Available at WWW.MATHEMATICSWEB.ORG POWERED BY SCIENCE DIRECT

Stochastic Processes and their Applications 103 (2003) 293–310

stochastic processes and their applications

www.elsevier.com/locate/spa

Dynamic programming for ergodic control with partial observations

V.S. Borkar*

School of Technology and Computer Science, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India

Received 18 February 2001; received in revised form 11 April 2002; accepted 17 July 2002

Abstract

A dynamic programming principle is derived for a discrete time Markov control process taking values in a finite dimensional space, with ergodic cost and partial observations. This uses the embedding of the process into another for which an accessible atom exists and hence a coupling argument can be used. In turn, this is used for deriving a martingale dynamic programming principle for ergodic control of partially observed diffusion processes, by 'lifting' appropriate estimates from a discrete time problem associated with it to the continuous time problem.

1. Introduction

Recently, the author derived the dynamic programming principle for ergodic (or 'average cost') control of finite Markov chains with partial observations (Borkar, 2000a). This article extends this result first to a more general state space, viz., a finite dimensional Euclidean space, and then to continuous time diffusions for which the statement will be in the framework of the 'martingale approach' of Davis and Varaiya (1973), Rishel (1970), and Striebel (1984). In the first case, the approach is based on Athreya–Ney–Nummelin construction of pseudo-atoms (Athreya and Ney, 1978; Nummelin, 1978) as described in Meyn and Tweedie (1993, pp. 100–104), which allows us to adapt the coupling argument of Borkar (2000a). In the latter case, the derivation is via an embedded discrete control problem. See Bhatt and Borkar (1996), Borkar (1999), and Borkar (2000b) for earlier work on this problem.

* Tel.: +91-22-215-2971x2293; fax: +91-22-215-2110.
E-mail address: borkar@tifr.res.in (V.S. Borkar).

PII: S0304-4149(02)00190-4

In the next section, we set up the formalism for the discrete problem. Section 3 follows Borkar (2000a) to derive the dynamic programming principle using a coupling argument, facilitated at this level of generality by the Athreya–Ney–Nummelin construction of pseudo-atoms. Section 4 introduces the continuous time problem. Section 5 introduces an embedded discrete time problem that goes with it and uses it to derive the martingale dynamic programming principle for the continuous problem.

2. The discrete problem

Let S, W, U denote Polish spaces representing, resp., the state, observation and control spaces, with the additional restrictions that S be a finite dimensional Euclidean space and U compact. We shall denote by $\mathscr{P}(\cdots)$ the Polish space of probability measures on '$\cdots$' with the Prohorov topology (Borkar, 1995, Chapter 2). Let $\{X_n\}$ be an S-valued controlled Markov chain with associated U-valued control process $\{Z_n\}$ and W-valued observation process $\{Y_n\}$. The controlled transition kernel is given by the map

$$(x,u) \in S \times U \to p(x,u,\mathrm{d}z,\mathrm{d}y) \in \mathscr{P}(S \times W),$$

assumed to be continuous. Let λ denote the Lebesgue measure on S. We assume the existence of $\eta \in \mathscr{P}(W)$ and $\varphi \in C_{\mathrm{b}}(S \times U \times S \times W)$ such that $p(x,u,\mathrm{d}z,\mathrm{d}y) = \varphi(x,u,z,y)\lambda(\mathrm{d}z)\eta(\mathrm{d}y)$, with $\varphi(\cdot) > 0$. Thus,

$$P(X_{n+1} \in A, Y_{n+1} \in A'/X_m, Z_m, Y_m, m \leqslant n) = \int_{A'} \int_A \varphi(X_n, Z_n, z, y)\lambda(\mathrm{d}z)\eta(\mathrm{d}y) \quad (1)$$

for Borel $A \subset S, A' \subset W$. Call $\{Z_n\}$ strict sense admissible if it is adapted to $\sigma(Y_m, m \leqslant n), n \geqslant 0$. The ergodic control problem under partial observations in its original form is to minimize over all such $\{Z_n\}$ the 'ergodic cost'

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} E[k(X_m, Z_m)] \quad (2)$$

for a prescribed $k \in C_{\mathrm{b}}(S \times U)$. Define $\bar{k} \in C_{\mathrm{b}}(\mathscr{P}(S), U)$ by $\bar{k}(\mu,u) = \int k(x,u)\mu(\mathrm{d}x)$, $\mu \in \mathscr{P}(S), u \in U$. Then, (2) equals

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} E[\bar{k}(\pi_m, Z_m)], \quad (3)$$

where π_n is the regular conditional law of X_n given $\sigma(Y_m, Z_m, m \leqslant n), n \geqslant 0$. Standard Bayes arguments show that $\{\pi_n\}$ is given recursively by the nonlinear filter

$$\pi_{n+1}(\mathrm{d}z) = \frac{\int \pi_n(\mathrm{d}x)\varphi(x, Z_n, z, Y_{n+1})\lambda(\mathrm{d}z)}{\int\int \pi_n(\mathrm{d}x)\lambda(\mathrm{d}z')\varphi(x, Z_n, z', Y_{n+1})}, \quad n \geqslant 0. \quad (4)$$

This allows one to consider the equivalent complete observations ergodic control problem, the so-called 'separated' control problem of controlling the $\mathscr{P}(S)$-valued controlled Markov process $\{\pi_n\}$ evolving according to (4), over strict sense admissible $\{Z_n\}$, so as to minimize the ergodic cost (3).

For later technical convenience, we assume that Y_0 is deterministic, so that $\pi_0 =$ the law of X_0. This causes no loss of generality, as we can always condition on Y_0 a priori. Following Borkar (1993), we may exhibit $\{X_n\}$ as a noise-driven dynamical system

$$X_{n+1} = F(X_n, Z_n, \xi_{n+1}), \quad n \geqslant 0, \tag{5}$$

where $F : S \times U \times [0,1] \to S$ is measurable and $\{\xi_n\}$ are i.i.d. uniformly distributed on $[0,1]$. (This may require an augmentation of the underlying probability space (Borkar, 1993).)

We shall now reproduce in law these processes on a more convenient probability space, the so-called 'canonical' space. Thus, let $\Omega = [0,1]^\infty \times U^\infty \times W^\infty \times S$ and let $\mathscr{F}$ denote the corresponding product Borel σ-field. Let (u, v, w, x) denote a typical element of Ω with $u = [u_0, u_1, \ldots] \in [0,1]^\infty$, $v = [v_0, v_1, \ldots] \in U^\infty$, $w = [w_0, w_1, \ldots] \in W^\infty$ and $x \in S$. Let $u^n = [u_0, u_1, \ldots, u_n]$ and define v^n, w^n analogously for $n \geqslant 0$. By the definition of strict sense admissible controls,

$$Z_n = \psi_n(Y_0, \ldots, Y_n), \quad n \geqslant 0$$

for some $\psi_n : W^{n+1} \to U$. Define

$$\bar{\psi}_n(w^n) = [\psi_0(w^0), \psi^1(w^1), \ldots, \psi_n(w^n)], \quad n \geqslant 0. \tag{6}$$

Let ℓ^n denote the Lebesgue measure on $[0,1]^n$ and η^n the product measure $\eta \times \eta \times \cdots \times \eta$ (n times) for $n \geqslant 1$. Define a probability measure P_0 on $(\Omega, \mathscr{F})$ by: If $B_1 \subset [0,1]^{n+1}$, $B_2 \subset U^{n+1}$, $B_3 \subset W^{n+1}$, $B_4 \subset S$ are Borel sets, then

$$P_0\left(\prod_{i=1}^{4} B_i\right) = \ell^{n+1}(B_1)\eta^{n+1}(B_3 \cap \{w^n\colon\ \bar{\psi}_n(w^n) \in B_2\})\pi_0(B_4).$$

Define $\{\xi_n, Z_n, Y_n\}$, X_0 canonically on $(\Omega, \mathscr{F}, P_0)$ by

$$\xi_n((u, v, w, x)) = u_n,$$

$$Z_n((u, v, w, x)) = v_n,$$

$$Y_n((u, v, w, x)) = w_n,$$

$$X_0((u, v, w, x)) = x$$

for $n \geqslant 0$. Then under P_0,

- $\{\xi_n\}$ are i.i.d. uniform on $[0,1]$,
- $\{Y_n\}$ are i.i.d. with law η,
- X_0 has law π_0,
- $(\{\xi_n\}, \{Y_n\}, X_0)$ are an independent family, and,
- $\{Z_n\}$ is specified by (6).

