Solvability of Forward-Backward SDEs and the Nodal Set of Hamilton-Jacobi-Bellman Equations

Jin Ma[1] and Jiongmin Yong[2]

Abstract. In this paper, the solvability of a class of forward-backward stochastic differential equations (SDEs for short) over an arbitrarily prescribed time duration is studied. We design a stochastic relaxed control problem, with both drift and diffusion all being controlled, so that the solvability problem is converted to a problem of finding the nodal set of the viscosity solution to a certain Hamilton-Jacobi-Bellman equation. Our method overcomes the fatal difficulty encountered in the traditional contraction mapping theorem approach to the existence theorem of such SDEs.

Keywords. forward-backward stochastic differential equations, stochastic control, relaxed control, viscosity solutions, nodal set.

AMS Mathematics subject classification. 35K15, 49L20, 49L25, 60H10, 93E20.

[1] Department of Mathematics, Purdue University, West Lafayette, IN 47907.

[2] IMA, University of Minnesota, Minneapolis, MN 55455; on leave from Department of Mathematics, Fudan University, Shanghai 200433, China. Part of the work was done while this author was visiting Purdue University, and he would like to thank Professors L. D. Berkovitz and J. Douglas for their hospitality. Also, partial support from NSF of China and Fok Ying Tung Education Foundation is acknowledged.

§1. Introduction.

This paper studies the solvability (or the existence theorem) of the adapted solutions to a certain class of forward-backward stochastic differential equations. Let $(\Omega, \mathcal{F}, P; \mathcal{F}_t)$ be a filtered probability space satisfying the *usual conditions* (see §2). Suppose that on this probability space a d-dimensional $\{\mathcal{F}_t\}$-Brownian motion $\{W_t\}_{t\geq 0}$ is given. Consider the following forward-backward stochastic differential equations (SDE for short):

$$
X_t = x + \int_0^t b(X_s, Y_s, Z_s)ds + \int_0^t \sigma(X_s, Y_s, Z_s)dW_s, \tag{1.1}
$$

$$
Y_t = g(X_T) + \int_t^T \widehat{b}(X_s, Y_s, Z_s)ds + \int_t^T \widehat{\sigma}(X_s, Y_s, Z_s)dW_s, \tag{1.2}
$$

where (X, Y, Z) takes value in $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{m\times d}$ and $b, \widehat{b}, \sigma, \widehat{\sigma}, g$ are some smooth functions with appropriate dimensions; $T > 0$ is a prescribed constant which is called the *time duration*. Our objective is to find a triple (X, Y, Z) which is $\{\mathcal{F}_t\}$-adapted, square integrable, such that the equations (1.1)–(1.2) are satisfied on $[0, T]$. One should note that it is the extra process Z that makes it possible for (1.1)–(1.2) to have an adapted solution (cf. [27,28]).

The forward-backward SDEs of this kind was first introduced by Bismut [3] for characterizing the duality in optimal stochastic control and the stochastic Pontryagin Maximum Principle, in which the adjoint equation is a backward SDE. The SDE has been brought into strong attention recently because of its appeal not only in optimal control theory but also in mathematical finance and partial differential equations (cf. [11,28] and the references therein). However, all the existing results regarding the existence and uniqueness of an adapted solution require that the product of Lipschitz constants of the coefficients and the time duration T be small enough (see, for example, [1,28]). The restriction is simply due to the usual scheme of Picard iteration and the contraction mapping theorem. Thus, because of the "forward-backward" nature, one falls into a fatal difficulty when the time duration is large. In fact, [1] provided a counterexample showing that for some special kind of forward-backward SDEs, the adapted solution may fail to exist when the product of the Lipschitz constant and the time duration is larger than one. Therefore, the solvability of such an SDE over an arbitrarily prescribed time interval becomes an interesting issue, and sometimes is even crucial (for instance, when one studies the so-called decoupling problem in optimal stochastic control theory), but so far remains open.

In this paper we reformulate the above "forward-backward" SDE and consider its solutions in a *wider* sense. Namely, we allow the component Z to be a suitable adapted measure-valued process and allow the underlying probability space to change when necessary. This relaxation nontrivially contains the ordinary adapted solution as a special case. We then convert the problem of finding adapted solutions to the forward-backward SDEs to a problem of, roughly speaking, finding the "zero"-point (called nodal set) of the viscosity solution of a certain Hamilton-Jacobi-Bellman (HJB for short) equation. In fact, this HJB equation is exactly the one that the value function of the properly designed optimal relaxed stochastic control problem should satisfy.

Intuitively, our scheme will work if the following two problems can be solved.

Problem 1. For any $x \in \mathbb{R}^n$, there exists a $y \in \mathbb{R}^m$ such that $v(0, x, y) = 0$, where v is the viscosity solution of a certain HJB equation;

Problem 2. (a) For any $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$, there exists an optimal relaxed control which attains its value function $V(0, x, y)$;

(b) $v(s, x, y) = V(s, x, y)$ for all $(s, x, y) \in [0, T] \times \mathbb{R}^n \times \mathbb{R}^m$.

It turns out that if both Problems 1 and 2 above can be solved, then the optimal trajectory (X, Y), together with the optimal relaxed control process Z, will be an adapted solution to the forward-backward equation (1.1)–(1.2).

The main advantage of this scheme is that both Problems 1 and 2 are attackable. In fact, Problem 2-(a) is an existence theorem of the optimal relaxed controls, which should always be true in principle; while Problem 2-(b) follows from the so-called "Chattering Lemma", which is also one of the main features of stochastic relaxed control. However, we should point out here that in our case, these problems are by no means trivial because we have to deal with a system in which both drift and diffusion coefficients contain the control; to our knowledge, no existing results can be found in the literature regarding these issues. As for Problem 1, it is basically a problem of the *existence* of the nodal set for a certain HJB-equation. The notion of a "nodal set" was first introduced in [7] for studying the eigenfunctions of some linear elliptic equations (see [10] also). Recently, this notion was used for the study of the general solutions to elliptic and parabolic equations ([17,24]). Also, it is further related to the so-called singular set, zero-set and partial regularity of solutions to PDEs (see [5,6,14] and references cited therein). Therefore, one should have at least some clue as to how to approach such a problem.

In summary, our scheme overcomes the difficulty encountered in the contraction mapping approach, and provides a novel strategy for us to study the original solvability problem. In a forthcoming paper, we will use this method to investigate the once "mysterious" part Z in the previous works on backward (or forward-backward) SDEs (cf. [27,28]), which comes from the martingale representation theorem; from our point of view, it is nothing but some function of the optimal state and the optimal control. On the other hand, we also hope that this correspondence will raise some intersting questions in the study of (the existence of) the nodal sets in partial differential equations, via the existence of the solution to forward-backward SDEs.

This paper is organized as follows. In section 2 we collect some necessary notation and preliminaries and state our main results. Sections 3 is devoted to the existence of the optimal relaxed controls and the so-called Chattering Lemma. In section 4 we prove the equivalent relations among the solvability of forward-backward SDEs, solvability of optimal relaxed control and the existence of the nodal set of the related HJB-equations, which lead to the proof of our main theorem. A special class of forward-backward SDEs, raised from mathematical finance, is considered in sections 5 and 6. The solvability of such SDEs over an arbitrary time interval is proved using our new approach. Finally, some discussions and concluding remarks are made in section 7.

§2. Preliminaries and the Statement of Main Results.

§2.1. Basic Notations.

Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and B be a Banach space. Denote by $L^2([0,T]\times\Omega; B)$ the space of all measurable, square integrable processes defined on $[0,T]$ with values in B. If $B = H$ is a Hilbert space, then $L^2([0,T]\times\Omega; H)$ is also a Hilbert spaces with the usual inner product. Throughout this paper, we also assume that all the probability spaces will be filtered such that the quartuple $(\Omega, \mathcal{F}, P; \mathcal{F}_t)$ satisfies the *usual conditions* (that is, $\{\mathcal{F}_t\}$ is right-continuous and $\{\mathcal{F}_0\}$ contains all the P-null sets in $\mathcal{F}$), and carries a d-dimensional $\{\mathcal{F}_t\}$-Brownian motion $\{W_t : t \geq 0\}$. Let $\overline{\mathbb{R}^k}$ denote the one-point compactification of the Euclidean space $\mathbb{R}^k$. Then $\overline{\mathbb{R}^k}$ is a compact metric space. We also denote the set of all $l_1 \times l_2$ matrices $A = (a_{i,j})$ by $\mathbb{R}^{l_1\times l_2}$, and $|A|^2 \triangleq \sum_{i,j} a_{i,j}^2 = \mathrm{tr}AA^T$.

Now let U be a compact metric space. Denote by $\mathcal{P}(U)$ the space of all probability measures on $\mathcal{B}(U)$, the Borel σ-algebra of U. We endow the space $\mathcal{P}(U)$ with the Prohorov metric (cf. [12]). Then $\mathcal{P}(U)$ is also a compact metric space. Let $\mathcal{M}([0,T]\times U)$ be the space of all $\mathcal{P}(U)$-valued functions $\mu_t(\cdot)$, $t \geq 0$, such that $\mu_\cdot(A)$ is Borel measurable for all $A \in \mathcal{B}(U)$. For $\mu \in \mathcal{M}([0,T]\times U)$, define $\lambda_\mu(t,A) = \int_0^t \mu_s(A)ds$ for $t \in [0,T]$, $A \in \mathcal{B}(U)$. It is easy to see that λ_μ enjoys the following properties:

(i) $\lambda_\mu(0,A) = 0, \quad \forall A \in \mathcal{B}(U)$;

(ii) $\lambda_\mu(t,U) = t, \quad \forall t \geq 0$;

(iii) $\lambda_\mu(t,\cdot)$ is a measure on $\mathcal{B}(U)$ with total mass no more than t;

(iv) $\lambda_\mu(t,A)$ is nondecreasing in t for all $A \in \mathcal{B}(U)$;

(v) $\sup_{A\in\mathcal{B}(U)} |\lambda_\mu(s,A) - \lambda_\mu(t,A)| = |s-t|$.

In what follows, we denote $\mathcal{L}([0,T]\times U)$ to be the space of all the mappings $\lambda : [0,T]\times\mathcal{B}(U)\to[0,T]$ satisfying (i)—(v) above. It is known that for any $\lambda\in\mathcal{L}([0,T]\times U)$, there exists a $\mu\in\mathcal{M}([0,T]\times U)$ such that $\lambda=\lambda_\mu$ (cf. [22]). Therefore, the compactness of $\mathcal{P}(U)$ and $[0,T]$ renders $\mathcal{L}([0,T]\times U)$ a compact metric space (see [16] for a proof).

Furthermore, if a probability space $(\Omega,\mathcal{F},P)$ is given, then we denote $\mathbb{M}(\Omega)$ (resp. $\mathbb{L}(\Omega)$) to be the totality of all $\mathcal{M}([0,T]\times U)$ (resp. $\mathcal{L}([0,T]\times U)$)-valued random variables $\mu(\cdot)$ (resp. $\lambda(\cdot)$) defined on $(\Omega,\mathcal{F},P)$, such that for each $A\in\mathcal{B}(U)$, the process $\mu_\cdot(A,\cdot)$ is $\{\mathcal{F}_t\}$-adapted. It can also be checked (cf. [22], for example) that for any $\lambda\in\mathbb{L}(\Omega)$, there exists a $\mu\in\mathbb{M}(\Omega)$, such that for P-a.e. $\omega\in\Omega$,

$$\lambda(t,A,\omega)=\int_0^t\mu_s(A,\omega)ds,\qquad\forall A\in\mathcal{B}(U),\ \forall t\ge0. \tag{2.1}$$

Thus, we see that there exists a one-to-one correspondence between $\mathbb{M}(\Omega)$ and $\mathbb{L}(\Omega)$.

Since the representation (2.1) is important in our discussion, we now take a closer look at the process μ. First, we sketch the construction of μ (cf. [22]). Let $\overline{U}^n\subset\mathcal{B}(U)$ be a finite partition of U such that $\overline{U}^{n+1}$ is "finer" than $\overline{U}^n$; namely, if $A\in\overline{U}^{n+1}$ and $A'\in\overline{U}^n$ such that $A\cap A'\ne\emptyset$, then $A\subseteq A'$. We assume that the diameters of the sets of $\overline{U}^n$ is less than $1/n$ and define $\overline{U}=\bigcup_n\overline{U}^n$. For each $A\in\overline{U}$ and $\omega\in\Omega$, we define the left-derivative of λ in the variable t by

$$\mu_t(A,\omega)=\lim_{\delta\to0}\frac{1}{\delta}[\lambda(t,A,\omega)-\lambda(t-\delta,A,\omega)], \tag{2.2}$$

whenever the limit exists. Since by condition (v) above, $\lambda(\cdot,A,\omega)$ is absolutely continuous, the limit in (2.2) exists for a.e. $t\in[0,T]$, with the null set depending on A and ω. Since $\overline{U}$ is countable, we can find a $dP\otimes dt$-null set, say $\mathcal{N}\subseteq[0,T]\times\Omega$, such that the limit exists for all $A\in\overline{U}$, $\forall(t,\omega)\notin\mathcal{N}$. It is easily seen that for $(t,\omega)\notin\mathcal{N}$, μ defines a probability measure on the algebra generated by $\overline{U}$; by using the well-known Carathéodory Extension Theorem, we can extend this probability measure uniquely to $\sigma(\overline{U})=\mathcal{B}(U)$. Finally, for $(t,\omega)\in\mathcal{N}$ we define μ to be any element in $\mathcal{P}(U)$. Clearly, the process $\mu_\cdot(A,\cdot)$ defined in this way will be progressively measurable for each $A\in\mathcal{B}(U)$, since λ is.

