Optimality necessary conditions in singular stochastic control problems with nonsmooth data ☆

K. Bahlali [a], F. Chighoub [b], B. Djehiche [c], B. Mezerdi [d,*]

[a] UFR Sciences, UTV, B.P. 132, 83957 La Garde, Cedex, France
[b] Laboratory of Applied Mathematics, University Mohamed Khider, PO Box 145, Biskra 07000, Algeria
[c] Dept. of Mathematics, Royal Institute of Technology, S 100 44, Stockholm, Sweden
[d] Laboratory of Applied Mathematics, University Med Khider, PO Box 145, Biskra 07000, Algeria

ARTICLE INFO

ABSTRACT

The present paper studies the stochastic maximum principle in singular optimal control, where the state is governed by a stochastic differential equation with nonsmooth 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 nondegenerate 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.

© 2009 Elsevier Inc. All rights reserved.

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 a problem in which the state evolves according to the d-dimensional stochastic differential equation

$$\begin{cases} dx_t = b(t, x_t, u_t)\, dt + \sigma(t, x_t)\, dB_t + G_t\, d\xi_t, & \text{for } t \in [0, T], \\ x_0 = \alpha, \end{cases} \tag{1.1}$$

and the expected cost has the form

$$J(u, \xi) = E\left[\int_0^T f(t, x_t, u_t)\, dt + \int_0^T k_t\, d\xi_t + g(x_T) \right]. \tag{1.2}$$

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 has been produced on the subject, including Beněs, 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 a 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 the 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 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) has been studied, with differentiability assumptions on the data weakened. The first result has been derived by Mezerdi [22], in the case of an SDE with a nonsmooth 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. See also [7,8,11,21,24] for other versions of the maximum principle.

Our aim in this paper is to extend the stochastic maximum principle in singular optimal control, to the case where the coefficients b, σ, 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 nondegenerate case and the Bouleau–Hirsch flow property in the degenerate case.

2. Assumptions and preliminaries

Let (Ω, 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 nonempty subset of $\mathbb{R}^n$ and $A_2 = ([0, \infty))^m$. U_1 is the class of measurable, adapted processes $u : [0, T] \times \Omega \to A_1$, and U_2 is the class of measurable, adapted processes $\xi : [0, T] \times \Omega \to A_2$.

Definition 2.1. An admissible control is a pair (u, ξ) of measurable $A_1 \times A_2$-valued, F_t-adapted processes, such that ξ is of bounded variation, nondecreasing 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

$$\begin{cases} dx_t = b(t, x_t, u_t)\, dt + \sigma(t, x_t)\, dB_t + G_t\, d\xi_t, & \text{for } t \in [0, T], \\ x_0 = \alpha. \end{cases} \tag{2.1}$$

Since $d\xi_t$ may be singular with respect to Lebesgue measure dt, we call ξ 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

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

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

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

$$\left| b(t, x, a) - b(t, y, a) \right| + \left| \sigma(t, x) - \sigma(t, y) \right| \leqslant M|x - y|, \tag{2.3}$$

$$\left| f(t, x, a) - f(t, y, a) \right| + \left| g(x) - g(y) \right| \leqslant M|x - y|, \tag{2.4}$$

$$b(t, x, a) \text{ and } f(t, x, a) \text{ are continuous in } a \text{ uniformly in } (t, x), \tag{2.5}$$

and

> G, k are continuous and bounded. (2.6)

The problem is to 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 for 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 $p > 0$,

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

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

Since b, σ^j (the jth column of the matrix σ), 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, σ_x, f_x and g_x any Borel measurable functions such that

$$\partial_x b(t, x, a) = b_x(t, x, a) \quad dx\text{-a.e.},$$

$$\partial_x f(t, x, a) = f_x(t, x, a) \quad dx\text{-a.e.},$$

$$\partial_x \sigma(t, x) = \sigma_x(t, x) \quad dx\text{-a.e.},$$

$$\partial_x g(x) = g_x(x) \quad dx\text{-a.e.}$$

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).

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

Theorem 2.1. *(See Krylov [20].) Let (Ω, F, F_t, P) be a filtered probability space, $(B_t)_{t \geqslant 0}$ be a d-dimensional Brownian motion and $b : \Omega \times \mathbb{R}_+ \to \mathbb{R}^d$, $\sigma : \Omega \times \mathbb{R}_+ \to \mathbb{R}^d \otimes \mathbb{R}^d$ be bounded adapted processes such that: $\exists c > 0, \forall \zeta \in \mathbb{R}^d, \forall (t, x) \in [0, T] \times \mathbb{R}^d,$ $\zeta^* \sigma \sigma^* \zeta \geqslant c|\zeta|^2$. Let*

$$x_t = x + \int_0^t b(s, \omega) \, dt + \int_0^t \sigma(s, \omega) \, dB_s,$$

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

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

where K is a constant and $B(0, M)$ is the ball of center 0 and radius M.

Lemma 2.1 *(Ekeland variational principle). (See [15].) Let (S, d) be metric space and $\rho : S \to \mathbb{R} \cup \{+\infty\}$ be lower-semicontinuous and bounded from below. For $\varepsilon \geqslant 0$, suppose $u^\varepsilon \in S$ satisfies $\rho(u^\varepsilon) \leqslant \inf_{u \in S} \rho(u) + \varepsilon$. Then for any $\lambda > 0$, there exists $u^\lambda \in S$ such that*

$$\rho(u^\lambda) \leqslant \rho(u^\varepsilon),$$

$$d(u^\lambda, u^\varepsilon) \leqslant \lambda,$$

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

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, ξ), $(v, \eta) \in U$, we set

$$d_1(u, v) = P \otimes dt \{ (\omega, t) \in \Omega \times [0, T], \ v(\omega, t) \neq u(\omega, t) \}, \tag{2.7}$$

$$d_2(\xi, \eta) = \left(E\left[\sup_{0 \leqslant t \leqslant T} |\xi_t - \eta_t|^2 \right] \right)^{\frac{1}{2}}, \tag{2.8}$$

$$d\big((u, \xi), (v, \eta)\big) = d_1(u, v) + d_2(\xi, \eta), \tag{2.9}$$

where $P \otimes dt$ is the product measure of P with the Lebesgue measure dt.

Lemma 2.2.

(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 nondegenerate case

In this section, we assume the following condition:

$$\exists c > 0, \ \forall \zeta \in \mathbb{R}^d, \ \forall (t, x) \in [0, T] \times \mathbb{R}^d, \quad \zeta^* \sigma(t, x) \sigma^*(t, x) \zeta \geqslant c |\zeta|^2. \tag{3.1}$$

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

$$p_t := E\left[\int_t^T \Phi^*(s, t).f_x(s, \hat{x}_s, \hat{u}_s)\, ds + \Phi^*(T, t).g_x(\hat{x}_T)/F_t\right], \tag{3.2}$$

such that for all $a \in A_1$ and $\eta \in U_2$

$$0 \leqslant H(t, \hat{x}_t, a, p_t) - H(t, \hat{x}_t, \hat{u}_t, p_t) \quad dt\text{-a.e.}, \ P\text{-a.s.}, \tag{3.3}$$

and

$$0 \leqslant E\int_0^T \left(k_t + G_t^* p_t\right) d(\eta - \hat{\xi})_t, \tag{3.4}$$

where the Hamiltonian H associated to the control problem is

$$H(t, x, u, p) = p.b(t, x, u) + f(t, x, u), \tag{3.5}$$

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

$$\begin{cases} d\Phi(s, t) = b_x(s, \hat{x}_s, \hat{u}_s).\Phi(s, t)\, ds + \displaystyle\sum_{1 \leqslant j \leqslant d} \sigma_x^j(s, \hat{x}_s).\Phi(s, t)\, dB_s^j, \\ \Phi(t, t) = Id. \end{cases} \tag{3.6}$$

Here Φ^* denotes the transpose of the matrix Φ.