Define $\{X_n\}$ recursively by (5). By construction, it is a controlled Markov chain satisfying

$$P(X_{n+1} \in A / X_m, Z_m, m \leqslant n) = \int_A \bar{\varphi}(X_n, Z_n, z)\lambda(\mathrm{d}z),$$

where $\bar{\varphi}(x,u,z) = \int \varphi(x,u,z,y)\eta(\mathrm{d}y)$. For $n \geqslant 0$, let $\mathscr{F}_n = \sigma(X_m, Z_m, Y_m, \xi_m, m \leqslant n)$ and let P_{0n} be the restriction of P_0 to $(\Omega, \mathscr{F}_n)$ for $n \geqslant 0$. Define a new probability measure P on $(\Omega, \mathscr{F})$ as follows. If P_n denotes its restriction to $(\Omega, \mathscr{F}_n)$, then $P_n \ll P_{0n}$ with

$$\Lambda_n \triangleq \frac{\mathrm{d}P_n}{\mathrm{d}P_{0n}} = \prod_{m=0}^{n-1} \frac{\varphi(X_m, Z_m, X_{m+1}, Y_{m+1})}{\bar{\varphi}(X_m, Z_m, X_{m+1})}, \quad n \geqslant 0. \tag{7}$$

It is easily verified that $(\Lambda_n, \mathscr{F}_n)$ is a nonnegative martingale with mean 1 and therefore the above defines in a consistent and unique manner a probability measure P on $(\Omega, \bigvee_n \mathscr{F}_n)$. Since $\mathscr{F} = \bigvee_n \mathscr{F}_n$ by construction, we are through. Furthermore, under P, $\{X_n, Y_n, Z_n, \xi_n\}$ have the same joint law as the corresponding processes we started with.

This construction permits us to define wide sense admissible controls as in Fleming and Pardoux (1982). Intuitively, this relaxation allows for controls that incorporate extraneous randomness which does not, however, use any information that it should not. Formally, $\{Z_n\}$ is said to be a wide sense admissible control if for each n, $(Z_m, Y_m, m \leqslant n)$ is independent of $\{\xi_m\}$, $X_0, \{Y_m, m > n\}$ under P_0. This includes in particular strict sense admissible controls. See Fleming and Pardoux (1982) for a full justification of this relaxation, which carries over in toto to the present framework. Our 'relaxed' partially observed control problem then is to minimize (3) over all wide sense admissible $\{Z_n\}$. Under P_0, the laws of $X_0, \{\xi_n\}, \{Y_n\}$ are fixed and $\{X_n\}$ gets specified by (5) once $\{Z_n\}$ is. Thus, the above framework is specified in law by specifying the regular conditional law of $\{Z_n\}$ given $\{Y_n\}$, or equivalently, the joint law of $(\{Z_n\}, \{Y_n\})$ where the marginal for $\{Y_n\}$ remains fixed. Thus we may refer to either of these as *the* wide sense admissible control. Denote by Φ the set of all wide sense admissible controls, with a typical element denoted by $\{Z_n\}$ by abuse of notation.

We shall also make the following stability assumption: There exist functions $h, \mathscr{V} \in C(S)$ satisfying $h \geqslant 1, \lim_{\|x\| \to \infty} h(x) = \lim_{\|x\| \to \infty} \mathscr{V}(x) = \infty$, such that under any wide sense admissible $\{Z_n\}$,

$$E[\mathscr{V}(X_{n+1})/\mathscr{F}_n] - \mathscr{V}(X_n) \leqslant -h(X_n) + CI_B(X_n), \tag{8}$$

where $C > 0$ and $B = \{x \in S : \|x\| \leqslant R\}$ for some $R > 0$. Then B is compact nonempty, with $\lambda(B) > 0$. Let $\tau_B = \min\{n \geqslant 0 : X_n \in B\}$ ($= \infty$ if the r.h.s. is empty). Then it is well known that

$$E[\tau_B / X_0 = x] = \mathrm{O}(\mathscr{V}(x)). \tag{9}$$

(See, e.g., Meyn and Tweedie, 1993, Chapter 14, p. 338.) In particular, it is finite everywhere. With an eye on later developments, we define $\mathscr{P}_0(S) = \{\mu \in \mathscr{P}(S) \colon \int \mathscr{V}\,\mathrm{d}\mu < \infty\}$. Note that this will be a proper subset of $\mathscr{P}(S)$. Furthermore, by (8), $E[\mathscr{V}(X_{n+1})] = E[\int \mathscr{V}\,\mathrm{d}\pi_{n+1}] \leqslant E[\mathscr{V}(X_n)] +$ a constant $= E[\int \mathscr{V}\,\mathrm{d}\pi_n] +$ a constant, whence it follows that $\pi_0 \in \mathscr{P}_0(S) \Rightarrow \pi_n \in \mathscr{P}_0(S)\ \forall\ n$, a.s. Thus we may suppose that $\pi_0 \in \mathscr{P}_0(S)$ and view $\{\pi_n\}$ as a process in $\mathscr{P}_0(S)$.

We shall further assume:

(†) Under all wide sense admissible controls,

$$\limsup_{n \to \infty} \frac{E[\mathscr{V}(X_n)]}{n} = 0.$$

3. The vanishing discount limit

In preparation for the vanishing discount argument to be used later, we introduce the family of discounted cost problems indexed by the discount factor $\alpha > 0$, wherein one seeks to minimize over all wide sense admissible $\{Z_n\}$ the discounted cost

$$J_\alpha(\{Z_n\},\pi) = E\left[\sum_{m=0}^{\infty} \alpha^m k(X_m, Z_m)/\pi_0 = \pi\right]$$
$$= E\left[\sum_{m=0}^{\infty} \alpha^m \bar{k}(\pi_m, Z_m)/\pi_0 = \pi\right].$$

The associated value function

$$V_\alpha(\pi) = \inf_{\Phi} J_\alpha(\{Z_n\},\pi),$$

then satisfies the dynamic programming equation (Hernández-Lerma and Lasserre, 1999)

$$V_\alpha(\pi) = \min_u \left[\bar{k}(\pi,u) + \alpha \int \phi(\pi,u,\mathrm{d}\pi')V_\alpha(\pi')\right], \qquad \pi \in \mathscr{P}_0(S), \tag{10}$$

where $(\pi,u) \in \mathscr{P}_0(S) \times U \to \phi(\pi,u,\mathrm{d}\pi') \in \mathscr{P}(\mathscr{P}_0(S))$ is the controlled transition kernel of the $\mathscr{P}_0(S)$-valued controlled Markov chain $\{\pi_n\}$. From our hypotheses on $p(\cdot)$, it is easily verified that $\phi(\cdot,\cdot,\mathrm{d}\pi')$ is a continuous map. Furthermore, if $v(\pi)$ attains the minimum on the r.h.s. for some measurable $v : \mathscr{P}(S) \to U$, then $Z_n = v(\pi_n)$, $n \geqslant 0$, defines an optimal control and conversely, if $\{Z_n\}$ is optimal, then Z_n attains the minimum on the r.h.s., a.s. with respect to the law of π_n. The existence of a $v(\cdot)$ as above is guaranteed by a standard measurable selection theorem (Wagner, 1977). See Hernández-Lerma and Lasserre (1996) for a detailed account of these developments.

