Abstract. The present paper studies the stochastic maximum principle in singular optimal control, where the state is governed by a stochastic differential equation with non smooth coefficients, allowing both classical control and singular control. The proof of the main result is based on the approximation of the initial problem, by a sequence of control problems with smooth coefficients. We then apply Ekeland’s variational principle for this approximating sequence of control problems, in order to establish necessary conditions satisfied by a sequence of near optimal controls. Finally, we prove the convergence of the scheme, using Krylov’s inequality in the non degenerate case and the Bouleau-Hirsch flow property in the degenerate one. The adjoint process obtained is given by means of distributional derivatives of the coefficients.

Key words. stochastic differential equation, stochastic control, maximum principle, singular control, distributional derivative, adjoint process, variational principle.


1. Introduction. We consider stochastic control problems of nonlinear systems, where the control variable has two-components, the first being absolutely continuous and the second singular. More precisely, we study the stochastic maximum principle in optimal control for problem in which the state evolves according to the $d$—dimensional stochastic differential equation

$$\begin{cases}
    dx_t = b(t, x_t, u_t) \, dt + \sigma(t, x_t) \, dB_t + G_t \, d\xi_t, & \text{for } t \in [0, T], \\
    x_0 = \alpha,
\end{cases}$$

and the expected cost has the form

$$J(u, \xi) = E \left[ \int_0^T f(t, x_t, u_t) \, dt + \int_0^T k_t \, d\xi_t + g(x_T) \right].$$

Singular control problems have numerous applications. They appear in mathematical finance, e.g in the problem of optimal consumption investment, with transaction costs (see Davis, Norman [14], Shreve, Soner [25]). A huge literature have been produced on the subject, including Benës, Shepp, and Witsenhausen [6], Chow, Menaldi, and Robin [12], Karatzas, Shreve [19], Davis, Norman [14], Haussmann, Suo [17], [18]. See [17] for a complete list of references on the subject. The approaches used in these papers, are mainly based on dynamic programming. It was shown in particular that the value function is solution of a variational inequality, and the optimal state is a reflected diffusion at the free boundary. Note that in [17], the authors...
apply the compactification method to show existence of an optimal relaxed singular control.

The other major approach to study singular control problems is the investigation for necessary conditions satisfied by an optimal control. The first version of the stochastic maximum principle that covers singular control problems was obtained by Cadenillas and Haussmann [10], in which they consider linear dynamics, convex cost criterion and convex state constraints. The method used in [10] is based on the known principle of convex analysis, related to the minimization of convex, Gâteaux differentiable functionals defined on a convex closed set.

A first order weak maximum principle has been derived by Bahlali and Chala [1], in which convex perturbations are used for both absolutely continuous and singular components. A second order stochastic maximum principle for nonlinear SDEs with a controlled diffusion matrix was obtained by Bahlali and Mezerdi [4], extending the Peng’s maximum principle [23] to singular control problems. This result is based on two perturbations of the optimal control, the first is a spike variation, on the absolutely continuous component of the control, and the second one is convex on the singular component. A similar approach has been used by Bahlali et al. [2] to study the relaxed stochastic maximum principle in the case of uncontrolled diffusion coefficient.

On the other hand, the stochastic maximum principle for classical control problems(without the singular part) have been studied, with differentiability assumptions on the data weakened. The first result has been derived by Mezerdi [22], in the case of a SDE with a non smooth drift, by using Clarke generalized gradients and stable convergence of probability measures. In [3] [5], the authors extend the classical stochastic maximum principle to the case where the coefficients of the diffusion process are only Lipschitz continuous. The adjoint process obtained is given by means of generalized derivatives of the coefficients.

Our aim in this paper is to extend the stochastic maximum principle in singular optimal control to the case where the coefficients $b, \sigma, f$ and $g$ are Lipschitz continuous in the state variable. The main result is proved via an approximation scheme of the initial control problem by a sequence of control problems where the data are smooth functions. Ekeland’s variational principle is then applied to derive necessary conditions for near optimality satisfied by a sequence of near optimal controls. The convergence of the approximation scheme is obtained by using Krylov’s estimate in the non degenerate case and the Bouleau Hirsch flow property in the degenerate case.

2. Assumptions and preliminaries. Let $(\Omega, F, F_t, P)$ be a filtered probability space, satisfying the usual conditions, on which a $d$-dimensional Brownian motion $(B_t)$ is defined with the filtration $(F_t)$. Let $T$ be a strictly positive real number, $A_1$ is a non empty subset of $\mathbb{R}^n$ and $A_2 = ([0, \infty))^m$. $U_1$ is the class of measurable, adapted processes $u : [0, T] \times \Omega \rightarrow A_1$, and $U_2$ is the class of measurable, adapted processes $\xi : [0, T] \times \Omega \rightarrow A_2$.

**Definition 2.1.** An admissible control is a pair $(u, \xi)$ of measurable $A_1 \times A_2$-valued, $F_t$-adapted processes, such that $\xi$ is of bounded variation, non decreasing left-continuous with right limits and $\xi_0 = 0$.

We denote by $U = U_1 \times U_2$ the set of all admissible controls.

For $(u, \xi) \in U$, suppose that the state $x_t = x_t^{(u, \xi)} \in \mathbb{R}^d$ is described by the equation

$$\begin{align*}
   \begin{cases}
   dx_t = b(t, x_t, u_t) \, dt + \sigma(t, x_t) \, dB_t + G_t \, d\xi_t, & \text{for } t \in [0, T], \\
   x_0 = \alpha,
   \end{cases}
\end{align*}$$

(2.1)
Since $dt$ may be singular with respect to Lebesgue measure $dt$, we call $\xi$ the singular part of the control and the process $u$ its absolutely continuous part. Suppose we are given a cost functional $J(u, \xi)$ of the form

$$ J(u, \xi) = E \left[ \int_0^T f(t, x_t, u_t) dt + \int_0^T k_t d\xi_t + g(x_T) \right], \quad (2.2) $$

where $b: [0, T] \times \mathbb{R}^d \times A_1 \rightarrow \mathbb{R}^d$, $\sigma: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d \otimes \mathbb{R}^d$, $f: [0, T] \times \mathbb{R}^d \times A_1 \rightarrow \mathbb{R}$, $g: \mathbb{R}^d \rightarrow \mathbb{R}$, $G: [0, T] \rightarrow \mathbb{R}^d \otimes \mathbb{R}^m$, and $k: [0, T] \rightarrow ((0, \infty))^m$.

Assume that $b, \sigma, f$ and $g$ are Borel measurable, bounded functions and there exist $M > 0$, such that for all $(t, x, y, a)$ in $\mathbb{R}^+ \times \mathbb{R}^d \times \mathbb{R}^d \times A_1$

$$ |b(t, x, a) - b(t, y, a)| + |\sigma(t, x) - \sigma(t, y)| \leq M |x - y|, \quad (2.3) $$
$$ |f(t, x, a) - f(t, y, a)| + |g(x) - g(y)| \leq M |x - y|, \quad (2.4) $$

$b(t, x, a)$ and $f(t, x, a)$ are continuous in $a$ uniformly in $(t, x)$, and

$$ G, k \text{ are continuous and bounded.} \quad (2.6) $$

Find $(\hat{u}, \hat{\xi}) \in U$ such that

$$ J(\hat{u}, \hat{\xi}) = \min_{(u, \xi) \in U} J(u, \xi). $$

Any $(\hat{u}, \hat{\xi})$ satisfying the above property is called an optimal control of problem (2.1), (2.2). The corresponding state process $\hat{x}$ is called the optimal state process.

Under the above hypothesis, the SDE (2.1) has a unique strong solution $x_t$, such that for any $p > 0$,

$$ E \left[ \sup_{0 \leq t \leq T} |x_t|^p \right] < +\infty. $$

Moreover the cost functional is well defined from $U$ into $\mathbb{R}$.

Since $b, \sigma^j$ (the $j$th column of the matrix $\sigma$), $f$ and $g$ are Lipschitz continuous functions in the state variable, then they are differentiable almost everywhere in the sense of Lebesgue measure (Rademacher Theorem see [13]). Let us denote by $b_x, \sigma_x$, $f_x$ and $g_x$ any Borel measurable functions such that

$$ \partial_x b(t, x, a) = b_x(t, x, a) \quad dx-a.e., $$
$$ \partial_x f(t, x, a) = f_x(t, x, a) \quad dx-a.e., $$
$$ \partial_x \sigma(t, x) = \sigma_x(t, x) \quad dx-a.e., $$
$$ \partial_x g(x) = g_x(x) \quad dx-a.e. $$

It is clear that these almost everywhere derivatives are bounded by the Lipschitz constant $M$. Finally, assume that $b_x(t, x, a)$ and $f_x(t, x, a)$ are continuous in $a$ uniformly in $(t, x)$. 

