mathematics

MDPI

Article

Optimal Control with Partially Observed Regime Switching: Discounted and Average Payoffs

Beatris Adriana Escobedo-Trujillo [1,*], Javier Garrido-Meléndez [1], Gerardo Alcalá [2] and J. D. Revuelta-Acosta [1]

[1] Facultad de Ingeniería Campus Coatzacoalcos, Universidad Veracruzana, Coatzacoalcos 96535, Veracruz, Mexico; jgarrido@uv.mx (J.G.-M.); jrevuelta@uv.mx (J.D.R.-A.)
[2] Centro de Investigación en Recursos Energéticos y Sustentables, Universidad Veracruzana, Coatzacoalcos 96535, Veracruz, Mexico; galcala@uv.mx
* Correspondence: bescobedo@uv.mx

Abstract: We consider an optimal control problem with the discounted and average payoff. The reward rate (or cost rate) can be unbounded from above and below, and a Markovian switching stochastic differential equation gives the state variable dynamic. Markovian switching is represented by a hidden continuous-time Markov chain that can only be observed in Gaussian white noise. Our general aim is to give conditions for the existence of optimal Markov stationary controls. This fact generalizes the conditions that ensure the existence of optimal control policies for optimal control problems completely observed. We use standard dynamic programming techniques and the method of hidden Markov model filtering to achieve our goals. As applications of our results, we study the discounted linear quadratic regulator (LQR) problem, the ergodic LQR problem for the modeled quarter-car suspension, the average LQR problem for the modeled quarter-car suspension with damp, and an explicit application for an optimal pollution control.

Keywords: ergodicity; filtering theory; hidden Markov models; partial observation; Wonham filter

MSC: 49N05; 49N10; 49N30; 49N90; 93C41

check for updates

Citation: Escobedo-Trujillo, B.A.; Garrido-Meléndez, J.; Alcalá, G.; Revuelta-Acosta, J.D. Optimal Control with Partially Observed Regime Switching: Discounted and Average Payoffs. *Mathematics* **2022**, *10*, 2073. https://doi.org/10.3390/math10122073

Academic Editor: Panagiotis-Christos Vassiliou

Received: 18 May 2022
Accepted: 10 June 2022
Published: 15 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, there has been more attention to a class of optimal control problems where the dynamic systems are governed means switching diffusions in which the switching is modeled by a continuous-time Markov chain (ψ) with unobservable hidden states (also known as partially observed optimal control problems). In these problems, an observable process y whose outcomes are "influenced" by the outcomes of ψ in a known way is assumed. Since ψ cannot be observed directly, the goal is to learn about ψ by observing y. Following the last mentioned, this article concerns with an optimal control problem with discounted and ergodic payoff in which the dynamic system $x(t)$ evolves according to a Markovian regime-switching diffusion $dx(t) = f(x(t), \psi(t))dt + \sigma(x(t), \psi(t))dW(t)$ for given continuous functions f and σ. The reward rate is allowed to be unbounded from above and from below. In this paper, the Wonham filter to estimate the states of the Markov chain from the observable evolution of a given process (y) is used. As a result, the original system $x(t)$ is converted to a completely observable one $\overline{x}(t)$.

Our main results extend the dynamic programming technique to this family of stochastic optimal control problems with reward (or cost) rate per unit of time unbounded and Markovian regime-switching diffusions. The regime switching is modeled by a continuous-time Markov chain (ψ) with unobservable states. Early works include research on an optimal control problem with an ergodic payoff, considering that the dynamic system evolves according to Markovian switching diffusions. However, this diffusion does not depend on a hidden Markov chain [1]. Research on deriving the dynamic programming

principle for a partially observed optimal control problem in which the dynamic system is governed by a discrete-time Markov control process taking values in a finite-dimensional space has also been proposed [2]. Finally, one paper studied the optimal control with Markovian switching that is completely observable and rewards rate unbounded [3]. As an application of our results, we study the discounted linear quadratic regulator (LQR) problem, the ergodic LQR problem for the modeled quarter-car suspension, the average (ergodic) LQR problem for the modeled quarter-car suspension with damp, and an explicit application for an optimal pollution control. Other applications with bounded payoff different from those studied in this work are found in [4–6].

The objective of the theory of controlled regime-switching diffusions is to model controlled diffusion systems whose dynamics are affected by discrete phenomena. In these systems, the discrete phenomena are modeled by a Markov chain in continuous time, whose states represent the discrete phenomenon involved. There is an extensive list of references dealing with the case of completely observable stochastic optimal control in which a switching diffusion governs the stochastic systems. A literature review includes the textbooks [7,8] and the papers [9–14], with several applications, including optimization portfolios, wireless communication systems, and wind turbines, among others.

Generally, to solve unobserved optimal control problems, where the dynamic systems are governed by a hidden Markovian switching diffusion, it is necessary to transform them into completely observed ones, which in our case is done using a Wonham filter.

This Wonham filter estimates the hidden state of the Markov chain from the observable evolution of the process y. When these estimates are replaced in the original system, this becomes a completely observable system [15,16] and ([17], Section 22.3). The numerical results for Wonham's filter are given in [18].

The paper is organized as follows: in Section 1, an introduction is given. In Section 2, the main assumptions are given. In this section, the partially observable system is converted into an observable system. The conditions to ensure the existence of optimal solutions for the optimal control problem with discounted payoff are given in Section 3. In Section 4, the conditions to ensure the existence of optimal solutions for the optimal control problem with average payoff are deduced. To illustrate our results, four applications are developed: an application on a linear quadratic regulator (LQR) with discounted payoff (Section 5); the development of a model of a quarter-car suspension LQR with an average payoff (Section 6); the study of an optimal control of a vehicle active suspension system with damp (Section 7); and an explicit application for an optimal pollution control (Section 8).

2. Formulation of the Problem

This work focuses on controlled hybrid stochastic differential Equations (HSDE) under partial observation. To explain this, first, we consider the stochastic differential equations of the form:

$$\mathrm{d}x(t) = b(x(t), \psi(t), u(t))\mathrm{d}t + \sigma(x(t), \psi(t))\mathrm{d}W(t), \quad x(0) = x_0, \ \psi(0) = i, \tag{1}$$

where $b : \mathbb{R}^n \times E \times \mathcal{U} \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times E \to \mathbb{R}^{n\times d}$ in (1) depend on a finite state and time-continuous irreducible and aperiodic Markov chain $\psi(\cdot)$ taking values in $E = \{1, \ldots, N\}$. For all $i, j \in E$ the transition probabilities are given by:

$$\mathbb{P}(\psi(s+t)) = j \mid \psi(s) = i = \begin{cases} q_{ij}t + o(t), & \text{if } i \neq j, \\ 1 + q_{ii}t + o(t), & \end{cases}$$

where the constants $q_{ij} \geq 0$ are the transition rates from i to j and satisfy that $q_{ii}(x) = -\sum_{i\neq j} q_{ij}(x)$, the transition matrix is denoted by $Q = \{q_{ij}\}_{i,j=1,2,\ldots,N}$. The control component is $u(t) \in \mathcal{U}$ with $\mathcal{U}$ a compact set of $\mathbb{R}^m$, and W is a d-dimensional standard Brownian motion independent of $\psi(\cdot)$. Throughout the work, it is considered that both the Markov chain $\psi(\cdot)$ and the Brownian motion W are defined on a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ that satisfies the usual conditions.

Until now, the switching diffusion (1) seems to be formulated as a classical switching diffusion, as in [11–14,19], among others. However, we propose that the process ψ is a hidden Markov chain, i.e., at any given instant of time, the exact state of the Markov chain $\psi(\cdot)$ cannot be observed directly. Instead, we can only observe the process y given by:

$$dy(t) = h(\psi(t))\mathrm{d}t + \sigma_0 \mathrm{d}B(t), \quad y(0) = 0, \tag{2}$$

whose dynamics depends on the value of $\psi(\cdot)$. In Equation (2), $h : E \to \mathbb{R}$ is a bounded function, whereas B is a one-dimensional Brownian motion independent of W and ψ, and σ_0 is a positive constant.

Under partial observation, the best way to work is through nonlinear filtering. This technique studies the conditional distribution of $\psi(t)$ given the observed data accumulated up to time t, namely:

$$\Psi_i(t) = \mathbb{P}(\psi(t) = i \mid \sigma_1(y(s),\ 0 \leq s \leq t)), \quad \forall i \in E, \tag{3}$$

where $\sigma_1(y(s),\ 0 \leq s \leq t))$ is the σ_1-algebra generated by the process $y(t)$ and $\sum_{i=1}^{N} \Psi_i(t) = 1$. Taking into account the following notation:

$$\begin{aligned} h^T(\Psi) &= (h(1), h(2), \ldots, h(N)), \\ \Psi^T(t) &= (\Psi_1(t), \ldots, \Psi_N(t)), \\ \mathrm{diag}(h) &= \mathrm{diag}(h(1), \ldots, h(N)), \end{aligned}$$

and using the Wonham filtering techniques, we know that the process Ψ in (3) satisfies the following Equation (see for instance [15] or ([17], Section 22.3)):

$$\begin{aligned} \mathrm{d}\Psi(t) = &\Big[Q\Psi(t) - \sigma_0^{-2} h^T(\Psi(t))\Big(\mathrm{diag}(h) - h^T(\Psi(t))I_N\Big)\Psi(t)\Big]\mathrm{d}t \\ &+\sigma_0^{-2}\Big(\mathrm{diag}(h) - h^T(\Psi(t))I_N\Big)\Psi(t)\mathrm{d}y(t), \end{aligned} \tag{4}$$

where I_N is the $N \times N$ identity matrix. If we introduce the process:

$$\mathrm{d}w_0(t) = \sigma_0^{-1}(\mathrm{d}y(t) - h^T(\Psi(t))dt),$$

then Equation (4) can be rewritten as:

$$\mathrm{d}\Psi(t) = Q\Psi(t)\mathrm{d}t + \sigma_0^{-1}\Big(\mathrm{diag}(h) - h^T(\Psi(t))I_N\Big)\Psi(t)\mathrm{d}w_0(t). \tag{5}$$

Remark 1. *Note that the unique solution of* (5) *exists up to an explosion time* τ *(see, for instance [20]). However,* $\tau = \infty$ *a.s. since* $\Psi_i(t) \leq 1$ *for all* $t < \tau$ *and* $\forall i \in E$.

At this point, we have defined the controlled HSDE with partial observation. To fulfill the objective of this work, that is, to solve an optimal control problem with the discounted and average payoff with partial observation, we will transform this problem into one with complete observation (see for instance [5,6,16]). First, we will establish the following notational convention.

For the coefficients $b : \mathbb{R}^n \times E \times \mathcal{U} \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times E \to \mathbb{R}^{n \times d}$

$$\begin{aligned} b(x(t), \psi(t), u(t)) &= (b_1(x(t), \psi(t), u(t)), \ldots, b_n(x(t), \psi(t), u(t))), \\ \sigma(x(t), \psi(t)) &= \{\sigma_{kl}(x(t), \psi(t))\}_{k=1,\ldots,n; l=1,\ldots,d}, \end{aligned}$$

we have their filtered estimates:

$$\overline{b}_k(x(t),\Psi(t),u(t)) = \sum_{i=1}^{N} \Psi_i(t) b_k(x(t),i,u(t)), \tag{6}$$

$$\overline{\sigma}_{kl}(x(t),\Psi(t)) = \sum_{i=1}^{N} \Psi_i(t) \sigma_{kl}(x(t),i), \tag{7}$$

and with equalities (6)–(7), we establish the new coefficients:

$$\overline{b}(x(t),\Psi(t),u(t)) = (\overline{b}_1(x(t),\Psi(t),u(t)),\ldots,\overline{b}_n(x(t),\Psi(t),u(t))),$$
$$\overline{\sigma}(x(t),\Psi(t)) = \{\overline{\sigma}_{kl}(x(t),\Psi(t))\}_{k=1,\ldots,n;l=1,\ldots,d}$$

With the use of above functions and Equation (1), we introduce the components of a new diffusion process as:

$$\mathrm{d}x_k(t) = \overline{b}_k(x(t),\Psi(t),u(t))\mathrm{d}t + \sum_{l=1}^{d} \overline{\sigma}_{kl}(x_k(t),\Psi(t))\mathrm{d}W_l(t), \quad x(0) = x_0, \tag{8}$$

and therefore, we obtain from (5) and (8) the following controlled system with complete observation:

$$\begin{cases} \mathrm{d}x(t) = \overline{b}(x(t),\Psi(t),u(t))\mathrm{d}t + \overline{\sigma}(x(t),\Psi(t))\mathrm{d}W(t), \\ \mathrm{d}\Psi(t) = Q\Psi(t)\mathrm{d}t + \sigma_0^{-1}\big(\mathrm{diag}(h) - h^T(\Psi(t))I_N\big)\Psi(t)\mathrm{d}w_0(t), \end{cases} \tag{9}$$

where $(x(t),\Psi(t)) \in \mathbb{R}^n \times S_N$ with:

$$S_N = \{\Psi = (\Psi_1,\ldots,\Psi_N) \in \mathbb{R}^N |\ \Psi_i(t) > 0,\ \sum_{i=1}^{N} \Psi_i(t) = 1\}.$$

Throughout this work, we will use the following Assumption 1.

