APPLICATIONES MATHEMATICAE
27,4 (2000), pp. 455–464

V. S. BORKAR (Mumbai)

THE VALUE FUNCTION IN ERGODIC CONTROL OF DIFFUSION PROCESSES WITH PARTIAL OBSERVATIONS II

Abstract. The problem of minimizing the ergodic or time-averaged cost for a controlled diffusion with partial observations can be recast as an equivalent control problem for the associated nonlinear filter. In analogy with the completely observed case, one may seek the value function for this problem as the vanishing discount limit of value functions for the associated discounted cost problems. This passage is justified here for the scalar case under a stability hypothesis, leading in particular to a "martingale" formulation of the dynamic programming principle.

1. Introduction. The usual approach to control of partially observed diffusions is via the equivalent "separated" control problem for the associated nonlinear filter. This approach has proved quite successful for the finite horizon and infinite horizon discounted costs. (See, e.g., Borkar (1989), Chapter V.) For the average cost or "ergodic" control problem, however, a completely satisfactory treatment is still lacking. While the existence of optimal controls in appropriate classes of controls is known in many cases (see, e.g., Bhatt and Borkar (1996)), the characterization of optimality via dynamic programming has not yet been fully developed. Limited results are available in Bhatt and Borkar (1996) which takes a convex duality approach, and in Borkar (1999) where the vanishing discount limit is justified for the limited class of the so-called "asymptotically flat" diffusions of Basak, Borkar and Ghosh (1997). The aim here is to present another special case where this limit can be justified, that of stable scalar diffusions.

2000 *Mathematics Subject Classification*: Primary 93E20.

Key words and phrases: ergodic control, partial observations, scalar diffusions, value function, vanishing discount limit.

Research partially supported by grant No. DST/MS/III-045/96 from the Department of Science and Technology, Government of India.

Specifically, consider the scalar diffusion $X(\cdot)$ and an associated observation process $Y(\cdot) = [Y_1(\cdot), \ldots, Y_m(\cdot)]^{\mathrm{T}}$ described by the stochastic differential equations

$$X(t) = X_0 + \int_0^t m(X(s), u(s))\, ds + \int_0^t \sigma(X(s))\, dW(s), \tag{1.1}$$

$$Y(t) = \int_0^t h(X(s))\, ds + W'(t), \tag{1.2}$$

for $t \geq 0$. Here,

(i) $m(\cdot,\cdot) : \mathbb{R} \times U \to \mathbb{R}$, for a prescribed compact metric space U, is a bounded continuous map, Lipschitz in its first argument uniformly w.r.t. the second,

(ii) $\sigma(\cdot) : \mathbb{R} \to \mathbb{R}^+$ is bounded Lipschitz and uniformly bounded away from zero,

(iii) $h(\cdot) : \mathbb{R} \to \mathbb{R}^m$ is bounded twice continuously differentiable with bounded first and second derivatives,

(iv) X_0 is a random variable with prescribed law $\pi_0 \in \mathcal{P}(R)$ (here and later, for a Polish space S, $\mathcal{P}(S)$:= the Polish space of probability measures on S with the Prokhorov topology),

(v) $W(\cdot), W'(\cdot)$ are resp. one and m-dimensional standard Brownian motions and $(X_0, W(\cdot), W'(\cdot))$ are independent,

(vi) $u(\cdot) : \mathbb{R}^+ \to U$ is a "control" process with measurable paths, adapted to $\{\mathcal{G}_t\}$:= the natural filtration of $Y(\cdot)$. Call such $u(\cdot)$ "strict sense admissible" controls. (More generally, $u(\cdot)$ is said to be "admissible" if for $t \geq s$, $W(t) - W(s)$ is independent of $(u(y), W(y), y \leq s, X_0)$.)

