Représentation probabiliste d'équations HJB pour le contrôle optimal de processus à sauts, EDSR (équations différentielles stochastiques rétrogrades) et calcul stochastique.

Elena Bandini

Politecnico di Milano
Graduate School in Mathematical Models and Methods in Engineering

Host Department : Department of Mathematics "F. Brioschi"

Université Paris-Saclay
École doctorale de mathématiques Hadamard (EDMH, ED 574)

Établissement d'inscription : ENSTA ParisTech
Laboratoire d'accueil : Unité de Mathématiques Appliquées

Ph.D THESIS IN MATHEMATICAL MODELS AND METHODS FOR ENGINEERING

and

THÈSE DE DOCTORAT EN MATHÉMATIQUES APPLIQUÉES

Elena BANDINI

Probabilistic Representation of HJB Equations for Optimal Control of Jump Processes, BSDEs and Related Stochastic Calculus

Ph.D Thesis defended on April 7, 2016 at Politecnico di Milano (Cycle XXVIII)

Reviewers :	Giulia DI NUNNO	(University of Oslo)
	Saïd HAMADÈNE	(Université du Mans)
	Agnès SULEM	(INRIA Rocquencourt)

Committee :

Fausto GOZZI	(LUISS Roma) President of the Jury
Marco FUHRMAN	(Politecnico di Milano) Thesis Co-Director
Francesco RUSSO	(ENSTA ParisTech) Thesis Co-Director
Gianmario TESSITORE	(Università degli Studi di Milano-Bicocca) Examinator

The chairs of the Doctoral Programmes :	Roberto LUCCHETTI	(Politecnico di Milano)
	Frédéric PAULIN	(École Polytechnique)

NNT : 2016SACLY005

Alla mia famiglia

Abstract

In the present document we treat three different topics related to stochastic optimal control and stochastic calculus, pivoting on the notion of backward stochastic differential equation (BSDE) driven by a random measure.

The three first chapters of the thesis deal with optimal control for different classes of non-diffusive Markov processes, in finite or infinite horizon. In each case, the value function, which is the unique solution to an integro-differential Hamilton-Jacobi-Bellman (HJB) equation, is probabilistically represented as the unique solution of a suitable BSDE. In the first chapter we control a class of semi-Markov processes on finite horizon; the second chapter is devoted to the optimal control of pure jump Markov processes, while in the third chapter we consider the case of controlled piecewise deterministic Markov processes (PDMPs) on infinite horizon. In the second and third chapters the HJB equations associated to the optimal control problems are fully nonlinear. Those situations arise when the laws of the controlled processes are not absolutely continuous with respect to the law of a given, uncontrolled, process. Since the corresponding HJB equations are fully nonlinear, they cannot be represented by classical BSDEs. In these cases we have obtained nonlinear Feynman-Kac representation formulae by generalizing the control randomization method introduced in Kharroubi and Pham (2015) for classical diffusions. This approach allows us to relate the value function with a BSDE driven by a random measure, whose solution has a sign constraint on one of its components. Moreover, the value function of the original non-dominated control problem turns out to coincide with the value function of an auxiliary dominated control problem, expressed in terms of equivalent changes of probability measures.

In the fourth chapter we study a backward stochastic differential equation on finite horizon driven by an integer-valued random measure μ on $\mathbb{R}_+ \times E$, where E is a Lusin space, with compensator $\nu(dt\, dx) = dA_t\, \phi_t(dx)$. The generator of this equation satisfies a uniform Lipschitz condition with respect to the unknown processes. In the literature, well-posedness results for BSDEs in this general setting have only been established when A is continuous or deterministic. We provide an existence and uniqueness theorem for the general case, i.e. when A is a right-continuous

nondecreasing predictable process. Those results are relevant, for example, in the framework of control problems related to PDMPs. Indeed, when μ is the jump measure of a PDMP on a bounded domain, then A is predictable and discontinuous.

Finally, in the two last chapters of the thesis we deal with stochastic calculus for general discontinuous processes. In the fifth chapter we systematically develop stochastic calculus via regularization in the case of jump processes, and we carry on the investigations of the so-called weak Dirichlet processes in the discontinuous case. Such a process X is the sum of a local martingale and an adapted process A such that $[N, A] = 0$, for any continuous local martingale N. Given a function $u : [0, T] \times \mathbb{R} \to \mathbb{R}$, which is of class $C^{0,1}$ (or sometimes less), we provide a chain rule type expansion for $u(t, X_t)$, which constitutes a generalization of Itô's lemma being valid when u is of class $C^{1,2}$. This calculus is applied in the sixth chapter to the theory of BSDEs driven by random measures. In several situations, when the underlying forward process X is a special semimartingale, or, even more generally, a special weak Dirichlet process, we identify the solutions (Y, Z, U) of the considered BSDEs via the process X and the solution u to an associated integro-partial differential equation.

Key words: Backward stochastic differential equation (BSDE), stochastic optimal control, Hamilton-Jacobi-Bellman equation, nonlinear Feynman-Kac formula, constrained BSDE, random measures and compensators, pure jump processes, piecewise deterministic Markov processes, semi-Markov processes, stochastic calculus via regularization, weak Dirichlet processes.

Résumé

Dans le présent document on aborde trois divers thèmes liés au contrôle et au calcul stochastiques, qui s'appuient sur la notion d'équation différentielle stochastique rétrograde (EDSR) dirigée par une mesure aléatoire.

Les trois premiers chapitres de la thèse traitent des problèmes de contrôle optimal pour différentes catégories de processus markoviens non-diffusifs, à horizon fini ou infini. Dans chaque cas, la fonction valeur, qui est l'unique solution d'une équation intégro-différentielle de Hamilton-Jacobi-Bellman (HJB), est représentée comme l'unique solution d'une EDSR appropriée. Dans le premier chapitre, nous contrôlons une classe de processus semi-markoviens à horizon fini; le deuxième chapitre est consacré au contrôle optimal de processus markoviens de saut pur, tandis qu'au troisième chapitre, nous examinons le cas de processus markoviens déterministes par morceaux (PDMPs) à horizon infini. Dans les deuxième et troisième chapitres les équations d'HJB associées au contrôle optimal sont complètement non-linéaires. Cette situation survient lorsque les lois des processus contrôlés ne sont pas absolument continues par rapport à la loi d'un processus donné. Étant les équations d'HJB correspondantes complètement non-linéaires, ces équations ne peuvent pas être représentées par des EDSRs classiques. Dans ces cadre, nous avons obtenu des formules de Feynman-Kac non linéaires en généralisant la méthode de la randomisation du contrôle introduite par Kharroubi et Pham (2015) pour les diffusions classiques. Ces techniques nous permettent de relier la fonction valeur du problème de contrôle à une EDSR dirigée par une mesure aléatoire, dont une composante de la solution subit une contrainte de signe. En plus, on démontre que la fonction valeur du problème de contrôle originel non dominé coïncide avec la fonction valeur d'un problème de contrôle dominé auxiliaire, exprimé en termes de changements mesures équivalentes de probabilité.

Dans le quatrième chapitre, nous étudions une équation différentielle stochastique rétrograde à horizon fini, dirigée par une mesure aléatoire à valeurs entières μ sur $\mathbb{R}_+ \times E$, où E est un espace lusinien, avec compensateur de la forme $\nu(dt\, dx) = dA_t\, \phi_t(dx)$. Le générateur de cette équation satisfait une condition de Lipschitz uniforme par rapport aux inconnues. Dans la littérature, l'existence et unicité pour des

EDSRs dans ce cadre ont été établis seulement lorsque A est continu ou déterministe. Nous fournissons un théorème d'existence et d'unicité même lorsque A est un processus prévisible, non décroissant, continu à droite. Ce résultat s'applique, par exemple, au cas du contrôle lié aux PDMPs. En effet, quand μ est la mesure de saut d'un PDMP sur un domaine borné, A est prévisible et discontinu.

Enfin, dans les deux derniers chapitres de la thèse nous traitons le calcul stochastique pour des processus discontinus généraux. Dans le cinquième chapitre, nous développons le calcul stochastique via régularisations des processus à sauts qui ne sont pas nécessairement des semimartingales. En particulier nous poursuivons l'étude des processus dénommés de Dirichlet faibles, dans le cadre discontinu. Un tel processus X est la somme d'une martingale locale et d'un processus adapté A tel que $[N, A] = 0$, pour toute martingale locale continue N. Pour une fonction $u : [0, T] \times \mathbb{R} \to \mathbb{R}$ de classe $C^{0,1}$ (ou parfois moins), on exprime un développement de $u(t, X_t)$, dans l'esprit d'une généralisation du lemme d'Itô, lequel vaut lorsque u est de classe $C^{1,2}$. Le calcul est appliqué dans le sixième chapitre à la théorie des EDSRs dirigées par des mesures aléatoires. Dans de nombreuses situations, lorsque le processus sous-jacent X est une semimartingale spéciale, ou plus généralement, un processus de Dirichlet spécial faible, nous identifions les solutions des EDSRs considérées via le processus X et la solution u d'une équation aux dérivées partielles intégro-différentielle associée.

Mots clés: Équations différentielles stochastiques rétrogrades (EDSR), contrôle optimal stochastique, équations d'Hamilton-Jacobi-Bellman, formule de Feynman-Kac non linéaire, EDSR avec contraintes, mesures aléatoires et compensateurs, processus de saut pur, processus markoviens déterministes par morceaux, processus semi-markoviens, calcul stochastique via régularization, processus de Dirichlet faibles.

Acknowledgments

I would like to take this opportunity to express my sincere gratitude to my advisors Prof. Marco Fuhrman and Prof. Francesco Russo, for devoting much of their time to the development of the present Ph.D. thesis, for the many suggestions, as well as for their continuous support and attention. They gave me the possibility to appreciate different areas of research in stochastic analysis, always leading me towards those subjects which turned out to be the best fitted for my research interests. I also wish to thank Dott. Fulvia Confortola for her help and for her precious advices.

I would thank Prof. Huyên Pham for giving me the possibility to work on a cutting-edge topic of stochastic analysis, which results in the article [**6**] and in the work in preparation [**5**], whose formulations unfortunately were premature to be part of this doctoral dissertation. I wish also to thank Prof. Jean Jacod for his kindness and willingness; it has been an incredible honour for me to have the possibility to discuss stochastic analysis with him.

I am very grateful to my three referees Prof. Giulia Di Nunno, Prof. Saïd Hamadène and Prof. Agnès Sulem. I thank them for agreeing to make the reports on the thesis and for their interest in my work. I would also thank Prof. Fausto Gozzi and Prof. Gianmario Tessitore for agreeing to partecipate to the jury of my thesis.

Finally, I would like to thank all the people who made the completion of the present Ph.D. thesis possible.

Contents

Introduction

In the present introductory chapter we provide a general overview of the subsequent chapters of the doctoral dissertation. All the main results of the thesis are here recalled; for the sake of brevity, we will do not set out the technical assumptions in detail, instead we refer to later chapters for the precise statements. We also give only general references, while a detailed analysis on the technical aspects will be developed in the body of the document.

Brief overview and general references on optimal control problems, BSDEs and discontinuous stochastic processes

In this Ph.D. thesis we deal with stochastic processes and the associated optimal control problems. We consider stochastic dynamical systems, where a random noise affects the system evolution. Introducing a functional cost which depends on the state and on the control variable, we are interested in minimizing its expected value over all possible realizations of the noise process. There exists a large literature on stochastic control problems of this type; we mention among others the monographs by Krylov [**89**], Bensoussan [**13**], Yong and Zhou [**132**], Fleming and Soner [**65**], Pham [**107**]. In the present work we focus on optimal control problems of stochastic processes with jumps. An important class of those processes is determined starting from the so-called marked point processes. Marked point processes are related to the martingale theory by means of the concept of compensator, which describes the local dynamics of a marked point process. Martingale methods in the theory of point processes go back to Watanabe [**130**], who discovered the martingale characterization of Poisson processes, but the first systematic treatment of a general marked point process using martingales was given by Brémaud [**18**]. The martingale definition of compensator gives the basis to construct a martingale calculus which has the same power as Itô calculus for diffusions, see Jacod's book [**77**].

In the past few years, many different methods have been developed to solve optimal control problems of the type mentioned above. In our work we consider the approach based on the theory of backward stochastic differential equations, BSDEs for short. BSDEs are stochastic differential equations with a final condition rather than an initial condition. This subject started with the paper [**98**] by Pardoux and Peng, where the authors first solved general nonlinear BSDEs driven by the Wiener process. Afterwards, a systematic theory has been developed for diffusive BSDEs, see for instance El Karoui and Mazliak [**52**], El Karoui, Peng and Quenez [**53**], Pardoux [**96**], [**97**]. Many generalizations have also been considered where the Brownian motion was replaced by more general processes. Backward equations driven by a Brownian motion and a Poisson random measure have been studied for instance in Tang and Li [**128**], Barles, Buckdahn and Pardoux [**10**], Royer [**113**], Kharroubi, Ma, Pham and Zhang [**87**], Øksendal, Sulem and Zhang [**94**], in view of various applications including stochastic maximum principle, partial differential equations of nonlocal type, quasi-variational inequalities and impulse control. There are instead few results on BSDEs driven by more general random measures, among which we recall for instance Xia [**131**], Jeanblanc, Mania, Santacroce and Schweizer [**80**], Confortola, Fuhrman and Jacod [**29**]. In most cases, the authors deal with BSDEs with jumps with a random compensator which is absolutely continuous with respect to a deterministic measure, that can be reduced to a Poisson measure by a Girsanov change of probability, see for instance Becherer [**12**], Crépey and Matoussi [**33**], Kazi-Tani, Possamai and Zhou [**83**], [**84**].

I. Feynman-Kac formula for nonlinear HJB equations

I.1. State of the art. We fix our attention on BSDEs whose random dependence is guided by a forward Markov process, typically a solution of a stochastic differential equation. Those equations are commonly called forward BSDEs; since Peng [**101**] and Pardoux and Peng [**99**], it is well-known that forward BSDEs provide a probabilistic representation (nonlinear Feynman-Kac formula) for a class of semilinear parabolic partial differential equations. Let $T < \infty$ be a finite time horizon and consider the filtered space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, where $\mathbb{F}$ is the canonical $\mathbb{P}$-completed filtration associated with a d-dimensional Brownian motion $W = (W_t)_{t\in[0,T]}$. We suppose $\mathcal{F} = \mathcal{F}_T$. Let $t \in [0,T]$ and $x \in \mathbb{R}^n$; a forward-backward stochastic differential equation on $[t,T]$ is a problem of the following type:

$$\begin{cases} X_s = x + \int_t^s b(r,X_r)dr + \int_t^s \eta(r,X_r)dW_r \\ Y_s = g(X_T) + \int_s^T l(r,X_r,Y_r,Z_r)dr - \int_s^T Z_r dW_r, \end{cases} \tag{1}$$

where $b\colon [0,T]\times\mathbb{R}^n \to \mathbb{R}^n$, $\eta\colon [0,T]\times\mathbb{R}^n \to \mathbb{R}^{n\times d}$, $l\colon [0,T]\times\mathbb{R}^n\times\mathbb{R}\times\mathbb{R}^d \to \mathbb{R}$, and $g\colon \mathbb{R}^n \to \mathbb{R}$ are Borel measurable functions. Then, it is well-known that, under suitable assumptions on the coefficients, the above forward-backward equation admits a unique solution $\{(X_s^{t,x}, Y_s^{t,x}, Z_s^{t,x}),\ t \le s \le T\}$ for any $(t,x) \in [0,T]\times\mathbb{R}^n$. Moreover, $Y_t^{t,x}$ is deterministic, therefore we can define the function

$$v(t,x) := Y_t^{t,x}, \qquad \text{for all } (t,x) \in [0,T]\times\mathbb{R}^n,$$

which turns out to be a viscosity solution to the following partial differential equation:

$$\begin{cases} \frac{\partial v}{\partial t}(t,x) + \mathcal{L}v(t,x) + l\big(t,x,v(t,x),\eta^T(t,x)D_x v(t,x)\big) = 0, & (t,x) \in [0,T) \times \mathbb{R}^n, \\ v(T,x) = g(x), & x \in \mathbb{R}^n, \end{cases}$$

where the operator $\mathcal{L}$ is given by

$$\mathcal{L}v = \langle b, D_x v\rangle + \frac{1}{2}\mathrm{tr}\big(\eta\,\eta^T D_x^2 v\big). \tag{2}$$

Let us now consider the following fully nonlinear PDE of Hamilton-Jacobi-Bellman (HJB) type

$$\frac{\partial v}{\partial t} + \sup_{a \in A}\bigg(\langle h(x,a), D_x v\rangle + \frac{1}{2}\mathrm{tr}\big(\sigma\sigma^T(x,a)D_x^2 v\big) + f(x,a)\bigg) \;=\; 0, \tag{3}$$

on $[0,T) \times \mathbb{R}^d$, together with the terminal condition

$$v(T,x) \;=\; g(x), \qquad x \in \mathbb{R}^d,$$

where A is a subset of $\mathbb{R}^q$, and $h\colon \mathbb{R}^n \times A \to \mathbb{R}^n$, $\sigma\colon \mathbb{R}^n \times A \to \mathbb{R}^{n\times d}$, $f\colon \mathbb{R}^n \times A \to \mathbb{R}$ are Borel measurable functions. As it is well-known, see for example Pham [**107**], the above equation is the dynamic programming equation of a stochastic control problem whose value function is given by

$$v(t,x) \;:=\; \sup_{\alpha}\mathbb{E}\bigg[\int_t^T f(X_s^{t,x,\alpha},\alpha_s) + g(X_T^{t,x,\alpha})\bigg], \tag{4}$$

where $X^{t,x,\alpha}$ is the controlled state process starting at time $t \in [0,T]$ from $x \in \mathbb{R}^d$, which evolves on $[t,T]$ according to the stochastic equation

$$X_s^{t,x,\alpha} \;=\; x + \int_t^s h(X_r^{t,x,\alpha},\alpha_r)dr + \int_t^s \sigma(X_r^{t,x,\alpha},\alpha_r)\,dW_r, \tag{5}$$

where α is a predictable control process valued in A. Notice that, if $\sigma(x)$ does not depend on $a \in A$ and $\sigma\sigma^T(x)$ is of full rank, then the above HJB equation can be written as

$$\frac{\partial v}{\partial t} + \frac{1}{2}\mathrm{tr}\big(\sigma\sigma^T(x)D_x^2 v\big) + F(x,\sigma^T(x)D_x v) \;=\; 0, \tag{6}$$

where $F(x,z) = \sup_{a\in A}[f(x,a) + \langle\theta(x,a),z\rangle]$ is the θ-Fenchel-Legendre transform of f and $\theta(x,a) = \sigma^T(x)(\sigma\sigma^T(x))^{-1}h(x,a)$ is a solution to $\sigma(x)\theta(x,a) = h(x,a)$. Then, since F depends on $\sigma^T D_x v$, from [**99**] we know that the semilinear PDE (6) admits a nonlinear Feynman-Kac formula through a Markovian forward-backward stochastic differential equation.

Starting from Peng [**103**], the BSDEs approach to the optimal control problem has been deeply investigated in the diffusive case; we mention for instance [**107**], Ma and Yong [**93**], [**132**], and [**53**]. However, all those results require that only the drift coefficient of the stochastic equation depends on the control parameter and that $\sigma\sigma^T(x)$ is of full rank, so that the HJB equation is a second-order semilinear partial differential equation and the nonlinear Feyman-Kac formula is obtained as we explained above. The general case with possibly degenerate controlled diffusion coefficient $\sigma(x,a)$, associated to a fully nonlinear HJB equation, has only recently been completely solved by Kharroubi and Pham [**88**]. We also mention that a first step in this direction was made by Soner, Touzi, and Zhang [**124**], where however the theory of second-order BSDEs (2BSDEs) was used rather than the standard theory of backward stochastic differential equations. 2BSDEs are backward stochastic differential equations formulated under a non-dominated family of singular probability measures, so that their theory relies on tools from quasi-sure analysis. On the other hand, according to [**88**], it is enough to consider a backward stochastic differential equation with jumps, where the jumps are constrained to be nonpositive, formulated under a single probability measure, as in the standard theory of BSDEs.

Let us describe informally the approach presented in [**88**], which we will call *control randomization method*; for greater generality and precise statements we refer to the original paper of Kharroubi and Pham. In [**88**] the forward-backward system associated to the HJB equation (3) is constructed as follows: the forward equation, starting at time $t \in [0,T]$ from $(x,a) \in \mathbb{R}^d \times A$, evolves on $[t,T]$ according to the system of equations

$$\begin{aligned} X_s^{t,x,a} &= x + \int_t^s h(X_r^{t,x,a}, I_r^{t,a})\, dr + \int_t^s \sigma(X_r^{t,x,a}, I_r^{t,a})\, dW_r, \\ I_s^{t,a} &= a + \int_t^s \int_A (b - I_{r^-}^{t,a})\, \mu(dr\, db). \end{aligned}$$

Its form is deduced from the controlled state dynamics (5) randomizing the state process $X^{t,x,\alpha}$, i.e., introducing, in place of the control α, a pure-jump (uncontrolled) process I, driven by a Poisson random measure μ on $\mathbb{R}_+ \times A$ independent of W, with intensity measure $\lambda(db)dt$, where λ is a finite measure on $(A, \mathcal{B}(A))$, with full topological support. W and μ are defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, where $\mathbb{F}$ is the completion of the natural filtration generated by W and μ themselves. Regarding the backward equation, as expected, it is driven by the Brownian motion W and the Poisson random measure μ, namely it is a BSDE with jumps with terminal condition $g(X_T^{t,x,a})$ and generator $f(X_\cdot^{t,x,a}, I_\cdot^{t,a})$, as it is natural from the expression of the HJB equation. The backward equation is also characterized by a constraint on the jump component, which turns out to be a crucial aspect of the theory introduced in [**88**], and requires the presence of an increasing process K in the BSDE. This latter process is reminiscent of the one arising in the reflected BSDE theory, see El Karoui et al. [**51**], where however K has to fulfill the Skorohod condition, namely is only active to prevent Y from passing below the obstacle. In conclusion, the backward stochastic differential equation has the following form:

$$\begin{aligned} Y_s^{t,x,a} \;&=\; g(X_T^{t,x,a}) + \int_s^T f(X_r^{t,x,a}, I_r^{t,a})\, dr + K_T^{t,x,a} - K_s^{t,x,a} \\ &\quad - \int_s^T Z_r^{t,x,a}\, dW_r - \int_s^T \int_A L_r^{t,x,a}(b)\, \mu(dr\, db), \qquad t \leq s \leq T, \ a.s. \end{aligned} \tag{7}$$

together with the jump constraint

$$L_s^{t,x,a}(b) \;\leq\; 0, \qquad d\mathbb{P} \otimes ds \otimes \lambda(db) \ a.e. \tag{8}$$

Notice that the presence of the increasing process K in the backward equation does not guarantee the uniqueness of the solution. For this reason, as in the theory of reflected BSDEs, in [**88**] the authors look only for the minimal solution (Y, Z, L, K) to the above BSDE, in the sense that for any other solution $(\bar{Y}, \bar{Z}, \bar{L}, \bar{K})$ we must have $Y \leq \bar{Y}$. The existence of the minimal solution is based on a penalization approach and on the monotonic limit theorem of Peng [**104**].

The nonlinear Feynman-Kac formula becomes

$$v(t, x, a) \ := \ Y_t^{t,x,a}, \qquad (t, x, a) \in [0, T] \times \mathbb{R}^d \times A.$$

Observe that the value function v should not depend on a, but only on (t, x). The function v turns out to be independent of the variable a, as a consequence of the A-nonpositive jump constraint. Indeed, the constraint (8) implies that

$$\mathbb{E}\bigg[\int_t^{t+h} \int_A [v(s, X_s^{t,x,a}, b) - v(s, X_s^{t,x,a}, I_{s^-}^{t,a})]^+ \lambda(db)\, ds\bigg] = 0$$

for any $h > 0$. If v is continuous, by sending h to zero in the above equality divided by h (and by dominated convergence theorem), we can obtain from the mean-value theorem that

$$\int_A [v(s, x, b) - v(s, x, a)]^+ \lambda(db) = 0,$$

from which we see that v does not depend on a. However, it is not clear a priori that the function v is continuous, therefore, in [**88**], the rigorous proof relies on fine viscosity solutions arguments and on mild conditions on λ and A, as the assumptions that the interior set of A is connected and that A is the closure of its interior. In the end, in [**88**] it is proved that the function v does not depend on the variable a in the interior of A and that the viscosity solution to equation (3) admits the probabilistic representation formula

$$v(t, x) \ := \ Y_t^{t,x,a}, \qquad (t, x) \in [0, T] \times \mathbb{R}^d$$

for any a in the interior of A.

In [**88**] another probabilistic representation is also provided, called dual representation, for the solution v to (3). More precisely, let $\mathcal{V}$ be the set of predictable processes $\nu \colon \Omega \times [0, T] \times A \to (0, \infty)$ which are essentially bounded, and consider the probability measure $\mathbb{P}^\nu$ equivalent to $\mathbb{P}$ on $(\Omega, \mathcal{F}_T)$ with Radon-Nikodym density:

$$\frac{d\mathbb{P}^\nu}{d\mathbb{P}}\bigg|_{\mathcal{F}_t} = L_t^\nu \ := \ \mathcal{E}_t\bigg(\int_0^\cdot \int_A (\nu_s(b) - 1)\tilde{\mu}(ds\, db)\bigg),$$

where $\mathcal{E}_t(\cdot)$ is the Doléans-Dade exponential, and $\tilde{\mu}(ds\, db)$ is the compensated random measure $\mu(ds\, db) - \lambda(db)\, ds$. Notice that W remains a Brownian motion under $\mathbb{P}^\nu$, and the effect of the probability measure $\mathbb{P}^\nu$, by Girsanov's Theorem, is to change the compensator $\lambda(db)\, ds$ of μ under $\mathbb{P}$ to $\nu_s(b)\, \lambda(db)\, ds$ under $\mathbb{P}^\nu$. The dual representation reads:

$$v(t, x) \ = \ Y_t^{t,x,a} \ = \ \sup_{\nu \in \mathcal{V}} \mathbb{E}^\nu\bigg[g(X_T^{t,x,a}) + \int_t^T f(X_s^{t,x,a}, I_s^{t,a}) ds\bigg], \tag{9}$$

where $\mathbb{E}^\nu$ denotes the expectation with respect to $\mathbb{P}^\nu$.

The control randomization method has been applied to many cases in the framework of optimal switching and impulse control problems, see Elie and Kharroubi

[**54**], [**55**], [**56**], Kharroubi, Ma, Pham and Zhang [**87**], and developed with extensions and applications, see Cosso and Chokroun [**25**], Cosso, Fuhrman and Pham [**31**], and Fuhrman and Pham [**67**]. In all the above mentioned cases the controlled processes are diffusions constructed as solutions to stochastic differential equations of Itô type.

Differently to the diffusive framework, the BSDE approach to optimal control of non-diffusive processes is not very traditional. Indeed, there exists a large literature on optimal control of marked point processes (see Brémaud [**18**], Elliott [**57**] as general references), but there are relatively few results on their connections with BSDEs. This gap has been partially filled by Confortola and Fuhrman [**28**] in the case of optimal control for pure jump processes, where a probabilistic representation for the value function is provided by means of a BSDE driven by a suitable random measure. In [**28**] conditions are imposed to guarantee that the set of controlled probability laws is absolutely continuous with respect to the law of a given, uncontrolled, process. This gives a natural extension to the non-diffusive framework of the well-known diffusive case where only the drift coefficient of the stochastic equation depends on the control parameter.

In **Chapter 1** we extend the approach of [**28**] to the optimal control problem of semi-Markov processes. For a semi-Markov process X, the Markovian structure can be recovered by considering the pair of processes (X, θ), where θ_s denotes the duration period in the state X_s up to moment s. However, the pair (X, θ) is not pure jump. This prevents to apply in this context the results of [**28**], and requires an ad hoc treatment.

We are also interested in the more general case when the laws of the controlled processes form a non-dominated model, and consequently the HJB equation is fully nonlinear. Indeed, non-diffusive control problems of this type are very frequent in applications, even when the state space is finite. In **Chapter 2** we provide a Feynman-Kac representation formula for the value function of an optimal control problem for pure jump Markov processes, in a general non-dominated framework. **Chapter 3** is then devoted to generalize previous results to the case of a control problem for piecewise deterministic Markov processes. This latter class of processes includes in particular the family of semi-Markov processes. The results in Chapters 2 and 3 are achieved adapting the control randomization method developed in [**88**] for classical diffusions.

In the next paragraphs we describe the contents of Chapters 1, 2, 3.

I.2. Optimal control of semi-Markov processes. In **Chapter 1** we study optimal control problems for a class of semi-Markov processes, and we provide a Feynman-Kac representation formula for the value function by means of a suitable class of BSDEs.

A semi-Markov process on a general state space E can be seen as a two dimensional, time-homogeneous, process $(X_s, \theta_s)_{s\geq 0}$, strongly Markovian with respect to its natural filtration $\mathbb{F}$. The pair $(X_s, \theta_s)_{s\geq 0}$ is associated to a family of probability measures $\mathbb{P}^{x,\vartheta}$ for $x \in E$, $\vartheta \in [0,\infty)$, such that $\mathbb{P}^{x,\vartheta}(X_0 = x, \theta_0 = \vartheta) = 1$. The process (X,θ) is constructed starting from a jump rate function $\lambda(x,\vartheta)$ and a jump measure $A \mapsto Q(x,\vartheta,A)$ on E, depending on $x \in E$ and $\vartheta \geq 0$. If the process starts from (x,ϑ) at time $t=0$, then the distribution of its first jump time T_1 under $\mathbb{P}^{x,\vartheta}$ is

$$\mathbb{P}^{x,\vartheta}(T_1 > s) = \exp\left(-\int_{\vartheta}^{\vartheta+s} \lambda(x,r)\,dr\right), \tag{10}$$

and the conditional probability that X is in A immediately after a jump at time $T_1 = s$ is

$$\mathbb{P}^{x,\vartheta}(X_{T_1} \in A \,|\, T_1 = s) = Q(x,s,A).$$

The component θ, called the age process, is defined as

$$\theta_s = \begin{cases} \theta_0 + s & \text{if}\, X_p = X_s \quad \forall\, 0 \leqslant p \leqslant s,\ p,s \in \mathbb{R}, \\ s - \sup\{\, p:\ 0 \leqslant p \leqslant s,\ X_p \neq X_s\} & \text{otherwise.} \end{cases}$$

We notice that the component X alone is not a Markov process. The existence of a semi-Markov process of the type above is a well known fact, see for instance Stone [**125**]. Our main restriction is that the jump rate function λ is uniformly bounded, which implies that the process X is non explosive. Denoting by T_n the jump times of X, we consider the marked point process (T_n, X_{T_n}) with the associated integer-valued random measure $p(dt\,dy) = \sum_{n\geq 1} \delta_{(T_n, X_{T_n})}$ on $(0,\infty)\times E$, where δ indicates the Dirac measure. The compensator $\tilde{p}$ of p has the form $\tilde{p}(ds\,dy) = \lambda(X_{s-},\theta_{s-})\,Q(X_{s-},\theta_{s-},dy)\,ds$.

We focus on optimal intensity-control problem for the semi-Markov process introduced above. This is formulated in a classical way by means of a change of probability measure, see e.g. El Karoui [**49**], Elliott [**57**], Brémaud [**18**]. In our formulation we admit control actions that can depend not only on the state process X but also on the length of time θ the process has remained in that state. This approach can be found for instance in Chitopekar [**24**] and in [**125**]. The class of admissible control processes, denoted by $\mathcal{A}$, contains all the predictable processes $(u_s)_{s\in[0,\,T]}$ with values in U. For every fixed $t \in [0,\,T]$ and $(x,\vartheta) \in E \times [0,\infty)$, we define the value function of the optimal control problem as

$$V(t,x,\vartheta) \;\; = \inf_{u(\cdot)\in\mathcal{A}} \mathbb{E}_{u,t}^{x,\vartheta}\left[\int_0^{T-t} l(t+s, X_s, \theta_s, u_s)\,ds + g(X_{T-t}, \theta_{T-t})\right],$$

where g, l are given real functions. Here $\mathbb{E}_{u,t}^{x,\vartheta}$ denotes the expectation with respect to another probability $\mathbb{P}_{u,t}^{x,\vartheta}$, depending on t and on the control process u, and constructed in such a way that the compensator under $\mathbb{P}_{u,t}^{x,a}$ is $r(t+s, X_{s-}, \theta_{s-}, y, u_s)$ $\lambda(X_{s-},\theta_{s-})\,Q(X_{s-},\theta_{s-},dy)\,ds$, where r is some given measurable function.

Our approach to this control problem consists in introducing a family of BSDEs parametrized by $(t,x,\vartheta) \in [0,T]\times E\times[0,\infty)$, on $[0,\,T-t]$:

$$Y_{s,t}^{x,\vartheta} + \int_s^{T-t}\int_E Z_{\sigma,t}^{x,\vartheta}(y)\,q(d\sigma\,dy) = g(X_{T-t},\theta_{T-t}) + \int_s^{T-t} f\Big(t+\sigma, X_\sigma, \theta_\sigma, Z_{\sigma,t}^{x,\vartheta}(\cdot)\Big)\,d\sigma, \tag{11}$$

where $q(ds\,dy)$ denotes the compensated random measure $p(ds\,dy) - \tilde{p}(ds\,dy)$. The generator of (11) is the Hamiltonian function:

$$f(s,x,\vartheta,z(\cdot)) = \inf_{u\in U}\Big\{\, l(s,x,\vartheta,u) + \int_E z(y)(r(s,x,\vartheta,y,u)-1)\,\lambda(x,\vartheta)\,Q(x,\vartheta,dy)\Big\}. \tag{12}$$

Under appropriate assumptions, the previous optimal control problem has a solution, and the corresponding value function and optimal control can be represented by means of the solution to the BSDE (11). In order to prove the existence of an optimal control we need to require that the infimum in the definition of f is achieved. We define the (possibly empty) sets

$$\begin{aligned}\Gamma(s,x,\vartheta,z(\cdot)) = \Big\{\, u\in U : f(s,x,\vartheta,z(\cdot)) &= l(s,x,\vartheta,u)\\ &+ \int_E z(y)(r(s,x,\vartheta,y,u)-1)\,\lambda(x,\vartheta)\,Q(x,\vartheta,dy)\Big\}\end{aligned} \tag{13}$$

and we assume that the following condition holds.

Hypothesis 1. The sets Γ in (13) are non empty; moreover, for every fixed $t\in[0,\,T]$ and $(x,\vartheta)\in S$, one can find a predictable process $u^{*\,t,x,\vartheta}(\cdot)$ with values in U satisfying

$$u_s^{*\,t,x,\vartheta} \in \Gamma(t+s, X_{s-}, \theta_{s-}, Z_{s,t}^{x,\vartheta}(\cdot)), \qquad \mathbb{P}^{x,\vartheta}\text{-a.s. } \forall s\in[0,\,T-t]. \tag{14}$$

Theorem 2. *Assume that Hypothesis 1 holds. Then, under suitable measurability and integrability conditions on r, l and g, $u^{*\,t,x,\vartheta}(\cdot)$ is an optimal control for the control problem starting from (x,ϑ) at time zero with time horizon $T-t$. Moreover, $Y_{0,t}^{x,\vartheta}$ coincides with the value function, i.e.*

$$Y_{0,t}^{x,\vartheta} = J(t,x,\vartheta,u^{*\,t,x,\vartheta}(\cdot)).$$

At this point we solve a nonlinear variant of the Kolmogorov equation for the process (X,θ) by means of the BSDEs approach. The integro-differential infinitesimal generator associated to the process (X,θ) (which is time-homogeneous, Markov, but not pure jump) has the form

$$\tilde{\mathcal{L}}\psi(x,\vartheta) := \partial_\vartheta\psi(x,\vartheta) + \int_K [\psi(y,0)-\psi(x,\vartheta)]\,\lambda(x,\vartheta)\,Q(x,\vartheta,dy), \quad (x,\vartheta)\in E\times[0,\infty).$$

The differential term ∂_θ does not allow to study the associated nonlinear Kolmogorov equation proceeding as in the pure jump Markov processes framework considered in [**28**]. On the other hand, the two dimensional Markov process $(X_s,\theta_s)_{s\geqslant 0}$ belongs to the larger class of piecewise deterministic Markov processes (PDMPs) introduced by Davis in [**35**], and studied in the optimal control framework by several authors, see Section I.4 below and references therein. Taking into account the specific structure of the semi-Markov processes, we present a reformulation of the Kolmogorov equation which allows us to consider solutions in a classical sense. Indeed, since the second component of the process $(X_s,\,\theta_s)_{s\geqslant 0}$ is linear in s, we introduce the formal directional derivative operator

$$(Dv)(t,x,\vartheta) := \lim_{h\downarrow 0}\frac{v(t+h,x,\vartheta+h)-v(t,x,\vartheta)}{h},$$

and we consider the following nonlinear Kolmogorov equation

$$\begin{cases} Dv(t,x,\vartheta) + \mathcal{L}v(t,x,\vartheta) + f(t,x,\vartheta,v(t,x,\vartheta),v(t,\cdot,0) - v(t,x,\vartheta)) = 0, \\ \hfill t \in [0,T],\ x \in E,\ \vartheta \in [0,\infty), \\ v(T,x,\vartheta) = g(x,\vartheta), \end{cases} \tag{15}$$

where

$$\mathcal{L}\psi(x,\vartheta) := \int_E [\psi(y,0) - \psi(x,\vartheta)]\,\lambda(x,\vartheta)\,Q(x,\vartheta,dy), \qquad (x,\vartheta) \in E \times [0,\infty).$$

We look for a solution v such that the map $t \mapsto v(t,x,t+c)$ is absolutely continuous on $[0,T]$, for all constants $c \in [-T,\,+\infty)$. While it is easy to prove well-posedness of (15) under boundedness assumptions on f and g, we show that there exists a unique solution under much weaker conditions related to the distribution of the process (X,θ). This is achieved by defining a formula of Itô type, involving the directional derivative operator D, for the composition of the process $(X_s,\,\theta_s)_{s\geqslant 0}$ with functions v smooth enough. In conclusion we have the following result.

Theorem 3. *Under suitable measurability and integrability conditions on f and g, the nonlinear Kolmogorov equation* (15) *has a unique solution $v(t,x,\vartheta)$. Moreover, for every fixed $t \in [0,\,T]$, for every $(x,\vartheta) \in E \times [0,\,\infty)$ and $s \in [0,\,T-t]$,*

$$Y^{x,\vartheta}_{s,t} = v(t+s, X_{s-}, \theta_{s-}), \tag{16}$$

$$Z^{x,\vartheta}_{s,t}(y) = v(t+s,y,0) - v(t+s,X_{s-},\theta_{s-}), \tag{17}$$

so that in particular $v(t,x,\vartheta) = Y^{x,\vartheta}_{0,t}$.

At this point, we go back to the original control problem and we observe that the associated Hamilton-Jacobi-Bellman equation has the form (15) with f given by the Hamiltonian function (12). Then, taking into account Theorems 2 and 3, we are able to identify the HJB solution $v(t,x,\vartheta)$, constructed probabilistically via BSDEs, with the value function.

Corollary 4. *Assume that Hypothesis 1 holds. Then, under suitable measurability and integrability conditions on r, l and g, the value function coincides with $v(t,x,\vartheta)$, i.e.*

$$J(t,x,\vartheta,u^{*\,t,x,\vartheta}(\cdot)) = v(t,x,\vartheta) = Y^{x,\vartheta}_{0,t}.$$

I.3. Optimal control of pure jump processes. In **Chapter 2** we study a classical finite-horizon optimal control problem for continuous-time pure jump Markov processes. For the value function of this problem, we prove a nonlinear Feynman-Kac formula by extending in a suitable way the control randomization method in [**88**].

We consider controlled pure jump Markov processes taking values in a Lusin space $(E,\mathcal{E})$. They are obtained starting from a rate measure $\lambda(x,a,B)$ defined for $x \in E$, $a \in A$, $B \in \mathcal{E}$, where A is a space of control actions equipped with its σ-algebra $\mathcal{A}$. These Markov processes are controlled by choosing a feedback control law, namely a measurable function $\alpha : [0,\infty)\times E \to A$, such that $\alpha(t,x) \in A$ is the control action selected at time t if the system is in state x. The controlled Markov process X is then simply the one corresponding to the rate transition measure $\lambda(x,\alpha(t,x),B)$.

We denote by $\mathbb{P}_\alpha^{t,x}$ the corresponding law, where t, x are the initial time and starting point. For convenience, we base this "weak construction" on the well-posedness of the martingale problem for multivariate (marked) point processes studied in Jacod [**75**]. Indeed, on a canonical space Ω, we define an E-valued random variable E_0 and a marked point process $(T_n, E_n)_{n\geq 1}$ with values in $E \times (0, \infty]$, with corresponding random measure

$$p(dt\,dy) = \sum_{n\geq 1} 1_{\{T_n<\infty\}}\,\delta_{(T_n,E_n)}(dt\,dy).$$

The process X is constructed by setting $X_t = E_n$ for every $t \in [T_n,\, T_{n+1})$. Moreover, for all $s \geq 0$ we define $\mathcal{F}_s = \mathcal{G}_s \vee \sigma(E_0)$, where $\mathcal{G}_t$ denotes the σ-algebra generated by the marked point process up to time $t > 0$. Then, according to Theorem 3.6 in [**75**], the law $\mathbb{P}_\alpha^{t,x}$ is the unique probability measure on $(\Omega, \mathcal{F}_\infty)$ such that its restriction to $\mathcal{F}_0$ is the Dirac measure concentrated at x, and the $(\mathcal{F}_t)_{t\geq 0}$-compensator of the measure p is the random measure $\lambda(X_{s-}, \alpha(s, X_{s-}), dy)\,ds$.

The value function of the corresponding control problem with finite time horizon $T > 0$ is defined as:

$$V(t,x) = \sup_\alpha \mathbb{E}_\alpha^{t,x}\left[\int_t^T f(s,\, X_s, \alpha(s, X_s))\,ds + g(X_T)\right], \quad t \in [0,T],\, x \in E, \tag{18}$$

where $\mathbb{E}_\alpha^{t,x}$ denotes the expectation with respect to $\mathbb{P}_\alpha^{t,x}$, and f, g are given real functions, defined respectively on $[0,T] \times E \times A$ and on E, and representing the running cost and the terminal cost. We consider the case when the costs f ad g are bounded and

$$\sup_{(x,a)\in E\times A} \lambda(x,a,E) < \infty. \tag{19}$$

The optimal control problem is associated to the following first-order fully nonlinear integro-differential HJB equation on $[0,\, T] \times E$:

$$\begin{cases} -\frac{\partial v}{\partial t}(t,x) &= \sup_{a\in A}\left(\int_E (v(t,y) - v(t,x))\,\lambda(x,a,dy) + f(t,x,a)\right), \\ v(T,x) &= g(x). \end{cases} \tag{20}$$

Notice that the integral operator in the HJB equation allows for easy notions of solutions, that avoid the use of the theory of viscosity solutions. Indeed, under suitable measurability assumptions, a bounded function $v : [0,\, T] \times E \to \mathbb{R}$ is a solution to (20) if the terminal condition holds, (20) holds almost surely on $[0,\, T]$, and $t \mapsto v(t,x)$ is absolutely continuous in $[0,\, T]$.

For the HJB equation (20) we present a classical result on existence and uniqueness of the solution and the identification property with the value function V. The compactness of the space of control actions A, usually needed to ensure the existence of an optimal control (see Pliska [**108**]), is not asked here. This is possible by using a different measurable selection result requiring however lower-semicontinuity conditions, that may be found for instance in Bertsekas and Shreve [**15**]. We have the following result.

Theorem 5. *Assume that λ has the Feller property and satisfies* (19)*, and that f, g are bouded and lower-semicontinuous functions. Then there exists a unique solution*

$v \in LSC_b([0,T] \times E)$ to the HJB equation, and it coincides with the value function V.

At this point, in order to relate the value function $V(t,x)$ to an appropriate BSDE, we implement the control randomization method in [**88**] in the pure jump framework. Finding the correct formulation required some efforts; in particular we could not mimic the works on control randomization in the diffusive framework, where the controlled process is defined as the solution to a stochastic differential equation.

In a first step, for any initial time $t \geq 0$ and starting point $x \in E$, we replace $(X_s, \alpha(s, X_s))$ by an (uncontrolled) Markovian pair of pure jump stochastic processes (X_s, I_s), in such a way that the process I is a Poisson process with values in the space of control actions A, with an intensity measure $\lambda_0(db)$ which is arbitrary but finite and with full support. The construction of such a pair of pure jump processes relies on the well-posedness of the martingale problem for marked point processes recalled before, and is obtained by assigning a rate transition measure on $E \times A$ of the form:

$$\lambda_0(db)\, \delta_x(dy) + \lambda(x, a, dy)\, \delta_a(db).$$

Next we formulate an auxiliary optimal control problem where we control the intensity of the process I: for any predictable, bounded and positive random field $\nu_t(b)$, by means of a theorem of Girsanov type we construct a probability measure $\mathbb{P}_\nu^{t,x,a}$ under which the compensator of I is the random measure $\nu_t(b)\, \lambda_0(db)\, dt$ (under $\mathbb{P}_\nu^{t,x,a}$ the law of X also changes) and then we maximize the functional

$$\mathbb{E}_\nu^{t,x,a}\left[g(X_T) + \int_t^T f(s,\, X_s,\, I_s)\, ds\right],$$

over all possible choices of the process ν. Following the terminology of [**88**], this will be called the *dual* control problem. Its value function, denoted $V^*(t,x,a)$, also depends a priori on the starting point $a \in A$ of the process I, and the family $\{\mathbb{P}_\nu^{t,x,a}\}_\nu$ is a dominated model.

At this point, we can introduce a BSDE that represents $V^*(t,x,a)$. It is an equation on the time interval $[t,T]$ of the form

$$\begin{aligned} Y_s^{t,x,a} &= g(X_T) + \int_s^T f(r, X_r, I_r)\, dr + K_T^{t,x,a} - K_s^{t,x,a} \\ &\quad - \int_s^T \int_{E\times A} Z_r^{t,x,a}(y,\, b)\, q(dr\, dy\, db) - \int_s^T \int_A Z_r^{t,x,a}(X_r,\, b)\, \lambda_0(db)\, dr, \end{aligned} \tag{21}$$

with unknown triple $(Y^{t,x,a}, Z^{t,x,a}, K^{t,x,a})$, where q is the compensated random measure associated to (X,I), Z is a predictable random field and K a predictable increasing càdlàg process, where we additionally add the sign constraint

$$Z_s^{t,x,a}(X_{s-}, b) \leqslant 0. \tag{22}$$

Under the previous conditions, this equation has a unique minimal solution (Y,Z,K) in a certain class of processes, and a dual representation formula holds.

Theorem 6. *For all $(t,x,a) \in [0,\,T] \times E \times A$ there exists a unique minimal solution $(Y^{t,x,a},\, Z^{t,x,a},\, K^{t,x,a})$ to (21)-(22). Moreover, for all $s \in [t,T]$, $Y_s^{t,x,a}$ has the explicit representation: $\mathbb{P}^{t,x,a}$-a.s.,*

$$Y_s^{t,x,a} = \operatorname*{ess\,sup}_{\nu \in \mathcal{V}} \mathbb{E}_\nu^{t,x,a}\left[g(X_T) + \int_s^T f(r, X_r, I_r)\, dr \,\middle|\, \mathcal{F}_s \right], \quad s \in [t,T]. \tag{23}$$

In particular, setting $s = t$, we have the following representation formula for the value function of the dual control problem:

$$V^*(t,x,a) = Y_t^{t,x,a}, \qquad (t,x,a) \in [0,T] \times E \times A. \tag{24}$$

The proof of this result relies on a penalization approach and a monotonic passage to the limit. More precisely, we introduce the following family of BSDEs with jumps indexed by $n \geqslant 1$ on $[t,\,T]$:

$$\begin{aligned} Y_s^{n,t,x,a} \;\; &= \;\; g(X_T) + \int_s^T f(r, X_r, I_r)\, dr + K_T^{n,t,x,a} - K_s^{n,t,x,a} \\ &\quad - \int_s^T \int_{E\times A} Z_r^{n,t,x,a}(y,b)\, q(dr\, dy\, db) - \int_s^T \int_A Z_r^{n,t,x,a}(X_r,b)\, \lambda_0(db)\, dr, \end{aligned} \tag{25}$$

where $K^{n,t,x,a}$ is the nondecreasing process defined by

$$K_s^{n,t,x,a} = n \int_t^s \int_A [Z_r^{n,t,x,a}(X_r,b)]^+ \, \lambda_0(db)\, dr.$$

Here $[u]^+$ denotes the positive part of u. The existence and uniqueness of a solution $(Y^{n,t,x,a}, Z^{n,t,x,a})$ to the BSDE (25) relies on a standard procedure, based on a fixed point argument and on integral representation results for martingales. Notice that the use of the filtration $(\mathcal{F})_{t\geq 0}$ introduced above is essential, since it involves application of martingale representation theorems for multivariate point processes (see e.g. Theorem 5.4 in [**75**]). The first component of this solution turns out to satisfy

$$Y_s^{n,t,x,a} = \operatorname*{ess\,sup}_{\nu \in \mathcal{V}^n} \mathbb{E}_\nu^{t,x,a}\left[g(X_T) + \int_s^T f(r, X_r, I_r)\, dr \,\middle|\, \mathcal{F}_s \right], \tag{26}$$

where $\mathcal{V}^n$ denotes the subset of controls ν bounded by n. Since the sets $\mathcal{V}^n$ are nested, we have that $(Y^{n,t,x,a})_n$ increasingly converges to $Y^{t,x,a}$ as n goes to infinity. Together with uniform estimates on $(Z^{n,t,x,a}, K^{n,t,x,a})_n$, this allows a monotonic passage into the limit and gives the existence of the minimal solution to the constrained BSDE (21)-(22). Finally, from (26), by control-theoretic considerations we also get the dual representation formula (23) for the minimal solution $Y^{t,x,a}$.

At this point, we need to relate the original optimal control problem with the dual one.

We start by proving that the dual value function does not depend on a. To this end, denoted $v^n(t,x,a) := Y_t^{n,t,x,a}$ and $\bar v(t,x,a) := V^*(t,x,a)$, we consider the penalized HJB equation in the integral form satisfied by $Y^{n,t,x,a}$:

$$\left\{ \begin{array}{lll} -\partial_t v^n(t,x,a) & = \int_E \left(v^n(t,y,a) - v^n(t,x,a)\right)\, \lambda(x,a,dy) & \\ & \quad + f(t,x,a) + n \int_A [v^n(t,x,b) - v^n(t,x,a)]^+ \lambda_0(db), & \\ v^n(T,x,a) & = g(x). & \end{array} \right. \tag{27}$$

Passing to the limit in (27) when n goes to infinity, taking into account that $\bar{v}$ is right-continuous, we get

$$\int_A [\bar{v}(t,x,b) - \bar{v}(t,x,a)]^+ \lambda_0(db) = 0$$

and by further arguments this finally allows to conclude that $\bar{v}(t,x,a) = \bar{v}(t,x)$.

Then, going back to the penalized HJB equation (27) and passing to the limit, we see that $\bar{v}$ is a classical supersolution of (20). In particular $\bar{v}$ is greater than the unique solution to the HJB equation. By control-theoretic considerations we also prove that $\bar{v}$ is smaller than the value function V. We conclude that the value function of the dual optimal control problem coincides with the value function of the original control problem.

Theorem 7. *Let v be the unique solution to the Hamilton-Jacobi-Bellman equation provided by Theorem 5. Then for every $(t,x,a) \in [0,T] \times E \times A$, the nonlinear Feynman-Kac formula holds:*

$$v(t,x) = V(t,x) = Y_t^{t,x,a}.$$

In particular, the value function V of the optimal control problem defined in (18) *and the dual value function V^* defined in* (24) *coincide.*

I.4. Optimal control of PDMPs. In **Chapter 3** we prove that the value function in an infinite-horizon optimal control problem for piecewise deterministic Markov processes (PDMPs) can be represented by means of an appropriate constrained BSDE. As in Chapter 2, this is obtained by suitably extending the control randomization method in [**88**]. Compared to the pure jump case, the PDMPs context is more involved and requires different techniques. In particular, the presence of the controlled flow in the PDMP's dynamics and the corresponding differential operator in the HJB equation suggest to use the theory of viscosity solutions. In addition, we consider discounted infinite-horizon optimal control problems, where the payoff is a cost to be minimized. Such problems are very traditional for PDMPs, see e.g. Davis [**35**], Costa and Dufour [**32**], Guo and Hérnandez-Lerma [**72**]; moreover the finite-horizon case can be brought back to the infinite-horizon case by means of a standard transformation, see Chapter 3 in [**35**]. The infinite-horizon character of the optimal control problems complicates the tractation via the BSDE techniques, since it leads to deal with BSDEs over an infinite time horizon as well.

We consider controlled PDMPs on a general measurable state space $(E, \mathcal{E})$. These processes are obtained starting from a continuous deterministic flow $\phi^\beta(t,x)$, $(t,x) \in [0,\infty) \times E$, depending on the choice of a function $\beta(t)$ taking values on the space of control actions $(A, \mathcal{A})$, and from a jump rate $\lambda(x,a)$ and a transition measure $Q(x,a,dy)$ on E, depending both on $(x,a) \in E \times A$. We select the control strategy among the set of *piecewise open-loop policies*, i.e., measurable functions that depend only on the last jump time and post jump position. This kind of approach is habitual in the literature, see for instance Almudevar [**1**], Davis [**34**], Bauerle and Rieder [**11**], Lenhart and Yamada [**91**], Dempster [**40**]. Roughly speaking, at each jump time T_n, we choose an open loop control α_n depending on the initial condition X_{T_n} to be used until the next jump time. A control α in this class of admissible control laws, denoted by $\mathcal{A}_{ad}$, has the explicit form

$$\alpha_t = \sum_{n=1}^{\infty} \alpha_n(t - T_n, X_{T_n})\, 1_{[T_n,\, T_{n+1})}(t), \tag{28}$$

and the controlled process X is

$$X_t = \phi^{\alpha_n}(t - T_n, X_{T_n}), \quad t \in [T_n,\, T_{n+1}).$$

For any $x \in E$ and $\alpha \in \mathcal{A}_{ad}$, $\mathbb{P}^x_\alpha$ indicates the probability measure such that, for every $n \geq 1$, the conditional survivor function of the jump time T_{n+1} and the distribution of the post jump position $X_{T_{n+1}}$ on $\{T_n < \infty\}$ are

$$\begin{aligned}
&\mathbb{P}^x_\alpha(T_{n+1} > s \,|\, \mathcal{F}_{T_n}) = \exp\left(-\int_{T_n}^{s} \lambda(\phi^{\alpha_n}(r - T_n, X_{T_n}), \alpha_n(r - T_n, X_{T_n}))\, dr\right),\\
&\mathbb{P}^x_\alpha(X_{T_{n+1}} \in B |\, \mathcal{F}_{T_n},\, T_{n+1}) = Q(\phi^{\alpha_n}(T_{n+1} - T_n, X_{T_n}), \alpha_n(T_{n+1} - T_n, X_{T_n}), B).
\end{aligned}$$

The corresponding value function, depending on $x \in E$, is defined as:

$$\begin{aligned}
V(x) &= \inf_{\alpha \in \mathcal{A}_{ad}} \mathbb{E}^x_\alpha \left[\int_0^\infty e^{-\delta\, s}\, f(X_s, \alpha_s)\, ds\right] \qquad (29)\\
&= \inf_{\alpha \in \mathcal{A}_{ad}} \mathbb{E}^x_\alpha \left[\int_0^\infty e^{-\delta\, s} \sum_{n \in \mathbb{N}} f(\phi^{\alpha_n}(s - T_n, X_{T_n}), \alpha_n(s - T_n, X_{T_n}))\, 1_{[T_n, T_{n+1})}(s)\, ds\right],
\end{aligned}$$

where $\mathbb{E}^x_\alpha$ indicates the expectation with respect to $\mathbb{P}^x_\alpha$, f is a given real function on $E \times A$ representing the running cost, and $\delta \in (0,\, \infty)$ is a discounting factor. We assume that λ and f are bounded functions, uniformly continuous, and Q is a Feller stochastic kernel.

When E is an open subset of $\mathbb{R}^d$, and $h(x, a)$ is a bounded Lipschitz continuous function, $\phi^\alpha(t, x)$ is defined as the unique solution of the ordinary differential equation

$$\dot{x}(t) = h(x(t), \alpha(t)), \quad x(0) = x \in E.$$

In this case, according to Davis and Farid [**36**], under the compactness assumption for the space of control actions A, the value function V is the unique continuous viscosity solution on $[0,\, \infty) \times E$ to the fully-nonlinear, integro-differential HJB equation

$$\delta v(x) = \sup_{a \in A}\left(h(x, a) \cdot \nabla v(x) + \lambda(x, a) \int_E (v(y) - v(x))\, Q(x, a, dy)\right) \quad x \in E. \tag{30}$$

Our main goal is to represent the value function $V(x)$ by means of an appropriate backward stochastic differential equation. To this end, we implement the control randomization method in the PDMPs framework. The first step consists in replacing, for any starting point $x \in E$, the state trajectory and the associated control process (X_s, α_s) by an uncontrolled PDMP (X_s, I_s). The process (X, I) takes values on $E \times A$, and is constructed in a canonical way by assigning a new triplet of local characteristics. The compensator corresponding to (X, I) is the random measure

$$\tilde{p}(ds\, dy\, db) = \lambda_0(db)\, \delta_x(dy)\, ds + \lambda(x, a)\, Q(x, a, dy)\, \delta_a(db)\, ds.$$

In particular, I is a Poisson process with values in the space of control actions A, with an arbitrary intensity $\lambda_0(db)$ finite and with full topological support. For any fixed starting point (x,a) in $E\times A$, $\mathbb{P}^{x,a}$ denotes the unique solution to the martingale problem for marked point processes on $E \times A$, corresponding to $\tilde p$ and (x,a). The trajectories of the process X are then constructed as above, with the help of the deterministic flow associated to the vector field h.

At this point, we define a dual control problem, where we control the intensity of the process I. To this end, we consider the class of predictable, bounded and positive random fields $\nu_t(b)$, and we construct a probability measure $\mathbb{P}^{x,a}_\nu$ under which the compensator of I is the random measure $\nu_s(db)\,\lambda_0(db)\,ds$. The dual control problem consists then in minimizing over all admissible ν the functional

$$J(x,a,\nu) = \mathbb{E}^{x,a}_\nu\left[\int_0^\infty e^{-\delta\, s}\, f(X_s, I_s)\, ds\right]. \tag{31}$$

The dual value function $V^*(x,a) = \inf_{\nu\in\mathcal{V}} J(x,a,\nu)$ can be represented by means of a BSDE over *infinite horizon*, of the form

$$\begin{aligned} Y^{x,a}_s &= Y^{x,a}_T - \delta \int_s^T Y^{x,a}_r\, dr + \int_s^T f(X_r, I_r)\, dr - (K^{x,a}_T - K^{x,a}_s) \qquad (32)\\ &- \int_s^T \int_A Z^{x,a}_r(X_r,\, b)\, \lambda_0(db)\, dr - \int_s^T \int_{E\times A} Z^{x,a}_r(y,\, b)\, q(dr\, dy\, db), \quad 0 \leqslant s \leqslant T < \infty, \end{aligned}$$

with the sign constraint

$$Z^{x,a}_s(X_{s-}, b) \geqslant 0. \tag{33}$$

Under suitable conditions, equation (32)-(33) has a unique maximal (not minimal since the payoff is a cost to be minimized) solution (Y,Z,K) in a certain class of processes, and Y admits a dual representation formula.

Theorem 8. *For every $(x,a) \in E\times A$, there exists a unique maximal solution to the BSDE with partially nonnegative jumps* (32)-(33)*. Moreover, $Y^{x,a}$ has the explicit representation:*

$$Y^{x,a}_s = \operatorname*{ess\,inf}_{\nu\in\mathcal{V}} \mathbb{E}^{x,a}_\nu\left[\int_s^\infty e^{-\delta(r-s)}\, f(X_r, I_r)\, dr\Big|\mathcal{F}_s\right], \ \forall\ s \geqslant 0. \tag{34}$$

In particular, setting $s=0$, we have the following dual representation formula:

$$V^*(x,a) = Y^{x,a}_0, \quad (x,a) \in E\times A. \tag{35}$$

The proof of this result relies as usual on a penalization approach and a monotonic passage to the limit. However, since we deal with infinite-horizon equations, we need to implement an additional approximating step, where we introduce a family of penalized BSDEs depending on a finite horizon $T>0$. More precisely, for $n\geq 1$, we consider the following family of penalized BSDEs on $[0,\,\infty)$:

$$\begin{aligned} Y^{n,x,a}_s &= Y^{n,x,a}_T - \delta\int_s^T Y^{n,x,a}_r\, dr + \int_s^T f(X_r, I_r)\, dr\\ &\quad -n\int_s^T \int_A [Z^{n,x,a}_r(X_r,b)]^-\, \lambda_0(db)\, dr - \int_s^T\int_A Z^{n,x,a}_r(X_r,b)\, \lambda_0(db)\, dr \end{aligned}$$

$$-\int_s^T \int_{E\times A} Z_r^{n,x,a}(y,b)\, q(dr\,dy\,db), \quad 0 \leqslant s \leqslant T < \infty, \tag{36}$$

where $[z]^- = \max(-z, 0)$ denotes the negative part of z. In order to study the well-posedness of equation (36), we introduce a second family of penalized BSDEs, also parametrized by $T > 0$, and with zero final cost:

$$\begin{aligned} Y_s^{T,n,x,a} &= -\delta \int_s^T Y_r^{T,n,x,a}\, dr + \int_s^T f(X_r, I_r)\, dr \\ &\quad - n \int_s^T \int_A [Z_r^{T,n,x,a}(X_r, b)]^- \,\lambda_0(db)\, dr \\ &\quad - \int_s^T \int_A Z_r^{T,n,x,a}(X_r, b)\, \lambda_0(db)\, dr \\ &\quad - \int_s^T \int_{E\times A} Z_r^{T,n,x,a}(y,b)\, q(dr\,dy\,db), \quad 0 \leqslant s \leqslant T. \end{aligned} \tag{37}$$

The existence of a unique solution $(Y^{T,n}, Z^{T,n})$ to (37) is a well known fact, and relies as usual on fixed point arguments. We prove that the sequence $(Y^T)_{T>0}$ converges $\mathbb{P}^{x,a}$-a.s. to some process Y, uniformly on compact subsets of $\mathbb{R}_+$, and that, for any $S > 0$, the sequence $(Z^{n,T}|_{[0,\,S]})_{T>S}$ converges to some process $Z^n|_{[0,\,S]}$ in a suitable sense. This allows to pass to the limit in (37), and, the time S being arbitrary, to conclude that $(Y^n,\, Z^n)$ is the unique solution to (36). The process Y^n satisfies the dual representation formula:

$$Y_s^{n,x,a} = \operatorname*{ess\,inf}_{\nu\in\mathcal{V}^n} \mathbb{E}_\nu^{x,a}\left[\int_s^\infty e^{-\delta\,(r-s)}\, f(X_r, I_r)\, dr \middle| \mathcal{F}_s\right], \quad s \geqslant 0, \tag{38}$$

where $\mathcal{V}^n$ denotes the subset of controls ν bounded by n.

By (38) we see that $(Y^n)_n$ increasingly converges to Y as n goes to infinity. Moreover we provide uniform estimates on $(Z^n|_{[0,\,S]}, K^n|_{[0,\,S]})_n$ for every $S > 0$. Then we monotonically pass into the limit in (36) and we get the existence of the (unique) maximal solution (Y, Z, K) to the constrained BSDE (32)-(33), for which we also prove the dual representation formula (34).

Finally, we show that the maximal solution to (32)-(33) at the initial time also provides a Feynman-Kac representation of the value function (29) of our original optimal control problem for PDMPs. To this end we introduce the deterministic real function on $E \times A$

$$v(x,a) := Y_0^{x,a}. \tag{39}$$

We have the following result.

Theorem 9. *The function v in* (39) *does not depend on the variable a:*

$$v(x,a) = v(x,a'), \quad \forall a, a' \in A,$$

for all $x \in E$. Let us define

$$v(x) = v(x,a), \quad \forall x \in E,$$

for any $a \in A$. Then v is a viscosity solution to (30).

Notice that the concept of viscosity solution we use does not require continuity properties; this is usually called discontinuous viscosity solution.

The fact that the function v in (39) is independent on its last component (which is a consequence of the A-nonnegative constrained jumps) has a key role in the derivation of the viscosity solution properties of v, and the proof of this feature constitutes a relevant task. Differently from [**88**] and the related papers in the diffusive context, this is obtained exclusively by means of control-theoretic techniques and relies on the identification formula (35). By avoiding the use of viscosity theory tools, no additional hypothesis is required on the space of controls A, which can therefore be very general. The non-dependence of v on a is a consequence of the following result.

Proposition 10. *Fix $x \in E$, $a, a' \in A$, and $\nu \in \mathcal{V}$. Then, there exists a sequence $(\nu^\varepsilon)_\varepsilon \in \mathcal{V}$ such that*

$$\lim_{\varepsilon \to 0^+} J(x, a', \nu^\varepsilon) = J(x, a, \nu). \tag{40}$$

Indeed identity (40) implies that $V^*(x, a') \leq J(x, a, \nu)$, for every $x \in E$, $a, a' \in A$. By the arbitrariness of ν it follows that

$$V^*(x, a') \leq V^*(x, a)$$

and, exchanging the roles of a and a', this allows to conclude that $V^*(x, a) = v(x, a)$ does not depend on a.

Once we get that V^* (and therefore v) does not depend on a, we show that it actually provides a viscosity solution to the HJB equation (30). Differently to the previous literature, we give a direct proof of the viscosity solution property of v, which avoid to resort to a penalized HJB equation. This is achieved by generalizing to the setting of the dual control problem the classical proof that allows to derive the HJB equation from the dynamic programming principle. As a preliminary step, we need to give an identification result of the following form.

Lemma 11. *The function v is such that, for any $(x, a) \in E \times A$, we have*

$$Y_s^{x,a} = v(X_s, I_s), \quad s \geqslant 0 \quad d\mathbb{P}^{x,a} \otimes ds \text{ -a.e.} \tag{41}$$

Identification (41) is proved by showing an analogous result for Y^n, and using the convergence of Y^n to Y provided in Theorem 8. This result follows from the Markov property of the state process (X, I), and relies on an iterative construction of the solution of standard BSDEs inspired by El Karoui, Peng and Quenez [**53**].

Finally, to conclude that $v(x)$ actually gives the unique solution to the HJB equation we need to use a comparison theorem for viscosity sub and supersolutions to the equation (30). Under an additional assumption on λ and Q (see condition **(HλQ')**), and the compactness of A, the above mentioned comparison theorem insures that v is the unique viscosity solution to (30), which coincides therefore to the value function V. This yields in particular the nonlinear Feynman-Kac formula for V, as well as the equality between the value functions of the primal and the dual control problems.

Corollary 12. *Assume that A is compact, and that Hypothesis* **(HλQ')** *holds. Then the value function V of the optimal control problem defined in* (29) *admits the non-linear Feynman-Kac representation formula:*

$$V(x) = Y_0^{x,a}, \quad (x,a) \in E \times A.$$

Moreover, $V(x) = V^*(x,a)$.

II. BSDEs driven by general random measures, possibly non quasi-left continuous

As we have already mentioned, BSDEs with discontinuous driving terms have been considered by many authors, among which Barles, Buckdahn and Pardoux [**10**], El Karoui and Huang [**50**], Xia [**131**], Becherer [**12**], Carbone, Ferrario, Santacroce [**22**], Cohen and Elliott [**26**], Jeanblanc, Mania, Santacroce and Schweizer [**80**], Confortola, Fuhrman and Jacod [**29**]. In all the papers cited above, and more generally in the literature on BSDEs, the generator of the backward stochastic differential equation, usually denoted by f, is integrated with respect to a measure dA, where A is a nondecreasing continuous (or deterministic and right-continuous as in [**26**]) process. In **Chapter 4** we provide an existence and uniqueness result for the general case, i.e. when A is a right-continuous nondecreasing predictable process..

More precisely, consider a finite horizon $T > 0$, a Lusin space $(E, \mathcal{E})$ and a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, with $(\mathcal{F}_t)_{t \geq 0}$ right continuous. We denote by $\mathcal{P}$ the predictable σ-field on $\Omega \times [0,T]$. In Chapter 4 we study the backward stochastic differential equation

$$Y_t = \xi + \int_{(t,T]} f(s, Y_{s-}, Z_s(\cdot))\, dA_s - \int_{(t,T]} \int_E Z_s(x)\, (\mu - \nu)(ds, dx), \quad 0 \leq t \leq T, \tag{42}$$

where μ is an integer valued random measure on $\mathbb{R}_+ \times E$ with compensator $\nu(dt, dx) = dA_t\, \phi_t(dx)$, with A a right-continuous nondecreasing predictable process such that $A_0 = 0$, and ϕ is a transition probability from $(\Omega \times [0,T], \mathcal{P})$ into $(E, \mathcal{E})$. We suppose, without loss of generality, that ν satisfies $\nu(\{t\} \times dx) \leq 1$ identically, so that $\Delta A_t \leq 1$.

For such general BSDE the existence and uniqueness results were at disposal only in particular frameworks, see e.g. [**26**] for the deterministic case, and counter-examples were provided in the general case, see Section 4.3 in [**29**]. For this reason, the existence and uniqueness result is not a trivial extension of known results, and we have to impose an additional technical assumption, which is of course violated by the counter-example presented in [**29**].

Let us give some definitions. For any $\beta \geq 0$, $\mathcal{E}^\beta$ denotes the Doléans-Dade exponential of the process βA, namely

$$\mathcal{E}_t^\beta \;=\; e^{\beta A_t} \prod_{0 < s \leq t} (1 + \beta\, \Delta A_s)\, e^{-\beta\, \Delta A_s}. \tag{43}$$

By $\mathbb{H}^2_\beta(0,T)$ we indicate the set of pairs (Y,Z) such that $Y\colon \Omega\times[0,T]\to\mathbb{R}$ is an adapted càdlàg process satisfying

$$\|Y\|^2_{\mathbb{H}^2_{\beta,Y}(0,T)} \;:=\; \mathbb{E}\bigg[\int_{(0,T]} \mathcal{E}^\beta_t |Y_{t-}|^2\,dA_t\bigg] < \infty, \tag{44}$$

and $Z\colon \Omega\times[0,T]\times E\to\mathbb{R}$ is a predictable random field satisfying

$$\begin{aligned}\|Z\|^2_{\mathbb{H}^2_{\beta,Z}(0,T)} \;:=\; \mathbb{E}\bigg[\int_{(0,T]} \mathcal{E}^\beta_t \int_E \big|Z_t(x)-\hat Z_t\big|^2\,\nu(dt,dx)\\ +\sum_{0<t\le T}\mathcal{E}^\beta_t\big|\hat Z_t\big|^2\big(1-\Delta A_t\big)\bigg] \;<\; \infty,\end{aligned} \tag{45}$$

where

$$\hat Z_t \;=\; \int_E Z_t(x)\,\nu(\{t\}\times dx), \qquad 0\le t\le T.$$

Definition 13. A solution to equation (42) with data (β,ξ,f) is a pair $(Y,Z)\in\mathbb{H}^2_\beta(0,T)$ satisfying equation (42). We say that equation (42) admits a unique solution if, given two solutions $(Y,Z),(Y',Z')\in\mathbb{H}^2_\beta(0,T)$, we have $(Y,Z)=(Y',Z')$ in $\mathbb{H}^2_\beta(0,T)$.

Notice that, given a solution (Y,Z) to equation (42) with data (β,ξ,f), the process $(Z_t 1_{[0,T]}(t))_{t\ge0}$ belongs to the space $\mathcal{G}^2(\mu)$ introduced in Jacod's book [**77**]. In particular, the stochastic integral $\int_{(t,T]}\int_E Z_s(x)\,(\mu-\nu)(ds,dx)$ in (42) is well-defined, and the process $M_t := \int_{(0,t]}\int_E Z_s(x)(\mu-\nu)(ds,dx)$, $t\in[0,T]$, is a square integrable martingale.

Suitable measurability and integrability conditions are imposed on ξ and on f, and f is also asked to verify a uniform Lipschitz condition of the form:

$$\begin{aligned}&|f(\omega,t,y',\zeta')-f(\omega,t,y,\zeta)| \le L_y|y'-y|\\ &+L_z\bigg(\int_E\bigg|\zeta'(x)-\zeta(x)-\Delta A_t(\omega)\int_E\big(\zeta'(z)-\zeta(z)\big)\phi_{\omega,t}(dz)\bigg|^2\phi_{\omega,t}(dx)\\ &+\Delta A_t(\omega)\big(1-\Delta A_t(\omega)\big)\bigg|\int_E(\zeta'(x)-\zeta(x))\,\phi_{\omega,t}(dx)\bigg|^2\bigg)^{1/2},\end{aligned} \tag{46}$$

for some $L_y, L_z\ge 0$. As usual, in order to prove the well-posedness of the BSDE (42) we give a preliminary result, where the existence and uniqueness of the equation is provided where f does not depend on (y,ζ).

Lemma 14. *Consider a triple (β,ξ,f) and suppose that $f=f(\omega,t)$ does not depend on (y,ζ). Then, there exists a unique solution $(Y,Z)\in\mathbb{H}^2_\beta(0,T)$ to equation (42) with data (β,ξ,f). Moreover, the following identity holds:*

$$\begin{aligned}&\mathbb{E}\big[\mathcal{E}^\beta_t\,|Y_t|^2\big]+\beta\,\mathbb{E}\bigg[\int_{(t,T]}\mathcal{E}^\beta_s\,(1+\beta\Delta A_s)^{-1}\,|Y_{s-}|^2\,dA_s\bigg]\\ &+\mathbb{E}\bigg[\int_{(t,T]}\mathcal{E}^\beta_s\int_E\big|Z_s(x)-\hat Z_s\big|^2\,\nu(ds,dx)+\sum_{t<s\le T}\mathcal{E}^\beta_s\big|\hat Z_s\big|^2\big(1-\Delta A_s\big)\bigg]\end{aligned}$$

$$= \mathbb{E}\big[\mathcal{E}^\beta_T\,|\xi|^2\big] + 2\,\mathbb{E}\bigg[\int_{(t,T]} \mathcal{E}^\beta_s\, Y_{s-}\, f_s\, dA_s\bigg] - \mathbb{E}\bigg[\sum_{t<s\leq T} \mathcal{E}^\beta_s\,|f_s|^2\,|\Delta A_s|^2\bigg], \quad (47)$$

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

The proof of Lemma 14 is based on the martingale representation theorem for marked point processes given in [**75**]. In order to prove the existence and uniqueness results we take into account that $M_t := \int_{(t,T]}\int_E Z_s(y)\,(\mu-\nu)(ds\,dy)$ is a square integrable martingale if and only if $Z \in \mathcal{G}^2_{\mathrm{loc}}(\mu)$, and that

$$\langle M, M\rangle_T \;=\; \int_{(0,T]}\int_E \big|Z_t(x) - \hat{Z}_t\big|^2\,\nu(dt,dx) + \sum_{0<t\leq T}\big|\hat{Z}_t\big|^2\big(1-\Delta A_t\big),$$

see Theorem B.22). Properties of the Doléans-Dade exponential $\mathcal{E}^\beta$ are also exploited, in particular we use that $d\mathcal{E}^\beta_s = \beta\,\mathcal{E}^\beta_{s-}\,dA_s$ and that $\mathcal{E}^\beta_{s-} = \mathcal{E}^\beta_s\,(1+\beta\Delta A_s)^{-1}$.

Identity (47) plays a fundamental role to get our main result, which reads as follows.

Theorem 15. *Suppose that there exists* $\varepsilon \in (0,1)$ *such that*

$$2\,L_y^2\,|\Delta A_t|^2 \;\leq\; 1-\varepsilon, \qquad \mathbb{P}\text{-}a.s.,\ \forall\, t\in[0,T]. \quad (48)$$

Then there exists a unique solution $(Y,Z) \in \mathbb{H}^2_\beta(0,T)$ *to equation* (42) *with data* (β,ξ,f), *for every* β *such that*

$$\beta \geq \bar{\beta}_t \quad \mathbb{P}\text{-a.s.},\ \forall\, t\in[0,T],$$

where $(\bar{\beta}_t)_{t\in[0,T]}$ *is a strictly positive predictable process depending only on* ε, ΔA, L_z *and* L_y.

The proof of Theorem 15 is based on Lemma 14, and is quite technical. Notice that in [**26**] the same condition (48) is imposed. As mentioned earlier, in that paper the authors study a class of BSDEs with a generator f integrated with respect to a deterministic (rather than predictable) right-continuous nondecreasing process A, and provide an existence and uniqueness result for this class of BSDEs. However, the proof in [**26**] relies heavily on the assumption that A is deterministic, and can not be extended to the case where A is predictable, which therefore requires a completely different procedure.

II.1. Motivation and future applications. The results in Theorem 15 could be employed to solve, by means of the BSDEs theory, optimal control problems of PDMPs on state spaces with boundary. We recall that the BSDEs approach to optimal control for PDMPs is implemented in Chapter 3 by means of the control randomization method. However, in that chapter only the case of PDMPs taking values in open state spaces is considered. Indeed in those cases the compensator $\nu(ds\,dy) = dA_s\,\phi_t(dy)$ of the random measure associated to the PDMP is quasi-left continuous, and a fairly complete theory was developed in the literature for BSDEs driven by such random measures. On the contrary, PDMP's jumps at the boundary of the domain correspond to predictable discontinuities for the process A. BSDEs driven by random measures of this type belong to the class of equations (42) mentioned before, for which, to our knowledge, Theorem 15 constitutes the only general well-posedness result at disposal in literature.

More precisely, consider a PDMP X on a general state space E with boundary ∂E. The jump dynamics of X in the interior of the domain is described by the transition probability measure $Q : E \times \mathcal{E} \to E$ and the jump rate measure $\lambda : E \to \mathbb{R}_+$ introduced in Chapter 3. In addition, a forced jump occurs every time the process reaches the active boundary $\Gamma \in \partial E$ (for the precise definition of Γ see page 61 in [**35**]). In this case, the process immediately jumps back to the interior of the domain accordingly to a transition probability measure $R : \partial E \times \mathcal{E} \to E$. The compensator of the integer-valued random measure associated to X then admits the form

$$\tilde{p}(ds\, dy) = \lambda(X_{s-})\, Q(X_{s-}, dy)\, ds + R(X_{s-}, dy)\, dp^*_s,$$

where

$$p^*_s = \sum_{n=1}^{\infty} 1_{\{s \geq T_n\}}\, 1_{\{X_{T_n-} \in \Gamma\}}$$

is the process counting the number of jumps of X from the active boundary $\Gamma \in \partial E$. In particular, the compensator can be rewritten as

$$\tilde{p}(ds\, dy) = dA_s\, \phi(X_{s-}, dy),$$

where $\phi(X_{s-}, dy) := Q(X_{s-}, dy)\, 1_{\{X_{s-} \in E\}} + R(X_{s-}, dy)\, 1_{\{X_{s-} \in \Gamma\}}$, and

$$A_s := \lambda(X_{s-})\, ds + dp^*_s$$

is a predictable and discontinuous process, with jumps $\Delta A_s = 1_{\{X_{s-} \in \Gamma\}}$.

In this context condition (48) in Theorem 15 reads

$$L_y \;<\; \frac{1}{\sqrt{2}}. \tag{49}$$

This is the only additional condition required in order to have a unique solution to a BSDE of the form (42) driven by the random measure associated to a PDMP. In particular, Theorem 15 does not impose any condition on L_z, i.e. on the Lipschitz constant of f with respect to its last argument. This is particularly important in the study of control problems related to PDMPs by means of BSDEs methods: in this case indeed $L_y = 0$ and condition (49) is automatically satisfied. This fact opens to the possibility of extending the control randomization method developed in Chapter 3 also in the case of optimal control of PDMPs with bounded domain. This will be the subject of a future work.

III. Weak Dirichlet processes and BSDEs driven by a random measure

III.1. State of the art. Stochastic calculus via regularization was essentially known in the case of continuous integrators X, see e.g. Russo and Vallois [**116**], [**117**]. A survey on basic elements of the calculus, can be found in Russo and Vallois [**121**]; it applies mainly in the case when X is not a semimartigale. In the framework of calculus via regularizations, a complete theory has been developed. In particular stochastic differential equations were studied, Itô formulae for processes with finite quadratic (and more general) variations were provided. In Flandoli and Russo [**63**] were given Itô-Wentzell type formulae, and generalizations to the case of Banach space type integrators are considered for instance in Di Girolami and Russo [**44**]. The notion of covariation $[X, Y]$ (resp. quadratic variation $[X, X]$) for two processes X, Y (resp. a process X) has been introduced in the framework of regularizations (see Russo and Vallois [**119**]) and of discretization as well (see Föllmer [**66**]). For instance, if X is a finite quadratic variation continuous process, an Itô formula has been proved for the expansion of $F(X_t)$, when $F \in C^2$, see [**119**]. When F is of class C^1 and X a reversible semimartingale, an Itô expansion was established in Russo and Vallois [**120**]. An important notion in calculus via regularizations is the one of *Dirichlet process* (with respect to a given filtration $(\mathcal{F}_t)$). The notion of Dirichlet process is a generalization of the concept of semimartingale, and was introduced by [**66**] and Bertoin [**14**] in the discretization framework. The analogue of the Doob-Meyer decomposition for a Dirichlet process is that it is the sum of a local martingale M and an adapted process A with zero quadratic variation. Here A is the generalization of a bounded variation process. The concept of $(\mathcal{F}_t)$-weak Dirichlet process (or simply weak Dirichlet process) was later introduced in Errami and Russo [**58**] and Gozzi and Russo [**71**] and applications to stochastic control were considered in Gozzi and Russo [**70**]. Such a process is defined as the sum of a local martingale M and an adapted $(\mathcal{F}_t)$-orthogonal process A, in the sense that $[A, N] = 0$ for every continuous local martingale N. An $(\mathcal{F}_t)$-weak Dirichlet process constitutes a natural generalization of the notion of the one of $(\mathcal{F}_t)$-Dirichlet process. An useful chain rule was established for $F(t, X_t)$ when F belongs to class $C^{0,1}$ and X is a weak Dirichlet process (with finite quadratic variation), see [**71**]. Such a process is indeed again a weak Dirichlet process (with possibly no finite quadratic variation).

As far as calculus via regularizations when X is a càdlàg integrator process only a few steps were done: we refer in particular to [**119**], Russo and Vallois [**118**], and the book of Di Nunno, Øksendal and Proske [**45**], see Chapter 15 and references therein. For instance no Itô type formulae have been established and in the discretization framework only few chain rule results are available for $F(X)$, when $F(X)$ is not a semimartingale. In that direction two peculiar results are available: the expansion of $F(X_t)$ when X is a reversible semimartingale and F is of class C^1 with some Hölder conditions on the derivatives (see Errami, Russo and Vallois [**59**]) and a chain rule for $F(X_t)$ when X is a weak Dirichlet (càdlàg) process and F is of class C^1, see Coquet, Jakubowsky, Mémin and Slomińsky [**30**]. The work in [**59**] has been continued by several authors, see e.g. Eisenbaum [**47**] and references therein, expanding the remainder making use of local time type processes.

In fact, the notion of $(\mathcal{F}_t)$-Dirichlet process does not fit to the framework of calculus with respect to jump processes. Indeed, requiring a process A to be of zero quadratic variation imposes that A is continuous. On the other hand, a bounded variation process with jumps has a non zero finite quadratic variation, so the generalization of the semimartingale is not necessarily represented by the notion of Dirichlet process. The property of weak Dirichlet process turns out to be a correct generalization of the one of semimartingale in the discontinuous framework. This concept was extended to the case of jumps processes in the significant work [**30**], by using the discretizations techniques.

III.2. Stochastic calculus via regularization and weak Dirichlet processes with jumps. In **Chapter 5** we extend, in a systematic way, stochastic calculus via regularizations to the case of jump processes, and we carry on the investigations of the so called weak Dirichlet processes in the discontinuous case.

The first basic objective consists in developing a calculus via regularization in the case of finite quadratic variation càdlàg processes. To this end, we revisit the definitions given by [**119**] concerning forward integrals (resp. covariations). Let X and Y be two càdlàg processes. The stochastic integral $\int_0^{\cdot} Y_s \, d^- X_s$ and the covariation $[Y, X]$ are defined as the uniform convergence in probability (u.c.p.) limit of the expressions

$$I^{-ucp}(\varepsilon, t, Y, dX) = \int_{(0,\,t]} Y(s) \, \frac{X((s+\varepsilon)\wedge t) - X(s)}{\varepsilon} \, ds, \tag{50}$$

$$[Y, X]^{ucp}_{\varepsilon}(t) = \int_{(0,\,t]} \frac{(Y((s+\varepsilon)\wedge t) - Y(s))(X((s+\varepsilon)\wedge t) - X(s))}{\varepsilon} \, ds. \tag{51}$$

That convergence ensures that the limiting objects are càdlàg, since the approximating expressions have the same property. For instance a càdlàg process X will be called *finite quadratic variation process* whenever the limit (which will be denoted by $[X, X]$) of

$$[X, X]^{ucp}_{\varepsilon}(t) := \int_{(0,\,t]} \frac{(X((s+\varepsilon)\wedge t) - X(s))^2}{\varepsilon} \, ds \tag{52}$$

exists u.c.p. In [**119**], the authors introduced a slightly different approximation of $[X, X]$ when X is continuous, namely

$$C_{\varepsilon}(X, X)(t) := \int_{(0,\,t]} \frac{(X((s+\varepsilon) - X(s))^2}{\varepsilon} \, ds. \tag{53}$$

When the u.c.p. limit of $C_{\varepsilon}(X, X)$ exists, it is automatically a continuous process, since the approximating processes are continuous. For this reason, when X is a jump process, the choice of approximation (53) would not be suitable, since its quadratic variation is expected to be a jump process. In that case, the u.c.p. convergence of (52) can be shown to be equivalent with a notion of convergence which is associated with the a.s. convergence (up to subsequences) in measure of $C_{\varepsilon}(X, X)(t) \, dt$. Both formulations will be used in the development of the calculus.

For a càdlàg finite quadratic variation process X, we establish, via regularization techniques, an Itô formula for $C^{1,2}$ functions of X of the following form.

Proposition 16. *Let X be a finite quadratic variation càdlàg process and $F : [0,\, T] \times \mathbb{R} \to \mathbb{R}$ a function of class $C^{1,2}$. Then*

$$F(t, X_t) = F(0, X_0) + \int_0^t \partial_s F(s, X_s) \, ds + \int_0^t \partial_x F(s, X_s) \, d^- X_s$$

$$+\frac{1}{2}\int_0^t \partial^2_{xx}F(s, X_{s-})\, d[X, X]^c_s$$
$$+\sum_{s\leq t}[F(s, X_s) - F(s, X_{s-}) - \partial_x F(s, X_{s-})\, \Delta\, X_s]. \tag{54}$$

From Proposition 16 will easily follow an Itô formula under weaker regularity conditions on F. Notice that a similar formula was stated in [**59**], using a discretization definition of the covariation, when F is time-homogeneous.

Proposition 17. *Let $F : [0, T] \times \mathbb{R} \to \mathbb{R}$ be a function of class C^1 such that $\partial_x F$ is Hölder continuous with respect to the second variable for some $\lambda \in [0, 1)$. Let (X_t) be a reversible semimartingale, satisfying*

$$\sum_{0<s\leq t} |\Delta X_s|^{1+\lambda} < \infty \quad \text{a.s.}$$

Then

$$F(t, X_t) = F(0, X_0) + \int_0^t \partial_s F(s, X_s)\, ds + \int_0^t \partial_x F(s, X_{s-})\, dX_s + \frac{1}{2}\, [\partial_x F(\cdot, X),\, X]_t$$
$$+ J(F, X)(t),$$

where

$$J(F, X)(t) = \sum_{0<s\leq t}\left[F(s, X_s) - F(s, X_{s-}) - \frac{\partial_x F(s, X_s) + \partial_x F(s, X_{s-})}{2}\, \Delta X_s\right].$$

The proof of Proposition 16 is based on an accurate separation between the neighborhood of "big" and "small" jumps, where specific tools are used. To this end, a fundamental role is played by the two following lemmata, the second one based on Lemma 1, Chapter 3, in Billingsley [**16**].

Lemma 18. *Let Y_t be a càdlàg function with values in $\mathbb{R}^n$. Let $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a uniformly continuos function on each compact, such that $\phi(y, y) = 0$ for every $y \in \mathbb{R}^n$. Let $0 \leq t_1 \leq t_2 \leq ... \leq t_N \leq T$. We have*

$$\sum_{i=1}^{N} \frac{1}{\varepsilon}\int_{t_i-\varepsilon}^{t_i} 1_{]0,\, s]}(t)\, \phi(Y_{(t+\varepsilon)\wedge s}, Y_t)\, dt \overset{\varepsilon\to 0}{\longrightarrow} \sum_{i=1}^{N} 1_{]0,\, s]}(t_i)\, \phi(Y_{t_i}, Y_{t_i-}), \tag{55}$$

uniformly in $s \in [0, T]$.

Lemma 19. *Let X be a càdlàg (càglàd) real process. Let $\gamma > 0$, $t_0, t_1 \in \mathbb{R}$ and $I = [t_0,\, t_1]$ be a subinterval of $[0, T]$ such that*

$$|\Delta X_t|^2 \leq \gamma^2, \quad \forall t \in I. \tag{56}$$

Then there is $\varepsilon_0 > 0$ such that

$$\sup_{\substack{a,\, t\in I\\ |a-t|\leq\varepsilon_0}} |X_a - X_t| \leq 3\gamma.$$

Another significant tool for our scopes is a Lemma of Dini type in the case of càdlàg functions, which reads as follows.

Lemma 20. *Let* $(G_n,\, n \in \mathbb{N})$ *be a sequence of continuous increasing functions, let* G *(resp.* F*) from* $[0, T]$ *to* $\mathbb{R}$ *be a càdlàg (resp. continuous) function. We set* $F_n = G_n + G$ *and suppose that* $F_n \to F$ *pointwise. Then*

$$\limsup_{n\to\infty} \sup_{s\in[0,\,T]} |F_n(s) - F(s)| \le 2 \sup_{s\in[0,\,T]} |G(s)|.$$

The second target of the chapter consists in investigating weak Dirichlet jump processes. Contrarily to the continuous case, the decomposition $X = M + A$ is generally not unique. We introduce the notion of a *special weak Dirichlet process* with respect to some filtration $(\mathcal{F}_t)$. Such a process is a weak Dirichlet process admitting a decomposition

$$X = M^c + M^d + A, \tag{57}$$

where M^c is a continuous local martingale, M^d is a purely discontinuous local martingale, and A is an $(\mathcal{F}_t)$-orthogonal, predictable process. Supposing that $A_0 = M^d_0 = 0$, the decomposition (57) is unique. In that case the decomposition (57) will be called the *canonical decomposition* of X. We remark that a continuous weak Dirichlet process is special weak Dirichlet.

In the sequel we will denote by μ^X the jump measure associated to X, and by ν^X its compensator. We will also indicate by $\mathbb{D}^{ucp}$ the set of all adapted càdlàg processes equipped with the topology of the uniform convergence in probability (u.c.p.), by $\mathcal{A}$ (resp $\mathcal{A}_{\mathrm{loc}}$) the collection of all adapted processes with integrable variation (resp. with locally integrable variation), and by $\mathcal{A}^+$ (resp $\mathcal{A}^+_{\mathrm{loc}}$) the collection of all adapted integrable increasing (resp. adapted locally integrable) processes.

We start by giving an expansion of $F(t, X_t)$ where F is of class $C^{0,1}$ and X is a càdlàg weak Dirichlet process of finite quadratic variation. The process $(F(t, X_t))$ turns out to be again a weak Dirichlet process, however not necessarily of finite quadratic variation.

Theorem 21. *Let* $X = M + A$ *be a càdlàg weak Dirichlet process of finite quadratic variation. Then, for every* $F : [0, T] \times \mathbb{R} \to \mathbb{R}$ *of class* $C^{0,1}$*, we have*

$$\begin{aligned} F(t, X_t) &= F(0, X_0) + \int_0^t \partial_x F(s, X_{s-})\, dM_s \\ &+ \int_{(0,\,t]\times\mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}))\, 1_{\{|x|\le 1\}}\, (\mu^X - \nu^X)(ds\, dx) \\ &- \int_{(0,\,t]\times\mathbb{R}} x\, \partial_x F(s, X_{s-})\, 1_{\{|x|\le 1\}}\, (\mu^X - \nu^X)(ds\, dx) \\ &+ \int_{(0,\,t]\times\mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}) - x\, \partial_x F(s, X_{s-}))\, 1_{\{|x|>1\}}\, \mu^X(ds\, dx) + \Gamma^F(t), \end{aligned} \tag{58}$$

where $\Gamma^F : C^{0,1} \to \mathbb{D}^{ucp}$ *is a continuous linear map, such that, for every* $F \in C^{0,1}$*, it fulfills the following properties.*

(a) $[\Gamma^F, N] = 0$ *for every* N *continuous local martingale.*

(b) *If* A *is predictable, then* Γ^F *is predictable.*

Starting from Theorem 21, we are able to provide an analogous chain rule when X and $(F(t, X_t))$ are both special weak Dirichlet processes. This constitutes our main result. We make use of the following conditions.

$$\int_{(0,\cdot]\times\mathbb{R}} |F(t, X_{t-} + x) - F(t, X_{t-}) - x\,\partial_x F(t, X_{t-})|\,1_{\{|x|>1\}}\,\mu^X(dt\,dx) \in \mathcal{A}^+_{\text{loc}}, \quad (59)$$

$$\int_{(0,\cdot]\times\mathbb{R}} |x|\,1_{\{|x|>1\}}\,\mu^X(dt\,dx) \ \in \mathcal{A}^+_{\text{loc}}. \quad (60)$$

Theorem 22. *Let X be a special weak Dirichlet process of finite quadratic variation with its canonical decomposition $X = M^c + M^d + A$. Assume that condition* (59) *holds. Then, for every $F : [0, T] \times \mathbb{R} \to \mathbb{R}$ of class $C^{0,1}$, we have*

(1) *$Y_t = F(t, X_t)$ is a special weak Dirichlet process, with decomposition $Y = M^F + A^F$, where*

$$M^F_t = F(0, X_0) + \int_0^t \partial_x F(s, X_s)\,d(M^c + M^d)_s$$
$$+ \int_{(0,\,t]\times\mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}) - x\,\partial_x F(s, X_{s-}))\,(\mu^X - \nu^X)(ds\,dx),$$

and $A^F : C^{0,1} \to \mathbb{D}^{ucp}$ is a linear map such that, for every $F \in C^{0,1}$, A^F is a predictable $(\mathcal{F}_t)$-orthogonal process.

(2) *If in addition condition* (60) *holds, M^F reduces to*

$$M^F_t = F(0, X_0) + \int_0^t \partial_x F(s, X_s)\,dM^c_s$$
$$+ \int_{(0,\,t]\times\mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}))\,(\mu^X - \nu^X)(ds\,dx).$$

We remark that a first important step in this sense was done in [**30**], where X belongs to a bit different class of special weak Dirichlet jump processes (of finite energy) and F does not depend on time and has bounded derivative. In [**30**] the authors show that $F(X)$ is again a special weak Dirichlet process. There the underlying process has finite energy, which requires a control of the expectation of the approximating sequences of the quadratic variation. On the other hand, our techniques do not require that type of control. Moreover, the integrability condition (59) that we ask on $F(t, X_t)$ in order to get the chain rule in Theorem 22 is automatically verified under the hypothesis on the first-order derivative considered in [**30**].

In some circumstances a chain rule may hold even when F is only continuous if we know a priori some information of $(F(t, X_t))$. No assumption are required in this case on the càdlàg process X.

Proposition 23. *Let X be an adapted càdlàg process. Let $F : [0, T] \times \mathbb{R} \to \mathbb{R}$ be a continuous function such that the following holds.*

(i) *$F(t, X_t) = B_t + A'_t$, where B has bounded variation and A' is a continuous $(\mathcal{F}_t)$-orthogonal process;*

(ii) $\int_{(0,\,\cdot]\times\mathbb{R}} |F(s, X_{s-} + x) - F(s, X_{s-})|\,\mu^X(ds\,dx) \ \in \mathcal{A}^+_{\text{loc}}$.

Then $F(t, X_t)$ is a special weak Dirichlet process with decomposition

$$F(t, X_t) = F(0, X_0) + \int_{(0,\,t]\times\mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}))\,(\mu^X - \nu^X)(ds\,dx) + A^F(t), \tag{61}$$

where A^F is a predictable $(\mathcal{F}_t)$-orthogonal process.

Finally, we also introduce a subclass of weak Dirichlet processes, called *particular*. A particular weak Dirichlet process X admits a decomposition $X = M + A$, where M is an $(\mathcal{F}_t)$-local martingale, and $A = V + A'$, with V a bounded variation adapted process and A' a continuous adapted process $(\mathcal{F}_t)$-orthogonal process such that $A'_0 = 0$. Those processes inherit some of the semimartingales features: as in the semimartingale case, the particular weak Dirichlet processes admit a (unique) canonical decomposition when $\int_{(0,\,\cdot]\times\mathbb{R}} |x|\,1_{\{|x|>1\}}\,\mu(dt\,dx) \in \mathcal{A}^+_{\mathrm{loc}}$. and an integral representation holds. Under that condition, those particular processes are indeed special weak Dirichlet processes.

III.3. Application to BSDEs driven by a random measure. In **Chapter 6** we apply the stochastic calculus developed in Chapter 5, and we provide an identification result for the solution of a forward backward stochastic differential equation driven by a random measure, when the underlying process X is of weak Dirichlet type. Indeed, given a solution (Y, Z, U) to this forward BSDE, often Y appears to be of the type $u(t, X_t)$ where u is a deterministic function; by using the stochastic calculus with respect to weak Dirichlet processes, we are able to identify also Z and U in terms of u.

More precisely, fix a finite time horizon $T > 0$ and let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ be a given filtered probability space, where $(\mathcal{F}_t)_{t\geq 0}$ satisfies the usual conditions. We will focus on general BSDEs of the type

$$\begin{aligned} Y_t = \xi + \int_{(t,\,T]} \tilde{g}(s,\, Y_{s-},\, Z_s)\,d\zeta_s + \int_{(t,\,T]\times\mathbb{R}} \tilde{f}(s,\, e,\, Y_{s-},\, U_s(e))\,\lambda(ds\,de) \\ - \int_{(t,\,T]} Z_s\,dM_s - \int_{(t,\,T]\times\mathbb{R}} U_s(e)\,(\mu - \nu)(ds\,de). \end{aligned} \tag{62}$$

Here μ is a random measure on $[0,\,T] \times \mathbb{R}$ with compensator ν. Besides μ and ν appear three driving random elements: a continuous martingale M, a non-decreasing adapted continuous process ζ, and a predictable random measure λ on $[0, T] \times \mathbb{R}$, equipped with the usual product σ-fields. The other data of equation (62) are a square integrable random variable ξ, and two measurable functions $\tilde{g} : \Omega \times [0, T] \times \mathbb{R}^2 \to \mathbb{R}$, $\tilde{f} : \Omega \times [0, T] \times \mathbb{R}^3 \to \mathbb{R}$.

The Brownian context of Pardoux-Peng [**99**] appears as a particular case, setting $\mu = \lambda = 0$, $\zeta_s \equiv s$. There M is a standard Brownian motion and ξ is measurable with respect to the Brownian σ-field at terminal time. In that case the unknown can be reduced to (Y, Z), since U can be arbitrarily chosen. Another important subcase of (62) arises when only the purely discontinuous driving term appears, i.e. M and ζ vanish. A significant example is represented by BSDEs driven by the random measure associated to a pure jump process, as in Chapter 2, or to a piecewise deterministic Markov process, as in Chapter 3.

When the random dependence of $\tilde{f}$ and $\tilde{g}$ is provided by a Markov solution X of a forward SDE, and ξ is a real function of X at the terminal time T, equation (62) becomes a forward BSDE. As we have recalled in Section I, this generally constitutes a stochastic representation of a partial integro-differential equation (PIDE). Indeed, solutions of forward BSDEs generate solutions of PIDEs in the viscosity sense. More precisely, for each given couple $(t, x) \in [0, T] \times \mathbb{R}$, consider an underlying process X given by the solution $X^{t,x}$ of an SDE starting at x at time t. Let $(Y^{t,x}, Z^{t,x}, U^{t,x})$ be a family of solutions of the forward BSDE. In that case, under reasonable general assumptions, the function $v(t, x) := Y_t^{t,x}$ is a viscosity solution of the related PIDE. A demanding task consists in characterizing the pair $(Z, U) := (Z^{t,x}, U^{t,x})$, in term of v; this is generally called the *identification problem* of (Z, U). In the continuous case, this was for instance the object of Fuhrman and Tessitore [**68**]: the authors show that if $v \in C^{0,1}$, then $Z_s = \partial_x v(s, X_s)$; under more general assumptions, they also associate Z with a generalized gradient of v. At our knowledge, in the discontinuous case, the problem of the identification of the martingale integrands pair (Z, U) has not been deeply investigated, except for particular situations, as for instance the purely discontinuous case treated in Confortola and Fuhrman [**28**].

In Chapter 6 we discuss the mentioned identification problem in a quite general framework by means of the calculus related to weak Dirichlet processes. When Y is a deterministic function v of a special semimartingale X (or more generally a special weak Dirichlet process with finite quadratic variation), related in a specific way to the random measure μ, we apply the chain rule in Theorem 22 in order to identify the pair (Z, U).

We fix an integer-valued random measure μ on $[0, T] \times \mathbb{R}$, with compensator ν. We suppose, without loss of generality, that ν satisfies $\nu(\{t\} \times dx) \leq 1$ identically. We set

$$\begin{aligned} D &= \{(\omega, t) : \mu(\omega, \{t\} \times \mathbb{R}) > 0\}, \\ J &= \{(\omega, t) : \nu(\omega, \{t\} \times \mathbb{R}) > 0\}, \\ K &= \{(\omega, t) : \nu(\omega, \{t\} \times \mathbb{R}) = 1\}. \end{aligned}$$

We will ask the following condition on μ.

Hypothesis 24.

(i) $D = K \cup (\cup_n [[T_n^i]])$ up to an evanescent set, where $(T_n^i)_n$ are totally inaccessible times such that $[[T_n^i]] \cap [[T_m^i]] = \emptyset$, $n \neq m$.

(ii) For every predictable time S such that $[[S]] \subset K$, $\nu(\{S\}, de) = \mu(\{S\}, de)$ a.s.

With respect to a generic process X, we will consider the following assumption in relation to μ.

Hypothesis 25. $X = X^i + X^p$, where X^p is a càdlàg predictable process satisfying $\{\Delta X^p \neq 0\} \subset J$, and X^i is a càdlàg quasi-left continuous adapted process satisfying

$\{\Delta X^i \neq 0\} \subset D$. Moreover, there exists a predictable measurable map $\tilde{\gamma} : \Omega \times]0,\, T] \times \mathbb{R} \to \mathbb{R}$ such that

$$\Delta X^i_t(\omega)\, 1_{]0,\,T]}(t) = \tilde{\gamma}(\omega, t, \cdot) \quad d\mathbb{P}\, \mu(ds, de)\text{-a.e.} \tag{63}$$

The hypothesis below will concern a pair of processes (X, Y).

Hypothesis 26. X is a special weak Dirichlet process of finite quadratic variation, satisfying condition (60). $Y_t = v(t,\, X_t)$ for some (deterministic) function $v : [0, T] \times \mathbb{R} \to \mathbb{R}$ of class $C^{0,1}$ such that $F = v$ and X verify condition (59).

We have the following result.

Proposition 27. *Let μ satisfy Hypothesis 24. Let X be a process verifying Hypothesis 25 with decomposition $X = X^i + X^p$, where $\tilde{\gamma}$ is the predictable process which relates μ and X^i in agreement with* (63). *Let (Y, Z, U) be a solution to the BSDE* (62) *such that the pair (X, Y) satisfies Hypothesis 26 with corresponding function v. Let X^c denote the continuous local martingale M^c of X given in the canonical decomposition* (57). *If $U_s - (v(s, X_{s-} + \tilde{\gamma}(s, \cdot)) - v(s, X_{s-})) \in \mathcal{G}^2_{\mathrm{loc}}(\mu)$, then the pair (Z, U) fulfills*

$$Z_s = \partial_x v(s, X_s)\, \frac{d\langle X^c, M\rangle_s}{d\langle M\rangle_s} \quad d\mathbb{P}\, d\langle M\rangle_s\text{-a.e.}, \tag{64}$$

$$U_s - (v(s, X_{s-} + \tilde{\gamma}(s, \cdot)) - v(s, X_{s-})) = l_s\, 1_K(s) \quad d\mathbb{P}\, \nu(ds\, de)\text{-a.e.}, \tag{65}$$

where l is a predictable process.

In the purely discontinuous framework, i.e. when in the BSDE (62) M and ζ vanish, we make use of the chain rule (61) in Proposition 23, which allows, for a general càdlàg process X, to express $v(t, X_t)$ without requiring any differentiability on v. In particular Proposition 23 does not ask X to be a special weak Dirichlet process, provided we have some a priori information on the structure of $v(t, X_t)$. We need the following condition on a pair of processes (X, Y).

Hypothesis 28.

(i) $Y = B + A'$, with B a bounded variation process and A' a continuous $(\mathcal{F}_t)$-orthogonal process;

(ii) $Y_t = v(t,\, X_t)$ for some continuous deterministic function $v : [0, T] \times \mathbb{R} \to \mathbb{R}$, satisfying the integrability condition

$$\int_{(0,\,\cdot]\times\mathbb{R}} |v(t, X_{t-} + x) - v(t, X_{t-})|\, \mu^X(dt\, dx) \ \in \mathcal{A}^+_{\mathrm{loc}}. \tag{66}$$

The identification in that case reads as follows.

Proposition 29. *Let μ satisfy Hypothesis 24. Let X verify Hypothesis 25 with decomposition $X = X^i + X^p$, where $\tilde{\gamma}$ is the predictable process which relates μ and X^i in agreement with* (63). *Let (Y, U) be a solution to the BSDE* (62) *with $M = 0$ and $\zeta = 0$, such that (X, Y) satisfies Hypothesis 28 with corresponding function v.*

If in addition $U_s - (v(s, X_{s-} + \tilde{\gamma}(s, \cdot)) - v(s, X_{s-})) \in \mathcal{G}^2_{\mathrm{loc}}(\mu)$, then there exists a predictable process l_s such that

$$U_s - (v(s, X_{s-} + \tilde{\gamma}(s, \cdot)) - v(s, X_{s-})) = l_s\, 1_K(s) \quad d\mathbb{P}\, \nu(ds\, de)\text{-a.e.} \tag{67}$$

We remark that in most of the literature on BSDEs, the measures ν, λ and ζ of equation (62) are non-atomic in time. As we have underlined in Chapter 4, a challenging case arises when one or more of those predictable processes have jumps in time. Our approach to the identification problem also applies to forward BSDEs presenting predictable jumps. As an example, we provide an identification result for a BSDE driven by the random measure μ associated to a PDMP taking values in a bounded real interval.

Further remarks and future developments of the thesis

We take the occasion to emphasize that every proof reported in the thesis is new; on the other hand, when a known result is needed, we give references to where a proof can be found. We also underline that **Chapter 1** is based on Bandini and Confortola [**4**], **Chapter 2** on Bandini and Fuhrman [**7**], **Chapter 3** on Bandini [**3**], **Chapter 4** on Bandini [**2**], **Chapters 5** and **6** respectively on Bandini and Russo [**9**] and [**8**].

Some challenging issues arising in this work are left for future research. First of all, as recalled in Section II.1, our existence and uniqueness result for BSDEs driven by general, possibly non quasi-left continuous, random measures opens to the possibility of studying optimal control problems for PDMPs with bounded state spaces by means of BSDEs techniques. This could allow to provide nonlinear Feynman-Kac representation formulae for the value functions of those control problems. In particular, combining ideas from Chapters 3 and 4, it might turn out that the value function of the optimal control problem of a PDMP with a bounded state space solves a backward stochastic differential equation with constrained jumps. Notice that it would be interesting to apply to this context the identification results obtained in Chapter 6, which are already conceived for BSDEs driven by random measures with possible predictable jumps. Another challenging development might consist in extending the results obtained in Chapter 2 to a non-Markovian pure jump framework. Optimal control problems for non-Markovian stochastic differential equations driven by a Brownian motion have been recently studied with the BSDEs techniques by means of the control randomization approach, see Fuhrman and Pham [**67**]. In this context the constrained BSDE characterizing the value function can be seen as a path-dependent version of the HJB equation. Notice that the control randomization method does not rely on the path-dependent HJB equation associated by dynamic programming principle to the value function in the non-Markovian context. This allows to circumvent delicate issues of dynamic programming (as originally studied in El Karoui [**49**] for general non-Markovian stochastic control problems), viscosity solutions and comparison principles for fully nonlinear path-dependent PDEs, as recently studied in Peng [**106**], Ekren, Keller, Touzi and Zhang [**48**] and Tang and Zhang [**127**], see also Fabbri, Gozzi and Swiech [**61**] for HJB equations in infinite dimension arising typically for stochastic systems with delays. This suggests in particular an original approach to derive the HJB equation for the value function of stochastic control problem from the BSDE representation, hence without dynamic programming principle. The generalization of these results to the jump case has not yet been investigated, and could be obtained by mixing the methodology in [**67**] with the specific theory for optimal control of pure jump processes developed in Chapter 2. Finally, we emphasize that the chain rule type expansions provided in Chapter 5 may be helpful to get verification theorems for stochastic optimal control problems of general jump processes. In the diffusive context, this was done in Gozzi and Russo [**70**] which treated optimal control problems of continuous processes without control in the diffusion. Those verification theorems have the advantage of requiring less regularity of the value function than the classical ones, which need instead C^1 regularity in time and C^2 in space (see e.g. Fleming and Soner [**65**]), and they can be applied also to problems with pathwise optimality and optimality in probability. It would be also judicious to generalize our results of Chapters 5 and 6 to the case of path-dependent càdlàg processes. In the case of path-dependent continuous processes, a first step for extending the chain rules of Chapter 5 was done in [**43**].

Chapter 1

Optimal control of semi-Markov processes with a BSDE approach

1.1. Introduction

In this chapter we study optimal control problems for a class of semi-Markov processes using a suitable class of backward stochastic differential equations, driven by the random measure associated to the semi-Markov process itself.

Let us briefly describe our framework. Our starting point is a semi-Markov pure jump process X on a general state space E. It is constructed starting from a jump rate function $\lambda(x, \vartheta)$ and a jump measure $A \mapsto Q(x, \vartheta, A)$ on E, depending on $x \in E$ and $\vartheta \geq 0$. Our approach is to consider a semi-Markov pure jump process as a two dimensional time-homogeneous and strong Markov process $\{(X_s, \theta_s),\ s \geq 0\}$ with its natural filtration $\mathcal{F}$ and a family of probabilities $\mathbb{P}^{x,\vartheta}$ for $x \in E$, $\vartheta \in [0, \infty)$ such that $\mathbb{P}^{x,\vartheta}(X_0 = x, \theta_0 = \vartheta) = 1$. If the process starts from (x, ϑ) at time $t = 0$ then the distribution of its first jump time T_1 under $\mathbb{P}^{x,\vartheta}$ is described by the formula

$$\mathbb{P}^{x,\vartheta}(T_1 > s) = \exp\left(-\int_{\vartheta}^{\vartheta+s} \lambda(x, r)\, dr\right), \tag{1.1}$$

and the conditional probability that the process is in A immediately after a jump at time $T_1 = s$ is

$$\mathbb{P}^{x,\vartheta}(X_{T_1} \in A \,|\, T_1 = s) = Q(x, s, A).$$

X_s is called the state of the process at time s, and θ_s is the duration period in this state up to moment s:

$$\theta_s = \begin{cases} \theta_0 + s & \text{if } X_p = X_s \quad \forall\, 0 \leqslant p \leqslant s,\ p, s \in \mathbb{R}, \\ s - \sup\{\, p :\ 0 \leqslant p \leqslant s,\ X_p \neq X_s\} & \text{otherwise.} \end{cases}$$

We note that X alone is not a Markov process. We limit ourselves to the case of a semi-Markov process X such that the survivor function of T_1 under $\mathbb{P}^{x,0}$ is absolutely continuous and admits a hazard rate function λ as in (1.1). The holding times of the process are not necessarily exponentially distributed and can be infinite with positive probability. Our main restriction is that the jump rate function λ is uniformly bounded, which implies that the process X is non explosive. Denoting by T_n the jump times of X, we consider the marked point process (T_n, X_{T_n}) and the associated random measure $p(dt\, dy) = \sum_n \delta_{(T_n, X_{T_n})}$ on $(0,\infty)\times E$, where δ denotes the Dirac measure. The dual predictable projection $\tilde{p}$ of p (shortly, the compensator) has the following explicit expression

$$\tilde{p}(ds\, dy) = \lambda(X_{s-},\theta_{s-})\, Q(X_{s-},\theta_{s-},dy)\, ds.$$

In Section 1.3 we address an optimal intensity-control problem for the semi-Markov process. This is formulated in a classical way by means of a change of probability measure, see e.g. El Karoui [**49**], Elliott [**57**] and Brémaud [**18**]. We define a class $\mathcal{A}$ of admissible control processes $(u_s)_{s\in[0,\,T]}$; for every fixed $t\in[0,\,T]$ and $(x,\vartheta)\in E\times[0,\infty)$, the cost to be minimized and the corresponding value function are

$$\begin{aligned} J(t,x,\vartheta,u(\cdot)) &= \mathbb{E}_{u,t}^{x,\vartheta}\left[\int_0^{T-t} l(t+s,X_s,\theta_s,u_s)\, ds + g(X_{T-t},\theta_{T-t})\right], \\ v(t,x,\vartheta) &= \inf_{u(\cdot)\in\mathcal{A}} J(t,x,\vartheta,u(\cdot)), \end{aligned}$$

where g, l are given real functions. Here $\mathbb{E}_{u,t}^{x,\vartheta}$ denotes the expectation with respect to another probability $\mathbb{P}_{u,t}^{x,\vartheta}$, depending on t and on the control process u and constructed in such a way that the compensator under $\mathbb{P}_{u,t}^{x,\vartheta}$ equals $r(t+s,X_{s-},\theta_{s-},y,u_s)$ $\lambda(X_{s-},\theta_{s-})\ Q(X_{s-},\theta_{s-},dy)\, ds$, for some function r given in advance as another datum of the control problem. Since the process $(X_s,\theta_s)_{s\geq 0}$ we want to control is time-homogeneous and starts from (x,ϑ) at time $s=0$, we introduce a temporal translation which allows to define the cost functional for all $t\in[0,T]$. For more details see Remark 1.3.2.

Our approach to this control problem consists in introducing a family of BSDEs parametrized by $(t,x,\vartheta)\in[0,T]\times E\times[0,\infty)$:

$$\begin{aligned} & Y_{s,t}^{x,\vartheta} + \int_s^{T-t}\int_E Z_{\sigma,t}^{x,\vartheta}(y)\, q(d\sigma\, dy) \\ & = g(X_{T-t},\theta_{T-t}) + \int_s^{T-t} f\Big(t+\sigma, X_\sigma,\theta_\sigma, Z_{\sigma,t}^{x,\vartheta}(\cdot)\Big)\, d\sigma, \end{aligned} \tag{1.2}$$

$s\in[0,\,T-t]$, where the generator is given by the Hamiltonian function f defined for every $s\in[0,\,T]$, $(x,\vartheta)\in E\times[0,\,+\infty)$, $z\in L^2(E,\mathcal{E},\lambda(x,\vartheta)\,Q(x,\vartheta,dy))$, as

$$f(s,x,\vartheta,z(\cdot)) = \inf_{u\in U}\Big\{\, l(s,x,\vartheta,u) + \int_E z(y)(r(s,x,\vartheta,y,u)-1)\lambda(x,\vartheta)\, Q(x,\vartheta,dy)\,\Big\}. \tag{1.3}$$

Under appropriate assumptions 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 (1.2).