We shall need to compare $V_\alpha(\cdot)$ for two different values of its argument. With this in mind, we construct on a common probability space two S-valued Markov chains as above with a common control process $\{Z_n\} \in \Phi$, but different initial laws, say $\tilde{\pi}$ and $\hat{\pi}$. This is done by a small modification of the construction of the previous section. Note that the specification of $\{Z_n\} \in \Phi$ for initial law $\tilde{\pi}$ entails the specification of its joint law with $\{Y_n\}$ under P_0, assumed to satisfy the independence or conditional independence conditions stipulated in the definition of wide sense admissible controls. Denote this joint law by $\zeta(\mathrm{d}y^\infty, \mathrm{d}u^\infty) \in \mathscr{P}(W^\infty \times U^\infty)$. Define

$$\bar{\Omega} = ([0,1]^\infty \times S) \times ([0,1]^\infty \times S) \times U^\infty \times W^\infty \times W^\infty$$

with $\bar{\mathscr{F}}$ = the corresponding product σ-field and $\bar{P}_0$ the probability measure on $(\bar{\Omega},\bar{\mathscr{F}})$ defined by

$$\bar{P}_0((\mathrm{d}\tilde{u}^\infty \times \mathrm{d}\tilde{x}) \times (\mathrm{d}\hat{u}^\infty \times \mathrm{d}\hat{x}) \times \mathrm{d}z^\infty \times \mathrm{d}\tilde{y}^\infty \times \mathrm{d}\hat{y}^\infty)$$
$$= \ell^\infty(\mathrm{d}\tilde{u}^\infty)\tilde{\pi}(\mathrm{d}\tilde{x})\ell^\infty(\mathrm{d}\hat{u}^\infty)\hat{\pi}(\mathrm{d}\hat{x})\zeta(\mathrm{d}\tilde{y}^\infty, \mathrm{d}z^\infty)\eta^\infty(\mathrm{d}\hat{y}^\infty).$$

On $(\bar{\Omega},\bar{\mathscr{F}},\bar{P}_0)$, define processes $\{\tilde{\xi}_n\}$, $\{\hat{\xi}_n\}$, $\{Z_n\}$, $\{\tilde{Y}_n\}$, $\{\hat{Y}_n\}$, and random variables $\tilde{X}_0$, $\hat{X}_0$ canonically as follows: For $\omega = (\tilde{u}^\infty, \tilde{x}, \hat{u}^\infty, \hat{x}, z^\infty, \tilde{y}^\infty, \hat{y}^\infty)$, let $\tilde{\xi}_n(\omega) = \tilde{u}_n$,

$\hat{\xi}_n(\omega)=\hat{u}_n$, $\tilde{X}_0(\omega)=\tilde{x}$, $\hat{X}_0(\omega)=\hat{x}$, $Z_n(\omega)=z_n$, $\tilde{Y}_n(\omega)=\tilde{y}_n$, $\hat{Y}_n(\omega)=\hat{y}_n$, $n \geqslant 0$. Define $\{\tilde{X}_n\}$, $\{\hat{X}_n\}$ recursively by

$$\tilde{X}_{n+1} = F(\tilde{X}_n, Z_n, \tilde{\xi}_{n+1}),$$

$$\hat{X}_{n+1} = F(\hat{X}_n, Z_n, \hat{\xi}_{n+1})$$

for $n \geqslant 0$. Let $\Gamma_n = \sigma(\hat{X}_m, \tilde{X}_m, \hat{Y}_m, \tilde{Y}_m, \hat{\xi}_m, \tilde{\xi}_m, Z_m, m \leqslant n)$. Then $\bar{\mathscr{F}} = \bigvee_n \Gamma_n$. Define a new probability measure $\bar{P}$ on $(\bar{\Omega}, \bar{\mathscr{F}})$ as follows: If $\bar{P}_n, \bar{P}_{0n}$ are the restrictions of $\bar{P}, \bar{P}_0$, resp., to $(\bar{\Omega}, \Gamma_n)$, then $\bar{P}_n \ll \bar{P}_{0n}$ with

$$\bar{\Lambda}_n \triangleq \frac{\mathrm{d}\bar{P}_n}{\mathrm{d}\bar{P}_{0n}} = \prod_{m=0}^{n-1} \frac{\varphi(\tilde{X}_m, Z_m, \tilde{X}_{m+1}, \tilde{Y}_{m+1})}{\bar{\varphi}(\tilde{X}_m, Z_m, \tilde{X}_{m+1})} \frac{\varphi(\hat{X}_m, Z_m, \hat{X}_{m+1}, \hat{Y}_{m+1})}{\bar{\varphi}(\hat{X}_m, Z_m, \hat{X}_{m+1})},$$

$n \geqslant 0$. Then the controlled Markov chains $\{\tilde{X}_n\}$, $\{\hat{X}_n\}$ defined on $(\bar{\Omega}, \bar{\mathscr{F}}, \bar{P})$ form the desired pair in so far as their initial laws are, resp., $\tilde{\pi}, \hat{\pi}$, and they are governed by a "common" $\{Z_n\}$ which is wide sense admissible for both. A rigorous justification is given in Lemma 3.1 of Borkar (2000a). For later use, we also introduce the notation

$$\bar{Y}_n = (\tilde{Y}_n, \hat{Y}_n), \qquad \bar{X}_n = (\tilde{X}_n, \hat{X}_n), \quad n \geqslant 0.$$

What the foregoing achieves is to identify each wide sense control for $\tilde{\pi}$ with one for $\hat{\pi}$. This identification may be many-one. We next combine this with the Athreya–Ney–Nummelin construction. For this, note that $\{\bar{X}_n\}$ is an $H \triangleq S^2$-valued controlled Markov chain with U-valued control $\{Z_n\}$ and W^2-valued observation process $\{\bar{Y}_n\}$. Let the controlled transition kernel be denoted by $\bar{p}(x, u, \mathrm{d}x' \times \mathrm{d}y') \in \mathscr{P}(S^2 \times W^2)$ for $x \in H, u \in U$.

Define $G = B^2$ and define $v \in \mathscr{P}(H)$ by $v(A) = (\lambda \times \lambda)(A \cap G)/\lambda(B)^2$ for Borel $A \subset H$. Let $\delta = \frac{1}{2}(\inf_{x \in B, u \in U, y \in B} \bar{\varphi}(x, u, y)\lambda(B))^2 > 0$. Then, for any Borel $A \subset H$,

$$\bar{p}(x, u, A \times W^2) \geqslant \delta I_G\{x\} v(A), \tag{11}$$

which is the 'minorization condition' of Meyn and Tweedie (1993, p. 102) in our context. This allows us to adapt the Athreya–Ney–Nummelin construction from Meyn and Tweedie (1993, pp. 102–105), as described next.

Let $H^* = H \times \{0,1\}$, endowed with its Borel σ-field. For any measure μ on H, define a measure μ^* on H^* as follows: For Borel $A \subset H$, let $A_0 = A \times \{0\}$, $A_1 = A \times \{1\}$. (A similar notation will be followed, in what follows, with other sets in place of A.) Then

$$\mu^*(A_0) = (1 - \delta)\mu(A \cap G) + \mu(A \cap G^{\mathrm{c}}),$$

$$\mu^*(A_1) = \delta\mu(A \cap G).$$

Note that $\mu^*(A_0 \cup A_1) = \mu(A)$ and for $A \subset G^{\mathrm{c}}$, $\mu^*(A_0) = \mu(A)$. For a measure μ on $H \times W^2$, on the other hand, we define the measure μ^* on $H^* \times W^2$ by

$$\mu^*(A_0 \times D) = (1 - \delta)\mu((A \cap G) \times D) + \mu((A \cap G^{\mathrm{c}}) \times D),$$

$$\mu^*(A_1 \times D) = \delta\mu((A \cap G) \times D).$$

for A, A_0, A_1 as above, $D \subset W^2$ Borel.

Also, for a measurable map $f: H \times U \to \mathbb{R}$, define $f^*: H^* \times U \to \mathbb{R}$ by: for $(x^1, x^2, i) \in H \times \{0,1\}$, $u \in U$,

$$f^*((x^1, x^2, i), u) = f((x^1, x^2), u).$$

Likewise, for $f: H^n \times U^m \to \mathbb{R}$ with $n, m \geqslant 1$, define $f^*: (H^*)^n \times U^m \to \mathbb{R}$, by

$$\begin{aligned} &f^*((x_1^1, x_1^2, i_1), \ldots, (x_n^1, x_n^2, i_n), u_1, \ldots, u_m) \\ &\quad = f((x_1^1, x_1^2), \ldots, (x_n^1, x_n^2), u_1, \ldots, u_n). \end{aligned}$$

On a convenient (e.g., 'canonical') probability space $(\Omega^*, \mathscr{F}^*, P^*)$, define an H^*-valued controlled Markov chain $\{X_n^*, i_n^*\}$ (where $X_n^* = (\tilde{X}_n^*, \hat{X}_n^*)$) with U-valued control process $\{Z_n^*\}$ and W^2-valued observation process $\{Y_n^*\}$, so that:

- the controlled transition kernel of $\{X_n^*, i_n^*, Y_n^*\}$ is given by: for $x = (x_0, i_0) \in H^*$,

$$\begin{aligned} q(x, u, \mathrm{d}x' \times \mathrm{d}y) &= \bar{p}^*(x_0, u, \mathrm{d}x' \times \mathrm{d}y), && x \in H_0 - G_0 \\ &= \tfrac{1}{(1-\delta)}(\bar{p}^*(x_0, u, \mathrm{d}x' \times \mathrm{d}y) - \delta v^*(\mathrm{d}x')\eta^2(\mathrm{d}y)), && x \in G_0 \\ &= v^*(\mathrm{d}x')\eta^2(\mathrm{d}y), && x \in H_1, \end{aligned}$$

