduce that for $\lambda \in (0, 1)$

$$|a + b|^p = \left|\lambda \frac{a}{\lambda} + (1 - \lambda) \frac{b}{1 - \lambda}\right|^p \leq \lambda \frac{|a|^p}{\lambda} + (1 - \lambda) \frac{|b|^p}{1 - \lambda}$$

$$= \lambda^{1-p} |a|^p + (1 - \lambda)^{1-p} |b|^p$$

Setting $\lambda^{1-p} = 1 + \epsilon$, we have the thesis. \hfill \square

Next we prove the following useful a priori estimates:

Lemma 3 For every $\epsilon > 0$ let $c_\epsilon = \left(1 - \left(\frac{1}{1+\epsilon}\right)^{1-p}\right)^{1-p}$.

Suppose that Hypothesis 1 holds with $\beta > 1 + \frac{c_\epsilon}{1+\epsilon} + p L' + (p - 1)(L (1 + \epsilon)) \frac{1}{1-p}$. Then there exist two constants $C_1$ and $C_2$ only depending on $(\beta, p, L, L', \epsilon)$, such that any pair $(Y, Z)$ in $L^p_\beta$ which solves (4.1) satisfies

$$\mathbb{E}|Y_t|^p e^{\beta A_t} \leq C_1 \mathbb{E}\left(|\xi|^p e^{\beta A_T} + \int_0^T \int_K |f(s, x, 0, 0)|^p e^{\beta A_s} v(ds, dy)\right) \tag{4.13}$$

$$\mathbb{E} \int_0^T \int_K (|Y|^p + |Z(s, x)|^p) e^{\beta A_s} v(ds, dx) \leq C_2 \mathbb{E}\left(|\xi|^p e^{\beta A_T} + \int_0^T \int_K |f(s, x, 0, 0)|^p e^{\beta A_s} v(ds, dy)\right) \tag{4.14}$$

Proof Inequality (4.12) with $a = Y_{s-} + Z(s, x)$ and $b = -Y_{s-}$ holds

$$|Z(s, x)|^p \leq (1 + \epsilon)|Y_{s-} + Z(s, x)|^p + c_\epsilon |Y_{s-}|^p.$$
It follows that
\[ |Y_{s-} + Z(s, x)|^p - |Y_{s-}|^p \geq \frac{1}{1+\epsilon} |Z(s, x)|^p - \left( 1 + \frac{c\epsilon}{1+\epsilon} \right) |Y_{s-}|^p. \]

So (4.11), and the fact that \( \phi_t, \omega_t(K) = 1 \) yield almost surely,
\[
|Y_t|^p e^{\beta A_t} + \frac{1}{1+\epsilon} \int_t^T \int_K |Z(s, x)|^p e^{\beta A_s} \mu(ds, dx) + \beta \int_t^T |Y_s|^p e^{\beta A_s} dA_s \\
\leq |\xi|^p e^{\beta A_T} + \left( 1 + \frac{c\epsilon}{1+\epsilon} \right) \int_t^T |Y_s|^p e^{\beta A_s} dN_s \\
+ p \int_t^T \int_K |Y_s|^{p-1} |f(s, x, Y_s, Z_s(\cdot))| e^{\beta A_s} v(ds, dx). \tag{4.15}
\]

Since the process \( |Z(s, x)|^p e^{\beta A_s} \) satisfies condition (3.4), the stochastic integral \( \int_t^T \int_K |Z(s, x)|^p e^{\beta A_s} (\mu(ds, dx) - v(ds, dx)) \) is a martingale. So taking the expectation in (4.15) yields
\[
\mathbb{E}[|Y_t|^p e^{\beta A_t}] \\
\quad + \mathbb{E} \int_t^T \int_K \left( \frac{1}{1+\epsilon} |Z(s, x)|^p + \left( \beta - 1 - \frac{c\epsilon}{1+\epsilon} \right) |Y_s|^p \right) e^{\beta A_s} v(ds, dx) \\
\leq \mathbb{E} \left( |\xi|^p e^{\beta A_T} + \int_t^T \int_K p |Y_s|^{p-1} |f(s, x, Y_s, Z_s(\cdot))| e^{\beta A_s} v(ds, dx) \right) \tag{4.16}
\]

From the Lipschitz condition of \( f \) and thanks to Young’s inequality, we have that
\[
\mathbb{E}[|Y_t|^p e^{\beta A_t}] + \mathbb{E} \int_t^T \int_K \left[ \frac{1}{1+\epsilon} |Z(s, x)|^p \right. \\
\quad + \left( \beta - 1 - \frac{c\epsilon}{1+\epsilon} \right) |Y_s|^p \left. \right] e^{\beta A_s} v(ds, dx) \\
\leq \mathbb{E} \left( |\xi|^p e^{\beta A_T} + \int_t^T \int_K p |Y_s|^{p-1} \left( |f(s, x, 0, 0)| + L'|Y_s| \right) e^{\beta A_s} v(ds, dx) \right) \\
\quad + \mathbb{E} \left( \int_t^T p |Y_s|^{p-1} L|Z_s|_{L^p(\phi)} e^{\beta A_s} dA_s \right) \\
\leq \mathbb{E}[|\xi|^p e^{\beta A_T}] + \mathbb{E} \int_t^T \int_K \left( \frac{p-1}{\gamma} \right)^{p-1} |f(s, x, 0, 0)|^p e^{\beta A_s} v(ds, dx) \\
\quad + \mathbb{E} \int_t^T \int_K \left( pL' + (p-1) \left( L^p \left( \frac{1+\epsilon}{\alpha} \right) \right)^{\frac{1}{p-1}} + \gamma \right) |Y_s|^p + \alpha \frac{1}{1+\epsilon} |Z_s|^p e^{\beta A_s} v(ds, dx),
\]
with $\alpha, \gamma > 0$. If we choose $\alpha \in (0, 1)$ and $\gamma = \frac{1}{2} \left[ \beta - 1 - \frac{c_\epsilon}{1+\epsilon} - pL' - (p - 1) \left( \frac{L^p (1+\epsilon)}{\alpha} \right)^{\frac{1}{p-1}} \right]$, when $(Y, Z) \in L^p_\beta$, this implies

$$
E|Y_t|^p e^{\beta A_t} + \frac{1 - \alpha}{1 + \epsilon} E \int_t^T \int_K |Z(s, x)|^p e^{\beta A_s} \nu(ds, dx) + E \int_t^T \int_K \frac{1}{2} \left( \beta - 1 - \frac{c_\epsilon}{1+\epsilon} - pL' - (p - 1)(2L^p(1+\epsilon))^{\frac{1}{p-1}} \right) |Y_s|^p \nu(ds, dx)
$$

$$
\leq E \left( |\xi|^p e^{\beta A_T} + \left( \frac{p-1}{\gamma} \right)^{p-1} \int_t^T \int_K |f(s, x, Y_s, Z(s, x)) - Z(s, x)|^p e^{\beta A_s} \nu(ds, dx) \right),
$$
giving us both (4.13) and (4.14).

\[ \square \]

### 4.1.2 Existence and uniqueness

In this section, we will give an existence and uniqueness result for Eq. (4.1).

**Theorem 2** Let be $\epsilon > 0$ and $c_\epsilon$ as in Lemma 2. Suppose that $\xi$ and the generator $f$ satisfy Hypothesis 1 and $\beta > 1 + \frac{c_\epsilon}{1+\epsilon} + pL' + (p - 1)((L + 1)^p(1+\epsilon))^{\frac{1}{p-1}}$. Then, there exists a unique pair $(Y, Z)$ in $L^p_\beta$ which solves BSDE (4.1).

**Remark 3** We highlight that Hypothesis 1 depends on the parameters $p$ and $\beta$. To ensure existence of a solution for fixed $p$, it has to be satisfied with $\beta$ chosen large enough.

We note that if $p$ tends to 1 the parameter $\beta$ gets arbitrarily large. If we relax the integrability assumption on the terminal condition $\xi$ we have to impose, as consequence, a stronger condition on $A_T$.

We recall finally that our $L^p$ assumptions are not in general comparable with the $L^1$ assumptions in [19] which involve suitable doubly weighted spaces.