Given a bounded continuous "running cost" $k : \mathbb{R} \times U \to \mathbb{R}$, the ergodic control problem under partial observations seeks to minimize over all strict sense admissible controls the "ergodic" or average cost

$$\limsup_{t\to\infty} \frac{1}{t} \int_0^t E[k(X(s), u(s))]\, ds. \tag{1.3}$$

We consider the weak formulation of this problem described in Chapter I of Borkar (1989) and the relaxed control framework. The latter in particular implies the following: We suppose that $U = \mathcal{P}(Q)$ for a compact metric space Q and that there exist bounded continuous $\overline{m} : \mathbb{R} \times Q \to \mathbb{R}$, Lipschitz in the first argument uniformly w.r.t. the second, and bounded continuous $\overline{k} : \mathbb{R} \times Q \to \mathbb{R}$, such that

$$m(x, u) = \int \overline{m}(x, y)\, u(dy), \qquad k(x, u) = \int \overline{k}(x, y)\, u(dy) \qquad \forall x, u.$$

See Borkar (1989), Chapter I, for background.

The "stability" assumption is the following: Say $u(\cdot)$ is a "Markov control" if $u(t) = v(X(t))$ for all t and a measurable $v : \mathbb{R} \to U$. (This is *not* strict sense admissible in general.) Equation (1.1) has a unique strong solution $X(\cdot)$ under any Markov control (see, e.g., Borkar (1989), pp. 10–12). We assume that any such $X(\cdot)$ is stable, i.e., positive recurrent (Bhattacharya (1981)), under all Markov $u(\cdot) \sim v(X(\cdot))$. By abuse of notation, we may refer to the map $v(\cdot)$ itself as the Markov control.

Consequences of stability are given in Section 3, following a description of the separated control problem in the next section. The final section establishes the vanishing discount limit, leading to a martingale formulation of the dynamic programming principle in the spirit of Striebel (1984).

2. The separated control problem. Let $\{\mathcal{F}_t\}$:= the natural filtration of $(u(\cdot), Y(\cdot))$ and $\{\widehat{\mathcal{F}}_t\}$:= the natural filtration of $(u(\cdot)$, $Y(\cdot)$, $X(\cdot)$, $W(\cdot)$, $W'(\cdot))$. Let $\pi_t \in \mathcal{P}(\mathbb{R})$ be the regular conditional law of $X(t)$ given $\mathcal{F}_t$. Let $\pi_t(f) = \int f \, d\pi_t$ for $f : \mathbb{R} \to \mathbb{R}$ measurable (when defined) and

$$L_u f(x) = \frac{1}{2}\sigma^2(x)\frac{d^2 f}{dx^2}(x) + m(x,u)\frac{df}{dx}(x), \qquad x \in \mathbb{R},\ u \in U,$$

for $f \in C^2(\mathbb{R})$. The evolution of $\{\pi_t\}$ is given by the nonlinear filter (see, e.g., Borkar (1989), Section V.1)

$$\begin{aligned}(2.1) \qquad \pi_t(f) = \pi_0(f) &+ \int_0^t \pi_s(L_{u(s)} f)\, ds \\ &+ \int_0^t \langle \pi_s(hf) - \pi_s(f)\pi_s(h), d\widehat{Y}(s)\rangle, \qquad t \geq 0,\end{aligned}$$

where the "innovations process" $\widehat{Y}(t) := Y(t) - \int_0^t \pi_s(h)\, ds$, $t \geq 0$, is a standard Brownian motion in $\mathbb{R}^m$. Rewrite (1.3) as

$$(2.2) \qquad \limsup_{t\to\infty} \frac{1}{t}\int_0^t E[\pi_s(k(\cdot, u(s)))]\, ds.$$

The "separated" control problem equivalent to the one above is to control $\{\pi_t\}$ given by (2.1) over strict sense admissible $u(\cdot)$, so as to minimize the cost (2.2). For well-posedness issues concerning (2.1), see Borkar (1989), Section V.1. (Assumption (iii) above plays a role here. See Fleming and Pardoux (1982) for weaker conditions.)