- with

$$\begin{aligned} P^*((X_0^*, i_0^*) \in A_0, Y_0^* \in A', Z_0^* \in A'') &= (1-\delta) P(\bar{X}_0 \in A \cap G, \bar{Y}_0 \in A', Z_0 \in A'') \\ &\quad + P(\bar{X}_0 \in A \cap G^{\mathrm{c}}, \bar{Y}_0 \in A', Z_0 \in A''), \\ P^*((X_0^*, i_0^*) \in A_1, Y_0^* \in A', Z_0^* \in A'') &= \delta P(\bar{X}_0 \in A \cap G, \bar{Y}_0 \in A', Z_0 \in A''), \end{aligned}$$

  for $A \subset H, A' \subset W^2, A'' \subset U$ Borel,
- and

$$\begin{aligned} &P^*(Z_n^* \in A / (X_m^*, i_m^*, Y_m^*) = (x_m, i_m, y_m), m \leqslant n, Z_k^* = z_k, k < n) \\ &\quad = P(Z_n \in A / (\bar{X}_m, \bar{Y}_m) = (x_m, y_m), m \leqslant n, Z_k = z_k, k < n), \end{aligned}$$

  for $n \geqslant 1$.

Let $\Gamma_n^* = \sigma(X_m^*, i_m^*, Y_m^*, Z_m^*, m \leqslant n), n \geqslant 0$.

Lemma 3.1. *G_1 is an accessible atom of $\{X_n^*, i_n^*\}$ in the sense of Meyn and Tweedie (1993, p. 100).*

This follows as in Meyn and Tweedie (1993, pp. 104–105).

Lemma 3.2. *For any Borel $A^i \subset H, D^i \subset W^2, Q^i \subset U$, $0 \leqslant i \leqslant n$, $n \geqslant 0$,*

$$\begin{aligned} &P^*\left(((X_0^*, i_0^*, Y_0^*, Z_0^*), \ldots, (X_n^*, i_n^*, Y_n^*, Z_n^*)) \in \prod_{i=0}^{n} (A_0^i \cup A_1^i) \times D^i \times Q^i\right) \\ &\quad = P\left(((\bar{X}_0, \bar{Y}_0, Z_0), \ldots, (\bar{X}_n, \bar{Y}_n, Z_n)) \in \prod_{i=0}^{n} A^i \times D^i \times Q^i\right). \end{aligned} \quad (12)$$

Proof. The proof is by induction on n. The case $n=0$ is included in our construction explicitly. Suppose the claim holds for some $n \geqslant 0$. From our definition of $q(\cdot)$, it follows exactly as in Theorem 5.1.3, Meyn and Tweedie (1993, p. 104), that

$$P^*((X^*_{n+1}, Y^*_{n+1}) \in (A_0^{n+1} \cup A_1^{n+1}) \times D^{n+1}/\Gamma^*_n) = P((\bar{X}_n, \bar{Y}_n) \in A^{n+1} \times D^{n+1}/\Gamma_n)$$

for Borel $A^{n+1} \subset H, D^{n+1} \subset W^2$. Multiply the l.h.s. (resp., r.h.s.) of (12) by the l.h.s. (resp., r.h.s.) of the above. Taking expectations and using the induction hypothesis,

$$\begin{aligned}
&P^* \Bigg(((X^*_0, i^*_0, Y^*_0), \ldots, (X^*_{n+1}, i^*_{n+1}, Y^*_{n+1})) \\
&\quad \in \prod_{i=0}^{n+1} (A_0^i \cup A_1^i) \times D^i, (Z^*_0, \ldots, Z^*_n) \in \prod_{i=0}^{n} Q^i \Bigg) \\
&\quad = P \left(((\bar{X}_0, \bar{Y}_0), \ldots, (\bar{X}_{n+1}, \bar{Y}_{n+1})) \in \prod_{i=0}^{n+1} A^i \times D^i, (Z_0, \ldots, Z_n) \in \prod_{i=0}^{n} Q^i \right).
\end{aligned}$$

The induction step is then completed using our specification above of the regular conditional law of Z^*_{n+1} given $((X^*_m, i^*_m, Y^*_m), m \leqslant n, Z^*_m, m < n)$ for $n \geqslant 1$. $\square$

Let $\tau = \min\{n \geqslant 0 : (X^*_n, i^*_n) \in G_1\}$.

Lemma 3.3. *For any* $(x,i) = ((x_1,x_2),i) \in H^*$,

$$E[\tau/(X^*_0, i^*_0) = (x,i)] = O(\mathscr{V}(x_1) + \mathscr{V}(x_2)).$$

Proof. Define $\bar{\mathscr{V}}(x_1,x_2,i) \triangleq \mathscr{V}(x_1) + \mathscr{V}(x_2)$ and $\bar{h}(x_1,x_2,i) \triangleq h(x_1) + h(x_2)$ for $x_1,x_2 \in S, i \in \{0,1\}$. From (8), it then follows that

$$E[\bar{\mathscr{V}}(X^*_{n+1}, i^*_{n+1})/\Gamma^*_n] - \bar{\mathscr{V}}(X^*_n, i^*_n) \leqslant -\bar{h}^*(X^*_n, i^*_n) + C^*,$$

for a suitable constant C^*. It follows that a stochastic Liapunov condition similar to (8) holds for the chain (X^*_n, i^*_n) with $\bar{\mathscr{V}}$, $\bar{h}$ + a suitable constant, replacing $\mathscr{V}, h$, resp. The claim then is the counterpart of (9) and follows by standard arguments as in Theorem 14.2.3 of Meyn and Tweedie (1993, p. 338). $\square$

From now on we largely mimick the arguments of Borkar (2000a).

Lemma 3.4. *For a suitable constant* $\bar{K} < \infty$,

$$|V_\alpha(\tilde{\pi}) - V_\alpha(\hat{\pi})| \leqslant \bar{K} \left(\int \mathscr{V}\, \mathrm{d}\tilde{\pi} + \int \mathscr{V}\, \mathrm{d}\hat{\pi} \right).$$

Proof. Without loss of generality, let $V_\alpha(\tilde{\pi}) \geqslant V_\alpha(\hat{\pi})$, the other case being handled by a symmetric argument. Let $\{\hat{Z}_n\}$ be an optimal wide sense admissible process for initial law $\hat{\pi}$. (The existence of this is established in Borkar, 1989, Section V.3.)

Then

$$|V_\alpha(\tilde{\pi}) - V_\alpha(\hat{\pi})| = V_\alpha(\tilde{\pi}) - V_\alpha(\hat{\pi}) \leqslant J_\alpha(\{\hat{Z}_n\},\tilde{\pi}) - J_\alpha(\{\hat{Z}_n\},\hat{\pi}),$$

where in order to interpret the r.h.s., we use the above construction of processes with a common control process but different initial laws. Let $\hat{k}(x_1,x_2,z) = k(x_1,z) - k(x_2,z)$. Then

$$\begin{aligned}|V_\alpha(\tilde{\pi}) - V_\alpha(\hat{\pi})| &\leqslant \sup_{\Phi} |J_\alpha(\{Z_n\},\tilde{\pi}) - J_\alpha(\{Z_n\},\hat{\pi})| \\ &\leqslant \sup_{\Phi} \left|\sum_{m=0}^{\infty} \alpha^m E[k(\tilde{X}_m, Z_m) - k(\hat{X}_m, Z_m)]\right| \\ &\leqslant \sup_{\Phi} \left|\sum_{m=0}^{\infty} \alpha^m E[\hat{k}^*(X_m^*, Z_m^*)]\right| \\ &= \sup_{\Phi} \left|\sum_{m=0}^{\infty} \alpha^m E[\hat{k}^*(X_m^*, Z_m^*) I\{\tau \geqslant m\}],\right.\end{aligned}$$

where the last step follows from the fact that $\tilde{X}^*_{\tau+m}, \hat{X}^*_{\tau+m}$ for $m \geqslant 1$ have the same law conditioned on Γ^*_τ and thus $E[\hat{k}^*(X_m^*, Z_m^*) I\{\tau < m\}] = 0$. Hence, for any $K \geqslant |k(\cdot,\cdot)|$,

$$\begin{aligned}|V_\alpha(\tilde{\pi}) - V_\alpha(\hat{\pi})| &\leqslant 2KE[\tau] \\ &\leqslant \hat{K}E[\mathscr{V}(\hat{X}_0) + \mathscr{V}(\tilde{X}_0)],\end{aligned}$$

by the preceding lemma, for a suitable constant $\hat{K}$. The claim follows. □

Fix $\pi^* \in \mathscr{P}_0(S)$ and define $\bar{V}_\alpha(\pi) = V_\alpha(\pi) - V_\alpha(\pi^*)$. From (10), we have

$$\bar{V}_\alpha(\pi) = \min_u \left[\bar{k}(\pi,u) + \alpha \int \phi(\pi,u,\mathrm{d}\pi')\bar{V}_\alpha(\pi') - (1-\alpha)V_\alpha(\pi^*)\right].$$

It is easy to see that $(1-\alpha)V_\alpha(\pi^*)$ is bounded. Thus, we can find $\alpha(n) \to 1$ such that $(1-\alpha(n))V_{\alpha(n)}(\pi^*) \to \gamma$ for some $\gamma \in \mathbb{R}$. Let $\hat{V}(\pi) = \limsup_{n\to\infty} \bar{V}_{\alpha(n)}(\pi)$, $\tilde{V}(\pi) = \liminf_{n\to\infty} \bar{V}_{\alpha(n)}(\pi)$.