Assumption 1.

(a) *The control set $\mathcal{U}$ is compact.*

(b) *$b : \mathbb{R}^n \times E \times \mathcal{U} \to \mathbb{R}^n$ is a continuous function that satisfies the Lipschitz continuous property on x uniformly in $(i,u) \in E \times \mathcal{U}$, that is, there exists a constant $C_1 > 0$ such that:*

$$\max_{(i,u)\in E\times U} \|b(x,i,u) - b(y,i,u)\| \le C_1\|x-y\|.$$

(c) *There exists constants $C_2, C_3 > 0$ such that, $\sigma : \mathbb{R}^n \times E \to \mathbb{R}^{n\times d}$ satisfies:*

$$\|\sigma(x,i) - \sigma(y,i)\| \le C_2\|x-y\| \quad \text{and} \quad x^T\sigma(x,i)\sigma^T(x,i)x \ge C_3\|x\|^2$$

for all $x, y \in \mathbb{R}^n$ and for all $i \in E$.

(d) *There exists $C_4, C_5 > 0$ with:*

$$\|\sigma(x,i)\| \le C_4(1+\|x\|) \text{ and } \|b(x,i,u)\| \le C_5(1+\|x\|)$$

for $i \in E$ and $u \in \mathcal{U}$.

Under Assumption 1 and taking into account Remark 1, we know that the system (9) has a unique solution.

For $x \in \mathbb{R}^n$, we denote by $\nabla \nu_x$ and $\mathbb{H}_x$ the gradient and the Hessian matrix of x, respectively, and $\langle \cdot, \cdot \rangle$ the scalar product. For a sufficiently smooth real-valued function $\nu : \mathbb{R}^n \times \mathbb{R}^N \to \mathbb{R}$. Let:

$$\begin{aligned}\mathbb{L}^{u,\Psi}\nu(x,\Psi) &:= \left\langle \nabla \nu_x, \overline{b}(x,\Psi,u) \right\rangle + \frac{1}{2} Tr\Big[(\mathbb{H}_x \nu) a(x,\Psi)\Big] \\ &\quad + \langle \nabla \nu_\Psi, Q\Psi(t) \rangle + \frac{1}{2\sigma_0^2} Tr\Big[(\mathbb{H}_\Psi \nu((x,\Psi))) A_2(\Psi(t))\Big]\end{aligned}$$

with

$$a(x,\Psi) = \overline{\sigma}(x,\Psi)\overline{\sigma}(x,\Psi)^T,$$

$$A_2(\Psi(t)) = [\Big(\mathrm{diag}(h) - h^T(\Psi(t)) I_N\Big)\Psi(t)][\Big(\mathrm{diag}(h) - h^T(\Psi(t)) I_N\Big)\Psi(t)]^T,$$

the operator associated with Equation (9). In order to carry out the aim of this work, we define the control policies.

Definition 1. *A function of the form $u(t) := f(t, x(t), \Psi(t))$ for some measurable function $f : [0, \infty) \times \mathbb{R}^n \times S_N \to \mathcal{U}$, is called a Markov policy, whereas $u(t) := f(x(t), \Psi(t))$ for some measurable function $f : \mathbb{R}^n \times S_N \to \mathcal{U}$ is said to be a stationary Markov policy. The stationary Markov policies set is denote by $\mathbb{F}$.*

The following assumption represents a Lyapunov-like condition.

Assumption 2. *There exists a function $(w \geq 1) \in \mathrm{C}^2(\mathbb{R}^n \times S_N)$, and constants $p \geq q > 0$, such that:*

(i) $\lim_{|x| \to \infty} w(x,\Psi) = +\infty$, *and*
(ii) $\mathbb{L}^{u,\Psi} w(x,\Psi) \leq -qw(x,\Psi) + p$ *for each* $u \in \mathcal{U}$ *and* $(x,\Psi) \in \mathbb{R}^n \times S_N$.

It is important to point out that since the $\psi(\cdot)$ is irreducible and aperiodic, we can ensure the existence of a unique invariant measure for the Markov–Feller process $(x^f(\cdot), \Psi(\cdot))$ (see [21,22]). Moreover, the Assumption 2 allows us to conclude that the Markov process $(x^f(\cdot), \Psi(\cdot))$, where $f \in \mathbb{F}$ is positive recurrent and there exists a unique invariant probability measure $\mu_f(dx, \Psi)$ for which is satisfied:

$$\mu_f(w) := \int_{\mathbb{R}^n \times S_N} w(x,\Psi) \mu_f(dx, d\Psi) < \infty. \tag{10}$$

Note that for every $f \in \mathbb{F}$, the measure μ_f belongs to the space defined as follows.

Definition 2. *The w-norm is defined as:*

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

where ν is the real-valued measurable function on $\mathbb{R}^n \times S_N$ and w is the Lyapunov function given in Assumption 2. The normed linear space of real-valued measurable functions ν with finite w-norm is denoted by $\mathcal{B}_w(\mathbb{R}^n \times S_N)$. Moreover, the normed linear space of finite signed measures μ on $\mathbb{R}^n \times S_N$ such that:

$$\| \mu \|_w := \int_{\mathbb{R}^n} w(x,\Psi) \mid \mu \mid (dx, d\Psi) < \infty,$$

where $\mid \mu \mid$ is the total variation of μ is denoted by $\mathbb{M}_w(\mathbb{R}^n \times S_N)$.

Remark 2. *For each $\nu \in \mathcal{B}_w(\mathbb{R}^n \times S_N)$ and $\mu \in \mathbb{M}_w(\mathbb{R}^n \times S_N)$, we get:*

$$\Big| \int \nu(x,\Psi) \mu(dx, d\Psi) \Big| \leq \| \nu \|_w \int w(x,\Psi) \mid \mu \mid (dx, d\Psi) = \| \nu \|_w \| \mu \|_w < \infty,$$

that is, the integral $\int v(x,\Psi)\mu(dx,\Psi)$ *is finite.*

The next result will be useful later.

Lemma 1. *The condition* (ii) *in Assumption 2 implies that:*
(a) $\mathbb{E}^{x,\Psi,f}[w(x(t),\Psi(t))] \leq e^{-qt}w(x,\Psi) + \frac{p}{q}(1-e^{-qt})$;
(b) $\lim_{t\to\infty}\frac{1}{t}\mathbb{E}^{x,\Psi,f}[w(x(t),\Psi(t))] = 0$ *for all* $f \in \mathbb{F}$, $(x,\Psi) \in \mathbb{R}^n \times S_N$, *and* $t \geq 0$;
(c) $\mu_f(w) \leq \frac{p}{q}$ *for all* $h \in \mathbb{F}$.

Proof. (a) After applying Dynkin's formula to the function $e^{qt}w$, we use case (ii) of Assumption 2 to get:

$$\begin{aligned}\mathbb{E}^{x,\Psi,f}[e^{qt}w(x(t),\Psi(t)] &= w(x,\Psi_0) + \mathbb{E}^{x,\Psi,f}\Big[\int_0^t e^{qs}[\mathbb{L}^{u,\Psi}w(x(s),\Psi(s)) + qw(x(s),\Psi(s))]ds\Big]\\ &\leq w(x,\Psi_0) + \mathbb{E}^{x,\Psi,f}\Big[\int_0^t e^{qs}pds\Big]\\ &\leq w(x,\Psi_0) + \frac{p}{q}(e^{qt}-1).\end{aligned} \tag{11}$$

Finally, if we multiply the inequality (12) by e^{-qt}, we obtain the result. To prove (b), it is enough take the limit from the inequality (12). Integrating both sides of (12) with respect to the invariant probability μ_f, we obtain $\mu_f(w) \leq e^{-qt}\mu_f(w) + \frac{p}{q}(1-e^{-qt})$, i.e., $\mu_f(w) \leq p/q$; thus, the result (c) follows. □

In this work, the *reward rate* is a measurable function $r : \mathbb{R}^n \times E \times \mathcal{U} \to \mathbb{R}$ that satisfies the following conditions:

Assumption 3.

(a) *The function* $r(x,i,u)$ *is continuous on* $\mathbb{R}^n \times E \times \mathcal{U}$; *moreover, for each* $R > 0$, *there exists a constant* $K(R) > 0$ *such that:*

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

i.e., r is locally Lipschitz in x uniformly with respect to $i \in E$ *and* $u \in U$.

(b) $r(\cdot,\cdot,u)$ *is in the normed linear space of real-valued functions* $\mathcal{B}_w(\mathbb{R}^n \times E)$ *uniformly in u; that is, there exists* $M > 0$ *such that for all* $(x,i) \in \mathbb{R}^n \times E$:

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

Notation. The rate reward $r : \mathbb{R}^n \times E \times \mathcal{U} \to \mathbb{R}$ is vector form is given by:

$$r^T(x,\Psi,u) = (r(x,1,u), r(x,2,u), \ldots, r(x,N,u)),$$

and its estimation is:

$$\bar{r}(x,\Psi(t),u) = \Psi^T(t)r(x,\Psi,u) = \sum_{i=1}^{N}\Psi_i(t)r(x,i,u). \tag{12}$$

Henceforth, for each stationary Markov policy $f \in \mathbb{F}$, we write:

$$\bar{r}(x,\Psi,f) := \bar{r}(x,\Psi,f(x,i)).$$

3. The Discounted Case

The objective of this section is to give conditions that guarantee the existence of discounted optimal policies for the α-discounted payoff criterion we are concerned with.

Definition 3. *Let r be as in Assumption 3 and α a positive constant. Given a stationary Markov policy $f \in \mathbb{F}$ and an initial state $x(0) = x, \Psi(0) = \Psi$, the total expected discount payoff (or discounted payoff, for short) is defined as:*

$$V_\alpha(x, \Psi, f) := \mathbb{E}^{x,\Psi,f}\Big[\int_0^\infty e^{-\alpha t}\bar{r}(x(t), \Psi(t), f)dt\Big].$$

Observe that the value function does not depend on the time at which the optimal control problem is studied to get the stationarity of the problem.

The following result shows a bound of the total expected discount payoff given in Definition 3. We will omit its proof because it is a direct consequence of Assumption 3 and inequality in Lemma 1a.

Proposition 1. *Suppose that Assumptions 2 and 3b hold. Then, for each x in $\mathbb{R}^n$, $\Psi \in S_N$ and $f \in \mathbb{F}$ we have:*

$$\sup_{f\in\mathbb{F}}|V_\alpha(x, \Psi, f)| \leq M(\alpha)w(x, \Psi) \text{ with } M(\alpha) := M\frac{\alpha + d}{\alpha c}.$$

implying that α-discounted payoff $V_\alpha(\cdot, \cdot, f)$, belongs to the space $\mathcal{B}_w(\mathbb{R}^n \times S_N)$. Here, q and p are as in Assumption 2 and M is the constant in Assumption 3b.