Following Fleming and Pardoux (1982), we enlarge the class of controls under consideration as follows: Let $(\Omega, \mathcal{F}, P)$ denote the underlying probability space. Without loss of generality, let $\mathcal{F} = \bigvee_{t\geq 0} \widehat{\mathcal{F}}_t$. Define on $(\Omega, \mathcal{F})$ another probability measure P_0 as follows. If $P_{0t}, \overline{P}_t$ are the restrictions to

$(\Omega, \widehat{\mathcal{F}}_t)$ of P_0, P_t respectively, then

$$\frac{dP_t}{dP_{0t}} = \exp\left(\int_0^t \langle h(X(s)), dY(s)\rangle - \frac{1}{2}\int_0^t \|h(X(s))\|^2\, ds\right), \qquad t \geq 0.$$

Under P_0, $Y(\cdot)$ is a standard Brownian motion in $\mathbb{R}^m$, independent of X_0, $W(\cdot)$. Call $u(\cdot)$ "wide sense admissible" if under P_0, for $t \geq s$, $Y(t) - Y(s)$ is independent of $\{X_0, W(\cdot), u(y),\ Y(y),\ y \leq s\}$. (Note that this includes strict sense admissible controls.) The problem now is to minimize (2.2) over all such $u(\cdot)$.

To summarize, we consider the "separated control problem" defined by:

• the $\mathcal{P}(\mathbb{R})$-valued controlled Markov process $\{\pi_t\}$, whose evolution is described by (2.1),

• U-valued control process $u(\cdot)$ assumed to be wide sense admissible in the sense defined above,

• the objective of minimizing the associated "ergodic" cost defined by (2.2) over all such $u(\cdot)$.

Note that we are considering the so-called weak formulation of the control problem (Borkar (1989), Chapter I), i.e., a "solution" for the above control system is a pair of processes $\{\pi_t, u(t)\}$ satisfying the foregoing on some probability space. Call it a "stationary pair" if they form a jointly stationary process in $\mathcal{P}(\mathbb{R}) \times U$ and an optimal stationary pair if the corresponding ergodic cost (2.2) (wherein the *lim inf* will perforce be a *lim*) is the least attainable.

3. Consequences of stability. Let $(X(\cdot), u(\cdot))$ be as in (2.1). Define Markov controls $v_{\rm m}, v_{\rm M}$ such that $v_{\rm m}(x) \in \operatorname{Argmin}(m(x,\cdot))$, $v_{\rm M}(x) \in \operatorname{Argmax}(m(x,\cdot))$ for all x. This is possible by a standard measurable selection theorem (see, e.g., Borkar (1989), p. 20). Let $X_{\rm m}(\cdot), X_{\rm M}(\cdot)$ be corresponding solutions to (1.1) on the same probability space as $X(\cdot)$, with the same $W(\cdot)$ and initial condition X_0. By the comparison theorem of Ikeda and Watanabe (1981), pp. 352–353,

$$X_{\rm m}(t) \leq X(t) \leq X_{\rm M}(t) \quad \forall t \text{ a.s.} \tag{3.1}$$

Thus for any $y > 0$,

$$P(|X(t)| \geq y) \leq P(|X_{\rm m}(t)| \geq y) + P(|X_{\rm M}(t)| \geq y).$$

Since $X_{\rm m}(\cdot), X_{\rm M}(\cdot)$ are stable, this implies tightness of the laws of $X(t)$, $t \geq 0$, and therefore of the laws of π_t, $t \geq 0$, by Lemma 3.6, pp. 126–127, of Borkar (1989). It then follows as in Lemmas 3.1, 3.2 of Bhatt and Borkar (1996) that an optimal stationary pair $\{\pi_t, u(t)\}$ exists.

