A further remark on dynamic programming for partially observed Markov processes

Vivek S. Borkar[a,*,1], Amarjit Budhiraja[b]

[a] *School of Technology and Computer Science, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*
[b] *Department of Statistics, University of North Carolina, Chapel Hill, NC 27599-3260, USA*

Abstract

In (Stochastic Process. Appl. 103 (2003) 293), a pair of dynamic programming *inequalities* were derived for the 'separated' ergodic control problem for partially observed Markov processes, using the 'vanishing discount' argument. In this note, we strengthen these results to derive a single dynamic programming equation for the same.
© 2004 Elsevier B.V. All rights reserved.

Keywords: Controlled Markov processes; Dynamic programming; Partial observations; Ergodic cost; Vanishing discount; Pseudo-atom

1. Introduction

In Borkar (2003), one of us extended the 'vanishing discount' argument for dynamic programming of ergodic control of Markov chains on a discrete state space to the problem of ergodic control of partially observed Markov processes in a finite-dimensional Euclidean space. Specifically, a pair of dynamic programming *inequalities* is derived. These in turn yield necessary/sufficient conditions for optimality. The aim of the present note is to show that under some additional conditions, one can in fact replace these by a single dynamic programming equation which is the exact counterpart of the corresponding equation for completely observed Markov chains on a discrete state space.

---

* Corresponding author. Tel.: +91-2222804614; fax: +91-2222804610.
E-mail addresses: borkar@tifr.res.in (V.S. Borkar), budhiraj@email.unc.edu (A. Budhiraja).

The study of dynamic programming equation for ergodic control under partial obser-
vations goes back to Platzman (1980). This and some of the subsequent works (e.g.,
Fernández-Gaucherand et al. (1990)) treat the discrete state space case under somewhat
nontransparent conditions, e.g., involving reachable sets of probability measures for the
nonlinear filter. Also, discreteness of the state space is crucial. (Fernández-Gaucherand
et al. (1990) also includes a comparison of various results available till that point.)
A major development was (Stettner, 1993; Runggaldier and Stettner, 1994), where the
dynamic programming equations were derived under more transparent conditions. But
even those conditions (see, e.g., assumption (A.6), Runggaldier and Stettner, 1994,
p. 151) are rather restrictive. To illustrate this point, consider an (uncontrolled) system

$$X_{n+1} = aX_n + \xi_{n+1}, \quad n \geqslant 0,$$

where $|a| < 1$ (for stability) and $\{\xi_n\}$ are i.i.d. normal with zero mean and unit vari-
ance. Let $\varphi(\cdot)$ denote the Gaussian probability density with zero mean and unit vari-
ance. One of the implications of the aforementioned condition (A.6) from Runggaldier
and Stettner, 1994, (if true) would be that for a suitable $\Delta > 0$,

$$\sup_{x,x'} \frac{\int_{[-1,1]} \varphi(y - ax)\,\mathrm{d}y}{\int_{[-1,1]} \varphi(y - ax')\,\mathrm{d}y} \geqslant \Delta.$$

This is clearly impossible, as the l.h.s. $\to 0$ as $\|x\| \to \infty$ while x' is kept fixed. Thus
it seems that this condition is reasonable only for a compact state space.

For the finite state case, these results were improved upon in Borkar (2000), which
managed to derive these equations in a more general set-up. (The 'aperiodicity' condi-
tion in Borkar (2003) can in fact be dropped with a slight modification of the proof.)
These results used a coupling argument as a key ingredient to show the bounded-
ness of renormalized discounted value functions in the vanishing discount limit. Thus
it depended critically on the finiteness of the state space. The article Borkar (2003)
mentioned above managed to extend this to a general state space by using the Athreya–
Ney–Nummelin construction of a 'pseudo-atom' described in Meyn and Tweedie (1993),
Chapter 5. In a general state space, however, boundedness of the renormalized dis-
counted value functions is not enough, one also needs their equicontinuity to enable an
application of the Arzela–Ascoli theorem. In absence of an equicontinuity result, Borkar
(2003) could establish only a *pair of dynamic programming inequalities* rather than
a single dynamic programming equation. The present note fills in this lacuna by pro-
viding this last step, so that the result is now complete. To compare with Runggaldier
and Stettner (1994) again, while we do assume a density w.r.t. an underlying probabil-
ity measure for the controlled transition kernel (see, however, a possible generalization
pointed out at the end of Section 4), this appears less restrictive than assumption (A.6)
of Runggaldier and Stettner (1994).

The next section recalls the problem framework and the underlying assumptions.
Section 3 recalls the vanishing discount argument of Borkar (2003). Section 4 de-
scribes the key step in the vanishing discount limit that facilitates the above, viz., an
equicontinuity result for the renormalized discounted value function. This leads to the

dynamic programming equation we seek. Section 5 states the implications to ergodic control of partially observed diffusions along the lines of Borkar (2003).

2. The control problem

Let Polish spaces S, W, U denote respectively, 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 the Polish space '$\cdots$' 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 an $\eta \in \mathscr{P}(W)$ and $\varphi \in C_b(S \times U \times S \times W)$ 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

$$P(X_{n+1} \in A, Y_{n+1} \in A' / X_m, Z_m, Y_m, m \leqslant n)$$

$$= \int_{A'} \int_A \varphi(X_n, Z_n, z, y)\lambda(\mathrm{d}z)\eta(\mathrm{d}y) \tag{1}$$

for Borel $A \subset S$, $A' \subset W$. Let $\bar{\varphi}(x, u, z) \triangleq \int \varphi(x, u, z, y)\eta(\mathrm{d}y)$. Note that $\bar{\varphi}(x, u, z) > 0$.

