Morfismos, Vol. 16, No. 2, 2012, pp. 51–68

Variance optimality for controlled Markov-modulated diffusions *

Beatris Adriana Escobedo-Trujillo
Carlos Octavio Rivera-Blanco

Abstract

This paper concerns controlled Markov-modulated diffusions. Our main objective is to give conditions for the existence of optimal policies for the limiting average variance criterion. To this end, we use the fact that the family of average reward optimal policies is nonempty. Then, within this family, we search policies that minimize the limiting average variance.

2010 Mathematics Subject Classification: 93E20, 60J60.
Keywords and phrases: variance optimality criterion, controlled Markov-modulated diffusions, exponential ergodicity.

1 Introduction

Using the fact that the family of average reward optimal policies is nonempty (see [5] for details), in this paper we study the existence of stationary policies that minimize the limiting average variance in the class of average optimal policies. Under our assumptions we extend to controlled switching diffusions the results in [6] on discrete-time Markov control processes.

A diffusion with Markovian switchings (also known as a piecewise diffusions, switching diffusions, or Markov-modulated diffusions) is a stochastic differential equation with coefficients depending on a continuous-time irreducible finite-state homogeneous Markov chain. The motivation to study switching diffusions is that recent studies suggest that

*The research of the first author (BAET) was supported by CONACyT scholarship 167588.

such processes are more general and appropriate for a wide variety of applications not covered by standard Markovs diffusion models. For related references see [1, 4, 5, 12, 14, 15, 16].

The existence of optimal policies for Markov-modulated diffusions with the average reward criterion has been previously studied in the literature; see, for instance [3, 4, 5]. But, as is well known, this criterion is very underselective because an average reward optimal policy may have an arbitrarily bad behavior for large but finite lengths of time. To avoid this situation, many authors consider more sensitive criteria such as the limiting average variance criterion; see [6, 7, 9, 10, 11, 13].

The paper is organized as follows. Section 2 introduces our assumptions, which lead to the notion of w and w^2-exponential ergodicity, a crucial tool for our results. In Section 3 we define the average optimality criterion we are interested in, and we summarize some results from [5] on the existence of solutions to the average reward problem, which is essentially our point of departure to analyze variance optimality. In Section 4 we define the variance optimality criterion and we prove the existence of variance optimal policies. In Section 5 we prove the Theorem 4.4 which states that the limiting average variance equals a constant independient of the initial state. Our results are illustrated with an example in Section 6.

2 Model Definition and Ergodic Properties

The control system we are concerned with is the controlled Markov-modulated diffusion process

$$dx(t) = b(x(t), \psi(t), u(t))dt + \sigma(x(t), \psi(t))dW(t), \tag{1}$$

for $t \geq 0$, $x(0) = x$, and $\psi(0) = i$, with coefficients depending on a continuous-time irreducible Markov chain $\psi(\cdot)$ with a finite state space $E = \{1, 2, \ldots, N\}$, and transition probabilities

$$\mathcal{P}(\psi(s+t) = j | \psi(s) = i) = q_{ij}t + o(t). \tag{2}$$

For states $i \neq j$ the number $q_{ij} \geq 0$ is the transition rate from i to j, while $q_{ii} := -\sum_{j \neq i} q_{ij}$. Moreover, in (1), $b : \mathbb{R}^n \times E \times U \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times E \to \mathbb{R}^{n \times d}$ are given functions, and $W(\cdot)$ is a d-dimensional standard Brownian motion independent of $\psi(\cdot)$. The stochastic process $u(\cdot)$ is a $U-$valued process called a control process, and the set $U \subset \mathbb{R}^m$ is called the control (or action) space.

Notation.

- For $x \in \mathbb{R}^n$ and a matrix A, we use the usual Euclidean norms

$$|x|^2 := \sum_k x_k^2 \text{ and } |A|^2 := \sum_{k,l} A_{k,l}^2.$$

  We will also denote by A' the transpose of a square matrix A.

- We use $P^{x,i,u}(t,\cdot)$ to denote the transition probability of the process $(x(\cdot),\psi(\cdot))$, i.e.,

$$P^{x,i,u}(t, B \times J) \quad := \quad P((x(t),\psi(t)) \in B \times J | x(0) = x, \psi(0) = i)$$

  for every Borel set $B \subset \mathbb{R}^n$ and $J \subset E$. The associated conditional expectation is written $\mathbb{E}^{x,i,u}(\cdot)$.

Assumption 2.1.

(*a*) *The control set U is compact.*

(*b*) *$b(x,i,u)$ is continuous on $\mathbb{R}^n \times E \times U$, and $x \mapsto b(x,i,u)$ satisfies a Lipschitz condition uniformly in $(i,u) \in E \times U$; that is, there exist a positive constant K_1 such that*

$$\max_{(i,u)\in E\times U} |b(x,i,u) - b(y,i,u)| \le K_1|x-y| \text{ for all } x,y \in \mathbb{R}^n.$$

(*c*) *There exists a positive constant K_2 such that, for each $i \in E$ and $x, y \in \mathbb{R}^n$,*

$$|\sigma(x,i) - \sigma(y,i)| \le K_2|x-y|,$$

(*d*) *There exists a positive constant K_3 such that the matrix $a(x,i) := \sigma(x,i)\sigma'(x,i)$ satisfies that, for each $i \in E$ and $x, y \in \mathbb{R}^n$,*

$$x'a(y,i)x \ge K_3|x|^2 \quad (\text{uniform ellipticity}).$$

Remark 2.2. *The Lipschitz conditions on b and σ in Assumption 2.1 imply that b and σ satisfy a linear growth condition. That is, there exists a constant $\overline{C} \ge K_1 + K_2$ such that for all $x \in \mathbb{R}^n$*

$$\sup_{(u,i)\in U\times E} (|b(x,i,u)|) + |\sigma(x,i))| \le \overline{C}(1+|x|).$$

Control policies. For our present purposes, we can restrict ourselves to consider stationary Markov policies, defined as follows.

