République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOHAMED KHIDER, BISKRA

FACULTÉ des SCIENCES EXACTES et des SCIENCES de la NATURE et de la VIE

DEPARTEMENT DE MATHEMATIQUES

Thèse présentée en vue de l'obtention du Diplôme de

Doctorat en Sciences Mathématiques

Option: **Probabilités et Statistiques**

Par

Saloua Labed

Titre

Calcul Stochastique et Optimisation Dynamique des Processus Aléatoires

Soutenue le:

Devant le jury composé de:

Abdelhakim Necir	**Pr**	**UMK Biskra**	**Président**
Brahim Mezerdi	**Pr**	**UMK Biskra**	**Rapporteur**
Djamel Meraghni	**Pr**	**UMK Biskra**	**Examinateur**
Khaled Bahlali	**MC(A)**	**Univ. Toulon**	**Examinateur**
Salah eddine Rebiai	**Pr**	**UMB Batna 2**	**Examinateur**
Khaled Melkemi	**Pr**	**UMB Batna 2**	**Examinateur**

Dedicace

Je dédie ce travail :

À la mémoire de ma mère,

À mon cher père,

À ma chère tante,

À mes frères et sœurs,

À tous ceux et celles qui m'ont encouragé, entouré de leur soutien durant les moments difficiles avec tant d'amour et de compréhension.

Un grand merci à tout qui m'a conduit à ce jour mémorable.

REMERCIEMENTS

En premier lieu, je tiens à remercier mon directeur de thèse, le Pr. Brahim Mezerdi pour la confiance qu'il m'a accordée en acceptant d'encadrer ce travail doctoral, par ses multiples conseils. Enfin, j'ai été extrêmement sensible à ses qualités humaines d'écoute et de compréhension tout au long de ce travail doctoral.

Je n'oublierai jamais l'aide, y compris financière, qui m'a été prodiguée par le Laboratoire de Mathématiques Appliquées (LMA) de Biskra.

Mes remerciements s'adressent au Pr. Abdelhakim Necir, directeur du LMA de l'université de Biskra, qui a accepté d'être le président du jury de ma thèse.

Je souhaite aussi adresser ma gratitude au Pr. Salah Eddine Rebiai de l'université de Batna, pour avoir accepté d'examiner mon travail et de faire partie du jury.

Je tiens également à exprimer ma reconnaissance à Dr. Djamel Meraghni, Maître de Conférences à l'université de Biskra, pour avoir accepté de faire partie de ce jury.

J'exprime ma gratitude à Dr. Khaled Bahlali , Maître de conférences à l'université de Toulon, France, pour avoir participé à ce jury.

Je remercie aussi chaleureusement le Pr. Khaled Melkemi de l'université de Batna, d'avoir accepté de participer à ce jury.

Un grand merci s'adresse tout particulièrement à Mr Boubakeur Labed Maître de conférence à l'université de Biskra et à Mr Mokhtar Hafayed Maître de conférence à l'université de Biskra pour ses précieuses aides.

Finalement, je remercie ma famille de son soutien moral. Pour terminer, je remercie tout mes collègues de travaille et amis.

Contents

Introduction

In this thesis, we are interested in optimality necessary conditions for control problems of systems evolving according to the stochastic differential equation

$$dx\left(t\right)=b\left(t,x\left(t\right),u\left(t\right)\right)dt+\sigma\left(t,x\left(t\right),u\left(t\right)\right)dW_t,\quad x(0)=x_0$$

on some filtered probability space $(\Omega,\mathcal{F},(\mathcal{F}_t)_t,P)$, where b and σ are deterministic functions, $(W_t,t\geq 0)$ is Brownian motion, x_0 is the initial state and $u\left(t\right)$ stands for the control variable. Our control problem consists in minimizing a cost functional of the form

$$J(u)=E\left[\int_0^1 h\left(t,x\left(t\right),u\left(t\right)\right)dt+g(x\left(1\right))\right],$$

over the class $\mathcal{U}$ of admissible controls, that is adapted processes, with values in some compact metric space $\mathbb{A}$, called the action space.

Let us first speak quickly about the optimization problems. One of the principal approaches in solving optimization problems is to derive a set of necessary conditions that must be satisfied by any optimal solution. For example, in obtaining an optimum of a finite-dimensional function, one relies on the zero-derivative condition (for the unconstrained case) or the Kuhn-Tucker condition (for the constrained case), which are necessary conditions for optimality. These necessary conditions become sufficient under certain convexity conditions on the objective/constraint functions. But in the problems of optimal control, it become an optimization problems in infinite-dimensional spaces; therefore these problems are substantially difficult to solve.

A control $\bar{u}$ is called optimal if it satisfies

$$J(\bar{u}) = \inf \{J(u), u \in \mathcal{U}\}.$$

If, moreover, $\bar{u}$ is in $\mathcal{U}$, it is called strict. Existence of such a strict control or an optimal control in $\mathcal{U}$ follows from the convexity of the image of the action space by the map $(b(t,x,.), \sigma^2(t,x,.), h(t,x,.))$, called the Filipov-type convexity condition, see [13, 23, 27, 37, 43]. Without this convexity condition an optimal control does not necessarily exist in $\mathcal{U}$, this set is not equipped with a compact topology. The idea is then to introduce a larger class $\mathcal{R}$ of control processes, in which the controller chooses at time t a probability measure $q_t(da)$ on the control set $\mathcal{U}$, rather than an element $u_t \in \mathcal{U}$. These are called relaxed controls and have a richer topological structure, for which the control problem becomes solvable and the SDE will have the form

$$dx(t) = \int_A b(t, x(t), a)\, q_t(da)dt + \int_A \sigma(t, x(t), a)\, M(da, dt), \quad x(0) = x_0,$$

where $M(da, dt)$ is an orthogonal continuous martingale measure, whose intensity is the relaxed control $q_t(da)dt$, and his corresponding cost is given by

$$J(q) = E\left[\int_0^1 \int_A h(t, x(t), a)\, q_t(da)dt + g(x(1))\right].$$

The relaxed control problem finds its interest in two essential points. The first is that an optimal solution exists. Fleming [27] derived an existence result of an optimal relaxed control for systems with uncontrolled diffusion coefficient. The existence of an optimal solution, where the drift and the diffusion coefficients depend explicitly on the relaxed control variable, has been solved by El Karoui et al.[23], see also [37, 36]. The relaxed optimal control in this general case is shown to be Markovian. See also [10] for an alternative proof of the existence of an optimal relaxed control based on Skorokhod selection theorem.

The second advantage of the use of relaxed controls is that it is a generalization of the strict control problem, in the sense that both control problems have the same value function. Indeed, if $q_t(da) = \delta_{u_t}(da)$ is a Dirac measure charging u_t for each t, we get a strict control as a particular case of the relaxed one.

Motivated by the existence of an optimal relaxed control, various versions of the stochastic maximum principle have been proved. The first result in this direction has been established in [51], where a stochastic maximum principle for relaxed controls, in the case of uncontrolled diffusion coefficient has been given by using the first order adjoint process (see also [9] the extension to singular control problems). The case of a controlled diffusion coefficient has been treated in [10], by using Ekeland's variational principle and an approximation scheme, by using the first and second order adjoint processes. Let us point out that a different relaxation has been used in [3, 1], where the drift and diffusion coefficient have been replaced by their relaxed counterparts. Their relaxed state process is linear in the control variable and is different from ours, in the sense that in our case we relax the infinitesimal generator instead of relaxing directly the state process. Then, we obtain a maximum principle of the Pontryagin type.

The maximum principle of Pontryagin type is formulated and derived by the Russian mathematician Lev Pontryagin and his students in the 1950s. This principle used in optimal control theory, he is truly a milestone of optimal control theory. He find that any optimal control along with the optimal state trajectory must solve the so-called Hamiltonian system, it can also be called a forward-backward differential equation, where we can compare it with the stochastic case, a maximum condition of a function called the Hamiltonian. Its proof is historically based on maximizing the Hamiltonian. The initial application of this principle was to the maximization of the terminal speed of a rocket. However, as it was subsequently mostly used for minimization of a performance index it has here been referred to as the minimum principle. The mathematical significance of the maximum principle lies in that maximizing the Hamiltonian is much easier than the original control problem that is infinite-dimensional. Another approach of the Pontryagin type is a Peng-type.

The aim of the present this work is to obtain a Peng-type general stochastic maximum principle for relaxed controls, using directly the spike perturbation. Our method differs from the one used in [10], in the sense that we don't use neither the approximation procedure nor Ekeland's variational principle. We use a spike variation method directly on the relaxed optimal control. Then, we derive the variational equation from the state equation and the variational inequality from the inequality

$$J\left(q^{\theta}\right)-J\left(q\right)\geq 0.$$

As for strict controls, the first order expansion of $J\left(q^{\theta}\right)$ is not sufficient to obtain a necessary optimality condition. One has to consider the second-order terms (with respect to the state) in the expansion of $J\left(q^{\theta}\right)-J\left(q\right)$. Although the second-order terms are quadratic with respect to the state variable, a so called second-order variational equation and second-order variational inequality are introduced. By using a suitable predictable representation theorem for martingale measures [55], we obtain the corresponding first and second-order adjoint equations, which are linear backward stochastic differential equations driven by the optimal martingale measure. This could be seen as one of the novelties of this work.

This thesis is organized as follows. In the first chapter, we begin by given a definition and basic properties of martingale measures and we look about examples of martingale measures, then we go to in important result which is the representation of martingale measures, where we can discover that the intensity of martingale measures can be decompose, and the construction of martingale measures, without forget the representation of vector martingale measures. Finally, we set two essentially results, which are the stability theorem for martingale measures and the approximation by the integral of Brownian motion, which they have big applications, since we set here a famous lemma Known by the name of chattering lemma. Let us point out that in this work, we interest exactly to orthogonal continuous martingale measures.

In the second chapter, we are interesting to give the general stochastic maximum principle for control problems, and we refer the interested reader to the famous references Peng [56], Young Zho [60]. Here we have first state of the stochastic maximum principle, which contain the adjoint equations, maximum principle and stochastic Hamiltonian systems, then we go to the proof of the maximum principle which is rather lengthy and technical, we need Taylor expansions and duality analysis, then the completion of the proof.

In the last chapter, we present our result which is the generalization of the second chapter result in the case that here we have stochastic differential equations driven by orthogonal martingale measures. But before this, may we speak about the cases which lead to relax our problem, for this we begin by setting the control problem which decompose to the strict control problem and relaxed control problem, then we present a predictable representation for martingale measures and a representation of relaxed control problems. Finally, present our main result. We obtain a maximum principle of the Pontriagin type for relaxed controls, extending the well known Peng stochastic maximum principle to the class of measure-valued controls.

Chapter 1

Martingale measures and basic properties

Martingale measure theory was introduced by JB Walsh in 1984 [59]. The idea was to construct a stochastic calculus for two parameter "space-time" processes having a martingale property in the time variable and a measure property in space. Martingale measure arise in the representation of processes whose quadratic variation is the integral of a space-time function.

1.1 Definition and basic properties of martingale measures

Considering set functions on $\mathbb{R}^{d+1}$ with all coordinates treated symmetrically, we choose one coordinate to be the "time" and the other coordinates to be the "space".

Let us begin with some remarks on random set functions and vector-valued measures. Let $(E, \mathcal{E})$ a Lusin space, i.e a measurable space homeomorphic to a Borel subset of the line. (this includes all Euclidean space and, more generally, all Polish spaces).

We consider a function $U(A,\omega)$ defined on $\mathcal{A}\times\Omega$, where $\mathcal{A}$ is a subring of $\mathcal{E}$ which satisfies

$$\|U(A)\|_2^2 = E\left[U(A)^2\right] < \infty; \quad \forall A \in \mathcal{A}.$$

Suppose that U is finitely additive

$$\text{if } A\cap B = \varnothing \Rightarrow U(A)+U(B) = U(A\cup B) \ \text{ a.s. } \ \forall A \text{ and } B \text{ in } \mathcal{A}.$$

In most interesting cases U will not be countably additive if we consider it as a real-valued set function. However, it may become countably additive if we consider it as a set function with values in $\mathrm{L}^2(\Omega,\mathcal{F},P)$. Let $\|U(A)\|_2 = E\left[U(A)^2\right]^{\frac{1}{2}}$ be the L^2-norm of $U(A)$.

We will say that the map U is σ-finite when there exists an increasing sequence (E_n) of E such that

1. $\bigcup_n E = E$;
2. $\forall n$, $\mathcal{E}_n = \mathcal{E}_{|E_n} \subseteq \mathcal{A}$;
3. $sup\left\{\|U(A)\|_2, A\in\mathcal{E}_n\right\} < \infty$.

Define a set function μ by

$$\mu(A) = \|U(A)\|_2^2.$$

A σ-finite additive set function U is countably additive on $\mathcal{E}_n$ (as an L^2-valued set function) iff

$$A_j \in \mathcal{E}_n, \forall n, A_j \downarrow \varnothing \Rightarrow \lim_{j\to\infty}\mu(A_j) = 0. \tag{1.1}$$

If U is countably additive on $\mathcal{E}_n$, $\forall n$, we can make a trivial further extension: if $A\in\mathcal{E}_n$, set $U(A) = \lim_{n\to\infty} U(A\cap E_n)$ if the limit exists in L^2, and let $U(A)$ be undefined. This leaves U unchanged on each $\mathcal{E}_n$, but may change its values on some sets $A\in\mathcal{E}$ which are not any $\mathcal{E}_n$. We will assume below that all our countably additive set functions have been extended in this way. We will say that such a U is σ-finite L^2-valued measure.

Definition 1.1.1 *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space satisfying the usual condition (El Karoui et Méleard [22]). $\{M_t(A), t \geqq 0, A \in \mathcal{A}\}$ is a $\mathcal{F}_t-$martingale measure if and only if*

1) $M_0 = 0, \quad \forall A \in \mathcal{A}$

2) $\{M_t(A), t \geqq 0\}$ *is a $\mathcal{F}_t$-martingale,* $\forall A \in \mathcal{A}$

3) $\forall t > 0, M_t(.)$ *is a L^2-valeud σ-finite measure.*

Remark 1.1.1 *When we integrate over dx for fixed t- this is the Bochner integral- and over dt for fixed sets A - this is the Ito integral. The problem facing us now is to integrate over dx and dt at the same time.*

There are two rather different classes of martingale measures which have been popular, orthogonal martingale measures and martingale measures with a nuclear covariance.

Definition 1.1.2 *A martingale measure M is orthogonal if, for any two disjoint sets A and B in $\mathcal{A}$, the martingales $\{M_t(A), t \geq 0\}$ and $\{M_t(B), t \geq 0\}$ are orthogonal.*

Equivalently, M is orthogonal if the product $M_t(A) M_t(B)$ is a martingale for any two disjoint sets A and B. This is in turn equivalent to having $\langle M(A), M(B)\rangle_t$, the predictable process of bounded variation, vanish.

Definition 1.1.3 *A martingale measure M has nuclear covariance if there exists a finite measure η on $(E, \mathcal{E})$ and a complete ortho-normal system (ϕ_k) in $L^2(E, \mathcal{E}, \eta)$ such that $\eta(A) = 0 \Rightarrow \mu(A) = 0$ for all $A \in \mathcal{E}$ and*

$$\sum_k E\left(M_t(\phi_k)^2\right) < 0$$

where $M_t(\phi_k) = \int \phi_k(x) M_t(dx)$ is a Bochner integral.

1.1.1 Worthy Measures

Unfortunately, it is not possible to construct a stochastic integral with respect to all martingale measures, so we will need to add some conditions. There are rather strong, and, though sufficient, are doubtless not necessary. However, they are satisfied for both orthogonal martingale measures and those with a nuclear covariance.

Let M be a σ-finite martingale measure. By restricting ourselves to one of the E_n, if necessary, we can assume that M is finite. We shall also restrict ourselves to a fixed time interval $[0,T]$.

Definition 1.1.4 *The covariance function of M is*

$$\bar{Q}_t(A,B) = \langle M(A), M(B)\rangle_t .$$

Note that $\bar{Q}_t$ is symmetric in A and B and biadditive: for fixed A, $\bar{Q}_t(A,.)$ and $\bar{Q}_t(.,A)$ are additive set function. Indeed, if $B \cap C = \varnothing$,

$$\begin{aligned}\bar{Q}_t(A, B\cap C) &= \langle M(A), M(B)+M(C)\rangle_t \\ &= \langle M(A), M(B)\rangle_t + \langle M(A), M(C)\rangle_t \\ &= \bar{Q}_t(A,B) + \bar{Q}_t(A,C).\end{aligned}$$

Moreover, by the general theory,

$$\left|\bar{Q}_t(A,B)\right| \leq Q_t(A,A)^{1/2} Q_t(B,B)^{1/2} .$$

A set $A \times B \times (s,t] \subset E \times E \times \mathbb{R}_+$ will be called a rectangles. Define a set function Q on rectangles by

$$Q(A \times B \times (s,t]) = \bar{Q}_t(A,B) - \bar{Q}_s(A,B),$$

and extend Q by additivity to finite disjoint unions of rectangles, i.e. if $A_i \times B_i \times (s_i, t_i]$

are disjoint, $i = 1, ..., n$ set

$$Q\left(\bigcup_{i=1}^{n} A_i \times B_i \times (s_i, t_i]\right) = \sum_{i=1}^{n} \left[\bar{Q}_{t_i}(A_i, B_i) - \bar{Q}_{s_i}(A_i, B_i)\right].$$

Definition 1.1.5 *A signed measure $K(dx, dy, ds)$ on $\mathcal{E} \times \mathcal{E} \times \mathcal{B}$ is positive definite if for each bounded measurable function f for which the integral makes sense,*

$$\int_{E \times E \times \mathbb{R}_+} f(x, s) f(y, s) K(dx, dy, ds) \geq 0.$$

For such a positive definite signed measure K, define

$$(f, g)_K = \int_{E \times E \times \mathbb{R}_+} f(x, s) g(y, s) K(dx, dy, ds) \geq 0.$$

Note that $(f, f)_K \geq 0$ by the last inequality.

We are led to the following definition.

Definition 1.1.6 *A martingale measure M is worthy if there exist a random $\sigma-$finite measure $K(\Lambda, w)$, $\Lambda \in \mathcal{E} \times \mathcal{E} \times \mathcal{B}$, $w \in \Omega$, such that*

i) *K is positive definite and symmetric in x and y,*

ii) *for fixed A, B, $\{K(A \times B \times (0, t]), t \geq 0\}$ is predictable,*

iii) *for all n, $E\{K(E_n \times E_n \times (0, T])\} < \infty$,*

iv) *for any rectangle Λ, $|Q(\Lambda)| \leq K(\Lambda)$.*

We call K the dominating measure of M.

Remark 1.1.2 *1. The requirement that K be symmetric is no restriction see [59] for more detaile.*

2. Both orthogonal martingale measures and those with nuclear covariance are worthy. But, we will show it below only for orthogonal martingale measures.

If M is worthy with covariance Q and dominating measure K, then $K+Q$ is a positive set function. The $\sigma-$field $\mathcal{E}$ is separable, so that we can first restrict ourselves to a countable subalgebra of $\mathcal{E} \times \mathcal{E} \times \mathcal{B}$ upon which $Q(., w)$ is finitely additive for a.e. w. Then $K+Q$ is a positive finitely additive set function by the measure $2K$, and hence can be extended to a signed measure on $E \times E {\times} \mathcal{B}$, and the total variation of Q satisfies

$$|Q|(\Lambda) \leq K(\Lambda)$$

for all $\mathcal{E} \times \mathcal{E} \times \mathcal{B}$.

Let

$$\triangle(E) = \{(x, x) : x \in E\},$$

be the diagonal of E.

Proposition 1.1.1 *A worthy martingale measure is orthogonal iff Q is support by* $\triangle(E) \times \mathbb{R}_+$.

Proof. $Q(A \times B \times (0, t]) = \langle M(A), M(B) \rangle_t$.

If M is orthogonal and $A \cap B = \varnothing$, this vanishes hence

$$|Q|[(A \times B - \triangle(E)) \times \mathbb{R}_+] = 0,$$

i.e. $\sup pQ \subset \triangle(E) \times \mathbb{R}_+$. Conversely, if this vanishes for all disjoint A and B, M is evidently orthogonal. ■

Definition 1.1.7 *If M is a martingale measure and if, moreover, for all A of $\mathcal{A}$, the map $t \to M_t(A)$ is continuous, we will say that M is continuous.*

We can associate with each set A of $\mathcal{A}$ the increasing process $\langle M(A)\rangle$ of the martingale $\{M_t(A), t \geqq 0\}$. The process can be regularized in a positive measure on $\mathbb{R}_+ \times E$, in the following sense

Theorem 1.1.1 *(Walsh [59]) If M is a $\mathcal{F}_t-$orthogonal martingale measure, there exists a random σ-finite positive measure $\upsilon(ds, dx)$ on $\mathbb{R}_+ \times E$, $\mathcal{F}_t-$predictable, such that for each A of $\mathcal{A}$ the process $(\upsilon((0,t] \times A))_t$ is predictable, and satisfies*

$$\forall A \in \mathcal{A}, \forall t > 0, \qquad \upsilon((0,t] \times A) = \langle M(A)\rangle_t \quad P\text{-}a.s.$$

If M is continuous, υ is continuous. The measure υ is called the intensity of M.

Remark 1.1.3 *1) We have*

$$\forall A, B \in \mathcal{A}, \forall t > 0, \qquad \langle M(A), M(B)\rangle_t = \langle M(A \cap B)\rangle_t = \upsilon((0,t] \times A \cap B) \quad P\text{-}a.s.$$

The measure υ characterizes thus completely all quadratic variations of the orthogonal martingale measure M.

2) In the following, measures on $\mathbb{R}_+ \times E$ are positive and σ-finite.

1.1.2 Stochastic integrals

Let M be a worthy martingale measure on the Lusin space $(E \times \mathcal{E})$, and let Q_M and K_M be its covariation and dominating measures respectively. This definition of the stochastic integral may look unfamiliar at first, but it merely following Ito's construction in a different setting.

In the classical case, one constructs the stochastic integral as a process rather than as a random variable. That is, one construct $\left\{\int_0^t f dW, t \geq 0\right\}$ simultaneously for all t, one can then say that the integral is a martingale, for instance. The analogue of "martingale" in this setting is "martingale measure". According, they define this stochastic integral as a martingale measure.

Recall that we are restricting ourselves to a finite time interval $(0, T]$ and to one of the E_n, so that M is finite. As usual, they first define the integral for elementary functions, then for simple functions, and then for all functions in a certain class by a functional completion argument.

Definition 1.1.8 *(Walsh [59]) A function $f(x, s, w)$ is elementary if it is of form*

$$f(x, s, w) = X(w)\, 1_{(a,b]}(s)\, 1_A(x),$$

where $0 \leq a < t$, X is bounded and $\mathcal{F}_a$-measurable, and $A \in \mathcal{E}$. f is simple if it is a finite sum of elementary function. We denote the class of simple function by $\mathcal{S}$.

Definition 1.1.9 *The predictable σ-field $\mathcal{P}$ on $\Omega \times E \times \mathbb{R}_+$ is the σ-field generated by $\mathcal{S}$. A function is predictable if it is $\mathcal{P}$-measurable.*

They define a norm $\|.\|_M$ on the predictable functions by

$$\|f\|_M = E\left\{(|f|, |f|)_K\right\}^{1/2}.$$

Note that they have used the absolute value of f to define $\|f\|_{M'}$ so that

$$(f, f)_Q \leq \|f\|_M^2.$$

Let $\mathcal{P}_M$ be the class of all predictable f for which $\|f\|_M < \infty$.

Proposition 1.1.2 *Let $f \in \mathcal{P}_M$ and let $A = \{(x, s) : |f(x, s)| \geq \epsilon\}$. Then*

$$E\left\{K(A \times E \times [0, T])\right\} \leq \frac{1}{\epsilon} \|f\|_M\, E\left\{K(E \times E \times [0, T])\right\}.$$

Proof.

$$\begin{aligned}\epsilon E\left\{K\left(A\times E\times[0,T]\right)\right\} &\leq E\left\{\int|f(x,t)|\,K(dx,dy,dt)\right\}=E\left\{(|f|,1)_K\right\}\\ &\leq E\left\{(|f|,|f|)_K^{1/2}\,K(E\times E\times[0,T])\right\}\\ &\leq \|f\|_M\,E\left\{K(E\times E\times[0,T])\right\}^{1/2}\end{aligned}$$

where we have used Schawartz's inequality in two forms. ∎

Proposition 1.1.3 *$\mathcal{S}$ is dense in $\mathcal{P}_M$.*

Proof. If $f\in\mathcal{P}_M$, let

$$f_N(x,s)=\begin{cases} f(x,s) & \text{if } |f(x,t)|<N\\ 0 & \text{otherwise}\end{cases},$$

then

$$\|f-f_N\|_M=E\left\{\int|f(x,s)-f_N(x,s)|\,|f(y,s)-f_N(y,s)|\,K(dx,dy,ds)\right\}$$

which goes to zero by monotone convergence. Thus the bounded functions are dense. If f is bounded step function, i.e. if there exist $0\leq t_0<t_1<...<t_n$ such that $t\to f(x,t)$ is constant on each (t_j,t_{j+1}), then f can be uniformly approximated by simple functions. It remains to show that the step function are dense in the bounded functions.

To simplify our notation, let us suppose that $K(E\times E\times ds)$ is absolutely continuous with respect to Lebesgue measure. [We can always make a preliminary time change to assure this.] If $f(x,s,w)$ is bounded and predictable, set

$$f_n(x,s,w)=2^{-n}\int_{(k-1)2^{-n}}^{k2^{-n}}f(x,u,w)\,du \text{ if } k2^{-n}\leq s\leq(k+1)\,2^{-n},$$

fix w and x. Then $f_n(x,s,w)\to f(x,s,w)$ for a.e. s by either the martingale convergence theorem or Lebesgue's differentiation theorem. It follows easily that $\|f-f_N\|_M\to 0$. ∎

Now the integral can be constructed with a minimum of interruption. If

$$f(x,s,w) = X(w)\, 1_{(a,b]}(s)\, 1_A(x)$$

is an elementary function, define a martingale measure $f.M$ by

$$f.M_t(B) = X(w)\left(M_{t\wedge b}(A\cap B) - M_{t\wedge a}(A\cap B)\right). \tag{1.2}$$

Lemma 1.1.1 *$f.M$ is a worthy martingale measure. Its covariance and dominating measures $Q_{f.M}$ and $K_{f.M}$ are given by*

$$Q_{f.M}(dx,dy,ds) = f(x,s)\, f(y,s)\, Q_M(dx,dy,ds) \tag{1.3}$$

$$K_{f.M}(dx,dy,ds) = |f(x,s)\, f(y,s)|\, K_M(dx,dy,ds). \tag{1.4}$$

Moreover

$$E\left\{f.M_t(B)^2\right\} \le \|f\|_M^2 \text{ for all } B\in\mathcal{E}, t\le T. \tag{1.5}$$

Proof. $f.M_t(B)$ is adapted since $X \in \mathcal{F}_a$; it is square integrable, and a martingale. $B \to f.M_t(B)$ is countably additive (in L^2), which is clear from (1.2). Moreover

$$\begin{aligned} &f.M_t(B)\, f.M_t(C) - \int_{B\times C\times[0,t]} f(x,s)\, f(y,s)\, Q_M(dx,dy,ds) \\ &= X^2\left[\left(M_{t\wedge b}(A\cap B) - M_{t\wedge a}(A\cap B)\right)\left(M_{t\wedge b}(A\cap C) - M_{t\wedge a}(A\cap C)\right)\right. \\ &\quad \left. - \langle M(A\cap B), M(A\cap C)\rangle_{t\wedge b} + \langle M(A\cap B), M(A\cap C)\rangle_{t\wedge a}\right] \end{aligned}$$

which is a martingale. This proves (1.3), and (1.4) follows immediately since $K_{f.M}$ is positive and positive definite. (1.5) then follows easily. ■

We now define $f.M$ for $f \in \mathcal{S}$ by linearity.

Suppose now that $f \in \mathcal{P}_M$. By Proposition 1.1.3 there exist $f_n \in \mathcal{S}$ such that $\|f - f_n\|_M \to 0$. By (1.5), if $A \in \mathcal{E}$ and $t \leq T$,

$$E\left\{(f_m.M_t(A) - f_n.M_t(A))^2\right\} \leq \|f_m - f_n\|_M \to 0$$

as $m, n \to \infty$. It follows that $(f_n.M_t(A))$ is Cauchy in $L^2(\Omega, \mathcal{F}, P)$, so that it converge in L^2 to a martingale which we shall call $f.M_t(A)$. The limit is independent of the sequence (f_n).

