SIAM REVIEW
Vol. 20, No. 1, January 1978

AN INTRODUCTORY APPROACH TO DUALITY IN OPTIMAL STOCHASTIC CONTROL*

JEAN-MICHEL BISMUT†

Abstract. The purpose of this paper is to compare the results which have been recently obtained in optimal stochastic control. Various maximum principles are shown to derive from a general Pontryagin principle for Ito equations. Other applications of duality to optimal stochastic control are given.

Duality is an essential tool in optimization theory. Its use has been systematized by Rockafellar in a large variety of situations: in finite dimensional spaces and in infinite dimensional spaces in duality. Rockafellar has applied in [25] the instruments of convex analysis to deterministic control. Lions applies them in [22] to the control of partial differential equations. All of these approaches give maximum principles for the considered optimization problems which, in deterministic control of differential equations are equivalent to the Pontryagin principle. It is then quite natural to apply this approach to optimal stochastic control. Already Kushner, in [17], [18], [20], has introduced stochastic Lagrange multipliers. Rishel in [24] and Davis and Varaiya in [10] use the stochastic process theory to give necessary and sufficient conditions for a control to be an optimum. More recently, Haussmann used Neustadt's theory of extremals in [15] to derive necessary conditions of optimality for stochastic control.

Our purpose in this paper is to give an account of the duality methods which we have developed in our previous work [2]–[9] on optimal stochastic control and to compare it systematically to the results which have been given on the subject by Davis and Varaiya [10]. Kushner [17], [18], [20], and Wonham [28]. This paper does not directly treat the topics covered in the review of Fleming [14], which is concerned with the control of Markov processes. The methods used here apply to more general processes than the Markov processes, but do not give the strong results which are obtainable only for Markov processes.

In the first section we describe the general system which we consider to be defined by an Ito evolution equation, in which the control enters in stochastic open-loop form and where the state is observable.

In the second section we discuss various problems related to the observation problem and prove that it is legitimate to consider a fixed information pattern not depending on the control. In particular, we prove that the separation principle of Wonham [28], who describes a system where the state is not completely observed, and the control of Girsanov densities of Benes [1] can be put in the general framework of § 1.

In the third section we give the main results on duality which were obtained in [2] and [3], and discuss the generalized stochastic maximum principle for Ito equations, and its connections with Kushner's maximum principle [17]–[20], the dynamic programming equation, Davis and Varaiya's necessary and sufficient conditions [10]. In particular it is shown in a worked example that the dual state variable p is itself a solution of a dual stochastic control problem: this last result is

* Received by the editors October 22, 1975, and in revised form August 18, 1976.
† 191, rue d'Alésia, 75014 Paris, France.

classical in general duality theory. Finally a case is discussed where it is possible to define two distinct dual problems associated with a given stochastic control problem and the relations between the dual state variables is exhibited.

The proofs of the main results are not given. The reader is referred to the referenced papers.

1. A description of the system. In this section we briefly describe the general form of the stochastic system which we are going to control. More specifically, our equation will be an Ito stochastic differential equation of the form (1.1) below. The justification for considering an Ito equation and *not* a system in its Girsanov transformed form (see § 2.3 below for a definition of this term) as in [1], [10], [11], [12] will be given in the next sections.

We now introduce the standard probability framework commonly used in current probability literature. $(\Omega, \mathscr{F}, P)$ is a complete probability space, endowed with an increasing sequence of complete sub-σ-fields $\mathscr{F}_t$ which we assume to be right-continuous, i.e.,

$$\mathscr{F}_t = \bigcap_{t'>t} \mathscr{F}_{t'}.$$

$\mathscr{F}_t$ represents the information at time t.

w is an m-dimensional Brownian motion on $(\Omega, \mathscr{F}, P)$ adapted to $\mathscr{F}_t$. U is a compact metric space.

(L, b, σ) are functions defined on $\Omega \times R^+ \times R^d \times U$ with values respectively in R, R^d and $R^d \otimes R^m$, measurable in (ω, t, x) and continuous in u, and such that:

(a) b and σ are uniformly Lipschitz in x.

(b) $|L|$ grows at most $|x|^2$ when $|x| \to +\infty$, and this uniformly in (ω, t, u).

(c) b and σ grow at most as $|x|$ when $|x| \to +\infty$, and this uniformly in (ω, t, u).

(d) For each (x, u), these functions define adapted measurable processes, i.e., $(\omega, t) \to (L, b, \sigma)\ (\omega, t, x, u)$ is measurable, and for each t, $\omega \to (L, b, \sigma)(\omega, t, x, u)$ is $\mathscr{F}_t$-measurable.

u is now taken to be measurable and adapted to $\{\mathscr{F}_t\}_{t \geqq 0}$ with values in U. u will then be a stochastic process. We will not discuss here some of the technical points of measurability given in [3]. The only important thing is that we will *not* have a problem with the choice of the measurable version which we will take for u.

We consider then the equation:

$$\begin{aligned} dx &= b(\omega, t, x, u)\, dt + \sigma(\omega, t, x, u) \cdot dw, \\ x(0) &= x_0, \end{aligned} \tag{1.1}$$

where x_0 is $\mathscr{F}_0$ measurable and square integrable.

u is *not* taken here in feedback form, but is supposed directly to be adapted to $\{\mathscr{F}_t\}_{t \geqq 0}$($u$ is then a function of (ω, t).) Then, by standard approximation methods, equation (1.1) has one and only one solution adapted to $\{\mathscr{F}_t\}_{t \geqq 0}$, in the sense of Ito, i.e., for each ω, which also defines a trajectory of the Brownian motion w, there is one and only one trajectory of x corresponding to this ω.

We want to find u minimizing

$$E \int_0^T L(\omega, t, x_t, u_t)\, dt + E(\Phi(\omega, x_T)). \tag{1.2}$$

where Φ is a function $\mathcal{F}_T$-measurable in ω and continuous in x, growing with x at most as $|x|^2$.