α-discounted optimal problem. The optimal control problem with discounted payoff consists of finding a policy $f^* \in \mathbb{F}$ such that:

$$V_\alpha^*(x, \Psi) = V_\alpha(x, \Psi, f^*) = \sup_{f\in\mathbb{F}} V_\alpha(x, \Psi, f). \tag{13}$$

The function $V_\alpha^*(x, \Psi)$ is referred to as *the optimal discount payoff*, whereas the policy $f^* \in \mathbb{F}$ is called *the discounted optimal*.

Definition 4. *We say that a function $v \in C^2(\mathbb{R}^n \times S_N) \cap \mathcal{B}_w(\mathbb{R}^n \times S_N)$, and a policy $f^* \in \mathbb{F}$ verify (are a solution of) the* α-discounted payoff optimality equations *(or Hamilton–Jacobi–Bellman (HJB) equation) if, for every $x \in \mathbb{R}^n$ and $\Psi \in S_N$:*

$$\alpha v(x, \Psi) = \bar{r}(x, \Psi, f^*) + \mathbb{L}^{f^*,\Psi}v(x, \Psi) \tag{14}$$

$$= \sup_{f\in\mathbb{F}}\Big\{\bar{r}(x, \Psi, f) + \mathbb{L}^{f,\Psi}v(x, \Psi)\Big\}. \tag{15}$$

Proposition 2. *If Assumptions 1, 2, and 3 hold, then:*

(a) There exists a function v in $C^2(\mathbb{R}^n \times S_N) \cap \mathcal{B}_w(\mathbb{R}^n \times S_N)$ and a policy $f^* \in \mathbb{F}$, such that (14) and (15) hold.

(b) The function v coincides with $V_\alpha^*(x, \Psi)$ in (13).

(c) A policy $f^* \in \mathbb{F}$ is an α-discount optimal if and only if (14) and (15) are satisfied.

Proof.

(a) Theorem 3.2 in [23] ensures that the value function $V_\alpha(x,\Psi)$ defined in (13) considering $\Psi \equiv 0$ is the unique solution of the HJB Equation (14) in $C^2(\mathbb{R}^n) \cap \mathcal{B}_w(\mathbb{R}^n)$. The existence of a function v in $C^2(\mathbb{R}^n \times S_N) \cap \mathcal{B}_w(\mathbb{R}^n \times S_N)$ and a policy $f^* \in \mathbb{F}$, such that (14) and (15) hold, follows from Theorem 3.1 and 3.2 in [23] for each $\Psi \in S_N$ fixed.

(b) By Dynkin's formula for all $(x,\Psi) \in \mathbb{R}^n \times S_N$, $f \in \mathbb{F}$ and $t \geq 0$:

$$\mathbb{E}^{x,\Psi,f}[e^{-\alpha t}v(x(t),\Psi(t))] = v(x,\Psi) + \mathbb{E}^{x,\Psi,f}\Big[\int_0^T \mathbb{L}^{f,\Psi}\Big[e^{-\alpha t}v(x(t),\Psi(t))dt\Big] \tag{16}$$

Observe that:

$$\begin{aligned}\mathbb{L}^{f,\Psi}\Big[e^{-\alpha t}v(x(t),\Psi(t))\Big] &= -\alpha e^{-\alpha t}v(x,\Psi)\\ &+ e^{-\alpha t}\bar{b}(x,\Psi,f)v_x(x,\Psi)\\ &+ e^{-\alpha t}\frac{1}{2}Tr(a(x,\Psi))v_{xx}(x,\Psi)\\ &= e^{-\alpha t}[-\alpha v(x(t),\Psi(t)) + \mathbb{L}^{f,\Psi}v(x(t),\Psi(t))].\end{aligned}$$

Therefore, the right-hand member of (16) equals:

$$\mathbb{E}^{x,\Psi,f}[e^{-\alpha t}v(x(t),\Psi(t))] = v(x,\Psi) + \mathbb{E}^{x,\Psi,f}\Big[e^{-\alpha t}(\mathbb{L}^{f,\Psi}v(x(t),\Psi(t)) - \alpha v(x(t),\Psi(t)))dt\Big]$$

and from (15):

$$\mathbb{E}^{x,\Psi,f}[e^{-\alpha t}v(x(t),\Psi(t))] \leq v(x,\Psi) - \mathbb{E}^{x,\Psi,f}\left[\int_0^T e^{-\alpha t}\bar{r}(x(t),\Psi(t),f)dt\right].$$

This yields:

$$v(x,\Psi) \geq \mathbb{E}^{x,\Psi,f}\left[\int_0^t [e^{-\alpha t}\bar{r}(x(t),\Psi(t),f)dt\right] + \mathbb{E}^{x,\Psi,f}[e^{-\alpha t}v(x(t),\Psi(t))].$$

Now, as a consequence of v is in $\mathcal{B}_w(\mathbb{R}^n \times S_N)$ and Lemma 1 (a),(b), we have that:

$$\begin{aligned}|\mathbb{E}^{x,\Psi,f}[e^{-\alpha t}v(x(t),\Psi(t))]| &\leq \mathbb{E}^{x,\Psi,f}[[e^{-\alpha t}\|v\|_w w(x(t),\Psi(t))]\\ &\leq e^{-\alpha t}\|v\|_w \mathbb{E}^{x,\Psi,f}w(x(t),\Psi(t))\\ &\leq e^{-\alpha t}\|v\|_w\left[e^{-qT}w(x,\Psi) + \frac{p}{q}(1-e^{-qT})\right] \text{ (by Lemma 1(a))}\\ &\to 0 \quad \text{as} \quad t \to \infty.\end{aligned}$$

Therefore:

$$v(x,\Psi) \geq \mathbb{E}^{x,\Psi,f}\left[\int_0^\infty [e^{-\alpha s}\bar{r}(x(s),\Psi(s),f)ds\right] = V_\alpha(x,\Psi,f) \quad \text{for all } f \in \mathbb{F}.$$

Thus, $v(x,\Psi) \geq V_\alpha(x,\Psi,f)$. In particular, if we take $f^* \in \mathbb{F}$ satisfying (14) and proceed as above, we get:

$$v(x,\Psi) = V_\alpha^*(x,\Psi,f^*).$$

(c) *The if part.* Suppose that $f^* \in \mathbb{F}$ satisfies Equations (14) and (15). Then, proceeding as in part (b), we obtain that $f^* \in \mathbb{F}$ is an optimal policy.
The only if part. By mimic the same procedure of part (b), we can obtain that for any $f \in \mathbb{F}$ fixed:

$$\alpha V_\alpha(x,\Psi,f) = \bar{r}(x,\Psi,f) + \mathbb{L}^{f,\Psi}V_\alpha(x,\Psi,f); \quad \text{for all } x \in \mathbb{R}^n, \Psi \in S_N. \tag{17}$$

On the other hand, by part (b) we can assert that:

$$\alpha v(x,\Psi) = \sup_{f\in\mathbb{F}}\{\bar{r}(x,\Psi,f) + \mathbb{L}^{f,\Psi}v(x,\Psi)\}; \quad \text{for all } x\in\mathbb{R}^n, \Psi\in S_N. \tag{18}$$

Now let $f^*\in\mathbb{F}$ be an optimal policy, so that $V_\alpha(x,\Psi,f^*) = v(x,\Psi)$. Then, we get the result from (17) and (18).

□

Remark 3. *Briefly, Proposition 2 says that if the HJB-Equations (14) and (15) admit a solution $v\in C^2(\mathbb{R}^n\times S_N)\cap\mathcal{B}_w(\mathbb{R}^n\times S_N)$, then v is the optimal discount payoff (13) to the switching Markovian stochastic control problem with a discounted payoff completely observed, and $f^*\in\mathbb{F}$ is an optimal stationary policy.*

4. Average Optimality Criteria

As in (10), let $\mu_f(\nu) := \int_{\mathbb{R}^n}\nu(x,\Psi)\mu_f(dx,\Psi)$ for every $\nu\in\mathcal{B}_w(\mathbb{R}^n\times S_N)$.

Assumption 4. *Let $(x(t),\Psi(t))$ be the solution of the hidden Markovian-switching diffusion (1)–(4). Then, we suppose that there exist positive constants C and δ such that:*

$$\sup_{f\in\mathbb{F}}|\mathbb{E}^{x,\Psi,f}[\nu(x(t),\Psi(t))] - \mu_f(\nu)| \le Ce^{-\delta t}\parallel\nu\parallel_w w(x,\Psi) \tag{19}$$

*for all $(x,\Psi)\in\mathbb{R}^n\times S_N$, $\nu\in\mathcal{B}_w(\mathbb{R}^n\times S_N)$, and $t\ge 0$. That is, we assume that the process $(x(t),\Psi(t))$ is uniformly w-*exponentially ergodic.

Next, we define the long-run average optimality criterion.

Definition 5. *For each $f\in\mathbb{M}$, $(x,\Psi)\in\mathbb{R}^n\times S_N$, and $T\ge 0$, let:*

$$J_T(x,\Psi,f) := \mathbb{E}^{x,\Psi,f}\Big[\int_0^T \bar{r}(t,x(t),\Psi(t),f)dt\Big]. \tag{20}$$

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

$$J(x,\Psi,f) := \liminf_{T\to\infty}\frac{1}{T}J_T(x,\Psi,f). \tag{21}$$

The function:

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

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

Remark 4. *In some optimal control problems, the limit of $J_T(x,\Phi,f)/T$ as $T\to\infty$ might not exist. To avoid this difficulty, in optimal control problems, it defines the average payoff as a liminf as in (21), which be interpreted as the worst average payoff that is to be maximized.*

For each $f\in\mathbb{F}$, let:

$$J(f) := \mu_f(\bar{r}(\cdot,\Psi,f)) = \int_{\mathbb{R}^n}\bar{r}(x,\Psi,f)\mu_f(dx,d\Psi). \tag{22}$$

with μ_f as in (10). Now, observe that J_T defined in (20) can be expressed as:

$$J_T(x,\Psi,f) = TJ(f) + \int_0^T[\mathbb{E}^{x,\Psi,f}\bar{r}(x(t),\Psi(t),f) - J(f)]dt, \tag{23}$$

therefore, multiplying (23) by $\frac{1}{T}$ and letting $T \to \infty$ we obtain, by (19):

$$J(x,\Psi,f) = \lim_{T\to\infty} \frac{1}{T} J_T(x,\Psi,f) = J(f) \ \text{ for all } \ (x,\Psi) \in \mathbb{R}^n \times S_N. \tag{24}$$

Moreover, by the definition (22) of $J(f)$, the Assumption 3b, and (10):

$$|J(f)| \leq \int_{\mathbb{R}^n} |\, \bar{r}(x(t),\Psi(t),f) \,|\, \mu_f(dx,d\Psi) \leq M \cdot \mu_f(w) < \infty \ \text{ for all } f \in \mathbb{F}.$$

Therefore, by Lemma 1c:

$$\sup_{f\in\mathbb{F}} |J(f)| \leq M \cdot \mu_f(w) \leq M \cdot \frac{p}{q}, \tag{25}$$

thus, the reward $J(f)$ is uniformly bounded on $\mathbb{F}$. From (24) and (25) we obtain that the following:

$$J^* := \sup_{f\in\mathbb{F}} J(f) = \sup_{f\in\mathbb{F}} J(x,\Phi,f) = J^*(x,\Phi) \ \text{ for all } (x,\Phi) \in \mathbb{R}^n \times S_N \tag{26}$$

has a finite value.

Thus, under the Assumptions 1, 2, and 4, it follows from (19) (w-exponential ergodicity) and (22) that the long-run expected average reward (21) coincides with the constant $J(f)$ for every $f \in \mathbb{F}$. Indeed, note that J_T defined in (20) can be expressed as:

$$J_T(x,\Psi,f) = TJ(f) + \int_0^T [\mathbb{E}^{x,\Psi,f}_t \bar{r}(x(t),\Psi(t),f) - J(f)]dt.$$

Definition 6. *(a) A pair (J,v) consisting of a constant $J \in \mathbb{R}$ and a function $v \in C^2(\mathbb{R}^n \times S_N) \cap \mathcal{B}_w(\mathbb{R}^n \times S_N)$ is said to be a solution of the* average reward HJB-equation *if:*

$$J = \max_{u\in U}[\bar{r}(x,\Psi,u) + \mathbb{L}^{u,\Psi} v(x,\Psi)] \ \text{ for all } \ (x,\Psi) \in \mathbb{R}^n \times S_N. \tag{27}$$

(b) If a stationary policy $f \in \mathbb{F}$ attains the maximum in (27), that is:

$$J = \bar{r}(x,\Psi,f) + \mathbb{L}^{f,\Psi} v(x,\Psi)] \ \text{ for all } \ (x,\Psi) \in \mathbb{R}^n \times S_N, \tag{28}$$

then f is called a canonical policy.