Theorem 1.1.2 *If $f \in \mathcal{P}_M$, then $f.M$ is a worthy martingale measure. It is orthogonal if M is. Its covariance and dominating measures respectively are given by*

$$Q_{f.M}(dx, dy, ds) = f(x, s) f(y, s) Q_M(dx, dy, ds), \tag{1.6}$$

$$K_{f.M}(dx, dy, ds) = |f(x, s) f(y, s)| K_M(dx, dy, ds). \tag{1.7}$$

Moreover, if $g \in \mathcal{P}_M$ and $A, B \in \mathcal{E}$, then

$$\langle f.M(A), g.M(B)\rangle_t = \int_{A\times B\times[0,t]} f(x, s) g(y, s) Q_M(dx, dy, ds) \tag{1.8}$$

$$E\left\{f.M_t(A)^2\right\} \leq \|f\|_M^2. \tag{1.9}$$

Proof. $f.M(A)$ is the L^2 limit of the martingales $f_n.M(A)$, and is hence a square-integrable martingale. For each n

$$f_n.M_t(A) f_n.M_t(B) - \int_{A\times B\times[0,t]} f_n(x, s) f_n(y, s) Q_M(dx, dy, ds) \tag{1.10}$$

is a martingale. $f_n.M_t(A)$ and $f_n.M_t(B)$ each converge in L^2, hence their product converges in L^1. Moreover

$$E\left\{\left|\int_{A\times B\times[0,t]}\left(f_n(x,s)\,f_n(y,s)-f(x,s)\,f(y,s)\right)Q_M(dx,dy,ds)\right|\right\}$$
$$\leq E\left\{\int_{E\times E\times[0,T]}|f_n(x)|\,|f_n(y)-f(y)|\,K_M(dx,dy,ds)\right\}$$
$$+E\left\{\int_{E\times E\times[0,T]}|f_n(x)-f(x)|\,|f(y)|\,K_M(dx,dy,ds)\right\}$$
$$\leq E\left\{(|f_n|,|f-f_n|)_K+(|f-f_n|,|f|)_K\right\}$$
$$\leq \left(\|f_n\|_M+\|f\|_M\right)\|f-f_n\|_M\to 0$$

we use Schwartz in the last inequality. Thus the expression (1.10) converge in L^1 to

$$f.M_t(A)\,f.M_t(B)-\int_{A\times B\times[0,t]}f(x,s)\,f(y,s)\,Q_M(dx,dy,ds)$$

which is therefore a martingale. The latter integral, being predictable, must therefore equal $\langle f.M(A), f.M(B)\rangle_t$, which verifies (1.6), and (1.7) follows.

This see that $f.M_t(A)$ is a martingale measure, we must check countable additivity. If $A_n\subset E$, $A_n\downarrow\varnothing$, then

$$E\left\{f.M_t(A_n)^2\right\}\leq E\left\{\int_{A_n\times A_n\times[0,t]}|f(x,s)\,f(y,s)|\,K(dx,dy,ds)\right\}$$

which goes to zero by monotone convergence.

If M is orthogonal, Q_M sits on $\triangle(E)\times[0,T]$, hence, by (1.6), so does $Q_{f.M}$. Then, $f.M$ is orthogonal. ■

Now that the stochastic integral is defined as a martingale measure, we define the usual stochastic integral by

$$\int_{A\times[0,t]}f\,dM=f.M_t(A)$$

and

$$\int_{E\times[0,t]}f\,dM=f.M_t(E)$$

while

$$\int f dM = \lim_{t\to\infty} f.M_t(E).$$

When it is necessary we will indicate the variables of integration. For instance

$$\int_{A\times[0,t]} f(x,s)\, dM(dx,ds) \text{ and } \int_A \int_{[0,t]} f(x,s)\, dM_{xs}$$

both denote $f.M_t(A)$.

It is frequently necessary to change the order of integration in iterated stochastic integrals. Here is a form of stochastic Fubini's theorem which will be useful.

Let $(G, \mathcal{G}, \mu)$ be a finite measure space and let M be a martingale with dominating measure K.

Theorem 1.1.3 *Let $f(x,s,w,\lambda)$, $x \in E$, $s \geq 0$, $w \in \Omega$, $\lambda \in G$ be a $\mathcal{P} \times \mathcal{G}$-measurable function. Suppose that*

$$E\left\{\int_{E\times E\times[0,T]\times G} |f(x,s,w,\lambda) f(y,s,w,\lambda)|\, K(dx,dy,ds)\,\mu(d\lambda)\right\} < \infty.$$

Then

$$\int_G \left[\int_{E\times[0,t]} f(x,s,\lambda)\, M(dx,ds)\right] \mu(d\lambda) = \int_{E\times[0,t]} \left[\int_G f(x,s,\lambda)\,\mu(d\lambda)\right] M(dx,ds).$$

Proof. See Walsh [59] ■

This property characterizes continuous orthogonal martingale measures, in the following sense. From new, when we say martingale measures it means that we speak about orthogonal continuous martingale measures.

Corollary 1.1.1 *Let M be an orthogonal martingale measure on E and $\upsilon(ds, dx)$ a random continuous positive measure on $\mathbb{R}_+ \times E$. Then M is a continuous martingale measure with intensity υ if and only if*

$$E\left(\exp\left[\int_0^t \int_E f(s,x) M(ds,dx) - 1/2 \int_{(0,t]\times E} f^2(s,x)\, \upsilon(ds,dx)\right]\right) = 1 \qquad (1.11)$$
$$\forall f \in L_\upsilon^2.$$

Proof. The condition is clearly necessary.

Conversely, let us consider $f \in L_\upsilon^2$ and the following function F

$$F(w,u,x) = \theta f(w,u,x)\, \mathbf{1}_{]s,t]}(u)\, \mathbf{1}_{G_s}(w),$$

where $G_s \in \mathcal{F}_s$, $0 \leqq s < t$, $\theta \in \mathbb{R}$.

The condition (1.11) implies that

$$E\left(\exp\left[\theta \mathbf{1}_{G_s}(M_t(f) - M_s(f)) - \mathbf{1}_{G_s}\tfrac{\theta^2}{2}\int_s^t \int_E f^2(u,x)\, \upsilon(du,dx)\right]\right) = 1 \text{ i.e}$$
$$E\left(\mathbf{1}_{G_s}\exp\left[\theta(M_t(f) - M_s(f)) - \tfrac{\theta^2}{2}\int_s^t \int_E f^2(u,x)\, \upsilon(du,dx)\right]\right) = P(G_s).$$

Then, for $f \in L_\upsilon^2$, $M_t(f)$ is a continuous martingale with quadratic variation , according to the result of Jacod and Memin [40] about the characterization of continuous martingales.

■

1.2 Examples of martingale measures

1.2.1 Finite space

Let us suppose that E is a finite space $\{a_1, a_2, ..., a_n\}$. A martingale measure is uniquely determined by the n-orthogonal square integrable martingales $(M_t(\{a_i\}))_{i=1}^n$.
Conversely, let $m_t^1, ..., m_t^n$ be n-orthogonal martingales with increasing processes $(C_t^i)_{i=1}^n$; then the mapping

$$M_t(A) = \sum_{i=1}^n m_t^i \delta_{\{a_i\}}(A)$$

defines a martingale measure on E with intensity $dC_t^i \delta_{\{a_i\}}(da)$, since

$$\begin{aligned} \langle M(da) \rangle_t &= \left\langle \sum_{i=1}^n m_t^i \delta_{\{a_i\}}(da) \right\rangle_t \\ &= \sum_{i=1}^n \delta_{\{a_i\}}(da) \left\langle m_t^i \right\rangle_t \\ &= \sum_{i=1}^n dC_t^i \delta_{\{a_i\}}(da). \end{aligned}$$

1.2.2 More generally

Proposition 1.2.1 *Let E be a Lusin space and $(u_s)_{s \geqq 0}$ an E-valued predictable process. Let us consider moreover a square integrable martingale m_t with quadratic variation process C_t. Let*

$$M_t(A) = \int_0^t \mathbf{1}_A(u_s) dm_s, \tag{1.12}$$

for $A \in \mathcal{E}$, then $\{M_t(A), t \geqq 0, A \in \mathcal{A}\}$ is a martingale measure with intensity equal to $\delta_{u_s}(da)dC_s$. If m is continuous, M is continuous.

Conversely, all martingale measures with intensity $\delta_{u_s}(da)dC_s$ are of the form (1.12), with $m_t = M_t(E)$.

Proof. We get immediately that M_t is a martingale measure and

$$\begin{aligned}\langle M(da)\rangle_t &= \left\langle \int_0^t \mathbf{1}_{\{da\}}(u_s)dm_s \right\rangle \\ &= \int_0^t \left(\mathbf{1}_{\{da\}}(u_s)\right)^2 d\langle m\rangle_s \\ &= \int_0^t \mathbf{1}_{\{da\}}(u_s)d\langle m\rangle_s \\ &= \int_0^t \delta_{u_s}(da)dC_s\end{aligned}$$

since $\mathbf{1}_A(u_s) = \begin{cases} 1 & \text{if } u_s \in A \\ 0 & \text{if not} \end{cases} = \delta_{u_s}(A)$, then the intensity of $\{M_t(A), t \geqq 0, A \in \mathcal{A}\}$ is $\delta_{u_s}(da)dC_s$.

Conversely; let us study the difference $M_t(A) - M_t(f\mathbf{1}_E)$, $A \in \mathcal{E}$, where $f(\omega, s) = \mathbf{1}_A(u_s(\omega))$.

Let us remark that

$$\begin{aligned}M_t(f\mathbf{1}_E) &= \int_E \int_0^t (\mathbf{1}_A(u_s)\mathbf{1}_E(u_s))\, M(da, ds) = \int_E \int_0^t \mathbf{1}_{A\cap E}(u_s)M(da, ds) \\ &= \int_E \int_0^t \mathbf{1}_A(u_s)M(da, ds) = \int_0^t \mathbf{1}_A(u_s)\int_E M(da, ds) \\ &= \int_0^t \mathbf{1}_A(u_s)M(E, ds) = \int_0^t \mathbf{1}_A(u_s)dm_s;\end{aligned}$$

because $m_s = M_s(E)$ and f is not depending on a.

$M_t(A) - M_t(f\mathbf{1}_E)$ is a martingale with increasing process

$$\begin{aligned}\langle M(A) - M(f\mathbf{1}_E)\rangle_t &= \int_E \int_0^t (\mathbf{1}_A(u_s) - f(s))^2\, \delta_{u_s}(da)dC_s \\ &= \int_0^t \left(\mathbf{1}_A(u_s) - 2\mathbf{1}_A(u_s)f(s) + f^2(s)\right) dC_s \\ &= \int_0^t (\mathbf{1}_A(u_s) - f(s))^2\, dC_s = 0\end{aligned}$$

then

$$M_t(A) = M_t(f\mathbf{1}_E) = \int_0^t \mathbf{1}_A(u_s)dm_s \quad P - p.s$$

■

1.2.3 White noises

As the Brownian motion in the theory of continuous martingales, there exist fundamental martingale measure: white noises. Let us consider a centered Gaussian measure W on $(\mathbb{R}_+ \times E, \mathfrak{B}(\mathbb{R}_+) \otimes \mathcal{E}, \mu)$, where μ is a positive $\sigma-$finite measure on $\mathbb{R}_+ \times E$, defined by

$$\forall h \in L^2_\mu, \qquad E(\exp W(h)) = \exp\left(\frac{1}{2}\int_{\mathbb{R}_+ \times E} h^2(y)\,\mu(dy)\right). \tag{1.13}$$

A construction of such a measure is given by Neveu [53].

The process $B_t(A) = W((0,t] \times A)$, defined for the state $A \in \mathcal{A}$ which satisfy

$$\mu = ((0,t] \times A) < \infty, \forall t > 0,$$

is then a Gaussian process with independent increments and intensity μ, with cadlag trajectories. It is easy to show that $\{B_t(A), t \geqq 0, A \in \mathcal{A}\}$ is a martingale measure with a deterministic intensity, with respect to its natural filtration. When μ is continuous, its continuity is proven according to Corollary 1.1.1 and the characterization (1.13).

Definition 1.2.1 *When the measure μ is continuous, the family $\{B_t(A), t \geqq 0, A \in \mathcal{A}\}$ is called white noise with intensity μ.*

White noises are completely determined by the deterministic nature of their intensity.

Proposition 1.2.2 *Let $\{M_t(A), t \geqq 0, A \in \mathcal{A}\}$ be a $\mathcal{F}_t-$martingale measure with a deterministic continuous intensity υ. Then, M is a white noise (with respect to its natural filtration)*

1.2.4 Image martingale measures

Definition 1.2.2 *$(E, \mathcal{E})$ and $(U, \mathcal{U})$ are two Lusin spaces. Let N be a martingale measure with intensity $\upsilon(ds, dx)$ on $\Omega \times \mathbb{R}_+ \times U$ and $\phi(w, s, u)$ a $\mathcal{P} \otimes \mathcal{U}-$measurable $E-$valued process.*

Let

$$M_t(w, B) = \int_0^t \int_U \mathbf{1}_B(\phi(w, s, u)) N(w, ds, du).$$

$\{M_t(B), t \geq 0, B \in \mathcal{E}\}$ defines a martingale measure with intensity μ, where μ is given by

$$\mu((0, t] \times B) = \int_{(0,t]} \int_U \mathbf{1}_B(\phi(s, u)) \upsilon(ds, du).$$

M is called image martingale measure of N under ϕ. Let us remark that N is continuous, M is also continuous.

1.3 Representation of martingale measures

1.3.1 Intensity decomposition. Construction of martingale measures

We will prove first that the form $q_t(dx)\, dk_t$ for a martingale measure intensity is not a restrictive assumption.

Lemma 1.3.1 *Let $\upsilon(dt, du)$ be a random predictable $\sigma-$finite measure. υ can be decomposed as follows;*

$$\upsilon(dt, du) = q_t(dx)\, dk_t$$

where k_t is a random predictable increasing process and $(q_t(dx)\, dk_t)_{t \geq 0}$ is a predictable family of random $\sigma-$finite measures.

Proof. We will use the notation of section 2.

If υ is a finite measure, the lemma is well known. Otherwise, there exists a $\mathcal{P}\otimes\mathcal{E}-$measurable function $W : \Omega \times \mathbb{R}_+ \times E \to (0, \infty)$ such that

$$\upsilon'(dt, dx) = \upsilon(dt, dx) .W(t, x)$$

is finite. Then we can decompose

$$\upsilon'(dt, dx) = q'_t(dx) dk_t;$$

the result follows by setting

$$q_t(dx) = W(t, x)^{-1} .q'_t(dx) .$$

■

Remark 1.3.1 *This decomposition is not unique, and it is always possible to assume that the process k_t is increasing, for example by replacing k_t by $k_t + t$. In the following, we will use this decomposition of the intensity in which the time coordinate plays a special role, and we will denote the intensities of martingale measures in the form $q_t(dx) dk_t$, with an increasing processes $(k_t)_{t\geq 0}$.*

An important result is that is always possible to give a representation of the random measures as image measures of deterministic measures (cf. A.V Skorohod [58], N. Elkaroui and J.P. Lepeltier [20], B. Grigelionis [33])

Theorem 1.3.1 *Let $(q_t(dx))_{t\geq 0}$ be a predictable family of random $\sigma-$finite measures, defined on a Lusin space $(E, \mathcal{E})$.*

Let us also consider a Lusin space $(U,\mathcal{U})$ and a deterministic diffuse $\sigma-$finite measure λ on U which satisfies

$$q_t(E) \leq \lambda(U) \quad \forall t \in \mathbb{R}_+, \forall w \in \Omega.$$

Then there exists a predictable process $\varphi(t,u)$, with values in $E \cup \{\delta\}$, (δ is the cemetery point), such that

$$q_t(A) = \int_U \mathbf{1}_A(\varphi(t,u))\lambda(du) \quad \forall A \in \mathcal{E}, \forall w \in \Omega \tag{1.14}$$

and a predictable kernel from E to U. $Q(t,x,du)$ which satisfies

$$\int_U \mathbf{1}_B(u) f(\varphi(t,u))\lambda(du) = \int_E f(x) Q(t,x,B)\, q_t(dx) \tag{1.15}$$

$\forall w \in \Omega, \forall f$ *measurable positive,* $\forall B \in \mathcal{U}$.
The kernel $Q(t,x,.)$ is the conditional law of u with respect to the $\sigma-$field generated by φ.

According to this theorem, the existence of a continuous martingale measure with intensity $q_t(dx)\,dk_t$, follows immediately from the existence of a white noise, as the construction will show it. When k_t is deterministic, the martingale measure is given as image measure of white noise, and the general case follows by using a time-change.

Theorem 1.3.2 *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered space and υ a random positive continuous $\sigma-$finite measure, satisfying*

$$\upsilon(dt,dx) = q_t(dx)\,dk_t, \qquad \begin{cases} (k_t) & \text{continuous and increasing} \\ (q_t) & \text{predictable.} \end{cases}$$

There exist on an extension $\hat{\Omega} = \left(\Omega \times \tilde{\Omega}, \mathcal{F} \times \tilde{\mathcal{F}}, (\mathcal{F}_t \times \tilde{\mathcal{F}}_t)_{t\geq 0}, P \times \tilde{P}\right)$ a continuous martingale measure N with intensity υ, obtained as time-changed image measure of a white noise.
Moreover, N is orthogonal to each continuous $(\mathcal{F}_t, P)$ martingale measure M.

Proof. i) Let us assume that k_t is deterministic.
We can build on an auxiliary space $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t\geq 0}, \tilde{P})$ a white noise B with intensity $\lambda(du)\,dk_t$, where λ satisfies the assumpositions of Theorem 1.3.1. On the extension $\left(\hat{\Omega}, \hat{\mathcal{F}}, (\hat{\mathcal{F}}_t)_{t\geq 0}, \hat{P}\right) =$

$\left(\Omega \times \tilde{\Omega}, \mathcal{F} \times \tilde{\mathcal{F}}, (\mathcal{F}_t \times \tilde{\mathcal{F}}_t)_{t \geq 0}, P \times \tilde{P}\right)$, B is a continuous martingale measure with a deterministic intensity and then a $(\hat{\mathcal{F}}_t)-$white noise (Proposition 1.2.2). Let $\varphi(t,u)$ be the predictable process satisfying (1.14). It is clear that φ is $\hat{\mathcal{P}} \otimes \mathcal{U}$ measurable, $\hat{\mathcal{P}}$ being the predictable $\sigma-$field on the extension $\hat{\Omega}$.

By the definition 1.2.2 and (1.14), the family

$$N_t(w, w', A) = \int_0^t \int_U \mathbf{1}_A(\varphi(w,s,u)) B(w', ds, du), \quad A \in \mathcal{E},$$

is a continuous martingale measure with intensity

$$\int_0^t \int_U \mathbf{1}_A(\varphi(w,s,u)) \lambda(du)\, dk_s = v\left((0,t] \times A\right).$$

Moreover, B and each $(\mathcal{F}_t, P)-$martingale measure M are orthogonal (by construction, M is again in a $\hat{\mathcal{F}}_t-$martingale measure). We verify that each predictable step function , the martingale measure $\int_0^t \int_U h(\varphi(s,u)) B(ds, du)$ and M are orthogonal, and that this property is more generally satisfied for h in $L^2(dP \otimes q_t(dx)\, dk_t)$. That implies immediately the orthogonality for M and N.

ii) If k_t is not deterministic, let us consider $\tau_t = \inf\{s > 0, k_s \geq t\}$. τ_t is then the increasing inverse of k_t. We can consider the $\sigma-$finite random measure $\gamma(dt, dx) = q_{\tau_t}(dx)\, dt$, where q_τ is predictable (for the filtration $\mathcal{F}_{\tau_t}$).

According to i), we construct a white noise B with intensity $\lambda(du)\, dt$, φ a predictable process (for $\mathcal{F}_{\tau_t}$), such that

$$N_t(A) = \int_0^t \int_U \mathbf{1}_A(\varphi(w,s,u)) B(ds, du), \quad \text{defines for } t \geq 0,\ A \in \mathcal{E},$$

a $\mathcal{F}_{\tau_t}-$martingale measure, with intensity $\gamma(dt, dx)$.

Let us now consider the $\mathcal{F}_t-$martingale measure $\{M_t(A), t \geqq 0, A \in \mathcal{A}\}$ defined by $M_t(A) = N_{k_t}(A)$. The intensity of M is then $q_t(dx)\, dk_t$, since

$$\langle M(A)\rangle_t = \int_0^{k_t}\int_E \mathbf{1}_A(x) q_{\tau_s}(dx)\, ds = \int_0^t \int_E \mathbf{1}_A(x)\, q_u(dx)\, dk_u.$$

■

1.3.2 Extension and representation of martingale measures as image measures of a white noise

Martingale measures can be described as time changed image measures for white noises. To obtain this property, it is necessary to use an extension result, (this idea is due to Funaki [32]), and the following theorem is thus fundamental

Theorem 1.3.3 *Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered space, E and $\tilde{E}$ two Lusin spaces and M a continuous martingale measure with intensity $q_t(dx)\, dk_t$ on $\mathbb{R}_+ \times E$, where k_t is a continuous increasing process and $(q_t(dx))_{t\geq 0}$ is a $\mathcal{F}_t-$predictable family of random measures.*
Let $r_t(x, d\tilde{x})$ be a predictable probability transition kernel from E to $\tilde{E}$ and define the predictable $\sigma-$finite measure $p_t(dx, d\tilde{x})$ on $\mathbb{R}_+ \times E \times \tilde{E}$ as follows:

$$p_t(dx, d\tilde{x}) = q_t(dx)\, r_t(x, d\tilde{x}).$$

Then there exists on an extension $\left(\Omega \times \tilde{\Omega}, \mathcal{F} \otimes \tilde{\mathcal{F}}, P \otimes \tilde{P}\right)$ a continuous martingale measure $\tilde{M}_t(dx, d\tilde{x})$ with intensity $dk_t p_t(dx, d\tilde{x})$ and whose projection on $\mathbb{R}_+ \times E$ is M, i.e.

$$\tilde{M}_t\left(A \times \tilde{E}, (w, \tilde{w})\right) = M_t(A, w), \quad \forall A \in \mathcal{A}, (w, \tilde{w}) \in \Omega \times \tilde{\Omega}, \forall t \geq 0.$$

Proof. Let N be the continuous martingale measure on $E \times \tilde{E}$, built on an auxiliary space $\left(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t\geq 0}, \tilde{P}\right)$ with intensity $dk_t p_t\,(dx, d\tilde{x})$ such that N and each $\mathcal{F}_t-$martingale measure are orthogonal (Theorem 1.1.2).

Let us consider the mapping

$$\tilde{M}_t\,(C) = \int_0^t \int_E r_s\,(x, C)\, M\,(ds, dx) + \int_0^t \int_{E\times\tilde{E}} \left[\mathbf{1}_C\,(x, \tilde{x}) - r_s\,(x, C)\right] N\,(ds, dx, d\tilde{x})$$

$\forall C \in \mathcal{E} \otimes \tilde{\mathcal{E}}$, where $r_s\,(x, C) = \int_{\tilde{E}} \mathbf{1}_C\,(x, \tilde{x})\, r_s\,(x, d\tilde{x})\,.$

The two terms on the right of the above equality are orthogonal continuous martingale measures. $\left\{\tilde{M}_t\,(C)\,, t \geq 0, C \in \mathcal{E} \otimes \tilde{\mathcal{E}}\right\}$ is then a continuous martingale measure with intensity given by

$$\begin{aligned}
&\int_{(0,t]} dk_s \left[\int_E r_s^2\,(x, C)\, q_s\,(dx) + \int_{E\times\tilde{E}} p_s\,(dx, d\tilde{x})\, \left(\mathbf{1}_C\,(x, \tilde{x}) - r_s\,(x, C)\right)^2\right] \\
&= \int_{(0,t]} dk_s \left[\int_E r_s^2\,(x, C)\, q_s\,(dx) + \int_{E\times\tilde{E}} q_s\,(dx)\, r_s\,(x, d\tilde{x})\, \left(\mathbf{1}_C\,(x, \tilde{x}) + r_s^2\,(x, C) - 2r_s\,(x, C)\, \mathbf{1}_C\,(x, \tilde{x})\right)\right] \\
&= \int_{(0,t]} dk_s \int_E r_s\,(x, C)\, q_s\,(dx) \;\; (\; r_s\,(x, .) \text{ is a probability}) \\
&= \int_{(0,t]} dk_s p_s\,(C)\,.
\end{aligned}$$

b) Let us assume that C is in $\mathcal{E}$

$$\mathbf{1}_C\,(x) - \int_{\tilde{E}} r_s\,(x, d\tilde{x})\, \mathbf{1}_C\,(x) = 0 \;\text{ and then }\; \tilde{M}_t(C) = M_t(C).$$

■

This result can be applied to continuous square integrable martingales, by interpreting them as degenerated martingale measures.

Corollary 1.3.1 *Let n_t be a continuous square integrable martingale with increasing process*

$$\langle n\rangle_t(w) = \int_0^t \int_E \sigma^2(w,s,x)\, q_s(w,dx)\, dk_s$$

(k_t) being a predictable family of random measures and $\sigma(s,x)$ a function of $L^2(q_s(dx)dk_s)$. We assume moreover that $n_o = 0$.

There exists on an extension a continuous martingale measure N with intensity $\sigma^2(s,x)q_s(dx)dk_s$ such that

$$n_t = N_t(E).$$

Proof. See [22] ∎

Using Theorem 1.3.3, we can now state that each martingale measure is representable as time-changed image martingale measure of a white noise. An application of this result is given in Méléard, Roelly-Coppoletta [48]; [49]: it allows to give a sense to a stochastic differential equation in the space of vector measures with values in $L^2(\Omega)$ for a certain class of measure-valued branching processes.

Theorem 1.3.4 *Let M be a continuous martingale measure on $(\Omega,\mathcal{F},(\mathcal{F}_t)_{t\geq 0},P)$ with intensity $q_t(dx)\, dk_t$. Let λ be the diffuse $\sigma-$finite measure and φ be the predictable process given in Theorem 1.3.1.*

1. *If (k_t) is deterministic, there exist an extension $(\hat{\Omega},\hat{\mathcal{F}},\hat{\mathcal{F}}_t,\hat{P})$ of $(\Omega,\mathcal{F},\mathcal{F}_t,P)$ and a white noise $B_t(\hat{w},du)$ with intensity $\lambda(du)dk_t$ such that:*

$$\forall f \in L^2(q_s(dx)dk_s), \quad M_t(f) = \int_0^t \int_U f(\varphi(s,u))\, B(ds,du).$$

2. *In the general case, M is a time-changed image martingale measure of a white noise.*

Proof. We use the predictable kernel $Q_t(x,du)$ defined in Theorem 1.3.1 by (1.15). We consider the measure $p_t(dx,du) = Q_t(x,du)\,q_t(dx)$, it satisfies

$$\forall f \in \mathcal{E}, A \in \mathcal{U}, \quad \int_U \mathbf{1}_B(\varphi(t,u))\,\mathbf{1}_A(u)\lambda(du) = \int_{E\times U} \mathbf{1}_B(x)\,\mathbf{1}_A(u)p_t(dx,du)$$

According to Theorem 1.3.3, we build on $E \times U$ a continuous martingale measure $\hat{M}$ with intensity $p_t(dx,du)\,dk_t$ and whose projection onto E is M. The martingale measure $N(dt,du) = \int_E \hat{M}(dt,dx,du)$ has thus the intensity

$$\int_E Q_t(x,du)\,q_t(dx)\,dk_t = dk_t \mathbf{1}_{\{\varphi(t,u)\neq\delta\}}\lambda(du), \quad \delta \text{ cemetery point.}$$

N_t is not a white noise, because its intensity is not deterministic. We build then on an auxiliary space a white noise $W_t(du)$ with intensity $\lambda(du)dk_t$ and we consider the martingale measure

$$B_t(du) = N_t(du) + \mathbf{1}_{\{\delta\}}(\varphi(t,u))\,W_t(du).$$

Then, B is a continuous martingale measure with deterministic intensity and is therefore a white noise (Proposition 1.2.2).

1. Let f be in $L^2(q_s(dx)dk_s)$, then $f \circ \varphi$ belongs to $L^2(dk_t\lambda(du))$ and

$$\begin{aligned}
\int_0^t \int_U f(\varphi(s,u))\,B(ds,du) &= \int_0^t \int_U f(\varphi(s,u))\,N(ds,du) \\
&\quad + \int_0^t \int_U f(\varphi(s,u))\,\mathbf{1}_{\{\delta\}}(\varphi(t,u))\,W(ds,du) \\
&= \int_0^t \int_U f(\varphi(s,u))\,N(ds,du) \\
&= \int_0^t \int_U f(\varphi(s,u)) \int_E \hat{M}(ds,dx,du) \\
&= \int_0^t \int_E \int_U f(\varphi(s,u))\,\hat{M}(ds,dx,du).
\end{aligned}$$

We want to compare this quantity to

$$\int_0^t \int_E f(x)\, M(ds, dx) = \int_0^t \int_E \int_U f(x)\, \hat{M}(ds, dx, du).$$

Then

$$\begin{aligned}
& E\left[\left(\int_0^t \int_E \int_U f(\varphi(s,u))\, \hat{M}(ds, dx, du) - \int_0^t \int_E \int_U f(x)\, \hat{M}(ds, dx, du)\right)^2\right] \\
&= E\left[\int_0^t \int_E \int_U (f(\varphi(s,u)) - f(x))^2\, Q_s(x, du)\, q_s(dx)\, dk_s\right] \\
&= E\left[\int_0^t \int_E \int_U (f(\varphi(s,u)) - f(\varphi(s,u)))^2\, \lambda(du) dk_s\right] = 0.
\end{aligned}$$

Thus

$$\begin{aligned}
\int_0^t \int_E \int_U f(\varphi(s,u))\, \hat{M}(ds, dx, du) &= \int_0^t \int_E \int_U f(x)\, \hat{M}(ds, dx, du) \\
&= \int_0^t \int_E f(x)\, M(ds, dx) \quad P\text{-a.s}
\end{aligned}$$

2. The proof of the generalization is similar to the proof of theorem 1.3.2 (ii).

■

1.3.3 Representation of vector martingale measures

The first theorem of this section gives a representation of vector martingale measures in terms of orthogonal martingale measures, which generalizes the representation theorem for continuous martingales in terms of Brownian motions.