2. The observation problem. The formulation given previously seems to be quite unsatisfactory. First, it deals with Ito equations, and this seems to be a severe limitation if we think that the Girsanov transformation (see § 2.3) allows us to control equations which do not have solutions in the Ito sense, but only in a weaker sense specified later in § 2.3.

A second aspect of the problem is the way in which information enters the system and the way the controller takes it into account. First we assume that the σ-fields $\mathcal{F}_t$ do *not* depend on the choice of the control, i.e., the information available at time t does not vary with the control u. This is clearly counterintuitive, as many examples of stochastic control have shown. Second, we take control in "stochastic open-loop" form. We follow Lindquist's terminology in [21], i.e., we do not allow u to depend explicitly on x, if only because, unless we put strict Lipschitz conditions on u, equation (1.1) would no longer make sense, or only in the Girsanov sense (see § 2.3), and this is a situation which we do *not* want.

2.1. A first justification. It is quite obvious mathematically that the variation of the σ-fields with u makes sense intuitively: the "observation" time t of a driver driving his car clearly depends on the past controls—here, the past speeds. However, a strong variation of these σ-fields makes the problem hopeless in the mathematical sense. The σ-fields are a very fragile mathematical construction and the technical problems would be untractable. We have shown in [7] that in nonclassical information patterns where information depends on control, an optimal control may well not exist. Moreover, the formulation given in § 1 has the advantage of defining the problem of optimal control as a problem of minimization of a functional on a *well-defined* set of measurable functions.

However, to prove that our definition of the problem makes sense, we will show that many classical problems of optimal stochastic control can be put in this form.

2.2. The separation principle. The separation principle, proved by Wonham in [28], for the standard linear regulator with a noise on the observation, gives the important result that the optimal control can be taken as a function of the best estimate of the state variable. We show here very simply that the model (1.1)–(1.2) with full observation applies also in this case.

We consider an equation:

$$dx = (Ax + Bu)\,dt + dw, \qquad x(0) = x_0, \tag{2.1}$$

where

- w is a d-dimensional white noise of covariance $Q_t\,dt$. We assume $\|Q_t\|$ to be bounded uniformly.
- x_0 is a Gaussian vector, independent of w'. $\hat{x}_0$ is the mean of x_0.

The observation process z is

$$dz = Cx\,dt + d\eta, \qquad z(0) = 0, \tag{2.2}$$

where z is in R^n, η is an n-dimensional Brownian motion, independent of (w, x_0).

We want to minimize

$$E\int_0^T K(t, x, u)\,dt$$

with u_t a function only of $(z_s; s \leqq t)$.

Then let (x_1, x_2) be defined by

$$\begin{aligned} &dx_1 = Ax_1\,dt + dw,\\ &x_1(0) = x_0,\\ &dx_2 = (Ax_2 + Bu)\,dt,\\ &x_2(0) = 0, \end{aligned} \tag{2.3}$$

and (z_1, z_2) be defined by

$$\begin{aligned} &dz_1 = Cx_1\,dt + d\eta,\\ &z_1(0) = 0,\\ &dz_2 = Cx_2\,dt,\\ &z_2(0) = 0. \end{aligned} \tag{2.4}$$

At time t, the controller knows the past values of z and the past controls. He knows then the past values of (z, z_2) and, because $z_1 = z - z_2$, he knows also the past values of z_1.

We take as *an assumption* that the only information the controller has about the system is the past values of z_1. This information does *not* depend on the control.

Let us write the Kalman–Bucy filtering equation for $\hat{x}_1$, which is the best estimate of x_1:

$$\begin{aligned} &d\hat{x}_1 = A\hat{x}_1\,dt + K_t\,dM_t,\\ &\hat{x}_1(0) = \hat{x}_0, \end{aligned} \tag{2.5}$$

where M_t is the innovation process

$$z_{1_t} - \int_0^t C\hat{x}_{1_s}\,ds$$

which is a Brownian motion (for this result, we refer to [16] or [19]).

All the processes being jointly Gaussian, $x_{1_t} - \hat{x}_{1_t}$ is independent of the σ-field generated by $(z_{1_s} | s \leqq t)$. Moreover, because we have:

$$\begin{aligned} &dz_1 = C\hat{x}_1\,dt + dM,\\ &z_1(0) = 0, \end{aligned} \tag{2.6}$$

the σ-fields generated by M and z_1 are identical.

Finally, if $\hat{x}$ is given by

$$\begin{aligned} &d\hat{x} = (A\hat{x} + Bu)\,dt + dM,\\ &\hat{x}(0) = \hat{x}_0, \end{aligned} \tag{2.7}$$

we can write

$$E\int_0^T K(t, x_t, u_t)\,dt = E\int_0^T K(t, \hat{x}_t + x_t - \hat{x}_t, u_t)\,dt. \tag{2.8}$$

$x_t - \hat{x}_t$ being independent of $\hat{x}_t$ and u_t, and not depending even implicitly on u, the criterion (2.8) can be written as

$$E\int_0^T L(t, \hat{x}_t, u_t)\,dt. \tag{2.9}$$

The system (2.7), (2.9) is precisely the type considered in § 1. The facts had been pointed out by Wonham in [28] under the name of separation principle. Let us insist on the fact that the reason why the separation principle works is essentially that the information does not depend on the choice of the control.

If we take u directly in feedback form relative to z, then we can use the Girsanov transformation, defined in § 2.3, to treat the problem as a stochastic open loop problem. The admissible controls are generally not the same in both cases.

2.3. The Girsanov transformation. We now come to the classical problem where weak diffusions are considered in the formalism of Benes [1].

We take an equation of type

$$\begin{aligned} dx &= b(t, x, u)\,dt + dw, \\ x(0) &= 0, \end{aligned} \tag{2.10}$$

where w is a d dimensional Brownian motion.