We need more information for the process μ; for instance, we need to know whether the law of μ is completely determined by that of λ. In fact, it is the exceptional set $\mathcal{N}$ that makes it difficult to conclude that the law of λ completely determines the finite distributions of μ. We have nevertheless the following lemma; its proof is deferred to §3.

Lemma 2.1 *Suppose that $\lambda\in\mathbb{L}(\Omega)$ and $\lambda'\in\mathbb{L}(\Omega')$ such that they are identical in law as $\mathcal{L}([0,T]\times U)$-valued random variables; let $\mu\in\mathbb{M}(\Omega)$ and $\mu'\in\mathbb{M}(\Omega')$ be such that*

$$\begin{aligned}\lambda(t,A,\omega)&=\int_0^t\mu_s(A,\omega)ds,\ \text{a.s. } P,\\ \lambda'(t,A,\omega')&=\int_0^t\mu'_s(A,\omega')ds,\ \text{a.s. } P',\end{aligned} \tag{2.3}$$

for $(t, A) \in [0,T] \times \mathcal{B}(U)$*. Suppose that for* P-a.e. $\omega \in \Omega$*, the following hold:*

(1) for every $A \in \mathcal{B}(U)$*,* $\mu_{\cdot}(A, \omega)$ *is left-continuous, on* $(0, T]$*;*

(2) $\lim_{t \downarrow 0} \mu_t(A, \omega) = \mu_0(A, \omega)$ *for all* $A \in \mathcal{B}(U)$*.*

Then modulo a modification for μ' *at* $t = 0$*,* $\mu'_{\cdot}(A, \omega')$ *also satisfies (1), (2), for* P'-a.s. $\omega' \in \Omega'$ *and* μ *and* μ' *are identical in law as* $\mathcal{M}([0,T] \times U)$*-valued random variables.*

§2.2. Relaxed Control Problems, Value Functions.

Our relaxed control problem is an extension of the usual ones (cf. [15,16,22], for example); namely, we allow the control to also enter the diffusion coefficient. We formulate our control problem precisely as follows. Let us first define the admissible relaxed control set.

Definition 2.2. *The admissible relaxed control set, denoted by* $\mathcal{U}^R$*, is the set of all six-tuples* $\mathcal{A} = (\Omega, \mathcal{F}, P, \mathcal{F}_t, W, \mu)$*, in which* $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ *is a filtered probability space,* W *is an* $\{\mathcal{F}_t\}$*-Brownian motion and* $\mu \in \mathbb{M}(\Omega)$*.*

Remark 2.3. By a slight abuse of notation, we will often denote the generic element in $\mathcal{U}^R$ by μ instead of specifying all six components when the context is clear.

It is known that, under our definition, an ordinary U-valued $\{\mathcal{F}_t\}$-adapted process Z can always be represented as a relaxed control, by simply setting $\mu_t(du, \omega) = \delta_{Z_t(\omega)}(u)du$; in other words, $\mu_t(\cdot, \omega)$ is a δ-measure supported at $Z_t(\omega) \in U$ (cf. [4,15,16,22]). In this case, for any continuous function $g : U \to \mathbb{R}^n$, we have

$$\int_U g(u)\mu_t(du, \omega) = \int_U g(u)\delta_{Z_t(\omega)}(u)du = g(Z_t(\omega)). \tag{2.4}$$

Thus, we may imbed the set of all U-valued $\{\mathcal{F}_t\}$-adapted processes (called regular control later), denoted by $\mathcal{U}$, into $\mathcal{U}^R$ in an obvious way.

Now we suppose that an admissible relaxed control $\mathcal{A} = (\Omega, \mathcal{F}, P, \mathcal{F}_t, W, \mu)$ is given, the state equation is given by

$$\xi_t = \eta_s + \int_s^t \int_U b(\xi_r, u)\mu_r(du)dr + \int_s^t \int_U \sigma(\xi_r, u)\mu_r(du)dW_r, \qquad t \in [s, T], \tag{2.5}$$

where η_s is the initial state, which is assumed to be $\mathcal{F}_s$-measurable; and $s \in [0,T]$ is the starting time. The solution of (2.5), whenever it exists, will be denoted by $\xi_t(s, \eta_s, \mathcal{A})$, $t \in [s, T]$. We shall consider the following Mayer-type cost functional:

$$J(s, \eta_s; \mathcal{A}) = E^P[G(\xi_T(s, \eta_s, \mathcal{A}))], \tag{2.6}$$

The objective of a decision maker is then to choose a suitable admissible relaxed control $\mathcal{A} = (\Omega, \mathcal{F}, P; \mathcal{F}_t, W, \mu) \in \mathcal{U}^R$ so as to minimize the cost functional (2.6). We define the value function of the optimal relaxed control problem as follows:

$$V(s,x) = \inf_{\mathcal{A} \in \mathcal{U}^R} J(s,x;\mathcal{A}), \qquad (s,x) \in [0,T] \times \mathbb{R}^n. \tag{2.7}$$

Remark 2.4. Since we will not require the diffusion coefficient σ to be non-degenerate, the cost functional can be taken as Lagrange-type or even Bolza-type by simply adding an extra "state". Our results in §3 and §4 will also hold for these cases.

We shall prove in §4 that the value function $V(t,x)$ is the unique viscosity solution of the following HJB equation:

$$\begin{cases} V_t + \inf_{u \in U} \{ \langle V_x, b(x,u) \rangle + \frac{1}{2} \mathrm{tr}\, (\sigma(x,u)\sigma(x,u)^T V_{xx}) \} = 0, \\ V \big|_{t=T} = G(x). \end{cases} \tag{2.8}$$

Here, V_x and V_{xx} stand for the gradient and the Hessian of V in x. For the definition and basic results of viscosity solutions of HJB equations, see [8]. Next, for any continuous function $W(t,x)$, we define the nodal set $\mathcal{N}(W)$ of W by

$$\mathcal{N}(W) = \{(t,x) \mid W(t,x) = 0\}. \tag{2.9}$$

As pointed out in the introduction, this notion was first introduced in [7] for eigenfunctions of elliptic differential operators. Recently, this notion was also used in the study of general solutions of elliptic and parabolic equations. Here, we apply it to the viscosity solutions of HJB equations.

§2.3. Forward-Backward Stochastic Differential Equations.

We now turn to forward-backward stochastic equations. Let the probability space $(\Omega, \mathcal{F}, P; \mathcal{F}_t)$ be given such that a d-dimensional Brownian motion W is defined on it. Consider the following forward-backward SDE in a wider sense:

$$X_t = x + \int_0^t \int_U b(X_s, Y_s, u)\mu_s(du)ds + \int_0^t \int_U \sigma(X_s, Y_s, u)\mu_s(du)dW_s, \tag{2.10}$$

$$Y_t = g(X_T) + \int_t^T \int_U \widehat{b}(X_s, Y_s, u)\mu_s(du)ds + \int_t^T \int_U \widehat{\sigma}(X_s, Y_s, u)\mu_s(du)dW_s, \tag{2.11}$$

where the coefficients $b, \widehat{b}, \sigma, \widehat{\sigma}$ satisfy appropriate conditions (see (A.1) in §3), $g : \mathbb{R}^n \to \mathbb{R}^m$ is continuous. Moreover, X takes value in $\mathbb{R}^n$, Y takes value in $\mathbb{R}^m$ and U is a compact subset of $\overline{\mathbb{R}^{m \times d}}$.

Definition 2.5. *A triple (X, Y, μ) is called an adapted solution to (2.10)–(2.11), if (X, Y) is an $\mathbb{R}^n \times \mathbb{R}^m$-valued, $\{\mathcal{F}_t\}$ -adapted square integrable process and $\mu \in \mathbb{M}(\Omega)$ such that (X, Y, μ) satisfies (2.10)–(2.11) P-almost surely.*

Definition 2.6. *Let $T > 0$ be any given number. The forward-backward SDE (2.10)–(2.11) is called solvable over $[0, T]$ if there exists a probability space $(\Omega, \mathcal{F}, P; \mathcal{F}_t)$ on which is defined an $\{\mathcal{F}_t\}$-Brownian motion W, such that (2.10)–(2.11) has an adapted solution (X, Y, μ) defined on the interval $[0, T]$.*

It is clear that our definition of the adapted solution is an extension of the usual ones (cf. [27,28]). In fact, as we already mentioned in §2.2, an ordinary U-valued, $\{\mathcal{F}_t\}$-adapted process can always be represented as a δ-measure-valued process. Therefore, if (X, Y, Z) is an ordinary adapted solution, then (X, Y, δ_Z) is an adapted solution in the wider sense of Definition 2.6. Thus, under some restrictive conditions on the data and if the time duration T is *small enough* (cf. [28]), we know that the forward-backward equation is solvable in our sense.

Our purpose is to investigate the solvability on any time interval $[0, T]$. One of our main results of this paper can be stated as follows. For simplicity, we omit the precise statement of the assumptions (see §4 for details).

Theorem 2.7. *For any $T > 0$ and $x \in \mathbb{R}^n$, the forward-backward SDE (2.10)–(2.11) is solvable on $[0, T]$ if and only if the nodal set $\mathcal{N}(v)$ of v contains the point $(0, x, y)$ for some $y \in \mathbb{R}^m$, where v is the unique viscosity solutions of the following HJB equation:*

$$\begin{cases} v_t + H(x, v_x, v_y, v_{xx}, v_{xy}, v_{yy}) = 0, \\ v(T, x, y) = (y - g(x))^2, \end{cases} \tag{2.12}$$

where H is given by

$$\begin{aligned} H(x, v_x, v_y, v_{xx}, v_{xy}, v_{yy}) = \inf_{u \in U} \{ & \langle v_x, b(x, y, u) \rangle + \langle v_y, \widehat{b}(x, y, u) \rangle \\ & + \frac{1}{2} \mathrm{tr}\, (\sigma(x, y, u) \sigma(x, y, u)^T v_{xx} + \mathrm{tr}\, (\sigma(x, y, u) \widehat{\sigma}(x, y, u)^T v_{xy} \\ & + \frac{1}{2} \mathrm{tr}\, (\widehat{\sigma}(x, y, u) \widehat{\sigma}(x, y, u)^T v_{yy}) \}. \end{aligned}$$

The above theorem gives an equivalent relation between the solvability of the SDE (2.10)–(2.11) and some property of the nodal set $\mathcal{N}(v)$. In §§5–6, for an important class of forward-backward SDEs, we will characterize the nodal set of the corresponding value function, which will give the solvability of this kind of SDEs.

§3. Existence of Optimal Relaxed Controls.

In this section, we will study the optimal relaxed control problem stated in §2.2. Let us first give a proof of Lemma 2.1.

Proof of Lemma 2.1. First, note that since both λ and λ' are continuous, the equalities in (2.3) will hold for "all t, a.s. P." (resp. "all t, a.s. P'."). Thus we may assume without loss of generality that the equalities in (2.3) hold for all t and all ω (resp. all ω'). Next, observe that if $\mu_\cdot(A,\omega)$ is left-continuous for all $A \in \mathcal{B}(U)$ and $\omega \in \Omega$, then the left-derivative of $\lambda(t,A,\omega)$ in variable t exists and equals $\mu(t,A,\omega)$ for all $(t,A,\omega) \in (0,T] \times \mathcal{B}(U) \times \Omega$. Therefore, by the construction of μ and μ' and the assumption that λ and λ' are identical in law, we have

$$
\begin{aligned}
&P'\{\lim_{\delta\downarrow 0}\frac{1}{\delta}[\lambda'(t,A,\cdot)-\lambda'(t-\delta,A,\cdot)] = \mu'(t,A,\cdot), \forall t \in (0,T]\} \\
=&P\{\lim_{\delta\downarrow 0}\frac{1}{\delta}[\lambda(t,A,\cdot)-\lambda(t-\delta,A,\cdot)] = \mu(t,A,\cdot), \forall t \in (0,T]\} = 1,
\end{aligned}
\tag{3.1}
$$

for all $A \in \mathcal{B}(U)$; and by changing the value for μ' at $t=0$ if necessary, we have

$$
P'\{\lim_{\delta\downarrow 0}\frac{1}{\delta}\lambda'(\delta,A,\cdot) = \mu'(0,A,\cdot)\} = P\{\lim_{\delta\downarrow 0}\frac{1}{\delta}\lambda(\delta,A,\cdot) = \mu(0,A,\cdot)\} = 1,
$$

for all $A \in \mathcal{B}(U)$. Moreover, for fixed $A \in \mathcal{B}(U)$, the finite distributions of the processes $\mu_\cdot(A,\cdot)$ and $\mu'_\cdot(A,\cdot)$, which are now completely determined by the law of λ and λ', are the same. By a standard Monotone Class argument, we see that for each $f \in C(U)$, the processes

$$
\begin{aligned}
\mu_f(t,\omega) &\stackrel{\Delta}{=} \int_U f(u)\mu_t(du,\omega), \qquad (t,\omega) \in [0,T]\times\Omega, \\
\mu'_f(t,\omega') &\stackrel{\Delta}{=} \int_U f(u)\mu'_t(du,\omega'), \qquad (t,\omega') \in [0,T]\times\Omega'
\end{aligned}
\tag{3.2}
$$

are identical in law. Finally, we can take a countable dense subset $\{f_i\} \subset C(U)$ which separates the points in $\mathcal{P}(U)$, such that every μ_{f_i} and μ'_{f_i} are identical in law, $i=1,2,\cdots$; whence μ and μ' are identical in law as $\mathcal{M}([0,T]\times U)$-valued random variables. $\square$

Now, for the state equation (2.5) and the cost functional (2.6), we make the following assumptions.