Backward equations driven by random measures have been studied in many papers, within Tang and Li [**128**], Barles, Buckdahn and Pardoux [**10**], Royer [**114**], Kharroubi, Ma, Pham and Zhang [**87**], Xia [**131**], and more recently Becherer [**12**], Crépey and Matoussi [**33**], Kazi-Tani, Possamaï and Zhou [**84**], [**83**], Confortola and Fuhrman [**27**], [**28**]. In many of them, among which [**128**], [**10**], [**114**] and [**87**], the stochastic equations are driven by a Wiener process and a Poisson process. A more general result on BSDEs driven by random measures is given by [**131**], but in this case the generator f depends on the process Z in a specific way and this condition prevents a direct application to optimal control problems. In [**12**], [**33**], [**84**], [**83**], the authors deal with BSDEs with jumps with a random compensator more general than the compensator of a Poisson random measure; here are involved random compensators which are absolutely continuous with respect to a deterministic measure, that can be reduced to a Poisson measure by a Girsanov change of probability. Finally, in [**27**] have been recently studied BSDEs driven by a random measure related to a pure jump process, and in [**28**] the pure jump Markov case is considered.

Our backward equation (1.2) is driven by a random measure associated to a two dimensional Markov process (X, θ), and his compensator is a stochastic random measure with a non-dominated intensity as in [**28**]. Even if the associated process is not pure jump, the existence, uniqueness and continuous dependence on the data for the BSDE (1.2) can be deduced extending in a straightforward way the results in [**28**].

Concerning the optimal control of semi-Markov processes, the case of a finite number of states has been studied in Chitopekar [**24**], Howard [**74**], Jewell [**81**], Osaki [**95**], while the case of arbitrary state space is considered in Ross [**112**], Gihman and Skorohod [**69**], and Stone [**125**]. As in [**24**] and in [**125**], in our formulation we admit control actions that can depend not only on the state process but also on the length of time the process has remained in that state. The approach based on BSDEs is classical in the diffusive context and is also present in the literature in the case of BSDEs with jumps, see as instance Lim and Quenez [**92**]. However, it seems to us be pursued here for the first time in the case of the semi-Markov processes. It allows to treat in a unified way a large class of control problems, where the state space is general and the running and final cost are not necessarily bounded. We remark that, comparing with [**125**], the controlled processes we deal with have laws absolutely continuous with respect to a given, uncontrolled process; see also a more detailed comment in Remark 1.3.3 below. Moreover, in [**125**] optimal control problems for semi-Markov processes are studied in the case of infinite time horizon.

In Section 1.4 we solve a nonlinear variant of the Kolmogorov equation for the process (X, θ), with the BSDEs approach. The process (X, θ) is time-homogeneous and Markov, but is not a pure jump process. In particular it has the integro-differential infinitesimal generator

$$\tilde{\mathcal{L}}\psi(x,\vartheta) := \partial_\vartheta \psi(x,\vartheta) + \int_E [\psi(y,0) - \psi(x,\vartheta)]\, \lambda(x,\vartheta)\, Q(x,\vartheta,dy), \quad (x,\vartheta) \in E \times [0,\infty).$$

The additional differential term ∂_ϑ does not allow to study the associated nonlinear Kolmogorov equation proceeding as in the pure jump Markov processes framework (see [**28**]). On the other hand, the two dimensional Markov process $(X_s, \theta_s)_{s\geqslant 0}$ belongs to the larger class of piecewise-deterministic Markov processes (PDMPs) introduced by Davis in [**35**], and studied in the optimal control framework by several authors, within Davis and Farid [**36**], Vermes [**129**], Dempster [**40**], Lenhart and Yamada [**91**]. Moreover, we deal with a very specific PDMP: taking into account the particular structure of semi-Markov processes, we present a reformulation of the Kolmogorov equation which allows us to consider solutions in a classical sense. In particular, we notice that the second component of the process $(X_s,\, \theta_s)_{s\geqslant 0}$ is linear in s. This fact suggests to introduce the formal directional derivative operator

$$(Dv)(t,x,\vartheta) := \lim_{h\downarrow 0} \frac{v(t+h,x,\vartheta+h)-v(t,x,\vartheta)}{h},$$

and to consider the following nonlinear Kolmogorov equation

$$\begin{cases} Dv(t,x,\vartheta)+\mathcal{L}v(t,x,\vartheta)+f(t,x,\vartheta,v(t,x,\vartheta),v(t,\cdot,0)-v(t,x,\vartheta))=0, \\ \hfill t\in[0,T],\ x\in E,\ \vartheta\in[0,\infty), \\ v(T,x,\vartheta)=g(x,\vartheta), \end{cases} \tag{1.4}$$

where

$$\mathcal{L}\psi(x,\vartheta) := \int_E [\psi(y,0)-\psi(x,\vartheta)]\,\lambda(x,\vartheta)\,Q(x,\vartheta,dy), \qquad (x,\vartheta)\in E\times[0,\infty).$$

Then we look for a solution v such that the map $t\mapsto v(t,x,t+c)$ is absolutely continuous on $[0,T]$, for all constants $c\in[-T,+\infty)$. The functions f,g in (1.4) are given. While it is easy to prove well-posedness of (1.4) under boundedness assumptions, we achieve the purpose of finding a unique solution under much weaker conditions related to the distribution of the process (X,θ): see Theorem 1.4.7. To this end we need to define a formula of Itô type, involving the directional derivative operator D, for the composition of the process $(X_s,\, \theta_s)_{s\geqslant 0}$ with functions v smooth enough (see Lemma 1.4.2 below).

We construct the solution v by means of a family of BSDEs of the form (1.2). By the results above there exists a unique solution $(Y^{x,\vartheta}_{s,t}, Z^{x,\vartheta}_{s,t})_{s\in[0,T-t]}$ and the estimates on the BSDEs are used to prove well-posedness of (1.4). As a by-product we also obtain the representation formulae

$$v(t,x,\vartheta)=Y^{x,\vartheta}_{0,t}, \quad Y^{x,\vartheta}_{s,t}=v(t+s,X_s,\theta_s), \quad Z^{x,\vartheta}_{s,t}(y)=v(t+s,y,0)-v(t+s,X_{s-},\theta_{s-}),$$

which are sometimes called, at least in the diffusive case, non linear Feynman-Kac formulae.

Finally we can go back to the original control problem and observe that the associated Hamilton-Jacobi-Bellman equation has the form (1.4) where f is the Hamiltonian function (1.3). By previous results we are able to identify the HJB solution $v(t,x,\vartheta)$, constructed probabilistically via BSDEs, with the value function.

1.2. Notation, preliminaries and basic assumptions

1.2.1. Semi-Markov Jump Processes. We recall the definition of a semi-Markov process, as given, for instance, in [**69**]. More precisely we will deal with a semi-Markov process with infinite lifetime (i.e. non explosive). Suppose we are given a measurable space $(E, \mathcal{E})$, a set Ω and two functions $X : \Omega \times [0, \infty) \to E$, $\theta : \Omega \times [0, \infty) \to [0, \infty)$. For every $t \geq 0$, we denote by $\mathcal{F}_t$ the σ-algebra $\sigma((X_s, \theta_s), s \in [0, t])$. We suppose that for every $x \in E$ and $\vartheta \in [0, \infty)$, a probability $\mathbb{P}^{x,\vartheta}$ is given on $(\Omega, \mathcal{F}_{[0,\infty)})$ and the following conditions hold.

(1) $\mathcal{E}$ contains all one-point sets. Δ denotes a point not included in E.
(2) $\mathbb{P}^{x,\vartheta}(X_0 = x, \theta_0 = \vartheta) = 1$ for every $x \in E$, $\vartheta \in [0, \infty)$.
(3) For every s, $p \geqslant 0$ and $A \in \mathcal{E}$ the function $(x, \vartheta) \mapsto \mathbb{P}^{x,\vartheta}(X_s \in A, \theta_s \leqslant p)$ is $\mathcal{E} \otimes \mathcal{B}^+$-measurable.
(4) For every $0 \leq t \leq s$, $p \geqslant 0$, and $A \in \mathcal{E}$ we have $\mathbb{P}^{x,\vartheta}(X_s \in A, \theta_s \leqslant p \,|\, \mathcal{F}_t) = \mathbb{P}^{X_t,\theta_t}(X_s \in A, \theta_s \leqslant p)$, $\mathbb{P}^{x,\vartheta}$-a.s.
(5) All the trajectories of the process X have right limits when E is given the discrete topology (the one where all subsets are open). This is equivalent to require that for every $\omega \in \Omega$ and $t \geq 0$ there exists $\delta > 0$ such that $X_s(\omega) = X_t(\omega)$ for $s \in [t, t + \delta]$.
(6) All the trajectories of the process a are continuous from the right piecewise linear functions. For every $\omega \in \Omega$, if $[\alpha, \beta)$ is the interval of linearity of $\theta_\cdot(\omega)$ then $\theta_s(\omega) = \theta_\alpha(\omega) + s - \alpha$ and $X_\alpha(\omega) = X_s(\omega)$; if β is a discontinuity point of $\theta_\cdot(\omega)$ then $\theta_{\beta+}(\omega) = 0$ and $X_\beta(\omega) \neq X_{\beta-}(\omega)$.
(7) For every $\omega \in \Omega$ the number of jumps of the trajectory $t \mapsto X_t(\omega)$ is finite on every bounded interval.

X_s is called the *state* of the process at time s, θ_s is the *duration period* in this state up to moment s. Also we call X_s the *phase* and θ_s the *age* or the *time component* of a semi-Markov process. X is a non explosive process because of condition (7). We note, moreover, that the two-dimensional process (X, θ) is a strong Markov process with time-homogeneous transition probabilities because of conditions (2), (3), and (4). It has right-continuous sample paths because of conditions (1), (5) and (6), and it is not a pure jump Markov process, but only a PDMP.

The class of semi-Markov processes we consider in the chapter will be described by means of a special form of joint law R under $\mathbb{P}^{x,\vartheta}$ of the first jump time T_1, and the corresponding position X_{T_1}. To proceed formally, we fix $X_0 = x \in E$ and define the first jump time

$$T_1 = \inf\{p > 0 : \; X_p \neq x\},$$

with the convention that $T_1 = +\infty$ if the indicated set is empty.
We introduce $S := E \times [0, +\infty)$ an we denote by $\mathcal{S}$ the smallest σ-algebra containing all sets of $\mathcal{E} \otimes \mathcal{B}([0, +\infty))$. (Here and in the following $\mathcal{B}(\Lambda)$ denotes the Borel σ-algebra of a topological space Λ). Take an extra point $\Delta \notin E$ and define $X_\infty(\omega) = \Delta$ for all $\omega \in \Omega$, so that $X_{T_1} : \Omega \to E \cup \{\Delta\}$ is well defined. Then on the extended space $S \cup \{(\Delta, \infty)\}$ we consider the smallest σ-algebra, denoted by $\mathcal{S}^{\mathrm{enl}}$, containing $\{(\Delta, \infty)\}$ and all sets of $\mathcal{E} \otimes \mathcal{B}([0, +\infty))$. Then (X_{T_1}, T_1) is a random variable with values in $(S \cup \{(\Delta, \infty)\}, \mathcal{S}^{\mathrm{enl}})$. Its law under $\mathbb{P}^{x,\vartheta}$ will be denoted by $R(x, \vartheta, \cdot)$.

We will assume that R is constructed from two given functions denoted by λ and Q. More precisely we assume the following.

Hypothesis 1.2.1. There exist two functions

$$\lambda : S \to [0,\infty) \text{ and } Q : S \times \mathcal{E} \to [0,1]$$

such that

(i) $(x,\vartheta) \mapsto \lambda(x,\vartheta)$ is $\mathcal{S}$-measurable;
(ii) $\sup_{(x,\vartheta)\in S} \lambda(x,\vartheta) \leqslant C \in \mathbb{R}^+$;
(iii) $(x,\vartheta) \mapsto Q(x,\vartheta,A)$ is $\mathcal{S}$-measurable $\forall A \in \mathcal{E}$;
(iv) $A \mapsto Q(x,\vartheta,A)$ is a probability measure on $\mathcal{E}$ for all $(x,\,\vartheta) \in S$.

We define a function H on $E \times [0,\infty]$ by

$$H(x,s) := 1 - e^{-\int_0^s \lambda(x,r)dr}. \tag{1.5}$$

Given λ and Q, we will require that for the semi-Markov process X we have, for every $(x,\vartheta) \in S$ and for $A \in \mathcal{E}$, $0 \le c < d \le \infty$,

$$\begin{aligned} R(x,\vartheta,A\times(c,d)) &= \frac{1}{1-H(x,\vartheta)} \int_c^d Q(x,s,A) \frac{d}{d\,s} H(x,\vartheta+s)\,ds \\ &= \int_c^d Q(x,s,A)\, \lambda(x,\vartheta+s)\, \exp\left(-\int_\vartheta^{\vartheta+s} \lambda(x,r)\,dr\right) ds, \end{aligned} \tag{1.6}$$

where R was described above as the law of (X_{T_1},T_1) under $\mathbb{P}^{x,\vartheta}$. The existence of a semi-Markov process satisfying (1.6) is a well known fact, see for instance [**125**] Theorem 2.1, where it is proved that X is in addition a strong Markov process. The nonexplosive character of X is made possible by Hypothesis 1.2.1-(ii).

We note that our data only consist initially in a measurable space $(E,\mathcal{E})$ ($\mathcal{E}$ contains all singleton subsets of E), and in two functions λ, Q satisfying Hypothesis 1.2.1. The semi-Markov process X can be constructed in an arbitrary way provided (1.6) holds.

Remark 1.2.2.

(1) Note that (1.6) completely specifies the probability measure $R(x,\vartheta,\cdot)$ on $(S \cup \{(\Delta,\,\infty)\},\mathcal{S}^{\text{enl}})$: indeed simple computations show that, for $s \ge 0$,

$$\begin{aligned} \mathbb{P}^{x,\vartheta}(T_1 \in (s,\infty]) &= 1 - R(x,\vartheta,E\times(0,s]) \\ &= \exp\left(-\int_\vartheta^{\vartheta+s} \lambda(x,r)\,dr\right), \end{aligned} \tag{1.7}$$

and we clearly have

$$\mathbb{P}^{x,\vartheta}(T_1 = \infty) \quad = \quad R(x,\vartheta,\{(\Delta,\infty)\}) = \exp\left(-\textstyle\int_\vartheta^\infty \lambda(x,r)\,dr\right).$$

Moreover, the kernel R is well defined, because $H(x,\vartheta) < 1$ for all $(x,\vartheta) \in S$ by Hypothesis 1.2.1-(ii).

(2) The data λ and Q have themselves a probabilistic interpretation. In fact if in (1.7) we set $\vartheta = 0$ we obtain

$$\mathbb{P}^{x,0}(T_1 > s) = \exp\left(-\int_0^s \lambda(x,r)\,dr\right) = 1 - H(x,s). \tag{1.8}$$

This means that under $\mathbb{P}^{x,0}$ the law of T_1 is described by the distribution function H, and

$$\lambda(x,\vartheta) = \frac{\frac{\partial H}{\partial \vartheta}(x,\vartheta)}{1 - H(x,\vartheta)}.$$

Then $\lambda(x,\vartheta)$ is the jump rate of the process X given that it has been in state x for a time ϑ.
Moreover, the probability $Q(x,s,\cdot)$ can be interpreted as the conditional probability that X_{T_1} is in $A \in \mathcal{E}$ given that $T_1 = s$; more precisely,

$$\mathbb{P}^{x,\vartheta}(X_{T_1} \in A, T_1 < \infty \,|\, T_1) = Q(x, T_1, A)\, 1_{T_1 < \infty}, \qquad \mathbb{P}^{x,\vartheta} - a.s.$$

(3) In [**69**] the following observation is made: starting from $T_0 = t$ define inductively $T_{n+1} = \inf\{s > T_n \,:\, X_s \neq X_{T_n}\}$, with the convention that $T_{n+1} = \infty$ if the indicated set is empty; then, under the probability $\mathbb{P}^{x,\vartheta}$, the sequence of the successive states of the semi-Markov X is a Markov chain, as in the case of Markov processes. However, while for the latter the duration period in the state depends only on this state and it is necessarily exponentially distributed, in the case of a semi Markov process the duration period depends also on the state into which the process moves and the distribution of the duration period may be arbitrary.

(4) In [**69**] is also proved that the sequence $(X_{T_n}, T_n)_{n\geq 0}$ is a discrete-time Markov process in $(S \cup \{(\Delta, \infty)\}, \mathcal{S}^{\mathrm{enl}})$ with transition kernel R, provided we extend the definition of R making the state (Δ, ∞) absorbing, i.e. we define

$$R(\Delta, \infty, S) = 0, \qquad R(\Delta, \infty, \{(\Delta, \infty)\}) = 1.$$

Note that $(X_{T_n}, T_n)_{n\geq 0}$ is time-homogeneous.

This fact allows for a simple description of the process X. Suppose one starts with a discrete-time Markov process $(\tau_n, \xi_n)_{n\geq 0}$ in S with transition probability kernel R and a given starting point $(x,\vartheta) \in S$ (conceptually, trajectories of such a process are easy to simulate). One can then define a process Y in E setting $Y_t = \sum_{n=0}^N \xi_n 1_{[\tau_n, \tau_{n+1})}(t)$, where $N = \sup\{n \geq 0 : \tau_n \leqslant \infty\}$. Then Y has the same law as the process X under $\mathbb{P}^{x,\vartheta}$.

(5) We stress that (1.5) limits ourselves to deal with a class of semi-Markov processes for which the survivor function T_1 under $\mathbb{P}^{x,0}$ admits a hazard rate function λ.

$\square$

1.2.2. BSDEs driven by a Semi-Markov Process. Let be given a measurable space $(E, \mathcal{E})$, a transition measure Q on E and a given positive function λ, satisfying Hypothesis 1.2.1. Let X be the associated semi-Markov process constructed out of them as described in Section 1.2.1. We fix a deterministic terminal time $T > 0$ and a pair $(x, \vartheta) \in S$, and we look at all processes under the probability $\mathbb{P}^{x,\vartheta}$. We denote by $\mathcal{F}$ the natural filtration $(\mathcal{F}_t)_{t\in[0,\infty)}$ of X. Conditions 1, 5 and 6 above imply that the filtration $\mathcal{F}$ is right continuous (see [**18**], Appendix A2, Theorem T26). The predictable σ-algebra (respectively, the progressive σ-algebra) on $\Omega \times [0, \infty)$ is denoted by $\mathcal{P}$ (respectively, by $Prog$). The same symbols also denote the restriction to $\Omega \times [0, T]$.

We define a sequence $(T_n)_{n\geqslant 1}$ of random variables with values in $[0, \infty]$, setting

$$T_0(\omega) = 0, \quad T_{n+1}(\omega) = \inf\{s \geqslant T_n(\omega) : \, X_s(\omega) \neq X_{T_n}(\omega)\}, \tag{1.9}$$

with the convention that $T_{n+1}(\omega) = \infty$ if the indicated set is empty. Being X a jump process we have $T_n(\omega) \leqslant T_{n+1}(\omega)$ if $T_{n+1}(\omega) < \infty$, while the non explosion of X means that $T_{n+1}(\omega) \to \infty$. We stress the fact that $(T_n)_{n\geqslant 1}$ coincide by definition with the time jumps of the two dimensional process (X, θ).

For $\omega \in \Omega$ we define a random measure on $([0, \infty) \times E, \mathcal{B}[0, \infty) \otimes \mathcal{E})$ setting

$$p(\omega, C) = \sum_{n\geqslant 1} 1_{\{(T_n(\omega),\, X_{T_n}(\omega))\in C\}}, \qquad C \in \mathcal{B}[0, \infty) \otimes \mathcal{E}. \tag{1.10}$$

The random measure $\lambda(X_{s-}, \theta_{s-})\, Q(X_{s-}, \theta_{s-}, dy)\, ds$ is called the compensator, or the dual predictable projection, of $p(ds, dy)$. We are interested in the following family of backward equations driven by the compensated random measure $q(ds\, dy) = p(ds\, dy) - \lambda(X_{s-}, \theta_{s-})\, Q(X_{s-}, \theta_{s-}, dy)\, ds$ and parametrized by (x, ϑ): $\mathbb{P}^{x,\vartheta}$-a.s.,

$$Y_s + \int_s^T \int_E Z_r(y)\, q(dr\, dy) = g(X_T, \theta_T) + \int_s^T f\Big(r, X_r, \theta_r, Y_r, Z_r(\cdot)\Big)\, dr, \qquad s \in [0, T]. \tag{1.11}$$

We consider the following assumptions on the data f and g.

Hypothesis 1.2.3.

(1) The final condition $g : S \to \mathbb{R}$ is $\mathcal{S}$-measurable and $\mathbb{E}^{x,a}\left[|g(X_T, \theta_T)|^2\right] < \infty$.

(2) The generator f is such that

(i) for every $s \in [0, T]$, $(x, \vartheta) \in S$, $r \in \mathbb{R}$, f is a mapping $f(s, x, \vartheta, r, \cdot) : \mathcal{L}^2(E, \mathcal{E}, \lambda(x, \vartheta)\, Q(x, \vartheta, dy)) \to \mathbb{R}$;

(ii) for every bounded and $\mathcal{E}$-measurable $z : E \to \mathbb{R}$ the mapping

$$(s, x, \vartheta, r) \mapsto f(s, x, \vartheta, r, z(\cdot)) \tag{1.12}$$

is $\mathcal{B}([0, T]) \otimes \mathcal{S} \otimes \mathcal{B}(\mathbb{R})$-measurable.

(iii) There exist $L \geqslant 0$, $L' \geqslant 0$ such that for every $s \in [0, T]$, $(x, \vartheta) \in S$, $r, r' \in \mathbb{R}, z, z' \in \mathcal{L}^2(E, \mathcal{E}, \lambda(x, \vartheta)\, Q(x, \vartheta, dy))$ we have

$$\begin{aligned} &\left|f(s, x, \vartheta, r, z(\cdot)) - f(s, x, \vartheta, r', z'(\cdot))\right| \\ &\leqslant L' \left|r - r'\right| + L \left(\int_E \left|z(y) - z'(y)\right|^2 \lambda(x, \vartheta)\, Q(x, \vartheta, dy)\right)^{1/2}. \end{aligned} \tag{1.13}$$

(iv) We have

$$\mathbb{E}^{x,\vartheta}\left[\int_0^T |f(s,X_s,\theta_s,0,0)|^2\,ds\right] < \infty. \tag{1.14}$$

Remark 1.2.4. Assumptions (i), (ii), and (iii) imply the following measurability properties of $f(s, X_s, \theta_s, Y_s, Z_s(\cdot))$:

- if $Z \in \mathcal{L}^2(p)$, then the mapping
$$(\omega, s, y) \mapsto f(s, X_{s-}(\omega), \theta_{s-}(\omega), y, Z_s(\omega, \cdot))$$
is $\mathcal{P} \otimes \mathcal{B}(\mathbb{R})$-measurable;
- if, in addition, Y is a $Prog$-measurable process, then
$$(\omega, s) \mapsto f(s, X_{s-}(\omega), \theta_{s-}(\omega), Y_s(\omega), Z_s(\omega, \cdot))$$
is $Prog$-measurable.

□

We introduce the space $\mathbb{M}^{x,\vartheta}$ of the processes (Y, Z) on $[0, T]$ such that Y is real-valued and $Prog$-measurable, $Z : \Omega \times E \to \mathbb{R}$ is $\mathcal{P} \otimes \mathcal{E}$-measurable, and

$$||(Y,Z)||^2_{\mathbb{M}^{x,\vartheta}} := \mathbb{E}^{x,\vartheta}\left[\int_0^T |Y_s|^2\,ds + \int_0^T\int_E |Z_s(y)|^2\,\lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds\right] < \infty.$$

The space $\mathbb{M}^{x,\vartheta}$ endowed with this norm is a Banach space, provided we identify pairs of processes whose difference has norm zero.

Theorem 1.2.5. *Suppose that Hypothesis 1.2.3 holds for some $(x, \vartheta) \in S$. Then there exists a unique pair (Y, Z) in $\mathbb{M}^{x,\vartheta}$ which solves the BSDE* (1.11)*. Let moreover (Y', Z') be another solution in $\mathbb{M}^{x,\vartheta}$ to the BSDE* (1.11) *associated with the driver f' and final datum g'. Then*

$$\begin{aligned}
&\sup_{s\in[0,T]} \mathbb{E}^{x,\vartheta}\left[|Y_s - Y'_s|^2\right] + \mathbb{E}^{x,\vartheta}\left[\int_0^T |Y_s - Y'_s|^2 ds\right] \\
&+ \mathbb{E}^{x,\vartheta}\left[\int_0^T\int_E |Z_s(y) - Z'_s(y)|^2 \lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds\right] \\
&\leqslant C\mathbb{E}^{x,\vartheta}\left[|g(X_T) - g'(X_T)|^2\right] \\
&+ C\mathbb{E}^{x,\vartheta}\left[\int_0^T |f(s,X_s,\theta_s,Y'_s,Z'_s(\cdot)) - f'(s,X_s,\theta_s,Y'_s,Z'_s(\cdot))|^2 ds\right],
\end{aligned} \tag{1.15}$$

where C is a constant depending on T, L, L'.

Remark 1.2.6. The construction of a solution to the BSDE (1.11) is based on the integral representation theorem of marked point process martingales (see, e.g., [**35**]), and on a fixed-point argument. Similar results of well-posedness for BSDEs driven by random measures can be found in literature, see, in particular, the theorems given in [**28**], Section 3, and in [**12**]. Notice that these results can not be a priori straight applied to our framework: in [**12**] are involved random compensators which are absolutely continuous with respect to a deterministic measure, instead in our case the compensator is a stochastic random measure with a non-dominated intensity; [**28**] apply to BSDEs driven by a random measure associated to a pure jump Markov process, while the two dimensional process (X, θ) is Markov but not pure jump. Nevertheless, under Hypothesis 1.2.3, Theorem 3.4 and Proposition 3.5 in [**28**] can be extended to our framework without additional difficulties. The proofs turn out to be very similar to those of the mentioned results, and we do not report them here to alleviate the presentation. □

1.3. Optimal control of semi-Markov processes

1.3.1. Formulation of the problem. In this section we consider again a measurable space $(E, \mathcal{E})$, a transition measure Q and a function λ satisfying Hypothesis 1.2.1. The data specifying the optimal control problem we will address to are an action (or decision) space U, a running cost function l, a terminal cost function g, a (deterministic, finite) time horizon $T > 0$ and another function r specifying the effect of the control process. We define an admissible control process, or simply a control, as a predictable process $(u_s)_{s\in[0,\,T]}$ with values in U. The set of admissible control processes is denoted by $\mathcal{A}$. We will make the following assumptions:

Hypothesis 1.3.1.

(1) $(U, \mathcal{U})$ is a measurable space.

(2) The function $r : [0,\,T] \times S \times E \times U \to \mathbb{R}$ is $\mathcal{B}([0,\,T]) \otimes \mathcal{S} \otimes \mathcal{E} \otimes \mathcal{U}$-measurable and there exists a constant $C_r > 1$ such that,

$$0 \leqslant r(t, x, \vartheta, y, u) \leqslant C_r, \qquad t \in [0,\,T],\ (x, \vartheta) \in S,\ y \in E,\ u \in U. \tag{1.16}$$

(3) The function $g : S \to \mathbb{R}$ is $\mathcal{S}$-measurable, and for all fixed $t \in [0,\,T]$,

$$\mathbb{E}^{x,\vartheta}\left[|g(X_{T-t}, \theta_{T-t})|^2\right] < \infty, \qquad \forall (x, \vartheta) \in S. \tag{1.17}$$

(4) The function $l : [0,\,T] \times S \times U \to \mathbb{R}$ is $\mathcal{B}([0\ T]) \otimes \mathcal{S} \otimes \mathcal{U}$-measurable and there exists $\alpha > 1$ such that, for every fixed $t \in [0,\,T]$, for every $(x, \vartheta) \in S$ and $u(\cdot) \in \mathcal{A}$,

$$\begin{aligned} &\inf_{u\in U} l(t, x, \vartheta, u) > \infty; \\ &\mathbb{E}^{x,\vartheta}\left[\textstyle\int_0^{T-t} |\inf_{u\in U} l(t+s, X_s, \theta_s, u)|^2\, ds\right] < \infty, \\ &\mathbb{E}^{x,\vartheta}\left[\textstyle\int_0^{T-t} |l(t+s, X_s, \theta_s, u_s)|\, ds\right]^\alpha < \infty. \end{aligned} \tag{1.18}$$

To any $(t, x, \vartheta) \in [0,\,T] \times S$ and any control $u(\cdot) \in \mathcal{A}$ we associate a probability measure $\mathbb{P}^{x,\vartheta}_{u,t}$ by a change of measure of Girsanov type, as we now describe. Recalling the definition of the jump times T_n in (1.9), we define, for every fixed $t \in [0,\,T]$,

$$L^t_s = \exp\left(\int_0^s \int_E (1 - r(t+\sigma, X_\sigma, \theta_\sigma, y, u_\sigma))\, \lambda(X_\sigma, \theta_\sigma)\, Q(X_\sigma, \theta_\sigma, dy)\, d\sigma\right) \cdot$$
$$\cdot \prod_{n\geqslant 1 : T_n \leqslant s} r(t + T_n, X_{T_n-}, \theta_{T_n-}, X_{T_n}, u_{T_n}),$$

for all $s \in [0, T-t]$, with the convention that the last product equals 1 if there are no indices $n \geqslant 1$ satisfying $T_n \leqslant s$. As a consequence of the boundedness assumption on Q and λ it can be proved, using for instance Lemma 4.2 in [**27**], or [**18**] Chapter VIII Theorem T11, that for every fixed $t \in [0, T]$ and for every $\gamma > 1$ we have

$$\mathbb{E}^{x,\vartheta}\left[\left|L^t_{T-t}\right|^\gamma\right] < \infty, \qquad \mathbb{E}^{x,\vartheta}\left[L^t_{T-t}\right] = 1, \tag{1.19}$$

and therefore the process L^t is a martingale (relative to $\mathbb{P}^{x,\vartheta}$ and $\mathcal{F}$). Defining a probability $\mathbb{P}^{x,\vartheta}_{u,t}(d\omega) = L^t_{T-t}(\omega)\, \mathbb{P}^{x,\vartheta}(d\omega)$, we introduce the cost functional corresponding to $u(\cdot) \in \mathcal{A}$ as

$$J(t,x,\vartheta,u(\cdot)) = \mathbb{E}^{x,a}_{u,t}\left[\int_0^{T-t} l(t+s, X_s, \theta_s, u_s)\, ds + g(X_{T-t}, \theta_{T-t})\right], \tag{1.20}$$

where $\mathbb{E}^{x,\vartheta}_{u,t}$ denotes the expectation under $\mathbb{P}^{x,\vartheta}_{u,t}$. Taking into account (1.17), (1.18) and (1.19), and using Hölder inequality it is easily seen that the cost is finite for every admissible control. The control problem starting at (x,ϑ) at time $s = 0$ with terminal time $s = T - t$ consists in minimizing $J(t,x,\vartheta,\cdot)$ over $\mathcal{A}$. We finally introduce the value function

$$v(t,x,\vartheta) = \inf_{u(\cdot)\in\mathcal{A}} J(t,x,\vartheta,u(\cdot)), \qquad t \in [0, T],\ (x,\vartheta) \in S.$$

The previous formulation of the optimal control problem by means of change of probability measure is classical (see e.g. [**49**], [**57**], [**18**]). Some comments may be useful at this point.

Remark 1.3.2.

1. The particular form of cost functional (1.20) is due to the fact that the time-homogeneous Markov process $(X_s, \theta_s)_{s\geqslant 0}$ satisfies
$$\mathbb{P}^{x,\vartheta}(X_0 = x,\, \theta_0 = \vartheta) = 1;$$
the introduction of the temporal translation in the first component allows us to define $J(t,x,\vartheta,u(\cdot))$ for all $t \in [0, T]$.
2. We recall (see e.g. [**18**], Appendix A2, Theorem T34) that a process u is $\mathcal{F}$-predictable if and only if it admits the representation
$$u_s(\omega) = \sum_{n\geqslant 0} u^{(n)}_s(\omega)\, 1_{(T_n(\omega), T_{n+1}(\omega)]}(s)$$
where for each $(\omega, s) \mapsto u^{(n)}_s(\omega)$ is $\mathcal{F}_{[0,\,T_n]}\otimes\mathcal{B}(\mathbb{R}^+)$-measurable, with $\mathcal{F}_{[0,\,T_n]} = \sigma(T_i, X_{T_i},\, 0 \leqslant i \leqslant n)$ (see e.g. [**18**], Appendix A2, Theorem T30). Thus the fact that controls are predictable processes admits the following interpretation: at each time T_n (i.e. immediately after a jump) the controller, having observed the random variables T_i, X_{T_i}, $(0 \leqslant i \leqslant n)$, chooses his current action, and updates her/his decisions only at time T_{n+1}.
3. It can be proved (see [**75**] Theorem 4.5) that the compensator of $p(ds\, dy)$ under $\mathbb{P}^{x,\vartheta}_{u,t}$ is
$$r(t+s, X_{s-}, \theta_{s-}, y, u_s)\, \lambda(X_{s-}, \theta_{s-})\, Q(X_{s-}, \theta_{s-}, dy)\, ds,$$

whereas the compensator of $p(ds\,dy)$ under $\mathbb{P}^{x,\vartheta}$ was $\lambda(X_{s-},\theta_{s-})\,Q(X_{s-},\theta_{s-},dy)$ ds. This explains that the choice of a given control $u(\cdot)$ affects the stochastic system multiplying its compensator by $r(t+s,x,\vartheta,y,u_s)$.

4. We call control law an arbitrary measurable function $\underline{\mathrm{u}} : [0,\,T]\times S \to U$. Given a control law one can define an admissible control u setting $u_s = \underline{\mathrm{u}}(s,X_{s-},\theta_{s-})$.
   Controls of this form are called feedback controls. For a feedback control the compensator of $p(ds\,dy)$ is $r(t+s,X_{s-},\theta_{s-},y,\underline{\mathrm{u}}(s,X_{s-},\theta_{s-}))\;\lambda(X_{s-},\theta_{s-})$ $Q(X_{s-},\theta_{s-},dy)\;ds$ under $\mathbb{P}^{x,\vartheta}_{u,t}$. Thus, the process (X,θ) under the optimal probability is a two-dimensional Markov process corresponding to the transition measure
   $$r(t+s,x,\vartheta,y,\underline{\mathrm{u}}(s,x,\vartheta))\,\lambda(x,\vartheta)\,Q(x,\vartheta,dy)$$
   instead of $\lambda(x,\vartheta)\,Q(x,\vartheta,dy)$. However, even if the optimal control is in the feedback form, the optimal process is not, in general, time-homogeneous since the control law may depend on time. In this case, according to the definition given in Section 1.2, the process X under the optimal probability is not a semi-Markov process.

$\square$

Remark 1.3.3. Our formulation of the optimal control should be compared with another approach (see e.g. [**125**]). In [**125**] is given a family of jump measures on E $\{Q(x,b,\cdot),\,b\in B\}$ with B some index set endowed with a topology. In the so called *strong formulation* a control u is an ordered pair of functions (λ',β) with $\lambda' : S\to\mathbb{R}^+$, $\beta : S\to B$ such that

$$\begin{aligned}&\lambda' \text{ and } \beta \text{ are } \mathcal{S}-\text{measurable};\\ &\forall x\in E,\,\exists\, t(x)>0:\; \textstyle\int_0^{t(x)} \lambda'(x,r)\,dr<\infty;\\ &Q(\cdot,\beta,A) \text{ is } \mathcal{B}^+\text{-measurable } \forall A\in\mathcal{E}.\end{aligned}$$

If $\mathcal{A}$ is the class of controls which satisfies the above conditions, then a control $u=(\lambda',\beta)\in\mathcal{A}$ determines a controlled process X^u in the following manner. Let

$$H^u(x,s) := 1-e^{-\int_0^s \lambda'(x,r)\,dr},\quad \forall(x,s)\in S,$$

and suppose that $(X_0^u,\theta_0^u)=(x,\vartheta)$. Then at time 0, the process starts in state x and remains there a random time $S_1>0$, such that

$$\mathbb{P}^{x,\vartheta}\{S_1\leqslant s\} = \frac{H^u(x,\vartheta+s)-H^u(x,\vartheta)}{1-H^u(x,\vartheta)}. \tag{1.21}$$

At time S_1 the process transitions to the state $X^u_{S_1}$, where

$$\mathbb{P}^{x,\vartheta}\left\{X^u_{S_1}\in A|S_1\right\} = Q(x,\beta(x,S_1),A).$$

The process stays in state $X^u_{S_1}$ for a random time $S_2>0$ such that

$$\mathbb{P}^{x,\vartheta}\left\{S_2\leqslant s|S_1,\,X^u_{S_1}\right\} = H^u(X^u_{S_1},s)$$

and then at time S_1+S_2 transitions to $X^u_{S_1+S_2}$, where

$$\mathbb{P}^{x,\vartheta}\left\{X^u_{S_1+S_2}\in A|S_1,\,X^u_{S_1},\,S_2\right\} = Q(X^u_{S_1},\beta(X^u_{S_1},S_2),A).$$

We remark that the process X^u constructed in this way turns out to be semi-Markov.

We also mention that the class of control problems specified by the initial data λ' and β is in general larger that the one we address in this chapter. This can be seen noticing that in our framework all the controlled processes have laws which are absolutely continuous with respect to a single uncontrolled process (the one corresponding to $r \equiv 1$) whereas this might not be the case for the rate measures $\lambda'(x,\vartheta)\, Q(x, \beta(x,\vartheta), A)$ when $u = (\lambda', \beta)$ ranges in the set of all possible control laws. □

1.3.2. BSDEs and the synthesis of the optimal control. We next proceed to solve the optimal control problem formulated above. A basic role is played by the BSDE: for every fixed $t \in [0, T]$, $\mathbb{P}^{x,\vartheta}$-a.s.

$$\begin{aligned} Y^{x,\vartheta}_{s,t} &+ \int_s^{T-t} \int_E Z^{x,\vartheta}_{\sigma,t}(y)\, q(d\sigma\, dy) \\ &= g(X_{T-t}, \theta_{T-t}) + \int_s^{T-t} f\Big(t+\sigma, X_\sigma, \theta_\sigma, Z^{x,\vartheta}_{\sigma,t}(\cdot)\Big) d\sigma, \end{aligned} \tag{1.22}$$

$\forall s \in [0, T-t]$, with terminal condition given by the terminal cost g and generator given by the Hamiltonian function f defined for every $s \in [0, T]$, $(x,\vartheta) \in S$, $z \in L^2(E, \mathcal{E}, \lambda(x,\vartheta)\, Q(x,\vartheta, dy))$, as

$$f(s,x,\vartheta,z(\cdot)) = \inf_{u\in U} \Big\{ l(s,x,\vartheta,u) + \int_E z(y)(r(s,x,\vartheta,y,u)-1)\lambda(x,\vartheta)Q(x,\vartheta,dy) \Big\}. \tag{1.23}$$

In (1.22) the superscript (x,ϑ) denotes the starting point at time $s=0$ of the process $(X_s, \theta_s)_{s\geqslant 0}$, while the dependence of Y and Z on the parameter t is related to the temporal horizon of the considered optimal control problem. For every $t \in [0\ T]$, we look for a process $Y^{x,\vartheta}_{s,t}(\omega)$ adapted and càdlàg and a process $Z^{x,\vartheta}_{s,t}(\omega, y)$ $\mathcal{P}\otimes\mathcal{E}$-measurable satisfying the integrability conditions

$$\begin{aligned} &\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t} \left|Y^{x,\vartheta}_{s,t}\right|^2 ds\right] < \infty, \\ &\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t} \int_E \left|Z^{x,\vartheta}_{s,t}(y)\right|^2 \lambda(X_s,\theta_s)\, Q(X_s,\theta_s,dy)\, ds\right] < \infty. \end{aligned}$$

One can verify that, under Hypothesis 1.3.1 on the optimal control problem, all the assumptions of Hypothesis 1.2.3 hold true for the generator f and the terminal condition g in the BSDE (1.22). The only non trivial verification is the Lipschitz condition (1.13), which follows from the boundedness assumption (1.16). Indeed, for every $s \in [0, T]$, $(x,\vartheta) \in S$, z, $z' \in L^2(E, \mathcal{E}, \lambda(x,\vartheta)\, Q(x,\vartheta,dy))$,

$$\begin{aligned} &\int_E z(y)(r(s,x,\vartheta,y,u)) - 1)\, \lambda(x,\vartheta)\, Q(x,\vartheta,dy) \\ &\leqslant \int_E \left|z(y) - z'(y)\right|\, (r(s,x,\vartheta,y,u) - 1)\, \lambda(x,\vartheta)\, Q(x,\vartheta,dy) \\ &+ \int_E z'(y)(r(s,x,\vartheta,y,u) - 1)\, \lambda(x,\vartheta)\, Q(x,\vartheta,dy) \end{aligned}$$

$$\leqslant (C_r+1)\,(\lambda(x,\vartheta)\,Q(x,\vartheta,E))^{1/2}\cdot\left(\int_E \left|z(y)-z'(y)\right|^2\,\lambda(x,\vartheta)\,Q(x,\vartheta,dy)\right)^{1/2}$$
$$+\int_K z'(y)(r(s,x,\vartheta,y,u)-1)\,\lambda(x,\vartheta)\,Q(x,\vartheta,dy),$$

so that, adding $l(s,x,\vartheta,u)$ on both sides and taking the infimum over $u\in U$, it follows that

$$f(s,x,\vartheta,z)\leqslant L\left(\int_E \left|z(y)-z'(y)\right|^2\lambda(x,\vartheta)\,Q(x,\vartheta,dy)\right)^{1/2}+f(s,x,\vartheta,z'),\quad (1.24)$$

where $L:=(C_r+1)\sup_{(x,\vartheta)\in S}\sqrt{\lambda(x,\vartheta)}$ exchanging z and z' roles we obtain (1.13).

Then by Theorem 1.2.5, for every fixed $t\in[0,T]$, for every $(x,\vartheta)\in S$, there exists a unique solution of (1.22) $(Y^{x,\vartheta}_{s,t},Z^{x,\vartheta}_{s,t})_{s\in[0,T-t]}$, and $Y^{x,\vartheta}_{0,t}$ is deterministic. Moreover, we have the following result:

Proposition 1.3.4. *Assume that Hypotheses 1.3.1 hold. Then, for every $t\in[0,T]$, $(x,\vartheta)\in S$, and for every $u(\cdot)\in\mathcal{A}$,*

$$Y^{x,\vartheta}_{0,t}\leqslant J(t,x,\vartheta,u(\cdot)).$$

Proof. We consider the BSDE (1.22) at time $s=0$ and we apply the expected value $\mathbb{E}^{x,\vartheta}_{u,t}$ associated to the controlled probability $\mathbb{P}^{x,\vartheta}_{u,t}$. Since the $\mathbb{P}^{x,\vartheta}_{u,t}$-compensator of $p(dsdy)$ is
$r(t+s,X_{s-},\theta_{s-},y,u_s)\,\lambda(X_{s-},\theta_{s-})\,Q(X_{s-},\theta_{s-},dy)\,ds$, we have that

$$\begin{aligned}
&\mathbb{E}^{x,\vartheta}_{u,t}\left[\int_0^{T-t}\int_E Z^{x,\vartheta}_{s,t}(y)\,q(dsdy)\right]\\
&=\mathbb{E}^{x,\vartheta}_{u,t}\left[\int_0^{T-t}\int_E Z^{x,\vartheta}_{s,t}(y)\,p(dsdy)\right]\\
&-\mathbb{E}^{x,\vartheta}_{u,t}\left[\int_0^{T-t}\int_E Z^{x,\vartheta}_{s,t}(y)\,\lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds\right]\\
&=\mathbb{E}^{x,\vartheta}_{u,t}\left[\int_0^{T-t}\int_E Z^{x,\vartheta}_{s,t}(y)\,[r(t+s,X_s,\theta_s,y,u_s)-1]\,\lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds\right].
\end{aligned}$$

Then

$$\begin{aligned}
Y^{x,\vartheta}_{0,t}&=\mathbb{E}^{x,\vartheta}_{u,t}\left[g(X_{T-t},\theta_{T-t})\right]+\mathbb{E}^{x,\vartheta}_{u,t}\left[\int_0^{T-t} f(t+s,X_s,\theta_s,Z^{x,\vartheta}_{s,t}(\cdot))\,ds\right]\\
&-\mathbb{E}^{x,\vartheta}_{u,t}\left[\int_0^{T-t}\int_E Z^{x,\vartheta}_{s,t}(y)\,[r(t+s,X_s,\theta_s,y,u_s)-1]\,\lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds\right].
\end{aligned}$$

Adding and subtracting $\mathbb{E}^{x,\vartheta}_{u,t}\left[\int_0^{T-t} l(t+s,X_s,\theta_s,u_s)\,ds\right]$ on the right side we obtain the following relation:

$$\begin{aligned}
Y^{x,\vartheta}_{0,t}=&J(t,x,\vartheta,u(\cdot))\\
&+\mathbb{E}^{x,\vartheta}_{u,t}\left[\int_0^{T-t}\left[f(t+s,X_s,\theta_s,Z^{x,\vartheta}_{s,t}(\cdot))-l(t+s,X_s,\theta_s,u_s)\right]\,ds\right]
\end{aligned}$$

$$- \mathbb{E}_{u,t}^{x,\vartheta} \left[\int_0^{T-t} \int_E Z_{s,t}^{x,\vartheta}(\cdot) \, [r(t+s, X_s, \theta_s, y, u_s) - 1] \, \lambda(X_s, \theta_s) \, Q(X_s, \theta_s, dy) \, ds \right]. \tag{1.25}$$

By the definition of the Hamiltonian function f, the two last terms are non positive, and it follows that

$$Y_{0,t}^{x,\vartheta} \leqslant J(t, x, \vartheta, u(\cdot)), \qquad \forall u(\cdot) \in \mathcal{A}.$$

□

We define the following, possibly empty, set:

$$\begin{aligned} \Gamma(s, x, \vartheta, z(\cdot)) = \Big\{ & u \in U : f(s, x, \vartheta, z(\cdot)) = l(s, x, \vartheta, u) \\ & + \int_E z(y) \, (r(s, x, \vartheta, y, u) - 1) \, \lambda(x, \vartheta) \, Q(x, \vartheta, dy); \\ & s \in [0,\, T],\, (x, \vartheta) \in S,\, z \in L^2(E, \mathcal{E}, \lambda(x, \vartheta) \, Q(x, \vartheta, dy)) \Big\}. \end{aligned} \tag{1.26}$$

In order to prove the existence of an optimal control we need to require that the infimum in the definition of f is achieved. Namely we assume that

Hypothesis 1.3.5. The sets Γ introduced in (1.26) are non empty; moreover, for every fixed $t \in [0,\, T]$ and $(x, \vartheta) \in S$, one can find an $\mathcal{F}$-predictable process $u^{*\,t,x,\vartheta}(\cdot)$ with values in U satisfying

$$u_s^{*\,t,x,\vartheta} \in \Gamma(t+s, X_{s-}, \theta_{s-}, Z_{s,t}^{x,\vartheta}(\cdot)), \qquad \mathbb{P}^{x,\vartheta}\text{-a.s. } \forall s \in [0,\, T-t]. \tag{1.27}$$

Theorem 1.3.6. *Under Hypothesis 1.3.1 and 1.3.5 for every fixed $t \in [0,\, T]$ and $(x, \vartheta) \in S$, $u^{*\,t,x,\vartheta}(\cdot) \in \mathcal{A}$ is an optimal control for the control problem starting from (x, ϑ) at time zero with time horizon $T - t$. Moreover, $Y_{0,t}^{x,\vartheta}$ coincides with the value function, i.e. $Y_{0,t}^{x,\vartheta} = J(t, x, \vartheta, u^{*\,t,x,\vartheta}(\cdot))$.*

Proof. It follows immediately from the relation (1.25) and from the definition of the Hamiltonian function f. □

We recall that general conditions can be formulated for the existence of a process $u^{*\,t,x,\vartheta}(\cdot)$ satisfying (1.27), hence of an optimal control; this is done by means of an appropriate selection theorem, see e.g. Proposition 5.9 in [**28**].

We end this section with an example where the BSDE (1.22) can be explicitly solved and a closed form solution of an optimal control problem can be found.

Example 1.3.7. We consider a fixed time interval $[0,\, T]$ and a state space consisting of three states: $E = \{x_1, x_2, x_3, x_4\}$. We introduce $(T_n, \xi_n)_{n \geqslant 0}$ setting $(T_0, \xi_0) = (0, x_1)$, $(T_n, \xi_n) = (+\infty, x_1)$ if $n \geqslant 3$ and on (T_1, ξ_1) and (T_2, ξ_2) we make the following assumptions: ξ_1 takes values x_2 with probability 1, ξ_2 takes values x_3, x_4 with probability 1/2. This means that the system starts at time zero in a given state x_1, jumps into state x_2 with probability 1 at the random time T_1 and into state x_3 or x_4 with equal probability at the random time T_2. It has no jumps after. We take $U = [0,\, 2]$ and define the function r specifying the effects of the control process as $r(x_1, u) = r(x_2, u) = 1$, $r(x_3, u) = u$, $r(x_4, u) = 2 - u$, $u \in U$.

Moreover, the final cost g assumes the value 1 in $(x,\vartheta)=(x_4,T-T_2)$ and zero otherwise, and the running cost is defined as $l(s,x,\vartheta,u)=\frac{\alpha\, u}{2}\,\lambda(x,\vartheta)$, where $\alpha>0$ is a fixed parameter. The BSDE we want to solve takes the form:

$$\begin{aligned}
Y_s &+ \int_s^T \int_E Z_\sigma(y)\, p(d\sigma\, dy) = g(X_T,\, \theta_T)\\
&+ \int_s^T \inf_{u\in[0,\,2]} \Big\{ \frac{\alpha\, u}{2} + \int_E Z_\sigma(y)\, r(y,u)\, Q(X_\sigma,\theta_\sigma,dy)\Big\}\, \lambda(X_\sigma,\theta_\sigma) d\sigma \qquad (1.28)
\end{aligned}$$

that can be written as

$$\begin{aligned}
Y_s &+ \sum_{n\geqslant 1} Z_{T_n}(X_{T_n})\, 1_{\{s<T_n\leqslant T\}} = g(X_T,\, \theta_T)\\
&+ \int_s^T \inf_{u\in[0,\,2]} \Big\{ \frac{\alpha\, u}{2} + Z_\sigma(x_2)\Big\}\, \lambda(x_1,\vartheta+\sigma) 1_{\{0\leqslant\sigma<T_1\wedge T\}}\, d\sigma\\
&+ \int_s^T \inf_{u\in[0,\,2]} \Big\{ \frac{\alpha\, u}{2} + Z_\sigma(x_3)\frac{u}{2} + Z_\sigma(x_4)(1-\frac{u}{2})\Big\}\, \lambda(x_2,\sigma-T_1) 1_{\{T_1\leqslant\sigma<T_2\wedge T\}}\, d\sigma.
\end{aligned}$$

It is known by [**29**] that BSDEs of this type admit the following explicit solution $(Y_s, Z_s(\cdot))_{s\in[0,T]}$:

$$\begin{aligned}
Y_s &= y^0(s)1_{\{s<T_1\}} + y^1(s,T_1,\xi_1)\, 1_{\{T_1\leqslant s<T_2\}} + y^2(s,T_2,\xi_2,T_1,\xi_1)\, 1_{\{T_2\leqslant s\}},\\
Z_s(y) &= z^0(s,y)\, 1_{\{s\leqslant T_1\}} + z^1(s,y,T_1,\xi_1)\, 1_{\{T_1<s\leqslant T_2\}}, \quad y\in E.
\end{aligned}$$

To deduce y^0 and y^1 we reduce the BSDE to a system of two ordinary differential equation. To this end, it suffices to consider the following cases:

- $\omega\in\Omega$ such that $T<T_1(\omega)<T_2(\omega)$: (1.28) reduces to

$$\begin{aligned}
y^0(s) &= \int_s^T \inf_{u\in[0,\,2]} \Big\{ \frac{\alpha\, u}{2} + z^0(\sigma,x_2)\Big\}\, \lambda(x_1,\vartheta+\sigma)\, d\sigma\\
&= \int_s^T z^0(\sigma,x_2)\, \lambda(x_1,\vartheta+\sigma)\, d\sigma\\
&= \int_s^T (y^1(\sigma,\sigma,x_2)-y^0(\sigma))\, \lambda(x_1,\vartheta+\sigma)\, d\sigma; \qquad (1.29)
\end{aligned}$$

- $\omega\in\Omega$ such that $T_1(\omega)<T<T_2(\omega),\ s>T_1$: (1.28) reduces to

$$\begin{aligned}
&y^1(s,T_1,\xi_1)\\
&= \int_s^T \inf_{u\in[0,\,2]} \Big\{ \frac{\alpha\, u}{2} + z^1(\sigma,x_3,T_1,\xi_1)\frac{u}{2} + z^1(\sigma,x_4,T_1,\xi_1)(1-\frac{u}{2})\Big\}\, \lambda(\xi_1,\sigma-T_1)\, d\sigma\\
&= \int_s^T [z^1(\sigma,x_4,T_1,\xi_1)\wedge(\alpha+z^1(\sigma,x_3,T_1,\xi_1))]\, \lambda(\xi_1,\sigma-T_1)\, d\sigma\\
&= \int_s^T [(1\wedge\alpha)-y^1(\sigma,T_1,\xi_1)]\, \lambda(\xi_1,\sigma-T_1)\, d\sigma. \qquad (1.30)
\end{aligned}$$

Solving (1.29) and (1.30) we obtain

$$y^0(s) = (1\wedge\alpha)\left(1-e^{-\int_s^T \lambda(x_1,\vartheta+\sigma)\, d\sigma}\right)$$

$$- (1 \wedge \alpha)\, e^{-\int_s^T \lambda(x_1,\vartheta+\sigma)\,d\sigma} \int_s^T \lambda(x_1,\vartheta+\sigma)\, e^{\int_\sigma^T \lambda(x_1,\vartheta+z)\,dz} e^{-\int_\sigma^T \lambda(x_2,z-\sigma)\,dz}\, d\sigma\},$$

$$y^1(s,T_1,\xi_1) = (1\wedge\alpha)\left(1 - e^{-\int_s^T \lambda(\xi_1,\sigma-T_1)\,d\sigma}\right);$$

moreover,

$$\begin{aligned}
&y^2(s,T_2,\xi_2,T_1,\xi_1) = 1_{\{\xi_2=x_4\}},\\
&z^0(s,x_1) = z^0(s,x_3) = z^0(s,x_4) = 0,\\
&z^0(s,x_2) = y^1(s,s,x_2) - y^0(s),\\
&z^1(s,x_1,T_1,\xi_1) = z^1(s,x_2,T_1,\xi_1) = 0,\\
&z^1(s,x_3,T_1,\xi_1) = (1\wedge\alpha)\left(e^{-\int_s^T \lambda(\xi_1,\sigma-T_1)\,d\sigma} - 1\right),\\
&z^1(s,x_4,T_1,\xi_1) = 1 + z^1(s,x_3,T_1,\xi_1),
\end{aligned}$$

where z^0 and z^1 are obtained respectively from y^2, y^1 and y^1, y^0 by subtraction. The optimal cost is then given by $Y_0 = y^0(0)$. The optimal control is obtained during the computation of the Hamiltonian function: it is the process $u_s = 2\,1_{(T_1,T_2]}(s)$ if $\alpha \leqslant 1$, and the process $u_s = 0$ if $\alpha \geqslant 1$ (both are optimal if $\alpha = 1$).

1.4. Nonlinear variant of Kolmogorov equation

Throughout this section we still assume that a semi-Markov process X is given. It is constructed as in Section 1.2.1 by the rate function λ and the measure Q on E, and (X,θ) is the associated time-homogeneous Markov process. We assume that λ and Q satisfy Hypothesis 1.2.1.

It is our purpose to present here some nonlinear variants of the classical backward Kolmogorov equation associated to the Markov process (X,θ) and to show that their solution can be represented probabilistically by means of an appropriate BSDE of the type considered above.

We will suppose that two functions f and g are given, satisfying Hypothesis 1.2.3, and that moreover g verifies, for every fixed $t \in [0,\,T]$,

$$\mathbb{E}^{x,a}\left[|g(X_{T-t},\theta_{T-t})|^2\right] < \infty. \tag{1.31}$$

We define the operator

$$\mathcal{L}\psi(x,\vartheta) := \int_E [\psi(y,0) - \psi(x,\vartheta)]\,\lambda(x,\vartheta)\,Q(x,\vartheta,dy), \qquad (x,\vartheta) \in S, \tag{1.32}$$

for every measurable function $\psi : S \to \mathbb{R}$ for which the integral is well defined.
The equation

$$\begin{aligned}
v(t,x,\vartheta) &= g(x,\vartheta+T-t) + \int_t^T \mathcal{L}v(s,x,\vartheta+s-t)\,ds\\
&+ \int_t^T f(s,x,\vartheta+s-t, v(s,x,\vartheta+s-t), v(s,\cdot,0) - v(s,x,\vartheta+s-t))\,ds,
\end{aligned} \tag{1.33}$$

$t \in [0,\, T]$, $(x, \vartheta) \in S$, with unknown function $v : [0,\, T] \times S \to \mathbb{R}$ will be called the nonlinear Kolmogorov equation.

Equivalently, one requires that for every $x \in E$ and for all constant $c \in [-T,\, +\infty)$,

$$t \mapsto v(t, x, t + c) \text{ is absolutely continuous on } [0, T], \tag{1.34}$$

and

$$\begin{cases} Dv(t, x, \vartheta) + \mathcal{L}v(t, x, \vartheta) + f(t, x, \vartheta, v(t, x, \vartheta), v(t, \cdot, 0) - v(t, x, \vartheta)) = 0 \\ v(T, x, \vartheta) = g(x, \vartheta), \end{cases} \tag{1.35}$$

where D denotes the formal directional derivative operator

$$(Dv)(t, x, \vartheta) := \lim_{h \downarrow 0} \frac{v(t + h, x, \vartheta + h) - v(t, x, \vartheta)}{h}. \tag{1.36}$$

In other words, the presence of the directional derivative operator (1.36) allows us to understand the nonlinear Kolmogorov equation (1.35) in a classical sense. In particular, the first equality in (1.35) is understood to hold almost everywhere on $[0,\, T]$ outside of a dt-null set of points which can depend on (x, ϑ).

Under appropriate boundedness assumptions we have the following result:

Lemma 1.4.1. *Suppose that f and g verify Hypothesis 1.2.3 and that* (1.31) *holds; suppose, in addition, that*

$$\sup_{t \in [0,\, T],\, (x, \vartheta) \in S} \Big(|g(x, \vartheta)| + |f(t, x, \vartheta, 0, 0)| \Big) < \infty. \tag{1.37}$$

Then the nonlinear Kolmogorov equation (1.33) *has a unique solution v in the class of measurable bounded functions.*

Proof. The result follows as usual from a fixed-point argument, that we only sketch. Let us define a map Γ setting $v = \Gamma(w)$ where

$$\begin{aligned} v(t, x, \vartheta) &= g(x, \vartheta + T - t) + \int_t^T \mathcal{L}w(s, x, \vartheta + s - t)\, ds \\ &\quad + \int_t^T f(s, x, \vartheta + s - t, w(s, x, \vartheta + s - t), w(s, \cdot, 0) - w(s, x, \vartheta + s - t))\, ds. \end{aligned}$$

Using the Lipschitz character of f and Hypothesis 1.2.1-ii), one can show that, for some $\beta > 0$ sufficiently large, the above map is a contraction in the space of bounded measurable real functions on $[0,\, T] \times S$ endowed with the supremum norm:

$$||v||_* := \sup_{0 \leqslant t \leqslant T} \sup_{(x, \vartheta) \in S} e^{-\beta(T - t)} \left| v(t, x, \vartheta) \right|.$$

The unique fixed point of Γ gives the required solution. □

Our goal is now to remove the boundedness assumption (1.37). To this end we need to define a formula of Itô type for the composition of the process $(X_s,\, \theta_s)_{s \geqslant 0}$ with functions v smooth enough defined on $[0,\, T] \times S$. Taking into account the particular form of (1.33), and the fact that the second component of the process $(X_s,\, \theta_s)_{s \geqslant 0}$ is linear in s, the idea is to use in this formula the directional derivative operator D given by (1.36).

Lemma 1.4.2 (A formula of Itô type)**.** *Let consider functions* $v : [0,\, T] \times S \to \mathbb{R}$ *such that*

(i) $\forall\, x \in E$, $\forall\, c \in [-T,\, +\infty)$, *the map* $t \mapsto v(t, x, t+c)$ *is absolutely continuous on* $[0,\, T]$, *with directional derivative* D *given by* (1.36)*;*

(ii) *for fixed* $t \in [0,\, T]$, $\{v(t+s, y, 0) - v(t+s, X_{s-}, \theta_{s-}),\ s \in [0,\, T-t],\ y \in E\}$ *belongs to* $\mathcal{L}^1_{loc}(p)$.

Then $\mathbb{P}^{x,\vartheta}$*-a.s., for every* $t \in [0,\, T]$,

$$v(T, X_{T-t}, \theta_{T-t}) - v(t, x, \vartheta) = \int_0^{T-t} Dv(t+s, X_s, \theta_s)\, ds + \int_0^{T-t} \mathcal{L}v(t+s, X_s, \theta_s)\, ds + \int_0^{T-t} \int_E (v(t+s, y, 0) - v(t+s, X_{s-}, \theta_{s-}))\; q(ds, dy), \tag{1.38}$$

where the stochastic integral is a local martingale.

Proof. We proceed by reasoning as in the proof of Theorem 26.14 in [**35**]. We consider a function $v : [0,\, T] \times S \to \mathbb{R}$ satisfying (i) and (ii), and we denote by N_t the number of jumps in the interval $[0,\, t]$:

$$N_t = \sum_{n \geqslant 1} 1_{\{T_n \leqslant t\}}.$$

We have

$$\begin{aligned} v(T, X_T, \theta_T) - v(0, x, \vartheta) &= v(T, X_T, \theta_T) - v(T_{N_T}, X_{T_{N_T}}, \theta_{T_{N_T}}) \\ &\quad + \sum_{n=2}^{N_T} \big\{ v(T_n, X_{T_n}, \theta_{T_n}) - v(T_{n-1}, X_{T_{n-1}}, \theta_{T_{n-1}}) \big\} \\ &\quad + v(T_1, X_{T_1}, \theta_{T_1}) - v(0, x, \vartheta). \end{aligned}$$

Noticing that $X_{T_{n-}} = X_{T_{n-1}}$ for all $n \in [1,\, N_T]$, $X_T = X_{T_{N_T}}$, and that $\theta_{T_n} = 0$ for all $n \in [1,\, N_T]$, $\theta_{T_{1-}} = \vartheta + T_1$, and $\theta_{T_{n-}} = T_n - T_{n-1}$ for all $n \in [2,\, N_T]$, we have

$$v(T, X_T, \theta_T) - v(0, x, \vartheta) = I + II + III,$$

where

$$\begin{aligned} I &= (v(T_1, X_{T_1}, 0) - v(T_1, X_{T_1-}, \theta_{T_1-})) + (v(T_1, x, \vartheta + T_1) - v(0, x, \vartheta)) =: I' + I'', \\ II &= \sum_{n=2}^{N_T} (v(T_n, X_{T_n}, 0) - v(T_n, X_{T_n-}, \theta_{T_n-}) \\ &\quad + \sum_{n=2}^{N_T} (v(T_n, X_{T_{n-1}}, T_n - T_{n-1}) - v(T_{n-1}, X_{T_{n-1}}, 0))) \\ &\quad =: II' + II'', \\ III &= v(T, X_T, T - T_N) - v(T_N, X_{T_N}, 0). \end{aligned}$$

Let H denote the $\mathcal{P}\otimes\mathcal{E}$-measurable process

$$H_s(y) = v(s,y,0) - v(s,X_{s-},\theta_{s-}),$$

with the convention $X_{0-}=X_0$, $\theta_{0-}=\vartheta_0$. We have

$$\begin{aligned} I' + II' &= \sum_{n\geqslant 1: T_n\leqslant T} (v(T_n, X_{T_n}, 0) - v(T_n, X_{T_n-}, \theta_{T_n-})) \\ &= \sum_{n\geqslant 1: T_n\leqslant T} H_{T_n}(X_{T_N}) = \int_0^T\int_E H_s(y)\, p(ds,dy). \end{aligned}$$

On the other hand, since v satisfies (i) and recalling the definition 1.36 of the directional derivative operator D,

$$\begin{aligned} &I'' + II'' + III \\ &= \int_0^{T_1} \lim_{h\to 0} \frac{v(0+hs, x, \vartheta + hs) - v(0,x,\vartheta)}{h}\, ds \\ &+ \sum_{n\geqslant 2: T_n\leqslant T} \int_{T_{n-1}}^{T_n} \lim_{h\to 0} \frac{v(T_{n-1} + h(s-T_{n-1}), X_{T_{n-1}}, \theta_{T_{n-1}} + h(s-T_{n-1})) - v(T_{n-1}, X_{T_{n-1}}, \theta_{T_{n-1}})}{h}\, ds \\ &+ \int_{T_{N_T}}^T \lim_{h\to 0} \frac{v(T_{N_T} + h(s-T_{N_T}), X_{T_{N_T}}, \theta_{T_{N_T}} + h(s-T_{N_T})) - v(T_{N_T}, X_{T_{N_T}}, \theta_{T_{N_T}})}{h}\, ds \\ &= \int_0^T Dv(s, X_s, \theta_s)\, ds. \end{aligned}$$

Then $\mathbb{P}^{x,\vartheta}$-a.s.,

$$\begin{aligned} &v(T, X_T, a_T) - v(0,x,\vartheta) \\ &= \int_0^T Dv(s,X_s,\theta_s)\, ds + \int_0^T\int_E (v(s,y,0) - v(s,X_{s-},\theta_{s-}))\; p(ds,dy) \\ &= \int_0^T Dv(s,X_s,\theta_s)\, ds + \int_0^T \mathcal{L}v(s,X_s,\theta_s)\, ds \\ &+ \int_0^T\int_E (v(s,y,0) - v(s,X_{s-},\theta_{s-}))\; q(ds,dy), \end{aligned}$$

where the second equality is obtained using the identity $q(dt\,dy) = p(dt\,dy) - \lambda(X_{t-},\theta_{t-})\; Q(X_{t-},\theta_{t-},dy)\,dt$ together with the definition (1.32) of the operator $\mathcal{L}$.

Finally, applying a shift in time, i.e. considering for every $t\in[0,\,T]$ the differential of the process $v(s+t, X_{s-},\theta_{s-})$ with respect to $s\in[0,\,T-t]$, the previous formula becomes: $\mathbb{P}^{x,\vartheta}$-a.s., for every $t\in[0,\,T]$,

$$\begin{aligned} v(T-t, X_T, \theta_T) - v(t,x,\vartheta) &= \int_0^{T-t} Dv(s+t, X_s,\theta_s)\, ds + \int_0^{T-t} \mathcal{L}v(s+t,X_s,\theta_s)\, ds \\ &+ \int_0^{T-t}\int_E (v(s+t,y,0) - v(s+t,X_{s-},\theta_{s-}))\; q(ds,dy), \end{aligned}$$

where the stochastic integral is a local martingale thanks to condition (ii). $\square$

We will call (1.38) the Itô formula for $v(t+s,\cdot,\cdot)\circ(X_s,\theta_s)_{s\in[0,\,T-t]}$. In differential notation:

$$\begin{aligned}dv(t+s,X_{s-},\theta_{s-}) &= Dv(t+s,X_{s-},\theta_{s-})\,ds \;+\; \mathcal{L}v(t+s,X_{s-},\theta_{s-})\,ds\\ &\quad+\int_E (v(t+s,y,0)-v(t+s,X_{s-},\theta_{s-}))\;q(ds,dy).\end{aligned}$$

Remark 1.4.3. With respect to the classical Itô formula, we underline that in (1.38) we have

- the directional derivative operator D instead of the usual time derivative;
- the temporal translation in the first component of v, i.e. we consider the differential of the process
  $v(t{+}s,X_{s-},\theta_{s-})$ with respect to $s\in[0,\,T{-}t]$. Indeed, the time-homogeneous Markov process $(X_s,\theta_s)_{s\geqslant 0}$ satisfies
  $$\mathbb{P}^{x,\vartheta}(X_0=x,\,\theta_0=\vartheta)=1,$$
  and the temporal translation in the first component allows us to consider $dv(t,X_t,\theta_t)$ for all $t\in[0,\,T]$.

□

We go back to consider the Kolmogorov equation (1.33) in a more general setting. More precisely, on the functions f, g we will only ask that they satisfy Hypothesis 1.2.3 for every $(x,\vartheta)\in S$ and that (1.31) holds.

Definition 1.4.4. We say that a measurable function $v:[0,\,T]\times S\to\mathbb{R}$ is a solution of the nonlinear Kolmogorov equation (1.33), if, for every fixed $t\in[0,\,T]$, $(x,\vartheta)\in S$,

1. $\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\int_E |v(t+s,y,0)-v(t+s,X_s,\theta_s)|^2\,\lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds\right]<\infty$;
2. $\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}|v(t+s,X_s,\theta_s)|^2\,ds\right]<\infty$;
3. (1.33) is satisfied.

Remark 1.4.5. Condition 1. is equivalent to the fact that $v(t+s,y,0)-v(t+s,X_{s-},\theta_{s-})$ belongs to $\mathcal{L}^2(p)$. Conditions 1. and 2. together are equivalent to the fact that the pair

$$\{v(t+s,X_s,\theta_s),\,v(t+s,y,0)-v(t+s,X_{s-},\theta_{s-});\,s\in[0,\,T-t],\,y\in E\}$$

belongs to the space $\mathbb{M}^{x,\vartheta}$; in particular they hold true for every measurable bounded function v. □

Remark 1.4.6. We need to verify the well-posedness of equation (1.33) for a function v satisfying the condition 1. and 2. above. We start by noticing that, for every $(x,\vartheta)\in S$, $\mathbb{P}^{x,\vartheta}$-a.s.,

$$\int_0^T\int_E |v(s,y,0)-v(s,X_s,\theta_s)|^2\,\lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds+\int_0^T |v(s,X_s,\theta_s)|^2\,ds<\infty.$$

By the law (1.7) of the first jump it follows that the set $\{\omega \in \Omega : T_1(\omega) > T\}$ has positive $\mathbb{P}^{x,\vartheta}$ probability, and on this set we have $X_{s-}(\omega) = x$, $\theta_{s-}(\omega) = \vartheta + s$. Taking such an ω we get

$$\int_0^T \int_E |v(s,y,0) - v(s,x,\vartheta+s)|^2 \, \lambda(x,\vartheta+s)\, Q(x,\vartheta+s,dy)\, ds \\ + \int_0^T |v(s,x,\vartheta+s)|^2 \, ds < \infty, \ \forall (x,\vartheta) \in S.$$

Since $\sup_{(x,\vartheta)\in S} \lambda(x,\vartheta) < \infty$ by assumption, Hölder's inequality implies that

$$\begin{aligned}
&\int_0^T |\mathcal{L}(v(s,x,\vartheta+s))| \, ds \\
&\leqslant \int_0^T \int_E |v(s,y,0) - v(s,x,\vartheta+s)| \, \lambda(x,\vartheta+s)\, Q(x,\vartheta+s,dy)\, ds \\
&\leqslant c \left(\int_0^T \int_E |v(s,y,0) - v(s,x,\vartheta+s)|^2 \, \lambda(x,\vartheta+s)\, Q(x,\vartheta+s,dy)\, ds \right)^{1/2} \\
&< \infty
\end{aligned}$$

for some constant c and for all $(x,\vartheta) \in S$. Similarly, since

$$\mathbb{E}^{x,a} \left[\int_0^T |f(s,X_s,\theta_s,0,0)|^2 \, ds \right] < \infty,$$

and arguing again on the jump time T_1, we deduce that

$$\int_0^T |f(s,x,\vartheta+s,0,0)|^2 \, ds < \infty, \ \forall (x,\vartheta) \in S;$$

finally, from the Lipschitz conditions on f we can conclude that

$$\begin{aligned}
&\int_0^T |f(s,x,\vartheta+s, v(s,x,\vartheta+s), v(s,\cdot,0) - v(s,x,\vartheta+s))| \, ds \\
&\leqslant c_1 \left(\int_0^T |f(s,x,\vartheta+s,0,0)|^2 \, ds \right)^{1/2} + c_2 \left(\int_0^T |v(s,x,\vartheta+s)|^2 \, ds \right)^{1/2} \\
&+ c_3 \left(\int_0^T \int_E |v(s,y,0) - v(s,x,\vartheta+s)|^2 \, \lambda(x,\vartheta+s)\, Q(x,\vartheta+s,dy)\, ds \right)^{1/2} \\
&< \infty
\end{aligned}$$

for some constants c_i, $i = 1,2,3$, and for all $(x,\vartheta) \in S$. Therefore, all terms occurring in equation (1.33) are well defined. □

For every fixed $t \in [0,\, T]$ and $(x,\vartheta) \in S$, we consider now a BSDE of the form

$$\begin{aligned}
Y_{s,t}^{x,\vartheta} + \int_s^{T-t} \int_E Z_{r,t}^{x,\vartheta}(y)\, q(dr\, dy) &= g(X_{T-t}, \theta_{T-t}) \\
&+ \int_s^{T-t} f\Big(t+r, X_{r-}, \theta_{r-}, Y_{r,t}^{x,\vartheta}, Z_{r,t}^{x,\vartheta}(\cdot)\Big)\, dr, \ \ s \in [0,\, T-t]. \qquad (1.39)
\end{aligned}$$

Then there exists a unique solution $(Y^{x,\vartheta}_{s,t}, Z^{x,\vartheta}_{s,t}(\cdot))_{s\in[0,T-t]}$, in the sense of Theorem 1.2.5, and $Y^{x,\vartheta}_{0,t}$ is deterministic. We are ready to state the main result of this section.

Theorem 1.4.7. *Suppose that f, g satisfy Hypothesis 1.2.3 for every $(x,\vartheta)\in S$ and that* (1.31) *holds. Then the nonlinear Kolmogorov equation* (1.33) *has a unique solution $v(t,x,\vartheta)$ in the sense of Definition 1.4.4.*

Moreover, for every fixed $t\in[0,T]$, for every $(x,\vartheta)\in S$ and $s\in[0,T-t]$ we have

$$Y^{x,\vartheta}_{s,t} = v(t+s, X_{s-}, \theta_{s-}), \tag{1.40}$$

$$Z^{x,\vartheta}_{s,t}(y) = v(t+s, y, 0) - v(t+s, X_{s-}, \theta_{s-}), \tag{1.41}$$

so that in particular $v(t,x,\vartheta) = Y^{x,\vartheta}_{0,t}$.

Remark 1.4.8. The equalities (1.40) and (1.41) are understood as follows.

- $\mathbb{P}^{x,\vartheta}$-a.s., equality (1.40) holds for all $s\in[0,T-t]$. The trajectories of $(X_s)_{s\in[0,T-t]}$ are piecewise constant and càdlàg, while the trajectories of $(\theta_s)_{s\in[0,T-t]}$ are piecewise linear in s (with unitary slope) and càdlàg; moreover the processes $(X_s)_{s\in[0,T-t]}$ and $(\theta_s)_{s\in[0,T-t]}$ have the same jump times $(T_n)_{n\geqslant 1}$. Then the equality (1.40) is equivalent to the condition
$$\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\left|Y^{x,\vartheta}_{s,t} - v(t+s,X_s,\theta_s)\right|^2 ds\right] = 0.$$
- The equality (1.41) holds for all (ω,s,y) with respect to the measure $\lambda(X_{s-}(\omega),\theta_{s-}(\omega))\,Q(X_{s-}(\omega),\theta_{s-}(\omega),dy)\,\mathbb{P}^{x,\vartheta}(d\omega)ds$, i.e.,

$$\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\int_E\left|Z^{x,\vartheta}_{s,t}(y) - v(t+s,y,0) + v(t+s,X_s,\theta_s)\right|^2 \lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds\right]$$
$$= 0.$$

□

Proof. *Uniqueness.* Let v be a solution of the nonlinear Kolmogorov equation (1.33). It follows from equality (1.33) itself that for every $x\in E$ and every $\tau\in[-T,+\infty)$, $t\mapsto v(t,x,t+\tau)$ is absolutely continuous on $[0,T]$. Indeed, applying in (1.33) the change of variable $\tau := \vartheta - t$, we obtain $\forall t\in[0,T]$, $\forall \tau\in[-T,+\infty)$,

$$\begin{aligned} v(t,x,t+\tau) &= g(x,T+\tau) + \int_t^T \mathcal{L}v(s,x,s+\tau)\,ds \\ &\quad + \int_t^T f(s,x,s+\tau, v(s,x,s+\tau), v(s,\cdot,0) - v(s,x,s+\tau))\,ds. \end{aligned}$$

Then, since by assumption the process $v(t+s,y,0) - v(t+s,X_{s-},\theta_{s-})$ belongs to $\mathcal{L}^2(p)$, we are in a position to apply the Itô formula (1.38) to the process $v(t+s,X_{s-},\theta_{s-})$, $s\in[0,T-t]$. We get: $\mathbb{P}^{x,\vartheta}$-a.s.,

$$v(t+s,X_{s-},\theta_{s-}) = v(t,x,\vartheta) + \int_0^s Dv(t+r,X_r,\theta_r)\,dr + \int_0^s \mathcal{L}v(t+r,X_r,\theta_r)\,dr$$

$$+ \int_0^s \int_E (v(t+r, y, 0) - v(t+r, X_r, \theta_r))\, q(dr, dy), \quad s \in [0, T-t].$$

We know that v satisfies (1.35); moreover the process X has piecewise constant trajectories, the process θ has linear trajectories in s, and they have the same time jumps. Then, $\mathbb{P}^{x,\vartheta}$-a.s.,

$$\begin{aligned} &Dv(t+s, X_{s-}, \theta_{s-}) + \mathcal{L}v(t+s, X_{s-}, \theta_{s-}) \\ &+ f(t+s, X_{s-}, \theta_{s-}, v(t+s, X_{s-}, \theta_{s-}), v(t+s, \cdot, 0) - v(t+s, X_{s-}, \theta_{s-})) = 0, \end{aligned}$$

for almost $s \in [0, T-t]$. In particular, $\mathbb{P}^{x,\vartheta}$-a.s.,

$$\begin{aligned} &v(t+s, X_{s-}, \theta_{s-}) \\ &= v(t, x, \vartheta) + \int_0^s \int_E (v(t+r, y, 0) - v(t+r, X_{r-}, \theta_{r-}))\, q(dr, dy) \\ &- \int_0^s f(t+r, X_r, \theta_r, v(t+s, X_s, \theta_s), v(t+r, y, 0) - v(t+r, X_r, \theta_r))\, dr, \end{aligned}$$

for $s \in [0, T-t]$. Since $v(T, x, \vartheta) = g(x, \vartheta)$ for all $(x, \vartheta) \in S$, by simple computations we can prove that, $\forall s \in [0, T-t]$,

$$\begin{aligned} &v(t+s, X_{s-}, \theta_{s-}) + \int_s^{T-t} \int_E (v(t+r, y, 0) - v(t+r, X_{r-}, \theta_{r-}))\, q(dr, dy) \\ &= g(X_{T-t}, \theta_{T-t}) \\ &+ \int_s^{T-t} f(t+r, X_r, a_r, v(t+r, X_r, \theta_r), v(t+r, y, 0) - v(t+r, X_r, \theta_r))\, dr. \end{aligned}$$

Since the pairs $(Y^{x,\vartheta}_{s,t}, Z^{x,\vartheta}_{s,t}(\cdot))_{s\in[0,T-t]}$ and $(v(t+s, X_{s-}, \theta_{s-})\ , v(t+s, y, 0) - v(t+s, X_{s-}, \theta_{s-}))_{s\in[0,T-t]}$ are both solutions to the same BSDE under $\mathbb{P}^{x,\vartheta}$, they coincide as members of the space $\mathbb{M}^{x,\vartheta}$. It follows that equalities (1.40) and (1.41) hold. In particular, $v(t, x, \vartheta) = Y^{x,\vartheta}_{0,t}$, and this yields the uniqueness of the solution.

Existence. We proceed by an approximation argument, following the same lines of the proof of Theorem 4.4 in [**28**]. We recall that, by Theorem 1.2.5, for every fixed $t \in [0, T]$, the BSDE (1.39) has a unique solution $(Y^{x,\vartheta}_{s,t}, Z^{x,\vartheta}_{s,t}(\cdot))_{s\in[0,T-t]}$ for every $(x, \vartheta) \in S$; moreover, $Y^{x,\vartheta}_{0,t}$ is deterministic, i.e., there exists a real number, denoted by $v(t, x, \vartheta)$, such that $\mathbb{P}^{x,\vartheta}(Y^{x,\vartheta}_{0,t} = v(t, x, \vartheta)) = 1$. At this point, we set $f^n = (f \wedge n) \vee (-n)$ and $g^n = (g \wedge n) \vee (-n)$ as the truncations of f and g at level n. By Lemma 1.4.1, for $t \in [0, T]$, $(x, \vartheta) \in S$, equation

$$\begin{aligned} v^n(t, x, \vartheta) &= g^n(x, \vartheta + T - t) + \int_t^T \mathcal{L}v^n(s, x, \vartheta + s - t)\, ds \qquad (1.42) \\ &+ \int_t^T f^n(s, x, \vartheta + s - t, v^n(s, x, \vartheta + s - t), v^n(s, \cdot, 0) - v^n(s, x, \vartheta + s - t))\, ds. \end{aligned}$$

admits a unique bounded measurable solution v^n. In particular, the first part of the proof yield the following identifications:

$$v^n(t, x, \vartheta) = Y^{x,\vartheta,n}_{0,t},$$

$$
\begin{aligned}
v^n(t+s, X_{s-}, \theta_{s-}) &= Y_{s,t}^{x,\vartheta,n},\\
v^n(t+s, y, 0) - v^n(t+s, X_{s-}, \theta_{s-}) &= Z_{s,t}^{x,\vartheta,n}(y),
\end{aligned}
$$

in the sense of Remark 1.4.8, where $(Y_{s,t}^{x,\vartheta,n}, Z_{s,t}^{x,\vartheta,n}(\cdot))_{s\in[0,\,T-t]}$ is the unique solution to the BSDE

$$
\begin{aligned}
&Y_{s,t}^{x,\vartheta,n} + \int_s^{T-t}\int_E Z_{r,t}^{x,\vartheta,n}(y)\, q(dr\, dy)\\
&\quad = g^n(X_{T-t}, \theta_{T-t}) + \int_s^{T-t} f^n\left(t+r, X_r, \theta_r, Y_{r,t}^{x,\vartheta,n}, Z_{r,t}^{x,\vartheta,n}(\cdot)\right) dr,
\end{aligned}
$$

for all $s \in [0,\, T-t]$. Recalling (1.39) and applying Theorem 1.2.5, we deduce that, for some constant c,

$$
\begin{aligned}
&\sup_{s\in[0,\,T-t]} \mathbb{E}^{x,\vartheta}\left[|Y_{s,t}^{x,\vartheta} - Y_{s,t}^{x,\vartheta,n}|^2\right] + \mathbb{E}^{x,\vartheta}\left[\int_0^{T-t} |Y_{s,t}^{x,\vartheta} - Y_{s,t}^{x,\vartheta,n}|^2 ds\right]\\
&+ \mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\int_E |Z_{s,t}^{x,\vartheta}(y) - Z_{s,t}^{x,\vartheta,n}(y)|^2 \lambda(X_s,\theta_s)\, Q(X_s,\theta_s,dy)\, ds\right]\\
&\leqslant c\mathbb{E}^{x,\vartheta}\left[|g(X_{T-t},\theta_{T-t}) - g^n(X_{T-t},\theta_{T-t})|^2\right]\\
&+ c\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t} |f(t+s, X_s, \theta_s, Y_{s,t}^{x,\vartheta}, Z_{s,t}^{x,\vartheta}(\cdot)) - f^n(t+s, X_s, \theta_s, Y_{s,t}^{x,\vartheta}, Z_{s,t}^{x,\vartheta}(\cdot))|^2 ds\right]\\
&\longrightarrow 0,
\end{aligned}
\tag{1.43}
$$

where the two final terms tend to zero by monotone convergence. In particular (1.43) yields

$$
|v(t,x,\vartheta) - v^n(t,x,\vartheta)|^2 = |Y_{0,t}^{x,\vartheta} - Y_{0,t}^{x,\vartheta,n}|^2 \leqslant \sup_{s\in[0,\,T-t]} \mathbb{E}^{x,\vartheta}\left[|Y_{s,t}^{x,\vartheta} - Y_{s,t}^{x,\vartheta,n}|^2\right] \longrightarrow 0,
$$

and therefore v is a measurable function. At this point, applying the Fatou Lemma we get

$$
\begin{aligned}
&\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t} \left|Y_{s,t}^{x,\vartheta} - v(t+s, X_s, \theta_s)\right|^2 ds\right]\\
&+ \mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\int_E \left|Z_{s,t}^{x,\vartheta}(y) - v(t+s,y,0) + v(t+s,X_s,\theta_s)\right|^2 \lambda(X_s,\theta_s)\, Q(X_s,\theta_s,dy)\, ds\right]\\
&\leqslant \liminf_{n\to\infty} \mathbb{E}^{x,\vartheta}\left[\int_0^{T-t} \left|Y_{s,t}^{x,\vartheta} - v^n(t+s, X_s, \theta_s)\right|^2 ds\right]\\
&+ \liminf_{n\to\infty} \mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\int_E \left|Z_{s,t}^{x,\vartheta}(y) - v^n(t+s,y,0) + v^n(t+s,X_s,\theta_s)\right|^2 \lambda(X_s,\theta_s)\, Q(X_s,\theta_s,dy)\, ds\right]\\
&= \liminf_{n\to\infty} \mathbb{E}^{x,\vartheta}\left[\int_0^{T-t} \left|Y_{s,t}^{x,\vartheta} - Y_{s,t}^{x,\vartheta,n}\right|^2 ds\right]\\
&+ \liminf_{n\to\infty} \mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\int_E \left|Z_{s,t}^{x,\vartheta}(y) - Z_{s,t}^{x,\vartheta,n}(y)\right|^2 \lambda(X_s,\theta_s)\, Q(X_s,\theta_s,dy)\, ds\right] = 0
\end{aligned}
$$

by (1.43). The above calculations show that (1.40) and (1.41) hold. Moreover, they imply that

$$
\begin{aligned}
&\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t} |v(t+s,X_s,\theta_s)|^2\,ds\right] \\
&+ \mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\int_E |v(t+s,y,0)-v(t+s,X_s,\theta_s)|^2\,\lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds\right] \\
&= \mathbb{E}^{x,\vartheta}\left[\int_0^{T-t} \left|Y^{x,\vartheta}_{s,t}\right|^2 ds\right] \\
&+ \mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\int_E \left|Z^{x,\vartheta}_{s,t}(y)\right|^2 \lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds\right] \\
&< \infty,
\end{aligned}
$$

that accords to requirement of Definition 1.4.4.

It remains to show that v satisfies (1.33). This would follow from a passage to the limit in (1.42), provided we show that

$$\int_t^T \mathcal{L}v^n(s,x,\vartheta+s-t)\,ds \to \int_t^T \mathcal{L}v(s,x,\vartheta+s-t)\,ds, \tag{1.44}$$

and

$$
\begin{aligned}
&\int_t^T f^n(s,x,\vartheta+s-t,v^n(s,x,\vartheta+s-t),v^n(s,\cdot,0)-v^n(s,x,\vartheta+s-t))\,ds \\
&\to \int_t^T f(s,x,\vartheta+s-t,v(s,x,\vartheta+s-t),v(s,\cdot,0)-v(s,x,\vartheta+s-t))\,ds.
\end{aligned} \tag{1.45}
$$

To prove (1.44), we observe that

$$
\begin{aligned}
&\mathbb{E}^{x,\vartheta}\left|\int_0^{T-t}\mathcal{L}v(t+s,X_{s-},\theta_{s-})\,ds - \int_0^{T-t}\mathcal{L}v^n(t+s,X_{s-},\theta_{s-})\,ds\right| \\
&= \mathbb{E}^{x,\vartheta}\left|\int_0^{T-t}\int_E (Z^{x,\vartheta}_{s,t}-Z^{x,\vartheta,n}_{s,t})\,\lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds\right| \\
&\leqslant (T-t)^{1/2}\sup_{(x,\vartheta)}\sqrt{\lambda(x,\vartheta)}\cdot \\
&\quad\cdot\left(\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\int_E \left|Z^{x,\vartheta}_{s,t}-Z^{x,\vartheta,n}_{s,t}\right|\lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds\right]\right)^{1/2} \to 0,
\end{aligned}
$$

by (1.43). Then, for a subsequence (still denoted v^n) we get

$$\int_0^{T-t}\mathcal{L}v^n(t+s,X_s,\theta_s)\,ds \to \int_0^{T-t}\mathcal{L}v(t+s,X_s,\theta_s)\,ds, \quad \mathbb{P}^{x,\vartheta}\text{-a.s.}$$

Recalling the law (1.7) of the first jump T_1, we see that the set $\{\omega\in\Omega:\,T_1(\omega)>T\}$ has positive $\mathbb{P}^{x,\vartheta}$ probability, and on this set we have $X_{s-}(\omega)=x$, $\theta_{s-}(\omega)=\vartheta+s$. Choosing such an ω we have

$$\int_0^{T-t}\mathcal{L}v^n(t+s,x,\vartheta+s)ds \to \int_0^{T-t}\mathcal{L}v(t+s,x,\vartheta+s)ds,$$

i.e., by a translation of t in the temporal line,

$$\int_t^T \mathcal{L}v^n(s,x,\vartheta+s-t)ds \to \int_t^T \mathcal{L}v(s,x,\vartheta+s-t)ds.$$

To show (1.45), we compute

$$\begin{aligned}
&\mathbb{E}^{x,\vartheta}\left[\left|\int_0^{T-t} f(t+s,X_s,\theta_s,Y_{s,t}^{x,\vartheta},Z_{s,t}^{x,\vartheta}) - f^n(t+s,X_s,\theta_s,Y_{s,t}^{x,\vartheta,n},Z_{s,t}^{x,\vartheta,n}))\,ds\right|\right] \\
&\leqslant \mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\left|f(t+s,X_s,\theta_s,Y_{s,t}^{x,\vartheta},Z_{s,t}^{x,\vartheta}) - f^n(t+s,X_s,\theta_s,Y_{s,t}^{x,\vartheta},Z_{s,t}^{x,\vartheta})\right|\,ds\right] \\
&+ \mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\left|f^n(t+s,X_s,\theta_s,Y_{s,t}^{x,\vartheta},Z_{s,t}^{x,\vartheta}) - f^n(t+s,X_s,\theta_s,Y_{s,t}^{x,\vartheta,n},Z_{s,t}^{x,\vartheta,n})\right|\,ds\right].
\end{aligned}$$

The first integral term in the right-hand side tends to zero by monotone convergence. At this point, we notice that f^n is a truncation of f, and therefore it satisfies the Lipschitz condition (1.13) with the same constants L, L', independent of n. This yields the following estimate for the second integral:

$$\begin{aligned}
&L'\,\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\left|Y_{s,t}^{x,\vartheta}-Y_{s,t}^{x,\vartheta,n}\right|\,ds\right] \\
&+ L\,\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\Big(\int_E\left|Z_{s,t}^{x,\vartheta}(y)-Z_{s,t}^{x,\vartheta,n}(y)\right|^2\lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\Big)^{1/2}ds\right] \\
&\leqslant L'\left((T-t)\,\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\left|Y_{s,t}^{x,\vartheta}-Y_{s,t}^{x,\vartheta,n}\right|^2\,ds\right]\right)^{1/2} \\
&+ L\left((T-t)\,\mathbb{E}^{x,\vartheta}\left[\int_0^{T-t}\int_E\left|Z_{s,t}^{x,\vartheta}(y)-Z_{s,t}^{x,\vartheta,n}(y)\right|^2\lambda(X_s,\theta_s)\,Q(X_s,\theta_s,dy)\,ds\right]\right)^{1/2},
\end{aligned}$$

which tends to zero, again by (1.43). Considering a subsequence (still denoted v^n) we get, $\mathbb{P}^{x,\vartheta}$-a.s.,

$$\begin{aligned}
&\int_0^{T-t} f^n(t+s,X_s,\theta_s,v^n(t+s,X_s,\theta_s),v^n(t+s,y,0)-v^n(t+s,X_s,\theta_s))\,ds \\
&\quad\to \int_0^{T-t} f(t+s,X_s,\theta_s,v(t+s,X_s,\theta_s),v(t+s,y,0)-v(t+s,X_s,\theta_s))\,ds.
\end{aligned}$$

Choosing also in this case an ω in the set $\{\omega\in\Omega:\,T_1(\omega)>T\}$, we find

$$\begin{aligned}
&\int_0^{T-t} f^n(t+s,x,\vartheta+s,v^n(t+s,x,\vartheta+s),v^n(t+s,y,0)-v^n(t+s,x,\vartheta+s))\,ds \\
&\to \int_0^{T-t} f(t+s,x,\vartheta+s,v(t+s,x,\vartheta+s),v(t+s,y,0)-v(t+s,x,\vartheta+s))\,ds,
\end{aligned}$$

and a change of temporal variable allows to prove that (1.33) holds, and to conclude the proof. □

We finally introduce the Hamilton-Jacobi-Bellman (HJB) equation associated to the control problem considered in Section 1.3: for every $t \in [0, T]$ and $(x, \vartheta) \in S$,

$$\begin{aligned} v(t, x, \vartheta) = g(x, \vartheta + T - t) + \int_t^T \mathcal{L}v(s, x, \vartheta + s - t)\, ds \\ + \int_t^T f(s, x, \vartheta + s - t, v(s, \cdot, 0) - v(s, x, \vartheta + s - t))\, ds, \end{aligned} \tag{1.46}$$

where $\mathcal{L}$ denotes the operator introduced in (1.32), f is the Hamiltonian function defined by (1.23) and g is the terminal cost. Since (1.46) is a nonlinear Kolmogorov equation of the form (1.33), we can apply Theorem 1.4.7 and conclude that the value function and an optimal control law can be represented by means of the HJB solution $v(t, x, \vartheta)$.

Corollary 1.4.9. *Let Hypotheses 1.3.1 and 1.3.5 hold. For every fixed $t \in [0, T]$, for every $(x, \vartheta) \in S$ and $s \in [0, T - t]$, there exists a unique solution v to the HJB equation* (1.46)*, satisfying*

$$\begin{aligned} v(t + s, X_{s-}, \theta_{s-}) &= Y^{x,\vartheta}_{s,t}, \\ v(t + s, y, 0) - v(t + s, X_{s-}, \theta_{s-}) &= Z^{x,\vartheta}_{s,t}(y), \end{aligned}$$

where the above equalities are understood as explained in Remark 1.4.8. In particular an optimal control is given by the formula

$$u^{*\,t,x,\vartheta}_s \in \Gamma(t + s, X_{s-}, \theta_{s-}, v(t + s, \cdot, 0) - v(t + s, X_{s-}, \theta_{s-})),$$

while the value function coincides with $v(t, x, \vartheta)$, i.e.

$$J(t, x, \vartheta, u^{*\,t,x,\vartheta}(\cdot)) = v(t, x, \vartheta) = Y^{x,\vartheta}_{0,t}.$$

Chapter 2

Constrained BSDEs representation of the value function for optimal control of pure jump Markov processes

2.1. Introduction

In this chapter we prove that the value function in a classical optimal control problem for pure jump Markov processes can be represented by means of an appropriate backward stochastic differential equation, that we introduce and for which we prove an existence and uniqueness result.

We start by describing our setting in an informal way. A pure jump Markov process X in a general measurable state space $(E, \mathcal{E})$ can be described by means of a rate transition measure, or intensity measure, $\nu(t, x, B)$ defined for $t \geq 0$, $x \in E$, $B \in \mathcal{E}$. The process starts at time $t \geq 0$ from some initial point $x \in E$ and stays there up to a random time T_1 such that

$$\mathbb{P}(T_1 > s) = \exp\left(-\int_t^s \nu(r, x, E)\, dr\right), \qquad s \geq t.$$

At time T_1, the process jumps to a new point X_{T_1} chosen with probability (conditionally to T_1) $\nu(T_1, x, \cdot)/\nu(T_1, x, E)$ and then it stays again at X_{T_1} up to another random time T_2 such that

$$\mathbb{P}(T_2 > s \mid T_1, X_{T_1}) = \exp\left(-\int_{T_1}^s \nu(r, X_{T_1}, E)\, dr\right), \qquad s \geq T_1,$$

and so on.

A controlled pure jump Markov process is obtained starting from a rate measure $\lambda(x, a, B)$ defined for $x \in E$, $a \in A$, $B \in \mathcal{E}$, i.e., depending on a control parameter a taking values in a measurable space of control actions $(A, \mathcal{A})$. A natural way to control a Markov process is to choose a feedback control law, which is a measurable function $\alpha : [0, \infty) \times E \to A$. $\alpha(t, x) \in A$ is the control action selected at time t if the system is in state x. The controlled Markov process X is simply the one corresponding to the rate transition measure $\lambda(x, \alpha(t, x), B)$. Let us denote by $\mathbb{P}_\alpha^{t,x}$ the corresponding law, where t, x are the initial time and starting point.

We note that an alternative construction of (controlled or uncontrolled) Markov processes consists in defining them as solutions to stochastic equations driven by some noise (for instance, by a Poisson process) and with appropriate coefficients depending on a control process. In the context of pure jump processes, our approach based on the introduction of the controlled rate measure $\lambda(x, a, B)$ often leads to more general results and it is more natural in several contexts.

In the classical finite horizon control problem one seeks to maximize over all control laws α a functional of the form

$$J(t, x, \alpha) = \mathbb{E}_\alpha^{t,x} \left[\int_t^T f(s,\ X_s, \alpha(s, X_s))\, ds + g(X_T) \right], \tag{2.1}$$

where a deterministic finite horizon $T > 0$ is given and f, g are given real functions, defined on $[0, T] \times E \times A$ and E, representing the running cost and the terminal cost, respectively. The value function of the control problem is defined in the usual way:

$$V(t, x) = \sup_\alpha J(t, x, \alpha), \qquad t \in [0, T],\ x \in E. \tag{2.2}$$

We will only consider the case when the controlled rate measure λ and the costs f, g are bounded. Then, under some technical assumptions, V is known to be the unique solution on $[0, T] \times E$ to the Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{cases} -\frac{\partial v}{\partial t}(t, x) & = & \sup_{a \in A} \left(\int_E (v(t, y) - v(t, x))\, \lambda(x, a, dy) + f(t, x, a) \right), \\ v(T, x) & = & g(x), \end{cases} \tag{2.3}$$

and if the supremum is attained at some $\alpha(t, x) \in A$ depending measurably on (t, x) then α is an optimal feedback law. Note that the right-hand side of (2.3) is an integral operator: this allows for easy notions of solutions to the HJB equation, that do not in particular need the use of the theory of viscosity solutions.

Our purpose is to relate the value function $V(t, x)$ to an appropriate BSDE. We wish to extend to our framework the theory developed in the context of classical optimal control for diffusion processes, constructed as solutions to stochastic differential equations of Ito type driven by Browian motion, where representation formulae for the solution to the HJB equation exist and are often called non-linear Feyman-Kac formulae. The majority of those results requires that only the drift coefficient of the stochastic equation depends on the control parameter, so that in this case the HJB equation is a second-order semi-linear partial differential equation and the non-linear Feyman-Kac formula is well known, see e.g. El Karoui, Peng and Quenez [**53**]. Generally, in this case the laws of the corresponding controlled processes are all absolutely continuous with respect to the law of a given, uncontrolled process, so that they form a dominated model.

A natural extension to our framework could be obtained imposing conditions implying that the set of probability laws $\{\mathbb{P}_\alpha^{t,x}\}_\alpha$, when α varies over all feedback laws, is a dominated model. This is the point of view taken in Confortola and Fuhrman [**28**], where an appropriate BSDE is introduced and solved and a Feyman-Kac formula for the value function is proved in a restricted framework. This approach is also considered in Chapter 1 in the case of controlled semi-Markov processes and in Confortola and Fuhrman [**27**] in a non-Markovian context.

In the present chapter we want to consider the general case when $\{\mathbb{P}_\alpha^{t,x}\}_\alpha$ is not a dominated model. Even for finite state space E, by a proper choice of the measure $\lambda(x, a, B)$ it is easy to formulate quite natural control problems for which this is the case.

In the context of controlled diffusions, probabilistic formulae for the value function for non-dominated models have been discovered only in recent years. We note that in this case the HJB equation is a fully non-linear partial differential equation. To our knowledge, there are only a few available techniques. One possibility is to use the theory of second-order BSDEs, see for instance Cheridito, Soner, Touzi and Victoir [**23**], and Soner, Touzi and Zhang [**124**]. Another possibility relies on the use of the theory of G-expectations, see e.g. Peng [**105**]. Both theories have been largely developed by several authors. In this chapter we rather follow another approach which is presented in the paper Kharroubi and Pham [**88**] and was predated by similar results concerning optimal switching or optimal impuse control problems, see Elie and Kharroubi [**54**], [**55**], [**56**], Kharroubi, Ma, Pham and Zhang [**87**], and followed by some extensions and applications, see Fuhrman and Pham [**67**], Cosso and Choukroun [**25**], and Cosso, Fuhrman and Pham [**31**]. It consists in a *control randomization method* (not to be confused with the use of relaxed controls) which can be described informally as follows, in our framework of controlled pure jump Markov processes.

We note that for any choice of a feedback law α the pair of stochastic processes $(X_s, \alpha(s, X_s))$ represents the state trajectory and the associated control process. In a first step, for any initial time $t \geq 0$ and starting point $x \in E$, we replace it by an (uncontrolled) Markovian pair of pure jump stochastic processes (X_s, I_s), possibly constructed on a different probability space, in such a way that the process I is a Poisson process with values in the space of control actions A with an intensity measure $\lambda_0(da)$ which is arbitrary but finite and with full support. Next we formulate an auxiliary optimal control problem where we control the intensity of the process I: for any predictable, bounded and positive random field $\nu_t(a)$, by means of a theorem of Girsanov type we construct a probability measure $\mathbb{P}_\nu$ under which the compensator of I is the random measure $\nu_t(a)\,\lambda_0(da)\,dt$ (under $\mathbb{P}_\nu$ the law of X also changes) and then we maximize the functional

$$\mathbb{E}_\nu\left[g(X_T) + \int_t^T f(s,\, X_s,\, I_s)\,ds\right],$$

over all possible choices of the process ν. Following the terminology of [**88**], this will be called the *dual* control problem. Its value function, denoted $V^*(t,x,a)$, also depends *a priori* on the starting point $a \in A$ of the process I (in fact we should write $\mathbb{P}_\nu^{t,x,a}$ instead of $\mathbb{P}_\nu$, but in this discussion we drop this dependence for simplicity) and the family $\{\mathbb{P}_\nu\}_\nu$ is a dominated model. As in [**88**] we are able to show that the value functions for the original problem and the dual one are the same: $V(t,x) = V^*(t,x,a)$, so that the latter does not in fact depend on a. In particular we have replaced the original control problem by a dual one that corresponds to a dominated model and has the same value function. Moreover, we can introduce a well-posed BSDE that represents $V^*(t,x,a)$ (and hence $V(t,x)$). It is an equation on the time interval $[t,T]$ of the form

$$\begin{aligned} Y_s &= g(X_T) + \int_s^T f(r, X_r, I_r)\, dr + K_T - K_s \\ &\quad - \int_s^T \int_{E\times A} Z_r(y,\, b)\, q(dr\, dy\, db) - \int_s^T \int_A Z_r(X_r,\, b)\, \lambda_0(db)\, dr, \end{aligned} \tag{2.4}$$

with unknown triple (Y,Z,K) (depending also on (t,x,a)), where q is the compensated random measure associated to (X,I), Z is a predictable random field and K a predictable increasing càdlàg process, where we additionally add the sign constraint

$$Z_s(X_{s-}, b) \leqslant 0. \tag{2.5}$$

It turns out that this equation has a unique minimal solution, in an appropriate sense, and that the value of the process Y at the initial time represents both the original and the dual value function:

$$Y_t = V(t,x) = V^*(t,x,a). \tag{2.6}$$

This is the desired BSDE representation of the value function for the original control problem and a Feyman-Kac formula for the general HJB equation (2.3).

The chapter is organized as follows. Section 2.2 is essentially devoted to lay down a setting where the classical optimal control problem (2.2) is solved by means of the corresponding HJB equation (2.3). We first recall the general construction of a Markov process given its rate transition measure. Having in mind to apply techniques based on BSDEs driven by random measures we need to work in a canonical setting and use a specific filtration, see Remark 2.2.2. Therefore the construction we present is based on the well-posedness of the martingale problem for multivariate (marked) point processes studied in Jacod [**75**] and it is exposed in detail. This general construction is then used to formulate in a precise way the optimal control problem for the jump Markov process and it is used again in the subsequent section when we define the pair (X,I) mentioned above. Still in Section 2.2, we present classical results on existence and uniqueness of the solution to the HJB equation (2.3) and its identification with the value function v. These results are similar to those in Pliska [**108**], a place where we could find a clear and complete exposition of all the basic theory and to which we refer for further references and related results. We note that the compactness of the space of control actions A, together with suitable upper-semicontinuity conditions of the coefficients of the control problem, is one of the standard assumptions needed to ensure the existence of an optimal control, which is usually constructed by means of an appropriate measurable selection theorem. Since our main aim was only to find a representation formula for the value function we wished to avoid the compactness condition. This was made possible by the use of a different measurable selection result, that however requires lower-semicontinuity conditions. Although this is not usual in the context of maximization problems, this turned out to be the right condition that allows to dispense with compactness assumptions and to prove well-posedness of the HJB equation and a verification theorem. A small variation of the proofs recovers the classical results in [**108**], and even with slightly weaker assumptions: see Remark 2.2.12 for a more detailed comparison.

In Section 2.3 we start to develop the control randomization method: we introduce the auxiliary process (X, I) and formulate the dual control problem under appropriate conditions. Finding the correct formulation required some efforts; in particular we could not mimic the approach of previous works on control randomization mentioned above, since we are not dealing with processes defined as solutions to stochastic equations.

In Section 2.4 we introduce the constrained BSDE (2.4)-(2.5) and we prove, under suitable conditions, that it has a unique minimal solution (Y, Z, K) in a certain class of processes. Moreover, the value of Y at the initial time coincides with the value function of the dual optimal control problem. This is the content of the first of our main results, Theorem 2.4.3. The proof relies on a penalization approach and a monotonic passage to the limit, and combines BSDE techniques with control-theoretic arguments: for instance, a "penalized" dual control problem is also introduced in order to obtain certain uniform upper bounds. In [**88**], in the context of diffusion processes, a more general result is proved, in the sense that the generator f may also depend on (Y, Z); similar generalizations are possible in our context as well, but they seem less motivated and in any case they are not needed for the applications to optimal control.

Finally, in Section 2.5 we prove the second of our main results, Theorem 2.5.1. It states that the initial value of the process Y in (2.4)-(2.5) coincides with the value function $v(t, x)$. As a consequence, the value function is the same for the original optimal control problem and for the dual one and we have the non-linear Feynman-Kac formula (2.6).

The assumptions in Theorem 2.5.1 are fairly general: the state space E and the control action space A are Borel spaces, the controlled kernel λ is bounded and has the Feller property, and the cost functions f, g are continuous and bounded. No compactness assumption is required. When E is finite or countable we have the special case of (continuous-time) controlled Markov chains. A large class of optimization problems for controlled Markovian queues falls under the scope of our result.

In recent years there has been much interest in numerical approximation of the value function in optimal control of Markov processes, see for instance the book Guo and Hérnandez-Lerma [**72**] in the discrete state case. The Feynman-Kac formula

(2.6) can be used to design algorithms based on numerical approximation of the solution to the constrained BSDE (2.4)-(2.5). Numerical schemes for this kind of equations have been proposed and analyzed in the context of diffusion processes, see Kharroubi, Langrené and Pham [**86**], [**85**]. We hope that the results in the present chapter may be used as a foundation for similar methods in the context of pure jump processes as well.

2.2. Pure jump controlled Markov processes

2.2.1. The construction of a jump Markov process given the rate transition measure. Let E be a Borel space, i.e., a topological space homeomorphic to a Borel subset of a compact metric space (some authors call it a Lusin space); in particular, E could be a Polish space. Let $\mathcal{E}$ denote the corresponding Borel σ-algebra.

We will often need to construct a Markov process in E with a given (time dependent) rate transition measure, or intensity measure, denoted by ν. With this terminology we mean that $B \mapsto \nu(t, x, B)$ is a nonnegative measure on $(E, \mathcal{E})$ for every $(t, x) \in [0, \infty) \times E$ and $(t, x) \mapsto \nu(t, x, B)$ is a Borel measurable function on $[0, \infty) \times E$ for every $B \in \mathcal{E}$. We assume that

$$\sup_{t \geq 0,\, x \in E} \nu(t, x, E) < \infty. \tag{2.7}$$

We recall the main steps in the construction of the corresponding Markov process. We note that (2.7) allows to construct a non-explosive process. Since ν depends on time the process will not be time-homogeneous in general. Although the existence of such a process is a well known fact, we need special care in the choice of the corresponding filtration, since this will be crucial when we solve associated BSDEs and implicitly apply a version of the martingale representation theorem in the sections that follow: see also Remark 2.2.2 below. So in the following we will use an explicit construction that we are going to describe. Many of the techniques we are going to use are borrowed from the theory of multivariate (marked) point processes. We will often follow [**75**], but we also refer the reader to the treatise Brandt and Last [**17**] for a more systematic exposition.

We start by constructing a suitable sample space to describe the jumping mechanism of the Markov process. Let Ω' denote the set of sequences $\omega' = (t_n, e_n)_{n \geq 1}$ in $((0, \infty) \times E) \cup \{(\infty, \Delta)\}$, where $\Delta \notin E$ is adjoined to E as an isolated point, satisfying in addition

$$t_n \leq t_{n+1}; \qquad t_n < \infty \implies t_n < t_{n+1}. \tag{2.8}$$

To describe the initial condition we will use the measurable space $(E, \mathcal{E})$. Finally, the sample space for the Markov process will be $\Omega = E \times \Omega'$. We define canonical functions $T_n : \Omega \to (0, \infty]$, $E_n : \Omega \to E \cup \{\Delta\}$ as follows: writing $\omega = (e, \omega')$ in the form $\omega = (e, t_1, e_1, t_2, e_2, \ldots)$ we set for $t \geq 0$ and for $n \geq 1$

$$T_n(\omega) = t_n, \qquad E_n(\omega) = e_n, \qquad T_\infty(\omega) = \lim_{n \to \infty} t_n, \qquad T_0(\omega) = 0, \qquad E_0(\omega) = e.$$

We also define $X : \Omega \times [0, \infty) \to E \cup \{\Delta\}$ setting

$$X_t = \begin{cases} 1_{[0,T_1]}(t)\, E_0 + \sum_{n\geq 1} 1_{(T_n,T_{n+1}]}(t)\, E_n & \text{for } t < T_\infty, \\ \Delta & \text{for } t \geq T_\infty. \end{cases}$$

$X_t = 1_{[0,T_1]}(t)\, E_0 + \sum_{n\geq 1} 1_{(T_n,T_{n+1}]}(t)\, E_n$ for $t < T_\infty$, $X_t = \Delta$ for $t \geq T_\infty$.

In Ω we introduce for all $t \geq 0$ the σ-algebras $\mathcal{G}_t = \sigma(N(s, A) : s \in (0, t], A \in \mathcal{E})$, i.e. generated by the counting processes defined as $N(s, A) = \sum_{n\geq 1} 1_{T_n \leq s} 1_{E_n \in A}$.

To take into account the initial condition we also introduce the filtration $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$, where $\mathcal{F}_0 = \mathcal{E} \otimes \{\emptyset, \Omega'\}$, and for all $t \geq 0$ $\mathcal{F}_t$ is the σ-algebra generated by $\mathcal{F}_0$ and $\mathcal{G}_t$. $\mathbb{F}$ is right-continuous and will be called the natural filtration. In the following all concepts of measurability for stochastic processes (adaptedness, predictability etc.) refer to $\mathbb{F}$. We denote by $\mathcal{F}_\infty$ the σ-algebra generated by all the σ-algebras $\mathcal{F}_t$. The symbol $\mathcal{P}$ denotes the σ-algebra of $\mathbb{F}$-predictable subsets of $[0, \infty) \times \Omega$.

The initial distribution of the process X will be described by a probability measure μ on $(E, \mathcal{E})$. Since $\mathcal{F}_0 = \{A \times \Omega' : A \in \mathcal{E}\}$ is isomorphic to $\mathcal{E}$, μ will be identified with a probability measure on $\mathcal{F}_0$, denoted by the same symbol (by abuse of notation) and such that $\mu(A \times \Omega') = \mu(A)$.

On the filtered sample space $(\Omega, \mathbb{F})$ we have so far introduced the canonical marked point process $(T_n, E_n)_{n\geq 1}$. The corresponding random measure p is, for any $\omega \in \Omega$, a σ-finite measure on $((0, \infty) \times E, \mathcal{B}((0, \infty)) \otimes \mathcal{E})$ defined as

$$p(\omega, ds\, dy) = \sum_{n\geq 1} 1_{T_n(\omega)<\infty}\, \delta_{(T_n(\omega),E_n(\omega))}(ds\, dy),$$

where δ_k denotes the Dirac measure at point $k \in (0, \infty) \times E$.

Now let ν denote a time-dependent rate transition measure as before, satisfying (2.7). We need to introduce the corresponding generator and transition semigroup as follows. We denote by $B_b(E)$ the space of $\mathcal{E}$-measurable bounded real functions on E and for $\phi \in B_b(E)$ we set

$$\mathcal{L}_t\phi(x) = \int_E (\phi(y) - \phi(x))\, \nu(t, x, dy), \qquad t \geq 0,\ x \in E.$$

For any $T \in (0, \infty)$ and $g \in B_b(E)$ we consider the Kolmogorov equation on $[0, T]\times E$:

$$\begin{cases} \frac{\partial v}{\partial s}(s, x) + \mathcal{L}_s v(s, x) = 0, \\ v(T, x) = g(x). \end{cases} \tag{2.9}$$

It is easily proved that there exists a unique measurable bounded function $v : [0, T]\times E$ such that $v(T, \cdot) = g$ on E and, for all $x \in E$, $s \mapsto v(s, x)$ is an absolutely continuous map on $[0, T]$ and the first equation in (2.9) holds for almost all $s \in [0, T]$ with respect to the Lebesgue measure. To verify this we first write (2.9) in the equivalent integral form

$$v(s, x) = g(x) + \int_s^T \mathcal{L}_r v(r, x)\, dr, \qquad s \in [0, T],\ x \in E.$$

Then, noting the inequality $|\mathcal{L}_t\phi(x)| \le 2\sup_{y\in E}|\phi(y)|\,\sup_{t\in[0,T],y\in E}\nu(t,y,E)$, a solution to the latter equation can be obtained by a standard fixed point argument in the space of bounded measurable real functions on $[0,T]\times E$ endowed with the supremum norm.

This allows to define the transition operator $P_{sT}: B_b(E)\to B_b(E)$, for $0\le s\le T$, letting $P_{sT}[g](x)=v(s,x)$, where v is the solution to (2.9) with terminal condition $g\in\mathcal{B}_b(E)$.

Proposition 2.2.1. *Let (2.7) hold and let us fix $t\in[0,\infty)$ and a probability measure μ on $(E,\mathcal{E})$.*

1. *There exists a unique probability measure on $(\Omega,\mathcal{F}_\infty)$, denoted by $\mathbb{P}^{t,\mu}$, such that its restriction to $\mathcal{F}_0$ is μ and the $\mathbb{F}$-compensator (or dual predictable projection) of the measure p under $\mathbb{P}^{t,\mu}$ is the random measure $\tilde p(ds\,dy) := 1_{[t,T_\infty)}(s)\,\nu(s,X_{s-},dy)\,ds$. Moreover, $\mathbb{P}^{t,\mu}(T_\infty=\infty)=1$.*
2. *In the probability space $\{\Omega,\mathcal{F}_\infty,\mathbb{P}^{t,\mu}\}$ the process X has distribution μ at time t and it is Markov on the time interval $[t,\infty)$ with respect to $\mathbb{F}$ with transition operator P_{sT}: explicitly, for every $t\le s\le T$ and for every $g\in\mathcal{B}_b(E)$,*
$$\mathbb{E}^{t,\mu}\left[g(X_T)\mid\mathcal{F}_s\right]=P_{sT}[g](X_s),\qquad \mathbb{P}^{t,\mu}-a.s.$$

Proof. Point 1 follows from a direct application of [**75**], Theorem 3.6. The non-explosion condition $\mathbb{P}^{t,\mu}(T_\infty=\infty)=1$ follows from the fact that λ is bounded.