$\mathscr{C}$ is the space of continuous functions x defined on R^+ with values in R^d, such that $x(0)=0$. $\mathscr{F}_t$ is the σ-field $\mathscr{B}(x_s | s \leqq t \wedge T)$, to which the negligible sets of $\mathscr{B}(x_s | s \leqq T)$ for the Brownian measure are added.

b is a function defined on $[0, T] \times \mathscr{C} \times U$ which is continuous in u for a given (t, x) in $[0, T] \times \mathscr{C}$, and which defines, for $u \in U$, a measurable adapted process. We assume that b is uniformly bounded (we could accept that $|b|$ grows at most linearly as $|x_t|$).

u is taken here in adapted feedback form relative to x.

We want to minimize

$$E\int_0^T L(t, x, u)\,dt \tag{2.11}$$

where L verifies the same assumptions as b.

To simplify, we assume that L does not depend on u. We are then going to make a transformation of the problem which will make the control u, which is feedback relative to x, an open loop control relative to a new state variable $\tilde{x}$.

First of all we must make sense of (2.10). We will then use the results of Stroock and Varadhan [27, Part 6]. Let P be the Brownian measure on $\mathscr{C}$. The Brownian motion is then noted x.

u is taken directly in feedback form relative to the state variable x (i.e., u_t is adapted to the σ-field $\mathscr{F}_t$). Generally, (2.10) does not have a solution in the sense of Ito, i.e., it is not possible to construct x from a given Brownian motion w. However, (2.10) has a unique weak solution, i.e., there is a unique measure Q^{b_u} on $\mathscr{C}$, such that

$$x_t - \int_0^t b(s, x, u)\, ds$$

is a Brownian motion for Q^{b_u} (x is the generating element of $\mathscr{C}$). Q^{b_u} has a density $\tilde{x}_t$ relative to P on each σ-field $\mathscr{F}_t$, given by

$$\tilde{x}_t = \left.\frac{dQ^{b_u}}{dP}\right|_{\mathscr{F}_t} = \exp\left\{\int_0^{t\wedge T} \langle b(s, x, u), dx_s\rangle - \frac{1}{2}\int_0^{t\wedge T} |b(s, x, u)|^2\, ds\right\} \tag{2.12}$$

(see [27], Thm. 6.2.). $\tilde{x}_t$ is the Girsanov density of Q^{b_u} relative to P on $\mathscr{F}_t$. Moreover, we have, by Ito's formula [23, Thm. 4, p. 111]

$$\begin{aligned} d\tilde{x} &= \tilde{x}_t b(t, x, u)\, dx, \\ \tilde{x}(0) &= 1, \end{aligned} \tag{2.13}$$

and the criterion (2.11) can be written as

$$E^P \tilde{x}_T A_T \tag{2.14}$$

where A_T is the random variable $\int_0^T L(t, x)\, dt$.

Now (2.13) is a true Ito equation of type (1.1). $\tilde{x}$ becomes the true state variable. u is in feedback form relative to x, and is *stochastic open-loop* relative to $\tilde{x}$. Finally, the criterion is of type (1.2). We are precisely in the situation of § 1.

It is clear now that the distinction between stochastic open-loop and feedback controls is intuitively founded, but can be mathematically side-stepped in many "classical" problems.

3. Duality in optimal stochastic control: The maximum principle. We come back to the problem defined in § 1 by Ito's equation (1.1) and the criterion (1.2). In deterministic control there is, under stringent conditions, an adjoint state variable which is a solution of a differential equation with terminal condition. This last fact does not raise any special difficulty. In stochastic control, we want to introduce a dual variable. However, for a dual variable to make sense, it should be observable, i.e., adapted to $\{\mathscr{F}_t\}_{t\geqq 0}$. But, a terminal condition will again appear, and this fact makes the problem more difficult than it is in deterministic control.

In [2] and [3] we have introduced duality methods in optimal stochastic control.

3.1. The maximum principle. Under very stringent convexity assumptions the generalized maximum principle for Ito equations can be written in the following way, as given in [2] and [3].

- $\mathscr{H}$ is defined by

$$\mathscr{H}(\omega, t, x, u, p, H) = \langle p, b(\omega, t, x, u)\rangle + \langle H, \sigma(\omega, t, x, u)\rangle - L(\omega, t, x, u). \tag{3.1}$$

- The maximum conditions can be written as

$$
\begin{aligned}
&dx = \frac{\partial \mathcal{H}}{\partial p}\,dt + \frac{\partial \mathcal{H}}{dH}\cdot dw,\\
&x(0) = x_0,\\
&dp = -\frac{\partial \mathcal{H}}{\partial x}\,dt + \frac{\partial \mathcal{H}}{\partial \sigma}\cdot dw + dM,\\
&p_T = -\frac{\partial \Phi}{\partial x_T},\\
&\max_u \mathcal{H}(\omega, t, x, u, p, H),
\end{aligned}
\tag{3.2}
$$

where

- p is an *adapted* right-continuous process,
- H is an adapted measurable process such that $E\int_0^T |H_t|^2\,dt < +\infty$,
- M is a square integrable martingale[1] such that $Mw_1, \cdots, Mw_m$ are martingales.

The existence of (x, p) satisfying (3.2) is a sufficient condition for x to be optimal when $\tilde{L}$ defined by

$$\tilde{L}(\omega, t, x, y, H) = \inf_{\substack{\{f(\omega,t,x,u)=y\\ \sigma(\omega,t,x,u)=H\}}} L(\omega, t, x, u)$$

(where $\tilde{L}$ is $+\infty$ when there is no u such that $(f(\omega, t, x, u), \sigma(\omega, t, x, u)) = (y, H)$) is convex in (x, y, H). This follows from Example II-3, Theorem IV-2, and Theorem V-1 in [3]. The existence of (x, p) satisfying (3.2) for x to be optimal is necessary under conditions specified in [2].

3.2. Relation with Kushner's maximum principle [20]. If all the randomness comes from w, then we know by a result of Ito (see [23, p. 135]) that M in (3.2) is 0.