3.2. Proof of the main result

Let φ be a nonnegative 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

$$b^n(t, x, a) = n^d \int_{\mathbb{R}^d} b(t, x - y, a)\varphi(ny)\, dy,$$

$$f^n(t, x, a) = n^d \int_{\mathbb{R}^d} f(t, x - y, a)\varphi(ny)\, dy,$$

$$\sigma^{j,n}(t, x) = n^d \int_{\mathbb{R}^d} \sigma^j(t, x - y)\varphi(ny)\, dy,$$

$$g^n(x) = n^d \int_{\mathbb{R}^d} g(x - y)\varphi(ny)\, dy.$$

Lemma 3.1.

(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 positive constant C, independent of t, x and n such that for every t in $[0, T]$*

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

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

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

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

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

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

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

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

$$\lim_{n \to \infty} \iint_{B(0,R) \times [0,T]} \sup_{a \in A} \left| b_x^n(t, x, a) - b_x(t, x, a) \right|^p dx\, dt = 0,$$

$$\lim_{n \to \infty} \iint_{B(0,R) \times [0,T]} \sup_{a \in A} \left| f_x^n(t, x, a) - f_x(t, x, a) \right|^p dx\, dt = 0.$$

Proof. Statements (1)–(3) are classical facts (see [16] for the proof).
(4) is proved as in [5] Lemma 2.3. □

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

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

The corresponding cost is given by

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

Lemma 3.2. *Let $(u, \xi) \in U$, x_t and y_t the solutions of (2.1) and (3.7) respectively corresponding to the control (u, ξ), then we have*

(1) $E[\sup_{0 \leqslant t \leqslant T} |x_t - y_t|^2] \leqslant M_1.(\epsilon_n)^2$, *where* $\epsilon_n = \frac{C}{n}$.
(2) $|J^n(u, \xi) - J(u, \xi)| \leqslant M_2.\epsilon_n$.

Proof. Since $x_t - y_t$ and $J^n(u, \xi) - J(u, \xi)$ do not depend on the singular part, then this lemma follows from standard arguments from stochastic calculus and Lemma 3.1. □

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 (3.7) and (3.8). However, by Lemma 3.2, we obtain the existence of $(\delta_n) \equiv (2M_2.\epsilon_n)$, a sequence of positive real numbers converging to 0, such that

$$J^n(\hat{u}, \hat{\xi}) \leqslant \inf_{(\upsilon, \eta) \in U} J^n(\upsilon, \eta) + \delta_n.$$

The control $(\hat{u}, \hat{\xi})$ is then δ_n-optimal for the perturbed control problem. According to Lemma 2.2, it is easy to see that $J^n(.,.)$ is continuous on $U = U_1 \times U_2$ endowed with the metric $d = d_1 + d_2$ defined by (2.9). By Ekeland's variational principle (Lemma 2.1) applied to $(\hat{u}, \hat{\xi})$ with $\lambda_n = \delta_n^{\frac{2}{3}}$, there exists an admissible control (u^n, ξ^n) such that

$$d\big((\hat{u}, \hat{\xi}), (u^n, \xi^n)\big) \leqslant \delta_n^{\frac{2}{3}},$$

and

$$J_\delta^n\big(u^n, \xi^n\big) \leqslant J_\delta^n(\upsilon, \eta), \quad \text{for all } (\upsilon, \eta) \in U,$$

where

$$J_\delta^n(\upsilon, \eta) = J^n(\upsilon, \eta) + \delta_n^{\frac{1}{3}} d\big((\upsilon, \eta), (u^n, \xi^n)\big).$$

This means that (u^n, ξ^n) is an optimal control for the perturbed system (3.7) with a new cost function J_δ^n. The controlled process x^n is defined as the unique solution to the stochastic differential equation,

$$\begin{cases} dx_t^n = b^n\big(t, x_t^n, u_t^n\big) dt + \sigma^n\big(t, x_t^n\big) dB_t + G_t \, d\xi_t^n, \\ y_0 = \alpha. \end{cases} \tag{3.9}$$

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

$$\begin{cases} d\Phi^n(s, t) = b_x^n\big(s, x_s^n, u_s^n\big).\Phi^n(s, t) \, ds + \displaystyle\sum_{1 \leqslant j \leqslant d} \sigma_x^{j,n}\big(s, x_s^n\big).\Phi^n(s, t) \, dB_s^j, \\ \\ \Phi^n(t, t) = Id. \end{cases} \tag{3.10}$$

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

Proposition 3.1. *For each integer n, there exist an admissible control (u^n, ξ^n) and an (F_t)-adapted process p_t^n given by*

$$p_t^n = E\left[\int\limits_t^T \Phi^{n,*}(s, t).f_x^n\big(s, x_s^n, u_s^n\big) \, ds + \Phi^{n,*}(T, t).g_x^n\big(x_T^n\big)/F_t \right], \tag{3.11}$$

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

$$E\big[H^n\big(t, x_t^n, \upsilon, p_t^n\big) - H^n\big(t, x_t^n, u_t^n, p_t^n\big)\big] \geqslant -\delta_n^{\frac{1}{3}}.M_1, \tag{3.12}$$

and

$$E \int\limits_0^T \big(k_t + G_t^* p_t^n\big) d\big(\eta - \xi^n\big)_t \geqslant -\delta_n^{\frac{1}{3}}.M_2, \tag{3.13}$$

for all $\upsilon \in A_1$, and $\eta \in U_2$. The Hamiltonian H^n is defined by

$$H^n(t, x, u, p) = p.b^n(t, x, u) + f^n(t, x, u). \tag{3.14}$$

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

$$\big(u_t^{n,\varepsilon}, \xi_t^n\big) = \begin{cases} (\upsilon, \xi_t^n), & t \in [t_0, t_0 + \varepsilon], \\ (u_t^n, \xi_t^n), & t \in [0, T]\backslash[t_0, t_0 + \varepsilon], \end{cases}$$

and

$$\big(u_t^n, \xi_t^{n,\varepsilon}\big) = \big(u_t^n, \xi_t^n + \varepsilon\big(\eta_t - \xi_t^n\big)\big).$$

Since (u_t^n, ξ_t^n) is optimal for the cost J_δ^n, then

$$0 \leqslant J_\delta^n\big(u_t^{n,\varepsilon}, \xi_t^n\big) - J_\delta^n\big(u_t^n, \xi_t^n\big)$$

and

$$0 \leqslant J_\delta^n\big(u_t^n, \xi_t^{n,\varepsilon}\big) - J_\delta^n\big(u_t^n, \xi_t^n\big).$$