The following theorem shows that if a policy satisfies the average reward HJB-equation, then it is an optimal average policy.

Theorem 1. *If Assumptions 1, 2, and 3 hold, then:*

(*i*) *The average reward HJB Equation (27) admits a unique solution (J,v), with $v \in C^2(\mathbb{R}^n \times S_N) \cap \mathcal{B}_w(\mathbb{R}^n \times S_N)$ satisfying $v(0,\Psi_0) = 0$ for some $\Psi_0 \in S_N$ fixed.*
(*ii*) *There exists a canonical policy.*
(*iii*) *The constant J in (27) equals J^* in (26).*
(*iv*) *There exists a stationary average optimal policy.*

Proof. (*i*) The steps for the proof of this incise are essentially the same given in proof of Theorem 6.4 in [24]; thus, we omit the proof.

(*ii*) Since $u \to r(\cdot,\cdot,u)$ and $u \to b(\cdot,\cdot,u)$ are continuous functions on the compact set $\mathcal{U}$, we obtain that $u \to \bar{r}(\cdot,\cdot,u) + \mathbb{L}^{u,\Psi} v(\cdot,\cdot)$ is a continuous function on $\mathcal{U}$; thus, the existence of a canonical policy $f \in \mathbb{F}$ follows from standard measurable selection theorems; see [25] (Theorem 12.2).

(iii) Observe that, by (27):

$$J \geq \bar{r}(x,\Psi,u) + \mathbb{L}^{u,\Psi} v(x,\Psi) \ \text{ for all } (x,\Psi) \in \mathbb{R}^n \times S_N \text{ and } \ u \in U. \tag{29}$$

Therefore, for any $f \in \mathbb{F}$, using Dynkin's formula and (29) we obtain:

$$\begin{aligned} \mathbb{E}^{x,\Psi,f} v(x(t),\Psi(t)) &= v(x,\Psi) + \mathbb{E}^{x,\Psi,f}\Big(\int_0^t \mathbb{L}^{f,\Psi} h(x(s),\Psi(s))ds\Big) \\ &\leq v(x,\Psi) + Jt - \mathbb{E}^{x,\Psi,f}\Big(\int_0^t \bar{r}(x(s),\Psi(s))ds\Big). \end{aligned} \tag{30}$$

Thus, multiplying by t^{-1} in (30) we have:

$$t^{-1} J_t(x,\Psi,f) \leq J + t^{-1} v(x,\Psi) - t^{-1}\mathbb{E}^{x,\Psi,f} v(x(t),\Psi(t)). \tag{31}$$

Consequently, letting $t \to \infty$ in (31), and using Lemma 1b and (24), we obtain:

$$J \geq J(f) \ \text{ for all } f \in \mathbb{F}.$$

To obtain the reverse inequality, similar arguments show that if:

$$J \leq \bar{r}(x,\Psi,u) + \mathbb{L}^{u,\Psi} v(x,\Psi) \ \text{ for all } \ (x,\Psi) \in \mathbb{R}^n \times S_N \text{ and } u \in U,$$

then $J \leq J(f)$ for all $f \in \mathbb{F}$. This last inequality together with (29) yields that if $f \in \mathbb{F}$ is a canonical policy, which satisfies (28), then we obtain that $J(f) = J$, and by (26):

$$J = J(f) = J^* = J^*(x,\Psi) \ \text{ for all } (x,\Psi) \in \mathbb{R}^n \times S_N. \tag{32}$$

(iv) Similar arguments to those given in (iii) lead us to that if $f \in \mathbb{F}$ is a canonical policy, then it is an average optimal. □

Theorem 1 indicates that if a policy satisfies the HJB Equation (27), then this policy is an optimal policy for the optimal control problem associated with the HJB equation. The difficulty with this approach is how to get a solution (J^*, v, f) of the HJB equation. The most common form of the solve the HJB equation is based on variants on the *vanishing discount approach* (see [11,24,26] for details).

Remark 5 ([1]). *In the optimality criteria known as bias optimality, overtaking optimality, sensitive discount optimality, and Blackwell optimality, the early returns and the asymptotic returns are both relevant; thus, to study them, we need first to analyze the discounted and average optimality criteria. These optimality criteria will be studied in future work.*

Remark 6.

- ***On Assumption 1, ([7], Theorems 3.17 and 3.18).*** *The uniform Lipschitz and linear growth conditions of b and σ ensure the existence and uniqueness of the global solution of the SDE with Markovian switching (1). The uniform Lipschitz condition $(\max_{(i,u)\in E\times U} \|b(x,i,u) - b(y,i,u)\| \leq C_1\|x-y\|$, $\|\sigma(x,i) - \sigma(y,i)\| \leq C_2\|x-y\|$) imply that the change rates of the functions $b(x,i,u)$ and $\sigma(x,i)$ are minor or equal to the change rate of a linear function of x. This gives, in particular, the continuity of b and σ in x for all $[t_0,\infty)$. Thus, the uniform Lipschitz condition excludes the functions b and σ that are discontinuous concerning x. It is important to note that although a function let continuous, it does not guarantee that it satisfies the uniform Lipschitz condition; for example, the continuous function $sin(x^2)$ does not satisfy this condition. Uniform Lipschitz condition can be replaced by the local Lipschitz condition. In fact, the local Lipschitz condition allows us to include a great variety of functions, such as functions $v \in C^2(\mathbb{R}^n \times E)$. However, the linear growth condition (Assumption 1 (d)) also excludes some important functions, such as $b(x,i) = -|x|^2 x + i$. Assumption 1 (d) is quite*

standard but may be restrictive for some applications. As far as the results of this paper are concerned, the uniform Lipschitz condition may be replaced by the weaker condition:

$$x^T b(x,i,u) + \frac{1}{2}||\sigma(x,i)||^2 \leq K(1+||x||^2), \quad \text{for all } (x,i) \in \mathbb{R}^n \times E, \tag{33}$$

where K is a positive constant. This last condition allows us to include many functions as the coefficients b and σ. For example:

$$b(x,i,u) = a(i)[x(t) - x^3(t)] + xg(u) \quad \sigma(x,i) = b(i)x^2(t)$$

with $a(i), b(i) > 0$ such that $b^2(i) \leq 2a(i)$ and for some continuous function $g : U \to \mathbb{R}$ given. It is possible to check that a diffusion process with the parameters given above satisfies the local Lipschitz condition but the linear growth condition is not satisfied. On the other hand, note that:

$$a(i)x[x - x^3] + x^2 g(u) + \frac{1}{2}b^2(i)x^4 \leq a(i)x^2 + x^2 g(u) \leq K(1+x^2)$$

with $K = \max_{(i,u)\in E\times U}\{a(i) + g(u)\}$ and a compact control set U. That is, the condition (33) is fulfilled. Thus, ([7], Theorem 3.18) guarantees that the SDE with Markovian switching with these coefficients has a unique global solution on $[t_0, \infty)$.

- ***On Assumption 2, ([7], Theorem 5.2).*** *This assumption guarantees the positive recurrence and the existence of an invariant measure $\mu_f(dx, \Psi)$ for the Markov–Feller process $(x(t), \Psi(t))$. Moreover, if this assumption holds together with the inequality $k(|x|^{\overline{p}}) \leq w(x,i)$ for positive numbers $k, \overline{p}, H$, then, the diffusion process (1) satisfies:*

$$limsup_{t\to\infty}\mathbb{E}|x(t)|^{\overline{p}} \leq H,$$

that is, $x(t)$ is asymptotically bounded in $\overline{p}$th moment. Some Lyapunov functions are, for example:

$$w(x,i) = k(i)|x|^{\overline{p}}, \quad k(i) > 0, \ \overline{p} \geq 2, \ \forall\ (x,i) \in \mathbb{R}^n \times E, \tag{34}$$

considering that the coefficients b and σ in (1) satisfy the Lipschitz condition and:

$$x^T b(x,i,u) + \frac{\overline{p}-1}{2}||\sigma(x,i)||^2 \leq B(i)||x||^2 + a, \tag{35}$$

with $a > 0$, and $B(i)$ be constants. In fact, using the inequality $a^c b^{1-c} \leq ac + b(1-c) \ \ \forall\ a, b \geq 0,\ c \in [0,1]$ and (35), we get:

$$
\begin{aligned}
\mathbb{L}^{u,\psi}w(x,i) &= k(i)\overline{p}||x||^{\overline{p}-1}b(x,i,u)+\frac{1}{2}k(i)\overline{p}(\overline{p}-1)||\sigma(x,i)||^2|x|^{\overline{p}-2}+\sum_{j=i}^{N}q_{ij}k(j)||x||^{\overline{p}}\\
&= \overline{p}k(i)||x||^{\overline{p}-2}\Big\{x^Tb(x,i,u)+\frac{\overline{p}-1}{2}||\sigma(x,i)||^2\Big\}+\sum_{j=i}^{N}q_{ij}k(j)||x||^{\overline{p}}\\
&\leq \overline{p}k(i)||x||^{\overline{p}-2}\{B(i)||x||^2+a\}+\sum_{j=i}^{N}q_{ij}k(j)||x||^{\overline{p}}\\
&\leq (\overline{p}B(i)k(i)+\sum_{j=i}^{N}q_{ij}k(j))||x||^{\overline{p}}+a\overline{p}k(i)||x||^{\overline{p}-2}\\
&= (\overline{p}B(i)k(i)+\sum_{j=i}^{N}q_{ij}k(j))||x||^{\overline{p}} \qquad (36)\\
&+ \Big[(a\overline{p}k(i))^{\overline{p}/2}\Big(\frac{2}{\lambda(i)}\Big)^{(\overline{p}-2)/2}\Big]^{2/\overline{p}}\Big[\frac{\lambda(i)}{2}||x||^{\overline{p}}\Big]^{(\overline{p}-2)/\overline{p}}\\
&\leq (\overline{p}B(i)k(i)+\sum_{j=i}^{N}q_{ij}k(j))||x||^{\overline{p}}+\frac{2}{\overline{p}}(a\overline{p}k(i))^{\overline{p}/2}\Big(\frac{2}{\lambda(i)}\Big)^{(\overline{p}-2)/2}\\
&+ \frac{\lambda(i)(\overline{p}-2)}{2\overline{p}}||x||^{p}\\
&\leq -\frac{\lambda(i)(\overline{p}+2)}{2\overline{p}}||x||^{\overline{p}}+\frac{2}{\overline{p}}(a\overline{p}k(i))^{\overline{p}/2}\Big(\frac{2}{\lambda(i)}\Big)^{(\overline{p}-2)/2}
\end{aligned}
$$

where $\lambda(i)=(\overline{p}B(i)k(i)+\sum_{j=i}^{N}q_{ij}k(j))$.
If we set:

$$q:=\min_{i\in E}\Big[\frac{\lambda(i)(\overline{p}+2)}{2\overline{p}}\Big]\quad p:=\max_{i\in E}\Big[\frac{2}{\overline{p}}(a\overline{p}k(i))^{\overline{p}/2}\Big(\frac{2}{\lambda(i)}\Big)^{(\overline{p}-2)/2}\Big],$$

then

$$\mathbb{L}^{u,\psi}w(x,i) \leq -q||x||^{\overline{p}}+p\leq -qw(x,i)+p.$$

Now, taking the Lyapunov function (34) we define:

$$w(x,\Psi)=\sum_{i=1}^{N}\Psi_iw(x,i)=\sum_{i=1}^{N}\Psi_ik(i)||x||^{\overline{p}}.$$