To prove point 2 we denote $v(s,x)=P_{sT}[g](x)$ the solution to the Kolmogorov equation (2.9) and note that

$$v(T,X_T)-v(s,X_s)=\int_s^T\frac{\partial v}{\partial r}(r,X_r)\,dr+\int_{(s,T]}\int_E(v(r,y)-v(r,X_{r-}))\,p(dr\,dy).$$

This identity is easily proved taking into account that X is constant among jump times and using the definition of the random measure p. Recalling the form of the $\mathbb{F}$-compensator $\tilde p$ of p under $\mathbb{P}^{t,\mu}$ we have, $\mathbb{P}^{t,\mu}$-a.s.,

$$\begin{aligned}
&\mathbb{E}^{t,\mu}\Big[\textstyle\int_{(s,T]}\int_E(v(r,y)-v(r,X_{r-}))\,p(dr\,dy)\mid\mathcal{F}_s\Big]\\
&\quad=\mathbb{E}^{t,\mu}\Big[\textstyle\int_{(s,T]}\int_E(v(r,y)-v(r,X_{r-}))\,\tilde p(dr\,dy)\mid\mathcal{F}_s\Big]\\
&\quad=\mathbb{E}^{t,\mu}\Big[\textstyle\int_{(s,T]}\int_E(v(r,y)-v(r,X_r))\,\nu(r,X_r,dy)\,dr\mid\mathcal{F}_s\Big]\\
&\quad=\mathbb{E}^{t,\mu}\Big[\textstyle\int_{(s,T]}\mathcal{L}_r v(r,X_r)\,dr\mid\mathcal{F}_s\Big]
\end{aligned}$$

and we finally obtain

$$\begin{aligned}
&\mathbb{E}^{t,\mu}\left[g(X_T)\mid\mathcal{F}_s\right]-P_{sT}[g](X_s)=\mathbb{E}^{t,\mu}[v(T,X_T)\mid\mathcal{F}_s]-v(s,X_s)\\
&\quad=\mathbb{E}^{t,\mu}\Big[\textstyle\int_s^T\Big(\frac{\partial v}{\partial r}(r,X_r)+\mathcal{L}_r v(r,X_r)\Big)\,dr\mid\mathcal{F}_s\Big]=0.
\end{aligned}$$

□

In the following we will mainly consider initial distributions μ concentrated at some point $x\in E$, i.e. $\mu=\delta_x$. In this case we use the notation $\mathbb{P}^{t,x}$ rather than $\mathbb{P}^{t,\delta_x}$. Note that, $\mathbb{P}^{t,x}$-a.s., we have $T_1>t$ and therefore $X_s=x$ for all $s\in[0,t]$.

Remark 2.2.2. Since the process X is $\mathbb{F}$-adapted, its natural filtration $\mathbb{F}^X = (\mathcal{F}_t^X)_{t\geq 0}$ defined by $\mathcal{F}_t^X = \sigma(X_s : s \in [0,t])$ is smaller than $\mathbb{F}$. The inclusion may be strict, and may remain such if we consider the corresponding completed filtrations. The reason is that the random variables E_n and E_{n+1} introduced above may coincide on a set of positive probability, for some n, and therefore knowledge of a trajectory of X does not allow to reconstruct the trajectory (T_n, E_n).

In order to have $\mathcal{F}_s = \mathcal{F}_s^X$ up to $\mathbb{P}^{t,\mu}$-null sets one could require that $\nu(t, x, \{x\}) = 0$, i.e. that T_n are in fact jump times of X, but this would impose unnecessary restrictions in some constructs that follow.

Clearly, the Markov property with respect to $\mathbb{F}$ implies the Markov property with respect to $\mathbb{F}^X$ as well.

2.2.2. Optimal control of pure jump Markov processes. In this section we formulate and solve an optimal control problem for a Markov process with a state space E, which is still assumed to be a Borel space with its Borel σ-algebra $\mathcal{E}$. The other data of the problem will be another Borel space A, endowed with its Borel σ-algebra $\mathcal{A}$ and called the space of control actions; a finite time horizon, i.e. a (deterministic) element $T \in (0,\infty)$; two real valued functions f and g, defined on $[0,T] \times E \times A$ and E and called running and terminal cost functions respectively; and finally a measure transition kernel λ from $(E \times A, \mathcal{E} \otimes \mathcal{A})$ to $(E, \mathcal{E})$: namely $B \mapsto \lambda(x,a,B)$ is a nonnegative measure on $(E,\mathcal{E})$ for every $(x,a) \in E \times A$ and $(x,a) \mapsto \lambda(x,a,B)$ is a Borel measurable function for every $B \in \mathcal{E}$. We assume that λ satisfies the following condition:

$$\sup_{x\in E, a\in A} \lambda(x,a,E) < \infty. \tag{2.10}$$

The requirement that $\lambda(x,a,\{x\}) = 0$ for all $x \in E$ and $a \in A$ is natural in many applications, but it is not needed. The kernel λ depending on the control parameter $a \in A$ plays the role of a controlled intensity measure for a controlled Markov process. Roughly speaking, we may control the dynamics of the process by changing its jump intensity dynamically. For a more precise definition, we first construct Ω, $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$, $\mathcal{F}_\infty$ as in the previous paragraph. Then we introduce the class of admissible control laws $\mathcal{A}_{ad}$ as the set of all Borel-measurable maps $\alpha : [0,T] \times E \to A$. To any such α we associate the rate transition measure $\nu^\alpha(t,x,dy) := \lambda(x, \alpha(t,x), dy)$.

For every starting time $t \in [0,T]$ and starting point $x \in E$, and for each $\alpha \in \mathcal{A}_{ad}$, we construct as in the previous paragraph the probability measure on $(\Omega, \mathcal{F}_\infty)$, that will be denoted $\mathbb{P}_\alpha^{t,x}$, corresponding to t, to the initial distribution concentrated at x and to the the rate transition measure ν^α. According to Proposition 2.2.1, under $\mathbb{P}_\alpha^{t,x}$ the process X is Markov with respect to $\mathbb{F}$ and satisfies $X_s = x$ for every $s \in [0,T]$; moreover, the restriction of the measure p to $(t,\infty) \times E$ admits the compensator $\lambda(X_{s-}, \alpha(s, X_{s-}), dy)\, ds$. Denoting by $\mathbb{E}_\alpha^{t,x}$ the expectation under $\mathbb{P}_\alpha^{t,x}$ we finally define, for $t \in [0,\, T]$, $x \in E$ and $\alpha \in \mathcal{A}_{ad}$, the gain functional

$$J(t,x,\alpha) = \mathbb{E}_\alpha^{t,x}\left[\int_t^T f(s,\, X_s, \alpha(s,X_s))\, ds + g(X_T)\right], \tag{2.11}$$

and the value function of the control problem

$$V(t,x) = \sup_{\alpha \in \mathcal{A}_{ad}} J(t,x,\alpha). \tag{2.12}$$

Since we will assume below that f and g are at least Borel-measurable and bounded, both J and V are well defined and bounded.

Remark 2.2.3. In this formulation the only control strategies that we consider are control laws of feedback type, i.e., the control action $\alpha(t,x)$ at time t only depends on t and on the state x for the controlled system at the same time. This is a natural and frequently adopted formulation. Different formulations are possible, but usually the corresponding value function is the same and, if an optimal control exists, it is of feedback type.

Remark 2.2.4. All the results that follows admit natural extensions to slightly more general cases. For instance, λ might depend on time, or the set of admissible control actions may depend on the present state (so admissible control laws should satisfy $\alpha(t,x) \in A(x)$, where $A(x)$ is a given subset of A) provided appropriate measurability conditions are satisfied. We limit ourselves to the previous setting in order to simplify the notation.

Let us consider the Hamilton-Jacobi-Bellman equation (for short, HJB equation) related to the optimal control problem: this is the following nonlinear integro-differential equation on $[0,T] \times E$:

$$-\frac{\partial v}{\partial t}(t,x) = \sup_{a \in A} \left(\mathcal{L}_E^a v(t,x) + f(t,x,a) \right), \tag{2.13}$$

$$v(T,x) = g(x), \tag{2.14}$$

where the operator $\mathcal{L}_E^a$ is defined by

$$\mathcal{L}_E^a \phi(x) = \int_E (\phi(y) - \phi(x))\, \lambda(x,a,dy) \tag{2.15}$$

for all $(t,x,a) \in [0,\, T] \times E \times A$ and every bounded Borel-measurable function $\phi : E \to \mathbb{R}$.

Definition 2.2.5. *We say that a Borel-measurable bounded function $v : [0,T] \times E \to \mathbb{R}$ is a solution to the HJB equation if the right-hand side of* (2.13) *is Borel-measurable and, for every $x \in E$,* (2.14) *holds, the map $t \mapsto v(t,x)$ is absolutely continuous in $[0,T]$ and* (2.13) *holds almost everywhere on $[0,T]$ (the null set of points where it possibly fails may depend on x).*

In the analysis of the HJB equation and the control problem we will use the following function spaces, defined for any metric space S:

(1) $C_b(S) = \{\phi : S \to \mathbb{R}$ continuous and bounded$\}$,

(2) $LSC_b(S) = \{\phi : S \to \mathbb{R}$ lower semi-continuous and bounded$\}$.

(3) $USC_b(S) = \{\phi : S \to \mathbb{R}$ upper semi-continuous and bounded$\}$.

$C_b(S)$, equipped with the supremum norm $\|\phi\|_\infty$, is a Banach space. $LSC_b(S)$ and $USC_b(S)$ are closed subsets of $C_b(S)$, hence complete metric spaces with the induced distance.

In the sequel we need the following classical selection theorem. For a proof we refer for instance to Bertsekas and Shreve [**15**], Propositions 7.33 and 7.34, where a more general statement can also be found.

Proposition 2.2.6. *Let U be a metric space, V a metric separable space. For $F : U \times V \to \mathbb{R}$ set*

$$F^*(u) = \sup_{v \in V} F(u,v), \qquad u \in U.$$

(1) *If $F \in USC_b(U \times V)$ and V is compact then $F^* \in USC_b(U)$ and there exists a Borel-measurable $\phi : U \to V$ such that*

$$F(u, \phi(u)) = F^*(u), \qquad u \in U.$$

(2) *If $F \in LSC_b(U \times V)$ then $F^* \in LSC_b(U)$ and for every $\epsilon > 0$ there exists a Borel-measurable $\phi_\epsilon : U \to V$ such that*

$$F(u, \phi_\epsilon(u)) \geq F^*(u) - \epsilon, \qquad u \in U.$$

Next we present a well-posedness result and a verification theorem for the HJB equation in the space $LSC_b([0,T] \times E)$, Theorems 2.2.7 and 2.2.10 below. The use of lower semi-continuous bounded functions was already commented in the introduction and will be useful for the results in Section 2.5. A small variation of our arguments also yields corresponding results in the class of upper semi-continuous functions, which are more natural when dealing with a maximization problem, see Theorems 2.2.8 and 2.2.11 that slightly generalize classical results. We first formulate the assumptions we need.

$$\lambda \text{ is a Feller transition kernel.} \tag{2.16}$$

We recall that this means that for every $\phi \in C_b(E)$ the function $(x,a) \to \int_E \phi(y)\, \lambda(x,a,dy)$ is continuous (hence it belongs to $C_b(E \times A)$ by (2.10)).

Next we will assume either that

$$f \in LSC_b([0,T] \times E \times A), \quad g \in LSC_b(E), \tag{2.17}$$

or

$$f \in USC_b([0,T] \times E \times A), \quad g \in USC_b(E) \text{ and } A \text{ is a compact metric space.} \tag{2.18}$$

Theorem 2.2.7. *Under the assumptions* (2.10), (2.16), (2.17) *there exists a unique solution $v \in LSC_b([0,T] \times E)$ to the HJB equation (in the sense of Definition 2.2.5).*

Proof. We first make a change of unknown function setting $\tilde{v}(t,x) = e^{-\Lambda t} v(t,x)$, where $\Lambda := \sup_{x \in E, a \in A} \lambda(x,a,E)$ is finite by (2.10). It is immediate to check that v is a solution to (2.13)-(2.14) if and only if $\tilde{v}$ is a solution to

$$-\frac{\partial \tilde{v}}{\partial t}(t,x) = \sup_{a \in A} \left(\mathcal{L}_E^a \tilde{v}(t,x) + e^{-\Lambda t} f(t,x,a) + \Lambda \tilde{v}(t,x) \right) \tag{2.19}$$

$$= \sup_{a\in A}\left(\int_E \tilde{v}(t,y)\,\lambda(x,a,dy) + (\Lambda - \lambda(x,a,E))\tilde{v}(t,x) + e^{-\Lambda t} f(t,x,a)\right),$$

$$\tilde{v}(T,x) = e^{-\Lambda T} g(x). \tag{2.20}$$

The notion of solution we adopt for (2.19)-(2.20) is completely analogous to Definition 2.2.5 and need not be repeated. We set $\Gamma_{\tilde{v}}(t,x) := \int_t^T \sup_{a\in A}\gamma_{\tilde{v}}(s,x,a)\,ds$ where

$$\gamma_{\tilde{v}}(t,x,a) := \int_E \tilde{v}(t,y)\,\lambda(x,a,dy) + (\Lambda - \lambda(x,a,E))\tilde{v}(t,x) + e^{-\Lambda t} f(t,x,a) \tag{2.21}$$

and note that solving (2.19)-(2.20) is equivalent to finding $\tilde{v} \in LSC_b([0,T]\times E)$ satisfying

$$\tilde{v}(t,x) = g(x) + \Gamma_{\tilde{v}}(t,x), \qquad t\in[0,T],\ x\in E.$$

We will prove that $\tilde{v} \mapsto g + \Gamma_{\tilde{v}}$ is a well defined map of $LSC_b([0,T]\times E)$ into itself and it has a unique fixed point, which is therefore the required solution.

Fix $\tilde{v} \in LSC_b([0,T]\times E)$. It follows easily from (2.10) that $\gamma_{\tilde{v}}$ is bounded and, if $\sup_{a\in A}\gamma_{\tilde{v}}(\cdot,\cdot,a)$ is Borel-measurable, $\Gamma_{\tilde{v}}$ is bounded as well. Next we prove that $\gamma_{\tilde{v}}$ and $\Gamma_{\tilde{v}}$ are lower semi-continuous. Note that $(x,a)\mapsto \Lambda - \lambda(x,a,E)$ continuous and *nonnegative* (this is the reason why we introduced the equation for $\tilde{v}$), so

$$(t,x,a)\mapsto (\Lambda - \lambda(x,a,E))\tilde{v}(t,x) + e^{-\Lambda t} f(t,x,a)$$

is in $LSC_b([0,T]\times E\times A)$. Since λ is Feller, it is known that the map

$$(t,x,a)\mapsto \int_E \tilde{v}(t,y)\,\lambda(x,a,dy) \tag{2.22}$$

is continuous when $\tilde{v} \in C_b([0,T]\times E)$ (see [**15**], Proposition 7.30). For general $\tilde{v} \in LSC_b([0,T]\times E)$, there exists a uniformly bounded and increasing sequence $\tilde{v}_n \in C_b([0,T]\times E)$ such that $\tilde{v}_n \to \tilde{v}$ pointwise (see [**15**], Lemma 7.14). From the Fatou Lemma we deduce that the map (2.22) is in $LSC_b([0,T]\times E\times A)$ and we conclude that $\gamma_{\tilde{v}} \in LSC_b([0,T]\times E\times A)$ as well. Therefore $\sup_{a\in A}\gamma_{\tilde{v}}(\cdot,\cdot,a)$, which equals the right-hand side of (2.19), is lower semi-continuous and hence Borel-measurable. To prove lower semi-continuity of $\Gamma_{\tilde{v}}$ suppose $(t_n,x_n)\to(t,x)$; then

$$\begin{aligned}\Gamma_{\tilde{v}}(t_n,x_n) - \Gamma_{\tilde{v}}(t,x) &= \int_{t_n}^t \sup_{a\in A}\gamma_{\tilde{v}}(s,x_n,a)\,ds \\ &\quad + \int_t^T (\sup_{a\in A}\gamma_{\tilde{v}}(s,x_n,a) - \sup_{a\in A}\gamma_{\tilde{v}}(s,x,a))\,ds \\ &\geq -|t-t_n|\,\|\gamma_{\tilde{v}}\|_\infty + \int_t^T (\sup_{a\in A}\gamma_{\tilde{v}}(s,x_n,a) - \sup_{a\in A}\gamma_{\tilde{v}}(s,x,a))\,ds.\end{aligned}$$

By the Fatou Lemma

$$\liminf_{n\to\infty}\Gamma_{\tilde{v}}(t_n,x_n) - \Gamma_{\tilde{v}}(t,x) \geq \int_t^T \liminf_{n\to\infty}(\sup_{a\in A}\gamma_{\tilde{v}}(s,x_n,a) - \sup_{a\in A}\gamma_{\tilde{v}}(s,x,a))\,ds \geq 0,$$

where in the last inequality we have used the lower semi-continuity of $\sup_{a\in A}\gamma_{\tilde{v}}(\cdot,\cdot,a)$.

Since we assume that $g \in LSC_b(E)$ we have thus checked that $\tilde{v} \mapsto g + \Gamma_{\tilde{v}}$ maps $LSC_b([0,T] \times E)$ into itself. To prove that it has a unique fixed point we note the easy estimate based on (2.10), valid for every $\tilde{v}', \tilde{v}'' \in LSC_b([0,T] \times E)$:

$$\begin{aligned}
&|\sup_{a\in A} \gamma_{\tilde{v}'}(t,x,a) - \sup_{a\in A} \gamma_{\tilde{v}''}(t,x,a)| \le \sup_{a\in A} |\gamma_{\tilde{v}'}(t,x,a) - \gamma_{\tilde{v}''}(t,x,a)| \\
&\qquad \le \sup_{a\in A} \left(\int_E |\tilde{v}'(t,y) - \tilde{v}''(t,y)|\, \lambda(x,a,dy) + |\tilde{v}'(t,x) - \tilde{v}''(t,x)|\, \lambda(x,a,E) \right) \\
&\qquad \le 2\Lambda \, \|\tilde{v}' - \tilde{v}''\|_\infty.
\end{aligned}$$

By a standard technique one proves that a suitable iteration of the map $\tilde{v} \mapsto g + \Gamma_{\tilde{v}}$ is a contraction with respect to the distance induced by the supremum norm, and hence that map has a unique fixed point. □

Theorem 2.2.8. *Under the assumptions* (2.10), (2.16), (2.18) *there exists a unique solution $v \in USC_b([0,T] \times E)$ to the HJB equation.*

Proof. The proof is almost the same as in the previous Theorem, replacing LSC_b with USC_b with obvious changes. We introduce $\tilde{v}$, $\gamma_{\tilde{v}}$ and $\Gamma_{\tilde{v}}$ as before and we prove in particular that $\gamma_{\tilde{v}} \in USC_b([0,T] \times E \times A)$. The only difference is that we can not immediately conclude that $\sup_{a\in A} \gamma_{\tilde{v}}(\cdot,\cdot,a)$ is upper semi-continuous as well. However, at this point we can apply point 1 of Proposition 2.2.6 choosing $U = [0,T] \times E$, $V = A$ and $F = \gamma_{\tilde{v}}$ and we deduce that in fact $\sup_{a\in A} \gamma_{\tilde{v}}(\cdot,\cdot,a) \in USC_b([0,T] \times E)$. The rest of the proof is the same. □

Corollary 2.2.9. *Under the assumptions* (2.10), (2.16), *if $f \in C_b([0,T] \times E \times A)$, $g \in C_b(E)$ and A is a compact metric space then the solution v to the HJB equation belongs to $C_b([0,T] \times E)$.*

The Corollary follows immediately from the two previous results. We proceed to a verification theorem for the HJB equation.

Theorem 2.2.10. *Under the assumptions* (2.10), (2.16), (2.17) *the unique solution $v \in LSC_b([0,T] \times E)$ to the HJB equation coincides with the value function V.*

Proof. Let us fix $(t,x) \in [0,T] \times E$. As in the proof of Proposition 2.2.1 we have the identity

$$g(X_T) - v(t,X_t) = \int_t^T \frac{\partial v}{\partial r}(r,X_r)\,dr + \int_{(t,T]} \int_E (v(r,y) - v(r,X_{r-}))\,p(dr\,dy),$$

which follows from the absolute continuity of $t \mapsto v(t,x)$, taking into account that X is constant among jump times and using the definition of the random measure p. Given an arbitrary admissible control $\alpha \in \mathcal{A}_{ad}$ we take the expectation with respect to the corresponding probability $\mathbb{P}^{t,x}_\alpha$. Recalling that the compensator under $\mathbb{P}^{t,x}$ is $1_{[t,\infty)}(s)\lambda(X_{s-},\, \alpha(s,X_{s-}),\, dy)\,ds$ we obtain

$$\begin{aligned}
\mathbb{E}^{t,x}_\alpha[g(X_T)] - v(t,X_t) \;&=\; \int_t^T \frac{\partial v}{\partial r}(r,X_r)\,dr \\
&\quad + \int_{(t,T]} \int_E (v(r,y) - v(r,X_{r-}))\,\lambda(X_{r-},\, \alpha(r,X_{r-}),\, dy)\,dr
\end{aligned}$$

$$= \int_t^T \left(\frac{\partial v}{\partial r}(r, X_r) + \mathcal{L}_E^{\alpha(r,X_r)} v(r, X_r) \right) dr.$$

Adding $\mathbb{E}_\alpha^{t,x} \int_t^T f(r,\, X_r, \alpha(r, X_r))\, dr$ to both sides and rearranging terms we obtain

$$v(t,x) = J(t,x,\alpha) - \mathbb{E}_\alpha^{t,x} \int_t^T \left\{ \frac{\partial v}{\partial r}(r, X_r) + \mathcal{L}_E^{\alpha(r,X_r)} v(r, X_r) + f(r,\, X_r, \alpha(r, X_r)) \right\} dr. \tag{2.23}$$

Recalling the HJB equation and taking into account that X has piecewise constant trajectories we conclude that the term in curly brackets $\{\ldots\}$ is nonpositive and therefore we have $v(t,x) \geq J(t,x,\alpha)$ for every admissible control.

Now we recall that in the proof of Theorem 2.2.7 we showed that the function $\gamma_{\tilde{v}}$ defined in (2.21) belongs to $LSC_b([0,T] \times E \times A)$. Therefore the function

$$F(t,x,a) := e^{\Lambda t} \gamma_{\tilde{v}}(t,x,a) = \mathcal{L}_E^a v(t,x) + f(t,x,a) + \Lambda v(t,x)$$

is also lower semi-continuous and bounded. Applying point 2 of Proposition 2.2.6 with $U = [0,T] \times E$ and $V = A$ we see that for every $\epsilon > 0$ there exists a Borel-measurable $\alpha_\epsilon : [0,T] \times E \to A$ such that $F(t,x,\alpha_\epsilon(t,x)) \geq \inf_{a \in A} F(t,x,a) - \epsilon$ for all $t \in [0,T]$, $x \in E$. Taking into account the HJB equation we conclude that for every $x \in E$ we have

$$\mathcal{L}_E^{\alpha_\epsilon(t,x)} v(t,x) + f(t,x,\alpha_\epsilon(t,x)) \geq -\frac{\partial v}{\partial t}(t,x) - \epsilon$$

for almost all $t \in [0,T]$. Noting that α_ϵ is an admissible control and choosing $\alpha = \alpha_\epsilon$ in (2.23) we obtain $v(t,x) \leq J(t,x,\alpha_\epsilon) + \epsilon(T-t)$. Since we know that $v(t,x) \geq J(t,x,\alpha)$ for every $\alpha \in \mathcal{A}_{ad}$ we conclude that v coincides with the value function V. □

Theorem 2.2.11. *If assumptions* (2.10), (2.16), (2.18) *hold, then the unique solution $v \in USC_b([0,T] \times E)$ to the HJB equation coincides with the value function V. Moreover there exists an optimal control α, which is given by any function satisfying*

$$\mathcal{L}_E^{\alpha(t,x)} v(t,x) + f(t,x,\alpha(t,x)) = \sup_{a \in A} \left(\mathcal{L}_E^a v(t,x) + f(t,x,a) \right). \tag{2.24}$$

Proof. We proceed as in the previous proof, but we can now apply point 2 of Proposition 2.2.6 to the function F and deduce that there exists a Borel-measurable $\alpha : [0,T] \times E \to A$ such that (2.24) holds. Any such control α is optimal: in fact we obtain for every $x \in E$,

$$\mathcal{L}_E^{\alpha(t,x)} v(t,x) + f(t,x,\alpha(t,x)) = -\frac{\partial v}{\partial t}(t,x)$$

for almost all $t \in [0,T]$ and so $v(t,x) = J(t,x,\alpha)$. □

Remark 2.2.12. As already mentioned, Theorems 2.2.8 and 2.2.11 are similar to classical results: compare for instance [**108**], Theorems 10, 12, 13, 14. In that paper the author solves the HJB equations by means of a general result on nonlinear semigroups of operators, and for this he requires some more functional-analytic structure, for instance he embeds the set of decision rules into a properly chosen topological vector space. He also has more stringent conditions of the kernel λ, for instance $\lambda(x, a, B)$ should be strictly positive and continuous in (x, a) for each fixed $B \in \mathcal{E}$.

2.3. Control randomization and dual optimal control problem

In this section we start to implement the control randomization method. In the first step, for any initial time $t \geq 0$ and starting point $x \in E$, we construct an (uncontrolled) Markovian pair of pure jump stochastic processes (X, I) with values in $E \times A$, by specifying its rate transition measure Λ as in (2.27) below. Next we formulate an auxiliary optimal control problem where, roughly speaking, we optimize a cost functional by modifying the intensity of the process I over a suitable family. This "dual" control problem will be studied in Section 2.4 by an approach based on BSDEs. In Section 2.5 we will prove that the dual value function coincides with the one introduced in the previous section.

2.3.1. A dual control system. Let E, A be Borel spaces with corresponding Borel σ-algebras $\mathcal{E}$, $\mathcal{A}$ and let λ be a measure transition kernel from $(E \times A, \mathcal{E} \otimes \mathcal{A})$ to $(E, \mathcal{E})$ as before. As another basic datum we suppose we are given a finite measure λ_0 on $(A, \mathcal{A})$ with full topological support, i.e., it is strictly positive on any non-empty open subset of A. Note that since A is metric separable such a measure can always be constructed, for instance supported on a dense discrete subset of A. We still assume (2.10), so we formulate the following assumption:

(Hλ) λ_0 is a finite measure on $(A, \mathcal{A})$ with full topological support and λ satisfies

$$\sup_{x \in E, a \in A} \lambda(x, a, E) < \infty. \tag{2.25}$$

We wish to construct a Markov process as in section 2.2.1, but with state space $E \times A$. Accordingly, let Ω' denote the set of sequences $\omega' = (t_n, e_n, a_n)_{n \geq 1}$ contained in $((0, \infty) \times E \times A) \cup \{(\infty, \Delta, \Delta')\}$, where $\Delta \notin E$ (respectively, $\Delta' \notin A$) is adjoined to E (respectively, to A) as an isolated point, satisfying (2.8) In the sample space $\Omega = E \times A \times \Omega'$ we define $T_n : \Omega \to (0, \infty]$, $E_n : \Omega \to E \cup \{\Delta\}$, $A_n : \Omega \to A \cup \{\Delta'\}$, as follows: writing $\omega = (e, a, \omega')$ in the form $\omega = (e, a, t_1, e_1, t_2, e_2, \ldots)$ we set for $t \geq 0$ and for $n \geq 1$

$$\begin{array}{llll} T_n(\omega) = t_n, & T_\infty(\omega) = \lim_{n \to \infty} t_n, & T_0(\omega) = 0, & \\ E_n(\omega) = e_n, & A_n(\omega) = a_n, & E_0(\omega) = e, & A_0(\omega) = a. \end{array}$$

We also define processes $X : \Omega \times [0, \infty) \to E \cup \{\Delta\}$, $I : \Omega \times [0, \infty) \to A \cup \{\Delta'\}$ setting

$$X_t = 1_{[0,T_1]}(t)\, E_0 + \sum_{n \geq 1} 1_{(T_n, T_{n+1}]}(t)\, E_n, \qquad I_t = 1_{[0,T_1]}(t)\, A_0 + \sum_{n \geq 1} 1_{(T_n, T_{n+1}]}(t)\, A_n,$$

for $t < T_\infty$, $X_t = \Delta$ and $I_t = \Delta'$ for $t \geq T_\infty$.

In Ω we introduce for all $t \geq 0$ the σ-algebras $\mathcal{G}_t = \sigma(N(s, B) : s \in (0, t], B \in \mathcal{E} \otimes \mathcal{A})$ generated by the counting processes $N(s, B) = \sum_{n \geq 1} 1_{T_n \leq s} 1_{(E_n, A_n) \in B}$ and the σ-algebra $\mathcal{F}_t$ generated by $\mathcal{F}_0$ and $\mathcal{G}_t$, where $\mathcal{F}_0 := \mathcal{E} \otimes \mathcal{A} \otimes \{\emptyset, \Omega'\}$. We still denote $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ and $\mathcal{P}$ the corresponding filtration and predictable σ-algebra. By abuse of notation we also denote by the same symbol the trace of $\mathcal{P}$ on subsets of the form $[0, T] \times \Omega$ or $[t, T] \times \Omega$, for deterministic times $0 \le t \le T < \infty$.

The random measure p is now defined on $(0, \infty) \times E \times A$ as

$$p(ds\, dy\, db) = \sum_{n \in \mathbb{N}} 1_{\{T_n < \infty\}}\, \delta_{\{T_n, E_n, A_n\}}(ds\, dy\, db). \tag{2.26}$$

By means of λ and λ_0 satisfying assumption **(Hλ)** we define a (time-independent) rate transition measure on $E \times A$ given by

$$\Lambda(x, a; dy\, db) = \lambda(x, a, dy)\, \delta_a(db) + \lambda_0(db)\, \delta_x(dy). \tag{2.27}$$

and the corresponding generator $\mathcal{L}$:

$$\begin{aligned} \mathcal{L}\varphi(x, a) \;&:=\; \int_{E \times A} (\varphi(y, b) - \varphi(x, a))\, \Lambda(x, a; dy\, db) \qquad\qquad (2.28)\\ &=\; \int_E (\varphi(y, a) - \varphi(x, a))\, \lambda(x, a, dy) \;+\; \int_A (\varphi(x, b) - \varphi(x, a))\, \lambda_0(db), \end{aligned}$$

for all $(x, a) \in E \times A$ and every function $\varphi \in B_b(E \times A)$.

Given any starting time $t \ge 0$ and starting point $(x, a) \in E \times A$, an application of Proposition 2.2.1 provides a probability measure on $(\Omega, \mathcal{F}_\infty)$, denoted by $\mathbb{P}^{t,x,a}$, such that (X, I) is a Markov process on the time interval $[t, \infty)$ with respect to $\mathbb{F}$ with transition probabilities associated to $\mathcal{L}$. Moreover, $\mathbb{P}^{t,x,a}$-a.s., $X_s = x$ and $I_s = a$ for all $s \in [0, t]$. Finally, the restriction of the measure p to $(t, \infty) \times E \times A$ admits as $\mathbb{F}$-compensator under $\mathbb{P}^{t,x,a}$ the random measure

$$\tilde{p}(ds\, dy\, db) := \lambda_0(db)\, \delta_{\{X_{s-}\}}(dy)\, ds + \lambda(X_{s-},\, I_{s-},\, dy)\, \delta_{\{I_{s-}\}}(db)\, ds.$$

We denote $q := p - \tilde{p}$ the compensated martingale measure associated to p.

Remark 2.3.1. Note that $\Lambda(x, a; \{x, a\}) = \lambda_0(\{a\}) + \lambda(x, a, \{x\})$. So even if we assumed that $\lambda(x, a, \{x\}) = 0$, in general the rate measure Λ would not satisfy the corresponding condition $\Lambda(x, a; \{x, a\}) = 0$. We remark that imposing the additional requirement that $\lambda_0(\{a\}) = 0$ is too restrictive since, due to the assumption that λ_0 has full support, it would rule out the important case when the space of control actions A is finite or countable.

2.3.2. The dual optimal control problem. We introduce a dual control problem associated to the process (X, I) and formulated in a weak form. For fixed (t, x, a), it consists in defining a family of probability measures $\{\mathbb{P}^{t,x,a}_\nu,\, \nu \in \mathcal{V}\}$ in the space $(\Omega, \mathcal{F}_\infty)$, all absolutely continuous with respect to $\mathbb{P}^{t,x,a}$, whose effect is to change the stochastic intensity of the process (X, I) (more precisely, under each $\mathbb{P}^{t,x,a}_\nu$ the compensator of the associated point process takes a desired form), with the aim of maximizing a cost depending on f, g. We note that $\{\mathbb{P}^{t,x,a}_\nu,\, \nu \in \mathcal{V}\}$ is a dominated family of probability measures. We proceed with precise definitions.

We still assume that **(Hλ)** holds. Let us define

$$\mathcal{V} \;=\; \{\nu : \Omega \times [0, \infty) \times A \to (0, \infty),\ \mathcal{P} \otimes \mathcal{A}\text{-measurable and bounded}\}.$$

For every $\nu \in \mathcal{V}$, we consider the predictable random measure

$$\tilde{p}^{\nu}(ds\,dy\,db) := \nu_s(b)\,\lambda_0(db)\,\delta_{\{X_{s-}\}}(dy)\,ds + \lambda(X_{s-},\,I_{s-},\,dy)\,\delta_{\{I_{s-}\}}(db)\,ds. \qquad (2.29)$$

Now we fix $t \in [0,T]$, $x \in E$, $a \in A$ and, with the help of a theorem of Girsanov type, we will show how to construct a probability measure on $(\Omega, \mathcal{F}_\infty)$, equivalent to $\mathbb{P}^{t,x,a}$, under which $\tilde{p}^{\nu}$ is the compensator of the measure p on $(0,T] \times E \times A$. By the Radon-Nikodym theorem one can find two nonnegative functions d_1, d_2 defined on $\Omega \times [0,\,\infty) \times E \times A$, measurable with respect to $\mathcal{P} \otimes \mathcal{E} \otimes \mathcal{A}$ such that

$$\begin{aligned}
\lambda_0(db)\,\delta_{\{X_{t-}\}}(dy)\,dt &= d_1(t,y,b)\,\tilde{p}(dt\,dy\,db)\\
\lambda(X_{t-},\,I_{t-},\,dy)\,\delta_{\{I_{t-}\}}(db)\,dt &= d_2(t,y,b)\,\tilde{p}(dt\,dy\,db),\\
d_1(t,y,b) + d_2(t,y,b) &= 1, \qquad \tilde{p}(dt\,dy\,db) - a.e.
\end{aligned}$$

and we have $d\tilde{p}^{\nu} = (\nu\,d_1 + d_2)\,d\tilde{p}$. For any $\nu \in \mathcal{V}$, consider then the Doléans-Dade exponential local martingale L^{ν} defined setting $L^{\nu}_s = 1$ for $s \in [0,t]$ and

$$\begin{aligned}
L^{\nu}_s &= \exp\bigg(\int_t^s \int_{E\times A} \log(\nu_r(b)\,d_1(r,y,b) \,+ d_2(r,y,b))\,p(dr\,dy\,db)\\
&\qquad - \int_t^s \int_A (\nu_r(b) - 1)\lambda_0(db)\,dr\bigg)\\
&= e^{\int_t^s \int_A (1-\nu_r(b))\lambda_0(db)\,dr} \prod_{n\geqslant 1: T_n \leqslant s} (\nu_{T_n}(A_n)\,d_1(T_n,E_n,A_n) + d_2(T_n,E_n,A_n))
\end{aligned}$$

for $s \in [t,T]$. When L^{ν} is a true martingale, i.e., $\mathbb{E}^{t,x,a}\,[L^{\nu}_T] = 1$, we can define a probability measure $\mathbb{P}^{t,x,a}_{\nu}$ equivalent to $\mathbb{P}^{t,x,a}$ on $(\Omega, \mathcal{F}_\infty)$ setting $\mathbb{P}^{t,x,a}_{\nu}(d\omega) = L^{\nu}_T(\omega)\mathbb{P}^{t,x,a}(d\omega)$. By the Girsanov theorem for point processes ([**75**], Theorem 4.5) the restriction of the random measure p to $(0,T] \times E \times A$ admits $\tilde{p}^{\nu} = (\nu\,d_1 + d_2)\,\tilde{p}$ as compensator under $\mathbb{P}^{t,x,a}_{\nu}$. We denote by $\mathbb{E}^{t,x,a}_{\nu}$ the expectation operator under $\mathbb{P}^{t,x,a}_{\nu}$ and by $q^{\nu} := p - \tilde{p}^{\nu}$ the compensated martingale measure of p under $\mathbb{P}^{t,x,a}_{\nu}$. The validity of the condition $\mathbb{E}^{t,x,a}\,[L^{\nu}_T] = 1$ under our assumptions, as well as other useful properties, are proved in the following proposition.

Lemma 2.3.2. *Let assumption* **(Hλ)** *hold. Then, for every $t \in [0,T]$, $x \in E$ and $\nu \in \mathcal{V}$, under the probability $\mathbb{P}^{t,x,a}$ the process L^{ν} is a martingale on $[0,T]$ and L^{ν}_T is square integrable.*

In addition, for every $\mathcal{P} \otimes \mathcal{E} \otimes \mathcal{A}$-measurable function $H : \Omega \times [t,T] \times E \times A \to \mathbb{R}$ such that $\mathbb{E}^{t,x,a}\left[\int_t^T \int_{E\times A} |H_s(y,b)|^2\,\tilde{p}(ds\,dy\,db)\right] < \infty$, the process

$$\int_t^{\cdot} \int_{E\times A} H_s(y,b)\,q^{\nu}(ds\,dy\,db)$$

is a $\mathbb{P}^{t,x,a}_{\nu}$-martingale on $[t,T]$.

Proof. The first part of the proof is inspired by Lemma 4.1 in [**88**]. In particular, since ν is bounded and $\lambda_0(A) < \infty$, we see that

$$S^{\nu}_T = \exp\left(\int_t^T \int_A |\nu_s(b) - 1|^2 \lambda_0(db)\,ds\right)$$

is bounded. Therefore, from Theorem 8, see also Theorem 9, in [**109**], follows the martingale property of L^ν together with its uniform integrability. Concerning the square integrability of L^ν_T, set $\ell(x,\lambda) := 2\ln(x\lambda + 1 - \lambda) - \ln(x^2\lambda + 1 - \lambda)$, for any $x \geq 0$ and $\lambda \in [0,1]$. From the definition of L^ν we have (recalling that $d_2(s,y,b) = 1 - d_1(s,y,b)$)

$$|L^\nu_T|^2 = L^{\nu^2}_T S^\nu_T \exp\left(\int_t^T \int_{E\times A} \ell(\nu_s(b), d_1(s,y,b))\, p(ds\,dy\,db)\right) \leq L^{\nu^2}_T S^\nu_T,$$

where the last inequality follows from the fact that ℓ is nonpositive. This entails that L^ν_T is square integrable.

Finally, let us fix a predictable function H such that

$$\mathbb{E}^{t,x,a}\left[\int_t^T \int_{E\times A} |H_s(y,b)|^2\, \tilde{p}(ds\,dy\,db)\right] < \infty.$$

The process $\int_t^\cdot \int_{E\times A} H_s(y,b)\, q^\nu(ds\,dy\,db)$ is a $\mathbb{P}^{t,x,a}_\nu$-local martingale, and the uniform integrability follows from the Burkholder-Davis-Gundy and Cauchy Schwarz inequalities, together with the square integrability of L^ν_T. □

To complete the formulation of the dual optimal control problem we specify the conditions that we will assume for the cost functions f, g:

(Hfg) $f \in B_b([0,T] \times E \times A)$ and $g \in B_b(E)$.

For every $t \in [0,T]$, $x \in E$, $a \in A$ and $\nu \in \mathcal{V}$ we finally introduce the dual gain functional

$$J(t,x,a,\nu) = \mathbb{E}^{t,x,a}_\nu\left[g(X_T) + \int_t^T f(s,\, X_s,\, I_s)ds\right],$$

and the dual value function

$$V^*(t,x,a) = \sup_{\nu\in\mathcal{V}} J(t,x,a,\nu). \tag{2.30}$$

Remark 2.3.3. Let us denote by $\{S_n\}$ (resp. $\{R_n\}$) the jump times of I (resp. of X), and by $\mu^I(ds\,db) = \sum_n \delta_{(S_n, I_{S_n})}(ds\,db)$ (resp. $\mu^X(ds\,dy) = \sum_n \delta_{(R_n, X_{R_n})}(ds\,dy)$) the corresponding random measure on $(0,\infty)\times A$ (resp. on $(0,\infty)\times E$).

An interpretation of the dual optimal control problem can be given as follows: under $\mathbb{P}^{t,x,a}$,

(i) the times $\{S_n\}$ e $\{R_n\}$ are disjoint;

(ii) the compensators of the random measures $\mu^I(ds\,db)$ and $\mu^X(ds\,dy)$ are

$$\tilde{\mu}^I(ds\,db) = \lambda_0(db)\, 1_{\{b\neq I_{s-}\}}\, ds, \quad \tilde{\mu}^X(ds\,dy) = \lambda(X_{s-}, I_{s-}, dy)\, 1_{\{y\neq X_{s-}\}}\, ds. \tag{2.31}$$

In particular, the effect of choosing ν is to change the intensity of the I-component.

To prove point (i), let us introduce the $\mathcal{P}$-measurable process $H : \Omega\times\mathbb{R}_+\times E\times A \to \mathbb{R}_+$ defined by

$$H_s(\omega,y,b) = (y - X_{s-}(\omega))^2(b - I_{s-}(\omega))^2. \tag{2.32}$$

We have

$$\mathbb{E}^{t,x,a}\left[\sum_n H_{T_n}(X_{T_n}, I_{T_n})\right] = \mathbb{E}^{t,x,a}\left[\int_0^\infty \int_E H_s(y, I_{s-})\,\lambda(X_{s-}, I_{s-}, dy)\,ds\right]$$
$$+\,\mathbb{E}^{t,x,a}\left[\int_0^\infty \int_A H_s(X_{s-}, b)\,\lambda_0(db)\,ds\right]$$
$$= 0.$$

Recalling (2.32), previous equality reads

$$\mathbb{E}^{t,x,a}\left[\sum_n (\Delta X_{T_n})^2\,(\Delta I_{T_n})^2\right] = 0,$$

i.e., for all $n \in \mathbb{N}$,

$$(\Delta X_{T_n})^2\,(\Delta I_{T_n})^2 = 0 \quad \mathbb{P}^{t,x,a}\text{-a.s.}$$

Therefore the jump times of X and I are disjoint.

Let now consider point (ii). Since, by (i), the jump times $\{S_n\}_{n\geqslant 1}$ and $\{R_n\}_{n\geqslant 1}$ are disjoint, for any $\mathbb{F}$-predictable processes $K : \Omega\times\mathbb{R}_+\times E \to \mathbb{R}_+$ and $J : \Omega\times\mathbb{R}_+\times A \to \mathbb{R}_+$, we have

$$\sum_n K_{R_n}(X_{R_n}) = \sum_n \tilde{K}_{T_n}(X_{T_n}), \qquad \sum_n J_{S_n}(I_{S_n}) = \sum_n \tilde{J}_{T_n}(I_{T_n})$$

where

$$\tilde{K}_s(y) = K_s(y)\,1_{\{y\neq X_{s-}\}}, \qquad \tilde{J}_s(b) = J_s(b)\,1_{\{b\neq I_{s-}\}}.$$

In particular, since $\tilde{K}_s(X_{s-}) = 0$ and $\tilde{J}_s(I_{s-}) = 0$ for all $s \in [0, T]$, we get

$$\mathbb{E}^{t,x,a}\left[\sum_n K_{R_n}(X_{R_n})\right] = \mathbb{E}^{t,x,a}\left[\sum_n \tilde{K}_{T_n}(X_{T_n})\right]$$
$$= \mathbb{E}^{t,x,a}\left[\int_0^\infty \int_E \tilde{K}_s(y)\,\lambda(X_{s-}, I_{s-}, dy)\,ds\right]$$
$$= \mathbb{E}^{t,x,a}\left[\int_0^\infty \int_E K_s(y)\,1_{\{y\neq X_{s-}\}}\,\lambda(X_{s-}, I_{s-}, dy)\,ds\right] \quad (2.33)$$

and

$$\mathbb{E}^{t,x,a}\left[\sum_n J_{S_n}(I_{S_n})\right] = \mathbb{E}^{t,x,a}\left[\sum_n \tilde{J}_{T_n}(I_{T_n})\right]$$
$$= \mathbb{E}^{t,x,a}\left[\int_0^\infty \int_A \tilde{J}_s(b)\,\lambda_0(db)\,ds\right]$$
$$= \mathbb{E}^{t,x,a}\left[\int_0^\infty \int_A J_s(b)\,1_{\{b\neq I_{s-}\}}\,\lambda_0(db)\,ds\right]. \quad (2.34)$$

Identities (2.34) and (2.33) show the validity of (2.31) under $\mathbb{P}^{t,x,a}$.

2.4. Constrained BSDE and the dual value function representation

In this section we introduce a BSDE, with a sign constrain on its martingale part, and prove existence and uniqueness of a minimal solution, in an appropriate sense. The BSDE is then used to give a representation formula for the dual value function introduced above.

Throughout this section we assume that the assumptions **(Hλ)** and **(Hfg)** are satisfied and we use the randomized control setting introduced above: $\Omega, \mathbb{F}, X, \mathbb{P}^{t,x,a}$ as well as the random measures $p, \tilde{p}, q$ are the same as in subsection 2.3.1. For any $(t, x, a) \in [0,\, T] \times E \times A$, we introduce the following notation.

- $\mathbf{L^2}(\lambda_0)$, the set of $\mathcal{A}$-measurable maps $\psi : A \to \mathbb{R}$ such that
$$|\psi|^2_{\mathbf{L^2}(\lambda_0)} \quad := \quad \int_A |\psi(b)|^2 \, \lambda_0(db) < \infty.$$
- $\mathbf{L^2_{t,x,a}}(\mathcal{F}_\tau)$, the set of $\mathcal{F}_\tau$-measurable random variable X such that $\mathbb{E}^{t,x,a}\left[|X|^2\right] < \infty$; here τ is an $\mathbb{F}$-stopping time with values in $[t, T]$.
- $\mathbf{S^2_{t,x,a}}$ the set of real valued càdlàg adapted processes $Y = (Y_s)_{t \leqslant s \leqslant T}$ such that
$$||Y||^2_{\mathbf{S^2_{t,x,a}}} := \mathbb{E}^{t,x,a}\left[\sup_{t \leqslant s \leqslant T} |Y_s|^2\right] < \infty.$$
- $\mathbf{L^2_{t,x,a}}(\mathrm{q})$, the set of $\mathcal{P} \otimes \mathcal{E} \otimes \mathcal{A}$-measurable maps $Z : \Omega \times [t, T] \times E \times A \to \mathbb{R}$ such that

$$\begin{aligned}
||Z||^2_{\mathbf{L^2_{t,x,a}}(q)} &:= \mathbb{E}^{t,x,a}\left[\int_t^T \int_{E \times A} |Z_s(y, b)|^2 \, \tilde{p}(ds\, dy\, db)\right] \\
&= \mathbb{E}^{t,x,a}\left[\int_t^T \int_E |Z_s(I_s, y)|^2 \, \lambda(X_s, I_s, dy)\, ds + \int_t^T \int_A |Z_s(X_s, b)|^2 \, \lambda_0(db)\, ds\right] < \infty.
\end{aligned}$$

- $\mathbf{K^2_{t,x,a}}$ the set of nondecreasing predictable processes $K = (K_s)_{t \leqslant s \leqslant T} \in \mathbf{S^2_{t,x,a}}$ with $K_t = 0$, with the induced norm
$$||K||^2_{\mathbf{K^2_{t,x,a}}} = \mathbb{E}^{t,x,a}\left[|K_T|^2\right].$$

We are interested in studying the following family of BSDEs parametrized by (t, x, a): $\mathbb{P}^{t,x,a}$-a.s.,

$$\begin{aligned}
Y_s^{t,x,a} &= g(X_T) + \int_s^T f(r, X_r, I_r)\, dr + K_T^{t,x,a} - K_s^{t,x,a} \\
&\quad - \int_s^T \int_{E \times A} Z_r^{t,x,a}(y,\, b)\, q(dr\, dy\, db) \\
&\quad - \int_s^T \int_A Z_r^{t,x,a}(X_r,\, b)\, \lambda_0(db)\, dr, \qquad s \in [t, T],
\end{aligned} \tag{2.35}$$

with the sign constraint

$$Z_s^{t,x,a}(X_{s-}, b) \leqslant 0, \qquad ds \otimes d\mathbb{P}^{t,x,a} \otimes \lambda_0(db) - \text{a.e. on } [t, T] \times \Omega \times A. \tag{2.36}$$

This constraint can be seen as a sign condition imposed on the jumps of the corresponding stochastic integral.

Definition 2.4.1. A solution to the equation (2.35)-(2.36) is a triple $(Y, Z, K) \in \mathbf{S^2_{t,x,a}} \times \mathbf{L^2_{t,x,a}}(\mathrm{q}) \times \mathbf{K^2_{t,x,a}}$ that satisfies (2.35)-(2.36).

A solution (Y, Z, K) is called minimal if for any other solution $(\tilde{Y}, \tilde{Z}, \tilde{K})$ we have, $\mathbb{P}^{t,x,a}$-a.s.,

$$Y_s \leqslant \tilde{Y}_s, \qquad s \in [t, T].$$

Proposition 2.4.2. *Under assumptions* $(\mathbf{H}\lambda)$ *and* **(Hfg)**, *for any* $(t, x, a) \in [0, T] \times E \times A$, *if there exists a minimal solution on* $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}^{t,x,a})$ *to the BSDE* (2.35)-(2.36), *then it is unique.*

Proof. Let (Y, Z, K) and (Y', Z', K') be two minimal solutions of (2.35)-(2.36). The component Y is unique by definition, and the difference between the two backward equations gives: $\mathbb{P}^{t,x,a}$-a.s.

$$\begin{aligned}
&\int_t^s \int_{E\times A} (Z_r(y,\, b) - Z'_r(y,\, b))\, p(dr\, dy\, db) \\
&= K_s - K'_s + \int_t^s \int_E (Z_r(y,\, I_{r-}) - Z'_r(y,\, I_{r-}))\, \lambda(X_{r-}, I_{r-}dy)\, dr, \quad \forall\, t \leqslant s \leqslant T.
\end{aligned}$$

The right hand is a predictable process, in particular it has no totally inaccessible jumps (see, e.g., Proposition 2.24, Chapter I, in Jacod and Shiryaev [**79**]), while the left side is a pure jump process with totally inaccessible jumps. This implies the uniqueness of the component Z, and as a consequence the component K is unique as well. □

We now state the main result of the section.

Theorem 2.4.3. *Under the assumptions* $(\mathbf{H}\lambda)$ *and* **(Hfg)**, *for all* $(t, x, a) \in [0, T] \times E \times A$ *there exists a unique minimal solution* $Y^{t,x,a}$ *to* (2.35)-(2.36). *Moreover, for all* $s \in [t, T]$, $Y^{t,x,a}_s$ *has the explicit representation:*

$$Y^{t,x,a}_s = \operatorname*{ess\,sup}_{\nu \in \mathcal{V}} \mathbb{E}^\nu \left[g(X_T) + \int_s^T f(r, X_r, I_r)\, dr \,\middle|\, \mathcal{F}_s \right], \quad s \in [t, T]. \tag{2.37}$$

In particular, setting $s = t$, *we have the following representation formula for the value function of the dual control problem:*

$$V^*(t, x, a) = Y^{t,x,a}_t, \qquad (t, x, a) \in [0, T] \times E \times A. \tag{2.38}$$

The rest of this section is devoted to prove Theorem 2.4.3. To this end we will use a penalization approach presented in the following subsections. Here we only note that for the solvability of the BSDE the use of the filtration $\mathbb{F}$ introduced above is essential, since it involves application of martingale representation theorems for multivariate point processes (see e.g. Theorem 5.4 in [**75**]).

2.4.1. Penalized BSDE and associated dual control problem. Let us consider the family of penalized BSDEs associated to (2.35)-(2.36), parametrized by the integer $n \geqslant 1$: $\mathbb{P}^{t,x,a}$-a.s.,

$$Y^{n,t,x,a}_s = g(X_T) + \int_s^T f(r, X_r, I_r)\, dr + K^{n,t,x,a}_T - K^{n,t,x,a}_s$$

$$
\begin{aligned}
&- \int_s^T \int_{E\times A} Z_r^{n,t,x,a}(y,b)\, q(dr\,dy\,db) \\
&- \int_s^T \int_A Z_r^{n,t,x,a}(X_r,b)\,\lambda_0(db)\,dr, \quad s\in[t,T],
\end{aligned} \tag{2.39}
$$

where K^n is the nondecreasing process in $\mathbf{K^2_{t,x,a}}$ defined by

$$
K_s^n = n \int_t^s \int_A [Z_r^n(X_r,b)]^+\,\lambda_0(db)\,dr.
$$

Here we denote by $[u]^+$ the positive part of u. The penalized BSDE (2.39) can be rewritten in the equivalent form: $\mathbb{P}^{t,x,a}$-a.s.,

$$
\begin{aligned}
Y_s^{n,t,x,a} &= g(X_T) + \int_s^T f^n(r,\, X_r,\, I_r,\, Z_r^{n,t,x,a}(X_r,\cdot))\,ds \\
&\qquad - \int_s^T \int_{E\times A} Z_r^{n,t,x,a}(y,b)\,q(dr\,dy\,db), \quad s\in[t,T].
\end{aligned}
$$

where the generator f^n is defined by

$$
f^n(t,x,a,\psi) := f(t,x,a) + \int_A \left\{ n\,[\psi(b)]^+ - \psi(b) \right\}\,\lambda_0(db), \tag{2.40}
$$

for all (t,x,a) in $[0,\,T]\times E\times A$, and $\psi\in\mathbf{L^2}(\lambda_0)$. We note that under **(Hλ)** and **(Hfg)** f^n is Lipschitz continuous in ψ with respect to the norm of $\mathbf{L^2}(\lambda_0)$, uniformly in (t,x,a), i.e., for every $n\in\mathbb{N}$ there exists a constant L_n depending only on n such that for every $(t,x,a)\in[0,\,T]\times E\times A$ and $\psi,\,\psi'\in\mathbf{L^2}(\lambda_0)$,

$$
|f^n(t,x,a,\psi') - f^n(t,x,a,\psi)| \leqslant L_n\,|\psi-\psi'|_{\mathbf{L^2}(\lambda_0)}.
$$

The use of the natural filtration $\mathbb{F}$ allows to use well known integral representation results for $\mathbb{F}$-martingales (see, e.g., Theorem 5.4 in [**75**]) and we have the following proposition, whose proof is standard and is therefore omitted (similar proofs can be found in [**131**] Theorem 3.2, [**12**] Proposition 3.2, [**28**] Theorem 3.4).

Proposition 2.4.4. *Let assumptions* **(Hλ)** *and* **(Hfg)** *hold. For every initial condition $(t,x,a)\in[0,\,T]\times E\times A$, and for every $n\in\mathbb{N}$, there exists a unique solution $(Y_s^{n,t,x,a}, Z_s^{n,t,x,a})_{s\in[t,T]}\in\mathbf{S^2_{t,x,a}}\times\mathbf{L^2_{t,x,a}}(\mathrm{q})$ satisfying the penalized BSDE* (2.39)*.*

Next we show that the solution to the penalized BSDE (2.39) provides an explicit representation of the value function of a corresponding dual control problem depending on n. This is the content of Lemma 2.4.5 which will allow to deduce some estimates uniform with respect to n.

For every $n\geqslant 1$, let $\mathcal{V}^n$ denote the subset of elements $\nu\in\mathcal{V}$ that take values in $(0,\,n]$.

Lemma 2.4.5. *Let assumptions* **(Hλ)** *and* **(Hfg)** *hold. For all $n\geq 1$ and $s\in[t,T]$,*

$$
Y_s^{n,t,x,a} = \operatorname*{ess\,sup}_{\nu\in\mathcal{V}^n} \mathbb{E}^\nu\left[g(X_T) + \int_s^T f(r,X_r,I_r)\,dr \,\middle|\, \mathcal{F}_s \right], \qquad \mathbb{P}^{t,x,a} - a.s. \tag{2.41}
$$

Proof. We fix $n \geq 1$ and for any $\nu \in \mathcal{V}^n$ we introduce the compensated martingale measure $q^\nu(ds\,dy\,db) = q(ds\,dy\,db) - (\nu_s(b) - 1)\,d_1(s,y,b)\,\tilde{p}(ds\,dy\,db)$ under $\mathbb{P}_\nu^{t,x,a}$. We see that the solution (Y^n, Z^n) to the BSDE (2.39) satisfies: $\mathbb{P}^{t,x,a}$-a.s.,

$$\begin{aligned} Y_s^n &= g(X_T) + \int_s^T f(r, X_r, I_r)\,dr + \int_s^T \int_A \{n[Z_r^n(X_r,b)]^+ - \nu_r(b)\,Z_r^n(X_r,b)\}\,\lambda_0(db)\,dr \\ &\quad - \int_s^T \int_{E\times A} Z_r^n(y,b)\,q^\nu(dr\,dy\,db), \qquad s \in [t,T]. \end{aligned} \tag{2.42}$$

By taking conditional expectation in (2.42) under $\mathbb{P}_\nu^{t,x,a}$ and applying Lemma 2.3.2 we get, for any $s \in [t,T]$,

$$\begin{aligned} Y_s^{n,t,x,a} &= \mathbb{E}^\nu\left[g(X_T) + \int_s^T f(r,X_r,I_r)\,dr \,\middle|\, \mathcal{F}_s\right] \qquad (2.43)\\ &\quad + \mathbb{E}^\nu\left[\int_s^T \int_A \{n[Z_r^{n,t,x,a}(X_r,b)]^+ - \nu_r(b)\,Z_r^{n,t,x,a}(X_r,b)\}\,\lambda_0(db)\,dr \,\middle|\, \mathcal{F}_s\right], \end{aligned}$$

$\mathbb{P}_\nu^{t,x,a}$-a.s. From the elementary numerical inequality: $n[u]^+ - \nu u \geqslant 0$ for all $u \in \mathbb{R}$, $\nu \in (0,n]$, we deduce by (2.43) that

$$Y_s^{n,t,x,a} \geqslant \operatorname*{ess\,sup}_{\nu\in\mathcal{V}^n} \mathbb{E}^\nu\left[g(X_T) + \int_s^T f(r,X_r,I_r)\,dr \,\middle|\, \mathcal{F}_s\right]. \tag{2.44}$$

On the other hand, for $\epsilon \in (0,\,1)$, let us consider the process $\nu^\epsilon \in \mathcal{V}^n$ defined by

$$\begin{aligned} \nu_s^\epsilon(b) =& n\,1_{\{Z_s^{n,t,x,a}(X_{s-},b)\geqslant 0\}} + \epsilon\,1_{\{-1<Z_s^{n,t,x,a}(X_{s-},b)<0\}} \\ &- \epsilon\,Z_s^{n,t,x,a}(X_{s-},b)^{-1}\,1_{\{Z_s^{n,t,x,a}(X_{s-},b)\leqslant -1\}}. \end{aligned}$$

By construction, we have

$$n[Z_s^{n,t,x,a}(X_{s-},b)]^+ - \nu_s^\epsilon(b)\,Z_s^{n,t,x,a}(X_{s-},b) \leqslant \epsilon, \qquad s \in [t,T],\ b \in A,$$

and thus for the choice of $\nu = \nu^\epsilon$ in (2.43):

$$\begin{aligned} Y_s^{n,t,x,a} &\leqslant \mathbb{E}_{\nu^\epsilon}^{t,x,a}\left[g(X_T) + \int_s^T f(r,X_r,I_r)\,dr \,\middle|\, \mathcal{F}_s\right] + \epsilon T|\lambda_0(A)| \\ &\leqslant \operatorname*{ess\,sup}_{\nu\in\mathcal{V}^n} \mathbb{E}^\nu\left[g(X_T) + \int_s^T f(r,X_r,I_r)\,dr \,\middle|\, \mathcal{F}_s\right] + \epsilon T|\lambda_0(A)|. \end{aligned}$$

Together with (2.44), this is enough to prove the required representation of Y^n. Note that we could not take $\nu_s(b) = n1_{\{Z_s^n(X_{s-},b)\geqslant 0\}}$, since this process does not belong to $\mathcal{V}^n$ because of the requirement of strict positivity. □

2.4.2. Limit behavior of the penalized BSDEs and conclusion of the proof of Theorem 2.4.3. As a consequence of the representation (2.41) we immediately obtain the following estimates:

Lemma 2.4.6. *Let assumptions* **(Hλ)** *and* **(Hfg)** *hold. There exists a constant C, depending only on T, f, g, such that for any $(t,x,a) \in [0,\,T] \times E \times A$ and $n \geq 1$, $\mathbb{P}^{t,x,a}$-a.s.,*

$$Y_s^{n,t,x,a} \leqslant Y_s^{n+1,t,x,a}, \qquad |Y_s^{n,t,x,a}| \leqslant C, \qquad s \in [t,T].$$

Proof. For fixed $s \in [t, T]$, the almost sure monotonicity of $Y^{n,t,x,a}$ follows from the representation formula (2.41), since by definition $\mathcal{V}^n \subset \mathcal{V}^{n+1}$; moreover, the same formula shows that we can take $C = ||g||_\infty + T\,||f||_\infty$. Finally, these inequalities hold for every $s \in [t, T]$ outside a null set, since the processes $Y^{n,t,x,a}$ are càdlàg. □

Moreover, the following a priori uniform estimate on the sequence $(Y^{n,t,x,a}$, $Z^{n,t,x,a}$, $K^{n,t,x,a})$ holds:

Lemma 2.4.7. *Let assumptions* **(Hλ)** *and* **(Hfg)** *hold. For all $(t, x, a) \in [0, T] \times E \times A$ and $n \in \mathbb{N}$, there exists a positive constant C' depending only on T, f, g such that*

$$||Y^{n,t,x,a}||^2_{\mathbf{S^2_{t,x,a}}} + ||Z^{n,t,x,a}||^2_{\mathbf{L^2_{t,x,a}}(q)} + ||K^{n,t,x,a}||^2_{\mathbf{K^2_{t,x,a}}} \leqslant C'. \tag{2.45}$$

Proof. In the following we omit for simplicity of notation the dependence on (t, x, a) for the triple $(Y^{n,t,x,a}, Z^{n,t,x,a}, K^{n,t,x,a})$. The estimate on Y^n follows immediately from the previous lemma:

$$||Y^n||^2_{\mathbf{S^2_{t,x,a}}} = \mathbb{E}^{t,x,a}\left[\sup_{s\in[t,T]} |Y^n_s|^2\right] \leqslant C^2. \tag{2.46}$$

Next we notice that, since K^n is continuous, the jumps of Y^n are given by the formula

$$\Delta Y^n_s = \int_{E\times A} Z^n_s(y, b)\, p(\{s\}, dy\, db).$$

The Itô formula applied to $|Y^n_t|^2$ gives:

$$\begin{aligned}
d|Y^n_r|^2 &= 2\, Y^n_{r-}\, dY^n_r + |\Delta Y^n_r|^2 \\
&= -2\, Y^n_{r-}\, f(X_{r-}, I_{r-})\, dr - 2\, Y^n_{r-}\, dK^n_r \\
&\quad +2\, Y^n_{r-} \int_{E\times A} Z^n_r(y, b)\, q(dr\, dy\, db) + 2\, Y^n_{r-} \int_A Z^n_r(X_{r-}, b)\, \lambda_0(db)\, dr \\
&\quad + \int_{E\times A} |Z^n_r(y, b)|^2\, p(\{r\}\, dy\, db).
\end{aligned} \tag{2.47}$$

Integrating (2.47) on $[s, T]$, for every $s \in [t, T]$, and recalling the elementary inequality $2ab \leqslant \frac{1}{\delta}a^2 + \delta b^2$ for any constant $\delta > 0$, and that

$$\mathbb{E}^{t,x,a}\left[\int_s^T \int_A |Z^n_r(X_{r-},\, b)|^2\, \lambda_0(db)\, dr\right] \leqslant \mathbb{E}^{t,x,a}\left[\int_s^T \int_{E\times A} |Z^n_r(y,\, b)|^2\, \tilde{p}(dr\, dy\, db)\right], \tag{2.48}$$

we have:

$$\begin{aligned}
&\mathbb{E}^{t,x,a}\left[|Y_s|^2\right] + \mathbb{E}^{t,x,a}\left[\int_s^T \int_{E\times A} |Z^n_r(y,\, b)|^2\, \tilde{p}(dr\, dy\, db)\right] \\
&\quad \leqslant \mathbb{E}^{t,x,a}\left[|g(X_T)|^2\right] \\
&\quad + \frac{1}{\beta}\mathbb{E}^{t,x,a}\left[\int_s^T |f(r, X_r, I_r)|^2\, dr\right] + \beta\mathbb{E}^{t,x,a}\left[\int_s^T |Y^n_r|^2\, dr\right] \\
&\quad + \frac{T\,\lambda_0(A)}{\gamma}\mathbb{E}^{t,x,a}\left[\int_s^T \int_{E\times A} |Z^n_r(y, b)|^2\, \tilde{p}(dr\, dy\, db)\right] + \gamma\mathbb{E}^{t,x,a}\left[\int_s^T |Y^n_r|^2\, dr\right]
\end{aligned}$$

$$+\frac{1}{\alpha}\mathbb{E}^{t,x,a}\left[\sup_{s\in[t,T]}|Y_s^n|^2\right]+\alpha\mathbb{E}^{t,x,a}\left[|K_T^n-K_s^n|^2\right],\quad s\in[t,T], \tag{2.49}$$

for some $\alpha,\beta,\gamma>0$, Now, from the equation (2.39) we obtain:

$$\begin{aligned}K_T^n-K_s^n &= Y_s^n-g(X_T)-\int_s^T f(r,X_r,I_r)dr\\ &\quad+\int_s^T\int_A Z_r^n(X_r,b)\,\lambda_0(db)\,dr\\ &\quad+\int_s^T\int_{E\times A} Z_r^n(y,b)\,q(dr\,dy\,db),\quad s\in[t,T].\end{aligned}$$

Next we note the equality

$$\begin{aligned}&\mathbb{E}^{t,x,a}\left[\left|\int_s^T\int_{E\times A} Z_r^n(y,\,b)\,q(dr\,dy\,db)\right|^2\right]\\ &=\mathbb{E}^{t,x,a}\left[\int_s^T\int_{E\times A}|Z_r^n(y,\,b)|^2\,p(dr\,dy\,db)\right]\\ &=\mathbb{E}^{t,x,a}\left[\int_s^T\int_{E\times A}|Z_r^n(y,\,b)|^2\,\tilde{p}(dr\,dy\,db)\right]\end{aligned}$$

that can be proved applying the Ito formula as before to the square of the martingale $u\mapsto\int_s^u\int_{E\times A}Z_r^n(y,\,b)\,q(dr\,dy\,db)$, $u\in[s,T]$ (or by considering its quadratic variation). Recalling again (2.48) we see that there exists some positive constant B such that

$$\begin{aligned}&\mathbb{E}^{t,x,a}\left[|K_T^n-K_s^n|^2\right]\\ &\leqslant B\Bigg(\mathbb{E}^{t,x,a}\left[|Y_s^n|^2\right]+\mathbb{E}^{t,x,a}\left[|g(X_T)|^2\right]+\mathbb{E}^{t,x,a}\left[\int_s^T|f(r,X_r,I_r)|^2\,dr\right]\\ &+\mathbb{E}^{t,x,a}\left[\int_s^T\int_{E\times A}|Z_r^n(y,b)|^2\,\tilde{p}(dr\,dy\,db)\right]\Bigg),\quad s\in[t,T].\end{aligned} \tag{2.50}$$

Plugging (2.50) into (2.49), and recalling the uniform estimation (2.46) on Y^n, we get

$$\begin{aligned}&(1-\alpha B)\,\mathbb{E}^{t,x,a}\left[|Y_s^n|^2\right]\\ &+\left(1-\left[\alpha B+\frac{T\,\lambda_0(A)}{\gamma}\right]\right)\,\mathbb{E}^{t,x,a}\left[\int_s^T\int_{E\times A}|Z_r^n(y,\,b)|^2\,\tilde{p}(dr\,dy\,db)\right]\\ &\leqslant(1+\alpha B)\,\mathbb{E}^{t,x,a}\left[|g(X_T)|^2\right]+\left(\alpha B+\frac{1}{\beta}\right)\,\mathbb{E}^{t,x,a}\left[\int_s^T|f(r,X_r,I_r)|^2\,dr\right]\\ &+\frac{C^2}{\alpha}+(\gamma+\beta)\,\mathbb{E}^{t,x,a}\left[\int_s^T|Y_r^n|^2\,dr\right],\quad s\in[t,T].\end{aligned}$$

Hence, by choosing $\alpha \in \left(0,\, \frac{1}{B}\right)$, $\gamma > \frac{T\,\lambda_0(A)}{1-\alpha B}$, $\beta > 0$, and applying Gromwall's lemma to $s \to \mathbb{E}^{t,x,a}\left[|Y_s^n|^2\right]$, we obtain:

$$
\begin{aligned}
&\sup_{s\in[t,\,T]} \mathbb{E}^{t,x,a}\left[|Y_s^n|^2\right] + \mathbb{E}^{t,x,a}\left[\int_t^T \int_{E\times A} |Z_s^n(y,\,b)|^2\, \tilde p(ds\, dy\, db)\right] \\
&\leqslant C'\left(\mathbb{E}^{t,x,a}\left[|g(X_T)|^2\right] + \mathbb{E}^{t,x,a}\left[\int_t^T |f(s, X_s, I_s)|^2\, ds\right] + C^2\right),
\end{aligned}
\tag{2.51}
$$

for some $C' > 0$ depending only on T, which gives the required uniform estimate for (Z^n) and also (K^n) by (2.50). □

We can finally present the conclusion of the proof of Theorem 2.4.3:

Proof. Let $(t, x, a) \in [0, T] \times E \times A$. We first show that (Y^n, Z^n, K^n) (we omit the dependence on (t, x, a) for simplicity of notation) solution to (2.39) converges in a suitable way to some process (Y, Z, K) solution to the constrained BSDE (2.35)-(2.36). By Lemma 2.4.6, $(Y^n)_n$ converges increasingly to some adapted process Y, which moreover satisfies $\mathbb{E}^{t,x,a}\left[\sup_{s\in[t,T]} |Y_s|^2\right] < \infty$ by the uniform estimate for $(Y^n)_n$ in Lemma 2.4.7 and Fatou's lemma. Furthermore, by the dominated convergence theorem, we also have $\mathbb{E}\int_0^T |Y_t^n - Y_t|^2 dt \to 0$. Next, we prove that there exists $(Z,\, K) \in \mathbf{L^2_{t,x,a}}(\mathrm{q}) \times \mathbf{K^2_{t,x,a}}$ with K predictable, such that

(i) Z is the weak limit of $(Z^n)_n$ in $\mathbf{L^2_{t,x,a}}(\mathrm{q})$;

(ii) K_τ is the weak limit of $(K_\tau^n)_n$ in $\mathbf{L^2_{t,x,a}}(\mathcal{F}_\tau)$, for any stopping time τ valued in $[t\ T]$;

(iii) $\mathbb{P}^{t,x,a}$-a.s.,

$$
\begin{aligned}
Y_s &= g(X_T) + \int_s^T f(r, X_r, I_r)\, dr + K_T - K_s \\
&- \int_s^T \int_{E\times A} Z_r(y,\, b)\, q(dr\, dy\, db) - \int_s^T \int_A Z_r(X_r,\, b)\, \lambda_0(db)\, dr,\ \ s \in [t, T],
\end{aligned}
$$

with

$$
Z_s(X_{s-}, b) \leqslant 0, \qquad ds \otimes d\mathbb{P}^{t,x,a} \otimes \lambda_0(db) - \text{a.e.}
$$

Let define the following mappings from $\mathbf{L^2_{t,x,a}}(\mathrm{q})$ to $\mathbf{L^2_{t,x,a}}(\mathcal{F}_\tau)$:

$$
\begin{aligned}
I_\tau^1 :&\quad Z \mapsto \int_t^\tau \int_{E\times A} Z_s(y, b)\, q(ds\, dy\, db), \\
I_\tau^2 :&\quad Z \mapsto \int_t^\tau \int_A Z_s(X_s, b)\, \lambda_0(db)\, ds,
\end{aligned}
$$

for each $\mathbb{F}$-stopping time τ with values in $[t, T]$. We wish to prove that $I_\tau^1 Z^n$ and $I_\tau^2 Z^n$ converge weakly in $\mathbf{L^2_{t,x,a}}(\mathcal{F}_\tau)$ to $I_\tau^1 Z$ and $I_\tau^2 Z$ respectively. Indeed, by the uniform estimates for $(Z^n)_n$ in Lemma 2.4.7, there exists a subsequence, denoted $(Z^{n_k})_k$, which converges weakly in $\mathbf{L^2_{t,x,a}}(\mathrm{q})$. Since I_1 and I_2 are linear continuous operators they are also weakly continuous so that we have $I^1_\tau Z^{n_k} \to I^1_\tau Z$ and $I^2_\tau Z^{n_k} \to I^2_\tau Z$ weakly in $\mathbf{L^2_{t,x,a}}(\mathcal{F}_\tau)$ as $k \to \infty$. Since we have from (2.39)

$$\begin{aligned} K^{n_k}_\tau &= -Y^{n_k}_\tau + Y^{n_k}_t - \int_t^\tau f(r, X_r, I_r)\, dr \\ &\quad + \int_t^\tau \int_A Z^{n_k}_r(X_r, b)\, \lambda_0(db)\, dr + \int_t^\tau \int_{E\times A} Z^{n_k}_r(y, b)\, q(dr\, dy\, db), \end{aligned}$$

we also obtain the weak convergence in $\mathbf{L^2_{t,x,a}}(\mathcal{F}_\tau)$ as $k \to \infty$

$$\begin{aligned} K^{n_k}_\tau \rightharpoonup K_\tau &:= -Y_\tau + Y_t - \int_t^\tau f(r, X_r, I_r)\, dr \\ &\quad + \int_t^\tau \int_A Z_r(X_r, b)\, \lambda_0(db)\, dr + \int_t^\tau \int_{E\times A} Z_r(y, b)\, q(dr\, dy\, db). \end{aligned} \tag{2.52}$$

Arguing as in Peng [**104**], proof of Theorem 2.1, Kharroubi, Ma, Pham and Zhang [**87**] Lemma 3.5, Essaky [**60**] Theorem 3.1, we see that K inherits from K^{n_k} the properties of having nondecreasing paths and of being square integrable and predictable. Finally, from Lemma 2.2 in [**104**] it follows that K and Y are càdlàg, so that $K^{t,x,a} \in \mathbf{K^2_{t,x,a}}$ and $Y^{t,x,a} \in \mathbf{S^2_{t,x,a}}$.

Notice that the processes Z and K in (2.52) are uniquely determined. Indeed, if (Z, K) and (Z', K') satisfy (2.52), then the predictable processes Z and Z' coincide at the jump times and can be identified almost surely with respect to $\tilde{p}(\omega, ds\, dy\, db)\mathbb{P}^{t,x,a}(d\omega)$ (a similar argument can be found in the proof of Proposition 2.4.2 to which we refer for more details). Finally, recalling that the jumps of p are totally inaccessible, we also obtain the uniqueness of the component K. The uniqueness of Z and K entails that all the sequences $(Z^n)_n$ and $(K^n)_n$ respectively converge (in the sense of points (i) and (ii) above) to Z and K.

It remains to show that the jump constraint (2.36) is satisfied. To this end, we consider the functional on $\mathbf{L^2_{t,x,a}}(\mathrm{q})$ given by

$$G: \quad Z \mapsto \mathbb{E}^{t,x,a}\left[\int_t^T \int_A [Z_s(X_{s-}, b)]^+\, \lambda_0(db)\, ds\right].$$

From uniform estimate (2.45), we see that $G(Z^n) \to 0$ as $n \to \infty$. Since G is convex and strongly continuous in the strong topology of $\mathbf{L^2_{t,x,a}}(\mathrm{q})$, then G is lower semicontinuous in the weak topology of $\mathbf{L^2_{t,x,a}}(\mathrm{q})$, see, e.g., Corollary 3.9 in Brezis [**19**]. Therefore, we find

$$G(Z) \leqslant \liminf_{n\to\infty} G(Z^n) = 0,$$

from which follows the validity of the jump constraint (2.36) on $[t, T]$. We have then showed that (Y, Z, K) is a solution to the constrained BSDE (2.35)-(2.36). It remains to prove that this is the minimal solution. To this end, fix $n \in \mathbb{N}$ and consider a triple $(\bar{Y}, \bar{Z}, \bar{K}) \in \mathbf{S^2_{t,x,a}} \times \mathbf{L^2_{t,x,a}}(\mathrm{q}) \times \mathbf{K^2_{t,x,a}}$ satisfying (2.35)-(2.36). For any $\nu \in \mathcal{V}^n$, by introducing the compensated martingale measure q^ν, we see that the solution $(\bar{Y}, \bar{Z}, \bar{K})$ satisfies: $\mathbb{P}^{t,x,a}$-a.s.,

$$\bar{Y}_s = g(X_T) + \int_s^T f(r, X_r, I_r)\, dr + \bar{K}_T - \bar{K}_s \tag{2.53}$$

$$- \int_s^T \int_{E \times A} \bar{Z}_r(y,\, b)\, q^\nu(dr\, dy\, db) - \int_s^T \int_A \nu_r(b)\, \bar{Z}_r(X_r,\, b)\, \lambda_0(db)\, dr \quad s \in [t, T].$$

By taking the expectation under $\mathbb{P}_\nu^{t,x,a}$ in (2.53), recalling Lemma 2.3.2, and that $\bar{K}$ is nondecreasing, we have

$$\begin{aligned} \bar{Y}_s \;&\geqslant\; \mathbb{E}^\nu \left[g(X_T) + \int_s^T f(r, X_r, I_r)\, dr \right] - \mathbb{E}^\nu \left[\int_s^T \int_A \nu_r(b)\, \bar{Z}_r(X_r,\, b)\, \lambda_0(db)\, dr \right] \\ &\geqslant\; \mathbb{E}^\nu \left[g(X_T) + \int_s^T f(r, X_r, I_r)\, dr \right] \qquad s \in [t, T], \end{aligned} \tag{2.54}$$

since ν is valued in $(0,\, n]$ and Z satisfies constraint (2.36). As ν is arbitrary in $\mathcal{V}^n$, we get from the representation formula (2.41) that $\bar{Y}_s \geqslant Y_s^n$, $\forall\, s \in [t, T]$, $\forall\, n \in \mathbb{N}$. In particular, $Y_s = \lim_{n\to\infty} Y_s^n \leqslant \bar{Y}_s$, i.e., the minimality property holds. The uniqueness of the minimal solution straightly follows from Proposition 2.4.2.

To conclude the proof, we argue on the limiting behavior of the dual representation for Y^n when n goes to infinity. Since $\mathcal{V}^n \subset \mathcal{V}$, it is clear from the representation (2.41) that, for all n and $s \in [t, T]$,

$$Y_s^n \leqslant \operatorname*{ess\,sup}_{\nu \in \mathcal{V}} \mathbb{E}^\nu \left[g(X_T) + \int_s^T f(r, X_r, I_r)\, dr \,\middle|\, \mathcal{F}_s \right].$$

Moreover, being Y the pointwise limit of Y^n, we deduce that

$$Y_s = \lim_{n\to\infty} Y_s^n \leqslant \operatorname*{ess\,sup}_{\nu \in \mathcal{V}} \mathbb{E}^\nu \left[g(X_T) + \int_s^T f(r, X_r, I_r)\, dr \,\middle|\, \mathcal{F}_s \right]. \tag{2.55}$$

On the other hand, for any $\nu \in \mathcal{V}$, introducing the compensated martingale measure q^ν under $\mathbb{P}^\nu$ as usual, we see that $(Y,\, Z,\, K)$ satisfies

$$\begin{aligned} Y_s = g(X_T) &+ \int_s^T f(r, X_r, I_r)\, dr + K_T - K_s \\ &- \int_s^T \int_{E \times A} Z_r(y,\, b)\, q^\nu(dr\, dy\, db) - \int_s^T \int_A Z_r(X_r,\, b)\, \nu_r(b) \lambda_0(db)\, dr, \quad s \in [t, T]. \end{aligned} \tag{2.56}$$

Arguing in the same way as in (2.54), we obtain

$$Y_s \;\geqslant\; \mathbb{E}^\nu \left[g(X_T) + \int_s^T f(r, X_r, I_r)\, dr \,\middle|\, \mathcal{F}_s \right],$$

so that $Y_s \geqslant \operatorname{ess\,sup}_{\nu \in \mathcal{V}} \mathbb{E}^\nu \left[g(X_T) + \int_s^T f(r, X_r, I_r)\, dr \,\middle|\, \mathcal{F}_s \right]$ by the arbitrariness of $\nu \in \mathcal{V}$. Together with (2.55) this gives the required equality. □

2.5. A BSDE representation for the value function

In this section we conclude the last step in the method of control randomization and we show that the minimal solution to the constrained BSDE (2.35)-(2.36) actually provides a non-linear Feynman-Kac representation of the solution to the Hamilton-Jacobi-Bellman (HJB) equation (2.13)-(2.14), that we re-write here:

$$-\frac{\partial v}{\partial t}(t,x) = \sup_{a\in A}\left(\mathcal{L}_E^a v(t,x) + f(t,x,a)\right), \qquad v(T,x) = g(x).$$

As a consequence of the dual representation in Theorem 2.4.3 it follows that the value function of the original optimal control problem can be identified with the dual one, which in particular turns out to be independent on the variable a.

For our result we need the following conditions:

$$\sup_{x\in E, a\in A} \lambda(x,a,E) < \infty, \tag{2.57}$$

$$\lambda \text{ is a Feller transition kernel,} \tag{2.58}$$

$$f \in C_b([0,T]\times E\times A), \quad g \in C_b(E). \tag{2.59}$$

We note that these assumptions are stronger that those required in Theorem 2.2.7 and therefore they imply that there exists a unique solution $v \in LSC_b([0,T]\times E)$ to the HJB equation in the sense of Definition 2.2.5. If, in addition, A is a compact metric space then $v \in C_b([0,T]\times E)$ by Corollary 2.2.9.

Let us consider again the Markov process (X,I) in $E\times A$ constructed in Section 2.3.1, with corresponding family of probability measures $\mathbb{P}^{t,x,a}$ and generator $\mathcal{L}$ introduced in (2.28). Since (2.57)-(2.59) are also stronger than **(Hλ)** and **(Hfg)**, by Theorem 2.4.3 there exists a unique solution to the BSDE (2.35)-(2.36).

Our main result is as follows:

Theorem 2.5.1. *Assume* (2.57), (2.58), (2.59). *Let v be the unique solution to the Hamilton-Jacobi-Bellman equation provided by Theorem 2.2.7. Then for every $(t,x,a)\in[0,\,T]\times E\times A$,*

$$v(t,x) = Y_t^{t,x,a},$$

where $Y^{t,x,a}$ is the first component of the minimal solution to the constrained BSDE with nonpositive jumps (2.35)-(2.36).

More generally, we have $\mathbb{P}^{t,x,a}$-a.s.,

$$v(s,X_s) = Y_s^{t,x,a}, \qquad s\in[t,T].$$

Finally, for the value function V of the optimal control problem defined in (2.12) *and the dual value function V^* defined in* (2.30) *we have the equalities*

$$V(t,x) = v(t,x) = Y_t^{t,x,a} = V^*(t,x,a).$$

In particular, the latter functions do not depend on a.

The rest of this section is devoted to prove Theorem 2.5.1.

2.5.1. A penalized HJB equation. Let us recall the penalized BSDE associated to (2.35)-(2.36): $\mathbb{P}^{t,x,a}$-a.s.,

$$Y_s^{n,t,x,a} = g(X_T) + \int_s^T f(r,X_r,I_r)\,ds - \int_s^T\int_{E\times A} Z_r^{n,t,x,a}(y,b)\,q(dr\,dy\,db) \tag{2.60}$$

$$+\int_s^T \int_A \left\{n\,[Z_r^{n,t,x,a}(X_r,b)]^+ - Z_r^{n,t,x,a}(X_r,b)\right\}\,\lambda_0(db)\,dr, \qquad s\in[t,T].$$

Let us now consider the parabolic semi-linear penalized integro-differential equation, of HJB type: for any $n\geq 1$,

$$\frac{\partial v^n}{\partial t}(t,x,a)+\int_A\{n\,[v^n(t,x,b)-v^n(t,x,a)]^+ \,-(v^n(t,x,b)-v^n(t,x,a))\}\,\lambda_0(db)$$
$$+\mathcal{L}v^n(t,x,a)+f(t,x,a)=0 \quad \text{on } [0,\,T)\times E\times A, \tag{2.61}$$
$$v^n(T,x,a)=g(x) \quad \text{on } E\times A, \tag{2.62}$$

The following lemma states that the solution of (2.61)-(2.62) can be represented probabilistically by means of the solution to the penalized BSDE (2.60):

Lemma 2.5.2. *Assume* (2.57)*,* (2.58)*,* (2.59)*. Then there exists a unique function* $v^n\in C_b([0,T]\times E\times A)$ *such that* $t\mapsto v^n(t,x,a)$ *is continuously differentiable on* $[0,T]$ *and* (2.61)-(2.62) *hold for every* $(t,x,a)\in[0,\,T)\times E\times A$*.*

Moreover, for every $(t,x,a)\in[0,\,T]\times E\times A$ *and for every* $n\in\mathbb{N}$*,*

$$Y_s^{n,t,x,a} \;=\; v^n(s,X_s,I_s) \tag{2.63}$$
$$Z_s^{n,t,x,a}(y,b) \;=\; v^n(s,y,b)-v^n(s,X_{s-},I_{s-}), \tag{2.64}$$

(to be understood as an equality between elements of the space $\mathbf{S^2_{t,x,a}}\times\mathbf{L^2_{t,x,a}}(\mathrm{q})$*) so that in particular* $v^n(t,x,a)=Y_t^{n,t,x,a}$*.*

Proof. We first note that $v^n\in C_b([0,T]\times E\times A)$ is the required solution if and only if

$$v^n(t,x,a)=g(x)+\int_t^T \mathcal{L}v^n(s,x,a)\,ds+\int_t^T f^n(s,x,a,v^n(s,x,\cdot)-v^n(s,x,a))\tag{2.65}$$

for $t\in[0,,T)$, $x\in E$, $a\in A$, where $f^n(t,x,a,\psi)$ is the map defined in (2.40). We use a fixed point argument, introducing a map Γ from $C_b([0,T]\times E\times A)$ to itself setting $v=\Gamma(w)$ where

$$v(t,x,a)=g(x)+\int_t^T \mathcal{L}w(s,x,a)\,ds+\int_t^T f^n(s,x,a,w(s,x,\cdot)-w(s,x,a))\,ds.$$

Using the boundedness assumptions on λ and λ_0 it can be shown by standard arguments that some iteration of the above map is a contraction in the space of bounded measurable real functions on $[0,\,T]\times E\times A$ endowed with the supremum norm and therefore the map Γ has a unique fixed point, which is the required solution v^n.

We finally prove the identifications (2.63)-(2.64). Since $v^n\in C_b([0,T]\times E\times A)$ we can apply the Itô formula to the process $v(s,X_s,I_s)$, $s\in[t,T]$, obtaining, $\mathbb{P}^{t,x,a}$-a.s.,

$$v^n(s,X_s,I_s)=v^n(t,x,a)+\int_t^s\left(\frac{\partial v^n}{\partial r}(r,X_r,I_r)+\mathcal{L}_r^I v^n(r,X_r,I_r)\right)\,dr$$
$$+\int_t^s\int_{E\times A}(v^n(r,y,b)-v^n(r,X_{r-},I_{r-}))\;q(dr\,dy\,db), \quad s\in[t,T].$$

Taking into account that v^n satisfies (2.61)-(2.62) and that (X, I) has piecewise constant trajectories, we obtain $\mathbb{P}^{t,x,a}$-a.s.,

$$\frac{\partial v^n}{\partial r}(r, X_r, I_r) + \mathcal{L}v^n(r, X_r, I_r) + f^n(r, X_r, I_r, v^n(r, X_r, \cdot) - v^n(r, X_r, I_r)) = 0,$$

for almost all $r \in [t, T]$. It follows that, $\mathbb{P}^{t,x,a}$-a.s.,

$$\begin{aligned} v^n(s, X_s, I_s) &= v^n(t, x, a) - \int_t^s f^n(r, X_r, I_r, v^n(r, X_r, \cdot) - v^n(r, X_r, I_r))\, dr \\ &\quad + \int_t^s \int_{E\times A} (v^n(r, y, b) - v(r, X_{r-}, I_{r-}))\; q(dr\, dy\, db), \quad s \in [t, T]. \end{aligned}$$

Since $v^n(T, x, a) = g(x)$ for all $(x, a) \in E \times A$, simple passages show that

$$\begin{aligned} v^n(s, X_s, I_s) &= g(X_T) + \int_t^s f^n(r, X_r, I_r, v^n(r, X_r, \cdot) - v^n(r, X_r, I_r))\, dr \\ &\quad - \int_t^s \int_{E\times A} (v^n(r, y, b) - v(r, X_{r-}, I_{r-}))\; q(dr\, dy\, db), \quad s \in [t, T]. \end{aligned}$$

Thus the pairs $(Y_s^{n,t,x,a}, Z_s^{n,t,x,a}(y, b))$ and $(v^n(s, X_s, I_s), v^n(s, y, b) - v^n(s, X_{s-}, I_{s-}))$ are both solutions to the same BSDE under $\mathbb{P}^{t,x,a}$, and thus they coincide as members of the space $\mathbf{S^2_{t,x,a}} \times \mathbf{L^2_{t,x,a}}(\mathrm{q})$. The required equalities (2.63)-(2.64) follow. In particular we have that $v^n(t, x, a) = Y_t^{n,t,x,a}$. □

2.5.2. Convergence of the penalized solutions and conclusion of the proof. We study the behavior of the functions v^n as $n \to \infty$. To this end we first show that they are bounded above by the solution to the HJB equation.

Lemma 2.5.3. *Assume* (2.57), (2.58), (2.59). *Let v denote the solution to the HJB equation as provided by Theorem 2.2.7 and v^n the solution to* (2.61)-(2.62) *as provided in Lemma 2.5.2. Then, for all $(t, x, a) \in [0, T] \times E \times A$ and $n \geq 1$,*

$$v(t, x) \geq v^n(t, x, a).$$

Proof. Let $v : [0, T] \times E \to \mathbb{R}$ be a solution to the HJB equation. As in the proof of Proposition 2.2.1 we have the identity

$$g(X_T) - v(t, X_t) = \int_t^T \frac{\partial v}{\partial r}(r, X_r)\, dr + \int_{(t,T]} \int_{E\times A} (v(r, y) - v(r, X_{r-}))\, p(dr\, dy\, db),$$

which follows from the absolute continuity of $t \mapsto v(t, x)$, taking into account that X is constant among jump times and using the definition of the random measure p defined in (2.26) and the fact that v depends on t, x only. Since v is a solution to the HJB equation we have, for all $x \in E$ $a \in A$,

$$-\frac{\partial v}{\partial t}(t, x) \geq \mathcal{L}_E^a v(t, x) + f(t, x, a) = \int_E (v(t, y) - v(t, x))\, \lambda(x, a, dy) + f(t, x, a),$$

almost surely on $[0, T]$. Taking into account that (X, I) has piecewise constant trajectories we obtain

$$g(X_T) - v(t, X_t) \leq \int_{(t,T]} \int_{E\times A} (v(r, y) - v(r, X_{r-}))\, p(dr\, dy\, db) \tag{2.66}$$

$$- \int_t^T \int_E (v(r,y) - v(r,X_r))\, \lambda(X_r, I_r, dy)\, dr - \int_t^T f(r, X_r, I_r)\, dr.$$

Then, for any $n \geq 1$ and $\nu \in \mathcal{V}^n$ let us consider the probability $\mathbb{P}_\nu^{t,x,a}$ introduced above and recall that under $\mathbb{P}_\nu^{t,x,a}$ the compensator of the random measure $p(dr\, dy\, db)$ is $\tilde{p}^\nu(dr\, dy\, db) = \nu_r(b)\, \lambda_0(db)\, \delta_{\{X_{r-}\}}(dy)\, dr + \lambda(X_{r-},\, I_{r-},\, dy)\, \delta_{\{I_{r-}\}}(db)\, dr$. Noting that $v(r,y) - v(r, X_{r-})$ is predictable, taking the expectation in (2.66) we obtain

$$\mathbb{E}_\nu^{t,x,a}[g(X_T)] - v(t,x) \leq -\mathbb{E}_\nu^{t,x,a} \int_t^T f(r, X_r, I_r)\, dr.$$

Since $\nu \in \mathcal{V}^n$ was arbitrary, and recalling (2.41), we conclude that

$$v(t,x) \geq \sup_{\nu \in \mathcal{V}^n} \mathbb{E}^\nu \left[g(X_T) + \int_t^T f(r, X_r, I_r)\, dr \right] = v^n(t,x,a).$$

□

From Lemma 2.5.2 we know that $v^n(t,x,a) = Y_t^{n,t,x,a}$, and from Lemma 2.4.6 we know that $v^n(t,x,a)$ is monotonically increasing and uniformly bounded. Therefore we can define

$$\bar{v}(t,x,a) := \lim_{n \to \infty} v^n(t,x,a), \qquad t \in [0,T],\ x \in E,\ a \in A.$$

$\bar{v}$ is bounded, and from Lemma 2.5.3 we deduce that $\bar{v} \leq v$. As an increasing limit of continuous functions, $\bar{v}$ is lower semi-continuous. Further properties of $\bar{v}$ are proved in the following lemma. In particular, (2.67) (or (2.68)) means that $\bar{v}$ is a supersolution to the HJB equation.

Lemma 2.5.4. *Assume (2.57), (2.58), (2.59) and let $\bar{v}$ be the increasing limit of v^n. Then $\bar{v}$ does not depend on a, i.e. $\bar{v}(t,x,a) = \bar{v}(t,x,b)$ for every $t \in [0,T]$, $x \in E$ and $a, b \in A$. Moreover, setting $\bar{v}(t,x) = \bar{v}(t,x,a)$ we have*

$$\bar{v}(t,x) - \bar{v}(t',x) \geq \int_t^{t'} (\mathcal{L}_E^a \bar{v}(s,x) + f(s,x,a))\, ds \tag{2.67}$$

for $0 \leq t \leq t' \leq T$, $x \in E$, $a \in A$. More generally, for arbitrary Borel-measurable $\alpha : [0,T] \to A$ we have

$$\bar{v}(t,x) - \bar{v}(t',x) \geq \int_t^{t'} (\mathcal{L}_E^{\alpha(s)} \bar{v}(s,x) + f(s,x,\alpha(s)))\, ds \tag{2.68}$$

for $0 \leq t \leq t' \leq T$, $x \in E$ and $a \in A$.

Proof. v^n satisfies the integral equation (2.65), namely

$$\begin{aligned} v^n(t,x,a) \quad &= \quad g(x) + \int_t^T \int_E (v^n(s,y,a) - v^n(s,x,a))\, \lambda(x,a,dy)\, ds \\ &\quad + \int_t^T f(s,x,a)\, ds + n \int_t^T \int_A [v^n(s,x,b) - v^n(s,x,a)]^+\, \lambda_0(db)\, ds. \end{aligned}$$

Since v^n is a bounded sequence in $C_b([0,T] \times E \times A)$ converging pointwise to $\bar{v}$, setting $t=0$, dividing by n and letting $n \to \infty$ we obtain

$$\int_0^T \int_A [\bar{v}(s,x,b) - \bar{v}(s,x,a)]^+ \, \lambda_0(db)\, ds = 0. \tag{2.69}$$

Next we claim that $\bar{v}$ is right-continuous in t on $[0,T)$, for fixed $x \in E$, $a \in A$. To prove this we first note that, neglecting the term with the positive part $[\dots]^+$ we have

$$\begin{aligned} v^n(t',x,a) - v^n(t,x,a) &\le -\int_t^{t'} \int_E (v^n(s,y,a) - v^n(s,x,a))\, \lambda(x,a,dy)\, ds \\ &\quad - \int_t^{t'} f(s,x,a)\, ds \\ &\le C_0(t'-t), \end{aligned} \tag{2.70}$$

for some constant $C_0 > 0$ and for all $0 \le t \le t' \le T$ and $n \ge 1$, where we have used again the fact that v^n is uniformly bounded. Now fix $t \in [0,T)$. Since, as already noticed, $\bar{v}$ is lower semi-continuous we have $\bar{v}(t,x,a) \le \liminf_{s \downarrow t} \bar{v}(s,x,a)$. The required right continuity follows if we can show that $\bar{v}(t,x,a) \ge \limsup_{s \downarrow t} \bar{v}(s,x,a)$. Suppose not. Then there exists $s_k \downarrow t$ such that $\bar{v}(s_k,x,a)$ tends to some limit $l > \bar{v}(t)$. It follows that $\bar{v}(s_k,x,a) - \bar{v}(t,x,a) > C_0(s_k - t)$ for some k sufficiently large, and therefore also $v^n(s_k,x,a) - v^n(t,x,a) > C_0(s_k - t)$ for some n sufficiently large, contradicting (2.70). This contradiction shows that $\bar{v}$ is right-continuous in t on $[0,T)$.

Then it follows from (2.69) that $\int_A [\bar{v}(t,x,b) - \bar{v}(t,x,a)]^+ \, \lambda_0(db) = 0$ for every $x \in E$, $a \in A$, $t \in [0,T]$. Therefore there exists $B \subset A$ (dependent on t,x,a) such that B is a Borel set with $\lambda_0(B) = 0$, and

$$\bar{v}(t,x,a) \quad \ge \quad \bar{v}(t,x,b'), \qquad b' \notin B. \tag{2.71}$$

Since λ_0 has full support, B cannot contain any open ball. So given an arbitrary $b \in A$ we can find a sequence $b_n \to b$, $b_n \notin B$. Writing (2.71) with b_n instead of b' and using the lower semi-continuity of $\bar{v}$ we deduce that $\bar{v}(t,x,a) \ge \liminf_n \bar{v}(t,x,b_n) \ge \bar{v}(t,x,b)$. Since a and b were arbitrary we finally conclude that $\bar{v}(t,x,a) = \bar{v}(t,x,b)$ for every $t \in [0,T]$, $x \in E$ and $a,b \in A$, so that $\bar{v}(t,x,a)$ does not depend on a and we can define $\bar{v}(t,x) = \bar{v}(t,x,a)$.

Passing to the limit as $n \to \infty$ in the first inequality of (2.70) we immediately obtain (2.67), so it remains to prove (2.68). Let $\mathcal{A}(\bar{v})$ denote the set of all Borel-measurable $\alpha : [0,T] \to A$ such that (2.68) holds, namely for every $0 \le t \le t' \le T$, $x \in E$, $a \in A$,

$$\begin{aligned} \bar{v}(t,x) - \bar{v}(t',x) &\ge \int_t^{t'} \int_E \bar{v}(s,y)\, \lambda(x,\alpha(s),dy)\, ds \\ &\quad - \int_t^{t'} \bar{v}(s,x)\, \lambda(x,\alpha(s),E)\, ds + \int_t^{t'} f(s,x,\alpha(s))\, ds. \end{aligned} \tag{2.72}$$

Suppose that $\alpha_n \in \mathcal{A}(\bar{v})$, $\alpha : [0,T] \to A$ is Borel-measurable and $\alpha_n(t) \to \alpha_n(t)$ for almost all $t \in [0,T]$. Note that

$$\int_E \bar{v}(t,y)\,\lambda(x,a,dy) = \lim_{n\to\infty} \int_E \bar{v}^n(t,y,a)\,\lambda(x,a,dy) \tag{2.73}$$

and the latter is an increasing limit. Since $v^n \in C_b([0,T]\times E\times A)$ and λ is Feller, for any $n \geq 1$ the functions in the right-hand side of (2.73) are continuous in (t,x,a) (see e.g. [**15**], Proposition 7.30) and therefore the left-hand side is a lower semicontinuous function of (t,x,a). It follows from this and the Fatou Lemma that

$$\begin{aligned} \int_t^{t'} \int_E \bar{v}(s,y)\,\lambda(x,\alpha(s),dy)\,ds &\leq \int_t^{t'} \liminf_{n\to\infty} \left[\int_E \bar{v}(s,y)\,\lambda(x,\alpha_n(s),dy)\right] ds \\ &\leq \liminf_{n\to\infty} \int_t^{t'} \int_E \bar{v}(s,y)\,\lambda(x,\alpha_n(s),dy)\,ds. \end{aligned}$$

Using this inequality and the continuity and boundedness of the maps $a \mapsto \lambda(x,a,E)$, $a \mapsto f(t,x,a)$ we see that assuming the validity of inequality (2.72) for α_n implies that it also holds for α, hence $\alpha \in \mathcal{A}(\bar{v})$.

Next we note that $\mathcal{A}(\bar{v})$ contains all piecewise constant functions of the form $\alpha(t) = \sum_{i=1}^k a_i 1_{[t_i,t_{i+1})}(t)$ with $k \geq 1$, $0 = t_1 < t_2 < \ldots < t_{k+1} = T$, $a_i \in A$: indeed, it is enough to write down (2.67) with $[t,t') = [t_i,t_{i+1})$ and sum up over i to get (2.68) for $\alpha(\cdot)$ and therefore conclude that $\alpha(\cdot) \in \mathcal{A}(\bar{v})$. Since we have already proved that the class $\mathcal{A}(\bar{v})$ is stable under almost sure pointwise limits it follows that $\mathcal{A}(\bar{v})$ contains all Borel-measurable functions $\alpha : [0,T] \to A$ as required. □

We are now ready to conclude the proof of our main result.

Proof of Theorem 2.5.1. We will prove the inequality

$$\bar{v}(t,x) \quad \geq \quad V(t,x), \qquad t \in [0,T], x \in E, \tag{2.74}$$

where $\bar{v} = \lim_{n\to\infty} v^n$ was introduced before Lemma 2.5.4. Since we know that $\bar{v} \leq v$ and, by Theorem 2.2.10, $v = V$ it follows from (2.74) that $\bar{v} = v = V$. Passing to the limit as $n \to \infty$ in (2.63) and recalling (2.38) all the other equalities follow immediately.

To prove (2.74) we fix $t \in [0,T]$, $x \in E$ and a Borel-measurable map $\alpha : [0,T] \times E \to A$, i.e. an element of $\mathcal{A}_{ad}$, the set of admissible control laws for the primal control problem, and denote by $\mathbb{P}_\alpha^{t,x}$ the associated probability measure on $(\Omega, \mathcal{F}_\infty)$, for the controlled system started at time t from point x, as in section 2.2.2. We will prove that $\bar{v}(t,x) \geq J(t,x,\alpha)$, the gain functional defined in (2.11). Recall that in Ω we had defined a canonical marked point process $(T_n,E_n)_{n\geq 1}$ and the associated random measure p. Fix $\omega \in \Omega$ and consider the points $T_n(\omega)$ lying in $(t,T]$, which we rename S_i; thus, $t < S_1 < \ldots S_k \leq T$, for some k (also depending on ω). Recalling that $\bar{v}(T,x) = g(x)$ we have

$$g(X_T) - \bar{v}(t,x) \quad = \quad g(X_T) - \bar{v}(S_k, X_{S_k}) + \sum_{i=1}^k [\bar{v}(S_i, X_{S_i}) - \bar{v}(S_i, X_{S_i-})]$$

$$+\sum_{i=2}^{k}[\bar v(S_i, X_{S_i-}) - \bar v(S_{i-1}, X_{S_{i-1}})] + \bar v(S_1, X_{S_1-}) - \bar v(t,x).$$

$\mathbb{P}_\alpha^{t,x}$-a.s we have $X_{S_i-} = X_{S_{i-1}}$ $(2 \le i \le k)$ and $X_{S_1-} = x$, so we obtain

$$\begin{aligned} g(X_T) - \bar v(t,x) &= g(X_T) - \bar v(S_k, X_{S_k}) + \sum_{i=1}^{k}[\bar v(S_i, X_{S_i}) - \bar v(S_i, X_{S_i-})] \\ &\quad + \sum_{i=2}^{k}[\bar v(S_i, X_{S_{i-1}}) - \bar v(S_{i-1}, X_{S_{i-1}})] + \bar v(S_1, x) - \bar v(t,x). \end{aligned}$$

The first sum can be written as

$$\sum_{i=1}^{k}[\bar v(S_i, X_{S_i}) - \bar v(S_i, X_{S_i-})] = \int_t^T \int_E [\bar v(s,y) - \bar v(s, X_{s-})]\, p(ds\, dy),$$

while the other can be estimated from above by repeated applications of (2.68), taking into account that X is constant in the intervals $(t, S_1]$, $(S_{i-1}, S_i]$ $(2 \le i \le k)$ and $(S_k, T]$:

$$\begin{aligned} &\bar v(S_i, X_{S_{i-1}}) - \bar v(S_{i-1}, X_{S_{i-1}}) \\ &\le -\int_{S_{i-1}}^{S_i} \Big(\mathcal{L}_E^{\alpha(s, X_{S_{i-1}})} \bar v(s, X_{S_{i-1}}) + f(s, X_{S_{i-1}}, \alpha(s, X_{S_{i-1}}))\Big)\, ds \\ &= -\int_{S_{i-1}}^{S_i} \Big(\mathcal{L}_E^{\alpha(s, X_s)} \bar v(s, X_s) + f(s, X_s, \alpha(s, X_s))\Big)\, ds \end{aligned}$$

for $2 \le i \le k$ and similar formulae for the intervals $(t, S_1]$, and $(S_k, T]$. We end up with

$$\begin{aligned} g(X_T) - \bar v(t,x) \;&\le\; \int_t^T \int_E [\bar v(s,y) - \bar v(s, X_{s-})]\, p(ds\, dy) \\ &\quad - \int_t^T \Big(\mathcal{L}_E^{\alpha(s, X_s)} \bar v(s, X_s) + f(s, X_s, \alpha(s, X_s))\Big)\, ds. \end{aligned}$$

Recalling that the compensator of the measure p under $\mathbb{P}_\alpha^{t,x}$ is $\lambda(X_{s-}, \alpha(s, X_{s-}), dy)$ $ds\ 1_{[t,\infty)}(s)$ we have, taking expectation,

$$\mathbb{E}_\alpha^{t,x} \int_t^T \int_E [\bar v(s,y) - \bar v(s, X_{s-})]\, p(ds\, dy) = \mathbb{E}_\alpha^{t,x} \int_t^T \mathcal{L}_E^{\alpha(s, X_s)} \bar v(s, X_s)\, ds,$$

which implies, by the previous inequality,

$$\mathbb{E}_\alpha^{t,x}[g(X_T)] - \bar v(t,x) \le -\mathbb{E}_\alpha^{t,x} \int_t^T f(s, X_s, \alpha(s, X_s))\, ds$$

and so $v(t,x) \ge J(t,x,\alpha)$. Since $\alpha \in \mathcal{A}_{ad}$ was arbitrary we conclude that $v(t,x) \ge V(t,x)$.

□

Chapter 3

Optimal control of Piecewise Deterministic Markov Processes and constrained BSDEs with nonnegative jumps

3.1. Introduction

The aim of the present chapter is to prove that the value function in an infinite-horizon optimal control problem for piecewise deterministic Markov processes (PDMPs) can be represented by means of an appropriate backward stochastic differential equation. Piecewise deterministic Markov processes, introduced in Davis [**35**], evolve through random jumps at random times, while the behavior between jumps is described by a deterministic flow. We consider optimal control problems of PDMPs where the control acts continuously on the jump dynamics and on the deterministic flow as well.

Let us start by describing our setting in an informal way. Let E be a Borel space and $\mathcal{E}$ the corresponding σ-algebra. A PDMP on $(E, \mathcal{E})$ can be described by means of three local characteristics, namely a continuous flow $\phi(t, x)$, a jump rate $\lambda(x)$, and a transition measure $Q(x, dy)$, according to which the location of the process at the jump time is chosen. The PDMP dynamic can be described as follows: starting from some initial point $x \in E$, the motion of the process follows the flow $\phi(t, x)$ until a random jump T_1, verifying

$$\mathbb{P}(T_1 > s) = \exp\left(-\int_0^s \lambda(\phi(r,x))\,dr\right), \quad s \geq 0.$$

At time T_1 the process jumps to a new point X_{T_1} selected with probability $Q(x, dy)$ (conditionally to T_1), and the motion restarts from this new point as before.

Now let us introduce a measurable space $(A, \mathcal{A})$, which will denote the space of control actions. A controlled PDMP is obtained starting from a jump rate $\lambda(x, a)$ and a transition measure $Q(x, a, dy)$, depending on an additional control parameter $a \in A$, and a continuous flow $\phi^\beta(t, x)$, depending on the choice of a measurable function $\beta(t)$ taking values on $(A, \mathcal{A})$. A natural way to control a PDMP is to chose a control strategy among the set of *piecewise open-loop policies*, i.e., measurable functions that depend only on the last jump time and post jump position. We can mention Almudevar [**1**], Bauerle and Rieder [**11**], Costa and Dufour [**32**], Davis [**35**], [**34**], Dempster [**40**], as a sample of works that use this kind of approach. Roughly speaking, at each jump time T_n, we choose an open loop control α_n depending on the initial condition X_{T_n} to be used until the next jump time. A control α in the class of admissible control laws $\mathcal{A}_{ad}$ has the explicit form

$$\alpha_t = \sum_{n=1}^{\infty} \alpha_n(t - T_n, X_{T_n})\, 1_{[T_n,\, T_{n+1})}(t), \tag{3.1}$$

and the controlled process X is

$$X_t = \phi^{\alpha_n}(t - T_n, E_n), \quad t \in [T_n,\, T_{n+1}).$$

We denote by $\mathbb{P}^x_\alpha$ the probability measure such that, for every $n > 1$, the conditional survivor function of the jump time T_{n+1} and the distribution of the post jump position $X_{T_{n+1}}$, are

$$\mathbb{P}^x_\alpha(T_{n+1} > s \,|\, \mathcal{F}_{T_n}) = \exp\left(-\int_{T_n}^s \lambda(\phi^{\alpha_n}(r - T_n, X_{T_n}), \alpha_n(r - T_n, X_{T_n}))\,dr\right),$$
$$\mathbb{P}^x_\alpha(X_{T_{n+1}} \in B |\, \mathcal{F}_{T_n},\, T_{n+1}) = Q(\phi^{\alpha_n}(T_{n+1} - T_n, X_{T_n}), \alpha_n(T_{n+1} - T_n, X_{T_n}), B),$$

on $\{T_n < \infty\}$.

In the classic infinite-horizon control problem one wants to minimize over all control laws α a functional cost of the form

$$J(x, \alpha) = \mathbb{E}^x_\alpha\left[\int_0^\infty e^{-\delta\, s}\, f(X_s, \alpha_s)\,ds\right] \tag{3.2}$$

where $\mathbb{E}^x_\alpha$ denotes the expectation under $\mathbb{P}^x_\alpha$, f is a given real function on $E \times A$ representing the running cost, and $\delta \in (0, \infty)$ is a discounting factor. The value function of the control problem is defined in the usual way:

$$V(x) = \inf_{\alpha \in \mathcal{A}_{ad}} J(x, \alpha), \quad x \in E. \tag{3.3}$$

Let now E be an open subset of $\mathbb{R}^d$, and $h(x, a)$ be a bounded Lipschitz continuous function such that $\phi^\alpha(t, x)$ is the unique solution of the ordinary differential equation

$$\dot{x}(t) = h(x(t), \alpha(t)), \quad x(0) = x \in E.$$

We will assume that λ and f are bounded functions, uniformly continuous, and Q is a Feller stochastic kernel. In this case, V is known to be the unique viscosity solution on $[0, \infty) \times E$ of the Hamilton-Jacobi-Bellman (HJB) equation

$$\delta v(x) = \sup_{a \in A} \left(h(x,a) \cdot \nabla v(x) + \lambda(x,a) \int_E (v(y) - v(x))\, Q(x,a,dy) \right), \quad x \in E. \tag{3.4}$$

The characterization of the optimal value function as the viscosity solution of the corresponding integro-differential HJB equation is an important approach to tackle the optimal control problem of PDMPs, and can be found for instance in Davis and Farid [**36**], Dempster and Ye [**41**], [**42**]. Alternatively, the control problem can be reformulated as a discrete-stage Markov decision model, where the stages are the jumps times of the process and the decision at each stage is the control function that solves a deterministic optimal control problem. The reduction of the optimal control problem to a discrete-time Markov decision process is exploited for instance in [**1**], [**11**], [**32**], [**35**], [**34**].

In the present chapter our aim is to represent the value function $V(x)$ by means of an appropriate BSDE. We are interested in the general case when $\{\mathbb{P}_\alpha^x\}_\alpha$ is a non-dominated model, which, roughly speaking, reflects the fully non-linear character of the HJB equation. This basic difficulty has prevented the effective use of BSDE techniques in the context of optimal control of PDMPs until now. In fact, we believe that this is the first time that this difficulty is coped with and this connection is established. It is our hope that the great development that BSDE theory has now gained will produce new results in the optimization theory of PDMPs. In the context of diffusions, probabilistic formulae for the value function for non-dominated models have been discovered only in the recent year. In this sense, a fundamental role is played by [**88**], where a new class of BSDEs with nonpositive jumps is introduced in order to provide a probabilistic formula, known as nonlinear Feynman-Kac formula, for fully nonlinear integro-partial differential equations, associated to the classical optimal control for diffusions. This approach was later applied to many cases within optimal switching and impulse control problems, see Elie and Kharroubi [**54**], [**55**], [**56**], Kharroubi, Ma, Pham and Zhang [**87**], and developed with extensions and applications, see Cosso and Chokroun [**25**], Cosso, Fuhrman and Pham [**31**], and Fuhrman and Pham [**67**]. In all the above mentioned cases the controlled processes are diffusions constructed as solutions to stochastic differential equations of Itô type driven by a Brownian motion.

We wish to extend to the PDMPs framework the theory developed in the context of optimal control for diffusions. The fundamental idea behind the derivation of the Feynman-Kac representation, borrowed from [**88**], concerns the so-called *randomization of the control*, that we are going to describe below in our framework. A first step in the generalization of this method to the non-diffusive processes context was done in Chapter 2, where a probabilistic representation for the value function associated to an optimal control problem for pure jump Markov processes was provided.

As in the pure jump case, also in the PDMPs framework the correct formulation of the randomization method requires some efforts, and can not be modelled on the diffusive case, since the controlled processes are not defined as solutions to stochastic differential equations. In addition, the presence of the controlled flow between jumps in the PDMP's dynamics makes the treatment more difficult and suggests to use the viscosity solution theory. Finally, we notice that we consider PDMPs with state space E with no boundary. This restriction is due to the fact that the presence of the boundary induces technical difficulties on the study of the associated BSDE, which would be driven by a non quasi-left continuous random measure, see Remark 3.2.3. For such general BSDEs the existence and uniqueness results were at disposal only in particular frameworks, see e.g. [**26**] for the deterministic case, and counter-examples were provided in the general case, see Section 4.3 in [**29**]. Only recently this problem was faced and solved in a general context in [**2**], were a technical condition is provided in order to achieve existence and uniqueness of the BSDE, see Chapter 4. The mentioned condition turns out to be verified in the case of control problems related to PDMPs with discontinuities at the boundary of the domain, see Remark 4.4.5. This fact opens to the possibility to apply the BSDEs techniques also in this context, which is left as a future development of the method.