We assume here that f, σ, Φ have bounded derivatives in x, do not depend on ω, that $L = 0$ and that σ does not depend on u. The dual state p is then a solution of

$$
\begin{aligned}
&dp = -(\langle p, b_x(t, x, u)\rangle + \langle H, \sigma_x(t, x)\rangle)\,dt + H\cdot dw,\\
&p_T = -\frac{\partial \Phi}{\partial x_T},
\end{aligned}
\tag{3.3}
$$

with $E\int_0^T |H_t|^2\,dt < +\infty$.

Moreover, if u is optimal, $dP \otimes dt$ a.s. for any u' in U, we have

$$\langle p_t, b(t, x, u_t)\rangle \geqq \langle p_t, b(t, x, u')\rangle \tag{3.4}$$

We will now formally prove that conditions (3.3) and (3.4) are precisely the necessary conditions given by Kushner in [20]. A simplified version of Theorem 3 in [20] can be stated in the following way:

[1] A martingale M is an integrable stochastic process such that for $t' \geqq t$, $M_t = E^{\mathcal{F}_t} M_{t'}$. In this paper, M_0 will always be assumed to be 0.

For $0 \leqq \tau \leqq t$ let $\rho(\tau, t)$ be the stochastic matrix given by the stochastic differential equations

$$d\rho(t, \tau) = b_x(t, x, u)\rho(t, \tau)\, dt + \sigma_x \rho(t, \tau)\, dw, \qquad (3.5)$$
$$\rho(\tau, \tau) = I.$$

Then if q_t is the process

$$q_t = -\rho * (T, t)\frac{\partial \Phi}{\partial x}(x_T) \qquad (3.6)$$

a necessary condition for u to be optimal is that $dP \otimes dt$ a.e., for any u' in U:

$$\langle E^{\mathscr{F}_t} q_t, b(t, x_t, u_t)\rangle \geqq \langle E^{\mathscr{F}_t} q_t, b(t, x_t, u')\rangle. \qquad (3.7)$$

To prove the equivalence between (3.5)–(3.7) and the system (3.3)–(3.4), we will show that the process

$$p_t = E^{\mathscr{F}_t} q_t \qquad (3.8)$$

is a solution of (3.3) when the state space is one dimensional. This last restriction is useful to avoid matrix manipulations.

From (3.5) we have

$$\rho(T, 0) = \rho(T, t)\rho(t, 0) \qquad (3.9)$$

or, equivalently,

$$\rho(T, t) = \rho(T, 0)\rho^{-1}(t, 0)$$

(ρ can be proved to be >0). By Ito's formula [23, Thm. 4, p. 111] if $\rho_t = \rho(t, 0)$, we have

$$d\rho_t^{-1} = -\frac{1}{\rho^2} d\rho + \frac{1}{\rho^3}(\sigma_x)^2 \rho^2\, dt, \qquad (3.10)$$
$$\rho^{-1}(0) = I,$$

or, equivalently,

$$d\rho_t^{-1} = \rho^{-1}((\sigma_x^2 - b_x)\, dt - \sigma_x\, dw), \qquad (3.11)$$
$$\rho_t^{-1}(0) = I.$$

The process $p_t = E^{\mathscr{F}_t} q_t$ can then be written as

$$p_t = -\rho_t^{-1} E^{\mathscr{F}_t} \rho_T \frac{\partial \Phi}{\partial x}(x_T). \qquad (3.12)$$

From (3.5), ρ_T can be proved to be square integrable, because f_x and σ_x are bounded. $(\rho_T(\partial\Phi/\partial x))(x_T)$ is then square integrable. Let M_t be the square integrable martingale:

$$M_t = -E^{\mathscr{F}_t} \rho_T \frac{\partial \Phi}{\partial x}(x_T). \qquad (3.13)$$

Then by the quoted Ito representation result [23, p. 135], M may be written as

$$M_t = \int_0^t H'_s \cdot dw + a \tag{3.14}$$

H' being an adapted measurable process such that:

$$E \int_0^T |H'_s|^2 \, ds < +\infty. \tag{3.15}$$

Then, $p_t = \rho_t^{-1} M_t$. By again applying Ito's formula, we have:

$$dp = \rho^{-1} \, dM + M \, d\rho^{-1} - H' \rho^{-1} \sigma_x \, dt, \tag{3.16}$$

or equivalently,

$$dp = \rho^{-1} H' \, dw + M\rho^{-1}((\sigma_x^2 - b_x) \, dt - \sigma_x \, dw) - \rho^{-1} H' \sigma_x \, dt, \tag{3.17}$$

equation (3.18) is equivalent to

$$dp = -(pb_x + \rho^{-1} H' \sigma_x - p\sigma_x^2) \, dt + (\rho^{-1} H' - p\sigma_x) \, dw. \tag{3.18}$$

Then, if H is defined by

$$H = \rho^{-1} H' - p\sigma_x, \tag{3.19}$$

equation (3.18) is equivalent to

$$\begin{aligned} dp &= -(pb_x + H\sigma_x) \, dt + H \cdot dw, \\ p_T &= -\partial\Phi/\partial x_T. \end{aligned} \tag{3.20}$$

It can then be proved that $\sup_{0 \leqq t \leqq T} |\rho_t^{-1}|^2$ and $\int_0^T |H'_t|^2 \, dt$ are both square-integrable. H is then such that $E \int_0^T |H_t|^2 < +\infty$. (3.20) is precisely (3.3).

We have now proved that Kushner's maximum principle derives from the generalized maximum principle.

3.3. Relation with the dynamic programming equation. We now show the relationship between the maximum principle stated in (3.2) and the dynamic programming equation. The material presented here will be extremely formal. Its purpose is mainly to explain the deep relations between the two approaches.