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'

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

for a prescribed $k \in C_b(S \times U)$. As in Borkar (2003), we shall consider a relaxation of this problem that allows for a larger class of controls, the so-called *wide sense admissible* controls. (This is an adaptation to the discrete time set-up of a notion originally introduced in the context of controlled diffusions in Fleming and Pardoux (1982).) To define these, we consider $\{X_n\}$ as being generated by a recursion

$$X_{n+1} = f(X_n, Z_n, \xi_{n+1}), \quad n \geqslant 0, \tag{3}$$

where $\{\xi_n\}$ are i.i.d., uniformly distributed on $[0, 1]$. This is always possible on a possibly augmented probability space by the results of Borkar (1993). We may then view all the above processes as being realized on the 'canonical' path space $(\Omega, \mathscr{F}, P)$. (See Borkar, 2003, for a detailed construction.) Define a new probability measure P_0 on $(\Omega, \mathscr{F})$ by: If P_n, P_{0n} denote restrictions of P, P_0, respectively, to $(\Omega, \mathscr{F}_n)$,

$\mathscr{F}_n \triangleq \sigma(X_m, Y_m, Z_m, \xi_m, m \leqslant n)$, $n \geqslant 0$, then $P_n \ll P_{0n}$ with

$$\Lambda_n \triangleq \frac{\mathrm{d}P_n}{\mathrm{d}P_{0n}} = \prod_{m=0}^{n-1} \frac{\varphi(X_m, Z_m, X_{m+1}, Y_{m+1})}{\bar{\varphi}(X_m, Z_m, X_{m+1})}, \quad n \geqslant 0. \tag{4}$$

We assume that

$$\int \bar{\varphi}(x, u, x') \left(\frac{\varphi(x, u, x', y)}{\bar{\varphi}(x, u, x')} \right)^{1+\varepsilon_0} \lambda(\mathrm{d}x')\eta(\mathrm{d}y) < \infty \tag{5}$$

for some $\varepsilon_0 > 0$. This ensures the uniform integrability of the ratio above w.r.t. $\bar{\varphi}(x, u, x')\lambda(\mathrm{d}x')\eta(\mathrm{d}y)$, a fact we use later. W.l.o.g., we may suppose that $\mathscr{F} = \vee_n \mathscr{F}_n$, whence the above defines a probability measure on $(\Omega, \mathscr{F})$ as desired. Note that if $\{Z_n\}$ is strict sense admissible then, under P_0, $\{Y_n\}$ are i.i.d. with law η, $(\{Y_m\}, \{\xi_m\}, X_0)$ is an independent family, and for each n, Y_n is independent of $X_m, m \leqslant n$; $Y_m, Z_m, m < n$. This motivates our definition of '*wide sense admissibility*'. The sequence $\{Z_n\}$ is said to be wide sense admissible if, under P_0, for each n, $(Z_m, Y_m, m \leqslant n)$ is independent of $\{\xi_m\}, X_0, \{Y_m, m > n\}$. Clearly this includes the strict sense admissible controls. Also, a wide sense admissible control is specified by specifying the joint law of $(\{Y_n\}, \{Z_n\})$. Thus we may identify the set of wide sense admissible controls with a subset Θ of $\mathscr{P}(U^\infty \times W^\infty)$ such that if $\Psi \in \Theta$ is the law of $(\{Z_n\}, \{Y_n\})$, then $\{Y_n\}$ are i.i.d. with law η and for each n, $Y_m, m > n$, is independent of $(Z_m, Y_m, m \leqslant n)$. Since both these conditions are preserved under convergence in $\mathscr{P}(U^\infty \times W^\infty)$, Θ is a closed set. Since U and therefore U^∞ is compact and the marginal on $\mathscr{P}(W^\infty)$ of $\Psi \in \mathscr{P}(U^\infty \times W^\infty)$ is fixed at η^∞, Θ is also tight and hence compact.

Letting π_n denote the regular conditional law of X_n given Y_m, Z_m, $m \leqslant n$, for $n \geqslant 0$, $\{\pi_n\}$ is given recursively by the nonlinear filter

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

$$\triangleq F(\pi_n, Z_n, Y_{n+1}), \quad n \geqslant 0. \tag{6}$$

Also, (2) can be written as

$$\limsup_{n \to \infty} \frac{1}{n} \sum_{m=0}^{n-1} E[\bar{k}(\pi_m, Z_m)], \tag{7}$$

where $\bar{k}(\mu, u) \triangleq \int k(x, u)\mu(\mathrm{d}x)$ for $\mu \in \mathscr{P}(S)$, $u \in U$. This allows us to consider the equivalent, so-called 'separated' control problem of controlling the $\mathscr{P}(S)$-valued controlled Markov process $\{\pi_n\}$ described by (6) over wide sense admissible $\{Z_n\}$ with the objective of minimizing (7).

We make the following assumption: There exist $\varepsilon_0 > 0$, $\mathscr{V} \in C(S)$ satisfying: $\lim_{\|x\| \to \infty} \mathscr{V}(x) = \infty$, and under any wide sense admissible $\{Z_n\}$,

$$E[\mathscr{V}(X_{n+1})|\mathscr{F}_n] - \mathscr{V}(X_n) \leqslant -\varepsilon_0 + CI_B(X_n), \tag{8}$$

where $C > 0$ and $B \triangleq \{x \in S : \|x\| \leqslant R\}$ for some $R > 0$. We strengthen this by further assuming that there exists $\hat{\mathscr{V}} \in C(S)$ satisfying $\lim_{\|x\| \to \infty} \hat{\mathscr{V}}(x) = \infty$, and under any

wide sense admissible $\{Z_n\}$,

$$E[\hat{\mathscr{V}}(X_{n+1})|\mathscr{F}_n] - \hat{\mathscr{V}}(X_n) \leqslant -\mathscr{V}(X_n) + \hat{C}I_{\hat{B}}(X_n), \tag{9}$$

$$\limsup_{n \to \infty} \frac{E[\hat{\mathscr{V}}(X_n)]}{n} = 0, \tag{10}$$

where $\hat{C} > 0$ and $\hat{B} \triangleq \{x \in S : \|x\| \leqslant \hat{R}\}$ for some $\hat{R} > 0$.