Definition 2.3. *Let $\mathbb{F}$ be the family of measurable functions $f : \mathbb{R}^n \times E \to U$. A control policy of the form $u(t) := f(x(t), \psi(t))$ for some $f \in \mathbb{F}$ and $t \geq 0$ is called a* stationary Markov policy*. Actually, by an abuse of terminology, f itself will be referred to as a stationary Markov policy.*

Infinitesimal generator. Let $C^2(\mathbb{R}^n \times E)$ be the space of real-valued continuous functions $\nu(x, i)$ on $\mathbb{R}^n \times E$, which are twice continuously differentiable in $x \in \mathbb{R}^n$ for each $i \in E$. For $\nu \in C^2(\mathbb{R}^n \times E)$ and $u \in U$, let

$$\mathcal{Q}\nu(x,i) := \sum_{j=1}^{N} q_{ij}\nu(x,j),$$

where $\mathcal{Q} = [q_{ij}]$ is the generator of the Markov chain $\psi(\cdot)$, and

$$\begin{aligned}\mathcal{L}^u\nu(x,i) \quad := \quad & \sum_{k=1}^{n} \frac{\partial \nu}{\partial x_k}(x,i) b_k(x,i,u) + \frac{1}{2}\sum_{k,l} a_{kl}(x,i)\frac{\partial^2 \nu}{\partial x_k \partial x_l}(x,i) \\ & + \mathcal{Q}\nu(x,i),\end{aligned}$$

where b_k is the $k-$th component of b, and a_{kl} is the $k, l-$component of the matrix $a(\cdot,\cdot)$ defined in Assumption 2.1(d).

For each $f \in \mathbb{F}$ and $(x, i) \in \mathbb{R}^n \times E$ let

$$\mathcal{L}^f\nu(x,i) := \mathcal{L}^{f(x,i)}\nu(x,i). \tag{3}$$

Under Assumption 2.1 for each stationary Markov policy $f \in \mathbb{F}$ there exists an almost surely unique strong solution of (1)-(2); see [12, page 88-90]. On the other hand, even though $x(t)$ itself is not necessarily Markov, it is well known that the joint process $(x(\cdot), \psi(\cdot))$ is Markov; see, for instance, [12, pp. 104-106]. The infinitesimal generator of the Markov process $(x(t), \psi(t))$ is $\mathcal{L}^f$ in (3) for each stationary Markov policy $f \in \mathbb{F}$; see [12, page 48].

Recurrence and ergodicity. For the variance optimality criterion, we require the following second order condition (a Lyapunov-like condition) that ensures the positive recurrence of the controlled Markov-modulated diffusion (1)-(2) (see [5], [14] and [15].)

Assumption 2.4. *There exists a function $w \in C^2(\mathbb{R}^n \times E)$, with $w \geq 1$, and constants $p \geq q > 0$ such that* (*i*) $\lim_{|x|\to\infty} w(x,i) = +\infty$ *for each* $i \in E$, *and*

(*ii*) *for each* $u \in U$ *and* $(x,i) \in \mathbb{R}^n \times E$

$$\mathcal{L}^u w^2(x,i) \leq -q w^2(x,i) + p. \tag{4}$$

Under the Assumption 2.4, for each $f \in \mathbb{F}$, the Markov process $(x(\cdot), \psi(\cdot))$ is Harris positive recurrent with a unique invariant probability measure $\mu_f(dx,i)$ (see [15]) for which

$$\mu_f(w^2) := \sum_{i=1}^{N} \int_{\mathbb{R}^n} w^2(x,i)\mu_f(dx,i) < \infty.$$

Definition 2.5. *Let* $\mathcal{B}_w(\mathbb{R}^n \times E)$ *be the normed linear space of real-valued measurable functions* ν *on* $\mathbb{R}^n \times E$ *with finite* $w-$*norm, which is defined as*

$$\| \nu \|_w := \sup_{(x,i)\in\mathbb{R}^n\times E} \frac{| \nu(x,i) |}{w(x,i)}.$$

Remark 2.6. *A consequence of Assumptions 2.1(d) and 2.4 (ii) states that the condition of second order (4) implies the following first order condition:*

$$\mathcal{L}^u w(x,i) \leq -q_1 w(x,i) + p_1, \tag{5}$$

for constants $q_1 = \frac{q}{2}$ *and* $p_1 = \frac{p}{2}$, *where* p *and* q *are the constants given in Assumption 2.4. For details, see [7, Proposition 2.3].*

Under Assumptions 2.1 and 2.4, Theorem 2.8 in [3] ensures that the controlled Markov-modulated diffusion (1)-(2) is uniformly $w-$*exponentially ergodic*, that is, there exist positive constants C and δ such that

$$\sup_{f\in\mathbb{F}} |\mathbb{E}^{x,i,f}[\nu(x(t),\psi(t))] - \mu_f(\nu)| \leq Ce^{-\delta t} \| \nu \|_w \, w(x,i) \tag{6}$$

for all $(x,i) \in \mathbb{R}^n \times E$, $\nu \in \mathcal{B}_w(\mathbb{R}^n \times E)$, and $t \geq 0$, where $\mu_f(\nu) := \sum_{i=1}^{N} \int_{\mathbb{R}^n} \nu(x,i)\mu_f(dx,i)$.