Considering that $w_x(x,\Psi)=\sum_{i=1}^{N}\Psi_ik(i)\overline{p}||x||^{\overline{p}-1}$, $w_{xx}(x,i)=\sum_{i=1}^{N}\Psi_ik(i)\overline{p}(\overline{p}-1)||x||^{\overline{p}-2}$, $\nabla w_\Psi(x,i)=[k(i),k(2),\ldots,k(n)]||x||^{\overline{p}}$ *and* $w_{\Psi\Psi}(x,\Psi)=0$; *a similar procedure to that given in (37) allows us to obtain that W is also a Lyapunov function. That is:*

$$\mathbb{L}^{u,\Psi}w(x,\Psi) \leq -q||x||^{\overline{p}}+p\leq -qw(x,\Psi)+p.$$

- ***On Assumption 3****. This assumption allows us that the reward rate (or cost rate) can be unbounded from above and below. For the Lyapunov function* $w(x,i)=k(i)|x|^{\overline{p}}$, *a reward rate of the form:*

$$r(x,i,u)=k(i)|x|^{\overline{p}}+h(u)$$

for some continuous function $h:U\to\mathbb{R}$ *satisfies the Assumption 3. In fact:*

$$|r(x,i,u)|\leq k(i)|x|^{\overline{p}}+\max_{u\in U}h(u)\leq (k(i)+\max_{u\in U}h(u))|x|^{\overline{p}}=Mw(x,i)$$

with $M = \max_{i\in E}\{k(i) + \max_{u\in U}\}$ *and* U *a compact set.*

- ***On Assumption 4.*** *This assumption indicates asymptotic behavior of* $x(t)$ *when* t *goes to infinite. Sufficient conditions for the w-exponentially ergodicity of the process* $(x(t), \psi(t))$ *can be seen in ([1], Theorem 2.8). In fact, in the proof of this theorem, Assumptions 1 and 2 are required. Note that, for the optimal control problem with discounted optimality criterion, the w-exponentially ergodicity of the process* $(x(t), \psi(t))$ *is not required. This assumption is only necessary to study the average reward optimality criterion.*

Remark 7. *In the following sections, our theoretical results are implemented in three applications. The dynamic system in the three applications evolves according to linear stochastic differential equations* $dx(t) = (A(i)x(t) + Bu(t))dt + \sigma dW(t)$*, namely, Assumption 1. The state numbers of the Markov chain is* 2*, that is,* $E = \{1, 2\}$*. The payoff rate is of the form* $r(x, i, u) = x^T R(i)x + u^T Su$ *with* $x \in \mathbb{R}^2$ *and* $u \in \mathcal{U} := [0, a1] \times [0, a2]$*,* $a1, a2 > 0$*. Taking* $w(x, i) = x^T R(i)x + 1$ *we get:*

$$\begin{aligned} |r(x,i,u)| &= |x^T R(i)x| + |u^T Su| \\ &\leq |x^T R(i)x| + |u^T Su||x^T R(i)x + 1| \\ &= max_{u\in\mathcal{U}}(|u^T Su| + 1)|x^T R(i)x + 1| \\ &= M_2 w(x,i) \end{aligned}$$

with $M_2 = max_{u\in\mathcal{U}}(|u^T Su| + 1)$*; thus, Assumption 3 also holds. A few calculations allow us to obtain the Assumption 2 with* $w(x, \Psi) = \sum_{i=1}^{2} \Psi_i(t) w(x, \psi(t)) = \sum_{i=1}^{2} \Psi_i(t)(x^T R(\psi(t))x + 1)$*. In fact:*

$$\begin{aligned} \mathbb{L}^{u,\Psi} w(x,\Psi) &= x^2[2A(i)[\Psi_1 R(1) + \Psi_2 R(2)] + R(1)\sum_{i=1}^{2} q_{i1}\Psi_i + R(2)\sum_{i=1}^{2} q_{i2}\Psi_i] \\ &+ x[R(1)\sum_{i=1}^{2} q_{i1}\Psi_i + R(2)\sum_{i=1}^{2} q_{i2}\Psi_i] \\ &+ \sigma^2[\Psi_1 R(1) + \Psi_2 R(2)]. \end{aligned} \tag{37}$$

Let $0 < q < -[2A(i)[\Psi_1 R(1) + \Psi_2 R(2)] + R(1)\sum_{i=1}^{2} q_{i1}\Psi_i + R(2)\sum_{i=1}^{2} q_{i2}\Psi_i]$*, and rewrite* $\mathbb{L}^{u,\Psi} w(x, \Psi)$ *as:*

$$\mathbb{L}^{u,\Psi} w(x,\Psi) = -qw(x,\Psi) + l(x,i,u).$$

where

$$\begin{aligned} l(x,i,u) &:= qw(x,\Psi) + x^2[2A(i)[\Psi_1 R(1) + \Psi_2 R(2)] + R(1)\sum_{i=1}^{2} q_{i1}\Psi_i + R(2)\sum_{i=1}^{2} q_{i2}\Psi_i] \\ &+ x[R(1)\sum_{i=1}^{2} q_{i1}\Psi_i + R(2)\sum_{i=1}^{2} q_{i2}\Psi_i] \\ &+ \sigma^2[\Psi_1 R(1) + \Psi_2 R(2)] \\ &\leq p, \end{aligned} \tag{38}$$

where the last inequality is obtained from fact that the function $l(x.i.u)$ *is continuous on the compact set* U *for all* $x \in \mathbb{R}$ *and that the term* $q + [2A(i)[\Psi_1 R(1) + \Psi_2 R(2)] + R(1)\sum_{i=1}^{2} q_{i1}\Psi_i + R(2)\sum_{i=1}^{2} q_{i2}\Psi_i]$ *is negative. Thus,* $\mathbb{L}^{u,\Psi} w(x, \Psi) = -qw(x, \Psi) + p$ *and Assumption 2b follows.*

5. Application 1: Discounted Linear Quadratic Regulator (LQR)

In this subsection, we consider the α-discounted linear quadratic regulator. To this end, we suppose that the dynamic system evolves according to the linear stochastic differential equations:

$$dx(t) = (\overline{A}(\Psi(t))x(t) + Bu(t))dt + \sigma dW(t). \tag{39}$$

with $\overline{A}(\Psi(t)) := \sum_{i=1}^{N} A(i)\Psi_i(t)$, $A : E \to \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $W(\cdot)$ is a m-dimensional Brownian motion, and σ is a positive constant. The expected cost is:

$$V_\alpha(x,\Psi,u) := \mathbb{E}^u_{x,\Psi}\left[\int_0^\infty e^{-\alpha s}\{x^T(s)\overline{D}(\Psi(s))x(s) + u^T\overline{R}(\Psi(s))u(s)\}ds\right].$$

where $\overline{D}(\Psi(t)) := \sum_{i=1}^{N} D(i)\Psi_i(t)$, $D : E \to \mathbb{R}^{n\times n}$, $\overline{R}(\Psi(t)) := \sum_{i=1}^{N} R(i)\Psi_i(t)$ and $R : E \to \mathbb{R}^{n\times n}$. The optimality equation or HJB-equation for the α-discounted partially observed LQR-optimal control problem is:

$$\alpha v(x,\Psi) = \min_{u\in U}\{x\overline{D}(\Psi(t))x^T + u^T\overline{R}(\Psi(t))u + \mathcal{L}^u vs.(x,\Psi)\}, \tag{40}$$

where the infinitesimal generator for the process $(x(t),\Psi(t))$ applied to $v(x,\Psi) \in C^{2,2}$ $(\mathbb{R}^n \times S_N)$ is:

$$\begin{aligned}\mathbb{L}^u vs.(x,\Psi) = (\overline{A}(\Psi)x + Bu)v_x(x,\Psi) + \frac{1}{2}[Tr(\sigma\sigma^T)]v_{xx}(x,\psi)\\ +\mathcal{Q}^T\Psi v_\Psi(x,\Psi,) + \frac{1}{2}v_{\Psi\Psi}(x,\Psi,)Tr[A_2]\end{aligned} \tag{41}$$

where

$$A_2 = [\sigma_0^{-1}\Big(\text{diag}(h) - h^T(\Psi(t))I_N\Big)\Psi(t)][\sigma_0^{-1}\Big(\text{diag}(h) - h^T(\Psi(t))I_N\Big)\Psi(t)]^T. \tag{42}$$

Note that, by minimizing (40) with respect to u, we find that the optimal control is the form:

$$f^*(x,\Psi) = -\frac{\overline{R}^{-1}(\Psi)}{2}B^T v_x. \tag{43}$$

By Proposition 2, if there exist a function $v \in C^{2,2}(\mathbb{R}^n \times S_N) \cup \mathcal{B}_w(\mathbb{R}^n \times S_N)$ and a policy $f^* \in \mathbb{F}$ such that (14) and (15) hold, then v coincides with the value function $v^*(x,\Psi) := \min_{u\in U} V_\alpha(x,\Psi,u)$ and $u(t) = f^*(x)$ is the α-discount optimal policy. Thus, we propose that the function $v \in C^{2,2}(\mathbb{R}^n \times S_N) \cup \mathcal{B}_w(\mathbb{R}^n \times S_N)$ that solves the HJB-Equation (40) has the form:

$$v(x,\Psi) = x^T Kx + n(\Psi) + c, \tag{44}$$

where $n : S_N \to \mathbb{R}$ is a twice differentiable continuous function, c is a constant, and K is a positive definite matrix. Inserting the derivative of $v(x,\Psi)$ in (43) we get the optimal control:

$$f^*(x,\Psi) = -\overline{R}^{-1}(\Psi)B^T K^T x, \tag{45}$$

where the equality (40) holds if the matrix K satisfies the algebraic Riccati equation:

$$\begin{aligned}\overline{A}^T(\Psi(t))K + K\overline{A}(\Psi(t)) - KB\overline{R}(\Psi(t))^{-1}B^T K\\ +\overline{D}(\Psi(t)) - \alpha K = 0,\end{aligned}$$

$$c = Tr[b(w(t))b^T(w(t))K]/\alpha$$

and $n(\cdot) \in \mathcal{C}^2(S_N)$ satisfies the partial differential equation:

$$Q^T\Psi(t)n'(\Psi(t)) + \frac{1}{2}Tr[A_2]n''(\Psi(t)) - \alpha n(\Psi(t)))I_n = 0, \ \ \forall \ \ \Psi(t) \in S_N,$$

where A_2 is as in (42), I_N is the identity matrix of $N \times N$, and $n^{'}$ and $n^{''}$ are the gradient and the Hessian of the n, respectively.

Simulation results. In the following figures, we assume that the Markov chain $\psi(t)$ has two states, namely, $E = \{1, 2\}$ and the dynamic system $x(t) \in \mathbb{R}^2$. We have computed the Wonham filter, the states of the dynamic system (39) $x(t) = [x_1(t), x_2(t)]^T$ with initial condition $x(0) = [10, 15]^T$, the value function (44), and the optimal control (45) for the following data: $\sigma = 1$, $\sigma_0 = 1$, $\alpha = 0.01$, $h(1) = 1$, $h(2) = 2$, $\Psi_1(0) = 0.5$, $\Psi_2(0) = 0.5$, $R_1 = 1$, $R_2 = 2$:

$$A(1) = \begin{bmatrix} -5 & 1 \\ 0 & -10 \end{bmatrix}, \ A(2) = \begin{bmatrix} -10 & 1 \\ 0 & -10 \end{bmatrix},$$

$$D(1) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ D(2) = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix},$$

and the transition matrix:

$$Q = \begin{bmatrix} -0.2 & 0.2 \\ 0.7 & -0.7 \end{bmatrix}.$$

To solve the Wonhan filter, we use the numerical method given in ([18], Section 8.4), considering that the Markov chain can only be observed through $dy(t) = h(\psi(t)) + \sigma_0 dB(t)$.