Lemma 3.5. *$\hat{V}$ satisfies*

$$\hat{V}(\pi) \leqslant \min_u \left[\bar{k}(\pi,u) - \gamma + \int \phi(\pi,u,\mathrm{d}\pi')\hat{V}(\pi')\right]. \tag{13}$$

Proof. Letting $n \to \infty$ in the above equation along $\alpha = \alpha(n)$, we have

$$\begin{aligned}\hat{V}(\pi) &= \inf_n \sup_{m \geqslant n} \min_u \left[\bar{k}(\pi,u) - \gamma + \int \phi(\pi,u,\mathrm{d}\pi')\bar{V}_{\alpha(m)}(\pi')\right] \\ &\leqslant \inf_n \min_u \sup_{m\geqslant n} \left[\bar{k}(\pi,u) - \gamma + \int \phi(\pi,u,\mathrm{d}\pi')\bar{V}_{\alpha(m)}(\pi')\right]\end{aligned}$$

$$= \min_u \inf_n \sup_{m \geqslant n} \left[\bar{k}(\pi, u) - \gamma + \int \phi(\pi, u, d\pi') \bar{V}_{\alpha(m)}(\pi') \right]$$

$$\leqslant \min_u \left[\bar{k}(\pi, u) - \gamma + \int \phi(\pi, u, d\pi') \hat{V}(\pi') \right].$$

The claim follows. □

Similarly we have:

Lemma 3.6. *$\tilde{V}$ satisfies*

$$\tilde{V}(\pi) \geqslant \min_u \left[\bar{k}(\pi, u) - \gamma + \int \phi(\pi, u, d\pi') \tilde{V}(\pi') \right]. \tag{14}$$

Proof. As above, we have

$$\tilde{V}(\pi) = \liminf_{n \to \infty} \min_u \left[\bar{k}(\pi, u) - \gamma + \int \phi(\pi, u, d\pi') \bar{V}_{\alpha(n)}(\pi') \right]$$

$$= \liminf_{n \to \infty} \left[\bar{k}(\pi, u_n) - \gamma + \int \phi(\pi, u_n, d\pi') \bar{V}_{\alpha(n)}(\pi') \right],$$

where u_n is the minimizer on the r.h.s. of the first equation. Fix π. By dropping to a subsequence if necessary, we may suppose that $\bar{V}_{\alpha(n)}(\pi) \to \tilde{V}(\pi)$ and $u_n \to u^*$ in U. By the preceding lemma $|\bar{V}_\alpha(\pi)| \leqslant K_1(1 + \int \mathscr{V} \, d\pi)$, $\pi \in \mathscr{P}_0(S)$. Thus, by Lemma 8.3.7 of Hernández-Lerma and Lasserre (1999, pp. 48–49), the last expression above is bounded from below by

$$\bar{k}(\pi, u^*) - \gamma + \int \phi(\pi, u^*, d\pi') \tilde{V}(\pi')$$

$$\geqslant \min_u \left[\bar{k}(\pi, u) - \gamma + \int \phi(\pi, u, d\pi') \tilde{V}(\pi') \right].$$

The claim follows. □

We also have:

Lemma 3.7. *γ is the optimal cost for the separated ergodic control problem, for all π_0.*

Proof. From (13), we have, under any wide sense admissible control $\{Z_n\}$,

$$E[\hat{V}(\pi_n)] \leqslant E[\bar{k}(\pi_n, Z_n)] - \gamma + E[\hat{V}(\pi_{n+1})], \quad n \geqslant 0.$$

Therefore,

$$\gamma \leqslant \frac{1}{n} \sum_{m=0}^{n-1} E[\bar{k}(\pi_n, Z_n)] + \frac{E[\hat{V}(\pi_n)] - \hat{V}(\pi_0)}{n}.$$

Then by (†),

$$\gamma \leqslant \liminf_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} E[\bar{k}(\pi_m, Z_m)]. \tag{15}$$

Next, let $Z_n = v(\pi_n)$, $n \geqslant 0$, where $v : \mathscr{P}_0(S) \to U$ is such that $v(\pi)$ attains the minimum on the r.h.s. of (14). (This is always possible by a standard measurable selection theorem (Wagner, 1977).) Argue as above to obtain, *under this choice of control*,

$$\gamma \geqslant \limsup_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} E[\bar{k}(\pi_m, Z_m)]. \tag{16}$$

Together with (15), this implies the result, along with the fact that the lim sup in (16) is, in fact, a limit and $Z_n = v(\pi_n)$, $n \geqslant 0$, is an optimal control process. □

We summarize the above observations as the following 'dynamic programming principle':

Theorem 3.1. *There exist* $\hat{V}, \tilde{V} : \mathscr{P}_0(S) \to \mathbb{R}$ *such that* (13) *and* (14) *hold and* $Z_n = v(\pi_n)$, $n \geqslant 0$, *for* $v(\cdot)$ *as above, is optimal. Conversely, if* $(\pi_n, Z_n = v(\pi_n))$, $n \geqslant 0$, *for some measurable* $v(\cdot)$ *is a stationary optimal solution with the law of* $\pi_n = \mu$, *then a.s. with respect to* μ, *equality holds in* (13) *and* $v(\pi)$ *attains the minimum on the r.h.s. of* (13). *Furthermore*, (16) *holds with equality and with* 'lim' *in place of* 'lim sup'.

Proof. The first claim is already contained in the foregoing. The second and third claims follow by standard arguments as in, e.g., the proof of Theorem 4.2 of Borkar (2000a). □.

4. The continuous time problem

In this section, we start afresh with notation. Thus, let $X(\cdot) = [X_1(\cdot), \ldots, X_d(\cdot)]^T$ be an $\mathbb{R}^d$-valued controlled diffusion controlled by a U-valued (U a compact metric space) control process $Z(\cdot)$, with an associated $\mathbb{R}^r$-valued observation process $Y(\cdot)$, defined on a probability space $(\Omega, \mathscr{F}, P)$ and described by the stochastic differential equations

$$X(t) = X_0 + \int_0^t m(X(s), Z(s))\,ds + \int_0^t \sigma(X(s))\,dB_1(s), \tag{17}$$

$$Y(t) = \int_0^t h(X(s))\,ds + B_2(t), \tag{18}$$

where:

- $m(\cdot,\cdot) : \mathbb{R}^d \times U \to \mathbb{R}^d$ is bounded continuous and Lipschitz in its first argument uniformly w.r.t. the second,
- $\sigma(\cdot) : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is bounded Lipschitz with the least eigenvalue of $\sigma(\cdot)\sigma(\cdot)^T$ uniformly bounded away from zero,

- X_0 has a prescribed law $\pi_0 \in \mathscr{P}(\mathbb{R}^d)$,
- $B_1(\cdot), B_2(\cdot)$ are independent, resp., d- and r-dimensional standard Brownian motions such that $(B_1(\cdot), B_2(\cdot), X_0)$ are independent,
- $Z(\cdot)$ is a U-valued control process adapted to the filtration $\sigma(Y(s), s \leqslant t), t \geqslant 0$,
- $h : \mathbb{R}^d \to \mathbb{R}^r$ is bounded continuous and twice continuously differentiable with bounded first and second partial derivatives.