3
Let us recall Krylov’s inequality and Ekeland’s variational principle, which will be used in the sequel.

**Theorem 2.2.** (Krylov [20]) Let \((\Omega, F, F_t, P)\) be a filtered probability space, \((B_t)_{t \geq 0}\) a d-dimensional Brownian motion, \(b: \Omega \times \mathbb{R}_+ \to \mathbb{R}^d, \sigma: \Omega \times \mathbb{R}_+ \to \mathbb{R}^d \otimes \mathbb{R}^d\) bounded adapted processes such that: \(\exists c > 0, \forall \zeta \in \mathbb{R}^d, \forall (t, x) \in [0, T] \times \mathbb{R}^d, \zeta^* \sigma \sigma^* \zeta \geq c |\zeta|^2\). Let

\[
x_t = x + \int_0^T b(t, \omega) \, dt + \int_0^T \sigma(t, \omega) \, dB_t,
\]

be an Itô process. Then for every Borel function \(f: \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}\) with support in \([0, T] \times B(0, \mathcal{M})\), the following inequality holds

\[
E \left[ \int_0^T |f(t, x_t)| \, dt \right] \leq K \left[ \int_{B(0, \mathcal{M})} \right. \left. \int_0^T |f(t, x)|^{d+1} \, dt \, dx \right]^{\frac{1}{d+1}},
\]

where \(K\) is a constant and \(B(0, \mathcal{M})\) is the ball of center \(0\) and radius \(\mathcal{M}\).

**Lemma 2.3.** (Ekeland variational principle [15]) Let \((S, \rho)\) be metric space and \(\rho: S \to \mathbb{R} \cup \{+\infty\}\) be lower-semicontinuous and bounded from below. For \(\varepsilon \geq 0\), suppose \(u^\varepsilon \in S\) satisfies \(\rho(u^\varepsilon) \leq \inf_{u \in S} \rho(u) + \varepsilon\). Then for any \(\lambda > 0\), there exists \(u^\lambda \in S\) such that

\[
\rho(u^\lambda) \leq \rho(u^\varepsilon),
\]

\[
d(u^\lambda, u^\varepsilon) \leq \lambda,
\]

\[
\rho(u^\lambda) \leq \rho(u) + \frac{\varepsilon}{\lambda} d(u, u^\lambda), \quad \text{for all } u \in S.
\]

To apply Ekeland’s variational principle to the control problem, we have to endow the set of controls with an appropriate metric. For any \((u, \xi), (v, \eta) \in U\), we set

\[
d_1(u, v) = \mathbb{P} \otimes dt \{ (\omega, t) \in \Omega \times [0, T], v(\omega, t) \neq u(\omega, t) \},
\]

\[
d_2(\xi, \eta) = \left( E \left[ \sup_{0 \leq t \leq T} |\xi_t - \eta_t|^2 \right] \right)^{\frac{1}{2}},
\]

\[
d((u, \xi), (v, \eta)) = d_1(u, v) + d_2(\xi, \eta).
\]

where \(\mathbb{P} \otimes dt\) is the product measure of \(\mathbb{P}\) with the Lebesgue measure \(dt\).

**Lemma 2.4.** (1) \((U, d)\) is a complete metric space.

(2) The cost functional \(J\) is continuous from \(U\) into \(\mathbb{R}\).

Proof. (1) It is clear that \((U_2, d_2)\) is a complete metric space. Moreover, it was shown in [19] that \((U_1, d_1)\) is a complete metric space. Hence \((U, d)\) is a complete metric space.

Item (2) is proved as in [22] [26]. □

3. The non degenerate case. In this section, we assume the following condition

\[
\exists c > 0, \forall \zeta \in \mathbb{R}^d, \forall (t, x) \in [0, T] \times \mathbb{R}^d, \zeta^* \sigma(t, x) \sigma^* (t, x) \zeta \geq c |\zeta|^2,
\]

(3.1)
3.1. The main result. The main result of this section is stated in the following Theorem.

**Theorem 3.1.** (Stochastic maximum principle) Let \((\hat{u}, \hat{\xi})\) be an optimal control for the controlled system (2.1), (2.2) and let \(\hat{x}\) be the corresponding optimal trajectory. Then there exists a measurable \(F_t\)-adapted process \(p_t\) satisfying

\[
p_t := -E \left[ \int_t^T \Phi^* (s,t) \cdot f_x (s,\hat{x}_s, \hat{u}_s) \, ds + \Phi^* (T,t) \cdot g_x (\hat{x}_T) / F_t \right],
\]

such that for all \(a \in A_1\) and \(\eta \in U_2\)

\[
0 \leq H (t,\hat{x}_t, a, p_t) - H (t,\hat{x}_t, \hat{u}_t, p_t) \quad dt\text{-a.e, } P\text{-a.s.},
\]

and

\[
0 \leq E \int_0^T (k_t + G^*_t p_t) \, d (\eta - \hat{\xi}),
\]

where the Hamiltonian \(H\) associated to the control problem is

\[
H (t,x,u,p) = p \cdot b(t,x,u) - f(t,x,u),
\]

and \(\Phi(s,t), (s \geq t)\) is the fundamental solution of the linear equation

\[
\begin{cases}
  d\Phi(s,t) = b_x (s,\hat{x}_s, \hat{u}_s) \cdot \Phi (s,t) \, ds + \sum_{1 \leq j \leq d} \sigma^j_x (s,\hat{x}_s) \cdot \Phi (s,t) \, dB^j_s, \\
  \Phi (t,t) = I_d.
\end{cases}
\]

Here * denotes the transpose.

3.2. Proof of the main result. Let \(\varphi\) be a non negative smooth function defined on \(\mathbb{R}^d\), with support in the unit ball such that \(\int_{\mathbb{R}^d} \varphi (y) \, dy = 1\). Define the following smooth functions by convolution

\[
\begin{align*}
  b^n (t, x, a) &= n^d \int_{\mathbb{R}^d} b (t, x - y, a) \varphi (ny) \, dy, \\
  f^n (t, x, a) &= n^d \int_{\mathbb{R}^d} f (t, x - y, a) \varphi (ny) \, dy, \\
  \sigma^{j,n} (t, x) &= n^d \int_{\mathbb{R}^d} \sigma^j (t,x - y) \varphi (ny) \, dy, \\
  g^n (x) &= n^d \int_{\mathbb{R}^d} g (x - y) \varphi (ny) \, dy.
\end{align*}
\]

**Lemma 3.2.** (1) The functions \(b^n (t, x, a), \sigma^{j,n} (t, x), f^n (t, x, a),\) and \(g^n (x)\) are Borel measurable bounded functions and Lipschitz continuous with constant \(K\) in \(x\).
(2) There exists a constant $C$ positive independent of $t$, $x$ and $n$ such that for every $t$ in $[0, T]$

$$|b^n(t, x, a) - b(t, x, a)| + |\sigma^{j,n}(t, x) - \sigma^j(t, x)| \leq \frac{C}{n},$$

$$|f^n(t, x, a) - f(t, x, a)| + |g^n(x) - g(x)| \leq \frac{C}{n}.$$

(3) The functions $b^n(t, x, a)$, $f^n(t, x, a)$, $\sigma^{j,n}(t, x)$ and $g^n(x)$ are $C^\infty$-functions in $x$, and for all $t$ in $[0, T]$, we have

$$\lim_{n \to \infty} b^n_x(t, x, a) = b_x(t, x, a) \quad dx\text{-a.e.},$$

$$\lim_{n \to \infty} f^n_x(t, x, a) = f_x(t, x, a) \quad dx\text{-a.e.},$$

$$\lim_{n \to \infty} \sigma^{j,n}_x(t, x) = \sigma^j_x(t, x) \quad dx\text{-a.e.},$$

$$\lim_{n \to \infty} g^n(x) = g(x) \quad dx\text{-a.e.}$$

(4) For every $p \geq 1$ and $R > 0$

$$\lim_{n \to \infty} \int_{B(0, R) \times [0, T]} \sup_{a \in A} |b^n_x(t, x, a) - b_x(t, x, a)|^p dxdt = 0,$$

$$\lim_{n \to \infty} \int_{B(0, R) \times [0, T]} \sup_{a \in A} |f^n_x(t, x, a) - f_x(t, x, a)|^p dxdt = 0.$$

Proof. Statements (1), (2) and (3) are classical facts (see [16] for the proof).

(4) is proved as in [20].