Now let $\tau(x) = \inf\{t \geq 0 : X(t) = x\}$. Define $\tau_{\rm m}(x), \tau_{\rm M}(x)$ analogously, with $X_{\rm m}(\cdot)$, resp. $X_{\rm M}(\cdot)$, in place of $X(\cdot)$. By (3.1), $E_x[\tau(0)] \leq E_x[\tau_{\rm M}(0)]$

(resp., $E_x[\tau_{\mathrm{m}}(0)]$) for $x \geq 0$ (resp., ≤ 0). Since $X_{\mathrm{m}}(\cdot), X_{\mathrm{M}}(\cdot)$ are stable, the right-hand side is bounded and so is $\phi(x) := \sup_{u(\cdot)} E_x[\tau(0)]$ for all x. Let $g : \mathbb{R} \to [0,1]$ be a smooth map with $g(0) = 0$ and $g(x) \to 1$ monotonically as $x \to \pm\infty$. Let $\beta \in (0,1)$ denote the optimal cost for the ergodic control problem that seeks to maximize

$$\limsup_{t\to\infty} \frac{1}{t} \int_0^t E[g(X(s))]\, ds.$$

The value function $V(\cdot) : \mathbb{R} \to \mathbb{R}$ for this problem (see Borkar (1989), Ch. VI) is given by

$$V(x) = \sup_{u(\cdot)\text{ admissible}} E_x\Big[\int_0^{\tau(0)} (g(X(s)) - \beta)\, ds\Big] \leq 2\phi(x). \tag{3.2}$$

Then $V(\cdot)$ is a C^2 solution to the associated Hamilton–Jacobi–Bellman equation (ibid.)

$$\frac{1}{2}\sigma^2(x)\frac{d^2V}{dx^2}(x) + \max_u \left(m(x,u)\frac{dV}{dx}(s)\right) = \beta - g(x).$$

For our choice of $g(\cdot)$, it follows that there exist $\varepsilon, a > 0$ such that

$$\max_u L_u V(x) \leq -\varepsilon \quad \text{ for } |x| > a. \tag{3.3}$$

REMARK. In conjunction with Ito's formula, (3.3) leads to: For $x \geq a$,

$$V(a) = E_x[V(x(\tau(a)))] \leq V(x) - \varepsilon E_x[\tau(a)].$$

Thus

$$\varepsilon\phi(a) + V(x) \geq \varepsilon(E_x[\tau(a)] + E_a[\tau(0)]) + V(a) = \varepsilon E_x[\tau(0)] + V(a),$$

leading to $V(x) \geq \varepsilon(\phi(x) - \phi(a)) + V(a)$. Together with its counterpart for $x \leq -a$ and (3.2), this shows that $V(\cdot), \phi(\cdot)$ have similar growth as $|x| \to \infty$.

Now consider independent scalar Brownian motions $W_1(\cdot), W_2(\cdot)$, and for $x_1, x_2 \in \mathbb{R}$, let $X_1(\cdot), X_2(\cdot)$ be the processes given by

$$X_i(t) = x_i + \int_0^t m(X_i(s), u(s))\, ds + \int_0^t \sigma(X_i(s))\, dW_i(s), \qquad t \geq 0,$$

for $i = 1, 2, u(\cdot)$ being a common control process admissible for both. Let $\xi = \inf\{t \geq 0 : X_1(t) = X_2(t)\}$.

LEMMA 3.1. $E[\xi] \leq K_1(V(x_1) + V(x_2))$.

Proof. Without loss of generality, suppose that $x_1 \geq x_2$. Define $\overline{X}_i(t)$, $t \geq 0$, $i = 1, 2$, by

$$\overline{X}_i(t) = x_i + \int_0^t m(\overline{X}_i(s), u_i(s))\, ds + \int_0^t \sigma(\overline{X}_i(s))\, dW_i(s), \qquad t \geq 0,$$

for $i = 1, 2$, with $u_1(\cdot) = v_M(\overline{X}_1(\cdot))$, $u_2(\cdot) = v_m(\overline{X}_2(\cdot))$. By the same comparison principle as in (3.1), it suffices to verify that

$$E[\overline{\xi}] \leq K_1(V(x_1) + V(x_2))$$

for $\overline{\xi} = \inf\{t \geq 0 : \overline{X}_1(t) = \overline{X}_2(t)\}$. From (3.3), it follows that $\overline{V}(x, y) = V(x) + V(y)$ serves as a stochastic Lyapunov function for $\overline{X}(\cdot) = (\overline{X}_1(\cdot), \overline{X}_2(\cdot))$, implying that the expected first hitting time thereof for any open ball in $\mathbb{R}^2$ is bounded (see, e.g., Bhattacharya (1981)). In particular, this is so if the ball is separated from the point (x_1, x_2) by the line $\{(x, y) : x = y\}$. The claim follows by standard arguments. ∎