We assume that all the randomness comes from w and that $\mathscr{F}_t$ is $\mathscr{B}(w_u | u \leqq t)$. We suppose that p_t can be written formally as $p(t, x_t)$. Then by Ito's formula we have, in the one dimensional case

$$\begin{aligned} &\frac{\partial p}{\partial t} + \frac{\partial p}{\partial x} b + \frac{1}{2} \frac{\partial^2 p}{\partial x^2} |\sigma|^2 = -\frac{\partial \mathscr{H}}{\partial x}, \\ &\frac{\partial p}{\partial x} \sigma = H. \end{aligned} \tag{3.21}$$

This equation is equivalent to

$$\frac{\partial p}{\partial t} + \frac{\partial p}{\partial x} b + \frac{1}{2} \frac{\partial^2 p}{\partial x^2} |\sigma|^2 = -\frac{\partial b}{\partial x} p - \frac{\partial \sigma}{\partial x} \sigma \frac{\partial p}{\partial x} + \frac{\partial L}{\partial x}. \tag{3.22}$$

Moreover, $\mathcal{H}$ is maximum in u, i.e., formally.

$$\frac{\partial b}{\partial u}p+\frac{\partial \sigma}{\partial u}\sigma\frac{\partial p}{\partial x}-\frac{\partial L}{\partial u}=0. \tag{3.23}$$

But, this is equivalent to saying that

$$\frac{\partial}{\partial u}\Big(pb(t,x,u)+\frac{1}{2}|\sigma|^2(t,x,u)\frac{\partial p}{\partial x}-L\Big)=0. \tag{3.24}$$

Let us now consider the dynamic programming payoff function:

$$V(t,x)=-\inf_u E_x\int_t^T L(s,x_s,u_s)\,ds. \tag{3.25}$$

Then, by [19, p. 105], we have

$$\frac{\partial V}{\partial t}=-\sup_u\Big(b(t,x,u)\frac{\partial V}{\partial x}+\frac{1}{2}|\sigma|^2(t,x,u)\frac{\partial^2 V}{\partial x^2}-L(t,x,u)\Big). \tag{3.26}$$

If we compare (3.22), (3.24), and (3.26), we see that, formally, p is equal to $\partial V/\partial x$. This is quite compatible with the idea that p is associated with the infinitesimal variations of the criterion when the state variable is perturbed by the addition of an adapted process. However, let us remember again that the control in § 1 is obtained in stochastic open-loop and that our problem cannot be directly compared with the problem that the dynamic programming equation is related to because a control in feedback form $u(t,x_t)$ is not necessarily in stochastic open-loop relative to u.

3.4. Relation with Davis and Varaiya maximum principle. We now compare (3.1)–(3.2) with the results of [10]. We again take problem (2.10), (2.11), which we change into (2.13) and (2.14). Let us apply (3.2). It is proved in [2] that the convexity assumptions which are necessary for rigorous application of the maximum principle are precisely verified for this case.

Then the maximum principle equations are

$$\begin{aligned}
&\mathcal{H}=\langle H,\tilde{x}b(t,x,u)\rangle,\\
&dp=-\langle H,b(t,x,u)\rangle\,dt+H\cdot dx,\\
&p_T=-\int_0^T L(t,x_t)\,dt,\\
&E\int_0^T |H_t|^2\,dt<+\infty,\\
&\max_u\langle H,b(t,x,u)\rangle,
\end{aligned} \tag{3.27}$$

(the term M is null by [23, p. 135]). Then, one can check that $p_t\tilde{x}_t$ is a martingale, i.e.,

$$p_t\tilde{x}_t=E^{\mathcal{F}_t}p_T\tilde{x}_T.$$

This implies

$$
(3.28)\qquad
\begin{aligned}
p_t &= -\frac{E^{\mathcal{F}_t}\tilde{x}_t \int_0^T L(s, x_s)\, ds}{\tilde{x}_t} \\
&= -\int_0^t L(s, x_s)\, ds - \frac{E^{\mathcal{F}_t}\tilde{x}_T \int_t^T L(s, x_s)\, ds}{\tilde{x}_t} \\
q_t &= -\left(p_t + \int_0^t L(s, x_s)\, ds\right)
\end{aligned}
$$

is then nothing else than

$$
(3.29)\qquad q_t = E_{u0}^{\mathcal{F}_t} \int_t^T L(s, x_s)\, ds = \inf_u E_u^{\mathcal{F}_t} \int_t^T L(s, x_s)\, ds,
$$

where the expectation $E_u^{\mathcal{F}_t}$ is taken with respect to the process defined by (2.10).

Moreover, by taking $H' = -H$, for the optimal u_0, we derive from (3.27);

$$
(3.30)\qquad q_t = q_0 - \int_0^t L(s, x)\, ds + \int_0^t \langle H', dx - b(s, x, u_0)\, ds\rangle
$$

and $\langle H', b(t, x, u)\rangle$ is minimum at the optimum.

But $dx - b(t, x, u_0)\, dt$ is a Brownian motion $\tilde{w}$ for the measure defined by $\tilde{x}\, dP$. Then

$$
(3.31)\qquad q_t = q_0 - \int_0^t L(s, x_s)\, ds + \int_0^t H'_s \cdot d\tilde{w}.
$$

We find then the Davis and Varaiya maximum principle [10], when L does not depend on u.

3.5. A worked example. We consider the equation

$$
(3.32)\qquad
\begin{aligned}
dx &= (Ax + Cu + f)\, dt + (Bx + Du + g) \cdot dw, \\
x(0) &= x_0,
\end{aligned}
$$

and the criterion

$$
(3.33)\qquad \frac{1}{2}\left(E \int_0^t (|Mx_t|^2 + \langle Nu_t, u_t\rangle)\, dt + E|M_1 x_T|^2\right)
$$

where A, C, B, D, M, N b, g are measurable adapted processes, where x_0 is $\mathcal{F}_0$ measurable and square integrable, M_1 is $\mathcal{F}_T$ measurable, N_t positive definite and uniformly elliptic, and where w is an m-dimensional Brownian motion. This example is completely treated in [8].

The growth conditions of § 1 are not verified, because u is taken to vary in a non compact vector space U. However, the maximum principle can be applied by an argument given in [8].