Remark. (i) Note that this is slightly more restrictive than the 'Liapunov condition' (8) assumed in Borkar (2003). (8) is the usual 'stochastic Liapunov' condition that ensures bounded mean hitting time of B. It does not, however, guarantee anything about the higher moments of this hitting time. Condition (9) ensures the square-integrability of this stopping time and thereby helps establish the uniform integrability of the 'coupling time' we discuss later. The condition is weaker than geometric ergodicity which ensures finiteness of certain exponential moments of the first hitting time of B (Meyn and Tweedie, 1993, Chapter 15). This would be true, e.g., if we replace (8), (9) with a single condition (9) with $a\hat{\mathscr{V}}(\cdot)(a > 0)$ replacing $\mathscr{V}$ on the right. See Tuominen and Tweedie (1994) for some related developments and additional insights. As an example, consider the scalar system

$$X_{n+1} = aX_n + G(X_n, Z_n) + \xi_{n+1}, \quad n \geqslant 0,$$

where $|a| < 1$, $\{\xi_n\}$ are i.i.d. normal with zero mean and unit variance and G is bounded continuous. The choice $\mathscr{V}(x) = \hat{\mathscr{V}}(x) = |x| + 1$ will do (implying in fact the stronger geometric ergodicity).

(ii) It will become clear (cf. the Appendix) that the strict positivity condition on φ can be replaced by the requirement that the same hold on $B \cup \hat{B}$ only. In this context, note that $\bar{\varphi}(x, u, z) = 0$ implies $\varphi(x, u, z, y) = 0$ for a.e. y, hence for all y by continuity, thus the likelihood ratio of the latter w.r.t. the former is well-defined.

Let

$$\mathscr{P}_0(S) \triangleq \left\{ \pi \in \mathscr{P}(S) : \int \hat{\mathscr{V}} \, \mathrm{d}\pi < \infty \right\}.$$

As observed in Borkar (2003), $\pi_0 \in \mathscr{P}_0(S)$ implies $\pi_n \in \mathscr{P}_0(S)$ a.s. for all n, enabling us to view $\{\pi_n\}$ as a process in $\mathscr{P}_0(S)$.

We shall denote by

$$\phi : (\pi, u) \in \mathscr{P}_0(S) \times U \to \phi(\pi, u, \mathrm{d}\pi') \in \mathscr{P}(\mathscr{P}_0(S))$$

the transition probability kernel of the controlled Markov chain $\{\pi_n\}$, defined by

$$\int g(\pi')\phi(\pi, u, \mathrm{d}\pi') = \iint g(F(\pi, u, y))\varphi(x, u, S, y)\pi(\mathrm{d}x)\eta(\mathrm{d}y) \tag{11}$$

for $g \in C_b(\mathscr{P}_0(S))$, where we write $\int_S \varphi(x, u, z, y)\lambda(\mathrm{d}z)$ as $\varphi(x, u, S, y)$ with a slight abuse of notation.

3. The vanishing discount limit

As in Borkar (2003), we view the ergodic control problem as the vanishing discount limit of the discounted cost control problem that seeks to minimize

$$E\left[\sum_{m=0}^{\infty} \alpha^m k(X_m, Z_m)\right] = E\left[\sum_{m=0}^{\infty} \alpha^m \bar{k}(\pi_m, Z_m)\right], \qquad (12)$$

where $\alpha \in (0,1)$ is the discount factor. Define the discounted value function $V_\alpha(\pi)$ as the infimum of (12) over all wide sense admissible controls when $\pi_0 = \pi$. This satisfies the dynamic programming equation (Borkar, 1989, Chapter V)

$$V_\alpha(\pi) = \min_u \left[\bar{k}(\pi, u) + \alpha \int \phi(\pi, u, d\pi') V_\alpha(\pi')\right], \qquad \pi \in \mathscr{P}(S). \qquad (13)$$

The renormalized value function $\bar{V}_\alpha(\pi)$ is defined as $\bar{V}_\alpha(\pi) = V_\alpha(\pi) - V_\alpha(\pi^*)$ for a fixed $\pi^* \in \mathscr{P}(S)$. Then by a simple algebraic manipulation from (13), we obtain

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

$$\pi \in \mathscr{P}(S). \qquad (14)$$

In Borkar (2003), it is shown that $\bar{V}_\alpha(\pi)$ remains bounded $\forall \pi$ as $\alpha \to 1$. This allows one to pass to the lim sup, respectively, lim inf, as $\alpha \to 1$ along a subsequence $\{\alpha(n)\}$ in (14) to obtain a pair of 'dynamic programming inequalities' for the ergodic control problem, satisfied respectively, by $V^*(\cdot) \triangleq \limsup_{n\to\infty} \bar{V}_{\alpha(n)}(\cdot)$ and $V_*(\cdot) \triangleq \liminf_{n\to\infty} \bar{V}_{\alpha(n)}(\cdot)$. See Borkar (2003) for details.

A key step in this procedure is the following: Given a quintuplet $\{\hat{X}_n, \hat{Y}_n, Z_n, \hat{\xi}_n, \hat{\pi}_n\}$ (with $\hat{\pi}_0 = \pi$) as above with a wide sense admissible $\{Z_n\}$ and another initial distribution π' (say), we construct on a common probability space another quintuplet $\{\tilde{X}_n, \tilde{Y}_n, Z_n, \tilde{\xi}_n, \tilde{\pi}_n\}$ with a *common* control process $\{Z_n\}$ that is wide sense admissible for both, and with $\tilde{\pi}_0 = \pi'$. Furthermore, on a possibly larger filtered probability space, there exist copies $\{\tilde{X}_n^*, \hat{X}_n^*, Z_n^*\}$ of $\{\tilde{X}_n, \hat{X}_n, Z_n\}$ and a stopping time τ (see (A.6) in the Appendix for definition of τ) such that $E[\tau]$ satisfies a bound

$$E[\tau] < K\left(\int \mathscr{V} d\pi + \int \mathscr{V} d\pi'\right), \qquad (15)$$

for some $K > 0$, uniformly w.r.t. the choice of $\{Z_n\}$ and furthermore, the regular conditional laws of $\hat{X}_{\tau+m}^*, \tilde{X}_{\tau+m}^*, m \geqslant 1$, given $\mathscr{F}_\tau$ coincide, where $\{\mathscr{F}_t\}$ is the underlying filtration. The construction is given in Borkar (2003), however, for the sake of completeness we provide the details in the Appendix. For notational ease we will suppress $*$ from the superscript for the processes $\{\tilde{X}_n^*, \hat{X}_n^*, Z_n^*\}$. We need the following additional fact facilitated by (9): Suppose π, π' are Dirac measures at x, x' respectively.