Figure 1 shows the solution of the filter Wonham equation and the states of the hidden Markov chain $\psi(t)$. As can be noted, in $t = 0.05$ s $\Psi_2(0.05) = \mathbb{P}(\psi(t) = 2 \mid y(s), 0 \leq s \leq 0.05) \geq \Psi_1(0.05)$, implying that the Markov chain with a higher probability to 0.5 is in state 2 in $t = 0.3$ ($\psi(0.3) = 2$). The evolution of the dynamic system (39) is given in Figure 2 (top); in this figure, we can note that the optimal control (45) moves the initial point $x(0) = [10, 15]^T$ to the point $[0, 0]^T$ in $t = 0.8$ s, indicating the good performance of the optimal control (45). The asymptotic behavior of the optimal control (45) is given in Figure 2 (bottom); this control stabilizes at zero around $t = 0.8$ s, since $x(t)$ also stabilizes at zero around $t = 0.8$ s.

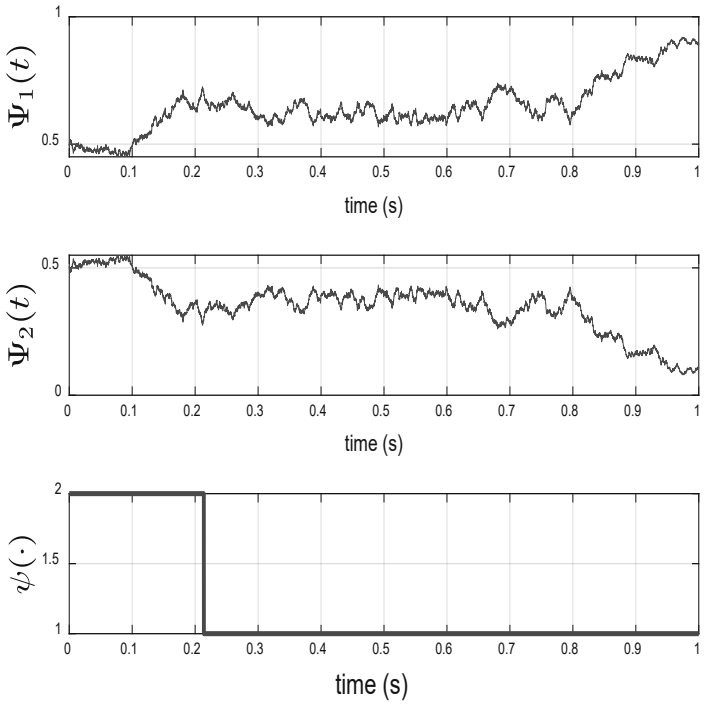

Figure 1. Wonham filter for the α-discounted LQR.

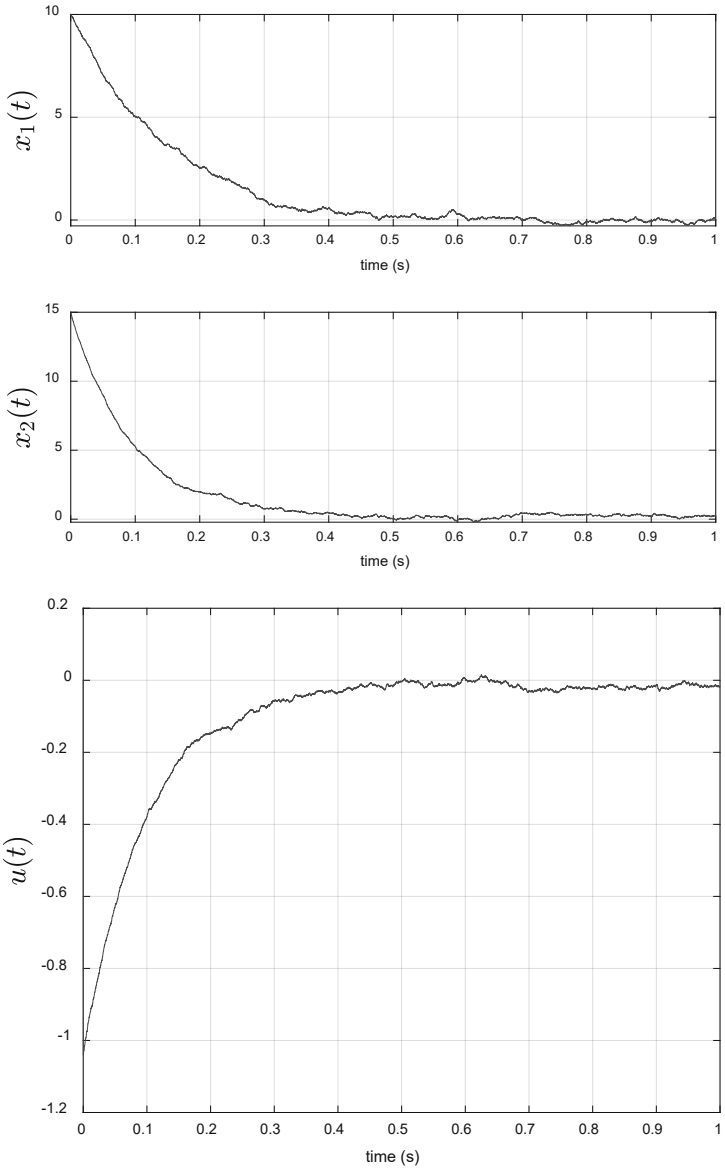

Figure 2. Asymptotic behavior of the state of dynamic system (**top**) and optimal control α-discount LQR (**bottom**).

6. Application 2: Average LQR: Modeling of a Quarter-Car Suspension

In this section, the basic quarter-car suspension model analyzed in [27] is considered, see Figure 3. The parameters are: the sprung mass (m_s), the unsprung mass (m_u), the suspension spring constant (k_s), and the tire spring constant (k). Let z_s, z_u, and z_r be the vertical displacements of the sprung mass, the unsprung mass, and the road profile, respectively. The equations of motion for this model are given by:

$$m_s z_s^{''}(t) = -k_s(z_s(t) - z_u(t)) - u(t), \quad (46)$$

$$m_u z_u^{''}(t) = k_s(z_s(t) - z_u(t)) - k(z_u(t) - z_r(t)) + u(t). \quad (47)$$

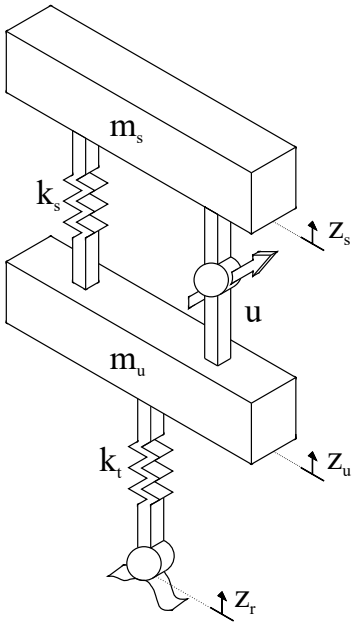

Figure 3. Schematic of a quarter-car suspension.

Now, defining $x_1(t) = z_s^{'}(t)$, $x_2(t) = z_u^{'}(t)$, $x_3(t) = z_s(t) - z_u(t)$, and $x_4(t) = z_u(t) - z_r$, the equations of motion (46) and (47) can be expressed in matrix form as:

$$dx(t) = (Ax(t) + Bu(t))dt + C_1 dz_r(t) \tag{48}$$

where $dx(t) = \begin{bmatrix} dx_1(t) \\ dx_2(t) \\ dx_3(t) \\ dx_4(t) \end{bmatrix}$, $A = \begin{bmatrix} 0 & 0 & \frac{k_s}{m_s} & 0 \\ 0 & 0 & \frac{k_s}{m_u} & \frac{k}{m_u} \\ 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$, $B = \begin{bmatrix} \frac{1}{m_s} \\ \frac{1}{m_s} \\ 0 \\ 0 \end{bmatrix}$, $C_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}$, and in the time domain, the road profile, $z_r(t)$, can be represented as the output of a linear first-order filter to white noise as follows:

$$dz_r(t) = -a(\psi(t))Vz_r(t)dt + \sigma_2 dW_1(t),$$

where V is the vehicle speed (assumed constant), σ_2 is a positive constant, and a is the road roughness coefficient depending on the type of road. Here, we assume that a depends on a hidden Markov chain, that is, $a(\psi(t))$ with $\psi(t) \in \{1, 2\}$. In our case, we consider that the dynamic system (48) evolves with additional white noise, that is:

$$dx(t) = (Ax(t) + Bu(t))dt + \sigma_1 dW(t) + C_1 dz_r(t) \tag{49}$$

The experts introduced the following performance index in order to trade off between the ride comfort and the handling while maintaining the constraint on suspension deflection:

$$\begin{aligned} J(x, \Psi, u) &= \lim_{T\to\infty} \frac{1}{T} \mathbb{E}^{x,\Psi,u} \Big[\int_0^T \Big[c_1 \frac{d^2 z_s}{d^2 t}^2 + c_2 [z_1(t) - z_u(t)]^2 \\ &+ c_3 [z_u(t) - z_r(t)]^2 + c_4 u(t)^2 \Big] dt \Big] \end{aligned} \tag{50}$$

Defining $y := \Big[\frac{d^2 z_s}{d^2 t}^2, [z_1(t) - z_u(t)]^2, [z_u(t) - z_r(t)]^2 \Big]$, $C := diag(c_1, c_2, c_3)$, and $R := [c_4]$, we can rewrite (50) as:

$$J(x, \Psi, u) = \lim_{T\to\infty} \frac{1}{T} \mathbb{E}^{x,\Psi,u} \Big[\int_0^T yCy^T + u^T(t)Ru(t)dt \Big] \tag{51}$$

Now, from the equations of motion in (46) and (47), note that $y = Mx + Nu$ with $M = \begin{bmatrix} 0 & 0 & \frac{k_s}{m_s} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$, and $N = \begin{bmatrix} -\frac{1}{m_s} \\ 0 \\ 0 \end{bmatrix}$. Thus, replacing this matrix form of y in (51) we can rewrite (50) again as:

$$J(x, \Psi, u) = \lim_{T \to \infty} \frac{1}{T} \mathbb{E}^{x,\Psi,u} \Big[\int_0^T (x^T Q_1 x + 2x^T Q_2 u + u^T R_1 u) dt \Big] \tag{52}$$

where $Q_1 = M^T C M$, $Q_2 = M^T C N$, $R_1 = N^T C N + R$.

The optimal control problem (OCP). The OCP in this application consists of finding $u^* \in U$ such that it minimizes the performance index (52) considering that the dynamic system evolves according to the stochastic differential Equation (49).

In the dynamic programming technique, we need the infinitesimal generator $\mathcal{L}^u$ of the process $(x(t), \Psi(t))$ applied to $v(x, \Psi, z_r) \in C^{2,2,2}(\mathbb{R}^n \times S_N \times \mathbb{R})$; in this case, this generator is:

$$\begin{aligned} \mathcal{L}^u vs.(x, \Psi, z_r) = & -a(\Psi(t)) v_{z_r}(x, \Psi, z_r) \\ & + (Ax + Bu) v_x(x, \Psi, z_r) \\ & + \mathcal{Q}^T \Psi v_\Psi(x, \Psi, z_r) \\ & + \frac{1}{2} Tr[\sigma_1 \sigma_1^T] v_{xx}(x, \Psi, z_r). \\ & + \frac{1}{2} Tr[\sigma_2 \sigma_2^T] v_{z_r z_r}(x, \Psi, z_r) \\ & + \frac{1}{2} v_{\Psi\Psi}(x, \Psi, z_r) Tr[A_2] \end{aligned} \tag{53}$$

where $A_2(\Psi(t)) = [\sigma_0^{-1}(\text{diag}(h) - h^T(\Psi(t)) I_N)\Psi(t)][\sigma_0^{-1}(\text{diag}(h) - h^T(\Psi(t)) I_N)\Psi(t)]^T$, whereas the Hamilton–Jacobi–Bellman Equation (or dynamic programming equation) associated with this problem is:

$$J = \max_{u \in U} [x^T Q_1 x + 2x^T Q_2 u + u^T R_1 u + \mathcal{L}^u vs.(x, \Psi, z_r)] \text{ for all } (x, \Psi) \in \mathbb{R}^n \times S_N, \tag{54}$$

see [28] for more details.