(A.1) The functions $b : \mathbb{R}^n \times U \to \mathbb{R}^n$; $\sigma : \mathbb{R}^n \times U \to \mathbb{R}^{n\times d}$ are continuous on $[0,T]\times U$; and differentiable in x, such that for some constant $C>0$,

$$
|b_x(x,u)| + |\sigma_x(x,u)| \le C, \quad \forall x \in \mathbb{R}^n,\ \forall u \in U;
\tag{3.3}
$$

$$(3.4) \qquad |b(0,u)| + |\sigma(0,u)| \le C, \quad \forall\, u \in U.$$

(A.2) The function $G : \mathbb{R}^n \to [0,\infty)$ is continuous.

It is evident that under our assumption (A.1), the SDE (2.5) has a pathwise unique solution on $[0,T]$ for any given admissible relaxed control $\mathcal{A} \in \mathcal{U}^R$, initial time $s \in [0,T]$ and initial state η_s. Furthermore, by a standard argument, one easily verifies the following moment estimates: for any integer $m > 0$, there exists a constant $C_{m,T} > 0$ depending only on b, σ, m and T such that

$$(3.5) \qquad E^P\{ \sup_{s\le t\le T} |\xi_t(s,\eta_s,\mathcal{A})|^{2m}\} \le C_{m,T} E|\eta_s|^{2m},$$

$$(3.6) \qquad E^P\{ \sup_{s\le t\le T} |\xi_t(s,\xi^1_s,\mathcal{A}) - \xi(t;s,\xi^2_s,\mathcal{A})|^{2m}\} \le C_{m,T} E|\xi^1_s - \xi^2_s|^{2m}.$$

The following two lemmas concerning relaxed control problem will be useful to our future discussion. Since our argument is independent of the initial state and initial time, we assume without loss of generality that $s = 0$ and that $\eta_0 = x$ is deterministic.

Lemma 3.1. *Suppose that an admissible relaxed control $\mathcal{A} = (\Omega, \mathcal{F}, P, \mathcal{F}_t, W, \mu) \in \mathcal{U}^R$ and an initial state $x \in \mathbb{R}^n$ be given. Then there exists a sequence of $\mu^{(k)} \in \mathbb{M}(\Omega)$ such that for each k, the path $\mu^{(k)}_\cdot(A,\omega)$ is continuous for all $A \in \mathcal{B}(U)$ and $\omega \in \Omega$; and that*

$$(3.7) \qquad E^P|\xi^{(k)} - \xi|^{*,2}_T \triangleq E^P\left\{ \sup_{0\le t\le T} |\xi^{(k)}_t - \xi_t|^2 \right\} \to 0, \qquad \text{as } k \to \infty,$$

where $\xi^{(k)}$ and ξ are the solutions of (2.5) with respect to $\mu^{(k)}$ and μ respectively.

Consequently, if we define $\mathcal{A}^{(k)} = (\Omega, \mathcal{F}, P, \mathcal{F}_t, W, \mu^{(k)})$, then as $k \to \infty$, we have $J(0,x;\mathcal{A}^{(k)}) \to J(0,x;\mathcal{A})$.

Proof. Since the second consequence follows directly from (3.7), we need only show the first assertion. Let $\mathcal{A} = (\Omega, \mathcal{F}, P, \mathcal{F}_t, W, \mu) \in \mathcal{U}^R$ and $\eta_0 = x$ be given. Define for each $k = 1, 2, \cdots$, and $A \in \mathcal{B}(U)$, $\omega \in \Omega$,

$$(3.8) \qquad \mu^{(k)}_t(A,\omega) = \begin{cases} \dfrac{1}{t}\displaystyle\int_0^t \mu_s(A,\omega)ds, & 0 < t < 2^{-k}; \\[2ex] 2^k \displaystyle\int_{t-2^{-k}}^t \mu_s(A,\omega)ds, & t \ge 2^{-k}. \end{cases}$$

It is clear that $\mu^{(k)}_t(A,\omega) \to \mu_t(A,\omega)$, a.e. $t \in [0,T]$, as $k \to \infty$ (with the null set depending on A and ω). However, if we borrow the notation $\overline{U}$ from §2.1, and the similar argument used in the proof of Lemma 2.1, we can find a $dP \otimes dt$-null set $\mathcal{N} \subset [0,T] \times \Omega$ such that

$\mu_t^{(k)}(A,\omega) \to \mu_t(A,\omega)$, for all $A \in \overline{U}$ and $(t,\omega) \notin \mathcal{N}$. Since $\overline{U}$ forms a topological basis of U and generates $\mathcal{B}(U)$, it is not hard to show by a Monotone-Class argument that if $(t,\omega) \notin \mathcal{N}$, then for any $f \in C(U;\mathbb{R}^n)$, we have

$$\int_U f(u)\mu_t^{(k)}(du,\omega) \to \int_U f(u)\mu_t(du,\omega), \qquad \text{as } k \to \infty. \tag{3.9}$$

Now let ξ be the solutions of (2.5) with respect to μ. By (3.9) and the condition (A.1), for $(t,\omega) \notin \mathcal{N}$, we must have

$$\begin{cases} \displaystyle\int_U b(\xi_s(\omega),u)\mu_t^{(k)}(du,\omega) \to \int_U b(\xi_s(\omega),u)\mu_t(du,\omega); \\ \displaystyle\int_U \sigma(\xi_s(\omega),u)\mu_t^{(k)}(du,\omega) \to \int_U \sigma(\xi_s(\omega),u)\mu_t(du,\omega), \end{cases} \tag{3.10}$$

as $k \to \infty$. Furthermore, by the Dominated Convergence Theorem, we see that

$$\begin{cases} \displaystyle E\left[\int_0^T \left|\int_U b(\xi_s,u)\mu_t^{(k)}(du) - \int_U b(\xi_s,u)\mu_t(du)\right|^2 dt\right] \to 0; \\ \displaystyle E\left[\int_0^T \left|\int_U \sigma(\xi_s,u)\mu_t^{(k)}(du) - \int_U \sigma(\xi_s,u)\mu_t(du)\right|^2 dt\right] \to 0, \end{cases} \tag{3.11}$$

as $k \to \infty$. Therefore, the conclusion (3.7) follows from a standard argument using the Burkholder-Davis-Gundy ([29]) and Gronwall inequalities. $\square$

Lemma 3.2. *Let the initial state $x \in \mathbb{R}^n$ be given. For any $\mathcal{A} = (\Omega, \mathcal{F}, P, \mathcal{F}_t, W, \mu) \in \mathcal{U}^R$, denote $W^{\mathcal{A}} = W$, $\lambda^{\mathcal{A}}(t,A,\omega) = \int_0^t \mu_s(A,\omega)ds$ and $\xi_t^{\mathcal{A}} = \xi_t(0,x,\mathcal{A})$, $t \in [0,T]$. Then the family $\{(W^{\mathcal{A}}, \lambda^{\mathcal{A}}, \xi^{\mathcal{A}}) : \mathcal{A} \in \mathcal{U}^R\}$ is tight, as a set of $C([0,T];\mathbb{R}^d) \times \mathcal{L}([0,T] \times U) \times C([0,T];\mathbb{R}^n)$-valued random variables.*

Proof. It suffices to show that each marginal is tight. Obviously, the family $\{W^{\mathcal{A}} : \mathcal{A} \in \mathcal{U}^R\}$ is tight because they are all Brownian motions. The family $\{\lambda^{\mathcal{A}} : \mathcal{A} \in \mathcal{U}^R\}$ is tight because $\mathcal{L}([0,T] \times U)$ is compact. Finally, since b, σ are both of at most linear growth, a similar argument as that in [17, §7] shows that

$$\begin{cases} \displaystyle\phi_{\mathcal{A}}(t) = \int_0^t \int_U a(\xi_r^{\mathcal{A}},u)\mu_r(du)dr; \\ \displaystyle\psi_{\mathcal{A}}(t) = \int_0^t \int_U \sigma(\xi_r^{\mathcal{A}},u)\mu_r(du)dW_r, \end{cases} \qquad \mathcal{A} \in \mathcal{U}^R \tag{3.12}$$

are tight, therefore $\{\xi^{\mathcal{A}}; \mathcal{A} \in \mathcal{U}^R\}$ is tight because $\xi_t^{\mathcal{A}} = x + \phi_{\mathcal{A}}(t) + \psi_{\mathcal{A}}(t)$ for each $t \in [s,T]$. The conclusion follows. $\square$

We now let $s \in [0,T]$ and $x \in \mathbb{R}^n$ be fixed. Without loss of generality, we further assume that $s = 0$. Let $\mathcal{A}^{(k)} = (\Omega^{(k)}, \mathcal{F}^{(k)}, P^{(k)}, \mathcal{F}_t^{(k)}, W^{(k)}, \mu^{(k)})$ be a minimizing sequence. Namely,

$$J(0,x;\mathcal{A}^{(k)}) \to V(0,x), \qquad \text{as } k \to \infty. \tag{3.13}$$

By Lemma 3.1, we may assume that each $\mu^{(k)}$ has "continuous paths" (that is, for each $A \in \mathcal{B}(U)$ and $P^{(k)}$-a.e. $\omega \in \Omega$, $\mu_\cdot^{(k)}(A,\omega)$ is continuous). Let us denote $\lambda^{(k)} = \lambda_{\mu^{(k)}}$ as in (2.1). By Lemma 3.2, we know that the family $\{W^{(k)}, \lambda^{(k)}, \xi^{(k)}\}$ is tight, so by the Skorohod Theorem (cf. [12]), there exists a probability space $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{P})$, on which is defined a sequence of processes $\{(\widehat{W}^{(k)}, \widehat{\lambda}^{(k)}, \widehat{\xi}^{(k)})\}_{k=1}^{\infty}$ and $(\widehat{W}, \widehat{\lambda}, \widehat{\xi})$, such that

(i) the triples $(\widehat{W}^{(k)}, \widehat{\lambda}^{(k)}, \widehat{\xi}^{(k)})$ and $(W^{(k)}, \lambda^{(k)}, \xi^{(k)})$ are identical in law, for $k = 1, 2, \cdots$; and

(ii) along a subsequence, may assume itself, $\{(\widehat{W}^{(k)}, \widehat{\lambda}^{(k)}, \widehat{\xi}^{(k)})\}$ converges $\widehat{P}$-almost surely to $(\widehat{W}, \widehat{\lambda}, \widehat{\xi})$ in the space $C([0,T];\mathbb{R}^d) \times \mathcal{L}([0,T] \times U) \times C([0,T];\mathbb{R}^n)$.

Since $\widehat{\lambda}^{(k)}$ and $\widehat{\lambda}$ are $\mathcal{L}([0,T] \times U)$-valued random variables, we know that there exist $\mathcal{M}([0,T] \times U)$-valued random variables $\widehat{\mu}^{(k)}$ and $\widehat{\mu}$, such that $\widehat{\lambda}^{(k)} = \lambda_{\widehat{\mu}^{(k)}}$ and $\widehat{\lambda} = \lambda_{\widehat{\mu}}$ (recall (2.1)). By Lemma 2.1, $\widehat{\mu}^{(k)}$ has the same law as $\mu^{(k)}$ for each k since each $\mu^{(k)}$ has "continuous paths". Furthermore, let us define $\widehat{\mathcal{F}}_t = \sigma\{(\widehat{\lambda}_s, \widehat{W}_s) : 0 \le s \le t\}$, then by noting that each $(\widehat{\lambda}^{(k)}, \widehat{W}^{(k)})$ has the some law as $(\lambda^{(k)}, W^{(k)})$ and that $(\widehat{\lambda}^{(k)}, \widehat{W}^{(k)})$ converges to $(\widehat{\lambda}, \widehat{W})$ $\widehat{P}$-almost surely, it is not hard to show that $\widehat{W}$ is an $\{\widehat{\mathcal{F}}_t\}$-Brownian motion. Also, by the definition of $\widehat{\mu}$, one can check that $\widehat{\mathcal{F}}_t = \sigma\{(\widehat{\mu}_s, \widehat{W}_s) : 0 \le s \le t\}$. Bearing these facts in mind, we now drop "$\widehat{\ }$" from the above expressions to simplify the notation. The following lemma will be crucial to the rest of this section.