**Proof** We start rewriting Eq. (4.1) in the following equivalent way

$$
Y_t + \int_t^T \int_K Z(s, x) (\mu(ds, dx) - \nu(ds, dx)) = \xi
$$

$$
+ \int_t^T \int_K f(s, x, Y_{s-}, Z(s, x)) - Z(s, x) \nu(ds, dx) \tag{4.17}
$$

This formulation singles out the “martingale increment”

$$
\int_t^T \int_K Z(s, x) (\mu(ds, dx) - \nu(ds, dx)).
$$

We define a mapping $\Phi$ from $L^p_\beta$ into itself such that $(Y, Z) \in L^p_\beta$ is a solution of BSDE (4.17) if and only if it is a fixed point of $\Phi$.
More precisely, given \((U, V) \in L^p_{\beta}\), let \((Y, Z) = \Phi(U, V)\) be the pair satisfying:

\[
Y_t + \int_t^T \int_K Z(s, y) (\mu(ds, dx) - \nu(ds, dx)) = \xi \\
+ \int_t^T \int_K [f(s, x, U_s, V_s(\cdot)) - V_s(\cdot)] \nu(ds, dx).
\]  

(4.18)

We show, first of all, that the pair \((Y, Z)\) exists. We start by observing that the random variable \(\int_0^T \int_K |f(s, x, U_s, V_s)| \nu(ds, dx)\) is \(p\) integrable. In fact, by the Lipschitz condition of \(f\) we have

\[
\int_t^T \int_K |f(s, x, U_s, V_s)| \nu(ds, dx) = \int_t^T \int_K e^{-\frac{\beta}{p} A_s} e^{\frac{\beta}{p} A_s} |f(s, x, U_s, V_s)| \nu(ds, dx) \\
\leq \int_t^T e^{-\frac{\beta}{p} A_s} e^{\frac{\beta}{p} A_s} \left[ L'|U_s| + L \left( \int_K |V_s|^p \phi_s(x) \right)^{\frac{1}{p}} + \int_K |f(s, x, 0, 0)| \phi_s(dx) \right] dA_s \\
\leq \left( \int_t^T e^{-\frac{\beta}{p} A_s} dA_s \right)^{\frac{p-1}{p}} \cdot \left[ L' \left( \int_t^T e^{\beta A_s} |U_s|^p dA_s \right)^{\frac{1}{p}} \\
+ L \left( \int_t^T \int_K e^{\beta A_s} |V_s|^p \nu(ds, dx) \right)^{\frac{1}{p}} \\
+ \left( \int_t^T \int_K e^{\beta A_s} |f(s, x, 0, 0)|^p \nu(ds, dx) \right)^{\frac{1}{p}} \right].
\]  

(4.19)

Since \(\frac{\beta}{p-1} \int_t^T e^{-\frac{\beta}{p-1} A_s} dA_s = e^{-\frac{\beta}{p-1} A_t} - e^{-\frac{\beta}{p-1} A_T} \leq e^{-\frac{\beta}{p-1} A_t}\), we arrive at

\[
\left( \int_t^T \int_K |f(s, x, U_s, V_s)| \nu(ds, dx) \right)^p \\
\leq C_p \left( \frac{p-1}{\beta} \right)^{p-1} e^{-\beta A_t} \\
\times \int_t^T \int_K e^{\beta A_s} ((L')^p|U_s|^p + L^p |V(s, x)|^p + |f(s, x, 0, 0)|^p) \nu(ds, dx) \\
\leq C(p, \beta, L', L) \int_t^T \int_K e^{\beta A_s} (|U_s|^p + |V(s, x)|^p \\
+ |f(s, x, 0, 0)|^p) \nu(ds, dx) < \infty,
\]

(4.20)

since \((U, V)\) are in \(L^p_{\beta}\) and (4.6) hold.

Consider the martingale \(\tilde{M}_t = e^{\mathcal{F}_t} \left[ \xi + \int_0^T \int_K [f(s, x, U_s, V_s) - V_s] \nu(ds, dx) \right], t \in [0, T]. \tilde{M}\) admits a right-continuous modification \(M\) (see, e.g., Corollary 2.48...
in [31]). Then, by the martingale representation Theorem 5.4 in [34] and noting that \( M \) is a \( p \)-integrable martingale there exists a predictable process \( Z \) with 
\[
\mathbb{E} \int_0^T \int_K |Z_s(x)| \nu(dx, dx) < \infty
\]
such that
\[
M_t = M_0 + \int_0^t \int_K Z(s, x) (\mu(ds, dx) - \nu(ds, dx)), \quad t \in [0, T].
\]

Define the process \( Y \) by
\[
Y_t = M_t - \int_0^t \int_K [f(s, x, U_s, V_s) - V_s] \nu(ds, dx), \quad t \in [0, T].
\]

By observing that \( Y_T = \xi \), we easily deduce that Eq. (4.18) is satisfied.

It remains to show that \((Y, Z) \in L^p_\beta\). It follows by (4.21) that
\[
Y_t = e^{\mathcal{F}_t} \left[ \xi + \int_t^T \int_K [f(s, x, U_s, V_s) - V_s] \nu(ds, dx) \right]
\]
and so, using (4.20), we obtain
\[
e^{\beta A_t |Y_t|^p} \leq 2^{p-1} e^{\beta A_t |e^{\mathcal{F}_t} \xi|^p}
\]
\[
+ 2^{p-1} e^{\beta A_t} \left| e^{\mathcal{F}_t} \int_t^T \int_K [f(s, x, U_s, V_s) - V_s] \nu(ds, dx) \right|^p
\]
\[
\leq 2^{p-1} e^{\beta A_t} \left[ e^{\beta A_T |\xi|^p} \right]
\]
\[
+ e^{\mathcal{F}_t} \left[ C_{(p, \beta, L', L, T)} \int_0^T \int_K e^{\beta A_t} (|U_s|^p + |V(s, x)|^p)
\]
\[
+ |f(s, x, 0, 0)|^p \nu(ds, dx) \right].
\]

Denoting by \( m_t \) the right-hand side of (4.22), we see that \( m \) is a martingale. In particular, for every stopping time \( \Sigma \) with values in \([0, T]\), we have
\[
\mathbb{E} e^{\beta A_\Sigma |Y_\Sigma|^p} \leq \mathbb{E} m_\Sigma = \mathbb{E} m_T < \infty
\]
by the optional stopping theorem. Next we define the increasing sequence of stopping times
\[
\Sigma_n = \inf \left\{ t \in [0, T] : \int_0^t \int_K e^{\beta A_s} (|Y_s|^p + |Z(s, x)|^p) \nu(ds, dx) > n \right\},
\]
with the convention \( \inf \emptyset = T \). The Itô formula (4.11) can be applied to \( Y, Z \) on the interval \([0, \Sigma_n]\). Hence, proceeding as in Lemma 3, we deduce
$$\mathbb{E} \int_0^T \int_K \left[ \frac{1}{2} \left( \beta - 1 - \frac{c_\epsilon}{1 + \epsilon} \right) |Y_s|^p + \frac{1}{1 + \epsilon} |Z(s, x)|^p \right] e^{\beta A_s} \, \nu(ds, dx)$$

$$\leq \mathbb{E} \left( |Y_{\Sigma_n}|^p e^{\beta A_{\Sigma_n}} \right) + c(\epsilon, \beta, p) \int_0^T \int_K |f(s, x, U_s, V_s) - V_s|^p e^{\beta A_s} \, \nu(ds, dx).$$

From (4.23) (with $\Sigma = \Sigma_n$), we deduce

$$\mathbb{E} \int_0^T \int_K e^{\beta A_s} (|Y_s|^p + |Z(s, x)|^p) \, \nu(dx, ds)$$

$$\leq c_1(\beta, \epsilon, p) \mathbb{E} e^{\beta A_T} |\xi|^p$$

$$+ c_2(\beta, \epsilon, p, L', L, T) \int_0^T \int_K e^{\beta A_s} (|U_s|^p + |V(s, x)|^p)$$

$$+ |f(s, x, 0, 0)|^p \, \nu(ds, dx).$$

(4.24)

By inequality (4.24) and by the definition of $\Sigma_n$, we deduce that for almost all $\omega \in \Omega$ there exists a $n_0(\omega) > 0$ such that for $n \geq n_0(\omega)$ $\Sigma_n(\omega) = T$, it follows that $\lim_n \Sigma_n = T$ a.s. Letting $n \to \infty$ in (4.24) we conclude that $(Y, Z) \in \mathcal{L}_\beta^p$.