Lemma 3.1. τ *is uniformly integrable as* x, x' *vary over a compact set and the control process varies over all wide sense admissible controls.*

Proof. Let $\hat{X}_0 = x$, $\tilde{X}_0 = x'$. Recall from Borkar (2003) (see also Appendix A below) that τ is in fact the first hitting time of the so called 'pseudo-atom'. As in Lemma 3.3 of Borkar (2003, p. 300) one can then show that

$$E[\tau] \leqslant K_1(\mathscr{V}(x) + \mathscr{V}(x')),$$

$$E\left[\sum_{m=0}^{\tau-1}(\mathscr{V}(\hat{X}_m) + \mathscr{V}(\tilde{X}_m))\right] \leqslant K_2(\hat{\mathscr{V}}(x) + \hat{\mathscr{V}}(x')),$$

for suitable constants $K_1, K_2 > 0$. Together these imply (see, e.g., Borkar, 1991, pp. 66–67) that

$$E[\tau^2] \leqslant K_3(\hat{\mathscr{V}}(x) + \hat{\mathscr{V}}(x')). \tag{16}$$

This implies the claim. $\quad\square$

Recall that $\hat{\pi}_0 = \pi$ and $\tilde{\pi}_0 = \pi'$. As in Borkar (2003), we then have

$$|\bar{V}_\alpha(\pi) - \bar{V}_\alpha(\pi')| = |V_\alpha(\pi) - V_\alpha(\pi')|$$

$$\leqslant \sup\left|\sum_m \alpha^m E[\bar{k}(\hat{\pi}_m, Z_m)] - \sum_m \alpha^m E[\bar{k}(\tilde{\pi}_m, Z_m)]\right|$$

$$\leqslant \sup\left|\sum_m \alpha^m E[k(\hat{X}_m, Z_m)] - \sum_m \alpha^m E[k(\tilde{X}_m, Z_m)]\right|$$

$$= \sup\left|E\left[\sum_{m=0}^{\tau} \alpha^m(k(\hat{X}_m, Z_m) - k(\tilde{X}_m, Z_m))\right]\right|$$

$$\leqslant \sup E\left[\left|\sum_{m=0}^{\tau} \alpha^m(k(\hat{X}_m, Z_m) - k(\tilde{X}_m, Z_m))\right|\right],$$

where the supremum throughout is over all wide sense admissible $\{Z_m\}$ and the second equality follows from the earlier observation regarding regular conditional laws of $\{\hat{X}_{\tau+m}\}, \{\tilde{X}_{\tau+m}\}$ given $\mathscr{F}_\tau$. Using the boundedness of k and (15), we conclude the boundedness of the l.h.s. uniformly in α as $\alpha \to 1$.

In the next section we stretch this argument a little further and show that $\bar{V}_\alpha, \alpha \in (0, 1)$, is in fact an equicontinuous family. Since it is pointwise bounded, the Arzela–Ascoli theorem implies that it is relatively compact in $C(\mathscr{P}_0(S))$.

4. Main results

From the above calculation, we have for $M \geqslant 1$,

$$|\bar{V}_\alpha(\pi) - \bar{V}_\alpha(\pi')| \leqslant \sup \left| E \left[\sum_{m=0}^{\tau} \alpha^m (k(\hat{X}_m, Z_m) - k(\tilde{X}_m, Z_m)) \right] \right|$$

$$\leqslant \sum_{m=0}^{M} \sup |E[k(\hat{X}_m, Z_m)] - E[k(\tilde{X}_m, Z_m)]|$$

$$+ K_4 \sup E[(\tau - M)^+], \tag{17}$$

for a suitable constant $K_4 > 0$. As before, each supremum is over all wide sense admissible controls. Fix $\delta > 0$ and take M large enough such that the second term on the right is less than $\delta/2$. This is possible by Lemma 3.1. In fact, (16) leads to

$$E[\tau^2] \leqslant K_3 \left(\int \hat{V}(x)\pi(\mathrm{d}x) + \int \hat{V}(x)\pi'(\mathrm{d}x) \right),$$

which ensures the required uniform integrability. We claim that the first term on the right in (17) can then be made smaller than $\delta/2$ if π, π' are close enough w.r.t. the Prohorov metric $\rho(\cdot, \cdot)$ on $\mathscr{P}(S)$ (Borkar, 1995, Chapter 2). For this purpose, consider *fixed* $\{Y_n\}, \{Z_n\}, \{\xi_n\}$ on $(\Omega, \mathscr{F}, P_0)$ and $\{X_n\}$ generated by (3) with the law of $X_0 = \pi$. Fix n, $0 \leqslant n \leqslant M$, and define

$$g(\pi, \Psi) \triangleq E_{\pi, \Psi}[k(X_n, Z_n)],$$

where $E_{\pi, \Psi}[\cdot]$ denotes expectation when the law of X_0 is π and the law of $(\{Z_m\}, \{Y_m\})$ (i.e., the wide sense admissible control) is $\Psi \in \Theta$.

Lemma 4.1. *g is bounded and continuous.*

Proof. $|g(\cdot)|$ is clearly bounded by any bound on $|k|$. For $j = 1, 2, \ldots, \infty$, let $\Psi_j \in \Theta$, $\pi_j \in \mathscr{P}(S)$, such that

$$\Psi_j \to \Psi_\infty,$$

$$\pi_j \to \pi_\infty,$$

as $j \uparrow \infty$. Let $(\{Z_n^j\}, \{Y_n^j\})$ be processes with laws Ψ_j, respectively, on some probability spaces. By enlarging the latter if necessary, construct on them also the independent random variables $\{\xi_n^j\}, X_n^0$ independent of $(\{Z_n^j\}, \{Y_n^j\})$, such that $\{\xi_n^j\}$ are i.i.d. uniformly distributed on $[0, 1]$ and X_n^0 has law π^j for $j = 1, 2, \ldots, \infty$. Let $\{X_n^j\}$ be specified by (3) for each j. By a standard monotone class argument, (1) translates into

$$E_0 \left[\left(h(X_{m+1}^j, Y_{m+1}^j) - \int \bar{\varphi}(X_m^j, Z_m^j, z) h(z, y) \lambda(\mathrm{d}z) \eta(\mathrm{d}y) \right) \right.$$