3 Average Optimality Criteria

Let $r : \mathbb{R}^n \times E \times U \to \mathbb{R}$ be a measurable function, which we call the *reward rate*. It satisfies the following conditions:

Assumption 3.1.

(*a*) *The function* $r(x,i,u)$ *is continuous on* $\mathbb{R}^n \times E \times U$ *and locally Lipschitz in* x *uniformly with respect to* $i \in E$ *and* $u \in U$*; that is, for each* $R > 0$*, there exists a constant* $K(R) > 0$ *such that*

$$\sup_{(i,u)\in E\times U} |r(x,i,u) - r(y,i,u)| \leq K(R)|x-y| \quad \text{for all } |x|,|y| \leq R.$$

(*b*) $r(\cdot,\cdot,u)$ *is in* $\mathcal{B}_w(\mathbb{R}^n \times E)$ *uniformly in* u*; that is, there exists* $M > 0$ *such that for each* $(x,i) \in \mathbb{R}^n \times E$

$$\sup_{u\in U} |r(x,i,u)| \leq Mw(x,i).$$

Notation. For each Markov policy $f \in \mathbb{F}$, $x \in \mathbb{R}^n$ and $i \in E$, we write

$$r(x,i,f) := r(x,i,f(x,i)).$$

The following definition concerns the long-run average optimality criterion.

Definition 3.2. *For each* $f \in \mathbb{F}$, $(x,i) \in \mathbb{R}^n \times E$, *and* $T \geq 0$, *let*

$$J_T(x,i,f) := \mathbb{E}^{x,i,f}\Big[\int_0^T r(x(t),\psi(t),f)dt\Big].$$

The long-run expected average reward *given the initial state* (x,i) *is*

$$J(x,i,f) := \liminf_{T\to\infty} \frac{1}{T} J_T(x,i,f). \tag{7}$$

The function

$$J^*(x,i) := \sup_{f\in\mathbb{F}} J(x,i,f) \quad \text{for all } (x,i) \in \mathbb{R}^n \times E$$

is referred to as the optimal gain *or* the optimal average reward. *If there is a policy* $f^* \in \mathbb{F}$ *for which* $J(x,i,f^*) = J^*(x,i)$ *for all* $(x,i) \in \mathbb{R}^n \times E$, *then* f^* *is called* average optimal.

Remark 3.3. *The following important results were proven in [3].*

(1) *The $w-$exponential ergodicity (6) gives that the long-run expected average reward (7) coincides with a constant $g(f)$, which is defined by*

$$g(f) := \mu_f(r(\cdot,\cdot,f)) = \sum_{i=1}^{N} \int_{\mathbb{R}^n} r(x,i,f)\mu_f(dx,i). \tag{8}$$

for every $f \in \mathbb{F}$. This is,

$$g(f) = J(x,i,f).$$

(2) *Under Assumptions 2.1, 2.4, and 3.1, Theorem 4.2 in [3] ensures the existence of optimal average reward policies. Denoting by g^* the optimal average reward and by $\mathbb{F}_{ao}$ the family of average optimal policies, we have:*

$$g^* := \sup_{f\in\mathbb{F}} g(f) = \sup_{f\in\mathbb{F}} J(x,i,f) \quad \textit{for all } (x,i) \in \mathbb{R}^n \times E.$$

(3) *We define, for each $f \in \mathbb{F}$, the* bias *of f as the function*

$$h_f(x,i) := \int_0^\infty [\mathbb{E}^{x,i,f} r(x(t),\psi(t),f) - g(f)]dt \ \textit{ for } (x,i) \in \mathbb{R}^n \times E. \tag{9}$$

Note that this function is finite-valued because (6) and the Assumption 3.1(b) give, for all $t \geq 0$,

$$| \mathbb{E}^{x,i,f} r(x(t),\psi(t),f) - g(f) | \leq e^{-\delta t} CMw(x). \tag{10}$$

Hence, by (9) and (10), the bias of f is such that

$$|h_f(x,i)| \leq \delta^{-1}CMw(x), \quad \textit{and so} \quad \| h_f(x,i) \|_w \leq \delta^{-1}CM.$$

This means that the bias h_f is a finite-valued function and, in fact, it is in $\mathcal{B}_w(\mathbb{R}^n \times E)$.

(4) *In addition, by Proposition 5.2 in [3] we known that for each $f \in \mathbb{F}$, the pair $(g(f), h_f)$ is the unique solution of the following Poisson equation*

$$g(f) = r(x,i,f) + \mathcal{L}^f h_f(x,i) \ \textit{ for } (x,i) \in \mathbb{R}^n \times E. \tag{11}$$

4 Variance Optimality

In this section we study the existence of a stationary policy that minimizes the limiting average variance in the class $\mathbb{F}_{ao}$ of average optimal policies.