We call such a $Z(\cdot)$ a strict sense admissible control. Let $\mathscr{F}_t$= the right-continuous completion of $\sigma(X(s), Y(s), Z(s), B_1(s), B_2(s), s \leqslant t)$ for $t \geqslant 0$. W.l.o.g., let $\mathscr{F} = \bigvee_t \mathscr{F}_t$. Let P_0 be a probability measure on $(\Omega, \mathscr{F})$ defined as follows: If P_t (resp., P_{0t}) denotes the restriction of P (resp., P_0) to $\mathscr{F}_t$ for $t \geqslant 0$, then $P_t \ll P_{0t}$ with

$$\Lambda(t) \triangleq \frac{\mathrm{d}P_t}{\mathrm{d}P_{0t}} = \mathrm{e}^{\int_0^t \langle h(X(s)), \mathrm{d}Y(s)\rangle - (1/2)\int_0^t \|h(X(s))\|^2 \, \mathrm{d}s}.$$

By Novikov's criterion (Ikeda and Watanabe, 1981, p. 142), this is a nonnegative martingale with mean one and therefore a legal family of Radon–Nikodym derivatives. By Girsanov's theorem (Ikeda and Watanabe, 1981, p. 178), it follows that under P_0, $Y(\cdot)$ is a Brownian motion independent of $(B_1(\cdot), X_0)$. Following Fleming and Pardoux (1982), we call $Z(\cdot)$ wide sense admissible if under P_0, for any $t \geqslant 0$, $Y(t+\cdot) - Y(t)$ is independent of $\{B_1(\cdot), X_0, Y(s), Z(s), s \leqslant t\}$. This is a larger class of control processes and subsumes the class of strict sense admissible controls introduced above. As in Fleming and Pardoux (1982) and much of the subsequent literature on control of partially observed diffusions, we shall work with wide sense admissible controls, referring the reader to Fleming and Pardoux (1982) for justification. Our aim then will be to minimize over all wide sense admissible controls the ergodic cost

$$\limsup_{t\to\infty} \frac{1}{t} \int_0^t E[k(X(s), Z(s))] \, \mathrm{d}s \tag{19}$$

for a prescribed $k \in C_b(\mathbb{R}^d \times U)$.

We make two further qualifications to the above formulation. The first is that we shall be using the relaxed control framework. That is, we suppose that $U = \mathscr{P}(\tilde{U})$ for a compact Polish space $\tilde{U}$ and $m_i(\cdot,\cdot), k(\cdot,\cdot)$ are of the form

$$m_i(x,u) = \int \tilde{m}_i(x,y)u(\mathrm{d}y),$$

$$k(x,u) = \int \tilde{k}(x,y)u(\mathrm{d}y),$$

for $x \in \mathbb{R}^d, u \in U$, where $m_i, k : \mathbb{R}^d \times U \to \mathbb{R}$ are bounded continuous, with the m_i's Lipschitz in the first argument uniformly w.r.t. the second. The second qualification is that we shall be using the weak formulation of this problem, i.e., we do not work with a fixed probability space, but consider the optimization problem over all probability spaces supporting the processes that fit the above description. For a definition and elaboration of this idea, see Borkar (1989, Chapter I).

As in Fleming and Pardoux (1982), we consider the equivalent 'separated control problem' of controlling the $\mathscr{P}(\mathbb{R}^d)$-valued controlled Markov process $\{\pi_t\}$ of regular conditional laws of $X(t)$ given the right-continuous completions of $\sigma(Y(s), Z(s), s \leqslant t)$, for $t \geqslant 0$. Letting $\pi(f)$ denote $\int f \, d\pi$ for bounded measurable $f : \mathbb{R}^d \to \mathbb{R}$ and $\pi \in \mathscr{P}(\mathbb{R}^d)$, the evolution of $\{\pi_t\}$ is given by the Fujisaki–Kallianpur–Kunita equations (Borkar, 1989)

$$\pi_t(f) = \pi_0(f) + \int_0^t \pi_s(\mathscr{L}_{Z(s)}(f)) \, ds + \int_0^t \langle \pi_s(hf) - \pi_s(f)\pi_s(h), d\hat{Y}(s) \rangle, \tag{20}$$

where

- $f \in C_b^2(\mathbb{R}^d)$ (= the space of bounded twice continuously differentiable maps $\mathbb{R}^d \to \mathbb{R}$ with bounded first and second partial derivatives),
- the operator $\mathscr{L}$ is defined by

$$\mathscr{L}_u(f)(x) = \sum_i m_i(x,u) \frac{\partial f}{\partial x_i}(x) + \frac{1}{2} \sum_{i,j,k} \sigma_{ik}(x)\sigma_{jk}(x) \frac{\partial^2 f}{\partial x_i \partial x_j}(x),$$

- the 'innovations process' $\hat{Y}(t) = Y(t) - \int_0^t \pi_s(h) \, ds$ is an r-dimensional standard Brownian motion under P, independent of $(X_0, B_1(\cdot))$.

See Borkar (1989, Chapter V), for well-posedness issues concerning (20). Our differentiability conditions on $h(\cdot)$ play an important role here. Cost (19) is equivalently written as

$$\limsup_{t \to \infty} \frac{1}{t} \int_0^t E[\pi_s(k(\cdot, Z(s)))] \, ds. \tag{21}$$

The stochastic Liapunov condition we assume in the continuous time case can be stated as follows: There exist $g \in C(\mathbb{R}^d)$, $\mathscr{V} \in C^2(\mathbb{R}^d)$ such that $g \geqslant 0$, $\lim_{\|x\| \to \infty} g(x) = \lim_{\|x\| \to \infty} \mathscr{V}(x) = \infty$ and

$$\mathscr{L}_u \mathscr{V}(x) \leqslant -g(x) + CI_B(x), \tag{22}$$

for C, B as before. Define $\mathscr{P}_0(\mathbb{R}^d) = \{\pi \in \mathscr{P}(\mathbb{R}^d) : \pi(\mathscr{V}) < \infty\}$. Arguing as in the discrete case, we have for $t > s$,

$$\begin{aligned} E\left[\int \mathscr{V} \, d\pi_t\right] &= E[\mathscr{V}(X_t)] \\ &\leqslant E[\mathscr{V}(X_s)] + \text{a constant} \\ &= E\left[\int \mathscr{V} \, d\pi_s\right] + \text{a constant}, \end{aligned}$$

which shows that $\pi_0 \in \mathscr{P}_0(\mathbb{R}^d) \Rightarrow \pi_t \in \mathscr{P}_0(\mathbb{R}^d)$ a.s. for all $t \geqslant 0$. We also assume that, under any wide sense admissible control,

$$\lim_{t\to\infty} \frac{E[\mathscr{V}(X(t))]}{t} = 0. \tag{23}$$

As before, we shall adopt the 'vanishing discount' argument. Thus we introduce the discounted cost under a wide sense admissible $\{Z(\cdot)\}$ and initial law π_0 to be

$$\begin{aligned} J_\alpha(\{Z(\cdot)\}, \pi_0) &= E\left[\int_0^\infty \mathrm{e}^{-\alpha t} k(X(t), Z(t))\,\mathrm{d}t\right] \\ &= E\left[\int_0^\infty \mathrm{e}^{-\alpha t} \pi_t(k(\cdot, Z(t)))\,\mathrm{d}t\right], \end{aligned}$$

where $\alpha > 0$ is the discount factor and the law of $X(0)$ is π_0. Define the associated value function

$$V_\alpha(\pi) = \inf J_\alpha(\{Z(\cdot)\}, \pi),$$

where the infimum is over all wide sense admissible controls.

Theorem 4.1. *For any* $t > 0$,

$$V_\alpha(\pi) = \inf\left[\int_0^t \mathrm{e}^{-\alpha s} \pi_s(k(\cdot, Z(s)))\,\mathrm{d}s + \mathrm{e}^{-\alpha t} V_\alpha(\pi_t)/\pi_0 = \pi\right], \tag{24}$$

where the infimum is over all wide sense admissible controls on $[0, t]$.

Corollary 4.1. $V_\alpha(\pi_t) - \int_0^t \mathrm{e}^{-\alpha s}\pi_s(k(\cdot, Z(s)))\,\mathrm{d}s$, $t \geqslant 0$, *is a submartingale under any wide sense admissible* $\{Z(\cdot)\}$ *and is a martingale if and only if* $(\pi_t, Z(t))$ *is an optimal pair.*

These are immediate from Borkar (1989, Theorem 2.3, p. 120). Once again, we need to compare $V_\alpha(\cdot)$ for two different values of its arguments, whence we need to construct on a common probability space two processes with a common control process, but with different initial laws (say, $\hat{\pi}$ and $\tilde{\pi}$). This proceeds exactly along the lines of the discrete case and has been introduced already in Borkar (1999). We shall briefly sketch the details.