$\mathcal{H}$ is given by

$$
(3.34)\qquad \mathcal{H} = -\tfrac{1}{2}(|Mx|^2 + \langle Nu, u\rangle) + \langle p, Ax + Cu + f\rangle + \langle H, Bx + Du + g\rangle.
$$

Then the maximum principle equations are

$$
(3.35)\qquad \begin{aligned}
&dp = (M^*Mx - A^*p - B^*H)\,dt + H\cdot dw + dM,\\
&p_T = -M_1^*M_1x_T,\\
&E\int_0^T |H_t|^2\,dt < +\infty,\\
&Nu = C^*p + D^*H.
\end{aligned}
$$

The questions of existence and uniqueness of solutions for (3.35) are solved in [8].

If we assume formally that p_t can be written as $-P_tx_t - r_t$ we find in [8] that P and r solve formally the following system of equations

$$
(3.36)\qquad \begin{aligned}
&dP + \{PA + A^*P + B^*PB + B^*\mathscr{H} + \mathscr{H}B - (B^*PD + PC + \mathscr{H}D)\\
&\quad (N + D^*PD)^{-1}(D^*PB + C^*P + D^*\mathscr{H}) + M^*M\}\,dt - \mathscr{H}\,dw - d\mathscr{M}\\
&\quad = 0,\\
&P_T = M_1^*M_1,\\
&\mathscr{M}w_i\ (i = 1, \cdots, m) \text{ is a martingale.}
\end{aligned}
$$

$$
(3.37)\qquad \begin{aligned}
&dr = \{(PC + B^*PD + \mathscr{H}D)(N + D^*PD)^{-1}C^* - A^*\}r\,dt\\
&\quad + [\{(PC + B^*PD + \mathscr{H}D)(N + D^*PD)^{-1}D^* - B^*\}\\
&\quad (Pg + h) - Pf - \mathscr{H}g]\,dt + h\cdot dw + dM',\\
&r_T = 0,\\
&M'w_i\ (i = 1, \cdots, m) \text{ is a martingale.}
\end{aligned}
$$

Conditions for existence and uniqueness of the solution of these equations are given in [8].

3.6. A second worked example. We assume that in § 3.5 M and M_1 have an inverse, and that $\|M_t^{-1}\|$ stays bounded. All the other terms are defined as in § 3.5. We consider equation

$$
(3.38)\qquad \begin{aligned}
&dp = \dot{p}\,ds + H\cdot dw + dM,\\
&p(0) = p_0,
\end{aligned}
$$

where $E\int_0^T (|\dot{p}_s|^2 + |H_s|^2)\,ds < +\infty$, and where M is a square integrable martingale such that $Mw_1, \cdots, Mw_m$ are martingales, and we want to minimize

$$
(3.39)\qquad \begin{aligned}
&E\Big(\langle p_0, x_0\rangle + \frac{1}{2}\int_0^T (|M^{*-1}(\dot{p} + A^*p + B^*H)|^2\\
&\qquad + \langle N^{-1}(C^*p + D^*H), C^*p + D^*H\rangle)\,dt + |M_1^{*-1}p_T|^2\Big).
\end{aligned}
$$

The control variables are then $\dot{p}$, H, *and* M.

The growth conditions of § 1 are not verified, because $\dot{p}$ and H vary in noncompact spaces. Moreover there is a new term M appearing in (3.38). It is proved in [2] and [3] that the maximum principle is applicable, except that there is no term M appearing in the dual equation.

Let us use equation (3.2) where x is the dual state variable, $\dot{p}$ and H are the control variables. $\mathscr{H}$ is then defined by

$$\begin{aligned}\mathscr{H} = -\tfrac{1}{2}(|M^{*-1}(\dot{p}+A^*p+B^*H)|^2 + \langle N^{-1}(C^*p+D^*H), C^*p+D^*H\rangle)\\ +\langle \dot{p}, x\rangle + \langle H', H\rangle\end{aligned} \tag{3.40}$$

If we maximize $\mathscr{H}$ in $\dot{p}$, we have

$$\dot{p} = M^*Mx - A^*p - B^*H. \tag{3.41}$$

If we maximize $\mathscr{H}$ in H we get

$$H' = BM^{-1}M^{*-1}(\dot{p} + A^*p + B^*H) + DN^{-1}(C^*p + D^*H), \tag{3.42}$$

and this may be written as

$$H' = Bx + DN^{-1}(C^*p + D^*H). \tag{3.43}$$

Finally, x will be the solution of

$$\begin{aligned} &dx = (Ax + CN^{-1}(C^*p + D^*H))\,dt + (Bx + DN^{-1}(C^*p + D^*H))\,dw,\\ &x(0) = x_0,\\ &x_T = -M_1^{-1}M_1^{*-1}p_T,\end{aligned} \tag{3.44}$$

(as previously indicated, there is now no term M in (3.44)). System (3.38), (3.41), (3.44) is nothing else than (3.32) and (3.35) (with $f = 0$ and $g = 0$).

We have found that, as in classical duality theory, the dual state p of the problem of § 3.5 is itself a solution of an optimization problem whose dual state is x. This point has been completely developed in [3].

3.7. Duality and dualities. We take again the example of § 2.3, but we now assume that L depends on the control variable also. In [1], Benes changes the initial formulation of the problem. If η_t is a one dimensional Brownian motion independent of z, criterion (2.11) may be written as:

$$E\left(\int_0^T L(t, x, u)\,dt + \eta_T\right) \tag{3.45}$$

because η_T has a null mean.

Benes then introduces a new state variable x' given by

$$\begin{aligned} &dx' = L(t, x, u)\,dt + d\eta,\\ &x'(0) = 0.\end{aligned} \tag{3.46}$$

The control problem is then equivalent to the minimization of

$$Ex'_T. \tag{3.47}$$

By applying the Girsanov transformation on the new state (x, x') variable, we come back to a problem of the kind treated in § 2.3. However, we have introduced a new random term η_t and the strategies are now mixed: they are not feedback strategies relative to x but to the enlarged state (x, x'). However, we side-step this difficulty here.