Lemma 3.3. For each $(t,\omega) \in [0,T] \times \Omega$, define

$$\begin{cases} M_b^{(k)}(t,\omega) = \displaystyle\int_U b(\xi_s(\omega),u)\mu_s^{(k)}(du,\omega), \quad M_b(t,\omega) = \int_U b(\xi_s(\omega),u)\mu_s(du,\omega); \\ M_\sigma^{(k)}(t,\omega) = \displaystyle\int_U \sigma(\xi_s(\omega),u)\mu_s^{(k)}(du,\omega), \quad M_\sigma(t,\omega) = \int_U \sigma(\xi_s(\omega),u)\mu_s(du,\omega). \end{cases} \tag{3.14}$$

Then, it holds that

$$M_b^{(k)} \xrightarrow{w} M_b, \quad M_\sigma^{(k)} \xrightarrow{w} M_\sigma, \quad \text{in } L^2([0,T] \times \Omega), \text{ as } k \to \infty. \tag{3.15}$$

Proof. Since $\lambda^{(k)}$ converges to λ, P-almost surely in $\mathcal{L}([0,T] \times U)$, we can find a subset $\Omega_0 \subseteq \Omega$ with $P(\Omega_0) = 1$, such that for any $g(\cdot,\cdot) \in C([0,T] \times U)$, one has

$$\lim_{n\to\infty} \int_0^T \int_U g(s,u)\mu_s^{(k)}(du,\omega) = \int_0^T \int_U g(s,u)\mu_s(du,\omega), \qquad \forall \omega \in \Omega_0. \tag{3.16}$$

Thus, if we set, for each $\omega \in \Omega$, $g_b^\omega(s,u) \triangleq b(\xi_s(\omega),u)$; $g_\sigma^\omega(s,u) \triangleq \sigma(\xi_s(\omega),u)$, then for any $\varphi \in L^2(\Omega; C([0,T])$, there is another Borel set in $\mathcal{F}$ with probability one, may assume Ω_0 itself, such that the functions $\varphi(\cdot,\omega)g_b^\omega(\cdot,\cdot), \varphi(\cdot,\omega)g_\sigma^\omega(\cdot,\cdot)$ all belong to $C([0,T]\times U)$ for each fixed $\omega \in \Omega_0$. Thus by definitions (3.14), we have from (3.16) that

$$\lim_{k\to\infty} \int_0^T \varphi(s,\omega) M_b^{(k)}(s,\omega)ds = \int_0^T \varphi(s,\omega) M_b(s,\omega)ds, \tag{3.17}$$

$$\lim_{k\to\infty} \int_0^T \varphi(s,\omega) M_\sigma^{(k)}(s,\omega)ds = \int_0^T \varphi(s,\omega) M_\sigma(s,\omega)ds, \tag{3.18}$$

for all $\omega \in \Omega_0$. By applying the Dominated Convergence Theorem, we obtain that

$$\langle \varphi, M_b^{(k)} \rangle \to \langle \varphi, M_b \rangle; \qquad \langle \varphi, M_\sigma^{(n)} \rangle \to \langle \varphi, M_\sigma \rangle, \qquad \text{as } k \to \infty, \tag{3.19}$$

where $\langle , \rangle$ denotes the inner product of $L^2([0,T]\times\Omega)$. This leads to our conclusion. $\square$

We now define for each $\tilde{\mu} \in \mathbb{M}(\Omega)$ a process

$$(M_b[\tilde{\mu}](t,\omega), M_\sigma[\tilde{\mu}](t,\omega)) \triangleq \left(\int_U b(\xi_t(\omega),u)\tilde{\mu}_t(du,\omega), \int_U \sigma(\xi_t(\omega),u)\tilde{\mu}_t(du,\omega) \right), \tag{3.20}$$

for $(t,\omega) \in [0,T]\times\Omega$; and $\mathcal{K} = \{(M_b[\tilde{\mu}], M_\sigma[\tilde{\mu}]) : \tilde{\mu} \in \mathbb{M}(\Omega)\}$. Then $\mathcal{K}$ is a convex set in $L^2([0,T]\times\Omega) \times L^2([0,T]\times\Omega)$, and $(M_a^{(k)}, M_\sigma^{(k)}), (M_a, M_\sigma) \in \mathcal{K}$. By Mazur's Theorem (cf. [16]), the weak closure of $\mathcal{K}$ equals its strong closure; Furthermore, for each integer $l > 0$ and $\varepsilon > 0$, there exists a finite set of numbers $\{\alpha_1, \cdots, \alpha_{N(l,\varepsilon)}\}$ satisfying $\alpha_i \geq 0$ and $\sum_i \alpha_1 = 1$, such that

$$\left\| \sum_{i=1}^{N(l,\varepsilon)} \alpha_i M_b^{(l+i)} - M_b \right\|^2_{L^2([0,T]\times\Omega)} + \left\| \sum_{i=1}^{N(l,\varepsilon)} \alpha_i M_\sigma^{(l+i)} - M_\sigma \right\|^2_{L^2([0,T]\times\Omega)} < \varepsilon. \tag{3.21}$$

Using these facts, we can now prove our main result of this section. Recall that $\mathcal{A}^{(k)} = (\Omega^{(k)}, \mathcal{F}^{(k)}, P^{(k)}, \mathcal{F}_t^{(k)}, W^{(k)}, \mu^{(k)})$ is a minimizing sequence (see (3.13)), $\xi^{(k)}$ is the corresponding state trajectory and $(W^{(k)}, \lambda^{(k)}, \xi^{(k)})$ converges to (W, λ, ξ) P-almost surely.

Theorem 3.4. *The limit process ξ of $\{\xi^{(k)}\}$ satisfies the following SDE:*

$$\xi_t = x + \int_0^t \int_U b(\xi_s,u)\mu_s(du)ds + \int_0^t \int_U \sigma(\xi_s,u)\mu_s(du)dW_s, \quad t \in [0,T]. \tag{3.22}$$

Consequently, the optimal relaxed control exists and $\mathcal{A} = (\Omega, \mathcal{F}, P, \mathcal{F}_t, W, \mu)$ is an optimal relaxed control.

Proof. First, for any function $\zeta : [0,T] \to \mathbb{R}^k$, we define $|\zeta|_t^* = \sup_{0\le s\le t} |\zeta_s|$ and $|\zeta|_t^{*,2} = (|\zeta|_t^*)^2$. Since $(\xi^{(k)}, W^{(k)}) \to (\xi, W)$ in $C([0,T];\mathbb{R}^n) \times C([0,T];\mathbb{R}^d)$, by the Dominated Convergence Theorem, we see that for any $\varepsilon > 0$, there exists an integer $N_0 = N_0(\varepsilon) > 0$, such that

$$E|\xi^{(k)} - \xi|_T^{*,2} + E|W^{(k)} - W|_T^{*,2} < \varepsilon, \qquad \forall k > N_0. \tag{3.23}$$

Now for such N_0 and ε, let $\overline{N} = N(N_0, \varepsilon)$ and $\{\alpha_1, \cdots, \alpha_{\overline{N}}\}$ be such that $\alpha_i \ge 0$; $\sum_i \alpha_i = 1$ and (3.21) holds, here and later on "$\sum_i$" means "$\sum_{i=1}^{\overline{N}}$".

Define, for each $i = 1, \cdots, \overline{N}$,

$$\begin{aligned}\Delta^i(W)(t) \triangleq & \int_0^t \int_U \sigma(\xi_s^{(N_0+i)}, u)\mu_s^{(N_0+i)}(du)dW_s^{(N_0+i)} \\ & - \int_0^t \int_U \sigma(\xi_s^{(N_0+i)}, u)\mu^{(N_0+i)}(du)dW_s.\end{aligned} \tag{3.24}$$

Then it is readily seen, by using the growth condition of σ (see (3.3)–(3.4)) and the moment estimation (3.5), that there exists a constant $K > 0$, depending only on σ, T and $\overline{N}$, such that

$$E|\sum_i \alpha_i \Delta^i(W)|_T^{*,2} < K\varepsilon, \tag{3.25}$$

We now denote all the constants depending only on b, σ, T and $\overline{N}$ by a generic one K which may vary line by line. By using the Lipschitz property of b and σ (see (3.3)) and the Burkholder-Davis-Gundy ([29])inequality, one can also show that

$$\begin{aligned}& E\left|\int_0^\cdot \sum_i \alpha_i \left[\int_U b(\xi_s^{(N_0+i)}, u)\mu_s^{(N_0+i)}(du) - \int_U b(\xi_s, u)\mu_s^{(N_0+i)}(du)\right] ds\right|_T^{*,2} \\ =& E\left|\int_0^t \sum_i \alpha_i \left[\int_U b(\xi_s^{(N_0+i)}, u)\mu_s^{(N_0+i)}(du) - M_b^{(N_0+i)}(s)\right] ds\right|_T^{*,2} < K\varepsilon,\end{aligned} \tag{3.26}$$

and

$$E\left|\int_0^\cdot \sum_i \alpha_i \left[\int \sigma(\xi_s^{(N_0+i)}, u)\mu_s^{(N_0+i)}(du) - M_\sigma^{(N_0+i)}(s)\right] dW_s\right|_T^{*,2} \le K\varepsilon. \tag{3.27}$$

Note that for each $k = 1, 2, \cdots$, the triple $(W^{(k)}, \mu^{(k)}, \xi^{(k)})$ satisfies the following equation (on the new probability space):

$$\xi_t^{(k)} = x + \int_0^t \int_U b(\xi_s^{(k)}, u)\mu_s^{(k)}(du)ds + \int_0^t \int_U \sigma(\xi_s^{(k)}, u)\mu_s^{(k)}(du)dW_s^{(k)}, \tag{3.28}$$

whence

$$\sum_i \alpha_i \xi_t^{(N_0+i)} = x + \int_0^t \sum_i \alpha_i \int_U b(\xi_s^{(N_0+i)}, u)\mu_s^{(N_0+i)}(du)ds \tag{3.29}$$
$$+ \int_0^t \sum_i \alpha_i \int_U \sigma(\xi_s^{(N_0+i)}, u)\mu_s^{(N_0+i)}(du)dW_s + \sum_i \alpha_i \Delta^i(W).$$

Since $f(x) = x^2$ is convex, it is easy to check that

$$E\left|\sum_i \alpha_i \xi^{(N_0+i)} - \xi\right|_T^{*,2} \leq \sum_i \alpha_i E|\xi^{(N_0+i)} - \xi|_T^{*,2} \leq \sum_i \alpha_i \varepsilon = \varepsilon. \tag{3.30}$$

Therefore, combining this with (3.24)–(3.27), we see from (3.28) that

$$\begin{aligned}
& E\left|\xi - x - \int_0^\cdot \int_U b(\xi_s, u)\mu_s(du)ds - \int_0^\cdot \int_U \sigma(\xi_s, u)\mu_s(du)dW_s\right|_T^{*,2} \\
= \; & E\left|\xi - x - \int_0^\cdot M_b(s)ds - \int_0^\cdot M_\sigma(s)dW(s)\right|_T^{*,2} \\
\leq \; & K\left\{ E\left|\sum_i \alpha_i \xi^{(N_0+i)} - \xi\right|_T^{*,2} \right. \\
& + E\left|\int_0^\cdot \sum_i \alpha_i \left[\int_U b(\xi_s^{(N_0+i)}, u)\mu_s^{(N_0+i)}(du) - M_b^{(N_0+i)}(s)\right] ds\right|_T^{*,2} \\
& + E\left|\int_0^\cdot \sum_i \alpha_i \left[\int_U \sigma(\xi_s^{(N_0+i)}, u)\mu_s^{(N_0+i)}(du) - M_\sigma^{(N_0+i)}(s)\right] ds\right|_T^{*,2} \\
& + E\left|\int_0^\cdot [\sum_i \alpha_i M_b^{(N_0+i)}(s) - M_b(s)]ds\right|_T^{*,2} \\
& \left. + E\left|\int_0^\cdot \sum_i \alpha_i M_\sigma^{(N_0+i)}(s) - M_\sigma(s)]dW_s\right|_T^{*,2} + E\left|\sum_i \alpha_i \Delta^i(W)\right|_T^{*,2} \right\} \\
\leq \; & K\left\{ \varepsilon + E\int_0^T |\sum_i \alpha_i M_b^{(N_0+i)}(s) - M_b(s)|^2 ds \right. \\
& \left. + E\int_0^T |\sum_i \alpha_i M_\sigma^{(N_0+i)}(s) - M_\sigma(s)|^2 ds \right\}.
\end{aligned} \tag{3.31}$$

But by the definition of $\alpha_i's$ and N_0, we see that the foregoing is less than $K\varepsilon$. Letting $\varepsilon \downarrow 0$, we obtain that ξ satisfies the (3.22). Finally, note that

$$J(0, x; \mu) = EG(\xi_T) = \lim_{n\to\infty} EG(\xi^{(k)}(T)) = V(0, x),$$

we see that μ is the optimal relaxed control. The theorem is proved. $\square$

Remark 3.5. It is clear that the argument in this section is independent of the initial time $s \in [0,T]$, thus the value function $V(s,x)$ is attainable by some relaxed control for any $(s,x) \in [0,T] \times \mathbb{R}^n$.

To conclude this section, let us prove the so-called "Chattering Lemma". This lemma tell us that any relaxed control can be approximated by ordinary controls in a suitable sense. Consequently, the value function of the ordinary control problem coincides with that of the relaxed control problem.

Theorem 3.6. *For any $\mathcal{A} = (\Omega, \mathcal{F}, P, \mathcal{F}_t, W, \mu) \in \mathcal{U}^R$, there exists a sequence of U-valued progressively measurable processes $\{Z^{(k)}\}_{k=1}^{\infty}$ defined on $(\Omega, \mathcal{F}, P)$, such that if we denote $\mathcal{A}^{(k)} = (\Omega, \mathcal{F}, P, \mathcal{F}, W, \delta_{Z^{(k)}})$, then*

$$\lim_{k\to\infty} J(s,x;\mathcal{A}^{(k)}) = J(s,x;\mathcal{A}). \tag{3.32}$$

Consequently, we have

$$V(s,x) = \inf_{\mathcal{A}\in\mathcal{U}^R} J(s,x;\mathcal{A}) = \inf_{\mathcal{A}\in\mathcal{U}} J(s,x;\mathcal{A}). \tag{3.33}$$

Proof. We will follow closely the scheme suggested by [16]. Let $\mathcal{A} = (\Omega, \mathcal{F}, P, \mathcal{F}_t, W, \mu)$ $\in \mathcal{U}^R$ be given. First, by Lemma 3.1, we again assume without loss of generality that μ has "continuous paths". Then, by Proposition 4.1 in [16], we can find a sequence of switching relaxed controls $\{\mu^{(k)}\}$ in $\mathbb{M}(\Omega)$, namely, for each $A \in \mathcal{B}(U)$ and $\omega \in \Omega$, $\mu_t^{(k)}(A,\omega)$ is a step function in t and we further modify it to satisfy the conditions (1) and (2) in Lemma 2.1, such that for

$$\lambda^{(k)}(t,A,\omega) \triangleq \int_0^t \mu_s^{(k)}(A,\omega)ds, \quad \lambda(t,A,\omega) \triangleq \int_0^t \mu_s(A,\omega)ds, \tag{3.34}$$

we have

$$\lim_{k\to\infty} \sup_{A\in\mathcal{B}(U)} \sup_{0\le t\le T} |\lambda^{(k)}(t,A,\omega) - \lambda(t,A,\omega)| = 0, \qquad P\text{-a.e. } \omega \in \Omega. \tag{3.35}$$

In other words, $\{\lambda^{(k)}\}$ converges to λ, P-almost surely in the space $\mathcal{L}([0,T]\times U)$ as $k \to \infty$.