Let us now informally describe the randomization method in the PDMPs framework. The first step, for any starting point $x \in E$, consists in replacing the state trajectory and the associated control process (X_s, α_s) by an (uncontrolled) PDMP (X_s, I_s), in such a way that I is a Poisson process with values in the space of control actions A, with an intensity $\lambda_0(db)$ which is arbitrary but finite and with full support, and X is suitably defined. In particular, the PDMP (X, I) is constructed in a different probability space by means of a new triplet of local characteristics and takes values on the enlarged space $E \times A$. Let us denote by $\mathbb{P}^{x,a}$ the corresponding law, where (x, a) is the starting point in $E \times A$. Then we formulate an auxiliary optimal control problem where we control the intensity of the process I: for any predictable, bounded and positive random field $\nu_t(b)$, by means of a theorem of Girsanov type, we construct a probability measure $\mathbb{P}^{x,a}_\nu$ under which the compensator of I is the random measure $\nu_t(db)\,\lambda_0(db)\,dt$ (under $\mathbb{P}^{x,a}_\nu$ the law of X is also changed) and we minimize the functional

$$J(x, a, \nu) = \mathbb{E}^{x,a}_\nu \left[\int_0^\infty e^{-\delta\, s}\, f(X_s, I_s)\, ds \right]. \tag{3.5}$$

over all possible choices of ν. This will be called the *dual* control problem. Notice that the family $\{\mathbb{P}^{x,a}_\nu\}_\nu$ is a dominated model. One of our main results states that the value function of the dual control problem, denoted as $V^*(x, a)$, can be represented by means of a well-posed constrained BSDE. The latter is an equation over an infinite horizon of the form

$$\begin{aligned} Y^{x,a}_s &= Y^{x,a}_T - \delta \int_s^T Y^{x,a}_r\, dr + \int_s^T f(X_r, I_r)\, dr - (K^{x,a}_T - K^{x,a}_s) \qquad (3.6)\\ &- \int_s^T \int_A Z^{x,a}_r(X_r,\, b)\, \lambda_0(db)\, dr - \int_s^T \int_{E\times A} Z^{x,a}_r(y,\, b)\, q(dr\, dy\, db), \quad 0 \leqslant s \leqslant T < \infty, \end{aligned}$$

with unknown triplet $(Y^{x,a}, Z^{x,a}, K^{x,a})$ where q is the compensated random measure associated to (X, I), $K^{x,a}$ is a predictable increasing càdlàg process, $Z^{x,a}$ is a predictable random field, where we additionally add the sign constraint

$$Z_s^{x,a}(X_{s-}, b) \geqslant 0. \tag{3.7}$$

The reference filtration is now the canonical one associated to the pair (X, I). We prove that this equation has a unique maximal solution, in an appropriate sense, and that the value of the process $Y^{x,a}$ at the initial time represents the dual value function:

$$Y_0^{x,a} = V^*(x, a). \tag{3.8}$$

Our main purpose is to show that the maximal solution to (3.6)-(3.7) at the initial time also provides a Feynman-Kac representation to the value function (3.3) of our original optimal control problem for PDMPs. To this end, we introduce the deterministic real function on $E \times A$

$$v(x, a) := Y_0^{x,a}, \tag{3.9}$$

and we prove that v is a viscosity solution to (3.4). By the uniqueness of the solution to the HJB equation (3.4) we conclude that the value of the process Y at the initial time represents both the original and the dual value function:

$$Y_0^{x,a} = V^*(x, a) = V(x). \tag{3.10}$$

Identity (3.10) is the desired BSDE representation of the value function for the original control problem and a Feynman-Kac formula for the general HJB equation (3.4).

Formula (3.10) can be used to design algorithms based on the numerical approximation of the solution to the constrained BSDE (3.6)-(3.7), and therefore to get probabilistic numerical approximations for the value function of the addressed optimal control problem. In the recent years there has been much interest in this problem, and numerical schemes for constrained BSDEs have been proposed and analyzed in the diffusive framework, see [**86**], [**85**]. We hope that our results may be used to get similar methods in the PDMPs context as well.

The chapter is organized as follows. Section 3.2 is dedicated to define a setting where the optimal control (3.3) is solved by means of the corresponding HJB equation (3.4). We start by recalling the construction of a PDMP given its local characteristics. In order to apply techniques based on BSDEs driven by general random measures, we work in a canonical setting and we use a specific filtration. The construction is based on the well-posedness of the martingale problem for multivariate marked point processes studied in Jacod [**75**], and is the object of Section 3.2.1. This general procedure is then applied in Section 3.2.2 to formulate in a precise way the optimal control problem we are interested in. At the end of Section 3.2.2 we recall a classical result on existence and uniqueness of the viscosity solution to the HJB equation (3.4), and its identification with the value function V, provided by Davis and Farid [**36**].

In Section 3.3 we start to develop the control randomization method. Given suitable local characteristics, we introduce an auxiliary process (X, I) on $E \times A$ by relying on the construction in Section 3.2.1, and we formulate a dual optimal control problem for it under suitable conditions. The formulation of the randomized process is very different from the diffusive framework, since our data are the local characteristics of the process rather than the coefficients of some stochastic differential equations solved by it. In particular, we need to choose a specific probability space under which the component I (independent to X) is a Poisson process.

In Section 3.4 we introduce the constrained BSDE (3.6)-(3.7) over infinite horizon. By a penalization approach, we prove that under suitable assumptions the above mentioned equation admits a unique maximal solution (Y, Z, K) in a certain class of processes. Moreover, the component Y at the initial time coincides with the value function V^* of the dual optimal control problem. This is the first of our main results, and is the object of Theorem 3.4.8.

Finally, in Section 3.5 we prove that the initial value of the maximal solution $Y^{x,a}$ to (3.6)-(3.7) provides a viscosity solution to (3.4). This is the second main result of the paper, which is stated in Theorem 3.5.1. As a consequence, by means of a comparison theorem for sub and supersolutions to first-order integro-partial differential equations, we get the desired nonlinear Feynman-Kac formula, as well as the equality between the value functions of the primal and the dual control problems, see Corollary 3.5.2. The proof of Theorem 3.5.1 is based on arguments from the viscosity theory, and combines BSDEs techniques with control-theoretic arguments. A relevant task is to derive the key property that the function v in (3.9) does not depend on a, as consequence of the A-nonnegative constrained jumps.

Recalling the identification in Theorem 3.4.8, we are able to give a direct proof of the non-dependence of v on a by means of control-theoretic techniques, see Proposition 3.5.6 and the comments below. This allows us to consider very general spaces A of control actions. Moreover, differently to the previous literature, we provide a direct proof of the viscosity solution property of v, which does not need to rely on a penalized HJB equation. This is achieved by generalizing to the setting of the dual control problem the proof that allows to derive the HJB equation from the dynamic programming principle, see Propositions 3.5.8 and 3.5.9.

3.2. Piecewise Deterministic controlled Markov Processes

3.2.1. The construction of a PDMP given its local characteristics. Given a topological space F, in the sequel $\mathcal{B}(F)$ will denote the Borel σ-field associated with F, and by $\mathbb{C}_b(F)$ the set of all bounded continuous functions on F. The Dirac measure concentrated at some point $x \in F$ will be denoted δ_x.

Let $(E, \mathcal{E})$ be a Borel measurable space. We will often need to construct a PDMP in E with a given triplet of local characteristics (ϕ, λ, Q). We assume that $\phi : \mathbb{R} \times E \to E$ is a continuous function, $\lambda : E \mapsto \mathbb{R}_+$ is a nonnegative continuous function satisfying

$$\sup_{x \in E} \lambda(x) < \infty, \tag{3.11}$$

and that Q maps E into the set of probability measures on $(E, \mathcal{E})$, and is a stochastic Feller kernel, i.e., for all $v \in \mathbb{C}_b(E)$, the map $x \mapsto \int_E v(y)\, Q(x, dy)$ $(x \in E)$ is continuous.

We recall the main steps of the construction of a PDMP given its local characteristics. The existence of a Markovian process associated with the triplet (ϕ, λ, Q) is a well known fact (see, e.g., [**35**], [**32**]). Nevertheless, we need special care in the choice of the corresponding filtration, since this will be crucial when we solve associated BSDEs and implicitly apply a version of the martingale representation theorem in the sections that follow. For this reason, in the following we will use an explicit construction that we are going to describe. Many of the techniques we are going to use are borrowed from the theory of multivariate (marked) point processes. We will often follow [**75**], but we also refer the reader to the treatise [**77**] for a more systematic exposition.

We start by constructing a suitable sample space to describe the jumping mechanism of the Markov process. Let Ω' denote the set of sequences $\omega' = (t_n, e_n)_{n \geq 1}$ in $((0, \infty) \times E) \cup \{(\infty, \Delta)\}$, where $\Delta \notin E$ is adjoined to E as an isolated point, satisfying in addition

$$t_n \leq t_{n+1}; \qquad t_n < \infty \implies t_n < t_{n+1}. \tag{3.12}$$

To describe the initial condition we will use the measurable space $(E, \mathcal{E})$. Finally, the sample space for the Markov process will be $\Omega = E \times \Omega'$. We define canonical functions $T_n : \Omega \to (0, \infty]$, $E_n : \Omega \to E \cup \{\Delta\}$ as follows: writing $\omega = (e, \omega')$ in the form $\omega = (e, t_1, e_1, t_2, e_2, \ldots)$ we set for $t \geq 0$ and for $n \geq 1$

$$T_n(\omega) = t_n, \qquad E_n(\omega) = e_n, \qquad T_\infty(\omega) = \lim_{n \to \infty} t_n, \qquad T_0(\omega) = 0, \qquad E_0(\omega) = e.$$

We also introduce the counting process $N(s, B) = \sum_{n \in \mathbb{N}} 1_{T_n \leq s} 1_{E_n \in B}$, and we define the process $X : \Omega \times [0, \infty) \to E \cup \Delta$ setting

$$X_t = \begin{cases} \phi(t - T_n, E_n) & \text{if } T_n \leq t < T_{n+1}, \text{ for } n \in \mathbb{N}, \\ \Delta & \text{if } t \geq T_\infty. \end{cases} \tag{3.13}$$

In Ω we introduce for all $t \geq 0$ the σ-algebras $\mathcal{G}_t = \sigma(N(s, B) \,:\, s \in (0, t], B \in \mathcal{E})$. To take into account the initial condition we also introduce the filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$, where $\mathcal{F}_0 = \mathcal{E} \otimes \{\emptyset, \Omega'\}$, and for all $t \geq 0$ $\mathcal{F}_t$ is the σ-algebra generated by $\mathcal{F}_0$ and $\mathcal{G}_t$. $\mathbb{F}$ is right-continuous and will be called the natural filtration. In the following all concepts of measurability for stochastic processes (adaptedness, predictability etc.) refer to $\mathbb{F}$. We denote by $\mathcal{F}_\infty$ the σ-algebra generated by all the σ-algebras $\mathcal{F}_t$. The symbol $\mathcal{P}$ denotes the σ-algebra of $\mathbb{F}$-predictable subsets of $[0, \infty) \times \Omega$.

On the filtered sample space $(\Omega, \mathbb{F})$ we have so far introduced the canonical marked point process $(T_n, E_n)_{n \geq 1}$. The corresponding random measure p is, for any $\omega \in \Omega$, a σ-finite measure on $((0, \infty) \times E, \mathcal{B}(0, \infty) \otimes \mathcal{E})$ defined as

$$p(\omega, ds\, dy) = \sum_{n \in \mathbb{N}} 1_{\{T_n(\omega) < \infty\}}\, \delta_{(T_n(\omega), E_n(\omega))}(ds\, dy), \tag{3.14}$$

where δ_k denotes the Dirac measure at point $k \in (0,\infty) \times E$. For notational convenience the dependence on ω will be suppressed and, instead of $p(\omega, ds\, dy)$, it will be written $p(ds\, dy)$.

Proposition 3.2.1. *Assume that* (3.11) *holds, and fix $x \in E$. Then there exists a unique probability measure on $(\Omega, \mathcal{F}_\infty)$, denoted by $\mathbb{P}^x$, such that its restriction to $\mathcal{F}_0$ is δ_x, and the $\mathbb{F}$-compensator of the measure p under $\mathbb{P}^x$ is the random measure*

$$\tilde{p}(ds\, dy) = \sum_{n\in\mathbb{N}} 1_{[T_n,\, T_{n+1})}(s)\, \lambda(\phi(s - T_n, E_n))\, Q(\phi(s - T_n, E_n), dy)\, ds.$$

Moreover, $\mathbb{P}^x(T_\infty = \infty) = 1$.

Proof. The result is a direct application of Theorem 3.6 in [**75**]. The fact that, $\mathbb{P}^x$-a.s., $T_\infty = \infty$ follows from the boundedness of λ, see Proposition 24.6 in [**35**]. □

For fixed $x \in E$, the sample path of the process (X_t) in (3.13) under $\mathbb{P}^x$ can be defined iteratively, by means of (ϕ, λ, Q), in the following way. Set

$$F(s,x) = \exp\left(-\int_0^s \lambda(\phi(r,x))\, dr\right),$$

we have

$$\mathbb{P}^x(T_1 > s) = F(s,x), \tag{3.15}$$
$$\mathbb{P}^x(X_{T_1} \in B |\, T_1) = Q(x,B), \tag{3.16}$$

on $\{T_1 < \infty\}$, and, for every $n > 1$,

$$\mathbb{P}^x(T_{n+1} > s \,|\, \mathcal{F}_{T_n}) = \exp\left(-\int_{T_n}^s \lambda(\phi(r - T_n, X_{T_n}))\, dr\right), \tag{3.17}$$
$$\mathbb{P}^x(X_{T_{n+1}} \in B |\, \mathcal{F}_{T_n},\, T_{n+1}) = Q(\phi(T_{n+1} - T_n, X_{T_n}), B), \tag{3.18}$$

on $\{T_n < \infty\}$.

Proposition 3.2.2. *In the probability space $\{\Omega, \mathcal{F}_\infty, \mathbb{P}^x\}$ the process X has distribution δ_x at time zero, and it is a homogeneous Markov process, i.e., for any $x \in E$, nonnegative times t, s, $t \le s$, and for every bounded measurable function f,*

$$\mathbb{E}^x[f(X_{t+s}) \,|\, \mathcal{F}_t] = P_s(f(X_t)), \tag{3.19}$$

where $P_t f(x) := \mathbb{E}^x[f(X_t)]$.

Proof. From (3.17), taking into account the semigroup property $\phi(t+s,x) = \phi(t, \phi(s,x))$, we have

$$\begin{aligned}
&\mathbb{P}^x(T_{n+1} > t+s \,|\, \mathcal{F}_t)\, 1_{\{t\in[T_n,\, T_{n+1})\}} \\
&= \frac{\mathbb{P}^x(T_{n+1} > t+s \,|\, \mathcal{F}_{T_n})}{\mathbb{P}^x(T_{n+1} > t \,|\, \mathcal{F}_{T_n})}\, 1_{\{t\in[T_n,\, T_{n+1})\}} \\
&= \exp\left(-\int_t^{t+s} \lambda(\phi(r - T_n, X_{T_n}))\, dr\right) 1_{\{t\in[T_n,\, T_{n+1})\}} \\
&= \exp\left(-\int_0^s \lambda(\phi(r + t - T_n, X_{T_n}))\, dr\right) 1_{\{t\in[T_n,\, T_{n+1})\}}
\end{aligned}$$

$$
\begin{aligned}
&= \exp\left(-\int_0^s \lambda(\phi(r, X_t))\, dr\right) 1_{\{t\in[T_n, T_{n+1})\}} \\
&= F(s, X_t)\, 1_{\{t\in[T_n, T_{n+1})\}}. \qquad (3.20)
\end{aligned}
$$

Hence, denoting $N_t = N(t, E)$, it follows from (3.20) that

$$
\mathbb{P}^x(T_{N_t+1} > t+s \,|\, \mathcal{F}_t) = F(s, X_t);
$$

in other words, conditional on $\mathcal{F}_t$, the jump time after t of a PDMP started at x has the same distribution as the first jump time of a PDMP started at X_t. Since the remaining interarrival times and postjump positions are independent on the past, we have shown that (3.19) holds for every bounded measurable function f. □

Remark 3.2.3. In the present chapter we restrict the analysis to the case of PDMPs on a domain E with no boundary. This choice is motivated by the fact that the presence of jumps at the boundary of the domain would induce discontinuities in the compensator of the random measure associated to the process. Since we have in mind to apply techniques based on BSDEs driven by the compensated random measure associated to the PDMP (see Section 3.4), this fact would considerably complicates the tractation.

More precisely, consider a PDMP on a state space E with boundary ∂E. In this case, when the process reaches the boundary, a forced jump occurs and the process immediately goes back to the interior of the domain. According to (26.2) in [**35**], the compensator of the counting measure p in (3.14) admits the form

$$
\tilde{p}(ds\, dy) = \lambda(X_{s-})\, Q(X_{s-}, dy)\, ds + dp^*_s\, R(X_{s-}, dy),
$$

where

$$
p^*_s = \sum_{n=1}^{\infty} 1_{\{s \geq T_n\}}\, 1_{\{X_{T_n-}\in\Gamma\}}
$$

is the process counting the number of jumps of X from the active boundary $\Gamma \in \partial E$ (for the precise definition of Γ see page 61 in [**35**]), and $R : \partial E \times \mathcal{E} \to E$ is the transition probability measure describing the distribution of the process after the forced jumps. In particular, the compensator $\tilde{p}$ can be rewritten as

$$
\tilde{p}(ds\, dy) = dA_s\, \phi(X_{s-}, dy),
$$

where $\phi(X_{s-}, dy) = Q(X_{s-}, dy)\, 1_{\{X_{s-}\in E\}} + R(X_{s-}, dy)\, 1_{\{X_{s-}\in\Gamma\}}$, and $A_s = \lambda(X_{s-})\, ds$ $+ dp^*_s$ is a predictable and discontinuous process, with jumps

$$
\Delta A_s = 1_{\{X_{s-}\in\Gamma\}}.
$$

The presence of these discontinuities in the compensator induces very technical difficulties in the study of the associated BSDE, see Chapter 4. The above mentioned case is left as a future improvement of the theory.

3.2.2. Optimal control of PDMPs. In the present section we aim at formulating an optimal control problem for piecewise deterministic Markov processes, and to discuss its solvability. The PDMP state space E will be an open subset of $\mathbb{R}^d$, and $\mathcal{E}$ the corresponding σ-algebra. In addition, we introduce a Borel space A, endowed with its σ-algebra $\mathcal{A}$, called the space of control actions. The additional hypothesis that A is compact is not necessary for the majority of the results, and will be explicitly asked whenever needed. The other data of the problem consist in three functions f, h and λ on $E\times A$, and in a probability transition Q from $(E\times A, \mathcal{E}\otimes\mathcal{A})$ to $(E, \mathcal{E})$, satisfying the following conditions.

(HhλQ)

(i) $h : E\times A \mapsto E$ is a bounded, uniformly continuous, function satisfying

$$\begin{cases} \forall x,\, x' \in E, \text{ and } \forall a, a' \in A, & |h(x,a) - h(x',a')| \leqslant L_h\,(|x-x'| + |a-a'|), \\ \forall x \in E \text{ and } \forall a \in A, & |h(x,a)| \leqslant M_h, \end{cases}$$

where L_h and M_h are constants independent of $a, a' \in A$, $x, x' \in E$.

(ii) $\lambda : E\times A \mapsto \mathbb{R}^+$ is a nonnegative bounded uniformly continuous function, satisfying

$$\sup_{(x,a)\in E\times A} \lambda(x,a) < \infty. \tag{3.21}$$

(iii) Q maps $E\times A$ into the set of probability measures on $(E, \mathcal{E})$, and is a stochastic Feller kernel. i.e., for all $v \in \mathbb{C}_b(E)$, the map $(x,a) \mapsto \int_{\mathbb{R}^d} v(y)\,Q(x,a,dy)$ is continuous (hence it belongs to $\mathbb{C}_b(E\times A)$).

(Hf) $f : E\times A \mapsto \mathbb{R}^+$ is a nonnegative bounded uniformly continuous function. In particular, there exists a positive constant M_f such that

$$0 \leqslant f(x,a) \leqslant M_f, \quad \forall\, x \in E,\, a \in A.$$

The requirement that $Q(x,a,\{x\}) = 0$ for all $x \in E$, $a \in A$ is natural in many applications, but here is not needed. h, λ and Q depend on the control parameter $a \in A$ and play respectively the role of and controlled drift, controlled jump rate and controlled probability transition. Roughly speaking, we may control the dynamics of the process by changing dynamically its deterministic drift, its jump intensity and its post jump distribution.

Let us give a more precise definition of the optimal control problem under study. To this end, we first construct Ω, $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$, $\mathcal{F}_\infty$ as in the previous paragraph.

We will consider the class of *piecewise open-loop controls*, first introduced in Vermes [**129**] and often adopted in this context, see for instance [**35**], [**32**], [**1**]. Let X be the (uncontrolled) process constructed in a canonical way from a marked point process (T_n, E_n) as in Section 3.2.1. The class of admissible control law $\mathcal{A}_{ad}$ is the set of all Borel-measurable maps $\alpha : [0,\, \infty) \times E \to A$, and the control applied to X is of the form:

$$\alpha_t = \sum_{n=1}^{\infty} \alpha_n(t - T_n, E_n)\, 1_{[T_n,\, T_{n+1})}(t). \tag{3.22}$$

In other words, at each jump time T_n, we choose an open loop control α_n depending on the initial condition E_n to be used until the next jump time.

By abuse of notation, we define the controlled process $X : \Omega \times [0, \infty) \to E \cup \{\Delta\}$ setting

$$X_t = \phi^{\alpha_n}(t - T_n, E_n), \quad t \in [T_n, T_{n+1}) \tag{3.23}$$

where $\phi^{\beta}(t, x)$ is the unique solution to the ordinary differential equation

$$\dot{x}(t) = h(x(t), \beta(t)), \quad x(0) = x \in E.$$

with β an $\mathcal{A}$-measurable function. Then, for every starting point $x \in E$ and for each $\alpha \in \mathcal{A}_{ad}$, by Proposition 3.2.1 there exists a unique probability measure on $(\Omega, \mathcal{F}_\infty)$, denoted by $\mathbb{P}^x_\alpha$, such that its restriction to $\mathcal{F}_0$ is δ_x, and the $\mathbb{F}$-compensator under $\mathbb{P}^x_\alpha$ of the measure $p(ds\, dy)$ is

$$\tilde{p}^\alpha(ds\, dy) = \sum_{n=1}^{\infty} 1_{[T_n, T_{n+1})}(s)\, \lambda(X_s, \alpha_n(s - T_n, E_n))\, Q(X_s, \alpha_n(s - T_n, E_n), dy)\, ds.$$

According to Proposition 3.2.2, under $\mathbb{P}^x_\alpha$ the process X in (3.23) is Markovian with respect to $\mathbb{F}$.

Denoting by $\mathbb{E}^x_\alpha$ the expectation under $\mathbb{P}^x_\alpha$, we finally define, for $x \in E$ and $\alpha \in \mathcal{A}_{ad}$, the functional cost

$$J(x, \alpha) = \mathbb{E}^x_\alpha \left[\int_0^\infty e^{-\delta\, s} f(X_s, \alpha_s)\, ds \right] \tag{3.24}$$

and the value function of the control problem

$$V(x) = \inf_{\alpha \in \mathcal{A}_{ad}} J(x, \alpha), \tag{3.25}$$

where $\delta \in (0, \infty)$ is a discounting factor that will be fixed from here on. By the boundedness assumption on f, both J and V are well defined and bounded.

Let us consider the Hamilton-Jacobi-Bellman equation (for short, HJB equation) associated to the optimal control problem: this is the following elliptic nonlinear equation on $[0, \infty) \times E$:

$$H^v(x, v, Dv) = 0, \tag{3.26}$$

where

$$H^\psi(z, v, p) = \sup_{a \in A} \left\{ \delta\, v - h(z, a) \cdot p - \int_E (\psi(y) - \psi(z))\, \lambda(z, a)\, Q(z, a, dy) - f(z, a) \right\}.$$

Remark 3.2.4. The HJB equation (3.26) can be rewritten as

$$\delta\, v(x) = \sup_{a \in A} \{\mathcal{L}^a v(x) + f(x, a)\} = 0, \tag{3.27}$$

where $\mathcal{L}^a$ is the operator depending on $a \in A$ defined as

$$\mathcal{L}^a v(x) := h(x, a) \cdot \nabla v(x) + \lambda(x, a) \int_E (v(y) - v(x))\, Q(x, a, dy). \tag{3.28}$$

Let us recall the following facts. Given a locally bounded function $z : E \to \mathbb{R}$, we define its lower semicontinuous (l.s.c. for short) envelope z_*, and its upper semicontinuous (u.s.c. for short) envelope z^*, by

$$z_*(x) = \liminf_{\substack{y \to x \\ y \in E}} z(y), \qquad z^*(x) = \limsup_{\substack{y \to x \\ y \in E}} z(y), \quad \text{for all } x \in E.$$

Definition 3.2.5. Viscosity solution to (3.26).

(i) A locally bounded u.s.c. function w on E is called a *viscosity supersolution* (resp. *viscosity subsolution*) of (3.26) if

$$H^w(x_0, w(x_0), D\varphi(x_0)) \geqslant \ (resp. \ \leqslant) \ 0.$$

for any $x_0 \in E$ and for any $\varphi \in C^1(E)$ such that

$$(u - \varphi)(x_0) = \min_E (u - \varphi) \ (resp. \ \max_E (u - \varphi)).$$

(ii) A function z on E is called a *viscosity solution* of (3.26) if it is locally bounded and its u.s.c. and l.s.c. envelopes are respectively subsolution and supersolution of (3.26).

The HJB equation (3.26) admits a unique continuous solution, which coincides with the value function V in (3.25). The following result is stated in Theorem 7.5 in [**36**].

Theorem 3.2.6. *Let* **(HhλQ)** *and* **(Hf)** *hold, and assume that A is compact. Then the value function V of the PDMPs optimal control problem is the unique continuous viscosity solution of* (3.26).

3.3. Control randomization and dual optimal control problem

In this section we start to implement the control randomization method. In the first step, for an initial time $t \geq 0$ and a starting point $x \in E$, we construct an (uncontrolled) Markovian pair of PDMPs (X, I) by specifying its local characteristics, see (3.29)-(3.30)-(3.31) below. Next we formulate an auxiliary optimal control problem where, roughly speaking, we optimize a functional cost by modifying the intensity of the process I over a suitable family.

This dual problem is studied in Section 3.4 by means of a suitable class of BSDEs. In Section 3.5 we will show that the same class of BSDEs provides a probabilistic representation of the value function introduced in the previous section. As a byproduct, we also get that the dual value function coincides with the one associated to the original optimal control problem.

3.3.1. A dual control system. Let E still denote an open subset of $\mathbb{R}^d$ with σ-algebra $\mathcal{E}$, and A be a Borel space with corresponding σ-algebra $\mathcal{A}$. Let moreover h, λ and Q be respectively two real functions on $E \times A$ and a probability transition from $(E \times A, \mathcal{E} \otimes \mathcal{A})$, satisfying **(Hh$\lambda$Q)** as before. We denote by $\phi(t, x, a)$ the unique solution to the ordinary differential equation

$$\dot{x}(t) = h(x(t), a), \quad x(0) = x \in E, \ a \in A.$$

In particular, $\phi(t,x,a)$ corresponds to the function $\phi^\beta(t,x)$, introduced in Section 3.2.2, when $\beta(t) \equiv a$. Let now introduce another finite measure λ_0 on $(A,\mathcal{A})$ satisfying the following assumption:

(Hλ_0) λ_0 is a finite measure on $(A,\mathcal{A})$ with full topological support.

The existence of such a measure is guaranteed by the fact that the space A is metric separable. We define

$$\tilde{\phi}(t,x,a) := (\phi(t,x,a) \quad a), \tag{3.29}$$

$$\tilde{\lambda}(x,a) := \lambda(x,a) + \lambda_0(A), \tag{3.30}$$

$$\tilde{Q}(x,a,dy\,db) := \frac{\lambda(x,a)\,Q(x,a,dy)\,\delta_a(db) + \lambda_0(db)\,\delta_x(dy)}{\tilde{\lambda}(x,a)}. \tag{3.31}$$

We wish to construct a PDMP (X,I) as in Section 3.2.1 but with enlarged state space $E \times A$ and local characteristics $(\tilde{\phi}, \tilde{\lambda}, \tilde{Q})$.

Firstly, we need to introduce a suitable sample space to describe the jump mechanism of the process (X,I) on $E \times A$. Accordingly, we set Ω' as the set of sequences $\omega' = (t_n, e_n, a_n)_{n \geq 1}$ contained in $((0,\,\infty) \times E \times A) \cup \{(\infty, \Delta, \Delta')\}$, where $\Delta \notin E$ (resp. $\Delta' \notin A$) is adjoined to E (resp. to A) as an isolated point, satisfying (3.12). In the sample space $\Omega = \Omega' \times E \times A$ we defined the random variables $T_n : \Omega \to (0,\,\infty]$, $E_n : \Omega \to E \cup \{\Delta\}$, $A_n : \Omega \to A \cup \{\Delta'\}$, as follows: writing $\omega = (e,a,\omega')$ in the form $\omega = (e,a,t_1,e_1,a_1,t_2,e_2,a_2,...)$ we set for $t \geq 0$ and for $n \geq 1$

$$T_n(\omega) = t_n, \qquad T_\infty(\omega) = \lim_{n\to\infty} t_n, \qquad T_0(\omega) = 0,$$

$$E_n(\omega) = e_n, \qquad A_n(\omega) = a_n, \qquad E_0(\omega) = e, \qquad A_0(\omega) = a.$$

We define the process (X,I) on $(E \times A) \cup \{\Delta, \Delta'\}$ setting

$$(X,I)_t = \begin{cases} (\phi(t-T_n, E_n, A_n), A_n) & \text{if } T_n \leq t < T_{n+1}, \text{ for } n \in \mathbb{N}, \\ (\Delta, \Delta') & \text{if } t \geq T_\infty. \end{cases} \tag{3.32}$$

In Ω we introduce for all $t \geq 0$ the σ-algebras $\mathcal{G}_t = \sigma(N(s,B) \,:\, s \in (0,t], B \in \mathcal{E} \otimes \mathcal{A})$ generated by the counting processes $N(s,A) = \sum_{n\in\mathbb{N}} 1_{T_n \leq s} 1_{E_n \in A}$ and the σ-algebra $\mathcal{F}_t$ generated by $\mathcal{F}_0$ and $\mathcal{G}_t$, where $\mathcal{F}_0 = \mathcal{E} \otimes \mathcal{A} \otimes \{\emptyset, \Omega'\}$. We still denote by $\mathbb{F} = (\mathcal{F}_t)_{t\geq 0}$ and $\mathcal{P}$ the corresponding filtration and predictable σ-algebra. The random measure p is now defined on $(0,\,\infty) \times E \times A$ as

$$p(ds\,dy\,db) = \sum_{n\in\mathbb{N}} 1_{\{T_n, E_n, A_n\}}(ds\,dy\,db). \tag{3.33}$$

Given any starting point $(x,a) \in E \times A$, by Proposition 3.2.1, there exists a unique probability measure on $(\Omega, \mathcal{F}_\infty)$, denoted by $\mathbb{P}^{x,a}$, such that its restriction to $\mathcal{F}_0$ is $\delta_{(x,a)}$ and the $\mathbb{F}$-compensator of the measure $p(ds\,dy\,db)$ under $\mathbb{P}^{x,a}$ is the random measure

$$\tilde{p}(ds\,dy\,db) = \sum_{n\in\mathbb{N}} 1_{[T_n,\,T_{n+1})}(s)\,\Lambda(\phi(s-T_n, E_n, A_n), A_n, dy\,db)\,ds,$$

where

$$\Lambda(x,a,dy\,db) = \lambda(x,a)\,Q(x,a,dy)\,\delta_a(db) + \lambda_0(db)\,\delta_x(dy), \quad \forall (x,a) \in E \times A.$$

We indicate by $q = p - \tilde{p}$ the compensated martingale measure associated to p.

As in Section 3.2.1, the sample path of a process (X, I) with values in $E \times A$, starting from a fixed initial point $(x, a) \in E \times A$ at time zero, can be defined iteratively by means of its local characteristics $(\tilde{h}, \tilde{\lambda}, \tilde{Q})$ in the following way. Set

$$F(s, x, a) = \exp\left(- \int_0^s (\lambda(\phi(r, x, a), a) + \lambda_0(A))\, dr \right),$$

we have

$$\mathbb{P}^{x,a}(T_1 > s) = F(s, x, a), \tag{3.34}$$

$$\mathbb{P}^{x,a}(X_{T_1} \in B, I_{T_1} \in C |\, T_1) = \tilde{Q}(x, B \times C), \tag{3.35}$$

on $\{T_1 < \infty\}$, and, for every $n > 1$,

$$\mathbb{P}^{x,a}(T_{n+1} > s \,|\, \mathcal{F}_{T_n}) = \exp\left(- \int_{T_n}^s (\lambda(\phi(r - T_n, X_{T_n}, I_{T_n}), I_{T_n}) + \lambda_0(A))\, dr \right), \tag{3.36}$$

$$\mathbb{P}^{x,a}(X_{T_{n+1}} \in B, I_{T_{n+1}} \in C |\, \mathcal{F}_{T_n}, T_{n+1}) = \tilde{Q}(\phi(T_{n+1} - T_n, X_{T_n}, I_{T_n}), I_{T_n}, B \times C), \tag{3.37}$$

on $\{T_n < \infty\}$.

Finally, an application of Proposition 3.2.2 provides that (X, I) is a Markov process on $[0, \infty)$ with respect to $\mathbb{F}$. For every real function taking values in $E \times A$, the infinitesimal generator is given by

$$\begin{aligned}\mathcal{L}\varphi(x, a) :=& h(x, a) \cdot \nabla_x \varphi(x, a) + \int_E (\varphi(y, a) - \varphi(x, a))\, \lambda(x, a)\, Q(x, a, dy) \\ &+ \int_A (\varphi(x, b) - \varphi(x, a))\, \lambda_0(db).\end{aligned}$$

For our purposes, it will be not necessary to specify the domain of the previous operator (for its formal definition we refer to Theorem 26.14 in [**35**]); in the sequel the operator L will be applied to test functions with suitable regularity.

3.3.2. The dual optimal control problem. We now introduce a dual optimal control problem associated to the process (X, I), and formulated in a weak form. For fixed (x, a), we consider a family of probability measures $\{\mathbb{P}_\nu^{x,a},\, \nu \in \mathcal{V}\}$ in the space $(\Omega, \mathcal{F}_\infty)$, whose effect is to change the stochastic intensity of the process (X, I).

Let us proceed with precise definitions. We still assume that **(HhλQ)**, **(Hλ_0)** and **(Hf)** hold. We recall that $\mathbb{F} = (\mathcal{F}_t)_{t \geqslant 0}$ is the augmentation of the natural filtration generated by p in (3.33). We define

$$\mathcal{V} = \{\nu : \Omega \times [0, \infty) \times A \to (0, \infty)\ \mathcal{P} \otimes \mathcal{A}\text{-measurable and bounded}\}.$$

For every $\nu \in \mathcal{V}$, we consider the predictable random measure

$$\begin{aligned}\tilde{p}^\nu(ds\, dy\, db) :=& \nu_s(b)\, \lambda_0(db)\, \delta_{\{X_{s-}\}}(dy)\, ds \\ &+ \lambda(X_{s-},\, I_{s-})\, Q(X_{s-},\, I_{s-},\, dy)\, \delta_{\{I_{s-}\}}(db)\, ds.\end{aligned} \tag{3.38}$$

In particular, by the Radon Nikodym theorem one can find two nonnegative functions d_1, d_2 defined on $\Omega \times [0, \infty) \times E \times A$, $\mathcal{P} \otimes \mathcal{E} \otimes \mathcal{A}$, such that

$$\begin{aligned}\lambda_0(db)\, \delta_{\{X_{t-}\}}(dy)\, dt &= d_1(t, y, b)\, \tilde{p}(dt\, dy\, db)\\ \lambda(X_{t-}, I_{t-}, dy)\, \delta_{\{I_{t-}\}}(db)\, dt &= d_2(t, y, b)\, \tilde{p}(dt\, dy\, db),\\ d_1(t, y, b) + d_2(t, y, b) &= 1, \qquad \tilde{p}(dt\, dy\, db) - a.e.\end{aligned}$$

and we have $d\tilde{p}^\nu = (\nu\, d_1 + d_2)\, d\tilde{p}$. For any $\nu \in \mathcal{V}$, consider then the Doléans-Dade exponential local martingale L^ν defined setting

$$\begin{aligned}L^\nu_s = \exp\bigg(&\int_0^s \int_{E\times A} \log(\nu_r(b)\, d_1(r, y, b) \,+ d_2(r, y, b))\, p(dr\, dy\, db)\\ &- \int_0^s \int_A (\nu_r(b) - 1)\lambda_0(db)\, dr\bigg)\\ = e^{\int_0^s \int_A (1-\nu_r(b))\lambda_0(db)\, dr} &\prod_{n\geqslant 1: T_n \leqslant s} (\nu_{T_n}(A_n)\, d_1(T_n, E_n, A_n) + d_2(T_n, E_n, A_n)),\end{aligned} \tag{3.39}$$

for $s \geq 0$. When $(L^\nu_t)_{t\geq 0}$ is a true martingale, for every time $T > 0$ we can define a probability measure $\mathbb{P}^{x,a}_{\nu,T}$ equivalent to $\mathbb{P}^{x,a}$ on $(\Omega, \mathcal{F}_T)$ setting

$$\mathbb{P}^{x,a}_{\nu,T}(d\omega) = L^\nu_T(\omega)\, \mathbb{P}^{x,a}(d\omega). \tag{3.40}$$

By the Girsanov theorem for point processes (see Theorem 4.5 in [**75**]) the restriction of the random measure p to $(0, T] \times E \times A$ admits $\tilde{p}^\nu = (\nu\, d_1 + d_2)\, \tilde{p}$ as compensator under $\mathbb{P}^{x,a}_{\nu,T}$. We set $q^\nu := p - \tilde{p}^\nu$. and we denote by $\mathbb{E}^{x,a}_{\nu,T}$ the expectation operator under $\mathbb{P}^{x,a}_{\nu,T}$. Previous considerations are formalized in the following Lemma, which is a direct consequence of Lemma 2.3.2.

Lemma 3.3.1. *Let assumptions* **(HhλQ)** *and* **(Hλ_0)** *hold. Then, for every $(x, a) \in E \times A$ and $\nu \in \mathcal{V}$, under the probability $\mathbb{P}^{x,a}$, the process $(L^\nu_t)_{t\geq 0}$ is a martingale. Moreover, for every time $T > 0$, L^ν_T is square integrable, and, for every $\mathcal{P}_T \otimes \mathcal{E} \otimes \mathcal{A}$-measurable function $H : \Omega \times [0, T] \times E \times A \to \mathbb{R}$ such that $\mathbb{E}^{x,a}\left[\int_0^T \int_{E\times A} |H_s(y, b)|^2\, \tilde{p}(ds\, dy\, db)\right] < \infty$, the process $\int_0^\cdot \int_{E\times A} H_s(y, b)\, q^\nu(ds\, dy\, db)$ is a $\mathbb{P}^{x,a}_{\nu,T}$-martingale on $[0, T]$.*

We aim at extending the previous construction to the infinite horizon, in order to get a suitable probability measure on $(\Omega, \mathcal{F}_\infty)$. We have the following result.

Proposition 3.3.2. *Let assumptions* **(HhλQ)** *and* **(Hλ_0)** *hold. Then, for every $(x, a) \in E \times A$ and $\nu \in \mathcal{V}$, there exists a unique probability $\mathbb{P}^{x,a}_\nu$ on $(\Omega, \mathcal{F}_\infty)$, under which the random measure $\tilde{p}^\nu$ in* (3.38) *is the compensator of the measure p in* (3.33) *on $(0, \infty) \times E \times A$. Moreover, for any time $T > 0$, the restriction of $\mathbb{P}^{x,a}_\nu$ on $(\Omega, \mathcal{F}_T)$ is given by the probability measure $\mathbb{P}^{x,a}_{\nu,T}$ in* (3.40)*.*

Proof. For simplicity, in the sequel we will drop the dependence of $\mathbb{P}^{x,a}$ and $\mathbb{P}^{x,a}_\nu$ on (x, a), which will be denoted respectively by $\mathbb{P}$ and $\mathbb{P}^\nu$.

We notice that $\mathcal{F}_{T_n} = \sigma(T_1, E_1, A_1, ..., T_n, E_n, A_n)$ defines an increasing family of sub σ-fields of $\mathcal{F}_\infty$ such that $\mathcal{F}_\infty$ is generated by $\bigcup_n \mathcal{F}_{T_n}$. The idea is then to provide a family $\{\mathbb{P}_n^\nu\}_n$ of probability measures on $(\Omega, \mathcal{F}_{T_n})$ under which $\tilde{p}^\nu$ is the compensator of the measure p on $(0,\, T_n] \times E \times A$, and which is consistent (i.e., $\mathbb{P}_{n+1}^\nu\big|_{\mathcal{F}_{T_n}} = \mathbb{P}_n^\nu$). Indeed, if we have at disposal such a family of probabilities, we can naturally define on $\bigcup_n \mathcal{F}_{T_n}$ a set function $\mathbb{P}^\nu$ verifying the desired property, by setting $\mathbb{P}^\nu(B) := \mathbb{P}_n^\nu(B)$ for every $B \in \mathcal{F}_{T_n}$, $n \geq 1$. Finally, to conclude we would need to show that $\mathbb{P}^\nu$ is countably additive on $\bigcup_n \mathcal{F}_{T_n}$, and therefore can be extended uniquely to $\mathcal{F}_\infty$.

Let us proceed by steps. For every $n \in \mathbb{N}$, we set

$$d\mathbb{P}_n^\nu := L_{T_n}^\nu \, d\mathbb{P} \quad \text{on } (\Omega, \mathcal{F}_{T_n}), \tag{3.41}$$

where L^ν is given by (3.39). Notice that, for every $n \in \mathbb{N}$, the probability $\mathbb{P}_n^\nu$ is well defined. Indeed, recalling the boundedness properties of ν and λ_0, we have

$$\begin{aligned} L_{T_n}^\nu &= e^{\int_0^{T_n} \int_A (1-\nu_r(b))\lambda_0(db)\, dr} \prod_{k=1}^n (\nu_{T_k}(A_k)\, d_1(T_k, E_k, A_k) + d_2(T_k, E_k, A_k)) \\ &\leq (||\nu||_\infty)^n \, e^{\lambda_0(A)\, T_n}, \end{aligned} \tag{3.42}$$

and since T_n is exponentially distributed (see (3.17)), we get

$$\mathbb{E}\left[L_{T_n}^\nu\right] \leq (||\nu||_\infty)^n \, \mathbb{E}\left[e^{\lambda_0(A)\, T_n}\right] < \infty.$$

Then, arguing as in the proof of the Girsanov theorem for point process (see, e.g., the comments after Theorem 4.5 in [**75**]), it can be proved that the restriction of the random measure p to $(0, T_n] \times E \times A$ admits $\tilde{p}^\nu = (\nu\, d_1 + d_2)\, \tilde{p}$ as compensator under $\mathbb{P}_n^\nu$. Moreover, $\{\mathbb{P}_n^\nu\}_n$ is a consistent family of probability measures on $(\Omega, \mathcal{F}_{T_n})$, namely

$$\mathbb{P}_{n+1}^\nu\big|_{\mathcal{F}_{T_n}} = \mathbb{P}_n^\nu, \quad n \in \mathbb{N}. \tag{3.43}$$

Indeed, taking into account definition (3.41), it is easy to see that identity (3.43) is equivalent to

$$\mathbb{E}\left[L_{T_n}^\nu | \mathcal{F}_{T_{n-1}}\right] = L_{T_{n-1}}^\nu, \quad n \in \mathbb{N}. \tag{3.44}$$

By Corollary 3.6, Chapter II, in Revuz and Yor [**111**], and taking into account the estimate (3.42), it follows that the process $(L_{t \wedge T_n}^\nu)_{t \geq 0}$ is a uniformly integrable martingale. Then, identity (3.44) follows from the optional stopping theorem for uniformly integrable martingales (see, e.g., Theorem 3.2, Chapter II, in [**111**]).

At this point, we define the following probability measure on $\bigcup_n \mathcal{F}_{T_n}$:

$$\mathbb{P}^\nu(B) := \mathbb{P}_n^\nu(B), \quad B \in \mathcal{F}_{T_n},\ n \in \mathbb{N}. \tag{3.45}$$

In order to get the desired probability measure on $(\Omega, \mathcal{F}_\infty)$, we need to show that $\mathbb{P}^\nu$ in (3.45) is σ-additive on $\bigcup_n \mathcal{F}_{T_n}$: in this case, $\mathbb{P}^\nu$ can indeed be extended uniquely to $\mathcal{F}_\infty$, see Theorem 6.1 in Jacod and Protter [**78**].

Let us then prove that $\mathbb{P}^\nu$ in (3.45) is countably additive on $\bigcup_n \mathcal{F}_{T_n}$. To this end, let us introduce the product space $\tilde{E}_\Delta^{\mathbb{N}} := (E \times A \times [0,\, \infty) \cup \{(\Delta, \Delta', \infty)\})^{\mathbb{N}}$, with associated Borel σ-algebra $\tilde{\mathcal{E}}_\Delta^{\mathbb{N}\otimes}$. For every $n \in \mathbb{N}$, we define the following probability measure on $(\tilde{E}_\Delta^n, \tilde{\mathcal{E}}_\Delta^{n\otimes})$:

$$\mathbb{Q}_n^\nu(A) := \mathbb{P}_n^\nu(\omega : \pi_n(\omega) \in A), \quad A \in \tilde{E}_\Delta^n, \tag{3.46}$$

where $\pi_n = (T_1, E_1, A_1, ..., T_n, E_n, A_n)$. The consistency property (3.43) of the family $(\mathbb{P}_n^\nu)_n$ implies that

$$\mathbb{Q}_{n+1}^\nu(A \times \tilde{E}_\Delta) = \mathbb{Q}_{n+1}^\nu(A), \quad A \in \tilde{E}_\Delta^n. \tag{3.47}$$

Let now define

$$\begin{aligned} &\mathcal{A} := \{A \times \tilde{E}_\Delta \times \tilde{E}_\Delta \times ... : A \in \tilde{E}_\Delta^n,\ n \geq 0\}, \\ &\mathbb{Q}^\nu(A \times \tilde{E}_\Delta \times \tilde{E}_\Delta \times ...) := \mathbb{Q}_n^\nu(A), \quad A \in \tilde{E}_\Delta^n, n \geq 0. \end{aligned} \tag{3.48}$$

By the Kolmogorov extension theorem for product spaces (see Theorem 1.1.10 in Strook and Varadhan [**126**]), it follows that $\mathbb{Q}^\nu$ is σ-additive on $\mathcal{A}$. Then, collecting (3.45), (3.46) and (3.48), it is easy to see that the σ-additivity of $\mathbb{Q}^\nu$ on $\mathcal{A}$ implies the σ-additivity of $\mathbb{P}^\nu$ on $\bigcup_n \mathcal{F}_{T_n}$.

Finally, we need to show that

$$\mathbb{P}^\nu\big|_{\mathcal{F}_T} = L_T^\nu\,\mathbb{P} \quad \forall T > 0,$$

or, equivalently, that

$$\mathbb{E}[L_T^\nu\,\psi] = \mathbb{E}^\nu[\psi] \quad \forall\psi\ \mathcal{F}_T\text{-measurable function.}$$

To this end, fix $T > 0$, and let ψ be a $\mathcal{F}_{T\wedge T_n}$-measurable bounded function. In particular, ψ is $\mathcal{F}_{T\wedge T_m}$-measurable, for every $m \geq n$. Since by definition $\mathbb{P}^\nu\big|_{\mathcal{F}_{T_n}} = L_{T_n}^\nu\,\mathbb{P}, n \in \mathbb{N}$, we have

$$\begin{aligned} \mathbb{E}^\nu[\psi] &= \mathbb{E}[L_{T_m}^\nu\,\psi] \\ &= \mathbb{E}[\mathbb{E}[L_{T_m}^\nu\,\psi|\mathcal{F}_{T\wedge T_m}]] \\ &= \mathbb{E}[\psi\,\mathbb{E}[L_{T_m}^\nu|\mathcal{F}_{T\wedge T_m}]] \\ &= \mathbb{E}[\psi\,L_{T\wedge T_m}^\nu] \quad \forall m \geq n. \end{aligned}$$

Since $L_{T\wedge T_m}^\nu \underset{m\to\infty}{\longrightarrow} L_T^\nu$ a.s., and $(L_s^\nu)_{s\in[0,T]}$ is a uniformly integrable martingale, by Theorem 3.1, Chapter II, in [**111**], we get

$$\mathbb{E}^\nu[\psi] = \lim_{m\to\infty} \mathbb{E}[L_{T\wedge T_m}^\nu\,\psi] = \mathbb{E}[L_T^\nu\,\psi], \quad \forall\psi \in \bigcup_n \mathcal{F}_{T\wedge T_n}.$$

Then, by the monotone class theorem, recalling that $\bigvee_n \mathcal{F}_{T\wedge T_n} = \mathcal{F}_{\bigvee_n \mathcal{F}_{T\wedge T_n}}$ (see, e.g., Corollary 3.5, point 6, in He, Wang and Yan [**73**]), we get

$$\mathbb{E}^\nu[\psi] = \mathbb{E}[L_T^\nu\,\psi], \quad \forall\psi \in \bigvee_n \mathcal{F}_{T\wedge T_n} = \mathcal{F}_{\bigvee_n \mathcal{F}_{T\wedge T_n}} = \mathcal{F}_T.$$

This concludes the proof. □

Finally, for every $x \in E$, $a \in A$ and $\nu \in \mathcal{V}$, we introduce the dual functional cost

$$J(x, a, \nu) := \mathbb{E}_\nu^{x,a}\left[\int_0^\infty e^{-\delta\,t}\,f(X_t, I_t)\,dt\right], \tag{3.49}$$

and the dual value function

$$V^*(x, a) := \inf_{\nu\in\mathcal{V}} J(x, a, \nu), \tag{3.50}$$

where $\delta > 0$ in (3.49) is the discount factor introduced in Section 3.2.2.

3.4. Constrained BSDEs and the dual value function representation

In this section we introduce a BSDE with a sign constrain on its martingale part, for which we prove the existence and uniqueness of a maximal solution, in an appropriate sense. This constrained BSDE is then used to give a probabilistic representation formula for the dual value function introduced in (3.50).

Throughout this section we still assume that **(HhλQ)**, **(Hλ_0)** and **(Hf)** hold. The random measures p, $\tilde{p}$ and q, as well as the dual control setting $\Omega, \mathbb{F}, (X, I), \mathbb{P}^{x,a}$, are the same as in Section 3.3.1. We recall that $\mathbb{F} = (\mathcal{F}_t)_{t \geqslant 0}$ is the augmentation of the natural filtration generated by p, and that $\mathcal{P}_T$, $T > 0$, denotes the σ-field of $\mathbb{F}$-predictable subsets of $[0, T] \times \Omega$.

For any $(x, a) \in E \times A$ we introduce the following notation.

- $\mathbf{L^2_{x,a}}(\mathcal{F}_\tau)$, the set of $\mathcal{F}_\tau$-measurable random variables ξ such that $\mathbb{E}^{x,a}\left[|\xi|^2\right] < \infty$; here $\tau \geqslant 0$ is an $\mathbb{F}$-stopping time.
- $\mathbf{S^\infty}$ the set of real-valued càdlàg adapted processes $Y = (Y_t)_{t \geqslant 0}$ which are uniformly bounded.
- $\mathbf{S^2_{x,a}(0, T)}$, $T > 0$, the set of real-valued càdlàg adapted processes $Y = (Y_t)_{0 \leqslant t \leqslant T}$ satisfying
$$||Y||_{\mathbf{S^2_{x,a}(0,\,T)}} := \mathbb{E}^{x,a}\left[\sup_{0 \leqslant t \leqslant T} |Y_t|^2\right] < \infty.$$
- $\mathbf{L^2_{x,a}(0, T)}$, $T > 0$, the set of real-valued progressive processes $Y = (Y_t)_{0 \leqslant t \leqslant T}$ such that
$$||Y||^2_{\mathbf{L^2_{x,a}(0,\,T)}} := \mathbb{E}^{x,a}\left[\int_0^T |Y_t|^2\, dt\right] < \infty.$$
We also define $\mathbf{L^2_{x,a,loc}} := \cap_{T>0} \mathbf{L^2_{x,a}(0, T)}$.
- $\mathbf{L^2_{x,a}}(\mathrm{q}; \mathbf{0, T})$, $T > 0$, the set of $\mathcal{P}_T \otimes \mathcal{B}(E) \otimes \mathcal{A}$-measurable maps $Z : \Omega \times [0, T] \times E \times A \to \mathbb{R}$ such that
$$\begin{aligned} ||Z||^2_{\mathbf{L^2_{x,a}}(\mathrm{q};\mathbf{0,\,T})} &:= \mathbb{E}^{x,a}\Big[\int_0^T \int_{E \times A} |Z_t(y, b)|^2\, \tilde{p}(dt\, dy\, db)\Big] \\ &= \mathbb{E}^{x,a}\Big[\int_0^T \int_E |Z_t(y, I_t)|^2\, \lambda(X_t, I_t)\, Q(X_t, I_t, dy)\, dt\Big] \\ &+ \mathbb{E}^{x,a}\Big[\int_0^T \int_A |Z_t(X_t, b)|^2\, \lambda_0(db)\, dt\Big] < \infty. \end{aligned}$$
We also define $\mathbf{L^2_{x,a,loc}}(\mathrm{q}) := \cap_{T>0} \mathbf{L^2_{x,a}}(\mathrm{q}; \mathbf{0, T})$.
- $\mathbf{L^2}(\lambda_0)$, the set of $\mathcal{A}$-measurable maps $\psi : A \to \mathbb{R}$ such that
$$|\psi|^2_{\mathbf{L^2}(\lambda_0)} := \int_A |\psi(b)|^2\, \lambda_0(db) < \infty.$$

- $\mathbf{L^2_{x,a}}(\lambda_0; \mathbf{0, T})$, $T > 0$, the set of $\mathcal{P}_T \otimes \mathcal{A}$-measurable maps $W : \Omega \times [0, T] \times A \to \mathbb{R}$ such that

$$|W|^2_{\mathbf{L^2_{x,a}}(\lambda_0;\mathbf{0,T})} := \mathbb{E}^{x,a}\left[\int_0^T \int_A |W_t(b)|^2 \, \lambda_0(db)\, dt\right] < \infty.$$

We also define $\mathbf{L^2_{x,a,loc}}(\lambda_0) := \cap_{T>0} \mathbf{L^2_{x,a}}(\lambda_0; \mathbf{0, T})$.

- $\mathbf{K^2_{x,a}}(\mathbf{0, T})$, $T > 0$, the set of nondecreasing càdlàg predictable processes $K = (K_t)_{0 \leqslant t \leqslant T}$ such that $K_0 = 0$ and $\mathbb{E}^{x,a}\left[|K_T|^2\right] < \infty$. We also define $\mathbf{K^2_{x,a,loc}} := \cap_{T>0} \mathbf{K^2_{x,a}}(\mathbf{0, T})$.

We are interested in studying the following family of BSDEs with partially nonnegative jumps over an infinite horizon, parametrized by (x, a): $\mathbb{P}^{x,a}$-a.s.,

$$Y^{x,a}_s = Y^{x,a}_T - \delta \int_s^T Y^{x,a}_r \, dr + \int_s^T f(X_r, I_r)\, dr - (K^{x,a}_T - K^{x,a}_s) \tag{3.51}$$
$$- \int_s^T \int_A Z^{x,a}_r(X_r, b)\, \lambda_0(db)\, dr - \int_s^T \int_{E\times A} Z^{x,a}_r(y, b)\, q(dr\, dy\, db), \quad 0 \leqslant s \leqslant T < \infty,$$

with

$$Z^{x,a}_s(X_{s-}, b) \geqslant 0, \qquad ds \otimes d\mathbb{P}^{x,a} \otimes \lambda_0(db), \text{-a.e. on } [0, \infty) \times \Omega \times A, \tag{3.52}$$

where δ is the positive parameter introduced in Section 3.2.2.

We look for a *maximal solution* $(Y^{x,a}, Z^{x,a}, K^{x,a}) \in \mathbf{S}^\infty \times \mathbf{L^2_{x,a,loc}}(\mathrm{q}) \times \mathbf{K^2_{x,a,loc}}$ to (3.51)-(3.52), in the sense that for any other solution $(\tilde{Y}, \tilde{Z}, \tilde{K}) \in \mathbf{S}^\infty \times \mathbf{L^2_{x,a,loc}}(\mathrm{q}) \times \mathbf{K^2_{x,a,loc}}$ to (3.51)-(3.52), we have $Y^{x,a}_t \geqslant \tilde{Y}_t$, $\mathbb{P}^{x,a}$-a.s., for all $t \geqslant 0$.

Proposition 3.4.1. *Let Hypotheses* **(HhλQ)**, **(Hλ_0)** *and* **(Hf)** *hold. Then, for any $(x, a) \in E \times A$, there exists at most one maximal solution $(Y^{x,a}, Z^{x,a}, K^{x,a}) \in \mathbf{S}^\infty \times \mathbf{L^2_{x,a,loc}}(\mathrm{q}) \times \mathbf{K^2_{x,a,loc}}$ to the BSDE with partially nonnegative jumps* (3.51)-(3.52).

Proof. Let (Y, Z, K) and (Y', Z', K') be two maximal solutions of (3.51)-(3.52). By definition, we clearly have the uniqueness of the component Y. Regarding the other components, taking the difference between the two backward equations we obtain: $\mathbb{P}^{x,a}$-a.s.

$$\begin{aligned} 0 &= -(K_t - K'_t) - \int_0^t \int_A (Z_s(X_s, b) - Z'_s(X_s, b))\, \lambda_0(db)\, ds \\ &\quad - \int_0^t \int_{E\times A} (Z_s(y, b) - Z'_s(y, b))\, q(ds\, dy\, db), \quad 0 \leqslant t \leqslant T < \infty, \end{aligned}$$

that can be rewritten as

$$\int_0^t \int_{E\times A} (Z_s(y, b) - Z'_s(y, b))\, p(ds\, dy\, db) = -(K_t - K'_t) \tag{3.53}$$
$$+ \int_0^t \int_E (Z_s(y, I_s) - Z'_s(y, I_s))\, \lambda(X_s, I_s)\, Q(X_s, I_s, dy)\, ds, \quad 0 \leqslant t \leqslant T < \infty.$$

The right-hand side of (3.53) is a predictable process, therefore it has no totally inaccessible jumps (see, e.g., Proposition 2.24, Chapter I, in [**79**]); on the other hand, the left side is a pure jump process with totally inaccessible jumps. This implies that $Z = Z'$, and as a consequence the component K is unique as well. □

In the sequel, we prove by a penalization approach the existence of the maximal solution to (3.51)-(3.52). In particular, this will provide a probabilistic representation of the dual value function V^* introduced in Section 3.3.2.

3.4.1. Penalized BSDE and associated dual control problem. Let us introduce the family of penalized BSDEs on $[0, \infty)$ associated to (3.51)-(3.52), parametrized by the integer $n \geqslant 1$: $\mathbb{P}^{x,a}$-a.s.,

$$\begin{aligned} Y_s^{n,x,a} &= Y_T^{n,x,a} - \delta \int_s^T Y_r^{n,x,a}\, dr + \int_s^T f(X_r, I_r)\, dr \\ &\quad -n \int_s^T \int_A [Z_r^{n,x,a}(X_r, b)]^- \,\lambda_0(db)\, dr - \int_s^T \int_A Z_r^{n,x,a}(X_r, b)\, \lambda_0(db)\, dr \\ &\quad - \int_s^T \int_{E\times A} Z_r^{n,x,a}(y, b)\, q(dr\, dy\, db), \quad 0 \leqslant s \leqslant T < \infty, \end{aligned} \tag{3.54}$$

where $[z]^- = \max(-z, 0)$ denotes the negative part of z.

We shall prove that there exists a unique solution to equation (3.54), and provide an explicit representation to (3.54) in terms of a family of dual control problems. To this end, we start by considering, for fixed $T > 0$, the family of BSDEs on $[0, T]$: $\mathbb{P}^{x,a}$-a.s.,

$$\begin{aligned} Y_s^{T,n,x,a} &= -\delta \int_s^T Y_r^{T,n,x,a}\, dr + \int_s^T f(X_r, I_r)\, dr \\ &\quad -n \int_s^T \int_A [Z_r^{T,n,x,a}(X_r, b)]^- \,\lambda_0(db)\, dr - \int_s^T \int_A Z_r^{T,n,x,a}(X_r, b)\, \lambda_0(db)\, dr \\ &\quad - \int_s^T \int_{E\times A} Z_r^{T,n,x,a}(y, b)\, q(dr\, dy\, db), \quad 0 \leqslant s \leqslant T, \end{aligned} \tag{3.55}$$

with zero final cost at time $T > 0$.

Remark 3.4.2. The penalized BSDE (3.55) can be rewritten in the equivalent form: $\mathbb{P}^{x,a}$-a.s.,

$$Y_s^{T,n,x,a} = \int_s^T f^n(X_r,\, I_r,\, Y_r^{T,n,x,a},\, Z_r^{T,n,x,a})\, ds - \int_s^T \int_{E\times A} Z_r^{T,n,x,a}(y, b)\, q(dr\, dy\, db),$$

$s \in [0, T]$, where the generator f^n is defined by

$$f^n(x, a, u, \psi) := f(x, a) - \delta u - \int_A \left\{ n\, [\psi(a)]^- + \psi(b) \right\}\, \lambda_0(db), \tag{3.56}$$

for all $(x, a, u, \psi) \in E \times A \times \mathbb{R} \times \mathbf{L^2}(\lambda_0)$.

We notice that, under Hypotheses **(HhλQ)**, **(Hλ_0)** and **(Hf)**, f^n is Lipschitz continuous in ψ with respect to the norm of $\mathbf{L^2}(\lambda_0)$, uniformly in (x, a, u), i.e., for every $n \in \mathbb{N}$, there exists a constant L_n, depending only on n, such that for every $(x, a, u) \in E \times A \times \mathbb{R}$ and ψ, $\psi' \in \mathbf{L^2}(\lambda_0)$,

$$|f^n(x, a, u, \psi') - f^n(x, a, u, \psi)| \leqslant L^n |\psi - \psi'|_{\mathbf{L^2}(\lambda_0)}.$$

For every integer $n \geqslant 1$, let $\mathcal{V}^n$ denote the subset of elements $\nu \in \mathcal{V}$ valued in $(0, n]$.

Proposition 3.4.3. *Let Hypotheses* **(HhλQ)**, **(Hλ_0)** *and* **(Hf)** *hold. For every* $(x, a, n, T) \in E \times A \times \mathbb{N} \times (0, \infty)$, *there exists a unique solution* $(Y^{T,n,x,a}, Z^{T,n,x,a}) \in \mathbf{S}^\infty \times \mathbf{L}^{\mathbf{2}}_{\boldsymbol{x,a}}(\mathrm{q}; \mathbf{0}, \mathbf{T})$ *to* (3.55). *Moreover, the following uniform estimate holds:* $\mathbb{P}^{x,a}$-*a.s.,*

$$Y_s^{T,n,x,a} \leqslant \frac{M_f}{\delta}, \quad \forall\, s \in [0, T]. \tag{3.57}$$

Proof. The existence and uniqueness of a solution $(Y^{T,n,x,a}, Z^{T,n,x,a}) \in \mathbf{S}^{\mathbf{2}}_{\mathbf{x,a}}(\mathbf{0}, \mathbf{T}) \times \mathbf{L}^{\mathbf{2}}_{\mathbf{x,a}}(\mathrm{q}; \mathbf{0}, \mathbf{T})$ to (3.55) is based on a fixed point argument, and uses integral representation results for $\mathbb{F}$-martingales, with $\mathbb{F}$ the natural filtration (see, e.g., Theorem 5.4 in [**75**]). This procedure is standard and we omit it (similar proofs can be found in the proofs of Theorem 3.2 in [**131**], Proposition 3.2 in [**12**], Theorem 3.4 in [**28**]). It remains to prove uniform estimate (3.57). To this end, let us apply Itô's formula to $e^{-\delta\, r}\, Y_r^{T,n,x,a}$ between s and T. We get: $\mathbb{P}^{x,a}$-a.s.

$$\begin{aligned} Y_s^{T,n,x,a} &= \int_s^T e^{-\delta\,(r-s)}\, f(X_r, I_r)\, dr - \int_s^T \int_{E\times A} e^{-\delta\,(r-s)}\, Z_r^{T,n,x,a}(y, b)\, q(dr\, dy\, db) \\ &- \int_s^T \int_A e^{-\delta\,(r-s)}\, \{n[Z_r^{T,n,x,a}(X_r, b)]^- + Z_r^{T,n,x,a}(X_r, b)\}\, \lambda_0(db)\, dr, \quad s \in [0, T]. \end{aligned} \tag{3.58}$$

Now for any $\nu \in \mathcal{V}^n$, let us introduce the compensated martingale measure $q^\nu(ds\, dy\, db) = q(ds\, dy\, db) - (\nu_s(b) - 1)\, d_1(s, y, b)\, \tilde{p}(ds\, dy\, db)$ under $\mathbb{P}^{x,a}_\nu$. Taking the expectation in (3.58) under $\mathbb{P}^{x,a}_\nu$, conditional to $\mathcal{F}_s$, and since $Z^{T,n,x,a}$ is in $\mathbf{L}^{\mathbf{2}}_{\mathbf{x,a}}(\mathrm{q}; \mathbf{0}, \mathbf{T})$, from Lemma 3.3.1 we get that, $\mathbb{P}^{x,a}$-a.s.,

$$\begin{aligned} &Y_s^{T,n,x,a} \\ &= -\mathbb{E}^{x,a}_\nu \left[\int_s^T \int_A e^{-\delta\,(r-s)}\, \{n[Z_r^{T,n,x,a}(X_r, b)]^- + \nu_r(b)\, Z_r^{T,n,x,a}(X_r, b)\}\, \lambda_0(db)\, dr \Big| \mathcal{F}_s \right] \\ &+ \mathbb{E}^{x,a}_\nu \left[\int_s^T e^{-\delta\,(r-s)}\, f(X_r, I_r)\, dr \Big| \mathcal{F}_s \right], \quad s \in [0, T]. \end{aligned} \tag{3.59}$$

From the elementary numerical inequality: $n[z]^- + \nu z \geqslant 0$ for all $z \in \mathbb{R}$, $\nu \in (0, n]$, we deduce by (3.59) that, for all $\nu \in \mathcal{V}^n$,

$$Y_s^{T,n,x,a} \leqslant \mathbb{E}^{x,a}_\nu \left[\int_s^T e^{-\delta\,(r-s)}\, f(X_r, I_r)\, dr \Big| \mathcal{F}_s \right], \quad s \in [0, T].$$

Therefore, $\mathbb{P}^{x,a}$-a.s.,

$$Y_s^{T,n,x,a} \leqslant \mathbb{E}^{x,a}_\nu \left[\int_s^\infty e^{-\delta\,(r-s)}\, |f(X_r, I_r)|\, dr \Big| \mathcal{F}_s \right] \leqslant \frac{M_f}{\delta}, \quad s \in [0, T].$$

□

Proposition 3.4.4. *Let Hypotheses* **(HhλQ)**, **(Hλ_0)** *and* **(Hf)** *hold. Then, for every* $(x,a,n) \in E \times A \times \mathbb{N}$, *there exists a unique solution* $(Y^{n,x,a}, Z^{n,x,a}) \in \mathbf{S}^\infty \times \mathbf{L}^2_{x,a,\mathbf{loc}}(\mathrm{q})$ *to* (3.54).

Proof. *Uniqueness.* Fix $n \in \mathbb{N}$, $(x,a) \in E\times A$, and consider two solutions $(Y^1, Z^1) = (Y^{1,n,x,a}, Z^{1,n,x,a})$, $(Y^2, Z^2) = (Y^{2,n,x,a}, Z^{2,n,x,a}) \in \mathbf{S}^\infty \times \mathbf{L}^2_{\mathbf{x,a,loc}}(\mathrm{q})$ of (3.54). Set $\bar{Y} = Y^2 - Y^1$, $\bar{Z} = Z^2 - Z^1$. Let $0 \leqslant s \leqslant T < \infty$. Then, an application of Itô's formula to $e^{-2\,\delta\, r}|\bar{Y}_r|^2$ between s and T yields: $\mathbb{P}^{x,a}$-a.s.,

$$
\begin{aligned}
e^{-2\,\delta\, s}|\bar{Y}_s|^2 = e^{-2\,\delta\, T}|\bar{Y}_T|^2 \\
& - 2n \int_s^T \int_A e^{-2\,\delta\, r}\, \bar{Y}_r\, \{[Z_r^2(X_s,b)]^- - [Z_r^1(X_s,b)]^-\}\, \lambda_0(db)\, dr \\
& - 2 \int_s^T \int_A e^{-2\,\delta\, r}\, \bar{Y}_r\, \bar{Z}_r(X_s,b)\, \lambda_0(db)\, dr \\
& - 2 \int_s^T \int_{E\times A} e^{-2\,\delta\, r}\, \bar{Y}_r\, \bar{Z}_r(y,b)\, q(dr\, dy\, db) \\
& - \int_s^T \int_{E\times A} e^{-2\,\delta\, r}\, |\bar{Z}_r(y,b)|^2\, p(dr\, dy\, db).
\end{aligned}
\tag{3.60}
$$

Notice that

$$
\begin{aligned}
& -n \int_s^T \int_A e^{-\delta\,(r-s)}\, \bar{Y}_r\, \{[Z_r^2(X_r,b)]^- - [Z_r^1(X_r,b)]^-\}\, \lambda_0(db)\, dr \\
& = \int_s^T \int_A e^{-\delta\,(r-s)}\, \bar{Y}_r\, \{Z_r^2(X_r,b) - Z_r^1(X_r,b)\}\, \nu_r^\varepsilon\, \lambda_0(db)\, dr \\
& \quad -\varepsilon \int_s^T \int_A e^{-\delta\,(r-s)}\, \bar{Y}_r\, \{Z_r^2(X_r,b) - Z_r^1(X_r,b)\}\, 1_{\{|\bar{Y}_r| \leqslant 1\}} \cdot \\
& \qquad\qquad \cdot 1_{\{[Z_r^2(X_r,b)]^- = [Z_r^1(X_r,b)]^-,\, |\bar{Z}_r(X_r,b)| \leqslant 1\}}\, \lambda_0(db)\, dr \\
& \quad -\varepsilon \int_s^T \int_A e^{-\delta\,(r-s)}\, 1_{\{|\bar{Y}_r| > 1\}}\, 1_{\{[Z_r^2(X_r,b)]^- = [Z_r^1(X_r,b)]^-,\, |\bar{Z}_r(X_r,b)| > 1\}}\, \lambda_0(db)\, dr,
\end{aligned}
$$

where $\nu^\varepsilon : \mathbb{R}_+ \times \Omega \times A$ is given by

$$
\begin{aligned}
\nu_r^\varepsilon(b) \;=\; & -n\, \frac{[Z_r^2(X_r,b)]^- - [Z_r^1(X_r,b)]^-}{\bar{Z}_r(X_r,b)}\, 1_{\{[Z_r^2(X_r,b)]^- - [Z_r^1(X_r,b)]^- \neq 0\}} \\
& + \varepsilon\, 1_{\{|\bar{Y}_r| \leqslant 1\}}\, 1_{\{[Z_r^2(X_r,b)]^- = [Z_r^1(X_r,b)]^-,\, |\bar{Z}_r(X_r,b)| \leqslant 1\}} \\
& + \varepsilon\, (\bar{Y}_r)^{-1}\, (\bar{Z}_r(X_r^{x,a},b))^{-1}\, 1_{\{|\bar{Y}_r| > 1\}}\, 1_{\{[Z_r^2(X_r^{x,a},b)]^- = [Z_r^1(X_s^{x,a},b)]^-,\, |\bar{Z}_r(X_r^{x,a},b)| > 1\}},
\end{aligned}
\tag{3.61}
$$

for arbitrary $\varepsilon \in (0, 1)$. In particular, ν^ε is a $\mathcal{P} \otimes \mathcal{A}$-measurable map satisfying $\nu_r^\varepsilon(b) \in [\varepsilon, n]$, $dr \otimes d\mathbb{P}^{x,a} \otimes \lambda_0(db)$-almost everywhere. Consider the probability measure $\mathbb{P}^{x,a}_{\nu^\varepsilon}$ on $(\Omega, \mathcal{F}_\infty)$, whose restriction to $(\Omega, \mathcal{F}_T)$ has Radon-Nikodym density:

$$
L_s^{\nu^\varepsilon} := \mathcal{E}\left(\int_0^\cdot \int_{E\times A} (\nu_t^\varepsilon(b)\, d_1(t,y,b) + d_2(t,y,b) - 1)\, q(dt\, dy\, db) \right)_s
\tag{3.62}
$$

for all $0 \leqslant s \leqslant T$, where $\mathcal{E}(\cdot)_s$ is the Doléans-Dade exponential. The existence of such a probability is guaranteed by Proposition 3.3.2. From Lemma 3.3.1 it follows that $(L_s^{\nu^\varepsilon})_{s\in[0,T]}$ is a uniformly integrable martingale. Moreover, $L_T^{\nu^\varepsilon} \in \mathbf{L^p}(\mathcal{F}_T)$, for any $p \geqslant 1$. Under the probability measure $\mathbb{P}^{x,a}_{\nu^\varepsilon}$, by Girsanov's theorem, the compensator of p on $[0, T] \times E \times A$ is $(\nu_s^\varepsilon(b)\, d_1(s,y,b) + d_2(s,y,b))\, \tilde{p}(ds\, dy\, db)$. We denote by $q^{\nu^\varepsilon}(ds\, dy\, db) := p(ds\, dy\, db) - (\nu_s^\varepsilon(b)\, d_1(s,y,b) + d_2(s,y,b))\, \tilde{p}(ds\, dy\, db)$ the compensated martingale measure of p under $\mathbb{P}^{x,a}_{\nu^\varepsilon}$. Therefore equation (3.60) becomes: $\mathbb{P}^{x,a}$-a.s.,

$$e^{-2\delta s}|\bar{Y}_s|^2 \leqslant e^{-2\delta T}|\bar{Y}_T|^2 - 2\int_s^T\int_A e^{-2\delta r}\, \bar{Y}_r\, \bar{Z}_r(X_s,b)\, q^{\nu^\varepsilon}(ds\, dy\, db) + 2\,\frac{\varepsilon}{\delta}\,\lambda_0(A),$$

for all $\varepsilon \in (0, 1)$. Moreover, from the arbitrariness of ε, we obtain

$$e^{-2\delta s}|\bar{Y}_s|^2 \leqslant e^{-2\delta T}|\bar{Y}_T|^2 - 2\int_s^T\int_A e^{-2\delta r}\, \bar{Y}_r\, \bar{Z}_r(X_s,b)\, q^{\nu^\varepsilon}(ds\, dy\, db). \tag{3.63}$$

From Lemma 3.3.1, we see that the stochastic integral in (3.63) is a martingale, so that, taking the expectation $\mathbb{E}^{x,a}_{\nu^\varepsilon}$, conditional on $\mathcal{F}_s$, with respect to $\mathbb{P}^{x,a}_{\nu^\varepsilon}$, we achieve

$$e^{-2\delta s}|\bar{Y}_s|^2 \quad \leqslant \quad e^{-2\delta T}\, \mathbb{E}^{x,a}_{\nu^\varepsilon}[|\bar{Y}_T|^2|\mathcal{F}_s]. \tag{3.64}$$

In particular, $(e^{-2\delta s}|\bar{Y}_s|^2)_{t\geqslant 0}$ is a submartingale. Since $\bar{Y}$ is uniformly bounded, we see that $(e^{-2\delta s}|\bar{Y}_s|^2)_{t\geqslant 0}$ is a uniformly integrable submartingale, therefore $e^{-2\delta s}|\bar{Y}_s|^2 \to \xi_\infty \in \mathbf{L^1}(\Omega, \mathcal{F}, \mathbb{P}^{x,a}_{\nu^\varepsilon})$, as $s \to \infty$. Using again the boundedness of $\bar{Y}$, we obtain that $\xi_\infty = 0$, which implies $\bar{Y} = 0$. Finally, plugging $\bar{Y} = 0$ into (3.60) we conclude that $\bar{Z} = 0$.

Existence. Fix $(x, a, n) \in E \times A \times \mathbb{N}$. For $T > 0$, let $(Y^{T,n,x,a}, Z^{T,n,x,a}) = (Y^T, Z^T)$ denote the unique solution to the penalized BSDE (3.55) on $[0, T]$.
Step 1. Convergence of $(Y^T)_T$. Let $T, T' > 0$, with $T < T'$, and $s \in [0, T]$. We have

$$|Y_s^{T'} - Y_s^T|^2 \leqslant e^{-2\delta(T-s)}\, \mathbb{E}^{x,a}_{\nu^\varepsilon}\left[|Y_T^{T'} - Y_T^T|^2|\mathcal{F}_s\right] \overset{T\to\infty}{\longrightarrow} 0, \tag{3.65}$$

where the convergence result follows from (3.57). Let us now consider the sequence of real-valued càdlàg adapted processes $(Y^T)_T$. It follows from (3.65) that, for any $t \geqslant 0$, the sequence $(Y_t^T(\omega))_T$ is Cauchy for almost every ω, so that it converges $\mathbb{P}^{x,a}$-a.s. to some $\mathcal{F}_t$-measurable random variable Y_t, which is bounded from the right-hand side of (3.57). Moreover, using again (3.65) and (3.57), we see that, for any $0 \leqslant S < T \wedge T'$, with $T, T' > 0$, we have

$$\sup_{0\leqslant t\leqslant S}|Y_t^{T'} - Y_t^T| \leqslant e^{-\delta(T\wedge T'-S)}\,\frac{M_f}{\delta} \overset{T,T'\to\infty}{\longrightarrow} 0. \tag{3.66}$$

In other words, the sequence $(Y^T)_{T>0}$ converges $\mathbb{P}^{x,a}$-a.s. to Y uniformly on compact subsets of $\mathbb{R}_+$. Since each Y^T is a càdlàg process, it follows that Y is càdlàg, as well. Finally, from estimate (3.57) we see that Y is uniformly bounded and therefore belongs to $\mathbf{S}^\infty$.

Step 2. Convergence of $(Z^T)_T$. Let $S, T, T' > 0$, with $S < T < T'$. Then, applying Itó's formula to $e^{-2\delta s}|Y_t^{T'} - Y_t^T|^2$ between 0 and S, and taking the expectation, we find

$$
\begin{aligned}
&\mathbb{E}^{x,a}\left[\int_0^S \int_{E\times A} e^{-2\,\delta\,r}\,|Z_r^{T'}(y,b) - Z_r^T(y,b)|^2\,\tilde{p}(dr\,dy\,db)\right]\\
&= e^{-2\,\delta\,S}\mathbb{E}^{x,a}\left[|Y_S^{T'} - Y_S^T|^2\right] - |Y_0^{T'} - Y_0^T|^2\\
&-2n\,\mathbb{E}^{x,a}\left[\int_0^S \int_A e^{-2\,\delta\,r}\,(Y_r^{T'} - Y_r^T)\,\{[Z_r^2(X_r,b)]^- - [Z_r^1(X_r,b)]^-\}\,\lambda_0(db)\,dr\right]\\
&-2\mathbb{E}^{x,a}\left[\int_0^S \int_A e^{-2\,\delta\,r}\,(Y_r^{T'} - Y_r^T)\,(Z_r^{T'}(X_r,b) - Z_r^T(X_r,b))\,\lambda_0(db)\,dr\right].
\end{aligned}
$$

Recalling the elementary inequality $bc \leqslant b^2 + c^2/4$, for any $b,\ c \in \mathbb{R}$, we get

$$
\begin{aligned}
&\mathbb{E}^{x,a}\left[\int_0^S \int_{E\times A} e^{-2\,\delta\,r}\,|Z_r^{T'}(y,b) - Z_r^T(y,b)|^2\,\tilde{p}(dr\,dy\,db)\right]\\
&\leqslant e^{-2\,\delta\,S}\mathbb{E}^{x,a}\left[|Y_S^{T'} - Y_S^T|^2\right] + 4(n^2+1)\,\lambda_0(A)\,\mathbb{E}^{x,a}\left[\int_0^S e^{-2\,\delta\,r}\,|Y_r^{T'} - Y_r^T|^2\,dr\right]\\
&+\frac{1}{4}\mathbb{E}^{x,a}\left[\int_0^S \int_A e^{-2\,\delta\,r}\,|[Z_r^2(X_r,b)]^- - [Z_r^1(X_r,b)]^-|^2\,\lambda_0(db)\,dr\right]\\
&+\frac{1}{4}\mathbb{E}^{x,a}\left[\int_0^S \int_A e^{-2\,\delta\,r}\,|Z_r^{T'}(X_r,b) - Z_r^T(X_r,b)|^2\,\lambda_0(db)\,dr\right].
\end{aligned}
$$

Multiplying the previous inequality by $e^{2\,\delta\,s}$, and recalling the form of the compensator $\tilde{p}$, we get

$$
\begin{aligned}
&\frac{1}{2}\mathbb{E}^{x,a}\left[\int_0^S \int_{E\times A} e^{-2\,\delta\,r}\,|Z_r^{T'}(y,b) - Z_r^T(y,b)|^2\,\tilde{p}(dr\,dy\,db)\right]\\
&\leqslant e^{-2\,\delta\,S}\mathbb{E}^{x,a}\left[|Y_S^{T'} - Y_S^T|^2\right] + 4(n^2+1)\,\lambda_0(A)\,\mathbb{E}^{x,a}\left[\int_0^S e^{-2\,\delta\,r}\,|Y_r^{T'} - Y_r^T|^2\,dr\right]\\
&\overset{T,T'\to\infty}{\longrightarrow} 0,
\end{aligned}
$$

where the convergence to zero follows from estimate (3.66). Then, for any $S > 0$, we see that $(Z^T_{|[0,\,S]})_{T>S}$ is a Cauchy sequence in the Hilbert space $\mathbf{L^2_{x,a}}(\mathrm{q};\mathbf{0},\,\mathbf{S})$. Therefore, we deduce that there exists $\tilde{Z}^S \in \mathbf{L^2_{x,a}}(\mathrm{q};\mathbf{0},\,\mathbf{S})$ such that $(Z^T_{|[0,\,S]})_{T>S}$ converges to $\tilde{Z}^S$ in $\mathbf{L^2_{x,a}}(\mathrm{q};\mathbf{0},\,\mathbf{S})$, i.e.,

$$
\mathbb{E}^{x,a}\left[\int_0^S \int_{E\times A} e^{-2\,\delta\,r}\,|Z_r^T(y,b) - \tilde{Z}_r^S(y,b)|^2\,\tilde{p}(dr\,dy\,db)\right] \overset{T\to\infty}{\longrightarrow} 0.
$$

Notice that $\tilde{Z}^{S'}_{|[0,\,S]} = \tilde{Z}^S$, for any $0 \leqslant S \leqslant S' < \infty$. Indeed, $\tilde{Z}^{S'}_{|[0,\,S]}$, as $\tilde{Z}^S$, is the limit in $\mathbf{L^2_{x,a}}(\mathrm{q};\mathbf{0},\,\mathbf{S})$ of $(Z^T_{|[0,\,S]})_{T>S}$. Hence, we define $Z_s = \tilde{Z}^S_s$ for all $s \in [0,\,S]$ and for any $S > 0$. Observe that $Z \in \mathbf{L^2_{x,a,loc}}(\mathrm{q})$. Moreover, for any $S > 0$, $(Z^T_{|[0,\,S]})_{T>S}$ converges to $Z_{|[0,\,S]}$ in $\mathbf{L^2_{x,a}}(\mathrm{q};\mathbf{0},\,\mathbf{S})$, i.e.,

$$
\mathbb{E}^{x,a}\left[\int_0^S \int_{E\times A} e^{-2\,\delta\,r}\,|Z_r^T(y,b) - Z_r(y,b)|^2\,\tilde{p}(dr\,dy\,db)\right] \overset{T\to\infty}{\longrightarrow} 0. \tag{3.67}
$$

Now, fix $S \in [0,\, T]$ and consider the BSDE satisfied by (Y^T, Z^T) on $[0,\, S]$: $\mathbb{P}^{x,a}$-a.s.,

$$\begin{aligned} Y_t^T &= Y_S^T - \delta \int_t^S Y_r^T \, dr + \int_t^S f(X_r, I_r)\, dr \\ &\quad -n \int_t^S \int_A [Z_r^T(X_r, b)]^- \,\lambda_0(db)\, dr - \int_t^S \int_A Z_r^T(X_r, b)\, \lambda_0(db)\, dr, \\ &\quad - \int_t^S \int_{E\times A} Z_r^T(y, b)\, q(dr\, dy\, db), \qquad 0 \leqslant t \leqslant S. \end{aligned}$$

From (3.67) and (3.66), we can pass to the limit in the above BSDE by letting $T \to \infty$ keeping S fixed. Then we deduce that (Y, Z) solves the penalized BSDE (3.54) on $[0,\, S]$. Since S is arbitrary, it follows that (Y, Z) solves equation (3.54) on $[0,\, \infty)$. □

The penalized BSDE (3.54) can be represented by means of a suitable family of dual control problems.

Lemma 3.4.5. *Let Hypotheses* **(HhλQ)**, **(Hλ_0)** *and* **(Hf)** *hold. Then, for every* $(x, a, n) \in E \times A \times \mathbb{N}$, $\mathbb{P}^{x,a}$*-a.s., the solution* $(Y^{n,x,a}, Z^{n,x,a})$ *to* (3.54) *admits the following explicit representation:*

$$Y_s^{n,x,a} = \operatorname*{ess\,inf}_{\nu \in \mathcal{V}^n} \mathbb{E}_\nu^{x,a} \left[\int_s^\infty e^{-\delta\,(r-s)} f(X_r, I_r)\, dr \Big| \mathcal{F}_s \right], \quad s \geqslant 0. \tag{3.68}$$

Proof. Fix $n \in \mathbb{N}$, and for any $\nu \in \mathcal{V}^n$, let us introduce the compensated martingale measure $q^\nu(ds\, dy\, db) = q(ds\, dy\, db) - (\nu_s(b) - 1)\, d_1(s, y, b)\, \tilde p(ds\, dy\, db)$ under $\mathbb{P}_\nu^{x,a}$. Fix $T \geqslant s$ and apply Itô's formula to $e^{-\delta\, r}\, Y_r^{n,x,a}$ between s and T. Then we obtain:

$$\begin{aligned} Y_s^{n,x,a} &= e^{-\delta\,(T-s)}\, Y_T^{n,x,a} + \int_s^T e^{-\delta\,(r-s)} f(X_r, I_r)\, dr \\ &\quad - \int_s^T \int_A e^{-\delta\,(r-s)} \{ n [Z_r^{n,x,a}(X_r, b)]^- + \nu_r(a)\, Z_r^{n,x,a}(X_r, b) \}\, \lambda_0(db)\, dr \\ &\quad - \int_s^T \int_{E\times A} e^{-\delta\,(r-s)}\, Z_r^{n,x,a}(y, b)\, q^\nu(dr\, dy\, db), \quad s \in [t,\, T]. \end{aligned} \tag{3.69}$$

Taking the expectation in (3.69) under $\mathbb{P}_\nu^{x,a}$, conditional to $\mathcal{F}_s$, and since by Proposition 3.4.4 $Z^{n,x,a}$ is in $\mathbf{L^2_{loc,x,a}}(\mathrm{q})$, we get from Lemma 3.3.1 that, $\mathbb{P}^{x,a}$-a.s.,

$$\begin{aligned} Y_s^{n,x,a} &= \mathbb{E}_\nu^{x,a} \left[e^{-\delta\,(T-s)}\, Y_T^{n,x,a} + \int_s^T e^{-\delta\,(r-s)} f(X_r, I_r)\, dr \Big| \mathcal{F}_s \right] \\ &\quad - \mathbb{E}_\nu^{x,a} \left[\int_s^T \int_A e^{-\delta\,(r-s)} \{ n [Z_r^{n,x,a}(X_r, b)]^- + \nu_r(a)\, Z_r^{n,x,a}(X_r, b) \}\, \lambda_0(db)\, dr \Big| \mathcal{F}_s \right]. \end{aligned} \tag{3.70}$$

From the elementary numerical inequality: $n[z]^- + \nu z \geqslant 0$ for all $z \in \mathbb{R}$, $\nu \in (0,\, n]$, we deduce by (3.70) that, for all $\nu \in \mathcal{V}^n$,

$$Y_s^{n,x,a} \leqslant \mathbb{E}_\nu^{x,a} \left[e^{-\delta\,(T-s)}\, Y_T^{n,x,a} + \int_s^T e^{-\delta\,(r-s)} f(X_r, I_r)\, dr \Big| \mathcal{F}_s \right]$$

$$\leqslant \mathbb{E}_{\nu}^{x,a}\left[e^{-\delta(T-s)}\,Y_T^{n,x,a} + \int_s^\infty e^{-\delta(r-s)}\,f(X_r, I_r)\,dr\Big|\mathcal{F}_s\right].$$

Since $Y^{n,x,a}$ is in $\mathbf{S}^\infty$ by Proposition 3.4.4, sending $T \to \infty$, we obtain from the conditional version of Lebesgue dominated convergence theorem that

$$Y_s^{n,x,a} \quad \leqslant \quad \mathbb{E}_{\nu}^{x,a}\left[\int_s^\infty e^{-\delta(r-s)}\,f(X_r, I_r)\,dr\Big|\mathcal{F}_s\right],$$

for all $\nu \in \mathcal{V}^n$. Therefore,

$$Y_s^{n,x,a} \leqslant \operatorname*{ess\,inf}_{\nu\in\mathcal{V}^n} \mathbb{E}_{\nu}^{x,a}\left[\int_s^\infty e^{-\delta(r-s)}\,f(X_r, I_r)\,dr\Big|\mathcal{F}_s\right]. \tag{3.71}$$

On the other hand, for $\varepsilon \in (0, 1)$, let us consider the process $\nu^\varepsilon \in \mathcal{V}^n$ defined by:

$$\nu_s^\varepsilon(b) = n\,1_{\{Z_s^{n,x,an}(X_{s-},b)\leqslant 0\}} + \varepsilon\,1_{\{0<Z_s^{n,x,a}(X_{s-},b)<1\}} + \varepsilon\,Z_s^{n,x,a}(X_{s-},b)^{-1}\,1_{\{Z_s^{n,x,a}(X_{s-},b)\geqslant 1\}}$$

(notice that we can not take $\nu_s(b) = n1_{\{Z_s^n(X_{s-},b)\leqslant 0\}}$, since this process does not belong to $\mathcal{V}^n$ because of the requirement of strict positivity). By construction, we have

$$n[Z_s^n(X_{s-},b)]^- + \nu_s^\varepsilon(b)\,Z_s^n(X_{s-},b) \leqslant \varepsilon, \quad s \geqslant 0,\, b \in A,$$

and thus for this choice of $\nu = \nu^\varepsilon$ in (3.70):

$$\begin{aligned} Y_s^{n,x,a} \geqslant{} & \mathbb{E}_{\nu^\varepsilon}^{x,a}\left[e^{-\delta(T-s)}\,Y_T^{n,x,a} + \int_s^T e^{-\delta(r-s)}\,f(X_r, I_r)\,dr\Big|\mathcal{F}_s\right] \\ & - \varepsilon\,\frac{1 - e^{-\delta(T-s)}}{\delta}\,\lambda_0(A). \end{aligned}$$

Letting $T \to \infty$, since f is bounded by M_f and $Y^{n,x,a}$ is in $\mathbf{S}^\infty$, it follows from the conditional version of Lebesgue dominated convergence theorem that

$$\begin{aligned} Y_s^{n,x,a} \quad & \geqslant \quad \mathbb{E}_{\nu^\varepsilon}^{x,a}\left[\int_s^\infty e^{-\delta(r-s)}\,f(X_r, I_r)\,dr\Big|\mathcal{F}_s\right] - \frac{\varepsilon}{\delta}\,\lambda_0(A), \\ & \geqslant \quad \operatorname*{ess\,inf}_{\nu\in\mathcal{V}^n} \mathbb{E}_{\nu}^{x,a}\left[\int_s^\infty e^{-\delta(r-s)}\,f(X_r, I_r)\,dr\Big|\mathcal{F}_s\right] - \frac{\varepsilon}{\delta}\,\lambda_0(A). \end{aligned}$$

From the arbitrariness of ε, together with (3.71), this is enough to prove the required representation of $Y^{n,x,a}$. □

Let us define

$$K_t^{n,x,a} := n\int_0^t \int_A [Z_s^{n,x,a}(X_s, b)]^-\,\lambda_0(db)\,ds, \quad t \geqslant 0.$$

The following a priori uniform estimate on the sequence $(Z^{n,x,a}, K^{n,x,a})_{n\geqslant 0}$ holds.

Lemma 3.4.6. *Assume that hypotheses* **(HhλQ)**, **(Hλ_0)** *and* **(Hf)** *hold. For every $(x, a, n) \in E \times A \times \mathbb{N}$, and for every $T > 0$, there exists a constant C depending only on M_f, δ and T such that*

$$||Z^{n,x,a}||^2_{\mathbf{L}^2_{x,a}(\mathrm{q};\mathbf{0},\mathbf{T})} + ||K^{n,x,a}||^2_{\mathbf{K}^2_{x,a}(\mathbf{0},\mathbf{T})} \leqslant C. \tag{3.72}$$

Proof. In what follows we shall denote by $C > 0$ a generic positive constant depending on M_f, δ and T, which may vary from line to line. Fix $T > 0$ and apply Itô's formula to $|Y_r^{n,x,a}|^2$ between 0 and T. Noticing that $K^{n,x,a}$ is continuous and $\Delta Y_r^{n,x,a} = \int_{E\times A} Z_r^{n,x,a}(y,b)\, p(\{r\}\, dy\, db)$, we get: $\mathbb{P}^{x,a}$-a.s.,

$$\begin{aligned}
\mathbb{E}^{x,a}\left[|Y_0^{n,x,a}|^2\right] &= \mathbb{E}^{x,a}\left[|Y_T^{n,x,a}|^2\right] - 2\mathbb{E}^{x,a}\left[\int_0^T |Y_r^{n,x,a}|^2\, dr\right] \\
&\quad - 2\mathbb{E}^{x,a}\left[\int_s^T Y_r^{n,x,a}\, dK_r^{n,x,a}\right] + 2\mathbb{E}^{x,a}\left[\int_0^T Y_r^{n,x,a}\, f(X_r, I_r)\, dr\right] \\
&\quad - 2\mathbb{E}^{x,a}\left[\int_0^T \int_A Y_r^{n,x,a}\, Z_r^{n,x,a}(X_r, b)\, \lambda_0(db)\, dr\right] \\
&\quad - \mathbb{E}^{x,a}\left[\int_0^T \int_{E\times A} |Z_r^{n,x,a}(y,b)|^2\, \tilde{p}(dr\, dy\, db)\right].
\end{aligned}$$

Set now $C_Y := \frac{M_f}{\delta}$. Recalling the uniform estimate (3.57) on Y^n, and using elementary inequalities, we get

$$\begin{aligned}
&\mathbb{E}^{x,a}\left[\int_0^T \int_{E\times A} |Z_s^{n,x,a}(y,\, b)|^2\, \tilde{p}(ds\, dy\, db)\right] \\
&\leqslant C_Y^2 + 2\,T\,C_Y^2 + 2\,T\,C_Y\,M_f + 2\,C_Y\,T\,\mathbb{E}^{x,a}\left[|K_T^{n,x,a}|\right] \\
&\quad + \frac{C_Y}{\alpha}\,T\,\lambda_0(A) + \alpha\, C_Y\, \mathbb{E}^{x,a}\left[\int_0^T \int_A |Z_r^{n,x,a}(X_s, b)|^2\, \lambda_0(db)\, dr\right], \qquad (3.73)
\end{aligned}$$

for any $\alpha > 0$. At this point, from relation (3.54), we obtain:

$$\begin{aligned}
K_T^{n,x,a} &= Y_0^{n,x,a} - Y_T^{n,x,a} - \delta \int_0^T \int_A Y_s^{n,x,a}\, ds \\
&\quad + \int_0^T f(X_s, I_s) ds + \int_0^T \int_A Z_s^{n,x,a}(X_s, b)\, \lambda_0(db)\, ds \\
&\quad + \int_0^T \int_{E\times A} Z_s^{n,x,a}(y,b)\, q(ds\, dy\, db). \qquad (3.74)
\end{aligned}$$

Then, using the inequality $2bc \leqslant \frac{1}{\beta} b^2 + \beta c^2$, for any $\beta > 0$, and taking the expected value we have

$$\begin{aligned}
2\,\mathbb{E}^{x,a}\left[|K_T^{n,x,a}|\right] &\leqslant 2\,\delta\, C_Y\, T + 2\, M_f\, T + \frac{T}{\beta}\,\lambda_0(A) \\
&\quad + \beta\,\mathbb{E}^{x,a}\left[\int_0^T \int_A |Z_s^{n,x,a}(X_s, b)|^2\, \lambda_0(db)\, ds\right]. \qquad (3.75)
\end{aligned}$$

Plugging (3.75) into (3.73), we get

$$\begin{aligned}
&\mathbb{E}^{x,a}\left[\int_0^T \int_{E\times A} |Z_s^{n,x,a}(y,\, b)|^2\, \tilde{p}(ds\, dy\, db)\right] \\
&\leqslant C + C_Y\,(2\,T\,\beta + \alpha) \int_0^T \int_A |Z_s^{n,x,a}(X_s, b)|^2\, \lambda_0(db)\, ds.
\end{aligned}$$

Hence, choosing $\alpha + 2\,T\,\beta = \frac{1}{2\,C_Y}$, we get

$$\frac{1}{2}\,\mathbb{E}^{x,a}\left[\int_0^T \int_{E\times A} |Z_s^{n,x,a}(y,\,b)|^2\,\tilde p(ds\,dy\,db)\right] \;\leqslant\; C,$$

which gives the required uniform estimate for $(Z^{n,x,a})$, and also $(K^{n,x,a})$ by (3.74). □

3.4.2. BSDE representation of the dual value function. In order to prove the main result of this section we give the following preliminary result.

Lemma 3.4.7. *Assume that Hypotheses* **(HhλQ)**, **(Hλ_0)** *and* **(Hf)** *hold. For every $(x,a)\in E\times A$, let $(Y^{x,a},Z^{x,a},K^{x,a})\in \mathbf{S}^\infty\times \mathbf{L}^2_{x,a,loc}(\mathrm{q})\times \mathbf{K}^2_{x,a,loc}$ be a solution to the BSDE with partially nonnegative jumps* (3.51)-(3.52). *Then,*

$$Y_s^{x,a} \leqslant \operatorname*{ess\,inf}_{\nu\in\mathcal{V}} \mathbb{E}_\nu^{x,a}\left[\int_s^\infty e^{-\delta(r-s)}\,f(X_r,I_r)\,dr\Big|\mathcal{F}_s\right],\ \forall\ s\geqslant 0. \tag{3.76}$$

Proof. Let $(x,a)\in E\times A$, and consider a triplet $(Y^{x,a},Z^{x,a},K^{x,a})\in \mathbf{S}^\infty\times\mathbf{L}^2_{x,a,loc}(\mathrm{q})\times \mathbf{K}^2_{x,a,loc}$ satisfying (3.51)-(3.52). Applying Itô's formula to $e^{-\delta\,r}\,Y_r^{x,a}$ between s and $T>s$, and recalling that $K^{x,a}$ is nondecreasing, we have

$$\begin{aligned} Y_s^{x,a} &\leqslant e^{-\delta\,(T-s)}\,Y_T^{x,a} + \int_s^T e^{-\delta\,(r-s)}\,f(X_r,I_r)\,dr\\ &\quad - \int_s^T\int_A e^{-\delta\,(r-s)}\,Z_r^{x,a}(X_r,b)\,\lambda_0(db)\,dr\\ &\quad - \int_s^T\int_{E\times A} e^{-\delta\,(r-s)}\,Z_r^{x,a}(y,b)\,\tilde q(dr\,dy\,db),\quad 0\leqslant s\leqslant T<\infty. \end{aligned} \tag{3.77}$$

Then for any $\nu\in\mathcal{V}$, let us introduce the compensated martingale measure $q^\nu(ds\,dy\,db) = q(ds\,dy\,db) - (\nu_s(b)-1)\,d_1(s,y,b)\,\tilde p(ds\,dy\,db)$ under $\mathbb{P}_\nu^{x,a}$. Taking expectation in (3.77) under $\mathbb{P}_\nu^{x,a}$, conditional to $\mathcal{F}_s$, and recalling that $Z^{x,a}$ is in $\mathbf{L}^2_{x,a,loc}(\mathrm{q})$, we get from Lemma 3.3.1 that, $\mathbb{P}^{x,a}$-a.s.,

$$\begin{aligned} Y_s^{x,a} &\leqslant \mathbb{E}_\nu^{x,a}\left[e^{-\delta\,(T-s)}\,Y_T^{x,a} + \int_s^T e^{-\delta\,(r-s)}\,f(X_r,I_r)\,dr\Big|\mathcal{F}_s\right]\\ &\quad -\mathbb{E}_\nu^{x,a}\left[\int_s^T\int_A e^{-\delta\,(r-s)}\,\nu_r(a)\,\bar Z_r^{x,a}(X_r,b)\,\lambda_0(db)\,dr\Big|\mathcal{F}_s\right]. \end{aligned} \tag{3.78}$$

Furthermore, since ν is strictly positive and $Z^{x,a}$ satisfies the nonnegative constraint (3.52), from inequality (3.78) we get

$$\begin{aligned} Y_s^{x,a} &\leqslant \mathbb{E}_\nu^{x,a}\left[e^{-\delta\,(T-s)}\,Y_T^{x,a} + \int_s^T e^{-\delta\,(r-s)}\,f(X_r,I_r)\,dr\Big|\mathcal{F}_s\right]\\ &\leqslant \mathbb{E}_\nu^{x,a}\left[e^{-\delta\,(T-s)}\,Y_T^{x,a} + \int_s^\infty e^{-\delta\,(r-s)}\,f(X_r,I_r)\,dr\Big|\mathcal{F}_s\right]. \end{aligned}$$

Finally, sending $T \to \infty$ and recalling that $Y^{x,a}$ is in $\mathbf{S}^\infty$, the conditional version of Lebesgue dominated convergence theorem yields

$$Y_s^{x,a} \quad \leqslant \quad \mathbb{E}_\nu^{x,a}\left[\int_s^\infty e^{-\delta\,(r-s)}\, f(X_r, I_r)\, dr \Big| \mathcal{F}_s\right]$$

for all $\nu \in \mathcal{V}$, and the conclusion follows from the arbitrariness of $\nu \in \mathcal{V}$, . $\square$

Now we are ready to state the main result of the section.

Theorem 3.4.8. *Under Hypotheses* **(HhλQ)**, **(Hλ_0)** *and* **(Hf)**, *for every $(x,a) \in E \times A$, there exists a unique maximal solution $(Y^{x,a}, Z^{x,a}, K^{x,a}) \in \mathbf{S}^\infty \times \mathbf{L}^{\mathbf{2}}_{\boldsymbol{x,a,\mathrm{loc}}}(\mathrm{q}) \times \mathbf{K}^{\mathbf{2}}_{\boldsymbol{x,a,\mathrm{loc}}}$ to the BSDE with partially nonnegative jumps* (3.51)-(3.52). *In particular,*

(i) *$Y^{x,a}$ is the nondecreasing limit of $(Y^{n,x,a})_n$;*

(ii) *$Z^{x,a}$ is the weak limit of $(Z^{n,x,a})_n$ in $\mathbf{L}^{\mathbf{2}}_{\boldsymbol{x,a,\mathrm{loc}}}(\mathrm{q})$;*

(iii) *$K_s^{x,a}$ is the weak limit of $(K_s^{n,x,a})_n$ in $\mathbf{L}^{\mathbf{2}}(\mathcal{F}_s)$, for any $s \geqslant 0$;*

Moreover, $Y^{x,a}$ has the explicit representation:

$$Y_s^{x,a} = \operatorname*{ess\,inf}_{\nu\in\mathcal{V}} \mathbb{E}_\nu^{x,a}\left[\int_s^\infty e^{-\delta(r-s)}\, f(X_r, I_r)\, dr \Big| \mathcal{F}_s\right],\ \forall\ s \geqslant 0. \tag{3.79}$$

In particular, setting $s = 0$, we have the following representation formula for the value function of the dual control problem:

$$V^*(x,a) = Y_0^{x,a}, \quad (x,a) \in E \times A. \tag{3.80}$$

Proof. Let $(x,a) \in E \times A$ be fixed. From the representation formula (3.68) it follows that $Y_s^n \geqslant Y_s^{n+1}$ for all $s \geqslant 0$ and all $n \in \mathbb{N}$, since by definition $\mathcal{V}^n \subset \mathcal{V}^{n+1}$ and $(Y^n)_n$ are càdlàg processes. Moreover, recalling the boundedness of f , from (3.68) we see that $(Y^n)_n$ is lower-bounded by a constant which does not depend n. Then $(Y^{n,x,a})_n \in \mathbf{S}^\infty$ converges decreasingly to some adapted process $Y^{x,a}$, which is moreover uniformly bounded by Fatou's lemma. Furthermore, for every $T > 0$, the Lebesgue's dominated convergence theorem insures that the convergence of $(Y^{n,x,a})_n$ to Y also holds in $\mathbf{L^2(0,\ T)}$.

Let us fix $T \geqslant 0$. By the uniform estimates in Lemma 3.4.6, the sequence $(Z^{n,x,a}_{|[0,\,T]})_n$ is bounded in the Hilbert space $\mathbf{L}^{\mathbf{2}}_{\mathbf{x,a}}(\mathrm{q}; \mathbf{0},\ \mathbf{T})$. Then, we can extract a subsequence which weakly converges to some Z^T in $\mathbf{L}^{\mathbf{2}}_{\mathbf{x,a}}(\mathrm{q}; \mathbf{0},\ \mathbf{T})$. Let then define the following mappings

$$\begin{aligned} I_\tau^1 := \qquad & Z \longmapsto \int_0^\tau \int_{E\times A} Z_s(y,b)\, q(ds\, dy\, db) \\ & \mathbf{L}^{\mathbf{2}}_{\mathbf{x,a}}(\mathrm{q}; \mathbf{0},\ \mathbf{T}) \longrightarrow \mathbf{L}^{\mathbf{2}}(\mathcal{F}_\tau), \end{aligned}$$

$$\begin{aligned} I_\tau^2 := \qquad & Z(X_s, \cdot) \longmapsto \int_0^\tau \int_A Z_s(X_s, b)\, \lambda_0(db)\, ds \\ & \mathbf{L}^{\mathbf{2}}_{\mathbf{x,a}}(\lambda_0; \mathbf{0},\ \mathbf{T}) \longrightarrow \mathbf{L}^{\mathbf{2}}(\mathcal{F}_\tau), \end{aligned}$$

for every stopping time $0 \leqslant \tau \leqslant T$. We notice that I_τ^1 (resp. I_τ^2) defines a linear continuous operator (hence weakly continuous) from $\mathbf{L}^{\mathbf{2}}_{\mathbf{x,a}}(\mathrm{q}; \mathbf{0},\ \mathbf{T})$ (resp. $\mathbf{L}^{\mathbf{2}}_{\mathbf{x,a}}(\lambda_0; \mathbf{0},\ \mathbf{T})$)

to $\mathbf{L^2}(\mathcal{F}_\tau)$. Therefore $I^1_\tau Z^{n,x,a}_{|[0,\,T]}$ (resp., $I^2_\tau Z^{n,x,a}_{|[0,\,T]}(X,\cdot)$) weakly converges to $I^1_\tau \tilde Z^T$ (resp., $I^2_\tau \tilde Z^T(X,\cdot)$) in $\mathbf{L^2}(\mathcal{F}_\tau)$. Since

$$\begin{aligned}
K^{n,x,a}_\tau &= Y^{n,x,a}_\tau - Y^{n,x,a}_0 - \delta \int_0^\tau Y^{n,x,a}_r\,dr + \int_0^\tau f(X_r, I_r)\,dr \\
&\quad - \int_0^\tau \int_A Z^{n,x,a}_r(X_r, b)\,\lambda_0(db)\,dr \\
&\quad - \int_0^\tau \int_{E\times A} Z^{n,x,a}_r(y, b)\,q(dr\,dy\,db),\ \forall\,\tau \in [0,\,T],
\end{aligned}$$

we also have the following weak convergence in $\mathbf{L^2}(\mathcal{F}_\tau)$:

$$\begin{aligned}
K^{n,x,a}_\tau \rightharpoonup \tilde K^T_\tau &:= Y^{x,a}_\tau - Y^{x,a}_0 - \delta \int_0^\tau Y^{x,a}_r\,dr + \int_0^\tau f(X_r, I_r)\,dr \\
&\quad - \int_0^\tau \int_A Z^{x,a}_r(X_r, b)\,\lambda_0(db)\,dr \\
&\quad - \int_0^\tau \int_{E\times A} Z^{x,a}_r(y, b)\,q(dr\,dy\,db),\ \forall\,\tau \in [0,\,T].
\end{aligned}$$

Since the process $(K^{n,x,a}_s)_{s\in[0,\,T]}$ is nondecreasing and predictable and $K^{n,x,a}_0 = 0$, the limit process $\tilde K^T_\tau$ on $[0,\,T]$ remains nondecreasing and predictable with $\mathbb{E}^{x,a}\left[|\tilde K^T_T|^2\right] < \infty$ and $\tilde K^T_0 = 0$. Moreover, by Lemma 2.2. in Peng [**104**], $\tilde K^T_\tau$ and $\tilde Y^T_\tau$ are càdlàg, therefore $\tilde K^T_\tau \in \mathbf{K^2_{x,a}}(\mathbf{0},\,\mathbf{T})$ and $\tilde Y^T_\tau \in \mathbf{S}^\infty$.

Then we notice that $\tilde Z^{T'}_{|[0,\,T]} = \tilde Z^T$, $\tilde K^{T'}_{|[0,\,T]} = \tilde K^T$, for any $0 \leqslant T \leqslant T' < \infty$. Indeed, for $i = 1, 2$, $I^i\,\tilde Z^{T'}_{|[0,\,T]}$, as $I^i\,\tilde Z^T$, is the weak limit in $\mathbf{L^2}(\mathcal{F}_s)$ of $(I^i\,Z^{n,x,a}_{|[0,\,T]})_{n\geqslant 0}$, while $\tilde K^{T'}_{|[0,\,T]}$, as $\tilde K^T$, is the weak limit in $\mathbf{L^2}(\mathcal{F}_s)$ of $(K^{n,x,a}_{|[0,\,T]})_{n\geqslant 0}$, for every $s \in [0,\,T]$. Hence, we define $Z^{x,a}_s = \tilde Z^T_s$, $K^{x,a}_s = \tilde K^T_s$ for all $s \in [0,\,T]$ and for any $T > 0$. Observe that $Z^{x,a} \in \mathbf{L^2_{x,a,loc}}(\mathrm{q})$ and $K^{x,a} \in \mathbf{K^2_{x,a,loc}}$. Moreover, for any $T > 0$, for $i = 1, 2$, $(I^i\,Z^{n,x,a}_{|[0,\,T]})_{n\geqslant 0}$ weakly converges to $I^i\,Z^{x,a}_{|[0,\,T]}$ in $\mathbf{L^2}(\mathcal{F}_s)$, and $(K^{n,x,a}_{|[0,\,T]})_{n\geqslant 0}$ weakly converges to $K^{x,a}_{|[0,\,T]}$ in $\mathbf{L^2}(\mathcal{F}_s)$, for $s \in [0,\,T]$. In conclusion, we have: $\mathbb{P}^{x,a}$-a.s.,

$$\begin{aligned}
Y^{x,a}_s &= Y^{x,a}_T - \delta \int_s^T Y^{x,a}_r\,dr + \int_s^T f(X_r, I_r)\,dr - (K^{x,a}_T - K^{x,a,\delta}_s) \\
&\quad - \int_s^T \int_A Z^{x,a}_r(X_r, b)\,\lambda_0(db)\,dr \\
&\quad - \int_s^T \int_{E\times A} Z^{x,a}_r(y, b)\,q(dr\,dy\,db), \quad 0 \leqslant s \leqslant T.
\end{aligned}$$

Since T is arbitrary, it follows that $(Y^{x,a}, Z^{x,a}, K^{x,a})$ solves equation (3.51) on $[0,\,\infty)$.

To show that the jump constraint (3.52) is satisfied, we consider the functional $G : \mathbf{L^2_{x,a}}(\lambda_0; \mathbf{0},\,\mathbf{T}) \to \mathbb{R}$ given by

$$G(V(\cdot)) \;:=\; \mathbb{E}\left[\int_0^T \int_A [V_s(b)]^-\,\lambda_0(db)\,ds\right], \quad \forall\, V \in \mathbf{L^2_{x,a}}(\lambda_0; \mathbf{0},\,\mathbf{T}).$$

Notice that $G(Z^{n,x,a}(X,\cdot)) = \mathbb{E}^{x,a}\left[K_T^{n,x,a}/n\right]$, for any $n \in \mathbb{N}$. From uniform estimate (3.72), we see that $G(Z^{n,x,a}(X,\cdot)) \to 0$ as $n \to \infty$. Since G is convex and strongly continuous in the strong topology of $\mathbf{L^2_{x,a}}(\lambda_0; \mathbf{0},\ \mathbf{T})$, then G is lower semicontinuous in the weak topology of $\mathbf{L^2_{x,a}}(\lambda_0; \mathbf{0},\ \mathbf{T})$, see, e.g., Corollary 3.9 in Brezis [**19**]. Therefore, we find

$$G(Z^{x,a}(X,\cdot)) \leqslant \liminf_{n\to\infty} G(Z^{n,x,a}(X,\cdot)) = 0,$$

which implies the validity of jump constraint (3.52) on $[0,\, T]$, and the conclusion follows from the arbitrary of T.

Hence, $(Y^{x,a}, Z^{x,a}, K^{x,a})$ is a solution to the constrained BSDE (3.51)-(3.52) on $[0,\,\infty)$.

It remains to prove the representation formula (3.79) and the maximality property for $Y^{x,a}$. Firstly, since by definition $\mathcal{V}^n \subset \mathcal{V}$ for all $n \in \mathbb{N}$, it is clear from representation formula (3.68) that

$$\begin{aligned} Y_s^{n,x,a} &= \operatorname*{ess\,inf}_{\nu\in\mathcal{V}^n} \mathbb{E}_\nu^{x,a}\left[\int_s^\infty e^{-\delta\,(r-s)}\, f(X_r, I_r)\, dr \Big| \mathcal{F}_s\right] \\ &\geqslant \operatorname*{ess\,inf}_{\nu\in\mathcal{V}} \mathbb{E}_\nu^{x,a}\left[\int_s^\infty e^{-\delta\,(r-s)}\, f(X_r, I_r)\, dr \Big| \mathcal{F}_s\right], \end{aligned}$$

for all $n \in \mathbb{N}$, for all $s \geqslant 0$. Moreover, being $Y^{x,a}$ the pointwise limit of $Y^{n,x,a}$, we deduce that

$$Y_s^{x,a} \geqslant \operatorname*{ess\,inf}_{\nu\in\mathcal{V}} \mathbb{E}_\nu^{x,a}\left[\int_s^\infty e^{-\delta\,(r-s)}\, f(X_r, I_r)\, dr | \mathcal{F}_s\right], \quad s \geqslant 0. \tag{3.81}$$

On the other hand, $Y^{x,a}$ satisfies the opposite inequality (3.76) from Lemma 3.4.7, and thus we achieve the representation formula (3.79).

Finally, to show that $Y^{x,a}$ is the maximal solution, let consider a triplet $(\bar{Y}^{x,a}, \bar{Z}^{x,a}, \bar{K}^{x,a}) \in \mathbf{S}^\infty \times \mathbf{L^2_{x,a,loc}}(\mathrm{q}) \times \mathbf{K^2_{x,a,loc}}$ solution to (3.51)-(3.52). By Lemma 3.4.7, $(\bar{Y}^{x,a}, \bar{Z}^{x,a}, \bar{K}^{x,a})$ satisfies inequality (3.76). Then, from the representation formula (3.79) it follows that $\bar{Y}_s^{x,a} \leqslant Y_s^{x,a}$, $\forall\, s \geqslant 0$, $\mathbb{P}^{x,a}$-a.s., i.e., the maximality property holds. The uniqueness of the maximal solution directly follows from Proposition 3.4.1. □

3.5. A BSDE representation for the value function

Our main purpose is to show how maximal solutions to BSDEs with nonnegative jumps of the form (3.51)-(3.52) provide actually a Feynman-Kac representation to the value function V associated to our optimal control problem for PDMPs. We know from Theorem 3.4.8 that, under Hypotheses **(HhλQ)**, **(Hλ_0)** and **(Hf)** , there exists a unique maximal solution $(Y^{x,a}, Z^{x,a}, K^{x,a})$ on $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}^{x,a})$ to (3.51)-(3.52). Let us introduce a deterministic function $v : E \times A \to \mathbb{R}$ as

$$v(x,a) := Y_0^{x,a}, \quad (x,a) \in E \times A. \tag{3.82}$$

Our main result is as follows:

Theorem 3.5.1. *Assume that Hypotheses* **(HhλQ)**, **(Hλ_0)**, *and* **(Hf)** *hold. Then the function v in* (3.82) *does not depend on the variable a:*

$$v(x,a) = v(x,a'), \quad \forall a, a' \in A,$$

for all $x \in E$. Let us define by misuse of notation the function v on E by

$$v(x) = v(x,a), \quad \forall x \in E,$$

for any $a \in A$. Then v is a (discontinuous) viscosity solution to (3.26).