For $n \in \mathbb{N}^*$, let us consider the sequence of perturbed control problems obtained by replacing $b$, $\sigma$, $f$ and $g$ by $b^n$, $\sigma^n$, $f^n$ and $g^n$. Let us denote $y$ the solution of the controlled stochastic differential equation.

$$\begin{cases}
\frac{dy_t}{dt} = b^n(t, y_t, u_t) dt + \sigma^n(t, y_t) dB_t + G_t d\xi_t, \\
y_0 = \alpha,
\end{cases} \quad (3.7)$$

The corresponding cost is given by

$$J^n(u, \xi) = E \left[ \int_0^T f^n(t, y_t, u_t) dt + \int_0^T k_t d\xi_t + g^n(y_T) \right], \quad (3.8)$$

Lemma 3.3. Let $(u, \xi) \in U$, $x_t$ and $y_t$ the solutions of (2.1) and (3.9) respectively corresponding to the control $(u, \xi)$, then we have

(1) $E \left[ \sup_{0 \leq t \leq T} |x_t - y_t|^2 \right] \leq M_1 (\epsilon_n)^2$, where $\epsilon_n = \frac{C}{n}$.

(2) $|J^n(u, \xi) - J(u, \xi)| \leq M_2 \epsilon_n$.

Proof. Since $x_t - y_t$ and $J^n(u, \xi) - J(u, \xi)$ does not depend on the singular part, then This lemma follows from standard arguments from stochastic calculus and lemma 3.2. □
Let us suppose that \((\hat{u}, \hat{\xi}) \in U\) is an optimal control for the initial control problem (2.1) and (2.2). Note that \((\hat{u}, \hat{\xi})\) is not necessarily optimal for the perturbed control problem (3.9) and (3.10). However, by Lemma 3.6 we obtain the existence of \((\delta_n) \equiv (2M_2, e_n)\) a sequence of positive real numbers converging to 0, such that

\[
J^n(\hat{u}, \hat{\xi}) \leq \inf_{(v, \eta) \in U} J^n(v, \eta) + \delta_n.
\]

The control \((\hat{u}, \hat{\xi})\) will be \(\delta_n\)-optimal for the perturbed control problem. According to Lemma 3.5, it is easy to see that \(J^n(\cdots)\) is continuous on \(U = U_1 \times U_2\) endowed with the metric \(d = d_1 + d_2\) defined by (3.8). By Ekeland’s variational principle (lemma 3.4) applied to \((\hat{u}, \hat{\xi})\) with \(\lambda_n = \frac{\delta_n^2}{2}\), there exist an admissible control \((u^n, \xi^n)\) such that

\[
d\left((\hat{u}, \hat{\xi}), (u^n, \xi^n)\right) \leq \frac{\delta_n^2}{2},
\]

and

\[
J^n_b(u^n, \xi^n) \leq J^n_b(v, \eta), \quad \text{for all } (v, \eta) \in U,
\]

where

\[
J^n_b(v, \eta) = J^n(v, \eta) + \delta^n_b d((v, \eta), (u^n, \xi^n)).
\]

This means that \((u^n, \xi^n)\) is an optimal control for the perturbed system (3.9) with a new cost function \(J^n_b\). The controlled process \(x^n\) is defined as the unique solution to the stochastic differential equation,

\[
\begin{cases}
    dx^n_t = b^n(t, x^n_t, u^n_t) \, dt + \sigma^n(t, x^n_t) \, dB_t + G_t \, d\xi^n_t, \\
y_0 = \alpha.
\end{cases}
\]  

(3.9)

We consider \(\Phi^n(s, t) \ (s \geq t)\), the fundamental solution of the linear stochastic differential equation

\[
\begin{cases}
    d\Phi^n(s, t) = b^n_s(s, x^n_s, u^n_s) \cdot \Phi^n(s, t) \, ds + \sum_{1 \leq j \leq d} \sigma^n_{x^j}(s, x^n_s) \cdot \Phi^n(s, t) \, dB^j_s, \\
\Phi^n(t, t) = I_d.
\end{cases}
\]  

(3.10)

Note that \(b^n_s, \sigma^n_{x^j} \ (j = 1, \ldots, d)\) are respectively the matrices of first order partial derivatives of \(b^n, \sigma^n_{x^j} \ (j = 1, \ldots, d)\) with respect to \(x\).

**Proposition 3.4.** For each integer \(n\), there exists an admissible control \((u^n, \xi^n)\) and a \((F_1)\)-adapted process \(p^n_t\) given by

\[
p^n_t = -E \left[ \int_t^T \Phi^{n, \ast}(s, t) \cdot f^n_s(s, x^n_s, u^n_s) \, ds + \Phi^{n, \ast}(T, t) \cdot g^n_T(x^n_T) / F_t \right],
\]

(3.11)

and a Lebesgue null set \(N\) such that for \(t \in N^c\)

\[
E[H^n(t, x^n_t, v, p^n_t) - H^n(t, x^n_t, u^n_t, p^n_t)] \geq -\frac{\delta_n^2}{2} M_1,
\]

(3.12)
and

$$E \int_0^T (k_t + G_t^* p^n_t) \, d (\eta - \xi^n)_t \geq -\delta_n^i \cdot M_2.$$  \hfill (3.13)

for all \( v \in A_1 \), and \( \eta \in U_2 \). The Hamiltonian \( H^n \) is defined by

$$H^n (t, x, u, p) = p \cdot b^n (t, x, u) - f^n (t, x, u).$$  \hfill (3.14)

Here \( * \) denotes the transpose.

Proof. According to the optimality of \((u^n, \xi^n)\) for the perturbed system with cost function \( J^n \), we can use the spike variation method to derive a maximum principle for \((u^n, \xi^n)\). Let \( t_0 \in [0, T], v \in A_1 \) and \( \eta \in U_2 \), for any \( \varepsilon > 0 \), define the two perturbations \((u_t^{n, \varepsilon}, \xi_t^{n, \varepsilon})\) and \((u_t^n, \xi_t^n)\) by

\[
(u_t^{n, \varepsilon}, \xi_t^{n, \varepsilon}) = \begin{cases} (v, \xi_t^n) & t \in [t_0, t_0 + \varepsilon], \\ (u_t^n, \xi_t^n) & t \in [0, T] \setminus [t_0, t_0 + \varepsilon]. \end{cases}
\]

and

\[
(u_t^n, \xi_t^n) = (u_t^n, \xi_t^n + \varepsilon (\eta_t - \xi_t^n))
\]

Since \((u_t^n, \xi_t^n)\) is optimal for the cost \( J^n \), then

\[
0 \leq J^n (u_t^{n, \varepsilon}, \xi_t^{n, \varepsilon}) - J^n (u_t^n, \xi_t^n)
\]

and

\[
0 \leq J^n (u_t^n, \xi_t^{n, \varepsilon}) - J^n (u_t^n, \xi_t^n)
\]

this imply that

\[
0 \leq J^n (u_t^{n, \varepsilon}, \xi_t^{n, \varepsilon}) - J^n (u_t^n, \xi_t^n) + \delta_n^i \cdot d_1 (u_t^n, u_t^{n, \varepsilon}),
\]

and

\[
0 \leq J^n (u_t^n, \xi_t^{n, \varepsilon}) - J^n (u_t^n, \xi_t^n) + \delta_n^i \cdot d_2 (\xi_t^n, \xi_t^{n, \varepsilon}),
\]

using the definitions of \( d_1 \) and \( d_2 \) it holds that

\[
0 \leq J^n (u_t^{n, \varepsilon}, \xi_t^{n, \varepsilon}) - J^n (u_t^n, \xi_t^n) + \delta_n^i \cdot M_1 \varepsilon,
\]

and

\[
0 \leq J^n (u_t^n, \xi_t^{n, \varepsilon}) - J^n (u_t^n, \xi_t^n) + \delta_n^i \cdot M_2 \varepsilon,
\]

where \( M_i \) \((i = 1, 2)\) is a positive constant. From inequalities (3.17) and (3.18) respectively we use the same method as in subsection 3.3 in [2] to obtain respectively (3.14) and (3.15).

We use a transformation that makes it possible to apply Krylov’s estimate for diffusion processes. Define dynamics 

\[
\mathbf{b} : [0, T] \times \mathbb{R}^d \times A_1 \to \mathbb{R}^d, \quad \mathbf{b} : [0, T] \times \mathbb{R}^d \times A_1 \to 
\]
where we get from line to line. approximating sequence adjoint process (3.13) and the associated Hamiltonian (3.16), corresponding to the process (3.1) and the associated Hamiltonian (3.4), corresponding to the optimal state

$\text{Lipschitz continuous with constant } d_{i} : [0, T] \times \mathbb{R}^{d} \to \mathbb{R}^{d}$.

By squaring, taking expectations and using Burkholder-Davis-Gundy inequality, Proof

Note that, Lemma 3.5. The following estimates hold

Let $z$ the unique solution of

$$
\begin{cases}
    dz_t = \tilde{b}(t, z_t, u_t) dt + \tilde{\sigma}(t, z_t) dB_t, \\
    z_0 = \alpha.
\end{cases}
$$

(3.17)

This implies that $x = z + \int_{0}^{t} G_s d\xi_s$ solves the SDE (2.1) with data $(b, \sigma)$.