This implies that

$$0 \leqslant J^n\big(u_t^{n,\varepsilon}, \xi_t^n\big) - J^n\big(u_t^n, \xi_t^n\big) + \delta_n^{\frac{1}{3}}.d_1\big(u_t^n, u_t^{n,\varepsilon}\big),$$

and

$$0 \leqslant J^n\big(u_t^n, \xi_t^{n,\varepsilon}\big) - J^n\big(u_t^n, \xi_t^n\big) + \delta_n^{\frac{1}{3}}.d_2\big(\xi_t^n, \xi_t^{n,\varepsilon}\big).$$

Using the definitions of d_1 and d_2 it holds that

$$0 \leqslant J^n\big(u_t^{n,\varepsilon}, \xi_t^n\big) - J^n\big(u_t^n, \xi_t^n\big) + \delta_n^{\frac{1}{3}}.M_1\varepsilon, \tag{3.15}$$

and

$$0 \leqslant J^n\big(u_t^n, \xi_t^{n,\varepsilon}\big) - J^n\big(u_t^n, \xi_t^n\big) + \delta_n^{\frac{1}{3}}.M_2\varepsilon, \tag{3.16}$$

where M_i $(i = 1, 2)$ is a positive constant. From inequalities (3.15) and (3.16) respectively we use the same method as in Section 3.3 in [2] to obtain respectively (3.12) and (3.13). $\square$

We use a transformation that makes it possible to apply Krylov's estimate for diffusion processes. Define the coefficients $\bar{b} : [0, T] \times \mathbb{R}^d \times A_1 \to \mathbb{R}^d$, $\bar{b}^n : [0, T] \times \mathbb{R}^d \times A_1 \to \mathbb{R}^d$, $\bar{\sigma} : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$, and $\bar{\sigma}^n : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^d$, by

$$\bar{b}(t, x, a) = b\left(t, x + \int_0^t G_s \, d\xi_s, a\right),$$

$$\bar{b}^n(t, x, a) = b^n\left(t, x + \int_0^t G_s \, d\xi_s, a\right),$$

$$\bar{\sigma}(t, x) = \sigma\left(t, x + \int_0^t G_s \, d\xi_s\right),$$

$$\bar{\sigma}^n(t, x) = \sigma^n\left(t, x + \int_0^t G_s \, d\xi_s\right).$$

Let z be the unique solution of

$$\begin{cases} dz_t = \bar{b}(t, z_t, u_t) \, dt + \bar{\sigma}(t, z_t) \, dB_t, \\ z_0 = \alpha. \end{cases} \tag{3.17}$$

This implies that $x_t = z_t + \int_0^t G_s \, d\xi_s$ solves the SDE (2.1) with data (b, σ).
Similary, let z^n be the unique solution of

$$\begin{cases} dz_t^n = \bar{b}^n\big(t, z_t^n, u_t\big) \, dt + \bar{\sigma}^n\big(t, z_t^n\big) \, dB_t, \\ z_0^n = \alpha. \end{cases} \tag{3.18}$$

Then $x_t^n = z_t^n + \int_0^t G_s \, d\xi_s$ solves the SDE (3.7) with data (b^n, σ^n).

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

Lemma 3.3. *The following estimates*

$$\lim_{n \to +\infty} E\left[\sup_{0 \leqslant t \leqslant T} \big|x_t^n - \hat{x}_t\big|^2\right] = 0, \tag{3.19}$$

$$\lim_{n \to +\infty} E\left[\sup_{t \leqslant s \leqslant T} \big|\Phi^n(s, t) - \Phi(s, t)\big|^2\right] = 0, \tag{3.20}$$

$$\lim_{n \to +\infty} E\left[\sup_{0 \leqslant t \leqslant T} \big|p_t^n - p_t\big|^2\right] = 0, \tag{3.21}$$

$$\lim_{n \to +\infty} E\big[\big|H^n\big(t, x_t^n, u_t^n, p_t^n\big) - H(t, \hat{x}_t, \hat{u}_t, p_t)\big|\big] = 0, \tag{3.22}$$

hold at least for a subsequence.

Φ_t, p_t and H are determined respectively by the solution of (3.6)*, the adjoint process* (3.2) *and the associated Hamiltonian* (3.5)*, corresponding to the optimal state process $\hat{x}_t$. Φ_t^n, p_t^n and H^n are determined respectively by the solution* (3.10)*, the adjoint process* (3.11) *and the associated Hamiltonian* (3.14)*, corresponding to the approximating sequence x_t^n, given by* (3.9)*.*

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

$$E\left[\sup_{0\leqslant t\leqslant T}\left|x_t^n-\hat{x}_t\right|^2\right]\leqslant C\left(A_1^n+A_2^n+A_3^n+M.\left(d_2\left(\xi^n,\hat{\xi}\right)\right)^2\right),$$

where M is a positive constant, and

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

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

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

By using the boundedness 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 σ^n are Lipschitz in the state variable, then

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

Finally, we conclude from Lemma 3.2 that $\lim_{n\to+\infty}A_3^n=0$. Then by Gronwall Lemma, we obtain (3.19).

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

$$E\left[\sup_{t\leqslant s\leqslant T}\left|\Phi^n(s,t)-\Phi(s,t)\right|^2\right]\leqslant CE\left[\sup_{t\leqslant s\leqslant T}\left|\Phi^n(s,t)\right|^4\right]^{\frac{1}{2}}\left\{E\left[\int_0^T\left|b_x^n\left(t,x_t^n,u_t^n\right)-b_x(t,\hat{x}_t,\hat{u}_t)\right|^4\,dt\right]^{\frac{1}{2}}\right.$$

$$\left.+\sum_{1\leqslant j\leqslant d}E\left[\int_0^T\left|\sigma_x^{j,n}\left(t,x_t^n\right)-\sigma_x^j(t,\hat{x}_t)\right|^4\,dt\right]^{\frac{1}{2}}\right\}.$$

Since the coefficients in the linear stochastic differential equation (3.10) are bounded, it is easy to see that $E[\sup_{s\leqslant t\leqslant T}|\Phi^n(s,t)|^4]<+\infty$. To obtain the desired result it is sufficient to prove that

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

$$\lim_{n\to+\infty}E\left[\int_0^T\left|\sigma_x^{j,n}\left(t,x_t^n\right)-\sigma_x^j(t,\hat{x}_t)\right|^4\,dt\right]=0,\quad\text{for }j=1,\ldots,d.$$

We have

$$E\left[\int_0^T\left|b_x^n\left(t,x_t^n,u_t^n\right)-b_x(t,\hat{x}_t,\hat{u}_t)\right|^4\,dt\right]\leqslant C\left(I_1^n+I_2^n\right),$$

where

$$I_1^n=E\left[\int_0^T\left|b_x^n\left(t,x_t^n,u_t^n\right)-b_x^n\left(t,x_t^n,\hat{u}_t\right)\right|^4\chi_{\{u^n\neq\hat{u}\}}(t)\,dt\right],$$

$$I_2^n=E\left[\int_0^T\left|b_x^n\left(t,x_t^n,\hat{u}_t\right)-b_x(t,\hat{x}_t,\hat{u}_t)\right|^4\,dt\right].$$