Finally, we prove that the map $\Phi$ is a contraction. Let $(U^i, V^i), i = 1, 2$, be elements of $L_\beta^p$ and let $(Y^i, Z^i) = \Phi(U^i, V^i)$. Denote $\overline{Y} = Y^1 - Y^2$, $\overline{Z} = Z^1 - Z^2$, $\overline{U} = U^1 - U^2$, $\overline{V} = V^1 - V^2$, $\overline{f}_s = f(s, x, U^1_s, V^1_s) - f(s, x, U^2_s, V^2_s) - \overline{V}$. Lemma 3 applies to $\overline{Y}, \overline{Z}$. Noting that $\overline{Y}_T = 0$, we obtain

$$\mathbb{E} e^{\beta A_t} |\overline{Y}_t|^p$$

$$+ \mathbb{E} \int_t^T \int_K \left( \frac{1}{1 + \epsilon} |\overline{Z}(s, x)|^p + \left( \beta - 1 - \frac{c_\epsilon}{1 + \epsilon} \right) |\overline{Y}_s|^p \right) e^{\beta A_s} \, \nu(ds, dx)$$

$$\leq p \mathbb{E} \int_t^T \int_K e^{\beta A_s} |\overline{Y}_s|^{p-1} \left[ |\overline{Z}(s, x)| + |\overline{V}(s, x)| \right] \, \nu(ds, dx), \quad t \in [0, T].$$

From the Lipschitz conditions of $f$ and elementary inequalities, it follows that

$$\mathbb{E} \int_0^T \int_K \left[ \left( \beta - 1 - \frac{c_\epsilon}{1 + \epsilon} \right) |\overline{Y}_s|^p + \frac{1}{1 + \epsilon} |\overline{Z}(s, x)|^p \right] e^{\beta A_s} \, \nu(ds, dx)$$

$$\leq p(L + 1) \mathbb{E} \int_t^T e^{\beta A_s} |\overline{Y}_s|^{p-1} \left( \int_K |\overline{V}(s, x)|^p \, \phi_s(dx) \right)^{1/p} \, dA_s$$

$$+ pL \mathbb{E} \int_t^T \int_K e^{\beta A_s} |\overline{Y}_s|^{p-1} |\overline{U}_s| \, \nu(ds, dx)$$

$$\leq \frac{\alpha}{1 + \epsilon} \mathbb{E} \int_t^T \int_K e^{\beta A_s} |\overline{V}(s, x)|^p \, \nu(ds, dx)$$

$$+ (p - 1) \left( (L + 1)^p \cdot \frac{1 + \epsilon}{\alpha} \right)^{p-1} \mathbb{E} \int_t^T \int_K e^{\beta A_s} |\overline{Y}_s|^p \, \nu(ds, dx)$$

$$+ \mathbb{E} \int_t^T \int_K e^{\beta A_s} |\overline{Y}_s|^p \, \nu(ds, dx).$$
for every $\alpha > 0$, $\gamma > 0$. This can be written

$$
\left[ \beta - 1 - \frac{c_\epsilon}{1 + \epsilon} - (p - 1) \left( \frac{(L + 1)^p(1 + \epsilon)}{\alpha} \right)^{\frac{1}{p-1}} \right]
\times \mathbb{E} \int_0^T \int_K |Y_s|^p \nu(ds, dx) + \mathbb{E} \int_0^T \int_K \left[ \frac{1}{1 + \epsilon} |Z(s, x)|^p \right] e^{B_{\alpha}s} \nu(ds, dx)
\leq \mathbb{E} \int_0^T \int_K \left[ L' \left( \frac{1}{\gamma} \right)^{p-1} |U_s|^p + \frac{\alpha}{1 + \epsilon} |V(s, x)|^p \right] e^{B_{\alpha}s} \nu(ds, dx).$$

By the assumption on $\beta$, it is possible to find $\alpha \in (0, 1)$ such that

$$
\beta > 1 + \frac{c_\epsilon}{1 + \epsilon} + (p - 1) \left( \frac{(L + 1)^p(1 + \epsilon)}{\alpha} \right)^{\frac{1}{p-1}} + pL' \frac{1}{\sqrt{ \alpha}}.
$$

If $L' = 0$, we see that $\Phi$ is an $\alpha$-contraction on $L^p_{\beta}$ endowed with the equivalent norm

$$
(Y, Z) \mapsto \left[ \beta - 1 - \frac{c_\epsilon}{1 + \epsilon} - (p - 1) \left( \frac{(L + 1)^p(1 + \epsilon)}{\alpha} \right)^{\frac{1}{p-1}} \right]
\times \mathbb{E} \int_0^T \int_K |Y_s|^p \nu(ds, dx) + \mathbb{E} \int_0^T \int_K \left[ \frac{1}{1 + \epsilon} |Z(s, x)|^p \right] e^{B_{\alpha}s} \nu(ds, dx).
$$

(4.25)

If $L' > 0$, we choose $\gamma = 1/ \sqrt{ \alpha}$ and obtain

$$
\mathbb{E} \int_0^T \int_K \left[ \frac{L'}{\sqrt{ \alpha}} |Y_s|^p, \nu(ds, dx) + \frac{1}{1 + \epsilon} |Z(s, x)|^p \right] e^{B_{\alpha}s} \nu(ds, dx)
\leq \mathbb{E} \int_0^T \int_K \left[ L' \left( \frac{1}{\sqrt{ \alpha}} \right)^{p-1} |U_s|^p + \frac{1}{1 + \epsilon} |V(s, x)|^p \right] e^{B_{\alpha}s} \nu(ds, dx)
= \alpha \mathbb{E} \int_0^T \int_K \left[ \frac{L'}{\sqrt{ \alpha}} |Y|^p + \frac{1}{1 + \epsilon} |Z|^p \right] e^{B_{\alpha}s} \nu(ds, dx),
$$

so that $\Phi$ is an $\alpha$-contraction on $L^p_{\beta}$ endowed with the equivalent norm

$$
(Y, Z) \mapsto \mathbb{E} \int_0^T \int_K \left[ \frac{L'}{\sqrt{ \alpha}} |Y|^p + \frac{1}{1 + \epsilon} |Z|^p \right] e^{B_{\alpha}s} \nu(ds, dx).
$$
In all cases, there exists a unique fixed point which is the required unique solution to BSDE (4.17).

**Remark 4** Under Hypothesis 1, we have existence of the solution to BSDE (4.1), in the sense of Definition 1. In contrast, the uniqueness holds in the smaller subclasses $L^p_{F}$, but it is not guaranteed within the class of all possible solutions as shown by the following example.

Consider a univariate point process. The space $K = \{\Delta\}$ is a singleton, and $N_t = 1_{\{S \leq t\}}$, where $S$ is a variable with values in $(0, T) \cup \{\infty\}$. The filtration $(\mathcal{F}_t)$ is still the one generated by $N$, and $G$ denotes the law of $S$, whereas $g(t) = G((t, \infty])$.

We suppose that $G$ has no atom, but is supported by $[0, T]$. We have $A_t = a(t \wedge S)$, where $a(t) = -\log g(t)$ is increasing, finite for $t < v$ and infinite if $t \geq v$, where $v = \inf(t : g(t) = 0) \leq T$ is the right end point of the support of the measure $G$.

Consider the following equation with $\xi = 0$ and $f(t, x, y, z) = z$:

$$
Y_t + \int_{(t,T]} Z_s (dN_s - dA_s) = 0.
$$

(4.26)

The pair $(Y_t, Z_t)$ with $Y_t = w e^{A_t} 1_{[t < S]}$ and $Z_t = -Y_t -$ is solution for any $w \in \mathbb{R}$ (see [19], Proposition 11). The only solution in $L^p_{F}$ is $Y \equiv 0$, $Z \equiv 0$.