To conclude that $v(x)$ actually provides the unique solution to (3.26) (and therefore coincides with the value function V by Theorem 3.2.6), we need to use a comparison theorem for viscosity sub and supersolutions to the fully nonlinear integro-partial differential equations of HJB type. To this end, we introduce the following additional condition on Q and λ.

(HλQ')

(i) $\sup_{(x,a)\in E\times A} \int_E |y-x|\, \lambda(x,a)\, Q(x,a,dy) < \infty$;

(ii) $\exists\ c, C > 0$: for every $\psi \in W^{1,\infty}(E)$, $\psi(0) = 0$, for every $K \subset E$ compact set, and $x_1, x_2 \in E$, $a \in A$,

$$\Big| \int_{K+x_1} \psi(y-x_1)\, \lambda(x_1,a)\, (Q(x_1,a,dy) - \int_{K+x_2} \psi(y-x_2)\, \lambda(x_2,a)\, Q(x_2,a,dy)\Big|$$
$$\leqslant c||\nabla\psi||_\infty ||x_1-x_2||,$$

and

$$\Big| \int_{K^c+x_1} \psi(y-x_1)\, \lambda(x_1,a)\, Q(x_1,a,dy) - \int_{K^c+x_2} \psi(y-x_2)\, \lambda(x_2,a)\, Q(x_2,a,dy)\Big|$$
$$\leqslant C||\nabla\psi||_\infty ||x_1-x_2||.$$

Corollary 3.5.2. *Let Hypotheses* **(HhλQ)**, **(Hλ_0)**, **(HλQ')** *and* **(Hf)** *hold, and assume that A is compact. Then the value function V of the optimal control problem defined in* (3.25) *admits the Feynman-Kac representation formula:*

$$V(x) = Y_0^{x,a}, \quad (x,a) \in E \times A.$$

Moreover, the value function V coincides with the dual value function V^* defined in (3.50), *namely*

$$V(x) = V^*(x,a) = Y_0^{x,a}, \quad (x,a) \in E \times A. \tag{3.83}$$

Proof. Under the additional assumption **(HλQ')**, a comparison theorem for viscosity super and subsolutions for elliptic IPDEs of the form (3.26) holds, see Theorem IV.1 in Sayah [**123**]. Then, it follows from Theorem 3.5.1 that the function v in (3.82) is the unique viscosity soluton to (3.26), and it is continuous. In particular, by Theorem 3.2.6, v coincides with the value function V of the PDMPs optimal control problem, which admits therefore the probabilistic representation (3.5.2). Finally, Theorem 3.4.8 implies that the dual value function V^* coincides with the value function V of the original control problem, so that (3.83) holds. $\square$

The rest of the chapter is devoted to prove Theorem 3.5.1.

3.5.1. The identification property of the penalized BSDE. For every $n \in \mathbb{N}$ let us introduce the deterministic function v^n defined on $E \times A$ by

$$v^n(x,a) = Y_0^{n,x,a}, \quad (x,a) \in E \times A. \tag{3.84}$$

We investigate the properties of the function v^n. Firstly, it straightly follows from (3.84) and (3.57) that

$$|v^n(x,a)| \leqslant \frac{M_f}{\delta}, \quad \forall\, (x,a) \in E \times A.$$

Moreover, we have the following result.

Lemma 3.5.3. *Under Hypotheses* **(HhλQ)**, **(Hλ_0)** *and* **(Hf)**, *for any $n \in \mathbb{N}$, the function v^n is such that, for any $(x,a) \in E \times A$, we have*

$$Y_s^{n,x,a} = v^n(X_s, I_s), \quad s \geqslant 0 \quad d\mathbb{P}^{x,a} \otimes ds \text{ -a.e.} \tag{3.85}$$

Remark 3.5.4. When the pair of Markov processes (X, I) is the unique strong solution to some system of stochastic differential equations, (X, I) often satisfies a stochastic flow property, and the fact that $Y_s^{n,x,a}$ is a deterministic function of (X_s, I_s) straight follows from the uniqueness of the BSDE (see, e.g., Remark 2.4 in Barles, Buckdahn and Pardoux [**10**]). In our framework, we deal with the local characteristics of the state process (X, I) rather than with the stochastic differential equation solved by it. As a consequence, a stochastic flow property for (X, I) is no more directly available. The idea is then to prove the identification (3.85) using an iterative construction of the solution of standard BSDEs. This alternative approach is based on the fact that, when f does not depend on y, z, the desired identification follows from the Markov property of the state process (X, I), and it is inspired by the proof of the Theorem 4.1. in El Karoui, Peng and Quenez [**53**].

Proof. Fix $(x, a, n) \in E \times A \times \mathbb{N}$. Let $(Y^n, Z^n) = (Y^{n,x,a}, Z^{n,x,a})$ be the solution to the penalized BSDE (3.54). From Proposition 3.4.4 we know that there exists a sequence $(Y^{n,T}, Z^{n,T})_T = (Y^{n,T,x,a}, Z^{n,T,x,a})_T$ in $\mathbf{S}^\infty \times \mathbf{L^2_{x,a,loc}}(\mathrm{q})$ such that, when T goes to infinity, $(Y^{n,T})_T$ converges $\mathbb{P}^{x,a}$-a.s. to (Y^n) and $(Z^{n,T})_T$ converges to (Z^n) in $\mathbf{L^2_{x,a,loc}}(\mathrm{q})$. Let now fix $T, S > 0$, $S < T$, and consider the BSDE solved by $(Y^{n,T}, Z^{n,T})$ on $[0, S]$:

$$\begin{aligned} Y_t^{n,T} \quad &= \quad Y_S^{n,T} - \delta \int_t^S Y_r^{n,T}\, dr + \int_t^S f(X_r, I_r)\, dr \\ &\quad -n \int_t^S \int_A [Z_r^{n,T}(X_r, b)]^- \,\lambda_0(db)\, dr - \int_t^S \int_A Z_r^{n,T}(X_r, b)\, \lambda_0(db)\, dr, \\ &\quad - \int_t^S \int_{E\times A} Z_r^{n,T}(y, b)\, q(dr\, dy\, db), \qquad 0 \leqslant t \leqslant S. \end{aligned}$$

Then, it follows from Proposition 3.4.3 that there exists a sequence $(Y^{n,T,k}, Z^{n,T,k})_k = (Y^{n,T,k,x,a}, Z^{n,T,k,x,a})_k$ in $\mathbf{L^2_{x,a}(0,S)} \times \mathbf{L^2_{x,a}}(\mathrm{q}, \mathbf{0}, \mathbf{S})$ converging to $(Y^{n,T}, Z^{n,T})$ in $\mathbf{L^2_{x,a}(0,S)} \times \mathbf{L^2_{x,a}}(\mathrm{q}, \mathbf{0}, \mathbf{S})$, such that $(Y^{n,T,0}, Z^{n,T,0}) = (0, 0)$ and

$$Y_t^{n,T,k+1} \quad = \quad Y_S^{n,T,k} - \delta \int_t^S Y_r^{n,T,k}\, dr + \int_t^S f(X_r, I_r)\, dr$$

$$-n \int_t^S \int_A [Z_r^{n,T,k}(X_r, b)]^- \, \lambda_0(db)\, dr - \int_t^S \int_A Z_r^{n,T,k}(X_r, b)\, \lambda_0(db)\, dr,$$
$$- \int_t^S \int_{E\times A} Z_r^{n,T,k+1}(y, b)\, q(dr\, dy\, db), \qquad 0 \leqslant t \leqslant S.$$

Let us define

$$v^{n,T}(x,a) := Y_0^{n,T}, \quad v^{n,T,k}(x,a) := Y_0^{n,T,k}.$$

We start by noticing that, for $k = 0$, we have, $\mathbb{P}^{x,a}$-a.s.,

$$Y_t^{n,T,1} = \mathbb{E}^{x,a}\left[\int_t^S f(X_r, I_r)\, dr \Big| \mathcal{F}_t\right], \quad t \in [0,\, S].$$

Then, from the Markov property of (X, I) we get

$$Y_t^{n,T,1} = v^{n,T,1}(X_t, I_t), \quad d\mathbb{P}^{x,a} \otimes dt \text{ -a.e.} \tag{3.86}$$

Furthermore, identification (3.86) implies

$$Z_t^{n,T,1}(y,b) = v^{n,T,1}(X_{t-}, I_{t-}) - v^{n,T,1}(y,b), \tag{3.87}$$

where (3.87) has to be understood as an equality (almost everywhere) between elements of the space $\mathbf{L^2_{x,a}}(\mathrm{q}; \mathbf{0}, \mathbf{S})$. At this point we consider the inductive step: $1 \leqslant k \in \mathbb{N}$, and assume that, $\mathbb{P}^{x,a}$-a.s.,

$$\begin{aligned} Y_t^{n,T,k} &= v^{n,T,k}(X_t, I_t) \\ Z_t^{n,T,k}(y,b) &= v^{n,T,k}(y,b) - v^{n,T,k}(X_{t-}, I_{t-}). \end{aligned}$$

Then

$$\begin{aligned} Y_t^{n,T,k+1} &= \mathbb{E}^{x,a}\Big[v_\delta^{n,T,k}(X_S, I_S) - \delta \int_t^S v^{n,T,k}(X_r, I_r)\, dr + \int_t^S f(X_r, I_r)\, dr \\ &\quad - n \int_t^S \int_A [v^{n,T,k}(X_t, b) - v^{n,T,k}(X_t, I_t)]^- \, \lambda_0(db)\, dr \\ &\quad - \int_t^S \int_A v^{n,T,k}(X_t, b) - v^{n,T,k}(X_t, I_t)\, \lambda_0(db)\, dr \Big| \mathcal{F}_t\Big], \qquad 0 \leqslant t \leqslant S. \end{aligned}$$

Using again the Markov property of (X, I), we achieve that

$$Y_t^{n,T,k+1} = v^{n,T,k+1}(X_t, I_t), \quad d\mathbb{P}^{x,a} \otimes dt \text{ -a.e.} \tag{3.88}$$

Then, applying the Itô formula to $|Y_t^{n,T,k} - Y_t^{n,T}|^2$ and taking the supremum of t between 0 and S, one can show that

$$\mathbb{E}^{x,a}\left[\sup_{0\leqslant t\leqslant S}\left|Y_t^{n,T,k} - Y_t^{\delta,n,T}\right|^2\right] \to 0 \quad \text{as } k \text{ goes to infinity.}$$

Therefore, $v^{n,T,k}(x,a) \to v^{n,T}(x,a)$ as k goes to infinity, for all $(x,a) \in E \times A$, from which it follows that

$$Y_t^{n,T,x,a} = v^{n,T}(X_t, I_t), \quad d\mathbb{P}^{x,a} \otimes dt \text{ -a.e.} \tag{3.89}$$

Finally, from (3.66) we have that $(Y^{n,T,x,a})_T$ converges $\mathbb{P}^{x,a}$-a.s. to $(Y^{n,x,a})$ uniformly on compact sets of $\mathbb{R}$. Thus, $v^{n,T}(x,a) \to v^n(x,a)$ as T goes to infinity, for all $(x,a) \in$

$E \times A$, and this gives the requested identification $Y_t^{n,x,a} = v^n(X_t, I_t)$, $d\mathbb{P}^{x,a} \otimes dt$ - a.e. □

Remark 3.5.5. By Proposition 3.4.1, the maximal solution to the constrained BSDE (3.51)-(3.52) is the pointwise limit of the solution to the penalized BSDE (3.54). Then, as a byproduct of Lemma 3.5.3 we have the following identification for v: $\mathbb{P}^{x,a}$-a.s.,

$$v(X_s, I_s) = Y_s^{x,a}, \quad (x,a) \in E \times A,\ s \geqslant 0. \tag{3.90}$$

3.5.2. The non-dependence of the function v on the variable a. We claim that the function v in 3.82 does not depend on its last argument:

$$v(x,a) = v(x,a'), \quad a, a' \in A, \quad \text{for any } x \in E. \tag{3.91}$$

We recall that, by (3.80) and (3.82), v coincides with the value function V^* of the dual control problem introduced in Section 3.3.2. Therefore, (3.91) holds if we prove that $V^*(x,a)$ does not depend on a. This is insured by the following result.

Proposition 3.5.6. *Assume that Hypotheses* **(HhλQ)**, **(Hλ_0)** *and* **(Hf)** *hold. Fix $x \in E$, $a, a' \in A$, and $\nu \in \mathcal{V}$. Then, there exists a sequence $(\nu^\varepsilon)_\varepsilon \in \mathcal{V}$ such that*

$$\lim_{\varepsilon \to 0^+} J(x, a', \nu^\varepsilon) = J(x, a, \nu). \tag{3.92}$$

Proof. See Section 3.5.4. □

Identity (3.92) implies that

$$V^*(x,a') \leq J(x,a,\nu) \quad x \in E,\ a, a' \in A,$$

and by the arbitrariness of ν one can conclude that

$$V^*(x,a') \leq V^*(x,a) \quad x \in E,\ a, a' \in A.$$

In other words $V^*(x,a) = v(x,a)$ does not depend on a, and (3.91) holds.

3.5.3. Viscosity properties of the function v. Taking into account (3.91), by misuse of notation, we define the function v on E by

$$v(x) := v(x,a), \quad \forall x \in E, \quad \text{for any } a \in A. \tag{3.93}$$

We shall prove that the function v in (3.93) provides a viscosity solution to (3.26). We separate the proof of viscosity subsolution and supersolution properties, which are different. In particular the supersolution property is more delicate and should take into account the maximality property of $Y^{x,a}$.

Remark 3.5.7. Identity (3.90) in Remark 3.5.5 gives

$$v(X_s) = Y_s^{x,a}, \quad \forall x \in E,\ s \geqslant 0, \quad \text{for any } a \in A. \tag{3.94}$$

Proof of the viscosity subsolution property to (3.26).

Proposition 3.5.8. *Let assumptions* **(HhλQ)**, **(Hλ_0)** *and* **(Hf)** *hold. Then, the function v in (3.93) is a viscosity subsolution to (3.26).*

Proof. Let $\bar{x} \in E$, and let $\varphi \in C^1(E)$ be a test function such that

$$0 = (v^* - \varphi)(\bar{x}) = \max_{x \in E}(v^* - \varphi)(x). \tag{3.95}$$

By the definition of $v^*(\bar{x})$, there exists a sequence $(x_m)_m$ in E such that

$$x_m \to \bar{x} \text{ and } v(x_m) \to v^*(\bar{x})$$

when m goes to infinity. By the continuity of φ and by (3.95) it follows that

$$\gamma_m := \varphi(x_m) - v(x_m) \to 0,$$

when m goes to infinity. Let η be a fixed positive constant and $\tau_m := \inf\{t \geqslant 0 : |\phi(t, x_m) - x_m| \geqslant \eta\}$. Let moreover $(h_m)_m$ be a strictly positive sequence such that

$$h_m \to 0 \text{ and } \frac{\gamma_m}{h_m} \to 0,$$

when m goes to infinity.

We notice that there exists $M \in \mathbb{N}$ such that, for every $m > M$, $h_m \wedge \tau_m = h_m$. Let us introduce $\bar{\tau} := \inf\{t \geqslant 0 : |\phi(t, \bar{x}) - \bar{x}| \geqslant \eta\}$. Clearly $\bar{\tau} > 0$. We show that it does not exists a subsequence τ_{n_k} of τ_n such that $\tau_{n_k} \to \tau_0 \in [0, \bar{\tau})$. Indeed, let $\tau_{n_k} \to \tau_0 \in [0, \bar{\tau})$. In particular $|\phi(\tau_{n_k}, \bar{x}) - \bar{x}| \geqslant \eta$. Then, by the continuity of ϕ it follows that $|\phi(\tau_0, \bar{x}) - \bar{x}| \geqslant \eta$, and this is in contradiction with the definition of $\bar{\tau}$.

Let now fix $a \in A$, and let $Y^{x_m,a}$ be the unique maximal solution to (3.51)-(3.52) under $\mathbb{P}^{x_m,a}$. We apply the Itô formula to $e^{-\delta t}\, Y_t^{x_m,a}$ between 0 and $\theta_m := \tau_m \wedge h_m \wedge T_1$, where T_1 denotes the first jump time of (X, I). Using the identification (3.94), from the constraint (3.52) and the fact that K is a nondecreasing process it follows that $\mathbb{P}^{x_m,a}$-a.s.,

$$\begin{aligned} v(x_m) \quad &\leqslant \quad e^{-\delta\theta_m}\, v(X_{\theta_m}) + \int_0^{\theta_m} e^{-\delta r}\, f(X_r, I_r)\, dr \\ &\qquad - \int_0^{\theta_m} e^{-\delta r} \int_E (v(y) - v(X_r))\, \tilde{q}(dr\, dy), \end{aligned}$$

where $\tilde{q}(dr\, dy) = p(dr\, dy) - \lambda(X_r, I_r)\, Q(X_r, I_r, dy)\; dr$. In particular

$$v(x_m) \leqslant \mathbb{E}^{x_m,a}\left[e^{-\delta\theta_m}\, v(X_{\theta_m}) + \int_0^{\theta_m} e^{-\delta r}\, f(X_r, I_r)\, dr\right].$$

Equation (3.95) implies that $v \leqslant v^* \leqslant \varphi$, and therefore

$$\varphi(x_m) - \gamma_m \leqslant \mathbb{E}^{x_m,a}\left[e^{-\delta\theta_m}\, \varphi(X_{\theta_m}) + \int_0^{\theta_m} e^{-\delta r}\, f(X_r, I_r)\, dr\right].$$

At this point, applying Itô's formula to $e^{-\delta r}\, \varphi(X_r)$ between 0 and θ_m, we get

$$-\frac{\gamma_m}{h_m} + \mathbb{E}^{x_m,a}\left[\int_0^{\theta_m} \frac{1}{h_m}\, e^{-\delta r}\, [\delta\, \varphi(X_r) - \mathcal{L}^{I_r}\varphi(X_r) - f(X_r, I_r)]\, dr\right] \leqslant 0, \tag{3.96}$$

where $\mathcal{L}^{I_r}\varphi(X_r) = \int_E (\varphi(y) - \varphi(X_r))\, \lambda(X_r, I_r)\, Q(X_r, I_r, dy)$. Now we notice that, $\mathbb{P}^{x_m,a}$-a.s., $(X_r, I_r) = (\phi(r, x_m), a)$ for $r \in [0,\, \theta_m]$. Taking into account the continuity of the map $(y,b) \mapsto \delta\,\varphi(y) - \mathcal{L}^b\varphi(y) - f(y,b)$, we see that for any $\varepsilon > 0$,

$$-\frac{\gamma_m}{h_m} + (\varepsilon + \delta\,\varphi(x_m) - \mathcal{L}^a\varphi(x_m) - f(x_m,a))\,\mathbb{E}^{x_m,a}\left[\frac{\theta_m\, e^{-\delta\,\theta_m}}{h_m}\right] \leqslant 0, \tag{3.97}$$

Let $f_{T_1}(s)$ denote the distribution density of T_1 under $\mathbb{P}^{x_m,a}$, see (3.34). Taking $m > M$, we have

$$\begin{aligned}\mathbb{E}^{x_m,a}\left[\frac{g(\theta_m)}{h_m}\right] &= \frac{1}{h_m}\int_0^{h_m} s\, e^{-\delta\, s}\, f_{T_1}(s)\, ds + \frac{h_m\, e^{-\delta\, h_m}}{h_m}\, \mathbb{P}^{x_m,a}[T_1 > h_m]\\ &= \frac{1}{h_m}\int_0^{h_m} s\, e^{-\delta\, s}\,(\lambda(\phi(r,x_m),a) + \lambda_0(A))\, e^{-\int_0^s(\lambda(\phi(r,x_m),a)+\lambda_0(A))\,dr}\, ds\\ &\quad + e^{-\delta\, h_m}\, e^{-\int_0^{h_m}(\lambda(\phi(r,x_m),a)+\lambda_0(A))\,dr}.\end{aligned} \tag{3.98}$$

By the boundedness of λ and λ_0, it is easy to see that the two terms in the right-hand side of (3.98) converge respectively to zero and one when m goes to infinity. Thus, passing into the limit in (3.97) as m goes to infinity, we obtain

$$\delta\,\varphi(\bar{x}) - \mathcal{L}^a\varphi(\bar{x}) - f(\bar{x},a) \leqslant 0.$$

From the arbitrariness of $a \in A$ we conclude that v is a viscosity subsolution to (3.26) in the sense of Definition 3.2.5. □

Proof of the viscosity supersolution property to (3.26)**.**

Proposition 3.5.9. *Let assumptions* **(HhλQ)**, **(Hλ_0)**, *and* **(Hf)** *hold. Then, the function v in* (3.93) *is a viscosity supersolution to* (3.26)*.*

Proof. Let $\bar{x} \in E$, and let $\varphi \in C^1(E)$ be a test function such that

$$0 = (v_* - \varphi)(\bar{x}) = \min_{x\in E}(v_* - \varphi)(x). \tag{3.99}$$

Notice that we can assume w.l.o.g. that $\bar{x}$ is strict minimum of $v_* - \varphi$. As a matter of fact, one can subtract to φ a positive cut-off function which behaves as $|x-\bar{x}|^2$ when $|x-\bar{x}|^2$ is small, and that regularly converges to 1 as $|x-\bar{x}|^2$ increases to 1.

Then, for every $\eta > 0$, we can define

$$0 < \beta(\eta) := \inf_{x\notin B(\bar{x},\eta)}(v_* - \varphi)(x). \tag{3.100}$$

We will show the result by contradiction. Assume thus that

$$H^\varphi(\bar{x},\varphi,\nabla\varphi) < 0.$$

Then by the continuity of H, there exists $\eta > 0$, $\beta(\eta) > 0$ and $\varepsilon \in (0, \beta(\eta)\delta]$ such that

$$H^\varphi(y,\varphi,\nabla\varphi) \leqslant -\varepsilon,$$

for all $y \in B(\bar{x},\eta) = \{y \in E : |\bar{x} - y| < \eta\}$. By definition of $v_*(\bar{x})$, there exists a sequence $(x_m)_m$ taking values in $B(\bar{x},\eta)$ such that

$$x_m \to \bar{x} \text{ and } v(x_m) \to v_*(\bar{x})$$

when m goes to infinity. By the continuity of φ and by (3.99) it follows that

$$\gamma_m := v(x_m) - \varphi(x_m) \to 0,$$

when m goes to infinity. Let fix $T > 0$ and define $\theta := \tau \wedge T$, where $\tau = \inf\{t \geqslant 0 : X_t \notin B(\bar{x}, \eta)\}$.

At this point, let us fix $a \in A$, and consider the solution $Y^{n,x_m,a}$ to the penalized (3.54), under the probability $\mathbb{P}^{x_m,a}$. Notice that

$$\mathbb{P}^{x_m,a}\{\tau = 0\} = \mathbb{P}^{x_m,a}\{X_0 \notin B(\bar{x}, \eta)\} = 0.$$

We apply the Itô formula to $e^{-\delta t}\, Y_t^{n,x_m,a}$ between 0 and θ. Then, proceeding as in the proof of Lemma 3.4.5 we get the following inequality:

$$Y_0^{n,x_m,a} \geqslant \inf_{\nu \in \mathcal{V}^n} \mathbb{E}_\nu^{x_m,a}\left[e^{-\delta\theta}\, Y_\theta^{n,x_m,a} + \int_0^\theta e^{-\delta r}\, f(X_r, I_r)\, dr\right]. \tag{3.101}$$

Since $Y^{n,x_m,a}$ converges decreasingly to the maximal solution $Y^{x_m,a}$ to the constrained BSDE (3.51)-(3.52), and recalling the identification (3.94), inequality (3.101) leads to the corresponding inequality for $v(x_m)$:

$$v(x_m) \geqslant \inf_{\nu \in \mathcal{V}} \mathbb{E}_\nu^{x_m,a}\left[e^{-\delta\theta}\, v(X_\theta) + \int_0^\theta e^{-\delta r}\, f(X_r, I_r)\, dr\right].$$

In particular, there exists a strictly positive, predictable and bounded function ν_m such that

$$v(x_m) \geqslant \mathbb{E}_{\nu_m}^{x_m,a}\left[e^{-\delta\theta}\, v(X_\theta) + \int_0^\theta e^{-\delta r}\, f(X_r, I_r)\, dr\right] - \frac{\varepsilon}{2\delta}. \tag{3.102}$$

Now, from equation (3.99) and (3.100) we get

$$\varphi(x_m) + \gamma_m \geqslant \mathbb{E}_{\nu_m}^{x_m,a}\left[e^{-\delta\theta}\, \varphi(X_\theta) + \beta\, e^{-\delta\theta}\, 1_{\{\tau \leqslant T\}} + \int_0^\theta e^{-\delta r}\, f(X_r, I_r)\, dr\right] - \frac{\varepsilon}{2\,\delta}.$$

At this point, applying Itô's formula to $e^{-\delta r}\, \varphi(X_r)$ between 0 and θ, we get

$$\begin{aligned}\gamma_m + \mathbb{E}_{\nu_m}^{x_m,a}&\left[\int_0^\theta e^{-\delta r}\left[\delta\, \varphi(X_r) - \mathcal{L}^{I_r}\varphi(X_r) - f(X_r, I_r)\right] dr - \beta\, e^{-\delta\theta}\, 1_{\{\tau \leqslant T\}}\right] \\ &+ \frac{\varepsilon}{2} \geqslant 0,\end{aligned} \tag{3.103}$$

where $\mathcal{L}^{I_r}\varphi(X_r) = \int_E (\varphi(y) - \varphi(X_r))\, \lambda(X_r, I_r)\, Q(X_r, I_r, dy)$. Noticing that, for $r \in [0, \theta]$,

$$\begin{aligned}\delta\, \varphi(X_r) - \mathcal{L}^{I_r}\varphi(X_r) - f(X_r, I_r) &\leqslant \delta\, \varphi(X_r) - \inf_{b \in A}\{\mathcal{L}^b\varphi(X_r) - f(X_r, b)\} \\ &= H^\varphi(X_r, \varphi, \nabla\varphi) \\ &\leqslant -\varepsilon,\end{aligned}$$

from (3.103) we obtain

$$\begin{aligned}0 &\leqslant \gamma_m + \frac{\varepsilon}{2\,\delta} + \mathbb{E}_{\nu_m}^{x_m,a}\left[-\varepsilon \int_0^\theta e^{-\delta r}\, dr - \beta\, e^{-\delta\theta}\, 1_{\{\tau \leqslant T\}}\right] \\ &= \gamma_m - \frac{\varepsilon}{2\,\delta} + \mathbb{E}_{\nu_m}^{x_m,a}\left[\left(\frac{\varepsilon}{\delta} - \beta\right) e^{-\delta\theta} 1_{\{\tau \leqslant T\}} + \frac{\varepsilon}{\delta} e^{-\delta\theta}\, 1_{\{\tau > T\}}\right] \\ &\leqslant \gamma_m - \frac{\varepsilon}{2\,\delta} + \frac{\varepsilon}{\delta}\, \mathbb{E}_{\nu_m}^{x_m,a}\left[e^{-\delta\theta}\, 1_{\{\tau > T\}}\right]\end{aligned}$$

$$
\begin{aligned}
&= \gamma_m - \frac{\varepsilon}{2\,\delta} + \frac{\varepsilon}{\delta}\,\mathbb{E}_{\nu_m}^{x_m,a}\left[e^{-\delta T}\,1_{\{\tau>T\}}\right]\\
&\leqslant \gamma_m - \frac{\varepsilon}{2\,\delta} + e^{-\delta T}.
\end{aligned}
$$

Letting T and m go to infinity we achieve the contradiction: $0 \leqslant -\frac{\varepsilon}{2\,\delta}$. □

3.5.4. Proof of Proposition 3.5.6. We start by giving a technical result. In the sequel, Π^{n_1,n_2} and Γ^{n_1,n_2} will denote respectively the random sequences $(T_{n_1}, E_{n_1}, A_{n_1}, T_{n_1+1}, E_{n_1+1}, A_{n_1+1}, ..., T_{n_2}, E_{n_2}, A_{n_2})$ and $(T_{n_1}, A_{n_1}, T_{n_1+1}, A_{n_1+1}, ..., T_{n_2}, A_{n_2})$, n_1, $n_2 \in \mathbb{N}\setminus\{0\}$, $n_1 \leq n_2$, where $(T_k,\ E_k,\ A_k)_{k\geq 1}$ denotes the sequence of random variables introduced in Section 3.3.1.

Lemma 3.5.10. *Assume that Hypotheses* **(HhλQ)**, **(Hλ_0)** *and* **(Hf)** *hold. Let $\nu^n : \Omega\times\mathbb{R}_+\times(\mathbb{R}_+\times A)^n\times A \to (0,\infty)$, $n > 1$ (resp. $\nu^0 : \Omega\times\mathbb{R}_+\times A\to(0,\infty)$), be some $\mathcal{P}\otimes\mathcal{B}((\mathbb{R}_+\times A)^n)\otimes\mathcal{A}$-measurable maps, uniformly bounded with respect to n (resp. a bounded $\mathcal{P}\otimes\mathcal{A}$-measurable map). Let moreover $g : \Omega\times A\to(0,\infty)$ be a bounded $\mathcal{A}$-measurable map, and set*

$$
\nu_t(b) = \nu_t^0(b)\,1_{\{t\leqslant T_1\}} + \sum_{n=1}^{\infty}\nu_t^n(\Gamma^{1,n},b)\,1_{\{T_n<t\leqslant T_{n+1}\}}, \tag{3.104}
$$

$$
\nu_t'(b) = g(b)\,1_{\{t\leqslant T_1\}} + \nu_t^0(b)\,1_{\{T_1<t\leqslant T_2\}} + \sum_{n=2}^{\infty}\nu_t^{n-1}(\Gamma^{2,n},b)\,1_{\{T_n<t\leqslant T_{n+1}\}}. \tag{3.105}
$$

Fix $x\in E$, $a, a'\in A$. Then, for every $n>1$, for every $\mathcal{B}((\mathbb{R}_+\times E\times A)^n)$-measurable function $F : (\mathbb{R}_+\times E\times A)^n\to\mathbb{R}$,

$$
\mathbb{E}_{\nu'}^{x,a'}\left[F(\Pi^{1,n})|\mathcal{F}_{T_1}\right] = \left.\frac{\mathbb{E}_{\nu}^{x,a}\left[1_{\{T_1>\tau\}}\,F(\tau,\chi,\xi,\Pi^{1,n-1})\right]}{\mathbb{P}_{\nu}^{x,a}(T_1>\tau)}\right|_{\tau=T_1,\chi=X_1,\xi=A_1}. \tag{3.106}
$$

Proof of the Lemma. Taking into account (3.36), (3.37), and (3.105), we have: for all $r\geqslant T_1$,

$$
\begin{aligned}
&\mathbb{P}_{\nu'}^{x,a'}\left[T_2>r, E_2\in F, A_2\in C|\mathcal{F}_{T_1}\right]\\
&= \int_r^\infty\int_F \exp\left(-\int_{T_1}^s\lambda(\phi(t-T_1,E_1,A_1),A_1)\,dt - \int_{T_1}^s\int_A \nu_t^0(b)\,\lambda_0(db)\,dt\right)\cdot\\
&\cdot\lambda(\phi(s-T_1,E_1,A_1),A_1)\,Q(\phi(s-T_1,E_1,A_1),A_1,dy)\,ds\\
&+ \int_r^\infty\int_C \exp\left(-\int_{T_1}^s\lambda(\phi(t-T_1,E_1,A_1),A_1)\,dt - \int_{T_1}^s\int_A \nu_t^0(b)\,\lambda_0(db)\,dt\right)\cdot\\
&\cdot\nu_s^0(b)\,\lambda_0(db)\,ds,
\end{aligned} \tag{3.107}
$$

and, for all $r\geqslant T_n$, $n>2$,

$$
\begin{aligned}
&\mathbb{P}_{\nu'}^{x,a}\left[T_{n+1}>r, E_{n+1}\in F, A_{n+1}\in C|\mathcal{F}_{T_n}\right]\\
&= \int_r^\infty\int_F \exp\left(-\int_{T_n}^s\lambda(\phi(t-T_n,E_n,A_n),A_n)\,dt\right)\cdot\\
&\cdot\exp\left(-\int_{T_n}^s\int_A \nu_t^{n-1}(\Gamma^{2,n},b)\,\lambda_0(db)\,dt\right)\cdot
\end{aligned}
$$

$$\cdot\, \lambda(\phi(s-T_n,E_n,A_n),A_n)\, Q(\phi(s-T_n,E_n,A_n),A_n,dy)\, ds$$
$$+\int_r^\infty \int_C \exp\bigg(-\int_{T_n}^s \lambda(\phi(t-T_n,E_n,A_n),A_n)\, dt\bigg)\cdot$$
$$\cdot \exp\bigg(-\int_{T_n}^s\int_A \nu_t^{n-1}(\Gamma^{2,n},b)\,\lambda_0(db)\,dt\bigg)\nu_s^{n-1}(\Gamma^{2,n},b)\,\lambda_0(db)\,ds. \qquad (3.108)$$

We will prove identity (3.106) by induction. Let us start by showing that (3.106) holds in the case $n=2$, namely that, for every $\mathcal{B}((\mathbb{R}_+\times E\times A)^2)$-measurable function $F:(\mathbb{R}_+\times E\times A)^2\to\mathbb{R}$,

$$\mathbb{E}_{\nu'}^{x,a'}\left[F(\Pi^{1,2})|\mathcal{F}_{T_1}\right]=\frac{\mathbb{E}_\nu^{x,a}\left[1_{\{T_1>\tau\}}\,F(\tau,\chi,\xi,\Pi^{1,1})\right]}{\mathbb{P}_\nu^{x,a}(T_1>\tau)}\bigg|_{\tau=T_1,\chi=X_1,\xi=A_1}. \qquad (3.109)$$

From (3.107) we get

$$\mathbb{E}_{\nu'}^{x,a'}\left[F(\Pi^{1,2})|\mathcal{F}_{T_1}\right]=\mathbb{E}_{\nu'}^{x,a'}\left[F(T_1,E_1,A_1,T_2,E_2,A_2)|\mathcal{F}_{T_1}\right]$$
$$=\int_{T_1}^\infty\int_E F(T_1,E_1,A_1,s,y,A_1)\cdot$$
$$\cdot\exp\bigg(-\int_{T_1}^s\lambda(\phi(t-T_1,E_1,A_1),A_1)\,dt-\int_{T_1}^s\int_A\nu_t^0(b)\,\lambda_0(db)\,dt\bigg)\cdot$$
$$\cdot\,\lambda(\phi(s-T_1,E_1,A_1),A_1)\,Q(\phi(s-T_1,E_1,A_1),A_1,dy)\,ds$$
$$+\int_{T_1}^\infty\int_A F(T_1,E_1,A_1,s,\phi(s-T_1,E_1,A_1),b)\cdot$$
$$\cdot\exp\bigg(-\int_{T_1}^s\lambda(\phi(t-T_1,E_1,A_1),A_1)\,dt-\int_{T_1}^s\int_A\nu_t^0(b)\,\lambda_0(db)\,dt\bigg)\;\nu_s^0(b)\,\lambda_0(db)\,ds.$$

On the other hand,

$$\mathbb{P}_\nu^{x,a}(T_1>\tau)=\exp\bigg(-\int_0^\tau\lambda(\phi(t-\tau,\chi,\xi),\xi)\,dt-\int_0^\tau\int_A\nu_t^0(b)\,\lambda_0(db)\,dt\bigg),$$

and

$$\mathbb{E}_\nu^{x,a}\left[1_{\{T_1>\tau\}}\,F(\tau,\chi,\xi,\Pi^{1,1})\right]=\mathbb{E}_\nu^{x,a}\left[1_{\{T_1>\tau\}}\,F(\tau,\chi,\xi,T_1,E_1,A_1)\right]$$
$$=\int_\tau^\infty\int_E 1_{\{s>\tau\}}\,F(\tau,\chi,\xi,s,y,\xi)\cdot$$
$$\cdot\exp\bigg(-\int_0^s\lambda(\phi(t-\tau,\chi,\xi),\xi)\,dt-\int_0^s\int_A\nu_t^0(b)\,\lambda_0(db)\,dt\bigg)\cdot$$
$$\cdot\,\lambda(\phi(s-\tau,\chi,\xi),\xi)\,Q(\phi(s-\tau,\chi,\xi),\xi,dy)\,ds$$
$$+\int_\tau^\infty\int_A 1_{\{s>\tau\}}\,F(\tau,\chi,\xi,s,\phi(s-\tau,\chi,\xi),b)\cdot$$
$$\cdot\exp\bigg(-\int_0^s\lambda(\phi(t-\tau,\chi,\xi),\xi)\,dt-\int_0^s\int_A\nu_t^0(b)\,\lambda_0(db)\,dt\bigg)\;\nu_s^0(b)\,\lambda_0(db)\,ds.$$

Therefore,

$$\frac{\mathbb{E}_\nu^{x,a}\left[1_{\{T_1>\tau\}}\,F(\tau,\chi,\xi,\Pi^{1,1})\right]}{\mathbb{P}_\nu^{x,a}(T_1>\tau)}$$

$$
\begin{aligned}
&= \exp\left(\int_0^\tau \lambda(\phi(t-\tau,\chi,\xi),\xi)\,dt + \int_0^\tau \int_A \nu_t^0(b)\,\lambda_0(db)\,dt\right)\cdot \\
&\cdot \int_\tau^\infty \int_E 1_{\{s>\tau\}}\,F(\tau,\chi,\xi,s,y,\xi)\cdot \\
&\cdot \exp\left(-\int_0^s \lambda(\phi(t-\tau,\chi,\xi),\xi)\,dt - \int_0^s \int_A \nu_t^0(b)\,\lambda_0(db)\,dt\right)\cdot \\
&\cdot \lambda(\phi(s-\tau,\chi,\xi),\xi)\,Q(\phi(s-\tau,\chi,\xi),\xi,dy)\,ds \\
&+ \exp\left(\int_0^\tau \lambda(\phi(t-\tau,\chi,\xi),\xi)\,dt + \int_0^\tau \int_A \nu_t^0(b)\,\lambda_0(db)\,dt\right)\cdot \\
&\cdot \int_\tau^\infty \int_A 1_{\{s>\tau\}}\,F(\tau,\chi,\xi,s,\phi(s-\tau,\chi,\xi),b)\cdot \\
&\cdot \exp\left(-\int_0^s \lambda(\phi(t-\tau,\chi,\xi),\xi)\,dt - \int_0^s \int_A \nu_t^0(b)\,\lambda_0(db)\,dt\right)\,\nu_s^0(b)\,\lambda_0(db)\,ds \\
&= \int_\tau^\infty \int_E 1_{\{s>\tau\}}\,F(\tau,\chi,\xi,s,y,\xi)\cdot \\
&\cdot \exp\left(-\int_\tau^s \lambda(\phi(t-\tau,\chi,\xi),\xi)\,dt - \int_\tau^s \int_A \nu_t^0(b)\,\lambda_0(db)\,dt\right)\cdot \\
&\cdot \lambda(\phi(s-\tau,\chi,\xi),\xi)\,Q(\phi(s-\tau,\chi,\xi),\xi,dy)\,ds \\
&+ \int_\tau^\infty \int_A 1_{\{s>\tau\}}\,F(\tau,\chi,\xi,s,\phi(s-\tau,\chi,\xi),b)\cdot \\
&\cdot \exp\left(-\int_\tau^s \lambda(\phi(t-\tau,\chi,\xi),\xi)\,dt - \int_\tau^s \int_A \nu_t^0(b)\,\lambda_0(db)\,dt\right)\,\nu_s^0(b)\,\lambda_0(db)\,ds,
\end{aligned}
$$

and (3.109) follows.

Assume now that (3.106) holds for $n-1$, namely that, for every $\mathcal{B}((\mathbb{R}_+ \times E \times A)^{n-1})$-measurable function $F : (\mathbb{R}_+ \times E \times A)^{n-1} \to \mathbb{R}$,

$$
\mathbb{E}_{\nu'}^{x,a'}\left[F(\Pi^{1,n-1})|\mathcal{F}_{T_1}\right] = \frac{\mathbb{E}_{\nu}^{x,a}\left[1_{\{T_1>\tau\}}\,F(\tau,\chi,\xi,\Pi^{1,n-2})\right]}{\mathbb{P}_{\nu}^{x,a}(T_1>\tau)}\bigg|_{\tau=T_1,\chi=X_1,\,\xi=A_1}. \tag{3.110}
$$

We have to prove that (3.110) implies that, for every $\mathcal{B}((\mathbb{R}_+ \times E \times A)^n)$-measurable function $F : (\mathbb{R}_+ \times E \times A)^n \to \mathbb{R}$,

$$
\mathbb{E}_{\nu'}^{x,a'}\left[F(\Pi^{1,n})|\mathcal{F}_{T_1}\right] = \frac{\mathbb{E}_{\nu}^{x,a}\left[1_{\{T_1>\tau\}}\,F(\tau,\chi,\xi,\Pi^{1,n-1})\right]}{\mathbb{P}_{\nu}^{x,a}(T_1>\tau)}\bigg|_{\tau=T_1,\chi=X_1,\,\xi=A_1}. \tag{3.111}
$$

Using (3.108), we get

$$
\begin{aligned}
&\mathbb{E}_{\nu'}^{x,a'}\left[F(\Pi^{1,n})|\mathcal{F}_{T_1}\right] \\
&= \mathbb{E}_{\nu'}^{x,a'}\left[\mathbb{E}_{\nu'_\varepsilon}^{x,a'}\left[F(\Pi^{1,n})|\mathcal{F}_{T_{n-1}}\right]|\mathcal{F}_{T_1}\right] \\
&= \mathbb{E}_{\nu'}^{x,a'}\bigg[\int_{T_{n-1}}^\infty \int_E F(\Pi^{1,n-1},s,y,A_{n-1})\cdot \\
&\cdot \exp\bigg(-\int_{T_{n-1}}^s \lambda(\phi(t-T_{n-1},E_{n-1},A_{n-1}),A_{n-1})\,dt
\end{aligned}
$$

$$
\begin{aligned}
&\qquad - \int_{T_{n-1}}^{s} \int_A \nu_t^{n-2}(\Gamma^{1,n-1}, b)\, \lambda_0(db)\, dt\Big) \cdot \\
&\cdot \lambda(\phi(s - T_{n-1}, E_{n-1}, A_{n-1}), A_{n-1})\, Q(\phi(s - T_{n-1}, E_{n-1}, A_{n-1}), A_{n-1}, dy)\, ds \\
&+ \int_{T_{n-1}}^{\infty} \int_A F(\Pi^{1,n-1}, s, \phi(s - T_{n-1}, E_{n-1}, A_{n-1}), b) \cdot \\
&\cdot \exp\Big(- \int_{T_{n-1}}^{s} \lambda(\phi(t - T_{n-1}, E_{n-1}, A_{n-1}), A_{n-1})\, dt \\
&\qquad - \int_{T_{n-1}}^{s} \int_A \nu_t^{n-2}(\Gamma^{1,n-1}, b)\, \lambda_0(db)\, dt\Big)\, \nu_s^{n-2}(\Gamma^{1,n-1}, b)\, \lambda_0(db)\, ds \Big| \mathcal{F}_{T_1}\Big].
\end{aligned}
\tag{3.112}
$$

At this point we observe that the term in the conditional expectation in the right-hand side of (3.112) only depends on the random sequence $\Pi^{1,n-1}$. For any sequence of random variables $(S_i, W_i, V_i)_{i \in [1,n-1]}$ with values in $([0, \infty) \times E \times A)^{n-1}$, $S_{i-1} \leq S_i$ for every $i \in [1, n-1]$, we set

$$
\begin{aligned}
&\psi(S_1, W_1, V_1, ..., S_{n-1}, W_{n-1}, V_{n-1}) := \\
&\int_{S_{n-1}}^{\infty} \int_E F(S_1, W_1, ..., V_{n-1}, S_{n-1}, W_{n-1}, s, y, V_{n-1}) \cdot \\
&\cdot \exp\Big(- \int_{S_{n-1}}^{s} \lambda(\phi(t - S_{n-1}, W_{n-1}, V_{n-1}), V_{n-1})\, dt \\
&\qquad - \int_{S_{n-1}}^{s} \int_A \nu_t^{n-2}(S_1, V_1, ..., S_{n-1}, V_{n-1}, b)\, \lambda_0(db)\, dt\Big) \cdot \\
&\cdot \lambda(\phi(s - S_{n-1}, W_{n-1}, V_{n-1}), V_{n-1})\, Q(\phi(s - S_{n-1}, W_{n-1}, V_{n-1}), V_{n-1}, dy)\, ds \\
&+ \int_{S_{n-1}}^{\infty} \int_A F(S_1, W_1, V_1, ..., S_{n-1}, W_{n-1}, V_{n-1}, , s, \phi(s - S_{n-1}, W_{n-1}, V_{n-1}), b) \cdot \\
&\cdot \exp\Big(- \int_{S_{n-1}}^{s} \lambda(\phi(t - S_{n-1}, W_{n-1}, V_{n-1}), V_{n-1})\, dt \\
&\qquad - \int_{S_{n-1}}^{s} \int_A \nu_t^{n-2}(S_1, V_1, ..., S_{n-1}, V_{n-1}, b)\, \lambda_0(db)\, dt\Big) \cdot \\
&\cdot \nu_s^{n-2}(S_1, V_1, ..., S_{n-1}, V_{n-1}, b)\, \lambda_0(db)\, ds.
\end{aligned}
$$

Identity (3.112) can be rewritten as

$$
\mathbb{E}_{\nu'}^{x,a'}\left[F(\Pi^{1,n}) | \mathcal{F}_{T_1}\right] = \mathbb{E}_{\nu'}^{x,a'}\left[\psi(\Pi^{1,n-1}) \Big| \mathcal{F}_{T_1}\right]. \tag{3.113}
$$

Then, by applying the inductive step (3.110), we get

$$
\begin{aligned}
&\mathbb{E}_{\nu'}^{x,a'}\left[F(\Pi^{1,n}) | \mathcal{F}_{T_1}\right] \\
&= \mathbb{E}_{\nu'}^{x,a'}\left[\psi(\Pi^{1,n-1}) \Big| \mathcal{F}_{T_1}\right] \\
&= (\mathbb{P}_{\nu}^{x,a}[T_1 > \tau])^{-1}\, \mathbb{E}_{\nu}^{x,a}\left[1_{\{T_1 > \tau\}}\, \psi(\tau, \chi, \xi, \Pi^{1,n-2})\right]\Big|_{\tau = T_1, \chi = X_1, \xi = A_1}.
\end{aligned}
\tag{3.114}
$$

Since

$$
\begin{aligned}
&\psi(\tau,\chi,\xi,\Pi^{1,n-2})\\
&=\int_{T_{n-2}}^{\infty}\int_E F(\tau,\chi,\xi,\Pi^{1,n-2},s,y,A_{n-2})\cdot\\
&\cdot\exp\bigg(-\int_{T_{n-2}}^{s}\lambda(\phi(t-T_{n-2},E_{n-2},A_{n-2}),A_{n-2})\,dt-\int_{T_{n-2}}^{s}\int_A\nu_t^{n-2}(\Gamma^{1,n-2},b)\,\lambda_0(db)\,dt\bigg)\cdot\\
&\cdot\lambda(\phi(s-T_{n-2},E_{n-2},A_{n-2}),A_{n-2})\,Q(\phi(s-T_{n-2},E_{n-2},A_{n-2}),A_{n-2},dy)\,ds\\
&+\int_{T_{n-2}}^{\infty}\int_A F(\tau,\chi,\xi,\Pi^{1,n-2},s,\phi(s-T_{n-2},E_{n-2},A_{n-2}),b)\cdot\\
&\cdot\exp\bigg(-\int_{T_{n-2}}^{s}\lambda(\phi(t-T_{n-2},E_{n-2},A_{n-2}),A_1)\,dt-\int_{T_{n-2}}^{s}\int_A\nu_t^{n-2}(\Gamma^{1,n-2},b)\,\lambda_0(db)\,dt\bigg)\cdot\\
&\cdot\nu_s^{n-2}(\Gamma^{1,n-2},b)\,\lambda_0(db)\,ds\\
&=\mathbb{E}_\nu^{x,a}[F(\tau,\chi,\xi,\Pi^{1,n-1})|\mathcal{F}_{T_{n-2}}],
\end{aligned}
$$

identity (3.114) can be rewritten as

$$
\begin{aligned}
&\mathbb{E}_{\nu'}^{x,a'}\left[F(\Pi^{1,n})|\mathcal{F}_{T_1}\right]\\
&=(\mathbb{P}_\nu^{x,a}[T_1>\tau])^{-1}\,\mathbb{E}_\nu^{x,a}\Big[1_{\{T_1>\tau\}}\,\mathbb{E}_\nu^{x,a}[F(\tau,\chi,\xi,\Pi^{1,n-1})|\mathcal{F}_{T_{n-2}}]\Big]\Big|_{\tau=T_1,\chi=X_1,\xi=A_1}\\
&=\frac{\mathbb{E}_\nu^{x,a}\left[1_{\{T_1>\tau\}}\,F(\tau,\chi,\xi,\Pi^{1,n-1})\right]}{\mathbb{P}_\nu^{x,a}(T_1>\tau)}\bigg|_{\tau=T_1,\,\chi=E_1,\,\xi=A_1}.
\end{aligned}
\tag{3.115}
$$

This concludes the proof of the Lemma. □

Proof of Proposition 3.5.6. We start by noticing that,

$$
J(x,a,\nu)=\mathbb{E}_\nu^{x,a}\left[F(T_1,E_1,A_1,T_2,E_2,A_2,...)\right],
$$

where

$$
\begin{aligned}
&F(T_1,E_1,A_1,T_2,E_2,A_2,...)\\
&=\int_0^\infty e^{-\delta t}f(X_t,I_t)\,dt\\
&=\int_0^{T_1}e^{-\delta t}f(\phi(t,X_0,I_0),I_0)\,dt+\sum_{n=2}^{\infty}\int_{T_{n-1}}^{T_n}e^{-\delta t}f(\phi(t-T_{n-1},E_{n-1},A_{n-1}),A_{n-1})\,dt.
\end{aligned}
\tag{3.116}
$$

We aim at constructing a sequence of controls $(\nu^\varepsilon)_\varepsilon\in\mathcal{V}$ such that

$$
\begin{aligned}
J(x,a',\nu^\varepsilon)&=\mathbb{E}_{\nu^\varepsilon}^{x,a'}\left[F(T_1,E_1,A_1,T_2,E_2,A_2,...)\right]\\
&\underset{\varepsilon\to0}{\longrightarrow}\mathbb{E}_\nu^{x,a}\left[F(T_1,E_1,A_1,T_2,E_2,A_2,...)\right]=J(x,a,\nu).
\end{aligned}
\tag{3.117}
$$

Since $\nu\in\mathcal{V}$, then there exists a $\mathbb{P}^{x,a}$-null set $\mathcal{N}$ such that ν admits the representation

$$
\nu_t(b)=\nu_t^0(b)\,1_{\{t\leqslant T_1\}}+\sum_{n=1}^{\infty}\nu_t^n(T_1,A_1,T_2,A_2,...,T_n,A_n,b)\,1_{\{T_n<t\leqslant T_{n+1}\}}
\tag{3.118}
$$

for all $(\omega, t) \in \Omega \times \mathbb{R}_+$, $\omega \notin \mathcal{N}$, for some $\nu^n : \Omega \times \mathbb{R}_+ \times (\mathbb{R}_+ \times A)^n \times A \to (0, \infty)$, $n > 1$ (resp. $\nu^0 : \Omega \times \mathbb{R}_+ \times A \to (0, \infty)$) $\mathcal{P} \otimes \mathcal{B}((\mathbb{R}_+ \times A)^n) \otimes \mathcal{A}$-measurable maps, uniformly bounded with respect to n (resp. bounded $\mathcal{P} \otimes \mathcal{A}$-measurable map), see, e.g., Definition 26.3 in [**35**].

Let $\bar{B}(a, \varepsilon)$ be the closed ball centered in a with radius ε. We notice that $\varepsilon \mapsto \lambda_0(\bar{B}(a, \varepsilon))$ defines a nonnegative, right-continuous, nondecreasing function, satisfying

$$\lambda_0(\bar{B}(a, 0)) = \lambda_0(\{a\}) \geq 0, \quad \lambda_0(\bar{B}(a, \varepsilon)) > 0 \quad \forall \varepsilon > 0.$$

If $\lambda_0(\{a\}) > 0$, we set $h(\varepsilon) = \varepsilon$ for every $\varepsilon > 0$. Otherwise, if $\lambda_0(\{a\}) = 0$, we define h as the right inverse function of $\varepsilon \mapsto \lambda_0(\bar{B}(a, \varepsilon))$, namely

$$h(p) = \inf\{\varepsilon > 0 : \lambda_0(\bar{B}(a, \varepsilon)) \geq p\}, \; p \geq 0.$$

From Lemma 1.37 in [**73**] the following property holds:

$$\forall p \geq 0, \quad \lambda_0(\bar{B}(a, h(p))) \geq p. \tag{3.119}$$

At this point, we introduce the following family of processes, parametrized by ε:

$$\begin{aligned} \nu_t^\varepsilon(b) = {} & \frac{1}{\varepsilon} \frac{1}{\lambda_0(\bar{B}(a, h(\varepsilon)))} 1_{\{b \in \bar{B}(a, h(\varepsilon))\}} 1_{\{t \leqslant T_1\}} + \nu_t^0(b) \, 1_{\{T_1 < t \leqslant T_2\}} \\ & + \sum_{n=2}^\infty \nu_t^{n-1}(T_2, A_2, ..., T_n, A_n, b) \, 1_{\{T_n < t \leqslant T_{n+1}\}}. \end{aligned} \tag{3.120}$$

With this choice, for all $r > 0$,

$$\begin{aligned} & \mathbb{P}_{\nu^\varepsilon}^{x,a'}(T_1 > r, E_1 \in F, A_1 \in C) \\ & = \int_r^\infty \int_F \exp\left(-\int_0^s \lambda(\phi(t, x, a'), a') \, dt - \frac{s}{\varepsilon}\right) \lambda(\phi(s, x, a'), a') \, Q(\phi(s, x, a'), a', dy) \, ds \\ & + \int_r^\infty \int_C \exp\left(-\int_0^s \lambda(\phi(t, x, a'), a') \, dt - \frac{s}{\varepsilon}\right) \frac{1}{\varepsilon} \frac{1}{\lambda_0(\bar{B}(a, h(\varepsilon)))} 1_{\{b \in \bar{B}(a, h(\varepsilon))\}} \, \lambda_0(db) \, ds. \end{aligned} \tag{3.121}$$

To prove (3.117), it is enough to show that, for every $k > 1$,

$$\mathbb{E}_{\nu^\varepsilon}^{x,a'}[\bar{F}(\Pi^{1,k})] \underset{\varepsilon \to 0}{\longrightarrow} \mathbb{E}_\nu^{x,a}[\bar{F}(\Pi^{1,k})], \tag{3.122}$$

where

$$\begin{aligned} \bar{F}(S_1, W_1, V_1, ..., S_k, W_k, V_k) = {} & \int_0^{S_1} e^{-\delta t} f(\phi(t, X_0, I_0), I_0) \, dt \\ & + \sum_{n=2}^k \int_{S_{n-1}}^{S_n} e^{-\delta t} f(\phi(t - S_{n-1}, W_{n-1}, V_{n-1}), V_{n-1}) \, dt, \end{aligned} \tag{3.123}$$

for any sequence of random variables $(S_n, W_n, V_n)_{n \in [1,k]}$ with values in $([0, \infty) \times E \times A)^n$, with $S_{n-1} \leq S_n$ for every n.

As a matter of fact, the remaining term

$$R(\varepsilon, k) := \mathbb{E}_{\nu^\varepsilon}^{x,a'}\left[\int_{T_k}^\infty e^{-\delta t} f(\phi(t - T_{n-1}, E_{n-1}, A_{n-1}), A_{n-1}) \, dt\right]$$

converges to zero, uniformly in ε, as k goes to infinity. To see it, we notice that

$$|R(\varepsilon,k)| \leq \frac{M_f}{\delta}\,\mathbb{E}^{x,a'}_{\nu^\varepsilon}\left[e^{-\delta T_k}\right] = \frac{M_f}{\delta}\,\mathbb{E}^{x,a'}\left[L^{\nu^\varepsilon}_{T_k}\,e^{-\delta T_k}\right], \tag{3.124}$$

where, L^ν is the Doléans-Dade exponential local martingale defined in (3.39). Taking into account (3.120) and (3.119), we get

$$\mathbb{E}^{x,a'}\left[L^{\nu^\varepsilon}_{T_k}\,e^{-\delta T_k}\right] \leq \mathbb{E}^{x,a'}\left[\frac{e^{T_1\,\lambda_0(A)}\,e^{-T_1\,\frac{1}{\varepsilon}}}{\varepsilon^2}\,L^{\bar\nu}_{T_k}\,e^{-\delta T_k}\right] \leq \frac{4}{e^2}\,\mathbb{E}^{x,a'}\left[\frac{e^{T_1\,\lambda_0(A)}}{T_1^2}\,L^{\bar\nu}_{T_k}\,e^{-\delta T_k}\right]$$

where

$$\bar\nu(t,b) := 1_{\{t\leqslant T_1\}} + \nu^0_t(b)\,1_{\{T_1<t\leqslant T_2\}} + \sum_{n=2}^\infty \nu^{n-1}_t(T_2,A_2,...,T_n,A_n,b)\,1_{\{T_n<t\leqslant T_{n+1}\}}.$$

Since $\bar\nu \in \mathcal{V}$, by Proposition 3.3.2 there exists a unique probability $\mathbb{P}^{x,a'}_{\bar\nu}$ on $(\Omega,\mathcal{F}_\infty)$ such that its restriction on $(\Omega,\mathcal{F}_{T_k})$ is $L^{\bar\nu}_{T_k}\,\mathbb{P}^{x,a'}$. Then (3.124) reads

$$|R(\varepsilon,k)| \leq \frac{4\,M_f}{\delta\,e^2}\,\mathbb{E}^{x,a'}_{\bar\nu}\left[\frac{e^{T_1\,\lambda_0(A)}}{T_1^2}\,e^{-\delta T_k}\right], \tag{3.125}$$

and the conclusion follows by the Lebesgue dominated convergence theorem.

Let us now prove (3.122). By Lemma 3.5.10, taking into account (3.121), we achieve

$$\begin{aligned}
&\mathbb{E}^{x,a'}_{\nu^\varepsilon}[\bar F(\Pi^{1,k})]\\
&= \mathbb{E}^{x,a'}_{\nu^\varepsilon}\left[\mathbb{E}^{x,a'}_{\nu^\varepsilon}\left[\bar F(\Pi^{1,k})\Big|\,\mathcal{F}_{T_1}\right]\right]\\
&= \mathbb{E}^{x,a'}_{\nu^\varepsilon}\left[\frac{\mathbb{E}^{x,a}_{\nu}\left[1_{\{T_1>\tau\}}\,\bar F(s,y,b,\Pi^{1,k-1})\right]}{\mathbb{P}^{x,a}_{\nu}(T_1>\tau)}\bigg|_{s=T_1,\,y=X_1,\,b=A_1}\right]\\
&= \int_0^\infty\int_E \frac{\mathbb{E}^{x,a}_{\nu}\left[1_{\{T_1>s\}}\,\bar F(s,y,a',\Pi^{1,k-1})\right]}{\mathbb{P}^{x,a}_{\nu}(T_1>s)}\cdot\\
&\cdot\exp\left(-\int_0^s \lambda(\phi(t,x,a'),a')\,dt - \frac{s}{\varepsilon}\right)\,\lambda(\phi(s,x,a'),a')\,Q(\phi(s,x,a'),a',dy)\,ds\\
&+ \int_0^\infty\int_A \frac{\mathbb{E}^{x,a}_{\nu}\left[1_{\{T_1>s\}}\,\bar F(s,\phi(s,x,a'),b,\Pi^{1,k-1})\right]}{\mathbb{P}^{x,a}_{\nu}(T_1>s)}\cdot\\
&\cdot\exp\left(-\int_0^s \lambda(\phi(t,x,a'),a')\,dt - \frac{s}{\varepsilon}\right)\,\frac{1}{\varepsilon}\,\frac{1}{\lambda_0(\bar B(a,h(\varepsilon)))}1_{\{b\in\bar B(a,h(\varepsilon))\}}\,\lambda_0(db)\,ds.
\end{aligned} \tag{3.126}$$

At this point, we set

$$\varphi(s,y,b) := \frac{\mathbb{E}^{x,a}_{\nu}\left[1_{\{T_1>s\}}\,\bar F(s,y,b,\Pi^{1,k-1})\right]}{\mathbb{P}^{x,a}_{\nu}(T_1>s)}, \quad s\in[0,\infty),\ y\in E,\ b\in A. \tag{3.127}$$

Notice that, for every $(y,b)\in E\times A$,

$$\bar F(s,y,b,\Pi^{1,k-1}) = \int_0^s e^{-\delta t}f(\phi(t,X_0,I_0),I_0)\,dt + \int_s^{T_1} e^{-\delta t}f(\phi(t-s,y,b),b)\,dt$$

$$+\sum_{n=2}^{k-1}\int_{T_{n-1}}^{T_n} e^{-\delta t} f(\phi(t-T_{n-1},E_{n-1},A_{n-1}),A_{n-1})\,dt,$$

so that

$$|\varphi(s,y,b)| \leq \frac{M_f}{\delta}. \tag{3.128}$$

Identity (3.126) becomes

$$\begin{aligned}
&\mathbb{E}_{\nu^\varepsilon}^{x,a'}[\bar F(\Pi^{1,k})]\\
&= \int_0^\infty \int_E \varphi(s,y,a')\,\exp\left(-\int_0^s \lambda(\phi(t,x,a'),a')\,dt - \frac{s}{\varepsilon}\right)\cdot\\
&\cdot\,\lambda(\phi(s,x,a'),a')\,Q(\phi(s,x,a'),a',dy)\,ds\\
&+ \int_0^\infty \int_A \varphi(s,\phi(s,x,a'),b)\,\exp\left(-\int_0^s \lambda(\phi(t,x,a'),a')\,dt - \frac{s}{\varepsilon}\right)\cdot\\
&\cdot\,\frac{1}{\varepsilon}\,\frac{1}{\lambda_0(\bar B(a,h(\varepsilon)))} 1_{\{b\in \bar B(a,h(\varepsilon))\}}\,\lambda_0(db)\,ds\\
&=: I_1(\varepsilon) + I_2(\varepsilon).
\end{aligned}$$

Using the change of variable $s = \varepsilon\, z$, we have

$$\begin{aligned}
I_1(\varepsilon) &= \int_0^\infty \int_E f_\varepsilon(z,y)\,\lambda(\phi(\varepsilon\, z,x,a'),a')\,Q(\phi(\varepsilon\, z,x,a'),a',dy)\; dz,\\
I_2(\varepsilon) &= \int_0^\infty \int_A g_\varepsilon(z,b)\,\lambda_0(db)\,dz,
\end{aligned}$$

where

$$\begin{aligned}
f_\varepsilon(z,y) &:= \varepsilon\,\varphi(\varepsilon\, z,y,a')\,\exp\left(-\int_0^{\varepsilon\, z} \lambda(\phi(t,x,a'),a')\,dt - z\right),\\
g_\varepsilon(z,b) &:= \varphi(\varepsilon\, z,\phi(\varepsilon\, z,x,a'),b)\,\exp\left(-\int_0^{\varepsilon\, z} \lambda(\phi(t,x,a'),a')\,dt - z\right)\cdot\\
&\quad\cdot\,\frac{1}{\lambda_0(\bar B(a,h(\varepsilon)))}\,1_{\{b\in \bar B(a,h(\varepsilon))\}}.
\end{aligned}$$

Exploiting the continuity properties of λ, Q, ϕ and f, we get

$$I_2(\varepsilon) \underset{\varepsilon\to 0}{\longrightarrow} \varphi(0,x,a), \tag{3.129}$$

where we have used that $\phi(0,x,b) = x$ for every $b \in A$. On the other hand, from the estimate (3.128), it follows that

$$|f_\varepsilon(z,y)| \leq \frac{M_f}{\delta}\, e^{-z}\,\varepsilon.$$

Therefore

$$|I_1(\varepsilon)| \leq \frac{M_f}{\delta}\,\varepsilon\,||\lambda||_\infty \int_0^\infty e^{-z}\,dz = \frac{M_f}{\delta}\,\varepsilon\,||\lambda||_\infty \underset{\varepsilon\to 0}{\longrightarrow} 0. \tag{3.130}$$

Collecting (3.130) and (3.129), we conclude that

$$\mathbb{E}_{\nu^\varepsilon}^{x,a'}[\bar F(\Pi^{1,k})] \underset{\varepsilon\to 0}{\longrightarrow}\; \varphi(0,x,a). \tag{3.131}$$

Recalling the definitions of φ and $\bar{F}$ given respectively in (3.127) and (3.123), we see that

$$
\begin{aligned}
&\varphi(0,x,a)\\
&= (\mathbb{P}_\nu^{x,a}(T_1 > 0))^{-1}\,\mathbb{E}_\nu^{x,a}\left[1_{\{T_1>0\}}\,\bar{F}(0,x,a,\Pi^{1,k-1})\right]\\
&= \mathbb{E}_\nu^{x,a}\left[\bar{F}(0,x,a,\Pi^{1,k-1})\right]\\
&= \mathbb{E}_\nu^{x,a}\left[\int_0^{T_1} e^{-\delta t} f(\phi(t,x,a),a)\,dt + \sum_{n=2}^{k}\int_{T_{n-2}}^{T_{n-1}} e^{-\delta t} f(\phi(t-T_{n-1},E_{n-1},A_{n-1}),A_{n-1})\,dt\right]\\
&= \mathbb{E}_\nu^{x,a}\left[\bar{F}(\Pi^{1,k})\right],
\end{aligned}
$$

and this concludes the proof. $\square$

Chapter 4

BSDEs driven by a general random measure, possibly non quasi-left-continuous

4.1. Introduction

Backward stochastic differential equations have been deeply studied since the seminal paper Pardoux and Peng [**98**]. In [**98**], as well as in many subsequent papers, the driving term was a Brownian motion. BSDEs with a discontinuous driving term have also been studied, see, among others, Buchdahn and Pardoux [**21**], Tang and Li [**128**], Bares, Buckdahn and Pardoux [**10**], El Karoui and Huang [**50**], Xia [**131**], Becherer [**12**], Carbone, Ferrario and Santacroce [**22**], Cohen and Elliott [**26**], Jeanblanc, Mania, Santacroce and Schweizer [**80**], Confortola, Fuhrman and Jacod [**29**].

In all the papers cited above, and more generally in the literature on BSDEs, the generator (or driver) of the backward stochastic differential equation, usually denoted by f, is integrated with respect to a measure dA, where A is a nondecreasing continuous (or deterministic and right-continuous as in [**26**]) process. The general case, i.e. A is a right-continuous nondecreasing predictable process, is addressed in this chapter. It is worth mentioning that Section 4.3 in [**29**] provides a *counter-example* to existence for such general backward stochastic differential equations. For this reason, the existence and uniqueness result (Theorem 4.4.1) is not a trivial extension of known results. Indeed, in Theorem 4.4.1 we have to impose an additional technical assumption, which is violated by the counter-example presented in [**29**] (see Remark 4.4.3(ii)). This latter assumption reads as follows: there exists $\varepsilon \in (0,1)$ such that (notice that $\Delta A_t \leq 1$)

$$2\, L_y^2\, |\Delta A_t|^2 \;\leq\; 1-\varepsilon, \qquad \mathbb{P}\text{-a.s.}, \forall\, t \in [0,T], \tag{4.1}$$

where L_y is the Lipschitz constant of f with respect to y. As mentioned earlier, in [**26**] the authors study a class of BSDEs with a generator f integrated with respect to a *deterministic* (rather than predictable) right-continuous nondecreasing process A, even if this class is driven by a countable sequence of square-integrable martingales, rather than just a random measure. They provide an existence and uniqueness result for this class of BSDEs, see Theorem 6.1 in [**26**], where the same condition (4.1) is imposed (see Remark 4.4.3(i)). However, the proof of Theorem 6.1 in [**26**] relies heavily on the assumption that A is deterministic, and it can not be extended to the case where A is predictable, which therefore requires a completely different proof.

The results obtained in this chapter can be particularly useful in the study of control problems related to piecewise deterministic Markov processes by means of BSDEs methods, see Remark 4.4.5.

The chapter is organized as follows: in Section 4.2 we introduce the random measure μ and we fix the notation. In Section 4.3 we provide the definition of solution to the backward stochastic differential equation and we solve it in the case where $f = f(t, \omega)$ is independent of y and z (Lemma 4.3.6). Finally, in Section 4.4 we prove the main result (Theorem 4.4.1) of this chapter, i.e. the existence and uniqueness for our backward stochastic differential equation.

4.2. Preliminaries

Consider a finite time horizon $T \in (0, \infty)$, a Lusin space $(E, \mathcal{E})$, and a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, with $(\mathcal{F}_t)_{t \geq 0}$ right-continuous. We denote by $\mathcal{P}$ the predictable σ-field on $\Omega \times [0, T]$. In the sequel, given a measurable space $(G, \mathcal{G})$, we say that a function on the product space $\Omega \times [0, T] \times G$ is predictable if it is $\mathcal{P} \otimes \mathcal{G}$-measurable.

Let μ be an integer-valued random measure on $\mathbb{R}_+ \times E$. In the sequel we use a martingale representation theorem for the random measure μ, namely Theorem 5.4 in Jacod [**75**]. For this reason, we suppose that $(\mathcal{F}_t)_{t \geq 0}$ is the natural filtration of μ, i.e. the smallest right-continuous filtration in which μ is optional. We also assume that μ is a discrete random measure, i.e. the sections of the set $D = \{(\omega, t) \colon \mu(\omega, \{t\} \times E) = 1\}$ are finite on every finite interval. However, the results of this chapter (in particular, Theorem 4.4.1) are still valid for more general random measure μ for which a martingale representation theorem holds (see Remark 4.4.4 for more details).

We denote by ν the $(\mathcal{F}_t)_{t \geq 0}$-compensator of μ. Then, ν can be disintegrated as follows

$$\nu(\omega, dt\, dx) \;=\; dA_t(\omega)\, \phi_{\omega,t}(dx), \tag{4.2}$$

where A is a right-continuous nondecreasing predictable process such that $A_0 = 0$, and ϕ is a transition probability from $(\Omega \times [0, T], \mathcal{P})$ into $(E, \mathcal{E})$. We suppose, without loss of generality, that ν satisfies $\nu(\{t\} \times dx) \leq 1$ identically, so that $\Delta A_t \leq 1$. We define A^c as $A^c_t = A_t - \sum_{0 < s \leq t} \Delta A_s$, $\nu^c(dt\, dx) = 1_{J^c \times E}\, \nu(dt, dx)$, $\nu^d(dt\, dx) = \nu(dt\, dx) - \nu^c(dt\, dx) = 1_{J \times E}\, \nu(dt, dx)$, where $J = \{(\omega, t) \colon \nu(\omega, \{t\} \times dx) > 0\}$.

We denote by $\mathcal{B}(E)$ the set of all Borel measurable functions on E. Given a measurable function $Z \colon \Omega \times [0, T] \times E \to \mathbb{R}$, we write $Z_{\omega,t}(x) = Z(\omega, t, x)$, so that

$Z_{\omega,t}$, often abbreviated as Z_t or $Z_t(\cdot)$, is an element of $\mathcal{B}(E)$. For any $\beta \geq 0$ we also denote by $\mathcal{E}^\beta$ the Doléans-Dade exponential of the process βA, which is given by

$$\mathcal{E}_t^\beta = e^{\beta A_t} \prod_{0<s\leq t} (1 + \beta\, \Delta A_s)\, e^{-\beta\, \Delta A_s}. \tag{4.3}$$

4.3. The backward stochastic differential equation

The backward stochastic differential equation driven by the random measure μ is characterized by a triple (β, ξ, f), where $\beta > 0$ is a positive real number, and:

- $\xi\colon \Omega \to \mathbb{R}$, the *terminal condition*, is an $\mathcal{F}_T$-measurable random variable satisfying $\mathbb{E}[\mathcal{E}_T^\beta |\xi|^2] < \infty$;
- $f\colon \Omega \times [0,T] \times \mathbb{R} \times \mathcal{B}(E) \to \mathbb{R}$, the *generator*, is such that:
  (i) for any $y \in \mathbb{R}$ and $Z\colon \Omega \times [0,T] \times E \to \mathbb{R}$ predictable $\Longrightarrow f(\omega, t, y, Z_{\omega,t}(\cdot))$ predictable;
  (ii) for some nonnegative constants L_y, L_z, we have

$$\begin{aligned} &|f(\omega,t,y',\zeta') - f(\omega,t,y,\zeta)| \leq L_y|y'-y| \\ &+ L_z\bigg(\int_E \bigg|\zeta'(x) - \zeta(x) - \Delta A_t(\omega) \int_E \big(\zeta'(z) - \zeta(z)\big)\phi_{\omega,t}(dz)\bigg|^2 \phi_{\omega,t}(dx) \\ &+ \Delta A_t(\omega)\big(1 - \Delta A_t(\omega)\big)\bigg| \int_E (\zeta'(x) - \zeta(x))\, \phi_{\omega,t}(dx)\bigg|^2\bigg)^{1/2}, \end{aligned} \tag{4.4}$$

  for all $(\omega, t) \in \Omega \times [0,T]$, $y, y' \in \mathbb{R}$, $\zeta, \zeta' \in L^2(E, \mathcal{E}, \phi_{\omega,t}(dx))$;
  (iii) $\mathbb{E}[(1 + \sum_{0<t\leq T} |\Delta A_t|^2) \int_0^T \mathcal{E}_t^\beta |f(t,0,0)|^2\, dA_t] < \infty$.

Remark 4.3.1. The measurability condition (i) on f is somehow awkward, however it seems to be unavoidable. Indeed, we notice that the same condition is imposed in [**29**], assumption (2.8), and a similar condition is imposed in [**27**], assumption (3.2). We also observe that at page 4 of [**29**], the authors provide some examples of assumptions on f which imply the measurability condition (i) above (see in particular assumption (2.10) in [**29**]). □

Given (β, ξ, f), the backward stochastic differential equation takes the following form

$$Y_t = \xi + \int_{(t,T]} f(s, Y_{s-}, Z_s(\cdot))\, dA_s - \int_{(t,T]} \int_E Z_s(x)\, (\mu - \nu)(ds\, dx), \quad 0 \leq t \leq T. \tag{4.5}$$

Definition 4.3.2. For every $\beta \geq 0$, we define $\mathbb{H}_\beta^2(0,T)$ as the set of pairs (Y, Z) such that:

- $Y\colon \Omega \times [0,T] \to \mathbb{R}$ is an adapted càdlàg process satisfying

$$\|Y\|^2_{\mathbb{H}^2_{\beta,Y}(0,T)} := \mathbb{E}\bigg[\int_{(0,T]} \mathcal{E}_t^\beta |Y_{t-}|^2\, dA_t\bigg] < \infty; \tag{4.6}$$

- $Z\colon \Omega \times [0,T] \times E \to \mathbb{R}$ is a predictable process satisfying

$$\|Z\|^2_{\mathbb{H}^2_{\beta,Z}(0,T)} := \mathbb{E}\bigg[\int_{(0,T]} \mathcal{E}_t^\beta \int_E \big|Z_t(x) - \hat{Z}_t\big|^2\, \nu(dt\, dx)$$

$$+ \sum_{0<t\leq T} \mathcal{E}_t^\beta \big|\hat{Z}_t\big|^2 \big(1-\Delta A_t\big)\Bigg] \;<\; \infty, \tag{4.7}$$

where

$$\hat{Z}_t \;=\; \int_E Z_t(x)\,\nu(\{t\}\times dx), \qquad 0\leq t\leq T.$$

For every $(Y,Z)\in\mathbb{H}^2_\beta(0,T)$, we denote

$$\|(Y,Z)\|^2_{\mathbb{H}^2_\beta(0,T)} := \|Y\|^2_{\mathbb{H}^2_{\beta,Y}(0,T)} + \|Z\|^2_{\mathbb{H}^2_{\beta,Z}(0,T)}.$$

Remark 4.3.3. (i) Notice that the space $\mathbb{H}^2_\beta(0,T)$, endowed with the topology induced by $\|\cdot\|_{\mathbb{H}^2_\beta(0,T)}$, is an Hilbert space, provided we identify pairs of processes $(Y,Z),(Y',Z')$ satisfying $\|(Y-Y',Z-Z')\|_{\mathbb{H}^2_\beta(0,T)}=0$.

(ii) Suppose that there exists $\gamma\in(0,1]$ such that $\Delta A_t\leq 1-\gamma$, for all $t\in[0,T]$, $\mathbb{P}$-a.s.. Then Z belongs to $\mathbb{H}^2_{\beta,Z}(0,T)$ if and only if $\sqrt{\mathcal{E}^\beta}Z$ is in $L^2(\Omega\times[0,T]\times E,\mathcal{P}\otimes\mathcal{E},\mathbb{P}\otimes\nu(dt\,dx))$, i.e.

$$\mathbb{E}\bigg[\int_{(0,T]} \mathcal{E}_t^\beta \int_E \big|Z_t(x)\big|^2\,\nu(dt\,dx)\bigg] \;<\; \infty.$$

□

Definition 4.3.4. A solution to equation (4.5) with data (β,ξ,f) is a pair $(Y,Z)\in\mathbb{H}^2_\beta(0,T)$ satisfying equation (4.5). We say that equation (4.5) admits a unique solution if, given two solutions $(Y,Z),(Y',Z')\in\mathbb{H}^2_\beta(0,T)$, we have $(Y,Z)=(Y',Z')$ in $\mathbb{H}^2_\beta(0,T)$.

Remark 4.3.5. Notice that, given a solution (Y,Z) to equation (4.5) with data (β,ξ,f), we have (recalling that $\beta\geq 0$, so that $\mathcal{E}_t^\beta\geq 1$)

$$\begin{aligned}&\mathbb{E}\bigg[\int_{(0,T]}\int_E \big|Z_t(x)-\hat{Z}_t\big|^2\,\nu(dt\,dx) + \sum_{0<t\leq T}\big|\hat{Z}_t\big|^2\big(1-\Delta A_t\big)\bigg]\\ &= \|Z\|^2_{\mathbb{H}^2_{0,Z}(0,T)} \leq \|Z\|^2_{\mathbb{H}^2_{\beta,Z}(0,T)} < \infty.\end{aligned}$$

This implies that the process $(Z_t 1_{[0,T]}(t))_{t\geq 0}$ belongs to $\mathcal{G}^2(\mu)$, see (3.62) and Proposition 3.71-(a) in Jacod's book [**77**]. In particular, the stochastic integral $\int_{(t,T]}\int_E Z_s(x)\,(\mu-\nu)(ds\,dx)$ in (4.5) is well-defined, and the process

$$M_t := \int_{(0,t]}\int_E Z_s(x)(\mu-\nu)(ds\,dx), \quad t\in[0,T],$$

is a square integrable martingale (see Proposition 3.66 in [**77**]). □

Lemma 4.3.6. *Consider a triple (β,ξ,f) and suppose that $f=f(\omega,t)$ does not depend on (y,ζ). Then, there exists a unique solution $(Y,Z)\in\mathbb{H}^2_\beta(0,T)$ to equation (4.5) with data (β,ξ,f). Moreover, the following identity holds:*

$$\mathbb{E}\big[\mathcal{E}_t^\beta\,|Y_t|^2\big] + \beta\,\mathbb{E}\bigg[\int_{(t,T]}\mathcal{E}_s^\beta\,(1+\beta\Delta A_s)^{-1}\,|Y_{s-}|^2\,dA_s\bigg]$$

$$+\,\mathbb{E}\Big[\int_{(t,T]} \mathcal{E}_s^\beta \int_E \big|Z_s(x)-\hat{Z}_s\big|^2\,\nu(ds\,dx) + \sum_{t<s\le T} \mathcal{E}_s^\beta \big|\hat{Z}_s\big|^2\big(1-\Delta A_s\big)\Big]$$
$$=\ \mathbb{E}\big[\mathcal{E}_T^\beta\,|\xi|^2\big] + 2\,\mathbb{E}\Big[\int_{(t,T]} \mathcal{E}_s^\beta\,Y_{s-}\,f_s\,dA_s\Big] - \mathbb{E}\Big[\sum_{t<s\le T} \mathcal{E}_s^\beta\,|f_s|^2\,|\Delta A_s|^2\Big], \quad (4.8)$$

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

Proof. *Uniqueness.* It is enough to prove that equation (4.5) with data $(\beta,0,0)$ has the unique (in the sense of Definition 4.3.4) solution $(Y,Z)=(0,0)$. Let (Y,Z) be a solution to equation (4.5) with data $(\beta,0,0)$. Since the stochastic integral in (4.5) is a square integrable martingale (see Remark 4.3.5), taking the conditional expectation with respect to $\mathcal{F}_t$ we obtain, $\mathbb{P}$-a.s., $Y_t=0$, for all $t\in[0,T]$. This proves the claim for the component Y and shows that the martingale $M_t := \int_{(0,t]}\int_E Z_s(x)(\mu-\nu)(ds\,dx) = 0$, $\mathbb{P}$-a.s., for all $t\in[0,T]$. Therefore, the predictable bracket $\langle M,M\rangle_T=0$, $\mathbb{P}$-a.s., where we recall that (see Proposition 3.71-(a) in [**77**])

$$\langle M,M\rangle_T\ =\ \int_{(0,T]}\int_E \big|Z_t(x)-\hat{Z}_t\big|^2\,\nu(dt\,dx) + \sum_{0<t\le T}\big|\hat{Z}_t\big|^2\big(1-\Delta A_t\big).$$

This concludes the proof, since $\|Z\|^2_{\mathbb{H}^2_{\beta,Z}(0,T)} \le \mathbb{E}\big[\mathcal{E}_T^\beta\langle M,M\rangle_T\big]=0$.

Identity (4.8). Let (Y,Z) be a solution to equation (4.5) with data (β,ξ,f). From Itô's formula applied to $\mathcal{E}_s^\beta\,|Y_s|^2$ it follows that (recall that $d\mathcal{E}_s^\beta = \beta\,\mathcal{E}_{s-}^\beta\,dA_s$)

$$\begin{aligned}
d(\mathcal{E}_s^\beta\,|Y_s|^2) &= \mathcal{E}_{s-}^\beta\,d|Y_s|^2 + |Y_{s-}|^2\,d\mathcal{E}_s^\beta + \Delta\mathcal{E}_s^\beta\,\Delta|Y_s|^2\\
&= \mathcal{E}_{s-}^\beta\,d|Y_s|^2 + |Y_{s-}|^2\,d\mathcal{E}_s^\beta + (\mathcal{E}_s - \mathcal{E}_{s-}^\beta)\,d|Y_s|^2\\
&= \mathcal{E}_s^\beta\,d|Y_s|^2 + |Y_{s-}|^2\,d\mathcal{E}_s^\beta\\
&= 2\,\mathcal{E}_s^\beta\,Y_{s-}\,dY_s + \mathcal{E}_s^\beta\,(\Delta Y_s)^2 + \beta\,\mathcal{E}_{s-}^\beta\,|Y_{s-}|^2\,dA_s\\
&= 2\,\mathcal{E}_s^\beta\,Y_{s-}\,dY_s + \mathcal{E}_s^\beta\,(\Delta Y_s)^2 + \beta\,\mathcal{E}_s^\beta\,(1+\beta\Delta A_s)^{-1}\,|Y_{s-}|^2\,dA_s, \qquad (4.9)
\end{aligned}$$

where the last equality follows from the identity $\mathcal{E}_{s-}^\beta = \mathcal{E}_s^\beta(1+\beta\Delta A_s)^{-1}$. Integrating (4.9) on the interval $[t,\,T]$, we obtain

$$\begin{aligned}
\mathcal{E}_t^\beta\,|Y_t|^2\ &=\ \mathcal{E}_T^\beta\,|\xi|^2 + 2\int_{(t,T]}\mathcal{E}_s^\beta\,Y_{s-}\,f_s\,dA_s - 2\int_{(t,T]}\mathcal{E}_s^\beta\,Y_{s-}\int_E Z_s(x)\,(\mu-\nu)(ds\,dx)\\
&\quad - \sum_{t<s\le T}\mathcal{E}_s^\beta\,(\Delta Y_s)^2 - \beta\int_{(t,T]}\mathcal{E}_s^\beta\,(1+\beta\Delta A_s)^{-1}\,|Y_{s-}|^2\,dA_s. \qquad (4.10)
\end{aligned}$$

Now, notice that

$$\Delta Y_s = \int_E Z_s(x)\,(\mu-\nu)(\{s\}\times dx) - f_s\,\Delta A_s. \qquad (4.11)$$

Thus

$$|\Delta Y_s|^2 = \bigg|\int_E Z_s(x)\,(\mu-\nu)(\{s\}\times dx)\bigg|^2 + |f_s|^2|\Delta A_s|^2$$

$$- 2 f_s \Delta A_s \int_E Z_s(x)\,(\mu-\nu)(\{s\}\times dx). \tag{4.12}$$

Plugging (4.12) into (4.10), we find

$$\begin{aligned}
&\mathcal{E}_t^\beta\,|Y_t|^2 + \beta\int_{(t,T]} \mathcal{E}_s^\beta\,(1+\beta\Delta A_s)^{-1}\,|Y_{s-}|^2\,dA_s \\
&+ \sum_{t<s\le T} \mathcal{E}_s^\beta \left|\int_E Z_s(x)\,(\mu-\nu)(\{s\}\times dx)\right|^2 \\
&= \mathcal{E}_T^\beta\,|\xi|^2 + 2\int_{(t,T]} \mathcal{E}_s^\beta\,Y_{s-}\,f_s\,dA_s - 2\int_{(t,T]} \mathcal{E}_s^\beta\,Y_{s-}\int_E Z_s(x)\,(\mu-\nu)(ds\,dx) \\
&- \sum_{t<s\le T} \mathcal{E}_s^\beta\,|f_s|^2\,|\Delta A_s|^2 + 2\sum_{t<s\le T} \mathcal{E}_s^\beta\,f_s\,\Delta A_s \int_E Z_s(x)\,(\mu-\nu)(\{s\}\times dx).
\end{aligned} \tag{4.13}$$

Notice that

$$\begin{aligned}
&\mathbb{E}\Big[\sum_{t<s\le T} \mathcal{E}_s^\beta \left|\int_E Z_s(x)\,(\mu-\nu)(\{s\}\times dx)\right|^2\Big] \\
&= \mathbb{E}\Big[\int_{(t,T]} \mathcal{E}_s^\beta \int_E \big|Z_s(x) - \hat{Z}_s\big|^2\,\nu(ds,dx) + \sum_{t<s\le T} \mathcal{E}_s^\beta \big|\hat{Z}_s\big|^2\big(1-\Delta A_s\big)\Big].
\end{aligned} \tag{4.14}$$

We also observe that the two stochastic integrals

$$\begin{aligned}
M_t^1 &:= \int_{(0,t]} \mathcal{E}_s^\beta\,Y_{s-}\int_E Z_s(x)\,(\mu-\nu)(ds\,dx) \\
M_t^2 &:= \sum_{0<s\le t} \mathcal{E}_s^\beta\,f_s\,\Delta A_s \int_E Z_s(x)\,(\mu-\nu)(\{s\}\times dx)
\end{aligned}$$

are martingales. Therefore, taking the expectation in (4.13) and using (4.14), we end up with (4.8).

Existence. Consider the martingale $\tilde{M}_t := \mathbb{E}[\xi + \int_{(0,T]} f_s\,dA_s|\mathcal{F}_t]$, $t\in[0,T]$. Let M be a right-continuous modification of $\tilde{M}$. Then, by the martingale representation Theorem 5.4 in [**75**] and Proposition 3.66 in [**77**] (noting that M is a square integrable martingale), there exists a predictable process $Z\colon \Omega\times[0,T]\times E\to\mathbb{R}$ such that

$$\mathbb{E}\Big[\int_{(0,T]}\int_E \big|Z_t(x)-\hat{Z}_t\big|^2\,\nu(dt\,dx) + \sum_{0<t\le T}\big|\hat{Z}_t\big|^2\big(1-\Delta A_t\big)\Big] < \infty$$

and

$$M_t = M_0 + \int_{(0,t]}\int_E Z_s(x)\,(\mu-\nu)(ds\,dx), \qquad t\in[0,\,T]. \tag{4.15}$$

Set

$$Y_t = M_t - \int_{(0,t]} f_s\,dA_s, \qquad t\in[0,T]. \tag{4.16}$$

Using the representation (4.15) of M, and noting that $Y_T=\xi$, we see that Y satisfies (4.5). When $\beta>0$, it remains to show that Y satisfies (4.6) and Z satisfies (4.7).

To this end, let us define the increasing sequence of stopping times

$$S_k = \inf\Big\{t \in (0,\,T] : \int_{(0,t]} \mathcal{E}^\beta_s\,|Y_{s-}|^2\,dA_s + \int_{(0,t]} \mathcal{E}^\beta_s \int_E \big|Z_s(x) - \hat{Z}_s\big|^2\,\nu(ds\,dx) + \sum_{0<s\le t} \mathcal{E}^\beta_s\big|\hat{Z}_s\big|^2\big(1-\Delta A_s\big) > k\Big\}$$

with the convention $\inf\emptyset = T$. Computing the Itô differential $d(\mathcal{E}^\beta_s\,|Y_s|^2)$ on the interval $[0,\,S_k]$ and proceeding as in the derivation of identity (4.8), we find

$$\begin{aligned}
&\mathbb{E}\left[\int_{(0,S_k]} \mathcal{E}^\beta_s \int_E \big|Z_s(x) - \hat{Z}_s\big|^2\,\nu(ds\,dx) + \sum_{0<s\le S_k} \mathcal{E}^\beta_s\big|\hat{Z}_s\big|^2\big(1-\Delta A_s\big)\right] \\
&+ \beta\,\mathbb{E}\left[\int_{(0,S_k]} \mathcal{E}^\beta_s\,(1+\beta\Delta A_s)^{-1}\,|Y_{s-}|^2\,dA_s\right] \\
&\le \mathbb{E}\left[\mathcal{E}^\beta_{S_k}|Y_{S_k}|^2\right] + 2\mathbb{E}\left[\int_{(0,S_k]} \mathcal{E}^\beta_s\,Y_{s-}\,f_s\,dA_s\right].
\end{aligned} \tag{4.17}$$

Let us now prove the following inequality (recall that we are assuming $\beta > 0$)

$$\mathcal{E}^\beta_t\Big(\int_{(t,T]} |f_s|\,dA_s\Big)^2 \;\le\; \Big(\frac{1}{\beta} + \beta\sum_{t<s\le T}|\Delta A_s|^2\Big)\int_{(t,T]} \mathcal{E}^\beta_s\,|f_s|^2\,dA_s. \tag{4.18}$$

Set, for all $s \in [0,\,T]$,

$$\begin{aligned}
\bar{A}_s &:= \frac{\beta}{2}A^c_s + \sum_{0<r\le s,\,\Delta A_r\neq 0}\big(\sqrt{1+\beta\Delta A_r} - 1\big), \\
\underline{A}_s &:= -\frac{\beta}{2}A^c_s - \sum_{0<r\le s,\,\Delta A_r\neq 0}\frac{\sqrt{1+\beta\Delta A_r} - 1}{\sqrt{1+\beta\Delta A_r}}.
\end{aligned}$$

Denote by $\bar{\mathcal{E}}$ (resp. $\underline{\mathcal{E}}$) the Doléans-Dade exponential of the process $\bar{A}$ (resp. $\underline{A}$). Using Proposition 6.4 in [**77**] we see that

$$1 \;=\; \underline{\mathcal{E}}_s\,\bar{\mathcal{E}}_s, \qquad (\bar{\mathcal{E}}_s)^2 \;=\; \mathcal{E}^\beta_s, \qquad \forall\, s \in [0,\,T]. \tag{4.19}$$

Then, we conclude that

$$\begin{aligned}
\mathcal{E}^\beta_t\Big(\int_{(t,T]} |f_s|\,dA_s\Big)^2 &= \mathcal{E}^\beta_t\Big(\int_{(t,T]} \underline{\mathcal{E}}_{s-}\,\bar{\mathcal{E}}_{s-}\,|f_s|\,dA_s\Big)^2 \\
&\le \Big(\frac{1}{\beta} + \beta\sum_{t<s\le T}|\Delta A_s|^2\Big)\int_{(t,T]} \mathcal{E}^\beta_s\,|f_s|^2\,dA_s,
\end{aligned}$$

where we used the inequality $\mathcal{E}^\beta_{s-} \le \mathcal{E}^\beta_s$ (which follows from (4.3)) and

$$\begin{aligned}
\mathcal{E}^\beta_t \int_{(t,T]} (\underline{\mathcal{E}}_{s-})^2\,dA_s &= \mathcal{E}^\beta_t\,\frac{(\underline{\mathcal{E}}_t)^2 - (\underline{\mathcal{E}}_T)^2}{\beta} + \mathcal{E}^\beta_t\,\beta\sum_{t<s\le T}(\underline{\mathcal{E}}_{s-})^2\,\frac{|\Delta A_s|^2}{1+\beta\,\Delta A_s} \\
&\le \frac{1}{\beta} + \beta\sum_{t<s\le T}|\Delta A_s|^2,
\end{aligned}$$

where the last inequality follows from $\frac{1}{1+\beta\Delta A_s} \leq 1$ and identities (4.19). Now, using (4.16) and (4.18) we obtain

$$\begin{aligned}
\mathcal{E}_t^\beta\,|Y_t|^2 &= \mathcal{E}_t^\beta \left|\mathbb{E}\Big[\xi + \int_{(t,T]} f_s\,dA_s\Big|\mathcal{F}_t\Big]\right|^2 \\
&\leq 2\,\mathbb{E}\big[\mathcal{E}_t^\beta\,|\xi|^2\big|\mathcal{F}_t\big] + 2\,\mathbb{E}\Big[\mathcal{E}_t^\beta\,\Big(\int_{(t,T]}|f_s|\,dA_s\Big)^2\Big|\mathcal{F}_t\Big] \\
&\leq 2\,\mathbb{E}\Big[\mathcal{E}_T^\beta\,|\xi|^2 + \Big(\frac{1}{\beta} + \beta\sum_{0<s\leq T}|\Delta A_s|^2\Big)\int_{(0,T]}\mathcal{E}_s^\beta\,|f_s|^2\,dA_s\Big|\mathcal{F}_t\Big]. \qquad (4.20)
\end{aligned}$$

Denote by m_t a right-continuous modification of the right-hand side of (4.20). We see that $m = (m_t)_{t\in[0,T]}$ is a uniformly integrable martingale. In particular for every stopping time S with values in $[0,\,T]$, we have, by Doob's optional stopping theorem,

$$\mathbb{E}\left[\mathcal{E}_S^\beta|Y_S|^2\right] \leq \mathbb{E}\,[m_S] \leq \mathbb{E}\,[m_T] < \infty. \qquad (4.21)$$

Notice that $(1+\beta\Delta A_s)^{-1} \geq \frac{1}{1+\beta}$ $\mathbb{P}$-a.s. Using the inequality $2ab \leq \gamma a^2 + \frac{1}{\gamma}b^2$ with $\gamma = \frac{\beta}{2(1+\beta)}$, and plugging (4.21) (with $S = S_k$) into (4.17), we find the estimate

$$\begin{aligned}
&\frac{\beta}{2(1+\beta)}\mathbb{E}\left[\int_{(0,S_k]}\mathcal{E}_s^\beta|Y_{s-}|^2\,dA_s\right] \\
&+\ \mathbb{E}\left[\int_{(0,S_k]}\mathcal{E}_s^\beta\int_E\big|Z_s(x)-\hat{Z}_s\big|^2\,\nu(ds\,dx) + \sum_{0<s\leq S_k}\mathcal{E}_s^\beta\big|\hat{Z}_s\big|^2\big(1-\Delta A_s\big)\right] \\
&\leq\ 2\,\mathbb{E}\big[\mathcal{E}_T^\beta\,|\xi|^2\big] + 2\,\mathbb{E}\Big[\Big(\frac{1}{\beta} + \beta\sum_{0<s\leq T}|\Delta A_s|^2\Big)\Big(\int_{(0,T]}\mathcal{E}_s^\beta\,|f_s|^2\,dA_s\Big)\Big].
\end{aligned}$$

From the above inequality we deduce that

$$\begin{aligned}
&\mathbb{E}\left[\int_{(0,S_k]}\mathcal{E}_s^\beta\,|Y_{s-}|^2\,dA_s\right] \\
&+\mathbb{E}\Big[\int_{(0,S_k]}\mathcal{E}_s^\beta\int_E\big|Z_s(x)-\hat{Z}_s\big|^2\,\nu(ds\,dx) + \sum_{0<s\leq S_k}\mathcal{E}_s^\beta\big|\hat{Z}_s\big|^2\big(1-\Delta A_s\big)\Big] \\
&\leq c(\beta)\,\left(\mathbb{E}\left[\mathcal{E}_T^\beta\,|\xi|^2\right] + \mathbb{E}\left[\Big(\frac{1}{\beta} + \beta\sum_{0<s\leq T}|\Delta A_s|^2\Big)\int_{(0,T]}\mathcal{E}_s^\beta\,|f_s|^2\,dA_s\right]\right), \qquad (4.22)
\end{aligned}$$

where $c(\beta) = 2 + \frac{4(1+\beta)}{\beta}$. Setting $S = \lim_k S_k$ we deduce

$$\begin{aligned}
&\mathbb{E}\left[\int_{(0,S]}\mathcal{E}_s^\beta\,|Y_{s-}|^2\,dA_s\right] \\
&+\mathbb{E}\Big[\int_{(0,S]}\mathcal{E}_s^\beta\int_E\big|Z_s(x)-\hat{Z}_s\big|^2\,\nu(ds\,dx) + \sum_{0<s\leq S}\mathcal{E}_s^\beta\big|\hat{Z}_s\big|^2\big(1-\Delta A_s\big)\Big] \\
&< \infty, \quad \mathbb{P}\text{-a.s.},
\end{aligned}$$

which implies $S = T$, $\mathbb{P}$-a.s., by the definition of S_k. Letting $k \to \infty$ in (4.22), we conclude that Y satisfies (4.6) and Z satisfies (4.7), so that $(Y, Z) \in \mathbb{H}^2_\beta(0, T)$. $\square$

4.4. Main result

Theorem 4.4.1. *Suppose that there exists $\varepsilon \in (0, 1)$ such that*

$$2\, L_y^2\, |\Delta A_t|^2 \;\leq\; 1 - \varepsilon, \qquad \mathbb{P}\text{-}a.s.,\ \forall\, t \in [0, T]. \tag{4.23}$$

Then there exists a unique solution $(Y, Z) \in \mathbb{H}^2_\beta(0, T)$ to equation (4.5) with data (β, ξ, f), for every β satisfying

$$\beta \;\geq\; \frac{\frac{L_y^2}{\hat{L}_{z,t}^2} + \frac{2\,\hat{L}_{z,t}^2}{1-\delta+2\,\hat{L}_{z,t}^2\,\Delta A_t}}{1 - \Delta A_t\Big(\frac{L_y^2}{\hat{L}_{z,t}^2} + \frac{2\,\hat{L}_{z,t}^2}{1-\delta+2\,\hat{L}_{z,t}^2\,\Delta A_t}\Big)}, \qquad \mathbb{P}\text{-}a.s.,\ \forall\, t \in [0, T], \tag{4.24}$$

for some $\delta \in (0, \varepsilon)$ and strictly positive predictable process $(\hat{L}_{z,t})_{t\in[0,T]}$ given by

$$\hat{L}_{z,t}^2 \;=\; \max\bigg(L_z^2 + \delta,\, \frac{(1-\delta)\, L_y}{\sqrt{2(1-\delta)} - 2\, L_y\, \Delta A_t}\bigg). \tag{4.25}$$

Remark 4.4.2. (i) Notice that when condition (4.23) holds the right-hand side of (4.24) is a well-defined nonnegative real number, so that there always exists some $\beta \geq 0$ which satisfies (4.24).

(ii) Observe that in Theorem 4.4.1 there is no condition on L_z, i.e. on the Lipschitz constant of f with respect to its last argument. $\square$

Proof of Theorem 4.4.1. The proof is based on a fixed point argument that we now describe. Let us consider the function $\Phi : \mathbb{H}^2_\beta(0, T) \to \mathbb{H}^2_\beta(0, T)$, mapping (U, V) to (Y, Z) as follows:

$$Y_t \;=\; \xi + \int_{(t,T]} f(t, U_{s-}, V_s)\, dA_s - \int_{(t,T]}\int_E Z_s(x)\,(\mu - \nu)(ds\, dx), \tag{4.26}$$

for all $t \in [0,\, T]$. By Lemma 4.3.6 there exists a unique $(Y, Z) \in \mathbb{H}^2_\beta(0, T)$ satisfying (4.26), so that Φ is a well-defined map. We then see that (Y, Z) is a solution in $\mathbb{H}^2_\beta(0, T)$ to the BSDE (4.5) with data (β, ξ, f) if and only if it is a fixed point of Φ.

Let us prove that Φ is a contraction when β is large enough. Let $(U^i, V^i) \in \mathbb{H}^2_\beta(0, T)$, $i = 1, 2$, and set $(Y^i, Z^i) = \Phi(U^i, V^i)$. Denote $\bar{Y} = Y^1 - Y^2$, $\bar{Z} = Z^1 - Z^2$, $\bar{U} = U^1 - U^2$, $\bar{V} = V^1 - V^2$, $\bar{f}_s = f(s, U^1_{s-}, V^1_s) - f(s, U^2_{s-}, V^2_s)$. Notice that

$$\bar{Y}_t \;=\; \int_{(t,T]} \bar{f}_s\, dA_s - \int_{(t,T]}\int_E \bar{Z}_s(x)\,(\mu - \nu)(ds, dx), \qquad 0 \leq t \leq T. \tag{4.27}$$

Then, identity (4.8), with $t = 0$, becomes (noting that $\mathbb{E}[\mathcal{E}^\beta_0 |\bar{Y}_0|^2]$ is nonnegative)

$$\beta\, \mathbb{E}\bigg[\int_{(0,T]} \mathcal{E}^\beta_s\, (1 + \beta\Delta A_s)^{-1}\, |\bar{Y}_{s-}|^2\, dA_s\bigg]$$
$$+\, \mathbb{E}\bigg[\int_{(0,T]} \mathcal{E}^\beta_s \int_E \big|\bar{Z}_s(x) - \hat{\bar{Z}}_s\big|^2\, \nu(ds\, dx) + \sum_{0<s\leq T} \mathcal{E}^\beta_s \big|\hat{\bar{Z}}_s\big|^2 \big(1 - \Delta A_s\big)\bigg]$$

$$\leq\ 2\,\mathbb{E}\bigg[\int_{(0,T]}\mathcal{E}_s^\beta\,\bar{Y}_{s-}\,\bar{f}_s\,dA_s\bigg]-\mathbb{E}\bigg[\sum_{0<s\leq T}\mathcal{E}_s^\beta\,|\bar{f}_s|^2\,|\Delta A_s|^2\bigg]. \tag{4.28}$$

From the standard inequality $2ab\leq\frac{1}{\alpha}a^2+\alpha b^2$, $\forall\,a,b\in\mathbb{R}$ and $\alpha>0$, we obtain, for any strictly positive predictable processes $(c_s)_{s\in[0,T]}$ and $(d_s)_{s\in[0,T]}$,

$$\begin{aligned}2\,\mathbb{E}\bigg[\int_{(0,T]}\mathcal{E}_s^\beta\,\bar{Y}_{s-}\,\bar{f}_s\,dA_s\bigg]&\leq\mathbb{E}\bigg[\int_{(0,T]}\frac{1}{c_s}\mathcal{E}_s^\beta\,|\bar{Y}_{s-}|^2\,dA_s^c\bigg]\\&\quad+\mathbb{E}\bigg[\sum_{0<s\leq T}\frac{1}{d_s}\,\mathcal{E}_s^\beta\,|\bar{Y}_{s-}|^2\,\Delta A_s\bigg]\\&\quad+\mathbb{E}\bigg[\int_{(0,T]}c_s\,\mathcal{E}_s^\beta\,|\bar{f}_s|^2\,dA_s^c\bigg]+\mathbb{E}\bigg[\sum_{0<s\leq T}d_s\,\mathcal{E}_s^\beta\,|\bar{f}_s|^2\,\Delta A_s\bigg].\end{aligned}$$

Therefore (4.28) becomes

$$\begin{aligned}&\mathbb{E}\bigg[\int_{(0,T]}\Big(\beta-\frac{1}{c_s}\Big)\mathcal{E}_s^\beta|\bar{Y}_{s-}|^2\,dA_s^c\bigg]\\&+\mathbb{E}\bigg[\sum_{0<s\leq T}\Big(\beta\,(1+\beta\Delta A_s)^{-1}-\frac{1}{d_s}\Big)\,\mathcal{E}_s^\beta\,|\bar{Y}_{s-}|^2\,\Delta A_s\bigg]\\&+\mathbb{E}\bigg[\int_{(0,T]}\mathcal{E}_s^\beta\int_E\big|\bar{Z}_s(x)-\hat{\bar{Z}}_s\big|^2\,\nu(ds\,dx)+\sum_{0<s\leq T}\mathcal{E}_s^\beta\big|\hat{\bar{Z}}_s\big|^2\big(1-\Delta A_s\big)\bigg]\\&\leq\mathbb{E}\bigg[\int_{(0,T]}c_s\,\mathcal{E}_s^\beta\,|\bar{f}_s|^2\,dA_s^c\bigg]+\mathbb{E}\bigg[\sum_{0<s\leq T}\big(d_s-\Delta A_s\big)\,\mathcal{E}_s^\beta\,|\bar{f}_s|^2\,\Delta A_s\bigg].\end{aligned} \tag{4.29}$$

Now, by the Lipschitz property (4.4) of f, we see that for any predictable process $(\hat{L}_{z,s})_{s\in[0,T]}$, satisfying $\hat{L}_{z,s}>L_z$, $\mathbb{P}$-a.s. for every $s\in[0,\,T]$, we have

$$|\bar{f}_s|^2\leq 2L_y^2|\bar{U}_{s-}|^2+2\hat{L}_{z,s}^2\bigg(\int_E\big|\bar{V}_s(x)-\hat{\bar{V}}_s\big|^2\,\phi_s(dx)+1_{\{\Delta A_s\neq 0\}}\frac{1-\Delta A_s}{\Delta A_s}\big|\hat{\bar{V}}_s\big|^2\bigg), \tag{4.30}$$

for all $s\in[0,\,T]$. For later use, fix $\delta\in(0,\varepsilon)$ and take $(\hat{L}_{z,s})_{s\in[0,T]}$ given by (4.25). Notice that the two components inside the maximum in (4.25) are nonnegative (the first being always strictly positive, the second being zero if $L_y=0$) and uniformly bounded, as it follows from condition (4.23). Plugging inequality (4.30) into (4.29), and using the following identity for $\bar{Z}$ (and the analogous one for $\bar{V}$)

$$\begin{aligned}&\mathbb{E}\bigg[\int_{(0,T]}\mathcal{E}_s^\beta\int_E\big|\bar{Z}_s(x)-\hat{\bar{Z}}_s\big|^2\,\nu(ds\,dx)+\sum_{0<s\leq T}\mathcal{E}_s^\beta\big|\hat{\bar{Z}}_s\big|^2\big(1-\Delta A_s\big)\bigg]\\&=\ \mathbb{E}\bigg[\int_{(0,T]}\mathcal{E}_s^\beta\int_E|\bar{Z}_s(x)|^2\,\nu^c(ds\,dx)\bigg]+\mathbb{E}\bigg[\sum_{0<s\leq T}\mathcal{E}_s^\beta\big(\widehat{|\bar{Z}_s|^2}-|\hat{\bar{Z}}_s|^2\big)\bigg],\end{aligned}$$

we obtain

$$\mathbb{E}\bigg[\int_{(0,T]}\Big(\beta-\frac{1}{c_s}\Big)\mathcal{E}_s^\beta|\bar{Y}_{s-}|^2\,dA_s^c\bigg]$$

$$+\,\mathbb{E}\bigg[\sum_{0<s\le T}\Big(\beta\,(1+\beta\Delta A_s)^{-1}-\frac{1}{d_s}\Big)\,\mathcal{E}_s^\beta\,|\bar{Y}_{s-}|^2\,\Delta A_s\bigg]$$
$$+\,\mathbb{E}\bigg[\int_{(0,T]}\mathcal{E}_s^\beta\int_E|\bar{Z}_s(x)|^2\,\nu^c(ds\,dx)\bigg]+\mathbb{E}\bigg[\sum_{0<s\le T}\mathcal{E}_s^\beta\big(\widehat{|\bar{Z}_s|^2}-|\hat{\bar{Z}}_s|^2\big)\bigg]$$
$$\le\ 2\,L_y^2\,\mathbb{E}\bigg[\int_{(0,T]}c_s\,\mathcal{E}_s^\beta|\bar{U}_{s-}|^2\,dA_s^c\bigg]+2\,\mathbb{E}\bigg[\int_{(0,T]}c_s\,\hat{L}_{z,s}^2\,\mathcal{E}_s^\beta\int_E|\bar{V}_s(x)|^2\,\nu^c(ds\,dx)\bigg]$$
$$+\,2\,L_y^2\,\mathbb{E}\bigg[\sum_{0<s\le T}\big(d_s-\Delta A_s\big)\,\mathcal{E}_s^\beta\,|\bar{U}_{s-}|^2\,\Delta A_s\bigg]$$
$$+\,2\,\mathbb{E}\bigg[\sum_{0<s\le T}\big(d_s-\Delta A_s\big)\,\hat{L}_{z,s}^2\,\mathcal{E}_s^\beta\,\big(\widehat{|\bar{V}_s|^2}-|\hat{\bar{V}}_s|^2\big)\bigg].\qquad(4.31)$$

Set $b_s := \min(\beta-\frac{1}{c_s},\beta(1+\beta\Delta A_s)^{-1}-\frac{1}{d_s})$ and $a_s := 2\hat{L}_{z,s}^2\max(c_s,d_s-\Delta A_s)$, $s\in[0,\,T]$. Then, inequality (4.31) can be rewritten as (recalling that $\hat{L}_{z,s}>0$)

$$\mathbb{E}\bigg[\int_{(0,T]}b_s\,\mathcal{E}_s^\beta|\bar{Y}_{s-}|^2\,dA_s^c\bigg]+\mathbb{E}\bigg[\sum_{0<s\le T}b_s\,\mathcal{E}_s^\beta\,|\bar{Y}_{s-}|^2\,\Delta A_s\bigg]$$
$$+\,\mathbb{E}\bigg[\int_{(0,T]}\mathcal{E}_s^\beta\int_E|\bar{Z}_s(x)|^2\,\nu^c(ds\,dx)\bigg]+\mathbb{E}\bigg[\sum_{0<s\le T}\mathcal{E}_s^\beta\big(\widehat{|\bar{Z}_s|^2}-|\hat{\bar{Z}}_s|^2\big)\bigg]$$
$$\le\ \mathbb{E}\bigg[\int_{(0,T]}\frac{L_y^2}{\hat{L}_{z,s}^2}\,a_s\,\mathcal{E}_s^\beta|\bar{U}_{s-}|^2\,dA_s^c\bigg]+\mathbb{E}\bigg[\sum_{0<s\le T}\frac{L_y^2}{\hat{L}_{z,s}^2}\,a_s\,\mathcal{E}_s^\beta\,|\bar{U}_{s-}|^2\,\Delta A_s\bigg]$$
$$+\,\mathbb{E}\bigg[\int_{(0,T]}a_s\,\mathcal{E}_s^\beta\int_E|\bar{V}_s(x)|^2\,\nu^c(ds\,dx)\bigg]+\mathbb{E}\bigg[\sum_{0<s\le T}a_s\,\mathcal{E}_s^\beta\,\big(\widehat{|\bar{V}_s|^2}-|\hat{\bar{V}}_s|^2\big)\bigg].\qquad(4.32)$$

It follows from (4.32) that Φ is a contraction if:

(i) there exists $\alpha\in(0,1)$ such that $a_s\le\alpha$, $\mathbb{P}$-a.s. for every $s\in[0,\,T]$;

(ii) $\frac{L_y^2}{\hat{L}_{z,s}^2}\le b_s$, $\mathbb{P}$-a.s. for every $s\in[0,\,T]$.

Let us prove that (i) and (ii) hold. Condition (i) is equivalent to ask that there exists $\alpha\in(0,1)$ such that, for all $s\in[0,\,T]$,

$$c_s\ \le\ \frac{1-\alpha}{2\,\hat{L}_{z,s}^2},\qquad d_s\ \le\ \frac{1-\alpha}{2\,\hat{L}_{z,s}^2}+\Delta A_s,\quad \mathbb{P}\text{-a.s.}$$

Then we choose $\alpha=\delta$, where $\delta\in(0,\varepsilon)$ was fixed in the statement of the theorem, and c_s,d_s given by

$$c_s\ =\ \frac{1-\delta}{2\,\hat{L}_{z,s}^2},\qquad d_s\ =\ \frac{1-\delta}{2\,\hat{L}_{z,s}^2}+\Delta A_s,\qquad(4.33)$$

for all $s\in[0,\,T]$, so that (i) holds true. Concerning (ii), we have, for all $s\in[0,\,T]$,

$$\min\Big(\beta-\frac{1}{c_s},\beta(1+\beta\Delta A_s)^{-1}-\frac{1}{d_s}\Big)\ \ge\ \frac{L_y^2}{\hat{L}_{z,s}^2},$$

which becomes

$$\beta \;\geq\; \frac{L_y^2}{\hat{L}_{z,s}^2} + \frac{1}{c_s}, \qquad \beta \;\geq\; \frac{\frac{L_y^2}{\hat{L}_{z,s}^2} + \frac{1}{d_s}}{1 - \Delta A_s\Big(\frac{L_y^2}{\hat{L}_{z,s}^2} + \frac{1}{d_s}\Big)}, \tag{4.34}$$

where for the last inequality we need to impose the additional condition

$$1 - \Delta A_s\bigg(\frac{L_y^2}{\hat{L}_{z,s}^2} + \frac{1}{d_s}\bigg) \;>\; 0.$$

This latter inequality can be rewritten as

$$L_y^2\,\Delta A_s \;<\; \hat{L}_{z,s}^2\bigg(1 - \frac{\Delta A_s}{d_s}\bigg) \;=\; \frac{(1-\delta)\,\hat{L}_{z,s}^2}{1 - \delta + 2\,\hat{L}_{z,s}^2\,\Delta A_s}, \tag{4.35}$$

where the last equality follows from the definition of d_s in (4.33). From (4.25), and since in particular

$$\hat{L}_{z,s}^2 \;\geq\; \frac{(1-\delta)\,L_y}{\sqrt{2(1-\delta)} - 2\,L_y\,\Delta A_s} \;>\; \frac{(1-\delta)\,L_y^2\,\Delta A_s}{1 - \delta - 2\,L_y^2\,|\Delta A_s|^2}, \qquad \mathbb{P}\text{-a.s.}, \forall\, s \in [0,\,T],$$

it follows that inequality (4.35) holds. Finally, concerning (4.34), we begin noting that

$$\frac{L_y^2}{\hat{L}_{z,s}^2} + \frac{1}{c_s} < \frac{\frac{L_y^2}{\hat{L}_{z,s}^2} + \frac{1}{d_s}}{1 - \Delta A_s\Big(\frac{L_y^2}{\hat{L}_{z,s}^2} + \frac{1}{d_s}\Big)},$$

as it can be shown using (4.33). Now, let us denote

$$\frac{\frac{L_y^2}{\hat{L}_{z,s}^2} + \frac{1}{d_s}}{1 - \Delta A_s\Big(\frac{L_y^2}{\hat{L}_{z,s}^2} + \frac{1}{d_s}\Big)} \;=\; H_s(\hat{L}_{z,s}^2),$$

where, for every $s \in [0,\,T]$,

$$H_s(\ell) \;=\; \frac{h_s(\ell)}{1 - \Delta A_s\,h_s(\ell)}, \qquad h_s(\ell) \;=\; \frac{L_y^2}{\ell} + \frac{2\,\ell}{1 - \delta + 2\,\ell\,\Delta A_s}, \qquad \ell > 0.$$

Notice that H_s attains its minimum at $\ell_s^* = \frac{(1-\delta)\,L_y}{\sqrt{2(1-\delta)} - 2\,L_y\,\Delta A_s}$. This explains the expression of the second component inside the maximum in (4.25). In conclusion, given $(\hat{L}_{z,s})_{s\in[0,\,T]}$ as in (4.25) we obtain a lower bound for β from the second inequality in (4.34), which corresponds to (4.24). □

Remark 4.4.3. (i) In [**26**] the authors study a class of BSDEs driven by a countable sequence of square-integrable martingales, with a generator f integrated with respect to a right-continuous nondecreasing process A as in (4.5). Similarly to our setting, A is not necessarily continuous, however in [**26**] it is supposed to be *deterministic* (instead of predictable). Theorem 6.1 in [**26**] provides an existence and uniqueness result for the class of BSDEs studied in [**26**] under the following assumption ($2\,L_{y,t}^2$ corresponds to c_t and ΔA_t corresponds to $\Delta\mu_t$ in the notation of [**26**]):

$$2\,L_{y,t}^2\,|\Delta A_t|^2 \;<\; 1, \qquad \forall\, t \in [0,\,T], \tag{4.36}$$

where $L_{y,t}$ is a measurable deterministic function uniformly bounded such that (4.4) holds with $L_{y,t}$ in place of L_y. As showed at the beginning of the proof of Theorem 6.1 in [**26**], if (4.36) holds (and A is as in [**26**]), then there exists $\varepsilon \in (0,1)$ such that

$$2\, L_{y,t}^2\, |\Delta A_t|^2 \;\leq\; 1-\varepsilon, \qquad \forall\, t \in [0,\, T]. \tag{4.37}$$

This proves that when condition (4.36) holds then (4.37) is also valid, since in our setting we can take $L_{y,t} \equiv L_y$.