Next, for any l, let $\{u_1, \cdots, u_m\}$ be a 2^{-l}-net of U, and $V_1, \cdots, V_m \in \mathcal{B}(U)$ be a partition of U such that

$$|u_i - u| < 2^{-l}, \qquad \forall u \in V_i. \tag{3.36}$$

Then by Theorem 4 in [16], we can find for each $\mu^{(k)}$ a sequence of ordinary controls $\{Z^{(k,l)}\}$, taking values only in $\{u_1, \cdots, u_m\}$ and satisfying the conditions (1) and (2) in Lemma 2.1, such that if we denote

$$\lambda^{(k,l)}(t, A, \omega) \triangleq \int_0^t \delta_{Z_s(\omega)}(A) ds, \qquad (t, A, \omega) \in [0, T] \times \mathcal{B}(U) \times \Omega, \tag{3.37}$$

then $\{\lambda^{(k,l)}\}$ converges to $\lambda^{(k)}$, P-almost surely in $\mathcal{L}([0,T] \times U)$, as $l \to \infty$, for each k. Therefore, by a diagonalization scheme, we can find a subsequence of $\{\lambda^{(k,l)}\}$, still denoted by $\{\lambda^{(k)}\}$, such that each $\lambda^{(k)}$ is a representation of ordinary control, and that $\{\lambda^{(k)}\}$ converges P-almost surely to λ in the space $\mathcal{L}([0,T] \times U)$, as $k \to \infty$.

We now consider the SDE

$$\begin{aligned} \xi_t^{(k)} &= \xi_0 + \int_0^t b(\xi_s^{(k)}, Z_s^{(k)}) ds + \int_0^t \sigma(\xi_s^{(k)}, Z_s^{(k)}) dW_s \\ &= \xi_0 + \int_0^t \int_U b(\xi_s^{(k)}, u) \delta_{Z_s^{(k)}}(u) du ds + \int_0^t \int_U \sigma(\xi_s^{(k)}, u) \delta_{Z_s^{(k)}}(u) du dW_s. \end{aligned} \tag{3.38}$$

By Lemma 3.2, the sequence $\{\xi^{(k)}\}$ is tight, whence the family $\{(\xi^{(k)}, \lambda^{(k)}, W)\}$ is tight. Thus by Skorohod's Theorem ([11]), there exists another probability space $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{P})$, on which is defined a Brownian motion $\widehat{W}$, a pair $(\widehat{\xi}, \widehat{\lambda})$, and a sequence $\{(\widehat{\xi}^{(k)}, \widehat{\lambda}^{(k)})\}$, $k = 1, 2, \cdots$, such that $(\widehat{\xi}^{(k)}, \widehat{\lambda}^{(k)})$ and $(\xi^{(k)}, \lambda^{(k)})$ are identical in law for each k, and that $\{(\widehat{\xi}^{(k)}, \widehat{\lambda}^{(k)})\}$ converges $\widehat{P}$-almost surely to $(\widehat{\xi}, \widehat{\lambda})$ in $C([0,T]; \mathbb{R}^n) \times \mathcal{L}([0,T] \times U)$ as $k \to \infty$. It is also clear that $\widehat{\lambda}$ will have the same law as λ.

Note that each $\widehat{\lambda}^{(k)}$ can be written as

$$\widehat{\lambda}^{(k)}(t, A, \widehat{\omega}) = \widehat{\lambda}_{\widehat{\mu}^{(k)}}(t, A, \widehat{\omega}) = \int_0^t \widehat{\mu}_s^{(k)}(A, \omega) ds, \tag{3.39}$$

for some $\widehat{\mu}^{(k)} \in \mathbb{M}(\widehat{\Omega})$, and similarly, $\widehat{\lambda}$ can be written as $\widehat{\lambda}_{\widehat{\mu}}$ in the manner form as (3.39). Thanks to Lemma 2.1, we see that $\widehat{\mu}^{(k)}$ and $\mu^{(k)}$, $\widehat{\mu}$ and μ are all have the some law since they are at least left-continuous with right-limits. Therefore, each $\widehat{\lambda}^{(k)}$ will have the form

$$\widehat{\lambda}^{(k)}(t, A, \widehat{\omega}) = \int_0^t \delta_{\widehat{Z}^{(k)}(s, \widehat{\omega})}(A) ds,$$

for some U-valued process $\widehat{Z}^{(k)}$ defined on $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{P})$. Furthermore, on the probability space $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{P})$, $(\widehat{\xi}^{(k)}, \widehat{\mu}^{(k)}, \widehat{W})$ will satisfy (3.38), with $(\xi^{(k)}, \mu^{(k)}, W)$ being replaced by $(\widehat{\xi}^{(k)}, \widehat{\mu}^{(k)}, \widehat{W})$ since they are identical in law.

Finally, using the same technique as that in Theorem 3.3, we can prove that $\widehat{\xi}$ satisfies the SDE

$$\widehat{\xi}_t = \int_0^t \int_U b(\widehat{\xi}_s, u) \widehat{\mu}_s(du) ds + \int_0^t \int_U \sigma(\widehat{\xi}_s, u) \widehat{\mu}_s(du) d\widehat{W}_s,$$

whence $\widehat{\xi}$ and ξ are identical in law. Therefore, if we define $\widehat{\mathcal{A}}^{(k)} = (\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{P}, \widehat{\mathcal{F}_t}, \widehat{W}, \widehat{\mu}^{(k)})$ and $\widehat{\mathcal{A}} = (\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{P}, \widehat{\mathcal{F}_t}, \widehat{W}, \widehat{\mu})$, with $\widehat{\mathcal{F}}$ being properly defined to measure $\mu^{(k)}$, $\widehat{\mu}$ and $\widehat{W}$, then we have shown that for any $(s,x) \in [0,T] \times \mathbb{R}^n$,

$$J(s,x;\widehat{\mathcal{A}}^{(k)}) = J(s,x;\mathcal{A}^{(k)}); \quad J(s,x;\widehat{\mathcal{A}}) = J(s,x;\mathcal{A}). \tag{3.40}$$

This, together with the above argument, obviously leads to the first part of the theorem. The second assertion is obvious, so the theorem is proved. □

§4. The Equivalent Relations.

We now investigate the necessary and sufficient conditions for the solvability of (2.10)–(2.11). Let $T > 0$ be arbitrarily given. Since we are only looking for adapted solutions, let us assume *a priori* that (X, Y, μ) is adapted. In this case, we can rewrite, for any $t \geq \tau \geq 0$, the equation (2.11) as

$$Y_t = Y_\tau - \int_\tau^t \int_U b(X_s, Y_s, u)\mu_s(du)ds - \int_\tau^t \int_U \widehat{\sigma}(X_s, Y_s, u)\mu_s(du)dW_s. \tag{4.1}$$

In particular, setting $\tau = 0$ and combine with (2.10), we see that solving (2.10)–(2.11) becomes the following problem:

Problem (SDE). For any $x \in \mathbb{R}^n$, find a probability space $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$, an $\mathbb{R}^m$-valued $\mathcal{F}_0$-measurable random variable Y_0, an $\{\mathcal{F}_t\}$-Brownian motion defined on this space; and an $\{\mathcal{F}_t\}$-adapted, $\mathcal{P}(U)$-valued process μ such that the following SDE with terminal condition:

$$X_t = x + \int_0^t \int_U b(X_s, Y_s, u)\mu_s(du)ds + \int_0^t \int_U \sigma(X_s, Y_s, u)\mu_s(du)dW_s, \tag{4.2}$$

$$Y_t = Y_0 - \int_0^t \int_U \widehat{b}(X_s, Y_s, u)\mu_s(du)ds - \int_0^t \int_U \widehat{\sigma}(X_s, Y_s, u)\mu_s(du)dW_s; \tag{4.3}$$

$$Y_T = g(X_T), \tag{4.4}$$

has an $\{\mathcal{F}_t\}$-adapted solution (X, Y) on $(\Omega, \mathcal{F}, P)$.

Remark 4.1. One should note that Problem (SDE) is not easy in general, since it is essentially a two-point boundary value problem for SDEs. It is quite possible that it *does not* have any solution of any kind.

To derive our equivalent relation we design the following stochastic relaxed control problem. For any $\eta \in \mathbb{R}^{n+m}$, let $\eta = (x^T, y^T)^T$, where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ and the superscript " T " means transpose. To simplify notation, we henceforth write $\eta = (x, y)$. Let $\xi \triangleq (X, Y)$; also define $\tilde{b} = (b, -\widehat{b})$ and $\tilde{\sigma} = (\sigma, -\widehat{\sigma})$ in a similar way.

Rewrite (4.2) and (4.3) as a single *state equation* starting from an arbitrary initial time $s \in [0, T]$ and initial state η_0 which is assumed to be $\mathcal{F}_s$-measurable:

$$\xi_t = \eta_s + \int_s^t \int_U \tilde{b}(\xi_r, u)\mu_r(du)dr + \int_s^t \int_U \tilde{\sigma}(\xi_r, u)\mu_r(du)dW_r. \tag{4.5}$$

Define $G(\eta) = |g(x) - y|^2$, then obviously G satisfies condition (A.2) in §3. Next, we define the cost functional to be:

$$J(s, \eta_s; \mathcal{A}) = E^P G(\xi_T(s, \eta_s, \mathcal{A})), \tag{4.6}$$

where $\mathcal{A} = (\Omega, \mathcal{F}, P, \mathcal{F}_t, W, \mu)$. If we use the notation $\mathcal{U}^R$ in §2.2, then the value function for the optimal (relaxed) control problem is defined as

$$V(s, x, y) = \inf_{\mathcal{A} \in \mathcal{U}^R} J(s, x, y; \mathcal{A}). \tag{4.7}$$

Intuitively, by introducing the above stochastic relaxed control problem, the Problem (SDE) should be replaced by the following two problems.

Problem (N) For any $x \in \mathbb{R}^n$, find a $y \in \mathbb{R}^m$ such that $V(0, x, y) = 0$;

Problem (C) For any $(x, y) \in \mathbb{R}^{n+m}$, find an $\mathcal{A}^*_{x,y} \in \mathcal{U}^R$, such that

$$J(0, x, y; \mathcal{A}^*_{x,y}) = V(0, x, y). \tag{4.8}$$

We now validate the above idea. First, note that by Theorem 3.6, Problem (C) is always solvable under our assumptions (A.1)–(A.2), we have the following equivalence relations.

Theorem 4.2. Assume that the assumptions (A.1)–(A.2) hold. Then, the following statements are equivalent:

(1) The forward-backward SDE (2.10)–(2.11) is solvable over $[0, T]$;

(2) The Problem (SDE) is solvable;

(3) The Problem (N) is solvable.

Proof. (1)⇒(2). Let $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ be the probability space on which the adapted solution (X, Y, μ) of (2.10)–(2.11) is defined. Rewrite (2.11) in a forward version as (4.1).

Since Y is adapted, Y_0 is $\mathcal{F}_0$-measurable. The terminal condition (4.4) is trivially true, so (1) implies (2).

(2)$\Rightarrow$(3). Let $x \in \mathbb{R}^n$ be given. Suppose $(\Omega, \mathcal{F}, P, \mathcal{F}_t)$ is the probability space on which the solution of Problem (SDE) is solvable. Let (X, Y, μ) be the solution of (4.2)—(4.4). Then we must have $J(0, x, Y_0; \mathcal{A}) = 0$. Now let $\pi \in \mathcal{P}(\mathbb{R}^m)$ be the distribution of Y_0 and P_y be the regular conditional probability of P given $Y_0 = y$. Denote $\mathcal{A}^y = (\Omega, \mathcal{F}, P_y, \mathcal{F}_t, W, \mu)$, then it is clear that $\mathcal{A}^y \in \mathcal{U}^R$ and for π-a.e. $y \in \mathbb{R}^m$, $\xi_t(0, x, Y_0; \mathcal{A}) = \xi_t(0, x, y; \mathcal{A}^y)$ for all $t \in [0, T]$, P_y-a.s. , where $\xi = (X, Y)$. Therefore,

$$\begin{aligned} J(0, x, Y_0; \mathcal{A}) &= E^P[g(X_T(0, x, Y_0; \mathcal{A})) - Y_T(0, x, Y_0; \mathcal{A})]^2 \\ &= \int_{\mathbb{R}^m} E^{P_y}[g(X_T(0, x, y; \mathcal{A}^y)) - Y_T(0, x, y; \mathcal{A}^y)]^2 \pi(dy) \\ &= \int_{\mathbb{R}^m} J(0, x, y; \mathcal{A}^y) \pi(dy). \end{aligned} \tag{4.9}$$

Since $J(0, x, Y_0; \mathcal{A}) = 0$ and $J(0, x, y; \mathcal{A}^y) \geq 0$ for any $y \in \mathbb{R}^m$, we have $J(0, x, y, \mathcal{A}^y) = 0$, for π-a.e. $y \in \mathbb{R}^m$. Finally, since $V(0, x, y) \geq 0$ for all $(x, y) \in \mathbb{R}^{n+m}$, we obtain that $V(0, x, y) = 0$, for π-a.e. $y \in \mathbb{R}^n$. Evidently this implies the solvability of Problem (N).