The following lemma gives a useful estimate:

LEMMA 3.2. *For $T > 0$, the law of $X(t)$, $t \in (0, T]$, conditioned on $X_0 = x$, has a density $p(t, x, \cdot)$ satisfying the estimates*

$$C_1 t^{-1/2} \exp(-C_2|y - x|^2/t) \leq p(t, x, y) \leq C_3 t^{-1/2} \exp(-C_4|y - x|^2/t),$$

where C_1, C_2, C_3, C_4 are constants that depend on T.

Proof. For controls of the type $u(t) = v(X(t), t)$, $t \geq 0$, with a measurable $v : \mathbb{R} \times \mathbb{R}^+ \to U$, these are the estimates of Aronson (1967). The general case follows from the fact that the one-dimensional marginals of $X(\cdot)$ under any admissible $u(\cdot)$ can be mimicked by $u(\cdot)$ of the above type (Bhatt and Borkar (1996), p. 1552). ∎

COROLLARY 3.1. *For any $s \geq 0$, the conditional law of $X(s+t)$, $t \in [0, T]$, given $\widehat{\mathcal{F}}_s$ has a density satisfying the above estimates with x replaced by $X(s)$ throughout.*

Proof. Combine the above with Theorem 1.6, p. 13, of Borkar (1989). ∎

Let $\mathcal{P}_e(\mathbb{R}) = \{\mu \in \mathcal{P}(\mathbb{R}) : \int e^{ax}\, \mu(dx) < \infty \text{ for all } a > 0\}$.

For $x > 0$, $\phi(x) = E[\tau_{\mathrm{M}}(0)/X_{\mathrm{M}}(0) = x]$. Thus $\phi(\cdot)$ satisfies: $\phi(0) = 0$ and

$$\frac{1}{2}\sigma^2(x)\frac{d^2\phi}{dx^2}(x) + m(x, v_M(x))\frac{d\phi}{dx}(x) = -1 \quad \text{ on } (0, \infty).$$

Explicit solution of this o.d.e. shows that $\phi(\cdot)$ has at most exponential growth on $[0,\infty)$. A symmetric argument shows the same for $E[\tau_{\mathrm{m}}(0)/X_{\mathrm{m}}(0) = x]$ on $(-\infty, 0]$. Thus $V(\cdot)$ has at most exponential growth and hence $\int V(x)\, \mu(dx) < \infty$ for $\mu \in \mathcal{P}_e(\mathbb{R})$. What is more, by Corollary 3.1 above, if $\pi_0 \in \mathcal{P}_e(\mathbb{R})$, $E[V(X(t))] < \infty$ for $t \geq 0$, implying $E[\pi_t(V)] < \infty$, or

$\pi_t(V) < \infty$ a.s. A similar argument shows that $\pi_0 \in \mathcal{P}_e(\mathbb{R}) \Rightarrow \pi_t \in \mathcal{P}_e(\mathbb{R})$ a.s. for $t \geq 0$, allowing one to view $\{\pi_t\}$ as a $\mathcal{P}_e(\mathbb{R})$-valued process.

4. The vanishing discount limit. The associated discounted cost problem with discount factor $\alpha > 0$ is to minimize over all wide sense admissible $u(\cdot)$ the discounted cost

$$J_\alpha(u(\cdot), \pi_0) = E\Big[\int_0^\infty e^{-\alpha t} k(X(t), u(t))\, dt\Big] = E\Big[\int_0^\infty e^{-\alpha t} \pi_t(k(\cdot, u(t)))\, dt\Big].$$

For $\pi \in \mathcal{P}(\mathbb{R})$ define the discounted value function

$$\psi_\alpha(\pi) = \inf E\Big[\int_0^\infty e^{-\alpha t} \pi_t(k(\cdot, u(t)))\, dt / \pi_0 = \pi\Big],$$

where the infimum is over all wide sense admissible $u(\cdot)$. This infimum is, in fact, a minimum—see Borkar (1989), Chapter V. We shall need a bound on $|\psi_\alpha(\pi) - \psi_\alpha(\pi')|$ for $\pi \neq \pi'$. For this purpose, we first construct on a common probability space two solutions to (2.1), (2.2) with different initial laws, but a "common" wide sense admissible $u(\cdot)$ as follows. (We closely follow Borkar (1999).)