$$\left. g((X_0^j, Y_0^j, Z_0^j), \ldots, (X_m^j, Y_m^j, Z_m^j)) \Lambda_{m+1} \right] = 0, \tag{18}$$

for all $h \in C_b(S \times W), g \in C_b((S \times W \times U)^{m+1})$. From the continuity of φ, it follows that the laws of $(X_m^j, Y_m^j, Z_m^j, \ m \geqslant 0)$ remain tight as the law of X_0^j varies over any tight family (in particular, over $\{\pi^j, j \geqslant 1\}$) and the law of $(\{Z_m\}, \{Y_m\})$ varies over Θ. Also, Λ_m for each m remains uniformly integrable thanks to (5). Hence (18) is preserved under convergence in joint law of the processes involved. It follows that any limit point in law of the joint law of $(X_m^j, Y_m^j, Z_m^j, \ m \geqslant 0)$ as $j \to \infty$ satisfies (18), therefore (1), with the initial law $= \pi^\infty$. That is, the joint law of $(X_m^j, Y_m^j, Z_m^j, \ m \geqslant 0)$ converges to that of $(X_m^\infty, Y_m^\infty, Z_m^\infty, \ m \geqslant 0)$ as $j \to \infty$. Thus the joint law of $(X_m, Y_m, Z_m, \ m \geqslant 0)$ varies continuously with the law of $X_0, (\{Z_m\}, \{Y_m\})$, as the latter vary over $\mathscr{P}(S) \times \Theta$. The claim follows. □

Corollary 4.1. $g(\cdot, \Psi), \ \Psi \in \Theta,$ *is an equicontinuous family.*

In particular, this implies that we can make the first term on the r.h.s. of (17) less than $\delta/2$ by making $\rho(\pi, \pi')$ small enough. Thus we have:

Corollary 4.2. $\bar{V}_\alpha, \alpha \in (0,1),$ *is an equicontinuous, pointwise bounded family in* $C(\mathscr{P}_0(S))$.

Thus by the Arzela–Ascoli theorem, it is relatively compact in $C(\mathscr{P}_0(S))$. Note that the term $(1-\alpha)V_\alpha(\pi^*)$ is bounded uniformly in $\alpha \in (0,1)$. Thus we may take a subsequence $\alpha(n) \to 1$ in (14) such that

$$\bar{V}_{\alpha(n)} \to V^* \quad \text{in } C(\mathscr{P}_0(S)),$$

$$(1-\alpha(n))V_{\alpha(n)}(\pi^*) \to \beta \quad \text{in } \mathscr{R}.$$

From our continuity assumption on φ and Scheffe's theorem (Borkar, 1995, Theorem 2.3.3, p. 26), it follows that the map

$$(\pi, u) \to \left(\int \varphi(x, u, S, y) \pi(dx) \right) \eta(dy)$$

is continuous w.r.t. the total variation norm. A standard argument then shows that the expression in square brackets on the r.h.s. of (14) converges to

$$\left[\bar{k}(\pi, u) - \beta + \int \phi(\pi, u, d\pi') V^*(\pi') \right]$$

uniformly in u. Thus passing to the limit in (14), one has the dynamic programming equation for ergodic control of the separated control problem:

$$V^*(\pi) = \min_u \left[\bar{k}(\pi, u) - \beta + \int \phi(\pi, u, d\pi') V^*(\pi') \right]. \tag{19}$$

By a standard measurable selection theorem (Wagner, 1977), there exists a measurable $v : \mathscr{P}_0(S) \to U$ such that $v(\pi)$ attains the minimum on the r.h.s. for each π. Then we have:

Theorem 4.1. (i) *The dynamic programming equation* (19) *has a solution* $(V^*(\cdot), \beta) \in C(\mathscr{P}_0(S)) \times \mathscr{R}$ *where β is the optimal cost regardless of the initial law.*

(ii) $Z_n = v(\pi_n)$, $n \geqslant 0$, *for* $v(\cdot)$ *as above is optimal. Conversely, if* $(\pi_n, Z_n, \ n \geqslant 0)$ *is an optimal stationary pair, then*

$$Z_n \in \text{Arg min} \left(\bar{k}(\pi_n, \cdot) + \int \phi(\pi_n, \cdot, \mathrm{d}\pi') V^*(\pi') \right)$$

a.s. w.r.t. the law of π_n.

This is proved exactly as in Borkar (2000, pp. 680–1).

Remark. The following simple generalization of the foregoing is worth noting: Suppose we replace (1) by

$$P(X_{n+1} \in A, Y_{n+1} \in A' / X_m, Z_m, Y_m, \ m \leqslant n)$$

$$= \int_{A'} \int_A \kappa(X_n, Z_n, y, \mathrm{d}z) \eta(\mathrm{d}y)$$

for a suitable probability kernel

$$(x, z, y) \rightarrow \kappa(x, z, y, \mathrm{d}z') \in \mathscr{P}(S)$$

and assume that for $\bar{\kappa}(x, z, \mathrm{d}z') \triangleq \int \kappa(x, z, y, \mathrm{d}z') \eta(\mathrm{d}y)$, we have

$$\kappa(x, z, y, \mathrm{d}z') \ll \bar{\kappa}(x, z, \mathrm{d}z') \ \forall y$$

and

$$\zeta(x, z, y, z') \triangleq \frac{\mathrm{d}\kappa(x, z, y, \mathrm{d}z')}{\mathrm{d}\bar{\kappa}(x, z, \mathrm{d}z')}$$

is continuous in its arguments and satisfies:

$$\iint (\zeta(x, z, y, z'))^{1+\varepsilon_0} \bar{\kappa}(x, z, \mathrm{d}z') \eta(\mathrm{d}y) < \infty$$

for some $\varepsilon_0 > 0$. (This replaces (5).) The nonlinear filter then becomes

$$\pi_{n+1}(\mathrm{d}z) = \frac{\int \pi_n(\mathrm{d}x) \kappa(x, Z_n, Y_{n+1}, \mathrm{d}z)}{\int \int \pi_n(\mathrm{d}x) \kappa(x, Z_n, Y_{n+1}, \mathrm{d}z')}$$

$$\triangleq F(\pi_n, Z_n, Y_{n+1}), \quad n \geqslant 0.$$

The rest of the development goes through exactly as before, except for the pseudo-atom construction described in Appendix A. The latter needs an additional 'minorization condition' along the lines of Meyn and Tweedie (1993, p. 102).