(3)$\Rightarrow$(1). Suppose that Problem (N) is solvable. Then for any $x \in \mathbb{R}^n$, choose $y \in \mathbb{R}^m$, such that $V(0, x, y) = 0$. Now for such (x, y), let $\mathcal{A}^*_{x,y} \in \mathcal{U}^R$ be an optimal relaxed control, and (X^*, Y^*) be the optimal trajectory (*i.e.*, the solution of (4.2)–(4.3)). Then we must have

$$E^P|g(X^*_T) - Y^*_T|^2 = J(0, x, y; \mathcal{A}^*) = V(0, x, y) = 0. \tag{4.10}$$

Namely the terminal condition (4.4) is also satisfied. Rewrite (4.3) backwardly as (2.11), we see that (X^*, Y^*, μ^*) solves forward-backward equation on $[0, T]$. $\square$

We now consider the following HJB-equation. Recall that $\xi = (x, y)$.

$$0 = v_t + \min_{u \in U} \left\{ \langle\, v_\xi, \tilde{a}(\xi, u) \,\rangle + \frac{1}{2} \mathrm{tr}[\tilde{\sigma}(\xi, u) \tilde{\sigma}(\xi, u)^T v_{\xi\xi}] \right\}, \tag{4.11}$$

with the terminal condition

$$v(T, x, y) = |g(x) - y|^2. \tag{4.12}$$

It is well known (cf. [8]) that under our assumptions on the data, the above HJB-equation (4.11)–(4.12) has a unique continuous viscosity solution. Denote this solution by $v(t, x, y)$. We have the following theorem.

Theorem 4.3. *Let (A.1)–(A.2) hold. Then, the forward-backward SDE (2.10)–(2.11) is solvable over* $[0,T]$ *if and only if for any* $x \in \mathbb{R}^n$*, there exists a* $y \in \mathbb{R}^m$ *such that* $v(0,x,y)=0$*, where* v *is the viscosity solution of (4.11)–(4.12). In another word, the nodal set* $\mathcal{N}(v)$ *of* v *contains the point* $(0,x,y)$ *for some* $y \in \mathbb{R}^m$*.*

Proof. Consider the ordinary optimal control problem with state equation:

$$
(4.13)\qquad \begin{cases} X_t = x + \int_s^t b(X_r,Y_r,Z_r)dr + \int_s^t \sigma(X_r,Y_r,Z_r)dW_r; \\ Y_t = y + \int_s^t \widehat{b}(X_r,Y_r,Z_r)dr + \int_s^t \widehat{\sigma}(X_r,Y_r,Z_r)dW_r, \end{cases} \qquad t \in [s,T],
$$

with cost functional:

$$
(4.14)\qquad J(s,x,y;Z) = E^P|g(X_T) - Y_T|^2.
$$

The value function is defined to be

$$
(4.15)\qquad V^0(s,x,y) = \inf J(s,x,y;Z),
$$

where the infimum is taken over all the U-valued adapted process Z and all the underlying probability spaces. By the Chattering Lemma (Theorem 3.6), we have $V^R(s,x,y) = V^0(s,x,y)$ for all $(s,x,y) \in [0,T]\times\mathbb{R}^n\times\mathbb{R}^m$. On the other hand, it is known that the value function of the ordinary stochastic control problem is a viscosity solution of the HJB-equation (4.11)–(4.12) (cf. [8]). Since such a viscosity solution is unique, we must have

$$
(4.16)\qquad V^R(s,x,y) = V^0(s,x,y) = v(s,x,y).
$$

Thus the theorem follows from Theorem 4.2. □

§5. Solvability of a Class of Forward-Backward SDEs.

In this section, we try to solve a class of forward-backward SDEs by using the results of previous sections. Suppose a probability space $(\Omega,\mathcal{F},P;\mathcal{F}_t)$ and an $\{\mathcal{F}_t\}$-Brownian motion W is given. Consider the following forward-backward SDE:

$$
(5.1)\qquad \begin{cases} X_t = x + \int_0^t b(X_s,Y_s)ds + \int_0^t \sigma(X_s,Y_s)dW_s, \\ Y_t = g(X_T) + \int_t^T \widehat{b}(X_s,Y_s)ds + \int_t^T \int_U \widehat{\sigma}(X_s,Y_s,u)\mu_s(du)dW_s, \end{cases} \qquad t\in[0,T],
$$

where U is a compact subset of $\overline{\mathbb{R}^{n\times m}}$.

For simplicity, let us assume that all the processes and functions appeared in (5.1) are scalar valued; the higher dimensional case is basically the same (see Remark 6.6). We claim that if the filtration $\{\mathcal{F}_t\}$ is actually generated by the Brownian motion W, then the SDE of the type (5.1) is actually equivalent to the following form which is of special interest in mathematical finance (see, for example [11] and the references therein).

$$
\begin{cases} X_t = x + \int_0^t b(X_s, Y_s)ds + \int_0^t \sigma(X_s, Y_s)dW_s, \\ Y_t = E\left\{ g(X_T) + \int_t^T \widehat{b}(X_s, Y_s)ds \,\middle|\, \mathcal{F}_t \right\}, \end{cases} \qquad t \in [0, T]. \tag{5.2}
$$

To verify our claim, we first note that if (X, Y, Z) is an adapted solution of (5.1) in the usual sense ([27,28]), then by simply taking the condition expectation $E\{\cdot|\mathcal{F}_t\}$ on both sides of the second equation for each $t \in [0, T]$, we see immediately that (X, Y) is an adapted solution of (5.2). Conversely, suppose that (5.2) has an adapted solution (X, Y) which is square integrable, we shall prove that there exists an $\{\mathcal{F}_t\}$-adapted, square integrable process Z such that (X, Y, δ_Z) is a solution of a forward-backward SDE of the type (5.1). To see this, consider the square integrable martingale

$$
M_t = E\left\{ g(X_T) + \int_0^T \widehat{b}(X_s, Y_s)ds \,\middle|\, \mathcal{F}_t \right\}, \qquad t \geq 0. \tag{5.3}
$$

By the martingale representation theorem (see for example, [19, p.182]), there exists a square integrable, $\{\mathcal{F}_t\}$-adapted process Z such that

$$
M_t = M_0 + \int_0^t Z_s dW_s, \qquad \forall t \geq 0. \tag{5.4}
$$

Since

$$
\begin{aligned} & g(X_T) + \int_0^T \widehat{b}(X_s, Y_s)ds = M_0 + \int_0^T Z_s dW_s, \\ & Y_t = M_t - \int_0^t \widehat{b}(X_s, Y_s)ds, \end{aligned} \tag{5.5}
$$

a simple computation leads to that

$$
Y_t = g(X_T) + \int_t^T \widehat{b}(X_s, Y_s)ds - \int_t^T Z_s dW_s, \tag{5.6}
$$

which is exactly of the form (5.1) with $\widehat{\sigma}(x,y,z)\equiv z$. It is obvious that our equation (5.1) is nontrivially more general.

We now use our scheme designed in the previous sections to solve the SDE (5.1). Let Z_t take values in some compact set $U\subset \mathbb{R}$, which will be determined later. Then, we formulate an stochastic optimal relaxed control problem with the state equation

$$
\left\{\begin{aligned}
X_t &= x+\int_s^t b(X_r,Y_r)dr+\int_s^t \sigma(X_r,Y_r)dW_r,\\
Y_t &= y-\int_s^t \widehat{b}(X_r,Y_r)dr-\int_s^t\int_U \widehat{\sigma}(X_r,Y_r,u)\mu_r(du)dW_r,
\end{aligned}\right.\qquad t\in[s,T], \tag{5.7}
$$

and the cost functional

$$
J(s,x,y;\mu)=E|Y_T-g(X_T)|^2. \tag{5.8}
$$

We define the value function $V(s,x,y)$ as

$$
V(s,x,y)=\inf_{\mu} J(s,x,y;\mu). \tag{5.9}
$$

Then, we know that $V(s,x,y)$ is the unique viscosity solution of the following HJB-equation:

$$
\left\{\begin{aligned}
&V_t+H(x,y,V_x,V_y,V_{xx},V_{xy},V_{yy})=0,\\
&V(T,x,y)=(y-g(x))^2,
\end{aligned}\right. \tag{5.10}
$$

where

$$
\begin{aligned}
H(x,y,V_x,V_y,V_{xx},V_{xy},V_{yy}) &= b(x,y)V_x-\widehat{b}(x,y)V_y+\frac{1}{2}\sigma(x,y)^2V_{xx}\\
&\quad+\inf_{z\in U}\{-\sigma(x,y)\widehat{\sigma}(x,y,z)V_{xy}+\frac{1}{2}\widehat{\sigma}(x,y,z)^2V_{yy}\}.
\end{aligned} \tag{5.11}
$$

Theorem 4.3 tells us that the solvability of (5.1) is equivalent to the following problem: *For each $x\in\mathbb{R}$, find a $y\in\mathbb{R}$, such that*

$$
V(0,x,y)=0. \tag{5.12}
$$

Recall that (see [7,10,17,24]), the nodal set $\mathcal{N}(V)$ of V is defined to be

$$
\mathcal{N}(V)=\{(s,x,y)\in[0,\infty)\times\mathbb{R}^2 \mid V(s,x,y)=0\}. \tag{5.13}
$$

Thus, (5.1) is solvable if and only if the nodal set $\mathcal{N}(V)$ of V intersects each set of form $\{(0,x)\} \times \mathbb{R}$. Here, we should note that the value function actually depends on the the time duration $T > 0$. Now, let us set

$$v(t,x,y) = V(T-t,x,y), \qquad (t,x,y) \in [0,T] \times \mathbb{R} \times \mathbb{R}. \tag{5.14}$$

Then, v is the unique viscosity solution of the following ([8]):

$$\begin{cases} v_t - \frac{1}{2}\sigma(x,y)^2 v_{xx} - b(x,y)v_x + \widehat{b}(x,y)v_y \\ \qquad\qquad - \inf_{z \in U}\{-\sigma(x,y)\widehat{\sigma}(x,y,z)v_{xy} + \frac{1}{2}\widehat{\sigma}(x,y,z)^2 v_{yy}\} = 0, \\ v(0,x,y) = (y-g(x))^2. \end{cases} \tag{5.15}$$

The advantage of (5.15) is that this problem is posed on $(0,\infty)\times\mathbb{R}\times\mathbb{R}$, and the solvability of (5.1) over any $[0,T]$ is equivalent to the following statement: *For any* $(t,x) \in [0,\infty)\times\mathbb{R}$, *there exists* a $y \in \mathbb{R}$, *such that*

$$v(t,x,y) = 0. \tag{5.16}$$

Clearly, one way to do this is to find a function $\theta : [0,\infty) \times \mathbb{R} \to \mathbb{R}$, such that

$$v(t,x,\theta(t,x)) = 0, \qquad \forall (x,t) \in [0,\infty) \times \mathbb{R}; \tag{5.17}$$

or equivalently,

$$\{(t,x,\theta(t,x)) \bigm| (t,x) \in [0,\infty) \times \mathbb{R}\} \subset \mathcal{N}(v). \tag{5.18}$$

Sometimes, such a hypersurface $y = \theta(t,x)$ is called a nodal surface of V ([7]). In the rest of this section, we are going to give some intuitive arguments of constructing such a nodal set. The existence of such a nodal set implies the solvability of the forward backward SDE (5.1).

Suppose $v(t,x,y)$ is a classical solution of (5.15) and $\theta(t,x)$ is an undetermined smooth function with

$$\theta(0,x) = g(x), \qquad x \in \mathbb{R}. \tag{5.19}$$

We define

$$w(t,x) = v(t,x,\theta(t,x)), \qquad (t,x) \in (0,\infty) \times \mathbb{R}. \tag{5.20}$$

Then,

$$w(0,x)=0, \qquad x\in\mathbb{R}. \tag{5.21}$$

On the other hand, by (5.20), we have, at $(t,x,\theta(t,x))$, that

$$\begin{cases} w_t = v_t + v_y\theta_t, \\ w_x = v_x + v_y\theta_x \\ w_{xx} = v_{xx} + 2v_{xy}\theta_x + v_{yy}\theta_x^2 + v_y\theta_{xx}. \end{cases} \tag{5.22}$$

Then, by the equation in (5.15), we obtain

$$\begin{aligned} 0 =& w_t - \frac{1}{2}\sigma^2 w_{xx} - bw_x - (\theta_t - \frac{1}{2}\sigma^2\theta_{xx} - b\theta_x - \widehat{b})v_y \\ &- \inf_{z\in U}\{(\widehat{\sigma}+\sigma\theta_x)(-\sigma v_{xy} + \frac{1}{2}(\widehat{\sigma}-\sigma\theta_x)v_{yy})\}. \end{aligned} \tag{5.23}$$

Here, v_y, v_{xy} and v_{yy} are evaluated at $(t,x,\theta(t,x))$, $b, \widehat{b}$ and σ at $(x,\theta(t,x))$ and $\widehat{\sigma}$ at $(t,x,\theta(t,x),z)$. Now, we take the function $\theta(t,x)$ to be the classical solution of the following problem: (assuming, for the time being, such a solution exists)

$$\begin{cases} \theta_t - \frac{1}{2}\sigma(x,\theta)^2\theta_{xx} - b(x,\theta)\theta_x - \widehat{b}(x,\theta) = 0, \\ \theta(0,x) = g(x). \end{cases} \tag{5.24}$$

Further, we assume that

$$0 \in \{\widehat{\sigma}(x,\theta(t,x),z) + \sigma(x,\theta(t,x))\theta_x(t,x) \bigm| z\in U\}. \tag{5.25}$$

Then, from (5.23)–(5.25), we see that

$$w_t - \frac{1}{2}\sigma^2 w_{xx} - bw_x \le 0. \tag{5.26}$$

Hence, by (5.21) and maximum principle ([23]),

$$w(t,x)\le 0, \qquad (t,x)\in(0,\infty)\times\mathbb{R}. \tag{5.27}$$

However, by definition we know that $w(t,x)$ is nonnegative. Hence, we obtain

$$v(t,x,\theta(t,x)) \equiv w(t,x) = 0, \qquad (t,x)\in(0,\infty)\times\mathbb{R}. \tag{5.28}$$

Thus, $\theta(t,x)$ is a nodal surface of v.