Similarly, let $z^n$ the unique solution of

$$
\begin{cases}
    dz^n_t = \tilde{b}^n(t, z^n_t, u_t) dt + \tilde{\sigma}^n(t, z^n_t) dB_t, \\
    z^n_0 = \alpha.
\end{cases}
$$

(3.18)

Then $x^n = z^n + \int_{0}^{t} G_s d\xi_s$ solves the SDE (3.9) with data $(b^n, \sigma^n)$.

Note that, $\tilde{b}, \tilde{b}^n, \tilde{\sigma}^j$, and $\tilde{\sigma}^{j,n}$ $(j = 1, ..., d)$ are measurable bounded functions and Lipschitz continuous with constant $M$ in $x$. We conclude that the generalized derivatives (in the sense of distributions) $\tilde{b}_x, \tilde{b}^n_x, \tilde{\sigma}^j_x$, and $\tilde{\sigma}^{j,n}_x$ $(j = 1, ..., d)$ are well defined.

**Lemma 3.5.** The following estimates hold

$$
\begin{align*}
    \lim_{n \to +\infty} E \left[ \sup_{0 \leq t \leq T} |x^n_t - \hat{x}_t|^2 \right] &= 0, \\
    \lim_{n \to +\infty} E \left[ \sup_{0 \leq s \leq T} |\Phi^n(s, t) - \Phi(s, t)|^2 \right] &= 0, \\
    \lim_{n \to +\infty} E \left[ \sup_{0 \leq t \leq T} |p^n_t - p_t|^2 \right] &= 0, \\
    \lim_{n \to +\infty} E \left[ |H^n(t, x^n_t, u^n_t, p^n_t) - H(t, \hat{x}_t, \hat{u}_t, p_t)| \right] &= 0,
\end{align*}
$$

(3.19) \hspace{1cm} (3.20) \hspace{1cm} (3.21) \hspace{1cm} (3.22)

where $\Phi_t$, $p_t$ and $H$ are determined respectively by the solution of (3.5), the adjoint process (3.1) and the associated Hamiltonian (3.4), corresponding to the optimal state process $\hat{x}_t$. $\Phi^n_t$, $p^n_t$ and $H^n$ are determined respectively by the solution (3.12), the adjoint process (3.13) and the associated Hamiltonian (3.16), corresponding to the approximating sequence $x^n_t$, given by (3.11).

**Proof.** In what follows, $C$ represents a generic constant, which can be different from line to line.

By squaring, taking expectations and using Burkholder-Davis-Gundy inequality, we get

$$
E \left[ \sup_{0 \leq t \leq T} |x^n_t - \hat{x}_t|^2 \right] \leq C \left( A^n_1 + A^n_2 + A^n_3 + M. \left( d_2 (\xi^n, \xi) \right)^2 \right),
$$

where $A^n_1, A^n_2, A^n_3$, and $M$ are constants dependent on $n$.
where $M$ is a positive constant, and

\[ A_1^n = E \left[ \int_0^t \left| b^n (s, x^n_s, u^n_s) - b^n (s, x^n_s, \hat{u}_s) \right|^2 \chi_{\{u^n \neq \bar{u}\}} (s) \, ds \right], \]

\[ A_2^n = E \left[ \int_0^t \left| b^n (s, x^n_s, \hat{u}_s) - b^n (s, \hat{x}_s, \hat{u}_s) \right|^2 + |\sigma^n (s, x^n_s) - \sigma^n (s, \hat{x}_s)|^2 \, ds \right], \]

\[ A_3^n = E \left[ \int_0^t \left| b^n (s, \hat{x}_s, \hat{u}_s) - b (s, \hat{x}_s, \hat{u}_s) \right|^2 + |\sigma^n (s, \hat{x}_s) - \sigma (s, \hat{x}_s)|^2 \, ds \right]. \]

By using the boundness of the coefficient $b^n$ and the fact that $d_1 (u^n, \hat{u}) \to 0$ as $n \to +\infty$, we have $\lim_{n \to +\infty} A_1^n = 0$. Since $b^n$ and $\sigma^n$ are Lipschitz in the state variable, then

\[ A_2^n \leq CE \left[ \int_0^t \sup_{0 \leq r \leq s} |x^n_r - \hat{x}_r|^2 \, ds \right]. \]

Finally, we conclude from the Lemma 3.2 that $\lim_{n \to +\infty} A_3^n = 0$. Then by Gronwall Lemma, we obtain (3.21).

Again, using standard arguments based on Burkholder-Davis-Gundy, Schwartz inequalities and Gronwall Lemma, we easily check that

\[ E \left[ \sup_{t \leq s \leq T} |\Phi^n (s, t) - \Phi (s, t)|^2 \right] \leq \]

\[ CE \left[ \sup_{t \leq s \leq T} |\Phi^n (s, t)|^4 \right] \frac{1}{2} \left\{ E \left[ \int_0^T |b^n_x (t, x^n_t, u^n_t) - b_x (t, \hat{x}_t, \hat{u}_t)|^4 \, dt \right] \right\} \]

\[ + \sum_{1 \leq j \leq d} E \left[ \int_0^T |\sigma^{j,n}_x (t, x^n_t) - \sigma^j_x (t, \hat{x}_t)|^4 \, dt \right] \frac{1}{2} \}, \]

Since the coefficients in the linear stochastic differential equation (3.12) are bounded, it is easy to see that $E \left[ \sup_{s \leq t \leq T} |\Phi^n (s, t)|^4 \right] < +\infty$. To obtain the desired result it is sufficient to prove that

\[ \lim_{n \to +\infty} E \left[ \int_0^T \left| b^n_x (t, x^n_t, u^n_t) - b_x (t, \hat{x}_t, \hat{u}_t) \right|^4 \, dt \right] = 0, \]

\[ \lim_{n \to +\infty} E \left[ \int_0^T \left| \sigma^{j,n}_x (t, x^n_t) - \sigma^j_x (t, \hat{x}_t) \right|^4 \, dt \right] = 0, \text{ for } j = 1, \ldots, d, \]

we have, $E \left[ \int_0^T \left| b^n_x (t, x^n_t, u^n_t) - b_x (t, \hat{x}_t, \hat{u}_t) \right|^4 \, dt \right] \leq C \left( I_1^n + I_2^n \right)$, where

\[ I_1^n = E \left[ \int_0^T \left| b^n_x (t, x^n_t, u^n_t) - b^n_x (t, x^n_t, \hat{u}_t) \right|^4 \chi_{\{u^n \neq \bar{u}\}} (t) \, dt \right], \]

\[ I_2^n = E \left[ \int_0^T \left| b^n_x (t, x^n_t, \hat{u}_t) - b^n_x (t, \hat{x}_t, \hat{u}_t) \right|^4 \, dt \right]. \]
First, in view of the boundness of the derivative \( b^n_t \) by the Lipschitz constant and the fact that \( d_1 (w^n, \bar{\bar{\xi}}) \to 0 \) as \( n \to +\infty \), we obtain \( \lim_{n \to +\infty} I^n_1 = 0 \). Next, let \( k \geq 1 \) be a fixed integer, we then get

\[
\lim_{n \to +\infty} I^n_2 \leq \liminf_n \{ J^n_1 + J^n_2 + J^n_3 \},
\]

where

\[
J^n_1 = E \left[ \int_0^T \left| b^n_x (t, x^n_t, \bar{\bar{\xi}}) - b^n_x (t, x^n_t, \bar{\bar{\xi}}) \right|^4 dt \right],
\]

\[
J^n_2 = E \left[ \int_0^T \left| b^n_x (t, x^n_t, \bar{\bar{\xi}}) - b^n_x (t, \hat{x}_t, \bar{\bar{\xi}}) \right|^4 dt \right],
\]

\[
J^n_3 = E \left[ \int_0^T \left| b^n_x (t, \hat{x}_t, \bar{\bar{\xi}}) - b_x (t, \hat{x}_t, \bar{\bar{\xi}}) \right|^4 dt \right].
\]

Now, let \( \bar{\tilde{\xi}} \) (resp. \( z^n \)) denotes the unique solution of the SDE (3.19) (resp. (3.20)) corresponding to \( (\hat{\bar{\bar{\xi}}}, \bar{\tilde{\xi}}) \) (resp. \( (u^n, \xi^n) \)), then it holds that

\[
J^n_1 = E \left[ \int_0^T \left| b^n_x (t, z^n_t, \hat{\bar{\bar{\xi}}} t) - b^n_x (t, z^n_t, \hat{\bar{\bar{\xi}}} t) \right|^4 dt \right],
\]

and

\[
J^n_3 = E \left[ \int_0^T \left| b^n_x (t, \hat{\bar{\bar{\xi}}} t, \hat{\bar{\bar{\xi}}} t) - b_x (t, \hat{\bar{\bar{\xi}}} t, \hat{\bar{\bar{\xi}}} t) \right|^4 dt \right].
\]