Theorem 1.3.5 *Let $(M^i)_{i=1}^n$ be n continuous martingale measures on a Lusin space E, with intensities*

$$\left\langle M^i(\varphi), M^j(\psi)\right\rangle_t = \int_0^t \int_E \varphi(x)\, \psi(x)\, a_{ij}(s,x)\, q_s(dx) dk_s$$

where

$$a_{ij}(s,x)=\sum_{k=1}^{n}\sigma_{ik}(s,x)\,\sigma_{kj}(s,x)\,,$$

$\forall i,k\in\{1,...,n\}$, $\sigma_{ik}(s,x)\in L^2(q_s(dx)dk_s),(k_t)$ *is a continuous increasing process,* $(q_t(dx))$ *is a predictable process of random finite measures.*
There exists on an extension n *continuous orthogonal martingale measures* $\left(\hat{M}_s^i(dx)\right)_{i=1}^n$ *with intensity* $q_s(dx)dk_s$ *which satisfy*

$$M_t^i(\varphi)=\sum_{k=1}^{n}\int_0^t\int_E\varphi(x)\,\sigma_{ik}(s,x)\,\hat{M}^k(ds,dx)\quad\forall i\in\{1,...,n\}\,.$$

Proof. This theorem is proven with the same method as in [39].
We can suppose that $\sigma(s,x)=a^{1/2}(s,x)$ is the symmetric square root of $a(s,x)$ and define

$$\tilde{\sigma}(s,x)=\lim_{\epsilon\downarrow 0}\ a^{1/2}(s,x)\,(a(s,x)+\epsilon I)^{-1}\,,\quad\forall(s,x)\in\mathbb{R}_+\times E.$$

We have

$$\sigma(s,x)\,\tilde{\sigma}(s,x)=\tilde{\sigma}(s,x)\,\sigma(s,x)=E_R(s,x)\,,$$

where $E_R(s,x)$ is the orthogonal projection onto range $a(s,x)\left(\mathbb{R}^d\right)$ and denote $E_N(s,x)=I-E_R(s,x)$.
We define then, for $i\in\{1,...,n\}$, the continuous martingale measure

$$\hat{M}_s^i(f)=\sum_{k=1}^{n}\int_0^t\tilde{\sigma}_{ik}(s,x)\,f(x)\,M^k(ds,dx)+\sum_{k=1}^{n}\int_0^t\int_E E_N(s,x)\,f(x)\,\tilde{M}^k(ds,dx)$$

where $\left(\tilde{M}^k\right)_{k=0}^n$ are n continuous orthogonal martingale measures with intensity $q_s(dx)dk_s$ built on an auxiliary space. It is therefore easy to verify that

$$\left\langle\hat{M}^i(f)\,,\hat{M}^j(g)\right\rangle_t=\delta_{ij}\int_0^t\int_E f(x)\,g(x)\,q_s(dx)dk_s\quad\forall f,g\in L^2(q_s(dx)dk_s)$$

and that

$$\sum_{k=1}^{n}\int_{0}^{t}\int_{E} f\left(x\right)\sigma_{ik}\left(s,x\right)\hat{M}^{k}\left(ds,dx\right)=M_{t}^{i}\left(f\right).$$

■

(The calculations are carried out in the book of Ikeda and Watanabe [39] p. 90.).

Corollary 1.3.2 *If we use the notations and the result of Theorem* 1.3.4*, and if the process* (k_t) *is deterministic, we can represent the martingale measures* $\left(M^i\right)_{i=1}^n$ *with* n *orthogonal white noises* $\left(B^i\right)_{i=1}^n$ *by*

$$M_{t}^{i}\left(f\right)=\sum_{k=1}^{n}\int_{0}^{t}\int_{U} f\left(\varphi\left(s,u\right)\right)\sigma_{ik}\left(s,\varphi\left(s,u\right)\right)B^{k}\left(ds,du\right).$$

A very interesting problem is to obtain a similar representation theorem for vector square integrable martingales $(m_t^i)_{i=1}^n$ whose quadratic variation process has the special form

$$\left\langle m^{i},m^{j}\right\rangle_{t}=\int_{0}^{t}\int_{E} a_{ij}\left(s,x\right)q_{s}(dx)dk_{s}$$

(where a is a quadratic matrix). The aim is to represent them in terms of orthogonal martingale measures with intensity $q_s(dx)dk_s$. It will be used in particular to describe solutions of martingale problems. To obtain this result, we need an extension property, which generalizes to vector martingales the extension property obtained in corollary 1.3.1 for the dimension one.

Proposition 1.3.1 *Let* $(m_t^i)_{i=1}^n$ *be* n *continuous square integrable martingales such that* $m_0^i=0$*. We assume that the quadratic variation process corresponding to* m_i *and* m_j *is*

$$\left\langle m^{i},m^{j}\right\rangle_{t}=\int_{0}^{t}\int_{E} a_{ij}\left(s,x\right)q_{s}(dx)dk_{s},$$

where: $a\left(s,x\right)=\sigma\left(s,x\right)\sigma^{*}\left(s,x\right)$ *is a* $\mathcal{P}\otimes\mathcal{E}$ *measurable matrix such that*

$$a_{ij}\left(s,x\right)\in L^{2}(q_{s}(dx)dk_{s}),\quad\forall i,j\in\left\{1,...,n\right\},$$

$(k_t)_{t\geq 0}$ is a continuous increasing process, $(q_t(dx))_{t\geq 0}$ is a predictable finite measure-valued process.

Then on an extension, there exist n continuous martingale measures $(M_s^i(dx))_{i=1}^n$ such that $\forall B, C \in \mathcal{E}$,

$$\left\langle M^i(B), M^j(C)\right\rangle_t = \int_0^t \int_E \mathbf{1}_B(x)\,\mathbf{1}_C(x)\,a_{ij}(s,x)\,q_s(dx)dk_s$$

and $M_t^i(E) = m_t^i, \quad \forall t \geq 0.$

Proof.

a) We suppose first that the symmetric matrix $\triangle(s) = \left(\int a_{ij}(s,x)\,q_s(dx)\right)_{\substack{1\leq i\leq n\\ 1\leq j\leq n}}$ is invertible. Let us denote by $\delta(s)$ its inverse. For f in $L^2(q_s(dx)dk_s)$, we will denote $Q(s,f)$ the symmetric matrix $\left(\int a_{ij}(s,x)\,f(x)\,q_s(dx)\right)_{\substack{1\leq i\leq n\\ 1\leq j\leq n}}$, $Q(s,1) = \triangle(s)$.

It is easy to build on a larger space n martingale measures $\left(\hat{N}^i\right)_{i=1}^n$ which satisfy

$$\left\langle \hat{N}^i(f), \hat{N}^j(g)\right\rangle_t = \int_0^t \int_E f(x)\,g(x)\,a_{ij}(s,x)\,q_s(dx)dk_s, \forall f,g \in L^2(q_s(dx)dk_s).$$

In fact, we can define on $\mathbb{R}_+ \times E \times \{1,...,n\}$ a martingale measure N with intensity $\sum_{k=1}^n q_s(dx)dk_s\delta_{\{i\}}(dj)$ (see Theorem 1.3.2) and construct the martingale measures $\left(\hat{N}_s^i(dx)\right)_{i=1}^n$ as follows

$$\forall A \in \mathcal{E}, \hat{N}_s^i(A) = \sum_{k=1}^n \int_0^t \int_A \sigma_{ik}(s,x)\,N(ds,dx,\{k\}).$$

We may take therefore

$$i \in \{1,...,n\}, t\geq 0, f \in L^2(q_s(dx)dk_s),$$

$$M_s^i(f) = \sum_{k=1}^n \int_0^t (Q(s,f)\,\delta(s))_{ik}\,dm_s^k + \sum_{k=1}^n \int_0^t \int_E (f(x)\,I - Q(s,f)\,\delta(s))_{ik}\,\hat{N}^k(ds,dx)$$

(I identity matrix of $\mathcal{M}_n(\mathbb{R})$).

It is immediate to verify that $M_t^i(E) = m_t^i$, since $Q(s,E) = \triangle(s)$. Let us calculate the intensity of $(M^i)_{i=1}^n$: For every f and g in $L^2(q_s(dx)dk_s)$, we set

$$\begin{aligned}
&\langle M^i(f), M^j(g)\rangle_t = \sum_{k,l=1}^n \int_0^t (Q(s,f)\delta(s))_{ik} (Q(s,f)\delta(s))_{jl} \int_E a_{kl}(s,x)\, q_s(dx)dk_s \\
&+\sum_{k,l=1}^n \int_0^t \int_E (f(x) I - Q(s,f)\delta(s))_{ik} (g(x) I - Q(s,g)\delta(s))_{jl}\, a_{kl}(s,x)\, q_s(dx)dk_s \\
&= \int_0^t [Q(s,f)\delta(s)\triangle(s)(Q(s,g)\delta(s))^*]_{ij} (Q(s,f)\delta(s))_{jl}\, dk_s \\
&\quad + \int_0^t \int_E [(f(x) I - Q(s,f)\delta(s))\, a(s,x)\, (g(x) I - Q(s,g)\delta(s))^*]_{ij}\, q_s(dx)dk_s
\end{aligned}$$

$Q(s,.)$ and $\delta(s)$ are symmetric matrices for every s in $\mathbb{R}_+$. Thus,

$$\begin{aligned}
Q(s,f)\delta(s)\triangle(s)(Q(s,g)\delta(s))^* &= Q(s,f)\delta(s)\triangle(s)\delta(s)Q(s,g) \\
&= Q(s,f)\delta(s)Q(s,g)
\end{aligned}$$

and,

$$\begin{aligned}
&\int_E [(f(x) I - Q(s,f)\delta(s))\, a(s,x)\, (g(x) I - Q(s,g)\delta(s))^*]\, q_s(dx) \\
&= \int_E [(f(x) I - Q(s,f)\delta(s))\, a(s,x)\, (g(x) I - \delta(s)Q(s,g))]\, q_s(dx) \\
&= \int_E [f(x)g(x)a(s,x) - f(x)a(s,x)\delta(s)Q(s,g) - Q(s,f)\delta(s)g(x)a(s,x) \\
&\qquad + Q(s,f)\delta(s)a(s,x)\delta(s)Q(s,g)]\, q_s(dx) \\
&= \int_E f(x)g(x)a(s,x)\, q_s(dx) - Q(s,f)\delta(s)Q(s,g).
\end{aligned}$$

So, $\langle M^i(f), M^j(g)\rangle_t = \int_0^t \int_E f(x)g(x)a_{ij}(s,x)\, q_s(dx)dk_s$.

b) When $\triangle(s)$ is not invertible, we use a method similar to that one of Ikeda and

Watanabe [39]: We introduce the symmetric matrix $\tilde{\delta}(s)$ which satisfies

$$\forall s \in \mathbb{R}_+, \tilde{\delta}(s) \triangle (s) = \triangle (s) \tilde{\delta}(s) = E_R(s),$$

where $E_R(s)$ is the orthogonal projection onto range $\triangle(s)\mathbb{R}^d$. We have

$$I - E_R(s) = E_N(s), \quad \text{with} \quad E_N(s) \triangle (s) = 0, \quad \tilde{\delta}(s) \triangle (s) \tilde{\delta}(s) = 0$$

Let us consider now

$$M_t^i(f) = \sum_{k=1}^n \int_0^t \left(Q(s,f) \tilde{\delta}(s) \right)_{ik} dm_s^k + \sum_{k=1}^n \int_0^t \int_E \left(f(x) I - Q(s,f) \tilde{\delta}(s) \right)_{ik} \hat{N}^k(ds,dx).$$

We get

$$\begin{aligned} M_t^i(E) &= \sum_{k=1}^n \int_0^t \left(\triangle(s) \tilde{\delta}(s) \right)_{ik} dm_s^k + \sum_{k=1}^n \int_0^t \int_E \left(I - \triangle(s) \tilde{\delta}(s) \right)_{ik} \hat{N}^k(ds,dx) \\ &= \sum_{k=1}^n \int_0^t (E_R(s))_{ik} dm_s^k + \sum_{k=1}^n \int_0^t \int_E (E_N(s))_{ik} \hat{N}^k(ds,dx) \\ &= m_t^i - \sum_{k=1}^n \int_0^t (E_N(s))_{ik} dm_s^k + \sum_{k=1}^n \int_0^t \int_E (E_N(s))_{ik} \hat{N}^k(ds,dx). \end{aligned}$$

The two right-hand terms have the intensity

$$\begin{aligned} &\sum_{k,l=1}^n \int_0^t (E_N(s))_{ik} (E_N(s))_{jl} \int_E a_{kl}(s,x) q_s(dx) dk_s \\ &= \sum_{l=1}^n \int_0^t (E_N(s) \triangle (s))_{il} (E_N(s))_{jl} dk_s = 0 \end{aligned}$$

and thus vanish.

We verify easily, with an analogous calculation, that the quadratic variation $\langle M^i(f), M^j(g) \rangle_t$

is

$$\int_0^t \int_E f(x) g(x) a_{ij}(s,x) q_s(dx) dk_s.$$

■

Let us give now this theorem, which is obtained immediately by application of Theorem 1.3.5 and Proposition 1.3.1

Theorem 1.3.6 *Let $(m_t^i)_{i=1}^n$ be n continuous square integrable martingales, with (matrix valued) quadratic variation process*

$$\langle m^i, m^j \rangle_t = \int_0^t \int_E a_{ij}(s,x) q_s(dx) dk_s.$$

There exist on an extension n continuous orthogonal martingale measures $\left(\hat{M}_s^i(dx)\right)_{i=1}^n$ with intensity $q_s(dx)dk_s$ which satisfy

$$m_t^i = \sum_{k=1}^n \int_0^t \int_E \sigma_{ik}(s,x) \hat{M}^k(ds,dx), \forall i \in \{1,...,n\}.$$

1.4 Stability theorem for martingale measures

Theorem 1.4.1 *Let M be an orthogonal continuous martingale measure defined on $\Omega \times [0,T] \times E$, with intensity*

$$\upsilon(da,dt) = q_t(da)dk_t.$$

Let us consider a sequence of random predictable measures $(\upsilon^n)_{n\in\mathbb{N}}$ converging weakly to υ on $E \times [0,T]$ P almost surely, such that

$$\upsilon^n(E \times .) = \upsilon(E \times .), \ a.s.$$

Then there exists on an extension of probability space a sequence of orthogonal continuous martingale measures M^n defined on $E \times [0,T]$ with intensity υ^n, such that

For each predictable bounded function φ from $\Omega \times [0,T] \times E$ to $\mathbb{R}$, continuous in the $E-$variable,

$$\lim_{n \to +\infty} E\left[\left(M_t^n(\varphi) - M_t(\varphi)\right)^2\right] = 0.$$

Since $\upsilon^n (E \times .) = \upsilon (E \times .)$, υ^n can be decomposed as $\upsilon^n(da, dt) = q_t^n(da)dk_t$, where $q_t^n(E) = 1$.

To obtained this theorem, we shall prove thanks to a generalization of the Skorohod representation theorem the existence of a sequence of random measures m^n on $E \times E \times [0,T]$ satisfying

$$m^n(dx, dy, dt) = m_t^{-n}(dx, dy)\, dk_t;$$

$$m_t^{-n}(dx, E) = q_t^n(dx),$$

$$m_t^{-n}(E, dy) = q_t(dx),$$

and converging weakly to a measure carried only by the diagonal.

To prove this theorem we need

Lemma 1.4.1 *Under the hypotheses of Theorem* 1.4.1*, for almost all w, there exists a sequence of random probability measures on $E \times E \times [0,T]$, $m^n(w, da, da', dt)$, satisfying*

$$m^n(w, E, da', dt) = \upsilon(w, da', dt) = q_t(w, da')dk_t$$

$$m^n(w, da, E, dt) = \upsilon^n(w, da, dt) = q_t^n(w, da)dk_t$$

and converging weakly to a random probability measure $m(w, da, da', dt)$ on $E \times E \times [0,T]$, such that

$$m(w, da, da', dt) = \upsilon(w, da, dt)\, \delta_a(da').$$

Proof. Fix $w \in \Omega$ such that $(q_t^n(w, da)dk_t)$ converges weakly to $q_t(w, da)dk_t$ on $E \times [0,T]$. Thanks to a generalization of Skorohod's representation theorem, one can construct an auxiliary probability space $\tilde{\Omega}$ and random variables $X_w^n(\tilde{w})$, $X_w^\infty(\tilde{w})$, $T_w(\tilde{w})$ with values respectively in E, E, $[0,T]$, depending measurable on w, such that:

$(X_w^n(\tilde{w}), T_w(\tilde{w}))$ has law $q_t^n(w,da)dk_t$, $(X_w^\infty(\tilde{w}), T_w(\tilde{w}))$ has law $q_t(w,da)dk_t$ and $(X_w^n(\tilde{w}))$ converges for each $\tilde{w}$ of $\tilde{\Omega}$ to $X_w^\infty(\tilde{w})$.

Then $(X_w^n(\tilde{w}), X_w^\infty(\tilde{w}), T_w(\tilde{w}))$ converges (everywhere) to $(X_w^\infty(\tilde{w}), X_w^\infty(\tilde{w}), T_w(\tilde{w}))$.

The law m^n of (X_w^n, X_w^∞, T_w) answers the problem. ■

Remark 1.4.1 *The time-martingale of m^n being the predictable measure dk_t, the dual predictable projection of m^n can be disintegrated in the form $Q_t^n(da, da')\,dk_t$, where Q^n is a predictable kernel. (It suffices to apply the disintegration theorem for dual predictable projections of random measures [41]).*

We will then have for each $\mathcal{P}\otimes\mathcal{E}\otimes\mathcal{E}$ measurable function f defined on $\Omega\times[0,T]\times E\times E$,

$$E\left[\int_{[0,T]\times E^2} f(s,a,a')\, m^n(da,da',ds)\right] = E\left[\int_{[0,T]\times E^2} f(s,a,a')\, Q_s^n(da,da')\, dk_s\right].$$

Moreover, the second martingale of m^n being the predictable process (q_t),

$$Q_t^n(E,da') = q_t(da') \quad dk_t\otimes P \quad a.s.$$

In the same way, the predictable measure m is disintegrated in the form $Q_t(da,da')\,dk_t$, where Q is predictable kernel.

Now, we proof the Theorem 1.4.1, and we use in this proof the Theorem 1.3.3.

Proof. We naturally use the martingale measure $\hat{M}^n$ constructed above.

The first martingale measure of $\hat{M}^n$ defines an orthogonal continuous martingale measure M^n on $E\times[0,T]$ as follows

$$\begin{aligned}&\forall A\in\mathcal{E}, \forall t\in[0,1],\\ &M_t^n(A) = \int_0^t\int_E \hat{Q}_s^n(A,a')\, M(da',ds) + \int_0^t\int_{E\times E}\left(\mathbf{1}_A(a) - \hat{Q}_s^n(A,a')\right) R(da,da',ds).\end{aligned}$$

The sequence of martingale measures (M^n) approximates the martingale measure M in the sense of Theorem 1.4.1:

Consider first a continuous bounded function φ defined on E.

According to the above results, we have

$$M_t^n(\varphi) = \int_0^t \int_{E\times E} \varphi(a)\, \hat{M}^n(da, da', ds)\,; \quad M_t(\varphi) = \int_0^t \int_{E\times E} \varphi(a')\, \hat{M}^n(da, da', ds)\,.$$

Thus,

$$\begin{aligned} E\left[(M_t^n(\varphi) - M_t(\varphi))^2\right] &= E\left[\left(\int_0^t \int_{E\times E} (\varphi(a) - \varphi(a'))\, \hat{M}^n(da, da', ds)\right)^2\right] \\ &= E\left[\int_0^t \int_{E\times E} (\varphi(a) - \varphi(a'))^2\, Q_s^n(da, da')\, dk_s\right] \\ &= E\left[\int_0^t \int_{E\times E} (\varphi(a) - \varphi(a'))^2\, m^n(da, da', ds)\right] \end{aligned}$$

by definition of the predictable projection of m^n.

According to Lemma 1.4.1, this tends to 0 when n tend to infinity, using the boundedness of φ and Lebesgue's dominated convergence theorem.

The generalization of this result to predictable function ϕ continuous in the $E-$variable is obtained thanks to the following result: it is proved in [40] that the weak topology on $\mathcal{R}$ (space of Radon measures on $E \times [0,T]$ whose projection on $[0,T]$ is the Lebesgue measure) is the same as the stable topology, i.e. the topology where the convergence is required for measurable bounded functions continuous in the $E-$variable. ■

Remark 1.4.2 **a)** *Moreover, by Doob's inequality, we obtain that*

$$E\left[\sup_{t\leq T} (M_t^n(\varphi) - M_t(\varphi))^2\right]$$

tends to 0 *when* n *tends to infinity.*

b) *The above construction implies in particular that*

$$\forall n \in \mathbb{N}, \forall t \in [0,T]\,, \quad M_t^n(E) = M_t(E)\,.$$

1.5 Approximation by the stochastic integral of a Brownian motion

An amusing application of the next theorem is the following: if E is a compact set, each continuous martingale measure can be obtained as a limit in $L^2(\Omega)$ of sequence of time-changed stochastic integrals with respect to a single Brownian motion.

Theorem 1.5.1 *We assume that the Lusin space E is a compact set. Let M be a continuous orthogonal martingale measure with intensity $q_t(da)dt$ on $E \times [0,1]$.*
Then, there exists a sequence of predictable $E-$valued processes $\left(u^k(s)\right)_{k\in\mathbb{N}}$ and a Brownian motion W defined on an extension of the probability space Ω, such that
$\forall t \in [0,1], \forall \phi$ *a continuous bounded function from E to $\mathbb{R}$,*

$$\lim_{k\to+\infty} E\left[\left(M_t(\phi) - \int_0^t \phi\left(u^k(s)\right) dW_s\right)^2\right] = 0.$$

This theorem derives from Theorem 1.4.1 and a fundamental approximation lemma obtained first for deterministic measures and then generalized for random measures [27], [24], known under the name of chattering lemma.
A similar result for more general intensity $q_t(da)dk_t$ (k is continuous) can be deduced by time change.

Lemma 1.5.1 *(**Chattering lemma**) Let (q_t) be a predictable process, with values in the space of probability measure on E. Then there exists a sequence of predictable processes $\left(u^k(t)\right)_{k\in\mathbb{N}}$ with values in E such that the sequence of random measures $\left(\delta_{u_t^k}(da)\, dt\right)$ converge weakly to $q_t(da)\, dt, P-a.s$, where k tends to $+\infty$.*

Proof. We will prove here the Theorem 1.5.1.
Let M be an orthogonal continuous martingale measure with intensity $q_t(da)\, dt$, defined on $E \times [0,1]$, where E is a compact set. We have see, in chattering lemma that the

random measure $q_t(da)\,dt$ defined on $\Omega \times E \times [0,1]$ is approximated by a sequence of atomic measures of the form $\delta_{u_t^k}(da)\,dt$, for the weak topology on the space of measures on $E \times [0,1]$, and this being true for almost all w of Ω.

By Theorem 1.4.1 there exists a sequence of orthogonal continuous martingale measure M^k with intensity $\delta_{u_t^k}(da)\,dt$ which converges to M.

The martingale measure M^k can be represented as stochastic integrals with respect to the same Brownian motion. Indeed, we have in Proposition 1.2.1, that for each A of E,

$$M_t^k(A) = \int_0^t \mathbf{1}_A\left(u^k\right) dm_s^k,$$

where m^k is the continuous martingale defined by $m_t^k = M_t^k(E)$.

But we have noted in Remark 1.4.2.b), that for each k of $\mathbb{N}$, $M_t^k(E) = M_t(E)$.

$M_.(E)$ is a continuous $\mathcal{F}_t-$martingale with quadratic variation process t, thus it is a $\mathcal{F}_t$ Brownian motion (independent of k) that we shall denote by W.

We have finally obtained that for each continuous bounded function φ

$$\forall t \in [0,1],\, M_t^k(\varphi) = \int_0^t \varphi\left(u_s^k\right) dW_s,$$

and that

$$\lim_{k\to+\infty} E\left[\left(M_t(\varphi) - \int_0^t \varphi\left(u_s^k\right) dW_s\right)^2\right] = 0.$$

■

Remark 1.5.1 *Since, for each function φ, $M_.(\varphi)$ is a continuous martingale with increasing process*

$$\int_0^t \left(\int_E \varphi^2(a)\, q_s(da)\right) ds,$$

$M_.(\varphi)$ can be represented as a stochastic integral with respect to Brownian motion N^φ in the form

$$M_t(\varphi) = \int_0^t V_s^\varphi dN_s^\varphi, \text{ où } V_s^\varphi = \left(\int_E \varphi^2(a)\, q_s(da)\right)^{1/2}.$$

It is clear that V^{φ} *is not linear in* φ *and the Brownian motion* N^{φ} *depends on* φ*.*

The interest of Theorem 1.5.1 is to give an approximation of $M_t(\varphi)$ in $L^2(\Omega)$ by stochastic integrals with respect to a "canonical" Brownian motion, that is not depending on the function φ.

Remark 1.5.2 *In the case where the time martingale of the intensity of the martingale measure is not the Lebesgue measure, but a random measure* dk_t*,* k *being continuous and increasing, a similar result can be obtained thanks to time change. Let us consider the inverse of* k_t

$$\tau_t = \inf\{s > 0, k_s \geq t\},$$

τ_t *is continuous and increasing from* $[0,1]$ *to* $[0,1]$*.*
The random function N *defined on* $\Omega \times \mathcal{E} \times [0,1]$ *by*

$$N_t(A) = M_{\tau_t}(A)$$

is then a $\mathcal{F}_t-$*martingale measure with intensity* $q_{\tau_t}(da)\,dt$*.*

According to Theorem 1.5.1, one can define a $\mathcal{F}_{\tau_t}-$Brownian motion $\tilde{W}$ and a sequence of $E-$valued predictable processes $\left(\tilde{u}^k\right)$ (for the filtration $\mathcal{F}_{\tau_t}$) such that for each event A of $\mathcal{E}$, the sequence $\int_0^t \mathbf{1}_A\left(\tilde{u}_s^k\right) d\tilde{W}_s$ converges (for each t) in $L^2(\Omega)$ into $N_t(A)$.

Remark 1.5.3 *We deduce from it that for each continuous bounded function ,*

$$\lim_{k\to+\infty} E\left[\left(M_t(\varphi) - \int_0^{k_t} \varphi\left(\tilde{u}_s^k\right) d\tilde{W}_s\right)^2\right] = 0.$$

Chapter 2

A general stochastic maximum principle for control problems

In this chapter we study the following type of stochastic optimal control problem. Minimize a cost function

$$J(u(.)) = E\int_0^T f(t,x(t),u(t))\,dt + h(x(T))$$

subject to

$$\begin{cases} dx(t) = b(t,x(t),u(t))\,dt + \sigma(t,x(t),u(t))\,dW(t) \\ x(0) = x_0, \end{cases}$$

in the above, $u(t)$ is the control variable valued in a subset of $\mathbb{R}^k$, $x(t)$ is the state variable, $W(t)$ is an $m-$dimensional standard Brownian motion, and f, h, b, σ are given maps. The object is to obtain a necessary condition, called the maximum principle, for optimal control.

There are many works concerning this subject (see [15], [37], [38], [44]). A difficulty is treating the case where the diffusion coefficient σ contains the control variable u. Bensoussan [14], [15] studied such a case. The maximum principle he obtained is of local condition, and his method depends heavily on the control being convex. In this problem, since the control domain is not necessarily convex, we must obtain the maximum principle in its global form. A classical way of treating such a problem is to use the "spike variation method" [57].

2.1 Statement of the Stochastic Maximum Principle

First recall the strong formulation of the stochastic optimal control problem, then introduce some assumptions.

Let $\left(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P\right)$ be a given filtered probability space satisfying the usual conditions, on which an $m-$dimensional standard Brownian motion $W(t)$ (with $W(0) = 0$) is given. We consider the following stochastic controlled system

$$\begin{cases} dx(t) = b(t, x(t), u(t))\, dt + \sigma(t, x(t), u(t))\, dW(t) \\ x(0) = x_0, \end{cases}, t \in [0, T], \tag{2.1}$$

with the cost functional

$$J(u(.)) = E\left[\int_0^T f(t, x(t), u(t))\, dt + h(x(T))\right]. \tag{2.2}$$

In the above, $b : [0, T] \times \mathbb{R}^n \times U \to \mathbb{R}^n$, $\sigma : [0, T] \times \mathbb{R}^n \times U \to \mathbb{R}^{n\times m}$, $f : [0, T] \times \mathbb{R}^n \times U \to \mathbb{R}$, and $h : \mathbb{R}^n \to \mathbb{R}$. We define

$$\begin{cases} b(t,x,u) = \begin{pmatrix} b^1(t,x,u) \\ \vdots \\ b^n(t,x,u) \end{pmatrix}, \\ \sigma(t,x,u) = \left(\sigma^1(t,x,u), \ldots, \sigma^n(t,x,u)\right), \\ \sigma^j(t,x,u) = \begin{pmatrix} \sigma^{1j}(t,x,u) \\ \vdots \\ \sigma^{nj}(t,x,u) \end{pmatrix}, 1 \leq j \leq m. \end{cases} \tag{2.3}$$

Let us make the following assumptions

(S$_0$) $(\mathcal{F}_t)_{t\geq 0}$ is the natural filtration generated by $W(t)$, augmented by all the $P-$null sets in $\mathcal{F}$.

(S$_1$) (U,d) is a separable metric space and $T > 0$.