(ii) Section 4.3 in [**29**] provides a counter-example to existence for BSDE (4.5) when A is discontinuous, as it can be the case in our setting; the rest of the paper [**29**] studies BSDE (4.5) with A continuous. Let us check that the counter-example proposed in [**29**] does not satisfy condition (4.23). In [**29**] the process A is a pure jump process with a single jump of size $p \in (0,1)$ at a deterministic time $t \in (0,T]$. The Lipschitz constant of f with respect to y is $L_y = \frac{1}{p}$. Then

$$2\, L_y^2\, |\Delta A_t|^2 \;=\; 2$$

if t is the jump time of A, so that condition (4.23) is violated. □

Remark 4.4.4. Suppose that μ is an integer-valued random measure on $\mathbb{R}_+ \times E$ not necessarily discrete. Then ν can still be disintegrated as follows

$$\nu(\omega, dt\, dx) \;=\; dA_t(\omega)\, \phi_{\omega,t}(dx),$$

where A is a right-continuous nondecreasing predictable process such that $A_0 = 0$, but ϕ is in general only a transition measure (instead of transition probability) from $(\Omega \times [0,T], \mathcal{P})$ into $(E, \mathcal{E})$. Notice that when μ is discrete one can choose ϕ to be a transition probability, therefore $\phi(E) = 1$ and $\nu(\{t\} \times E) = \Delta A_t$ (a property used in the previous sections). When μ is not discrete, let us suppose that ν^d can be disintegrated as follows

$$\nu^d(\omega, dt\, dx) \;=\; \Delta A_t(\omega)\, \phi^d_{\omega,t}(dx), \qquad \phi^d_{\omega,t}(E) \;=\; 1, \tag{4.38}$$

where ϕ^d is a transition *probability* from $(\Omega \times [0,T], \mathcal{P})$ into $(E, \mathcal{E})$. In particular $\nu^d(\{t\} \times E) = \Delta A_t$. Then, when (4.38) and a martingale representation theorem for μ hold, all the results of this chapter are still valid and can be proved proceeding along the same lines. As an example, (4.38) holds when μ is the jump measure of a Lévy process, indeed in this case ΔA_t is identically zero. □

Remark 4.4.5. As an application of the results presented in this chapter, suppose that μ is the jump measure of a piecewise deterministic Markov process X with values in E. We follow the notation introduced in [**35**], Chapter 2, Section 24 and 26. Denoted by $(T_n)_n$ the jump times of the process X, the random measure μ can be written as

$$\mu(dt\, dx) = \sum_{n=1}^{\infty} \delta_{(T_n,\, X_{T_n})}(dt\, dx).$$

Moreover, according to (26.2) in [**35**], the compensator of μ has the form

$$\nu(\omega, dt\, dx) = (\lambda(X_{t-}(\omega))\, dt + dp^*_t(\omega))\, P(X_{t-}(\omega), dx), \tag{4.39}$$

where $P : \bar{E} \times \mathcal{E} \to E$ and $\lambda : E \to \mathbb{R}_+$ are respectively the transition probability measure and the jump rate of the process X, and

$$p_t^* = \sum_{n=1}^{\infty} 1_{\{t \geq T_n\}} 1_{\{X_{T_n-} \in \Gamma\}}$$

is the process counting the number of jumps of X from the active boundary $\Gamma \subset \partial E$ (for the precise definition of Γ see page 61 in [**35**]).

From (4.39) we see that decomposition (4.2) for ν holds with $dA_t(\omega) = \lambda(X_{t-}(\omega))\, dt + dp_t^*(\omega)$ and $\phi_{\omega,t}(dx) = P(X_{t-}(\omega), dx)$. In particular, A is predictable (not deterministic) and discontinuous, with jumps $\Delta A_t = 1_{\{X_{t-} \in \Gamma\}}$. In this case condition (4.23) can be written as

$$L_y \; < \; \frac{1}{\sqrt{2}}. \tag{4.40}$$

The fact that the above condition is only on L_y, rather than on L_z, is particularly important in the study of control problems related to PDMPs by means of BSDEs methods. This latter turns out to be technically involved and is the subject of a work in progress by the author, where the methodology developed in Chapter 3 is extended in suitable way to the case of PDMPs on a state space with boundary. Here, we just say that when control problems are considered then $L_y = 0$ and condition (4.40) is automatically satisfied. We also emphasize that, as expected, the main difficulties arise from the presence of discontinuities at the boundary of the domain. $\square$

Chapter 5

Weak Dirichlet processes with jumps

5.1. Introduction

The present chapter extends stochastic calculus via regularization to the case of jump processes, and carries on the investigations of the so called weak Dirichlet processes in the discontinuous case. This calculus will be applied in Chapter 6, where we provide the identification of the solution of a forward backward stochastic differential equation driven by a random measure, when the underlying process is of weak Dirichlet type.

Stochastic calculus via regularization was essentially known in the case of continuous integrators X, see e.g. Russo and Vallois [**116**], [**117**], with a survey in [**121**]. In this case a fairly complete theory was developed, see for instance Itô formulae for processes with finite quadratic (and more general) variations, stochastic differential equations, Itô-Wentzell type formulae in Flandoli and Russo [**63**], and generalizations to the case of Banach space type integrators given in Di Girolami and Russo [**44**]. The notion of covariation $[X, Y]$ (resp. quadratic variation $[X, X]$) for two processes X, Y (resp. a process X) has been introduced in the framework of regularizations (see Russo and Vallois [**119**]) and of discretizations as well (see Föllmer [**66**]). Even if there is no direct theorem relating the two approaches, they coincide in all the examples considered in the literature. If X is a finite quadratic variation continuous process, an Itô formula has been proved for the expansion of $F(X_t)$, when $F \in C^2$, see [**119**]; this constitutes the counterpart of the related result for discretizations, see [**66**]. Moreover, for F of class C^1 and X a reversible semimartingale, an Itô expansion has been established in Russo and Vallois [**120**].

When F is less regular than C^1, the Itô formula can be replaced by a Fukushima-Dirichlet decomposition for X *weak Dirichlet process* (with respect to a given filtration $(\mathcal{F}_t)$). The notion of Dirichlet process is a familiar generalization of the concept of semimartingale, and was introduced by [**66**] and Bertoin [**14**] in the discretization framework. The analogue of the Doob-Meyer decomposition for a Dirichlet process is that it is the sum of a local martingale M and an adapted process A with zero quadratic variation. Here A is the generalization of a bounded variation process. However, requiring A to have zero quadratic variation imposes that A is continuous, see Lemma 5.3.9; since a bounded variation process with jumps has a non zero finite quadratic variation, the generalization of the semimartingale in the jump case is not necessarily represented by the notion of Dirichlet process. A natural generalization should then at least include the possibility that A is a bounded variation process with jumps. The concept of $(\mathcal{F}_t)$-weak Dirichlet process was later introduced in Errami and Russo [**58**] and Gozzi and Russo [**71**] for a continuous process X, and applications to stochastic control were considered in Gozzi and Russo [**70**]. Such a process is defined as the sum of a local martingale M and an adapted process A such that $[A, N] = 0$ for every continuous local martingale N. This notion turns out to be a correct generalization of the semimartingale notion in the discontinuous framework, and is extended to the case of jumps processes in the significant work Coquet, Jakubowsky, Mémin and Slomińsky [**30**], by using the discretizations techniques. In the continuous case, a chain rule was established for $F(t, X_t)$ when F belongs to class $C^{0,1}$ and X is a weak Dirichlet process, see [**71**]. Such a process is indeed again a weak Dirichlet process (with possibly no finite quadratic variation). Towards calculus in the jump case only few steps were done in [**119**], Russo and Vallois [**118**], and several other authors, see Chapter 15 of the book of Di Nunno, Øksendal and Proske [**45**] and references therein. For instance no Itô type formulae have been established in the framework of regularization and in the discretization framework only very few chain rule results are available for $F(X)$, when $F(X)$ is not a semimartingale. In that direction two peculiar results are available: the expansion of $F(X_t)$ when X is a reversible semimartingale and F is of class C^1 with some Hölder conditions on the derivatives (see Errami, Russo and Vallois [**59**]) and a chain rule for $F(X_t)$ when X is a weak Dirichlet (càdlàg) process and F is of class C^1, see [**30**]. The work in [**59**] has been continued by several authors, see e.g. Eisenbaum [**47**] and references therein, expanding the remainder making use of local time type processes. A systematic study of that calculus was missing and in this chapter we fill out this gap.

Let us now go through the description of the main results of the chapter. As we have already mentioned, our first basic objective consists in developing a calculus via regularization in the case of finite quadratic variation càdlàg processes. To this end, we revisit the definitions given by [**119**] concerning forward integrals (resp. covariations). Those objects are introduced as u.c.p. (uniform convergence in probability) limit of the expressions of the type (5.12) (resp. (5.13)). That convergence ensures that the limiting objects are càdlàg, since the approximating expressions have the same property. For instance a càdlàg process X will be called *finite quadratic variation process* whenever the limit (which will be denoted by $[X, X]$) of

$$[X, X]^{ucp}_{\varepsilon}(t) := \int_{]0,\, t]} \frac{(X((s+\varepsilon) \wedge t) - X(s))^2}{\varepsilon}\, ds, \tag{5.1}$$

exists u.c.p. In [**119**], the authors introduced a slightly different approximation of $[X, X]$ when X is continuous, namely

$$C_\varepsilon(X, X)(t) := \int_{]0,\, t]} \frac{(X((s+\varepsilon) - X(s))^2}{\varepsilon}\, ds. \tag{5.2}$$

When the u.c.p. limit of $C_\varepsilon(X, X)$ exists, it is automatically a continuous process, since the approximating processes are continuous. For this reason, when X is a jump process, the choice of approximation (5.2) would not be suitable, since its quadratic variation is expected to be a jump process. In that case, the u.c.p. convergence of (5.1) can be shown to be equivalent with a notion of convergence which is associated with the a.s. convergence (up to subsequences) in measure of $C_\varepsilon(X, X)(t)\, dt$, see Section 5.4. Both formulations will be used in the development of the calculus.

For a càdlàg finite quadratic variation process X, we establish, via regularization techniques, an Itô formula for $C^{1,2}$ functions of X. This is the object of Proposition 5.5.1, whose proof is based on an accurate separation between the neighborhood of "big" and "small" jumps, where specific tools are used, see for instance the preliminary results Lemma 5.3.11 and Lemma 5.3.12. Another significant instrument is a Lemma of Dini type in the case of càdlàg functions, see Lemma 5.3.15. Finally, from Proposition 5.5.1 easily follows an Itô formula under weaker regularity conditions on F, see Proposition 5.5.2. We remark that a similar formula was stated in [**59**], using a discretization definition of the covariation, when F is time-homogeneous.

The second target of the chapter consists in investigating weak Dirichlet jump processes. Contrarily to the continuous case, the decomposition $X = M + A$ is generally not unique. We introduce the notion of *special weak Dirichlet process* with respect to some filtration $(\mathcal{F}_t)$. Such a process is a weak Dirichlet process admitting a decomposition $X = M + A$, where M is an $(\mathcal{F}_t)$-local martingale and where the "orthogonal" process A is predictable. The decomposition of a special weak Dirichlet process is unique, see Proposition 5.6.8. Such a process constitutes a generalization of the notion of semimartingale in the framework of weak Dirichlet processes. We remark that a continuous weak Dirichlet process is a special weak Dirichlet.

Two significant results are Theorem 5.6.14 and Theorem 5.6.26. They both concern expansions of $F(t, X_t)$ where F is of class $C^{0,1}$ and X is a weak Dirichlet process of finite quadratic variation. Theorem 5.6.14 states that $F(t, X_t)$ will be again a weak Dirichlet process, however not necessarily of finite quadratic variation. Theorem 5.6.26 concerns the cases when X and $(F(t, X_t))_t$ are special weak Dirichlet processes. A first significant step in this sense was done in [**30**], where X belongs to a bit different class of special weak Dirichlet jump processes (of finite energy) and F does not depend on time and has bounded derivative. They show that $F(X)$ is again a special weak Dirichlet process. In [**30**] the underlying process has finite energy, which requires a control of the expectation of the approximating sequences of the quadratic variation. On the other hand, our techniques do not require that type of control. Moreover, the integrability condition (5.134) that we ask on $F(t, X_t)$ in order to get the chain rule in Theorem 5.6.26 is automatically verified under the hypothesis on the first-order derivative considered in [**30**], see Remark 5.6.25. In some cases a chain rule may hold even when F is only continuous if we know a priori some information of $(F(t,X_t))$. This is provided by Proposition 5.6.28 and does not require any assumption on the càdlàg process X. This applies for instance to the case when X is a pure jump process, see Remark 5.6.30.

In the present chapter we also introduce a subclass of weak Dirichlet processes, called *particular*, see Definition 5.6.16. Those processes inherit some of the semimartingales features: as in the semimartingale case, the particular weak Dirichlet processes admit an integral representation (see Proposition 5.6.19) and a (unique) canonical decomposition holds when $|x|\,1_{\{|x|>1\}} * \mu \in \mathcal{A}^+_{loc}$. Under that conditions, those particular processes are indeed special weak Dirichlet processes, see Proposition 5.6.18 and 5.6.19.

The chapter is organized as follows. In Section 5.2 we introduce the notations and we recall some basic results on the stochastic integration with respect to integer-valued random measures associated to càdlàg processes. In Section 5.3 we give some preliminary results to the development of the calculus via regularization with jumps; additional comments and technical results on calculus via regularizations in the discontinuous framework are reported in Section 5.4. Section 5.5 is devoted to the proof of a $C^{1,2}$ Itô formula for càdlàg processes. Finally, Section 5.6 concerns the study of weak Dirichlet processes, and presents the expansions of $F(t,X_t)$ for X weak Dirichlet, when F is of class $C^{0,1}$.

5.2. Preliminaries and basic notations

In what follows, we are given a probability space $(\Omega,\mathcal{F},\mathbb{P})$, a positive horizon T and a filtration $\mathcal{F}=(\mathcal{F}_t)_{t\geq 0}$. Given a topological space E, in the sequel $\mathcal{B}(E)$ will denote the Borel σ-field associated with E. $\mathcal{P}$ (resp. $\tilde{\mathcal{P}}=\mathcal{P}\otimes\mathcal{B}(\mathbb{R})$) will designate the predictable σ-field on $\Omega\times[0,T]$ (resp. on $\tilde{\Omega}=\Omega\times[0,T]\times\mathbb{R}$). Analogously, we set $\mathcal{O}$ (resp. $\tilde{\mathcal{O}}=\mathcal{O}\otimes\mathcal{B}(\mathbb{R})$) the optional σ-field on $\Omega\times[0,T]$ (resp. on $\tilde{\Omega}$). The symbols $\mathbb{D}^{ucp}$ and $\mathbb{L}^{ucp}$ will denote the space of all adapted càdlàg and càglàd processes endowed with the u.c.p. (uniform convergence in probability) topology. By convention, any càdlàg process defined on $[0,\,T]$ is extended on $\mathbb{R}_+$ by continuity.

We will also indicate by $\mathcal{A}$ (resp $\mathcal{A}_{\rm loc}$) the collection of all adapted processes with integrable variation (resp. with locally integrable variation), and by $\mathcal{A}^+$ (resp $\mathcal{A}^+_{\rm loc}$) the collection of all adapted integrable increasing (resp. adapted locally integrable) processes. The significance of locally is the usual one which refers to localization by stopping times, see e.g. (0.39) of Jacod's book [**77**].

We will indicate by $C^{1,2}$ (resp. $C^{0,1}$) the space of all functions

$$u:[0,\,T]\times\mathbb{R}\to\mathbb{R},\quad (t,x)\mapsto u(t,x)$$

that are continuous together their derivatives $\partial_t u$, $\partial_x u$, $\partial_{xx} u$ (resp. $\partial_x u$). $C^{1,2}$ is equipped with the topology of uniform convergence on each compact of u, $\partial_x u$, $\partial_{xx} u$, $\partial_t u$; $C^{0,1}$ is equipped with the same topology on each compact of u and $\partial_x u$.

5.2.1. Càdlàg processes and the associated random measures. The concept of random measure allows a very tractable description of the jumps of a càdlàg process. We recall here the main definitions and some properties that we will extensively use in the following; for a more detailed discussion on this topic and the unexplained notations see Appendices A and B.

For any $X = (X_t)$ adapted real valued càdlàg process on $[0, T]$, we call jump measure of X the integer-valued random measure on $\mathbb{R}_+ \times \mathbb{R}$ defined as

$$\mu^X(\omega, dt\, dx) := \sum_{s \in]0, T]} 1_{\{\Delta X_s(\omega) \neq 0\}}\, \delta_{(s,\, \Delta X_s(\omega))}(dt\, dx). \tag{5.3}$$

Remark 5.2.1. The jump measure μ^X acts in the following way: for any positive function $W \in \tilde{\mathcal{O}}$ we have

$$\sum_{s \in]0, T]} 1_{\{\Delta X_s \neq 0\}} W_s(\cdot, \Delta X_s) = \int_{]0,T] \times \mathbb{R}} W_s(\cdot, x)\, \mu^X(\cdot, ds\, dx).$$

In the sequel we will make often use of the following assumption on the processes X:

$$\sum_{s \in]0, T]} |\Delta X_s|^2 < \infty, \text{ a.s.} \tag{5.4}$$

Adapting the definition of locally bounded process stated before Theorem 15, Chapter IV, in [**110**], to the processes indexed by $[0, T]$, we can state the following.

Definition 5.2.2. A process $(X_t)_{t \in [0, T]}$ is locally bounded if there exists a sequence of stopping times $(\tau_n)_{n \geq 1}$ in $[0, T] \cup \{+\infty\}$ increasing to ∞ a.s., such that $(X_{\tau_n \wedge t}\, 1_{\{\tau_n > 0\}})_{t \in [0, T]}$ is bounded.

Remark 5.2.3.

(i) Any càglàd process is locally bounded, see the lines above Theorem 15, Chapter IV, in [**110**].

(ii) Let X be a càdlàg process satisfying condition (5.4). Set $(Y_t)_{t \in [0, T]} = (X_{t-}, \sum_{s<t} |\Delta X_s|^2)_{t \in [0, T]}$. The process Y is càglàd, therefore locally bounded by item (i). In particular, we can fix a sequence of stopping times $(\tau_n)_{n \geq 1}$ in $[0, T] \cup \{+\infty\}$ increasing to ∞ a.s., such that $(Y_{\tau_n \wedge t}\, 1_{\{\tau_n > 0\}})_{t \in [0, T]}$ is bounded.

Proposition 5.2.4. *Let $p = 1, 2$. Let X be a real-valued càdlàg process on $[0, T]$ satisfying*

$$\sum_{s \in]0, T]} |\Delta X_s|^p < \infty, \text{ a.s.}$$

Then

$$\int_{]0,\, t] \times \mathbb{R}} |x|^p\, 1_{\{|x| \leq 1\}}\, \mu^X(ds\, dx) \;\in \mathcal{A}^+_{\mathrm{loc}}. \tag{5.5}$$

Proof. Set $Y_t = \sum_{s<t} |\Delta X_s|^p$. The process Y is càglàd, therefore locally bounded; in particular, we can fix a sequence of stopping times $(\tau_n)_{n \geq 1}$ in $[0, T] \cup \{+\infty\}$

increasing to ∞ a.s., such that $(Y_{\tau_n \wedge t} 1_{\{\tau_n>0\}})_{t\in[0,T]}$ is bounded. Fix $\tau = \tau_n$, and let M such that $\sup_{t\in[0,T]} |Y_{t\wedge\tau} 1_{\{\tau>0\}}| \le M$. We have

$$\begin{aligned}
&\mathbb{E}\left[\int_{]0,\,t\wedge\tau]\times\mathbb{R}} |x|^p 1_{\{|x|\le 1\}}\ \mu^X(ds,\,dx)\right]\\
&= \mathbb{E}\left[\sum_{0<s<t\wedge\tau} |\Delta X_s|^p 1_{\{|\Delta X_s|\le 1\}} 1_{\{\tau>0\}} + |\Delta X_{t\wedge\tau}|^p 1_{\{|\Delta X_{t\wedge\tau}|\le 1\}} 1_{\{\tau>0\}}\right]\\
&\le M+1,
\end{aligned}$$

and thus (5.5) holds. $\square$

Corollary 5.2.5. *Let X be a càdlàg process satisfying condition* (5.4). *Then*

$$x\, 1_{\{|x|\le 1\}} \in \mathcal{G}^2_{\mathrm{loc}}(\mu^X). \tag{5.6}$$

In particular the stochastic integral

$$\int_{]0,\,t]\times\mathbb{R}} x\, 1_{\{|x|\le 1\}}\, (\mu^X - \nu^X)(ds\, dx) \tag{5.7}$$

is well-defined and defines a purely discontinuous square integrable local martingale.

Proof. Property (5.6) is a direct application of Proposition 5.2.4 with $p = 2$, and Lemma B.21-2. The second part of the result follows by (5.6) and Theorem B.22. $\square$

Remark 5.2.6. Let $\varphi : \Omega \times [0,T] \times \mathbb{R} \to \mathbb{R}$ be a $\tilde{\mathcal{P}}$-measurable function and A a $\tilde{\mathcal{P}}$-measurable subset of $\Omega \times [0,\,T] \times \mathbb{R}$, such that

$$|\varphi|\, 1_A * \mu^X \in \mathcal{A}^+_{\mathrm{loc}}, \tag{5.8}$$

$$|\varphi|^2\, 1_{A^c} * \mu^X \in \mathcal{A}^+_{\mathrm{loc}}. \tag{5.9}$$

Then the process φ belongs to $\mathcal{G}^1_{\mathrm{loc}}(\mu^X)$.

As a matter of fact, (5.8) and Proposition B.18 give that $\varphi\, 1_A$ belongs to $\mathcal{G}^1_{\mathrm{loc}}(\mu^X)$. On the other hand, (5.9), together with Lemma B.21-2), implies that $\varphi\, 1_{A^c}$ belongs to $\mathcal{G}^2_{\mathrm{loc}}(\mu^X) \subset \mathcal{G}^1_{\mathrm{loc}}(\mu^X)$.

Proposition 5.2.7. *Let X be a càdlàg process on* $[0,\,T]$ *satisfying condition* (5.4)*, and let F be a function of class $C^{1,2}$. Then*

$$|(F(s, X_{s-}+x) - F(s,X_{s-}) - x\,\partial_x F(s,X_{s-})|\, 1_{\{|x|\le 1\}} * \mu^X \in \mathcal{A}^+_{\mathrm{loc}}.$$

Proof. Let $(\tau_n)_{n\ge 1}$ be the sequence of stopping times introduced in Remark 5.2.3-(ii) for the process $Y_t = (X_{t-}, \sum_{s<t} |\Delta X_s|^2)$. Fix $\tau = \tau_n$, and let M such that $\sup_{t\in[0,T]} |Y_{t\wedge\tau} 1_{\{\tau>0\}}| \le M$. So, by an obvious Taylor expansion, taking into account Remark 5.2.1, we have

$$\begin{aligned}
&\mathbb{E}\left[\int_{]0,\,t\wedge\tau]\times\mathbb{R}} |(F(s,X_{s-}+x) - F(s,X_{s-}) - x\,\partial_x F(s,X_{s-})|\, 1_{\{|x|\le 1\}}\, \mu^X(ds,\,dx)\right]\\
&= \mathbb{E}\left[\sum_{0<s\le t\wedge\tau} [F(s,X_s) - F(s,X_{s-}) - \partial_x F(s,X_{s-})\,\Delta X_s]\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left[\sum_{0<s\le t\wedge\tau} (\Delta X_s)^2\, 1_{\{\tau>0\}} \int_0^1 [\partial^2_{xx}F(s, X_{s-} + a\,\Delta X_s) - \partial^2_{xx}F(s, X_{s-})]\, da\right]\\
&\le 2 \sup_{\substack{y\in[-M,\, M]\\ t\in[0,\, T]}} |\partial^2_{xx}F|(t,\, y)\cdot\\
&\quad \cdot \mathbb{E}\left[\sum_{0<s<t\wedge\tau} |\Delta X_s|^2\, 1_{\{|\Delta X_s|\le 1\}}\, 1_{\{\tau>0\}} + |\Delta X_\tau|^2\, 1_{\{|\Delta X_\tau|\le 1\}}\, 1_{\{\tau>0\}}\right]\\
&\le 2 \sup_{\substack{y\in[-M,\, M]\\ t\in[0,\, T]}} |\partial^2_{xx}F|(t,\, y)\, \cdot (M+1),
\end{aligned}
$$

and this concludes the proof. □

Proposition 5.2.8. *Let X be a càdlàg process on $[0,\,T]$ satisfying condition (5.4), and let F be a function of class $C^{0,1}$. Then*

$$|(F(s, X_{s-}+x) - F(s, X_{s-})|^2\, 1_{\{|x|\le 1\}} * \mu^X \in \mathcal{A}^+_{\rm loc}, \tag{5.10}$$

$$|x\, \partial_x F(s, X_{s-})|^2\, 1_{\{|x|\le 1\}} * \mu^X \in \mathcal{A}^+_{\rm loc}. \tag{5.11}$$

Proof. Proceeding as in the proof of Proposition 5.2.7, we consider the sequence of stopping times $(\tau_n)_{n\ge 1}$ defined in Remark 5.2.3-(ii) for the process $Y_t = (X_{t-}, \sum_{s<t} |\Delta X_s|^2)$. Fix $\tau = \tau_n$, and let M such that $\sup_{t\in[0,\,T]} |Y_{t\wedge\tau}\, 1_{\{\tau>0\}}| \le M$. For any $t \in [0, T]$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\int_{[0,\, t\wedge\tau]\times\mathbb{R}} |(F(s, X_{s-}+x) - F(s, X_{s-})|^2\, 1_{\{|x|\le 1\}}\, \mu^X(ds,\, dx)\right]\\
&\le \sup_{\substack{y\in[-M,\, M]\\ t\in[0,\, T]}} |\partial_x F|^2(t,\, y)\cdot\\
&\quad \cdot \mathbb{E}\left[\sum_{s<t\wedge\tau} |\Delta X_s|^2\, 1_{\{|\Delta X_s|\le 1\}} 1_{\{\tau>0\}} + |\Delta X_\tau|^2\, 1_{\{|\Delta X_\tau|\le 1\}}\, 1_{\{\tau>0\}}\right]\\
&\le \sup_{\substack{y\in[-M,M]\\ t\in[0,\, T]}} |\partial_x F|^2(t,\, y)\, \cdot (M+1),
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left[\int_{[0,\, t\wedge\tau]\times\mathbb{R}} |x\, \partial_x F(s, X_{s-})|^2\, 1_{\{|x|\le 1\}}\, \mu^X(ds,\, dx)\right]\\
&= \mathbb{E}\left[\int_{[0,\, t\wedge\tau]\times\mathbb{R}} |x|^2\, |\partial_x F|^2(t,\, X_{s-})\, 1_{\{|x|\le 1\}}\, \mu^X(ds,\, dx)\right]\\
&\le \sup_{\substack{y\in[-M,\, M]\\ t\in[0,\, T]}} |\partial_x F|^2(t,\, y)\cdot\\
&\quad \cdot \mathbb{E}\left[\sum_{s<t\wedge\tau} |\Delta X_s|^2\, 1_{\{|\Delta X_s|\le 1\}}\, 1_{\{\tau>0\}} + |\Delta X_\tau|^2\, 1_{\{|\Delta X_\tau|\le 1\}}\, 1_{\{\tau>0\}}\right]
\end{aligned}
$$

$$\leq \sup_{\substack{y\in[-M,\,M]\\ t\in[0,\,T]}} |\partial_x F|^2(t,\,y)\cdot(M+1).$$

□

5.3. Calculus via regularization with jumps

Let f and g be two functions defined on $\mathbb{R}$, and set

$$I^{-ucp}(\varepsilon,t,f,dg) = \int_{]0,\,t]} f(s)\,\frac{g((s+\varepsilon)\wedge t)-g(s)}{\varepsilon}\,ds, \tag{5.12}$$

$$[f,g]^{ucp}_{\varepsilon}(t) = \int_{]0,\,t]} \frac{(f((s+\varepsilon)\wedge t)-f(s))(g((s+\varepsilon)\wedge t)-g(s))}{\varepsilon}\,ds. \tag{5.13}$$

Notice that the function $I^{-ucp}(\varepsilon,t,f,dg)$ is càdlàg and admits the decomposition

$$I^{-ucp}(\varepsilon,t,f,dg) = \int_0^{(t-\varepsilon)_+} f(s)\,\frac{g(s+\varepsilon)-g(s)}{\varepsilon}\,ds + \int_{(t-\varepsilon)_+}^{t} f(s)\,\frac{g(t)-g(s)}{\varepsilon}\,ds. \tag{5.14}$$

Definition 5.3.1. Let X be a càdlàg process and Y be a process belonging to $L^1([0,\,T])$ a.s. Suppose that there exists a process $(I(t))_{t\in[0,\,T]}$ such that $(I^{-ucp}(\varepsilon,t,Y,dX))_{t\in[0,\,T]}$ converges u.c.p. to $(I(t))_{t\in[0,\,T]}$, namely

$$\lim_{\varepsilon\to 0}\mathbb{P}\left(\sup_{0\leq s\leq t}|I^{-ucp}(\varepsilon,t,Y,dX)-I(t)|>\alpha\right)=0 \qquad \text{for every } \alpha>0.$$

Then we will set $\int_{]0,\,t]} Y_s\,d^-X_s := I(t)$. That process will be called *the forward integral of Y with respect to X.*

Remark 5.3.2. In [**119**] a very similar notion of forward integral is considered:

$$I^{-RV}(\varepsilon,t,f,dg) = \int_{\mathbb{R}} f_{t]}(s)\,\frac{g_{t]}(s+\varepsilon)-g_{t]}(s)}{\varepsilon}\,ds,$$

with

$$f_{t]} = \begin{cases} f(0_+) & \text{if } x\leq 0,\\ f(x) & \text{if } 0<x\leq t,\\ f(t_+) & \text{if } x>t.\end{cases}$$

The u.c.p. limit of $I^{-RV}(\varepsilon,t,f,dg)$, when it exists, coincides with that of the process $I^{-ucp}(\varepsilon,t,f,dg)$. As a matter of fact, the process $I^{-RV}(\varepsilon,t,f,dg)$ is càdlàg and can be rewritten as

$$I^{-RV}(\varepsilon,t,f,dg) \;=\; I^{-ucp}(\varepsilon,t,f,dg) - f(0_+)\,\frac{1}{\varepsilon}\int_0^{\varepsilon}[g(s)-g(0_+)]\,ds. \tag{5.15}$$

In particular

$$\sup_{t\in[0,T]}[I^{-ucp}(\varepsilon,t,f,dg)-I^{-RV}(\varepsilon,t,f,dg)] = f(0_+)\,\frac{1}{\varepsilon}\int_0^{\varepsilon}[g(s)-g(0_+)]\,ds,$$

and therefore

$$\limsup_{\varepsilon\to 0}\sup_{t\in[0,T]}[I^{-RV}(\varepsilon,t,f,dg)-I^{-ucp}(\varepsilon,t,f,dg)]=0.$$

Proposition 5.3.3. *Let A be a càdlàg predictable process and Y be a process belonging to $L^1([0,T])$ a.s. Then the forward integral*

$$\int_{]0,\cdot]} Y_s\, d^- A_s,$$

when it exists, is a predictable process.

Proof. Since A is a càdlàg process, $A(t)=A(t+)$, and it follows from decomposition (5.14) that the process $I^{-ucp}(\varepsilon,t,f,dg)$ is predictable. By definition, the u.c.p stochastic integral, when it exists, is the u.c.p. limit of $I^{-ucp}(\varepsilon,t,f,dg)$ and it defines in particular a càdlàg process. Since the u.c.p. convergence preserves the predictability, the claim follows. □

Definition 5.3.4. Let X,Y be two càdlàg processes. Suppose the existence of a process $(\Gamma(t))_{t\geq 0}$ such that $[X,Y]^{ucp}_\varepsilon(t)$ converges u.c.p. to $(\Gamma(t))_{t\geq 0}$, namely

$$\lim_{\varepsilon\to 0}\mathbb{P}\left(\sup_{0\leq s\leq t}|[X,Y]^{ucp}_\varepsilon(t)-\Gamma(t)|>\alpha\right)=0 \qquad \text{for every } \alpha>0,$$

Then we will set $[X,Y]_t:=\Gamma(t)$. That process will be called *the covariation between X and Y*. In that case we say that *the covariation between X and Y exists*, and we symbolize it again by $[X,Y]$, if the sequence $[X,Y]^{ucp}_\varepsilon(t)$ converges u.c.p. to some process $(\Gamma(t))_{t\geq 0}$, namely

$$\lim_{\varepsilon\to 0}\mathbb{P}\left(\sup_{0\leq s\leq t}|[X,Y]^{ucp}_\varepsilon(t)-\Gamma(t)|>\alpha\right)=0 \qquad \text{for every } \alpha>0,$$

and in this case $[X,Y]_t:=\Gamma(t)$.

Definition 5.3.5. We say that a pair of càdlàg processes (X,Y) *admits all its mutual brackets* if $[X,X]$, $[X,Y]$, $[Y,Y]$ exist.

Definition 5.3.6. We say that a càdlàg process X *is finite quadratic variation* if $[X,X]$ exists.

Remark 5.3.7. Let X, Y be two càdlàg processes.

(1) By definition $[X,Y]$ is necessarily a càdlàg process.

(2) $[X,X]$ is an increasing process.

(3) $[X,X]^c$ denotes the continuous part of $[X,X]$.

Forward integrals and covariations generalize Itô integrals and the classical square brackets of semimartingales.

Proposition 5.3.8. *Let X,Y be two càdlàg semimartingales, M^1,M^2 two càdlàg local martingales, H,K two càdlàg adapted process. Then*

(i) $[X,Y]$ *exists and it is the usual bracket.*

(ii) $\int_{]0,\,\cdot]} H\, d^- X$ *is the usual stochastic integral* $\int_{]0,\,\cdot]} H_{s-} dX_s$.

(iii) $\left[\int_0^\cdot H_{s-}\, dM_s^1,\, \int_0^\cdot K_{s-}\, dM_s^2\right]$ *is the usual bracket and equals the process* $\int_0^\cdot H_{s-}\, K_{s-}\, d[M^1, M^2]_s$.

Proof. Items (i) and (ii) are consequences of Proposition 1.1 in [**119**] and Remark 5.3.2. Item (iii) follows from (i) and the corresponding properties for classical brackets of local martingales, see Theorem 29, Chapter 2 of [**110**]. □

Lemma 5.3.9. *Suppose that X is a càdlàg, finite quadratic variation process. Then*

(i) $\forall s \in [0,\, T],\ \ \Delta[X, X]_s = (\Delta X_s)^2$;

(ii) $[X, X]_s = [X, X]_s^c + \sum_{t \leq s} (\Delta X_t)^2 \quad \forall s \in [0,\, T]$, a.s.
In particular $\sum_{s \leq T} |\Delta X_s|^2 < \infty$ a.s.

Remark 5.3.10. Condition (5.4) holds for instance in the case of processes X of finite quadratic variation.

Proof. (i) Since X has finite quadratic variation, $[X, X]_\varepsilon^{ucp}$ converges u.c.p. to $[X, X]$. This implies the existence of a sequence (ε_n) such that $[X, X]_{\varepsilon_n}^{ucp}$ converges uniformly a.s. to $[X, X]$. We fix a realization ω outside a suitable null set, which will be omitted in the sequel. Let $\gamma > 0$. There is ε_0 such that

$$\varepsilon_n < \varepsilon_0 \Rightarrow |[X, X]_s - [X, X]_{\varepsilon_n}^{ucp}(s)| \leq \gamma, \quad \forall s \in [0,\, T]. \tag{5.16}$$

We fix $s \in]0,\, T]$. Let $\varepsilon_n < \varepsilon_0$. For every $\delta \in [0,\, s[$, we have

$$|[X, X]_s - [X, X]_{\varepsilon_n}^{ucp}(s - \delta)| \leq \gamma. \tag{5.17}$$

We need to show that the quantity

$$|[X, X]_s - [X, X]_{s-\delta} - (\Delta X_s)^2| \tag{5.18}$$

goes to zero, when $\delta \to 0$. For $\varepsilon := \varepsilon_n < \varepsilon_0$, (5.18) this is smaller or equal than

$$\begin{aligned}
&2\gamma + |[X, X]_\varepsilon^{ucp}(s) - [X, X]_\varepsilon^{ucp}(s - \delta) - (\Delta X_s)^2| \\
&= 2\gamma + \left| \frac{1}{\varepsilon} \int_{s-\varepsilon-\delta}^{s} (X_{(t+\varepsilon)\wedge s} - X_t)^2\, dt - \frac{1}{\varepsilon} \int_{s-\varepsilon-\delta}^{s-\delta} (X_{s-\delta} - X_t)^2\, dt - (\Delta X_s)^2 \right| \\
&\leq 2\gamma + \frac{1}{\varepsilon} \int_{s-\varepsilon-\delta}^{s-\delta} (X_{s-\delta} - X_t)^2\, dt + |I(\varepsilon,\, \delta,\, s)|, \quad \forall \delta \in [0,\, s[,
\end{aligned}$$

where

$$I(\varepsilon,\, \delta,\, s) = \frac{1}{\varepsilon} \int_{s-\varepsilon-\delta}^{s-\varepsilon} (X_{t+\varepsilon} - X_t)^2\, dt + \frac{1}{\varepsilon} \int_{s-\varepsilon}^{s} [(X_s - X_t)^2 - (\Delta X_s)^2]\, dt.$$

At this point, we have, , $\forall s \in [0,\, T]$,

$$|[X, X]_s - [X, X]_{s-\delta} - (\Delta X_s)^2| \leq 2\gamma + \frac{1}{\varepsilon} \int_{s-\varepsilon-\delta}^{s-\delta} (X_{s-\delta} - X_t)^2\, dt + |I(\varepsilon,\, \delta,\, s)|.$$

We take the $\limsup_{\delta \to 0}$ on both sides to get, since X is left continuous at s,

$$|\Delta[X, X]_s - (\Delta X_s)^2| \leq 2\gamma + \frac{1}{\varepsilon} \int_{s-\varepsilon}^{s} (X_{s-} - X_t)^2\, dt + \frac{1}{\varepsilon} \int_{s-\varepsilon}^{s} |(X_s - X_t)^2 - (\Delta X_s)^2|\, dt,$$

for $\varepsilon := \varepsilon_n < \varepsilon_0$. We take the limit when $n \to \infty$ and we get

$$|\Delta[X,X]_s - (\Delta X_s)^2| \le 2\gamma,$$

and this concludes the proof of (i).

(ii) We still work fixing a priori a realization ω. Set $Y_s = [X,X]_s,\ \ s \in [0,\, T]$. Since Y is an increasing càdlàg process, it can be decomposed as

$$Y_s = Y_s^c + \sum_{t \le s} \Delta Y_t, \quad \forall s \in [0,\, T], \quad \text{a.s.}$$

and the result follows from point (i). In particular, setting $s = T$, we get

$$\text{a.s. } \infty > [X,X]_T = [X,X]_T^c + \sum_{s \le T} (\Delta X_s)^2 \ge \sum_{s \le T} (\Delta X_s)^2.$$

□

We now state and prove some fundamental preliminary results, that we will deeply use in the sequel.

Lemma 5.3.11. *Let Y_t be a càdlàg function with values in $\mathbb{R}^n$. Let $\phi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be an equicontinuous function on each compact, such that $\phi(y,y) = 0$ for every $y \in \mathbb{R}^n$. Let $0 \le t_1 \le t_2 \le ... \le t_N \le T$. We have*

$$\sum_{i=1}^N \frac{1}{\varepsilon} \int_{t_i - \varepsilon}^{t_i} 1_{]0,\,s]}(t)\, \phi(Y_{(t+\varepsilon)\wedge s}, Y_t)\, dt \overset{\varepsilon \to 0}{\longrightarrow} \sum_{i=1}^N 1_{]0,\,s]}(t_i)\, \phi(Y_{t_i}, Y_{t_i-}), \tag{5.19}$$

uniformly in $s \in [0,\, T]$.

Proof. Without restriction of generality, we consider the case $n = 1$. Let us fix $\gamma > 0$. Taking into account that ϕ is equicontinuous on compacts, by definition of left and right limits, there exists $\delta > 0$ such that, for every $i \in \{1, ..., N\}$,

$$\ell < t_i,\ u > t_i,\ |\ell - t_i| \le \delta,\ |u - t_i| \le \delta \Rightarrow |\phi(Y_u, Y_\ell) - \phi(Y_{t_i}, Y_{t_i-})| < \gamma, \tag{5.20}$$

$$\begin{aligned}\ell_2 < \ell_1 < t_i,\ |\ell_1 - t_i| \le \delta,\ |\ell_2 - t_i| \le \delta \Rightarrow & |\phi(Y_{\ell_1}, Y_{\ell_2})| \\ & = |\phi(Y_{\ell_1}, Y_{\ell_2}) - \phi(Y_{t_i-}, Y_{t_i-})| < \gamma. \end{aligned} \tag{5.21}$$

Since the sum in (5.19) is finite, it is enough to show the uniform convergence in s of the integrals on $]t_i - \varepsilon,\, t_i]$, for a fixed $t_i \in [0,\, T]$, namely that

$$I(\varepsilon, s) := \frac{1}{\varepsilon} \int_{t_i - \varepsilon}^{t_i} 1_{]0,\,s]}(t)\, \phi(Y_{(t+\varepsilon)\wedge s}, Y_t)\, dt - 1_{]0,\,s]}(t_i)\, \phi(Y_{t_i}, Y_{t_i-}) \tag{5.22}$$

converges to zero uniformly in s, when ε goes to zero. Let thus fix $t_i \in [0,\, T]$, and choose $\varepsilon < \delta$. We distinguish the cases (i), (ii), (iii), (iv) concerning the position of s with respect to t_i.

(i) $s < t_i - \varepsilon$. (5.22) vanishes.

(ii) $s \in [t_i - \varepsilon,\, t_i[$. By (5.21) we get

$$|I(\varepsilon, s)| \le \frac{1}{\varepsilon} \int_{t_i - \varepsilon}^{t_i} |\phi(Y_s, Y_t)|\, dt \le \gamma.$$

(iii) $s \in [t_i, t_i + \varepsilon[$. By (5.20) we get

$$|I(\varepsilon, s)| \leq \frac{1}{\varepsilon} \int_{t_i - \varepsilon}^{t_i} |\phi(Y_{(t+\varepsilon)\wedge s}, Y_t) - \phi(Y_{t_i}, Y_{t_i -})| \, dt \leq \gamma.$$

(iv) $s \geq t_i + \varepsilon$. By (5.20) we get

$$|I(\varepsilon, s)| \leq \frac{1}{\varepsilon} \int_{t_i - \varepsilon}^{t_i} |\phi(Y_{t+\varepsilon}, Y_t) - \phi(Y_{t_i}, Y_{t_i -})| \, dt \leq \gamma.$$

Collecting all the cases above, we see that

$$\limsup_{\varepsilon \to 0} \sup_{s \in [0,\, T]} |I(\varepsilon, s)| \leq \gamma,$$

and letting γ go to zero we get the uniform convergence. □

Lemma 5.3.12. *Let X be a càdlàg (càglàd) real process. Let $\gamma > 0$, t_0, $t_1 \in \mathbb{R}$ and $I = [t_0,\, t_1]$ be a subinterval of $[0,\, T]$ such that*

$$|\Delta X_t|^2 \leq \gamma^2, \quad \forall t \in I. \tag{5.23}$$

Then there is $\varepsilon_0 > 0$ such that

$$\sup_{\substack{a,\, t \in I \\ |a-t| \leq \varepsilon_0}} |X_a - X_t| \leq 3\gamma.$$

Proof. We only treat the càdlàg case, the càglàd one is a consequence of an obvious time reversal argument. Also in this proof a realization ω will be fixed, but omitted. According to Lemma 1, Chapter 3, in [**16**], applied to $[t_0,\, t_1]$ replacing $[0,\, 1]$, there exist points

$$t_0 = s_0 < s_1 < ... < s_{l-1} < s_l = t_1$$

such that for every $j \in \{1, ..., l\}$

$$\sup_{d,\, u \in [s_{j-1},\, s_j[} |X_d - X_u| < \gamma. \tag{5.24}$$

Since X is càdlàg, we can choose ε_0 such that, $\forall j \in \{0,\, ...,\, l-1\}$,

$$|d - s_j| \leq \varepsilon_0 \quad \Rightarrow \quad |X_d - X_{s_j -}| \leq \gamma, \tag{5.25}$$

$$|u - s_j| \leq \varepsilon_0 \quad \Rightarrow \quad |X_u - X_{s_j}| \leq \gamma. \tag{5.26}$$

Let $t \in [s_{j-1},\, s_j[$ for some j and a such that $|t-a| \leq \varepsilon$ for $\varepsilon < \varepsilon_0$. Without restriction of generality we can take $t < a$. There are two cases.

(i) $a,\, t \in [s_{j-1},\, s_j[$. In this case, (5.24) gives

$$|X_a - X_t| < \gamma.$$

(ii) $s_{j-1} \leq t < s_j \leq a$. Then,

$$|X_a - X_t| \leq |X_a - X_{s_j}| + |X_{s_j} - X_{s_j -}| + |X_{s_j -} - X_t| \leq 3\gamma,$$

where the first absolute value is bounded by (5.26), the second by (5.23) and the third by (5.25).

□

Remark 5.3.13. Let $I = [t_0, t_1] \subset [0, T]$, let $\varepsilon > 0$. Let $t \in]t_0, t_1 - \varepsilon]$ and $s > t$. We will apply Lemma 5.3.12 to the couple (a, t), where $a = (t+\varepsilon) \wedge s$. Indeed $a \in I$ because $a \le t + \varepsilon \le t_1$.

Proposition 5.3.14. *Let (Z_t) be a càdlàg process, (V_t) be a bounded variation process. Then $[Z, V]_s$ exists and equals*

$$\sum_{t \le s} \Delta Z_t \, \Delta V_t, \quad \forall s \in [0, T].$$

In particular, V is a finite quadratic variation process.

Proof. We need to prove the u.c.p convergence to zero of

$$\frac{1}{\varepsilon} \int_{]0,\, s]} (Z_{(t+\varepsilon)\wedge s} - Z_t)(V_{(t+\varepsilon)\wedge s} - V_t)\, dt - \sum_{t \le s} \Delta Z_t \, \Delta V_t. \tag{5.27}$$

As usual the realization $\omega \in \Omega$ will be fixed, but often omitted. Let (t_i) be the enumeration of all the jumps of $Z(\omega)$ in $[0, T]$. We have

$$\lim_{i \to \infty} |\Delta Z_{t_i}(\omega)| = 0.$$

Indeed, if it were not the case, it would exists $a > 0$ and a subsequence (t_{i_l}) of (t_i) such that $|\Delta Z_{t_{i_l}}| \ge a$. This is not possible since a càdlàg function admits at most a finite number of jumps exceeding any $a > 0$, see considerations below Lemma 1, Chapter 2 of [**16**].

At this point, let $\gamma > 0$ and $N = N(\gamma)$ such that

$$n \ge N, \quad |\Delta Z_{t_n}| \le \gamma. \tag{5.28}$$

We introduce

$$A(\varepsilon, N) = \bigcup_{i=1}^{N}]t_i - \varepsilon, t_i], \quad B(\varepsilon, N) = \bigcup_{i=1}^{N}]t_{i-1}, t_i - \varepsilon], \tag{5.29}$$

and we decompose (5.27) into

$$I_A(\varepsilon, N, s) + I_{B1}(\varepsilon, N, s) + I_{B2}(\varepsilon, N, s) \tag{5.30}$$

where

$$\begin{aligned}
I_A(\varepsilon, N, s) &= \frac{1}{\varepsilon} \int_{]0,\, s] \cap A(\varepsilon, N)} (Z_{(t+\varepsilon)\wedge s} - Z_t)(V_{(t+\varepsilon)\wedge s} - V_t)\, dt \\
&\quad - \sum_{i=1}^{N} 1_{]0,\, s[}(t_i)\, \Delta Z_{t_i}\, \Delta V_{t_i}, \\
I_{B1}(\varepsilon, N, s) &= \frac{1}{\varepsilon} \int_{]0,\, s] \cap B(\varepsilon, N)} (Z_{(t+\varepsilon)\wedge s} - Z_t)((V_{(t+\varepsilon)\wedge s} - V_t)\, dt, \\
I_{B2}(N, s) &= - \sum_{i=N+1}^{\infty} 1_{]0,\, s[}(t_i)\, \Delta Z_{t_i}\, \Delta V_{t_i}.
\end{aligned}$$

Applying Lemma 5.3.11 to $Y = (Y^1, Y^2) = (Z, V)$ and $\phi(y_1, y_2) = (y_1^1 - y_2^1)(y_1^2 - y_2^2)$ we get

$$I_A(\varepsilon, N, s) \underset{\varepsilon \to 0}{\to} 0,$$

uniformly in s. On the other hand, for $t \in]t_{i-1},\, t_i - \varepsilon[$ and $s > t$, by Remark 5.3.13 we know that $(t+\varepsilon) \wedge s \in [t_{i-1},\, t_i]$. Therefore Lemma 5.3.12 with $X = Z$, applied successively to the intervals $I = [t_{i-1},\, t_i]$ implies that

$$\begin{aligned}
|I_{B1}(\varepsilon, N, s)| &= \frac{1}{\varepsilon} \int_{]0,\, s] \cap B(\varepsilon, N)} |Z_{(t+\varepsilon)\wedge s} - Z_t||V_{(t+\varepsilon)\wedge s} - V_t|\, dt \\
&\leq 3\,\gamma\, \frac{1}{\varepsilon} \int_{]0,\, s] \cap B(\varepsilon, N)} |V_{(t+\varepsilon)\wedge s} - V_t|\, dt \\
&\leq 3\,\gamma \int_{]0,\, s]} |V_{(t+\varepsilon)\wedge s} - V_t|\, \frac{dt}{\varepsilon} \\
&= 3\,\gamma \int_{]0,\, s]} \frac{dt}{\varepsilon} \int_{]t,\, (t+\varepsilon)\wedge s]} d\|V\|_r \\
&= 3\,\gamma \int_{]0,\, s]} d\|V\|_r \int_{[(r-\varepsilon)^+,\, r[} \frac{dt}{\varepsilon} \\
&\leq 3\,\gamma\, ||V||_T,
\end{aligned}$$

where $r \mapsto \|V\|_r$ denotes the total variation function of V. Finally, concerning $I_{B2}(N, s)$, by (5.28) we have

$$|I_{B2}(N, s)| \leq \gamma \sum_{i=N+1}^{\infty} 1_{]0,\, s[}(t_i)\; |\Delta V_{t_i}| \leq \gamma\, ||V||_T.$$

Therefore, collecting the previous estimations we get

$$\limsup_{\varepsilon \to 0} \sup_{s \in [0,\, T]} |I_A(\varepsilon, N, s) + I_{B1}(\varepsilon, N, s) + I_{B2}(N, s)| \leq 4\,\gamma\, ||V||_T,$$

and we conclude by the arbitrariness of $\gamma > 0$. □

Finally we give a generalization of Dini type lemma in the càdlàg case.

Lemma 5.3.15. *Let $(G_n,\, n \in \mathbb{N})$ be a sequence of continuous increasing functions, let G (resp. F) from $[0,\, T]$ to $\mathbb{R}$ be a càdlàg (resp. continuous) function. We set $F_n = G_n + G$ and suppose that $F_n \to F$ pointwise. Then*

$$\limsup_{n \to \infty} \sup_{s \in [0,\, T]} |F_n(s) - F(s)| \leq 2 \sup_{s \in [0,\, T]} |G(s)|.$$

Proof. Let $0 = t_0 < t_1 < ... < t_m = T$ such that $t_i = \frac{i}{m}, \quad i = 0, ..., m$. Let $\gamma > 0$. Let us fix $m \in \mathbb{N}$ such that $\delta\left(F,\, \frac{1}{m}\right) \leq \gamma$, where $\rho(F, \cdot)$ denotes the modulus of continuity of F. If $s \in [t_i,\, t_{i+1}]$, $0 \leq i \leq m-1$, we have

$$F_n(s) - F(s) \leq F_n(t_{i+1}) - F(s) + G(s) - G(t_{i+1}). \tag{5.31}$$

Now

$$F_n(t_{i+1}) - F(s) \leq F_n(t_{i+1}) - F(t_{i+1}) + F(t_{i+1}) - F(s)$$

$$\leq \delta\left(F, \frac{1}{m}\right) + F_n(t_{i+1}) - F(t_{i+1}). \tag{5.32}$$

From (5.31) and (5.32) it follows

$$\begin{aligned} F_n(s) - F(s) &\leq F_n(t_{i+1}) - F(t_{i+1}) + G(s) - G(t_{i+1}) + \delta\left(F, \frac{1}{m}\right) \\ &\leq 2||G||_\infty + \delta\left(F, \frac{1}{m}\right) + |F_n(t_{i+1}) - F(t_{i+1})|, \end{aligned} \tag{5.33}$$

where $||G||_\infty = \sup_{s\in[0,\,T]} |G(s)|$. Similarly,

$$F(s) - F_n(s) \geq -2||G||_\infty - \delta\left(F, \frac{1}{m}\right) - |F_n(t_i) - F(t_i)|. \tag{5.34}$$

So, collecting (5.33) and (5.34) we have $\forall s \in [t_i,\, t_{i+1}]$

$$|F_n(s) - F(s)| \leq 2||G||_\infty + \delta\left(F, \frac{1}{m}\right) + |F_n(t_i) - F(t_i)| + |F_n(t_{i+1}) - F(t_{i+1})|.$$

Consequently,

$$\sup_{s\in[0,\,T]} |F_n(s) - F(s)| \leq 2||G||_\infty + \delta\left(F, \frac{1}{m}\right) + \sum_{i=1}^{m} |F_n(t_i) - F(t_i)|. \tag{5.35}$$

Recalling that $F_n \to F$ pointwise, taking the lim sup in (5.35) we get

$$\limsup_{n\to\infty} \sup_{s\in[0,\,T]} |F_n(s) - F(s)| \leq 2||G||_\infty + \delta\left(F, \frac{1}{m}\right).$$

Since F is uniformly continuous and m is arbitrarily big, the result follows. □

5.4. Additional results on calculus via regularization

For every functions f, g defined on $\mathbb{R}$, let now set

$$\tilde{I}^-(\varepsilon, t, f, dg) = \int_{]0,\,t]} f(s)\, \frac{g(s+\varepsilon) - g(s)}{\varepsilon}\, ds, \tag{5.36}$$

$$C_\varepsilon(f,g)(t) = \frac{1}{\varepsilon} \int_{]0,\,t]} (f(s+\varepsilon) - f(s))(g(s+\varepsilon) - g(s))\, ds. \tag{5.37}$$

Definition 5.4.1. Assume that X, Y are two càdlàg processes. We say that *the forward integral of Y with respect to X exists in the pathwise sense*, if there exists some process $(I(t), t \geq 0)$ such that, for all subsequences (ε_n), there is a subsequence (ε_{n_k}) and a null set $\mathcal{N}$ with

$$\forall \omega \notin \mathcal{N}, \qquad \lim_{k\to\infty} |\tilde{I}^-(\varepsilon_{n_k}, t, Y, dX)(\omega) - I(t)(\omega)| = 0 \qquad \forall t \geq 0, \text{ a.s.}$$

Definition 5.4.2. Let X, Y be two càdlàg processes. *the covariation between X and Y (the quadratic variation of X) exists in the pathwise sense*, if there exists a càdlàg process $(\Gamma(t), t \geq 0)$ such that, for all subsequences (ε_n) there is a subsequence (ε_{n_k}) and a null set N:

$$\forall \omega \notin N, \qquad \lim_{k\to\infty} |C_{\varepsilon_{n_k}}(X,Y)(t)(\omega) - \Gamma(t)(\omega)| = 0 \qquad \forall t \geq 0, \text{ a.s.}$$

Proposition 5.4.3. *Let X, Y be two càdlàg processes. Then*

$$I^{-ucp}(\varepsilon, t, Y, dX) = \tilde{I}^{-}(\varepsilon, t, Y, dX) + R_1(\varepsilon, t) \tag{5.38}$$

$$[X, Y]^{ucp}_{\varepsilon}(t) = C_{\varepsilon}(X, Y)(t) + R_2(\varepsilon, t), \tag{5.39}$$

where

$$R_i(\varepsilon, t)(\omega) \underset{\varepsilon \to 0}{\longrightarrow} 0 \quad i = 1, 2,\ \forall t \in [0, T],\ \forall \omega \in \Omega. \tag{5.40}$$

Moreover, if X is continuous, then the convergence in (5.40) *holds u.c.p.*

Proof. We fix $t \in [0, T]$. Let $\gamma > 0$. The definition of right continuity in t insures that there exists $\delta > 0$ small enough such that

$$|X(t) - X(a)| \leq \gamma \quad \text{if } \ a - t < \delta,\ a > t,$$
$$|Y(t) - Y(a)| \leq \gamma \quad \text{if } \ a - t < \delta,\ a > t.$$

We start proving (5.38). From decomposition (5.14) and the definition of $\tilde{I}^{-}(\varepsilon, t, Y, dX)$ we get

$$\begin{aligned} I^{-ucp}(\varepsilon, t, Y, dX) - \tilde{I}^{-}(\varepsilon, t, Y, dX) &= \frac{1}{\varepsilon}\int_{(t-\varepsilon)_+}^{t} Y(s)\,[X(t) - X(s)]\,ds \\ &\quad -\frac{1}{\varepsilon}\int_{(t-\varepsilon)_+}^{t} Y(s)\,[X(s+\varepsilon) - X(s)]\,ds \\ &= \frac{1}{\varepsilon}\int_{(t-\varepsilon)_+}^{t} Y(s)\,[X(t) - X(s+\varepsilon)]\,ds =: R_1(\varepsilon, t). \end{aligned}$$

Choosing $\varepsilon < \delta$ we get

$$|R_1(\varepsilon, t)| \leq \gamma\, ||Y||_{\infty},$$

and since γ is arbitrary, we conclude that $R_1(\varepsilon, t) \to 0$ as ε goes to zero, for every $t \in [0, T]$.

It remains to show (5.39). To this end we evaluate

$$\begin{aligned} [X, Y]^{ucp}_{\varepsilon}(t) - C_{\varepsilon}(X, Y)(t) &= \frac{1}{\varepsilon}\int_{(t-\varepsilon)_+}^{t} [X(t) - X(s)]\,[Y(t) - Y(s)]\,ds \\ &\quad - \frac{1}{\varepsilon}\int_{(t-\varepsilon)_+}^{t} [X(s+\varepsilon) - X(s)]\,[Y(s+\varepsilon) - Y(s)]\,ds \\ &=: R_2(\varepsilon, t). \end{aligned}$$

We have

$$\begin{aligned} R_2(\varepsilon, t) &= \frac{1}{\varepsilon}\int_{(t-\varepsilon)_+}^{t} [X(t) - X(s)]\,[Y(t) - Y(s)]\,ds \\ &\quad - \frac{1}{\varepsilon}\int_{(t-\varepsilon)_+}^{t} [X(s+\varepsilon) - X(s)]\,[Y(t) - Y(s)]\,ds \\ &\quad + \frac{1}{\varepsilon}\int_{(t-\varepsilon)_+}^{t} [X(s+\varepsilon) - X(s)]\,[Y(t) - Y(s)]\,ds \\ &\quad - \frac{1}{\varepsilon}\int_{(t-\varepsilon)_+}^{t} [X(s+\varepsilon) - X(s)]\,[Y(s+\varepsilon) - Y(s)]\,ds \end{aligned}$$

$$= \frac{1}{\varepsilon} \int_{(t-\varepsilon)_+}^{t} [X(t) - X(s+\varepsilon)]\,[Y(t) - Y(s)]\, ds$$
$$+ \frac{1}{\varepsilon} \int_{(t-\varepsilon)_+}^{t} [X(s+\varepsilon) - X(s)]\,[Y(t) - Y(s+\varepsilon)]\, ds.$$

Choosing $\varepsilon < \delta$, the absolute value of previous expression is smaller than

$$2\,\gamma\,(||Y||_\infty + ||X||_\infty).$$

Since γ is arbitrary, $R_2(\varepsilon, t) \to 0$ as ε goes to zero, for every $t \in [0,\, T]$.
Suppose now that X is continuous. The expression of $R_2(\varepsilon, t)$ can be uniformly (in t) bounded by $2\rho(X, \varepsilon)\, \|Y\|_\infty$, where $\rho(X, \cdot)$ denotes the modulus of continuity of X; on the other hand $R_1(\varepsilon, t) \le 2\rho(X, \varepsilon)\, \|Y\|_\infty, \forall t \in [0, T]$. This concludes the proof of Proposition 5.4.3. □

Corollary 5.4.4. *Let X, Y be two càdlàg processes.*

1) *If the stochastic integral of Y with respect to X exists, then it exists in the pathwise sense. In particular, there is a null set $\mathcal{N}$ and, for any sequence $(\varepsilon_n) \downarrow 0$, a subsequence (ε_{n_k}) such that*

$$\tilde{I}^-(\varepsilon_{n_k}, t, Y, dX)(\omega) \underset{k\to\infty}{\longrightarrow} \left(\int_{]0,\,t]} Y_s\, d^- X_s \right)(\omega) \quad \forall t \in [0,\, T],\ \forall \omega \notin \mathcal{N}. \tag{5.41}$$

2) *If the covariation between X and Y exists, then it exists in the pathwise sense. In particular, there is a null set $\mathcal{N}$ and, for any sequence $(\varepsilon_n) \downarrow 0$, a subsequence (ε_{n_k}) such that*

$$C_{\varepsilon_{n_k}}(X, Y)(t)(\omega) \underset{k\to\infty}{\longrightarrow} [X, Y]_t\,(\omega) \quad \forall t \in [0,\, T],\ \forall \omega \notin \mathcal{N}. \tag{5.42}$$

Proof. The result is a direct application of Proposition 5.4.3. □

Lemma 5.4.5. *Let $g : [0,\, T] \to \mathbb{R}$ be a càglàd process, X be a càdlàg process such that the quadratic variation of X exists in the pathwise sense, see Definition 5.4.2. Setting (improperly) $[X, X] = \Gamma$, we have*

$$\int_0^s g_t\,(X_{(t+\varepsilon)\wedge s} - X_t)^2\, \frac{dt}{\varepsilon} \overset{\varepsilon\to 0}{\longrightarrow} \int_0^s g_t\, d[X, X]_t \quad \text{u.c.p.} \tag{5.43}$$

Proof. We have to prove that

$$\sup_{s\in[0,\,T]} \left| \int_0^s g_t\,(X_{(t+\varepsilon)\wedge s} - X_t)^2\, \frac{dt}{\varepsilon} - \int_0^s g_t\, d[X, X]_t \right| \overset{\mathbb{P}}{\longrightarrow} 0 \quad \text{as } \varepsilon \text{ goes to zero.} \tag{5.44}$$

Let ε_n be a sequence converging to zero. Since $[X, X]$ exists in the pathwise sense, there is a subsequence ε_{n_k}, that we still symbolize by ε_n, such that

$$C_{\varepsilon_n}(X, X)(t) \overset{n\to\infty}{\longrightarrow} [X, X]_t \quad \forall t \in [0,\, T] \text{ a.s.} \tag{5.45}$$

Let $\mathcal{N}$ be a null set such that

$$C_{\varepsilon_n}(X, X)(\omega, t) \overset{n\to\infty}{\longrightarrow} [X, X]_t(\omega) \quad \forall t \in [0,\, T],\ \forall \omega \notin \mathcal{N}. \tag{5.46}$$

From here on we fix $\omega \notin \mathcal{N}$. We have to prove that

$$\sup_{s\in[0,T]}\left|\int_0^s g_t\,(X_{(t+\varepsilon_n)\wedge s}-X_t)^2\,\frac{dt}{\varepsilon_n}-\int_0^s g_t\,d[X,X]_t\right| \overset{n\to\infty}{\longrightarrow} 0. \tag{5.47}$$

We will do it in two steps.

Step 1. We consider first the case of a càglàd process (g_t) with a finite number of jumps.

Let us fix $\gamma > 0$, $\varepsilon > 0$. We enumerate by $(t_i)_{i\geq 0}$ the set of jumps of $X(\omega)$ on $[0,\,T]$, union $\{T\}$. Without restriction of generality, we will assume that the jumps of (g_t) are included in $\{t_i\}_{i\geq 0}$. Let $N = N(\omega)$ such that

$$\sum_{i=N+1}^{\infty}|\Delta X_{t_i}|^2 \leq \gamma^2, \qquad \sum_{i=N+1}^{\infty}|\Delta g_{t_i}| = 0. \tag{5.48}$$

We define

$$\begin{aligned} A(\varepsilon,N) &= \bigcup_{i=1}^{N}]t_i-\varepsilon,\,t_i] \\ B(\varepsilon,N) &= [0,\,T]\setminus A(\varepsilon,N). \end{aligned}$$

The term inside the supremum in (5.44) can be written as

$$\frac{1}{\varepsilon}\int_{]0,\,s]} g_t\,(X_{(t+\varepsilon)\wedge s}-X_t)^2\,dt-\int_{]0,\,s]} g_t\,d[X,X]_t = J_1(s,\,\varepsilon)+J_2(s,\,\varepsilon)+J_3(s,\,\varepsilon),$$

where

$$\begin{aligned} J_1(\varepsilon,\,N,\,s) &= \frac{1}{\varepsilon}\int_{]0,\,s]\cap A(\varepsilon,N)} g_t\,(X_{(t+\varepsilon)\wedge s}-X_t)^2\,dt-\sum_{i=1}^{N}1_{]0,\,s]}(t_i)\,(\Delta X_{t_i})^2\,g_{t_i}, \\ J_2(\varepsilon,\,N,\,s) &= \frac{1}{\varepsilon}\int_{]0,\,s]\cap B(\varepsilon,N)} g_t\,(X_{t+\varepsilon}-X_t)^2\,dt \\ &\quad -\int_{]0,\,s]} g_t\,d[X,X]^c_t-\sum_{i=N+1}^{\infty}1_{]0,\,s]}(t_i)\,(\Delta X_{t_i})^2\,g_{t_i}, \\ J_3(\varepsilon,\,N,\,s) &= \frac{1}{\varepsilon}\int_{]0,\,s]\cap B(\varepsilon,N)} g_t\,\left[(X_{(t+\varepsilon)\wedge s}-X_t)^2-(X_{t+\varepsilon}-X_t)^2\right]\,dt. \end{aligned}$$

Applying Lemma 5.3.11 to $J_1(\varepsilon,\,N,\,s)$, with $Y=(Y^1,Y^2)=(t,X)$ and $\phi(y_1,y_2)=g_{y_2^1}(y_1^2-y_2^2)^2$, we get

$$\lim_{\varepsilon\to 0}\sup_{s\in[0,T]}|J_1(\varepsilon,\,N,\,s)|=0. \tag{5.49}$$

Concerning $J_3(\varepsilon,N,s)$, we have

$$\begin{aligned} &|J_3(\varepsilon,N,s)| \\ &=\left|\int_0^s g_t\,1_{B(\varepsilon,N)}(t)\,(X_{t+\varepsilon}-X_t)^2\,\frac{dt}{\varepsilon}-\int_0^s g_t\,1_{B(\varepsilon,N)}(t)\,(X_{(t+\varepsilon)\wedge s}-X_t)^2\,\frac{dt}{\varepsilon}\right| \\ &\leq \frac{||g||_\infty}{\varepsilon}\left(\int_{s-\varepsilon}^{s}1_{B(\varepsilon,N)}(t)\,(|X_{t+\varepsilon}-X_t|^2+|X_s-X_t|^2)\,\frac{dt}{\varepsilon}\right). \end{aligned}$$

We recall that

$$B(\varepsilon, N) = \bigcup_{i=1}^{N}]t_{i-1}, t_i - \varepsilon].$$

From Remark 5.3.13 it follows that, for every $t \in]t_{i-1}, t_i - \varepsilon]$ and $s > t$, $(t+\varepsilon) \wedge s \in [t_{i-1}, t_i]$. Therefore Lemma 5.3.12 applied successively to the intervals $[t_{i-1}, t_i]$ implies that

$$\limsup_{\varepsilon \to 0} \sup_{s \in [0,T]} |J_3(\varepsilon, N, s)| \le 18\gamma^2 \, ||g||_\infty. \tag{5.50}$$

It remains to evaluate the uniform limit of $J_2(\varepsilon_n, N, s)$. We start by showing that, for fixed $s \in [0, T]$, we have the pointwise convergence

$$\begin{aligned} J_2(\varepsilon_n, N, s) &= \frac{1}{\varepsilon_n} \int_{]0,\, s] \cap B(\varepsilon_n, N)} g_t \, (X_{t+\varepsilon_n} - X_t)^2 \, dt \\ &\quad - \int_{]0,\, s]} g_t \, d[X,X]^c_t - \sum_{i=N+1}^{\infty} 1_{]0,\, s]}(t_i) \, (\Delta X_{t_i})^2 \, g_{t_i} \\ &\underset{n\to\infty}{\to} 0, \quad \forall s \in [0,\, T]. \end{aligned} \tag{5.51}$$

We prove now that

$$\frac{dt}{\varepsilon_n} 1_{B(\varepsilon_n, N)}(t) \, (X_{t+\varepsilon_n} - X_t)^2 \Rightarrow d\Bigg(\sum_{\substack{t_i \le t \\ i=N+1}}^{\infty} (\Delta X_{t_i})^2 + [X,X]^c_t \Bigg). \tag{5.52}$$

It will be enough to show that, $\forall s \in [0, T]$,

$$\int_0^s \frac{dt}{\varepsilon_n} 1_{B(\varepsilon_n, N)}(t) \, (X_{t+\varepsilon_n} - X_t)^2 \to_{n\to\infty} \sum_{\substack{t_i \le s \\ i=N+1}}^{\infty} (\Delta X_{t_i})^2 + [X,X]^c_s. \tag{5.53}$$

By (5.45) and Lemma 5.3.9, we have

$$\int_0^s (X_{t+\varepsilon_n} - X_t)^2 \, \frac{dt}{\varepsilon_n} \overset{n\to\infty}{\longrightarrow} [X,X]^c_s + \sum_{t_i \le s} (\Delta X_{t_i})^2 \quad \forall s \in [0,\, T]. \tag{5.54}$$

On the other hand, we can show that

$$\int_0^s \frac{dt}{\varepsilon_n} 1_{A(\varepsilon_n, N)}(t) \, (X_{t+\varepsilon_n} - X_t)^2 \overset{n\to\infty}{\longrightarrow} \sum_{\substack{t_i \le s \\ i=1}}^{N} (\Delta X_{t_i})^2 \quad \forall s \in [0,\, T]. \tag{5.55}$$

Indeed

$$\begin{aligned} &\Big| \int_0^s \frac{dt}{\varepsilon_n} 1_{A(\varepsilon_n, N)}(t) \, (X_{t+\varepsilon_n} - X_t)^2 - \sum_{\substack{t_i \le s \\ i=1}}^{N} (\Delta X_{t_i})^2 \Big| \\ &\le \Big| \int_0^s \frac{dt}{\varepsilon_n} 1_{A(\varepsilon_n, N)}(t) \, (X_{(t+\varepsilon_n)\wedge s} - X_t)^2 - \sum_{\substack{t_i \le s \\ i=1}}^{N} (\Delta X_{t_i})^2 \Big| \\ &+ \Big| \int_0^s \frac{dt}{\varepsilon_n} 1_{A(\varepsilon_n, N)}(t) \, (X_{(t+\varepsilon_n)\wedge s} - X_t)^2 - \int_0^s \frac{dt}{\varepsilon_n} 1_{A(\varepsilon_n, N)}(t) \, (X_{t+\varepsilon_n} - X_t)^2 \Big| \end{aligned}$$

for all $s \in [0, T]$. The first addend converges to zero by Lemma 5.3.11 applied to $Y = X$ and $\phi(y) = (y_1 - y_2)^2$. The second one converges to zero by similar arguments as those we have used to prove Proposition 5.4.3. This establishes (5.55). Subtracting (5.54) and (5.55), we get (5.53), and so (5.52).

We remark that the left-hand side of (5.52) are positive measures. Moreover, we notice that $t \mapsto g_t(\omega)$ is μ-continuous, where μ is the measure on the right-hand side of (5.52). At this point, Portmanteau theorem and (5.52) insure that $J_2(\varepsilon_n, N, s)$ converges to zero as n goes to infinity, for every $s \in [0, T]$.

Finally, we control the convergence of $J_2(\varepsilon_n, N, s)$, uniformly in s. We make use of Lemma 5.3.15. We set

$$\begin{aligned}
G_n(s) &= \frac{1}{\varepsilon_n} \int_{]0,\, s]} 1_{B(\varepsilon_n, N)}(t)\, (X_{t+\varepsilon_n} - X_t)^2\, g_t\, dt, \\
F(s) &= \int_{]0,\, s]} g_t\, d[X, X]^c_t, \\
G(s) &= -\sum_{i=N+1}^{\infty} 1_{]0,\, s]}(t_i)\, (\Delta X_{t_i})^2\, g_{t_i}.
\end{aligned}$$

By (5.51), $F_n := G_n + G$ converges pointwise to F as n goes to infinity. Since G_n is continuous and increasing, F is continuous and G is càdlàg, Lemma 5.3.15 implies that

$$\limsup_{n \to \infty} \sup_{s \in [0,\, T]} |J_2(\varepsilon_n,\, N,\, s)| \leq 2\gamma^2\, ||g||_\infty. \tag{5.56}$$

Collecting (5.49), (5.50) and (5.56), it follows that

$$\limsup_{n \to \infty} \sup_{s \in [0,\, T]} \left| \int_0^s g_t\, (X_{(t+\varepsilon_n)\wedge s} - X_t)^2\, \frac{dt}{\varepsilon_n} - \int_0^s g_t\, d[X, X]_t \right| \leq 20\gamma^2\, ||g||_\infty.$$

Since γ is arbitrarily small, (5.47) follows.

Step 2. We treat now the case of a general càglàd process (g_t).

Let us fix $\gamma > 0$, $\varepsilon > 0$. Without restriction of generality, we can write $g_t = g_t^{\gamma, BV} + g_t^{\gamma}$, where $g_t^{\gamma, BV}$ is a process with a finite number of jumps and g_t^{γ} is such that $|\Delta g_t^{\gamma}| \leq \gamma$ for every $t \in [0, T]$. From Step 1, we have

$$I_s^{1,n} := \int_0^s g_t^{\gamma, BV}\, (X_{(t+\varepsilon_n)\wedge s} - X_t)^2\, \frac{dt}{\varepsilon_n} - \int_0^s g_t^{\gamma, BV}\, d[X, X]_t \tag{5.57}$$

converges to zero, uniformly in s, as n goes to infinity. Concerning (g_t^{γ}), by Lemma 5.3.12 we see that there exists $\bar{\varepsilon}_0 = \bar{\varepsilon}_0(\gamma)$ such that

$$\sup_{\substack{a,\, t \in I \\ |a-t| \leq \bar{\varepsilon}_0}} |g_a^{\gamma} - g_t^{\gamma}| \leq 3\gamma. \tag{5.58}$$

At this point, we introduce the càglàd process

$$g_t^{k, \gamma} = \sum_{i=0}^{2^k - 1} g^{\gamma}_{i\, 2^{-k} T}\, 1_{]i 2^{-k} T, (i+1) 2^{-k} T]}(t), \tag{5.59}$$

where k is such that $2^{-k} < \bar{\varepsilon}_0$. From (5.59), taking into account (5.58), we have

$$|g_t^\gamma - g_t^{k,\gamma}| = |g_t^\gamma 1_{]i2^{-k}T,(i+1)2^{-k}T]}(t) - g^\gamma_{i\,2^{-k}}| \le 3\gamma \quad \forall\, t \in [0, T]. \tag{5.60}$$

We set

$$I_s^{2,n} := \int_0^s (g_t^\gamma - g_t^{k,\gamma})\,(X_{(t+\varepsilon_n)\wedge s} - X_t)^2\,\frac{dt}{\varepsilon_n} - \int_0^s (g_t^\gamma - g_t^{k,\gamma})\,d[X,X]_t.$$

From (5.60)

$$\sup_{s\in[0,T]} |I_s^{2,n}| \le 3\gamma\,\Gamma$$

with

$$\Gamma = \sup_{n\in\mathbb{N}, s\in[0,T]} \left| \int_0^s (X_{(t+\varepsilon_n)\wedge s} - X_t)^2\,\frac{dt}{\varepsilon_n} \right| + [X,X]_T. \tag{5.61}$$

Notice that Γ is finite, since the term inside the absolute value in (5.61) converges uniformly by Step 1 with $g = 1$. On the other hand, by definition, $(g_t^{k,\gamma})$ has a finite number of jumps, therefore from Step 1 we get that

$$I_s^{3,n} = \int_0^s g_t^{k,\gamma}\,(X_{(t+\varepsilon_n)\wedge s} - X_t)^2\,\frac{dt}{\varepsilon_n} - \int_0^s g_t^{k,\gamma}\,d[X,X]_t \tag{5.62}$$

converges to zero, uniformly in s, as n goes to infinity. Finally, collecting all the terms, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \sup_{s\in[0,T]} \left| \int_0^s g_t\,(X_{(t+\varepsilon_n)\wedge s} - X_t)^2\,\frac{dt}{\varepsilon_n} - \int_0^s g_t\,d[X,X]_t \right| \\
&\le \limsup_{n\to\infty} \sup_{s\in[0,T]} |I_s^{1,n}| + \limsup_{n\to\infty} \sup_{s\in[0,T]} |I_s^{2,n}| + \limsup_{n\to\infty} \sup_{s\in[0,T]} |I_s^{3,n}| \\
&\le 3\,\gamma\Gamma.
\end{aligned} \tag{5.63}$$

and since γ is arbitrarily small, the result follows. □

Remark 5.4.6. Let X be a càdlàg processes. From Corollary 5.4.4 2) and Lemma 5.4.5 with $g = 1$, the following properties are equivalent:

- X is a finite quadratic variation process;
- $[X,X]$ exists in the pathwise sense.

Proposition 5.4.7. *Let X, Y be two càdlàg processes. The following properties are equivalent.*

(i) $[X,X]$, $[X,Y]$, $[Y,Y]$ *exist in the pathwise sense;*

(ii) *For all $(\varepsilon_n) \downarrow 0$ there is (ε_{n_k}) and a null set $\mathcal{N}$ such that, $\forall \omega \notin \mathcal{N}$,*

$$\begin{aligned}
dC_{\varepsilon_{n_k}}(X,Y)(\omega) &\underset{k\to\infty}{\longrightarrow} d[X,Y](\omega) \quad \text{weakly},\\
dC_{\varepsilon_{n_k}}(X,X)(\omega) &\underset{k\to\infty}{\longrightarrow} d[X,X](\omega) \quad \text{weakly},\\
dC_{\varepsilon_{n_k}}(Y,Y)(\omega) &\underset{k\to\infty}{\longrightarrow} d[Y,Y](\omega) \quad \text{weakly}.
\end{aligned}$$

(iii) *For every càglàd process* (g_t),

$$\lim_{\varepsilon\to 0}\int_0^s g_t\,\frac{(X((t+\varepsilon)\wedge s)-X(t))\,(Y((t+\varepsilon)\wedge s)-Y(t))}{\varepsilon}\,dt$$
$$=\int_0^s g_t\,d[X,Y]_t \quad \text{u.c.p.},$$
$$\lim_{\varepsilon\to 0}\int_0^s g_t\,\frac{(X((t+\varepsilon)\wedge s)-X(t))^2}{\varepsilon}\,dt$$
$$=\int_0^s g_t\,d[X,X]_t \quad \text{u.c.p.},$$
$$\lim_{\varepsilon\to 0}\int_0^s g_t\,\frac{(Y((t+\varepsilon)\wedge s)-Y(t))^2}{\varepsilon}\,dt$$
$$=\int_0^s g_t\,d[Y,Y]_t \quad \text{u.c.p.}$$

Proof. Without loss of generality, we first reduce to the case $g\geq 0$. Using polarity arguments of the type

$$[X+Y,X+Y]_t=[X,X]_t+[Y,Y]_t+2\,[X,Y]_t$$
$$[X+Y,X+Y]^{ucp}_\varepsilon(t)=[X,X]^{ucp}_\varepsilon(t)+[Y,Y]^{ucp}_\varepsilon(t)+2\,[X,Y]^{ucp}_\varepsilon(t),$$

we can reduce to the case $X=Y$.

(i) implies (iii) by Lemma 5.4.5.

(i) follows from (iii) choosing $g=1$ and Corollary 5.4.4 2).

(i) implies (ii) by Portmanteau theorem. □

Remark 5.4.8. Let X,Y be two càdlàg processes. The equivalence (i) $\Rightarrow$ (iii) in Proposition 5.4.7 with $g=1$ implies that the following are equivalent:

- (X,Y) admits all its mutual brackets;
- $[X,X]$, $[X,Y]$, $[Y,Y]$ exist in the pathwise sense.

Proposition 5.4.9. *Let X be a finite quadratic variation process. The following are equivalent.*

(i) *X is a weak Dirichlet process;*

(ii) *$X=M+A$, $[A,N]=0$ in the pathwise sense for every N continuous local martingale.*

Proof. (i) $\Rightarrow$ (ii) obviously. Assume now that (ii) holds. Taking into account Corollary 5.4.4 2), it is enough to prove that $[A,N]$ exists. Now, we recall that, whenever M and N are local martingale, $[M,N]$ exists by Proposition 5.3.8. Let N be a continuous local martingale. By Remark 5.4.6, $[X,X]$ and $[N,N]$ exist in the pathwise sense. By additivity and item (ii), $[X,N]=[M,N]$ exists in the pathwise sense. By Remark 5.4.8, (X,N) admits all its mutual brackets. Finally, by bilinearity

$$[A,N]=[X,N]-[M,N]=0.$$

□

5.5. Itô formula for $C^{1,2}$ functions

5.5.1. The basic formulae. We start with the Itô formula for finite quadratic variation processes in the sense of calculus via regularizations.

Proposition 5.5.1. *Let X be a finite quadratic variation càdlàg process and F : $[0, T] \times \mathbb{R} \to \mathbb{R}$ a function of class $C^{1,2}$. Then we have*

$$\begin{aligned} F(t, X_t) =&F(0, X_0) + \int_0^t \partial_s F(s, X_s)\, ds + \int_0^t \partial_x F(s, X_s)\, d^- X_s \\ &+ \frac{1}{2} \int_0^t \partial^2_{xx} F(s, X_{s-})\, d[X, X]^c_s \\ &+ \sum_{s \le t} [F(s, X_s) - F(s, X_{s-}) - \partial_x F(s, X_{s-})\, \Delta X_s]. \end{aligned} \tag{5.64}$$

Proof. Since X is a finite quadratic variation process, by Lemma 5.4.5, taking into account Definition 5.4.2 and Corollary 5.4.4-2), for a given càdlàg process (g_t) we have

$$\int_0^s g_t\, (X_{(t+\varepsilon)\wedge s} - X_t)^2\, \frac{dt}{\varepsilon} \overset{\varepsilon \to 0}{\longrightarrow} \int_0^s g_{t-}\, d[X, X]_t \quad \text{u.c.p.}$$

Setting $g_t = 1$ and $g_t = \frac{\partial^2_{xx} F(t,\, X_t)}{2}$, there exists a positive sequence ε_n such that

$$\lim_{n \to \infty} \int_0^s (X_{(t+\varepsilon_n)\wedge s} - X_t)^2\, \frac{dt}{\varepsilon_n} = [X, X]_s, \tag{5.65}$$

$$\lim_{n \to \infty} \int_0^s \frac{\partial^2_{xx} F(t,\, X_t)}{2}\, (X_{(t+\varepsilon_n)\wedge s} - X_t)^2\, \frac{dt}{\varepsilon_n} = \int_{]0,\, s]} \frac{\partial^2_{xx} F(t,\, X_{t-})}{2}\, d[X, X]_t, \tag{5.66}$$

uniformly in s, a.s. Let then $\mathcal{N}$ be a null set such that (5.65), (5.66) hold for every $\omega \notin \mathcal{N}$.

In the sequel we fix $\gamma > 0$, $\varepsilon > 0$, and $\omega \notin \mathcal{N}$, and we enumerate the jumps of $X(\omega)$ on $[0, T]$ by $(t_i)_{i \ge 0}$. Let $N = N(\omega)$ such that

$$\sum_{i=N+1}^{\infty} |\Delta X_{t_i}(\omega)|^2 \le \gamma^2. \tag{5.67}$$

From now on the dependence on ω will be often neglected. The quantity

$$J_0(\varepsilon, s) = \frac{1}{\varepsilon} \int_0^s [F((t + \varepsilon) \wedge s,\, X_{(t+\varepsilon)\wedge s}) - F(t,\, X_t)]\, dt, \quad s \in [0, T] \tag{5.68}$$

converges to $F(s, X_s) - F(0, X_0)$ uniformly in s. As a matter of fact, setting $Y_t = (t, X_t)$, we have

$$\begin{aligned} J_0(\varepsilon, s) &= \frac{1}{\varepsilon} \int_{[0,\, s[} F(Y_{(t+\varepsilon)\wedge s})\, dt - \frac{1}{\varepsilon} \int_{[0,\, s[} F(Y_t)\, dt \\ &= \frac{1}{\varepsilon} \int_{[\varepsilon,\, s+\varepsilon[} F(Y_{t\wedge s})\, dt - \frac{1}{\varepsilon} \int_{[0,\, s[} F(Y_t)\, dt \end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\varepsilon}\int_{[s,\,s+\varepsilon[} F(Y_{t\wedge s})\,dt - \frac{1}{\varepsilon}\int_{[0,\,\varepsilon[} F(Y_t)\,dt \\
&= F(Y_s) - \frac{1}{\varepsilon}\int_{[0,\,\varepsilon[} F(Y_t)\,dt \\
&\underset{\varepsilon\to 0}{\longrightarrow} F(Y_s) - F(Y_0) \quad \text{uniformly in } s. \qquad (5.69)
\end{aligned}
$$

As in (5.29), we define

$$
A(\varepsilon, N) = \bigcup_{i=1}^{N}]t_i - \varepsilon,\, t_i], \qquad (5.70)
$$

$$
B(\varepsilon, N) = \bigcup_{i=1}^{N}]t_{i-1},\, t_i - \varepsilon] = [0,\, T] \setminus A(\varepsilon, N). \qquad (5.71)
$$

$J_0(\varepsilon,\, s)$ can be also rewritten as

$$
J_0(\varepsilon,\, s) = J_A(\varepsilon,\, N,\, s) + J_B(\varepsilon,\, N,\, s), \qquad (5.72)
$$

where

$$
\begin{aligned}
J_A(\varepsilon,\, N,\, s) &= \frac{1}{\varepsilon}\int_0^s [F((t+\varepsilon)\wedge s,\, X_{(t+\varepsilon)\wedge s}) - F(t,\, X_t)]\, 1_{A(\varepsilon,N)}(t)\, dt, \quad (5.73)\\
J_B(\varepsilon,\, N,\, s) &= \frac{1}{\varepsilon}\int_0^s [F((t+\varepsilon)\wedge s,\, X_{(t+\varepsilon)\wedge s}) - F(t,\, X_t)]\, 1_{B(\varepsilon,N)}(t)\, dt. \quad (5.74)
\end{aligned}
$$

Applying Lemma 5.3.11 with $n = 2$ to $Y = (Y^1, Y^2) = (t, X)$ and $\phi(y_1, y_2) = F(y_1^1, y_1^2) - F(y_2^1, y_2^2)$, we have

$$
\begin{aligned}
J_A(\varepsilon,\, N,\, s) &= \sum_{i=1}^{N} \frac{1}{\varepsilon}\int_{t_i-\varepsilon}^{t_i} [F((t+\varepsilon)\wedge s,\, X_{(t+\varepsilon)\wedge s}) - F(t,\, X_t)]\, dt \\
&\overset{\varepsilon\to 0}{\longrightarrow} \sum_{i=1}^{N} 1_{]0,\,s]}(t_i)\, [F(t_i,\, X_{t_i}) - F(t_i,\, X_{t_i-})] \quad \text{uniformly in } s. \quad (5.75)
\end{aligned}
$$

Concerning $J_B(\varepsilon,\, N,\, s)$, it can be decomposed into the sum of the two terms

$$
\begin{aligned}
J_{B1}(\varepsilon,\, N,\, s) &= \frac{1}{\varepsilon}\int_0^s [F((t+\varepsilon)\wedge s,\, X_{(t+\varepsilon)\wedge s}) - F(t,\, X_{(t+\varepsilon)\wedge s})]\, 1_{B(\varepsilon,N)}(t)\, dt, \\
J_{B2}(\varepsilon,\, N,\, s) &= \frac{1}{\varepsilon}\int_0^s [F(t,\, X_{(t+\varepsilon)\wedge s}) - F(t,\, X_t)]\, 1_{B(\varepsilon,N)}(t)\, dt.
\end{aligned}
$$

Expanding in time we get

$$
J_{B1}(\varepsilon,\, N,\, s) = J_{B10}(\varepsilon,\, s) + J_{B11}(\varepsilon,\, N,\, s) + J_{B12}(\varepsilon,\, N,\, s) + J_{B13}(\varepsilon,\, N,\, s), \quad (5.76)
$$

where

$$
J_{B10}(\varepsilon,\, s) = \int_0^s \partial_t F(t,\, X_t)\, \frac{(t+\varepsilon)\wedge s - t}{\varepsilon}\, dt,
$$

$$
J_{B11}(\varepsilon,\, N,\, s) = -\sum_{i=1}^{N}\int_{t_i-\varepsilon}^{t_i} \partial_t F(t,\, X_t)\, \frac{(t+\varepsilon)\wedge s - t}{\varepsilon}\, dt,
$$

$$J_{B12}(\varepsilon, N, s) = \int_0^s R_1(\varepsilon, t, s)\, 1_{B(\varepsilon,N)}(t)\, \frac{(t+\varepsilon)\wedge s - t}{\varepsilon}\, dt,$$
$$J_{B13}(\varepsilon, N, s) = \int_0^s R_2(\varepsilon, t, s)\, 1_{B(\varepsilon,N)}(t)\, \frac{(t+\varepsilon)\wedge s - t}{\varepsilon}\, dt,$$

and

$$R_1(\varepsilon, t, s) = \int_0^1 [\partial_t F(t + a\,((t+\varepsilon)\wedge s - t),\, X_{(t+\varepsilon)\wedge s}) - \partial_t F(t,\, X_{(t+\varepsilon)\wedge s})]\, da, \tag{5.77}$$
$$R_2(\varepsilon, t, s) = \partial_t F(t,\, X_{(t+\varepsilon)\wedge s}) - \partial_t F(t,\, X_t). \tag{5.78}$$

A Taylor expansion in space up to second order gives

$$J_{B2}(\varepsilon, N, s) = J_{B20}(\varepsilon, s) + J_{B21}(\varepsilon, s) + J_{B22}(\varepsilon, N, s) + J_{B23}(\varepsilon, N, s), \tag{5.79}$$

where

$$J_{B20}(\varepsilon, s) = \frac{1}{\varepsilon}\int_0^s \partial_x F(t,\, X_t)\,(X_{(t+\varepsilon)\wedge s} - X_t)\, dt, \tag{5.80}$$
$$J_{B21}(\varepsilon, s) = \frac{1}{\varepsilon}\int_0^s \frac{\partial^2_{xx}F(t,\, X_t)}{2}\,(X_{(t+\varepsilon)\wedge s} - X_t)^2\, dt,$$
$$J_{B22}(\varepsilon, N, s) = -\frac{1}{\varepsilon}\sum_{i=1}^N \int_{t_i-\varepsilon}^{t_i} \Big[\partial_x F(t,\, X_t)\,(X_{(t+\varepsilon)\wedge s} - X_t) + \frac{\partial^2_{xx}F(t,\, X_t)}{2}\,(X_{(t+\varepsilon)\wedge s} - X_t)^2\Big]\, dt,$$
$$J_{B23}(\varepsilon, N, s) = \int_0^s R_3(\varepsilon, t, s)\, 1_{B(\varepsilon,N)}(t)\, \frac{(X_{(t+\varepsilon)\wedge s} - X_t)^2}{\varepsilon}\, dt,$$

and

$$R_3(\varepsilon, t, s) = \int_0^1 [\partial^2_{xx}F(t,\, X_t + a(X_{(t+\varepsilon)\wedge s} - X_t)) - \partial^2_{xx}F(t,\, X_t)]\, da. \tag{5.81}$$

Let us consider the term $J_{B22}(\varepsilon, N, s)$. Applying Lemma 5.3.11 with $n = 2$ to $Y = (Y^1, Y^2) = (t, X)$ and $\phi(y_1, y_2) = \partial_x F(y_2^1, y_2^2)(y_1^2 - y_2^2) + \partial^2_{xx}F(y_2^1, y_2^2)(y_1^2 - y_2^2)^2$, we get

$$\lim_{\varepsilon\to 0} J_{B22}(\varepsilon, N, s) = -\sum_{i=1}^N 1_{]0,\,s]}(t_i)\,\Big[\partial_x F(t_i,\, X_{t_{i-}})\,(X_{t_i} - X_{t_i-}) + \frac{\partial^2_{xx}F(t_i,\, X_{t_{i-}})}{2}(X_{t_i} - X_{t_i-})^2\Big] \tag{5.82}$$

uniformly in s. Moreover, the term $J_{B10}(\varepsilon, N, s)$ can be in

$$J_{B10}(\varepsilon, s) = \int_0^s \partial_t F(t,\, X_t)\, dt + J_{B10'}(\varepsilon, s) + J_{B10''}(\varepsilon, s), \tag{5.83}$$

with

$$J_{B10'}(\varepsilon, s) = \int_{s-\varepsilon}^s \partial_t F(t, X_t)\, \frac{s-t}{\varepsilon}\, dt, \tag{5.84}$$

$$J_{B10''}(\varepsilon,\, s) = -\int_{s-\varepsilon}^{s} \partial_t F(t, X_t)\, dt. \tag{5.85}$$

At this point we remark that identity (5.72) can be rewritten as

$$\begin{aligned} J_0(\varepsilon, s) &= J_A(\varepsilon, N, s) + \int_0^s \partial_t F(t,\, X_t)\, dt \\ &\quad + J_{B10'}(\varepsilon, s) + J_{B10''}(\varepsilon, s) + J_{B11}(\varepsilon, N, s) + J_{B12}(\varepsilon, N, s) + J_{B13}(\varepsilon, N, s) \\ &\quad + J_{B20}(\varepsilon, s) + J_{B21}(\varepsilon, s) + J_{B22}(\varepsilon, N, s) + J_{B23}(\varepsilon, N, s). \end{aligned} \tag{5.86}$$

Passing to the limit in (5.86) on both the left-hand and right-hand sides, uniformly in s, as ε goes to zero, taking into account convergences (5.69), (5.75), (5.82), we get

$$\begin{aligned} &F(s,\, X_s) - F(0,\, X_0) \\ &= \int_0^s \partial_t F(t,\, X_t)\, dt + \sum_{i=1}^N 1_{]0,\, s]}(t_i)\ \left[F(t_i,\, X_{t_i}) - F(t_i,\, X_{t_i-})\right] \\ &- \sum_{i=1}^N 1_{]0,\, s]}(t_i)\ \left[\partial_x F(t_i,\, X_{t_i-})\, (X_{t_i} - X_{t_i-}) - \frac{\partial^2_{xx} F(t_i,\, X_{t_i-})}{2} (X_{t_i} - X_{t_i-})^2\right] \\ &+ \lim_{\varepsilon \to 0} \left(J_{B20}(\varepsilon,\, N,\, s) + J_{B21}(\varepsilon,\, s) + L(\varepsilon, N, s)\right), \end{aligned} \tag{5.87}$$

where the previous limit is intended uniformly in s, and we have set

$$\begin{aligned} L(\varepsilon, N, s) := &J_{B10'}(\varepsilon,\, s) + J_{B10''}(\varepsilon,\, s) + J_{B11}(\varepsilon,\, N,\, s) + J_{B12}(\varepsilon,\, N,\, s) \\ &+ J_{B13}(\varepsilon,\, N,\, s) + J_{B23}(\varepsilon,\, N,\, s). \end{aligned}$$

We evaluate previous limit uniformly in s, for every $\omega \notin \mathcal{N}$. Without restriction of generality it is enough to show the uniform convergence in s for the subsequence ε_n introduced in (5.65)-(5.66), when $n \to \infty$.

According to (5.66), we get

$$\lim_{n \to \infty} J_{B21}(\varepsilon_n,\, s) = \int_{]0,\, s]} \frac{\partial^2_{xx} F(t,\, X_{t-})}{2}\, d[X, X]_t, \tag{5.88}$$

uniformly in s.

We now should discuss $J_{B12}(\varepsilon_n,\, N,\, s)$, $J_{B13}(\varepsilon_n,\, N,\, s)$ and $J_{B23}(\varepsilon_n,\, N,\, s)$. In the sequel, $\rho(f, \cdot)$ will denote the modulus of continuity of a function f, and by I_l the interval $[t_{l-1},\, t_l]$, $l \geq 0$. Since $\frac{(t+\varepsilon)\wedge s - t}{\varepsilon} \leq 1$ for every t, s, by Remark 5.3.13 we get

$$\begin{aligned} 1_{B(\varepsilon,N)}(t)\, |R_1(\varepsilon,\, t,\, s)| &\leq \rho\left(\partial_t F,\, \varepsilon\right), \\ 1_{B(\varepsilon,N)}(t)\, |R_2(\varepsilon,\, t,\, s)| &\leq \rho\Big(\partial_t F,\, \sup_l \sup_{\substack{t,a \in I_l \\ |t-a| \leq \varepsilon}} |X_a - X_t|)\Big), \\ 1_{B(\varepsilon,N)}(t)\, |R_3(\varepsilon,\, t,\, s)| &\leq \rho\Big(\partial^2_{xx} F,\, \sup_l \sup_{\substack{t,a \in I_l \\ |t-a| \leq \varepsilon}} |X_a - X_t|)\Big). \end{aligned}$$

Considering the two last inequalities, Lemma 5.3.12 applied successively to the intervals I_l implies

$$1_{B(\varepsilon,N)}(t)\,|R_2(\varepsilon,\,t,\,s)| \le \rho(\partial_t F,\,3\gamma),$$
$$1_{B(\varepsilon,N)}(t)\,|R_3(\varepsilon,\,t,\,s)| \le \rho(\partial^2_{xx} F,\,3\gamma).$$

Then, using again $\frac{(t+\varepsilon_n)\wedge s - t}{\varepsilon} \le 1$, we get

$$\sup_{s\in[0,\,T]} |J_{B12}(\varepsilon_n,\,N,\,s)| \le \rho(\partial_t F,\,\varepsilon_n)\cdot T,$$
$$\sup_{s\in[0,\,T]} |J_{B13}(\varepsilon_n,\,N,\,s)| \le \rho(\partial_t F,\,3\gamma)\cdot T,$$
$$\sup_{s\in[0,\,T]} |J_{B23}(\varepsilon_n,\,N,\,s)| \le \rho(\partial^2_{xx} F,\,3\gamma)\cdot \sup_{n\in\mathbb{N},s\in[0,\,T]} [X,X]^{ucp}_{\varepsilon_n}(s), \tag{5.89}$$

where we remark that the supremum in the right-hand side of (5.89) is finite taking into account (5.65). Therefore

$$\limsup_{n\to\infty} \sup_{s\in[0,\,T]} |J_{B23}(\varepsilon_n,\,N,\,s)| = \rho(\partial^2_{xx} F,\,3\gamma)\cdot \sup_{n\in\mathbb{N},s\in[0,\,T]} [X,X]^{ucp}_{\varepsilon_n}(s), \tag{5.90}$$
$$\limsup_{n\to\infty} \sup_{s\in[0,\,T]} |J_{B13}(\varepsilon_n,\,N,\,s)| = \rho(\partial_t F,\,3\gamma)\cdot T, \tag{5.91}$$

while

$$\lim_{n\to\infty} \sup_{s\in[0,\,T]} |J_{B12}(\varepsilon_n,\,N,\,s)| = 0. \tag{5.92}$$

Let now consider the terms $J_{B10'}(\varepsilon_n,\,s)$, $J_{B10''}(\varepsilon_n,\,s)$ and $J_{B11}(\varepsilon_n,\,N,\,s)$.

$$\sup_{s\in[0,\,T]} |J_{B10'}(\varepsilon_n,\,s)| \le \sup_{y\,\in\,\mathbb{K}^X(\omega)\times[0,\,T]} |\partial_t F(y)|\cdot\varepsilon_n,$$
$$\sup_{s\in[0,\,T]} |J_{B10''}(\varepsilon_n,\,s)| \le \sup_{y\,\in\,\mathbb{K}^X(\omega)\times[0,\,T]} |\partial_t F(y)|\cdot\varepsilon_n,$$
$$\sup_{s\in[0,\,T]} |J_{B11}(\varepsilon_n,\,N,\,s)| \le \sup_{y\,\in\,\mathbb{K}^X(\omega)\times[0,\,T]} |\partial_t F(y)|\,N\cdot\varepsilon_n,$$

where $\mathbb{K}^X(\omega)$ is the (compact) set $\{X_t(\omega),\,t\in[0,\,T]\}$. So, it follows

$$\begin{aligned}&\lim_{n\to\infty} \sup_{s\in[0,\,T]} |J_{B10'}(\varepsilon_n,\,s)|\\ &= \lim_{n\to\infty} \sup_{s\in[0,\,T]} |J_{B10''}(\varepsilon_n,\,s)|\\ &= \lim_{n\to\infty} \sup_{s\in[0,\,T]} |J_{B11}(\varepsilon_n,\,N,\,s)| = 0.\end{aligned} \tag{5.93}$$

Taking into account (5.93), (5.91), (5.90), and (5.88), we see that

$$\limsup_{n\to\infty} \sup_{s\in[0,\,T]} |L(\varepsilon_n,N,s)| = \rho(\partial^2_{xx} F,\,3\gamma)\cdot \sup_{n\in\mathbb{N},s\in[0,\,T]} [X,X]^{ucp}_{\varepsilon_n}(s) + \rho(\partial_t F,\,3\gamma)\cdot T. \tag{5.94}$$

Recalling that $J_{B20}(\varepsilon,s)$ in (5.80) is the ε-approximation of the forward integral $\int_0^t \partial_x F(s,X_s)\,d^-X_s$, to conclude it remains to show that

$$\sup_{s\in[0,\,T]} \big|J_{B20}(\varepsilon_n,\,s) - J(s)\big| \underset{n\to\infty}{\longrightarrow} 0 \quad \text{a.s.}, \tag{5.95}$$

where

$$
\begin{aligned}
J(s) = {} & F(s, X_s) - F(0, X_0) - \int_{]0,\, s]} \partial_t F(t, X_t)\, dt - \sum_{t \le s} [F(t, X_t) - F(t, X_{t-})] \\
& + \sum_{0<t\le s} \left[\partial_x F(t, X_{t-})\, (X_t - X_{t-}) + \frac{\partial^2_{xx} F(t, X_{t-})}{2}\, (X_t - X_{t-})^2 \right] \\
& - \frac{1}{2} \int_{]0,\, s]} \partial^2_{xx} F(t, X_{t-})\, d[X, X]_t.
\end{aligned} \tag{5.96}
$$

In particular this would imply that $\int_{]0,\, s]} \partial_x F(t, X_t)\, d^- X_t$ exists and equals $J(s)$. Taking into account (5.86), we have

$$
\begin{aligned}
J_{B20}(\varepsilon_n, s) = {} & J_0(\varepsilon_n, s) - J_A(\varepsilon_n, N, s) - \int_0^s \partial_t F(t, X_t)\, dt \\
& - L(\varepsilon_n, N, s) - J_{B21}(\varepsilon_n, s) - J_{B22}(\varepsilon_n, N, s).
\end{aligned} \tag{5.97}
$$

Taking into account (5.96) and (5.97), we see that the term inside the absolute value in (5.95) equals

$$
\begin{aligned}
& J_0(\varepsilon_n, s) - (F(s, X_s) - F(0, X_0)) \\
& - J_A(\varepsilon_n, N, s) + \sum_{i=1}^N 1_{]0,\, s]}(t_i)[F(t_i, X_{t_i}) - F(t_i, X_{t_i-})] \\
& - J_{B22}(\varepsilon_n, N, s) \\
& - \sum_{i=1}^N 1_{]0,\, s]}(t_i) \left[\partial_x F(t_i, X_{t_i-})\, (X_{t_i} - X_{t_i-}) + \frac{\partial^2_{xx} F(t_i, X_{t_i-})}{2}\, (X_{t_i} - X_{t_i-})^2 \right] \\
& - J_{B21}(\varepsilon_n, s) + \frac{1}{2} \int_{]0,\, s]} \partial^2_{xx} F(t, X_{t-})\, d[X, X]_t \\
& - L(\varepsilon_n, N, s) \\
& + \sum_{i=N+1}^\infty 1_{]0,\, s]}(t_i) \Big[F(t_i, X_{t_i}) - F(t_i, X_{t_i-}) - \partial_x F(t_i, X_{t_i-})\, (X_{t_i} - X_{t_i-}) \\
& \qquad\qquad - \frac{\partial^2_{xx} F(t_i, X_{t_i-})}{2}\, (X_{t_i} - X_{t_i-})^2 \Big].
\end{aligned}
$$

Taking into account (5.69), (5.75), (5.82), (5.92), (5.94),we have

$$
\begin{aligned}
& \limsup_{n \to \infty} \sup_{s \in [0,\, T]} |J_{B20}(\varepsilon_n, s) - J(s)| \\
& \le \limsup_{n \to \infty} \sup_{s \in [0,\, T]} |L(\varepsilon_n, N, s)| \\
& + \sup_{s \in [0,\, T]} \Big\{ \sum_{i=N+1}^\infty 1_{]0,\, s]}(t_i) \Big| F(t_i, X_{t_i}) - F(t_i, X_{t_i-}) - \partial_x F(t_i, X_{t_i-})\, \Delta X_{t_i} \\
& \qquad\qquad - \frac{\partial^2_{xx} F(t_i, X_{t_i-})}{2}\, (\Delta X_{t_i})^2 \Big| \Big\}
\end{aligned}
$$

$$
\begin{aligned}
&= \limsup_{n\to\infty} \sup_{s\in[0,T]} |L(\varepsilon_n, N, s)| \\
&+ \sup_{s\in[0,T]} \frac{(\Delta X_s)^2}{2} \sum_{i=N+1}^{\infty} 1_{]0,s]}(t_i) \left| \int_0^1 \partial^2_{xx} F(t_i, X_{t_i-} + a(\Delta X_{t_i}))\, da - \partial^2_{xx} F(t_i, X_{t_i-}) \right| \\
&\le \rho(\partial_t F, 3\gamma)\cdot T + \rho(\partial^2_{xx} F, 3\gamma) \sup_{n\in\mathbb{N}, s\in[0,T]} [X,X]^{ucp}_{\varepsilon_n}(s) + \gamma^2 \sup_{y\in\mathbb{K}^X(\omega)\times[0,T]} |\partial^2_{xx} F(y)|,
\end{aligned}
\tag{5.98}
$$

where the last term on the right-hand side of (5.98) is obtained using (5.67). Since γ is arbitrarily small, we conclude that

$$
\lim_{n\to\infty} \sup_{s\in[0,T]} |J_{B20}(\varepsilon_n, s) - J(s)| = 0, \quad \forall \omega \notin \mathcal{N}.
$$

This conclude the proof of the Itô formula. □

From Proposition 5.5.1, Proposition 5.3.8-ii), and by classical Banach-Steinhaus theory (see, e.g., [**46**], Theorem 1.18 pag 55) for F-type spaces, we have the following.

Proposition 5.5.2. *Let $F : [0, T] \times \mathbb{R} \to \mathbb{R}$ be a function of class C^1 such that $\partial_x F$ is Hölder continuous with respect to the second variable for some $\lambda \in [0, 1[$. Let $(X_t)_{t\in[0,T]}$ be a reversible semimartingale, satisfying moreover*

$$
\sum_{0<s\le t} |\Delta X_s|^{1+\lambda} < \infty \quad \text{a.s.}
$$

Then

$$
\begin{aligned}
F(t, X_t) &= F(0, X_0) + \int_0^t \partial_s F(s, X_s)\, ds + \int_0^t \partial_x F(s, X_{s-})\, dX_s \\
&+ \frac{1}{2} [\partial_x F(\cdot, X), X]_t + J(F, X)(t),
\end{aligned}
$$

where

$$
J(F, X)(t) = \sum_{0<s\le t} \left[F(s, X_s) - F(s, X_{s-}) - \frac{\partial_x F(s, X_s) + \partial_x F(s, X_{s-})}{2} \Delta X_s \right].
$$

Remark 5.5.3.

(i) Previous result can be easily extended to the case when X is multidimensional.

(ii) When F does not depend on time, previous statement was the object of [**59**], Theorem 3.8, example 3.3.1. In that case however, stochastic integrals and covariations were defined by discretizations means.

(iii) The proof of Proposition 5.5.2 follows the same lines as the one of Theorem 3.8. in [**59**].

5.5.2. Itô formula related to random measures. The object of the present section is to reexpress the statement of Proposition 5.5.1 making use of the jump measure μ^X associated with a càdlàg process X, recalled in Section 5.2.1. The compensator of $\mu^X(ds\,dy)$ is called the Lévy system of X, and will be denoted by $\nu^X(ds\,dy)$ (for more details see Chapter II, Section 1, in [**79**]); we also define

$$\hat{\nu}_t^X = \nu^X(\{t\},\, dy) \quad \text{for every } t \in [0,\, T]. \tag{5.99}$$

Corollary 5.5.4. *Let X be a finite quadratic variation càdlàg process and $F : [0,\, T] \times \mathbb{R} \to \mathbb{R}$ a function of class $C^{1,2}$. Then we have*

$$\begin{aligned}
F(t, X_t) &= F(0, X_0) + \int_0^t \partial_s F(s, X_s)\, ds + \int_0^t \partial_x F(s, X_s)\, d^- X_s \\
&\quad + \frac{1}{2}\int_0^t \partial^2_{xx} F(s, X_s)\, d[X,X]^c_s \\
&\quad + \int_{]0,\,t]\times\mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}))\, 1_{\{x \le 1\}}\, (\mu^X - \nu^X)(ds\, dx) \\
&\quad - \int_{]0,\,t]\times\mathbb{R}} x\, \partial_x F(s, X_{s-})\, 1_{\{x \le 1\}}\, (\mu^X - \nu^X)(ds\, dx) \\
&\quad + \int_{]0,\,t]\times\mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}) - x\, \partial_x F(s, X_{s-}))\, 1_{\{x > 1\}}\, \mu^X(ds\, dx) \\
&\quad + \int_{]0,\,t]\times\mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}) - x\, \partial_x F(s, X_{s-}))\, 1_{\{x \le 1\}}\, \nu^X(ds\, dx).
\end{aligned} \tag{5.100}$$

Proof. We set

$$\begin{aligned}
W_s(x) &= (F(s, X_{s-} + x) - F(s, X_{s-}) - x\, \partial_x F(s, X_{s-}))\, 1_{\{|x| \le 1\}}, \\
K_s(x) &= (F(s, X_{s-} + x) - F(s, X_{s-}))\, 1_{\{|x| \le 1\}}, \\
Y_s(x) &= x\, \partial_x F(s, X_{s-}) 1_{\{|x| \le 1\}}.
\end{aligned}$$

By Propositions 5.2.7, $|W| * \mu^X$ belongs to $\mathcal{A}^+_{\rm loc}$, while Proposition 5.2.8 insures that $K^2 * \mu^X$ and $Y^2 * \mu^X$ belong to $\mathcal{A}^+_{\rm loc}$. Then, Proposition B.18 implies that $W \in \mathcal{G}^1_{\rm loc}(\mu^X)$ and that the stochastic integral $W * (\mu^X - \nu^X)$ can be decomposed as $W * \mu^X - W * \nu^X$. On the other hand, since K, Y belong to $\mathcal{G}^2_{\rm loc}(\mu)$ (see Lemma B.21-2.) By Theorem B.22 it follows that K, Y belong to $\mathcal{G}^1_{\rm loc}(\mu^X)$ and that moreover $K * (\mu^X - \nu^X)$, $Y * (\mu^X - \nu^X)$ are purely discontinuous square integrable local martingales. □

5.6. About weak Dirichlet processes

5.6.1. Basic definitions. We consider again the filtration $(\mathcal{F}_t)_{t \ge 0}$ introduced at Section 5.2, which will be, without further mention, the underlying filtration.

Definition 5.6.1. Let X be an $(\mathcal{F}_t)$-adapted process. We say that X is $(\mathcal{F}_t)$-orthogonal if $[X, N] = 0$ for every N continuous local $(\mathcal{F}_t)$-martingale.

Remark 5.6.2. Basic examples of $(\mathcal{F}_t)$-orthogonal processes are purely discontinuous $(\mathcal{F}_t)$-local martingales, see Theorem A.6.

Proposition 5.6.3. *If M is a purely discontinuous $(\mathcal{F}_t)$-local martingale, then*

$$[M, M]_t = \sum_{s \le t} (\Delta M_s)^2.$$

Proof. The result follows from Theorem 5.2, Chapter I, in [**79**], and Proposition 5.3.8-(i). □

Definition 5.6.4. We say that an $(\mathcal{F}_t)$-adapted process X is a *Dirichlet process* if it admits a decomposition $X = M + A$, where M is a local martingale and A is a finite quadratic variation process with $[A, A] = 0$.

Definition 5.6.5. We say that X is an $(\mathcal{F}_t)$-adapted *weak Dirichlet process* if it admits a decomposition $X = M + A$, where M is a local martingale and the process A is $(\mathcal{F}_t)$-orthogonal.

Definition 5.6.6. We say that an $(\mathcal{F}_t)$-adapted process X is a *special weak Dirichlet process* if it admits a decomposition of the type above such that, in addition, A is predictable.

Remark 5.6.7. Obviously, a Dirichlet process is a special weak Dirichlet process.