Let $(\Omega, \mathcal{F}, P_0)$ be a probability space on which we have $\mathbb{R}$-valued random variables $\widehat{X}_0, \widetilde{X}_0$ with laws π, π' respectively, scalar Brownian motions $W_1(\cdot), W_2(\cdot)$ and m-dimensional Brownian motions $\widehat{Y}(\cdot)$, $\widetilde{Y}(\cdot)$, such that $[\widehat{X}_0, \widetilde{X}_0, W_1(\cdot), W_2(\cdot), \widehat{Y}(\cdot), \widetilde{Y}(\cdot)]$ is an independent family. Also defined on $(\Omega, \mathcal{F}, P_0)$ is a U-valued process $u(\cdot)$ with measurable sample paths, independent of $(\widehat{X}_0, \widetilde{X}_0, W_1(\cdot), W_2(\cdot), \widetilde{Y}(\cdot))$, and satisfying: For $t \geq s$, $\widehat{Y}(t) - \widehat{Y}(s)$ is independent of the foregoing and of $u(y)$, $\widehat{Y}(y)$, $y \leq s$. Let $\widehat{X}(\cdot), \widetilde{X}(\cdot)$ denote the solutions to (2.1) with initial conditions $\widehat{X}_0, \widetilde{X}_0$ and driving Brownian motions $W_1(\cdot)$, $W_2(\cdot)$ replacing $W(\cdot)$, respectively. Define $\mathcal{F}_t^* =$ the right-continuous completion of $\sigma(\widehat{X}(s), \widetilde{X}(s), \widehat{Y}(s), \widetilde{Y}(s), W_1(s), W_2(s), u(s), s \leq t)$, $t \geq 0$. Without any loss of generality, let $\mathcal{F} = \bigvee_t \mathcal{F}_t^*$. Define a new probability measure P on $(\Omega, \mathcal{F})$ as follows: Let P_t, P_{0t} be the restrictions of P, P_0 respectively to $(\Omega, \mathcal{F}_t^*)$ for $t \geq 0$. Then

$$\frac{dP_t}{dP_{0t}} = \exp\bigg(\int_0^t (\langle h(\widehat{X}(s)), d\widehat{Y}(s)\rangle + \langle h(\widetilde{X}(s)), d\widetilde{Y}(s)\rangle) - \frac{1}{2}\int_0^t (\|h(\widehat{X}(s))\|^2 + \|h(\widetilde{X}(s))\|^2)\, ds\bigg), \qquad t \geq 0.$$

Novikov's criterion (see, e.g., Ikeda and Watanabe (1981)) ensures that the right-hand side is a legal Radon–Nikodym derivative. By Girsanov's theorem (ibid.), under P,

$$\widehat{Y}(t) = \int_0^t h(\widehat{X}(s))\,ds + \widehat{W}(t), \qquad \widetilde{Y}(t) = \int_0^t h(\widetilde{X}(s))\,ds + \widetilde{W}(t),$$

for $t \geq 0$, where $\widehat{W}(\cdot), \widetilde{W}(\cdot)$ are m-dimensional Brownian motions and $(\widehat{X}_0, \widetilde{X}_0, W_1(\cdot), W_2(\cdot), \widehat{W}(\cdot), \widetilde{W}(\cdot))$ is an independent family. Further, $u(\cdot)$ is a wide sense admissible control for both $\widehat{X}(\cdot), \widetilde{X}(\cdot)$.