First, in view of the boundedness of the derivative b_x^n by the Lipschitz constant and the fact that $d_1(u^n,\hat{u})\to 0$ as $n\to+\infty$, we obtain $\lim_{n\to+\infty}I_1^n=0$. Next, let $k\geqslant 1$ be a fixed integer, then we get

$$\lim_{n \to +\infty} I_2^n \leqslant \overline{\lim_n} C.\big\{ J_1^n + J_2^n + J_3^n \big\},$$

where

$$J_1^n = E\left[\int_0^T \big| b_x^n(t, x_t^n, \hat{u}_t) - b_x^k(t, x_t^n, \hat{u}_t) \big|^4 dt \right],$$

$$J_2^n = E\left[\int_0^T \big| b_x^k(t, x_t^n, \hat{u}_t) - b_x^k(t, \hat{x}_t, \hat{u}_t) \big|^4 dt \right],$$

$$J_3^n = E\left[\int_0^T \big| b_x^k(t, \hat{x}_t, \hat{u}_t) - b_x(t, \hat{x}_t, \hat{u}_t) \big|^4 dt \right].$$

Now, let $\hat{z}$ (resp. z^n) denote the unique solution of the SDE (3.17) (resp. (3.18)) corresponding to $(\hat{u}, \hat{\xi})$ (resp. (u^n, ξ^n)), then it holds that

$$J_1^n = E\left[\int_0^T \big| \bar{b}_x^n(t, z_t^n, \hat{u}_t) - \bar{b}_x^k(t, z_t^n, \hat{u}_t) \big|^4 dt \right],$$

and

$$J_3^n = E\left[\int_0^T \big| \bar{b}_x^k(t, \hat{z}_t, \hat{u}_t) - \bar{b}_x(t, \hat{z}_t, \hat{u}_t) \big|^4 dt \right].$$

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

$$\overline{\lim_n} J_1^n \leqslant C E\left[\int_0^T \left(1 - w\left(\frac{t}{M}, \frac{\hat{z}_t}{M} \right) \right) dt \right] + C \overline{\lim_n} E\left[\int_0^T w\left(\frac{t}{M}, \frac{\hat{z}_t}{M} \right).\big| \bar{b}_x^n(t, z_t^n, \hat{u}_t) - \bar{b}_x^k(t, z_t^n, \hat{u}_t) \big|^4 dt \right].$$

Therefore without loss of generality, we may suppose that for all $n \in \mathbb{N}^*$, the functions $\bar{b}_x$, $\bar{\sigma}_x$, $\bar{b}_x^n$, and $\bar{\sigma}_x^n$ have compact support in $[0, T] \times B(0, M)$. Since the diffusion matrix $\bar{\sigma}^n$ is nondegenerate, then by applying Krylov's inequality, we obtain

$$\overline{\lim_n} J_1^n \leqslant C E\left[\int_0^T \left(1 - w\left(\frac{t}{M}, \frac{\hat{z}_t}{M} \right) \right) dt \right] + C \overline{\lim_n} \left\| \sup_{a \in A_1} \big| \bar{b}_x^n(t, x, a) - \bar{b}_x^k(t, x, a) \big|^4 \right\|_{d+1, M}.$$

Since b_x^n converges to b_x dx-a.e., it is simple to see that $\bar{b}_x^n$ converges to $\bar{b}_x$ dx-a.e. and

$$\overline{\lim_n} \left\| \sup_{a \in A_1} \big| \bar{b}_x^n(t, x, a) - \bar{b}_x^k(t, x, a) \big|^4 \right\|_{d+1, M} = 0.$$

Next, let M go to $+\infty$, then from the properties of the function $w(t, x)$ we have $\overline{\lim_n} J_1^n = 0$. Estimating J_3^n similarly, it holds that $\overline{\lim_n} J_3^n = 0$. By using the continuity of b_x^k in x, relation (3.19), and the Dominated Convergence Theorem, we deduce that $\overline{\lim_n} J_2^n = 0$. Hence $\lim_{n \to +\infty} I_1^n = 0$. Using the same technique, we prove that

$$\lim_{n \to +\infty} E\left[\int_0^T \big| \sigma_x^{j,n}(t, x_t^n) - \sigma_x^j(t, \hat{x}_t) \big|^4 dt \right] = 0, \quad \text{for } j = 1, \ldots, d.$$

Now, let us prove that $\lim_{n \to +\infty} E[\sup_{0 \leqslant t \leqslant T} |p_t^n - p_t|^2] = 0$. Clearly,

$$E\big[|p_t^n - p_t|^2 \big] \leqslant C\big(\alpha_1^n + \alpha_2^n \big), \tag{3.23}$$

where

$$\alpha_1^n = E\left[\int_t^T \big| \Phi^{n,*}(s, t).f_x^n(s, x_s^n, u_s^n) - \Phi^*(s, t).f_x(s, \hat{x}_s, \hat{u}_s) \big|^2 ds \right],$$

and

$$\alpha_2^n = E\big[\big|\Phi^{n,*}(T,t).g_x^n\big(x_T^n\big) - \Phi^*(T,t).g_x(\hat{x}_T)\big|^2\big].$$

Since f_x is bounded by the Lipschitz constant M, then applying Schwarz inequality, we get

$$\alpha_1^n \leqslant CE\bigg[\sup_{t\leqslant s\leqslant T}\big|\Phi^{n,*}(s,t)\big|^4\bigg]^{\frac{1}{2}}.E\bigg[\int\limits_0^T\big|f_x^n\big(s,x_s^n,u_s^n\big) - f_x(s,\hat{x}_s,\hat{u}_s)\big|^4\,ds\bigg]^{\frac{1}{2}} + CM.E\bigg[\sup_{t\leqslant s\leqslant T}\big|\Phi^{n,*}(s,t) - \Phi^*(s,t)\big|^2\bigg].$$

We use the continuity and the boundedness of derivatives f_x^n, f_x, relations (3.19), (3.20) 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 $f_x^n(s,x_s^n,u_s^n) - f_x(s,\hat{x}_s,\hat{u}_s)$, to conclude that $\lim_{n\to+\infty}\alpha_1^n = 0$.

On the other hand, since g_x is bounded by the Lipschitz constant, then applying Schwarz inequality we get

$$\alpha_2^n \leqslant C\big\{E\big[\big|\Phi^{n,*}(T,t)\big|^4\big]\big\}^{\frac{1}{2}}.\big\{E\big[\big|g_x^n\big(x_T^n\big) - g_x(\hat{x}_T)\big|^4\big]\big\}^{\frac{1}{2}} + CM.E\big[\big|\Phi^{n,*}(T,t) - \Phi^*(T,t)\big|^2\big].$$

Since, g_x^n and g_x are bounded by the Lipschitz constant and g_x^n converges to g_x, we conclude by using (3.19) and the Dominated Convergence Theorem that

$$\lim_{n\to+\infty}E\big[\big|g_x^n\big(x_T^n\big) - g_x(\hat{x}_T)\big|^4\big] = 0.$$

Then by using Burkholder–Davis–Gundy inequality together with (3.23), we obtain (3.21).