Proposition 5.6.8. *Let X be a special weak Dirichlet process of the type*

$$X = M^c + M^d + A, \tag{5.101}$$

where M^c is a continuous local martingale, and M^d is a purely discontinuous local martingale. Supposing that $A_0 = M_0^d = 0$, the decomposition (5.101) *is unique. In that case the decomposition $X = M^c + M^d + A$ will be called the* canonical decomposition *of X.*

Proof. Assume that we have two decompositions $X = M^c + M^d + A = M^{c'} + M^{d'} + A'$, with A and A' predictable, verifying $[A, N] = [A', N] = 0$ for every continuous local martingale N. We set $\tilde{A} = A - A'$, $\tilde{M}^c = M^c - M^{c'}$ and $\tilde{M}^d = M^d - M^{d'}$. By linearity, $\tilde{M}^c + \tilde{M}^d + \tilde{A} = 0$. We have

$$\begin{aligned} 0 &= [\tilde{M}^c + \tilde{M}^d + \tilde{A}, \tilde{M}^c] \\ &= [\tilde{M}^c, \tilde{M}^c] + [\tilde{M}^d, \tilde{M}^c] + [\tilde{A}, \tilde{M}^c] \\ &= [\tilde{M}^c, \tilde{M}^c], \end{aligned}$$

therefore $\tilde{M}^c = 0$ since $\tilde{M}^c$ is a continuous martingale. It follows in particular that $\tilde{A}$ is a predictable local martingale, hence a continuous local martingale, see e.g., the point 2) of the Remarks after Definition 7.11 in [**73**]. In particular

$$0 = [\tilde{M}^d, \tilde{M}^d] + [\tilde{A}, \tilde{M}^d] = [\tilde{M}^d, \tilde{M}^d]$$

and, since $\tilde{M}_0^d = 0$, we deduce that $\tilde{M}^d = 0$ and therefore $\tilde{A} = 0$. □

Remark 5.6.9. Every $(\mathcal{F}_t)$-special weak Dirichlet process is of the type (5.101). Indeed, every local martingale M can be decomposed as the sum of a continuous local martingale M^c and a purely discontinuous local martingale M^d, see Theorem 4.18, Chapter I, in [**79**].

Corollary 5.6.10. *Let X be an $(\mathcal{F}_t)$-special weak Dirichlet process. Then, for every $t \in [0, T]$,*

(i) $[X, X]_t = [M^c, M^c]_t + \sum_{s \le t} (\Delta X_t)^2$;

(ii) $[X, X]^c_t = [M^c, M^c]_t$.

Proof. (ii) follows from (i). Concerning (i), by the bilinearity of the covariation, and by the definitions of purely discontinuous local martingale (see Remark 5.6.2) and of special weak Dirichlet process, we have

$$\begin{aligned}
[X, X]_t &= [M^c, M^c]_t + [M^d, M^d]_t \\
&= [M^c, M^c]_t + \sum_{s \le t} (\Delta M^d_s)^2 \\
&= [M^c, M^c]_t + \sum_{s \le t} (\Delta X_s)^2,
\end{aligned}$$

where the second equality holds because of Proposition 5.6.3. □

We give a first relation between semimartingales and weak Dirichlet processes.

Proposition 5.6.11. *Let S be an $(\mathcal{F}_t)$-semimartingale which is a special weak Dirichlet process. Then S is a special semimartingale.*

Proof. Let $S = M^1 + V$ such that M^1 is a local martingale and V is a bounded variation process. Let moreover $S = M^2 + A$, where a predictable (F_t)-orthogonal process. Then $0 = V - A + M$, where $M = M^2 - M^1$. So A is a predictable semimartingale. By Corollary 8.7 in [**73**], A is a special semimartingale, and so by additivity S is a special semimartingale as well. □

5.6.2. Stability of weak Dirichlet processes under $C^{0,1}$ transformation. We begin with the $C^{1,2}$ stability.

Lemma 5.6.12. *Let $X = M + A$ be a càdlàg weak Dirichlet process of finite quadratic variation and $F : [0, T] \times \mathbb{R} \to \mathbb{R}$ be a $C^{1,2}$ real-valued function. Then*

$$\begin{aligned}
F(t, X_t) = F(0, X_0) &+ \int_0^t \partial_x F(s, X_{s-})\, dM_s \\
&+ \int_{]0,\, t] \times \mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-})) 1_{\{|x| \le 1\}}\, (\mu^X - \nu^X)(ds\, dx), \\
&- \int_{]0,\, t] \times \mathbb{R}} x\, \partial_x F(s, X_{s-})\, 1_{\{|x| \le 1\}}\, (\mu^X - \nu^X)(ds\, dx), \\
&+ \int_{]0,\, t] \times \mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}) - x\, \partial_x F(s, X_{s-})) 1_{\{|x| > 1\}}\, \mu^X(ds\, dx)
\end{aligned}$$

$$+ \Gamma^F(t), \tag{5.102}$$

where

$$\Gamma^F(t) := \int_0^t \partial_s F(s, X_s)\, ds + \int_0^t \partial_x F(s, X_s)\, d^- A_s + \int_0^t \partial^2_{xx} F(s, X_s)\, d[X, X]^c_s$$
$$+ \int_{]0,\, t] \times \mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}) - x\, \partial_x F(s, X_{s-})) 1_{\{|x| \le 1\}}\, \nu^X(ds\, dx). \tag{5.103}$$

Remark 5.6.13. Taking into account Proposition 5.3.3, we can observe that, if A is predictable, then Γ^F is a predictable process for any $F \in C^{1,2}$.

Proof. Expressions (5.102)-(5.103) follow by Corollary 5.5.4, in particular by (5.100). We remark that, since M is a local martingale and $\partial_x F(s, X_s)$ is a càdlàg process, by Proposition 5.3.8-(ii) we have

$$\begin{aligned}\int_0^t \partial_x F(s, X_s)\, d^- X_s &= \int_0^t \partial_x F(s, X_s)\, d^- M_s + \int_0^t \partial_x F(s, X_s)\, d^- A_s \\ &= \int_0^t \partial_x F(s, X_{s-})\, dM_s + \int_0^t \partial_x F(s, X_s)\, d^- A_s.\end{aligned}$$

□

Theorem 5.6.14. *Let $X = M + A$ be a càdlàg weak Dirichlet process of finite quadratic variation. Then, for every $F : [0, T] \times \mathbb{R} \to \mathbb{R}$ of class $C^{0,1}$, we have*

$$F(t, X_t) = F(0, X_0) + \int_0^t \partial_x F(s, X_{s-})\, dM_s$$
$$+ \int_{]0,\, t] \times \mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}))\, 1_{\{|x| \le 1\}}\, (\mu^X - \nu^X)(ds\, dx)$$
$$- \int_{]0,\, t] \times \mathbb{R}} x\, \partial_x F(s, X_{s-})\, 1_{\{|x| \le 1\}}\, (\mu^X - \nu^X)(ds\, dx)$$
$$+ \int_{]0,\, t] \times \mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}) - x\, \partial_x F(s, X_{s-}))\, 1_{\{|x| > 1\}}\, \mu^X(ds\, dx) + \Gamma^F(t), \tag{5.104}$$

where $\Gamma^F : C^{0,1} \to \mathbb{D}^{ucp}$ is a continuous linear map, such that its restriction to $C^{1,2}$ is given by (5.103)*. Moreover, for every $F \in C^{0,1}$, it fulfills the following properties.*

(a) $[\Gamma^F, N] = 0$ *for every N continuous local martingale.*

(b) *If A is predictable, then Γ^F is predictable.*

In particular point (a) implies that $F(s, X_s)$ is a weak Dirichlet process when X is a weak Dirichlet process.

Proof. In agreement with (5.104) we set

$$\Gamma^F(t) := F(t, X_t) - F(0, X_0) - \int_0^t \partial_x F(s, X_{s-})\, dM_s \tag{5.105}$$

$$- \int_{]0,\,t]\times\mathbb{R}} \{F(s, X_{s-}+x) - F(s, X_{s-}) - x\,\partial_x F(s, X_{s-})\}\, 1_{\{|x|>1\}}\, \mu^X(ds\,dx)$$
$$- \int_{]0,\,t]\times\mathbb{R}} \{F(s, X_{s-}+x) - F(s, X_{s-}) - x\,\partial_x F(s, X_{s-})\}\, 1_{\{|x|\le 1\}}\, (\mu^X - \nu^X)(ds\,dx).$$

We need first to prove that $C^{0,1} \supset F \mapsto \Gamma^F(t)$ is continuous with respect to the u.c.p. topology. For this we first observe that the map $F \mapsto F(t, X_t) - F(0, X_0)$ fulfills the mentioned continuity. Moreover, if $F^n \to F$ in $C^{0,1}$, then $\int_0^t (\partial_x F^n - \partial_x F)(s, X_{s-})\, dM_s$ converges to zero u.c.p. since $\partial_x F^n(s, X_{s-})$ converges to $\partial_x F(s, X_{s-})$ in $\mathbb{L}^{ucp}$, see Chapter II Section 4 in [**110**].

Let us consider the second line of (5.105). For almost all fixed ω, the process X has a finite number of jumps, $s_i = s_i(\omega), 1 \le i \le N(\omega)$, larger than one. Let $F^n \to F$ in $C^{0,1}$. Since the map is linear we can suppose that $F = 0$.

$$\begin{aligned}
&\sup_{0<t\le T} \bigg| \int_{]0,\,t]\times\mathbb{R}} \big\{F^n(s, X_{s-}(\omega)+x) - F^n(s, X_{s-}(\omega)) \\
&\qquad\qquad - x\,\partial_x F^n(s, X_{s-}(\omega))\big\}\, 1_{\{|x|>1\}}\, \mu^X(\omega, ds\,dx) \bigg| \\
&\le \int_{]0,\,T]\times\mathbb{R}} |F^n(s, X_{s-}(\omega)+x) - F^n(s, X_{s-}(\omega)) \\
&\qquad\qquad - x\,\partial_x F^n(s, X_{s-}(\omega))|\, 1_{\{|x|>1\}}\, \mu^X(\omega, ds\,dx) \\
&= \sum_{i=1}^{N(\omega)} |F^n(s_i, X_{s_i}(\omega)) - F^n(s_i, X_{s_i-}(\omega)) - \Delta X_{s_i}(\omega)\,\partial_x F^n(s_i, X_{s_i-}(\omega))|\; 1_{\{|\Delta X_{s_i}(\omega)|>1\}} \\
&\underset{n\to\infty}{\to} 0.
\end{aligned}$$

This shows in particular that

$$\int_{]0,\,\cdot]\times\mathbb{R}} \{F^n(s, X_{s-}(\omega)+x) - F^n(s, X_{s-}(\omega)) - x\,\partial_x F^n(s, X_{s-}(\omega))\}\, 1_{\{|x|>1\}}\, \mu^X(\omega, ds\,dx)$$
$$\to 0 \quad \text{u.c.p.}$$

and so the map defined by the second line in (5.105) is continuous.

Finally, the following proposition exploits the continuity properties of the last term in (5.105), and allows to conclude the continuity of the map $\Gamma^F : C^{0,1} \to \mathbb{D}^{ucp}$.

Proposition 5.6.15. *The map*

$$\begin{aligned}
I : C^{0,1} &\to \mathbb{D}^{ucp} \\
g &\mapsto \int_{]0,\cdot]\times\mathbb{R}} G^g\,(s,\, X_{s-},\, x)\, 1_{\{|x|\le 1\}}\, (\mu^X - \nu^X)(ds\,dx),
\end{aligned}$$

where

$$G^g\,(s,\, \xi,\, x) = g(s,\, \xi + x) - g(s, \xi) - x\,\partial_\xi g(s,\, \xi), \tag{5.106}$$

is continuous.

Proof (of the Proposition). We consider the sequence $(\tau_l)_{l\geq 1}$ of increasing stopping times introduced in Remark 5.2.3-(ii) for the process $Y_t = (X_{t-}, \sum_{s<t} |\Delta X_s|^2)$. Since $\Omega = \cup_l \{\omega : \tau_l(\omega) > T\}$ a.s., the result is proved if we show that, for every fixed $\tau = \tau_l$,

$$g \mapsto 1_{\{\tau>T\}}(\omega) \int_{]0,\,\cdot]\times\mathbb{R}} G^g(s,\, X_{s-},\, x)\, 1_{\{|x|\leq 1\}}\, (\mu^X - \nu^X)(ds\, dx)$$

is continuous. Let $g^n \to g$ in $C^{0,1}$. Then $G^{g^n} \to G^g$ in $C^0([0,\,T]\times\mathbb{R}^2)$. Since the map is linear we can suppose that $g = 0$. Let $\varepsilon_0 > 0$. We aim at showing that

$$\mathbb{P}\Bigg(\sup_{t\in[0,\,T]} \left|1_{\{\tau>T\}}(\omega) \int_{]0,\,t]\times\mathbb{R}} G^{g^n}(s,\, X_{s-},\, x)\, 1_{\{|x|\leq 1\}}\, (\mu^X - \nu^X)(ds\, dx)\right| > \varepsilon_0\Bigg) \underset{n\to\infty}{\longrightarrow} 0. \tag{5.107}$$

Let $W_s^n(x)$ (resp. by $\hat{W}_s^n$) denote the random field $G^{g^n}(s,\, X_{s-},\, x)\, 1_{\{|x|\leq 1\}}$ (resp. the process $\int_{\mathbb{R}} G^{g^n}(s,\, X_{s-},\, x)\, 1_{\{|x|\leq 1\}}\, \hat{\nu}^X(dx)$), and define

$$I_t^n := \int_{]0,\,t]\times\mathbb{R}} W_s^n(x)\, (\mu^X - \nu^X)(ds\, dx).$$

(5.107) will follow if we show that

$$\mathbb{P}\big(\sup_{t\in[0,\,T]} |I^n_{t\wedge\tau}| > \varepsilon_0\big) \underset{n\to\infty}{\longrightarrow} 0. \tag{5.108}$$

For every process $\phi = (\phi_t)_t$, we indicate the stopped process at τ by $\phi_t^\tau(\omega) := \phi_{t\wedge\tau(\omega)}(\omega)$. We have

$$(|W^n|^2 * \mu^X)^\tau \in \mathcal{A}^+. \tag{5.109}$$

As a matter of fact, let M such that $\sup_{t\in[0,\,T]} |Y_{t\wedge\tau}\, 1_{\{\tau>0\}}| \leq M$. Recalling Remark 5.2.1, an obvious Taylor expansion yields

$$\begin{aligned}
&\mathbb{E}\left[\int_{]0,\,t\wedge\tau]\times\mathbb{R}} |W_s^n(x)|^2\, \mu^X(ds,\, dx)\right]\\
&\leq 2 \sup_{\substack{y\in[-M,\,M]\\ t\in[0,\,T]}} |\partial_x g^n|^2(t,\, y)\cdot\\
&\quad\cdot \mathbb{E}\left[\sum_{0<s<\tau} |\Delta X_s|^2\, 1_{\{|\Delta X_s|\leq 1\}}\, 1_{\{\tau>0\}} + |\Delta X_\tau|^2\, 1_{\{|\Delta X_\tau|\leq 1\}}\, 1_{\{\tau>0\}}\right]\\
&\leq 2 \sup_{\substack{y\in[-M,\,M]\\ t\in[0,\,T]}} |\partial_x g^n|^2(t,\, y)\, \cdot (M+1).
\end{aligned} \tag{5.110}$$

It follows that $W^n\, 1_{[0,\,\tau]} \in \mathcal{G}^2(\mu^X)$ (see e.g. Lemma B.21-1., and consequently, by Proposition 3.66 of [**77**],

$$I^n_{t\wedge\tau} \text{ is a purely discontinuous square integrable martingale.} \tag{5.111}$$

On the other hand, $W^n \in \mathcal{G}^2_{\mathrm{loc}}(\mu^X)$, and by Theorem 11.12, point 3), in [**73**], it follows that

$$\langle I^n,\, I^n\rangle_t = \int_{]0,\,t]\times\mathbb{R}} |W^n_s(x)|^2\,\nu^X(ds\,dx) - \sum_{0<s\leq t} |\hat{W}^n_s|^2 \leq \int_{]0,\,t]\times\mathbb{R}} |W^n_s(x)|^2\,\nu^X(ds\,dx). \tag{5.112}$$

Taking into account (5.111), we can apply Doob inequality. Using estimates (5.110), (5.112) and (5.111), we get

$$\begin{aligned}\mathbb{P}\left[\sup_{t\in[0,\,T]} |I^n_{t\wedge\tau}| > \varepsilon_0\right] &\leq \frac{1}{\varepsilon_0^2}\,\mathbb{E}\left[|I^n_{T\wedge\tau}|^2\right]\\ &= \frac{1}{\varepsilon_0^2}\,\mathbb{E}\left[\langle I^n,\, I^n\rangle_{T\wedge\tau}\right]\\ &\leq \frac{2\,(M+1)}{\varepsilon_0^2}\sup_{\substack{y\in[-M,\,M]\\ t\in[0,\,T]}} |\partial_x g^n|^2(t,\,y).\end{aligned}$$

Therefore, since $\partial_x g^n \to 0$ in C^0 as n goes to infinity,

$$\lim_{n\to\infty}\mathbb{P}\left[\sup_{t\in[0,\,T]} |I^n_{t\wedge\tau}| > \varepsilon_0\right] = 0.$$

□

We continue the proof of Theorem 5.6.14. The restriction of the map Γ^F to $C^{1,2}$ is given by (5.103), taking into account (5.105) and Lemma 5.6.12. It remains to prove items (a) and (b).

(a) We have to prove that, for any continuous local martingale N, we have

$$\begin{aligned}&\Big[F(\cdot, X) - \int_0^\cdot \partial_x F(s, X_{s-})\,dM_s\\ &- \int_{]0,\,\cdot]\times\mathbb{R}} \{F(s, X_{s-}+x) - F(s, X_{s-}) - x\,\partial_x F(s, X_{s-})\}\,1_{\{|x|>1\}}\,\mu^X(ds\,dx)\\ &- \int_{]0,\,\cdot]\times\mathbb{R}} \{F(s, X_{s-}+x) - F(s, X_{s-}) - x\,\partial_x F(s, X_{s-})\}\,1_{\{|x|\leq 1\}}\,(\mu^X-\nu^X)(ds\,dx), N\Big]\\ &= 0.\end{aligned}$$

We set

$$\begin{aligned}Y_t &= \int_{]0,\,t]\times\mathbb{R}} W_s(x)\,1_{\{|x|\leq 1\}}\,(\mu^X-\nu^X)(ds\,dx),\\ Z_t &= \int_{]0,\,t]\times\mathbb{R}} W_s(x)\,1_{\{|x|>1\}}\,\mu^X(ds\,dx).\end{aligned}$$

with

$$W_s(x) = F(s, X_{s-}+x) - F(s, X_{s-}) - x\,\partial_x F(s, X_{s-}).$$

Since Z is a bounded variation process (X has almost surely a finite number of jumps larger than one) and N is continuous, Proposition 5.3.14 insures that

$$[Z, N] = 0.$$

By Proposition 5.2.8, $W^2 1_{\{|x|\le 1\}} * \mu^X \in \mathcal{A}^+_{\text{loc}}$, therefore $W 1_{\{|x|\le 1\}}$ belongs $\mathcal{G}^2_{\text{loc}}(\mu^X)$ as well, see Lemma B.21-2. In particular, by Theorem B.22-(iii), Y is a purely discontinuous (square integrable) local martingale. Recalling that an $(\mathcal{F}_t)$-local martingale, null at zero, is a purely discontinuous martingale if and only if it is $(\mathcal{F}_t)$-orthogonal (see Remark 5.6.2), from Proposition 5.3.8-(i) we have

$$[Y, N] = 0.$$

From Proposition 5.3.8-(iii), and the fact that $[M, N]$ is continuous, it follows that

$$\left[\int_0^{\cdot} \partial_x F(s, X_{s-})\, dM_s, N\right] = \int_0^{\cdot} \partial_x F(s, X_{s-})\, d\,[M, N]_s\,.$$

Therefore it remains to check that

$$[F(\cdot, X), N]_t = \int_0^{\cdot} \partial_x F(s, X_{s-})\, d\,[M, N]_s\,. \tag{5.113}$$

To this end, we evaluate the limit of

$$\begin{aligned}
&\frac{1}{\varepsilon}\int_0^t (F((s+\varepsilon)\wedge t, X_{(s+\varepsilon)\wedge t}) - F(s, X_s))\,(N_{(s+\varepsilon)\wedge t} - N_s)\,ds\\
&= \frac{1}{\varepsilon}\int_0^t (F((s+\varepsilon)\wedge t, X_{(s+\varepsilon)\wedge t}) - F((s+\varepsilon)\wedge t, X_s))\,(N_{(s+\varepsilon)\wedge t} - N_s)\,ds\\
&\quad + \frac{1}{\varepsilon}\int_0^t (F((s+\varepsilon)\wedge t, X_s) - F(s, X_s))\,(N_{(s+\varepsilon)\wedge t} - N_s)\,ds\\
&=: I_1(\varepsilon, t) + I_2(\varepsilon, t).
\end{aligned}$$

Concerning the term $I_1(\varepsilon, t)$, it can be decomposed as

$$I_1(\varepsilon, t) = I_{11}(\varepsilon, t) + I_{12}(\varepsilon, t) + I_{13}(\varepsilon, t),$$

where

$$\begin{aligned}
I_{11}(\varepsilon, t) &= \frac{1}{\varepsilon}\int_0^t \partial_x F(s, X_s)\,(N_{(s+\varepsilon)\wedge t} - N_s)(X_{(s+\varepsilon)\wedge t} - X_s)\,ds,\\
I_{12}(\varepsilon, t) &= \frac{1}{\varepsilon}\int_0^t (\partial_x F((s+\varepsilon)\wedge t, X_s) - \partial_x F(s, X_s))\cdot\\
&\qquad \cdot (N_{(s+\varepsilon)\wedge t} - N_s)(X_{(s+\varepsilon)\wedge t} - X_s)\;ds,\\
I_{13}(\varepsilon, t) &= \frac{1}{\varepsilon}\int_0^t \left(\int_0^1 (\partial_x F((s+\varepsilon)\wedge t, X_s + a(X_{(s+\varepsilon)\wedge t} - X_s))\right.\\
&\cdot \qquad \left. -\partial_x F((s+\varepsilon)\wedge t, X_s))\,da\right)(N_{(s+\varepsilon)\wedge t} - N_s)(X_{(s+\varepsilon)\wedge t} - X_s)\,ds.
\end{aligned}$$

Notice that the brackets $[X, X]$, $[X, N]$ and $[N, N]$ exist. Indeed, $[X, X]$ exists by definition, and $[N, N]$ exists by Proposition 5.3.8-(i). Concerning $[X, N]$, it can be decomposed as

$$[X, N] = [M, N] + [A, N],$$

where $[M, N]$ exists by Proposition 5.3.8-(i) and $[A, N] = 0$ by assumption, since A comes from the weak Dirichlet decomposition of X.

Then, from Corollary 5.4.4-2) and Proposition 5.4.7-(iii) we have

$$I_{11}(\varepsilon,\, t) \underset{\varepsilon\to 0}{\longrightarrow} \int_0^t \partial_x F(s, X_{s-})\, d[M,N]_s \quad \text{u.c.p.} \tag{5.114}$$

At this point, we have to prove the u.c.p. convergence to zero of the remaining terms $I_{12}(\varepsilon,t)$, $I_{13}(\varepsilon,t)$, $I_2(\varepsilon,t)$. First, since $\partial_x F$ is uniformly continuous on each compact, we have

$$|I_{12}(\varepsilon,\, t)| \leq \rho\Big(\partial_x F\Big|_{[0,\,T]\times\mathbb{K}^X};\, \varepsilon\Big)\,\sqrt{[X,X]^{ucp}_\varepsilon [N,N]^{ucp}_\varepsilon}, \tag{5.115}$$

where $\mathbb{K}^X$ is the (compact) set $\{X_t(\omega) : t \in [0,\, T]\}$. When ε goes to zero, the modulus of continuity component in (5.115) converges to zero a.s., while the remaining term u.c.p. converges to $\sqrt{[X,X]_t[N,N]_t}$ by definition. Therefore,

$$I_{12}(\varepsilon,\, t) \underset{\varepsilon\to 0}{\longrightarrow} 0 \quad \text{u.c.p.} \tag{5.116}$$

Let us then evaluate $I_{13}(t,\, \varepsilon)$. Since $[X,X]^{ucp}_\varepsilon$, $[N,N]^{ucp}_\varepsilon$ u.c.p. converge, there exists of a sequence (ε_n) such that $[X,X]^{ucp}_{\varepsilon_n}$, $[N,N]^{ucp}_{\varepsilon_n}$ converge uniformly a.s. respectively to $[X,X]$, $[N,N]$. We fix a realization ω outside a null set. Let $\gamma > 0$. We enumerate the jumps of $X(\omega)$ on $[0,\, T]$ by $(t_i)_{i\geq 0}$. Let $M = M(\omega)$ such that

$$\sum_{i=M+1}^{\infty} |\Delta X_{t_i}|^2 \leq \gamma^2.$$

We define

$$\begin{aligned} A(\varepsilon_n, M) &= \bigcup_{i=1}^{N}]t_i - \varepsilon,\, t_i] \\ B(\varepsilon_n, M) &= [0,\, T] \setminus A(\varepsilon_n, M). \end{aligned}$$

The term $I_{13}(\varepsilon_n,\, t)$ can be decomposed as the sum of two terms:

$$I^A_{13}(\varepsilon_n,\, t) = \sum_{i=1}^{M} \int_{t_i-\varepsilon_n}^{t_i} \frac{ds}{\varepsilon_n}\, 1_{]0,\,t]}(s)\, (X_{(s+\varepsilon_n)\wedge t} - X_s)(N_{(s+\varepsilon_n)\wedge t} - N_s)\cdot$$

$$\cdot \int_0^1 (\partial_x F((s+\varepsilon_n)\wedge t,\, X_s + a(X_{(s+\varepsilon_n)\wedge t} - X_s)) - \partial_x F((s+\varepsilon_n)\wedge t,\, X_s))\, da,$$

$$I^B_{13}(\varepsilon_n,\, t) = \frac{1}{\varepsilon_n}\int_{]0,\,t]} (X_{(s+\varepsilon_n)\wedge t} - X_s)(N_{(s+\varepsilon_n)\wedge t} - N_s)\, R^B(\varepsilon_n, s, t, M)\, ds,$$

with

$$R^B(\varepsilon_n, s, t, M) =$$

$$1_{B(\varepsilon_n,M)}(s) \int_0^1 [\partial_x F((s+\varepsilon_n)\wedge t,\, X_s + a(X_{(s+\varepsilon_n)\wedge t} - X_s)) - \partial_x F((s+\varepsilon_n)\wedge t,\, X_s)]\, da.$$

By Remark 5.3.13, we have for every s, t,

$$R^B(\varepsilon_n, s, t, M) \leq \rho\bigg(\partial_x F\Big|_{[0,\,T]\times\mathbb{K}^X},\, \sup_l \sup_{\substack{r,a\in[t_{l-1},\,t_l] \\ |r-a|\leq\varepsilon_n}} |X_a - X_r|\bigg),$$

so that Lemma 5.3.12 applied successively to the intervals $[t_{l-1}, t_l]$ implies

$$R^B(\varepsilon_n, s, t, M) \leq \rho\big(\partial_x F\big|_{[0,T]\times \mathbb{K}^X}, 3\gamma\big).$$

Then

$$|I_{13}^B(\varepsilon_n, t)| \leq \rho\big(\partial_x F\big|_{[0,T]\times \mathbb{K}^X}, 3\gamma\big)\sqrt{[N,N]^{ucp}_{\varepsilon_n}(T)\,[X,X]^{ucp}_{\varepsilon_n}(T)},$$

and we get

$$\limsup_{n\to\infty}\sup_{t\in[0,T]}|I_{13}^B(\varepsilon_n, t)| \leq \rho\big(\partial_x F\big|_{[0,T]\times \mathbb{K}^X}, 3\gamma\big)\sqrt{[N,N]_T\,[X,X]_T}. \tag{5.117}$$

Concerning $I_{13}^A(\varepsilon_n, t)$, we apply Lemma 5.3.11 to $Y = (Y^1, Y^2, Y^3) = (t, X, N)$ and

$$\phi(y_1, y_2) = (y_1^2 - y_2^2)\,(y_1^3 - y_2^3)\int_0^1 [\partial_x F(y_1^1,\, y_2^2 + a(y_1^2 - y_2^2)) - \partial_x F(y_1^1,\, y_2^2)]\,da.$$

Then $I_{13}^A(\varepsilon_n, t)$ converges uniformly in $t \in [0, T]$, as n goes to infinity, to

$$\sum_{i=1}^M 1_{]0,\,t]}(t_i)\,(X_{t_i}-X_{t_i-})(N_{t_i}-N_{t_i-})\int_0^1[\partial_x F(t_i,\, X_{t_i-}+a(X_{t_i}-X_{t_i-}))-\partial_x F(t_i,\, X_{t_i-})]\,da. \tag{5.118}$$

In particular, (5.118) equals zero since N is a continuous process. Then, recalling (5.117), we have

$$\limsup_{n\to\infty}\sup_{t\in[0,T]}|I_{13}(\varepsilon_n, t)| \leq \rho(\partial_x F, 3\,\gamma)\sqrt{[N,N]_T\,[X,X]_T},$$

and, by the arbitrariness of γ, we conclude that

$$\limsup_{n\to\infty}\sup_{t\in[0,T]}|I_{13}(\varepsilon_n, t)| = 0. \tag{5.119}$$

It remains to show the u.c.p. convergence to zero of $I_2(\varepsilon, t)$, as $\varepsilon \to 0$. To this end, let us write it as the sum of the two terms

$$I_{21}(\varepsilon, t) = \frac{1}{\varepsilon}\int_0^t (F(s+\varepsilon, X_s) - F(s, X_s))\,(N_{(s+\varepsilon)\wedge t} - N_s)\,ds,$$
$$I_{22}(\varepsilon, t) = \frac{1}{\varepsilon}\int_0^t (F((s+\varepsilon)\wedge t, X_s) - F(s+\varepsilon, X_s))\,(N_{(s+\varepsilon)\wedge t} - N_s)\,ds.$$

Concerning $I_{21}(\varepsilon, t)$, it can be written as

$$I_{21}(\varepsilon, t) = \int_{]0,\,t]} J_\varepsilon(r)\,dN_r \tag{5.120}$$

with

$$J_\varepsilon(r) = \int_{[(r-\varepsilon)_+,\,r[} \frac{F(s+\varepsilon, X_s) - F(s, X_s)}{\varepsilon}\,ds.$$

Since $J_\varepsilon(r) \to 0$ pointwise, it follows from the Lebesgue dominated convergence theorem that

$$\int_0^T J_\varepsilon^2(r)\,d\langle N,\, N\rangle_r \xrightarrow{\mathbb{P}} 0 \qquad \text{as } \varepsilon \to 0. \tag{5.121}$$

Therefore, according to [**82**], Problem 2.27 in Chapter 3,

$$\lim_{\varepsilon\to 0}\sup_{t\in[0,T]}|I_{21}(\varepsilon,t)|=0. \tag{5.122}$$

As far as $I_{22}(\varepsilon,t)$ is concerned, we have

$$\begin{aligned}|I_{22}(\varepsilon,\,t)| &\le \frac{1}{\varepsilon}\int_{t-\varepsilon}^{t}|F(t,X_s)-F(s+\varepsilon,X_s)|\,|N_t-N_s|\,ds\\ &\le 2\,\rho\big(F\big|_{[0,\,T]\times\mathbb{K}^X},\varepsilon\big)\,||N||_\infty\end{aligned}$$

and we get

$$\limsup_{\varepsilon\to 0}\sup_{t\in[0,T]}|I_{22}(\varepsilon,t)|=0. \tag{5.123}$$

This concludes the proof of item (a).

(b) Let F^n be a sequence of $C^{1,2}$ functions such that $F^n\to F$ and $\partial_x F^n\to\partial_x F$, uniformly on every compact subset. From Lemma 5.6.12, the process $\Gamma^{F^n}(t)$ in (5.103) equals

$$\begin{aligned}&\int_0^t \partial_s F^n(s,X_s)\,ds+\int_0^t \partial_x F^n(s,X_s)\,d^-A_s+\int_0^t \partial^2_{xx}F^n(s,X_s)\,d[X,X]^c_s\\ &+\int_{[0,\,t]\times\mathbb{R}}(F^n(s,X_{s-}+x)-F^n(s,X_{s-})-x\,\partial_x F^n(s,X_{s-}))\,1_{\{|x|\le 1\}}\nu^X(ds\,dx),\end{aligned}$$

which is predictable, see Remark 5.6.13. Since, by Theorem 5.6.14, point (a), the map $\Gamma^F: C^{0,1}\to\mathbb{D}^{ucp}$ is continuous, Γ^{F^n} converges to Γ^F u.c.p. Then Γ^F is predictable because it is the u.c.p. limit of predictable processes. □

5.6.3. A class of particular weak Dirichlet processes. The notion of Dirichlet process is a natural extension of the one of semimartingale only in the continuous case. Indeed, if X is a càdlàg process, which is also Dirichlet, then $X=M+A'$ with $[A',A']=0$, and therefore A' is continuous because of Lemma 5.3.9. This class does not include all the càdlàg semimartingale $S=M+V$, perturbed by a zero quadratic variation process A'. Indeed, if V is not continuous, $S+A'$ is not necessarily a Dirichlet process, even though X is a weak Dirichlet process. Notice that, in general, it is even not a special weak Dirichlet process, since V is generally not predictable.

We propose then the following natural extension of the semimartingale notion in the weak Dirichlet framework.

Definition 5.6.16. We say that X is an $(\mathcal{F}_t)$-*particular weak Dirichlet process* if it admits a decomposition $X=M+A$, where M is an $(\mathcal{F}_t)$-local martingale, $A=V+A'$ with V being a bounded variation adapted process and A' a continuous adapted $(\mathcal{F}_t)$-orthogonal process such that $A'_0=0$.

Remark 5.6.17.

(1) A particular weak Dirichlet process is a weak Dirichlet process. Indeed by Proposition 5.3.14 we have $[V,N]=0$, so

$$[A'+V,N]=[A',N]+[V,N]=0.$$

(2) There exist processes that are special weak Dirichlet and not particular weak Dirichlet. As a matter of fact, let for instance consider the deterministic process $A_t = 1_{\mathbb{Q}\cap[0,T]}(t)$. Then A is predictable and $[A, N] = 0$ for any N continuous local martingale, since, the fact that $A_t \equiv 0$ $d\mathbb{P}\, dt$ a.e. implies that $[A, N]^{ucp}_\varepsilon \equiv 0$. Moreover, since A is totally discontinuous, it can not have bounded variation, so that A is special weak Dirichlet but not particular weak Dirichlet.

In Propositions 5.6.18, 5.6.19 and Corollary 5.6.22 we extend some properties valid for semimartingales to the case of particular weak Dirichlet processes.

Proposition 5.6.18. *Let X be an $(\mathcal{F}_t)$-adapted càdlàg process satisfying assumption* (5.4). *X is a particular weak Dirichlet process if and only if there exist a continuous local martingale M^c, a predictable process $\alpha = \alpha^S + A'$, where α^S is predictable with bounded variation, A' is an $(\mathcal{F}_t)$-adapted continuous orthogonal process, $\alpha^S_0 = A'_0 = 0$, and*

$$X = M^c + \alpha + (x\, 1_{\{|x|\le 1\}}) * (\mu^X - \nu^X) + (x\, 1_{\{|x|>1\}}) * \mu^X. \tag{5.124}$$

In this case,

$$\Delta\alpha_t = \left(\int_{|x|\le 1} x\, \hat{\nu}^X_t(dx)\right), \quad t \in [0,\, T], \tag{5.125}$$

where $\hat{\nu}^X$ has been defined in (5.99).

Proof. If we suppose that decomposition (5.124) holds, then X is a particular weak Dirichlet process satisfying

$$X = M + V + A', \quad M = M^c + (x\, 1_{\{|x|\le 1\}}) * (\mu^X - \nu^X), \quad V = \alpha^S + (x\, 1_{\{|x|>1\}}) * \nu^X.$$

Conversely, suppose that $X = M+V+A'$ is a particular weak Dirichlet process. Since $S = M + V$ is a semimartingale, by Theorem 11.25 in [**73**], it can be decomposed as

$$S = S^c + \alpha^S + (x\, 1_{\{|x|\le 1\}}) * (\mu^S - \nu^S) + (x\, 1_{\{|x|>1\}}) * \mu^S. \tag{5.126}$$

In (5.126) μ^S is the jump measure of S and ν^S is the associated Lévy system, S^c is a continuous local martingale, and α^S is a predictable process with finite variation such that $\alpha^S_0 = 0$ and

$$\Delta\alpha^S_s = \left(\int_{|x|\le 1} x\, \hat{\nu}^S_s(dx)\right).$$

Consequently, since A' is adapted and continuous, with $A'_0 = 0$, we have

$$X = S + A' = S^c + (\alpha^S + A') + (x\, 1_{\{|x|\le 1\}}) * (\mu^X - \nu^X) + (x\, 1_{\{|x|>1\}}) * \mu^X$$

and (5.124) holds with $\alpha = \alpha^S + A'$ and $M^c = S^c$. The process α is $(\mathcal{F}_t)$-orthogonal. Indeed, for every $(\mathcal{F}_t)$-local martingale N, $[A', N] = 0$ and $[\alpha^S, N] = 0$ by Proposition 5.3.14. On the other hand, since $\Delta\alpha = \Delta\alpha^S$, (5.125) follows. □

The following condition on X will play a fundamental role in the sequel:

$$|x|\, 1_{\{|x|>1\}} * \mu^X \ \in \mathcal{A}^+_{\mathrm{loc}}. \tag{5.127}$$

Proposition 5.6.19. *Let X be an $(\mathcal{F}_t)$-particular weak Dirichlet process verifying condition* (5.4)*. X is a special weak Dirichlet process if and only if* (5.127) *holds.*

Proof. Suppose the validity of (5.127). We can decompose

$$(x\,1_{\{|x|>1\}}) * \mu^X = (x\,1_{\{|x|>1\}}) * (\mu^X - \nu^X) + (x\,1_{\{|x|>1\}}) * \nu^X.$$

Using the notation of (5.124), by additivity we get

$$X = M + A, \quad M = M^c + M^d, \quad A = \alpha + (x\,1_{\{|x|>1\}}) * \nu^X, \tag{5.128}$$

where $M^d = x * (\mu^X - \nu^X)$. In particular M and A are well-defined.

Since the process $\alpha + (x\,1_{\{|x|>1\}}) * \nu^X$ is predictable, given a local martingale N, $[A, N] = 0$ by Proposition 5.6.18 and again from the fact that $(x\,1_{\{|x|>1\}}) * \nu^X$ has bounded variation. Consequently X is a special weak Dirichlet process.

Conversely, let $X = M + V + A'$ be a particular weak Dirichlet process, with V bounded variation. We suppose that X is a special weak Dirichlet process. Since $[A', N] = 0$ for every continuous local martingale, then by additivity $X - A'$ is still a special weak Dirichlet process, A' being continuous adapted. But $X - A' = M + V$ is a semimartingale, and by Proposition 5.6.11 it is a special semimartingale. By Corollary 11.26 in [**73**],

$$|x|\,1_{\{|x|>1\}} * \mu^S \in \mathcal{A}^+_{\mathrm{loc}},$$

where μ^S is the jump measure of S. On the other hand, since A' is continuous, μ^S coincides with μ^X and (5.127) holds. □

We recall the following result on the stochastic integration theory, for a proof see Proposition B.30.

Proposition 5.6.20. *Let $W \in \mathcal{G}_{\mathrm{loc}}(\mu^X)$, and define $M^d_t = \int_{]0,t]\times\mathbb{R}} W_s(x)\,(\mu^X - \nu^X)(ds\,dx)$. Let moreover (Z_t) be a predictable process such that*

$$\sqrt{\sum_{s\le \cdot} Z_s^2 |\Delta M^d_s|^2} \in \mathcal{A}^+_{\mathrm{loc}}. \tag{5.129}$$

Then $\int_0^\cdot Z_s\,dM^d_s$ is a local martingale and equals

$$\int_{]0,\cdot]\times\mathbb{R}} Z_s\,W_s(x)\,(\mu^X - \nu^X)(ds\,dx). \tag{5.130}$$

Remark 5.6.21. Recalling that $\sqrt{[M,M]} \in \mathcal{A}^+_{\mathrm{loc}}$ for any local martingale M (see, e.g. Theorem 2.34 and Proposition 2.38 in [**77**]), condition (5.129) is verified for instance if Z is locally bounded.

Remark 5.6.22. Let X be a finite quadratic variation process of the type (5.124). Let $F : [0,T] \times \mathbb{R} \to \mathbb{R}$ be a $C^{0,1}$-real valued function with partial derivative $\partial_x F$. Then, formula (5.104) in Theorem 5.6.14 can be rewritten as

$$F(t, X_t) = F(0, X_0) + \int_0^t \partial_x F(s, X_s)\,dM^c_s$$

$$+\int_{]0,t]\times\mathbb{R}} (F(s,X_{s-}+x)-F(s,X_{s-}))\,1_{\{|x|\le 1\}}\,(\mu^X-\nu^X)(ds\,dx)$$
$$+\int_{]0,t]\times\mathbb{R}} (F(s,X_{s-}+x)-F(s,X_{s-})-x\,\partial_x F(s,X_{s-}))\,1_{\{|x|>1\}}\,\mu^X(ds\,dx)+\Gamma^F(t). \tag{5.131}$$

Indeed, setting

$$M^d_t=\int_{[0,t]\times\mathbb{R}} x\,1_{\{|x|\le 1\}}(\mu^X-\nu^X)(ds\,dx),$$

by Propositions 5.6.20, taking into account Remark 5.6.21, we have

$$\int_0^t \partial_x F(s,X_{s-})\,dM^d_s=\int_{]0,t]\times\mathbb{R}} x\,\partial_x F(s,X_{s-})\,1_{\{|x|\le 1\}}\,(\mu^X-\nu^X)(ds\,dx).$$

5.6.4. Stability of special weak Dirichlet processes under $C^{0,1}$ transformation. At this point, we investigate the stability properties of the class of special weak Dirichlet processes. We start with an important property.

Proposition 5.6.23. *Let X be an $(\mathcal{F}_t)$-special weak Dirichlet process with its canonical decomposition $X=M^c+M^d+A$. We suppose that conditions* (5.4), (5.127) *are verified. Then*

$$M^d_s=\int_{]0,s]\times\mathbb{R}} x\ (\mu^X-\nu^X)(dt\,dx). \tag{5.132}$$

Proof. Taking into account assumption (5.4), Corollary 5.2.5 together with condition (5.127) insures that the right-hand side of (5.132) is well-defined. By definition, it is the unique purely discontinuous local martingale whose jumps are indistinguishable from

$$\int_{\mathbb{R}} x\,\mu^X(\{t\},dx)-\int_{\mathbb{R}} x\,\nu^X(\{t\},dx).$$

It remains to prove that

$$\Delta M^d_t=\int_{\mathbb{R}} x\,\mu^X(\{t\},dx)-\int_{\mathbb{R}} x\,\nu^X(\{t\},dx),\ \text{up to indistinguishability.} \tag{5.133}$$

We have

$$\Delta M^d_t=\Delta X_t-\Delta A_t,\ t\ge 0,$$

Being A predictable, $\Delta A={}^p(\Delta A)$, see Corollary A.24. Now, by Corollary 1.29 in [**77**], for any local martingale L starting from zero, ${}^p(\Delta L)=0$; so for any predictable time τ we have

$$\begin{aligned}\Delta A_\tau\,1_{\{\tau<\infty\}}&=\mathbb{E}\left[\Delta X_\tau\,1_{\{\tau<\infty\}}|\mathcal{F}_{\tau-}\right]\\&=\mathbb{E}\left[\int_{\mathbb{R}} x\,\mu^X(\{\tau\},dx)\,1_{\{\tau<\infty\}}\middle|\mathcal{F}_{T-}\right]\\&=\int_{\mathbb{R}} x\,\nu^X(\{\tau\},dx)\,1_{\{\tau<\infty\}}\quad a.s.,\end{aligned}$$

where for the latter equality we have used Proposition B.11-b). Previous arguments make use of a small abuse of terminology. In order to get them rigorous one can take $\Omega_n \in \mathcal{F}_{\tau-}$ such that $\cup_n \Omega_n \cup \{\tau < \infty\} = \{\tau < \infty\}$ a.s.

The Predictable Section Theorem (see e.g. Proposition A.13) insures that ΔA_t and $\int_{\mathbb{R}} x\, \nu^X(\{t\}, dx)$ are indistinguishable. Since $\Delta X_t = \int_{\mathbb{R}} x\, \mu^X(\{t\}, dx)$, by additivity, (5.133) is established. □

Lemma 5.6.24. *Let X be a càdlàg process satisfying condition* (5.127)*. Let also $F : [0, T] \times \mathbb{R} \to \mathbb{R}$ be a function of class $C^{0,1}$ such that*

$$\int_{]0,\cdot]\times\mathbb{R}} |F(t, X_{t-}+x) - F(t, X_{t-}) - x\, \partial_x F(t, X_{t-})|\, 1_{\{|x|>1\}}\, \mu^X(dt\, dx) \in \mathcal{A}^+_{\rm loc}. \quad (5.134)$$

Then

$$\int_{]0,\cdot]\times\mathbb{R}} x\, \partial_x F(t, X_{t-}) 1_{\{|x|>1\}}\, \mu^X(dt\, dx) \ \in \mathcal{A}^+_{\rm loc}, \quad (5.135)$$

$$\int_{]0,\cdot]\times\mathbb{R}} |F(t, X_{t-} + x) - F(t, X_{t-})|\, 1_{\{|x|>1\}}\, \mu^X(dt\, dx) \in \mathcal{A}^+_{\rm loc}. \quad (5.136)$$

Remark 5.6.25. Condition (5.134) is automatically verified if X is a càdlàg process satisfying (5.127) and $F : [0, T] \times \mathbb{R} \to \mathbb{R}$ is a function of C^1 class with $\partial_x F$ bounded.

Proof. Condition (5.127) together the fact that the process $(\partial_x F(t, X_{t-}))$ is locally bounded implies (5.135); then condition (5.136) follows from (5.135) and (5.134). □

Theorem 5.6.26. *Let X be an $(\mathcal{F}_t)$-special weak Dirichlet process of finite quadratic variation with its canonical decomposition $X = M^c + M^d + A$. Assume that condition* (5.134) *holds. Then, for every $F : [0, T] \times \mathbb{R} \to \mathbb{R}$ of class $C^{0,1}$, we have*

(1) *$Y_t = F(t, X_t)$ is an $(\mathcal{F}_t)$-special weak Dirichlet process, with decomposition $Y = M^F + A^F$, where*

$$M^F_t = F(0, X_0) + \int_0^t \partial_x F(s, X_s)\, d(M^c + M^d)_s + \int_{]0,\,t]\times\mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}) - x\, \partial_x F(s, X_{s-}))\, (\mu^X - \nu^X)(ds\, dx),$$

and $A^F : C^{0,1} \to \mathbb{D}^{ucp}$ is a linear map such that, for every $F \in C^{0,1}$, A^F is a predictable $(\mathcal{F}_t)$-orthogonal process.

(2) *If moreover condition* (5.127) *holds, M^F reduces to*

$$M^F_t = F(0, X_0) + \int_0^t \partial_x F(s, X_s)\, dM^c_s + \int_{]0,\,t]\times\mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}))\, (\mu^X - \nu^X)(ds\, dx).$$

Proof. (1) For every F of class $C^{0,1}$, we set

$$A^F = \Gamma^F + \bar{V}^F, \quad (5.137)$$

where Γ^F has been defined in Theorem 5.6.14, and

$$\bar{V}_t^F := \int_{]0,\,t]\times\mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}) - x\,\partial_x F(s, X_{s-}))\, 1_{\{|x|>1\}}\, \nu^X(ds\,dx),$$

which is well defined by assumption (5.134).

The map $F \mapsto A^F$ is linear since $F \mapsto \Gamma^F$ and $F \mapsto \bar{V}^F$ are linear. Given $F \in C^{0,1}$, A^F is an $(\mathcal{F}_t)$-orthogonal process by Theorem 5.6.14, point (a), taking into account that $[\bar{V}^F, N] = 0$ by Proposition 5.3.14. Using decomposition (5.137), Theorem 5.6.14, point (b), and the fact that $\bar{V}$ is predictable, it follows that A^F is predictable.

(2) It remains to show that

$$\int_0^t \partial_x F(s, X_{s-})\, dM_s^d = \int_{]0,t]\times\mathbb{R}} x\,\partial_x F(s, X_{s-})\,(\mu^X - \nu^X)(ds\,dx).$$

This follows from Proposition 5.6.20 and Proposition 5.6.23, taking into account Remark 5.6.21. $\square$

Remark 5.6.27. In Theorem 5.6.26 condition (5.127) is verified for instance if X is a particular weak Dirichlet process, see Proposition 5.6.19.

5.6.5. The case of special weak Dirichlet processes without continuous local martingale. We end this section by considering the case of special weak Dirichlet processes with canonical decomposition $X = M + A$ where $M = M^d$ is a purely discontinuous local martingale. In particular there is no continuous martingale part. In this framework, under the assumptions of Theorem 5.6.26, if assumption (5.127) in verified, then item (2) of the theorem says that

$$F(t, X_t) = F(0, X_0) + \int_{]0,\,t]\times\mathbb{R}} (F(s, X_{s-}+x) - F(s, X_{s-}))\,(\mu^X - \nu^X)(ds\,dx) + A^F(t). \tag{5.138}$$

Since in the above formula no derivative appears, a natural question appears: is it possible to state a chain rule (5.138) when F is not of class $C^{0,1}$? '?

Indeed we have the following result, which does not suppose any weak Dirichlet structure on X.

Proposition 5.6.28. *Let X be an adapted càdlàg process. Let $F : [0, T] \times \mathbb{R} \to \mathbb{R}$ be a continuous function such that the following holds.*

(i) *$F(t, X_t) = B_t + A'_t$, where B has bounded variation and A' is a continuous $(\mathcal{F}_t)$-orthogonal process;*

(ii) $\int_{]0,\,\cdot]\times\mathbb{R}} |F(s, X_{s-} + x) - F(s, X_{s-})|\, \mu^X(ds\,dx) \ \in \mathcal{A}^+_{\mathrm{loc}}$.

Then $F(t, X_t)$ is an $(\mathcal{F}_t)$-special weak Dirichlet process with decomposition

$$F(t, X_t) = F(0, X_0) + \int_{]0,\,t]\times\mathbb{R}} (F(s, X_{s-}+x) - F(s, X_{s-}))\,(\mu^X - \nu^X)(ds\,dx) + A^F(t), \tag{5.139}$$

and A^F is a predictable $(\mathcal{F}_t)$-orthogonal process.

Remark 5.6.29.

(i) We remark that assumption (i) in Proposition 5.6.28 implies that $\sum_{s \leq T} |F(s, X_{s-} + \Delta X_s) - F(s, X_{s-})| < \infty$ a.s.

(ii) Condition (i) is always verified if $(F(s, X_s))$ is a bounded variation process. Indeed, in this case $B_t = \sum_{s \leq t} \Delta F(s, X_s)$ and $A'_t = F(t, X_t) - \sum_{s \leq t} \Delta F(s, X_s)$. The process A' is continuous by definition, and is $(\mathcal{F}_t)$-orthogonal being of finite variation, see Proposition 5.3.14. Moreover, since $(F(t, X_t))$ is of finite variation, the same holds for B.

Proof. By item (i) of Remark 5.6.29, the process $Y_t = \sum_{s \leq t} \Delta F(s, X_s)$ has bounded variation. Then, by item (ii) of Remark 5.6.29, one can always decompose $F(t, X_t)$ as

$$F(t, X_t) = \bar{B}_t + \bar{A}'_t,$$

where $\bar{B}$ and $\bar{A}'$ are respectively the bounded variation process and the continuous, $(\mathcal{F}_t)$-orthogonal process, given by

$$\bar{B}_t := \sum_{s \leq t} \Delta F(s, X_s), \tag{5.140}$$

$$\bar{A}'_t := B_t - \sum_{s \leq t} \Delta F(s, X_s) + A'_t. \tag{5.141}$$

Recalling the definition of the jump measure μ^X, and using condition (ii), we get

$$\begin{aligned}
\bar{B}_t &= F(t, X_{t-} + \Delta X_t) - F(t, X_{t-}) \\
&= \int_{]0,\, t] \times \mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}))\, \mu^X(ds\, dx) \\
&= \int_{]0,\, t] \times \mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}))\, (\mu^X - \nu^X)(ds\, dx) \\
&+ \int_{]0,\, t] \times \mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}))\, \nu^X(ds\, dx).
\end{aligned}$$

Finally, decomposition (5.139) holds with

$$A^F(t) := \bar{A}'_t + \int_{]0,\, t] \times \mathbb{R}} (F(s, X_{s-} + x) - F(s, X_{s-}))\, \nu^X(ds\, dx). \tag{5.142}$$

The process A^F in (5.142) is clearly predictable. The $(\mathcal{F}_t)$-orthogonality property of A^F follows from the orthogonality of A' and by Proposition 5.3.14, noticing that the integral term in (5.142) is a bounded variation process. □

Remark 5.6.30. Let (X_t) be a pure jump process, in the sense that $X_t = X_0 + \sum_{0 < s \leq t} \Delta X_s$, with a finite number of jumps on each compact. This happens for instance when X is generated by a marked point process (T_n, β_{T_n}) (see e.g. Chapter III, Section 2 b., in [**77**]), where $(T_n)_n$ are increasing random times such that

$$T_n \in]0,\, \infty[, \quad \lim_{n \to \infty} T_n = +\infty.$$

In that case, for any function F of C^0 class, we have

$$F(t, X_t) = F(0,\, X_0) + \sum_{s\le t}(F(s,\, X_{s-} + \Delta X_s) - F(s,\, X_{s-})),$$

so that item (i) in Proposition 5.6.28 holds with $B_t = F(0,\, X_0) + \sum_{s\le t}(F(s,\, X_{s-} + \Delta X_s) - F(s,\, X_{s-}))$, $A'_t = 0$. We suppose moreover that

$$\text{(ii')} \quad \int_{]0,\, t]\times\mathbb{R}} |F(s, X_{s-} + x) - F(s, X_{s-})| 1_{\{|x|>1\}}\, \mu^X(ds\, dx) \ \in \mathcal{A}^+_{\mathrm{loc}}.$$

In that case also item (ii) of Proposition 5.6.28 holds.

Indeed taking into account Definition 5.2.2 and Remark 5.2.3-(i), we consider a localizing sequence $(\tau_n)_{n\ge 1}$ for the process (X_{t-}), which is locally bounded. Fix $\tau = \tau_n$ and let M such that $\sup_{t\in[0,\, T]} |X_{(t-)\wedge\tau}\, 1_{\{\tau>0\}}| \le M$. We have a.s.

$$\begin{aligned}
&\sum_{0<s\le\tau\wedge T} 1_{\{|\Delta X_s|\le 1\}}\, |F(s, X_{s-} + \Delta X_s) - F(s, X_{s-})| \\
&= \sum_{0<s\le\tau\wedge T} 1_{\{|\Delta X_s|\le 1\}}\, 1_{\{\tau>0\}}\, |F(s, X_{s-} + \Delta X_s) - F(s, X_{s-})| \\
&\le 2 \sum_{0<s\le\tau\wedge T} \sup_{y\in[-(M+1),\, (M+1)]} |F(s, y)|\, 1_{\{\tau>0\}} < \infty.
\end{aligned}$$

When X fulfills condition (5.127), condition (ii)' holds for instance if $x \mapsto F(t, x)$ has linear growth, uniformly in t.

Chapter 6

Special weak Dirichlet processes and BSDEs driven by a random measure

6.1. Introduction

This chapter considers a forward BSDE driven by a random measure, when the underlying forward process X is a special semimartingale, or even more generally, a special weak Dirichlet process. Given a solution (Y, Z, U), often Y appears to be of the type $u(t, X_t)$ where u is a deterministic function. In this chapter we identify Z and U in terms of u, by applying the stochastic calculus with respect to (special) weak Dirichlet processes developed in Chapter 5.

Given some filtration $(\mathcal{F}_t)$, we recall that a special weak Dirichlet process is a process of the type $X = M + A$, where M is a $(\mathcal{F}_t)$-local martingale and A is a $(\mathcal{F}_t)$-predictable orthogonal process, see Definition 5.6.6. When A has bounded variation, then X is a special $(\mathcal{F}_t)$-semimartingale. The decomposition of a special weak Dirichlet process is unique, see Proposition 5.6.8. A significant result of Chapter 5 is the chain rule stated in Theorem 5.6.26, concerning the expansion of $F(t, X_t)$, where X is a special weak Dirichlet process of finite quadratic variation and F is of class $C^{0,1}$. If we know a priori that $F(t, X_t)$ is the sum of a bounded variation process and a continuous $(\mathcal{F}_t)$-orthogonal process, then the chain rule only requires F to be continuous; in that case no assumptions are required on the càdlàg process X, see Proposition 5.6.28.

As we have already mentioned, we will focus on forward BSDEs, which constitute a particular case of BSDEs in its general form. BSDEs have been deeply studied since the seminal paper [**98**] by Pardoux and Peng. In [**98**], as well as in many subsequent papers, the standard Brownian motion is the driving process (Brownian context) and the concept of BSDE is based on a non-linear martingale representation theorem with respect to the corresponding Brownian filtration. A recent monograph on the subject is Pardoux and Rascanu [**100**]. BSDEs driven by processes with jumps have also been investigated: two classes of such equations appear in the literature. The first one relates to BSDEs where the Brownian motion is replaced by a general càdlàg martingale M, see, among others, Buckdahn [**20**], El Karoui and Huang [**50**], Carbone, Ferrario and Santacroce [**22**]. An alternative version of BSDEs with a discontinuous driving term is the one associated to an integer-valued random measure μ, with corresponding compensator ν. In this case the BSDE is driven by a continuous martingale M and a compensated random measure $\mu - \nu$. In that equation naturally appears a purely discontinuous martingale which is a stochastic integral with respect to $\mu-\nu$, see, e.g., Xia [**131**], Buckdahn and Pardoux [**21**], Tang and Li [**128**]. A recent monograph on BSDEs driven by Poisson random measures is Delong [**39**]. Connections between the martingale and the random measure driven BSDEs are illustrated by Jeanblanc, Mania, Santacroce and Schweizer [**80**].

In this chapter we will focus on BSDEs driven by random measures (we will use the one-dimensional formalism for simplicity). Besides μ and ν appear three driving random elements: a continuous martingale M, a non-decreasing adapted continuous process ζ and a predictable random measure λ on $\Omega \times [0,T] \times \mathbb{R}$, equipped with the usual product σ-fields. Given a square integrable random variable ξ, and two measurable functions $\tilde{g} : \Omega \times [0,T] \times \mathbb{R}^2 \to \mathbb{R}$, $\tilde{f} : \Omega \times [0,T] \times \mathbb{R}^3 \to \mathbb{R}$, the equation takes the following form:

$$
\begin{aligned}
Y_t = \xi + \int_{]t,\,T]} \tilde{g}(s,\, Y_{s-},\, Z_s)\, d\zeta_s + \int_{]t,\,T]\times\mathbb{R}} \tilde{f}(s,\, e,\, Y_{s-},\, U_s(e))\, \lambda(ds\, de)\\
- \int_{]t,\,T]} Z_s\, dM_s - \int_{]t,\,T]\times\mathbb{R}} U_s(e)\, (\mu-\nu)(ds\, de). \qquad (6.1)
\end{aligned}
$$

As we have anticipated before, the unknown of (6.1) is a triplet (Y, Z, U) where Y, Z are adapted and U is a predictable random field. The Brownian context of Pardoux-Peng appears as a particular case, setting $\mu = \lambda = 0$, $\zeta_s \equiv s$. There M is a standard Brownian motion and ξ is measurable with respect to the Brownian σ-field at terminal time. In that case the unknown can be reduced to (Y, Z), since U can be arbitrarily chosen. Another significant subcase of (6.1) arises when only the purely discontinuous driving term appears, i.e. M and ζ vanish; under this simpler structure the related BSDE can be approached by an iterative method: a significant example is represented by BSDEs driven by a marked point process, as in Confortola, Fuhrman and Jacod [**29**].

When the random dependence of $\tilde{f}$ and $\tilde{g}$ is provided by a Markov solution X of a forward SDE, and ξ is a real function of X at the terminal time T, then the BSDE (6.1) is called a forward BSDE, the one that we have anticipated at the beginning. This generally constitutes a stochastic representation of a partial integro-differential equation (PIDE). In the Brownian case, when X is the solution of a classical SDE with diffusion coefficient σ, then the PIDE reduces to a semilinear parabolic PDE. If $v : [0,T] \times \mathbb{R} \times \mathbb{R}$ is a classical (smooth) solution of the mentioned PDE, then

$Y_s = v(s, X_s)$, $Z_s = \sigma(s, X_s)\,\partial_x v(s, X_s)$, generate a solution to the forward BSDE, see e.g. [**99**] and Peng [**101**], [**102**]. In the general case when the forward BSDEs are also driven by random measures, similar results have been established, for instance by Barles, Buckdahn and Pardoux [**10**], for the jump-diffusion case, and by Confortola and Fuhrman [**28**], for the purely discontinuous case, i.e. when no Brownian noise appears. In the context of martingale driven forward BSDEs, a first approach to the probabilistic representation has been carried on in Laachir and Russo [**90**].

Conversely, solutions of forward BSDEs generate solutions of PIDEs in the viscosity sense. More precisely, for each given couple $(t, x) \in [0, T] \times \mathbb{R}$, consider an underlying process X given by the solution $X^{t,x}$ of an SDE starting at x at time t. Let $(Y^{t,x}, Z^{t,x}, U^{t,x})$ be a family of solutions of the forward BSDE. In that case, under reasonable general assumptions, the function $v(t, x) := Y_t^{t,x}$ is a viscosity solution of the related PIDE. A demanding task consists in characterizing the couple $(Z, U) := (Z^{t,x}, U^{t,x})$, in terms of v; this is generally called the *identification problem* of (Z, U). In the continuous case, this was for instance the object of Fuhrman and Tessitore [**68**]: the authors show that if $v \in C^{0,1}$, then $Z_s = \partial_x v(s, X_s)$; under more general assumptions, they also associate Z with a generalized gradient of v. At our knowledge, in the discontinuous case, the problem of the identification of the martingale integrands pair (Z, U) has not been deeply investigated, except for particular situations, as for instance the one treated in [**28**].

In the present chapter we discuss the mentioned identification problem in a quite general framework by means of the calculus related to weak Dirichlet processes. When Y is a deterministic function v of a special semimartingale X, related in a specific way to the random measure μ, we apply the chain rule in Theorem 5.6.26 in order to identify the pair (Z, U). This is the object of Proposition 6.4.12. The result remains valid if X is a special weak Dirichlet process with finite quadratic variation. In the purely discontinuous framework, i.e. when in the BSDE (6.1) M and ζ vanish, we make use of the chain rule in Proposition 5.6.28, which, for a general càdlàg process X, allows to express $v(t, X_t)$ without requiring any differentiability on v. In particular Proposition 5.6.28 does not ask X to be a special weak Dirichlet process, provided we have some a priori information on the structure of $v(t, X_t)$. The identification in that case is stated in Proposition 6.4.18. We remark that in most of the literature on BSDEs, the measures ν, λ and ζ of equation (6.1) are non-atomic in time. A challenging case arises when one or more of those predictable processes have jumps in time. Well-posedness of BSDEs in that case has been partially discussed in Bandini [**2**] in the purely discontinuous case, and in a slightly different context by Cohen and Elliott [**26**], for BSDEs driven by a countable sequence of square-integrable martingales. Our approach to the identification problem also applies to forward BSDEs presenting predictable jumps.

The chapter is organized as follows. In Section 6.2 we fix the notations. In Section 6.3 we introduce a class of stochastic processes X related in a specific way to a given integer-valued random measure μ, and we provide some technical results on the related stochastic integration. Section 6.4 is devoted to solve the identification problem.

6.2. Notations

In what follows, we are given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ a positive horizon T and a filtration $(\mathcal{F}_t)_{t\geq 0}$, satisfying the usual conditions. Let $\mathcal{F} = \mathcal{F}_T$. Given a topological space E, in the sequel $\mathcal{B}(E)$ will denote the Borel σ-field associated with E. $\mathcal{P}$ (resp. $\tilde{\mathcal{P}} = \mathcal{P} \otimes \mathcal{B}(\mathbb{R})$) will denote the predictable σ-field on $\Omega \times [0, T]$ (resp. on $\tilde{\Omega} = \Omega \times [0, T] \times \mathbb{R}$). Analogously, we set $\mathcal{O}$ (resp. $\tilde{\mathcal{O}} = \mathcal{O} \otimes \mathcal{B}(\mathbb{R})$) as the optional σ-field on $\Omega \times [0, T]$ (resp. on $\tilde{\Omega}$). Moreover, $\tilde{\mathcal{F}}$ will be σ-field $\mathcal{F} \otimes \mathcal{B}([0, T] \times \mathbb{R})$, and we will indicate by $\mathcal{F}^{\mathbb{P}}$ the completion of $\mathcal{F}$ with the $\mathbb{P}$-null sets. We set $\tilde{\mathcal{F}}^{\mathbb{P}} = \mathcal{F}^{\mathbb{P}} \otimes \mathcal{B}([0, T] \times \mathbb{R})$. By default, all the stochastic processes will be considered with parameter $t \in [0,\, T]$. By convention, any càdlàg process defined on $[0,\, T]$ is extended to $\mathbb{R}_+$ by continuity.

A bounded variation process X on $[0,\, T]$ will be said to be with integrable variation if the expectation of its total variation is finite. $\mathcal{A}$ (resp. $\mathcal{A}_{\rm loc}$) will denote the collection of all adapted processes with integrable variation (resp. with locally integrable variation), and $\mathcal{A}^+$ (resp $\mathcal{A}^+_{\rm loc}$) the collection of all adapted integrable increasing (resp. adapted locally integrable) processes. The significance of locally is the usual one which refers to localization by stopping times, see e.g. (0.39) of Jacod's book [**77**].

We will indicate by $C^{0,1}$ the space of all functions

$$u : [0,\, T] \times \mathbb{R} \to \mathbb{R}, \quad (t, x) \mapsto u(t, x)$$

that are continuous together their derivative $\partial_x u$. $C^{0,1}$ is equipped with the topology of uniform convergence on each compact of u and $\partial_x u$.

The concept of random measure will be extensively used throughout the chapter. We refer the reader to Appendices A and B, where we have summarized the concepts needed in the following on the general theory of stochastic processes and on the stochastic integration with respect to random measures.

6.3. A class of stochastic processes X related in a specific way to an integer-valued random measure μ

Let μ be an integer-valued random measure on $[0,\, T] \times \mathbb{R}$, and ν a "good" version of the compensator of μ, as constructed in Proposition B.11-c). Set

$$\begin{aligned} D &= \{(\omega, t) : \mu(\omega, \{t\} \times \mathbb{R}) > 0\}, \\ J &= \{(\omega, t) : \nu(\omega, \{t\} \times \mathbb{R}) > 0\}, \\ K &= \{(\omega, t) : \nu(\omega, \{t\} \times \mathbb{R}) = 1\}. \end{aligned}$$

Remark 6.3.1. D is a thin set, J is the predictable support of D, and K is the largest predictable subset of D, see Proposition B.6 and Theorem B.10. The definition of predictable support of a random set is recalled in Definition A.25.

We formulate now an assumption on a generic càdlàg process X which will be related in the sequel to the integer-valued random measure μ.

Hypothesis 6.3.2. $X = X^i + X^p$, where X^i (resp. X^p) is a càdlàg quasi-left continuous adapted process (resp. càdlàg predictable process).

Proposition 6.3.3. *Let X be a càdlàg adapted process fulfilling Hypothesis 6.3.2. Then the two properties below hold.*

(i) $\Delta X^p \, 1_{\{\Delta X^i \neq 0\}} = 0$ *and* $\Delta X^i \, 1_{\{\Delta X^p \neq 0\}} = 0$, *up to an evanescent set.*

(ii) $\{\Delta X \neq 0\}$ *is the disjointed union of the random sets* $\{\Delta X^p \neq 0\}$ *and* $\{\Delta X^i \neq 0\}$.

Proof. (i) Recalling Propositions A.17 (resp. A.19), there exist a sequence of predictable times $(T_n^p)_n$ (resp. totally inaccessible times $(T_n^i)_n$) that exhausts the jumps of X^p (resp. X^i). On the other hand, $\Delta X^p_{T_n^i} = 0$ a.s. for every n, see Proposition A.17 (resp. $\Delta X^i_{T_n^p} = 0$ a.s. for every n, see Definition A.18), so that

$$\begin{aligned}\Delta X^i \, 1_{\{\Delta X^p \neq 0\}} &= \Delta X^i \, 1_{\cup_n [[T_n^p]]} = 0,\\ \Delta X^p \, 1_{\{\Delta X^i \neq 0\}} &= \Delta X^p \, 1_{\cup_n [[T_n^i]]} = 0.\end{aligned}$$

(ii) From point (i) we get

$$\begin{aligned}\{\Delta X \neq 0\} &= \{(\Delta X^i + \Delta X^p) \neq 0\}\\ &= \{(\Delta X^i \, 1_{\{\Delta X^p = 0\}} + \Delta X^p 1_{\{\Delta X^p \neq 0\}}) \neq 0\}\\ &= \{\Delta X^i 1_{\{\Delta X^p = 0\}} \neq 0\} \cup \{\Delta X^p \neq 0\}\\ &= \{\Delta X^i \neq 0\} \cup \{\Delta X^p \neq 0\}.\end{aligned}$$

□

Proposition 6.3.4. *Let X be a càdlàg adapted process satisfying Hypothesis 6.3.2. Then the properties below hold.*

(1) $\{(\omega, t) : \nu^X(\omega, \{t\} \times \mathbb{R}) > 0\} = \{\Delta X^p \neq 0\}$;

(2) $\{\Delta X^p \neq 0\}$ *is the largest predictable subset of* $\{\Delta X \neq 0\}$ *(up to an evanescent set).*

Proof. (1) $\{\Delta X \neq 0\}$ is the support of the random measure μ^X (see e.g. Proposition B.8). By Theorem B.10, the predictable support of $\{\Delta X \neq 0\}$ is given by $\{(\omega, t) : \nu^X(\{t\} \times \mathbb{R}) > 0\}$.

On the other hand, by Proposition 6.3.3-(ii), $\{\Delta X \neq 0\}$ is the disjointed union of $\{\Delta X^p \neq 0\}$ and $\{\Delta X^i \neq 0\}$. Since X^i is a càdlàg quasi-left continuous process, by Proposition A.26 we know that the predictable support of $\{\Delta X^i \neq 0\}$ is evanescent. By Definition A.25 of predictable support, taking into account the additivity of the predictable projection operator, ${}^p\left(1_{\{\Delta X \neq 0\}}\right) = 1_{\{\Delta X^p \neq 0\}}$, and this concludes the proof.

(2) By Proposition 6.3.3-(ii),

$$\{\Delta X^p \neq 0\} \subset \{\Delta X \neq 0\}. \tag{6.2}$$

Since $\{(\omega, t) : \nu^X(\{t\} \times \mathbb{R}) = 1\}$ is the largest predictable subset of $\{\Delta X \neq 0\}$ (see again Theorem B.10), it follows from point (1) and (6.2) that $\{\Delta X^p \neq 0\}$ coincides with $\{(\omega, t) : \nu^X(\{t\} \times \mathbb{R}) = 1\}$. □

Remark 6.3.5. We remark that item (2) in Proposition 6.3.4 has an interest in itself but will not be used in the sequel.

Proposition 6.3.6. *Let X satisfy Hypothesis 6.3.2 with decomposition $X = X^i + X^p$. Let moreover $(S_n)_n$ be a sequence of predictable times exhausting the jumps of X^p. Then*

$$\nu^X(\{S_n\}, dx) = \mu^X(\{S_n\}, dx) \text{ for any } n, \text{ a.s.} \tag{6.3}$$

Remark 6.3.7. Since $\{\Delta X^p \neq 0\}$ is a predictable thin set (see Definition A.4), the existence of a sequence of predictable times exhausting the jumps of X^p is a well-known fact, see Proposition A.17 and Definition A.1 for the definition of an exhausting sequence.

Proof. Let us fix n and let $(E_m)_m$ be a sequence of measurable subsets of $\mathbb{R}$ which is a π-class generating $\mathcal{B}(\mathbb{R})$. Since X^i is a càdlàg quasi-left continuous adapted process and S_n is a predictable time, then $\Delta X^i_{S_n} = 0$ a.s., see Definition A.18. This implies that $\Delta X_{S_n} = \Delta X^p_{S_n}$ a.s. by Hypothesis 6.3.2. Consequently, for every m we have

$$1_{E_m}(\Delta X^p_{S_n}) = 1_{E_m}(\Delta X_{S_n}) = \int_{\mathbb{R}} 1_{E_m}(x)\, \mu^X(\{S_n\}, dx) \text{ a.s.} \tag{6.4}$$

On the other hand, by Proposition B.11-b) and (6.4) we have

$$\begin{aligned}
\int_{\mathbb{R}} 1_{E_m}(x)\, \nu^X(\{S_n\}, dx) &= \mathbb{E}\left[\int_{\mathbb{R}} 1_{E_m}(x)\, \mu^X(\{S_n\}, dx) \Big| \mathcal{F}_{S_n-}\right] \\
&= \mathbb{E}\left[1_{E_m}(\Delta X^p_{S_n}) \Big| \mathcal{F}_{S_n-}\right] \\
&= 1_{E_m}(\Delta X^p_{S_n}) \text{ a.s.,}
\end{aligned}$$

where the latter equality follows from Corollary A.24. By (6.4), there exists a $\mathcal{P}$-measurable null set $\mathcal{N}_m$ such that

$$\int_{\mathbb{R}} 1_{E_m}(x)\, \nu^X(\{S_n\}, dx) = \int_{\mathbb{R}} 1_{E_m}(x)\, \mu^X(\{S_n\}, dx) \text{ for every } \omega \notin \mathcal{N}_m.$$

Define $\mathcal{N} = \cup_m \mathcal{N}_m$, then

$$\int_{\mathbb{R}} 1_{E_m}(x)\, \nu^X(\{S_n\}, dx) = \int_{\mathbb{R}} 1_{E_m}(x)\, \mu^X(\{S_n\}, dx) \text{ for every } m \text{ and } \omega \notin \mathcal{N}.$$

Then the claim follows by a monotone class argument, see Theorem 21, Chapter 1, in Dellacherie and Meyer [**38**]. □

We now recall an important notion of measure associated with μ, given in formula (3.10) in [**77**].

Definition 6.3.8. Let $(\tilde{\Omega}_n)$ be a partition of $\tilde{\Omega}$ constituted by elements of $\tilde{\mathcal{O}}$. $M^{\mathbb{P}}_{\mu}$ denotes the σ-finite measure on $(\tilde{\Omega}, \tilde{\mathcal{F}}^{\mathbb{P}})$, such that for every $W : \tilde{\Omega} \to \mathbb{R}$ positive, bounded, $\tilde{\mathcal{F}}^{\mathbb{P}}$-measurable function,

$$M^{\mathbb{P}}_{\mu}(W\, 1_{\tilde{\Omega}_n}) = \mathbb{E}\big[W\, 1_{\tilde{\Omega}_n} * \mu_T\big]. \tag{6.5}$$

Remark 6.3.9. Formally speaking we have $M^{\mathbb{P}}_{\mu}(d\omega, ds, de) = d\mathbb{P}(\omega)\, \mu(\omega, ds, de)$.

In the sequel we will formulate the following assumption for a generic càdlàg process Y with respect to the random measure μ.

Hypothesis 6.3.10. Y is a càdlàg adapted process satisfying $\{\Delta Y \neq 0\} \subset D$. Moreover, there exists a $\tilde{\mathcal{P}}$-measurable map $\tilde{\gamma} : \Omega \times]0,\, T] \times \mathbb{R} \to \mathbb{R}$ such that

$$\Delta Y_t(\omega)\, 1_{]0,\, T]}(t) = \tilde{\gamma}(\omega, t, \cdot) \quad dM^{\mathbb{P}}_{\mu}\text{-a.e.} \tag{6.6}$$

Example 6.3.11. Theorem 3.89 in [**77**] states an Itô formula which transforms a special semimartingale X into a special semimartingale $F(X_t)$ through a C^2 function $F : \mathbb{R} \to \mathbb{R}$. There the process $Y = X$ is supposed to fulfill Hypothesis 6.3.10.

Remark 6.3.12. Let us suppose that μ is the jump measure of a càdlàg process X. Hypothesis 6.3.10 holds for $Y = X$, with $\tilde{\gamma}(t, \omega, x) = x$.

The role of Hypothesis 6.3.10 is clarified by the following proposition.

Proposition 6.3.13. *Let Y be a càdlàg adapted process satisfying Hypothesis 6.3.10. Then, there exists a null set $\mathcal{N}$ such that, for every Borel function $\varphi : [0,\, T] \times \mathbb{R} \to \mathbb{R}_+$ satisfying $\varphi(s, 0) = 0$ for every $s \in [0,\, T]$, we have*

$$\sum_{0<s\leq T} \varphi(s, \Delta Y_s(\omega)) = \int_{]0,\, T]\times\mathbb{R}} \varphi(s, \tilde{\gamma}(\omega, s, e))\, \mu(\omega, ds\, de), \quad \omega \notin \mathcal{N}. \tag{6.7}$$

Proof. Taking into account that $\{\Delta Y \neq 0\} \subset D$ and the fact that $\varphi(s, 0) = 0$, it will be enough to prove that

$$\sum_{0<s\leq T} \varphi(s, \Delta Y_s(\omega))\, 1_D(\omega, s) = \int_{]0,\, T]\times\mathbb{R}} \varphi(s, \tilde{\gamma}(\omega, s, e))\, \mu(\omega, ds\, de), \quad \omega \notin \mathcal{N}, \tag{6.8}$$

for every Borel function $\varphi : [0,\, T] \times \mathbb{R} \to \mathbb{R}_+$.

Let $(I_m)_m$ be a sequence of subsets of $[0,\, T] \times \mathbb{R}$, which is a π-system generating $\mathcal{B}([0,\, T]) \otimes \mathcal{B}(\mathbb{R})$. Setting $\varphi_m(s, x) = 1_{I_m}(s, x)$, for every m we will show that

$$\sum_{0<s\leq T} \varphi_m(s, \Delta Y_s)\, 1_D(\cdot, s) = \int_{]0,\, T]\times\mathbb{R}} \varphi_m(s, \tilde{\gamma}(\cdot, s, e))\, \mu(\cdot, ds\, de), \quad \text{a.s.} \tag{6.9}$$

As a matter of fact, consider a bounded, $(\mathcal{F}_t)$-measurable function $\phi : \Omega \to \mathbb{R}_+$. Identity (6.9) holds if we show that the expectations of both sides against ϕ are equal. We write

$$\mathbb{E}\left[\phi \int_{]0,\, T]\times\mathbb{R}} \varphi_m(s, \tilde{\gamma}(\cdot, s, e))\, \mu(\cdot, ds\, de)\right]$$

$$
\begin{aligned}
&= \int_{\Omega\times]0,\,T]\times\mathbb{R}} d\mathbb{P}(\omega)\,\mu(\omega, ds\,de)\,\phi(\omega)\,\varphi_m(s, \tilde\gamma(\omega, s, e))\\
&= \int_{\Omega\times]0,\,T]\times\mathbb{R}} dM^{\mathbb{P}}_\mu(\omega, s, e)\,\phi(\omega)\,\varphi_m(s, \tilde\gamma(\omega, s, e))\\
&= \int_{\Omega\times]0,\,T]} dM^{\mathbb{P}}_\mu(\omega, s, y)\,\phi(\omega)\,\varphi_m(s, \Delta Y_s(\omega))\\
&= \int_{\Omega\times]0,\,T]\times\mathbb{R}} d\mathbb{P}(\omega)\,\mu(\omega, ds\,de)\,\phi(\omega)\,\varphi_m(s, \Delta Y_s(\omega))\\
&= \int_{\Omega} d\mathbb{P}(\omega)\,\phi(\omega) \sum_{0<s\leq T} 1_D(\omega, s)\,\varphi_m(s, \Delta Y_s(\omega)) \int_{\mathbb{R}} \delta_{\beta_s(\omega)}(dx)\\
&= \mathbb{E}\Bigg[\phi \sum_{0<s\leq T} 1_D(\cdot, s)\,\varphi_m(s, \Delta Y_s)\Bigg],
\end{aligned}
$$

where we have used the form of μ given by (B.3). Therefore, there exists a $\mathcal{P}$-null set $\mathcal{N}_m$ such that

$$
\sum_{0<s\leq T} \varphi_m(s, \Delta Y_s(\omega))\,1_D(\omega, s) = \int_{]0,\,T]\times\mathbb{R}} \varphi_m(s, \tilde\gamma(\cdot, s, e))\,\mu(\omega, ds\,de), \quad \omega \notin \mathcal{N}_m.
$$

Define $\mathcal{N} = \cup_m \mathcal{N}_m$, then for $\varphi = \varphi_m$ for every m we have

$$
\sum_{0<s\leq T} \varphi_m(s, \Delta Y_s(\omega))\,1_D(\omega, s) = \int_{]0,\,T]\times\mathbb{R}} \varphi_m(s, \tilde\gamma(\cdot, s, e))\,\mu(\omega, ds\,de), \quad \omega \notin \mathcal{N}.
$$

By a monotone class argument (see Theorem 21, Chapter 1, in [**38**]) the identity holds for every measurable bounded function $\varphi : [0,\,T]\times\mathbb{R} \to \mathbb{R}$, and therefore for every positive measurable function φ on $[0,\,T]\times\mathbb{R}$ as well. □

We consider an additional assumption on a generic adapted process Z.

Hypothesis 6.3.14. Z is a càdlàg predictable process satisfying $\{\Delta Z \neq 0\} \subset J$.

We have the following result.

Proposition 6.3.15. *Assume that X satisfies Hypothesis 6.3.2, with decomposition $X = X^i + X^p$, where X^i (resp. X^p) fulfills Hypothesis 6.3.10 (resp. Hypothesis 6.3.14). Then, there exists a null set $\mathcal{N}$ such that, for every Borel function $\varphi : [0,\,T]\times\mathbb{R} \to \mathbb{R}_+$ satisfying $\varphi(s, 0) = 0$, $s \in [0,\,T]$, we have, for every $\omega \notin \mathcal{N}$,*

$$
\int_{]0,\,T]\times\mathbb{R}} \varphi(s, x)\,\mu^X(\omega, ds\,dx) = \int_{]0,\,T]\times\mathbb{R}} \varphi(s, \tilde\gamma(\omega, s, e))\,\mu(\omega, ds\,de) + V^\varphi(\omega), \tag{6.10}
$$

with $V^\varphi(\omega) = \sum_{0<s\leq T} \varphi(s, \Delta X^p_s(\omega))$. In particular,

$$
\int_{]0,\,T]\times\mathbb{R}} \varphi(s, x)\,\mu^X(\omega, ds\,dx) \geq \int_{]0,\,T]\times\mathbb{R}} \varphi(s, \tilde\gamma(\omega, s, e))\,\mu(\omega, ds\,de) \quad \text{for every } \omega \notin \mathcal{N}. \tag{6.11}
$$

Identity (6.10) *still holds true when $\varphi : [0,\,T]\times\mathbb{R} \to \mathbb{R}$ and the left-hand side is finite.*

Proof. Let $\varphi : [0, T] \times \mathbb{R} \to \mathbb{R}_+$. Taking into account Proposition 6.3.3-(i) and the fact that $\varphi(s, 0) = 0$, we have, for almost all ω,

$$\begin{aligned}
&\sum_{0<s\leq T} \varphi(s, \Delta X_s(\omega))\\
&= \sum_{0<s\leq T} \varphi(s, \Delta X^i_s(\omega) + \Delta X^p_s(\omega))\, 1_{\{\Delta X^p = 0\}}(\omega, s)\\
&+ \sum_{0<s\leq T} \varphi(s, \Delta X^i_s(\omega) + \Delta X^p_s(\omega))\, 1_{\{\Delta X^p \neq 0\}}(\omega, s)\\
&= \sum_{0<s\leq T} \varphi(s, \Delta X^i_s(\omega))\, 1_{\{\Delta X^p = 0\}}(\omega, s) + \sum_{0<s\leq T} \varphi(s, \Delta X^p_s(\omega))\, 1_{\{\Delta X^p \neq 0\}}(\omega, s)\\
&= \sum_{0<s\leq T} \varphi(s, \Delta X^i_s(\omega)) + \sum_{0<s\leq T} \varphi(s, \Delta X^p_s(\omega)).
\end{aligned}$$

By Proposition 6.3.13 applied to $Y = X^i$, there exists a null set $\mathcal{N}$ such that, for every $\omega \notin \mathcal{N}$, previous expression gives

$$\begin{aligned}
&\int_{]0,\,T]\times\mathbb{R}} \varphi(s, x)\, \mu^X(\omega, ds\, dx)\\
&= \int_{]0,\,T]\times\mathbb{R}} \varphi(s, \tilde{\gamma}(\omega, s, e))\, \mu(\omega, ds\, de) + \sum_{0<s\leq T} \varphi(s, \Delta X^p_s(\omega)).
\end{aligned}$$

The second part of the statement holds decomposing $\varphi = \varphi^+ - \varphi^-$. □

Remark 6.3.16. The result in Proposition 6.3.15 still holds true if φ is a real-valued random function on $\Omega \times [0,\, T] \times \mathbb{R}$.

We will make the following assumption on μ.

Hypothesis 6.3.17.

(i) $D = K \cup (\cup_n [[T^i_n]])$ up to an evanescent set, where $(T^i_n)_n$ are totally inaccessible times such that $[[T^i_n]] \cap [[T^i_m]] = \emptyset$, $n \neq m$;

(ii) for every predictable time S such that $[[S]] \subset K$, $\nu(\{S\}, de) = \mu(\{S\}, de)$ a.s.

Remark 6.3.18. Hypothesis 6.3.17-(i) implies that $J = K$, up to an evanescent set, see Proposition B.13.

Remark 6.3.19. Let ν denote the compensator of μ.

(i) ν admits a disintegration of the type

$$\nu(\omega, ds\, de) = dA_s(\omega)\, \phi(\omega, s, de), \tag{6.12}$$

where ϕ is a random measure from $(\Omega \times [0,\, T], \mathcal{P})$ into $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and A is a right-continuous nondecreasing predictable process, such that $A_0 = 0$, see (B.1).

(ii) Given ν in the form (6.12), then the process A is continuous if and only if $D = \cup_n [[T_n^i]]$, where $(T_n^i)_n$ are totally inaccessible times, see, e.g., Assumption (A) in [**29**]. In this case it follows that $J = K = \emptyset$, and consequently Hypothesis 6.3.17 trivially holds.

For instance A in (6.12) is continuous when μ is a Poisson random measure, see, e.g., Chapter II, Section 4.b in [**79**].

We are ready to state the main result of the section.

Proposition 6.3.20. *Let μ satisfy Hypothesis 6.3.17. Assume that X satisfies Hypothesis 6.3.2, with decomposition $X = X^i + X^p$, where X^i (resp. X^p) fulfills Hypothesis 6.3.10 (resp. Hypothesis 6.3.14). Let $\varphi : \Omega \times [0, T] \times \mathbb{R} \to \mathbb{R}_+$ such that $\varphi(\omega, s, 0) = 0$ for every $s \in [0, T]$, up to indistinguishability, and assume that there exists a $\tilde{\mathcal{P}}$-measurable subset A of $\Omega \times [0, T] \times \mathbb{R}$ satisfying*

$$|\varphi|\, 1_A * \mu^X \in \mathcal{A}^+_{\mathrm{loc}}, \quad |\varphi|^2\, 1_{A^c} * \mu^X \in \mathcal{A}^+_{\mathrm{loc}}. \tag{6.13}$$

Then

$$\int_{]0,\, t] \times \mathbb{R}} \varphi(s, x)\, (\mu^X - \nu^X)(ds\, dx) = \int_{]0,\, t] \times \mathbb{R}} \varphi(s, \tilde{\gamma}(s, e))\, (\mu - \nu)(ds\, de) \quad \text{a.s.} \tag{6.14}$$

Remark 6.3.21. Under condition (6.13), Remark 5.2.6 and inequality (6.11) in Proposition 6.3.15 imply that $\varphi(s, x) \in \mathcal{G}^1_{\mathrm{loc}}(\mu^X)$ and $\varphi(s, \tilde{\gamma}(s, e)) \in \mathcal{G}^1_{\mathrm{loc}}(\mu)$. In particular the two stochastic integrals in (6.14) are well-defined.

Proof. Clearly the result holds if we show that φ verifies (6.14) under one of the two following assumptions:

(i) $|\varphi| * \mu^X \in \mathcal{A}^+_{\mathrm{loc}}$,

(ii) $|\varphi|^2 * \mu^X \in \mathcal{A}^+_{\mathrm{loc}}$.

By localization arguments, it is enough to show it when $|\varphi| * \mu^X \in \mathcal{A}^+$, $|\varphi|^2 * \mu^X \in \mathcal{A}^+$. Below we will consider the first case, the second case will follow from the first one by approaching φ with $\varphi(s, x)\, 1_{\varepsilon < |x| \leq 1/\varepsilon}\, 1_{s \in [0, T]}$ in $\mathcal{L}^2(\mu^X)$, and taking into account the fact that μ^X, restricted to $\varepsilon \leq |x| \leq 1/\varepsilon$, is finite, since μ^X is σ-finite.

Let us define

$$\begin{aligned} M_t &:= \int_{]0,\, t] \times \mathbb{R}} \varphi(\cdot, s, x)\, (\mu^X - \nu^X)(ds\, dx), \\ N_t &:= \int_{]0,\, t] \times \mathbb{R}} \varphi(\cdot, s, \tilde{\gamma}(\cdot, s, e))\, (\mu - \nu)(ds\, de). \end{aligned} \tag{6.15}$$

Notice that the processes M and N are purely discontinuous local martingales, see e.g. Definition B.16. We have to prove that M and N are indistinguishable. To this end, by Corollary A.9, it is enough to prove that $\Delta M = \Delta N$, up to an evanescent set. Observe that

$$\Delta M_s = \int_{\mathbb{R}} \varphi(\cdot, s, x)\, (\mu^X - \nu^X)(\{s\},\, dx)$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} \varphi(\cdot, s, x)\,(1 - 1_J(\cdot, s))\,(\mu^X - \nu^X)(\{s\}, dx) \\
&+ \int_{\mathbb{R}} \varphi(\cdot, s, x)\,1_J(\cdot, s)\,(\mu^X - \nu^X)(\{s\}, dx),
\end{aligned} \tag{6.16}
$$

and

$$
\begin{aligned}
\Delta N_s &= \int_{\mathbb{R}} \varphi(\cdot, s, \tilde{\gamma}(\cdot, s, e))\,(\mu - \nu)(\{s\}, de) \\
&= \int_{\mathbb{R}} \varphi(\cdot, s, \tilde{\gamma}(\cdot, s, e))\,1_J(\cdot, s)\,(\mu - \nu)(\{s\}, de) \\
&+ \int_{\mathbb{R}} \varphi(\cdot, s, \tilde{\gamma}(\cdot, s, e))\,(1 - 1_J(\cdot, s))\,(\mu - \nu)(\{s\}, de).
\end{aligned} \tag{6.17}
$$

By definition of J, for every ω and every s we have

$$
\nu(\omega, \{s\}, de)\,(1 - 1_J(\omega, s)) = 0. \tag{6.18}
$$

Moreover, since J is a predictable thin set, there exists a sequence of predictable times $(R_n)_n$ with disjoint graphs, such that $J = \cup_n [[R_n]]$. We recall that Hypothesis 6.3.17-(i) implies that $J = K$ up to an evanescent set, see Remark 6.3.18. By this fact, and taking into account Hypothesis 6.3.17-(ii), there exists a null set $\mathcal{N}$, such that, for every $n \in \mathbb{N}$, $\omega \notin \mathcal{N}$,

$$
\mu(\omega, \{R_n(\omega)\}, de)\,1_J(\omega, s) = \nu(\omega, \{R_n(\omega)\}, de)\,1_J(\omega, s).
$$

By additivity, it follows that for every $\omega \notin \mathcal{N}$, for every $s \in [0, T]$,

$$
\mu(\omega, \{s\}, de)\,1_J(\omega, s) = \nu(\omega, \{s\}, de)\,1_J(\omega, s). \tag{6.19}
$$

On the other hand, $\{\Delta X^p \neq 0\} \subset J$ by Hypothesis 6.3.14. Recalling that $\{\Delta X^p \neq 0\} = \{(\omega, s) : \nu^X(\{s\} \times \mathbb{R}) > 0\}$ (see Proposition 6.3.4-(1), for almost every ω, for every $s \in [0, T]$, we have

$$
\nu^X(\omega, \{s\}, dx)\,1_J(\omega, s) = \nu^X(\omega, \{s\}, dx)\,1_{\{\Delta X^p \neq 0\}}(\omega, s), \tag{6.20}
$$

so that

$$
\nu^X(\omega, \{s\}, dx)\,(1 - 1_J(\omega, s)) = \nu^X(\omega, \{s\}, dx)\,(1 - 1_{\{\Delta X^p \neq 0\}}(\omega, s)) = 0. \tag{6.21}
$$

Now notice that there always exists a sequence of predictable times exhausting the jumps of X^p, see Remark 6.3.7. By means of Proposition 6.3.6 we can prove, similarly as we did in order to establish (6.19), that for every $\omega \notin \mathcal{N}$ ($\mathcal{N}$ possibly enlarged), for every $s \in [0, T]$,

$$
\mu^X(\omega, \{s\}, dx)\,1_{\{\Delta X^p \neq 0\}}(\omega, s) = \nu^X(\omega, \{s\}, dx)\,1_{\{\Delta X^p \neq 0\}}(\omega, s). \tag{6.22}
$$

Finally, we notice that $\mu^X(\omega, \{s\}, dx)\,1_J(\omega, s) = \mu^X(\omega, \{s\}, dx)\,1_{J \cap \{\Delta X \neq 0\}}(\omega, s)$. Taking into account that X^i is a càdlàg quasi-left continuous process, by Definition A.18 we have

$$
\begin{aligned}
J \cap \{\Delta X \neq 0\} &= (\cup_n [[R_n]] \cap \{\Delta X^i \neq 0\}) \cup (\cup_n [[R_n]] \cap \{\Delta X^p \neq 0\}) \\
&= \cup_n [[R_n]] \cap \{\Delta X^p \neq 0\} = \{\Delta X^p \neq 0\}.
\end{aligned}
$$

This implies for every $\omega \notin \mathcal{N}$, and for every $s \in [0, T]$,

$$\begin{aligned}\mu^X(\omega, \{s\}, dx)\, 1_J(\omega, s) &= \mu^X(\omega, \{s\}, dx)\, 1_{J \cap \{\Delta X \neq 0\}}(\omega, s)\\ &= \mu^X(\omega, \{s\}, dx)\, 1_{\{\Delta X^p \neq 0\}}(\omega, s). \end{aligned} \tag{6.23}$$

Collecting (6.20), (6.22) and (6.23) we conclude that for every $\omega \notin \mathcal{N}$, for every $s \in [0, T]$,

$$\mu^X(\omega, \{s\}, dx)\, 1_J(\omega, s) = \nu^X(\omega, \{s\}, dx)\, 1_J(\omega, s). \tag{6.24}$$

Therefore, for every $\omega \notin \mathcal{N}$, for every $s \in [0, T]$, taking into account (6.18), (6.19), (6.21), (6.24), expressions (6.16) and (6.17) become

$$\Delta M_s = \int_{\mathbb{R}} \varphi(s, x)\, (1 - 1_J(s))\, \mu^X(\{s\},\, dx), \tag{6.25}$$

$$\Delta N_s = \int_{\mathbb{R}} \varphi(s, \tilde{\gamma}(s, e))\, (1 - 1_J(s))\, \mu(\{s\},\, de). \tag{6.26}$$

Now let us prove that, for every $s \in [0, T]$, $\Delta M_s(\omega) = \Delta N_s(\omega)$ for every $\omega \notin \mathcal{N}$, namely up to an evanescent set. Set

$$\varphi_s(\omega, t, x) := \varphi(\omega, t, x)\ (1 - 1_J(\omega, t))\, 1_{\{s\}}(t),$$

then ΔM_s and ΔN_s can be rewritten as

$$\begin{aligned}\Delta M_s(\omega) &= \int_{[0,\, T]\times\mathbb{R}} \varphi_s(\omega, t, x)\, \mu^X(\omega, dt\, dx),\\ \Delta N_s(\omega) &= \int_{[0,\, T]\times\mathbb{R}} \varphi_s(\omega, t, \tilde{\gamma}(\omega, t, e))\, \mu(\omega, dt\, de).\end{aligned}$$

Then, Proposition 6.3.15 applied to the process φ_s implies that (possibly enlarging the null set $\mathcal{N}$),

$$\int_{]0,\, T]\times\mathbb{R}} \varphi_s(\omega, t, x)\, \mu^X(\omega, dt\, dx) = \int_{]0,\, T]\times\mathbb{R}} \varphi_s(t, \tilde{\gamma}(\omega, t, e))\, \mu(\omega, dt\, de) + V^{\varphi_s}(\omega)$$

for every $\omega \notin \mathcal{N}$, or, equivalently, that

$$\int_{\mathbb{R}} \varphi(\omega, s, x)\, \mu^X(\omega, \{s\},\, dx) = \int_{\mathbb{R}} \varphi(\omega, s, \tilde{\gamma}(\omega, s, e))\, \mu(\omega, \{s\},\, de) + V^{\varphi_s}(\omega),$$

for every $\omega \notin \mathcal{N}$, where

$$V^{\varphi_s}(\omega) = \sum_{t \leq T} \varphi_s(\omega, t, \Delta X^p_t(\omega)) = \varphi(\omega, s, \Delta X^p_s(\omega))\ 1_{J^c \cap \{\Delta X^p \neq 0\}}(\omega, s). \tag{6.27}$$

Recalling that $\{\Delta X^p \neq 0\} \subset J$ by Hypothesis 6.3.14, it straightly follows from (6.27) that $V^{\varphi_s}(\omega)$ is zero. In particular, up to an evanescent set, we have

$$\int_{\mathbb{R}} \varphi(\omega, s, x)\, \mu^X(\omega, \{s\},\, dx) = \int_{\mathbb{R}} \varphi(s, \tilde{\gamma}(\omega, s, e))\, \mu(\omega, \{s\},\, de),$$

in other words $\Delta M = \Delta N$ up to an evanescent set, and this concludes the proof. □

We end the section focusing on the case when X is of jump-diffusion type.

Lemma 6.3.22. *Let μ satisfy Hypothesis 6.3.17. Let N be a continuous martingale, and B an increasing predictable càdlàg process, with $B_0 = 0$, such that $\{\Delta B \neq 0\} \subset J$. Let X be a process which is solution of equation*

$$X_t = X_0 + \int_0^t b(s, X_{s-})\, dB_s + \int_0^t \sigma(s, X_s)\, dN_s + \int_{]0,\, t] \times \mathbb{R}} \gamma(s, X_{s-}, e)\, (\mu - \nu)(ds\, de), \tag{6.28}$$

for some given Borel functions $b, \sigma : [0, T] \times \mathbb{R} \to \mathbb{R}$, and $\gamma : [0, T] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that

$$\int_0^t |b(s, X_{s-})|\, dB_s < \infty \text{ a.s.}, \tag{6.29}$$

$$\int_0^t |\sigma(s, X_s)|^2\, d[N, N]_s < \infty \text{ a.s.}, \tag{6.30}$$

$$(\omega, s, e) \mapsto \gamma(s, X_{s-}(\omega), e) \in \mathcal{G}^1_{\mathrm{loc}}(\mu). \tag{6.31}$$

Then X satisfies Hypothesis 6.3.2, with decomposition $X = X^i + X^p$, where

$$X^i_t = \int_{]0,\, t] \times \mathbb{R}} \gamma(s, X_{s-}, e)\, (\mu - \nu)(ds\, de), \tag{6.32}$$

$$X^p_t = X_0 + \int_0^t b(s, X_{s-})\, dB_s + \int_0^t \sigma(s, X_s)\, dN_s. \tag{6.33}$$

Moreover, the process X^i fulfills Hypothesis 6.3.10 with $\tilde{\gamma}(\omega, s, e) = \gamma(s, X_{s-}(\omega), e)\, (1 - 1_K(\omega, s))$, and the process X^p satisfies Hypothesis 6.3.14.

Proof. Since N is continuous, it straight follows from (6.33) that

$$\Delta X^p_s = b(s, X_{s-})\, \Delta B_s. \tag{6.34}$$

We remark that X^i in (6.32) has the same expression as N defined in (6.15) where the integrand $\varphi(\omega, s, \tilde{\gamma}(\omega, s, e))$ is replaced by $\gamma(s, X_{s-}(\omega), e)$. We recall that Hypothesis 6.3.17-(i) implies that $J = K$ up to an evanescent set, see Remark 6.3.18. Similarly as for (6.26), we get

$$\Delta X^i_s = \int_{\mathbb{R}} \gamma(s, X_{s-}, e)\; (1 - 1_K(s))\, \mu(\{s\},\, de). \tag{6.35}$$

Since by Hypothesis 6.3.17 $D \setminus K = \cup_n [[T^i_n]]$, $(T^i_n)_n$ being a sequence of totally inaccessible times with disjoint graphs, (6.35) can be rewritten as

$$\Delta X^i_s(\omega) = \gamma(s, X_{s-}(\omega), \beta_s(\omega))\; 1_{\cup_n [[T^i_n]]}(\omega, s). \tag{6.36}$$

We can easily show that the process X satisfies Hypothesis 6.3.2, namely X^p and X^i are respectively a càdlàg predictable process and a càdlàg quasi-left continuous adapted process. The fact that X^p is predictable straight follow from (6.33). Concerning X^i, let S be a predictable time; it is enough to prove that $\Delta X^i_S\, 1_{\{S < \infty\}} = 0$ a.s., see Definition A.18. Identity (6.36) gives

$$\Delta X^i_S(\omega)\, 1_{\{S<\infty\}} = \gamma(S, X_{S-}(\omega), \beta_S(\omega))\; 1_{\cup_n [[T^i_n]]}(\omega, S(\omega))\, 1_{\{S<\infty\}}. \tag{6.37}$$

Since the graphs of the totally inaccessible times T^i_n are disjoint, $1_{\cup_n [[T^i_n]]}(\omega, S(\omega))\; 1_{\{S<\infty\}} = \sum_n 1_{[[T^i_n]]}(\omega, S(\omega))\; 1_{\{S<\infty\}}$, and the conclusion follows by the definition of a totally inaccessible time, taking into account that S is a predictable time, see Remark A.15.

The process X^p in (6.33) satisfies Hypothesis 6.3.14. Indeed, by (6.34) we have

$$\{\Delta X^p \neq 0\} \subset \{\Delta B \neq 0\} \subset J = K. \tag{6.38}$$

Finally, we show that the process X^i in (6.32) fulfills Hypothesis 6.3.10 with $\tilde{\gamma}(\omega, s, e) = \gamma(s, X_{s-}(\omega), e)\,(1 - 1_K(\omega, s))$. First, the fact that $\{\Delta X^i \neq 0\} \subset D$ directly follows from (6.35). To prove $\Delta X^i_s(\omega) = \tilde{\gamma}(\omega, s, \cdot)$, $dM^{\mathbb{P}}_\mu(\omega, s)$-a.e. it is enough to show that

$$\mathbb{E}\left[\int_{]0,\,T]\times\mathbb{R}} \mu(\omega, ds\, de)\,|\tilde{\gamma}(\omega, s, e) - \Delta X^i_s(\omega)|\right] = 0.$$

To establish this, we see that by the structure of μ it follows that

$$\begin{aligned}&\mathbb{E}\left[\int_{]0,\,T]\times\mathbb{R}} \mu(\omega, ds\, de)\,|\tilde{\gamma}(\omega, s, e) - \Delta X^i_s(\omega)|\right]\\ &= \sum_{s\in]0,\,T]} \mathbb{E}\left[1_D(\cdot, s)\,|\tilde{\gamma}(\cdot, s, \beta_s(\cdot)) - \Delta X^i_s(\cdot)|\right],\end{aligned}$$

which vanishes taking into account (6.36). □

6.4. Application to BSDEs

6.4.1. About BSDEs driven by an integer-valued random measure. Let μ be an integer-valued random measure defined on $[0, T] \times \mathbb{R}$. Let M be a continuous process with $M_0 = 0$. Let $(\mathcal{F}_t)$ be the canonical filtration associated to μ and M, and suppose that M is an $(\mathcal{F}_t)$-local martingale. Let $\tilde{g} : \Omega \times [0, T] \times \mathbb{R}^2 \to \mathbb{R}$ and $\tilde{f} : \Omega \times [0, T] \times \mathbb{R}^3 \to \mathbb{R}$ be two measurable functions. The domain of $\tilde{f}$ (resp. $\tilde{g}$) is equipped with the σ-field $\mathcal{F} \otimes \mathcal{B}([0, T] \times \mathbb{R}^3)$ (resp. $\mathcal{F} \otimes \mathcal{B}([0, T] \times \mathbb{R}^2)$). Let λ be a predictable random measure on $[0, T] \times \mathbb{R}$. Let ζ be a non-decreasing adapted continuous process, and ξ a square integrable random variable. ν will denote a "good" version of the dual predictable projection of μ in the sense of Proposition B.11. In particular, $\nu(\omega, \{t\} \times \mathbb{R}) \leq 1$ identically.

We consider now the general BSDE

$$\begin{aligned}Y_t = \xi + \int_{]t,\,T]} \tilde{g}(s,\, Y_{s-},\, Z_s)\, d\zeta_s + \int_{]t,\,T]\times\mathbb{R}} \tilde{f}(s,\, e,\, Y_{s-},\, U_s(e))\, \lambda(ds\, de)\\ - \int_{]t,\,T]} Z_s\, dM_s - \int_{]t,\,T]\times\mathbb{R}} U_s(e)\, (\mu - \nu)(ds\, de)\end{aligned} \tag{6.39}$$

which constitutes equation (6.1) of the Introduction.

Remark 6.4.1. A general BSDE of type (6.39) is considered for instance in Xia [**131**] (see formula (1.1)), with the following restrictions on the random measures λ and ν:

$\lambda([0,\, T] \times \mathbb{R})$ is a bounded random variable, $\lambda([0,\, t] \times \mathbb{R})$ is continuous w.r.t. t,

$$\nu([0,\, t] \times \mathbb{R}) \ \text{ is continuous w.r.t. } \ t. \tag{6.40}$$

The author proves (see Theorem 3.2 in [**131**]) that under suitable assumptions on the coefficients $(\xi, \tilde{f}, \tilde{g})$ there exists a unique triplet of processes $(Y, Z, U) \in \mathcal{L}^2(\zeta\lambda) \times \mathcal{L}^2(M) \times \mathcal{L}^2(\mu)$, with $\mathbb{E}\big[\sup_{t\in[0,T]} Y_t^2\big] < \infty$, satisfying BSDE (6.39), where

$$\mathcal{L}^2(\zeta\lambda) := \Big\{(Y_t)_{t\in[0,T]} \text{ optional} : \mathbb{E}\Big[\int_0^T Y_s^2\, d\zeta_s\Big] + \mathbb{E}\Big[\int_0^T Y_s^2\, \lambda(ds, \mathbb{R})\Big] < \infty\Big\},$$
$$\mathcal{L}^2(M) := \Big\{(Z_t)_{t\in[0,T]} \text{ predictable} : \mathbb{E}\Big[\int_0^T Z_s^2\, d\langle M\rangle_s\Big] < \infty\Big\},$$

and $\mathcal{L}^2(\mu)$ is the space introduced in (B.20).

In the sequel we will consider stochastic processes related to the random measure μ in the following way.

Hypothesis 6.4.2. *X is an adapted càdlàg process verifying Hypothesis 6.3.2 with decomposition $X = X^i + X^p$, where X^i (resp. X^p) fulfills Hypothesis 6.3.10 with some predictable process $\tilde{\gamma}$ (resp. fulfills Hypothesis 6.3.14), with respect to the random measure μ.*

We consider some important examples.

Example 6.4.3. Let us focus on the BSDE

$$\begin{aligned} Y_t = g(X_T) &+ \int_{]t,\, T]} f(s,\, X_s,\, Y_s,\, Z_s,\, U_s(\cdot))\, ds \\ &- \int_{]t,\, T]} Z_s\, dW_s - \int_{]t,\, T]\times\mathbb{R}} U_s(e)\, (\mu - \nu)(ds\, de), \end{aligned} \tag{6.41}$$

which constitutes a particular case of the BSDE (6.39). This is considered for instance in Barles, Buckdahn and Pardoux [**10**]. Here W is a Brownian motion and $\mu(ds\, de)$ is a Poisson random measure with compensator

$$\nu(ds\, de) = \lambda(de)\, ds, \tag{6.42}$$

where λ is a Borel σ-finite measure on $\mathbb{R} \setminus \{0\}$ and

$$\int_{\mathbb{R}} (1 \wedge |e|^2)\, \lambda(de) < +\infty. \tag{6.43}$$

Poisson random measures have been introduced for instance in Chapter II, Section 4.b in [**79**]. The process X appearing in (6.41) is a Markov process satisfying the SDE

$$dX_s = b(X_s)\, ds + \sigma(X_s)\, dW_s + \int_{\mathbb{R}} \gamma(X_{s-}, e)\, (\mu - \nu)(ds\, de), \quad s \in [t,\, T], \tag{6.44}$$

where $b : \mathbb{R} \to \mathbb{R}$, $\sigma : \mathbb{R} \to \mathbb{R}$ are globally Lipschitz, and $\gamma : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is a measurable function such that, for some real K, and for all $e \in \mathbb{R}$,

$$\begin{cases} |\gamma(x, e)| \le K\, (1 \wedge |e|), \quad x \in \mathbb{R}, \\ |\gamma(x_1, e) - \gamma(x_2, e)| \le K\, |x_1 - x_2|\, (1 \wedge |e|) \quad x_1,\, x_2 \in \mathbb{R}. \end{cases} \tag{6.45}$$

For every starting point $x \in \mathbb{R}$ and initial time $t \in [0, T]$, there is a unique solution to (6.44) denoted $X^{t,x}$ (see [**10**], Section 1). Moreover, modulo suitable assumptions on the coefficients (g, f), it is proved that the BSDE (6.41) admits a unique solution $(Y, Z, U) \in \mathcal{S}^2 \times \mathcal{L}^2 \times \mathcal{L}^2(\mu)$, see Theorem 2.1 in [**10**], where

$$\begin{aligned}
\mathcal{S}^2 :&= \Big\{(Y_t)_{t\in[0,\,T]} \text{ adapted càdlàg} : \Big|\Big|\sup_{t\in[0,\,T]}|Y_t|\Big|\Big|_{L^2(\Omega)} < \infty\Big\},\\
\mathcal{L}^2 :&= \Big\{(Z_t)_{t\in[0,\,T]} \text{ predictable} : \mathbb{E}\Big[\int_0^T Z_s^2\,ds\Big] < \infty\Big\},\\
\mathcal{L}^2(\mu) :&= \Big\{(U_s(\cdot))_{s\in[0,\,T]} \text{ predictable random fields} :\\
&\qquad\qquad \mathbb{E}\Big[\int_{]0,\,T]\times\mathbb{R}} |U_s(e)|^2\,\nu(ds\,de)\Big] < \infty\Big\}.
\end{aligned}$$

When $X = X^{t,x}$ the solution (Y, Z) of (6.41) is denoted $(Y^{t,x}, Z^{t,x})$. In [**10**] it is proved that

$$u(t, x) := Y_t^{t,x}, \quad (t, x) \in [0, T] \times \mathbb{R}, \tag{6.46}$$

satisfies $Y_s^{t,x} = u(s, X_s^{t,x})$ for every $(t, x) \in [0, T] \times \mathbb{R}$, $s \in [t, T]$. □

Lemma 6.4.4. *Let μ and X be respectively the Poisson random measure and the stochastic process satisfying the SDE* (6.44) *in Example 6.4.3. Then μ satisfies Hypothesis 6.3.17 and X fulfills Hypothesis 6.4.2 with respect to μ, with decomposition $X = X^i + X^p$, where*

$$X_t^i = \int_{]0,\,t]\times\mathbb{R}} \gamma(X_{s-}, e)\,(\mu - \nu)(ds\,de), \tag{6.47}$$

$$X_t^p = \int_0^t b(X_s)\,ds + \int_0^t \sigma(X_s)\,dW_s. \tag{6.48}$$

In particular X^i satisfies Hypothesis 6.3.10 with $\tilde{\gamma}(\omega, s, e) = \gamma(X_{s-}(\omega), e)$.

Proof. Our aim is to apply Lemma 6.3.22. We start by noticing that ν in (6.42) is in the form (6.12) with $A_s = s$. Therefore Hypothesis 6.3.17 is verified, see Remark 6.3.19-(ii). On the other hand, the process X satisfies the stochastic differential equation (6.44), which is a particular case of (6.28) when $B_s = s$, $N_s = W_s$, and b, σ, γ are time homogeneous. b and σ verify (6.29), (6.30) since they have linear growth. Condition (6.31) can be verified using the characterization of $\mathcal{G}^1_{\mathrm{loc}}(\mu)$ in Theorem B.19. In that context, setting $W(\omega, s, e) = \gamma(s, X_{s-}(\omega), e)$, we get $\hat{W} = 0$, and we have to verify that $|W|^2 1_{\{|W|\leq 1\}} * \nu + |W| 1_{\{|W|>1\}} * \nu$ belongs to $\mathcal{A}^+_{\mathrm{loc}}$. This follows from (6.43) and (6.45).

Then, by Lemma 6.3.22, X verifies Hypothesis 6.3.2, with decomposition $X = X^i + X^p$, where X^i and X^p are given respectively by (6.47) and (6.48). Moreover, the process X^i fulfills Hypothesis 6.3.10 with $\tilde{\gamma}(\omega, s, e) = \gamma(X_{s-}(\omega), e)$, and the process X^p satisfies Hypothesis 6.3.14. □

When ζ and M vanish, BSDE (6.39) turns out to be driven only by a purely discontinuous martingale, and becomes

$$Y_t = \xi + \int_{]t,\,T]} \tilde{f}(s,\,\omega,\,e,\,Y_{s-},\,U_s(e))\,\lambda(ds\,de) - \int_{]t,\,T]\times\mathbb{R}} U_s(e)\,(\mu-\nu)(ds\,de). \quad (6.49)$$

Below we consider two significant cases, given respectively in Examples 6.4.5 and 6.4.7.

Example 6.4.5. In Confortola and Fuhrman [**28**] the authors study a BSDE driven by an integer-valued random measure μ associated to a given pure jump Markov process X, of the form

$$Y_t = g(X_T) + \int_{]t,\,T]} f(s,\,X_s,\,Y_s,\,U_s(\cdot))\,ds - \int_{]t,\,T]\times\mathbb{R}} U_s(e)\,(\mu-\nu)(ds\,de). \quad (6.50)$$

The underlying process X is generated by a marked point process (T_n, ζ_n), where $(T_n)_n$ are increasing random times such that $T_n \in]0,\,\infty[$, where either the times $(T_n)_n$ are a finite number or $\lim_{n\to\infty} T_n = +\infty$, and ζ_n are random variables in $\mathbb{R}$, see e.g. Chapter III, Section 2 b., in [**77**]. This means that X is a càdlàg process such that $X_t = \zeta_n$ for $t \in [T_n,\,T_{n+1}[$, for every $n \in \mathbb{N}$. In particular, X has a finite number of jumps on each compact. The associated integer-valued random measure μ is the sum of the Dirac measures concentrated at the marked point process (T_n, ζ_n), and can be written as

$$\mu(ds\,de) = \sum_{s\in[0,\,T]} 1_{\{X_{s-}\neq X_s\}}\,\delta_{(s,X_s)}(dt\,de). \quad (6.51)$$

Given a measure μ in the form (6.51), it is related to the jump measure μ^X in the following way: for every Borel subset A of $\mathbb{R}$,

$$\int_{]0,\,T]\times\mathbb{R}} 1_A(e - X_{s-})\,\mu(ds\,de) = \int_{]0,\,T]\times\mathbb{R}} 1_A(x)\,\mu^X(ds\,dx). \quad (6.52)$$

This is for instance explained in Example 3.22 in [**77**]. The pure jump process X then satisfies the equation

$$X_t = X_0 + \sum_{0<s\leq t} \Delta X_s = X_0 + \int_{]0,t]\times\mathbb{R}} (e - X_{s-})\,\mu(ds\,de). \quad (6.53)$$

The compensator of $\mu(ds\,de)$ is

$$\nu(ds\,de) = \lambda(s, X_{s-}, de)\,ds, \quad (6.54)$$

where λ the is the transition rate measure of the process satisfying

$$\sup_{t\in[0,T],\,x\in\mathbb{R}} \lambda(t,x,\mathbb{R}) < \infty, \quad (6.55)$$

see Section 2.1 in [**28**].