What this construction achieves is to identify each wide sense admissible control $u(\cdot)$ for π with one wide sense admissible control $u(\cdot)$ for π'. (This identification can be many-one.) By a symmetric argument that interchanges the roles of π, π', one may identify every wide sense admissible control for π' with one for π. Now suppose that $\psi_\alpha(\pi) \leq \psi_\alpha(\pi')$. Then for a wide sense admissible $u(\cdot)$ that is optimal for π for the α-discounted cost problem,

$$\begin{aligned}|\psi_\alpha(\pi) - \psi_\alpha(\pi')| &= \psi_\alpha(\pi') - \psi_\alpha(\pi) \leq J_\alpha(u(\cdot), \pi') - J_\alpha(u(\cdot), \pi)\\ &\leq \sup |J_\alpha(u(\cdot), \pi') - J_\alpha(u(\cdot), \pi)|\end{aligned}$$

where we use the above identification of controls and the supremum is correspondingly interpreted as being over appropriate wide sense admissible controls. If $\psi_\alpha(\pi) > \psi_\alpha(\pi')$, a symmetric argument works. We have proved:

LEMMA 4.1. $|\psi_\alpha(\pi) - \psi_\alpha(\pi')| \leq \sup |J_\alpha(u(\cdot), \pi) - J_\alpha(u(\cdot), \pi')|$.

Let $\pi, \pi' \in \mathcal{P}_e(\mathbb{R})$. Then this leads to:

LEMMA 4.2. *For a suitable constant* $K_0 > 0$,

$$|\psi_\alpha(\pi) - \psi_\alpha(\pi')| \leq K_0(\pi(V) + \pi'(V)).$$

Proof. Let $K > 0$ be a bound on $|k(\cdot, u)|$ and $K_1 > 0$ as in Lemma 3.1. Let $\xi = \inf\{t \geq 0 : \widehat{X}(t) = \widetilde{X}(t)\}$ and set $X'(t) = \widetilde{X}(t) I\{t \leq \xi\} + \widehat{X}(t) I\{t > \xi\}$. Then $X'(\cdot)$ satisfies (2.1) with the same $u(\cdot)$ as for $\widehat{X}(\cdot), \widetilde{X}(\cdot)$ and the driving Brownian motion

$$W(\cdot) = W_2(\cdot) I\{\cdot \leq \xi\} + (W_2(\xi) + W_1(\cdot) - W_1(\xi)) I\{\cdot > \xi\}.$$

Then

$$\begin{aligned}|\psi_\alpha(\pi) - \psi_\alpha(\pi')| &\leq \sup |J_\alpha(u(\cdot), \pi) - J_\alpha(u(\cdot), \pi')|\\ &\leq \sup \int_0^\infty e^{-\alpha t} E[|k(\widehat{X}(t), u(t)) - k(X'(t), u(t))|]\,dt\\ &\leq 2K \sup E[\xi] \leq 2KK_1 E[V(\widehat{X}_0) + V(\widetilde{X}_0)]\\ &= 2KK_1(\pi(V) + \pi'(V)). \ \blacksquare\end{aligned}$$

From Borkar (1989), Chapter V, we know that $\psi_\alpha(\cdot)$ satisfies the martingale dynamic programming principle: For $t \geq 0$,

$$e^{-\alpha t}\psi_\alpha(\pi_t) + \int_0^t e^{-\alpha s}\pi_s(k(\cdot, u(s)))\,ds$$

is an $\{\mathcal{F}_t\}$-submartingale. That is, for $t \geq s$,

$$\psi_\alpha(\pi_s) \leq E\Big[e^{-\alpha t}\psi_\alpha(\pi_t) + \int_s^t e^{-\alpha y}\pi_y(k(\cdot, u(y)))\,dy/\mathcal{F}_s\Big] \quad \text{a.s.} \tag{4.1}$$