Let $\mathscr{U}$ denote the space of measurable maps $\mu(\cdot) : [0, \infty) \to U$, with the coarsest topology that renders continuous the maps $\mu(\cdot) \in \mathscr{U} \to \int_0^T x(t) \int f\,\mathrm{d}\mu(t)\,\mathrm{d}t$ for all $T > 0$, $x(\cdot) \in L_2[0, T]$ and $f \in C_\mathrm{b}(U)$. This is known to be compact Polish. Let $\tilde{\Phi} \subset \mathscr{P}(C([0, \infty); \mathbb{R}^r) \times \mathscr{U})$ be defined as the set of laws of $(Y(t), Z(t)), t \in [0, \infty)$, where $Y(\cdot)$ is a standard Brownian motion in $\mathbb{R}^r$ and for each $t \in [0, \infty)$, $Y(t + \cdot) - Y(t)$ is independent of $\{Y(s), Z(s), s \leqslant t\}$. Since the marginal of $Y(\cdot)$ is fixed and $\mathscr{U}$ compact, $\tilde{\Phi}$ is tight, hence relatively compact by Prohorov's theorem. Since independence is preserved under weak convergence of probability measures, it follows that it is also closed, hence compact. Note that a wide sense admissible control on $[0, \infty)$ may be identified with an element of $\tilde{\Phi}$ and vice versa. Let $\zeta \in \tilde{\Phi}$.

Define

$$\Omega = C([0,\infty);\mathbb{R}^d) \times C([0,\infty);\mathbb{R}^d) \times C([0,\infty);\mathbb{R}^r) \times \mathscr{U}$$
$$\times C([0,\infty);\mathbb{R}^r) \times \mathbb{R}^d \times \mathbb{R}^d,$$

with the product Borel σ-field $\mathscr{G}$. Let κ^d (resp., κ^r) denote the law of d-dimensional (resp., r-dimensional) standard Brownian motion. On $(\Omega,\mathscr{G})$, define a probability measure P_0 by

$$\mathscr{P}(\mathrm{d}\hat{b} \times \mathrm{d}\tilde{b} \times \mathrm{d}\hat{y} \times \mathrm{d}z \times \mathrm{d}\tilde{y} \times \mathrm{d}\hat{x} \times \mathrm{d}\tilde{x})$$
$$= \kappa^d(\mathrm{d}\hat{b})\kappa^d(\mathrm{d}\tilde{b})\zeta(\mathrm{d}\hat{y}\,\mathrm{d}z)\kappa^r(\mathrm{d}\tilde{y})\hat{\pi}(\mathrm{d}\hat{x})\tilde{\pi}(\mathrm{d}\tilde{x}).$$

Define on $(\Omega,\mathscr{G},P_0)$ the Ω-valued random variable $(\hat{B}(\cdot),\hat{B}(\cdot),\hat{Y}(\cdot),Z(\cdot),\tilde{Y}(\cdot),\hat{X}_0,\tilde{X}_0)$ canonically. Define $\hat{X}(\cdot)$ (resp., $\tilde{X}(\cdot)$) by (17) with $\hat{X}_0,\hat{B}(\cdot)$ (resp., $\tilde{X}_0,\tilde{B}(\cdot)$) in place of $X_0,B_1(\cdot)$. Change the probability measure on $(\Omega,\mathscr{G})$ to P as follows: If P_{0t} (resp., P_t) are restrictions of P_0 (resp., P) to

$$\sigma((b_1(s),b_2(s),y_1(s),z(s),y_2(s),x_0,\bar{x}_0),s \leqslant t)$$

for $(b_1(\cdot),b_2(\cdot),y_1(\cdot),z(\cdot),y_2(\cdot),x_0,\bar{x}_0) \in \Omega$, $t \geqslant 0$, then $P_t \ll P_{0t}$ with

$$\frac{\mathrm{d}P_t}{\mathrm{d}P_{0t}} = \exp\left(\int_0^t (\langle h(\hat{X}(s)),\mathrm{d}\hat{Y}(s)\rangle + \langle h(\tilde{X}(s)),\mathrm{d}\tilde{Y}(s)\rangle)\right.$$
$$\left. - \frac{1}{2}\int_0^t (\|h(\hat{X}(s))\|^2 + \|h(\tilde{X}(s))\|^2)\,\mathrm{d}s\right), \qquad t \geqslant 0.$$

Then on $(\Omega,\mathscr{G},\mathscr{P})$, $\hat{X}(\cdot),\tilde{X}(\cdot)$ are processes governed by a common wide sense admissible control $Z(\cdot)$. (In Borkar (1999), one had $\hat{B}(\cdot)=\tilde{B}(\cdot)$. This is not necessary in the present approach.)

With this construction in place, we have, exactly as in the discrete case and in Borkar (1999),

$$|V_\alpha(\tilde{\pi}) - V_\alpha(\hat{\pi})| \leqslant \sup|J_\alpha(\{Z(\cdot)\},\tilde{\pi}) - J_\alpha(\{Z(\cdot)\},\hat{\pi})|, \tag{25}$$

where the infimum is over all wide sense admissible controls $\{Z(\cdot)\}$.

To show that this remains bounded as $\alpha \to 0$, we need to adapt the arguments of the discrete case, which we do via an embedded discrete time problem described next.

5. The embedded problem

Our embedded discrete time process will simply be $X(n)$, $n \geqslant 0$. To exhibit this as a controlled Markov process in $S = \mathbb{R}^d$ that fits the description of Sections 2 and 3, we start with some preliminaries. Let $\hat{\Phi} \subset \mathscr{P}(C([0,1];\mathbb{R}^r) \times \mathscr{U})$ be defined as the set of laws of $(Y(t),Z(t)), t \in [0,1]$, where $Y(\cdot)$ is a standard Brownian motion in $\mathbb{R}^r$ and for each $t \in [0,1]$, $Y(t+\cdot) - Y(t)$ is independent of $\{Y(s),Z(s),s \leqslant t\}$. As before, a wide sense admissible control on $[0,1]$ may be identified with an element of $\hat{\Phi}$ and vice versa. $\hat{\Phi}$ will then serve as the compact metric control space for our process $\{X(n)\}$, the compactness being established as above. (See also Borkar, 1989, Section V.3.) The controlled transition kernel $(x,u)\in\mathbb{R}^d\times\hat{\Phi}\to\hat{p}(x,u,\mathrm{d}y)\in\mathscr{P}(\mathbb{R}^d)$ is given by: $\hat{p}(x,u,\mathrm{d}y)$ is the law of $X(1)$ when $X(0)=x$ and the wide sense admissible control u is used on $[0,1]$. Define, for $x\in\mathbb{R}^d$, $u\in\hat{\Phi}$,

$$\hat{k}_\alpha(x,u)=E\left[\int_0^1 \mathrm{e}^{-\alpha s}k(X(s),\tilde{Z}(s))\,\mathrm{d}s/X(0)=x\right],$$

$$\hat{g}(x)=\inf_{\hat{\Phi}} E\left[\int_0^1 g(X(s))\,\mathrm{d}s/X(0)=x\right],$$

$$\hat{\alpha}=\mathrm{e}^{-\alpha},$$

where $\tilde{Z}(\cdot)$ is the actual realization of u. Then taking the control sequence $u_n\triangleq(Z(n+t),\ t\in[0,1])$ and running cost function $k'(\cdot,\cdot)\triangleq\hat{k}_\alpha(\cdot,\cdot)$, for the discrete time problem of Sections 2 and 3, we have

$$J_\alpha(\{Z(\cdot)\},\pi_0)=J_{\hat{\alpha}}(\{u_n\},\pi_0), \tag{26}$$

where the r.h.s. is in the sense of Sections 2 and 3.