## 5 An approximation scheme

In this section, we show that, under Assumption A and Hypothesis 1, the solution of Eq. (4.1) can be approximated by the solution of another BSDE driven by random measures with a finite number of jumps. Moreover, the solution of Eq. (4.1) can be obtained as limit of a finite sequence of deterministic differential equation.

For each (finite) integer $m \geq 1$, we set $T_m = S_m \wedge \inf(t : A_t \geq m)$ and we consider the BSDE

$$
Y^m_t + \int_t^T \int_K Z^m(s, x) \mu^m(dx, dx) = \xi^m
$$

$$
+ \int_t^T \int_K f(s, x, Y^m_s, Z^m_s(\cdot)) 1_{s \leq T_m} \nu^m(ds, dx)
$$

$$
\mu^m(ds, dx) = \mu(ds, dx) 1_{s \leq T_m}, \quad \nu^m(ds, dx) = \nu(ds, dx) 1_{s \leq T_m},
$$

$$
\xi^m = \xi 1_{T < T_m}.
$$

(5.1)

Then $\nu^m$ is the compensator of $\mu^m$, relative to $(\mathcal{F}_t)$ and also to the smaller filtration $(\mathcal{F}_t^{(m)} = \mathcal{F}_{t \wedge T_m})$ generated by $\mu^m$, whereas $\xi^m$ is $\mathcal{F}_T^{(m)}$-measurable.

We note that the generator $f(s, x, y, z) 1_{s \leq T_m}$ and the terminal condition $\xi^m$ satisfy Hypothesis 1 with respect to $(\mathcal{F}_t \wedge T_m)$. By Theorem 2, there exists a solution $(Y^m, Z^m)$ in $L^p_{F}$ for $\beta > 1 + \frac{c_\epsilon}{1+\epsilon} + pL' + (p - 1)((L + 1)p(1 + \epsilon))^{\frac{1}{p-1}}$. We may assume $Y^m_s = \xi^m = 0$, $Z^m_s = 0$ for $T_m < s \leq T$, so that BSDE (5.1) becomes

\[\square\] Springer
\[
Y_t^m + \int_t^T \int_K Z^m(s, x) \mu(ds, dx) = \xi^m
\]
\[
+ \int_t^T \int_K f(s, x, Y_s^m, Z^m(s, x)) 1_{s \leq T_m} \nu(ds, dx), \quad t \in [0, T],
\] (5.2)

and we have \( \mathbb{E}\left[ \int_0^T \int_K (|Y^m_s|^p + |Z(s, x)^m|^p) e^{\beta As} \nu(ds, dx) \right] < \infty \). Clearly, \( Y^m \) is adapted and \( Z^m \) is predictable also with respect to the natural filtration \( (\mathcal{F}_t) \) of \( \mu \).

**Proposition 1** Let \((Y^m, Z^m)\) and \((Y, Z)\) be the solutions in \( L_\beta^p \) to BSDEs (5.2) and (4.1), respectively. Then, if \( \beta > 1 + \frac{c_\mu}{1+\epsilon} + pL' + (p-1)((L+1)p(1+\epsilon))^{\frac{1}{p-1}} \), there exists a constant \( C \) independent of \( m \) such that

\[
\mathbb{E}\int_0^T \int_K (|Y_s - Y^m_s|^p + |Z(s, x) - Z^m(s, x)|^p) e^{\beta As} \nu(ds, dx) \leq C \mathbb{E}\left[ |\xi|^p e^{\beta AT} 1_{T_m < T} + \int_{T_m \wedge T} \int_K |f(s, x, 0, 0)|^p e^{\beta As} \nu(ds, dx) \right].
\]

**Proof** We define \( \tilde{Y} = Y - Y^m, \tilde{Z} = Z - Z^m \), and note that \((\tilde{Y}, \tilde{Z})\) is the solution, for \( t \in [0, T] \), to

\[
\tilde{Y}_t + \int_t^T \int_K \tilde{Z}(s, x) \mu(ds, dx) = \tilde{\xi} + \int_t^T \int_K \tilde{f}(s, x, \tilde{Y}_s, \tilde{Z}(s, x)) \nu(ds, dx),
\]

where we have set \( \tilde{f}(s, x, y, z) = f(s, x, Y_s^m + y, Z_s^m + z) - f(s, x, Y_s^m, Z_s^m) 1_{s \leq T_m} \) and \( \tilde{\xi} = \xi - \xi^m \). We deduce from the a priori estimate (4.14) that

\[
\mathbb{E}\int_0^T \int_K (|\tilde{Y}_s|^p + |\tilde{Z}(s, x)|^p) e^{\beta As} \nu(ds, dx) \leq C \mathbb{E}\left[ |\tilde{\xi}|^p e^{\beta AT} + \int_0^T |\tilde{f}(s, x, 0, 0)|^p e^{\beta As} \nu(ds, dx) \right] = C \mathbb{E}\left[ |\tilde{\xi}|^p e^{\beta AT} 1_{T_m < T} + \int_{T_m \wedge T} \int_K |f(s, x, 0, 0)|^p e^{\beta As} \nu(ds, dx) \right].
\]

where the last equality holds because \( Y_s^m = 0, \quad Z_s^m = 0 \) for \( T_m < s \leq T \). It follows that

\[
\mathbb{E}\int_0^T \int_K (|Y_s - Y^m_s|^p + |Z(s, x) - Z^m(s, x)|^p) e^{\beta As} \nu(ds, dx) \leq C \mathbb{E}\left[ |\xi|^p e^{\beta AT} 1_{T_m < T} + \int_{T_m \wedge T} \int_K |f(s, x, 0, 0)|^p e^{\beta As} \nu(ds, dx) \right].
\] (5.3)
Remark 5 Since the right-hand side in (5.3) tends to 0 as \( m \to \infty \), Proposition 1 ensures that it is possible to approximate the solution \((Y, Z)\) to Eq. (4.1), with the solution to BSDE (5.2) driven by random measures with a finite number of jumps. Moreover, it furnishes the error estimate for this approximation. In the case when \( A_T, \xi \) and \( f(s, x, 0, 0) \) are uniformly bounded, the approximation error can easily be expressed in terms of \( P(S_m < T) \). Under \( L_p^p \) integrability conditions in Hypothesis 1, a similar result can be obtained using Hölder inequality.

Our next step is to prove how is it possible to reduce the problem of solving Eq. (5.2) to solving a finite sequence of ordinary differential equation.

To this end, it is worth recalling that the filtration \((F_t)\) generated by the marked point process \( \mu \) has a very special structure, which reflects on adapted or predictable processes and hence on the solution \((Y^m, Z^m)\) of Eq. (5.2).

We need to introduce some notations, that might look complicated at first glance, but they indeed allow us to replace random elements by deterministic functions of all history \(((S_0, X_0), \ldots, (S_n, X_n))\) of the marked point process.

The process \((S_n, X_n)\) takes its values in the set \( S = ([0, T] \times K) \cup \{\infty, \Delta\} \). For any integer \( n \geq 0 \), we let \( H_n \) be the subset of \( S^{n+1} \) consisting in all \( D = ((t_0, x_0), \ldots, (t_n, x_n)) \) satisfying

\[
\begin{align*}
t_0 &= 0, \quad x_0 = \Delta, \quad t_{j+1} \geq t_j, \quad t_j \leq T \implies t_{j+1} > t_j, \\
t_j > T &\implies (t_j, x_j) = (\infty, \Delta).
\end{align*}
\]

We set \( D_{\text{max}}^\ast = t_n \) and endow \( H_n \) with its Borel \( \sigma \)-field \( \mathcal{H}_n \). We set \( S_0 = 0 \) and \( X_0 = \Delta \), so

\[
D_n = ((S_0, X_0), \ldots, (S_n, X_n))
\]

is a random element with values in \( H_n \), whose law is denoted as \( \Lambda_n \) (a probability measure on \((H_n, \mathcal{H}_n)\)).