Arguing as in [20], page 87, let \( w (t, x) \) be a continuous function such that \( w (t, x) = 0 \) if \( t^2 + x^2 \geq 1 \), and \( w (0, 0) = 1 \). Then for \( M > 0 \), we have

\[
\limsup_n J^n_1 \leq CE \left[ \int_0^T \left( 1 - w \left( \frac{t}{M}, \frac{\hat{\bar{\bar{\xi}}} t}{M} \right) \right) dt \right] + C \liminf_n E \left[ \int_0^T w \left( \frac{t}{M}, \frac{\hat{\bar{\bar{\xi}}} t}{M} \right) \left| b^n_x (t, z^n_t, \hat{\bar{\bar{\xi}}} t) - b^n_x (t, z^n_t, \hat{\bar{\bar{\xi}}} t) \right|^4 dt \right].
\]

Therefore without loss of generality, we may suppose that for all \( n \in \mathbb{N}^* \), the functions \( \bar{b}_x, \bar{\sigma}_x, \bar{\sigma}_x^0 \), and \( \bar{\sigma}_x^0 \) have compact support in \( [0, T] \times B (0, M) \). Since the diffusion matrix \( \bar{\sigma}_x^0 \) satisfies the non degeneracy condition with the same constant as \( \sigma \), then by applying Krylov’s inequality, we obtain

\[
\limsup_n J^n_1 \leq CE \left[ \int_0^T \left( 1 - w \left( \frac{t}{M}, \frac{\hat{\bar{\bar{\xi}}} t}{M} \right) \right) dt \right] + C \limsup_n \| \sup_{a \in A_1} \left| b^n_x (t, x, a) - \bar{b}_x (t, x, a) \right|^4 \|_{d_{+1, M}}.
\]

Since \( b^n_x \) converges to \( b_x \) \( dx \)-a.e., it is simple to see that \( \bar{b}_x^0 \) converges to \( \bar{b}_x \) \( dx \)-a.e. and

\[
\lim_{n \to +\infty} \sup_{a \in A_1} \left| b^n_x (t, x, a) - \bar{b}_x^0 (t, x, a) \right|^4 \|_{d_{+1, M}} = 0.
\]
Next, let $M$ goes to $+\infty$, then from the properties of the function $w(t, x)$ we have $\lim_{n \to +\infty} J^n_t = 0$. Estimating $J^n_3$ similarly, it holds that $\lim_{n \to +\infty} J^n_3 = 0$. We use the continuity of $b_k^1$ in $x$. From (3.21), and by using the Dominated convergence theorem we deduce that $\lim_{n \to +\infty} J^n_2 = 0$. Hence $\lim_{n \to +\infty} I^n_1 = 0$. Using the same technique, we prove that
\[
\lim_{n \to +\infty} E \left[ \int_0^T \left| \sigma_j^n(t, x^n_t) - \sigma_j^x(t, x_t) \right|^4 dt \right] = 0, \text{ for } j = 1, \ldots, d.
\]

Now, let us prove that $\lim_{n \to +\infty} E \left[ \sup_{0 \leq t \leq T} |p^n_t - p_1| \right] = 0$. Clearly,
\[
E \left[ |p^n_t - p_1| \right] \leq C (\alpha^n_1 + \alpha^n_2), \quad (3.23)
\]
where
\[
\alpha^n_1 = E \left[ \int_t^T \left| \Phi^{n,*}(s, t) \cdot f^n_x(s, x^n_s, u^n_s) - \Phi^x(s, t) \cdot f_x(s, x_s, u_s) \right|^2 ds \right],
\]
and
\[
\alpha^n_2 = E \left[ \Phi^{n,*}(T, t) \cdot g^n_x(x^n_T) - \Phi^*(T, t) \cdot g_x(x_T) \right]^2.
\]

Since $f_x$ is bounded by the Lipschitz constant $M$, and applying the Schwartz inequality, we get
\[
\alpha^n_2 \leq C E \left[ \sup_{t \leq s \leq T} |\Phi^{n,*}(s, t)|^4 \right] \frac{1}{2} E \left[ \int_0^T |f^n_x(s, x^n_s, u^n_s) - f_x(s, x_s, u_s)|^4 ds \right]^{\frac{1}{2}}
\]
\[
+ C M E \left[ \sup_{t \leq s \leq T} |\Phi^{n,*}(s, t) - \Phi^*(s, t)|^2 \right].
\]

Hence, by the continuity and the boundedness of derivatives $f^n_x$, $f_x$, relations (3.21), (3.22) and the fact that $d_1(u^n, \hat{u}) \to 0$ as $n \to \infty$, together with the Krylov’s inequality and the Dominated convergence theorem, for the term involving $f^n_x(s, x^n_s, u^n_s) - f_x(s, x_s, u_s)$, we get by sending $n$ to infinity $\lim_{n \to +\infty} \alpha^n_2 = 0$.

On the other hand, since $g_x$ is bounded by the Lipschitz constant, and applying the Schwartz inequality we get
\[
\alpha^n_2 \leq C \left\{ E \left[ |\Phi^{n,*}(T, t)|^4 \right] \right\}^{\frac{1}{2}} \left\{ E \left[ |g^n_x(x^n_T) - g_x(x_T)|^4 \right] \right\}^{\frac{1}{2}}
\]
\[
+ C M E \left[ |\Phi^{n,*}(T, t) - \Phi^*(T, t)|^2 \right],
\]

Since, $g^n_x$ and $g_x$ are bounded by the Lipschitz constant and $g^n_x$ converges to $g_x$, we conclude by (3.21) and the dominated convergence theorem that
\[
\lim_{n \to +\infty} E \left[ |g^n_x(x^n_T) - g_x(x_T)|^4 \right] = 0.
\]

From (3.25), then by using Burkholder-Davis-Gundy inequality, we obtain (3.23).
The Schwartz inequality gives

$$
E \left[ |H^n(t, x^n_t, u^n_t, p^n_t) - H(t, \hat{x}_t, \hat{u}_t, p_t)| \right] \leq \left\{ E |p^n_t - p_t|^2 \right\}^{\frac{1}{2}} \left\{ E |b^n(t, x^n_t, u^n_t)|^2 \right\}^{\frac{1}{2}} \nonumber
$$

$$
+ \left\{ E |b^n(t, x^n_t, u^n_t) - b(t, \hat{x}_t, \hat{u}_t)|^2 \right\}^{\frac{1}{2}} \left\{ E |p_t|^2 \right\}^{\frac{1}{2}} - E |f^n(t, x^n_t, u^n_t) - f(t, \hat{x}_t, \hat{u}_t)|. \nonumber
$$

Lemma 3.2 and (3.23) imply that the first expression in the right hand side converges to 0 as \( n \to +\infty \).

Next,

$$
E |b^n(t, x^n_t, u^n_t) - b(t, \hat{x}_t, \hat{u}_t)|^2 \leq C (\beta^n_1 + \beta^n_2 + \beta^n_3), \nonumber
$$

where

$$
\beta^n_1 = E \left[ |b^n(t, x^n_t, u^n_t) - b^n(t, x^n_t, \hat{u}_t)|^2 \chi_{\{u^n \neq \hat{u}\}}(t) \right], \nonumber
$$

$$
\beta^n_2 = E \left[ |b^n(t, x^n_t, \hat{u}_t) - b^n(t, \hat{x}_t, \hat{u}_t)|^2 \right], \nonumber
$$

$$
\beta^n_3 = E \left[ |b^n(t, \hat{x}_t, \hat{u}_t) - b(t, \hat{x}_t, \hat{u}_t)|^2 \right]. \nonumber
$$

The boundness of \( b^n \) and the fact that \( d_1(u^n, \hat{u}) \to 0 \), guarantee the convergence of \( \beta^n_1 \) to 0 as \( n \to +\infty \). By virtue of (3.21), and the dominated convergence theorem we get, \( \lim \beta^n_2 = 0 \). In view of the Lemma 3.2, we have \( \lim \beta^n_3 = 0 \).

The term \( E |f^n(t, x^n_t, u^n_t) - f(t, \hat{x}_t, \hat{u}_t)| \) can be treated by the same technique. \( \Box \)

Proof of Theorem 3.1. Let \( n \) goes to \( +\infty \), then from Proposition 3.7 and Lemma 3.8, we get

$$
E \left[ H(t, \hat{x}_t, v, p_t) - H(t, \hat{x}_t, \hat{u}_t, p_t) \right] \geq 0, \quad dt - a.e., \quad P - a.s., \nonumber
$$

$$
E \int_0^T (k_t + G_t^2 p_t) d\left( \eta - \hat{\eta} \right)_t \geq 0, \nonumber
$$

for every \( A_1 \)-valued \( F_t \)-measurable random variable \( v \), and \( \eta \in U_2 \).