Under suitable assumptions on the coefficients (g, f), Theorem 3.4 in [**28**] states that the BSDE (6.50) admits a unique solution $(Y, U) \in \mathcal{L}^2 \times \mathcal{L}^2(\mu)$, where $\mathcal{L}^2(\mu)$ and

$\mathcal{L}^2$ are the spaces introduced in Example 6.4.3. Theorem 4.4 in [**28**] shows moreover that there exists a measurable function $u : [0, T] \times \mathbb{R} \to \mathbb{R}$ such that

$$\forall\, e \in E,\ t \mapsto u(t, e) \text{ is absolutely continuous on } [0, T], \tag{6.56}$$

$$u(s, X_s) \in \mathcal{L}^2 \text{ and } u(s, e) - u(s, X_{s-}) \in \mathcal{L}^2(\mu), \quad s \in [0, T], \tag{6.57}$$

and the unique solution of the BSDE (6.50) can be represented as

$$Y_s = u(s, X_s), \quad s \in [0, T], \tag{6.58}$$

$$U_s(e) = u(s, e) - u(s, X_{s-}), \quad \lambda(s, X_{s-}, de)\, ds\text{-a.e.} \quad s \in [0, T]. \tag{6.59}$$

□

Lemma 6.4.6. *Let X and μ be respectively a pure jump Markov process and the corresponding integer-valued random measure as in Example 6.4.5. Then μ satisfies Hypothesis 6.3.17 and X fulfills Hypothesis 6.4.2 with decomposition $X = X^i$, $X^p = 0$. In particular, X^i satisfies Hypothesis 6.3.10 with $\tilde{\gamma}(\omega, s, e) = e - X_{s-}(\omega)$.*

Proof. Since ν in (6.54) is in the form (6.12) with $A_s = s$, Hypothesis 6.3.17 is verified, see Remark 6.3.19-(ii).

The process $X^i = X$ satisfies (6.53). Recalling the relation (6.52) between μ and μ^X, the continuity of the above mentioned process A also implies that $X = X^i$ is quasi-left continuous, see Corollary B.9. Finally, by definition of μ we have

$$\mathbb{E}\left[\int_{]0,\, T] \times \mathbb{R}} \mu(ds\, de)\, |(e - X_{s-}) - \Delta X_s|\right] = 0,$$

therefore X^i satisfies Hypothesis 6.3.10 with $\tilde{\gamma}(\omega, s, e) = e - X_{s-}(\omega)$. □

We start now describing the second example. In the recent paper Bandini [**2**], one studies the existence and uniqueness for a BSDE driven by a purely discontinuous martingale of the form

$$Y_t = \xi + \int_{]t,\, T]} \tilde{f}(s,\, Y_{s-},\, U_s(\cdot))\, dA_s - \int_{]t,\, T] \times \mathbb{R}} U_s(e)\, (\mu - \nu)(ds\, de). \tag{6.60}$$

Here $\mu(ds\, de)$ is an integer-valued random measure with compensator $\nu(ds\, de) = dA_s\, \phi_s(de)$, where ϕ is a probability kernel and A is a right-continuous nondecreasing predictable process, such that $\hat{\nu}_s(\mathbb{R}) = \Delta A_s \leq 1$ for every s. For any positive constant β, $\mathcal{E}^\beta$ will denote the Doléans-Dade exponential of the process βA. We consider the weighted spaces

$$\mathcal{L}^2_\beta(A) := \Big\{\text{adapted càdlàg processes } (Y_s)_{s \in [0,\, T]},\ \text{s.t.}\, \mathbb{E}\Big[\int_0^T \mathcal{E}^\beta_s |Y_{s-}|^2\, dA_s\Big] < \infty\Big\},$$

$$\mathcal{G}^2_\beta(\mu) := \Big\{\text{predictable processes } (U_s(\cdot))_{s \in [0,\, T]},\ \text{s.t.}$$

$$||U||^2_{\mathcal{G}^2_\beta(\mu)} := \mathbb{E}\Big[\int_{]0,\, T] \times \mathbb{R}} \mathcal{E}^\beta_s\, |U_s(e) - \hat{U}_s|^2\, \nu(ds\, de) + \sum_{s \in]0,\, T]} \mathcal{E}^\beta_s\, |\hat{U}_s|^2 (1 - \Delta A_s)\Big] < \infty\Big\}.$$

A solution to equation (6.60) with data $(\beta, \xi, \tilde{f})$ is a pair $(Y, Z) \in \mathcal{L}^2_\beta(A) \times \mathcal{G}^2_\beta(\mu)$ satisfying equation (6.60). We say that equation (6.60) admits a unique solution in

$\mathcal{L}^2_\beta(A) \times \mathcal{G}^2_\beta(\mu)$ if, given two solutions (Y, U), (Y', U'), we have $Y_t = Y'_t$ $d\mathbb{P} \otimes dA_t$-a.e. and $||U - U'||^2_{\mathcal{G}^2_\beta(\mu)} = 0$ (in particular $||U - U'||^2_{\mathcal{G}^2(\mu)} = 0$).

In [**2**] one requires suitable assumptions on the triplet $(\tilde{f}, \xi, \beta)$. In particular $\tilde{f}$ is of Lipschitz type in the third and fourth variable and ξ is a square integrable random variable with some weight. Moreover, the following technical assumption has to be fulfilled: there exists $\varepsilon \in]0,\, 1[$ such that

$$2\,|L_y|^2\,|\Delta A_t|^2 \leq 1 - \varepsilon, \quad \mathbb{P}\text{−a.s.}, \quad \forall\, t \in [0,\, T], \tag{6.61}$$

where L_y is the Lipschitz constant of $\tilde{f}$ with respect to y. Under these hypotheses, for β large enough, it can be proved that there exists a unique solution $(Y, U) \in \mathcal{L}^2_\beta(A) \times \mathcal{G}^2_\beta(\mu)$ to BSDE (6.60), see Theorem 4.1 in [**2**].

At this point some comments may be useful. Two random fields U and U' in $\mathcal{G}^2_{\text{loc}}(\mu)$ will be said to be equal if $U = U'$ $\mathbb{M}^{\mathbb{P}}_\nu$-a. e. (i.e., $d\mathbb{P}(\omega)\,\nu(\omega, dt\,de)$-a.e.).

Uniqueness in Theorem 4.1 in [**2**] means the following: if (Y, U), (Y', U') are solutions of the BSDE (6.60), then $Y = Y'$ and, by Proposition B.28, there is a predictable process (l_t) such that $U_t(e) - U'_t(e) = l_t\, 1_K(t)$, $d\mathbb{P}\,\nu(dt\,de)$-a.e. In other words, given a solution (Y, U_0) of BSDE (6.60), the class of all solutions will be given by the pairs (Y, U), where $U = l\, 1_K + U_0$ for some predictable process (l_t). In particular, if $K = \emptyset$, then the second component of the BSDE solution is unique in the smaller space $\mathcal{L}^2(\mu)$.

Example 6.4.7. Let us now consider a particular case of BSDE (6.60), namely a BSDE driven by the integer-valued random measure μ associated to a given Markov process X, of the form

$$Y_t = g(X_T) + \int_{]t,\,T]} f(s,\, X_{s-},\, Y_{s-},\, U_s(\cdot))\, dA_s - \int_{]t,\,T] \times \mathbb{R}} U_s(e)\,(\mu - \nu)(ds\,de). \tag{6.62}$$

We assume that X is a piecewise deterministic Markov process (PDMP) associated to the random measure μ, with values in the interval $]0, 1[$. Such a process has random jumps $(T_n)_n$ and a deterministic motion between jumps according to a drift $h :]0,\, 1[\to \mathbb{R}$ which is Lipschitz continuous. When the process reaches the boundary, it will instantaneously jump inside the interval. We will follow the notations in Davis [**35**], Chapter 2, Sections 24 and 26. For every $x \in]0, 1[$, we will express by $t_*(x)$ the first time such that the process X starting at x reaches 0 or 1. The behavior of X is described by a triplet of local characteristics (h, λ, P), where h is the drift introduced before, $\lambda :]0, 1[\to \mathbb{R}$ is a measurable function satisfying

$$\sup_{x \in]0,1[} |\lambda(x)| < \infty, \tag{6.63}$$

and P is a probability transition measure on $[0, 1] \times \mathcal{B}(]0, 1[)$, such that

$$\text{for some } \varepsilon > 0,\ P(x, B_\varepsilon) = 1 \text{ for } x \in \{0, 1\}, \text{ where } B_\varepsilon = \{x \in]0, 1[: t_*(x) > \varepsilon\}. \tag{6.64}$$

Set $N_t = \sum_{n \in \mathbb{N}} 1_{t \geq T_n}$. By Proposition 24.6 in [**35**], under conditions (6.63) and (6.64) we have

$$\mathbb{E}\,[N_t] < \infty \quad \forall\, t \in \mathbb{R}_+. \tag{6.65}$$

Notice that the PDMP X verifies the equation

$$X_t = X_0 + \int_0^t h(X_s)\, ds + \sum_{0<s\le t} \Delta X_s. \tag{6.66}$$

In particular X admits a finite number of jumps on each compact interval. By (26.9) in [**35**], the random measure μ is

$$\mu(ds\, de) = \sum_{n\in\mathbb{N}} 1_{\{X_{T_n}\in]0,\,1[\}} \delta_{(T_n,\, X_{T_n})}(ds\, de) = \sum_{0<s\le t} 1_{\{X_{s-}\neq X_s\}} \delta_{(s,\, X_s)}(ds\, de), \tag{6.67}$$

which is of the type of (6.51). This implies the validity of (6.52), so that (6.66) can be rewritten as

$$X_t = X_0 + \int_0^t h(X_s)\, ds + \int_{]0,\,t]\times]0,1[} (e - X_{s-})\, \mu(ds\, de).$$

In the following, by abuse of notations, μ will denote the trivial extension of previous measure to the real line. In particular (6.66) can be reexpressed as

$$X_t = X_0 + \int_0^t h(X_s)\, ds + \int_{]0,\,t]\times\mathbb{R}} (e - X_{s-})\, \mu(ds\, de). \tag{6.68}$$

The knowledge of $(h,\, \lambda,\, P)$ completely specifies the dynamics of X, see Section 24 in [**35**]. According to (26.2) in [**35**], the compensator of μ has the form

$$\nu(ds\, de) = (\lambda(X_{s-})\, ds + dp^*_s)\, P(X_{s-},\, de), \tag{6.69}$$

where

$$p^*_t = \sum_{n=1}^{\infty} 1_{\{t\ge T_n\}}\, 1_{\{X_{T_n-}\in\{0,1\}\}} \tag{6.70}$$

is the process counting the number of jumps of X from the boundary of its domain.

From (6.69) we can choose A_s and $\phi_s(de)$ such that $dA_s = \lambda(X_{s-})\, ds + dp^*_s$ and $\phi_s(de) = P(X_{s-}, de)$. In particular, A is predictable (not deterministic) and discontinuous, with jumps

$$\Delta A_s(\omega) = \hat\nu_s(\omega, \mathbb{R}) = \Delta p^*_s(\omega) = 1_{\{X_{s-}(\omega)\in\{0,1\}\}}. \tag{6.71}$$

Consequently, $\hat\nu_t(\omega, \mathbb{R}) > 0$ if and only if $\hat\nu_t(\omega, \mathbb{R}) = 1$, so that

$$J = \{(\omega, t) : \hat\nu_t(\omega, \mathbb{R}) > 0\} = \{(\omega, t) : \hat\nu_t(\omega, \mathbb{R}) = 1\} = K, \tag{6.72}$$

and

$$K = \{(\omega, t) : X_{t-}(\omega) \in \{0, 1\}\}. \tag{6.73}$$

□

Lemma 6.4.8. *Let X be the PDMP process considered in Example 6.4.7. Then*

$$\int_{]0,\,\cdot]\times\mathbb{R}} |e - X_{s-}|\, \nu(ds\, de) \in \mathcal{A}^+_{\mathrm{loc}}.$$

Proof. We start by noticing that

$$\int_{]0,\,T]\times\mathbb{R}} |e - X_{s-}|\,\nu(ds\,de) < \infty \quad \text{a.s.}$$

Indeed

$$\int_{]0,\,T]\times\mathbb{R}} |e - X_{s-}|\,\nu(ds\,de) = \int_{]0,\,T]\times]0,\,1[} |e - X_{s-}|\,(\lambda(X_{s-})\,ds + dp^*_s)\,P(X_{s-}, de)$$
$$\leq ||\lambda||_\infty\,(T + p^*_T).$$

For every $t \in [0,\,T]$ the jumps of the process

$$Y_t := \int_{]0,\,t]\times\mathbb{R}} |e - X_{s-}|\,\nu(ds\,de)$$

are given by

$$\Delta Y_t := \int_{]0,\,1[} |e - X_{t-}|\,\hat{\nu}_t(de) \leq \hat{\nu}_t(\mathbb{R}) \leq 1.$$

Since Y_t has bounded jumps, it is a locally bounded process and therefore it belongs to $\mathcal{A}^+_{\mathrm{loc}}$, see for instance the proof of Corollary at page 373 in [**110**]. □

Lemma 6.4.9. *Let μ and X be respectively the random measure and the associated PDMP satisfying equation* (6.68) *in Example 6.4.7. Assume in addition that there exists a function $\beta : \{0,1\} \to]0,1[$, such that*

$$X_s = \beta(X_{s-}) \quad \text{on } \{(\omega, s) : X_{s-}(\omega) \in \{0,1\}\}, \tag{6.74}$$

and

$$P(x,\,de) = \delta_{\beta(x)}(de) \quad \text{a.s.} \tag{6.75}$$

Then μ satisfies Hypothesis 6.3.17 and X fulfills Hypothesis 6.4.2 with decomposition $X = X^i + X^p$, with

$$X^i_t = \int_{]0,\,t]\times\mathbb{R}} (e - X_{s-})\,(\mu - \nu)(ds\,de), \tag{6.76}$$

$$X^p_t = X_0 + \int_0^t h(X_s)\,ds + \int_{]0,\,t]} \left(\int_{\mathbb{R}} (e - X_{s-})\,P(X_{s-},\,de)\right)(\lambda(X_{s-})\,ds + dp^*_s). \tag{6.77}$$

In particular X^i satisfies Hypothesis 6.3.10 with

$$\tilde{\gamma}(\omega, s, e) = (e - X_{s-}(\omega))\,1_{\{X_{s-}(\omega)\in]0,1[\}}(\omega, s).$$

Proof. Let us prove that Hypothesis 6.3.17-(i) holds. We recall that the measure μ was characterized by (6.67). We define $\mu^c := \mu\,1_{J^c}$, and $\nu^c := \nu\,1_{J^c}$. ν^c is the compensator of μ^c, see paragraph b) in [**76**]. Taking into account (6.69), (6.71) and (6.72), we have

$$\nu^c(ds\,de) = \lambda(X_{s-})\,P(X_{s-}, de)\,ds. \tag{6.78}$$

By Remark 6.3.19-(ii) we see that $D \cap J^c = \cup_n [[T^i_n]]$, $(T^i_n)_n$ totally inaccessible times. On the other hand, since by (6.72) $J = K$, we have $D = K \cup (D \cap J^c)$, therefore Hypothesis 6.3.17-(i) holds.

Let now consider Hypothesis 6.3.17-(ii). Taking into account (6.73), we have to prove that for every predictable time S such that $[[S]] \subset \{(\omega, t) : X_{t-}(\omega) \in \{0,1\}\}$,

$$\nu(\{S\}, de) = \mu(\{S\}, de) \quad \text{a.s.} \tag{6.79}$$

Let S be a predictable time satsifying $[[S]] \subset \{(\omega, t) : X_{t-}(\omega) \in \{0,1\}\}$. By (6.67), $\mu(\{S\}, de) = \delta_{X_S}(de)$, while from (6.69) we get $\nu(\{S\}, de) = P(X_{S-}, de)$. Therefore identity (6.79) can be rewritten as

$$P(X_{S-}, de) = \delta_{X_S}(de) \quad \text{a.s.} \tag{6.80}$$

Previous identity holds true under assumptions (6.74) and (6.75), and so Hypothesis 6.3.17-(ii) is established.

In order to prove the validity of Hypothesis 6.4.2, we will make use of Lemma 6.3.22. We recall that the process X satisfies the stochastic differential equation (6.68), which gives, taking into account Lemma 6.4.8,

$$\begin{aligned} X_t = X_0 &+ \int_0^t h(X_s)\, ds + \int_{]0,\, t]} \left(\int_{\mathbb{R}} (e - X_s)\, P(X_s,\, de) \right) \lambda(X_s)\, ds \\ &+ \int_{]0,\, t]} (\beta(X_{s-}) - X_{s-})\, dp^*_s + \int_{]0,\, t] \times \mathbb{R}} (e - X_{s-})\, (\mu - \nu)(ds\, de). \end{aligned} \tag{6.81}$$

We can show that previous equation is a particular case of (6.28). Indeed, we recall that, by (6.70) and (6.73), the support of the measure dp^* is included in K. We set $B_s = s + p^*(s)$ and $b(s,x) = \big(h(x) + \int_{\mathbb{R}} (e-x)\, \lambda(x)\, P(x,\, de)\big)\, 1_{K^c}(s) + (\beta(x) - x)\, 1_K(s)$. The reader can easily show that the sum of the first, second, and third integral in the right hand-side of (6.81) equals $\int_0^t b(s, X_{s-})\, dB_s$, provided we show that $\int_0^T |b(s, X_{s-})|\, dB_s$ is finite a.s. In fact we have

$$\begin{aligned} &\int_0^t |b(s, X_{s-})|\, dB_s \\ &\leq \int_0^t |h(X_s)|\, ds \\ &+ \int_{]0,\, t]} \left| \int_{\mathbb{R}} (e - X_{s-})\, \lambda(X_{s-})\, P(X_{s-},\, de)\, 1_{K^c}(s) + (\beta(X_{s-}) - X_{s-})\, 1_K(s) \right| dB_s \\ &= \int_0^t |h(X_s)|\, ds \\ &+ \int_{]0,\, t]} \left| \int_{\mathbb{R}} (e - X_{s-})\, P(X_{s-},\, de)\, (\lambda(X_{s-})\, 1_{K^c}(s) + 1_K(s)) \right| (ds + dp^*(s)) \\ &\leq \int_0^t |h(X_s)|\, ds + \int_{]0,\, t]} \int_{\mathbb{R}} |e - X_{s-}|\, \nu(ds,\, de). \end{aligned} \tag{6.82}$$

Recalling Lemma 6.4.8, and taking into account that h is locally bounded, we get that $\int_0^\cdot |b(s, X_{s-})|\, dB_s$ belongs to $\mathcal{A}^+_{\mathrm{loc}}$. Then, setting $N_s = 0$ and $\gamma(s,x,e) = e - x$, we see that X is a solution to equation (6.28).

Then, by Lemma 6.3.22, X satisfies Hypothesis 6.3.2, with decomposition $X = X^i + X^p$, where X^i and X^p are given respectively by (6.76) and (6.77). Moreover, the process X^i fulfills Hypothesis 6.3.10 with $\tilde{\gamma}(\omega, s, e) = (e - X_{s-}(\omega))\,(1 - 1_K(\omega, s)) = (e - X_{s-}(\omega))\,1_{\{X_{s-}(\omega) \in]0,1[\}}(\omega, s)$, and the process X^p satisfies Hypothesis 6.3.14. □

6.4.2. Identification of the BSDE's solution. We consider the following assumption on a couple (X, Y) of adapted processes.

Hypothesis 6.4.10. X is a special weak Dirichlet process of finite quadratic variation, satisfying condition (5.127). $Y_t = v(t, X_t)$ for some (deterministic) function $v : [0, T] \times \mathbb{R} \to \mathbb{R}$ of class $C^{0,1}$ such that $F = v$ and X verify condition (5.134).

Let us remark the following facts.

Proposition 6.4.11. *Assume that X is a process satisfying Hypothesis 6.3.2, with decomposition $X = X^i + X^p$, where X^i (resp. X^p) fulfills Hypothesis 6.3.10 (resp. Hypothesis 6.3.14), with respect to μ, with corresponding $\tilde{\gamma}$. Let in addition $v : [0, T] \times \mathbb{R} \to \mathbb{R}$ be a function of class $C^{0,1}$.*

(a) *If $\sum_{s \le T} |\Delta X_s|^2 < \infty$ a.s., then*

$$|v(s, X_{s-} + \tilde{\gamma}(s, e)) - v(s, X_{s-})|^2\, 1_{\{|\tilde{\gamma}(s,e)| \le 1\}} * \mu \in \mathcal{A}^+_{\mathrm{loc}}. \tag{6.83}$$

(b) *If X and $F = v$ satisfy conditions* (5.127) *and* (5.134), *then*

$$|v(s, X_{s-} + \tilde{\gamma}(s, e)) - v(s, X_{s-})|\, 1_{\{|\tilde{\gamma}(s,e)| > 1\}} * \mu \in \mathcal{A}^+_{\mathrm{loc}}. \tag{6.84}$$

(c) *If X and $F = v$ satisfy conditions* (5.127) *and* (5.134), *and moreover $\sum_{s \le T} |\Delta X_s|^2 < \infty$ a.s., then*

$$v(s, X_{s-} + \tilde{\gamma}(s, e)) - v(s, X_{s-}) \in \mathcal{G}^1_{\mathrm{loc}}(\mu).$$

Proof. Item (a) follows by Proposition 5.2.8 and inequality (6.11) in Proposition 6.3.15, with $\varphi(\omega, s, x) = |v(s, X_{s-}(\omega) + x) - v(s, X_{s-}(\omega))|^2\, 1_{\{|x| \le 1\}}$, allowing φ also depending on ω.

Item (b) is a consequence of (5.127) and (5.134) together with Lemma 5.6.24 and inequality (6.11) in Proposition 6.3.15, with $\varphi(\omega, s, x) = |v(s, X_{s-}(\omega) + x) - v(s, X_{s-}(\omega))|\; 1_{\{|x| > 1\}}$, allowing φ also depending on ω.

Finally, item (c) is a direct consequence of items (a), (b), and Remark 5.2.6, with $\varphi(\omega, s, e) = v(s, X_{s-}(\omega) + \tilde{\gamma}(\omega, s, e)) - v(s, X_{s-}(\omega))$ and $A = \{(\omega, s, e) : |\tilde{\gamma}(\omega, s, e)| > 1\}$. □

Proposition 6.4.12. *Let μ satisfy Hypothesis 6.3.17. Let X be a process verifying Hypothesis 6.4.2 with decomposition $X = X^i + X^p$, where $\tilde{\gamma}$ is the predictable process which relates μ and X^i in agreement with Hypothesis 6.3.10. Let (Y, Z, U) be a solution to the BSDE* (6.39) *such that the pair (X, Y) satisfies Hypothesis 6.4.10 with corresponding function v. Let X^c denote the continuous local martingale M^c of X given in the canonical decomposition* (5.101).

Then, the pair (Z, U) fulfills

$$Z_t = \partial_x v(t, X_t)\, \frac{d\langle X^c, M\rangle_t}{d\langle M\rangle_t} \quad d\mathbb{P}\, d\langle M\rangle_t\text{-a.e.}, \tag{6.85}$$

$$\int_{]0,\,t]\times\mathbb{R}} H_s(e)\,(\mu-\nu)(ds\,de) = 0, \quad \forall\, t \in]0,\,T],\ \text{a.s.}, \tag{6.86}$$

with

$$H_s(e) := U_s(e) - (v(s, X_{s-} + \tilde\gamma(s,e)) - v(s, X_{s-})). \tag{6.87}$$

If, in addition, $H \in \mathcal{G}^2_{\mathrm{loc}}(\mu)$,

$$\int_{]0,\,T]\times\mathbb{R}} |H_s(e) - \hat H_s\, 1_K(s)|^2\, \nu(ds\,de) = 0 \quad \text{a.s.} \tag{6.88}$$

Remark 6.4.13. Since the pair (X,Y) in Proposition 6.4.12 satisfies Hypothesis 6.4.10, then X and v in the statement satisfy (5.127) and (5.134). By Proposition 6.4.11-(c) it follows that $v(s, X_{s-}+\tilde\gamma(s,e))-v(s, X_{s-}) \in \mathcal{G}^1_{\mathrm{loc}}(\mu)$. Since $U \in \mathcal{G}^2_{\mathrm{loc}}(\mu) \subset \mathcal{G}^1_{\mathrm{loc}}(\mu)$, this yields $H \in \mathcal{G}^1_{\mathrm{loc}}(\mu)$.

Proof. By assumption, X is a special weak Dirichlet process satisfying condition (5.127), and $F = v$ is a function of class $C^{0,1}$ satisfying the integrability condition (5.134). So we are in the condition to apply Theorem 5.6.26 to $v(t,\, X_t)$. We get

$$\begin{aligned} v(t,\, X_t) &= v(0, X_0) + \int_{]0,\,t]\times\mathbb{R}} (v(s, X_{s-}+x) - v(s, X_{s-}))\,(\mu^X - \nu^X)(ds\,dx) \\ &\quad + \int_{]0,\,t]} \partial_x v(s, X_s)\, dX^c_s + A^v(t), \end{aligned} \tag{6.89}$$

where $A^v : C^{0,\,1} \to \mathbb{D}^{ucp}$ is a map such that, for every $v \in C^{0,\,1}$, A^v is a predictable orthogonal process. We set

$$\varphi(s,x) := v(s, X_{s-}+x) - v(s, X_{s-}).$$

Since X is of finite quadratic variation and verifies (5.127), and X and $F = v$ satisfy (5.134), by Proposition 5.2.8 and Lemma 5.6.24, we see that the process φ verifies condition (6.13) with $A = \{|x| > 1\}$. Moreover $\varphi(s,0) = 0$. Since μ verifies Hypothesis 6.3.17 and X verifies Hypothesis 6.4.2, we can apply Proposition 6.3.20 to $\varphi(s,x)$. Identity (6.89) becomes

$$\begin{aligned} v(t,\, X_t) &= v(0, X_0) + \int_{]0,\,t]\times\mathbb{R}} (v(s, X_{s-}+\tilde\gamma(s,e)) - v(s, X_{s-}))\,(\mu - \nu)(ds\,de) \\ &\quad + \int_{]0,\,t]} \partial_x v(s, X_s)\, dX^c_s + A^v(t). \end{aligned} \tag{6.90}$$

At this point we recall that the process $Y_t = v(t, X_t)$ fulfills the BSDE (6.39), which can be rewritten as

$$\begin{aligned} Y_t &= Y_0 + \int_{]0,\,t]} Z_s\, dM_s + \int_{]0,\,t]\times\mathbb{R}} U_s(e)\,(\mu-\nu)(ds\,de) \\ &\quad - \int_{]0,\,t]} \tilde g(s,\, Y_{s-},\, Z_s)\, d\zeta_s - \int_{]0,\,t]\times\mathbb{R}} \tilde f(s,\, e,\, Y_{s-},\, U_s(e))\, \lambda(ds\,de). \end{aligned} \tag{6.91}$$

By Proposition 5.6.8 the uniqueness of decomposition (6.90) yields identity (6.86) and

$$\int_{]0,\,t]} Z_s\, dM_s = \int_{]0,\,t]} \partial_x v(s, X_s)\, dX^c_s. \tag{6.92}$$

In particular, from (6.92) we get

$$
\begin{aligned}
0 &= \langle \int_{]0,\,t]} Z_s dM_s - \int_{]0,\,t]} \partial_x v(s, X_s)\, dX_s^c,\, M_t \rangle \\
&= \int_{]0,\,t]} Z_s d\langle M \rangle_s - \int_{]0,\,t]} \partial_x v(s, X_s)\, \frac{d\langle X^c,\, M\rangle_s}{d\langle M \rangle_s}\, d\langle M \rangle_s \\
&= \int_{]0,\,t]} \left(Z_s - \partial_x v(s, X_s)\, \frac{d\langle X^c,\, M\rangle_s}{d\langle M \rangle_s} \right) d\langle M \rangle_s,
\end{aligned}
$$

that gives identification (6.85).

If in addition we assume that $H \in \mathcal{G}^2_{\mathrm{loc}}(\mu)$, the predictable bracket at time t of the purely discontinuous martingale in identity (6.86) is well-defined, and equals

$$
\int_{]0,\,t]\times\mathbb{R}} |H_s(e) - \hat{H}_s\, 1_J(s)|^2\, \nu(ds\, de) + \sum_{s\in]0,\,t]} |\hat{H}_s|^2 (1 - \hat{\nu}_s(\mathbb{R}))\, 1_{J\setminus K}(s), \tag{6.93}
$$

see Theorem B.22, identity (B.25), and Remark B.23. The conclusion follows from the fact that under Hypothesis 6.3.17 we have $J = K$ up to an evanescent set, see Remark 6.3.18. □

We apply now previous result to the case of Example 6.4.3. We start with a preliminary result.

Lemma 6.4.14. *Let μ and X be respectively the Poisson random measure and the stochastic process satisfying the SDE* (6.44) *in Example 6.4.3. Let $u : [0,\, T]\times\mathbb{R} \to \mathbb{R}$ be a function of $C^{0,1}$ class such that $x \mapsto \partial_x u(s, x)$ has linear growth, uniformly in s. Then condition* (5.134) *holds for X and $F = u$.*

Proof. We have

$$
\begin{aligned}
&\int_{]0,\cdot]\times\mathbb{R}} |u(s, X_{s-} + x) - u(s, X_{s-}) - x\, \partial_x u(s, X_{s-})|\, 1_{\{|x|>1\}}\, \mu^X(ds\, dx) \\
&= \sum_{0<s\leq\cdot} |u(s, X_s) - u(s, X_{s-}) - \partial_x u(s, X_{s-})\, \Delta X_s|\, 1_{\{|\Delta X_s|>1\}} \\
&\leq \sum_{0<s\leq\cdot} |\Delta X_s|\, 1_{\{|\Delta X_s|>1\}} \left(\int_0^1 |\partial_x u(s, X_{s-} + a\, \Delta X_s)|\, da + \int_0^1 |\partial_x u(s, X_{s-})|\, da \right) \\
&\leq 2\, C \sum_{0<s\leq\cdot} |X_{s-}||\Delta X_s|\, 1_{\{|\Delta X_s|>1\}} + \sum_{s\leq t} |\Delta X_s|^2\, C\, 1_{\{|\Delta X_s|>1\}} \\
&= 2\, C \int_{]0,\cdot]\times\mathbb{R}} |X_{s-}|\, |x|\, 1_{\{|x|>1\}}\, \mu^X(ds\, dx) + \sum_{s\leq\cdot} |\Delta X_s|^2\, 1_{\{|\Delta X_s|>1\}}.
\end{aligned} \tag{6.94}
$$

Since X is of finite quadratic variation, the second term in the right-hand side of (6.94) is in $\mathcal{A}^+_{\mathrm{loc}}$ if and only if

$$
\sum_{s\in]0,\,\cdot]} |\Delta X_s|^2 \in \mathcal{A}^+_{\mathrm{loc}}, \tag{6.95}
$$

see Proposition 5.2.4 with $p = 2$. Since by (6.44) $\Delta X_s = \int_{\mathbb{R}} \gamma(X_{s-}, e)\, \mu(ds\, de)$, we have

$$\sum_{s \in]0,\,\cdot]} |\Delta X_s|^2 = \sum_{s \in]0,\,\cdot]} \left| \int_{\mathbb{R}} \gamma(X_{s-}, e)\, \mu(ds\, de) \right|^2 = \int_{]0,\cdot] \times \mathbb{R}} |\gamma(X_{s-}, e)|^2\, \mu(ds\, de),$$

and (6.95) reads

$$\int_{]0,\cdot] \times \mathbb{R}} |\gamma(X_{s-}, e)|^2\, \mu(ds\, de) \in \mathcal{A}^+_{\mathrm{loc}}. \tag{6.96}$$

Condition (6.96) holds because $|\gamma(x, e)| \leq K\,(1 \wedge |e|)$ for every $x \in \mathbb{R}$, $\int_{\mathbb{R}} (1 \wedge |e|^2)\, \lambda(de) < \infty$ (see, respectively, (6.45) and (6.43)), and taking into account the fact that the integrand in (6.96) is locally bounded.

Finally, the first term in the right-hand side of (6.94) belongs to $\mathcal{A}^+_{\mathrm{loc}}$ since X_{s-} is locally bounded (see e.g. the lines above Theorem 15, Chapter IV, in [**110**]) and X satisfies (5.127). The conclusion follows. □

We are ready to give the identification result in the framework of Example 6.4.3.

Corollary 6.4.15. *Let $(Y, Z, U) \in \mathcal{S}^2 \times \mathcal{L}^2 \times \mathcal{L}^2(\mu)$ be the unique solution to the BSDE* (6.41). *If the function u defined in* (6.46) *is of class $C^{0,1}$ such that $x \mapsto \partial_x u(t, x)$ has linear growth, uniformly in t, then the process (Z, U) satisfies*

$$Z_t = \partial_x u(t, X_t) \quad d\mathbb{P}\, dt\text{-a.e.}, \tag{6.97}$$

$$\int_{]0,\,t] \times \mathbb{R}} H_s(e)\, (\mu - \nu)(ds\, de) = 0, \quad \forall t \in]0, T], \text{ a.s.} \tag{6.98}$$

where

$$H_s(e) := U_s(e) - (u(s, X_{s-} + \gamma(s, X_{s-}, e)) - u(s, X_{s-})). \tag{6.99}$$

If in addition $H \in \mathcal{G}^2_{\mathrm{loc}}(\mu)$,

$$U_s(e) = u(s, X_{s-} + \gamma(s, X_{s-}, e)) - u(s, X_{s-}) \quad d\mathbb{P}\, \lambda(de)\, ds\text{-a.e.} \tag{6.100}$$

Proof. We aim at applying Proposition 6.4.12. By Lemma 6.4.4, μ satisfies Hypothesis 6.3.17 and X fulfills Hypothesis 6.4.2 with decomposition $X = X^i + X^p$, where X^i satisfies Hypothesis 6.3.10 with $\tilde{\gamma}(s, e) = \gamma(s, X_{s-}, e)$. Moreover, since X is a special semimartingale, it is of finite quadratic variation and (5.127) holds because of Corollary 11.26 in [**73**]. By Lemma 6.4.14, condition (5.134) holds for X and $F = u$, which implies that Hypothesis 6.4.10 is verified.

We can then apply Proposition 6.4.12: since $X^c = M = W$, (6.85) gives (6.97), while (6.86)-(6.87) with $\tilde{\gamma}(s, e) = \gamma(s, X_{s-}, e)$ yield (6.98)-(6.99). If in addition $H \in \mathcal{G}^2(\mu)$, since $\hat{H} = 0$ (ν is absolutely continuous with respect to the Lebesgue measure), (6.88) yields

$$\int_{]0,\,T] \times \mathbb{R}} |H_s(e)|^2\, \lambda(de)\, ds = 0, \tag{6.101}$$

and (6.100) follows. □

Remark 6.4.16. When the BSDE (6.41) is driven only by a standard Brownian motion, an identification result for Z analogous to (6.97) has been established by [**68**], even supposing only that f is Lipschitz with respect to Z.

Let us now consider a BSDE driven only by a purely discontinuous martingale, of the form (6.49). We formulate the following assumption for a couple of adapted processes (X, Y).

Hypothesis 6.4.17.

(i) $Y = B + A'$, with B a bounded variation process and A' a continuous $(\mathcal{F}_t)$-orthogonal process;

(ii) $Y_t = v(t, X_t)$ for some continuous deterministic function $v : [0, T] \times \mathbb{R} \to \mathbb{R}$, satisfying the integrability condition

$$\int_{]0,\,\cdot]\times\mathbb{R}} |v(t, X_{t-} + x) - v(t, X_{t-})|\, \mu^X(dt\, dx) \ \in \mathcal{A}^+_{\mathrm{loc}}. \tag{6.102}$$

We have the following result.

Proposition 6.4.18. *Let μ satisfy Hypothesis 6.3.17. Let X verify Hypothesis 6.4.2 with decomposition $X = X^i + X^p$, where $\tilde{\gamma}$ is the predictable process which relates μ and X^i in agreement with Hypothesis 6.3.10. Let (Y, U) be a solution to the BSDE* (6.49)*, such that (X, Y) satisfies Hypothesis 6.4.17 with corresponding function v.*

Then, the process U satisfies

$$\int_{]0,\,t]\times\mathbb{R}} H_s(e)\, (\mu - \nu)(ds\, de) = 0 \quad \forall t \in]0,\, T],\ \text{a.s.}, \tag{6.103}$$

with

$$H_s(e) := U_s(e) - (v(s, X_{s-} + \tilde{\gamma}(s, e)) - v(s, X_{s-})). \tag{6.104}$$

If in addition $H \in \mathcal{G}^2_{\mathrm{loc}}(\mu)$,

$$\int_{]0,\,T]\times\mathbb{R}} |H_s(e) - \hat{H}_s\, 1_K(s)|^2\, \nu(ds\, de) = 0 \quad \text{a.s.} \tag{6.105}$$

Remark 6.4.19. The assumption of continuity for $v(t, x)$ in Hypothesis 6.4.17-(ii) is somehow restrictive since it can be relaxed with respect to x. However our purpose is to illustrate the methodology and the assumption of continuity simplifies the proof.

Proof. By assumption, the couple (X, Y) satisfies Hypothesis 6.4.17 with corresponding function v. We are then in the condition to apply Proposition 5.6.28 to $v(t,\, X_t)$. We get

$$v(t,\, X_t) = v(0, X_0) + \int_{]0,\,t]\times\mathbb{R}} (v(s, X_{s-} + x) - v(s, X_{s-}))\, (\mu^X - \nu^X)(ds\, dx) + A^v(t), \tag{6.106}$$

where A^v is a predictable $(\mathcal{F}_t)$-orthogonal process. Set

$$\varphi(s, x) := v(s, X_{s-} + x) - v(s, X_{s-}).$$

By condition (ii) in Hypothesis 6.4.17, the process φ verifies condition (6.13) with $A = \Omega \times [0,\,T] \times \mathbb{R}$. Moreover $\varphi(s,0) = 0$. Since μ verifies Hypothesis 6.3.17, and X verifies Hypothesis 6.4.2 we can apply Proposition 6.3.20 to $\varphi(s,x)$. Identity (6.106) becomes

$$v(t,\,X_t) = v(0,X_0) + \int_{]0,\,t]\times\mathbb{R}} (v(s,X_{s-} + \tilde{\gamma}(s,e)) - v(s,X_{s-}))\,(\mu-\nu)(ds\,de) + A^v(t). \tag{6.107}$$

At this point we recall that the process $Y_t = v(t,X_t)$ fulfills the BSDE (6.49), which can be rewritten as

$$Y_t = Y_0 + \int_{]0,\,t]\times\mathbb{R}} U_s(e)\,(\mu-\nu)(ds\,de) - \int_{]0,\,t]\times\mathbb{R}} \tilde{f}(s,\,e,\,Y_{s-},\,U_s(e))\,\lambda(ds\,de). \tag{6.108}$$

By Proposition 5.6.8 the uniqueness of decomposition (6.107) yields identity (6.86). If in addition we assume that $H \in \mathcal{G}^2_{\mathrm{loc}}(\mu)$, the predictable bracket at time t of the purely discontinuous martingale in identity (6.86) is well-defined, and equals

$$\int_{]0,\,t]\times\mathbb{R}} |H_s(e) - \hat{H}_s\,1_J(s)|^2\,\nu(ds\,de) + \sum_{s\in]0,\,t]} |\hat{H}_s|^2(1-\hat{\nu}_s(\mathbb{R}))\,1_{J\setminus K}(s), \tag{6.109}$$

see Theorem B.22, identity (B.25), and Remark B.23. The conclusion follows from the fact that under Hypothesis 6.3.17 we have $J = K$, see Remark 6.3.18. □

Previous result can be applied to the framework of Example 6.4.5. We start with a preliminary observation.

Lemma 6.4.20. *Let X, μ be respectively the pure jump Markov process and the corresponding integer-valued random measure in Example 6.4.5. Let $u : [0,\,T] \times \mathbb{R} \to \mathbb{R}$ be a continuous function satisfying* (6.56)*,* (6.57) *and* (6.58)*. If we set $Y_t = u(t,X_t)$, then (X,Y) satisfies Hypothesis 6.4.17 with corresponding function u.*

Proof. From (6.53) and the fact that u is continuous, it follows that

$$u(t,X_t) = u(0,\,X_0) + \sum_{s\le t}(u(s,\,X_{s-} + \Delta X_s) - u(s,\,X_{s-})). \tag{6.110}$$

Obviously $Y_t = u(t,X_t)$ has a finite number of jumps on each compact. We have $\sum_{s\le t}|u(s,\,X_{s-} + \Delta X_s) - u(s,\,X_{s-})| < \infty$ a.s. for every $t \in \mathbb{R}_+$. Therefore, condition (i) in Hypothesis 6.4.17 holds with $B = u(0,\,X_0) + \sum_{s\le \cdot}(u(s,\,X_{s-} + \Delta X_s) - u(s,\,X_{s-}))$, $A' = 0$.

To verify the validity of condition (ii) of Hypothesis 6.4.17 with corresponding function $v = u$, we have to show that (6.102) holds with $v = u$. Denoting $||\lambda||_\infty = \sup_{t\in[0,\,T],\,x\in\mathbb{R}}|\lambda(t,x,\mathbb{R})|$, by (6.52) we have

$$\begin{aligned}
&\mathbb{E}\left[\int_{]0,\,T]\times\mathbb{R}} |u(s,X_{s-}+x) - u(s,\,X_{s-})|\,\mu^X(ds\,dx)\right]\\
&= \mathbb{E}\left[\int_{]0,\,T]\times\mathbb{R}} |u(s,e) - u(s,\,X_{s-})|\,\mu(ds\,de)\right]
\end{aligned}$$

$$= \mathbb{E}\left[\int_{]0,\,T]\times\mathbb{R}} |u(s,e) - u(s,\, X_{s-})|\,\lambda(s, X_{s-},\, de)\, ds\right]$$
$$\leq T\,||\lambda||_\infty^{1/2}\,||u(s,e) - u(s,\, X_{s-})||_{\mathcal{L}^2(\mu)}^{1/2}$$

and the conclusion follows since $u(s,e) - u(s, X_{s-}) \in \mathcal{L}^2(\mu)$ by (6.57). □

We have the following identification result in the framework of Example 6.4.5.

Corollary 6.4.21. *Let $(Y, U) \in \mathcal{L}^2 \times \mathcal{L}^2(\mu)$ be the unique solution to the BSDE* (6.50) *and X, u respectively the process and the function appearing in Example 6.4.5. Assume moreover that u is continuous. Then the process U satisfies*

$$U_t(e) = u(t,e) - u(t, X_{t-}) \quad d\mathbb{P}\,\lambda(t, X_{t-},\, de)\, dt\text{-a.e.} \tag{6.111}$$

Proof. We aim at applying Proposition 6.4.18. By Lemma 6.4.6, μ satisfies Hypothesis 6.3.17 and X fulfills Hypothesis 6.4.2 with decomposition $X = X^i$, $X^p = 0$, where X^i satisfies Hypothesis 6.3.10 with $\tilde{\gamma}(s,e) = e - X_{s-}$. Moreover, by Lemma 6.4.20, (X, Y) satisfies Hypothesis 6.4.17 with corresponding function $v = u$. We can then apply Proposition 6.4.18. We have

$$\begin{aligned} H_s(e) &:= U_s(e) - (u(s, X_{s-} + \tilde{\gamma}(s,e)) - u(s, X_{s-})) \\ &= U_s(e) - (u(s,e) - u(s, X_{s-})), \end{aligned} \tag{6.112}$$

which belongs to $\mathcal{L}^2(\mu)$, and therefore to $\mathcal{G}^2(\mu)$. Since moreover $\hat{H} = 0$ (ν is absolutely continuous with respect to the Lebesgue measure), (6.105) yields

$$\int_{]0,\,T]\times\mathbb{R}} |H_s(e)|^2\,\lambda(s, X_{s-},\, de)\, ds = 0, \text{ a.s.} \tag{6.113}$$

and (6.111) follows. □

Finally, we apply previous results to Example 6.4.7.

Lemma 6.4.22. *Let $(Y, U) \in \mathcal{L}^2 \times \mathcal{G}^2(\mu)$ be a solution to the BSDE* (6.62) *and X, u respectively the process and the function appearing in Example 6.4.7. Assume that $Y_t = u(t, X_t)$ for some continuous function $u : [0, T] \times \mathbb{R} \to \mathbb{R}$. Then (X, Y) satisfies Hypothesis 6.4.17 with corresponding function $v = u$.*

Proof. Since the process X has a finite number of jumps on each compact, the same holds for $Y_t = u(t, X_t)$. We set

$$B_t := \sum_{0<s\leq t} \Delta Y_s, \quad A'_t := Y_t - B_t. \tag{6.114}$$

Obviously B has bounded variation, and the process A' is continuous by definition. Since Y satisfies by assumption BSDE (6.62), for every local continuous martingale N we have

$$[Y, N]_t = \int_{]0,\,t]} f(s,\, X_{s-},\, Y_{s-},\, U_s(\cdot))\, d[A, N]_s - \left[\int_{]0,\,\cdot]\times\mathbb{R}} U_s(e)\,(\mu - \nu)(ds\, de), N\right]_t. \tag{6.115}$$

Since A is a predictable increasing process, therefore has bounded variation, $[A, N] = 0$ by Proposition 3.13 in [**9**]. The second term in (6.115) is zero because $\int_{]0,\,\cdot]\times\mathbb{R}} U_s(e)\,(\mu-\nu)(ds\,de)$ is a purely discontinuous martingale. Therefore (6.115) vanishes. Recalling that B has bounded variation, it also follows that $[B, N] = 0$, so that A' is a continuous $(\mathcal{F}_t)$-orthogonal process, and condition (i) in Hypothesis 6.4.17 holds.

It remains to show that $u(t, X_t)$ satisfies condition (6.102) with $v = u$. Since u is continuous, we have

$$\int_{]0,\,\cdot]\times\mathbb{R}} |u(s,\, X_{s-}+x) - u(s,\, X_{s-})|\, \mu^X(ds\,dx) = \sum_{0<s\leq \cdot} |u(s,\, X_s) - u(s,\, X_{s-})|$$

$$= \sum_{s\leq \cdot} |\Delta Y_s|. \tag{6.116}$$

The process Y takes values in the image of $[0,\, T]\times[0,1]$ with respect to u, which is a compact set. Therefore the jumps of Y are bounded, and (6.116) belongs to $\mathcal{A}^+_{\mathrm{loc}}$, see for instance the proof of Corollary at page 373 in [**110**]. □

Corollary 6.4.23. *Let $(Y, U) \in \mathcal{L}^2 \times \mathcal{G}^2(\mu)$ be a solution to the BSDE (6.62), and X the piecewise deterministic Markov process with local characteristics (h, λ, P) appearing in Example 6.4.7. Assume that $Y_t = u(t, X_t)$ for some continuous function u. Assume in addition that there exists a function $\beta : \{0, 1\} \to \mathbb{R}$, such that*

$$X_s = \beta(X_{s-}) \quad \text{on } \{(\omega, s) : X_{s-}(\omega) \in \{0, 1\}\}, \tag{6.117}$$

and that

$$P(x,\, de)\, 1_{\{x\in\{0,1\}\}}(s) = \delta_{\beta(x)}(de). \tag{6.118}$$

Then the process U satisfies

$$\int_{]0,\,t]\times\mathbb{R}} H_s(e)\,(\mu-\nu)(ds\,de) = 0 \quad \forall t \in]0,\, T], \text{a.s.}, \tag{6.119}$$

where

$$H_s(e) := (U_s(e) - (u(s, e) - u(s, X_{s-}))\, 1_{\{X_{s-}\in]0,1[\}}(s) + U_s(e)\, 1_{\{X_{s-}\in\{0,1\}\}}(s).$$

If in addition $H_s(e) \in \mathcal{G}^2_{\mathrm{loc}}(\mu)$,

$$U_s(e) = u(s, e) - u(s, X_{s-}) \quad d\mathbb{P}\,\lambda(X_{s-})\,P(X_{s-},\, de)\,ds\text{-a.e.} \tag{6.120}$$

Remark 6.4.24. If $H \in \mathcal{G}^2_{\mathrm{loc}}(\mu)$, the value of $U_s(\cdot)$ can be chosen on $K = \{(\omega, s) : X_{s-}(\omega) \in \{0, 1\}\}$ as an arbitrary $\mathcal{P}$-measurable process, see Proposition B.28.

Proof. We will apply Proposition 6.4.18. By Lemma 6.4.9, μ satisfies Hypothesis 6.3.17 and X fulfills Hypothesis 6.4.2 with decomposition $X = X^i + X^p$, where X^i satisfies Hypothesis 6.3.10 with $\tilde{\gamma}(\omega, s, e) = (e - X_{s-}(\omega))\, 1_{\{X_{s-}(\omega)\in]0,1[\}}(\omega, s)$. Moreover, by Lemma 6.4.22, Hypothesis 6.4.17 holds for (X, Y). We are then in condition to apply Proposition 6.4.18. Identity (6.103) yields

$$\int_{]0,\,t]\times\mathbb{R}} H_s(e)\,(\mu-\nu)(ds\,de) = 0 \quad \forall t \in [0,\, T],\ \text{a.s.}, \tag{6.121}$$

where

$$\begin{aligned}
H_s(e) &:= U_s(e) - [u(s, X_{s-} + \tilde{\gamma}(s,e)) - u(s, X_{s-})] \\
&= U_s(e) - [u(s, X_{s-} + (e - X_{s-})\, 1_{\{X_{s-} \in]0,1[\}}(s)) - u(s, X_{s-})] \\
&= [U_s(e) - (u(s,e) - u(s, X_{s-}))]\, 1_{\{X_{s-} \in]0,1[\}}(s) + U_s(e)\, 1_{\{X_{s-} \in \{0,1\}\}}(s), \\
&= [U_s(e) - (u(s,e) - u(s, X_{s-}))]\, 1_{K^c}(s) + U_s(e)\, 1_K(s), \qquad (6.122)
\end{aligned}$$

where in the latter equality we use the fact that $K = \{(\omega, s) : X_{s-}(\omega) \in \{0,1\}\}$.

It remains to prove (6.120). We recall that $\nu^c := \nu\, 1_{J^c}$ verifies $\nu^c(ds\, de) = \lambda(X_s)$ $P(X_s, de)\, ds$ by (6.78). We set $\nu^d := \nu\, 1_J$; since $J = K$, we have

$$\nu^d(ds\, de) = \nu(ds\, de)\, 1_K(s) = P(X_{s-},\, de)\, dp^*_s = \delta_{\beta(X_{s-})}(de)\, dp^*_s. \qquad (6.123)$$

If $H_s(e)$ belongs to $\mathcal{G}^2_{\mathrm{loc}}(\mu)$, recalling identity (B.32) in Remark B.23, identity (6.105) and (6.122) yield

$$\begin{aligned}
0 &= \int_{]0,\,T]\times\mathbb{R}} |H_s(e)|^2\, \nu^c(ds\, de) + \int_{]0,\,T]\times\mathbb{R}} |H_s(e) - \hat{H}_s\, 1_K(s)|^2\, \nu^d(ds\, de) \\
&= \int_{]0,\,T]\times\mathbb{R}} |U_s(e) - (u(s,e) - u(s, X_{s-}))|^2\, \nu^c(ds\, de) \\
&\quad + \int_{]0,\,T]\times\mathbb{R}} |U_s(e) - \hat{U}_s\, 1_K(s)|^2\, \nu^d(ds\, de). \qquad (6.124)
\end{aligned}$$

Taking into account condition (6.123), (6.71) and (6.73), we have

$$\hat{U}_s\, 1_K(s) = \int_{\mathbb{R}} U_s(e)\, \nu^d(\{s\}\, de) = \int_{\mathbb{R}} U_s(e))\, \delta_{\beta(X_{s-})}(de)\, 1_K(s) = U_s(\beta(X_{s-}))\, 1_K(s).$$

Consequently

$$\begin{aligned}
&\int_{]0,\,T]\times\mathbb{R}} |U_s(e) - \hat{U}_s\, 1_K(s)|^2\, \nu^d(ds\, de) \\
&= \int_{]0,\,T]\times\mathbb{R}} |U_s(e) - \hat{U}_s\, 1_K(s)|^2\, \delta_{\beta(X_{s-})}(de)\, dp^*_s = 0.
\end{aligned}$$

Therefore (6.124) gives simply

$$0 = \int_{]0,\,T]\times\mathbb{R}} |U_s(e) - (u(s,e)\,) - u(s, X_{s-})|^2\, \lambda(X_s)\, P(X_s, de)\, ds,$$

and (6.120) follows. □

Remark 6.4.25. In all the considered examples, the underlying process X was a Markov process which is a semimartingale. However, in the literature there are plenty of examples that are not semimartingales, even in the continuous case.

Let X be a solution of an SDE with distributional drift, see e.g. Flandoli, Russo and Wolf [**64**], Russo and Trutnau [**115**], Flandoli, Issoglio and Russo [**62**], of the type

$$dX_t = \beta(X_t)\, dt + dW_t, \qquad (6.125)$$

for a class of Schwartz distributions β. In particular in the one-dimensional case β is allowed to be the derivative of any continuous function. In this case X is not a semimartingale but only a Dirichlet process, so that, for $v \in C^{0,1}$, $v(t, X_t)$ is a weak Dirichlet process. Forward BSDEs related to a forward process X solving (6.125) have been studied for instance in Russo and Wurzer [**122**], when the terminal type is random.

Recalls on the general theory of stochastic processes

In this chapter we recall the main definitions and some properties of general theory of stochastic processes that we extensively use in our work; for a complete discussion on this topic we refer to Jacod and Shiryaev [**79**], Jacod [**77**] and He, Wang and Yan [**73**].

In what follows, we are given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a positive horizon T and a filtration $(\mathcal{F}_t)_{t \geq 0}$, satisfying the usual conditions. A random set will be a subset of $\Omega \times [0,\, T] \cup \{\infty\}$, and $[[\tau, \tau']]$ will denote the stochastic interval $\{(\omega, t) : t \in [0,\, T] \cup \{\infty\},\, \tau(\omega) \leq t \leq \tau'(\omega)\}$ associated to two stopping times τ, τ'. For a stopping time τ taking values in $[0,\, T] \cup \{\infty\}$, $\mathcal{F}_{\tau-}$ will denote the σ-field generated by $\mathcal{F}_0$ and the events $A \cap \{t < \tau\}$, where $t \in [0,\, T]$ and $A \in \mathcal{F}_t$, see (0.30) of [**77**]. We will denote by $\mathcal{P}$ (resp. $\tilde{\mathcal{P}} = \mathcal{P} \otimes \mathcal{B}(\mathbb{R})$) the predictable σ-field on $\Omega \times [0, T]$ (resp. on $\tilde{\Omega} = \Omega \times [0, T] \times \mathbb{R}$). Analogously, we set $\mathcal{O}$ (resp. $\tilde{\mathcal{O}} = \mathcal{O} \otimes \mathcal{B}(\mathbb{R})$) as the optional σ-field on $\Omega \times [0, T]$ (resp. on $\tilde{\Omega}$). In the sequel, a random set will be called predictable (resp. optional) if its restriction to $\Omega \times [0,\, T]$ is $\mathcal{P}$-measurable (resp. $\mathcal{O}$-measurable). Moreover, a stochastic process which is $\mathcal{P}$-measurable (resp. $\mathcal{O}$-measurable) will be called predictable (resp. optional).

We will also denote by $\mathcal{A}$ (resp. $\mathcal{A}_{\mathrm{loc}}$) the collection of all adapted processes with integrable variation (resp. with locally integrable variation), and by $\mathcal{A}^+$ (resp $\mathcal{A}^+_{\mathrm{loc}}$) the collection of all adapted integrable increasing (resp. adapted locally integrable) processes. The significance of locally is the usual one which refers to localization by stopping times, see e.g. (0.39) of [**77**].

Definition A.1 (Definition 1.30, Chapter I, in [**79**])**.** A random set A is called to be thin if it is of the form $A = \cup_n [[T_n]]$, where (T_n) is a sequence of stopping times; if moreover the sequence (T_n) satisfies $[[T_n]] \cap [[T_m]] = \emptyset$ for all $n \neq m$, it is called an exhausting sequence for A.

Remark A.2. Any optional random set whose sections are at most countable is thin in the sense of Definition A.1, see the comments below Definition 1.30, Chapter I, in [**79**].

Definition A.3 (Definition 1.15, [**73**])**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathcal{G}$ be a sub-σ-field of $\mathcal{F}$. A random variable ξ is called to be σ-integrable with respect to $\mathcal{G}$ if there exists $\Omega_n \in \mathcal{G}$, $\Omega_n \uparrow \Omega$ a.s. such that each $\xi\, 1_{\Omega_n}$ is integrable.

Definition A.4 (Definition 7.39 in [**73**])**.** An optional process $X = (X_t)$ is said to be thin if $\{\Delta X \neq 0\}$ is a thin set. A typical example of thin optional process is the jump ΔX of an adapted càdlàg process X.

Definition A.5 (Definition 7.33, in [**73**])**.** Let M and N be two local martingales. If $[M, N] = 0$, we say that M and N are mutually orthogonal.

The notion of purely discontinuous martingales appears for instance Definition 7.21, in [**73**]. Below we recall a useful characterization of such processes given in Theorem 7.34, in [**73**], the comments above and obvious localization arguments.

Theorem A.6. *Let M be a local martingale with $M_0 = 0$. Then M is purely discontinuous if and only if it is orthogonal to every continuous local martingale.*

Definition A.7 (Definition 1.10, Chapter I, in [**79**])**.** A random set A is called evanescent if the set $\{\omega : \exists\, t \in [0, T] \cup \{\infty\}$ with $(\omega, t) \in A\}$ is $\mathbb{P}$-null; two E-valued processes are called indistinguishable if the random set $\{X \neq Y\} = \{(\omega, t) : X_t(\omega) \neq Y_t(\omega)\}$ is evanescent, i.e., if almost all the paths of X and Y are the same.

Theorem A.8 (Theorem 4.18, Chapter I, in [**79**])**.** *Any local martingale M admits a unique (up to indistinguishability) decomposition*

$$M = M^c + M^d$$

where $M_0^d = 0$, M^c is a continuous local martingale and M^d is a purely discontinuous local martingale.

In the sequel $\mathcal{H}^{2,d}$ (resp. $\mathcal{H}^{2,d}_{\mathrm{loc}}$) will stand for the set of square integrable (resp. locally square integrable) purely discontinuous martingales.

Corollary A.9 (Corollary 4.19, Chapter I, in [**79**])**.** *Let M and N be two purely discontinuous local martingales having the same jumps $\Delta M = \Delta N$ (up to an evanescent set). Then M and N are indistinguishable.*

Proposition A.10 (Proposition 2.4-(a) and Proposition 2.6, Chapter I, in [**79**])**.** *If X is a predictable process, then ΔX is predictable. If moreover τ is a stopping time, then $X_\tau\, 1_{\{\tau < \infty\}}$ is $\mathcal{F}_{\tau-}$-measurable.*

A.1. Predictable and totally inaccessible stopping times

Definition A.11 (Definition 2.7, Chapter I, in [**79**])**.** A predictable time is a mapping $\tau : \Omega \to [0, T] \cup \{\infty\}$, such that the stochastic interval $[[0, \tau[[$ is predictable.

Remark A.12. If τ is a predictable (finite) time, then $[[\tau]] \in \mathcal{P}$, see e.g. the comments after Definition 2.7, Chapter I, in [**79**].

Proposition A.13 (Proposition 2.18-(b), Chapter I, in [**79**])**.** *If X and Y are two predictable processes satisfying $X_\tau = Y_\tau$ a.s. on $\{\tau < \infty\}$ for all predictable times τ, then X and Y are indistinguishable.*

Definition A.14 (Definition 2.20, Chapter I, in [**79**])**.** A stopping time τ is called totally inaccessible if $\mathbb{P}(\tau = S < \infty) = 0$ for all predictable time S.

Remark A.15. It straight follows from Definition A.14 that

$$1_{[[T^i]]}(\omega, T^p(\omega))\, 1_{\{T^i<\infty,\, T^p<\infty\}} = 0 \quad \text{a.s.} \tag{A.1}$$

for any totally inaccessible time T^i and predictable time T^p.

Indeed, taking the expectation of the left-hand side of (A.1) we get

$$\mathbb{E}\left[1_{[[T^i]]}(\cdot, T^p(\cdot))\, 1_{\{T^i<\infty,\, T^p<\infty\}}\right] = \mathbb{P}(\omega \in \Omega : T^i(\omega) = T^p(\omega) < \infty) = 0.$$

Lemma A.16 (Lemma 2.23, Chapter I, in [**79**])**.** *If A is a predictable thin set, then A admits an exhausting sequence of predictable times, namely there is a sequence (T_n) of predictable times whose graphs are pairwise disjoint, such that $A = \cup_n [[T_n]]$.*

Proposition A.17 (Proposition 2.24, Chapter I, in [**79**])**.** *If X is a càdlàg predictable process, there is a sequence of predictable times that exhausts the jumps of X. Furthermore, $\Delta X_\tau = 0$ a.s. on $\{\tau < \infty\}$ for all totally inaccessible time τ.*

Definition A.18 (Definition 2.25, Chapter I, in [**79**])**.** A càdlàg process X is quasi-left continuous if $\Delta X_\tau = 0$ a.s. on the set $\{\tau < \infty\}$ for every predictable time τ.

Proposition A.19 (Proposition 2.26, Chapter I, in [**79**])**.** *Let X be a càdlàg adapted process. X is quasi-left continuous if and only if there is a sequence of totally inaccessible times that exhausts the jumps of X.*

Theorem A.20 (Theorem 4.21, [**73**])**.** *For any adapted càdlàg process $X = (X_t)$ there exists a sequence $(T_n)_n$ of strictly positive stopping times satisfying the following conditions:*

(i) $\{\Delta X \neq 0\} \subset \cup_n [[T_n]]$;

(ii) *each T_n is predictable or totally inaccessible;*

(iii) $[[T_n]] \cap [[T_m]] = \emptyset$ *for every $m \neq n$.*

Theorem A.21 (Theorem 5.2, [**73**])**.** *Let X be a measurable process such that for every predictable time τ, X_τ is σ-integrable with respect to $\mathcal{F}_{\tau-}$. Then there exists a unique predictable process, called predictable projection, denoted by ${}^pX$, such that for every predictable time τ we have*

$$\mathbb{E}\left[X_\tau\, 1_{\{\tau<\infty\}} | \mathcal{F}_{\tau-}\right] = {}^pX_\tau\, 1_{\{\tau<\infty\}} \quad \text{a.s.}$$

Lemma A.22 (Lemma 1.37 in [**77**])**.** *Let A be an increasing predictable process with $A_0 = 0$. Then there exists a sequence of increasing stopping times (T_n), such that, $T_n(\omega) \uparrow +\infty$, and $A_{T_n \wedge T} \leq n$ for each n.*

Lemma A.23. *Let A be a predictable process such that $\sup_{t\le u}|A_t| < \infty$ a.s. $\forall u > 0$. Then, for every predictable time τ taking values in $]0, T] \cup \{+\infty\}$, we have that $A_\tau\, 1_{\{\tau<\infty\}}$ is σ-integrable with respect to $\mathcal{F}_{\tau-}$.*

Proof. We set $A^*_t = \sup_{s\le t} A_s$. A^* is a predictable and increasing process. Moreover $A_0 = 0$. By Lemma A.22 there exists a sequence of stopping times (T_n), such that $T_n \uparrow \tau = \inf\{t : A^*_t = \infty\} = \infty$, with $A^*_{T_n} \le n$ for each n. Let $\Omega_n = \{T_n \ge \tau\} \cap \{\tau < \infty\}$. Clearly $\cup_n \Omega_n = \{\tau < \infty\}$. Moreover

$$n \ge A^*_\tau\, 1_{\Omega_n} \in L^1.$$

By Theorem 56, Chapter IV, in [**37**], $\Omega_n \in \mathcal{F}_{\tau-}$, so the result follows. $\square$

Corollary A.24. *Let A be a predictable process such that $\sup_{t\le u}|A_t| < \infty$ a.s. $\forall u \in [0, T]$. Then its predictable projection exists and ${}^pA = A$.*

Proof. Let τ be a predictable time. By (1.5) in [**77**], $A_\tau 1_{\{\tau<\infty\}}$ is $\mathcal{F}_{\tau-}$-measurable. This, together with Lemma A.23, gives

$$\mathbb{E}\left[A_\tau\, 1_{\{\tau<\infty\}} | \mathcal{F}_{\tau-}\right] = A_\tau\, 1_{\{\tau<\infty\}} \quad \text{a.s.}$$

From Theorem A.21 we conclude that ${}^pA = A$. $\square$

Definition A.25 (Definition 2.32, Chapter I, in [**79**])**.** A random set A is called measurable if its restriction to $\Omega \times [0, T]$ is measurable. The predictable support of a measurable random set A is the predictable set $A' = \{{}^p(1_A) > 0\}$, which is defined up to an evanescent set.

Proposition A.26 (Proposition 2.35, Chapter I, in [**79**])**.** *Let X be a càdlàg adapted process. X is quasi-left continuous if and only if the predictable support of the random set $\{\Delta X \ne 0\}$ is evanescent.*

Remark A.27. For any totally inaccessible time T^i we have

$${}^p\left(1_{[[T^i]]}\, 1_{\{T^i<\infty\}}\right) = 0.$$

Indeed, by Theorem A.21, for every predictable time τ, we have

$${}^p\left(1_{[[T^i]]}(\tau)\, 1_{\{T^i<\infty\}}\right) 1_{\{\tau<\infty\}} = \mathbb{E}\left[1_{[[T^i]]}(\tau)\, 1_{\{T^i,\tau<\infty\}} | \mathcal{F}_{\tau-}\right]$$

which vanishes since $1_{[[T^i]]}(\tau)\, 1_{\{T^i,\tau<\infty\}} = 0$ a.s., see Remark A.15.

As we will see in the next section, the notion of predictable projection for a measurable process plays a fundamental role in the stochastic integration theory with respect to random measures. We have the following important result.

Theorem A.28 (Theorem 4.56, point c), Chapter I, in [**79**])**.** *Let H be an optional process with $H_0 = 0$. We have ${}^pH = 0$ and $\left[\sum_{s\le\cdot}|H_s|^2\right]^{1/2} \in \mathcal{A}^+_{\mathrm{loc}}$ if and only if there exists a local martingale M such that ΔM and H are indistinguishable.*

Appendix B

Random measures

In the present chapter some basic results on stochastic integration with respect to (nonnegative) random measures are recalled. These results are presented without proof, for a complete discussion on this topic see, e.g., Chapter II, Section 1, in [**79**], or Chapter XI, Section 1, in [**73**].

In what follows we refer to the notations introduced in **Appendix A**. $(E, \mathcal{E})$ will denote the measurable space constituted by $E = \mathbb{R}$ and its Borel σ-algebra $\mathcal{E}$. We remark however that the mentioned references consider the case when $(E, \mathcal{E})$ is any Blackwell space.

B.1. General random measures

Definition B.1 (Definition 1.3, Chapter II, in [**79**]). A random measure on $[0,T]\times E$ is a family $\mu = (\mu(\omega, dt\, de) : \omega \in \Omega)$ of measures on $([0,T] \times E, \mathcal{B}([0,T]) \otimes \mathcal{E})$ satisfying the following.

(1) For every $A \in \mathcal{B}([0,T])\otimes\mathcal{E}$, the map $\omega \mapsto \mu(\omega, A)$ is a (measurable) random variable.
(2) $\mu(\omega, \{0\} \times E) = 0$ identically.

Let μ be a random measure and $W \in \tilde{\mathcal{O}}$. Since $(t, e) \mapsto W_t(\omega, e)$ is $\mathcal{B}([0,T])\otimes\mathcal{E}$-measurable for each $\omega \in \Omega$, we can define the integral process $W * \mu$ by

$$W * \mu_t(\omega) = \int_{]0,\,t]\times E} W_s(\omega, e)\, \mu(\omega, ds\, de).$$

Remark B.2. We remark that for fixed ω, previous integral is a Lebesgue type integral. When W is positive (resp. negative), previous integral always exists but could be $+\infty$ (resp. $-\infty$).

In the sequel, given a random measure μ as before, we will often omit the reference to ω. In other words, we will write $\mu(dt\, de)$ instead of $\mu(\omega, dt\, de)$.

Definition B.3 (Definition 1.6, Chapter II, in [**79**]). (a) A random measure μ is called optional if the process $W * \mu$ is $\mathcal{O}$-measurable for every $W \in \tilde{\mathcal{O}}$. A random measure λ is called predictable if the process $W * \lambda$ is $\mathcal{P}$-measurable for every $W \in \tilde{\mathcal{P}}$.

(b) An optional random measure μ is called integrable if $1 * \mu \in \mathcal{A}^+$.

(c) An optional random measure μ is called $\tilde{\mathcal{P}}$-σ-finite if there exists a $\tilde{\mathcal{P}}$-measurable partition (A_n) of $\tilde{\Omega}$ such that each $1_{A_n} * \mu \in \mathcal{A}^+$.

Theorem B.4 (Theorem 1.8, Chapter II, in [**79**]). *Let μ be an optional $\tilde{\mathcal{P}}$-σ-finite random measure. There exists a random measure, called the compensator of μ and denoted by ν, which is unique up to a $\mathbb{P}$-null set, and which is characterized as being a predictable random measure satisfying*

$$\mathbb{E}[W * \nu_T] = \mathbb{E}[W * \mu_T],$$

for every nonnegative $W \in \tilde{\mathcal{P}}$. Moreover, there exists a predictable process $A \in \mathcal{A}^+$ and a kernel $\phi(\omega, t, de)$ from $(\Omega \times [0, T], \mathcal{P})$ into $(E, \mathcal{E})$ such that

$$\nu(\omega, dt\, de) = dA_t(\omega)\, \phi(\omega, t, de). \tag{B.1}$$

Of course, the disintegration (B.1) *is not unique.*

B.1.1. Integer-valued random measures.

Definition B.5 (Definition 1.13, Chapter II, in [**79**]). An integer-valued random measure is a random measure that satisfies the following properties.

(i) $\mu(\omega, \{t\} \times E) \leq 1$ identically;

(ii) for each $A \in [0, T] \times E$, $\mu(\cdot, A)$ takes values in $\mathbb{N}$.

(iii) μ is optional and $\tilde{\mathcal{P}}$-σ-finite.

Proposition B.6 (Proposition 1.14, Chapter II, in [**79**]). *Let μ be an integer-valued random measure. We set*

$$D = \{(\omega, s) |\, \mu(\omega, \{s\} \times E) = 1\}. \tag{B.2}$$

The random set D is thin (D is called the support of μ) and there exists an E-valued optional process β such that

$$\mu(\omega, dt\, de) = \sum_{s \geq 0} 1_D(\omega, s)\, \delta_{(s, \beta_s(\omega))}(dt\, de). \tag{B.3}$$

Remark B.7. Let μ be an integer-valued random measure, with associated support D and process β in the sense of (B.3). Then, for any $W \in \tilde{\mathcal{O}}$, we have

$$W * \mu_t = \sum_{s \in]0,\, t]} W_s(\beta_s)\, 1_D(s). \tag{B.4}$$

Proposition B.8 (Proposition 1.16, Chapter II, in [**79**]). *Let $X = (X_t)$ be an adapted càdlàg E-valued process. Then*

$$\mu^X(\omega, dt\, dx) = \sum_{s \in]0,\, T]} 1_{\{\Delta X_s(\omega) \neq 0\}}\, \delta_{(s,\, \Delta X_s(\omega))}(dt\, dx) \tag{B.5}$$

defines an integer-valued random measure on $[0, T] \times E$, and in the representation (B.3) *we have $D = \{\Delta X \neq 0\}$ and $\beta = \Delta X$.*

Corollary B.9 (Corollary 1.19, Section II, in [**79**]). *Let X be an adapted càdlàg process and μ^X be the measure associated to its jumps by* (B.5)*, and ν^X its compensator. Then X is quasi-left continuous if and only if there exists a version of ν^X that satisfies identically $\nu^X(\omega, \{s\}, de) = 0$.*

Theorem B.10 (Theorem 11.14 in [**73**]). *Let μ be the integer-valued random measure with support D, and let ν be its compensator. Set*

$$a = (a_t), \quad a_t = \nu(\{t\} \times E),\ t \geq 0, \tag{B.6}$$

$$J = \{a > 0\}, \tag{B.7}$$

$$K = \{a = 1\}. \tag{B.8}$$

Then a is a predictable thin process, $0 \leq a \leq 1$, J is the predictable support of D, and K is the largest predictable set contained in D (up to an evanescent set).

Proposition B.11 (Proposition 1.17, Chapter II, in [**79**]). *Let μ be an integer-valued random measure, ν its compensator, and $J = \{(\omega, t) : \nu(\omega, \{t\} \times E) > 0\}$.*

a) *J is a predictable thin set.*

b) *For all predictable times τ and nonnegative $W \in \tilde{\mathcal{P}}$ (or, equivalently, for every $W \in \tilde{\mathcal{P}}$ such that $\int_E W(\tau, e)\, \mu(\{\tau\}, de) 1_{\{\tau < \infty\}}$ exists)*

$$\int_E W_\tau(e)\, \nu(\{\tau\}, de) = \mathbb{E}\left[\int_E W_\tau(e)\, \mu(\{\tau\}, de)\middle|\mathcal{F}_{\tau-}\right] \text{ on } \{\tau < \infty\}. \tag{B.9}$$

c) *There is a version of ν such that $\nu(\omega, \{t\} \times E) \leq 1$ identically, and the thin set J is exhausted by a sequence of predictable times.*

Remark B.12. Because of the validity of property (B.9), the compensator ν is also called the dual predictable projection of μ.

Proposition B.13. *Let μ be an integer valued random measure with support D. Let J and K be the associated sets defined in* (B.7) *and* (B.8)*. If $D = K \cup (\cup_n [[S_n]])$, where $(S_n)_n$ are totally inaccessible times, then $J = K$ up to an evanescent set.*

Proof. We start by noticing some basic facts. From the definition of predictable support of a random set in Definition A.25, we have

$$1_J = {}^p(1_D). \tag{B.10}$$

Moreover, since K is predictable, by Corollary A.24 we get

$$^p(1_K) = 1_K; \tag{B.11}$$

on the other hand, by Remark A.27 the predictable projection of $1_{[[S_n]]}$ is zero since S_n is a totally inaccessible finite time. Consequently we obtain

$$^p(1_{\cup_n [[S_n]]}) = \sum_n {}^p(1_{[[S_n]]}) = 0. \tag{B.12}$$

Finally, identities (B.10), (B.11) and (B.12) imply

$$1_J = {}^p(1_D) = 1_K,$$

therefore $J = K$. □

B.2. Stochastic integrals with respect to a random measure

From here on μ will be an integer-valued random measure on $[0, T] \times E$, and ν a "good" version of the compensator of μ as constructed in Proposition B.11-(c).

We set $\hat{\nu}_t(de) = \nu(\{t\}, de)$ for all $t \in [0, T]$ and, for any $W \in \tilde{\mathcal{O}}$, we define

$$\hat{W}_t = \int_E W_t(x)\, \hat{\nu}_t(de), \quad t \geq 0, \tag{B.13}$$

$$\tilde{W}_t = \int_E W_t(x)\, \mu(\{t\}, de) - \int_E W_t(x)\, \hat{\nu}_t(de) = W_t(\beta_t)\, 1_D(t) - \hat{W}_t, \quad t \geq 0, \tag{B.14}$$

with the convention

$$\tilde{W}_t = +\infty \quad \text{if } \hat{W}_t \text{ is not defined.} \tag{B.15}$$

β and D in (B.14) are respectively the optional process and the support associated to μ, see Proposition B.6. For every $q \in [1, \infty[$, we also introduce the following linear spaces

$$\mathcal{G}^q(\mu) = \Big\{W \in \tilde{\mathcal{P}} : \Big[\sum_{s \leq \cdot} |\tilde{W}_s|^2\Big]^{q/2} \in \mathcal{A}^+\Big\}, \tag{B.16}$$

$$\mathcal{G}^q_{\mathrm{loc}}(\mu) = \Big\{W \in \tilde{\mathcal{P}} : \Big[\sum_{s \leq \cdot} |\tilde{W}_s|^2\Big]^{q/2} \in \mathcal{A}^+_{\mathrm{loc}}\Big\}. \tag{B.17}$$

We have $\mathcal{G}^q(\mu) \subset \mathcal{G}^{q'}(\mu)$ for every $q' \leq q$.

Remark B.14. The sets in (B.17) corresponding to $q = 1, 2$ coincide respectively with the spaces $\mathcal{G}(\mu)$ and $\mathcal{G}^2(\mu)$ introduced in [**73**], pages 301 and 304. In particular, under convention (B.15), any element $W \in \mathcal{G}^1_{\mathrm{loc}}(\mu)$ satisfies $|\hat{W}_t| < \infty$ for every $t \in [0, T]$.

Remark B.15. If $W \in \mathcal{G}^1_{\mathrm{loc}}(\mu)$, then exists a local martingale M such that ΔM and $\tilde{W}$ are indistinguishable.

This is a consequence of the fact that the predictable projection of $\tilde{W}$ is zero, see observations below Definition 1.27, Chapter II, in [**79**], and of Theorem A.28 with $H = \tilde{W}$.

Definition B.16 (Definition 1.27, point b), Chapter II, in [**79**])**.** If $W \in \mathcal{G}^1_{\mathrm{loc}}(\mu)$, we call stochastic integral of W with respect to $\mu - \nu$ and $W * (\mu - \nu)$ denotes any purely discontinuous local martingale M such that ΔM and $\tilde{W}$are indistinguishable.

Remark B.17. By Corollary A.9, if $W \in \mathcal{G}^1_{\mathrm{loc}}(\mu)$, all the stochastic integrals $W * (\mu - \nu)$ are equal up to indistinguishability.

Proposition B.18 (Proposition 1.28, Chapter II, in [**79**])**.** *Let $W \in \tilde{\mathcal{P}}$, such that $|W| * \mu \in \mathcal{A}^+_{\mathrm{loc}}$ (or equivalently, by Theorem B.4, $|W| * \nu \in \mathcal{A}^+_{\mathrm{loc}}$). Then $W \in \mathcal{G}^1_{\mathrm{loc}}(\mu)$ and*

$$W * (\mu - \nu) = W * \mu - W * \nu.$$

For any $W \in \tilde{\mathcal{P}}$, let now define the following two increasing (possibly infinite) predictable processes

$$C(W)_t = |W - \hat{W}|^2 * \nu_t + \sum_{s \le t} (1 - \hat{\nu}_s(E)) \, |\hat{W}_s|^2, \tag{B.18}$$
$$\bar{C}(W)_t = |W - \hat{W}| * \nu_t + \sum_{s \le t} (1 - \hat{\nu}_s(E)) \, |\hat{W}_s|.$$

The sets $\mathcal{G}^1_{\mathrm{loc}}(\mu)$ and $\mathcal{G}^2_{\mathrm{loc}}(\mu)$ can be characterized in the following way.

Theorem B.19 (Theorem 1.33, point c), Chapter II, in [**79**])**.** *Let $W \in \tilde{\mathcal{P}}$. Then W belongs to $\mathcal{G}^1_{\mathrm{loc}}(\mu)$ if and only if $C(W') + \bar{C}(W'')$ belongs to $\mathcal{A}^+_{\mathrm{loc}}$, where*

$$\begin{cases} W' = (W - \hat{W}) \, 1_{\{|W - \hat{W}| \le 1\}} + \hat{W} \, 1_{\{|\hat{W}| \le 1\}}, \\ W'' = (W - \hat{W}) \, 1_{\{|W - \hat{W}| > 1\}} + \hat{W} \, 1_{\{|\hat{W}| > 1\}}. \end{cases}$$

Proposition B.20 (Proposition 3.71 in [**77**])**.** *Let $W \in \tilde{\mathcal{P}}$. Then $W \in \mathcal{G}^2(\mu)$ if and only if $C(W) \in \mathcal{A}^+$.*

By Proposition B.20, the space $\mathcal{G}^2(\mu)$ can be rewritten as

$$\mathcal{G}^2(\mu) = \{W \in \tilde{\mathcal{P}} : \ ||W||_{\mathcal{G}^2(\mu)} < \infty\},$$

where

$$||W||^2_{\mathcal{G}^2(\mu)} := \mathbb{E}\,[C(W)] = \mathbb{E}\Big[\int_{]0,T] \times E} |W_s(e) - \hat{W}_s|^2 \, \nu(ds\,de) + \sum_{s \le T} |\hat{W}_s|^2 (1 - \hat{\nu}_s(E))\Big]. \tag{B.19}$$

Let us introduce the space

$$\mathcal{L}^2(\mu) := \{W \in \tilde{\mathcal{P}} : \ ||W||_{\mathcal{L}^2(\mu)} < \infty\} \tag{B.20}$$

with

$$||W||_{\mathcal{L}^2(\mu)} := \mathbb{E}\Big[\int_{]0,T] \times \mathbb{R}} |W_s(e)|^2 \, \nu(ds\,de)\Big].$$

We have the following result.

Lemma B.21.

(1) *If $W \in \mathcal{L}^2(\mu)$, then $W \in \mathcal{G}^2(\mu)$ and*

$$||W||^2_{\mathcal{G}^2(\mu)} \le ||W||^2_{\mathcal{L}^2(\mu)}. \tag{B.21}$$

(2) *If $|W|^2 * \mu \in \mathcal{A}^+_{\mathrm{loc}}$ then $W \in \mathcal{G}^2_{\mathrm{loc}}(\mu)$.*

Proof. Let $W \in \tilde{\mathcal{P}}$. For every $t \ge 0$, since $\hat{\nu}_t(\mathbb{R}) \le 1$, we have

$$\sum_{s \in]0,\,t]} |\hat{W}_s|^2 (1 - \hat{\nu}_s(E)) \le \sum_{s \le t} |\hat{W}_s|^2 \le \sum_{s \le t} \hat{\nu}_s(E) \int_E |W_s(e)|^2 \, \hat{\nu}_s(de) \le |W|^2 * \nu_t. \tag{B.22}$$

Assume now that moreover $W \in \mathcal{L}^2(\mu)$. Then (B.22), together with the triangle inequality, implies that

$$\mathbb{E}\Big[\sum_{s\in]0,T]} |\hat W_s|^2 (1-\hat\nu_s(E))\Big] < \infty, \qquad \mathbb{E}\Big[\int_{]0,T]\times E} |W_s(e)-\hat W_s|^2\,\nu(ds\,de)\Big] < \infty,$$

i.e., $W \in \mathcal{G}^2(\mu)$. Moreover, taking into account that

$$|\hat W|^2 * \nu_t = \sum_{s\le t} |\hat W_s|^2\,\hat\nu_s(E), \quad \forall t \ge 0, \tag{B.23}$$

the process $C(W)$ defined in (B.18) can be decomposed as

$$\begin{aligned} C(W)_t &= |W|^2 * \nu_t - 2\sum_{s\le t}|\hat W_s|^2 + \sum_{s\le t}|\hat W_s|^2\,\hat\nu_s(E) + \sum_{s\le t}|\hat W_s|^2\,(1-\hat\nu_s(E)) \\ &= |W|^2 * \nu_t - \sum_{s\le t}|\hat W_s|^2. \end{aligned} \tag{B.24}$$

In particular, we have

$$||W||^2_{\mathcal{G}^2(\mu)} = \mathbb{E}\Big[\int_{]0,T]\times\mathbb{R}} |W_s(e)|^2\,\nu(ds\,de) - \sum_{s\in]0,T]}|\hat W_s|^2\Big] \le ||W||^2_{\mathcal{L}^2(\mu)}.$$

This establishes point 1. Point 2. follows by usual localization arguments. □

Theorem B.22 (Theorem 11.21, point 3), in [**73**]). *Let $W \in \tilde{\mathcal{P}}$. The following properties are equivalent.*

(i) *W belongs to $\mathcal{G}^2_{\mathrm{loc}}(\mu)$.*

(ii) *$C(W)$ belongs to $\mathcal{A}^+_{\mathrm{loc}}$.*

(iii) *W belongs to $\mathcal{G}^1_{\mathrm{loc}}(\mu)$ and $W*(\mu-\nu)$ belongs to $\mathcal{H}^{2,d}_{\mathrm{loc}}$.*

In this case, we have

$$\langle W*(\mu-\nu), W*(\mu-\nu)\rangle_t = C(W)_t. \tag{B.25}$$

If in addition $|W|^2 * \nu_t \in \mathcal{A}^+_{\mathrm{loc}}$, then

$$\langle W*(\mu-\nu), W*(\mu-\nu)\rangle_t = |W|^2 * \nu_t - \sum_{s\le t}|\hat W_s|^2. \tag{B.26}$$

Remark B.23. Let $W \in \tilde{\mathcal{P}}$, and μ an integer-valued random measure with support D. We recall that the random sets J and K have been introduced in Theorem B.10. By definition of $\hat W$, J and K. We have

$$\hat W = \hat W\,1_J, \tag{B.27}$$

$$\hat\nu(E)\,1_K = 1_K, \tag{B.28}$$

$$1-\hat\nu(E) > 0 \quad \text{on } J\setminus K. \tag{B.29}$$

Taking into account (B.27), (B.28) and (B.29), we see that the quantity $C(W)$ in (B.18) can be rewritten as

$$C(W) = |W-\hat W\,1_J|^2 * \nu + \sum_{s\le \cdot}(1-\hat\nu_s(E))\,|\hat W_s|^2\,1_{J\setminus K}(s). \tag{B.30}$$

In the particular case of $K = J$, previous identity reduces to

$$C(W) = |W - \hat{W}\, 1_K|^2 * \nu. \tag{B.31}$$

Denoting $\nu^d = \nu\, 1_K$ and $\nu^c = \nu\, 1_{K^c}$, then

$$C(W) = \int_{]0,\,\cdot]\times\mathbb{R}} |W_s(e)|^2\, \nu^c(ds\, de) + \int_{]0,\,\cdot]\times\mathbb{R}} |W_s(e) - \hat{W}_s\, 1_K(s)|^2\, \nu^d(ds,\, de). \tag{B.32}$$

Remark B.24. It directly follows from (B.31) and from the definition of the $\mathcal{G}^2(\mu)$ seminorm (see (B.19)) that if $K = J$, then

$$||W||^2_{\mathcal{G}^2(\mu)} = ||W - \hat{W}\, 1_K||^2_{\mathcal{L}^2(\mu)} = ||W - \hat{W}||^2_{\mathcal{L}^2(\mu)}.$$

Proposition B.25. *Let (l_s) be a predictable process. Then $C(l\, 1_K) = 0$.*

Proof. By definition

$$(\widehat{l_s\, 1_K(s)}) = \int_E l_s\, 1_K(s)\, \hat{\nu}_s(de) = l_s\, 1_K(s)\, \hat{\nu}_s(E) = l_s\, 1_K(s), \tag{B.33}$$

where the latter equality follows from (B.28). Then (B.30) in Remark B.23 gives

$$C(l\, 1_K) = |l\, 1_K - l\, 1_K|^2 * \nu + \sum_{s\leq\cdot} (1 - \hat{\nu}_s(E))\, 1_K(s)\, |l_s|^2\, 1_{J\setminus K}(s) = 0.$$

□

Proposition B.26. *Let $W \in \tilde{\mathcal{P}}$. Then for any predictable process (l_s),*

$$C(W) = C(W + l\, 1_K).$$

Proof. We designate $W^0 = W + l\, 1_K$. Taking into account (B.33), we have

$$\hat{W}^0_s = (\widehat{W_s + l_s\, 1_K(s)}) = \hat{W}_s + l_s\, 1_K(s).$$

Then, recalling (B.30), we get

$$\begin{aligned} C(W^0) &= |W^0 - \hat{W}^0\, 1_J|^2 * \nu + \sum_{s\leq\cdot} (1 - \hat{\nu}_s(E))\, |\hat{W}^0_s|^2\, 1_{J\setminus K}(s) \\ &= |W + l\, 1_K - \hat{W}\, 1_J - l\, 1_K|^2 * \nu + \sum_{s\leq\cdot} (1 - \hat{\nu}_s(E))\, |\hat{W}_s + l_s\, 1_K(s)|^2\, 1_{J\setminus K}(s) \\ &= |W - \hat{W}\, 1_J|^2 * \nu + \sum_{s\leq\cdot} (1 - \hat{\nu}_s(E))\, |\hat{W}_s|^2\, 1_{J\setminus K}(s) = C(W). \end{aligned}$$

□

Corollary B.27. *Let $(l_s)_{s\in[0,\,T]}$ be a predictable process. If $W \in \mathcal{G}^2(\mu)$, then*

$$W + l\, 1_K \in \mathcal{G}^2(\mu), \tag{B.34}$$

and

$$||W + l\, 1_K||_{\mathcal{G}^2(\mu)} = ||W||_{\mathcal{G}^2(\mu)}. \tag{B.35}$$

Proof. (B.34) (resp. (B.35)) is a consequence of Proposition B.26 and Proposition B.20 (resp. formula (B.19)). □

Proposition B.28. *If $W \in \mathcal{G}^2(\mu)$ and $||W||_{\mathcal{G}^2(\mu)} = 0$, then*

$$||W - \hat{W}\, 1_K||_{\mathcal{L}^2(\mu)} = 0. \tag{B.36}$$

In particular, there is a predictable process (l_s) such that

$$W_s(e) = l_s\, 1_K(s), \quad \nu(ds\, de)\text{-a.e.}$$

Proof. Since $||W||_{\mathcal{G}^2(\mu)} = 0$, we have $C(W)_T = 0$ a.s., see (B.19). Recalling (B.30), this implies

$$\begin{cases} |W - \hat{W}\, 1_J|^2 * \nu = 0, \\ \sum_{s \leq \cdot} (1 - \hat{\nu}_s(E))\, |\hat{W}_s|^2\, 1_{J \setminus K}(s) = 0. \end{cases}$$

Since $1 - \hat{\nu}(E) > 0$ on $J \setminus K$ (see Remark B.23), previous identities imply

$$\begin{cases} |W - \hat{W}\, 1_J|^2 * \nu = 0, \\ \hat{W}\, 1_{J \setminus K} = 0, \end{cases}$$

which gives (B.36). □

Remark B.29. If $K = \emptyset$, then

$$||W||^2_{\mathcal{G}^2(\mu)} = 0 \text{ if and only if } ||W||^2_{\mathcal{L}^2(\mu)} = 0.$$

Indeed, by Proposition B.28, $K = \emptyset$ and $||W||^2_{\mathcal{G}^2(\mu)} = 0$ imply that $||W||^2_{\mathcal{L}^2(\mu)} = 0$. The opposite implication follows from the fact that $||W||^2_{\mathcal{G}^2(\mu)} \leq ||W||^2_{\mathcal{L}^2(\mu)}$, see Lemma B.21.

We end this section with an important result of the stochastic integration theory.

Proposition B.30. *Let $W \in \mathcal{G}^1_{\text{loc}}(\mu)$, and define $M_t = \int_{]0,t] \times \mathbb{R}} W_s(e)\, (\mu - \nu)(ds\, de)$. Let moreover (Z_t) be a predictable process such that*

$$\sqrt{\sum_{s \leq \cdot} Z_s^2 |\Delta M_s|^2} \in \mathcal{A}^+_{\text{loc}}. \tag{B.37}$$

Then $\int_0^\cdot Z_s\, dM_s$ is a local martingale and equals

$$\int_{]0,\cdot] \times \mathbb{R}} Z_s\, W_s(e)\, (\mu - \nu)(ds\, de). \tag{B.38}$$

Remark B.31. Since M is a local martingale, $\sqrt{[M,M]_t} \in \mathcal{A}^+_{\text{loc}}$, see e.g. Theorem 2.34 and Proposition 2.38 in [**77**]. Taking into account that M is a purely jump local martingale, by Proposition 5.3 in [**9**] this is equivalent to $\sqrt{\sum_{s \leq \cdot} |\Delta M_s|^2} \in \mathcal{A}^+_{\text{loc}}$. Then condition (B.37) is verified if for instance when Z is locally bounded.

Proof. The conclusion follows by the definition of the stochastic integral (B.38), see Definition B.16, provided we check the following three conditions.

(i) $\int_0^\cdot Z_s\, dM_s$ is a local martingale.

(ii) $\int_0^\cdot Z_s\, dM_s$ is a purely discontinuous local martingale; in agreement with Theorem A.6, we will show $[\int_0^\cdot Z_s\, dM_s, N] = 0$ for every N continuous local martingale vanishing at zero.

(iii) $\Delta\left(\int_0^\cdot Z_s\, dM_s\right)_t = \int_{\mathbb{R}} Z_t\, W_t(e)\, (\mu(\{t\}, de) - \nu(\{t\}, de)), \quad t \in [0,\, T].$

We prove now the validity of (i), (ii) and (iii).

Condition (B.37) is equivalent to $\sqrt{\int_0^t Z_s^2\, d[M,M]_s} \in \mathcal{A}_{\mathrm{loc}}^+$. According to Definition 2.46 in [**77**], $\int_0^t Z_s\, dM_s$ is the unique local martingale satisfying

$$\Delta\left(\int_0^\cdot Z_s\, dM_s\right)_t = Z_t\, \Delta M_t,\ \ t \in [0, T]. \tag{B.39}$$

This implies in particular item (i).

By Theorem 29, Chapter II, in [**110**], it follows that

$$\left[\int_0^\cdot Z_s\, dM_s, N\right] = \int_0^\cdot Z_s\, d[M,N]_s,$$

and item (ii) follows because M is orthogonal to N, see Theorem A.6.

Finally, by Definition B.16, taking into account (B.39), $\Delta\left(\int_0^\cdot Z_s\, dM_s\right)_t$ equals

$$Z_t\, \Delta M_t = \int_{\mathbb{R}} Z_t\, W_t(e)\, (\mu(\{t\}, de) - \nu(\{t\}, de))$$

for every $t \in [0,\, T]$, and this shows item (iii). $\square$

Bibliography

[1] Almudevar, A. A dynamic programming algorithm for the optimal control of piecewise deterministic markov processes. *SIAM Journal on Control and Optimization*, 40(2):525–539, 2001.

[2] Bandini, E. Existence and uniqueness for backward stochastic differential equations driven by a random measure. *Electronic Communications in Probability*, 20(71):1–13, 2015.

[3] Bandini, E. Optimal control of Piecewise Deterministic Markov Processes: a BSDE representation of the value function. *Preprint arXiv:1512.01659*, 2015.

[4] Bandini, E. & Confortola, F. Optimal control of semi-Markov processes with a backward stochastic differential equations approach. *Preprint, arXiv:1311.1063.*, 2013.

[5] Bandini, E. & Cosso, A. & Fuhrman, M. & Pham, H. Randomized Zakai equation and Bellman equation for partial observation control problem. *Work in preparation.*

[6] Bandini, E. & Cosso, A. & Fuhrman, M. & Pham, H. Randomization method and backward SDEs for optimal control of partially observed path-dependent stochastic systems. *Preprint arxiv:1511.09274*, 2015.

[7] Bandini, E. & Fuhrman, M. Constrained BSDEs representation of the value function in optimal control of pure jump Markov processes. *Preprint arXiv:1501.04362.*, 2015.

[8] Bandini, E. & Russo, F. Special weak Dirichlet processes and BSDEs driven by a random measure. *Preprint HAL*, https://hal.archives-ouvertes.fr/hal-01241076, 2015.

[9] Bandini, E. & Russo, F. Weak Dirichlet processes with jumps. *Preprint HAL*, https://hal-ensta.archives-ouvertes.fr/hal-01241073, 2015.

[10] Barles, G. & Buckdahn, R. & Pardoux, E. Backward stochastic differential equations and integral-partial differential equations. *Stochastic Rep.*, 60(1-2):57–83, 1997.

[11] Bauerle, N. & Rieder, U. Optimal control of piecewise deterministic markov processes with finite time horizon. *In Modern Trends in Controlled Stochastic Processes: Theory and Applications.*, pages 144–160, 2010.

[12] Becherer, D. Bounded solutions to backward SDEs with jumps for utility optimization and indifference hedging. *Ann. of App. Prob.*, 16:2027–2054, 2006.

[13] Bensoussan, A. *Stochastic control by functional analysis methods.* Studies in mathematics and its applications. Amsterdam North-Holland Pub. Co., 1982.

[14] Bertoin, J. Les processus de Dirichlet en tant qu'espace de Banach. *Stochastics*, 18(2):155–168, 1986.

[15] Bertsekas, D.P. & Shreve, S.E. *Stochastic optimal control. The discrete time case.*, volume 139 of *Mathematics in Science and Engineering.* Academic Press, 1978.

[16] Billingsley, P. *Convergence of Probability Measures.* Wiley Series in Probability and Statistics. John Wiley and Sons, INC., second edition, 1999.

[17] Brandt, A. & Last, G. . *Marked point processes on the real line. The dynamic approach.* Springer, 1995.

[18] Brémaud, P. *Point processes and queues, Martingale dynamics.* Series in Statistics. Springer, 1981.

[19] Brezis, H. *Functional Analysis. Sobolev Spaces and Partial Differential Equations.* Springer, 2010.

[20] Buckdahn, R. Backward stochastic differential equations driven by a martingale. *Unpublished*, 1993.

[21] Buckdahn, R. & Pardoux, E. BSDE's with jumps and associated integral-stochastic differential equations. *Preprint*, 1994.

[22] Carbone, R. & Ferrario, B. & Santacroce, M. Backward stochastic differential equations driven by càdlàg martingales. *Theory Probab. Appl.*, 52:304–314, 2008.

[23] Cheridito, P. & Soner, M. & Touzi, N. & Victoir, N. Second-order backward stochastic differential equations and fully nonlinear PDEs. *Communication in Pure and Applied Mathematics*, 60:1081–1110, 2007.

[24] Chitopekar, S. S. Continuous time Markovian sequential control processes. *SIAM J. Control*, 7:367–389, 1969.

[25] Choukroun, S. & Cosso, A. Backward SDE representation for stochastic Control Problems with Non Dominated Controlled Intensity. *To appear on Annals of Applied Probability.*, 2015.

[26] Cohen, S. & Elliott, R. . Existence, uniqueness and comparisons for BSDEs in general spaces. *The Annals of Probability*, 40:2264–2297, 2012.

[27] Confortola, F. & Fuhrman, M. Backward stochastic differential equations and optimal control of marked point processes. *SIAM Journal on Control and Optimization*, 124(51):3592–3623, 2013.

[28] Confortola, F. & Fuhrman, M. Backward stochastic differential equations associated to Markov jump processes and applications. *Stochastic Processes and their Applications*, 124:289–316, 2014.

[29] Confortola, F. & Fuhrman, M. & Jacod, J. Backward stochastic differential equations driven by a marked point process: an elementary approach, with an application to optimal control. *Preprint, arXiv:1407.0876*, 2014.

[30] Coquet, F. & Jakubowski, A. & Mémin, J. & Słomiński, L. Natural decomposition of processes and weak Dirichlet processes. In *In memoriam Paul-André Meyer: Séminaire de Probabilités XXXIX*, volume 1874 of *Lecture Notes in Math.*, pages 81–116. Springer, Berlin, 2006.

[31] Cosso, A. & Fuhrman, M. & Pham, H. Long time asymptotics for fully nonlinear Bellman equations: a Backward SDE approach. *Preprint arXiv 1410.1125*, 2014.

[32] Costa, O. L. & Dufour, F. *Continuous Average Control of Piecewise Deterministic Markov Processes.* Springer Briefs in Mathematics. Springer, 2013.

[33] Crépey, S. & Matoussi, A. Reflected and doubly reflected BSDEs with jumps. *Ann. of Appl. Prob.*, 18:2041–2069, 2008.

[34] Davis, M. H. A. Control of piecewise-deterministic processes via discrete-time dynamic programming. *In Stochastic differential systems (Bad Honnef, 1985), Lecture Notes in Control and Inform. Sci., Springer, Berlin*, 78:140–150, 1986.

[35] Davis, M. H. A. *Markov models and optimization.*, volume 49 of *Monographs on Statistics and Applied Probability.* Chapman & Hall., 1993.

[36] Davis, M. H. A. & Farid, M. Piecewise deterministic processes and viscosity solutions. *Stochastic Analysis, Control Optimization and Applications*, A Volume in Honor of W. H. Fleming on Occasion of His 70th Birthday:249–268, 1999.

[37] Dellacherie, C. & Meyer, P.-A. *Probabilités et potentiel.* Hermann, Paris, 1975. Chapitres I à IV, Édition entièrement refondue, Publications de l'Institut de Mathématique de l'Université de Strasbourg, No. XV, Actualités Scientifiques et Industrielles, No. 1372.

[38] Dellacherie, C. & Meyer, P.-A. *Probabilities and Potential.* Hermann, Paris, 1978. Chapters I to IV.

[39] Delong, L. *Backward stochastic differential equations with jumps and their actuarial and financial applications.* Springer, 2013.

[40] Dempster, M. A. H. Optimal control of piecewise deterministic Markov processes. *Applied stochastic analysis*, pages 303–325, 1989.

[41] Dempster, M. A. H. & Ye, J.J. Necessary and sufficient conditions for control of piecewise deterministic processes. *Stochastics and Stochastics Reports*, 40:125–145, 1992.

[42] Dempster, M. A. H. & Ye, J.J. Generalized Bellman-Hamilton-Jacob optimality conditions for a control problem with boundary conditions. *Applied Mathematics and Optimization*, 33(3):211–225, 1996.

[43] Di Girolami, C. & Russo, F. Generalized covariation and extended Fukushima decompositions for Banach space valued processes. Application to windows of Dirichlet processes. *Infinite Dimensional Analysis, Quantum Probability and Related Topics (IDA-QP)*, 15(2), 2012.

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

[45] Di Nunno, G. & Øksendal, B. & Proske, F. *Malliavin calculus for Lévy processes with applications to finance.* Universitext. Springer-Verlag, Berlin, 2009.

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

[47] Eisenbaum, N. Local time-space calculus for reversible semimartingales. In *Séminaire de Probabilités XL*, volume 1899 of *Lecture Notes in Math.*, pages 137–146. Springer, Berlin, 2007.

[48] Ekren, I. & Keller, K & Touzi, N. & Zhang, J. On viscosity solutions of path dependent pdes. *to appear on Annals of Probability.*, 2011.

[49] El Karoui, N. Les aspects probabilistes du contrôle stochastique. In *École d'été de probabilités de Saint-Flour, XII—1981*, volume 876 of *Lecture Notes in Math.*, pages 73–238. Springer, Berlin, 1981.

[50] El Karoui, N. & Huang, S. J. A general result of existence and uniqueness of backward stochastic differential equations. *Pitman Research Notes in Mathematics Series, Longman, Harlow*, 364:27–36, 1995-1996.

[51] El Karoui N. & Kapoundijan C. & Pardoux E. & Quenez M.C. Reflected solutions of backward SDEs, and related obstacle problems for PDEs. *Ann. Probab. 25*, pages 702–737, 1997.

[52] El Karoui, N. & Mazliak, L., editor. *Backward Stochastic Differential Equations.* Pitman Research Notes in Mathematics Series. Longman, 1997.

[53] El Karoui, N. & Peng, S. & Quenez, M.C. Backward stochastic differential equations in finance. *Math. Finance*, 7(1):1–71, 1997.

[54] Elie, R. & Kharroubi, I. Probabilistic representation and approximation for coupled systems of variational inequalities. *Statist. Probab. Lett.*, 80(17-18):1388–1396, 2010.

[55] Elie, R. & Kharroubi, I. Adding constraints to BSDEs with jumps: an alternative to multidimensional reflections. *ESAIM Probab. Stat.*, 18:233–250., 2014.

[56] Elie, R. & Kharroubi, I. BSDE representations for optimal switching problems with controlled volatility. *Stoch. Dyn.*, 14:1450003, 15 pp., 2014.

[57] Elliott, R.J. *Stochastic Calculus and its Applications.* Springer, 1982.

[58] Errami, M. & Russo, F. n-covariation, generalized Dirichlet processes and calculus with respect to finite cubic variation processes. *Stochastic Process. Appl.*, 104(2):259–299, 2003.

[59] Errami, M. & Russo, F. & Vallois, P. Itô's formula for $C^{1,\lambda}$-functions of a càdlàg process and related calculus. *Probab. Theory Related Fields*, 122(2):191–221, 2002.

[60] Essaky, E. H. Reflected backward stochastic differential equation with jumps and RCLL obstacle. *Bull. Sci. Math.*, 132(8):690–710, 2008.

[61] Fabbri, G. & Gozzi, F. & Swiech, A. *Stochastic Optimal Control in Infinite Dimensions: Dynamic Programming and HJB Equations.* book in preparation, draft.

[62] Flandoli, F. & Issoglio, E. & Russo, F. Multidimensional stochastic differential equations with distributional drift. *To appear: Transactions of the American Mathematical Society. Available at http://arxiv.org/abs/1401.6010*, 2014.

[63] Flandoli, F. & Russo, F. Generalized integration and stochastic ODEs. *Ann. Probab.*, 30(1):270–292, 2002.

[64] Flandoli, F. & Russo, F. & Wolf, J. Some SDEs with distributional drift. I. General calculus. *Osaka J. Math.*, 40(2):493–542, 2003.

[65] Fleming, W. H. & Soner, H. M. *Controlled Markov processes and viscosity solutions*, volume 25 of *Stochastic Modelling and Applied Probability.* Springer, New York, second edition, 2006.

[66] Föllmer, H. Calcul d'Itô sans probabilités. In *Seminar on Probability, XV (Univ. Strasbourg, Strasbourg, 1979/1980) (French)*, volume 850 of *Lecture Notes in Math.*, pages 143–150. Springer, Berlin, 1981.

[67] Fuhrman, M. & Pham, H. Randomized and backward SDE representation for optimal control of non-markovian SDEs. *Annals of Applied Probability*, 25(4):2134–2167, 2015.

[68] Fuhrman, M. & Tessitore, G. Generalized directional gradients, backward stochastic differential equations and mild solutions of semilinear parabolic equations. *Appl Math Optim*, 108:263–298, 2003.

[69] Gihman, I. I. & Skorohod, A. V. *The Theory of Stochastic Processes II.* Springer-Verlag Berlin Heidelberg New-York Tokyo, 1983.

[70] Gozzi, F. & Russo, F. Verification theorems for stochastic optimal control problems via a time dependent Fukushima-Dirichlet decomposition. *Stochastic Process. Appl.*, 116(11):1530–1562, 2006.

[71] Gozzi, F. & Russo, F. Weak Dirichlet processes with a stochastic control perspective. *Stochastic Process. Appl.*, 116(11):1563–1583, 2006.

[72] Guo, X. & Hernández-Lerma, O. *Continuous-time Markov decision processes. Theory and applications.*, volume 62 of *Stochastic Modelling and Applied Probability.* Springer, 2009.

[73] He, S. & Wang, J. & Yan, J. *Semimartingale theory and stochastic calculus.* Science Press Bejiing New York, 1992.

[74] Howard, R. A. *Dynamic Probabilistic Systems.* John Wiley, New York, 1971.

[75] Jacod, J. Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:235–253, 1974/1975.

[76] Jacod, J. Un théorème de représentation pour les martingales discontinues. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 34:225–244, 1976.

[77] Jacod, J. *Calcul Stochastique et Problèmes de martingales*, volume 714 of *Lecture Notes in Mathematics.* Springer, Berlin, 1979.

[78] Jacod, J. & Protter, P. E. *Probability essentials.* Springer Science & Business Media, 2003.

[79] Jacod, J. & Shiryaev, A. N. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, second edition, 2003.

[80] JEANBLANC, M. & MANIA, M. & SANTACROCE, M. & SCHWEIZER, M. Mean-variance hedging via stochastic control and BSDEs for general semimartingales. *Ann. Appl. Probab.*, 22(6):2388–2428, 2012.

[81] JEWELL, W. S. Markov-renewal programming I and II. *Operational Res.*, 11:938–971, 1963.

[82] KARATZAS, I. & SHREVE, S.E. *Brownian Motion and Stochastic Calculus*. Graduate Texts in Mathematics. Springer New York, 1991.

[83] KAZI-TANI, N. & POSSAMAÏ, D. & ZHOU, C. Second Order BSDEs with jumps. Existence and links with PIDEs. *Electronic Journal of Probability*, 20(65):1–31, 2015.

[84] KAZI-TANI, N. & POSSAMAÏ, D. & ZHOU, C. Second Order BSDEs with jumps. Formulation and uniqueness. *Ann. Appl. Probab.*, 25:2867–2908, 2015.

[85] KHARROUBI, I. & LANGRENÉ, N. & PHAM, H. Discrete time approximation of fully nonlinear HJB equations via BSDEs with nonpositive jumps. *Preprint arXiv:1311.4505.*, 2013.

[86] KHARROUBI, I. & LANGRENÉ, N. & PHAM, H. A numerical algorithm for fully nonlinear hjb equations: an approach by control randomization. *Monte Carlo Methods Appl.*, 20(2):145–165., 2014.

[87] KHARROUBI, I. & MA, J. & PHAM, H. & ZHANG, J. Backward SDEs with constrained jumps and quasi-variational inequalities. *Ann. Probab.*, 38(2):794–840, 2010.

[88] KHARROUBI, I. & PHAM, H. Feynman-Kac representation for Hamilton-Jacobi-Bellman IPDE. *Ann. Probab.*, 43(4):1823–1865, 2015.

[89] KRYLOV, N. V. *Controlled Diffusion Processes*, volume 14 of *Applications of Mathematics*. Springer-Verlag, Berlin-Heidelberg-New York edition, 1980.

[90] LAACHIR, I. & RUSSO, F. BSDEs, càdlàg martingale problems and mean-variance hedging under basis risk. *Preprint, arXiv 1411.6368v1*, 2014.

[91] LENHART, S. M. & YAMADA, N. Perron's method for viscosity solutions associated with piecewise-deterministic processes. *Funkcialaj Ekvacioj*, 34:173–186, 1991.

[92] LIM, T. & QUENEZ, M. C. Exponential utility maximization and indifference price in an incomplete market with defaults. *Preprint*, 2010.

[93] MA, J. & YONG, J. *Forward-backward stochastic differential equations and their applications*. Lectures Notes in Mathematics 1702. Springer, 1999.

[94] ØKSENDAL, B. & SULEM, A. & ZHANG, T. A stochastic HJB equation for optimal control of forward-backward SDEs. *Preprint arXiv:1312.1472*, 2013.

[95] OSAKI, S. & MINE, H. Linear programming algorithms for semi-Markovian decision processes. *J. Math. Anal.*, 22:356–381, 1968.

[96] PARDOUX É. Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order. *Progr. Probab. 42*, pages 79–127, 1998.

[97] PARDOUX, É. *Stochastic Analysis and related topics*, volume 42 of *Progress in Probability*, chapter Backward stochastic differential equations and viscosity solutions of systems of semilinear parabolic and elliptic PDEs of second order, pages 79–127. Birkhäuser, 1998.

[98] PARDOUX, É. & PENG, S. Adapted solution of a backward stochastic differential equation. *Systems Control Lett.*, 14(1):55–61, 1990.

[99] PARDOUX, É. & PENG, S. Backward Stochastic Differential Equations and Quasilinear Parabolic Partial Differential Equations. *Lecture Notes in CIS*, 176:200–217, 1992.

[100] PARDOUX, É. & RASCANU, A. *Stochastic differential equations, Backward SDEs, Partial differential equations*. Springer, 2014.

[101] PENG, S. Probabilistic Interpretation for Systems of Quasilinear Parabolic Partial Differential Equations. *Stochastics*, 37:61–74, 1991.

[102] PENG, S. A generalized dynamic programming principle and Hamilton-Jacobi-Bellman Equation. *Stochastics*, 38:119–134, 1992.

[103] PENG, S. Backward Stochastic Differential Equation and Its Application to Optimal Control. *Appl. Math. Optim.*, 27:125–144, 1993.

[104] Peng, S. Monotonic limit theorem for bsdes and non-linear doob-meyer decomposition. *Probability Theory and Related Fields.*, 16:225–234, 2000.

[105] Peng, S. G-expectation, G-Brownian motion and related stochastic calculus of Itô type. *Stochastic analysis and applications.*, 153:541–567, 2007.

[106] Peng, S. Note on viscosity solutions of path-dependent pde and g-martingales. *Preprint arXiv:1106.1144v1*, 2011.

[107] Pham, H. *Continuous-time and stochastic control and optimization with financial applications*, volume 61 of *Stochastic Modeling and Applied Probability*. Springer-Verlag Berlin, 2009.

[108] Pliska, S.R. Controlled jump processes. *Stochastic Processes and their Applications*, 3:259–282, 1975.

[109] Protter, P. & Shimbo, K. No arbitrage and general semimartingales. *IMS Collections, Markov Processes and Related Topics: A Festschrift for Thomas G. Kurtz.*, 4:267–283, 2008.

[110] Protter, P. E. *Stochastic Integration and Differential Equations*. Stochastic Modelling and Applied Probabilities. Springer, second edition, 2000.

[111] Revuz, D. & Yor, M. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.

[112] Ross, S. M. *Applied Probability Models with Optimization Applications*. Holden-Day, San Francisco, 1970.

[113] Royer, M. Backward stochastic differential equations with jumps and related non-linear expectations. *Stochastic Processes and their Applications 116*, 10:1358–1376, 2006.

[114] Royer, M. Backward stochastic differential equations with jumps and related nonlinear expectations. *Stochastic Processes and their Applications*, 10(116):1358–1376, 2006.

[115] Russo, F. & Trutnau, G. Some parabolic PDEs whose drift is an irregular random noise in space. *Ann. Probab.*, 35(6):2213–2262, 2007.

[116] Russo, F. & Vallois, P. Intégrales progressive, rétrograde et symétrique de processus non adaptés. *C. R. Acad. Sci. Paris Sér. I Math.*, 312(8):615–618, 1991.

[117] Russo, F. & Vallois, P. Forward, backward and symmetric stochastic integration. *Probab. Theory Related Fields*, 97(3):403–421, 1993.

[118] Russo, F. & Vallois, P. Noncausal stochastic integration for làd làg processes. In *Stochastic analysis and related topics (Oslo, 1992)*, volume 8 of *Stochastics Monogr.*, pages 227–263. Gordon and Breach, Montreux, 1993.

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

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

[121] Russo, F. & Vallois, P. Elements of stochastic calculus via regularization. In *Séminaire de Probabilités XL*, volume 1899 of *Lecture Notes in Math.*, pages 147–185. Springer, Berlin, 2007.

[122] Russo, F. & Wurzer, L. Elliptic PDEs with distributional drift and backward SDEs driven by a càdlàg martingale with random terminal time. *Preprint HAL INRIA 01023176*, 2014.

[123] Sayah, A. Equations d'Hamilton-Jacobi du premier ordre avec termes intégro-différentiels. Partie I: Unicité des solutions de viscosité. Partie II: Existence de solutions de viscosité. *Comm. in Partial Differential Equations*, 16(6-7):1075–1093, 1991.

[124] Soner, M. & Touzi, N. & Zhang, J. The wellposedness of second order backward SDEs. *Probability Theory and Related Fields.*, 153:149–190, 2011.

[125] Stone, L. D. Necessary and sufficient conditions for optimal control of semi-Markov jump processes. *SIAM J. Control Optim.*, 2(11):187–201, 1973.

[126] Stroock, D. W. & Varadhan, S. S. *Multidimensional diffusion processes*. Springer, 2007.

[127] TANG S. & ZHANG, F. Path-dependent optimal stochastic control and viscosity solution of associated bellman equations. *Preprint arxiv:1210.2078*, 2013.

[128] TANG, S. J. & LI, X. J. Necessary conditions for optimal control of stochastic systems with random jumps. *SIAM J. Control Optim.*, 32:1447–1475, 1994.

[129] VERMES, D. Optimal control of piecewise deterministic Markov process. *Stochastics*, 3(14):165–207, 1985.

[130] WATANABE, S. On discontinuous additive functionals and Lévy measures of a Markov process. *Japanese J. Math.*, 34:53–70, 1964.

[131] XIA, J. Backward stochastic differential equations with random measures. *Acta Mathematicae Applicatae Sinica*, 16(3):225–234, 2000.

[132] YONG, J. & ZHOU, X. Y. *Stochastic Controls: Hamiltonian systems and HJB equations*, volume 43 of *Applications of Mathematics*. Springer-Verlag, 1999.