Proposition 3. *Assume that $(x(t), z_r(t), \Psi(t))$ evolves according to (49). Then, the control that minimizes the long-run cost (52) is:*

$$f^*(x, \Psi, z_r) = -R_1^{-1}(Q_2^T + B^T K)^T x(t), \tag{55}$$

whereas the corresponding function v that solves the HJB Equation (54) is given by:

$$v(x, \Psi, z_r) = x^T K x + g(z_r) + n(\Psi)$$

where K is a positive semi-definite matrix that satisfies the Ricatti differential equation

$$\begin{aligned} K(A - BR_1^{-1}Q_2^T) + (A - BR_1^{-1}Q_2^T)K - KBR_1B^T P \\ (Q_1 - Q_2 R_1^{-1} Q_2^T) = 0, \end{aligned} \tag{56}$$

and $g(\cdot) \in C^2(\mathbb{R})$ satisfies the differential equation:

$$a(\Psi) g'(z_r) + \frac{1}{2} \sigma_2^2 g''(z_r) = 0, \tag{57}$$

and $n(\cdot) \in \mathcal{C}^2(S_N)$ satisfies the partial differential equation:

$$Q^T \Psi n'(\Psi(t)) + \frac{1}{2} Tr[A_2] n''(\Psi) = 0, \tag{58}$$

where A_2 is as in (41) and n' and n'' denote the gradient and the Hessian of the n, respectively. The optimal cost is given by:

$$J = Tr[\sigma_1 \sigma_1^T] K = J^*(x, \Psi) = \min_{u \in U} J(x, \Psi, u).$$

Proof. The HJB-equation for the partially observed LQR optimal control problem with $(x(t), \Psi(t))$ evolves according to (49) and finite cost (52) is (54), where $\mathcal{L}^u v(t, x, w, \Psi)$ is the infinitesimal generator given in (53). We are looking for a candidate solution $h \in \mathcal{C}^{2,2,2}(\mathbb{R}^n \times S_N \times \mathbb{R})$ to (54) in the form:

$$v(x, \Psi, z_r) = x^T K x + g(z_r) + n(\Psi), \tag{59}$$

for some continuous functions $g(\cdot) \in \mathcal{C}^2(\mathbb{R})$, $h(\cdot) \in \mathcal{C}^2(S_N)$ and K a positive semi-definite matrix. We assume that $g''(z_r) > 0$ for all $z_r \in \mathbb{R}$ and $n''(\Psi)$ is positive definite, so that the function $(x, \Psi, z_r) \to v(x, \Psi, z_r)$ is convex.

Now, the function $u \in U \to 2x^T Q_2 u + u^T R_1 u + B u v_x$ is strictly convex on the compact set U, and thus, attains its minimum at:

$$f^*(x, \Psi, z_r) = -\frac{1}{2} R^{-1}[-2x^T Q_2 - B h_x] = -R_1^{-1}(Q_2^T + B^T K)^T x(t). \tag{60}$$

Inserting $f^*(x, \Psi, z_r)$ and the partial derivatives of v with respect to x, z_r, and Ψ in the HJB-Equation (54), we obtain:

$$\begin{aligned} J &= x^T Q_1 x + 2x^T Q_2(-R_1^{-1}(Q_2^T + B^T K)^T x) \\ &+ (-R_1^{-1}(Q_2^T + B^T K)^T x)^T R_1(-R_1^{-1}(Q_2^T + B^T K)^T x) \\ &- a(\Psi(t)) g'(z_r) + (Ax + B(-R_1^{-1}(Q_2^T + B^T K)^T x)) 2Kx + \mathcal{Q}^T \Psi h'(\Psi) + + Tr[\sigma_1 \sigma_1^T] K \\ &+ \frac{1}{2} Tr[\sigma_2 \sigma_2^T] g''(z_r) + \frac{1}{2} h''(\Psi) Tr[A_2]. \end{aligned} \tag{61}$$

For equality (61) to hold, it is necessary that the functions g and h satisfy (57) and (58), respectively, and the matrix K satisfies the Ricatti differential Equation (56), whereas the constant $J = Tr[\sigma_1 \sigma_1^T] K$. Finally, from the Theorem 1, it follows that f^* is an optimal Markovian control and the value function $J_T^*(t, x, w, \Psi)$ is equal to (59). That is:

$$J^*(x, \Psi) = \min_{u \in U} J(x, \Psi, u) = J = Tr[\sigma_1 \sigma_1^T] K.$$

□

Simulation results. To solve the Wonhan filter, we use the numerical method given in ([18], Section 8.4), considering that the Markov chain $\psi(t)$ has two states that can only be observed through $dy(t) = h(\psi(t)) + \sigma_0 dB(t)$. The following data were used: $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_0 = 1$, $\alpha = 0.01$, $a(1) = 0.03$, $a(2) = 0.015$, $\Psi_1(0) = 0.5$, $\Psi_2(0) = 0.5$, $R = 1.0239 \times 10^{-5}$, $h(1) = -1$, $h(2) = 0.5$, $m_s = 329$ kg, $m_u = 51$ kg, $k_s = 4300$ N/m, $k = 210,000$ N/m, $V = 20$ m/s, $c_1 = 1$, $c_2 = c_3 = 1 \times 10^5$, $c_4 = 1 \times 10^{-6}$ and:

$$Q = \begin{bmatrix} -0.3 & 0.3 \\ 0.5 & -0.5 \end{bmatrix}.$$

The solution of the Wonham filter equation and the states of the hidden Markov chain $\psi(t)$ are shown in Figure 4. As can be noted, in $t = 1$ s, $\Psi_1(1) = \mathbb{P}(\psi(t) = 1 \mid y(s)$,

$0 \leq s \leq 1) \geq \Psi_2(1)$, implying that the Markov Chain with a probability greater than 0.5 is in state 1 at $t = 1$.

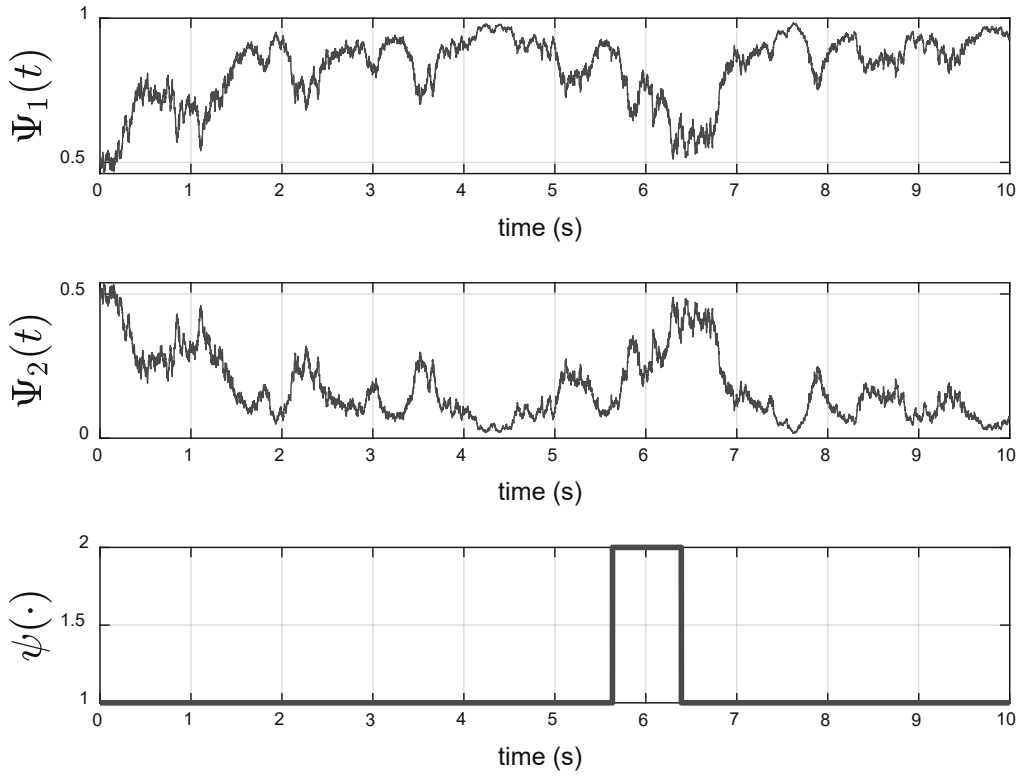

Figure 4. Wonham filter and hidden Markov chain (in t = 1 s).

The asymptotic behavior of the optimal control (55) is given in Figure 5 (bottom). It is interesting to note that this control minimizes the magnitude of the sprung mass velocity, $x_1 = z_s^{'}$ and unsprung mass velocity, $x_2 = z_u^{'}$ after $t = 9$ s, see Figure 5 (top). This behavior implies that the magnitude of the sprung mass acceleration, $x_1 = z_s^{''}$ and unsprung mass acceleration $x_2 = z_u^{'}$ are also minimized, considering that the stochastic differential equation that models the road profile depends on a hidden Markov chain. These results agree with the obtained by authors in [27]. These authors mentioned that two important objectives of a suspension system are ride comfort and handling performance. The ride comfort requires that the car body be isolated from road disturbances as much as possible to provide a good feeling for passengers. In practice, we are looking to minimize the acceleration of the sprung mass.

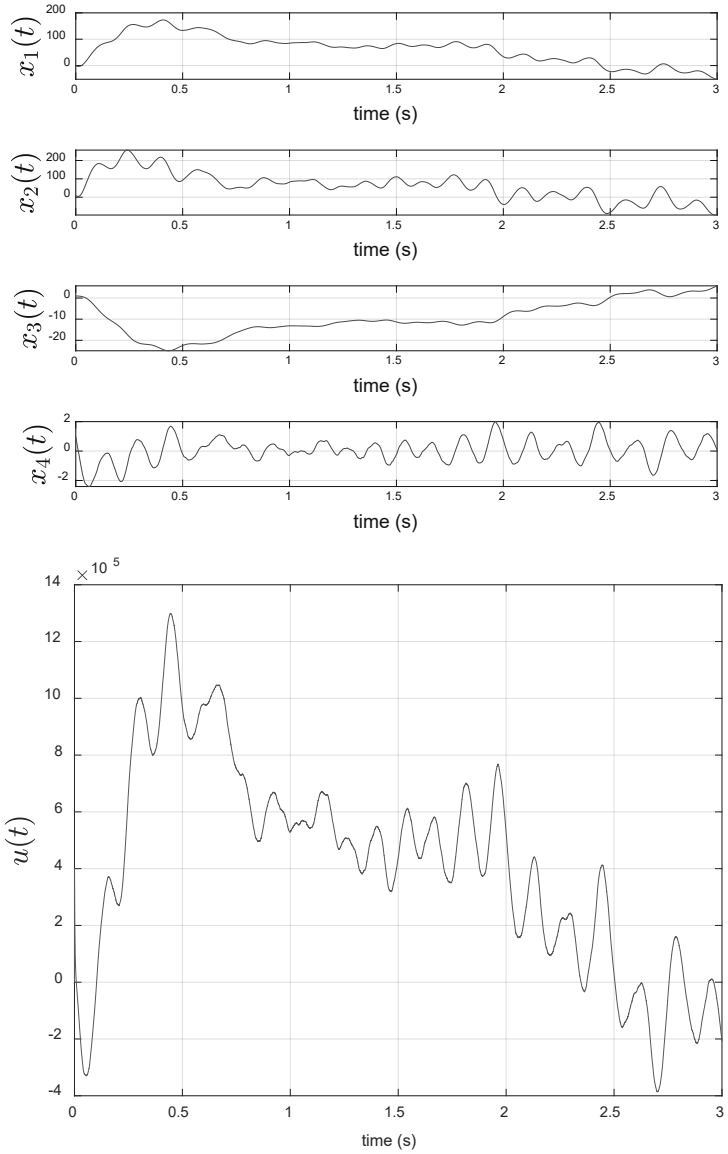

Figure 5. Asymptotic behavior of the state of dynamic system (**top**) and optimal control (**bottom**).