Let \( a \in A_1 \), then for every \( A_t \in F_t \)

$$
E \left[ \left( H(t, \hat{x}_t, a, p_t) - H(t, \hat{x}_t, \hat{u}_t, p_t) \right) \chi_{A_t} \right] \geq 0, \quad dt - a.e., \quad P - a.s., \nonumber
$$

which implies that

$$
E \left[ \left( H(t, \hat{x}_t, a, p_t) - H(t, \hat{x}_t, \hat{u}_t, p_t) \right) \chi_{A_t} / F_t \right] \geq 0 \nonumber
$$

Since \( H(t, \hat{x}_t, a, p_t) - H(t, \hat{x}_t, \hat{u}_t, p_t) \) is \( F_t \)-measurable, then the first variational inequality without expectations follows immediately. \( \Box \)

4. The Degenerate case. In this section we drop the uniform ellipticity condition on the diffusion matrix. It is clear that the method used in the last section will no longer be valid. To overcome this difficulty, the idea is to use a result of Bouleau and Hirsch [9], on the differentiability in the sense of distributions, of the solution of a SDE with Lipschitz coefficients, with respect to the initial data. This derivative is defined as the solution of a linear stochastic differential equation defined on an extension of the initial probability space.
Let $h$ be a continuous positive function on $\mathbb{R}^d$ such that $\int h(x)\,dx = 1$ and $\int |x|^2 h(x)\,dx < \infty$. We set
\[
D = \left\{ f \in L^2(h\,dx), \text{ such that } \frac{\partial f}{\partial x_j} \in L^2(h\,dx) \right\},
\]
where $\frac{\partial f}{\partial x_j}$ denotes the derivative in the distribution sense.

Equipped with the norm
\[
\|f\|_D = \left[ \int_{\mathbb{R}^d} f^2 h\,dx + \sum_{1 \leq j \leq d} \int_{\mathbb{R}^d} \left( \frac{\partial f}{\partial x_j} \right)^2 h\,dx \right]^{\frac{1}{2}},
\]
$D$ is a Hilbert space, which is a classical Dirichlet space (see [9]). Moreover $D$ is a subset of the Sobolev space $H^1_{loc}(\mathbb{R}^d)$.

Let $\bar{\Omega} = \mathbb{R}^d \times \Omega$, and $\bar{F}$ the Borel $\sigma$-field over $\bar{\Omega}$ and $\bar{P} = hdx \otimes P$. Let $\bar{B}_t(x, w) = B_t(w)$ and $\bar{F}_t$ the natural filtration of $\bar{B}_t$ augmented with $\bar{P}$-negligible sets of $\bar{F}$. It is clear that $(\bar{\Omega}, \bar{F}_t, (\bar{F}_t)_{t \geq 0}, \bar{P}, \bar{B}_t)$ is a Brownian motion. We introduce the process $\bar{x}_t$ defined on the enlarged space $(\bar{\Omega}, \bar{F}, (\bar{F}_t)_{t \geq 0}, \bar{P}, \bar{B}_t)$, which is the solution of the stochastic differential equation
\[
\begin{aligned}
&\left\{ \begin{array}{l}
d\bar{x}_t = b(t, \bar{x}_t, \bar{u}_t)\,dt + \sigma(t, \bar{x}_t)\,d\bar{B}_t + G_t\,d\xi_t, \text{ for } t \in [0, T], \\
\quad \bar{x}_0 = \alpha,
\end{array} \right.
\end{aligned}
\tag{4.1}
\]
associated to the control $(\bar{u}_t, \xi_t)(x, \omega) = (u_t, \xi_t)(\omega)$.

Since the coefficients are Lipschitz continuous and bounded, equations (4.1) has a unique $\bar{F}_t$-adapted solution. Equations (2.1) and (4.1) are almost the same except that uniqueness of the solution of (4.1) is slightly weaker, one can easily prove that the uniqueness implies that for each $t \geq 0$, $\bar{x}_t = x_t$, $\bar{P}$-a.s.

4.1. The main result. The main result of this section is stated in the following Theorem.

**Theorem 4.1.** *(Stochastic maximum principle)* Let $(\hat{u}, \hat{\xi})$ be an optimal control for the controlled system (2.1), (2.2) and let $\hat{x}$ be the corresponding optimal trajectory. Then there exists a measurable $\bar{F}_t$-adapted process $p_t$ satisfying
\[
p_t := -\mathbb{E}\left[ \int_{t}^{T} \Phi^*(s, t) \cdot f_x(s, \hat{x}_s, \hat{u}_s)\,ds + \Phi^*(T, t) \cdot G_T \cdot (\hat{x}_T) / \bar{F}_t \right],
\tag{4.2}
\]
such that for all $a \in A_1$ and $\eta \in U_2$
\[
0 \leq H(t, \hat{x}_t, a, p_t) - H(t, \hat{x}_t, \hat{u}_t, p_t) \quad dt\text{-a.e, } \bar{P}\text{-a.s.},
\tag{4.3}
\]
and
\[
0 \leq \tilde{\mathbb{E}}\int_{0}^{T} (k_t + G_t^* p_t)\,d(\eta - \hat{\xi}_t)
\tag{4.4}
\]
where the Hamiltonian \( H \) is defined by
\[
H(t,x,u,p) = p \cdot b(t,x,u) - f(t,x,u),
\]
and \( \Phi(s,t), (s \geq t) \) is the fundamental solution of the linear equation
\[
\begin{align*}
\{ & d\Phi_s = b_x(s,\tilde{x}_s,\tilde{u}_s) \cdot \Phi(s,t) \, ds + \sum_{1 \leq j \leq d} \sigma^j_x(s,\tilde{x}_s) \cdot \Phi(s,t) \, d\tilde{B}^j_s, \\
\Phi(t,t) &= Id.
\end{align*}
\]
Here * denotes the transpose.

4.2. Proof of the main result. Let \( \tilde{z}_t = \tilde{x}_t - \int_0^t G_s \, d\xi_s \) the unique solution of the SDE
\[
\begin{align*}
\left\{ & d\tilde{z}_t = \tilde{b}(t,\tilde{z}_t,\tilde{u}_t) \, dt + \tilde{\sigma}(t,\tilde{z}_t) \, d\tilde{B}_t, \\
\tilde{z}_0 &= \alpha.
\end{align*}
\]
on the enlarged space \( \left( \tilde{\Omega}, \tilde{\mathbb{F}}, \left( \tilde{\mathcal{F}}_t \right)_{t \geq 0}, \tilde{\mathcal{P}}, \tilde{\mathcal{B}}_t \right) \), where \( \tilde{b} \) and \( \tilde{\sigma} \) are defined in subsection 3.2.

**Theorem 4.2.** (The Bouleau-Hirsch flow property) For \( \tilde{\mathcal{P}} \)-almost every \( w \)
1. For all \( t \geq 0 \), \( \tilde{z}_t \) is in \( D^d \).
2. There exists a \( \tilde{\mathcal{F}}_t \)-adapted GL\( _d(\mathbb{R}) \)-valued continuous process \( \left( \tilde{\Phi}_t \right)_{t \geq 0} \) such that for every \( t \geq 0 \)
\[
\frac{\partial}{\partial x} (z^\alpha_t(w)) = \tilde{\Phi}_t(\alpha, w) \, dx \text{-a.e.},
\]
where \( \frac{\partial}{\partial x} \) denotes the derivative in the distribution sense.
3. The distributional derivative \( \tilde{\Phi}_t \) is the unique fundamental solution of the linear stochastic differential equation
\[
\begin{align*}
\{ & d\Phi(s,t) = \bar{b}_x(s,\tilde{z}_s,\tilde{u}_s) \cdot \Phi(s,t) \, ds + \sum_{1 \leq j \leq d} \bar{\sigma}_x^j(s,\tilde{z}_s) \cdot \Phi(s,t) \, d\tilde{B}^j_s, \ s \geq t, \\
\Phi(t,t) &= Id,
\end{align*}
\]
where \( \bar{b}_x \) and \( \bar{\sigma}_x^j \) are versions of the almost everywhere derivatives of \( \bar{b} \) and \( \bar{\sigma}^j \).
4. The image measure of \( \tilde{\mathcal{P}} \) by the map \( \tilde{z}_t \) is absolutely continuous with respect to the Lebesgue measure.