Remark 4.1. *We define the normed linear space $\mathcal{B}_{w^2}(\mathbb{R}^n \times E)$ of real-valued measurable functions ν on $\mathbb{R}^n \times E$ with finite w^2-norm, similarly to the normed linear spaces $\mathcal{B}_w(\mathbb{R}^n \times E)$, with w^2 in lieu of w.*

Remark 4.2. *Under the Assumptions 2.1, 2.4, and 3.1, the process $(x(\cdot), \psi(\cdot))$ is uniformly w^2-exponentially ergodic, i.e, there exist positive constants C and δ such that*

$$\sup_{f \in \mathbb{F}} |\mathbb{E}^{x,i,f}[\nu(x(t), \psi(t))] - \mu_f(\nu)| \leq Ce^{-\delta t} \parallel \nu \parallel_{w^2} w^2(x, i)$$

for all $(x, i) \in \mathbb{R}^n \times E$, $\nu \in \mathcal{B}_{w^2}(\mathbb{R}^n \times E)$, and $t \geq 0$. The proof of this result is similar to that given in [3, Theorem 2.8] with w^2 in lieu of w.

Definition 4.3. *For each f in $\mathbb{F}$, the limiting average variance of f given the initial state $(x, i) \in \mathbb{R}^n \times E$, is the function*

$$\sigma^2(x, i, f) := \lim_{T \to \infty} \frac{1}{T} \mathbb{E}^{x,i,f} \Big(\int_0^T r(x(t), \psi(t), f)dt - J_T(x, i, f) \Big)^2. \tag{12}$$

The following theorem, which is proved in Section 5, states that the limiting average variance equals a constant.

Theorem 4.4. *Under Assumptions 2.1 and 2.4, for each f in $\mathbb{F}$ and an arbitrary initial state $(x, i) \in \mathbb{R}^n \times E$, the limiting average variance $\sigma^2(x, i, f)$ equals the constant*

$$\sigma^2(f) := 2 \sum_{i=1}^{n} \int_{\mathbb{R}^n} (r(x, i, f) - g(f))h_f(x, i)\mu_f(dx, i). \tag{13}$$

The following definition concerns the variance optimality criterion.

Definition 4.5. *We say that a stationary policy f^* is variance optimal if $f^* \in \mathbb{F}_{ao}$ and, moreover,*

$$\sigma^2(f^*) = \min_{f \in \mathbb{F}_{ao}} \sigma^2(f). \tag{14}$$

We define for each $(x, i) \in \mathbb{R}^n \times E$ the set

$$U^*(x, i) := \{u \in U |\ g = r(x, i, u) + \mathcal{L}^u h(x, i)\},$$

with $g \in \mathbb{R}$ and $h \in C^2(\mathbb{R}^n \times E) \cap \mathcal{B}_w(\mathbb{R}^n \times E)$. By [3, Lemma 6.2] for each $(x, i) \in \mathbb{R}^n \times E$, $U^*(x, i)$ is a nonempty compact set.

Proposition 4.6. *Suppose that Assumptions 2.1, 2.4, and 3.1 are satisfied. Then*

(*i*) *There exist* $g, \sigma^2 \in \mathbb{R}$, $h \in C^2(\mathbb{R}^n \times E) \cap \mathcal{B}_w(\mathbb{R}^n \times E)$, *and* $\phi \in C^2(\mathbb{R}^n \times E) \cap \mathcal{B}_{w^2}(\mathbb{R}^n \times E)$, *that satisfy the system of equations*

$$g = \max_{u \in U} \{r(x, i, u) + \mathcal{L}^u h(x, i)\}, \tag{15}$$

$$\sigma^2 = \min_{u \in U^*(x,i)} \{2(r(x, i, u) - g)h(x, i) + \mathcal{L}^u \phi(x, i)\} \tag{16}$$

for all $(x, i) \in \mathbb{R}^n \times E$.

(*ii*) *A policy* f^* *in* $\mathbb{F}$ *is variance optimal if and only if* f^* *attains the maximum and the minimum in (15) and (16), respectively. The minimal limiting average variance* $\sigma^2(f^*)$ *equals* σ^2 *in (16).*

Proof. (*i*) The existence of a constant g and a function $h \in C^2(\mathbb{R}^n \times E) \cap \mathcal{B}_w(\mathbb{R}^n \times E)$ that satisfy (15) follows from Theorem 4.2 in [3]. The latter theorem also yields the existence of a stationary policy $f \in \mathbb{F}$ that attains the maximum in the right-hand side of (15), i.e.,

$$g = r(x, i, f) + \mathcal{L}^f h(x, i) \ \text{ for all } (x, i) \in \mathbb{R}^n \times E.$$

Now suppose that f is in $\mathbb{F}_{ao}$. Then by Proposition 5.3 in [3] the bias of f satisfies that

$$h_f(x, i) = h(x, i) - \mu_f(h) \ \text{ for all } (x, i) \in \mathbb{R}^n \times E. \tag{17}$$

Thus, using (8), the limiting average variance of f verifies that

$$\begin{aligned} \sigma^2(f) &= 2\sum_{i=1}^{n} \int_{\mathbb{R}^n} (r(x, i, f) - g(f))h_f(x, i)\mu_f(dx, i) \\ &= 2\sum_{i=1}^{n} \int_{\mathbb{R}^n} (r(x, i, f) - g)h(x, i)\mu_f(dx, i). \end{aligned} \tag{18}$$

This implies that $\sigma^2(f)$ is the expected average reward of the policy f when the reward rate is the function

$$r''(x,i,u) := 2(r(x,i,u) - g)h(x,i) \text{ for all } (x,i) \in \mathbb{R}^n \times E.$$