The process \( Y^m \) solution to (5.2) is an adapted càdlàg process, which is further continuous outside the times \( S_n \). Hence, for each \( 0 \leq n \leq m \) there is a Borel function \( y^n = y^n(t) \) on \( H_n \times [0, T] \) such that

\[
D_{\text{max}}^\ast = \infty \implies y^n(t) = 0
\]

\[
t \mapsto y^n(t) \text{ is continuous on } [0, T] \text{ and constant on } [0, T \land D_{\text{max}}^\ast]
\]

\[
S_n(\omega) \leq t < S_{n+1}(\omega), \quad t \leq T \implies Y(t) = y^n_{D_n(\omega)}(t),
\]

and we express this as \( Y \equiv (y^n)_{n=0}^m \). Also the component \( Z^m \) of the solution to (5.2) can be expressed as \( Z^m \equiv (z^n)_{n=0}^m \) where \( z^n = z^n_D(t, x) \) is a Borel function on \( H_n \times [0, T] \times E \) such that

\[
D_{\text{max}}^\ast = \infty \implies z^n(t, x) = 0
\]

\[
S_n(\omega) < t \leq S_{n+1}(\omega) \land T \implies Z(t, x) = z^n_{D_n(\omega)}(t, x).
\]

The generator \( f 1_{s \leq S_m} \) has a nice predictability property only after plugging in a predictable function \( Z \). This implies that, for any \( 0 \leq n \leq m \), and \( z^n = z^n_D(t, x) \), one
Lemma 4 Assume which is continuous in $t$ has a Borel function $f(z^n)^n$ such that (with $t \leq T$ below)

$$D^\text{max} = \infty \Rightarrow f(z^n)^n(t, x, y) = 0$$

$$S_n(\omega) < t \leq S_{n+1}(\omega), \; \xi(x) = w + z^n_{D_n(\omega)}(t, x) \Rightarrow f(\omega, t, x, y, \xi) = f(z^n)^n_{D_n(\omega)}(t, x, y, w). \quad (5.7)$$

The variable $\xi^m$ is $\mathcal{F}_T$-measurable; hence, for each $0 \leq n \leq m$ there is an $\mathcal{H}_n$-measurable map $D \mapsto u_D^n$ on $H_n$ with

$$D^\text{max} = \infty \Rightarrow u_D^n = 0$$

$$S_n(\omega) \leq T < S_{n+1}(\omega) \Rightarrow \xi^m(\omega) = u_D^n(\omega). \quad (5.8)$$

We state now a technical result, needed for proving Proposition 2.

Lemma 4 Assume (A). Then, for $n \geq 0$ the inequality $\mathbb{E}[e^{\beta A_T \wedge S_n+1}] < \infty$ implies

$$\mathbb{P}(S_{n+1} > T | \mathcal{F}_{S_n}) > 0 \; \text{a.s.} \quad (5.9)$$

In particular, if

$$\mathbb{E}[e^{\beta A_T}] < \infty, \quad (5.10)$$

then (5.9) holds true for every $n \geq 0$.

Proof Let $G^m_{D_n}(dt)$ be the conditional law of $S_{n+1}$ given $\mathcal{F}_{S_n}$. Let us introduce the corresponding cumulative distribution function $F_D(t) = G^m_{D}(0, t]$. Since we assume that the dual predictable projection $A$ of $\mu$ is continuous, we can take a version of $F_D$ which is continuous in $t$, and we have $\mathbb{P}$-a.s.,

$$A_t = A_{S_n} + \int_{S_n}^{t} \frac{1}{1 - F_{D_n}(s)} F_{D_n}(ds) = A_{S_n} - \log(1 - F_{D_n}(t)), \; S_n < t \leq S_{n+1}. \quad (5.11)$$

Since $F_D$ is continuous in $t$, the conditional law of $F_D(S_{n+1})$ given $\mathcal{F}_{S_n}$ is the uniform distribution on $(0, 1)$, so that in particular $\mathbb{E}[(1 - F_{D_n}(S_{n+1}))^{-1} | \mathcal{F}_{S_n}] = \infty$ a.s. Now suppose that (5.9) is violated for some $n$. Then there exists $Q \in \mathcal{F}_{S_n}$ with $\mathbb{P}(Q) > 0$ such that $\mathbb{P}(S_{n+1} \leq T | \mathcal{F}_{S_n}) = 1$ on $Q$. Then

$$\mathbb{P}(Q) = \mathbb{E}[1_Q \mathbb{P}(S_{n+1} \leq T | \mathcal{F}_{S_n})] = \mathbb{P}(Q \cap \{S_{n+1} \leq T\}),$$

which shows that $S_{n+1} \leq T$ a.s. on $Q$. It follows from (5.11) that

$$\mathbb{E}[e^{\beta A_T \wedge S_{n+1}}] \geq \mathbb{E}[1_Q e^{\beta A_T \wedge S_{n+1}}] \geq \mathbb{E}[1_Q e^{A_{S_{n+1}}}]$$

$$\geq \mathbb{E}[1_Q (1 - F_{D_n}(S_{n+1}))^{-1}] = \mathbb{E}[1_Q \mathbb{E}[(1 - F_{D_n}(S_{n+1}))^{-1} | \mathcal{F}_{S_n}]) = \infty,$$

contradicting the assumption. The first part of the lemma is therefore proved.
Next assume (5.10). Then, the conclusion follows from the statement proved above noting that
\[ \infty > \mathbb{E}[e^{\beta A_T}] \geq \mathbb{E}[e^{\beta A_{T \wedge S_n}}]. \]

\[ \square \]

**Proposition 2** \( Y^m \equiv (y^n)_{n=0}^m \) is a solution if and only if for \( P \)-almost all \( \omega \) we have:

\[ t \in [0, T] \Rightarrow y^m_{D_n(\omega)}(t) = u^m_{D_n(\omega)} = \xi^m(\omega). \] (5.12)

and for all \( n = 0, \ldots, m - 1 \) and \( t \in [0, T] \)

\[ y^n_{D_n(\omega)}(t) = u^n_{D_n(\omega)} + \int_t^T \int_K f(\{\hat{y}^{n+1}\}_{D_n(\omega)}(s, x, y^n_{D_n(\omega)}(s), -y^n_{D_n(\omega)}(s))) v^n_{D_n(\omega)}(ds, dx), \] (5.13)

where we set

\[ \hat{y}^{n+1} = (\hat{y}^{n+1}_{D_n}(t, x) : (D, t, x) \in H_n \times [0, T] \times E) : \hat{y}^{n+1}_{D_n}(t, x) = y^{n+1}_{D \cup \{(t, x)\}}(t) 1_{\{t > D_{\max}\}}. \] (5.14)

**Proof** The process \( A^m_{\tau^m} = w^m([0; \tau] \times E) \) satisfies \( A^m_{\tau^m} \leq m \); hence, Lemma 4 ensures that condition 5.9 holds. So we can apply Lemma 7 in [19] that with Assumption A give us the thesis.

\[ \square \]

**Remark 6** Approximation results of this type are interesting from a numerical point of view, in the sense that they might give rise to simple numerical ways for solving the BSDEs driven by random measures associated with marked point processes. We recall that in the \( L^2 \) framework there is no similar result (compare, e.g., [17,18]). In the paper [19], instead, where the \( L^1 \) theory is developed, the solution of BSDE is obtained with a localization method as limit of the sequence of processes \( (Y^m, Z^m) \) solution of the truncated BSDEs (5.2). But condition (5.9) is assumed. Here in the Lemma we provide a condition which implies it and we show that it is verified by the truncated marked point process associated with the random measure \( \mu^m \) defined in (5.2).

**6 Optimal control**

In this section, we solve an optimal control problem for a marked point process by using an approach, well known in the Brownian framework, based on the BSDEs. We assume that on \( (\Omega, \mathcal{F}, \mathbb{P}) \) is given a marked point process \( (S_n, X_n) \), associated with the random measure \( \mu(dt, dx) = \sum_{n \geq 1; s_n \leq t} \delta_{(S_n, X_n)}(dt, dx) \) and generating the filtration \( (\mathcal{F}_t) \). The compensator \( \nu \) of \( \mu \) with respect to this filtration admits the disintegration \( \nu(\omega, dt, dx) = dA_t(\omega) \phi_{\omega,t}(dx) \). We work under assumption that the
process $A$ is continuous (Assumption A). In particular, we suppose that $S_n \to \infty$ $\mathbb{P}$-a.s. We finally fix $X_0 \in K$ (deterministic), and we define

$$X_t = \sum_{n\geq 0} X_n 1_{[S_n, S_{n+1})}(t), \quad t \geq 0. \quad (6.1)$$

The data specifying the optimal control problem are an action (or decision) space $U$, and a function $r$ specifying the effect of the control process, a running cost function $l$, a terminal cost function $g$. We assume that these data satisfy the following conditions.