Now, consider the process \( y_t, t \geq 0 \), solution of the system valued in \( \mathbb{R}^d \), defined on the enlarged probability space \( \left( \bar{\Omega}, \bar{\mathbb{F}}, \left( \bar{\mathcal{F}}_t \right)_{t \geq 0}, \bar{\mathcal{P}}, \bar{\mathcal{B}}_t \right) \) by
\[
\begin{align*}
\left\{ & dy_t = b^n(t,y_t,u_t) \, dt + \sigma^n(t,y_t) \, d\tilde{B}_t + G_t \, d\xi_t, \\
y_0 &= \alpha,
\end{align*}
\]
and define the cost functional
\[
J^n(u_t) = \bar{E} \left[ \int_0^T f^n(t,y_t,u_t) \, dt + \int_0^T k_t d\xi_t + g^n(y_T) \right],
\]
15
where \( b^n, \sigma^n, f^n \) and \( g^n \) be the regularized functions of \( b, \sigma, f \) and \( g \).

The following result gives the estimates which relate the original control problem with the perturbed ones.

**Lemma 4.3.** Let \((x_t)\) and \((y_t)\) the solutions of (2.1) and (4.9) respectively, corresponding to an admissible control \((u, \xi)\). Then

1. \( E \sup_{0 \leq t \leq T} |x_t - y_t|^2 \leq M_1 \epsilon_n^2 \),
2. \( |J^n(u, \xi) - J(u, \xi)| \leq M_2 \epsilon_n \),

where \( \epsilon_n = \frac{C}{n} \), and \( M_1 \) and \( M_2 \) are positive constants.

Let \( (\hat{u}, \hat{\xi}) \) be an optimal control for the initial problem (2.1) and (2.2). Note that \( (\hat{u}, \hat{\xi}) \) is not necessarily optimal for the perturbed control problem (4.9) and (4.10).

However, according to lemma 4.3, there exists \( (\delta_n) \equiv (2M_2 \epsilon_n) \) a sequence of positive real numbers converging to 0, such that

\[
J^n(\hat{u}, \hat{\xi}) \leq \inf_{(v, \eta) \in U} J^n(v, \eta) + \delta_n.
\]

The functional \( J^n \) defined by (4.10) being continuous on \( U = U_1 \times U_2 \), with respect to the topology induced by the metric \( d'((u, \xi), (v, \eta)) = d'_1(u, v) + d'_2(\xi, \eta) \), where

\[
d'_1(u, v) = \bar{P} \otimes dt \{ (w, t) \in \bar{\Omega} \times [0, T], v(w, t) \neq u(w, t) \},
\]

\[
d'_2(\xi, \eta) = \left( \bar{E} \sup_{0 \leq t \leq T} |\xi_t - \eta_t|^2 \right)^\frac{1}{2},
\]

Then by applying Ekeland’s principle to \( J^n \) for \( (\hat{u}, \hat{\xi}) \) with \( \lambda_n = \frac{\delta_n^2}{2} \), there exists an admissible control \((u^n, \xi^n)\) such that

\[
d'((\hat{u}, \hat{\xi}), (u^n, \xi^n)) \leq \frac{\delta_n^2}{2},
\]

\[
J^n_\lambda(u^n, \xi^n) \leq J^n_\lambda(v, \eta), \text{ for any } (v, \eta) \in U,
\]

and \((u^n, \xi^n)\) is an optimal control for the perturbed system (4.9) with a new cost function

\[
J^n_\lambda(v, \eta) = J^n(v, \eta) + \delta_n^\frac{1}{2} d'((v, \eta), (u^n, \xi^n)).
\]

Denote by \( x^n \) the unique solution of (4.9) corresponding to \((u^n, \xi^n)\)

\[
\begin{align*}
dx^n_t &= b^n(t, x^n_t, u^n_t) \, dt + \sigma^n(t, x^n_t) \, dB_t + G_t d\xi^n_t, \\
x^n_0 &= \alpha,
\end{align*}
\]

(4.11)

The controlled process \( dz^n_t = dx^n_t - G_t d\xi^n_t \) is then defined as the solution to the stochastic differential equation

\[
\begin{align*}
dz^n_t &= \sigma^n(t, z^n_t, u^n_t) \, dt + \sigma^n(t, z^n_t) \, dB_t, \\
z^n_0 &= \alpha.
\end{align*}
\]

(4.12)
where $\vec{b}^n_t$ and $\vec{\sigma}^n$ are defined in subsection 3.2. Let $\Phi^n(s,t)$ $(s \geq t)$, be the fundamental solution of the linear equation

\[
\begin{aligned}
d\Phi^n(s,t) &= b^n_x(s, x^n_s, u^n_s) \cdot \Phi^n(s,t) \, ds + \sum_{1 \leq j \leq d} \sigma^n_{j,x}(s, x^n_s) \cdot \Phi^n(s,t) \, dB^n_j(s), \\
\Phi^n(t,t) &= Id.
\end{aligned}
\tag{4.13}
\]

**Proposition 4.4.** For each integer $n$, there exists an admissible control $(u^n_t, \xi^n_t)$ and a $(\bar{F}_t)$-adapted process $p^n_t$ given by

\[
p^n_t = -\mathbb{E} \left[ \int_t^T \Phi^n(s,t) \cdot f^n_x(s, x^n_s, u^n_s) \, ds + \Phi^n(s,t) \cdot g^n_x(x^n_t) \big/ \bar{F}_t \right],
\tag{4.14}
\]

and a Lebesgue null set $N$ such that for $t \in N^c$

\[
\mathbb{E} \left[ H^n(t, x^n_t, u^n_t, p^n_t) - H^n(t, x^n_t, u^n_t, p^n_t) \right] \geq -\delta_M^1 M_1,
\tag{4.15}
\]

and

\[
\mathbb{E} \left[ \int_t^T (k_t + G^*_t p^n_t) \, d(\eta - \xi^n_t) \right] \geq -\delta_M^2 M_2,
\tag{4.16}
\]

for all $v \in A_1$, and $\eta \in U_2$, where the Hamiltonian $H^n$ is defined by

\[
H^n(t, x, u) = p^n_t b^n(t, x, u) - f^n(t, x, u).
\tag{4.17}
\]

Here $^*$ denotes the transpose.

The proof goes as in section 3.2.

The proof of the main result is based on the following lemma.

**Lemma 4.5.** The following estimates hold

i) $\lim_{n \to +\infty} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |x^n_t - \hat{x}_t|^2 \right] = 0,$

ii) $\lim_{n \to +\infty} \mathbb{E} \left[ \sup_{s \leq t \leq T} |\Phi^n(s,t) - \Phi(s,t)|^2 \right] = 0,$

iii) $\lim_{n \to +\infty} \mathbb{E} \left[ \sup_{0 \leq t \leq T} |p^n_t - p_t|^2 \right] = 0,$

iv) $\lim_{n \to +\infty} \mathbb{E} \left[ \|H^n(t, x^n_t, u^n_t, p^n_t) - H(t, \hat{x}_t, \hat{u}_t, p_t)\| \right] = 0,$

where $\Phi_t$, $p_t$, and $H$ are determined by (4.6), (4.2), and (4.5), corresponding to the optimal solution $\hat{x}_t$. $\Phi^n_t$, $p^n_t$, and $H^n$ are determined by (4.13), (4.14) and (4.17), corresponding to the approximating sequence $x^n_t$, given by (4.11).

**Proof.** i) is proved as (3.21).

Let us prove ii)

Using Burkholder Davis Gundy, Schwartz inequalities and the Gronwall Lemma,
we obtain
\[
\tilde{E} \left[ \sup_{t \leq s \leq T} |\Phi^n(s, t) - \Phi(s, t)|^2 \right] \leq 
\]
\[
C \tilde{E} \left[ \sup_{t \leq s \leq T} |\Phi^0(s, t)|^4 \right]^{\frac{1}{2}} \left\{ \tilde{E} \left[ \int_0^T |b^n_x(t, x^n_t, \tilde{u}_t) - b_x(t, \tilde{x}_t, \tilde{u}_t)|^4 dt \right] \right\}^{\frac{1}{2}} 
+ \sum_{i \leq j \leq d} \tilde{E} \left[ \int_0^T |\sigma^n_{j,i}(t, x^n_t) - \sigma^j_x(t, \tilde{x}_t)|^4 dt \right]^{\frac{1}{2}} \right\}.
\]

Since the coefficients in the linear stochastic differential equation (4.13) are bounded, it is easy to see that \( \tilde{E} \left[ \sup_{t \leq s \leq T} |\Phi^0(s, t)|^4 \right] < +\infty \). To derive (4.19), it is sufficient to prove the following two assertions
\[
\tilde{E} \left[ \int_0^T |b^n_x(t, x^n_t, \tilde{u}_t) - b_x(t, \tilde{x}_t, \tilde{u}_t)|^4 dt \right] \to 0 \quad \text{as } n \to +\infty,
\]
and
\[
\tilde{E} \left[ \int_0^T |\sigma^n_{j,i}(t, x^n_t) - \sigma^j_x(t, \tilde{x}_t)|^4 dt \right] \to 0 \quad \text{as } n \to +\infty, \text{ for } j=1,2,\ldots,d.
\]