Hence, to find a solution of (16) we need to solve a new average reward control problem. This problem has the following components: the dynamic system (1), the action sets $U^*(x,i)$, and the reward rate $r''(x,i,u)$. Note that

$$\begin{aligned} |r''(x,i,u)| &= 2|(r(x,i,u) - g)h(x,i)| \\ &\leq 2(Mw(x,i) + g)|h(x,i)| \\ &\leq 2(Mw(x,i) + g)||h(x,i)||_w w(x,i) \\ &\leq 2w^2(x,i)||h||_w(M + g). \end{aligned}$$

where the first inequality is by assumption 3.1(b) and the second inequality holds since $h \in \mathcal{B}_w(\mathbb{R}^n \times E)$.
Therefore $r''(x,i,u)$ verifies the assumption 3.1(b) when w is replaced with w^2. The control problem with the above components satisfies the assumptions 2.1, 2.4 and 3.1 replacing w with w^2. Hence, by [3, Theorem 4.2], there exists (σ^2, ϕ), with $\sigma^2 \in \mathbb{R}$ and $\phi \in C^2(\mathbb{R}^n \times E) \cap \mathcal{B}_{w^2}(\mathbb{R}^n \times E)$, that satisfy the equation (16).

(ii) By [3, Theorem 4.2] there exists a policy $f^* \in \mathbb{F}$ that attains the maximum in (15). Note that a stationary policy f^* is in $\mathbb{F}_{ao}$ if and only if $f^*(x,i)$ is in $U^*(x,i)$ for all $(x,i) \in \mathbb{R}^n \times E$. Moreover, by [3, Lemma 6.2 (a)] , $U^*(x,i)$ is a compact set for each $(x,i) \in \mathbb{R} \times E$. Hence, by [3, Theorem 4.2] , f^* is variance optimal.

Now, by (16) we have

$$\sigma^2 \leq 2(r(x,i,f) - g)h(x,i) + \mathcal{L}^f \phi(x,i). \tag{19}$$

Then, by Dynkin's formula for diffusions with Markovian switchings [12, p. 48], for each $T > 0$ we obtain

$$\mathbb{E}^{x,i,f} \phi(x(T), \psi(T)) = \phi(x,i) + \mathbb{E}^{x,i,f}\Big[\int_0^T \mathcal{L}^f \phi(x(s), \psi(s))ds\Big],$$

and using (19)

$$\mathbb{E}^{x,i,f}\Big[\int_0^T 2(r(x(s), \psi(s), f) - g)h(x(s), \psi(s))ds\Big]$$

$$\geq \sigma^2 t + \phi(x,i) - \mathbb{E}^{x,i,f}\phi(x(T),\psi(T)).$$

Replacing (17) in the latter inequality gives

$$\mathbb{E}^{x,i,f}\Big[\int_0^T 2(r(x(s),\psi(s),f)-g)h_f(x(s),\psi(s)ds\Big]$$
$$+\mu_f(h)\mathbb{E}^{x,i,f}\Big[\int_0^T 2(r(x(s),\psi(s),f)-g)ds\Big]$$
$$\geq \sigma^2 T + \phi(x,i) - \mathbb{E}^{x,i,f}\phi(x(T),\psi(T)).$$

Thus, multiplying by T^{-1} both sides of this inequality and letting $T\to\infty$, it follows from (13) and the $w-$exponencial ergodicity (6), that for all $f\in\mathbb{F}_{ao}$,

$$\sigma^2(f)\geq\sigma^2.$$

Hence, $\inf_{f\in\mathbb{F}_{ao}}\sigma^2(f)\geq\sigma^2$. Now let us consider f^* in $\mathbb{F}_{ao}$ that minimizes (16) and proceeding as above we obtain that $\sigma^2(f^*)=\sigma^2$, and so

$$\sigma^2(f^*)=\sigma^2\leq\inf_{f\in\mathbb{F}_{ao}}\sigma^2(f). \tag{20}$$

This completes the proof of part (ii). □

Remark 4.7. *Equation (15) is called average reward Hamilton-Jacobi-Bellman equation, which is also known as the Bellman equation or the dynamic programming equation.*

5 Proof of Theorem 4.4

To prove Theorem 4.4 first note that the Dynkin's formula applied to h_f, the equation Poisson (11), and the ergodicity exponential (6) given that the total expected payoff of $f\in\mathbb{F}$ over the time interval $[0,T]$, when the initial state is $x\in\mathbb{R}^n$ (recall Definition 3.2) can be write as

$$J_T(x,i,f)=Tg(f)+h_f(x,i)+O(e^{-\delta T})$$

where $O(\cdot)$ is a residual term converging to zero as $t\to\infty$. Replacing this last equation in the limiting average variance (12), we obtain that

$$\sigma^2(x,i,f):=\lim_{T\to\infty}\frac{1}{T}\mathbb{E}^{x,i,f}\Big(\int_0^T r(x(t),\psi(t),f)dt-Tg(f)\Big)^2. \tag{21}$$

We define, for $f \in \mathbb{F}$ and $T \geq 0$

$$Y^f(T) := \int_0^T r(x(s), \psi(s), f)ds - Tg(f). \tag{22}$$

Then, replacing (22) in (21) we have

$$\sigma^2(x, i, f) := \lim_{T \to \infty} \frac{1}{T} \mathbb{E}^{x,i,f} \left[Y^f(T) \right]^2. \tag{23}$$

Seeing the equation (23) the first that comes to mind for of the Theorem 4.4 is to use the moment generating function associated to the process $Y^f(T)$. However, a key problem with moment generating functions is that moments and the moment generating function may not exist. By contrast, the *characteristic function* of the process $Y^f(T)$ always exists, and thus may be used instead.