**Hypothesis 3**

1. $(U, U)$ is a measurable space.
2. The functions $r : \Omega \times [0, T] \times K \times U \to \mathbb{R}$ and $l : \Omega \times [0, T] \times U \to \mathbb{R}$ are predictable, and there exists a constant $C_r > 1$, such that

$$0 \leq r(\omega, t, x, u) \leq Cr, \quad \omega \in \Omega, \; t \in [0, T], \; x \in K, \; u \in U, \quad (6.2)$$

and moreover

$$\mathbb{E} \int_0^T e^{\beta A_t} \sup_{u \in U} |l(t, u)|^p dA_t < \infty \quad (6.3)$$

3. The function $g : \Omega \times K \to \mathbb{R}$ is $\mathcal{F}_T \otimes K$-measurable and

$$\mathbb{E}[|g(X_T)|^p e^{\beta A_T}] < \infty \quad (6.4)$$

for some $\beta > 0$.

An admissible control process is any predictable process $(u_t)_{t \in [0, T]}$ with values in $U$. The set of admissible control processes is denoted $\mathcal{A}$.

The function $r$ models the effect of a control in the following sense. A controller will act to modify the measure $\mathbb{P}$ under which our system evolves, replacing it with the measure $\mathbb{P}_u$ on $(\Omega, \mathcal{F})$ for every control $u(\cdot) \in \mathcal{A}$ by a change of measure of Girsanov type, as we now describe. We define

$$\mathcal{E}_t^u := \exp \left( \int_0^t \int_K (1 - r(s, x, u_s)) \nu(ds, dx) \right) \prod_{n \geq 1 : S_n \leq t} r(S_n, X_n, u_{S_n}), \quad t \in [0, T],$$

with the convention that the last product equals 1 if there are no indices $n \geq 1$ satisfying $S_n \leq t$. $\mathcal{E}$ is a nonnegative supermartingale (see [34] Proposition 4.3). Moreover, the following result holds

**Lemma 5** ([Lemma 4.2, [17]]) Let $\gamma > 1$ and

$$\beta = \gamma + 1 + \frac{C_r^2}{\gamma - 1}. \quad (6.5)$$

If $\mathbb{E} \exp(\beta A_T) < \infty$, then we have $\sup_{t \in [0, T]} \mathbb{E}(\mathcal{E}_t^u)^\gamma < \infty$ and $\mathbb{E}(\mathcal{E}_T^u) = 1$ for all admissible controls $u$. 
Under the assumption of the lemma, the process $E$ is a martingale and this guarantees that $P_u$, defined as $P_u(d\omega) = E^u_T(\omega)P(d\omega)$, is a true probability. By Girsanov’s theorem for point processes (Theorem 4.5 in [34]), the predictable compensator of the measure $\mu$ under $P_u$ is

$$ν^u(dt, dx) = r(t, x, u_t) ν(dt, dx) = r(t, x, u_t) φ_t(dx) dA_t.$$

Therefore, we see that our controller is effectively modifying the rates of jumps of different sizes.

We finally define the cost associated with every $u(\cdot) ∈ A$ as

$$J(u(\cdot)) = E_u \left( ∫_0^T l(t, u_t) dA_t + g(X_T) \right),$$

where $E_u$ denotes the expectation under $P_u$. The control problem consists in minimizing $J(u(\cdot))$ over $A$.

**Remark 7**

1. Observe that

$$E_u ∫_0^T l(t, u_t) dA_t = E \left[ E^u_T ∫_0^T l(t, u_t) dA_t \right] ≤ (E(E^u_T)^q)^{1/q} \left[ E \left| ∫_0^T l(t, u_t) dA_t \right|^p \right]^{1/p} \leq C \left[ E ∫_0^T \sup_{u∈U} |l(t, u)|^p e^{βA_t} dA_t \right]^{1/p} < ∞,$$

Similarly

$$E_u g(X_T) = E E^u_T g(X_T) ≤ (E(E^u_T)^q)^{1/q} \left( E|g(X_T)|^p \right)^{1/p} ≤ c(E e^{βA_T} |g(X_T)|^p)^{1/p} < ∞,$$

so that under (6.3) and (6.4) the cost is finite for every admissible control.

2. We point out that the function $r$ can take the value zero, and this implies that the process $E$ is not necessarily strictly positive. Hence, the measures $P_u$ induced by the control are not equivalent to the original probability $P$ but are only absolutely continuous with respect to $P$.

3. Roughly speaking, we can say a controller attempts to modify the probabilities of different paths, in a dynamic way, to try and minimize the expected terminal cost $E[g(X_T)]$. However, using a control incurs a cost $l$, so the controller then needs to balance the benefits from increasing the probability of less costly outcomes (low values of $g(X_T)$) against the cost of controlling more actively.

To solve the optimal control problem, we introduce the BSDE

$$Y_t + ∫_t^T ∫_K Z(s, x) μ(dx, dx) = ξ + ∫_t^T f(s, Z_s(\cdot)) dA_s,$$  (6.6)
with terminal condition $g(X_T)$ being the terminal cost above, and with the generator $f$ being the Hamiltonian function defined below. This is Eq. (4.1), with $f$ only depending on $(\omega, t, \xi)$, and indeed, it comes from an equation of type (4.7) via transformation (4.10).

The Hamiltonian function $f$ is defined on $\Omega \times [0, T] \times B(E)$ as

$$f(\omega, t, \xi)(\cdot) = \left\{ \begin{array}{ll}
\inf_{u \in U} \left( l(\omega, t, u) + \int_K \xi(x) r(\omega, t, x, u) \phi_{\omega, t}(dx) \right) & \text{if } \int_K |\xi(x)| \phi_{\omega, t}(dx) < \infty \\
0 & \text{otherwise}
\end{array} \right.$$

(6.7)

We will assume that the infimum is in fact achieved, possibly at many points. Moreover, we need to verify that the generator of the BSDE satisfies the conditions required in Hypothesis 1, in particular the measurability property which does not follow from its definition. An appropriate assumption is the following one, since we will see below in Proposition 3 that it can be verified under quite general conditions.

**Hypothesis 4** For every predictable function $Z$ on $\Omega \times [0, T] \times E$, there exists a $U$-valued predictable process (i.e., an admissible control) $u_Z^*$ such that, $dA_t(\omega)P$-almost surely,

$$f(\omega, t, Z_{\omega, t}(\cdot)) = l(\omega, t, u_Z^*(\omega, t)) + \int_K Z_{\omega, t}(x) r(\omega, t, x, u_Z^*(\omega, t)) \phi_{\omega, t}(dx)$$

(6.8)

Now, it is easy to check that all the required assumptions of Hypothesis 1 for the solvability of BSDE (6.6) are satisfied. Using the boundedness assumption (6.2), it is easy to check that (4.5) is verified with $L' = 0$ and $L = C_r$. By (6.3), we also have

$$\mathbb{E} \int_0^T e^{\beta A_t} |f(t, 0)|^p dA_t = \mathbb{E} \int_0^T e^{\beta A_t} \left| \inf_{u \in U} l(\omega, t, u) \right|^p dA_t < +\infty$$

(6.9)

so that (4.6) holds. Assuming finally that (6.4) holds, by Theorem 2 the BSDE has a unique solution $(Y, Z) \in L^p_{\beta}$ if $\beta > 1 + \frac{c_\epsilon}{1 + \epsilon} + (p - 1)(2(C_r + 1)^p(1 + \epsilon))^{\frac{1}{p - 1}}$.

The corresponding admissible control $u^*_Z$, whose existence is required in Hypothesis 4, will be denoted as $u^*$.