5. Controlled diffusions

In this section we state without proof the implications to ergodic control of partially observed diffusions. The details exactly mimic those of Borkar (2003) and are therefore omitted. See Borkar (1989) for general background on controlled diffusions and partially observed controlled diffusions.

The controlled diffusion $X(\cdot) = [X_1(\cdot), \ldots, X_d(\cdot)]^{\mathrm{T}}$, assumed to be controlled by a $\mathscr{P}(U)$-valued 'relaxed' control process $Z(\cdot)$, and the associated $\mathscr{R}^r$-valued observation process $Y(\cdot)$, are described by the stochastic differential equations

$$X(t) = X_0 + \int_0^t m(X(s), Z(s))\, \mathrm{d}s + \int_0^t \sigma(X(s))\, \mathrm{d}B_1(s), \tag{20}$$

$$Y(t) = \int_0^t h(X(s))\mathrm{d}s + B_2(t), \tag{21}$$

where

- $m(\cdot, \cdot): \mathscr{R}^d \times \mathscr{P}(U) \to \mathscr{R}^d$ is of the form

  $$m(x, \mu) = \int \tilde{m}(x, u)\mathrm{d}\mu(u), \quad x \in \mathscr{R}^{\mathrm{d}}, \quad \mu \in \mathscr{P}(U),$$

  (componentwise integration) for some $\tilde{m}(\cdot, \cdot): \mathscr{R}^d \times U \to \mathscr{R}^d$ which is bounded continuous and Lipschitz in its first argument uniformly w.r.t. the second,
- $\sigma(\cdot): \mathscr{R}^d \to \mathscr{R}^{d \times d}$ is bounded Lipschitz with the least eigenvalue of $\sigma(\cdot)\sigma^{\mathrm{T}}(\cdot)$ uniformly bounded away from zero,
- X_0 has a prescribed law $\pi_0 \in \mathscr{P}(\mathscr{R}^d)$,
- $B_1(\cdot), B_2(\cdot)$ are independent, respectively, 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 wide sense admissible in the sense of Fleming and Pardoux (1982); Borkar (1989, Chapter V) or Borkar (2003),
- $h: \mathscr{R}^d \to \mathscr{R}^r$ is bounded continuous and twice continuously differentiable with bounded first and second partial derivatives.

The aim is to minimize the ergodic cost

$$\limsup_{t \to \infty} \frac{1}{t} \int_0^t E[k(X(s), Z(s))]\, \mathrm{d}s, \tag{22}$$

where $k \in C_b(\mathscr{R}^d \times \mathscr{P}(U))$ is of the form

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

for some $k \in C_b(\mathscr{R}^d \times U)$. We also assume the stability condition: There exist $\varepsilon_0 > 0$, $\mathscr{V}, \hat{\mathscr{V}} \in C^2(\mathscr{R}^d)$, such that

$$\lim_{\|x\| \to \infty} \mathscr{V}(x) = \infty,$$

$$\mathscr{L}_u \mathscr{V}(x) \leqslant -\varepsilon_0 + C I_B(x) \ \forall u,$$

$$\mathscr{L}_u \hat{\mathscr{V}}(x) \leqslant -\mathscr{V}(x) + \hat{C} I_{\hat{B}}(x) \ \forall u,$$

$$\lim_{t \to \infty} \frac{E[\hat{\mathscr{V}}(X(t))]}{t} = 0,$$

where $C, B, \hat{C}, \hat{B}$, are as before, the expectation $E[\,\cdot\,]$ is under any arbitrary wide sense admissible control, and

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

for $f \in C^2(\mathscr{R}^d)$.

As before, one considers the equivalent separated control problem of controlling the $\mathscr{P}(\mathscr{R}^d)$-valued process $\{\pi_t\}$ of regular conditional laws of $X(t)$ given $\mathscr{G}_t \triangleq$ the right-continuous completion of $\sigma(Y(s), Z(s), \ s \leqslant t)$ for $t \geqslant 0$. This evolves according to the Fujisaki–Kallianpur–Kunita equation

$$\pi_t(f) = \pi_0(f) + \int_0^t \pi_s(\mathscr{L}_{Z(s)}(f))\,\mathrm{d}s$$

$$+ \int_0^t \langle \pi_s(hf) - \pi_s(f)\pi_s(h), \mathrm{d}\hat{Y}(s)\rangle,$$

where

- $f \in C_b^2(\mathscr{R}^d)$ and $\pi(g) \triangleq \int g\,\mathrm{d}\pi$ for $g \in C_b(\mathscr{R}^d)$ and $\pi \in \mathscr{P}(\mathscr{R}^d)$,
- the 'innovations process' $\hat{Y}(t) \triangleq Y(t) - \int_0^t \pi_s(h)\,\mathrm{d}s, \ t \geqslant 0$, is an r-dimensional Brownian motion independent of $X_0, B_1(\cdot)$.

The cost (22) is written in the more convenient form

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

As in Borkar (2003), we observe that $\pi_0 \in \mathscr{P}_0(\mathscr{R}^d)$ implies $\pi_t \in \mathscr{P}_0(\mathscr{R}^d)$ a.s. for $t \geqslant 0$ and thus we may view $\{\pi_t\}$ as a $\mathscr{P}_0(\mathscr{R}^d)$-valued process. Our main result is:

Theorem 5.1. *There exist $V^* \in C(\mathscr{P}_0(\mathscr{R}^d))$, $\gamma \in \mathscr{R}^d$, such that γ is the optimal cost regardless of the initial law and for all $t > 0$,*

$$V^*(\pi) = \inf E\left[\int_0^t (\pi_s(k(\cdot, Z(s))) - \gamma)\,\mathrm{d}s + V^*(\pi_t)|\pi_0 = \pi\right],$$

where the infimum is over all wide sense admissible controls. In particular, $(V^*(\pi_t) + \int_0^t (\pi_s(k(\cdot, Z(s))) - \gamma)\,\mathrm{d}s, \mathscr{G}_t)$ is a submartingale and if it is a martingale, $(\pi_t, Z(t), t \geqslant 0)$ is an optimal pair. Conversely, it is a martingale if $(\pi_t, Z(t), t \geqslant 0)$ is an optimal stationary pair.