Now let us prove (3.22).

Apply Schwarz inequality to get

$$E\big[\big|H^n\big(t,x_t^n,u_t^n,p_t^n\big) - H(t,\hat{x}_t,\hat{u}_t,p_t)\big|\big] \leqslant \big\{E\big|p_t^n - p_t\big|^2\big\}^{\frac{1}{2}}\big\{E\big|b^n\big(t,x_t^n,u_t^n\big)\big|^2\big\}^{\frac{1}{2}}$$
$$+ \big\{E\big|b^n\big(t,x_t^n,u_t^n\big) - b(t,\hat{x}_t,\hat{u}_t)\big|^2\big\}^{\frac{1}{2}}\big\{E|p_t|^2\big\}^{\frac{1}{2}}$$
$$+ E\big|f^n\big(t,x_t^n,u_t^n\big) - f(t,\hat{x}_t,\hat{u}_t)\big|.$$

Lemma 3.1 and relation (3.21) imply that the first expression in the right-hand side converges to 0 as $n \to +\infty$.

Next,

$$E\big|b^n\big(t,x_t^n,u_t^n\big) - b(t,\hat{x}_t,\hat{u}_t)\big|^2 \leqslant C\big(\beta_1^n + \beta_2^n + \beta_3^n\big),$$

where

$$\beta_1^n = E\big[\big|b^n\big(t,x_t^n,u_t^n\big) - b^n\big(t,x_t^n,\hat{u}_t\big)\big|^2\chi_{\{u^n\neq\hat{u}\}}(t)\big],$$
$$\beta_2^n = E\big[\big|b^n\big(t,x_t^n,\hat{u}_t\big) - b^n(t,\hat{x}_t,\hat{u}_t)\big|^2\big],$$
$$\beta_3^n = E\big[\big|b^n(t,\hat{x}_t,\hat{u}_t) - b(t,\hat{x}_t,\hat{u}_t)\big|^2\big].$$

The boundedness of b^n and the fact that $d_1(u^n,\hat{u})\to_{n\to\infty}0$ (which implies the a.s. convergence to $\hat{u}$ of a subsequence which is still denoted by (u^n)), guarantee the convergence of β_1^n to 0 as $n \to +\infty$. By virtue of (3.19), and the Dominated Convergence Theorem, we get, $\lim_{n\to+\infty}\beta_2^n = 0$. In view of Lemma 3.1, we have $\lim_{n\to+\infty}\beta_3^n = 0$.

The term $E|f^n(t,x_t^n,u_t^n) - f(t,\hat{x}_t,\hat{u}_t)|$ can be treated by using the same technique. $\square$

Proof of Theorem 3.1. Let n go to $+\infty$, then from Proposition 3.1 and Lemma 3.3, we get

$$E\big[H(t,\hat{x}_t,v,p_t) - H(t,\hat{x}_t,\hat{u}_t,p_t)\big] \geqslant 0, \quad dt\text{-a.e.}, \ P\text{-a.s.},$$
$$E\int\limits_0^T\big(k_t + G_t^*p_t\big)\,d(\eta - \hat{\xi})_t \geqslant 0,$$

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

Let $a \in A_1$, then for every $A_t \in F_t$

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

which implies that

$$E\big[\big(H(t,\hat{x}_t,a,p_t) - H(t,\hat{x}_t,\hat{u}_t,p_t)\big)/F_t\big] \geqslant 0.$$

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. $\square$

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 an 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

$$\|f\|_D = \left[\int_{\mathbb{R}^d} f^2 h\,dx + \sum_{1 \leqslant j \leqslant d} \int_{\mathbb{R}^d} \left(\frac{\partial f}{\partial x_j} \right)^2 h\,dx \right]^{\frac{1}{2}},$$

D is a Hilbert space, which is a classical Dirichlet space (see [9]). Moreover D is a subset of the Sobolev space $H^1_{\mathrm{loc}}(\mathbb{R}^d)$.

Let $\widetilde{\Omega} = \mathbb{R}^d \times \Omega$, and $\widetilde{F}$ be the Borel σ-field over $\widetilde{\Omega}$ and $\widetilde{P} = h\,dx \otimes P$. Let $\widetilde{B}_t(x, w) = B_t(w)$ and $\widetilde{F}_t$ be the natural filtration of $\widetilde{B}_t$ augmented with $\widetilde{P}$-negligible sets of $\widetilde{F}$. It is clear that $(\widetilde{\Omega}, \widetilde{F}, (\widetilde{F}_t)_{t \geqslant 0}, \widetilde{P}, \widetilde{B}_t)$ is a Brownian motion. We introduce the process $\tilde{x}_t$ defined on the enlarged space $(\widetilde{\Omega}, \widetilde{F}, (\widetilde{F}_t)_{t \geqslant 0}, \widetilde{P}, \widetilde{B}_t)$, which is the solution of the stochastic differential equation

$$\begin{cases} d\tilde{x}_t = b(t, \tilde{x}_t, \tilde{u}_t)\,dt + \sigma(t, \tilde{x}_t)\,d\widetilde{B}_t + G_t\,d\tilde{\xi}_t, & \text{for } t \in [0, T], \\ \tilde{x}_0 = \alpha, \end{cases} \tag{4.1}$$

associated to the control $(\tilde{u}_t, \tilde{\xi}_t)(x, \omega) = (u_t, \xi_t)(\omega)$.

Since the coefficients are Lipschitz continuous and bounded, Eq. (4.1) has a unique $\widetilde{F}_t$-adapted solution. Eqs. (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 $t \geqslant 0$, $\tilde{x}_t = x_t$, $\widetilde{P}$-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 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*

$$p_t := \widetilde{E}\left[\int_t^T \Phi^*(s, t).f_x(s, \hat{x}_s, \hat{u}_s)\,ds + \Phi^*(T, t).g_x(\hat{x}_T)/\widetilde{F}_t \right], \tag{4.2}$$

such that for all $a \in A_1$ *and* $\eta \in U_2$

$$0 \leqslant H(t, \hat{x}_t, a, p_t) - H(t, \hat{x}_t, \hat{u}_t, p_t) \quad dt\text{-a.e., } \widetilde{P}\text{-a.s.,} \tag{4.3}$$

and

$$0 \leqslant \widetilde{E} \int_0^T \left(k_t + G_t^* p_t \right) d(\eta - \hat{\xi})_t, \tag{4.4}$$

where the Hamiltonian H is defined by

$$H(t, x, u, p) = p.b(t, x, u) + f(t, x, u), \tag{4.5}$$

and $\Phi(s, t)$, $(s \geqslant t)$ *is the fundamental solution of the linear equation*

$$\begin{cases} d\Phi_s = b_x(s, \hat{x}_s, \hat{u}_s).\Phi(s, t)\,ds + \sum_{1 \leqslant j \leqslant d} \sigma_x^j(s, \hat{x}_s).\Phi(s, t)\,d\widetilde{B}_s^j, \\ \Phi(t, t) = Id. \end{cases} \tag{4.6}$$

4.2. Proof of the main result

Let $\tilde{z}_t = \tilde{x}_t - \int_0^t G_s \, d\xi_s$ be the unique solution of the SDE

$$\begin{cases} d\tilde{z}_t = \bar{b}(t, \tilde{z}_t, u_t) \, dt + \bar{\sigma}(t, \tilde{z}_t) \, d\widetilde{B}_t, \\ \tilde{z}_0 = \alpha \end{cases} \tag{4.7}$$

on the enlarged space $(\widetilde{\Omega}, \widetilde{F}, (\widetilde{F}_t)_{t \geqslant 0}, \widetilde{P}, \widetilde{B}_t)$, where $\bar{b}$ and $\bar{\sigma}$ are defined in Section 3.