Hence, the main idea in the proof of Theorem 4.4 consists to use the characteristic function, $C^z(T) := e^{izY^f(T)}$, $z \in \mathbb{R}$, of the process $Y^f(T)$. To show, using Ito's formula, Poisson equation (11), and integration by parts that $C^z(T)$ satisfies a certain integral equation. Then, consider the Taylor series of the $C^z(t)$ and substitute into of the integral equation obtained to find a expression for $\mathbb{E}^{x,i,f} \left[Y^f(T) \right]^2$ which gives the result.

To begin, we need the following lemma.

Lemma 5.1. *Let $Y^f(\cdot)$ be as in (22), and let h_f be the bias function defined in (9). Then*

$$\lim_{T \to \infty} \frac{1}{T} \mathbb{E}^{x,i,f} [h_f(x(T), \psi(T)) Y^f(T)] = 0. \tag{24}$$

Proof. By Ito's Lemma for semimartingales (see [8] section 8.10, page 234, or [12] section 1.8, page 48),

$$\begin{aligned} h_f(x(T), \psi(T)) &= h_f(x, i) + \int_0^T \mathcal{L}^f h_f(x(s), \psi(s))ds \\ &- \int_0^T \sum_{0 \leq s < T} q_{\psi(s), \psi(s^+)} [h_f(x(s), \psi(s^+)) \\ &- h_f(x(s), \psi(s))])ds \\ &+ \int_0^T \sum_{k=1}^n \sum_{l=1}^d \frac{\partial h_f}{\partial x_k}(x(s), \psi(s)) \sigma_{kl}(x(s), \psi(s)) dW_l(s) \\ &+ \sum_{0 \leq s < T} [h_f(x(s), \psi(s^+)) - h_f(x(s), \psi(s))]. \end{aligned} \tag{25}$$

On the other hand, by Theorem 1 in [1] we obtain that

$$\begin{aligned}\sum_{0\leq s<T}[h_f(x(s),\psi(s^+))-h_f(x(s),\psi(s))] &= \int_0^T\sum_{j\in E}[h_f(x(s),j)\\ -h_f(x(s),\psi(s))](q_0-v)(ds,j) &+ \int_0^T\sum_{j\in E}q_{\psi(s),j}[h_f(x(s),j)\\ &- h_f(x(s),\psi(s))]ds, \end{aligned}\tag{26}$$

where q_0 is the jump measure of ψ and v is the compensator of q_0.

Replacing (26) in (25) and noting that $h_f(x,i)$ satisfies the Poisson equation (11), we have

$$\begin{aligned}h_f(x(T),\psi(T)) &= h_f(x,i)+\int_0^T(r(x(s),\psi(s),f)-g(f))ds\\ &+ \int_0^T\sum_{k=1}^n\sum_{l=1}^d\frac{\partial h_f}{\partial x_k}(x(s),\psi(s))\sigma_{kl}(x(s),\psi(s))dW_l(s)\\ &+ \int_0^T\sum_{j\in E}[h_f(x(s),j)-h_f(x(s),\psi(s))](q_0-v)(ds,j).\end{aligned}\tag{27}$$

For notational ease we define

$$N_T:=\int_0^T\sum_{k=1}^n\sum_{l=1}^d\frac{\partial h_f}{\partial x_k}(x(s),\psi(s))\sigma_{kl}(x(s),\psi(s))dW_l(s)$$

and

$$M_T:=\int_0^T\sum_{j\in E}[h_f(x(s),j)-h_f(x(s),\psi(s))](q_0-v)(ds,j).$$

Next we multiply by $h_f(x(T),\psi(T))$ on both sides of (27) and taking expectations we find that

$$\begin{aligned}\mathbb{E}^{x,i,f}[h_f(x(T),\psi(T))Y^f(T)] &= \mathbb{E}^{x,i,f}[h_f^2(x(T),\psi(T))]\\ &- \mathbb{E}^{x,i,f}[h_f(x(T),\psi(T))]h_f(x,i)\\ &- \mathbb{E}^{x,i,f}[h_f(x(T),\psi(T))N_T]\\ &- \mathbb{E}^{x,i,f}[h_f(x(T),\psi(T))M_T].\end{aligned}\tag{28}$$

From the w^2-exponential ergodicity, we have that

$$\mathbb{E}^{x,i,f}[h_f^2(x(T),\psi(T))]/T\to 0 \quad\text{and}\quad \mathbb{E}^{x,i,f}[h_f(x(T),\psi(T))]/T\to 0.\tag{29}$$

Now, by Remark 2.2 and Remark 3.3(3) we obtain that N_T is square integrable martingale and, moreover

$$\mathbb{E}^{x,i,f}\Big[\sum_{0\leq s<T}|h_f(x(s),\psi(s^+))-h_f(x(s),\psi(s))|\Big]<\infty.$$

Then, it follows from Theorem 26.12 in [2] that M_T is a martingale and, furthermore, it can shown that M_T is also a square integrable martingale. Then, applying the Cauchy-Schwarz inequality to the third and fourth summand of the right-hand side of (28) we get

$$\begin{aligned}\Big(\mathbb{E}^{x,i,f}\Big[h_f(x(T),\psi(T))N_T\Big]\Big)^2 &\leq \mathbb{E}^{x,i,f}[h_f^2(x(T),\psi(T))]\\ &\cdot \mathbb{E}^{x,i,f}[N_T^2],\end{aligned} \tag{30}$$

$$\begin{aligned}\Big(\mathbb{E}^{x,i,f}\Big[h_f(x(T),\psi(T))M_T\Big]\Big)^2 &\leq \mathbb{E}^{x,i,f}[h_f^2(x(T),\psi(T))]\\ &\cdot \mathbb{E}^{x,i,f}[M_T^2].\end{aligned} \tag{31}$$