Lemma 5.1. *The maps $\hat{p}(\cdot,\cdot,\mathrm{d}y)$ and $\hat{k}_{\cdot}(\cdot,\cdot)$ are continuous. Furthermore, $\hat{p}(x,u,\mathrm{d}y)$ has a strictly positive, continuous density w.r.t. the Lebesgue measure, which is bounded from above and below by expressions of the form $c_1\exp(-c_2\|y-x\|^2)$ for suitable constants $c_1,c_2>0$.*

Proof. The first claim follows from the fact that the law of $X(\cdot)$ varies continuously with (x,u). For fixed x, the continuous dependence on u is proved in Borkar (1989, Section V.3, pp. 122–132). (This is, in fact, the standard proof for existence of optimal wide sense admissible controls for the finite horizon problem.) The joint continuity in (x,u) requires only a minor modification of that argument. The second claim would follow for the special case of $Z(t)=v(X(t),t), t\geqslant 0$, ($v(\cdot,\cdot):\mathbb{R}^d\times[0,\infty)\to U$ measurable) from standard p.d.e. theory for nondegenerate linear parabolic equations (see, e.g., Ladyzenskaja et al., 1968), the Gaussian bounds from above and below being from Aronson (1967). The general case is reduced to this by the results of Borkar (1986) and Gy'ongi (1986) which show that the one-dimensional marginals of $X(\cdot)$ as above can be mimicked by those of another process $X'(\cdot)$ also governed by (17), but with $Z(\cdot)$ replaced by $v(X'(\cdot),\cdot)$ for a suitable measurable map v. □

Note that the latter part of this claim allows us to verify the minorization condition (11) in this case.

Lemma 5.2. $\lim_{\|x\|\to\infty}\hat{g}(x)=\infty$.

Proof. From the developments of Borkar (1989, Section V.3), we know that the inf in the definition of $\hat{g}(x)$ is in fact a minimum. Thus, let $\|x_n\|\to\infty$ and let $X^n(\cdot),Z^n(\cdot)$ be such that $X^n(0)=x_n$, $Z^n(\cdot)$ is wide sense admissible, and $\hat{g}(x_n)=$

$E[\int_0^1 g(X^n(s))\,\mathrm{d}s]$. Standard moment criteria (e.g., Billingsley, 1968, p. 95) show that the laws of $X^n(\cdot) - x_n, n \geqslant 1$, remain tight and therefore relatively compact in $\mathscr{P}(C([0,1];\mathbb{R}^d))$. Let $X^{n(k)}(\cdot) - x_{n(k)}$ converge in law. By Skorohod's theorem (Borkar, 1995, p. 22), there exist processes $\bar{X}^k(\cdot)$ that agree with $X^{n(k)}(\cdot) - x_{n(k)}$ in law for $k \geqslant 1$, and converge in $C([0,1];\mathbb{R}^d)$ a.s. Then $\bar{X}^k(\cdot) + x_{n(k)} \to \infty$ a.s., leading to

$$E\left[\int_0^1 g(\bar{X}^k(s) + x_{n(k)})\,\mathrm{d}s\right] = \hat{g}(x_{n(k)}) \to \infty.$$

The claim follows. □

By (22) and the Ito formula,

$$E[\mathscr{V}(X(n+1))/\mathscr{F}_n] - \mathscr{V}(X(n)) \leqslant -\hat{g}(X(n)) + C, \quad n \geqslant 1.$$

Thus, $\mathscr{V}$ serves as a Liapunov function for $X(n), n \geqslant 1$, just as it did in Sections 2 and 3.

In view of the above observations and (26), we may argue as in Section 3 to conclude that $\bar{V}_\alpha(\cdot) \triangleq V_\alpha(\cdot) - V_\alpha(\pi^*)$ remains bounded as $\alpha \to 0$. Let $\alpha(n) \downarrow 0$ be a sequence such that $\alpha(n)V_{\alpha(n)}(\pi^*) \to \gamma$ for some $\gamma \in \mathbb{R}$. (This is possible because $\alpha V_\alpha(\pi^*)$ remains bounded as $\alpha \to 0$, as can be easily verified.) Let $\hat{V}(\cdot) = \limsup_{n\to\infty} \bar{V}_{\alpha(n)}(\cdot)$, $\tilde{V}(\cdot) = \liminf_{n\to\infty} \bar{V}_{\alpha(n)}(\cdot)$. Then, subtracting $V_\alpha(\pi^*)$ from both sides of (24) and taking lim sup, resp. lim inf on both sides, we have the following counterparts of (13) and (14):

$$\hat{V}(\pi) \leqslant \inf\left[\int_0^t (\pi_s(k(\cdot, Z(s))) - \gamma)\,\mathrm{d}s + \hat{V}(\pi_t)/\pi_0 = \pi\right], \tag{27}$$

$$\tilde{V}(\pi) \geqslant \inf\left[\int_0^t (\pi_s(k(\cdot, Z(s))) - \gamma)\,\mathrm{d}s + \tilde{V}(\pi_t)/\pi_0 = \pi\right], \tag{28}$$

the infimum in each case being over wide sense admissible controls. Arguing exactly as in the discrete case, we then have:

Theorem 5.1. *There exist $\hat{V}(\cdot), \tilde{V}(\cdot) : \mathscr{P}_0(\mathbb{R}^d) \to \mathbb{R}$ and $\gamma \in \mathbb{R}$ such that (27) and (28) hold for all $t > 0$ and γ is the optimal cost. Furthermore, if the pair $(\pi_t, Z(t))$ is such that $\tilde{V}(\pi_t) + \int_0^t (\pi_s(k(\cdot, Z(s))) - \gamma)\,\mathrm{d}s$, $t \geqslant 0$, is an $\{\mathscr{F}_t\}$-supermartingale, then it is optimal. Conversely, if $(\pi_t, Z(t))$ is a stationary optimal pair, then the process $\hat{V}(\pi_t) + \int_0^t (\pi_s(k(\cdot, Z(s))) - \gamma)\,\mathrm{d}s$, $t \geqslant 0$, is an $\{\mathscr{F}_t\}$-martingale.*

This follows exactly as in the discrete case. It may be noted that existence of a stationary pair as above is established in Bhatt and Borkar (1996).

Acknowledgements

Work partially supported by a Grant for 'Nonlinear Studies' from the Indian Space Research Organization and the Defence Research and Development Organization, Government of India, administered through the Indian Institute of Science, Bangalore.

References

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

Athreya, K.B., Ney, P., 1978. A new approach to the limit theory of recurrent Markov chains. Trans. Amer. Math. Soc. 245, 493–501.

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

Billingsley, P., 1968. Convergence of Probability Measures. Wiley, New York.

Borkar, V.S., 1986. A remark on attainable distributions of controlled diffusions. Stochastics 18, 17–23.

Borkar, V.S., 1989. Optimal Control of Diffusion Processes. Pitman Research Notes in Maths, No. 203. Longman Scientific and Technical, Harlow, England.

Borkar, V.S., 1993. On white noise representations in stochastic realization theory. SIAM J. Control Optim. 31, 1093–1102.

Borkar, V.S., 1995. Probability Theory: An Advanced Course. Springer, New York.

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

Borkar, V.S., 2000a. Average cost dynamic programming equations for controlled Markov chains with partial observations. SIAM J. Control Optim. 39, 673–681.

Borkar, V.S. 2000b. The value function in ergodic control of diffusion processes with partial observations II. Appl. Math. 27, 455–464. (Correction note in Appl. Math., 28, 2001, 245–246).

Davis, M.H.A., Varaiya, P.P., 1973. Dynamic programming conditions for partially observable stochastic systems. SIAM J. Control Optim. 11, 226–261.

Fleming, W.H., Pardoux, E., 1982. Optimal control of partially observed diffusions. SIAM J. Control Optim. 20, 261–285.

Gy'ongi, I., 1986. Mimicking one dimensional marginal distributions of processes having an Ito differential. Prob. Theory Related Fields 71, 501–516.

Hernández-Lerma, O., Lasserre, J.B., 1996. Discrete-Time Markov Control Processes: Basic Optimality Criteria. Springer, New York.

Hernández-Lerma, O., Lasserre, J.B., 1999. Further Topics on Discrete-Time Markov Control Processes. Springer, New York.

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

Ladyzenskaja, O.A., Ural'seva, N.N., Solonnikov, V.A., 1968. Linear and Quasilinear Equations of Parabolic Type. Transactions of Mathematical Monographs 23. A.M.S., Providence, RI.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer, London.

Nummelin, E., 1978. A splitting technique for Harris recurrent chains. Z. Wahrscheinlichkeitstheorie Verw. Geb. 43, 309–318.

Rishel, R., 1970. Necessary and sufficient dynamic programming conditions for continuous time stochastic control. SIAM J. Control Optim. 8, 559–571.

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

Wagner, D.H., 1977. Survey of measurable selection theorems. SIAM J. Control Optim. 15, 859–903.