This follows exactly as in Theorem 5.1 of Borkar (2003) via an embedded discrete problem, the latter being handled as in the preceding section.

Appendix A

Here we briefly sketch the pseudo-atom construction which was referred to just before Lemma 3.1. Note that the control sequence $\{Z_n\}$ corresponds to some $\zeta(\mathrm{d}y^\infty, \mathrm{d}u^\infty) \in \Theta \subset \mathscr{P}(U^\infty \times W^\infty)$. Define

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

and let $\bar{\bar{\mathscr{F}}}$ be the natural product σ-field. Let $\bar{P}_0$ be a probability measure on $(\bar{\Omega}, \bar{\bar{\mathscr{F}}})$ defined as

$$\bar{P}_0((\mathrm{d}\tilde{u}^\infty \times \mathrm{d}\tilde{x}) \times (\mathrm{d}\hat{u}^\infty \times \mathrm{d}\hat{x}) \times \mathrm{d}z^\infty \times \mathrm{d}\tilde{y}^\infty \times \mathrm{d}\hat{y}^\infty)$$

$$\doteq \ell^\infty(\mathrm{d}\tilde{u}^\infty)\tilde{\pi}(\mathrm{d}\tilde{x})\ell^\infty(\mathrm{d}\hat{u}^\infty)\hat{\pi}(\mathrm{d}\hat{x})\zeta(\mathrm{d}\tilde{y}^\infty, \mathrm{d}z^\infty)\eta^\infty(\mathrm{d}\hat{y}^\infty).$$

Here ℓ denotes the Lebesgue measure. On $(\bar{\Omega}, \bar{\bar{\mathscr{F}}}, \bar{P}_0)$ define canonical processes $\{\tilde{\xi}_n\}, \{\hat{\xi}_n\}, \{Z_n\}, \{\tilde{Y}_n\}, \{\hat{Y}_n\}$ and variables $\tilde{X}_0, \hat{X}_0$ as follows. For $\omega = (\tilde{u}^\infty, \tilde{x}, \hat{u}^\infty, \hat{x}, z^\infty, \tilde{y}^\infty, \hat{y}^\infty)$

$$\tilde{\xi}_n(\omega) \doteq \tilde{u}_n, \quad \hat{\xi}_n(\omega) \doteq \hat{u}_n, \quad Z_n(\omega) \doteq z_n, \quad \tilde{Y}_n(\omega) \doteq \tilde{y}_n,$$

$$\hat{Y}_n(\omega) \doteq \hat{y}_n, \quad n \geqslant 0$$

and $\tilde{X}_0(\omega) \doteq \tilde{x}, \hat{X}_0(\omega) \doteq \hat{x}$. Define $\{\tilde{X}_n\}, \{\hat{X}_n\}$ recursively by

$$\tilde{X}_{n+1} \doteq F(\tilde{X}_n, Z_n, \tilde{\xi}_{n+1}), \quad \hat{X}_{n+1} \doteq F(\hat{X}_n, Z_n, \hat{\xi}_{n+1}), \quad n \geqslant 0.$$

Then by a change of measure technique one can obtain a measure $\bar{P}$ on $(\bar{\Omega}, \bar{\bar{\mathscr{F}}})$ (see Borkar, 2003, for details) such that $\{\tilde{X}_n\}$ and $\{\hat{X}_n\}$ defined on $(\bar{\Omega}, \bar{\bar{\mathscr{F}}}, \bar{P})$ are controlled Markov chains with the desired transition kernel and with initial laws $\tilde{\pi}$ and $\hat{\pi}$, respectively, and are driven by a common control sequence $\{Z_n\}$ that is wide sense admissible for both, with the corresponding observation processes $\{\tilde{Y}_n\}, \{\hat{Y}_n\}$. Let $\bar{Y}_n \doteq (\tilde{Y}_n, \hat{Y}_n)$ and $\bar{X}_n \doteq (\tilde{X}_n, \hat{X}_n)$. Then $\{\bar{X}_n\}$ is a $H \doteq S^2$ valued controlled Markov chain with U valued control $\{Z_n\}$ and W^2 valued observation sequence $\{\bar{Y}_n\}$. Denote the corresponding controlled transition kernel by $\bar{p}(x, u, \mathrm{d}x' \times \mathrm{d}y') \in \mathscr{P}(S^2 \times W^2)$; $(x, u) \in H \times U$. Let $G \doteq (B \cup \hat{B})^2$ and define $v \in \mathscr{P}(H)$ by

$$v(A) \doteq \frac{(\lambda \times \lambda)(A \cap G)}{\lambda(B \cup \hat{B})^2}, \quad A \in \mathscr{B}(H).$$

Let

$$\delta \doteq \frac{1}{2}\left(\inf_{x \in B \cup \hat{B}, u \in U, y \in B \cup \hat{B}} \bar{\varphi}(x, u, y)\lambda(B \cup \hat{B})\right)^2.$$

Note that by our assumption on φ, we have $\delta > 0$. Also,

$$\bar{p}(x, u, A \times W^2) \geqslant \delta I_G\{x\}v(A), \quad A \in \mathscr{B}(H), \quad (x, u) \in H \times U.$$

We are now ready for the pseudo-atom construction. Let $H^* \doteq H \times \{0,1\}$ endowed with its Borel σ-field.