Finally, the orthogonality property of martingale differences of N_T and M_T yields

$$\mathbb{E}^{x,i,f}[N_T^2]=O(T), \tag{32}$$

and

$$\mathbb{E}^{x,i,f}[M_T^2]=O(T). \tag{33}$$

Therefore, (24) follows from (29)-(33) □

Proof of the Theorem 4.4. For $z\in\mathbb{R}$ we define the characteristic function of process $Y^f(t)$ defined in (22) as

$$C^z(T):=e^{izY^f(T)}. \tag{34}$$

Note that

$$dC^z(T)=izdY^f(T)C^z(T)=iz[r(x(t),\psi(t),f)-g(f)]C^z(T)dt. \tag{35}$$

This implies

$$C^z(T)=1+iz\int_0^T[r(x(t),\psi(t),f)-g(f)]C^z(t)dt. \tag{36}$$

As the bias function h_f of $f \in \mathbb{F}$ satisfies the Poisson equation (11), from (36) we obtain

$$C^z(T) = 1 - iz\int_0^T \mathcal{L}^f h_f(x(t),\psi(t))C^z(t)dt. \tag{37}$$

Applying Ito's formula to h_f in the interval $[0,T]$ we obtain

$$\begin{aligned} dh_f(x(t),\psi(t)) &= \mathcal{L}^f h_f(x(t),\psi(t)) \\ &- \sum_{0\le t<T} q_{\psi(t),\psi(t^+)}[h_f(x(t),\psi(t^+)) - h_f(x(t),\psi(t))]) \\ &+ \sum_{k=1}^{n}\sum_{l=1}^{d} \frac{\partial h_f}{\partial x_k}(x(t),\psi(t))\sigma_{kl}(x(t),\psi(t))dW_l(t) \\ &+ \sum_{0\le t<T}[h_f(x(t),\psi(t^+)) - h_f(x(t),\psi(t))], \end{aligned} \tag{38}$$

and multiplication of (38) by $C^z(t)$ gives

$$\begin{aligned} &\mathcal{L}^f h_f(x(t),\psi(t))C^z(t) = dh_f(x(t),\psi(t))C^z(t) \\ &+ C^z(t)\sum_{0\le t<T} q_{\psi(t),\psi(t^+)}[h_f(x(t),\psi(t^+)) - h_f(x(t),\psi(t))]) \\ &- C^z(t)\sum_{k=1}^{n}\sum_{l=1}^{d} \frac{\partial h_f}{\partial x_k}(x(t),\psi(t))\sigma_{kl}(x(t),\psi(t))dW_l(t) \\ &- C^z(t)\sum_{0\le t<T}[h_f(x(t),\psi(t^+)) - h_f(x(t),\psi(t))]. \end{aligned} \tag{39}$$

Replacing (39) in (37) we have

$$\begin{aligned} C^z(T) &= 1 - iz\Big\{\int_0^T dh_f(x(t),\psi(t)C^z(t)dt \\ &+ \int_0^T C^z(t)\sum_{0\le t<T} q_{\psi(t),\psi(t^+)}[h_f(x(t),\psi(t^+)) - h_f(x(t),\psi(t))])dt \\ &- \int_0^T C^z(t)\sum_{k=1}^{n}\sum_{l=1}^{d} \frac{\partial h_f}{\partial x_k}(x(t),\psi(t))\sigma_{kl}(x(t),\psi(t))dW_l(t)dt \\ &- \int_0^T C^z(t)\sum_{0\le t<T}[h_f(x(t),\psi(t^+)) - h_f(x(t),\psi(t))]dt\Big\}. \end{aligned} \tag{40}$$

Using integration by parts we get

$$\int_0^T dh_f(x(t),\psi(t))C^z(t)dt = h_f(x(T),\psi(T))C^z(T) - h_f(x,i) + \int_0^T h_f(x(t),\psi(t))dC^z(t). \tag{41}$$

Therefore, replacing (41) in (40), taking expectations, and using the arguments in the proof of Lemma 5.1 we obtain

$$\mathbb{E}^{x,i,f}[C^z(T)] = 1 - iz\mathbb{E}^{x,i,f}\Big[h_f(x(T),\psi(T))C^z(T) - h_f(x,i) + \int_0^T h_f(x(t),\psi(t))dC^z(t)dt\Big]. \tag{42}$$

Finally, substituting (35) in (42) we obtain

$$\mathbb{E}^{x,i,f}[C^z(T)] = 1 - iz\mathbb{E}^{x,i,f}\Big[h_f(x(T),\psi(T)C^z(T) - h_f(x,i) + \int_0^T h_f(x(t),\psi(t))iz[r(x(t),\psi(t),f) - g(f)]C^z(t)dt\Big].$$

Consider now the Taylor series of $C^z(t)$ in the last equality

$$\mathbb{E}^{x,i,f}\Big[\sum_{k=0}^{\infty}\frac{(izY^f(T))^k)}{k!}\Big] = 1 - iz\mathbb{E}^{x,i,f}\Big[h_f(x(T),\psi(T))\Big(\sum_{k=0}^{\infty}\frac{(izY^f(T))^k)}{k!}\Big) - h_f(x,i) + \int_0^T h_f(x(t),\psi(t))iz[r(x(t),\psi(t),f) - g(f)]\Big(\sum_{k=0}^{\infty}\frac{(izY^f(T))^k)}{k!}\Big)dt\Big].$$