(S$_2$) The maps b, σ, f and h are measurable, and there exist a constant $L > 0$ and a modulus of continuity $\bar{w} : [0,\infty) \to [0,\infty)$ such that for $\varphi(t,x,u) = b(t,x,u)$, $\sigma(t,x,u)$, $f(t,x,u)$, $h(x)$, we have

$$\begin{cases} |\varphi(t,x,u) - \varphi(t,\hat{x},\hat{u})| \leq L|x-\hat{x}| + \bar{w}(d(u,\hat{u})), \\ \qquad \forall t \in [0,T], \quad x,\hat{x} \in \mathbb{R}^n, \quad u,\hat{u} \in U \\ |\varphi(t,0,u)| \leq L, \quad \forall (t,u) \in [0,T] \times U. \end{cases} \tag{2.4}$$

(S$_3$) The maps b, σ, f and h are C^2 in x. Moreover, there exists a constant $L > 0$ and a modulus of continuity $\bar{w} : [0,\infty) \to [0,\infty)$ such that for $\varphi = b, \sigma, f, h$, we have

$$\begin{cases} |\varphi_x(t,x,u) - \varphi_x(t,\hat{x},\hat{u})| \leq L|x-\hat{x}| + \bar{w}(d(u,\hat{u})), \\ |\varphi_{xx}(t,x,u) - \varphi_{xx}(t,\hat{x},\hat{u})| \leq \bar{w}(|x-\hat{x}| + d(u,\hat{u})), \\ \qquad \forall t \in [0,T], \quad x,\hat{x} \in \mathbb{R}^n, \quad u,\hat{u} \in U. \end{cases} \tag{2.5}$$

Assumption ($\mathbf{S}_0$) signifies that the system noise is the only source of uncertainty in the problem, and the past information about the noise is available to the controller. This assumption will be crucial below.
Now we define

$$\mathcal{U}[0,T] = \left\{ u : [0,T] \times \Omega \to U \,\middle|\, u \text{ is } (\mathcal{F}_t)_{t\geq 0} - \text{adapted} \right\}. \tag{2.6}$$

Given $u(.) \in \mathcal{U}[0,T]$, equation (2.1) is an SDE with random coefficients. From Yong-Zhou [60] (Chapter1, Section 6.4), they find that under ($\mathbf{S}_0$)-($\mathbf{S}_2$), for any $u(.) \in \mathcal{U}[0,T]$, the state equation (2.1) admits a unique solution $x(.) = x(., u(.))$ and the cost functional (2.2) is well-defined. In the case that $x(.)$ is the solution of (2.1) corresponding to $u(.) \in \mathcal{U}[0,T]$, we call $(x(.), u(.))$ an admissible pair, and $x(.)$ an admissible state process (trajectory). Our optimal control problem can be stated as follows.

Problem 2.1.1 *Minimize (2.2) over* $\mathcal{U}[0,T]$.

Any $\bar{u}(.) \in \mathcal{U}[0,T]$ satisfying

$$J(\bar{u}(.)) = \inf_{u(.)\in\mathcal{U}[0,T]} J(u(.)) \tag{2.7}$$

is called an optimal control. The corresponding $\bar{x}(.) = x(., \bar{u}(.))$ and $(\bar{x}(.), \bar{u}(.))$ are called an optimal state process/trajectory and optimal pair, respectively.
Notice that the strong formulation is adopted here for the optimal control problem see Yong-Zhou [60] (chapter 2, section 4.1 for more details). The next goal is to derive a set of necessary conditions for stochastic optimal controls, similar to the maximum principle for the deterministic case. To this end, we need some preparations.

2.1.1 Adjoint equations

In this subsection we will give adjoint equations involved in a stochastic maximum principle and the associated stochastic Hamiltonian system.

Recall that $\mathcal{S}^n = \left\{ A \in \mathbb{R}^{n\times n} \,\middle|\, A^T = A \right\}$ and $(x(.), u(.))$ be a given optimal pair.

We knew that in the deterministic case the adjoint variable $p(.)$ plays a central role in the maximum principle. The adjoint equation that $p(.)$ satisfies is a backward ordinary differential equation (meaning that the terminal value is specified). It is nevertheless equivalent to a forward equation if we reverse the time. In the stochastic case, however, one cannot simply reverse the time, as it may destroy the non anticipativeness of the solutions. Instead, we introduce the following terminal value problem for a stochastic differential equation

$$\begin{cases} dp(t) = -\Bigg[b_x(t, \bar{x}(t), \bar{u}(t))^* p(t) + \displaystyle\sum_{j=1}^{m} \sigma_x^j(t, \bar{x}(t), \bar{u}(t))^* q_j(t) \\ \qquad\qquad -f_x(t, \bar{x}(t), \bar{u}(t)) \Big] dt + q(t)\, dW(t), \quad t \in [0, T] \\ p(T) = -h_x(\bar{x}(T)). \end{cases} \tag{2.8}$$

Here the unknown is a pair of $(\mathcal{F}_t)_{t\geq 0}$ −adapted processes $(p(.), q(.))$. We call the above a backward stochastic differential equation (BSDE, for short). The key issue here is that the equation is to be solved backwards (since the terminal value is given); however, the solution $(p(.), q(.))$ is required to be $(\mathcal{F}_t)_{t\geq 0}$ −adapted. Any pair of processes $(p(.), q(.)) \in L^2_{\mathcal{F}}(0, T; \mathbb{R}^n) \times \left(L^2_{\mathcal{F}}(0, T; \mathbb{R}^n)\right)^m$ satisfying (2.8) is called an adapted solution of (2.8).

The adjoint variable $p(.)$ in the deterministic case corresponds to the so-called shadow price or the marginal value of the resource represented by the state variable in economic theory. The maximum principle is nothing but the so-called duality principle: Minimizing the total cost amounts to maximizing the total contribution of the marginal value. Nevertheless, in the stochastic situation, the controller has to balance carefully the scale of control and the degree of uncertainty if a control made is going to affect the volatility of the system (i.e., if the diffusion coefficient depends on the control variable). Therefore, the marginal value alone may not be able to fully characterize the trade-off between the cost and control gain in an uncertain environment. One has to introduce another variable to reflect the uncertainty or the risk factor in the system. This is done by introducing an additional adjoint equation as follows

$$\begin{cases} dP(t) = -\Big[b_x(t,\bar{x}(t),\bar{u}(t))^* P(t) + P(t)\, b_x(t,\bar{x}(t),\bar{u}(t)) \\ \qquad + \sum_{j=1}^{m} \sigma_x^j(t,\bar{x}(t),\bar{u}(t))^* P(t)\, \sigma_x^j(t,\bar{x}(t),\bar{u}(t)) \\ \qquad + \sum_{j=1}^{m} \sigma_x^j(t,\bar{x}(t),\bar{u}(t))^* Q_j(t) + Q_j(t)\, \sigma_x^j(t,\bar{x}(t),\bar{u}(t)) \\ \qquad + H_{xx}(t,\bar{x}(t),\bar{u}(t),p(t),q(t))\Big] dt + \sum_{j=1}^{m} Q_j(t)\, dW^j(t), \\ P(T) = -h_{xx}(\bar{x}(T)), \end{cases} \tag{2.9}$$

where the Hamiltonian H is defined by

$$H(t,x,u,p,q) = \langle p, b(t,x,u)\rangle + tr\left[q^T \sigma(t,x,u)\right] - f(t,x,u), \tag{2.10}$$
$$(t,x,u,p,q) \in [0,T] \times \mathbb{R}^n \times U \times \mathbb{R}^n \times \mathbb{R}^{n\times m},$$

and $(p(.), q(.))$ is the solution to (2.8). In the above (2.9), the unknown is again a pair of processes $(P(.), Q(.)) \in L^2_{\mathcal{F}}(0,T;\mathcal{S}^n) \times \left(L^2_{\mathcal{F}}(0,T;\mathcal{S}^n)\right)^m$.

Incidentally, the Hamiltonian $H(t,x,u,p,q)$ defined by (2.10) coincides with $H(t,x,u,p)$ defined by (2.8) when $\sigma = 0$.

Note that equation (2.9) is also a BSDE with matrix-valued unknowns. As with (2.8), under assumptions $(\mathbf{S}_0)$-$(\mathbf{S}_3)$, there exists a unique adapted solution $(P(.), Q(.))$ to (2.9). We refer to (2.8) (resp. (2.9)) as the first-order (resp. second-order) adjoint equations, and to $p(.)$ (resp. $P(.)$) as the first-order (resp. second-order) adjoint process. In what follows, if $(\bar{x}(.); \bar{u}(.))$ is an optimal (resp. admissible) pair, and $(p(.), q(.))$ and $(P(.), Q(.))$ are adapted solutions of (2.8) and (2.9), respectively. then $(\bar{x}(.), \bar{u}(.), p(.), q(.), P(.), Q(.))$ is called an optimal 6-tuple (resp. admissible 6-tuple).

2.1.2 Maximum principle and stochastic Hamiltonian systems

Now we are going to state the Pontryagin-type maximum principle for optimal stochastic controls. At first glance, it might be quite natural for one to expect that a stochastic maximum principle should maximize the Hamiltonian H defined by (2.10). Unfortunately, this is not true if the diffusion coefficient σ depends on the control.

Next, associated with an optimal 6-tuple $(\bar{x}(.), \bar{u}(.), p(.), q(.), P(.), Q(.))$, we define an $\mathcal{H}$-function

$$\begin{aligned}
\mathcal{H}(t,x,u) &= H(t,x,u,p(t),q(t)) - \tfrac{1}{2}tr\left[\sigma(t,\bar{x}(t),\bar{u}(t))^{*} P(t)\sigma(t,\bar{x}(t),\bar{u}(t))\right] \\
&\quad + \tfrac{1}{2}tr\left\{\left[\sigma(t,x,u)-\sigma(t,\bar{x}(t),\bar{u}(t))\right]^{*} P(t)\left[\sigma(t,x,u)-\sigma(t,\bar{x}(t),\bar{u}(t))\right]\right\} \\
&= \tfrac{1}{2}tr\left[\sigma(t,x,u)^{*} P(t)\sigma(t,x,u)\right] + \langle p, b(t,x,u)\rangle - f(t,x,u) \\
&\quad + tr\left[q(t)^{*}\sigma(t,x,u)\right] - tr\left[\sigma(t,x,u)^{*} P(t)\sigma(t,\bar{x}(t),\bar{u}(t))\right] \\
&= G(t,x,u,p(t),P(t)) + tr\left\{\sigma(t,x,u)^{*}\left[q(t)-P(t)\sigma(t,\bar{x}(t),\bar{u}(t))\right]\right\}.
\end{aligned} \tag{2.11}$$

Notice that an $\mathcal{H}$-function may be defined similarly associated with any admissible 6-tuple $(x(.), u(.), p(.), q(.), P(.), Q(.))$.

Theorem 2.1.1 *(Stochastic Maximum Principle) Let* ($\boldsymbol{S}_0$)-($\boldsymbol{S}_3$) *hold. Let* $(\bar{x}(.), \bar{u}(.))$ *be an optimal pair of Problem* 2.1.1. *Then there are pairs of processes*

$$\begin{cases} (p(.), q(.)) \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^n) \times \left(L^2_{\mathcal{F}}(0,T;\mathbb{R}^n)\right)^m \\ (P(.), Q(.)) \in L^2_{\mathcal{F}}(0,T;\mathcal{S}^n) \times \left(L^2_{\mathcal{F}}(0,T;\mathcal{S}^n)\right)^m \end{cases} \tag{2.12}$$

where

$$\begin{cases} q(.) = (q_1(.), ..., q_m(.)), \quad Q(.) = (Q_1(.), ..., Q_m(.)), \\ q_j(.) \in \left(L^2_{\mathcal{F}}(0,T;\mathbb{R}^n)\right), \quad Q_j(.) \in L^2_{\mathcal{F}}(0,T;\mathcal{S}^n), 1 \le j \le m, \end{cases} \tag{2.13}$$

satisfying the first-order and second-order adjoint equations (2.8) and (2.9), respectively, such that

$$\begin{gathered} H\left(t,\bar{x}\left(t\right),\bar{u}\left(t\right),p\left(t\right),q\left(t\right)\right)-H\left(t,\bar{x}\left(t\right),u,p\left(t\right),q\left(t\right)\right)\\ -\frac{1}{2}tr\left(\left[\sigma\left(t,\bar{x}\left(t\right),\bar{u}\left(t\right)\right)-\sigma\left(t,\bar{x}\left(t\right),u\right)\right]^{*}P\left(t\right)\right.\\ \left.\left[\sigma\left(t,\bar{x}\left(t\right),\bar{u}\left(t\right)\right)-\sigma\left(t,\bar{x}\left(t\right),u\right)\right]\right)\geq 0,\\ \forall u\in U,\quad a.e.\ t\in\left[0,T\right],\quad P-a.s., \end{gathered} \tag{2.14}$$

or, equivalently,

$$\mathcal{H}\left(t,\bar{x}\left(t\right),\bar{u}\left(t\right)\right)=\max_{u\in U}\mathcal{H}\left(t,\bar{x}\left(t\right),u\right),\quad a.e.\ t\in\left[0,T\right],\quad P-a.s. \tag{2.15}$$

The inequality (2.14) is called the variational inequality, and (2.15) is called the maximum condition. Note that the third term on the left-hand side of (2.14) reflects the risk adjustment, which must be present when σ depends on u.

Let us single out two important special cases.

Case 1. The diffusion does not contain the control variable, i.e.

$$\sigma\left(t,x,u\right)=\sigma\left(t,x\right),\quad\forall\left(t,x,u\right)\in\left[0,T\right]\times\mathbb{R}^{n}\times U \tag{2.16}$$

In this case, the maximum condition (2.15) reduces to

$$H\left(t,\bar{x}\left(t\right),\bar{u}\left(t\right),p\left(t\right),q\left(t\right)\right)=\max_{u\in U}H\left(t,\bar{x}\left(t\right),u,p\left(t\right),q\left(t\right)\right),\quad\text{a.e. } t\in\left[0,T\right],\quad P-a.s. \tag{2.17}$$

which is parallel to the deterministic case (no risk adjustment is required). We note that in this case, equation (2.9) for $(P(.), Q(.))$ is not needed. Thus, the twice differentiability of the functions b, σ, f and h in x is not necessary here.

Case 2. The control domain $U \subset \mathbb{R}^k$ is convex and all the coefficients are C^1 in u. Then

(2.14) implies

$$\langle H_u(t,\bar{x}(t),\bar{u}(t),p(t),q(t)),u-\bar{u}(t)\rangle \leq 0, \forall u\in U, \quad a.e.\ t\in[0,T], \quad P-a.s. \tag{2.18}$$

This is called a local form (in contrast to the global form (2.14) or (2.15)) of the maximum principle. Note that the local form does not involve the second-order adjoint process $P(.)$ either.

Analogous to the deterministic case, the system (2.1) along with its first-order adjoint system can be written as follows

$$\begin{cases} dx(t) = H_p(t,x(t),u(t),p(t),q(t))\,dt + H_q(t,x(t),u(t),p(t),q(t))\,dW(t), \\ dp(t) = -H_x(t,x(t),u(t),p(t),q(t))\,dt + q(t)\,dW(t), & t\in[0,T], \\ x(0)=x_{0,} \qquad P(T) = -h_x(x(T)). \end{cases} \tag{2.19}$$

The combination of (2.19), (2.9), and (2.14) (or (2.15)) is called an (extended) stochastic Hamiltonian system, with its solution being a 6-tuple $(x(.),u(.),p(.),q(.),P(.),Q(.))$. Therefore, we can rephrase Theorem 2.1.1 as the following.

Theorem 2.1.2 *Let* ($\boldsymbol{S}_0$)-($\boldsymbol{S}_3$) *hold. Let the precedent Problem* 2.1.1 *admit an optimal pair* $(\bar{x}(.),\bar{u}(.))$*. Then the optimal 6-tuple* $(\bar{x}(.),\bar{u}(.),p(.),q(.),P(.),Q(.))$ *of the precedent Problem* 2.1.1 *solves the stochastic Hamiltonian system (2.19), (2.9), and (2.14) (or (2.15)).*

It is seen from the above result that optimal control theory can be used to solve stochastic Hamiltonian systems. System (2.19) (with $u(.)$ given) is also called a forward-backward stochastic differential equation (FBSDE, for short).

2.2 Proof of the Maximum Principle

In this section we are going to give a proof of the stochastic maximum principle, Theorem 2.1.1. The idea is still the variational technique. However, due to the presence of the diffusion coefficient, which may contain the control variable, the method that works for deterministic case has to be substantially modified to fit the stochastic case.

2.2.1 Moment estimate

In this subsection we prove an elementary lemma, which will be useful in the sequel.

Lemma 2.2.1 *Let $Y(t) \in L^2_{\mathcal{F}}(0,T;\mathbb{R}^n)$ be the solution of the following*

$$\begin{cases} dY(t) = [A(t)Y(t) + \alpha(t)]\,dt + \sum_{j=1}^{m} [B^j(t)Y(t) + \beta^j(t)]\,dW^j(t) \\ Y(0) = Y_0 \end{cases} \tag{2.20}$$

where $A, B^j : [0,T] \times \Omega \to \mathbb{R}^{n\times m}$ and $\alpha, \beta^j : [0,T]\times\Omega \to \mathbb{R}^n$ are $(\mathcal{F}_t)_{t\geq 0}$adapted, and

$$\begin{cases} |A(t)|, |B^j(t)| \leq L, \quad a.e. \ \ t \in [0,T], P-a.s., 1 \leq j \leq m, \\ \int_0^T \left[E|\alpha(s)|^{2k}\right]^{\frac{1}{2k}} ds + \int_0^T \left[E|\beta^j(s)|^{2k}\right]^{\frac{1}{2k}} ds < \infty, 1 \leq j \leq m, \end{cases} \tag{2.21}$$

for some $k \geq 1$. Then

$$\sup_{t\in[0,T]} E|Y(t)|^{2k} \leq K\left[E|Y_0|^{2k} + \left(\int_0^T \left[E|\alpha(s)|^{2k}\right]^{\frac{1}{2k}} ds\right)^{2k} + \sum_{j=1}^{m}\left(\int_0^T \left[E\left|\beta^j(s)\right|^{2k}\right]^{\frac{1}{k}} ds\right)^{k}\right]. \tag{2.22}$$

Proof. For notational simplicity, we set the prove only in the case $m = 1$ (i.e., the Brownian motion $W(t)$ is one-dimensional). Thus, the index j in $B^j(.)$ and $\beta^j(.)$ will be dropped. We first assume that $\alpha(.)$ and $\beta(.)$ are bounded. Let $\epsilon > 0$ and define

$$\langle Y \rangle_\epsilon = \sqrt{|Y|^2 + \epsilon^2}, \qquad \forall Y \in \mathbb{R}^n. \tag{2.23}$$

Note that for any $\epsilon > 0$, the map $Y \to \langle Y \rangle_\epsilon$ is smooth and $\langle Y \rangle_\epsilon \to |Y|$ as $\epsilon \to 0$. The purpose of using such a function is to avoid some difficulties that might be encountered in differentiating functions like $|Y|^{2k}$ for noninteger k. Applying Itô's formula to $\langle Y \rangle_\epsilon$, we have

$$\begin{aligned} E\left\langle Y\left(t\right)\right\rangle_\epsilon^{2k} &\leq E\left\langle Y\left(0\right)\right\rangle_\epsilon^{2k} + 2kE\int_0^t \left\langle Y\left(s\right)\right\rangle_\epsilon^{2k-1}\left[\left|A\left(s\right)\right|\left\langle Y\left(s\right)\right\rangle_\epsilon + \left|\alpha\left(s\right)\right|\right]ds \\ &+k\left(2k-1\right)E\int_0^t \left\langle Y\left(s\right)\right\rangle_\epsilon^{2k-2}\left[\left|B\left(s\right)\right|\left\langle Y\left(s\right)\right\rangle_\epsilon + \left|\beta\left(s\right)\right|\right]^2 ds \\ &\leq E\left\langle Y\left(0\right)\right\rangle_\epsilon^{2k} + K_0 E\int_0^t \Big[\left\langle Y\left(s\right)\right\rangle_\epsilon^{2k} + \left|\alpha\left(s\right)\right|\left\langle Y\left(s\right)\right\rangle_\epsilon^{2k-1} \\ &+\left|\beta\left(s\right)\right|^2\left\langle Y\left(s\right)\right\rangle_\epsilon^{2k-2}\Big]ds. \end{aligned} \tag{2.24}$$

Here $K_0 = K_0\left(k, L\right)$ is independent of t. Applying Young's inequality, we obtain

$$E\left\langle Y\left(t\right)\right\rangle_\epsilon^{2k} \leq E\left\langle Y\left(0\right)\right\rangle_\epsilon^{2k} + KE\int_0^t \left[\left\langle Y\left(s\right)\right\rangle_\epsilon^{2k} + \left|\alpha\left(s\right)\right|^{2k} + \left|\beta\left(s\right)\right|^{2k}\right]ds. \tag{2.25}$$

Hence, it follows from Gronwall's inequality that

$$\begin{gathered} E\left\langle Y\left(t\right)\right\rangle_\epsilon^{2k} \leq K\left(E\left\langle Y\left(0\right)\right\rangle_\epsilon^{2k} + E\int_0^T \left[\left|\alpha\left(s\right)\right|^{2k} + \left|\beta\left(s\right)\right|^{2k}\right]ds\right), \\ t \in \left[0, T\right]. \end{gathered} \tag{2.26}$$

Here $K = K\left(L, k, T\right)$. Note that since we assume for the time being that $\alpha\left(.\right)$ and $\beta\left(.\right)$ are bounded, the above procedure goes through (otherwise the integration on the right-hand side of (2.25) may not exist; see (2.21)). Next, we want to refine the above estimate so that (2.22) will follow. To this end, note that (2.25) implies that its left-hand side is bounded uniformly in $t \in [0, T]$. Thus, we may set

$$\varphi\left(t\right) = \left[\sup_{0\leq s\leq t} E\left\langle Y\left(s\right)\right\rangle_\epsilon^{2k}\right]^{\frac{1}{2k}}, \quad t \in \left[0, T\right]. \tag{2.27}$$

We now return to (2.24), using (2.27). Define $\delta = (4K_0)^{-1}$. Then, for any $t \in [0, \delta]$, applying Holder's inequality and Young's inequality, we obtain

$$\begin{aligned} \varphi(t)^{2k} &\le \varphi(0)^{2k} + K_0 \left[\varphi(t)^{2k} t + \varphi(t)^{2k-1} \int_0^t \left(E|\alpha(s)|^{2k} \right)^{\frac{1}{2k}} ds + \varphi(t)^{2k-2} \int_0^t \left(E|\beta(s)|^{2k} \right)^{\frac{1}{k}} ds \right] \\ &\le \varphi(0)^{2k} + \tfrac{1}{2}\varphi(t)^{2k} + K_1 \left[\left\{ \int_0^t \left(E|\alpha(s)|^{2k} \right)^{\frac{1}{2k}} ds \right\}^{2k} + \left\{ \int_0^t \left(E|\beta(s)|^{2k} \right)^{\frac{1}{k}} ds \right\}^{k} \right]. \end{aligned} \tag{2.28}$$

The constant $K_1 = K_1(k, L, \delta)$ in (2.28) is independent of t. From (2.28), we obtain

$$\varphi(t)^{2k} \le 2\varphi(0)^{2k} + 2K_1 \left\{ \left(\int_0^t \left[E|\alpha(s)|^{2k} \right]^{\frac{1}{2k}} ds \right)^{2k} + \left(\int_0^t \left[E|\beta(s)|^{2k} \right]^{\frac{1}{k}} ds \right)^{k} \right\}, \quad \forall t \in [0, \delta]. \tag{2.29}$$

Now we can do the same thing on $[\delta, 2\delta]$ and on $[2\delta, 3\delta]$, and so on. Finally, we end up with

$$\varphi(T)^{2k} \le K \left\{ \varphi(0)^{2k} + \left(\int_0^T \left[E|\alpha(s)|^{2k} \right]^{\frac{1}{2k}} ds \right)^{2k} + \left(\int_0^T \left[E|\beta(s)|^{2k} \right]^{\frac{1}{k}} ds \right)^{k} \right\}, \tag{2.30}$$

with the constant $K = K(L, k, T, \delta)$. By (2.27), the definition of $\varphi(t)$, we conclude that

$$\sup_{t\in[0,T]} E\langle Y(s)\rangle_\epsilon^{2k} \le K \left\{ \langle Y(0)\rangle_\epsilon^{2k} + \left(\int_0^T \left[E|\alpha(s)|^{2k} \right]^{\frac{1}{2k}} ds \right)^{2k} + \left(\int_0^T \left[E|\beta(s)|^{2k} \right]^{\frac{1}{k}} ds \right)^{k} \right\}. \tag{2.31}$$

Letting $\epsilon \to 0$, we obtain (2.22). Finally, in the case that we only have (2.21) (instead of α and β being bounded), we can use the usual approximation. ■

2.2.2 Taylor expansions

The following elementary lemma will be used below.

Lemma 2.2.2 *Let $g \in C^2(\mathbb{R}^n)$. Then, for any $x, \bar{x} \in \mathbb{R}^n$,*

$$g(x) = g(\bar{x}) + \langle g_x(\bar{x}), x - \bar{x}\rangle + \int_0^1 \langle \theta g_{xx}(\theta\bar{x} + (1-\theta)x)(x - \bar{x}), x - \bar{x}\rangle \, d\theta. \tag{2.32}$$

Now, let $(\bar{x}(.), \bar{u}(.))$ be the given optimal pair. Then the following is satisfied

$$\begin{cases} d\bar{x}(t) = b(t, \bar{x}(t), \bar{u}(t))\,dt + \sigma(t, \bar{x}(t), \bar{u}(t))\,dW(t) \\ \bar{x}(0) = x_0, \end{cases}, t \in [0, T]. \tag{2.33}$$

Fix any $\bar{u}(.) \in \mathcal{U}[0, T]$ and $\epsilon > 0$. Define

$$u^{\epsilon}(t) = \begin{cases} \bar{u}(t), & t \in [0, T] \setminus E_{\epsilon}, \\ u(t), & t \in E_{\epsilon}, \end{cases} \tag{2.34}$$

where $E_{\epsilon} \subseteq [0, T]$ is a measurable set with $|E_{\epsilon}| = \epsilon$. Let $(x^{\epsilon}(.), u^{\epsilon}(.))$ satisfy the following

$$\begin{cases} dx^{\epsilon}(t) = b(t, x^{\epsilon}(t), u^{\epsilon}(t))\,dt + \sigma(t, x^{\epsilon}(t), u^{\epsilon}(t))\,dW(t) \\ x^{\epsilon}(0) = x_0, \qquad t \in [0, T]. \end{cases} \tag{2.35}$$

Next, for $\varphi = b^i, \sigma^{ij}, f\ (1 \leq i \leq n, 1 \leq j \leq m)$, we define

$$\begin{cases} \varphi_x(t) = \varphi_x(t, \bar{x}(t), \bar{u}(t)), \qquad \varphi_{xx}(t) = \varphi_{xx}(t, \bar{x}(t), \bar{u}(t)) \\ \delta\varphi(t) = \varphi(t, \bar{x}(t), u(t)) - \varphi(t, \bar{x}(t), \bar{u}(t)), \\ \delta\varphi_x(t) = \varphi_x(t, \bar{x}(t), u(t)) - \varphi_x(t, \bar{x}(t), \bar{u}(t)), \\ \delta\varphi_{xx}(t) = \varphi_{xx}(t, \bar{x}(t), u(t)) - \varphi_{xx}(t, \bar{x}(t), \bar{u}(t)). \end{cases} \tag{2.36}$$

Let $y^{\epsilon}(t)$ and $z^{\epsilon}(t)$ be respectively the solution of the following stochastic differential equations

$$\begin{cases} dy^{\epsilon}(t) = b_x(t)\,y^{\epsilon}(t)\,dt + \sum_{j=1}^{m} \left[\sigma_x^j(t)\,y^{\epsilon}(t) + \delta\sigma^j(t)\,\mathbf{1}_{E_{\varepsilon}}(t)\right] dW^j(t), \\ y^{\epsilon}(0) = 0, \qquad t \in [0, T], \end{cases} \tag{2.37}$$

and

$$\begin{cases} dz^{\epsilon}(t) = \left[b_x(t)\, z^{\epsilon}(t) + \delta b(t)\, \mathbf{1}_{E_\varepsilon}(t) + \frac{1}{2} b_{xx}(t)\, y^{\epsilon}(t)^2\right] dt \\ \qquad\qquad + \sum_{j=1}^{m} \left[\sigma_x^j(t)\, z^{\epsilon}(t) + \delta\sigma_x^j(t)\, y^{\epsilon}(t)\, \mathbf{1}_{E_\varepsilon}(t) + \frac{1}{2}\sigma_{xx}^j(t)\, y^{\epsilon}(t)^2\right] dW^j(t), \\ z^{\epsilon}(0) = 0, \qquad\qquad t \in [0,T], \end{cases} \tag{2.38}$$

where

$$\begin{aligned} b_{xx}(t)\, y^{\epsilon}(t)^2 &= \begin{pmatrix} tr\left(b_{xx}^1(t)\, y^{\epsilon}(t)\, y^{\epsilon}(t)^*\right) \\ \vdots \\ tr\left(b_{xx}^n(t)\, y^{\epsilon}(t)\, y^{\epsilon}(t)^*\right) \end{pmatrix}, \\ \sigma_{xx}^j(t)\, y^{\epsilon}(t)^2 &= \begin{pmatrix} tr\left(\sigma_{xx}^{1j}(t)\, y^{\epsilon}(t)\, y^{\epsilon}(t)^*\right) \\ \vdots \\ tr\left(\sigma_{xx}^{nj}(t)\, y^{\epsilon}(t)\, y^{\epsilon}(t)^*\right) \end{pmatrix}, 1 \leq j \leq m. \end{aligned} \tag{2.39}$$

The following result gives the Taylor expansion of the state with respect to the control perturbation.