§6. Nodal Sets of HJB Equations.

In this section, we will make the heuristic arguments given in the previous section rigorous.

First, we recall some standard notations. For any bounded or unbounded region $G \subseteq \mathbb{R}^n$, We let $C(\bar{G})$ be the set of all bounded continuous functions defined on $\bar{G}$ ($\overline{\mathbb{R}^n} = \mathbb{R}^n$) with the norm

$$\|w\|_{C(\bar{G})} = \max_{x \in \bar{G}} |w(x)|, \qquad \forall w \in C(\bar{G}).$$

Then, we let $C^2(\bar{G})$ be the set of all bounded twice continuously differentiable functions defined on $\bar{G}$ with the norm

$$\|w\|_{C^2(\bar{G})} = \|w\|_{C(\bar{G})} + \|w_x\|_{C(\bar{G})} + \|w_{xx}\|_{C(\bar{G})}, \qquad \forall w \in C^2(\bar{G}).$$

Here w_x and w_{xx} stand for the gradient and the Hessian of w, respectively. For $\alpha \in (0,1)$, we define $C^{2+\alpha}(\bar{G})$ be the set of all elements in $C^2(\bar{G})$ such that the second partial derivatives are Hölder continuous with the exponent α. The norm in $C^{2+\alpha}(\bar{G})$ is defined to be

$$\|w\|_{C^{2+\alpha}(\bar{G})} = \|w\|_{C^2(\bar{G})} + \sup_{x \neq x', x, x' \in \bar{G}} \frac{|w_{xx}(x) - w_{xx}(x')|}{|x - x'|^\alpha}.$$

Next, for any $T > 0$ and any bounded or unbounded region $G \subseteq \mathbb{R}^n$, denote $Q_T = (0,T) \times G$. Let $C^{2+\alpha, 1+\alpha/2}(\bar{Q}_T)$ be the space of all functions $\theta(t,x)$ which are differentiable in t and twice differentiable in x with θ_t and θ_{xx} being $\alpha/2$- and α-Hölder continuous in $(t,x) \in \bar{Q}_T$, respectively. In $C^{2+\alpha,1+\alpha/2}(\bar{Q}_T)$, we define the norm to be

$$\begin{aligned} \|\theta\|_{C^{2+\alpha,1+\frac{\alpha}{2}}(\bar{Q}_T)} &= \|\theta\|_{C(\bar{Q}_T)} + \|\theta_t\|_{C(\bar{Q}_T)} + \|\theta_x\|_{C(\bar{Q})} + \|\theta_{xx}\|_{C(\bar{Q})} \\ &+ \sup_{(t,x),(t'x') \in \bar{Q}, (t,x) \neq (t',x')} \frac{|\theta_t(t,x) - \theta_t(t',x')| + |\theta_{xx}(t,x) - \theta_{xx}(t',x')|}{(|x-x'|^2 + |t-t'|)^{\alpha/2}}. \end{aligned}$$

Now, let us make some hypotheses.

(H1) Functions $b, \widehat{b}, \sigma$ are $C^2(\mathbb{R}^2)$ and g is $C^{2+\alpha}(\mathbb{R})$ (for some $\alpha \in (0,1)$) with

$$\|b\|_{C^2(\mathbb{R}^2)} + \|\widehat{b}\|_{C^2(\mathbb{R}^2)} + \|\sigma\|_{C^2(\mathbb{R}^2)} + \|g\|_{C^{2+\alpha}(\mathbb{R})} \le C. \tag{6.1}$$

Moreover, there exists a constant $\nu > 0$, such that

$$\sigma(x,y)^2 \ge \nu, \qquad \forall (x,y) \in \mathbb{R}^2. \tag{6.2}$$

(H2) Function $\widehat{\sigma}$ is continuous. For each $z \in \mathbb{R}$, $\widehat{\sigma}(\cdot,\cdot,z)$ is in $C^2(\mathbb{R}^2)$ with

$$\|\widehat{\sigma}(\cdot,\cdot,z)\|_{C^2(\mathbb{R}^2)} \le C_R, \qquad \forall |z| \le R. \tag{6.3}$$

(H3) Function $\widehat{\sigma}$ satisfies

$$\{\widehat{\sigma}(x,y,z) \bigm| z \in \mathbb{R}\} = \mathbb{R}, \qquad \forall (x,y) \in \mathbb{R}^2. \tag{6.4}$$

Remark. The regularity of $b, \widehat{b}, \sigma$ and $\widehat{\sigma}$ might be relaxed. In order to to this, some arguments of [30] should be adopted. Also, the regularity of g can be relaxed as well. In this case, the solution θ of (5.24) will be less regular near $t = 0$. We also point out that (H1)–(H2) implies the state equation (5.7) satisfies (A.1) of §3 (for $z \in U$ with U being a compact set in $\mathbb{R}$).

The following result concerns the well-posedness of (5.24). A proof can be found in [9,23].

Lemma 6.1. *Let (H1) hold. Then, (5.24) admits a unique solution $\theta(t,x)$ in $C^{2+\alpha,1+\alpha/2}$ $([0,\infty)\times\mathbb{R}^2)$. In particular, for any $T>0$,*

$$\sup_{x\in\mathbb{R},t\in[0,T]} |\theta_x(t,x)| < \infty. \tag{6.5}$$

Now, we come up with the first main result of this section.

Theorem 6.2. *Let (H1)–(H3) hold. Let $v(t,x,y)$ be the viscosity solutions of (5.15) and $\theta(t,x)$ be the solution of (5.24). Then, the nodal set $\mathcal{N}(v)$ of $v(t,x,y)$ is given by*

$$\mathcal{N}(v) = \{(t,x,\theta(t,x)) \bigm| (t,x) \in [0,\infty)\times\mathbb{R}\}. \tag{6.6}$$

Proof. For any $\varepsilon > 0$, we introduce the following problem:

$$\begin{cases} v_t^\varepsilon - \dfrac{1}{2}\sigma(x,y)^2 v_{xx}^\varepsilon - b(x,y)v_x^\varepsilon + \widehat{b}(x,y)v_y^\varepsilon \\ \qquad\qquad - \inf\limits_{z\in U}\{-\sigma(x,y)\widehat{\sigma}(x,y,z)v_{xy}^\varepsilon + \dfrac{1}{2}\widehat{\sigma}(x,y,z)^2 v_{yy}^\varepsilon\} - \varepsilon v_{yy}^\varepsilon = 0, \\ v^\varepsilon(0,x,y) = (y-g(x))^2. \end{cases} \tag{6.7}$$

Here, we take U to be a compact set in $\mathbb{R}$, such that

$$\begin{gathered} \inf_{z\in U}\widehat{\sigma}(x,\theta(x,t),z) \le \sigma(x,\theta(t,x))\theta_x(t,x) \le \sup_{z\in U}\widehat{\sigma}(x,\theta(t,x),z), \\ \forall (t,x) \in [0,T]\times\mathbb{R}. \end{gathered} \tag{6.8}$$

This is possible due to (H3) and the boundedness of $\sigma(x,y)$ and $\theta_x(t,x)$ (see (6.1) and (6.5)). Then, we know that (6.7) is a nondegenerate fully nonlinear parabolic equation. From [21] (see [13] also), we know that there exists a classical solution v^ε of (6.7). On the other hand, this v^ε is the value function of the optimal stochastic control problem similar to (5.7)–(5.8) in which a term $\sqrt{\varepsilon/2}\,dW'_r$ is added in the second equation of (5.7) with W' being another Brownian motion independent of W. Thus, by [20], we can find a continuous function $K(t,x,y)>0$, independent of $\varepsilon>0$, such that

$$v^\varepsilon_{yy}(t,x,y)\le K(t,x,y),\qquad \forall (t,x,y)\in[0,\infty)\times\mathbb{R}^2,\quad \varepsilon>0. \tag{6.9}$$

Now, we set

$$w^\varepsilon(t,x)=v^\varepsilon(t,x,\theta(t,x)),\qquad (t,x)\in[0,\infty)\times\mathbb{R}. \tag{6.10}$$

Then, similar to (5.22), we have

$$\begin{aligned}
&w^\varepsilon_t-\frac{\sigma(x,\theta(t,x))^2}{2}w^\varepsilon_{xx}-b(x,\theta(t,x))w^\varepsilon_x\\
&=\inf_{z\in U}\{(\widehat\sigma(x,\theta(t,x),z)+\sigma(x,\theta(t,x))\theta_x(x,t))[-\sigma(x,\theta(t,x))v^\varepsilon_{xy}(t,x,\theta(t,x))\\
&\quad+\frac{1}{2}\big(\widehat\sigma(x,\theta(t,x),z)-\sigma(x,\theta(t,x))\theta_x(t,x)\big)v^\varepsilon_{yy}(t,x,\theta(t,x))]\}+\varepsilon v^\varepsilon_{yy}(x,t,\theta(t,x))\\
&\le\varepsilon K(t,x,\theta(x,t)).
\end{aligned} \tag{6.11}$$

Here, we have used the facts (6.8) and (6.9). Hence, the function $w^\varepsilon(t,x)$ defined by (6.10) satisfies the following (in the classical sense and thus in the viscosity sense):

$$\begin{cases} w^\varepsilon_t-\dfrac{\sigma(x,\theta(t,x))^2}{2}w^\varepsilon_{xx}-b(x,\theta(t,x))w^\varepsilon_x\le\varepsilon K(t,x,y),\\ w^\varepsilon(0,x)=(g(x)-\theta(0,x))^2=0.\end{cases} \tag{6.12}$$

On the other hand, by [8], we know that $v^\varepsilon(t,x,y)$ converges to the unique viscosity solution $v(t,x,y)$ of (5.15) uniformly in any compact sets. Thus, we see that $w^\varepsilon(t,x)$ converges to $w(t,x)=v(t,x,\theta(t,x))$ uniformly in any compact sets. Then, by [8] again, this $w(t,x)$ is a viscosity solution of

$$\begin{cases} w_t-\dfrac{\sigma(x,\theta(t,x))^2}{2}w_{xx}-b(x,\theta(t,x))w_x\le 0,\\ w(0,x)=0.\end{cases} \tag{6.13}$$

Therefore, by a comparison theorem, we must have $w(t,x)\le 0$. But, we know that $v(t,x,y)$ is nonnegative. Hence, we have

$$\mathcal{N}(v)\supseteq\{(t,x,\theta(t,x))\bigm|(t,x)\in[0,\infty)\times\mathbb{R}\}. \tag{6.14}$$

On the other hand, let us set

$$\varphi(t,x,y)=(y-\theta(t,x))^2,\qquad (t,x,y)\in[0,\infty)\times\mathbb{R}^2. \tag{6.15}$$

Then, by (5.24), we have

$$\begin{aligned}
&\varphi_t-\frac{\sigma(x,y)^2}{2}\varphi_{xx}-b(x,y)\varphi_x+\widehat{b}(x,t)\varphi_y\\
&\qquad-\inf_{z\in U}[-\sigma(x,y)\widehat{\sigma}(x,y,z)\varphi_{xy}+\frac{\sigma(x,y)^2}{2}\varphi_{yy}]\\
&=-\inf_{z\in U}[\sigma(x,y)\theta_x(t,x)+\widehat{\sigma}(x,y,z)]^2\le 0.
\end{aligned} \tag{6.16}$$

Also, we have $\varphi(0,x,y)=(y-g(x))^2$. Thus, $\varphi(t,x,y)$ is a viscosity subsolution of (5.15). Then, by a comparison theorem, we obtain that

$$(y-\theta(t,x))^2\le v(t,x,y),\qquad \forall(t,x,y)\in[0,\infty)\times\mathbb{R}^2. \tag{6.17}$$

This means that

$$\mathcal{N}(v)\subseteq\{(t,x,\theta(t,x))\bigm|(t,x)\in[0,\infty)\times\mathbb{R}\}. \tag{6.18}$$

Hence, (6.6) follows from (6.14) and (6.18). $\square$

The above result together with the results of §4 will give the following.

Theorem 6.3. *Let (H1)–(H3) hold. Then, for any $T>0$ and $x\in\mathbb{R}$, the forward-backward SDE (5.1) is solvable.*

We note that since the optimal (relaxed) controls are not necessarily unique, we do not have the uniqueness of the solutions of (5.1).

Next result is related the condition (6.4).

Theorem 6.4. *Let (H1)-(H2) hold. Suppose one of the following holds:*

(i) *The continuity of $\widehat{\sigma}$ in y is uniform in all its arguments. Moreover, the following holds:*

$$\inf_{x,z\in\mathbb{R}}[\sigma(x,g(x))g'(x)+\widehat{\sigma}(x,g(x),z)]^2>0; \tag{6.19}$$

(ii) *The following holds:*

$$\inf_{x,y,z\in\mathbb{R}}[\sigma(x,y)g'(x)+\widehat{\sigma}(x,y,z)]^2>0. \tag{6.20}$$

Then, there exists a $T_0>0$, such that for any $T\in(0,T_0]$, (5.1) is not solvable on $[0,T]$.