Equating second order terms in z we have

$$\mathbb{E}^{x,i,f}[Y^f(T)^2] = 2\mathbb{E}^{x,i,f}[h_f(x(T),\psi(T))Y^f(T)] + 2\mathbb{E}^{x,i,f}\Big[\int_0^T h_f(x(t),\psi(t))[r(x(t),\psi(t),f) - g(f)]dt\Big]. \tag{43}$$

It is easy prove that $h_f(x,i)[r(x,i,f(x,i)) - g(f)]$ is in $\mathbb{B}_{w^2}(\mathbb{R}^n \times E)$ and also that $x(\cdot)$ is w^2- exponentially ergodic (recall Remark 4.2). Hence multiplying (43) by $1/t$ and letting $t \to \infty$ the result (13) follows from Lemma 5.1 and the w^2-exponential ergodicity of $x(\cdot)$. □

6 An Example

Now we give an example to illustrate our results. This example is an extension of the one presented in [3]. Consider the scalar linear system

$$(44)\quad dx(t) = [b(\psi(t))x(t) + \beta u(t)]dt + \sigma dW(t), \quad x(0) = x, \quad \psi(0) = i,$$

where $b : E \to \mathbb{R}$ and the coefficients β, σ are given positive constants. The control $u(t)$ takes values in the compact set $U := [0, a]$, with $a > 0$. The controlled Markov-modulated diffusion (44) satisfies the Assumption 2.1.

Now let $r(x, i, u) := u$ be the reward rate. Choose a function $w(x, i)$ that satisfies Assumptions 2.4 and 3.1, respectively. Our goal is to find stationary policies $u(t) := f(x(t), \psi(t))$ that minimize the limiting average variance. To this end, first, we find stationary policies that optimize the long-run expected average reward and within this set we search variance optimal policies. To do this, we will use the equation (15), which in the present case takes the form

$$(45)\quad g = \max_{u\in[0,a]} \{u + h_x(x, i)[b(i)x + \beta u] + \frac{1}{2}\sigma^2 h_{xx}(x, i) + \mathcal{Q}h(x, i)\}.$$

For the particular case when $h_x(x, i) < -1/\beta$ for all $i \in E$ and $x \in \mathbb{R}^n$, the control policy $f^*(x, i) = 0$ is the unique policy that attains the maximum in (45) (hence, it is the unique average optimal policy). Consequently, by uniqueness of f^*, f^* it is also variance optimal.

Acknowledgement

We wish to thank Dr. Onésimo Hernández-Lerma for his valuable comments on an early version of this paper.

B. A. Escobedo-Trujillo
Civil Engineering Faculty
Universidad Veracruzana
Coatzacoalcos, Ver., México
Tel. (921) 218-77-83, 211-57-07
bescobedo@uv.mx

C. O. Rivera-Blanco
Centro de Investigación en Recursos Energéticos y Sustentables
Universidad Veracruzana
Coatzacoalcos, Ver., México
Tel. (921) 218-77-83, 211-57-07
crivera@uv.mx

References

[1] Bäuerle N; Rieder U., *Portfolio optimization with Markov-modulated stock prices and interest rates*, IEEE Trans. Automatic Control. **49** (2004), 442-447.

[2] Davis M. H. A., Markov Models and Optimization, Chapman & Hall, London, 1993.

[3] Escobedo-Trujillo B.A.; Hernández-Lerma O., *Overtaking optimality for controlled Markov-modulated diffusions*, J. Optimization. , (2011).

[4] Ghosh M.K.; Arapostathis A.; Marcus S.I., *Optimal control of switching diffusions with applications to flexible manufacturing systems*, SIAM J. Control Optim. **31** (1993), 1183-1204.

[5] Ghosh M.K.; Arapostathis A.; Marcus S.I., *Ergodic control of switching diffusions*, SIAM J. Control Optim. **35** (1997), 1962-1988.

[6] Hernández-Lerma O.; Vega-Amaya O.; Carrasco G., *Sample-path optimality and variance-minimization of average cost Markov control processes*, SIAM J. Control Optim. **38** (1999), 79-93.

[7] Jasso-Fuentes H.; Hernández-Lerma O., *Optimal ergodic control of Markov diffusion processes with minimum variance*, Stochastics. Electronic version.

[8] Klebaner F. C., Introduction to Stochastic Calculus with Applications, Imperial College Press, London, second edition, 2005.

[9] Mandl P., *On the variance in controlled Markov chains*, Kybernetika. (Prague) **7** (1971), 1-12.

[10] Mandl P., *An application of Ito's formula to stochastic contol systems*, Lecture Notes in Mathematics **294** (1972), 8-13.

[11] Mandl P., *A connection between controlled Markov chains and martingales*, Kybernetika. (Prague) **9** (1973), 237-241.

[12] Mao X.; Yuan C., Stochastic Differential Equations with Markovian Switching, Imperial College Press, London, 2006.

[13] Prieto-Rumeau T.; Hernández-Lerma O., *Variance minimization and overtaking optimality approach to continuous-time controlled Markov chains*, Math. Meth. Oper. Res. **70** (2009), 527-540.

[14] Yin G., Zhu C.; *On notion of weak stability and related issues of hybrid diffusion systems*, Nonlinear Analysis: Hybrid Systems **1** (2007), 173-187.

[15] Zhu C.; Yin G., *Asymptotic properties of hybrid diffusion systems*, SIAM J. Control Optim. **46** (2007), 1155-1179.

[16] Yin G.; Zhu C., Hybrid Switching Diffusions: Properties and Applications, Springer, New York, 2010.