Theorem 2.2.1 *Let ($\boldsymbol{S}_1$)-($\boldsymbol{S}_3$) hold. Then, for any $k \geq 1$,*

$$\sup_{t\in[0,T]} E\left|x^{\epsilon}(t) - \bar{x}(t)\right|^{2k} = O\left(\epsilon^k\right), \tag{2.40}$$

$$\sup_{t\in[0,T]} E\left|y^{\epsilon}(t)\right|^{2k} = O\left(\epsilon^k\right), \tag{2.41}$$

$$\sup_{t\in[0,T]} E\left|z^{\epsilon}(t)\right|^{2k} = O\left(\epsilon^{2k}\right), \tag{2.42}$$

$$\sup_{t\in[0,T]} E\left|x^{\epsilon}(t) - \bar{x}(t) - y^{\epsilon}(t)\right|^{2k} = O\left(\epsilon^{2k}\right), \tag{2.43}$$

$$\sup_{t\in[0,T]} E\left|x^{\epsilon}(t) - \bar{x}(t) - y^{\epsilon}(t) - z^{\epsilon}(t)\right|^{2k} = \circ\left(\epsilon^{2k}\right). \tag{2.44}$$

Moreover, the following expansion holds for the cost functional

$$\begin{aligned} J\left(u^{\epsilon}\left(.\right)\right) &= J\left(\bar{u}\left(.\right)\right) + E\left\langle h_x\left(\bar{x}\left(T\right)\right), y^{\epsilon}\left(T\right) + z^{\epsilon}\left(T\right)\right\rangle + \tfrac{1}{2}E\left\langle h_{xx}\left(\bar{x}\left(T\right)\right) y^{\epsilon}\left(T\right), y^{\epsilon}\left(T\right)\right\rangle \\ &+E\int_0^T \left[\left\langle f_x\left(t\right), y^{\epsilon}\left(t\right) + z^{\epsilon}\left(t\right)\right\rangle + \tfrac{1}{2}\left\langle f_{xx}\left(t\right) y^{\epsilon}\left(t\right), y^{\epsilon}\left(t\right)\right\rangle + \delta f\left(t\right)\mathbf{1}_{E_{\epsilon}}\left(t\right)\right] dt + \circ\left(\epsilon\right) \end{aligned} \tag{2.45}$$

Proof. The proof of the above theorem is rather lengthy and technical. For simplicity of presentation, we carry out the proof only for the case $n = m = 1$ (thus, the indices i and j will be omitted below).

1. Proof of (2.40) and (2.41) Let $\zeta^{\epsilon}\left(t\right) = x^{\epsilon}\left(t\right) - \bar{x}\left(t\right)$,, then we have

$$\begin{cases} d\zeta^{\epsilon}\left(t\right) = \left[\tilde{b}_x\left(t\right)\zeta^{\epsilon}\left(t\right) + \delta b\left(t\right)\mathbf{1}_{E_{\varepsilon}}\left(t\right)\right] dt + \left[\tilde{\sigma}_x^{\epsilon}\left(t\right)\zeta^{\epsilon}\left(t\right) + \delta\sigma\left(t\right)\mathbf{1}_{E_{\varepsilon}}\left(t\right)\right] dW\left(t\right), \\ \zeta^{\epsilon}\left(0\right) = 0, \qquad t \in \left[0, T\right], \end{cases} \tag{2.46}$$

where

$$\begin{cases} \tilde{b}_x^{\epsilon}\left(t\right) = \displaystyle\int_0^1 b_x\left(t, \bar{x}\left(t\right) + \theta\left(x^{\epsilon}\left(t\right) - \bar{x}\left(t\right)\right), u^{\epsilon}\left(t\right)\right) d\theta, \\ \tilde{\sigma}_x^{\epsilon}\left(t\right) = \displaystyle\int_0^1 \sigma_x\left(t, \bar{x}\left(t\right) + \theta\left(x^{\epsilon}\left(t\right) - \bar{x}\left(t\right)\right), u^{\epsilon}\left(t\right)\right) d\theta. \end{cases} \tag{2.47}$$

By Lemma 2.1.1, we obtain

$$\begin{aligned} \sup_{t\in[0,T]} E\left|\zeta^{\epsilon}\left(t\right)\right|^{2k} &\leq K\left(\int_0^T \left\{E\left|\delta b\left(s\right)\mathbf{1}_{E_{\varepsilon}}\left(s\right)\right|^{2k}\right\}^{\frac{1}{2k}} ds\right)^{2k} \\ &+K\left(\int_0^T \left\{E\left|\delta\sigma\left(s\right)\mathbf{1}_{E_{\varepsilon}}\left(s\right)\right|^{2k}\right\}^{\frac{1}{k}} ds\right)^{k} \\ &\leq K\left(\epsilon^{2k} + \epsilon^k\right) \leq K\epsilon^k. \end{aligned} \tag{2.48}$$

This proves (2.40). Similarly, we can prove (2.41).

2. Proof of (2.42) Form (2.38), Lemma 2.1.1 and (2.41), we have

$$\begin{aligned}\sup_{t\in[0,T]} E\left|z^{\epsilon}(t)\right|^{2k} &\leq K\left(\int_0^T\left\{E\left|\delta b(s)\mathbf{1}_{E_{\varepsilon}}(s)+\tfrac{1}{2}b_{xx}(s)y^{\epsilon}(s)^2\right|^{2k}\right\}^{\frac{1}{2k}}ds\right)^{2k}\\ &\quad+K\left(\int_0^T\left\{E\left|\delta\sigma_x(s)\mathbf{1}_{E_{\varepsilon}}(s)y^{\epsilon}(s)+\tfrac{1}{2}\sigma_{xx}(s)y^{\epsilon}(s)^2\right|^{2k}\right\}^{\frac{1}{k}}ds\right)^{k}\\ &\leq K\left(\int_0^T\left\{\mathbf{1}_{E_{\varepsilon}}(s)+\left(E\left|y^{\epsilon}(s)\right|^{4k}\right)^{\frac{1}{2k}}\right\}ds\right)^{2k}\\ &\quad+K\left(\int_0^T\left\{\mathbf{1}_{E_{\varepsilon}}(s)\left(E\left|y^{\epsilon}(s)\right|^{2k}\right)^{\frac{1}{k}}+\left(E\left|y^{\epsilon}(s)\right|^{4k}\right)^{\frac{1}{k}}\right\}ds\right)^{k}\\ &\leq K\epsilon^{2k}.\end{aligned} \tag{2.49}$$

This gives (2.42).

3. Proof of (2.43) Set

$$\eta^{\epsilon}(t)=x^{\epsilon}(t)-\bar{x}(t)-y^{\epsilon}(t)=\zeta^{\epsilon}(t)-y^{\epsilon}(t). \tag{2.50}$$

By (2.46) and (2.37), we have

$$\begin{aligned}d\eta^{\epsilon}(t)&=\left[\tilde{b}_x^{\epsilon}(t)\zeta^{\epsilon}(t)+\delta b(t)\mathbf{1}_{E_{\varepsilon}}(t)-b_x(t)y^{\epsilon}(t)\right]dt\\ &\quad+\left[\tilde{\sigma}_x^{\epsilon}(t)\zeta^{\epsilon}(t)-\sigma_x(t)y^{\epsilon}(t)(t)\right]dW(t)\\ &=\left[\tilde{b}_x^{\epsilon}(t)\eta^{\epsilon}(t)+\delta b(t)\mathbf{1}_{E_{\varepsilon}}(t)+\left(\tilde{b}_x^{\epsilon}(t)-b_x(t)\right)y^{\epsilon}(t)\right]dt\\ &\quad+\left[\tilde{\sigma}_x^{\epsilon}(t)\eta^{\epsilon}(t)+\left(\tilde{\sigma}_x^{\epsilon}(t)-\sigma_x(t)\right)y^{\epsilon}(t)(t)\right]dW(t).\end{aligned} \tag{2.51}$$

Thus, it follows from Lemma 2.1.1 that

$$\begin{aligned} E\left|\eta^{\epsilon}(t)\right|^{2k} &\leq K\left(\int_0^T\left\{E\left|\delta b(s)\,\mathbf{1}_{E_\varepsilon}(s)+\left(\tilde{b}_x^{\epsilon}(s)-b_x(s)\right)y^{\epsilon}(s)\right|^{2k}\right\}^{\frac{1}{2k}}ds\right)^{2k} \\ &\quad +K\left(\int_0^T\left\{E\left|\left(\tilde{\sigma}_x^{\epsilon}(s)-\sigma_x(s)\right)y^{\epsilon}(s)\right|^{2k}\right\}^{\frac{1}{k}}ds\right)^{k} \\ &\leq K\left(\epsilon+\int_0^T\left(E\left|y^{\epsilon}(s)\right|^{4k}\right)^{\frac{1}{4k}}\left(E\left|\tilde{b}_x^{\epsilon}(s)-b_x(s)\right|^{4k}\right)^{\frac{1}{4k}}ds\right)^{2k} \\ &\quad +K\left(\int_0^T\left(E\left|y^{\epsilon}(s)\right|^{4k}\right)^{\frac{1}{2k}}\left(E\left|\left(\tilde{\sigma}_x^{\epsilon}(s)-\sigma_x(s)\right)\right|^{4k}\right)^{\frac{1}{2k}}ds\right)^{k} \\ &\leq K\left[\epsilon^{2k}+\epsilon^{k}\left(\int_0^T\left(E\left|\tilde{b}_x^{\epsilon}(s)-b_x(s)\right|^{4k}\right)^{\frac{1}{4k}}ds\right)^{2k}\right] \\ &\quad +K\left[\epsilon^{k}\left(\int_0^T\left(E\left|\left(\tilde{\sigma}_x^{\epsilon}(s)-\sigma_x(s)\right)\right|^{4k}\right)^{\frac{1}{2k}}ds\right)^{k}\right]. \end{aligned} \tag{2.52}$$

Note that (by ($\mathbf{S}_3$) and (2.40))

$$\begin{aligned} \int_0^T\left(E\left|\tilde{b}_x^{\epsilon}(s)-b_x(s)\right|^{4k}\right)^{\frac{1}{4k}}ds &= \int_0^T\left[E\left|\int_0^1\left(b_x^{\epsilon}\left(s,\bar{x}(s)+\theta\left(x^{\epsilon}(s)-\bar{x}(s)\right),u^{\epsilon}(s)\right)\right)\right.\right. \\ &\qquad \left.\left.-b_x\left(s,\bar{x}(s),\bar{u}(s)\right)d\theta\right|^{4k}\right]^{\frac{1}{4k}}ds \\ &\leq K\int_0^T\left\{E\left|L\left|x^{\epsilon}(s)-\bar{x}(s)\right|+\delta b(s)\,\mathbf{1}_{E_\varepsilon}(s)\right|^{4k}\right\}^{\frac{1}{4k}}ds \\ &\leq K\left[\epsilon+\int_0^T\left\{E\left|x^{\epsilon}(s)-\bar{x}(s)\right|^{4k}\right\}^{\frac{1}{4k}}ds\right]\leq K\sqrt{\epsilon}. \end{aligned} \tag{2.53}$$

Similarly, we have

$$\int_0^T\left(E\left|\tilde{\sigma}_x^{\epsilon}(s)-\sigma_x(s)\right|^{4k}\right)^{\frac{1}{2k}}ds\leq K\epsilon, \tag{2.54}$$

then (2.43) follows from (2.52).

4. Proof of (2.44) Set

$$\begin{aligned}\xi^{\epsilon}(t) &= x^{\epsilon}(t) - \bar{x}(t) - y^{\epsilon}(t) - z^{\epsilon}(t) = \zeta^{\epsilon}(t) - y^{\epsilon}(t) - z^{\epsilon}(t) \\ &= \eta^{\epsilon}(t) - z^{\epsilon}(t).\end{aligned} \tag{2.54}$$

It is clear that

$$\begin{cases} d\xi^{\epsilon}(t) = B(t)\,dt + A(t)\,dW(t), \\ \xi^{\epsilon}(0) = 0, \end{cases} \tag{2.55}$$

where (noting (2.33)-(2.38) and Lemma 2.2.2)

$$\begin{aligned} B(t) &= b(t, x^{\epsilon}(t), u^{\epsilon}(t)) - b(t, \bar{x}(t), u^{\epsilon}(t)) - b_x(t)\left[y^{\epsilon}(t) + z^{\epsilon}(t)\right] \\ &\quad -\tfrac{1}{2} b_{xx}(t)\, y^{\epsilon}(t)^2 \\ &= b_x(t, \bar{x}(t), u^{\epsilon}(t))\, \zeta^{\epsilon}(t) + \tfrac{1}{2}\tilde{b}^{\epsilon}_{xx}(t)\, \zeta^{\epsilon}(t)^2 \\ &\quad - b_x(t)\left[y^{\epsilon}(t) + z^{\epsilon}(t)\right] - \tfrac{1}{2} b_{xx}(t)\, y^{\epsilon}(t)^2 \\ &= b_x(t)\, \xi^{\epsilon}(t) + \delta b_x(t)\, \mathbf{1}_{E_{\varepsilon}}(t)\, \zeta^{\epsilon}(t) \\ &\quad + \tfrac{1}{2}\left[\tilde{b}^{\epsilon}_{xx}(t) - b_{xx}(t, \bar{x}(t), u^{\epsilon}(t))\right] \zeta^{\epsilon}(t)^2 \\ &\quad + \tfrac{1}{2}\delta b_{xx}(t)\, \mathbf{1}_{E_{\varepsilon}}(t)\, \zeta^{\epsilon}(t)^2 + \tfrac{1}{2} b_{xx}(t)\left[\zeta^{\epsilon}(t)^2 - y^{\epsilon}(t)^2\right] \\ &= b_x(t)\, \xi^{\epsilon}(t) + \alpha^{\epsilon}(t), \end{aligned} \tag{2.56}$$

and

$$\begin{aligned}
A(t) &= \sigma\left(t, x^{\epsilon}(t), u^{\epsilon}(t)\right) - \sigma\left(t, \bar{x}(t), u^{\epsilon}(t)\right) - \sigma_x(t)\left[y^{\epsilon}(t) + z^{\epsilon}(t)\right] \\
&\qquad -\tfrac{1}{2}\sigma_{xx}(t)\, y^{\epsilon}(t)^2 - \delta\sigma_x(t)\, \mathbf{1}_{E_\varepsilon}(t)\, y^{\epsilon}(t) \\
&= \sigma_x\left(t, \bar{x}(t), u^{\epsilon}(t)\right)\zeta^{\epsilon}(t) + \tfrac{1}{2}\tilde{\sigma}^{\epsilon}_{xx}(t)\,\zeta^{\epsilon}(t)^2 - \sigma_x(t)\left[y^{\epsilon}(t) + z^{\epsilon}(t)\right] \\
&\qquad -\tfrac{1}{2}\sigma_{xx}(t)\, y^{\epsilon}(t)^2 - \delta\sigma_x(t)\, \mathbf{1}_{E_\varepsilon}(t)\, y^{\epsilon}(t) \\
&= \sigma_x(t)\,\xi^{\epsilon}(t) + \delta\sigma_x(t)\,\mathbf{1}_{E_\varepsilon}(t)\,\eta^{\epsilon}(t) \\
&\qquad +\tfrac{1}{2}\left[\tilde{\sigma}^{\epsilon}_{xx}(t) - \sigma_{xx}\left(t, \bar{x}(t), u^{\epsilon}(t)\right)\right]\zeta^{\epsilon}(t)^2 \\
&\qquad +\tfrac{1}{2}\delta\sigma_{xx}(t)\,\mathbf{1}_{E_\varepsilon}(t)\,\zeta^{\epsilon}(t)^2 + \tfrac{1}{2}\sigma_{xx}(t)\left[\zeta^{\epsilon}(t)^2 - y^{\epsilon}(t)^2\right] \\
&= \sigma_x(t)\,\xi^{\epsilon}(t) + \beta^{\epsilon}(t),
\end{aligned} \tag{2.57}$$

with

$$\begin{cases}
\tilde{b}^{\epsilon}_x(t) = 2\displaystyle\int_0^1 \theta b_{xx}\left(t, \theta\bar{x}(t) + (1-\theta)\, x^{\epsilon}(t), u^{\epsilon}(t)\right) d\theta, \\
\tilde{\sigma}^{\epsilon}_x(t) = 2\displaystyle\int_0^1 \theta \sigma_{xx}\left(t, \theta\bar{x}(t) + (1-\theta)\, x^{\epsilon}(t), u^{\epsilon}(t)\right) d\theta,
\end{cases} \tag{2.58}$$

and

$$\begin{cases}
\alpha^{\epsilon}(t) = \delta b_x(t)\,\mathbf{1}_{E_\varepsilon}(t)\,\zeta^{\epsilon}(t) + \tfrac{1}{2}\left[\tilde{b}^{\epsilon}_{xx}(t) - b_{xx}\left(t, \bar{x}(t), u^{\epsilon}(t)\right)\right]\zeta^{\epsilon}(t)^2 \\
\qquad +\tfrac{1}{2}\delta b_{xx}(t)\,\mathbf{1}_{E_\varepsilon}(t)\,\zeta^{\epsilon}(t)^2 + \tfrac{1}{2}b_{xx}(t)\left[\zeta^{\epsilon}(t)^2 - y^{\epsilon}(t)^2\right], \\
\beta^{\epsilon}(t) = \delta\sigma_x(t)\,\mathbf{1}_{E_\varepsilon}(t)\,\eta^{\epsilon}(t) + \tfrac{1}{2}\left[\tilde{\sigma}^{\epsilon}_{xx}(t) - \sigma_{xx}\left(t, \bar{x}(t), u^{\epsilon}(t)\right)\right]\zeta^{\epsilon}(t)^2 \\
\qquad +\tfrac{1}{2}\delta\sigma_{xx}(t)\,\mathbf{1}_{E_\varepsilon}(t)\,\zeta^{\epsilon}(t)^2 + \tfrac{1}{2}\sigma_{xx}(t)\left[\zeta^{\epsilon}(t)^2 - y^{\epsilon}(t)^2\right].
\end{cases} \tag{2.59}$$

In order to use Lemma 2.2.1, we need to estimate $\alpha^{\epsilon}(.)$ and $\beta^{\epsilon}(.)$. To this end, recall that $\bar{w}$ appearing in $(\mathbf{S}_3)$ is a modulus of continuity for $b_{xx}(t,.,u)$ (uniform in $t \in [0,T]$ and $u \in U$). Thus for any $\rho > 0$, there exists a constant $K_\rho > 0$ such that

$$\bar{w}(r) \leq \rho + rK_\rho, \qquad \forall r \geq 0. \tag{2.60}$$

Consequently,

$$\left|\tilde{b}^{\epsilon}_{xx}(t) - b_{xx}(t, \bar{x}(t), u^{\epsilon}(t))\right| \leq \bar{w}(|\zeta^{\epsilon}(t)|) \leq \rho + rK_{\rho}|\zeta^{\epsilon}(t)|. \tag{2.61}$$

Recalling (2.50) and (2.59), as well as (2.40)-(2.43), we can estimate $\alpha^{\epsilon}(.)$ as follows

$$\begin{aligned}
\int_0^T \left(E|\alpha^{\epsilon}(t)|^{2k}\right)^{\frac{1}{2k}} dt &\leq \int_0^T \Bigg\{\left(E|\delta b_x(t)\,\mathbf{1}_{E_\varepsilon}(t)\,\zeta^{\epsilon}(t)|^{2k}\right)^{\frac{1}{2k}} \\
&+ \left(E\left|\tfrac{1}{2}\left[\tilde{b}^{\epsilon}_{xx}(t) - b_{xx}(t, \bar{x}(t), u^{\epsilon}(t))\right]\zeta^{\epsilon}(t)^2\right|^{2k}\right)^{\frac{1}{2k}} \\
&+ \left(E\left|\tfrac{1}{2}\delta b_{xx}(t)\,\mathbf{1}_{E_\varepsilon}(t)\,\zeta^{\epsilon}(t)^2\right|^{2k}\right)^{\frac{1}{2k}} \\
&+ \left(E\left|\tfrac{1}{2}b_{xx}(t)\left[\zeta^{\epsilon}(t)^2 - y^{\epsilon}(t)^2\right]\right|^{2k}\right)^{\frac{1}{2k}}\Bigg\} dt \\
&\leq K\int_0^T \Bigg\{\mathbf{1}_{E_\varepsilon}(t)\left(E|\zeta^{\epsilon}(t)|^{2k}\right)^{\frac{1}{2k}} + \mathbf{1}_{E_\varepsilon}(t)\left(E|\zeta^{\epsilon}(t)|^{4k}\right)^{\frac{1}{2k}} \\
&+ \left(E[\rho + K_{\rho}|\zeta^{\epsilon}(t)|]^{4k}\right)^{\frac{1}{4k}}\left(E|\zeta^{\epsilon}(t)|^{8k}\right)^{\frac{1}{4k}} \\
&+ \left(E|\eta^{\epsilon}(t)|^{4k}\right)^{\frac{1}{4k}}\left(E|\zeta^{\epsilon}(t) + y^{\epsilon}(t)|^{4k}\right)^{\frac{1}{4k}}\Bigg\} dt \\
&\leq K\left\{\epsilon^{\frac{3}{2}} + \epsilon(\rho + \sqrt{\epsilon}K_{\rho}) + \epsilon^2 + \epsilon^{\frac{3}{2}}\right\}
\end{aligned} \tag{2.62}$$

this implies

$$\int_0^T \left(E|\alpha^{\epsilon}(t)|^{2k}\right)^{\frac{1}{2k}} dt = \circ(\epsilon).$$

Similar to (2.61), we have

$$|\tilde{\sigma}^{\epsilon}_{xx}(t) - \sigma_{xx}(t, \bar{x}(t), u^{\epsilon}(t))| \leq \bar{w}(|\zeta^{\epsilon}(t)|) \leq \rho + rK_{\rho}|\zeta^{\epsilon}(t)|. \tag{2.63}$$

As with (2.62), we can estimate $\beta^{\epsilon}(.)$ as follows.

$$\begin{aligned}\int_0^T \left(E\left|\beta^{\epsilon}(t)\right|^{2k}\right)^{\frac{1}{2k}} dt &\leq \int_0^T \left\{\left(E\left|\delta\sigma_x(t)\,\mathbf{1}_{E_\varepsilon}(t)\,\eta^{\epsilon}(t)\right|^{2k}\right)^{\frac{1}{k}}\right.\\ &+\left(E\left|\tfrac{1}{2}\left[\tilde{\sigma}^{\epsilon}_{xx}(t)-\sigma_{xx}(t,\bar{x}(t),u^{\epsilon}(t))\right]\zeta^{\epsilon}(t)^2\right|^{2k}\right)^{\frac{1}{k}}\\ &+\left(E\left|\tfrac{1}{2}\delta\sigma_{xx}(t)\,\mathbf{1}_{E_\varepsilon}(t)\,\zeta^{\epsilon}(t)^2\right|^{2k}\right)^{\frac{1}{k}}\\ &\left.+\left(E\left|\tfrac{1}{2}\sigma_{xx}(t)\left[\zeta^{\epsilon}(t)^2-y^{\epsilon}(t)^2\right]\right|^{2k}\right)^{\frac{1}{k}}\right\} dt\\ &\leq K\int_0^T\left\{\mathbf{1}_{E_\varepsilon}(t)\left(E\left|\eta^{\epsilon}(t)\right|^{2k}\right)^{\frac{1}{k}}+\mathbf{1}_{E_\varepsilon}(t)\left(E\left|\zeta^{\epsilon}(t)\right|^{4k}\right)^{\frac{1}{k}}\right.\\ &+\left(E\left[\rho+K_\rho\left|\zeta^{\epsilon}(t)\right|\right]^{4k}\right)^{\frac{1}{2k}}\left(E\left|\zeta^{\epsilon}(t)\right|^{8k}\right)^{\frac{1}{2k}}\\ &\left.+\left(E\left|\eta^{\epsilon}(t)\right|^{4k}\right)^{\frac{1}{2k}}\left(E\left|\zeta^{\epsilon}(t)+y^{\epsilon}(t)\right|^{4k}\right)^{\frac{1}{2k}}\right\} dt\\ &\leq K\left\{\epsilon^3+\epsilon^2\left(\rho^2+\epsilon K_\rho^2\right)+\epsilon^3+\epsilon^3\right\}\end{aligned} \tag{2.64}$$

thus

$$\int_0^T \left(E\left|\beta^{\epsilon}(t)\right|^{2k}\right)^{\frac{1}{k}} dt = \circ\left(\epsilon^2\right). \tag{2.65}$$

Then, by Lemma 2.2.1, we obtain (2.44).

5. Proof of (2.45) By Lemma 2.2.2, we have

$$\begin{aligned}J(u^{\epsilon}(.))-J(\bar{u}(.)) &= E\left[h(x^{\epsilon}(T))-h(\bar{x}(T))\right]\\ &+E\int_0^T\left[f(t,x^{\epsilon}(t),u^{\epsilon}(t))-f(t,\bar{x}(t),\bar{u}(t))\right]dt\\ &= E\left\langle h(\bar{x}(T)),\zeta^{\epsilon}(T)\right\rangle\\ &+E\int_0^1\left\langle \theta h_{xx}\left[\theta\bar{x}(T)+(1-\theta)x^{\epsilon}(T)\right]\zeta^{\epsilon}(T),\zeta^{\epsilon}(T)\right\rangle d\theta\\ &+E\int_0^T\left[\delta f(t)\,\mathbf{1}_{E_\varepsilon}(t)+\left\langle f_x(t,\bar{x}(t),u^{\epsilon}(t)),\zeta^{\epsilon}(t)\right\rangle\right.\\ &\left.+\left\langle \theta f_{xx}\left[t,\theta\bar{x}(t)+(1-\theta)x^{\epsilon}(t)\right]\zeta^{\epsilon}(t),\zeta^{\epsilon}(t)\right\rangle d\theta\right]dt.\end{aligned}$$

Now, recalling the definitions of ζ^{ϵ}, η^{ϵ} and ξ^{ϵ}, we have

$$\begin{aligned}
J(u^{\epsilon}(.)) - J(\bar{u}(.)) &= E\langle h_x(\bar{x}(T)), y^{\epsilon}(T) + z^{\epsilon}(T)\rangle + E\langle h_x(\bar{x}(T)), \xi^{\epsilon}(T)\rangle \\
&+\tfrac{1}{2}E\langle h_{xx}(\bar{x}(T))\, y^{\epsilon}(T), y^{\epsilon}(T)\rangle \\
&+\tfrac{1}{2}E\langle h_{xx}(\bar{x}(T))\, \eta^{\epsilon}(T), \zeta^{\epsilon}(T) + y^{\epsilon}(T)\rangle \\
&+E\int_0^1 \langle \theta\,[h_{xx}(\theta\bar{x}(T) + (1-\theta)\, x^{\epsilon}(T)) - h_{xx}(\bar{x}(T))]\,\zeta^{\epsilon}(T), \zeta^{\epsilon}(T)\rangle\, d\theta \\
&+E\int_0^T \{\delta f(t)\,\mathbf{1}_{E_\varepsilon}(t) + \langle \delta f_x(t), \zeta^{\epsilon}(t)\rangle\,\mathbf{1}_{E_\varepsilon}(t) \\
&+\langle f_x(t), y^{\epsilon}(t) + z^{\epsilon}(t)\rangle + \langle f_x(t), \xi^{\epsilon}(t)\rangle \\
&+\int_0^1 \langle \theta\,[f_{xx}(t, \theta\bar{x}(t) + (1-\theta)\, x^{\epsilon}(t)\, x^{\epsilon}(t), u^{\epsilon}(t)) \\
&-f_{xx}(t, \bar{x}(t), u^{\epsilon}(t))]\,\zeta^{\epsilon}(t), \zeta^{\epsilon}(t)\rangle\, d\theta \\
&+\tfrac{1}{2}\langle \delta f_{xx}(t)\,\zeta^{\epsilon}, \zeta^{\epsilon}\rangle\,\mathbf{1}_{E_\varepsilon}(t) \\
&+\tfrac{1}{2}\langle f_{xx}(t)\, y^{\epsilon}(t), y^{\epsilon}(t)\rangle + \tfrac{1}{2}\langle f_{xx}(t)\,\eta^{\epsilon}(t), \zeta^{\epsilon} + y^{\epsilon}(t)\rangle\big\}\, dt.
\end{aligned}$$

Then, by (2.40)-(2.44), we can show that

$$\begin{aligned}
J(u^{\epsilon}(.)) - J(\bar{u}(.)) &= E\langle h_x(\bar{x}(T)), y^{\epsilon}(T) + z^{\epsilon}(T)\rangle + \tfrac{1}{2}E\langle h_{xx}(\bar{x}(T))\, y^{\epsilon}(T), y^{\epsilon}(T)\rangle \\
&+E\int_0^T \big\{\langle f_x(t), y^{\epsilon}(t) + z^{\epsilon}(t)\rangle + \tfrac{1}{2}\langle f_{xx}(t)\, y^{\epsilon}(t), y^{\epsilon}(t)\rangle \\
&\delta f(t)\,\mathbf{1}_{E_\varepsilon}(t)\}\, dt + R(\epsilon) \qquad .
\end{aligned} \tag{2.66}$$

where $R(\epsilon)$ is of order $\circ(\epsilon)$. Hence, our conclusion follows.

■

2.2.3 Duality analysis and completion of the proof

From Theorem 2.2.1, we conclude that a necessary condition for a given optimal pair $(\bar{x}(.), \bar{u}(.))$ is the following

$$\begin{aligned} 0 \leq & \left\{E\left\langle h_x(\bar{x}(T)), y^\epsilon(T)+z^\epsilon(T)\right\rangle+\tfrac{1}{2}\left\langle h_{xx}(\bar{x}(T)) y^\epsilon(T), y^\epsilon(T)\right\rangle\right\} \\ & +E \int_0^T\left\{\left\langle f_x(t), y^\epsilon(t)+z^\epsilon(t)\right\rangle+\tfrac{1}{2}\left\langle f_{xx}(t) y^\epsilon(t), y^\epsilon(t)\right\rangle\right. \\ & \left.\delta f(t) \mathbf{1}_{E_\varepsilon}(t)\right\} dt+\circ(\epsilon), \qquad \forall u(.) \in \mathcal{U}[0, T], \forall \epsilon>0, \end{aligned} \tag{2.67}$$

where $y^\epsilon(.)$ and $z^\epsilon(.)$ are solutions to the (approximate) variational systems (2.37) and (2.36), respectively. As in the deterministic case, we are now in a position to get rid of $y^\epsilon(.)$ and $z^\epsilon(.)$, and then pass to the limit. To this end, we need some duality relations between the variational systems (2.37)-(2.38) and the adjoint equations (2.8) and (2.9).