**Theorem 5** Assume that Hypotheses 3 and 4 are satisfied and that

$$\mathbb{E} \exp \left( \frac{2p - 1}{p - 1} + (p - 1)C_r \left( \frac{p}{p - 1} \right)^2 \right)^{A_T} < \infty.$$  

(6.10)

Suppose also that there exists $\beta$ such that

$$\beta > 1 + \frac{c_\epsilon}{1 + \epsilon} + (p - 1)((C_r + 1)^p(1 + \epsilon))^{\frac{1}{p - 1}},$$
\[ \mathbb{E} \exp (\beta A_T) < \infty, \quad \mathbb{E}[|g(X_T)|^p e^{\beta A_T}] < \infty. \] (6.11)

Let \((Y, Z) \in L^p_\beta\) denote the solution to BSDE (6.6) and \(u^* = u^Z\) the corresponding admissible control. Then \(u^*(\cdot)\) is optimal and \(Y_0 = J(u^*(\cdot)) = \inf_{u(\cdot) \in \mathcal{A}} J(u(\cdot))\) is the optimal cost.

**Proof** Fix \(u(\cdot) \in \mathcal{A}\). Assumption (6.10) allows to apply Lemma 5 with \(\gamma = \frac{p}{p-1}\) and yields \(\mathbb{E}(\mathcal{E}_T^u)^{\frac{p}{p-1}} < \infty\). It follows that \(g(X_T)\) is integrable under \(\mathbb{P}_u\). Indeed, by (6.4)

\[ \mathbb{E}_u|g(X_T)| = \mathbb{E}|\mathcal{E}_T^u g(X_T)| \leq (\mathbb{E}(\mathcal{E}_T^u)^{\frac{p}{p-1}} (\mathbb{E}g(X_T)^p)^{1/p} < \infty. \]

We next show that under \(\mathbb{P}_u\) we have \(\mathbb{E}_u \int_0^T \int_K |Z(t, x)| v^u(dt, dx) < \infty\). First note that, by Hölder’s inequality,

\[
\int_0^T \int_K |Z(t, x)| v^u(dr, dx) = \int_0^T \int_K e^{-\frac{1}{p-1} \beta A_t} e^{\frac{1}{p-1} \beta A_t} |Z(t, x)| v(dr, dx) \\
\leq \left( \int_0^T e^{-\frac{1}{p-1} \beta A_t} dA_t \right)^{\frac{p-1}{p}} \left( \int_0^T \int_K e^{\beta A_t} |Z(t, x)|^p v(dr, dx) \right)^{1/p} \\
\leq \left( \frac{p-1}{\beta} \right)^{\frac{p-1}{p}} \left( \int_0^T \int_K e^{\beta A_t} |Z(t, x)|^p v(dr, dx) \right)^{1/p}.
\]

Therefore, using (6.2),

\[
\mathbb{E}_u \int_0^T \int_K |Z(t, x)| v^u(dr, dx) \\
= \mathbb{E}_u \int_0^T \int_K |Z(t, x)| r(t, x, u_t) v(dt, dx) \\
= \mathbb{E} \left[ \mathcal{E}_T^u \int_0^T \int_K |Z(t, x)| r(t, x, u_t) v(dt, dx) \right] \\
\leq (\mathbb{E}(\mathcal{E}_T^u)^{\frac{p}{p-1}})^{\frac{p-1}{p}} C_r \left( \frac{p-1}{\beta} \right)^{\frac{p-1}{p}} \left( \mathbb{E} \int_0^T \int_K e^{\beta A_t} |Z(t, x)|^p v(dr, dx) \right)^{1/p}
\]

and the right-hand side of the last inequality is finite, since \((Y, Z) \in L^p_\beta\).

By similar arguments, we also check that

\[
\mathbb{E}_u \int_0^T |f(t, Z_t(\cdot))| dA_t = \mathbb{E}\mathcal{E}_T^u \int_0^T |f(t, Z_t(\cdot))| dA_t \\
\leq \mathbb{E}\mathcal{E}_T^u C_r \left( \int_0^T \left( \int_K |Z(t, x)|^p \phi_{\omega, t}(dx) \right)^{1/p} + |f(t, 0)| \right) dA_t < \infty.
\]
Setting \( t = 0 \) and taking the \( \mathbb{P}_u \)-expectation in BSDE (6.6), we therefore obtain
\[
Y_0 + \mathbb{E}_u \left( \int_0^T \int_K Z(t, x) r(t, x, u_t) \nu(dt, dx) \right) = \mathbb{E}_u (g(X_T)) + \mathbb{E}_u \left( \int_0^T f(t, Z_t) \, dA_t \right).
\]

Adding \( \mathbb{E}_u \left( \int_0^T l(t, u_t) \, dA_t \right) \) to both sides, we finally obtain the equality
\[
Y_0 + \mathbb{E}_u \left( \int_0^T \left( l(t, u_t) + \int_K Z(t, x) r(t, x, u_t) \phi_t(dx) \right) \, dA_t \right) = J(u(\cdot)) + \mathbb{E}_u \left( \int_0^T f(t, Z_t) \, dA_t \right)
\]
\[
= J(u(\cdot)) + \mathbb{E}_u \left( \int_0^T \inf_{u \in U} \left( l(t, u) + \int_K Z(t, x) r(t, x, u) \phi_t(dx) \right) \, dA_t \right).
\]

This implies immediately the inequality \( Y_0 \leq J(u(\cdot)) \) for every admissible control, with an equality if \( u(\cdot) = u^*(\cdot) \).

Hypothesis 4 can be verified in specific situations when it is possible to compute explicitly the function \( u^Z \). General conditions for its validity can also be formulated using appropriate measurable selection theorems, as in the following proposition.

**Proposition 3** Let \( U \) be a topological space which is the union of countably many compact metrizable subsets of itself (for example \( \mathbb{R} \) or \( \mathbb{N} \)) with its Borel \( \sigma \)-field \( U \). Suppose that the functions \( r(\omega, t, x, \cdot), l(\omega, t, \cdot) \) are continuous on \( U \) for every \((\omega, t, x)\). If further \( r \) and \( l \) satisfy (6.2) and for \( dP \times dA_t \)-almost all \((\omega, t)\) and for every predictable function \( Z \) on \( \Omega \times [0, T] \times K \), there exists \( v \in U \) such that
\[
l(\omega, t, v) + \int_K Z_{\omega,t}(x) r(\omega, t, x, v) \phi_{\omega,t}(dx) = \inf_{u \in U} l(\omega, t, u) + \int_K Z_{\omega,t}(x) r(\omega, t, x, u) \phi_{\omega,t}(dx) \tag{6.12}
\]

then Hypothesis 4 is satisfied.

**Proof** For every predictable function \( Z \), set
\[
G^Z = \left\{ (\omega, t) : \int_K |Z(\omega, t, x)| \phi_{\omega,t}(dx) = \infty \right\}
\]
and define a map \( F : \Omega \times [0, T] \times Z \times U \to \mathbb{R} \) by
\[
F(\omega, t, Z, u) = \begin{cases} l(\omega, t, u) + \int_K Z(\omega, t, x) r(\omega, t, x, u) \phi_t(\omega, dx) & \text{if } (\omega, t) \notin G^Z, \\ 0 & \text{if } (\omega, t) \in G^Z. \end{cases}
\]
Then $F(\omega, t, Z, \cdot)$ is continuous for every $(\omega, t)$, $F(\omega, t, \cdot, u)$ is Lipschitz and $F(\cdot, \cdot, Z, u)$ is a predictable function on $\Omega \times [0, T]$. By Filippov’s implicit function theorem (see Theorem 21.3.4 in [15], or [26]), there exists a $U$-valued function $u^Z$ on $\Omega \times [0, T]$ such that $F(\omega, t, Z, u^Z(\omega, t)) = \inf_{u \in U} F(\omega, t, Z, u)$ for every $(\omega, t) \in \Omega \times [0, T]$ (so that (6.8) holds true for every $(\omega, t)$) and such that $u^Z$ is measurable with respect to the completion of the predictable $\sigma$-algebra in $\Omega \times [0, T]$ with respect to the measure $dA_t(\omega)\mathbb{P}(d\omega)$. After modification on a null set, the function $u^Z$ can be made predictable, and (6.8) still holds, as it is understood as an equality for $dA_t(\omega)\mathbb{P}(d\omega)$-almost all $(\omega, t)$.

\[\square\]

Remark 8

(i) If $U$ is a compact metrizable set, then the continuity assumption on $l$ and $r$ with respect to $u$ guarantees immediately that the infimum is attained as stated in 6.12 and the conclusion of Proposition 3 holds.