Theorem 4.2 *(The Bouleau–Hirsch flow property).*

(1) *For $\widetilde{P}$-almost every w and for all $t \geqslant 0$, $\tilde{z}_t$ is in D^d.*
(2) *There exists an $\widetilde{F}_t$-adapted $GL_d(\mathbb{R})$-valued continuous process $(\widetilde{\Phi}_t)_{t \geqslant 0}$ such that for every $t \geqslant 0$*

$$\frac{\partial}{\partial x}\big(z_t^\alpha(w)\big) = \widetilde{\Phi}_t(\alpha, w) \quad dx\text{-a.e.},$$

where $\frac{\partial}{\partial x}$ denotes the derivative in the distribution sense.
(3) *The distributional derivative $\widetilde{\Phi}_t$ is the unique solution of the linear stochastic differential equation*

$$\begin{cases} d\widetilde{\Phi}(s, t) = \bar{b}_x(s, \tilde{z}_s, \tilde{u}_s).\widetilde{\Phi}(s, t) \, ds + \sum_{1 \leqslant j \leqslant d} \bar{\sigma}_x^j(s, \tilde{z}_s).\widetilde{\Phi}(s, t) \, d\widetilde{B}_s^j, \quad s \geqslant t, \\ \widetilde{\Phi}(t, t) = Id, \end{cases} \tag{4.8}$$

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 $\widetilde{P}$ by the map $\tilde{z}_t$ is absolutely continuous with respect to the Lebesgue measure.*

Now, consider the process y_t, $t \geqslant 0$, defined on the enlarged probability space $(\widetilde{\Omega}, \widetilde{F}, (\widetilde{F}_t)_{t \geqslant 0}, \widetilde{P}, \widetilde{B}_t)$ which is a solution of

$$\begin{cases} dy_t = b^n(t, y_t, u_t) \, dt + \sigma^n(t, y_t) \, d\widetilde{B}_t + G_t \, d\xi_t, \\ y_0 = \alpha. \end{cases} \tag{4.9}$$

Define the cost functional accordingly

$$J^n(u_t) = \widetilde{E}\left[\int_0^T f^n(t, y_t, u_t) \, dt + \int_0^T k_t \, d\xi_t + g^n(y_T) \right], \tag{4.10}$$

where b^n, σ^n, f^n and g^n be the regularized functions of b, σ, f and g defined in Section 3.

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

Lemma 4.1. *Let (x_t) and (y_t) be the solutions of (2.1) and (4.9) respectively, corresponding to an admissible control (u, ξ). Then*

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

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 4.1, there exists $(\delta_n) \equiv (2M_2.\epsilon_n)$ a sequence of positive real numbers converging to 0, such that

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

The functional J^n defined by (4.10) is 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

$$d_1'(u, v) = \widetilde{P} \otimes dt\big\{(w, t) \in \widetilde{\Omega} \times [0, T], \ v(w, t) \neq u(w, t)\big\},$$

$$d_2'(\xi, \eta) = \left(\widetilde{E}\left[\sup_{0 \leqslant t \leqslant T} |\xi_t - \eta_t|^2 \right]\right)^{\frac{1}{2}}.$$

Then by applying Ekeland's principle to J^n and $(\hat{u}, \hat{\xi})$ with $\lambda_n = \delta_n^{\frac{2}{3}}$, there exists an admissible control (u^n, ξ^n) such that

$$d'\big((\hat{u}, \hat{\xi}), (u^n, \xi^n)\big) \leqslant \delta_n^{\frac{2}{3}},$$

$$J_\delta^n(u^n, \xi^n) \leqslant J_\delta^n(\upsilon, \eta), \quad \text{for any } (\upsilon, \eta) \in U.$$

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

$$J_\delta^n(\upsilon, \eta) = J^n(\upsilon, \eta) + \delta_n^{\frac{1}{3}} . d'\big((\upsilon, \eta), (u^n, \xi^n)\big).$$

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

$$\begin{cases} dx_t^n = b^n(t, x_t^n, u_t^n)\,dt + \sigma^n(t, x_t^n)\,d\widetilde{B}_t + G_t\,d\xi_t^n, \\ x_0^n = \alpha. \end{cases} \tag{4.11}$$

The controlled process $dz_t^n = dx_t^n - G_t\,d\xi_t^n$ is then defined as the solution to the stochastic differential equation

$$\begin{cases} dz_t^n = \bar{b}^n(t, z_t^n, u_t^n)\,dt + \bar{\sigma}^n(t, z_t^n)\,d\widetilde{B}_t, \\ z_0^n = \alpha, \end{cases} \tag{4.12}$$

where $\bar{b}^n$ and $\bar{\sigma}^n$ are defined in Section 3. Let $\Phi^n(s, t)$ $(s \geqslant t)$ be the fundamental solution of the linear equation

$$\begin{cases} d\Phi^n(s, t) = b_x^n(s, x_s^n, u_s^n).\Phi^n(s, t)\,ds + \displaystyle\sum_{1 \leqslant j \leqslant d} \sigma_x^{j,n}(s, x_s^n).\Phi^n(s, t)\,d\widetilde{B}_s^j, \\ \Phi^n(t, t) = Id. \end{cases} \tag{4.13}$$

Proposition 4.1. *For each integer n, there exist an admissible control (u^n, ξ^n) and an $(\widetilde{F}_t)$-adapted process p_t^n given by*

$$p_t^n = \widetilde{E}\left[\int_t^T \Phi^{n,*}(s, t).f_x^n(s, x_s^n, u_s^n)\,ds + \Phi^{n,*}(T, t).g_x^n(x_T^n)/\widetilde{F}_t \right], \tag{4.14}$$

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

$$\widetilde{E}\big[H^n(t, x_t^n, \upsilon, p_t^n) - H^n(t, x_t^n, u_t^n, p_t^n) \big] \geqslant -\delta_n^{\frac{1}{3}} . M_1, \tag{4.15}$$

and

$$\widetilde{E} \int_t^T (k_t + G_t^* p_t^n)d(\eta - \xi^n)_t \geqslant -\delta_n^{\frac{1}{3}} . M_2, \tag{4.16}$$

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

$$H^n(t, x, u, p) = p.b^n(t, x, u) + f^n(t, x, u). \tag{4.17}$$

The proof goes as in Section 3.
The proof of Theorem 4.1 is based on the following lemma.