Lemma 2.2.3 *Let ($\boldsymbol{S}_0$)-($\boldsymbol{S}_3$) hold. Let $y^\epsilon(.)$ and $z^\epsilon(.)$ be the solutions of (2.37) and (2.38), respectively. Let $(p(.), q(.))$ be the adapted solution of (2.8). Then*

$$E\left\langle p(T), y^\epsilon(T)\right\rangle=E \int_0^T\left[\left\langle f_x(t), y^\epsilon(t)\right\rangle+tr\left(q(t)^* \delta\sigma(t)\right) \mathbf{1}_{E_\varepsilon}(t)\right] dt, \tag{2.68}$$

and

$$\begin{aligned} E\left\langle p(T), z^\epsilon(T)\right\rangle= & E \int_0^T\left\{\left\langle f_x(t), z^\epsilon(t)\right\rangle\right. \\ & +\tfrac{1}{2}\left[\left\langle p(t), b_{xx}(t) y^\epsilon(t)^2\right\rangle+\sum_{j=1}^m\left\langle q_j(t), \sigma_{xx}^j(t) y^\epsilon(t)^2\right\rangle\right] \\ & \left.+\left[\left\langle p(t), \delta b(t)\right\rangle+\sum_{j=1}^m\left\langle q_j(t), \delta\sigma_x^j(t) y^\epsilon(t)\right\rangle\right] \mathbf{1}_{E_\varepsilon}(t)\right\} dt. \end{aligned} \tag{2.69}$$

Adding (2.68) and (2.69), and appealing to the Taylor expansions given in Theorem 2.2.1, we get

$$-E\left\langle h_x\left(\bar{x}\left(T\right)\right),y^\epsilon\left(T\right)+z^\epsilon\left(T\right)\right\rangle=E\int_0^T\left\{\left\langle f_x\left(t\right),y^\epsilon\left(t\right)+z^\epsilon\left(t\right)\right\rangle+\tfrac{1}{2}\left\langle p\left(t\right),b_{xx}\left(t\right)y^\epsilon\left(t\right)^2\right\rangle\right.$$
$$\left.+\left[\tfrac{1}{2}\sum_{j=1}^m\left\langle q_j\left(t\right),\sigma_{xx}^j\left(t\right)y^\epsilon\left(t\right)^2\right\rangle+\left\langle p\left(t\right),\delta b\left(t\right)\right\rangle+tr\left(q\left(t\right)^*\delta\sigma\left(t\right)\right)\right]\mathbf{1}_{E_\varepsilon}\left(t\right)\right\}dt+\circ\left(\epsilon\right) \tag{2.70}$$

Thus, by (2.44) and the optimality of $\bar{u}\left(.\right)$, we have

$$\begin{aligned}
&0\geq J\left(u^\epsilon\left(.\right)\right)-J\left(\bar{u}\left(.\right)\right)\\
&=-\tfrac{1}{2}E\left\langle h_{xx}\left(\bar{x}\left(T\right)\right)y^\epsilon\left(T\right),y^\epsilon\left(T\right)\right\rangle\\
&+\tfrac{1}{2}E\int_0^T\left\{-\left\langle f_{xx}\left(t\right)y^\epsilon\left(t\right),y^\epsilon\left(t\right)\right\rangle+\left\langle p\left(t\right),b_{xx}\left(t\right)y^\epsilon\left(t\right)^2\right\rangle\right.\\
&\left.+\sum_{j=1}^m\left\langle q_j\left(t\right),\sigma_{xx}^j\left(t\right)y^\epsilon\left(t\right)^2\right\rangle\right\}dt\\
&+E\int_0^T\left\{-\delta f\left(t+\right)\left\langle p\left(t\right),\delta b\left(t\right)\right\rangle\right.\\
&\left.+\sum_{j=1}^m\left\langle q_j\left(t\right),\delta\sigma^j\left(t\right)\right\rangle\right\}\mathbf{1}_{E_\varepsilon}\left(t\right)dt+\circ\left(\epsilon\right)\\
&=\tfrac{1}{2}E\left[tr\left(P\left(T\right)Y^\epsilon\left(T\right)\right)\right]\\
&+E\int_0^T\left\{\tfrac{1}{2}tr\left[H_{xx}\left(t\right)Y^\epsilon\left(t\right)\right]+\delta H\left(t\right)\mathbf{1}_{E_\varepsilon}\left(t\right)\right\}dt+\circ\left(\epsilon\right),
\end{aligned} \tag{2.71}$$

where

$$\begin{cases}
Y^\epsilon\left(t\right)=y^\epsilon\left(t\right)y^\epsilon\left(t\right)^*\\
H_{xx}\left(t\right)=H_{xx}\left(t,\bar{x}\left(t\right),\bar{u}\left(t\right),p\left(t\right),q\left(t\right)\right),\\
\delta H\left(t\right)=H\left(t,\bar{x}\left(t\right),u\left(t\right),p\left(t\right),q\left(t\right)\right)-H\left(t,\bar{x}\left(t\right),\bar{u}\left(t\right),p\left(t\right),q\left(t\right)\right)
\end{cases} \tag{2.72}$$

We see that (2.71) no longer contains the first-order terms in $y^\epsilon\left(.\right)$ and $z^\epsilon\left(.\right)$. But, unlike the deterministic case, there are left some second-order terms in $y^\epsilon\left(.\right)$, which are written in terms of the first-order in $Y^\epsilon\left(.\right)$. Hence, we want further to get rid of $Y^\epsilon\left(.\right)$. To this end, we need some duality relation between the equation satisfied by $Y^\epsilon\left(.\right)$ and the second-order adjoint equation (2.9) (which is exactly where the second-order adjoint equation comes in). Let us now derive the SDE satisfied by $Y^{\epsilon}(.)$. Applying Itô's formula to $y^{\epsilon}(t)\, y^{\epsilon}(t)^{*}$ and noting (2.38), one has

$$\begin{aligned}
&dY^{\epsilon}(t) = \Big\{ b_x(t)\, Y^{\epsilon}(t) + Y^{\epsilon}(t)\, b_x(t)^{*} \\
&+\sum_{j=1}^{m} \sigma_x^j(t)\, Y^{\epsilon}(t)\, \sigma_x^j(t)^{*} + \sum_{j=1}^{m} \delta\sigma^j(t)\, \delta\sigma^j(t)^{*}_{E_\varepsilon}\, \mathbf{1}(t) \\
&+\sum_{j=1}^{m} \left(\sigma_x^j(t)\, y^{\epsilon}(t)\, \sigma^j(t)^{*} + \delta\sigma^j(t)\, y^{\epsilon}(t)^{*}\, \delta\sigma_x^j(t)^{*}\right) \mathbf{1}_{E_\varepsilon}(t) \Big\}\, dt \\
&+\sum_{j=1}^{m} \left(\sigma_x^j(t)\, Y^{\epsilon}(t) + Y^{\epsilon}(t)\, \sigma_x^j(t)^{*}\right) dW^j(t) \\
&+\sum_{j=1}^{m} \left(\delta\sigma^j(t)\, y^{\epsilon}(t)^{*} + y^{\epsilon}(t)\, \delta\sigma^j(t)^{*}\right) \mathbf{1}_{E_\varepsilon}(t)\, dW^j(t).
\end{aligned} \tag{2.73}$$

To establish the duality relation between (2.73) and (2.9), we need the following lemma, whose proof follows directly from Itô's formula.

Lemma 2.2.4 *Let $Y(.), P(.) \in L^2_{\mathcal{F}}(0, T, \mathbb{R}^{n\times n})$ satisfy the following*

$$\begin{cases}
dY(t) = \Phi(t)\, dt + \sum_{j=1}^{m} \Psi_j(t)\, dW^j(t), \\
dP(t) = \Theta(t)\, dt + \sum_{j=1}^{m} Q_j(t)\, dW^j(t),
\end{cases} \tag{2.74}$$

with $\Phi(.), \Psi_j(.), \Theta(.)$ and $Q_j(.)$ all being elements in $L^2_{\mathcal{F}}(0, T, \mathbb{R}^{n\times n})$. Then

$$\begin{aligned}
&E\left\{ tr\left[P(t)\, Y(t)\right] - tr\left[P(0)\, Y(0)\right] \right\} \\
&= E\int_0^T \left\{ tr\left[\Theta(t)\, Y(t) + P(t)\, \Phi(t) + \sum_{j=1}^{m} Q_j(t)\, \Psi_j(t) \right] \right\} dt
\end{aligned} \tag{2.75}$$

Proof. (**Theorem 2.1.1**) Now we apply the above lemma to (2.73) and (2.9) to get the following (using Theorem 2.2.1, and noting $tr\,(AB) = tr\,(BA)$ and $Y\,(0) = 0$)

$$\begin{aligned} & E\left\{tr\left[P\left(t\right)Y^{\epsilon}\left(t\right)\right]\right\} \\ & = E\int_0^T tr\left(\delta\sigma\left(t\right)^{*} P\left(t\right)\delta\sigma\left(t\right)\mathbf{1}_{E_{\varepsilon}}\left(t\right) - H_{xx}\left(t\right)Y^{\epsilon}\left(t\right)\right)dt + \circ\left(\epsilon\right), \end{aligned} \tag{2.76}$$

where

$$\begin{cases} H_{xx}\left(t\right) = H_{xx}\left(t,\bar{x}\left(t\right),\bar{u}\left(t\right),p\left(t\right),q\left(t\right)\right), \\ \delta\sigma\left(t\right) = \sigma\left(t,\bar{x}\left(t\right),u\left(t\right)\right) - \sigma\left(t,\bar{x}\left(t\right),\bar{u}\left(t\right)\right). \end{cases}$$

Hence, (2.73) can be written as

$$\circ\left(\epsilon\right) \geq E\int_0^T \left(\delta H\left(t\right) + \frac{1}{2}tr\left[\delta\sigma\left(t\right)^{T} P\left(t\right)\delta\sigma\left(t\right)\right]\right)\mathbf{1}_{E_{\varepsilon}}\left(t\right)dt. \tag{2.77}$$

Then we can easily obtain the variational inequality (2.14). Easy manipulation shows that (2.14) is equivalent to (2.15). This completes the proof of Theorem 2.1.1. ■

Chapter 3

Maximum principle in optimal control of systems driven by martingale measures

In this chapter, first we present a generalization of the notion of control, of course, a control problem -and even a deterministic one- does not necessarily have a solution. Then we present our result.

3.1 Control problem

3.1.1 Strict control problem

The systems we wish to control are driven by the following n-dimensional stochastic differential equations of diffusion type, defined on some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$

$$dx(t) = b(t, x(t), u(t))dt + \sigma(t, x(t), u(t))dW_t, \qquad x(0) = x_0 \tag{3.1}$$

where, for each $t \in [0,1]$, the control u_t is in the action space $\mathcal{A}$, a compact set in $\mathbb{R}^k$, the drift term $b : \mathbb{R}_+ \times \mathbb{R}^n \times \mathcal{A} \to \mathbb{R}$, and diffusion coefficient $\sigma : \mathbb{R}_+ \times \mathbb{R}^n \times \mathcal{A} \to \mathbb{R}^n \otimes \mathbb{R}^m$ are bounded measurable and continuous in (x, a).

The finitesimal generator L, associate with (3.1), acting on functions f in $C_b^2(\mathbb{R}^n, \mathbb{R})$, is

$$Lf(t,x,u) = \frac{1}{2}\sum_{i,j}\left(a_{ij}\frac{\partial^2 f}{\partial x_i \partial x_j}\right)(t,x,u) + \sum_j \left(b_j \frac{\partial f}{\partial x_j}\right)(t,x,u) \tag{3.2}$$

where $a_{ij}(t,x,u)$ denotes the generic term of the symmetric matrix $\sigma\sigma^*(t,x,u)$. Let $\mathcal{U}$ denote the class of admissible controls, that is $(\mathcal{F}_t)_t$-adapted process with values in the action space $\mathcal{A}$. This class is nonempty since it contains constant controls.

The function to be minimized over such controls is

$$J(u) = E\left[\int_0^1 h(t, x(t), u(t))\, dt + g(x_1)\right], \tag{3.3}$$

where h and g are assumed to be real-valued, continuous, and bounded, respectively, on $\mathbb{R}_+ \times \mathbb{R}^n \times \mathcal{A}$ and on $\mathbb{R}^n$.

We now introduce the notion of strict control to (3.1).

Definition 3.1.1 *A strict control is the term* $\alpha = (\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P, u(t), x(t), x_0)$ *such that*

(1) $x_0 \in \mathbb{R}^n$ *is the initial data;*

(2) $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ *is a probability space equipped with a filtration* $(\mathcal{F}_t)_{t\geq 0}$ *satisfying the usual conditions;*

(3) $u(t)$ *is an A-valued process, progressively measurable with respect to* $(\mathcal{F}_t)$*;*

(4) $(x(t))$ *is* $\mathbb{R}^n$*-valued,* $\mathcal{F}_t$*-adapted, with continuous paths, such that*

$$f(x(t)) - f(x_0) - \int_0^t Lf(s, x(s), u(s))\, ds \quad \text{is a } P\text{-martingale,} \tag{3.4}$$

for each $f \in C_b^2$*, for each* $t > 0$*, where* L *is the infinitesimal generator of the diffusion*

$(x(t))$.

In fact, when the control $u(t)$ is constant, the conditions imposed above on the drift term and diffusion coefficient ensure that our martingale problem admits at least one solution, which implies weak existence of solutions of (3.1). The associated controls are called weak controls because of the possible change of the probability space and the Brownian motion with $u(t)$. When pathwise uniqueness holds for the controlled equation it is shown in El Karoui and al [23], that the weak and strong control problems are equivalent in the sense that they have the same value functions.

3.1.2 Relaxed control problem

The strict control problem as defined in the last section may fail to have an optimal solution. We begin by ad-hoc famous example taken from [46] in order to illustrate what we are going to do.

An example

Consider $U = \{-1, 1\}$ and consider piecewise continuous function $u : [0, 1] \to U$ (the controls).

The dynamic of the problem is given by the differential equation

$$\begin{cases} \dfrac{dx_t^u}{dt} = u(t) \\ x_0^u = 0 \end{cases}$$

and the cost associated to the problem is

$$J(u) = \int_0^1 (x_t^u)^2 \, dt$$

First claim: $\inf_u J(u) = 0$.

Indeed, consider an integer $n \in \mathbb{N}^*$ and take

$$u_n(t) = (-1)^k, \quad \text{if } \frac{k}{n} \leq t < \frac{k+1}{n} \text{ for } 0 \leq k \leq n-1.$$

Then, clearly, for all $t \in [0,1]$, $|x_t^{u_n}| \leq \dfrac{1}{n}$ and so $J(u) \leq \dfrac{1}{n^2}$.

Second claim: there is not an u such that $J(u) = 0$.

This is obvious as it would imply that $x_t^u = 0$, $\forall t$ and so $u_t = 0$ which is impossible.

If we analyze the previous example, we can understand where the trouble is: it is the fact that the sequence (u_n) lacks a limit in the space of controls, limit which should be the natural candidate to optimality. So we look for a space in which this limit exists.

Identify $u_n(t)$ with the Dirac measure on $U : \delta_{u_n(t)}(du)$. Set

$$q_n(dt, du) = \delta_{u_n(t)}(du)\, dt,$$

q_n is a measure over the space $[0,1] \times U$.

Lemma 3.1.1 *q_n converge weakly to*

$$\bar{q}(dt, du) = \frac{1}{2}\left[\delta_{-1} + \delta_1\right](du)\, dt.$$

Proof. Take f a continuous function on $[0,1] \times U$ (of course only the continuity over $[0,1]$ is meaningful).

One has

$$\int_{[0,1]\times U} f(t,u)\, q_n(dt, du) = \sum_{k=0}^{n-1} \int_{\frac{k}{n}}^{\frac{k+1}{n}} f\left(t, (-1)^k\right) dt.$$

Suppose first that n is even: $n = 2m$.

As $t \to f(t,1)$ and $t \to f(t,-1)$ are continuous over $[0,1]$, they are uniformly continuous.

Let $\epsilon > 0$. There is an $M > 0$ such that

$$\forall m \geq M, |f(t,u) - f(s,u)| < \epsilon \text{ if } |t-s| < \frac{1}{m}$$

where u is either 1 or -1.

Fix $m \geq M$. Then, for every $j = 0, ..., m-1$,

$$\left| \int_{\frac{2j}{2m}}^{\frac{2j+1}{2m}} f(t,u)\,dt - \int_{\frac{2j+1}{2m}}^{\frac{2j+2}{2m}} f(t,u)\,dt \right| < \frac{\epsilon}{2m}$$

one has

$$\sum_{j=0}^{m-1} \int_{\frac{2j}{2m}}^{\frac{2j+1}{2m}} f(t,u)\,dt + \sum_{j=0}^{m-1} \int_{\frac{2j+1}{2m}}^{\frac{2j+2}{2m}} f(t,u)\,dt = \int_0^1 f(t,u)\,dt$$

and

$$\left| \sum_{j=0}^{m-1} \int_{\frac{2j}{2m}}^{\frac{2j+1}{2m}} f(t,u)\,dt - \sum_{j=0}^{m-1} \int_{\frac{2j+1}{2m}}^{\frac{2j+2}{2m}} f(t,u)\,dt \right| < \frac{\epsilon}{2}$$

therefore,

$$\left| \sum_{j=0}^{m-1} \int_{\frac{2j}{2m}}^{\frac{2j+1}{2m}} f(t,u)\,dt - \frac{1}{2}\int_0^1 f(t,u)\,dt \right| < \frac{\epsilon}{2}$$

and

$$\left| \sum_{j=0}^{m-1} \int_{\frac{2j+1}{2m}}^{\frac{2j+2}{2m}} f(t,u)\,dt - \frac{1}{2}\int_0^1 f(t,u)\,dt \right| < \frac{\epsilon}{2}$$

so

$$\left| \sum_{k=0}^{2m-1} \int_{\frac{k}{2m}}^{\frac{k+1}{2m}} f\left(t,(-1)^k\right) dt - \frac{1}{2}\left[\int_0^1 f(t,1)\,dt + \int_0^1 f(t,-1)\,dt \right] \right| < \epsilon.$$

The case n odd is treated in the same way. ■

Now, we can define a "new" control problem associated to such a measure q, which is called a relaxed control.

Consider the dynamic

$$x^q(t) = x_0 + \int_0^t \int_U u q(ds,du)$$

and the associated cost is, as before,

$$J(q) = \int_0^1 (x^q(t))^2 \, dt.$$

Then it is clear that the previous problem is generalized by the present problem by taking measures q of the form

$$q(dt, du) = \delta_{u_t}(du)\, dt$$

moreover, if

$$\bar{q}(dt, du) = \frac{1}{2}[\delta_{-1} + \delta_1](du)\, dt$$

we have $J(\bar{q}) = 0$ and so the new problem has $\bar{q}$ as an optimal solution.

Remark 3.1.1 *We denote by $\mathcal{R}$ the collection of all relaxed controls.*

By a slight abuse of notation, we will often denote a relaxed control by q instead of specifying all the components.

Relaxed controls

We could want to take as controls all the measures $q(dt, du)$. However, for our purpose which is to prove existence of an optimal control, we have in mind to restrict to a compact space containing "classical" controls. This is why the following definition is set.

Definition 3.1.2 *Let $U \subset \mathbb{R}^k$. A relaxed control with values in U is a measure q over $[0, T] \times U$ such that the projection on $[0, T]$ is the Lebesgue measure.*

If there exists $u : [0, T] \to U$ such that

$$q(dt, du) = \delta_{u(t)}(du)\, dt,$$

q is identified with $(u(t))$ and said to be a control process.

We have an interesting decomposition of such a relaxed control.

Proposition 3.1.1 *Let q be a relaxed control with values in U. Then, for all $t \in [0,T]$, there exists a probability measure q_t over U such that*

$$q\left(dt, du\right) = dtq_t\left(du\right).$$

The proof is an application of Fubini theorem. The previous Proposition 3.1.1 allows us to better interpret what a relaxed control is. In a control process, at a time t, we assign the value $u\left(t\right)$. In a relaxed control, the value is "randomly" chosen over the space U with the probability distribution $q_t\left(du\right)$.

Another interest of Proposition 3.1.1 is that we can introduce a canonical decomposition of relaxed controls.

Definition 3.1.3 *Let $\mathcal{R}$ be the space of relaxed controls over U. Let $\alpha \in \mathcal{R}$. There exists by Proposition 3.1.1 a process (α_s) with values in the set of probability measures on U and such that*

$$\alpha\left(ds, du\right) = ds\alpha_s\left(du\right).$$

The process (q_t) defined on $\mathcal{R}$, which associates the process (α_t) to α is said the canonical process on $\mathcal{R}$.

The filtration $\mathcal{V}_t = (q_s, s \leq t)$ is said the canonical filtration.

Remark 3.1.2 *We can see that $\mathcal{V}_t$ is generated by relaxed controls q such that*

$$q_{[t,T]\times U}\left(ds, du\right) = \delta_{u_0}\left(du\right)dt$$

where u_0 is an arbitrarily fixed point in U.

Topology on the space $\mathcal{R}$

$\mathcal{R}$, as a set of measures, is classically equipped with the weak topology.

Definition 3.1.4 *A sequence (q_n) in $\mathcal{R}$ is said to converge to $q \in \mathcal{R}$ if for any continuous function with compact support f on $[0, T] \times U$,*

$$\int f(t,u)\, q_n(dt,du) \to \int f(t,u)\, q(dt,du).$$

This convergence is by definition only valid on continuous functions. However, as all the measures in $\mathcal{R}$ have the same marginal on $[0, T]$ (Lebesgue measure), it is possible to considerably improve it.

Proposition 3.1.2 *Suppose $q_n \to q$ in $\mathcal{R}$.*
Then, for every measurable function $f(t,u)$ such that $\forall t \in [0,T]$, $u \to f(t,u)$ is continuous, one has

$$\int f(t,u)\, q_n(dt,du) \to \int f(t,u)\, q(dt,du).$$

(stable convergence).

Finally, the following result makes clear that the set of relaxed controls has interesting compactness properties.

Proposition 3.1.3 *Suppose U is a compact set. Then $\mathcal{R}$ is compact.*

Now, we interest to the relaxed controls of SDE as solutions of a martingale problem for a diffusion process whose infinitesimal generator is integrated against the random measures defined over the action space of all controls. Let $\mathcal{V}$ be the set of Radon measures on $[0, 1] \times \mathcal{A}$ whose projections on $[0, 1]$ coincide with the Lebesgue measure dt. Equipped with the topology of stable convergence of measures, $\mathcal{V}$ is a compact metrizable space. Stable convergence is required for bounded measurable functions $h(t, a)$ such that for each fixed $t \in [0, 1]$, $h(t, .)$ is continuous.

Definition 3.1.5 *A relaxed control is the term* $q = (\Omega, \mathcal{F}, \mathcal{F}_t, P, W_t, q_t, x(t), x_0)$ *such that*

(1) $(\Omega, \mathcal{F}, \mathcal{F}_t, P)$ *is a filtered probability space satisfying the usual conditions;*

(2) (q_t) *is an* $P(\mathcal{A})$*-valued process, progressively measurable with respect to* $(\mathcal{F}_t)$*; and such for that for each* t*,* $1_{(0,1]}.q$ *is* $\mathcal{F}_t$*-measurable;*

(3) (x_t) *is* $\mathbb{R}^n$*-valued,* $\mathcal{F}_t$*-adapted, with continuous paths, such that* $x(0) = x_0$ *and*

$$f(x_t) - f(x_0) - \int_0^t \int_A Lf(s, x_s, a)\, q_s(w, da)ds \quad \text{is a } P\text{-martingale, for each } f \in C_b^2(\mathbb{R}^n, \mathbb{R}). \tag{3.5}$$

The cost function associated to a relaxed control q is defined as

$$J(u) = E\left[\int_0^1 \int_A h(t, x_t, a)\, q_t(da)dt + g(x_1)\right]. \tag{3.6}$$

The set $\mathcal{U}$ of strict controls is embedded into the set $\mathcal{R}$ of relaxed controls by the mapping

$$\Psi : u \in \mathcal{U} \to \Psi(u)(dt, da) = dt\delta_{u(t)}(da) \in \mathcal{R};$$

where δ_u is the Dirac measure at a single point u.

3.2 Formulation of the problem

3.2.1 Predictable representation for orthogonal martingale measures

We fixed a worthy martingale measure $M(da, dt)$ over a measurable space $(E, \mathcal{E})$ defined on the stochastic base $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ where the starting $\sigma-$field $\mathcal{F}_0$ is $P-$trivial and

$$\mathcal{F} = \vee_{t\geq 0}\mathcal{F}_t.$$

The set of $(M-)$integrable function $\mathcal{P}_M$ equals the closure of simple predictable functions on $\Omega \times [0, \infty) \times E$ with respect to the norm $(.,.)_K^{1/2}$. We restrict to orthogonal martingale measures.

We denote the set of square-integrable martingales over $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ by $\mathbf{M}^2$.

Proposition 3.2.1 *Let N be in $\mathbf{M}^2$. Then there exist a unique function $n \in \mathcal{P}_M$ such that*

$$N_t = N_0 + \int_0^t \int_E n(a, s)M(da, ds) + L_t,$$

where L is an L^2-martingale with $\left\langle L, \int_0^{\cdot} \int_E b(a, s)M(da, ds)\right\rangle = 0$ for every $b \in \mathcal{P}_M$.

Proof. See L. Overberk [55]. ■

3.2.2 Representation of relaxed controls

Since the set of a strict control $\mathcal{A}$ is compact, than the relaxed control can be given in the form of Sliding control or chattering control. The Sliding control is a relaxed control defined as

$$q_t = \sum_{i=1}^{n} \alpha_i(t)\delta_{u_i(t)}, u_i(t) \in \mathcal{A}, \qquad \alpha_i(t) \geq 0, \qquad \sum_{i=1}^{n} \alpha_i(t) = 1. \tag{3.7}$$

$$\text{If } u_i(s) = \gamma \text{ in } [r, r+\theta], \text{ than: } \sum_{i=1}^{n} \alpha_i(t)\delta_{\gamma(t)} = \delta_{\gamma(t)} \tag{3.8}$$

It is not difficult to show that the solution of the (relaxed) martingale problem (3.5) is the law of the solution of the following SDE

$$dx\,(t) = \sum_{i=1}^{n} b(t, x_t, u_i(t))\alpha_i(t)dt + \sum_{i=1}^{n} \sigma(t, x_t, u_i(t))\alpha_i(t)^{1/2}dW_t^i, \qquad x\,(0) = x_0 \tag{3.9}$$

where the W^i's are n-dimensional Brownian motions on an extension of the initial probability space. The process M defined by

$$M\,(A \times [0,t]) = \sum_{i=1}^{n} \int_0^t \alpha_i(s)^{1/2}\delta_{u_i(s)}(A)dW_s^i, \tag{3.10}$$

is in fact a strongly orthogonal continuous martingale measure (c.f Walsh [59], El Karoui and Méléard [22]) with intensity

$$q_t(da)dt = \sum \alpha_i(t)\delta_{u_i(t)}\,(da)\,dt.$$

Thus, the SDE in (3.10) can be expressed in terms of M and q as follows

$$dx\,(t) = \int_A b(t, x\,(t)\,, a)q_t(da)dt + \int_A \sigma(t, x\,(t)\,, a)M(da, dt). \tag{3.11}$$

The following theorem due to El Karoui and Méléard [22] shows in fact a general representation result for solution of the martingale problem (3.5) in terms of strongly orthogonal continuous martingale measures whose intensities are our relaxed controls.

Theorem 3.2.1 *(1) Let P be the solution of the martingale problem (3.5). Then P is the law of a d-dimensional adapted and continuous process X defined on an extension of the space $(\Omega, \mathcal{F}, \mathcal{F}_t)$ and solution of the following SDE starting at x_0*

$$dX_t^i = \int_A b_i(t, X_t, a) q_t(da) dt + \sum_{k=1}^{n} \int_A \sigma_{i,k}(t, X_t, a) M^k(da, dt), \tag{3.12}$$

where $M = \left(M^k\right)_{k=1}^n$ is a family of d-strongly orthogonal continuous martingale measures with intensity $q_t(da)dt$.

(2) If the coefficients b and σ are Lipschitz in x, uniformly in t and a, the SDE (3.12) has a unique pathwise solution.

3.3 Maximum principle for relaxed control problems

In this section we establish optimality necessary conditions for relaxed control problems, where the system is described by a SDE driven by an orthogonal continuous martingale measure and the admissible controls are measure-valued processes.

Recall the controlled SDE

$$dx(t) = \int_A b(t, x(t), a) q_t(da) dt + \int_A \sigma(t, x(t), a) M(da, dt), \qquad x(0) = x_0 \tag{3.13}$$

where $M(da, dt)$ is orthogonal continuous martingale measure whose intensity is the relaxed control $q_t(da)dt$. The corresponding cost is given by

$$J(q) = E\left[\int_0^1 \int_A h\left(t, x(t), a\right) q_t(da) dt + g(x(1))\right]. \tag{3.14}$$

We assume that the coefficients of the controlled equation satisfy the following hypothesis

($\mathbf{H}_1$) $b:\mathbb{R}_+\times\mathbb{R}^n\times\mathcal{A}\to\mathbb{R}$, $\sigma:\mathbb{R}_+\times\mathbb{R}^n\times\mathcal{A}\to\mathcal{M}_{n\times m}(\mathbb{R})$, and $h:\mathbb{R}_+\times\mathbb{R}^n\times\mathcal{A}\to\mathbb{R}$ are bounded measurable in (t,x,a) and twice continuously differentiable functions in x for each (t,a) and there exists a constant $C>0$ such that

$$|f(t,x,a)-f(t,y,a)|+|f_x(t,x,a)-f_x(t,y,a)|\leq C|x-y|, \tag{3.15}$$

f and their first and second derivatives are continuous in the control variable a, where f stands for one of the functions b,σ,h.