Let us prove the first limit. We have
\[
\tilde{E} \left[ \int_0^T |b^n_x(t, x^n_t, \tilde{u}_t) - b_x(t, \tilde{x}_t, \tilde{u}_t)|^4 dt \right] \leq C (I^n_1 + I^n_2 + I^n_3),
\]
where
\[
I^n_1 = \tilde{E} \left[ \int_0^T |b^n_x(t, x^n_t, \tilde{u}_t) - b_x(t, x^n_t, \tilde{u}_t)|^4 \chi_{(u^n \neq \tilde{u})} (t) dt \right],
\]
\[
I^n_2 = \tilde{E} \left[ \int_0^T |b^n_x(t, x^n_t, \tilde{u}_t) - b_x(t, x^n_t, \tilde{u}_t)|^4 dt \right],
\]
\[
I^n_3 = \tilde{E} \left[ \int_0^T |b_x(t, x^n_t, \tilde{u}_t) - b_x(t, \tilde{x}_t, \tilde{u}_t)|^4 dt \right].
\]

According to the boundedness of the derivative \( b^n_x \) by the Lipschitz constant and the fact that \( d^n_x (u^n, \tilde{u}) \to 0 \) as \( n \to +\infty \), we obtain \( \lim_{n \to +\infty} I^n_1 = 0 \).

Moreover, we have
\[
I^n_2 \leq \tilde{E} \left[ \int_0^T \sup_{0 \in A_1} |b^n_x(t, z^n_t, \alpha) - \bar{b}_x(t, z^n_t, \alpha)|^4 dt \right],
\]
\[
= \int \int \sup_{0 \in A_1} |b^n_x(t, y, \alpha) - \bar{b}_x(t, y, \alpha)|^4 \rho^n_\alpha (y) dy dt,
\]
18
where \( z^n_t \) denotes the unique solution of the SDE (3.20), corresponding to \((u^n, \xi^n)\), and \( \rho^n_t(y) \) its density with respect to the Lebesgue measure. Let us show

\[
\lim_{n \to +\infty} \int_{\mathbb{R}^d} \sup_{a \in A_1} \left| \bar{b}^n_x(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \rho^n_t(y) \, dy \, dt = 0.
\]

For each \( p > 0 \), \( \bar{E} \left[ \sup_{0 \leq t \leq T} |z^n_t|^p \right] < +\infty. \) Thus, \( \lim_{R \to +\infty} \bar{P} \left( \sup_{0 \leq t \leq T} |z^n_t| > R \right) = 0 \), then it is enough to show that for every \( R > 0 \),

\[
\lim_{n \to +\infty} \int_{B(0, R)} \sup_{a \in A_1} \left| \bar{b}^n_x(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \rho^n_t(y) \, dy = 0.
\]

According to Lemma 3.2, it is easy to see that

\[
\sup_{a \in A_1} \left| \bar{b}^n_x(t, y, a) - \bar{b}_x(t, y, a) \right|^4
= \sup_{a \in A_1} \left| b^n_x \left( t, y + \int_0^T G_t d\xi^n_t, a \right) - b_x \left( t, y + \int_0^T G_t d\xi^n_t, a \right) \right|^4 \to 0 \quad dy-a.e.,
\]

at least for a subsequence. Then by Egorov’s Theorem, for every \( \delta > 0 \), there exists a measurable set \( F \) with \( \lambda(F) < \delta \), such that \( \sup_{a \in A_1} \left| \bar{b}^n_x(t, y, a) - \bar{b}_x(t, y, a) \right| \) converges uniformly to 0 on the set \( F^c \). Note that, since the Lebesgue measure is regular, \( F \) may be chosen closed. This implies that

\[
\lim_{n \to +\infty} \int_{F^c} \sup_{a \in A_1} \left| \bar{b}^n_x(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \rho^n_t(y) \, dy
\leq \lim_{n \to +\infty} \left( \sup_{y \in F^c} \sup_{a \in A_1} \left| \bar{b}^n_x(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \right) = 0.
\]

Now, by using the boundness of the derivatives \( \bar{b}^n_x, \bar{b}_x \) we have

\[
\int_F \sup_{a \in A_1} \left| \bar{b}^n_x(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \rho^n_t(y) \, dy
= \bar{E} \left[ \sup_{a \in A_1} \left| \bar{b}^n_x(t, \hat{z}^n_t, a) - \bar{b}_x(t, \hat{z}^n_t, a) \right|^4 \chi_{\{\hat{z}^n_t \in F\}} \right]
\leq 2M^4 \bar{P} (\hat{z}^n_t \in F).
\]

According to (4.18), it is easy to see that \( \hat{z}^n_t = x^n_t - \int_0^t G_s d\xi^n_s \) converges to \( \hat{z}_t = \hat{x}_t - \int_0^t G_s d\hat{\xi}_s \) in probability, then in distribution. Applying the Portmanteau-Alexandrov theorem, we obtain

\[
\lim_{n} \int_F \sup_{a \in A_1} \left| \bar{b}^n_x(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \rho^n_t(y) \, dy \leq 2M^4 \lim \sup \bar{P} (\hat{z}^n_t \in F)
\leq 2M^4 \bar{P} (\hat{z}_t \in F)
= 2M^4 \int \rho_t(y) \, dy < \varepsilon.
\]
where \( \rho_t(y) \) denotes the density of \( \hat{z}_t \) with respect to Lebesgue measure.

Now, since

\[
\begin{align*}
\int \sup_{B((0, R))} \left| \nabla^n_x (t, y, a) - \nabla^n_x (t, y, a) \right|^4 \rho^n_t (y) \, dy \\
= \int \sup_{F} \left| \nabla^n_x (t, y, a) - \nabla^n_x (t, y, a) \right|^4 \rho^n_t (y) \, dy \\
+ \int \sup_{F} \left| \nabla^n_x (t, y, a) - \nabla^n_x (t, y, a) \right|^4 \rho^n_t (y) \, dy,
\end{align*}
\]

we get \( \lim_{n \to +\infty} I^n_2 = 0 \).

Let \( k \geq 0 \) be a fixed integer, then it holds that \( I_3^k \leq C \left( J_1^k + J_2^k + J_3^k \right) \), where

\[
\begin{align*}
J_1^k &= \mathbb{E} \left[ \int_0^T \left| b_x (t, x^n_t, \hat{u}_t) - b_x (t, x^n_t, \hat{u}_t) \right|^4 \, dt \right], \\
J_2^k &= \mathbb{E} \left[ \int_0^T \left| b_x^k (t, x^n_t, \hat{u}_t) - b_x^k (t, \hat{x}_t, \hat{u}_t) \right|^4 \, dt \right], \\
J_3^k &= \mathbb{E} \left[ \int_0^T \left| b_x^k (t, \hat{x}_t, \hat{u}_t) - b_x (t, \hat{x}_t, \hat{u}_t) \right|^4 \, dt \right].
\end{align*}
\]

Applying the same arguments used in the first limit (Egorov and Portmanteau-Alexandrov Theorems), we obtain that \( \lim_{n \to +\infty} J_1^k = 0 \). We use the continuity of \( b_x^k \) in \( x \) and the convergence in probability of \( x^n_T \) to \( \hat{x}_T \) to deduce that \( b_x^k (t, x^n_t, \hat{u}_t) \) converges to \( b_x^k (t, \hat{x}_t, \hat{u}_t) \) in probability as \( n \to +\infty \), and to conclude by using the dominated convergence theorem that \( \lim_{n \to +\infty} J_2^k = 0 \).

\[
J_3^k = \mathbb{E} \left[ \int_0^T \sup_{a \in A_1} \left| \nabla_x^k (t, \hat{z}_t, a) - \nabla_x (t, \hat{z}_t, a) \right|^4 \, dt \right]
\]

\[
= \int \int_0^T \sup_{a \in A_1} \left| \nabla_x^k (t, y, a) - \nabla_x (t, y, a) \right|^4 \rho_t (y) \, dy \, dt
\]

\( \nabla_x^k, \nabla_x \) being bounded, then by using the convergence of \( \nabla_x^k \) to \( \nabla_x \), and the dominated convergence theorem, we get \( \lim_{n \to +\infty} J_3^k = 0 \).

iii) and iv) are proved by using the same techniques as in ii) and lemma 3.5. \( \square \)

**Proof. of Theorem 4.1.** Use the Corollary 4.5 and Lemma 4.6. \( \square \)

**REFERENCES**