Lemma 4.2. *The following estimates hold*

(i) $\quad \displaystyle\lim_{n \to +\infty} \widetilde{E}\Big[\sup_{0 \leqslant t \leqslant T} |x_t^n - \hat{x}_t|^2 \Big] = 0;$ \hfill (4.18)

(ii) $\quad \displaystyle\lim_{n \to +\infty} \widetilde{E}\Big[\sup_{s \leqslant t \leqslant T} |\Phi^n(s, t) - \Phi(s, t)|^2 \Big] = 0;$ \hfill (4.19)

(iii) $\quad \displaystyle\lim_{n \to +\infty} \widetilde{E}\Big[\sup_{0 \leqslant t \leqslant T} |p_t^n - p_t|^2 \Big] = 0;$ \hfill (4.20)

(iv) $\quad \displaystyle\lim_{n \to +\infty} \widetilde{E}\big[\big| H^n(t, x_t^n, u_t^n, p_t^n) - H(t, \hat{x}_t, \hat{u}_t, p_t) \big| \big] = 0,$ \hfill (4.21)

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

Proof. (i) is proved as (3.19).

Let us prove (ii).

Using Burkholder–Davis–Gundy, Schwartz inequalities and Gronwall Lemma, we obtain

$$\widetilde{E}\Big[\sup_{t\leqslant s\leqslant T}\big|\Phi^n(s,t)-\Phi(s,t)\big|^2\Big]\leqslant C\widetilde{E}\Big[\sup_{t\leqslant s\leqslant T}\big|\Phi^n(s,t)\big|^4\Big]^{\frac{1}{2}}\Bigg\{\widetilde{E}\bigg[\int_0^T\big|b_x^n\big(t,x_t^n,\hat{u}_t\big)-b_x(t,\hat{x}_t,\hat{u}_t)\big|^4\,dt\bigg]^{\frac{1}{2}}$$

$$+\sum_{1\leqslant j\leqslant d}\widetilde{E}\bigg[\int_0^T\big|\sigma_x^{j,n}\big(t,x_t^n\big)-\sigma_x^j(t,\hat{x}_t)\big|^4\,dt\bigg]^{\frac{1}{2}}\Bigg\}.$$

Since the coefficients in the linear stochastic differential equation (4.13) are bounded, it is easy to see that $\widetilde{E}[\sup_{t\leqslant s\leqslant T}|\Phi^n(s,t)|^4]<+\infty$. To derive (4.19), it is sufficient to prove the following two assertions

$$\widetilde{E}\bigg[\int_0^T\big|b_x^n\big(t,x_t^n,\hat{u}_t\big)-b_x(t,\hat{x}_t,\hat{u}_t)\big|^4\,dt\bigg]\to 0\quad\text{as }n\to+\infty,$$

and

$$\widetilde{E}\bigg[\int_0^T\big|\sigma_x^{j,n}\big(t,x_t^n\big)-\sigma_x^j(t,\hat{x}_t)\big|^4\,dt\bigg]\to 0\quad\text{as }n\to+\infty,\text{ for }j=1,2,\ldots,d.$$

Let us prove the first limit. We have

$$\widetilde{E}\bigg[\int_0^T\big|b_x^n\big(t,x_t^n,u_t^n\big)-b_x(t,\hat{x}_t,\hat{u}_t)\big|^4\,dt\bigg]\leqslant C\big(I_1^n+I_2^n+I_3^n\big),$$

where

$$I_1^n=\widetilde{E}\bigg[\int_0^T\big|b_x^n\big(t,x_t^n,u_t^n\big)-b_x^n\big(t,x_t^n,\hat{u}_t\big)\big|^4\chi_{\{u^n\neq\hat{u}\}}(t)\,dt\bigg],$$

$$I_2^n=\widetilde{E}\bigg[\int_0^T\big|b_x^n\big(t,x_t^n,\hat{u}_t\big)-b_x\big(t,x_t^n,\hat{u}_t\big)\big|^4\,dt\bigg],$$

$$I_3^n=\widetilde{E}\bigg[\int_0^T\big|b_x\big(t,x_t^n,\hat{u}_t\big)-b_x(t,\hat{x}_t,\hat{u}_t)\big|^4\,dt\bigg].$$

According to the boundness of the derivative b_x^n by the Lipschitz constant and the fact that $d_1'(u^n,\hat{u})\to 0$ as $n\to+\infty$, we obtain $\lim_{n\to+\infty}I_1^n=0$.

Moreover, we have

$$I_2^n\leqslant\widetilde{E}\bigg[\int_0^T\sup_{a\in A_1}\big|\bar{b}_x^n\big(t,z_t^n,a\big)-\bar{b}_x\big(t,z_t^n,a\big)\big|^4\,dt\bigg]=\int_0^T\int_{\mathbb{R}^d}\sup_{a\in A_1}\big|\bar{b}_x^n(t,y,a)-\bar{b}_x(t,y,a)\big|^4\rho_t^n(y)\,dy\,dt,$$

where z_t^n denotes the unique solution of the SDE (4.12), corresponding to (u^n,ξ^n), and $\rho_t^n(y)$ its density with respect to the Lebesgue measure. Let us show

$$\lim_{n\to+\infty}\int_{\mathbb{R}^d}\sup_{a\in A_1}\big|\bar{b}_x^n(t,y,a)-\bar{b}_x(t,y,a)\big|^4\rho_t^n(y)\,dy\,dt=0.$$

For each $p>0$, we have $\widetilde{E}[\sup_{0\leqslant t\leqslant T}|z_t^n|^p]<+\infty$. Then, $\lim_{R\to\infty}\widetilde{P}(\sup_{0\leqslant t\leqslant T}|z_t^n|>R)=0$. So it is enough to show that for every $R>0$,

$$\lim_{n\to+\infty}\int_{B(0,R)}\sup_{a\in A_1}\big|\bar{b}_x^n(t,y,a)-\bar{b}_x(t,y,a)\big|^4\rho_t^n(y)\,dy=0.$$

According to Lemma 3.1, it is easy to see that

$$\sup_{a \in A_1} \left| \bar{b}_x^n(t, y, a) - \bar{b}_x(t, y, a) \right|^4 = \sup_{a \in A_1} \left| b_x^n \left(t, y + \int_0^T G_t \, d\xi_t^n, a \right) - b_x \left(t, y + \int_0^T G_t \, d\xi_t^n, a \right) \right|^4 \to 0 \quad dy\text{-a.e.,}$$

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 $\sup_{a \in A_1} |\bar{b}_x^n(t, y, a) - \bar{b}_x(t, y, a)|$ 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

$$\lim_{n \to +\infty} \int_{F^c} \sup_{a \in A_1} \left| \bar{b}_x^n(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \rho_t^n(y) \, dy \leqslant \lim_{n \to +\infty} \left(\sup_{y \in F^c} \sup_{a \in A_1} \left| \bar{b}_x^n(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \right) = 0.$$

Now, by using the boundness of the derivatives $\bar{b}_x^n$, $\bar{b}_x$ we have

$$\int_F \sup_{a \in A_1} \left| \bar{b}_x^n(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \rho_t^n(y) \, dy = \widetilde{E}\left[\sup_{a \in A_1} \left| \bar{b}_x^n(t, \hat{z}_t^n, a) - \bar{b}_x(t, \hat{z}_t^n, a) \right|^4 \chi_{\{\hat{z}_t^n \in F\}} \right] \leqslant 2M^4 \widetilde{P}\left(\hat{z}_t^n \in F \right).$$

According to (4.18), it is easy to see that $z_t^n = x_t^n - \int_0^t G_s \, d\xi_s^n$ 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

$$\lim_n \int_F \sup_{a \in A_1} \left| \bar{b}_x^n(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \rho_t^n(y) \, dy \leqslant 2M^4 \limsup \widetilde{P}\left(z_t^n \in F \right) \leqslant 2M^4 \widetilde{P}(\hat{z}_t \in F) = 2M^4 \int_F \rho_t(y) \, dy < \varepsilon,$$

where $\rho_t(y)$ denotes the density of $\hat{z}_t$ with respect to Lebesgue measure.