$b,\sigma,b_x,\sigma_x,h_x,g_x$ are bounded by $C(1+|x|)$.

$g:\mathbb{R}^n\to\mathbb{R}$ is bounded and twice continuously differentiable such that

$$|g(x)-g(y)|+|g_x(x)-g_x(y)|\leq C|x-y|. \tag{3.16}$$

Under the assumptions above, the controlled equation admits a unique strong solution such that for every $p\geq 1, E\left[\sup_{0\leq t\leq T}|x_t|^p\right]<M(p)$.

3.3.1 Preliminary results

The purpose of the stochastic maximum principle is to find necessary conditions for optimality satisfied by an optimal control. Due to the appearance of the control variable in $\sigma(.,.)$, the usual first order expansion approach can't work. Hence, we introduce a second-order expansion method, we proceed as the classical maximum principle (Peng [56]).

Suppose that $(x(.),q(.))$ is an optimal solution of the problem and let us introduce the strong perturbed relaxed control in the following way

$$q_t^{\theta}(A)=\begin{cases}\delta_v(A) & \text{if } t\in E\\ q_t(A) & \text{if } t\in E^c\end{cases} \tag{3.17}$$

where $E = \{r \leq t \leq r + \theta\}$, $0 \leq r < T$ is fixed and the $E^c = [0, T] \setminus E$, $\theta > 0$ is sufficiently small, and υ is an arbitrary $\mathcal{F}_r$-measurable random variable with values in $\mathcal{U}$, such that

$$\sup_{w \in \Omega} |\upsilon(w)| < \infty.$$

Let x_θ be the trajectory of the control system (3.13) corresponding to the control $q^\theta_.(A)$, which is the intensity of the orthogonal continuous martingale measures M^θ, we create it of the form

$$M_t^\theta(A) = \int_0^t \int_A \mathbf{1}_{[r,r+\theta]}(s)\delta_\nu(da)dW_s + \int_0^t \int_A \mathbf{1}_{[r,r+\theta]^C}(s)M(da, ds). \tag{3.18}$$

where $0 \leq r < T$ is fixed, $\theta > 0$ is sufficiently small, and υ is an arbitrary $\mathcal{F}_r$-measurable random variable with values in $\mathcal{U}$.

The variational inequality will be derived from the fact that

$$\lim_{\theta \to 0} \frac{1}{\theta}\left[J(q^\theta(.)) - J(q(.))\right] \geq 0, \tag{3.19}$$

to this end, we need the following estimation.

Lemma 3.3.1 *We assume* ($\boldsymbol{H}_1$), *then the following estimate holds*

$$E\left[\sup_{0 \leq t \leq T} |x_\theta(t) - x(t) - x_1(t) - x_2(t)|^2\right] \leq C(\theta)\theta^2 \tag{3.20}$$

where $\lim_{\theta \longrightarrow 0} C(\theta) = 0$ *and* $x_1(t)$, $x_2(t)$ *are solutions of the SDEs*

$$x_1(t) = \int_0^t \int_A \left[b(s, x_s, a)q_s^\theta(da) - b(s, x_s, a)q_s(da) + b_x(s, x_s, a)x_1(s)q_s(da)\right] ds$$
$$+ \int_0^t \int_A \left[\sigma(s, x_s, a)M^\theta(da, ds) - \sigma(s, x_s, a)M(da, ds) + \sigma_x(s, x_s, a)x_1(s)M(da, ds)\right] \tag{3.21}$$

$$\begin{aligned}
x_2(t) = &\int_0^t \int_A \left[\left(b_x(s,x_s,a)q_s^\theta(da) - b_x(s,x_s,a)q_s(da)\right) x_1(s)\right] ds \\
+&\int_0^t \int_A \left[b_x(s,x_s,a)x_2(s)q_s(da) + \tfrac{1}{2}b_{xx}(s,x_s,a)q_s(da)x_1(s)x_1(s)\right] ds \\
+&\int_0^t \int_A \left[\sigma_x(s,x_s,a)x_1(s)M^\theta(da,ds) - \sigma_x(s,x_s,a)x_1(s)M(da,ds)\right] \\
+&\int_0^t \int_A \left[\sigma_x(s,x_s,a)x_2(s) + \tfrac{1}{2}\sigma_{xx}(s,x_s,a)x_1(s)x_1(s)\right] M(da,ds).
\end{aligned} \tag{3.22}$$

Remark 3.3.1 *Equation (3.21) is called the first-order variational equation. It is the variational equation in the usual sense. (3.22) is called the second-order variational equation, without this equation we can not derive the variational inequality since σ depends explicitly on the control variable.*

Notation

1) For simplicity of the notations, we denote by

$$f\left(t, x(t), q_t\right) = \int_A f\left(t, x(t), a\right) q_t(da),$$

and f stands for b, σ, h and their first and second derivatives.

2) We will generically denote by C_k the positive constants that appear in the estimates below and may differ from line to line and from proof to proof.

Proof. The proof is inspired from [60], Theorem 4.4, page 128. We need to show that

$$E\left[\sup_{0\le t\le T} |x_1(t)|^2\right] \le C_k\theta, \tag{3.23}$$

$$E\left[\sup_{0\le t\le T} |x_2(t)|^2\right] \le C_k\theta^2. \tag{3.24}$$

We can write

$$
\begin{aligned}
E\left[|x_1(t)|^2\right] &\leq 4E\left|\int_0^t\int_A\left[b(s,x_s,a)q_s^\theta(da)-b(s,x_s,a)q_s(da)\right]ds\right|^2\\
&+4E\left|\int_0^t\int_A\left[\sigma(s,x_s,a)M^\theta(da,ds)-\sigma(s,x_s,a)M(da,ds)\right]\right|^2\\
&+4E\left|\int_0^t\int_A b_x(s,x_s,a)x_1(s)q_s(da)ds+\int_0^t\int_A\sigma_x(s,x_s,a)x_1(s)M(da,ds)\right|^2\\
&\leq E(I_1)+E(I_2)+E(I_3)
\end{aligned}
$$

Since q^θ is defined as in (3.17), then

$$
\begin{aligned}
E(I_1) &\leq 4E\int_0^t\left|\int_A\left[b(s,x_s,a)\delta_\upsilon(da)-b(s,x_s,a)q_s(da)\right]1_E\right|^2 ds\\
&\leq C_kE\int_r^{r+\theta}\left[|b(s,x_s,\upsilon)|^2+\int_A|b(s,x_s,a)|^2|q_s(da)|^2\right]ds\\
&\leq C_k\int_r^{r+\theta}E\left[1+|x(t)|^2\right]ds\\
&\leq C_k\int_r^{r+\theta}\left[1+E\left(\sup_{0\leq t\leq T}|x(t)|^2\right)\right]ds\leq C_k(1+\alpha)\theta.
\end{aligned}
$$

$$
\begin{aligned}
E(I_2) &\leq C_kE\left|\int_r^{r+\theta}\int_A\left[\sigma(s,x_s,a)\delta_\upsilon(da)dB_s-\sigma(s,x_s,a)M(da,ds)\right]\right|^2\\
&\leq C_kE\int_r^{r+\theta}\left[|\sigma(s,x_s,\upsilon)|^2ds+\int_A|\sigma(s,x_s,a)|^2q_s(da)ds\right]\\
&\leq C_k\int_r^{r+\theta}\left[1+E\left(\sup_{0\leq t\leq T}|x(t)|^2\right)\right]ds\leq C_k(1+\alpha)\theta
\end{aligned}
$$

$$
\begin{aligned}
E(I_3) &\leq C_kE\left[\int_0^t\int_A|b_x(s,x_s,a)|^2|x_1(s)|^2|q_s(da)|^2ds+\int_0^t\int_A|\sigma_x(s,x_s,a)||x_1(s)|^2q_s(da)ds\right]\\
&\leq C_kE\left(\int_0^t|x_1(s)|^2ds\right)\leq C_k\int_0^t E|x_1(s)|^2ds
\end{aligned}
$$

Then, we have

$$
E|x_1(s)|^2\leq C_kE\left(\int_0^t|x_1(s)|^2ds\right)+C_k(1+\alpha)\theta
$$

By Gronwall Lemma and Burkholder-Davis-Gundy's inequality, we have

$$E\left[\sup_{0\leq t\leq T}|x_1(t)|^2\right]\leq C_k\theta.$$

As precedently, we have

$$\begin{aligned}E\left[|x_2(t)|^2\right]&\leq 6E\left[\int_0^t\int_A[|b_x(s,x_s,a)x_2(s)q_s(da)ds|+|\sigma_x(s,x_s,a)x_2(s)M(da,ds)|]\right]^2\\&+3E\left[\int_0^t\int_A[|b_{xx}(s,x_s,a)x_1(s)x_1(s)|\,q_s(da)ds+|\sigma_{xx}(s,x_s,a)x_1(s)x_1(s)|\,M(da,ds)]\right]^2\\&+6E\int_0^t\left(\int_A\left|b_x(s,x_s,a)x_1(s)q_s^\theta(da)-b_x(s,x_s,a)x_1(s)q_s(da)\right|\right)^2ds\\&+6E\left[\int_0^t\int_A\left|\sigma_x(s,x_s,a)x_1(s)M^\theta(da,ds)-\sigma_x(s,x_s,a)x_1(s)M(da,ds)\right|\right]^2\end{aligned}$$

by (3.17), we have

$$\begin{aligned}E\left|x_2(s)^2\right|&\leq C_k\left(2\int_0^t E|x_2(s)|^2ds+4\int_r^{r+\theta}\theta ds+\int_0^t\theta^2ds\right)\\&\leq C_k\left(2\int_0^t E|x_2(s)|^2ds+4\int_r^{r+\theta}\theta ds+\int_0^T\theta^2ds\right)\\&\leq C_k\int_0^t E|x_2(s)|^2ds+C_k(4+T)\theta^2\end{aligned}$$

by Gronwall's and Burkholder-Davis-Gundy's inequalities, we obtained the inequalities (3.21), (3.22).

As in the proof of Lemma 1 in [56], set $x_3=x_1+x_2$, we have

$$\begin{aligned}b\left(t,x(t)+x_3(t),q_t^\theta\right)&=b\left(t,x(t),q_t^\theta\right)+b_x\left(t,x(t),q_t^\theta\right)x_3(t)\\&\quad+\int_0^1\int_0^1\lambda b_{xx}\left(t,x(t)+\lambda\theta x_3(t),q_t^\theta\right)d\lambda d\theta x_3(t)x_3(t)\\\sigma\left(t,x(t)+x_3(t),q_t^\theta\right)&=\sigma\left(t,x(t),q_t^\theta\right)+\sigma_x\left(t,x(t),q_t^\theta\right)x_3(t)\\&\quad+\int_0^1\int_0^1\lambda\sigma_{xx}\left(t,x(t)+\lambda\theta x_3(t),q_t^\theta\right)d\lambda d\theta x_3(t)x_3(t)\end{aligned}$$

then, we can write

$$\int_0^t b\left(s, x(s)+x_3(s), q_s^\theta\right) ds+\int_0^t \int_A \sigma\left(s, x(s)+x_3(s), a\right) M^\theta(da, ds)$$
$$=x(t)+x_1(t)+x_2(t)-x\left(0\right)+\int_0^t B^\theta(s) ds+\Lambda^\theta(t)$$

where

$$\begin{aligned}
B^\theta(s)= & \frac{1}{2} b_{xx}\left(s, x(s), q_s\right)\left(x_2(s) x_2(s)+2 x_1(s) x_2(s)\right) \\
& +\left(b_x\left(s, x(s), q_s^\theta\right)-b_x\left(s, x(s), q_s\right)\right) x_2(s) \\
& +\int_0^1 \int_0^1\left\{\left[\lambda b_{xx}\left(s, x(s)+\lambda \theta\left(x_1(s)+x_2(s)\right), q_t^\theta\right)\right] d \lambda d \theta\right. \\
& \left.\qquad\left(x_1(s)+x_2(s)\right)\left(x_1(s)+x_2(s)\right)\right\} \\
& -\int_0^1 \int_0^1\left\{\left[b_{xx}\left(s, x(s), q_s\right)\right] d \lambda d \theta\right. \\
& \left.\qquad\left(x_1(s)+x_2(s)\right)\left(x_1(s)+x_2(s)\right)\right\}
\end{aligned}$$

$$\begin{aligned}
\Lambda^\theta(t)= & \frac{1}{2} \int_0^t \int_A \sigma_{xx}\left(s, x(s), a\right)\left(x_2(s) x_2(s)+2 x_1(s) x_2(s)\right) M(da, ds) \\
& +\int_0^t \int_A \sigma_x\left(s, x(s), a\right) x_2(s) M^\theta(da, ds) \\
& -\int_0^t \int_A \sigma_x\left(s, x(s), a\right) x_2(s) M(da, ds) \\
& +\int_0^t \int_A \int_0^1 \int_0^1\left\{\lambda \sigma_{xx}\left(s, x(s)+\lambda \theta\left(x_1(s)+x_2(s)\right), a\right) d \lambda d \theta\right. \\
& \left.\qquad\left(x_1(s)+x_2(s)\right)\left(x_1(s)+x_2(s)\right)\right\} M^\theta(da, ds) \\
& -\int_0^t \int_A \int_0^1 \int_0^1\left\{\sigma_{xx}\left(s, x(s), a\right) d \lambda d \theta\left(x_1(s)+x_2(s)\right)\right. \\
& \left.\qquad\left(x_1(s)+x_2(s)\right)\right\} M(da, ds)
\end{aligned}$$

and we can drive

$$\begin{aligned} x_\theta(t) - x(t) - x_1(t) - x_2(t) = & \int_0^t \int_A \left[b\left(s, x_\theta(s), a\right) - b\left(s, x(s) + x_1(s) + x_2(s), a\right) \right] q_\theta(da)ds \\ & + \int_0^t \int_A \left[\sigma\left(s, x_\theta(s), a\right) - \sigma\left(s, x(s) + x_1(s) + x_2(s), a\right) \right] M^\theta(da, ds) \\ & + \int_0^t B^\theta(s)ds + \Lambda^\theta(t) \end{aligned}$$

since b and σ are Lipschitz then

$$\begin{aligned} E\left[|x_\theta(t) - x(t) - x_1(t) - x_2(t)|^2 \right] \leq & C_k \int_0^t E\left|x_\theta(t) - x(t) - x_1(t) - x_2(t)\right|^2 ds \\ & + C_k \int_0^t E\left|B^\theta(s)\right|^2 ds + C_k E\left|\Lambda^\theta(t)\right|^2 \end{aligned}$$

and since b and b_{xx} are bounded

$$\begin{aligned} E\left[\left|B^\theta(s)\right|^2 \right] \leq & C_k E\left|x_2(s)x_2(s)\right|^2 + C_k E\left|x_1(s)x_2(s)\right|^2 + C_k E\left|x_2(s)\right|^2 \\ & + C_k E\left|\left(x_1(s) + x_2(s)\right)\left(x_1(s) + x_2(s)\right)\right|^2 \end{aligned}$$

from (3.23) and (3.24), we can use Cauchy-schwarz's inequality, we have

$$E\left|B^\theta(s)\right|^2 \leq C_k\left(\theta^4 + \theta^3 + \theta^2\right).$$

Using the same think for Λ^θ, since σ_x and σ_{xx} are bounded and $\forall t > 0$, $M_t(.)$ is a L^2-valued $\sigma-$finite measure than

$$E\left|\Lambda^\theta(s)\right|^2 \leq C_k\left(\theta^4 + \theta^3 + \theta^2\right).$$

From the two last inequalities, we can conclude that

$$E\left[|x_\theta(t) - x(t) - x_1(t) - x_2(t)|^2 \right] \leq C_k \int_0^t E\left|x_\theta(t) - x(t) - x_1(t) - x_2(t)\right|^2 ds + C_k\left(\theta^4 + \theta^3 + \theta^2\right)$$

by Gronwall's lemma, we obtained the inequality

$$E\left[|x_\theta(t)-x(t)-x_1(t)-x_2(t)|^2\right]\leq C_k\left(\theta^4+\theta^3+\theta^2\right)\exp\left(C_kT\right).$$

Then, (3.20) follows from Burkholder-Davis-Gundy's inequalities. ■

We want now to derive a variational inequality which is become from the Taylor expansion and the cost functional with respect to the perturbation of the control variable.

Since q is an optimal relaxed control and from Lemma 3.3.1 we can derive.

Lemma 3.3.2 *Under (3.19), the assumption of Lemma 3.3.1, we have*

$$\begin{aligned}0\leq J\left(q^\theta\right)-J\left(q\right)\leq E&\left[\int_0^T\left(h\left(t,x(t),q_\theta(t)\right)-h\left(t,x(t),q(t)\right)\right)dt\right]\\+E&\left[g_x(x(T))\left(x_1(T)+x_2(T)\right)+\int_0^T h_x\left(t,x_t,q(t)\right)\left(x_1(t)+x_2(t)\right)dt\right]\\+\tfrac{1}{2}E&\left[g_{xx}(x(T))x_1(T)x_1(T)+\int_0^T h_{xx}\left(t,x(t),q(t)\right)x_1(t)x_1(t)dt\right]+o\left(\theta\right)\end{aligned}\tag{3.25}$$

Proof. Since (x,q) is optimal, we have

$$0\leq E\left[\int_0^T\left(h\left(t,x_\theta(t),q_\theta(t)\right)-h\left(t,x(t),q(t)\right)\right)dt\right]+E\left[g(x_\theta(T))-g(x(T))\right]$$

we use (3.21)

$$\begin{aligned}0\leq E\int_0^T&\left[h\left(t,x(t)+x_1(t)+x_2(t),q_\theta(t)\right)-h\left(t,x(t),q(t)\right)\right]dt\\&+E\left[g(x(T)+x_1(T)+x_2(T))-g(x(T))\right]+\circ\left(\theta\right)\end{aligned}\tag{3.26}$$

Then by Taylor expansion in the point x for $h\left(t,x+x_1+x_2,q_\theta\right)$ and $g\left(x+x_1+x_2\right)$, we have by (3.23) and (3.24), (3.26) can be rewritten as

$$\begin{aligned} 0 \le o(\theta) + \alpha(T) + E\int_0^T \left[h\left(t, x(t), q_\theta(t)\right) - h\left(t, x(t), q(t)\right)\right] dt \\ +E\int_0^T \left[h_x\left(t, x(t), q(t)\right)\left(x_1(t) + x_2(t)\right)\right] dt \\ +\tfrac{1}{2}E\int_0^T \left[h_{xx}\left(t, x(t), q(t)\right) x_1(t)x_1(t)\right] dt \\ +E\left[g_x(x(T)\left(x_1(T) + x_2(T)\right)\right] + \tfrac{1}{2}E\left[g_{xx}(x(T))x_1(T)x_1(T)\right] \end{aligned} \tag{3.27}$$

where $\alpha(T)$ is given by

$$\begin{aligned} \alpha(T) = E\int_0^T \left[h_x\left(t, x(t), q_\theta(t)\right) - h_x\left(t, x(t), q(t)\right)\right]\left(x_1(t) + x_2(t)\right) dt \\ +\tfrac{1}{2}E\int_0^T h_{xx}\left(t, x(t), q(t)\right)\left(x_1(t)x_2(t) + x_2(t)x_1(t) + x_2(t)x_2(t)\right) dt \\ +\tfrac{1}{2}E\int_0^T \left[\left(h_{xx}\left(t, x(t), q_\theta(t)\right) - h_{xx}\left(t, x(t), q(t)\right)\right)\left(x_1(t) + x_2(t)\right)\left(x_1(t) + x_2(t)\right)\right] dt \\ +\tfrac{1}{2}E\left[g_{xx}(x(T))\left(x_1(T)x_2(T) + x_2(T)x_1(T) + x_2(T)x_2(T)\right)\right] \end{aligned}$$

from q_θ definition and $(\mathbf{H}_1)$ assumption, we use (3.23), (3.24) and Cauchy-Schwarz's inequalities, then

$$\alpha(T) \le o(\theta)$$

Use this relation and (3.27) to complete the proof. ■

3.3.2 Adjoint processes and variational inequality

In this subsection, we will introduce the first and second order adjoint processes involved in the stochastic maximum principle and the associated stochastic Hamiltonian system. These are obtained from the first and second varaitional equations (3.21), (3.22) as well as (3.25).

First order terms

The first order estimation calculate the first order derivatives in (3.25). The linear term in (3.21) and (3.22) may treated in the following way (see [14]). Let ϕ_1 be the fundamental solution of the linear equation

$$\begin{cases} d\phi_1(t) = \int_A b_x(t,x(t),a)\phi_1(t)q_t(da)dt + \int_A \sigma_x(t,x(t),a)\phi_1(t)M(da,dt) \\ \phi_1(0) = I_d. \end{cases}$$

This equation is linear with bounded coefficients, then it have a strong unique solution. Moreover ϕ_1 is invertible and it inverse ψ_1 satisfies

$$\begin{cases} d\psi_1(t) = \int_A \left[\psi_1(t)\sigma_x(t,x(t),a)\sigma_x(t,x(t),a) - \psi_1(t)b_x(t,x(t),a)\right] q_t(da)dt \\ \qquad\qquad - \int_A \psi_1(t)\sigma_x(t,x(t),a)M(da,dt) \\ \psi_1(0) = I_d. \end{cases}$$

ϕ_1 and ψ_1 satisfy

$$E\left[\sup_{t\in[0,T]} \left|\phi_1(t)\right|^2\right] + E\left[\sup_{t\in[0,T]} \left|\psi_1(t)\right|^2\right] < \infty.$$

We introduce the following processes

$$\eta_1(t) = \psi_1(t)\left(x_1(t) + x_2(t)\right),$$

and

$$X_1 = \phi_1(T) g_x\left(x(T)\right) + \int_0^T \phi_1(s) \int_A h_x(s, x(s), a) q_s(da) ds$$
$$\zeta_1(t) = E\left(X_1/\mathcal{F}_t\right) - \int_0^t \phi_1(s) \int_A h_x(s, x(s), a) q_s(da) ds$$

then

$$E\left[g_x(x(T))\left(x_1(T) + x_2(T)\right)\right] = E\left[\phi_1(T) g_x\left(x(T)\right) \eta_1(T)\right] = E\left[\eta_1(T)\zeta_1(T)\right]$$

from the orthogonal martingale measure representation (Proposition 3.2.1) we have

$$E\left(X_1/\mathcal{F}_t\right) = E\left(X_1\right) + \int_0^t \int_A G_1(a, s) M\left(da, ds\right) + L_t,$$

where L is an L^2-martingale with $\left\langle L, \int_0^{\cdot} \int_E b(a, s) M(da, ds) \right\rangle = 0$ for every $b \in \mathcal{P}_M$ and such that $E\left[\langle L_t \rangle\right] < \infty$.

Applied Ito's formula to $\eta_1(t)\zeta_1(t)$ and we put

$$p_1(t) = \psi_1^*(t)\zeta_1(t), \tag{3.28}$$

$$Q_1(t) = \int_A \psi_1^*(t) G_1(t, a) q_t(da) - \int_A \sigma_x^*(t, x(t), a) q_t(da)) p_1(t) \tag{3.29}$$

moreover $p_1(t)$, $Q_1(t)$ satisfy

$$E\left[\sup_{0 \le t \le T} \left|p_1(t)\right|^2 + \sup_{0 \le t \le T} \left|Q_1(t)\right|^2\right] < \infty,$$

the process p_1 is called the first adjoint process.

We can derive

$$
\begin{aligned}
E\left[g_x(x(T))\left(x_1(T)+x_2(T)\right)\right] = & E\int_0^T\int_A p_1(t)\left(b(t,x(t),a)q_t^\theta(da)-b(t,x(t),a)q_t(da)\right)dt \\
& + E\int_0^T\int_A \left[Q_1(t)\left(\sigma(t,x(t),a)q_t^\theta(da)-\sigma(t,x(t),a)q_t(da)\right)\right]dt \\
& + \frac{1}{2}E\int_0^T\int_A p_1(t)b_{xx}(t,x(t),a)x_1(t)x_1(t)q_t(da)dt \\
& + \frac{1}{2}E\int_0^T\int_A Q_1(t)\sigma_{xx}(t,x(t),a)x_1(t)x_1(t)q_t(da)dt \\
& - E\int_0^T\int_A h_x(t,x(t),a)\left(x_1(t)+x_2(t)\right)q_t(da)dt \\
& + E\int_0^T\int_A p_1(t)\left[\left(b_x(t,x(t),a)q_t^\theta(da)-b_x(t,x(t),a)q_t(da)\right)\right]x_1(t)dt \\
& + E\int_0^T\int_A \left[Q_1(t)\left(\sigma_x(t,x(t),a)q_t^\theta(da)-\sigma_x(t,x(t),a)q_t(da)\right)\right]x_1(t)dt \\
& - E\int_0^T\int_A Q_1(t)\left[\sigma(t,x(t),a)+\sigma_x(t,x(t),a)x_1(t)\right]\delta_\nu(da)1_E(t)dt \\
& + E\int_0^T\int_A \psi_1(t)\left[\sigma(t,x(t),a)+\sigma_x(t,x(t),a)x_1(t)\right]\delta_\nu(da)1_E(t)d\left\langle B_t,L_t\right\rangle \\
& + o(\theta).
\end{aligned}
$$

To derive our variational inequality, we need to prove the following estimates in the last equality

$$
\begin{aligned}
E\int_0^T\int_A p_1(t)\left[\left(b_x(t,x(t),a)q_t^\theta(da)-b_x(t,x(t),a)q_t(da)\right)\right]x_1(t)dt &\leq C\theta, \\
E\int_0^T\int_A \left[Q_1(t)\left(\sigma_x(t,x(t),a)q_t^\theta(da)-\sigma_x(t,x(t),a)q_t(da)\right)\right]x_1(t)dt &\leq C\theta, \\
E\int_0^T\int_A Q_1(t)\left[\sigma(t,x(t),a)+\sigma_x(t,x(t),a)x_1(t)\right]\delta_\nu(da)1_E(t)dt &\leq C\theta
\end{aligned}
$$

and

$$
E\int_0^T\int_A \psi_1(t)\left[\sigma(t,x(t),a)+\sigma_x(t,x(t),a)x_1(t)\right]\delta_\nu(da)1_E(t)d\left\langle B_t,L_t\right\rangle \leq C\theta.
$$

By (3.23) and applying Young's inequality, the first inequality becomes

$$E\int_r^{r+\theta} p_1(t)\left[b_x(t,x(t),\nu)-b_x(t,x(t),q)\right]x_1(t)dt \leq C_k E\int_r^{r+\theta}\left\{\left[p_1(t)x_1(t)\right]^2+\left[b_x(t,x(t),\nu)-b_x(t,x(t),q)\right]^2\right\}dt$$
$$\leq C_k\left(\theta+E\int_r^{r+\theta}\left[1+\sup_{0\leq t\leq T}|x(t)|^2\right]dt\right)\leq C_k\theta$$

For the second and the third estimates, we use the same argument as in the first one. For the fourth term we use Kunita-Watanabe inequality

$$E\left[\int_r^{r+\theta}\int_A \psi_1(t)\left[\sigma(t,x(t),a)+\sigma_x(t,x(t),a)x_1(t)\right]\delta_\nu(da)d\left\langle B_t,L_t\right\rangle\right]\leq$$
$$E\left(\int_r^{r+\theta}\int_A \psi_1^2(t)\left[\sigma(t,x(t),a)+\sigma_x(t,x(t),a)x_1(t)\right]^2\delta_\nu(da)dt\right)^{1/2}E\left(\int_r^{r+\theta}\int_A \delta_\nu(da)d\left\langle L_t,L_t\right\rangle\right)^{1/2}$$
$$\leq C_k E\left(\int_r^{r+\theta}\psi_1^2(t)\left[\sigma^2(t,x(t),\nu)+\sigma_x^2(t,x(t),\nu)x_1^2(t)\right]dt\right)^{1/2}E\left(\langle L,L\rangle_{r+\theta}-\langle L,L\rangle_r\right)^{1/2}$$

Using the same arguments the inequality holds since $E\left[\langle L_t\rangle\right]<\infty$.

Let us now define the Hamiltonian

$$H(t,x,q,p,Q)=\int_A h(t,x,a)\,q(da)+p\int_A b(t,x,a)\,q(da)+Q\int_A \sigma(t,x,a)\,q(da),$$

Therefore, we use the value of $E\left[g_x(x(T))\left(x_1(T)+x_2(T)\right)\right]$ and the Hamiltonian definition, (3.25) can be rewritten

$$0\leq J\left(q^\theta\right)-J(q)\leq E\int_0^T\int_A\left[H\left(t,x(t),a,p_1(t),Q_1(t)\right)q_t^\theta(da)-H\left(t,x(t),a,p_1(t),Q_1(t)\right)q_t(da)\right]dt$$
$$+\tfrac{1}{2}E\int_0^T\int_A x_1(t)H_{xx}(x(t),a,p_1(t),Q_1(t))x_1^*(t)q_t(da)dt+\tfrac{1}{2}E\left[x_1(T)g_{xx}(x(T))x_1^*(T)\right]+o(\theta). \tag{3.30}$$

Second order terms

The second order estimation concerns the second order derivatives in (3.30). As in Peng [56], let $Z = x_1 x_1^*$. By Itô's formula we obtain

$$\begin{aligned} dZ(t) &= \int_A \left[Z(t)b_x^*(t,x(t),a) + b_x(t,x(t),a)Z(t)\right] q_t(da)dt + \mathbb{B}_\theta\left(t,x(t),a\right) \\ &+ \int_A \sigma_x(t,x(t),a)Z(t)\sigma_x^*(t,x(t),a)q_t(da)dt + \mathbb{A}_\theta\left(t,x(t),a\right)dt \\ &+ \int_A \left(Z(t)\sigma_x^*(t,x(t),a) + \sigma_x(t,x(t),a)Z(t)\right) M\left(da,dt\right) - \mathbb{B}\left(t,x(t),a\right). \end{aligned} \tag{3.31}$$

For simplicity of notations, we denote by

$$f\left(t\right) = \int_A f\left(t,x(t),a\right) q_t(da),\;\; f_\theta\left(t\right) = \int_A f\left(t,x(t),a\right) q_t^\theta(da)$$

in $\mathbb{A}_\theta$ and in $\mathbb{B}_\theta$, $\mathbb{B}$ by

$$f dM = \int_A f\left(t,x(t),a\right) M\left(da,dt\right),\;\; f_\theta dM^\theta = \int_A f\left(t,x(t),a\right) M^\theta\left(da,dt\right)$$

f stands for b,σ and their first derivatives.