(ii) We note that the function $f$ does not satisfy the requirement for the strict comparison principle (see [44]), since we allow $r$ to take the value zero. In solving the optimal control problem with the backward approach, we did not make direct use of comparison results (see proof of Theorem 5). If we require $r > 0$, we can solve the optimal control problem under the same assumptions of Proposition 3 except 6.12 by using the comparison principle for the backward equations (see [15], Section 21.3).

Acknowledgements

We wish to thank Prof. Jean Jacod for discussions on connections between BSDEs and point processes.

References

1. Almgren R, Chriss N (2001) Optimal execution of portfolio transactions. J Risk 3:5–39
2. Ankirchner S, Jeanblanc M, Kruse T (2014) BSDEs with singular terminal condition and a control problem with constraints. SIAM J Control Optim 52(2):893913
3. Aubin J-P, Frankowska H (1990) Set-valued analysis. Systems & control: foundations & applications, vol 2. Birkhäuser
4. Bahlali K, Eddahbi M, Essaky EH (2003) BSDE associated with Lévy processes and application to PDIE. Int J Stoch Anal 16(1):117
5. Bandini E (2015) Existence and uniqueness for backward stochastic differential equations driven by a random measure, possibly non quasi-left continuous. Electron Commun Probab 20
6. Bandini E, Confortola F Optimal control of semi-Markov processes with a backward stochastic differential equations approach. Preprint. arXiv:1311.1063
7. Barles G, Buckdahn R, Pardoux E (1997) Backward stochastic differential equations and integral-partial differential equations. Stoch Stoch Rep 60:57–83
8. Becherer D (2006) Bounded solutions to backward SDE’s with jumps for utility optimization and indifference hedging. Ann Appl Probab 16:2027–2054
9. Bismut J-M (1973) Conjugate convex functions in optimal stochastic control. J Math Anal Appl 44(2):384404
10. Briand P, Carmona R (2000) BSDEs with polynomial growth generators. J Appl Math Stoch Anal 13:207238
11. Briand P, Delyon B, Hu Y, Pardoux E, Stoica L (2003) Lp solutions of backward stochastic differential equations. Stoch Process Appl 108:109129
12. Carbone R, Ferrario B, Santacroce M (2008) Backward stochastic differential equations driven by càdlàg martingales. Theory Probab Appl 52:304–314
13. Cohen SN, Elliott RJ (2008) Solutions of backward stochastic differential equations on Markov chains. Commun Stoch Anal 2:251–262
14. Cohen SN, Elliott RJ (2010) Comparisons for backward stochastic differential equations on Markov chains and related no-arbitrage conditions. Ann Appl Probab 20:267–311
15. Cohen SN, Elliott RJ (2015) Stochastic calculus and applications, 2nd ed. Probability and its Applications. Springer, Cham, 2015. xxiii+666 pp
16. Cohen SN, Szpruch L (2012) On Markovian solution to Markov Chain BSDEs. Numer Algebra Control Optim 2:257–269
17. Confortola F, Fuhrman M (2013) Backward stochastic differential equations and optimal control of marked point processes. SIAM J Control Optim 51(5):3592–3623
18. Confortola F, Fuhrman M (2014) Backward stochastic differential equations associated to jump Markov processes and their applications. Stoch Process Their Appl 124:289–316
19. Confortola F, Fuhrman M, Jacod J Backward stochastic differential equation driven by a marked point process: an elementary approach with an application to optimal control. Ann Appl Probab, to appear
20. Crépey S, Matoussi A (2008) Reflected and doubly reflected BSDEs with jumps: a priori estimates and comparison. Ann Appl Probab 18:2041–2069
21. Davis MHA (1976) The representation of martingales of jump processes. SIAM J Control Optim 14(4):623–638
22. Davis MHA (1993) Markov models and optimization. Monographs on Statistics and Applied Probability, 49. Chapman & Hall
23. Eddahbi M, Fakhouri I, Ouknine Y (2017) \( L^p \) \((p \geq 2)\)-solutions of generalized BSDEs with jumps and monotone generator in a general filtration. Mod Stoch Theory Appl 4(1):2563
24. El Karoui N, Matoussi A, Ngoupeyou A (2016) Quadratic exponential semimartingales and application to BSDEs with jumps. arXiv preprint. arXiv:1603.06191
25. El Karoui N, Peng S, Quenez M-C (1997) Backward stochastic differential equations in finance. Math Finance 7:1–71
26. Filippov AF (1962) On certain questions in the theory of optimal control. Vestnik Moskov Univ Ser Mat Meh Astronom 2:2542 (1959). English trans. J Soc Indust Appl Math Ser A Control 1:7684
27. Forsyth P, Kennedy J, Tse S, Windcliff H (2012) Optimal trade execution: a mean quadratic variation approach. J Econ Dyn Control 36:19711991
28. Gatheral J, Schied A (2011) Optimal trade execution under geometric Brownian motion in the Almgren and Chriss framework. Int J Theor Appl Financ 14:353–368
29. Graewe P, Horst U (2017) Optimal trade execution with instantaneous price impact and stochastic resilience. SIAM J Control Optim 55(6):3707–3725
30. Graewe P, Horst U, Qiu J (2015) A non-Markovian liquidation problem and backward SPDEs with singular terminal conditions. SIAM J Control Optim 53(2):690–711
31. He S, Wang J, Yan Y (1992) Semimartingale theory and stochastic calculus. Science Press, Beijing
32. Li J, Wei Q (2014) Lp estimates for fully coupled FBSDEs with jumps. Stoch Process Appl 124(4):15821611
33. Jacob J (1974) Multivariate point processes; predictable projection, Radon–Nikodym derivatives, representation of martingales. Zeit für Wahr 31:235–253
34. (1979) Calcul stochastique et problèmes de martingales. Lecture Notes in Mathematics 714, Springer, Berlin (1979)
35. Jacob J, Mémin J (1981) Weak and strong solutions of stochastic differential equations: existence and stability. In: Williams R (ed) Stochastic integrals. Springer Verlag, Lectures Notes in Math. vol 851, pp 169-212
36. Kazi-Tani MN, Possamaï D, Zhou C (2015) Quadratic BSDEs with jumps: a fixed-point approach. Electron J Probab 20(66):128
37. Kazi-Tani N PD, Zhou C (2015) Quadratic BSDEs with jumps: related nonlinear expectations. Stoch Dyn 1500012
38. Kharroubi I, Lim T (2012) Progressive enlargement of filtrations and backward SDEs with jumps. Preprint
39. Kratz P, Schneborn T (2015) Portfolio liquidation in dark pools in continuous time. Math Finance 25(3):496544
40. Kruse T, Popier A (2016) BSDEs with monotone generator driven by Brownian and Poisson noises in a general filtration. Stochastics 88(4):491539
41. Kruse T, Popier A (2016) Minimal supersolutions for BSDEs with singular terminal condition and application to optimal position targeting. Stoch Process Appl 126(9):25542592
42. Nualart D, Schoutens W (2001) Backward stochastic differential equations and Feynman–Kac formula for Lévy processes, with applications in finance. Bernoulli 7(5):761–776
43. Pardoux E, Peng S (1990) Adapted solution of a backward stochastic differential equation. Syst Control Lett 14:55–61
44. Royer M (2006) Backward stochastic differential equations with jumps and related non-linear expectations. Stoch Proc Appl 116:13581376
45. Shen L, Elliott RJ (2011) Backward stochastic differential equations for a single jump process. Stoch Anal Appl 29:654–673
46. Situ R (1997) On solutions of backward stochastic differential equations with jumps and applications. Stoch Process Their Appl 66(2):209236
47. Tang SJ, Li XJ (1994) Necessary conditions for optimal control of stochastic systems with random jumps. SIAM J Control Optim 32:1447–1475
48. Xia J (2000) Backward stochastic differential equation with random measures. Acta Math Appl Sin (English Ser.) 16 (3)225–234
49. Yao S (2017) Lp solutions of backward stochastic differential equations with jumps. Stoch Process Appl 127(11):34653511

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