Now, since

$$\int_{B(0,R)} \sup_{a \in A_1} \left| \bar{b}_x^n(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \rho_t^n(y) \, dy = \int_F \sup_{a \in A_1} \left| \bar{b}_x^n(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \rho_t^n(y) \, dy$$

$$+ \int_{F^c} \sup_{a \in A_1} \left| \bar{b}_x^n(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \rho_t^n(y) \, dy,$$

we get $\lim_{n \to +\infty} I_2^n = 0$.

Let $k \geqslant 0$ be a fixed integer, then it holds that $I_3^n \leqslant C(J_1^k + J_2^k + J_3^k)$, where

$$J_1^k = \widetilde{E}\left[\int_0^T \left| b_x(t, x_t^n, \hat{u}_t) - b_x^k(t, x_t^n, \hat{u}_t) \right|^4 dt \right],$$

$$J_2^k = \widetilde{E}\left[\int_0^T \left| b_x^k(t, x_t^n, \hat{u}_t) - b_x^k(t, \hat{x}_t, \hat{u}_t) \right|^4 dt \right],$$

$$J_3^k = \widetilde{E}\left[\int_0^T \left| b_x^k(t, \hat{x}_t, \hat{u}_t) - b_x(t, \hat{x}_t, \hat{u}_t) \right|^4 dt \right].$$

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_T^n to $\hat{x}_T$ to deduce that $b_x^k(t, x_t^n, \hat{u}_t)$ 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 $\lim_{n \to +\infty} J_2^k = 0$.

$$J_3^k = \widetilde{E}\left[\int_0^T \sup_{a \in A_1} \left| \bar{b}_x^k(t, \hat{z}_t, a) - \bar{b}_x(t, \hat{z}_t, a) \right|^4 dt \right] = \int_0^T \int_{\mathbb{R}^d} \sup_{a \in A_1} \left| \bar{b}_x^k(t, y, a) - \bar{b}_x(t, y, a) \right|^4 \rho_t(y) \, dy \, dt.$$

$\bar{b}_x^k$, $\bar{b}_x$ being bounded, then by using the convergence of $\bar{b}_x^k$ to $\bar{b}_x$, 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 3.3. $\square$

Proof of Theorem 4.1. Use Proposition 4.1 and Lemma 4.2. $\square$

Acknowledgments

References

[1] S. Bahlali, A. Chala, The stochastic maximum principle in optimal control of singular diffusions with nonlinear coefficients, Random Oper. Stochastic Equations 13 (2005) 1–10.
[2] S. Bahlali, B. Djehiche, B. Mezerdi, The relaxed maximum principle in singular control of diffusions, SIAM J. Control Optim. 46 (2007) 427–444.
[3] K. Bahlali, B. Djehiche, B. Mezerdi, On the stochastic maximum principle in optimal control of degenerate diffusions with Lipschitz coefficients, Appl. Math. Optim. 56 (2007) 364–378.
[4] S. Bahlali, B. Mezerdi, A general stochastic maximum principle for singular control problems, Electron. J. Probab. 10 (2005) 988–1004.
[5] K. Bahlali, B. Mezerdi, Y. Ouknine, The maximum principle in optimal control of a diffusions with nonsmooth coefficients, Stoch. Stoch. Rep. 57 (1996) 303–316.
[6] V.E. Beněs, L.A. Shepp, H.S. Witsenhausen, Some solvable stochastic control problems, Stochastics 4 (1980) 39–83.
[7] A. Bensoussan, Lectures on Stochastic Control, Lecture Notes in Math., vol. 972, Springer, Berlin, 1983, 1–62 pp.
[8] J.M. Bismut, An introductory approach to duality in optimal stochastic control, SIAM Rev. 20 (1) (1978) 62–78.
[9] N. Bouleau, F. Hirsch, Sur la propriété du flot d'une équation différentielle stochastique, C. R. Acad. Sci. Paris 306 (1988) 421–424.
[10] A. Cadenillas, U.G. Haussmann, The stochastic maximum principle for a singular control problem, Stoch. Stoch. Rep. 49 (1994) 211–237.
[11] A. Cadenillas, I. Karatzas, The stochastic maximum principle for linear convex optimal control with random coefficients, SIAM J. Control Optim. 33 (2) (1995) 590–624.
[12] P.-L. Chow, J.-L. Menaldi, M. Robin, Additive control of stochastic linear systems with finite horizon, SIAM J. Control Optim. 23 (1985) 858–899.
[13] F.H. Clarke, Optimization and Nonsmooth Analysis, Wiley, New York, 1983.
[14] M.H.A. Davis, A. Norman, Portfolio selection with transaction costs, Math. Oper. Res. 15 (1990) 676–713.
[15] I. Ekeland, Non convex minimization problems, Bull. Amer. Math. Soc. (N.S.) 1 (1979) 443–474.
[16] H. Frankowska, The first order necessary conditions for nonsmooth variational and control problems, SIAM J. Control Optim. 22 (1) (1984) 1–12.
[17] U.G. Haussmann, W. Suo, Singular optimal stochastic controls I: Existence, SIAM J. Control Optim. 33 (1995) 916–936.
[18] U.G. Haussmann, W. Suo, Singular optimal stochastic controls II: Dynamic programming, SIAM J. Control Optim. 33 (1995) 937–959.
[19] I. Karatzas, S. Shreve, Connections between optimal stopping and stochastic control I: Monotone follower problem, SIAM J. Control Optim. 22 (1984) 856–877.
[20] N.V. Krylov, Controlled Diffusion Processes, Springer, Berlin, 1980.
[21] N.J. Kushner, Necessary conditions for continuous parameter stochastic optimization problems, SIAM J. Control Optim. 10 (1972) 550–565.
[22] B. Mezerdi, Necessary conditions for optimality for a diffusion with a nonsmooth drift, Stochastics 24 (1988) 305–326.
[23] S. Peng, A general stochastic maximum principle for optimal control problems, SIAM J. Control Optim. 28 (1990) 966–979.
[24] R.T. Rockafellar, Conjugate convex functions in optimal control problems and the calculus of variations, J. Math. Anal. Appl. 32 (1970) 174–222.
[25] S.E. Shreve, H.M. Soner, Optimal investment and consumption with transaction costs, Ann. Appl. Probab. 4 (3) (1994) 609–692.
[26] X.Y. Zhou, Stochastic near optimal controls: Necessary and sufficient conditions for near optimality, SIAM J. Control Optim. 36 (3) (1998) 929–947.