Then we have

$$\begin{aligned} &\mathbb{A}_\theta\left(t\right) q_t(da) = x_1(t)\left(b_\theta^*(t) - b^*(t)\right) + \left(b^\theta(t) - b(t)\right) x_1^*(t) - \sigma_x(t)x_1(t)\sigma^*(t) - \sigma(t)x_1^*(t)\sigma_x^*(t) \\ &\quad + \left[\left(\sigma_x(t)x_1(t)\sigma_\theta^*(t) + \sigma_\theta(t)x_1^*(t)\sigma_x^*(t)\right) - \left(\sigma_\theta(t)\sigma^*(t) + \sigma(t)\sigma_\theta^*(t)\right)\right] 1_{E^C}\left(t\right) + \sigma_\theta(t)\sigma_\theta^*(t) + \sigma(t)\sigma^*(t) \\ &\mathbb{B}_\theta\left(t\right) = \sigma_\theta(t)x_1^*(t) + x_1(t)\sigma_\theta^*(t)dM^\theta, \qquad \mathbb{B}\left(t\right) = \sigma(t)x_1^*(t) + x_1(t)\sigma^*(t)dM \end{aligned}$$

we remark that

$$\begin{aligned} &E\int_0^T \mathbb{A}_\theta\left(t\right)dt \le E\int_0^T \left[\left(\sigma_\theta(t)\sigma_\theta^*(t) + \sigma(t)\sigma^*(t)\right) - \left(\sigma_\theta(t)\sigma^*(t) + \sigma(t)\sigma_\theta^*(t)\right) 1_{E^C}\left(t\right)\right] dt + \circ\left(\theta\right) \\ &E\int_0^T \mathbb{B}_\theta\left(t\right) dM^\theta \le \circ\left(\theta\right) \text{ and } E\int_0^T \mathbb{B}_\theta\left(t\right) dM \le \circ\left(\theta\right). \end{aligned}$$

As in the first order estimation, we consider now the following symmetric matrix-valued linear equation associate to (3.31)

$$\begin{cases} d\phi_2(t) = \int_A \left[\phi_2(t)b_x^*(t,x(t),a) + b_x(t,x(t),a)\phi_2(t) + \sigma_x(t,x(t),a)\phi_2(t)\sigma_x^*(t,x(t),a)\right] q_t(da)dt \\ \qquad\qquad + \int_A \left(\phi_2(t)\sigma_x^*(t,x(t),a) + \sigma_x(t,x(t),a)\phi_2(t)\right) M(da,dt) \\ \phi_2(0) = I_d. \end{cases}$$

This equation is linear with bounded coefficients, hence it admit a unique strong solution. Moreover ϕ_2 is invertible and it inverse ψ_2 satisfies

$$\begin{cases} d\psi_2(t) = \int_A \left[\left(\sigma_x(t,x(t),a) + \sigma_x^*(t,x(t),a)\right)^2 \psi_2(t) - \psi_2(t)b_x^*(t,x(t),a)\right] q_t(da)dt \\ \qquad - \int_A \left[b_x(t,x(t),a)\psi_2(t) + \sigma_x(t,x(t),a)\psi_2(t)\sigma_x^*(t,x(t),a)\right] q_t(da)dt \\ \qquad - \left[\psi_2(t)\sigma_x^*(t,x(t),a) + \sigma_x(t,x(t),a)\psi_2(t)\right] M(da,dt) \\ \psi_2(0) = I_d. \end{cases}$$

It is easy to see that ϕ_2 and ψ_2 satisfy

$$E\left[\sup_{t\in[0,T]} \left|\phi_2(t)\right|^2\right] + E\left[\sup_{t\in[0,T]} \left|\psi_2(t)\right|^2\right] < \infty.$$

Using the same arguments as for the first order terms, we introduce the processes $\eta_2(t) = \psi_2(t)Z(t)$ and

$$X_2 = \phi_2^*(T)g_{xx}(x(T)) + \int_0^T \phi_2^*(s)\int_A H_{xx}(s,x(s),a)q_s(da)ds$$
$$\zeta_2(t) = E\left(X_2/\mathcal{F}_t\right) - \int_0^t \phi_2^*(s)\int_A H_{xx}(s,x(s),a)q_s(da)ds$$

We remark from these equality that

$$E\left[x_1(T)g_{xx}(x(T))x_1^*(T)\right] = E\left[\phi_2^*(T)g_{xx}(x(T))\eta_2(T)\right] = E\left[\eta_2(T)\zeta_2(T)\right]$$

The orthogonal martingale measure representation (Proposition 3.2.1) give us

$$E\left(X_2/\mathcal{F}_t\right) = E\left(X_2\right) + \int_0^t \int_A G_2(a,s) M(da,ds) + L'_t \tag{3.32}$$

where L' is an L^2-martingale with $\left\langle L', \int_0^{\cdot} \int_E b(a,s) M(da,ds) \right\rangle = 0$ for every $b \in \mathcal{P}_M$ and such that $E\left[\langle L'_t \rangle\right] < \infty$.

Apply Itô's formula to $\eta_2(t)\zeta_2(t)$, to obtain

$$\begin{gathered} E\left[x_1(T) g_{xx}(x(T)) x_1^*(T)\right] = -E\int_0^T \int_A x_1(t) H_{xx}(t,x(t),a) x_1^*(t) q_t(da) dt + \\ E\int_0^T \int_A tr\left[\left(\sigma(t,x(t),a) q_t^\theta(da) - \sigma(t,x(t),a) q_t(da)\right)^* p_2(t) \left(\sigma(t,x(t),a) q_t^\theta(da) - \sigma(t,x(t),a) q_t(da)\right)\right] dt \\ + o\left(\theta\right) \end{gathered} \tag{3.33}$$

where

$$p_2(t) = \psi_2^*(t) \zeta_2(t) \tag{3.34}$$

the process p_2 is called the second adjoint process.

Adjoint equations and the maximum principle

By applying Ito's formula to the adjoint processes p_1 in (3.28) and p_2 in (3.34), we obtain the first and second order adjoint equations, which have the forms

$$\begin{cases} -dp_1(t) = \displaystyle\int_A \left[b_x^*\left(t,x(t),a\right) p_1(t) + \sigma_x^*\left(t,x(t),a\right) Q_1(t) + h_x\left(t,x(t),a\right)\right] q_t(da) dt \\ \qquad\qquad\qquad - \displaystyle\int_A Q_1(t) M(da,dt) - \psi_1^*(t) dL_t \\ p_1(T) = g_x\left(x(T)\right). \end{cases} \tag{3.35}$$

with values in $\mathbb{R}^d$, where L is an L^2-martingale with $\left\langle L, \int_0^{\cdot} \int_E b(a,s) M(da,ds) \right\rangle = 0$ for every $b \in \mathcal{P}_M$, Q_1 is given by (3.29) with values in $\mathbb{R}^{d\times k}$. The adjoint equation that $p_1(.)$ satisfies is a linear backward stochastic differential equation. This BSDE has a unique adapted solution see Elkaroui, Peng and Quenez [25].

Using Itô's formula it is easy to see that p_2 is matrix valued and satisfies

$$\begin{cases} -dp_2(t) = \int_A \{b_x^*(t,x(t),a)\, p_2(t) + p_2(t) b_x(t,x(t),a) \\ \qquad\qquad +\sigma_x^*(t,x(t),a)\, Q_2(t) + Q_2(t)\sigma_x(t,x(t),a)\}\, q_t(da)dt \\ \qquad\qquad + \int_A \left[\sigma_x^*(t,x(t),a)\, p_2(t)\sigma_x(t,x(t),a) + H_{xx}(x(t),a,p_1(t),Q_1(t))\right] q_t(da)dt \\ \qquad\qquad - \int_A Q_2(t) M(da,dt) - \psi_2^*(t) dL_t' \\ p_2(T) = g_{xx}(x(T)), \end{cases} \tag{3.36}$$

where L' is given by (3.32) and Q_2 is given by

$$Q_2(t) = \int_A \left[\psi_2^*(t) G_2(t,a) - p_2(t)\sigma_x(t,x(t),a) + \sigma_x^*(t,x(t),a) p_2(t)\right] q_t(da) \tag{3.37}$$

Note that $p_2(.)$ is also a backward stochastic differential equation with matrix-valued unknowns. This BSDE have a unique adapted solution.

Remark 3.3.2 *$H_{xx}(x(t),q_t,p(t),Q(t))$ is the second derivative of the Hamiltonian H at x and it is given by*

$$H_{xx}(x(t),q_t,p(t),Q(t)) = h_{xx}(t,x(t),q_t) + p(t) b_{xx}(t,x(t),q_t) + Q(t)\sigma_{xx}(t,x(t),q_t).$$

We are ready now to state the main result.

Theorem 3.3.1 (*The stochastic maximum principle*) *Let q be an optimal control minimizing the cost J over $\mathcal{R}$ and x denotes the corresponding optimal trajectory. Then there are two unique couples of adapted processes (p_1, Q_1) and (p_2, Q_2) which are respectively solutions of the backward stochastic differential equations (3.35) and (3.36) such that*

$$\begin{aligned} 0 \le\ & H(t,x(t),\nu,p_1(t),Q_1(t)) - H(t,x(t),q_t,p_1(t),Q_1(t)) \\ & + \tfrac{1}{2} tr\left[(\sigma(t,x(t),\nu) - \sigma(t,x(t),q_t))^* p_2(t) (\sigma(t,x(t),\nu) - \sigma(t,x(t),q_t))\right] \end{aligned} \tag{3.38}$$

υ is an arbitrary $\mathcal{F}_r$-measurable random variable with values in $\mathcal{U}$, such that

$$\sup_{w\in\Omega} |\upsilon(w)| < \infty.$$

Proof. From (3.33), (3.30) can be rewritten

$$\begin{aligned} 0 \leq J\left(q^\theta\right) - J\left(q\right) \leq E\int_0^T \int_A \left[H\left(t, x(t), a, p_1(t), Q_1(t)\right) q_t^\theta(da) - H\left(t, x(t), a, p_1(t), Q_1(t)\right) q_t(da)\right] dt \\ + \tfrac{1}{2} E\int_0^T \int_A tr\left\{\left(\sigma^\theta(t, x(t), a) q_t^\theta(da) - \sigma(t, x(t), a) q_t(da)\right)^* p_2(t)\right. \\ \left.\left(\sigma^\theta(t, x(t), a) q_t^\theta(da) - \sigma(t, x(t), a) q_t(da)\right)\right\} dt + \circ\left(\theta\right). \end{aligned}$$

This equation is the variational inequation of the second order.

We use the definition of q_θ, the last variational inequality becomes

$$\begin{aligned} 0 \leq \tfrac{1}{\theta}\left(J\left(q^\theta\right) - J\left(q\right)\right) \leq \tfrac{1}{\theta} E\int_r^{r+\theta} \left[H\left(t, x(t), \nu, p_1(t), Q_1(t)\right) - H\left(t, x(t), q_t, p_1(t), Q_1(t)\right)\right] dt + o\left(\theta\right) \\ + \tfrac{1}{2\theta} E\int_r^{r+\theta} tr\left[\left(\sigma(t, x(t), \nu) - \sigma\left(t, x(t), q_t\right)\right)^* p_2(t)\left(\sigma(t, x(t), \nu) - \sigma(t, x(t), q_t)\right)\right] dt, \end{aligned}$$

Then, the desired result follows by letting θ going to zero. ∎

Conclusion

The original version of Pontryagin's maximum principle was for deterministic problems, with its key idea coming from the classical calculus of variations. In deriving the maximum principle, one first slightly perturbs an optimal control by means of the so-called spike variation, then considers the first-order term in a sort of Taylor expansion with respect to this perturbation. By sending the perturbation to zero, one obtains a kind of variational inequality. The final desired result (the maximum principle) then follows from the duality. If the diffusion terms also depend on the controls, we encounter an essential difficulty, we we try to do the same idea for control problems, thus the usual first-order variation method can not applied.

To surpass this difficulty, one needs to study both the first-order and second-order terms in the Taylor expansion of the spike variation and find a stochastic maximum principle involving a stochastic Hamiltonian system that consists of two forward-backward stochastic differential equations. Our result extends Peng's maximum principle to the class of measure valued controls.

Bibliography

[1] N. U. Ahmed, C. D. Charalambous, *Stochastic minimum principle for partially observed systems subject to continuous and jump diffusion processes and driven by relaxed controls, SIAM J.* Control and Optim., Vol. 51 (2013), no. 4, pp. 3235-3257.

[2] V. I. Arkin, M. T. Saksonov, *Nessary optimality conditions for stochastic differential equation,* Soviet. Math. Dokl. 20, pp 1-5, 1979.

[3] S. Bahlali, *Necessary and sufficient optimality conditions for relaxed and strict control problems.* SIAM J. Control Optim. 47, no. 4, 2078–2095, 2008.

[4] S. Bahlali, Doctoral thesis, Institut des Sciences Exactes, Université de Batna, 2002.

[5] S. Bahlali, B. Mezerdi, *Approximation in relaxed stochastic optimal control problems.* Conférence internationale sue les mathématiques appliquées et les Sciences de lingénieur, Casablanca 22-25 October 1998, pp. 274-277.

[6] S. Bahlali, B. Mezerdi, *Approximation in optimal control of diffusion processes, Rand. Operat. and Stocha. Equ. Vol. 8, N° 4, pp. 365-372, 2000.*

[7] S. Bahlali, B. Mezerdi, *Necessary conditions for optimality in relaxed stochastic control problems,* Stochastic and stoch, Rep, Vol 73 (3-4), pp. 201-218, 2001.

[8] S. Bahlali, B. Mezerdi, *A general stochastic maximum principle for singular control problems,* Electronic Journal of Probability, Vol. 10, paper n° 30, pp. 988-1004, 2005.

[9] S. Bahlali, B. Djehiche, B. Mezerdi, *The relaxed stochastic maximum principle in singular optimal control of diffusions.* SIAM J. Control Optim. 46 (2007), no. 2, 427–444

[10] S. Bahlali, B. Djehiche, B. Mezerdi, *Approximation and Optimality Necessary Conditions in relaxed Stochastic Control Problems,* J. App. Math. Stoch. Anal., Volume 2006, Article ID 72762, 1-23.

[11] S. Bahlali, B. Mezerdi, Y. Ouknine, *Maximum principle in optimal control of a diffusion with non mooth coefficients.* Stochastic & Stoch. Reports, vol. 57, pp. 303-316, 1996.

[12] S. Bahlali, B. Mezerdi, Y. Ouknine, *Pathwise uniqueness and approximation of solutions of stochastic differential equations. Sem. de proba, XXXII, Edit. J. Azema, P.A. Meyer, M. Yor, Lecture Notes in Math. 1686, pp. 166-187, Springer Verlag, 1998.*

[13] H. Becker and V. Mandrekar, *On the existence of optimal random controls,* J. Math. Mech. **18** (1969), no. 12, 1151-1166.

[14] A. Bensoussan, *Lectures on stochastic control, Nonlinear Filtering and Stochastic Control*, Lect. Notes in Math. 972, Springer, Berlin, 1982, 1-62.

[15] J. M. Bismut, *An introductory approach to duality in optimal stochastic control*, SIAM Rev., 20 (1978), pp. 62-78.

[16] S. Bochner, *Integration von Functionen, deren Werte die Elemente eines Vectorraumes sind,* in Collected Papers of Salomon Bochner (R.C. Gunning, ed.), Vol. 1, pp. 587-601, American Mathematical Society, Providence, Rhode Island 1982.

[17] N. Cho, *On a martingale problem and a relaxed control problem w.r.t. SDE*, J. Korean Math. Soc. **33**, No. 4, 7790 p. 777, 1996.

[18] I. Ekland, *On the variational principle,* J. Math. Anal., Vol. 47, pp. 324-353, 1974.

[19] N. El Karoui, M. Jeanblanc-Piqué, *Martingale measures and partially observable diffusions,* Prépublication n° 25 du Laboratoire de probabilités de l'université Paris 6.

[20] N. El Karoui, J. P. Lepeltier, *Représentation des processus ponctueis multivariés à l'aide d'un processus de Poisson,* Z. Wahrscheinlicheits theor. Verw. Geb. 39. 111-133 (1977).

[21] N. El Karoui, L. Mazliak, *Backward stochastic differential equations,* Addison Wesley, Longman, 1997.

[22] N. El Karoui and S. Méléard, *Martingale measures and stochastic calculus,* Proba. Th. and Rel. Fields **84** (1990), no. 1, 83-101.

[23] N. El Karoui, D. H. Nguyen, and M. Jeanblanc-Piqué, *Compactification methods in the control of degenerate diffusions: existence of an optimal control,* Stochastic **20** (1987), no. 3, 169-219.

[24] N. El Karoui, D. H. Nguyen, and M. Jeanblanc-Piqué, *Existence of an optimal markovian filter for the controle under partial observations,* SIAM J. of Control and Optimisation, 26 n° 4 (1988), pp. 1025-1061.

[25] N. El Karoui, S. Peng, M. C. Quenez, *Backward stochastic differential equation in finance,* Mathematical finance, Vol 7, No. 1 (January 1997), 1-71.

[26] R. J. Elliott, *Stochastic calculus and applications,* Springer Verlag, 1982.

[27] W. H. Fleming, *Generalized solutions in optimal stochastic control, Differential Games and Control Theory*, II Proceedings of 2^{nd} Conference, University of Rhode Island, Kingston, RI, 1976.

[28] W. H. Fleming, M. Nisio, *On the existence of optimal stochastic controls.* Math. Mech. 15, pp. 777-794, 1966.

[29] W. H. Fleming, M. Nisio, *On stochastic relaxed control for partially observed diffusions,* Nagoya Math. Journal 93, pp. 71-108. 1984.

[30] W. H. Fleming, R. W. Rishel, *Deterministic and stochastic control,* Springer Verlag, 1975.

[31] A. Friedmann, *Stochastic differential equations and applications,* Vol. 1, Academic Press, 1975.

[32] T. Funaki, *A certain class of diffusion processes associated with non linear parabolic equations,* Z.Wahescheinlichkeis theor, Verw, Geb, **67.** 331-348, 1984.

[33] B. Grigelionis, *On the representation of integer valued measures by means of stochastic integrals with respect to Poisson measures,* Litov, Mat. Sb. **11**, 93-108, 1971.

[34] U. G. Haussmann, *Existence of optimal Markovian controls for degenerate diffusions,* Stochastic Differential Systems, Lecture Notes in Control and Inform. Sci., vol. 78, Springer, Berlin, 1986, pp. 171-186.

[35] U. G. Haussmann, *A stochastic maximum principle for optimal control of diffusions,* Pitman Research Notes in Math, Series 151, 1986.

[36] U. G. Haussmann and J.-P. Lepeltier, *On the existence of optimal controls,* SIAM J. on Cont. Optim. **28** (1990), no. 4, 851-902.

[37] U. G. Haussmann, *General necessary conditions for optimal control of stochastic system,* Math. Programming Stud., 6 (1976), pp. 34-48.

[38] Y. Hu, *Maximum principle of optimal control for Markov processes*, in Proc. Symposium on Control Theory and Applications, PRC, September, 1987; Acta Math. Sinica, to appear.

[39] N. Ikeda, S. Watanabe, *Stochastic differential equations and diffusion processes,* Amsterdam/ Kodansha: North Holland Mathematical Library 1981.

[40] J. Jacod and J. Mémin, *Sur un type de convergence intermédiaire entre la convergence en loi et la convergence en probabilité,* Seminaire de Probabilités 1979/1980) , Lect. Notes in Math., Vol. 850, Springer, Berlin, 1981, pp. 529-546.

[41] J. Jacod, A. N. Shiryaev, *Limit Theorems for stochastic processes,* Springer.

[42] I. Karatzas, S. *Brownian motion and stochastic calculus*, Springer Verlag, 1989.

[43] H. J. Kushner, *Existence results for optimal stochastic controls,* J. of Optim. Theory and Appl. **15** (1975), 347-359.

[44] H. J. Kushner, *Necessary conditions for continuous parameter stochastic optimization problems*, SIAM J. Control, 10 (1972), pp. 550-565.

[45] S. Labed, B. Mezerdi, *The maximum principle in optimal control of systems driven by martingale measures,* Afrika Statistika, Vol. 12(1), 2017, pages 1001-1022.

[46] L. Mazliak, *An introduction to probabilistic methods in stochastic control,* Lectures at Pescara University, Laboratoire de probabilité, Université Paris VI, 75252 Paris, Cedex 05, France, April 96.

[47] S. Méléard, *Martingale measure approximation, application to the control of diffusions,* Prépublication 69, Labo. Probabilites, Université de Paris VI, 1992.

[48] S. Méléard, S. Roelly-Coppoletta, *A generalized equations for a continuous measure branching process,* To appear in: Proccedings " Stochastic PDE's and Applications", Trento 1988 Berlin Heidelberg New Yok, Springer.

[49] S. Méléard, S. Roelly-Coppoletta, *Discontinuous measure-valued branching processes and generalized stochastic equations.* Prépublication n° 9, du Laboratoire de Probabilités de l'Université Paris VI.

[50] S. Méléard, S. Roelly-Coppoletta, *Systèmes de particules et mesures martingales: un théorème de propagation du chaos,* Séminaire de Probabilités n° 22, L.N n° 1321, pp. 438-448, Springer, 1988.

[51] B. Mezerdi and S. Bahlali, *Necessary conditions for optimality in relaxed stochastic control problems,* Stochastics and Stoch. Rep. **73** (2002), no. 3-4, 201-218.

[52] B. Mezerdi and S. Bahlali, *Approximation in optimal control of diffusion processes.* Rand. Oper. Stoch. Equ. 8 (2000), no. 4, 365–372.

[53] J. Neveu, *Processus aléatoires gaussiens,* Séminaire de Mathématiques supérieures, 1968. Presses de l'Université de Montréal.

[54] B. ∅ksendal, *Stochastic differential equations, Springer, 1985.*

[55] L. Overbeck, *On the predictable representation property for superprocesses,* Séminaire de Probabilités, Vol. 29 (1995), Lect. Notes in Math. 1613, 108-116.

[56] S.G. Peng, *A general stochastic maximum principle for optimal control problems,* SIAM J. Control and Optim. **28** (1990), no. 4, 966-979.

[57] L. S. Pontragin, V. G. Boltynskll, R. V. Gamkrelidze, AND E. F. Mischenko, *The Mathematical Theory of Optimal Control Processes*, John Wiley, New York, 1962.

[58] A. V. Skorohod, *Studies in the theory of random processes,* Reading, Mass: Addison Wesley 1965.

[59] J. B. Walsh, *An introduction to stochastic partial differential equations,* In, Ecole d'été de probabilites de Saint-Flour XIV-1984. Berlin Heidelberg New York, Springer 1986.

[60] J. Young and X. Y. Zhou, *Stochastic Controls, Hamiltonian Systems and HJB Equations,* Vol. 43, Springer, New York, 1999.

Appendix

In this appendix we introduce the integral of Banach space valued functions, the so-called Bochner integral, define the corresponding Lebesgue and Sobolev spaces.

Bochner integral

The Bochner integral for functions landing in a separable Banach space. This integral was first introduced by Salomon Bochner in his 1933 paper Integration von Functionen [16]. It is a generalization of the Lebesgue integral.

If Y is a normed vector space and $M \subset Y$, then the Borel σ-algebra $\mathcal{B}(M)$ over M is the σ-algebra generated by the system of relatively open subsets of M. We write $\mathcal{B}_d = \mathcal{B}(\mathbb{R}^d)$ for the Borel σ-algebra over $\mathbb{R}^d$. The $d-$dimensional Lebesgue measure is denoted by dx. In one dimension we often write dt. The Lebesgue measure of $A \in \mathcal{B}_d$ is denoted by $|A|$. We define

$$\mathcal{N}_d = \{ N \in \mathcal{B}_d | \quad |A| = 0 \}$$

as the set of Borel measurable sets of measure zero. For $A \in \mathcal{B}_d$, a function $f : A \to Y$ is called (Borel-)measurable if $f^{-1}(B) \in \mathcal{B}(A)$ for all $B \in \mathcal{B}(Y)$. If $f : A \to Y$ is measurable, then $\|f\|$ is measurable as well, where $\|f\|(x) = \|f(x)\|$ for $x \in A$.

Throughout, let E be a complex Banach space. A function $f : \mathbb{R}^d \to X$ is called simple, if there are $N \in \mathbb{N}$, $A_n \in \mathcal{B}_d$ and $x_n \in E$ for $n = 1, ..., N$ such that

$$f = \sum_{n=1}^{N} 1_{A_n} x_n.$$

Observe that simple functions are measurable. We start with the integral over simple functions.

Definition 3.3.1 *Let* $f = \sum_{n=1}^{N} 1_{A_n} x_n$ *be a simple function with* $|A| < \infty$ *for:* $n = 1, ..., N$*. Then the Bochner integral of* f *is defined by*

$$\int_{\mathbb{R}^d} f dx = \int_{\mathbb{R}^d} f(x)\, dx = \sum_{n=1}^{N} |A_n|\, x_n \in E.$$

We note that the above integral is independent of the representation of the simple function f. It is further clear that the Bochner integral is linear on the vector space of simple functions whose support has finite measure. Moreover, as a consequence of the triangle inequality, for each simple function f we have the estimate

$$\left\| \int_{\mathbb{R}^d} f dx \right\| = \int_{\mathbb{R}^d} \|f\|\, dx \tag{A.1}$$

where the integral on the right-hand side is now the usual scalar-valued Lebesgue integral. As in the scalar case, we extend the Bochner integral to a larger class of function by taking limits of simple functions. As it turns out, besides measurability for this procedure a separability condition is necessary.

Lemma 3.3.3 *Let* $f : \mathbb{R}^d \to E$ *be a map. Then the following assertions are equivalent.*

a. *There is a sequence* $(f_k)_{k\in\mathbb{N}}$ *of simple functions* $f_k : \mathbb{R}^d \to E$ *such that* $f_k(x) \to f(x)$ *as* $k \to \infty$ *for all* $x \in \mathbb{R}^d$.

b. f *is measurable and* $f\left(\mathbb{R}^d\right) \subseteq E$ *is separable.*

If one of the assertions is true, then in ***a.*** *one can choose* $(f_k)_{k\in\mathbb{N}}$ *such that*

$$\|f_k(x)\| \leq 2\, \|f(x)\|\,, \text{ for all } x \in \mathbb{R}^d.$$

The lemma suggest the following notion.

Definition 3.3.2 *A map $f : \mathbb{R}^d \to E$ is called strongly measurable if there is a sequence $(f_k)_{k\in\mathbb{N}}$ of simple functions $f : \mathbb{R}^d \to E$ such that $f_k(x) \to f(x)$ as $k \to \infty$ for all $x \in \mathbb{R}^d$.*

For a strongly measurable f one would like to define the Bochner integral as a limit of Bochner integrals of simple functions. Fortunately, there is a simple criterion when this is possible.

Lemma 3.3.4 *Let $f : \mathbb{R}^d \to E$ be strongly measurable. Then the following assertions are equivalent.*

1. *There is a sequence of simple functions $(f_k)_{k\in\mathbb{N}}$ such that $f_k(x) \to f(x)$ as $k \to \infty$ for all $x \in \mathbb{R}^d$ and*
$$\lim_{k\to\infty} \int_{\mathbb{R}^d} \|f_k - f\| \, dx = 0.$$
2. *It holds that $\int_{\mathbb{R}^d} \|f\| \, dx < \infty$.*

If one of the assertions is true, then the limit $\lim\limits_{k\to\infty} \int_{\mathbb{R}^d} f_k dx$ exists in E and is independent of the sequence of simple functions $(f_k)_{k\in\mathbb{N}}$ as in 1.

Now we can define integrability and the Bochner integral for a large class of functions.

Definition 3.3.3 *A function $f : \mathbb{R}^d \to E$ is called Bochner integrable if it is strongly measurable and if $\int_{\mathbb{R}^d} \|f\| \, dx < \infty$. In this case one sets*

$$\int_{\mathbb{R}^d} f dx = \lim_{k\to\infty} \int_{\mathbb{R}^d} f_k dx$$

where $(f_k)_{k\in\mathbb{N}}$ is any sequence of simple functions as in 1. of the precedent Lemma. Furthermore, for $A \in \mathcal{B}_d$ a function $f : A \to E$ is called Bochner integrable if its extension f_0

by zero to $\mathbb{R}^d$ *is Bochner integrable, and in this case one defines*

$$\int_A f dx = \int_{\mathbb{R}^d} f_0 dx.$$

For $A \in \mathcal{B}_d$ *one finally sets*

$$\mathcal{L}(A, E) = \{ f : A \to E | \ f \text{ is integrable} \}.$$