We will consider a special case of the system considered in § 3.5:

$$
\begin{aligned}
&dx = (Ax + Cu)\,dt + dw,\\
&x(0) = x_0,
\end{aligned}
\tag{3.48}
$$

where x varies in m-dimensional vector space, w is an m-dimensional Brownian motion, and A, C, x_0 are constant.

The criterion is given by

$$
\frac{1}{2}E\int_0^T (|Mx_t|^2 + \langle Nu_t, u_t\rangle)\,dt
\tag{3.49}
$$

where M and N are constant.

We apply (3.35):

$$
\begin{aligned}
&dp = (M^*Mx - A^*p)\,dt + H\cdot dw,\\
&p_T = 0,\\
&Nu = C^*p.
\end{aligned}
\tag{3.50}
$$

In this case we find p in the form $-P_tx_t$ where P_t is a solution of a classical Riccati equation, and u_t is $-N^{-1}C^*P_tx_t$.

However, we may look at (3.48) and (3.49) as a special case of (2.10), (2.11), (3.46) and (3.47), by eventually randomizing the strategy. We will then use a second duality method for the standard linear regulator problem.

u is now taken in feedback form relative to (x, x'). Generally (2.10)–(3.46) does *not* have solutions in the Ito sense, but only in the Girsanov sense. The class of admissible controls is then completely changed: if a control u is adapted to (w, η) it is not necessarily adapted to (x, x') and conversely, if a control u is adapted to (x, x'), it is not necessarily adapted to (w, η). However, if u is a linear function of x, (3.48) has a solution in the sense of Ito, and u is also adapted to w. The linear feedback is, in this case, also a stochastic open-loop feedback relative to w. This fact will ensure that the optimal u found previously is also optimal in the class of feedback controls. Let us prove this fact.

The new state variable x' as in (3.46) is defined by:

$$
\begin{aligned}
&dx' = \tfrac{1}{2}(|Mx|^2 + \langle Nu, u\rangle)\,dt + d\eta,\\
&x'(0) = 0.
\end{aligned}
\tag{3.51}
$$

We take u in feedback form relative to (x, x'). To ensure the existence of the Girsanov density, we suppose that $u(x, x', t)$ is such that there is k and $k' \geqq 0$ with

$$
|u(x, x', t)| \leqq k|x_t| + k'.
\tag{3.52}
$$

Let $\tilde{x}$ be the solution of

$$
\begin{aligned}
&d\tilde{x} = \tilde{x}((Ax + Cu)\,dx + \tfrac{1}{2}(|Mx|^2 + \langle Nu, u\rangle)\,dx'),\\
&\tilde{x}(0) = 1,
\end{aligned}
\tag{3.53}
$$

where (x, x') is a Brownian motion. Then it is possible to prove that $E(\tilde{x}_t)=1$ for $t \geqq 0$ and that $\tilde{x}_t$ is the Girsanov density associated to (3.48)–(3.51) (the only difficulty is that here $|Mx|^2$ and $\langle Nu, u\rangle$ may grow as $|x|^2$ and not as $|x|$, as is usually required).

We then minimize

$$E(\tilde{x}_T x'_T). \tag{3.54}$$

Let us apply the maximum principle to (3.53)–(3.54) where u is restricted to verify (3.52), for a given (k, k') with

$$k > \sup_{0 \leqq t \leqq T} |N^{-1}C^*P_t|$$

and $k' > 0$. It is proved in [2] that for systems (3.52), (3.53), (3.54), the maximum principle is sufficient to ensure the optimality of a given control, although the growth conditions of § 1 are not verified.

Let $\mathcal{H}$ be the Hamiltonian

$$\mathcal{H} = \left(\langle H_1, Ax + Cu\rangle + \frac{H_2}{2}(|Mx|^2 + \langle Nu, u\rangle)\right)\tilde{x}. \tag{3.55}$$

Then, for u to be optimal, it is sufficient that u verifies (3.52) and that

$$\begin{aligned}
&H_2 \leqq 0,\\
&C^*H_1 + H_2 Nu = 0,\\
&d\tilde{p} = -\left(H_1(Ax + Cu) + \frac{H_2}{2}(|Mx|^2 + \langle Nu, u\rangle)\right) dt + H_1 \cdot dx + H_2 \cdot dx',\\
&\tilde{p}_T = -x'_T,\\
&|\tilde{p}_t| \leqq \lambda |x_t|^2 + \mu |x'_t| + \nu,\\
&E\int_0^T (|H_{1,t}|^2 + |H_{2,t}|^2)\, dt < +\infty,
\end{aligned} \tag{3.56}$$

(the growth condition on $\tilde{p}_t$ ensures the uniform integrability of $\tilde{p}_t\tilde{x}_t$ and the applicability of the maximum principle, although the growth conditions of § 1 are not verified). Let q_t be the process

$$q_t = -\frac{1}{2}E_u^{\mathcal{F}_t}\int_t^T (|Mx_t|^2 + \langle Nu_t, u_t\rangle)\, dt, \tag{3.57}$$

where u_t is equal to $-N^{-1}C^*P_tx_t$, and where the conditional expectation is taken relative to the process defined by (3.48). Then it is classical that

$$q_t = -\tfrac{1}{2}\langle P_t x_t, x_t\rangle. \tag{3.58}$$

We prove now that $\tilde{p}_t = q_t - x'_t$ is a solution of (3.56) with $H_1 = -Px$ and $H_2 = -1$. By (3.57),

$$q_t - \frac{1}{2}\int_0^t (|Mx|^2 + \langle Nu, u\rangle)\, dt$$

is a martingale relative to the process (3.48). By Ito's formula, and (3.58) we have

$$q_t = q_0 + \frac{1}{2}\int_0^t (|Mx_s|^2 + \langle Nu_s, u_s\rangle)\,ds - \int_0^t Px \cdot dw. \tag{3.59}$$

By (3.48) we also have

$$q_t = q_0 + \frac{1}{2}\int_0^t (|Mx_t|^2 + \langle Nu_t, u_t\rangle\,dt + \int_0^t Px \cdot (Ax + Cu)\,dt - \int_0^t Px\,dx. \tag{3.60}$$

It is then easily checked that $\tilde{p}_t$ satisfies the required equations. Moreover, we have

$$C^*H_1 + H_2Nu = 0. \tag{3.61}$$

u is then optimal among the feedback controls such that (3.52) holds for k large enough. (We have also proved that the optimal control does not depend on k!) This discussion underlines a simple example of some of the most difficult aspects of optimal stochastic control.