Proof. First, we let (i) holds. Then, there exist a $\delta > 0$ and $T_0 > 0$, such that

$$\inf_{t\in[0,T_0],x\in\mathbb{R}}[\sigma(x,y)\theta_x(t,x)) + \widehat\sigma(x,y,z)]^2 \geq \delta, \qquad \forall |y-g(x)| \leq \delta. \tag{6.21}$$

Then, we can find a function $h \in C^\infty(\mathbb{R}^2)$ such that

$$\begin{cases} 0 \leq h(x,y) \leq 1, \qquad (x,y) \in \mathbb{R}^2, \\ h(x,g(x)) = 1, \qquad \forall x \in \mathbb{R}, \\ h(x,y) = 0, \qquad \forall |y-g(x)| \geq \delta. \end{cases} \tag{6.22}$$

Now, we define

$$\psi(t,x,y) = \varepsilon t h(x,y) + (y-g(x))^2, \qquad (t,x,y) \in [0,\infty) \times \mathbb{R}^2, \tag{6.23}$$

with $\varepsilon > 0$ being undetermined. Similar to (6.16), we can obtain

$$\begin{aligned} &\psi_t - \frac{\sigma(x,y)^2}{2}\psi_{xx} - b(x,y)\psi_x + \widehat b(x,t)\psi_y \\ &\quad - \inf_{z\in U}[-\sigma(x,y)\widehat\sigma(x,y,z)\psi_{xy} + \frac{\sigma(x,y)^2}{2}\psi_{yy} \\ &= \varepsilon h(x,y) - \frac{\sigma(x,y)^2}{2}\varepsilon h_{xx}(x,y)t - b(x,y)\varepsilon h_x(x,y)t + \widehat b(x,y)\varepsilon h_y(x,y)t \\ &\quad - \inf_{z\in U}\{[\sigma(x,y)\theta_x(t,x) + \widehat\sigma(x,y,z)]^2 - \sigma(x,y)\widehat\sigma(x,y,z)\varepsilon h_{xy}(x,y)t \\ &\quad + \frac{\widehat\sigma(x,y,z)}{2}\varepsilon h_{yy}(x,y)t\} \leq 0, \end{aligned} \tag{6.24}$$

provided $\varepsilon > 0$ is sufficiently small and $(t,x) \in [0,T_0] \times \mathbb{R}$. Also, we have $\psi(0,x,y) = (y-g(x))^2$. Hence, it follows that

$$\varepsilon h(x,y)t + (y-\theta(t,x))^2 \leq v(t,x,y), \qquad (t,x,y) \in [0,T_0] \times \mathbb{R}^2. \tag{6.25}$$

This implies that

$$\mathcal{N}(v)\bigcap\{(0,T_0] \times \mathbb{R}^2\} = \emptyset. \tag{6.26}$$

Hence, (5.1) has no solutions on any $[0,T]$ with $T \in (0,T_0]$.

Now, in the case (ii) holds, we can similarly prove (6.26) by taking $\psi(t,x,y) = \varepsilon t + (y-g(x))^2$ for some sufficiently small $\varepsilon > 0$. $\square$

All the above results are sort of global. We next look at some local results. To this end, we let $x \in \mathbb{R}$. Define

$$T_x = \sup\{t > 0 \mid \inf_{z\in\mathbb{R}}[\sigma(x,\theta(t,x))\theta_x(t,x) + \widehat\sigma(x,\theta(t,x),z)]^2 = 0\}, \tag{6.27}$$

with $\theta(t,x)$ being the solution of (5.24). The following result gives a local version of Theorem 6.2.

Theorem 6.5. *Let (H1) and (H2) hold. Then,*

$$\mathcal{N}(v)\bigcap\{[0,T_x]\times\mathbb{R}^2\}=\{(t,x,\theta(t,x))\bigm| t\in[0,T_x]\ \},\qquad \forall x\in\mathbb{R}. \tag{6.28}$$

Consequently, for any $x\in\mathbb{R}$, (5.1) is solvable on $[0,T_x]$ and there exists a $\delta_0>0$, such that (5.1) is not solvable on $[0,T_x+\delta]$ for any $\delta\in(0,\delta_0]$.

The proof is clear.

Remark 6.6. Although we only discussed the case in which all the functions involved in (5.1) are scalar valued, it is not hard to see that our arguments are good for higher dimensions. In this case, some assumptions should be accordingly changed. For example, (6.2) and (6.4) should be replaced respectively by

$$\sigma(x,y)\sigma(x,y)^T\ge\nu I,\qquad \forall(x,y)\in\mathbb{R}^n\times\mathbb{R}^m, \tag{6.29}$$

and

$$\{\widehat{\sigma}(x,y,z)\bigm| z\in\mathbb{R}^\ell\}=\mathbb{R}^m,\qquad \forall(x,y)\in\mathbb{R}^n\times\mathbb{R}^m, \tag{6.30}$$

provided we assume x, y and z to be in $\mathbb{R}^n$, $\mathbb{R}^m$ and $\mathbb{R}^\ell$. We omit the exact statements here.

§7. Discussions.

In this section we discuss some interesting implications of our result obtained in the previous sections.

1. Discussion of the Hypothesis (H1)—(H3).

In order that the conclusion of Theorem 6.3 be true, we assumed several conditions on the data. Among them, (6.1), (6.3) are regularity conditions, and as we pointed out in the remark after (H1)—(H3), these conditions can be relaxed if the methodology of [30] is adopted. The condition (6.4) is in some sense standard as long as a backward equation is involved. In fact, even in the pure backward case (cf. [27,28]), this condition was also posed so that the process Z can be solved via the martingale representation theorem. From Theorem 6.4, we see that condition (6.4) is very close to a necessary condition for the solvability of (5.1) on any $[0,T]$ and any $x\in\mathbb{R}$. However, the seemingly artificial non-degeneracy condition of σ (6.2), which guarantees that the PDE (5.24) to have a classical solution, seems unremovable. A counterexample can be found in [1], in which $\sigma \equiv 0$ and the forward-backward equation with the type of (5.2) is not solvable for $T \geq 1$. Therefore, in order to make the forward-backward equation (5.1) solvable, it is necessary that the forward diffusion process is "random enough". We should say that the nondegeneracy of $\sigma(x, y)$ and the condition (6.4) represent the essential solvability feature of our forward-backward SDE (5.1). As a matter of fact, if, say, σ and $\widehat{\sigma}$ are both identically zero, then (5.1) is reduced to a two-point boundary value problem. We know that in general, it may have no solutions ([2]).

2. The Existence of Ordinary Adapted Solutions.

As we pointed out in §1, the only reason that we use the wider-sense (or relaxed) solution for the forward-backward SDEs is to guarantee that the optimal control exists, which is essential to our scheme. Therefore, it would be nice to know when the ordinary solution of the forward-backward equation exists. The result of [28] shows that under some restrictive conditions on data, the forward-backward equation (5.1) and (5.2) has a *unique* ordinary solution when T is small enough. Using the equivalent relations (Theorems 6.2 and 6.3), we see that this implies that the Problem (C) will have an ordinary optimal control which is even unique. However, when T is large, whether the relaxed control problem Problem (C) will have an ordinary optimal control is a quite challenging problem in general. But in some special cases, it is still workable. Let us take the forward-backward equation (5.1), with an assumption that $\widehat{\sigma}(x, y, z) \equiv z$, as an example.

Recall from (6.1) that g is bounded, then it is easily seen that there exists a compact set $U \subset \mathbb{R}$ such that for each $(t, x) \in [0, T] \times \mathbb{R}$.

$$0 \in \{z + \sigma(t, \theta(t, x))\theta_x(t, x) | z \in U\}.$$

From the proof of Theorem 6.3, this is sufficient for us to conclude that $v(t, x, \theta(t, x)) = 0$ for any $(t, x) \in [0, T] \times \mathbb{R}$.

Now let us take this compact set U to set up an ordinary control problem with the state equations (5.7) with $\widehat{\sigma}(x, y, z) \equiv z$, and cost functional (5.8). Note that now σ is linear in z, so one can show by a similar technique as we used in §3 and §4 that the ordinary optimal control exists (modulo a change of probability space). Therefore the SDE (5.1) will have an ordinary adapted solution (X, Y, Z), with Z taking value in a compact set! Such a result seems hard to be obtained by using the martingale representation theorem. It is evident that this scheme will also work for all forward-backward SDEs that are *linear* in Z, as long as a suitable compact set can be found *a priori* from the study of Problem (N).

3. Solvability of SDEs and Controllability.

Another interesting implication of our results, which does not seem possible to argue by using a contraction mapping theorem, is the non-solvability of the forward-backward SDEs. Theorem 6.4 provides a non-existence result which basically says that it is possible for some $T > 0$, (5.1) is solvable over $[0, T]$, but not solvable over any $[0, S]$ with $S < T$. In other words, even if the time duration is small, (5.1) can still be unsolvable if the coefficients are not well-matched. Therefore, it describes a deeper feature of the solvability of forward-backward SDEs that is so far undiscovered. One can check that all the known cases (cf. *e.g.*, [1,28]) in which an adapted solution exists are actually in the complement of this case (see condition (6.20)). This phenomenon is, however, quite natural if we look at it from other points of view; for instance, as a boundary value problem of linear ordinary differential equations (see [2], for example); or as a controllability problem. It is known (cf. [25]) that it is possible that the considered systems is *completely controllable* but not *small time controllable*. Therefore, there may be some kind of waiting time before hitting a target becoming possible.

4. The Nodal Sets.

To our best knowledge, in most of the literatures concerning nodal sets of the solutions to PDEs, people mainly focused on the estimates of the upper bound of the size of nodal sets in terms of Hausdorff measure (see [10,17,24] and also [5,6,14]). In this paper, we actually addressed the problem in terms of the *non-emptyness* and the *shape* of a nodal set. We have constructed such a nodal set for the viscosity solution of a certain class of HJB equations. We also believe that the study of the nodal sets for solutions to general nonlinear *degenerate* elliptic and parabolic PDEs would be very interesting; and some more investigation will be made along this line in our future publications.

References

[1] F. Antonelli, *Backward-Forward Stochastic Differential Equations, Ann. Appl. Prob.*, to appear.

[2] P. B. Bailey, L. F. Shampine and P. E. Waltman, *Nonlinear Two Point Boundary Value Problems*, Academic Press, New York, 1968.

[3] J. M. Bismut, *An introductory approach to duality in optimal stochastic control, SIAM Rev.*, 20 (1978), 62–78.

[4] V. S. Borkar, *Optimal Control of Diffusion Processes*, Longman Scientific & Technical, 1989.

[5] L. A. Caffarelli and A. Friedman, *Partial regularity of the zero-set of semilinear and superlinear elliptic equations, J. Diff. Eqn.*, 60 (1985), 420–433.

[6] P. Cannarsa and H. M. Soner, *On the singularities of the viscosity solutions to Hamilton-Jacobi-Bellman equations,* Indiana *Univ. Math. J.*, 36 (1987), 501–524.

[7] R. Courant and D. Hilbert, *Methods of Mathematical Physics, Vol.1*, Interscience, New York, 1953.

[8] M. G. Crandall, H. Ishii and P. L. Lions, *User's guide to viscosity solutions of second order partial differential equations, Bull. Amer. Math. Soc. (New Series)*, 27 (1992), 1–67.

[9] G. Dong, *Nonlinear Partial Differential Equations of Second Order*, Translation of Mathematical Monographs Vol.95, AMS, Providence, 1991.

[10] H. Donnelly and C. Fefferman, *Nodal sets of eigenfunctions on Riemannian manifolds, Invent. Math.*, 93 (1988), 161–183.

[11] D. Duffie, P.-Y. Geoffard, and C. Skiadas, *Efficient and Equilibrium Allocations with Stochastic Differential Utility,* Preprint.

[12] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, John Wiley & Sons, New York, 1986.

[13] L. C. Evans and S. Lenhart, *The parabolic Bellman equation,* Nonlinear *Anal. TMA*, 5 (1981), 765–773.

[14] W. H. Fleming, *The Cauchy problem for a nonlinear first order differential equation, J. Diff. Eqn.*, 5 (1969), 515–530.

[15] W. H. Fleming, *Generalized Solutions in Optimal Stochastic Control*, Differential games and control theory II, Lect. Notes in Pure and Appl. Math., **30**, Dekker, 1977.

[16] W. H. Fleming and M. Nisio, *On Stochastic Relaxed Control for Partially Observed Diffusions,* Nagoya *Math. J.*, 93 (1984), 71-108.

[17] R. Hardt and L Simon, *Nodal sets for solutions of elliptic equations, J. Diff. Geom.*, 30 (1989), 505–522.

[18] E. Hille and R. S. Philips, *Functional Analysis and Semigroups*, AMS, Providence, R.I., 1957.

[19] I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, Berlin, 1988.

[20] N. V. Krylov,*Controlled Diffusion Processes*, Springer-Verlag, 1980.

[21] N. V. Krylov, *Nonlinear Elliptic and Parabolic Equations of the Second Order*, Reidel, Dordrecht, Holland, 1987.

[22] H. J. Kushner, *Weak Convergence Methods and Singularly Perturbed Stochastic Control and Filtering Problems*, Berkhäuser, Berlin, 1990.

[23] O. A. Ladyzenskaja, V. A. Solonnikov and N. N. Ural'ceva, *Linear and Quasilinear Equations of Parabolic Type*, Amer. Math. Soc., Providence, R.I., 1968.

[24] F. H. Lin, *Nodal sets of solutions of elliptic and parabolic equations, Comm. Pure Appl. Math.*, 44 (1991), 287–308.

[25] J. L. Lions, *Exact controllability, stabilization and perturbations for distributed systems, SIAM Rev.*, 30 (1988), 1–68.

[26] J. Ma, *Discontinuous Reflection, and A Class of Singular Stochastic Control for Diffusions, Stochastics and Stochastics Reports*, in press.

[27] E. Pardoux and S. Peng, *Adapted Solution of a Backward Stochastic Differential Equation, Systems & Control Letters*, 14 (1990), 55–61.

[28] S. Peng, *Adapted Solution of Backward Stochastic Equations and Related Partial Differential Equations,* Preprint.

[29] P. Protter, *Stochastic Integration and Differential Equations, A New Approach*, Springer-Verlag, Berlin, 1990.

[30] M. V. Safonov, *On the classical solution of nonlinear elliptic equations of second order, Math. USSR Izv*, 33 (1989), 597–612.