For a set $A \in \mathscr{B}(H)$, we let $A_0 \doteq A \times \{0\}$ and $A_1 \doteq A \times \{1\}$. For every $\mu \in \mathscr{P}(H)$ we define a $\mu^* \in \mathscr{P}(H^*)$ as follows. For $A \in \mathscr{B}(H)$,

$$
\mu^*(A_0) \doteq (1 - \delta)\mu(AG) + \mu(A(G)^c),
$$
$$
\mu^*(A_1) \doteq \delta\mu(AG). \tag{A.1}
$$

Clearly, $\mu^*(A_0) + \mu^*(A_1) = \mu(A)$ and if $A \subset (G)^c$ then $\mu^*(A_0) = \mu(A)$. Similarly, for $\mu \in \mathscr{P}(H \times W^2)$ we define a $\mu^* \in \mathscr{P}(H^* \times W^2)$ as follows. For $A \in \mathscr{B}(H), D \in \mathscr{B}(W^2)$,

$$
\mu^*(A_0 \times D) \doteq (1 - \delta)\mu(AG \times D) + \mu(A(G)^c \times D),
$$
$$
\mu^*(A_1 \times D) \doteq \delta\mu(AG \times D). \tag{A.2}
$$

Clearly, $\mu^*(A_0 \times D) + \mu^*(A_1 \times D) = \mu(A \times D)$ and if $A \subset (G)^c$ then $\mu^*(A_0 \times D) = \mu(A \times D)$.

Given $f : H \times U \to \mathbb{R}$, define $f^* : H^* \times U \to \mathbb{R}$ as

$$
f^*((x^1, x^2, i), u) \doteq f((x^1, x^2), u), \quad (x^1, x^2) \in H, \quad i \in \{0, 1\}, \quad u \in U. \tag{A.3}
$$

In a similar way we define the extension of $f : H^n \times U^n \to \mathbb{R}$ to a function from $(H^*)^n \times U^n \to \mathbb{R}$ and denote it by f^*.

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

(1) The controlled transition kernel of (Υ_n, Y_n^*) is given as follows. For $\bar{z} \equiv (z, i) \in H^*$ and $u \in U$

$$
q(\bar{z}, u, \mathrm{d}x' \times \mathrm{d}y)
$$
$$
= \begin{cases} \bar{p}^*(z, u, \mathrm{d}x' \times \mathrm{d}y) & \text{if } \bar{z} \in H_0 \backslash G_0 \\ \dfrac{1}{1 - \delta}(\bar{p}^*(z, u, \mathrm{d}x' \times \mathrm{d}y) - \delta v^*(\mathrm{d}x')\eta^2(\mathrm{d}y)) & \text{if } \bar{z} \in G_0 \\ v^*(\mathrm{d}x')\eta^2(\mathrm{d}y) & \text{if } \bar{z} \in H_1, \end{cases} \tag{A.4}
$$

where $x' \equiv (x_0', j) \in H^*$.

(2) The initial distributions are given as follows:

$$
P^*(\Upsilon_0 \in A_0, Y_0^* \in A', Z_0^* \in A'') \doteq (1 - \delta)P(\bar{X}_0 \in AG, \bar{Y}_0 \in A', Z_0 \in A'')
$$
$$
+ P(\bar{X}_0 \in A(G)^c, \bar{Y}_0 \in A', Z_0 \in A'')
$$

$$
P^*(\Upsilon_0 \in A_1, Y_0^* \in A', Z_0^* \in A'') \doteq \delta P(\bar{X}_0 \in AG, \bar{Y}_0 \in A', Z_0 \in A''),
$$

for $A \in \mathscr{B}(H), \ A' \in \mathscr{B}(W^2), \ A'' \in \mathscr{B}(U)$.

(3) The control sequence $\{Z_n^*\}$ is given as follows. For $n \in \mathbb{N}, \ A' \in \mathscr{B}(U)$ and $(z_m, i_m, y_m, \alpha_m) \in H \times \{0, 1\} \times W^2 \times U$

$$
P^*(Z_n^* \in A' | (\Upsilon_m, Y_m^*) = (z_m, i_m, y_m), \ Z_{m-1}^* = \alpha_{m-1}, m \leqslant n)
$$
$$
= P(Z_n \in A' | (\bar{X}_m, \bar{Y}_m) = (z_m, y_m), \ Z_{m-1} = \alpha_{m-1}, m \leqslant n). \tag{A.5}
$$

The above construction ensures that the probability laws of the processes $\{X_n^*, Y_n^*, Z_n^*\}_{n \in \mathbb{N}_0}$ and $\{(\bar{X}_n, \bar{Y}_n), Z_n\}_{n \in \mathbb{N}_0}$ are the same. Furthermore, G_1 is an accessible atom of $\{Z_n\}$ in the sense of Meyn and Tweedie (1993). (See Borkar, 2003, for details.) Finally, define the hitting time of the "pseudo-atom" G_1 as

$$\tau \doteq \min\{n \geqslant 0 : \Upsilon_n \in G_1\}. \tag{A.6}$$

Properties of τ and the processes $\{\tilde{X}_n^*, \hat{X}_n^*, Y_n^*, Z_n^*\}_{n \in \mathbb{N}_0}$ stated above Lemma 3.1 are proved in Lemmas 3.1–3.3 of Borkar (2003).

References

Borkar, V.S., 1989. Optimal Control of Diffusion Processes. Pitman Research Notes in Mathematics, Vol. 203. Longman Scientific and Technical, Harlow, UK.

Borkar, V.S., 1991. Topics in Controlled Markov Chains. Pitman Research Notes in Mathematics, Vol. 240. Longman Scientific and Technical, Harlow, UK.

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

Borkar, V.S., 2003. Dynamic programming for ergodic control with partial observations. Stochastic Process. Appl. 103, 293–310.

Fernández-Gaucherand, E., Arapostathis, A., Marcus, S.I., 1990. Remarks on the existence of solutions to the average cost optimality equation in Markov decision processes. Systems Control Lett. 15, 425–432.

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

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

Platzman, L.K., 1980. Optimal infinite horizon control of finite probabilistic systems. SIAM J. Control Optim. 18, 362–380.

Runggaldier, W.J., Stettner, L., 1994. Approximations of Discrete Time Partially Observed Control Problems. In: Applied Maths Monographs, Vol. 6. Giardini Editori e Stampatori, Pisa.

Stettner, L., 1993. Ergodic control of partially observed Markov control processes with equivalent transition probabilities. Appl. Math. 22, 25–38.

Tuominen, P., Tweedie, R.L., 1994. Subgeometric rates of convergence of f-ergodic Markov chains. Adv. Appl. Probab. 26, 775–798.

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