3.8. The term *M*. We have seen in (3.1) that, in general, there is a term M which appears in the equation for p. When the only randomness comes from w, M can be taken to be 0. However, in the general case there may be other random elements coming into the system: the final time T may be random, the matrices in (3.32) may change at random times, etc. At each of these changes, there is, then, a reevaluation of predictions, which is registered by M. For the study of M, we refer to [2].

4. Conclusion. This paper gives only a brief review of the implications of duality of optimal stochastic control. The previous methods are readily extended to jump processes. This follows basically from the fact that it is the quadratic variation $\langle w, w\rangle$ of w which explains the working of the maximum principle more than the other properties of w ($\langle w, w\rangle$ is a predictable, right continuous increasing processs such that $w_t^2 - \langle w, w\rangle_t$ is a martingale.)

But $\langle w, w\rangle_t = t$ and this is also true for a Poisson martingale. Then, the functional framework will be exactly the same.

A basic difficulty in the formulation of problems of optimal stochastic control is to find the control in pure feedback form, i.e., the control must depend on the present value of the state and *not* on all the past states. This requires use of potential theory methods which fit a very large class of problems of control of diffusion processes, jump processes, jumping diffusions, etc. These results are given in [4] and [5].

Game situations are studied in [5], [6], [10], [12], [13]. An economic interpretation of the maximum principle has been given in [9].

Acknowledgment. The author is indebted to two referees for very helpful comments and suggestions.

REFERENCES

[1] E. BENES, *Existence of optimal stochastic control laws*, SIAM J. Control, 9, (1971), pp. 446–472.
[2] J. M. BISMUT, *Analyse convexe et probabilités*, Thèse, Faculté des Sciences de Paris, Paris, 1973.

[3] ———, *Conjugate convex functions in optimal stochastic control*, J. of Math. Anal. Appl., 44 (1973), pp. 384–404.

[4] ———, *Théorie probabiliste du contrôle des diffusions*, Mem. Amer. Math. Soc., 4 (1976), pp. 1–130.

[5] ———, *Control of jump processes and applications*, Bul. Soc. Math. France, to appear.

[6] ———, *Probability theory methods in zero-sum stochastic games*, SIAM J. Control, to appear.

[7] ———. *An example of interaction between information and control*, IEEE Trans. Automatic Control, 18 (1973), pp. 518–522.

[8] ———, *Linear quadratic optimal stochastic control with random coefficients*, SIAM J. Control, 14 (1976), pp. 419–444.

[9] ———, *Growth and optimal intertemporal allocation of risks*, J. Econom. Theory, 10 (1975), pp. 239–257.

[10] M. H. A. DAVIS AND P. P. VARAIYA, *Dynamic programming conditions for partially observable stochastic systems*, SIAM J. Control, 11 (1973), pp. 226–261.

[11] ———, *On the existence of optimal stochastic policies in stochastic control*, Ibid., 11 (1973), pp. 587–594.

[12] T. DUNCAN AND P. P. VARAIYA, *On the existence of a stochastic control system*, Ibid., 9 (1971), pp. 354–371.

[13] R. ELLIOTT, *The existence of value in stochastic differential games*, Ibid., 14 (1976), pp. 85–94.

[14] W. H. FLEMING, *Optimal continuous parameter stochastic control*, this Review, 11 (1969), pp. 470–509.

[15] U. G. HAUSSMANN, *General necessary conditions for optimal control of stochastic systems*, Mathematical Programming Study 6, North-Holland, Amsterdam, 1976, pp. 30–48.

[16] T. KAILATH, *A general likelihood-ratio formula for random signals in Gaussian noise*, IEEE Trans. Information Theory, 15 (1969), pp. 350–361.

[17] H. J. KUSHNER, *On the stochastic maximum principle: Fixed time of control*, J. Math. Anal. Appl., 11 (1965), pp. 78–92.

[18] ———, *On the stochastic maximum principle with average constraints*, J. Math. Anal. Appl., 12 (1965), pp. 13–26.

[19] ———, *Stochastic Stability and Control*, Academic Press, New York, 1967.

[20] ———, *Necessary conditions for continuous parameter stochastic optimization problems*, SIAM J. Control, 10 (1972), pp. 550–565.

[21] A. LINDQUIST, *On feedback control of linear stochastic systems*, Ibid., 11 (1973), pp. 323–343.

[22] J. L. LIONS, *Contrôle optimal des systèmes gouvernés par des équations aux dérivées partielles*, Dunod, Paris, 1968; English transl., Springer-Verlag, 1971.

[23] P. A. MEYER, *Intégrales stochastiques I, II, III*, Séminaire de Probabilités No. 1, Lecture Notes in Mathematics, No. 39, Springer-Verlag, Berlin-Heidelberg-New York, 1967, pp. 71–141.

[24] R. RISHEL, *Necessary and sufficient dynamic programming conditions for continuous time stochastic optimal control*, SIAM J. Control, 8 (1970), pp. 559–571.

[25] R. T. ROCKAFELLAR, *Conjugate convex functions in optimal control and the calculus of variations*, J. Math. Anal. Appl., 32, (1970), pp. 174–222.

[26] ———, *Convex Analysis*, Princeton Univ. Press, Princeton, NJ, 1970.

[27] D. W. STROOCK AND S. R. S. VARADHAN, *Diffusion process with continuous coefficients*, Comm. Pure Appl. Math., 22 (1969), pp. 345–400, 479–530.

[28] W. M. WONHAM *On the separation theorem of stochastic control*, SIAM J. Control, 6 (1968), pp. 312–326.