Fix $\pi^* \in \mathcal{P}_e(\mathbb{R})$ and let $\bar{\psi}_\alpha(\pi) = \psi_\alpha(\pi) - \psi_\alpha(\pi^*)$, $\psi(\pi) = \limsup_{\alpha\to 0}\bar{\psi}_\alpha(\pi)$ and $\Delta = \liminf_{\alpha\to 0}\alpha\psi_\alpha(\pi^*)$. Clearly, $|\Delta|$ is bounded by any bound on $|k(\cdot,\cdot)|$ and in view of Lemma 4.2, $\psi(\pi) = O(\pi(V)) = O(\pi(\phi))$. Rewrite (4.1) as

$$\bar{\psi}_\alpha(\pi_s) \leq E\Big[e^{-\alpha t}\bar{\psi}_\alpha(\pi_t) + \int_s^t e^{-\alpha y}(\pi_y(k(\cdot, u(y))) - \alpha\psi_\alpha(\pi^*))\,dy/\mathcal{F}_s\Big].$$

Taking lim sup as $\alpha \to 0$ on both sides, we get

$$\psi(\pi_s) \leq E\Big[\psi_\alpha(\pi_t) + \int_s^t (\pi_y(k(\cdot, u(y))) - \Delta)\,dy/\mathcal{F}_s\Big].$$

Thus we have:

THEOREM 4.1. *Under any wide sense admissible* $u(\cdot)$,

$$\psi(\pi_t) - \int_0^t (\pi_s(k(\cdot, u(s))) - \Delta)\,ds, \qquad t \geq 0,$$

is an $\{\mathcal{F}_t\}$*-submartingale. Further, if* $\{(\pi_t, u(t)) : t \geq 0\}$ *is a stationary pair under which it is a martingale, then it must be an optimal stationary pair and* Δ *the optimal cost.*

Proof. The first part is proved above. The second follows exactly as in Theorem 3.1, Borkar (1999). ■

This is a weak "verification theorem", weak because existence of a stationary pair as above is not guaranteed, even though existence of an optimal stationary pair is. This is so because a priori, Δ need not equal the optimal cost. However, one can show as in Theorem 3.1 of Borkar (1999) that it is less than or equal to the optimal cost. It is conjectured that if $(\pi_t, u(t))$, $t \geq 0$, in (4.1) is an optimal stationary pair with the law of $\pi_t = \mu \in \mathcal{P}(\mathcal{P}_e(\mathbb{R}))$, then Δ obtained as above is indeed the optimal cost for μ-a.s. π^*.

References

D. G. Aronson (1967), *Bounds for the fundamental solution of a parabolic equation*, Bull. Amer. Math. Soc. 73, 890–896.

G. K. Basak, V. S. Borkar and M. K. Ghosh (1997), *Ergodic control of degenerate diffusions*, Stochastic Anal. Appl. 15, 1–17.

A. G. Bhatt and V. S. Borkar (1996), *Occupation measures for controlled Markov processes: characterization and optimality*, Ann. Probab. 24, 1531–1562.

R. N. Bhattacharya (1981), *Asymptotic behaviour of several dimensional diffusions*, in: Stochastic Nonlinear Systems in Physics, Chemistry and Biology, L. Arnold and R. Lefever (eds.), Springer Ser. Synerg. 8, Springer, Berlin, 86–99.

V. S. Borkar (1989), *Optimal Control of Diffusion Processes*, Pitman Res. Notes Math. Ser. 203, Longman Sci. and Tech., Harlow.

V. S. Borkar (1999), *The value function in ergodic control of diffusion processes with partial observations*, Stochastics Stochastics Reports 67, 255–266.

W. F. Fleming and E. Pardoux (1982), *Optimal control of partially observed diffusions*, SIAM J. Control Optim. 20, 261–285.

N. Ikeda and S. Watanabe (1981), *Stochastic Differential Equations and Diffusion Processes*, North-Holland, Amsterdam, and Kodansha, Tokyo.

C. Striebel (1984), *Martingale methods for the optimal control of continuous time stochastic systems*, Stochastic Process. Appl. 18, 329–347.

Vivek S. Borkar
School of Technology and Computer Science
Tata Institute of Fundamental Research
Homi Bhabha Road
Mumbai 400 005, India
E-mail: borkar@tifr.res.in

Received on 14.1.2000;
revised version on 31.7.2000