7. Application 3: Optimal Control of a Vehicle Active Suspension System with Damp

The model analyzed in this subsection is given in [29]. In this application, a damp b_s is added to the quarter-car suspension given in Section 6, see Figure 6. The parameters in Figure 6 are: the sprung mass (m_s), the unsprung mass (m_u), the suspension spring constant (k_s), and the tire spring constant (k). Let z_s, z_u, and r be the vertical displacements of the sprung mass, the unsprung mass, and the road disturbance, respectively. The equations of motion are given by:

$$m_s z_s''(t) = -k_s(z_s(t) - z_u(t)) + b_s(z_u' - z_s') + u(t), \quad (62)$$

$$m_u z_u''(t) = k_s(z_s(t) - z_u(t)) - k(r(t) - z_u(t)) - b_s(z_u' - z_s') - u(t). \quad (63)$$

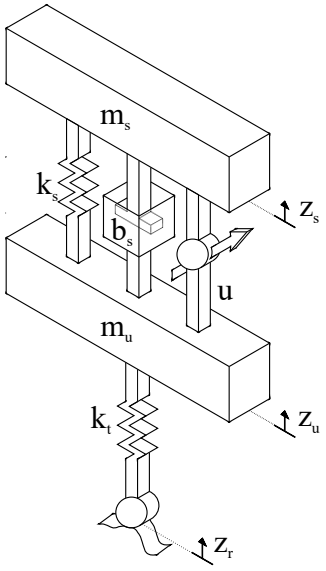

Figure 6. Quarter vehicle model of active suspension system.

Now, defining $x_1(t) = z_s(t)$, $x_2(t) = z_u(t)$, $x_3(t) = z_s'(t)$, and $x_4(t) = z_u'(t)$, the equations of motion in (62) and (63) can be expressed in matrix form as:

$$dx(t) = (Ax(t) + Bu(t))dt + Fr(t) \tag{64}$$

where $dx(t) = \begin{bmatrix} dx_1(t) \\ dx_2(t) \\ dx_3(t) \\ dx_4(t) \end{bmatrix}$, $A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{k_s}{m_s} & \frac{k_s}{m_s} & -\frac{k_s}{m_s} & \frac{k_s}{m_s} \\ \frac{k_s}{m_u} & -\frac{(k_s+k)}{m_u} & \frac{b_s}{m_u} & -\frac{b_s}{m_u} \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{m_s} \\ -\frac{1}{m_u} \end{bmatrix}$, $F = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{k}{m_u} \end{bmatrix}$,

and we assume that the road profile $r(t)$ is represented by a function with hidden Markovian switchings:

$$r(t) = \begin{cases} a(\psi(t))\{1 - cos(8\pi t)\}, & \tau_p \leq t \leq \tau_{p+1} \\ 0 \quad otherwise \end{cases} \tag{65}$$

where $a(1) = 0.05$ (road bump height is 10 cm), $a(2) = 0.025$ (road bump height is 16 cm), and τ_p, $p = 1, 2, \ldots$ are the random jump times of $\psi(t)$. In our case, we consider that the dynamic system (64) evolves with additional white noise, that is:

$$dx(t) = (Ax(t) + Bu(t) + Fr(t))dt + \sigma dW(t) \tag{66}$$

and we wish to minimize the discounted expected cost:

$$V_\alpha(x, \Psi, u) := \mathbb{E}^u_{x,\Psi}\left[\int_0^\infty e^{-\alpha s}\{x^T(s)Dx(s) + u^T(s)Ru(s)\}ds\right],$$

subject to (66) and (65). Considering the infinitesimal generator given in (53) with $z_r(t) \equiv r(t)$ and the Hamilton–Jacobi–Bellman equation associated as the following problem:

$$\alpha v(x, \Psi) = \max_{u \in U}[x^T Dx + u^T R_1 u + \mathcal{L}^u vs.(x, \Psi, r)] \text{ for all } (x, \Psi) \in \mathbb{R}^n \times S_N,$$

similar arguments to these given in Sections 5 and 6 allow us to find the optimal control f^* and the value function v^* for this setting. In fact:

$$v^*(x, \Psi) = x^T Kx + n(\Psi) + g(r) + c,$$

where $n : S_N \to \mathbb{R}$ is a twice differentiable continuous function, c is a constant, $g : \mathbb{R} \to \mathbb{R}$ is a twice differentiable continuous function, and K is a positive definite matrix. Inserting the derivative of $v(x, \Psi)$ in (43), we get the optimal control:

$$f^*(x, \Psi) = -\overline{R}^{-1}(\Psi)B^T K^T x, \tag{67}$$

where the matrix K satisfies the algebraic Riccati equation:

$$A^T K + KA - KBR^{-1}B^T K + D - \alpha K = 0,$$

$$c = Tr[\sigma\sigma^T K]/\alpha,$$

the function $g \in \mathcal{C}^2(\mathbb{R})$ satisfies the differential equation:

$$a(\Psi(t))g'(r) + \alpha g(r) = 0,$$

and $n(\cdot) \in \mathcal{C}^2(S_N)$ satisfies the partial differential equation:

$$Q^T\Psi(t)n'(\Psi(t)) + \frac{1}{2}Tr[A_2]n''(\Psi(t)) - \alpha n(\Psi(t)))I_4 = 0, \ \ \forall \ \ \Psi(t) \in S_N,$$

where A_2 is as in (42), I_4 is the identity matrix of 4×4, and $n^{'}$ and $n^{''}$ are the gradient and the Hessian of the n, respectively.

Simulation results. To solve the Wonhan filter, we use the numerical method given in ([18], Section 8.4) considering that the Markov chain $\psi(t)$ has two states and that can be only observed through $dy(t) = h(\psi(t)) + \sigma_0 dB(t)$. The following data were used: $\sigma = 1$, $\sigma_0 = 1$, $\alpha = 0.01$, $a(1) = 0.05$, $a(2) = 0.08$, $\Psi_1(0) = 0.4$, $\Psi_2(0) = 0.6$, $h(1) = 1, h(2) = 2$, $R = 1.0239 \times 10^{-5}$, $m_s = 300$ kg, $m_u = 60$ kg, $k_s = 1600$ N/m, $k = 190{,}000$ N/m, $b_s = 1000$ N/m, and:

$$Q = \begin{bmatrix} -0.2 & 0.2 \\ 0.4 & -0.4 \end{bmatrix}.$$

Figure 7 shows the solution of the Wonham filter equation and the states of the hidden Markov chain $\psi(t)$. As can be seen, in the time interval $[2, 4]$, $\Psi_1(1) = \mathbb{P}(\psi(t) = 1 \mid y(s),\ 0 \leq s \leq 1) \geq \Psi_2(1)$, implying that the Markov chain with a probability greater than 0.5 is in state 1.

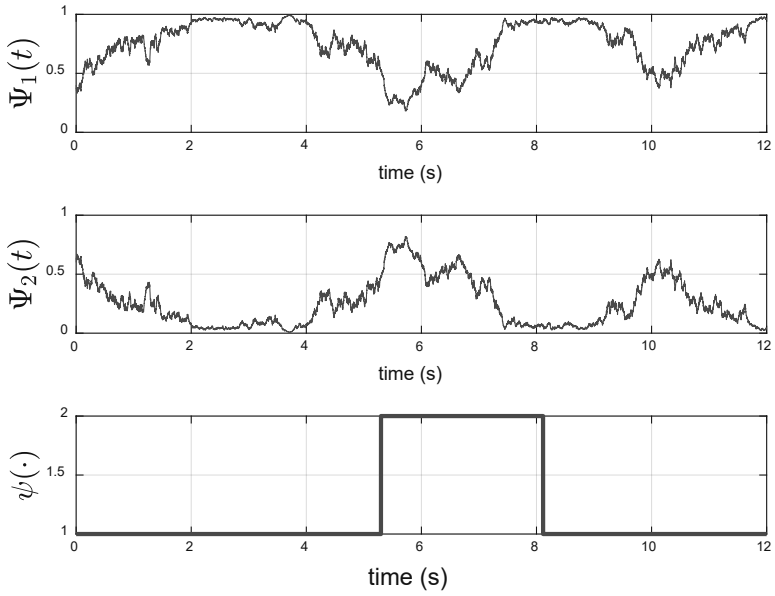

Figure 7. Wonham filter and hidden Markov chain (time interval [2, 4]).

The asymptotic behavior of the optimal control (67) is given in Figure 8 (bottom). It is interesting to note that this control minimizes the magnitude of the sprung mass, $x_1 = z_s$, and unsprung mass, $x_2 = z_u$, al well as their velocities, $x_3 = z_s^{'}$ and $x_4 = z_u^{'}$, after $t = 12$ s, see Figure 8 (top).

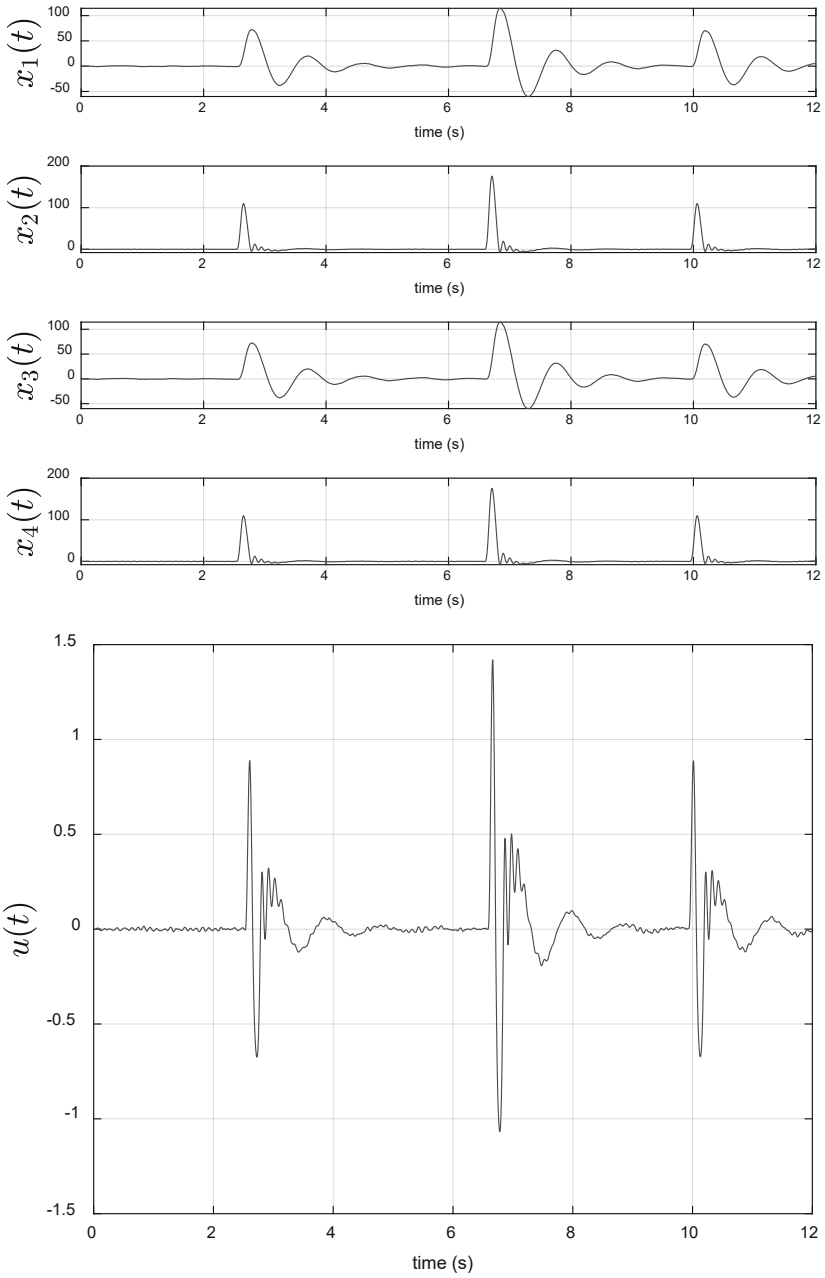

Figure 8. Asymptotic behavior of the state of the dynamic system (**top**) and optimal control (**bottom**).

8. Application 4: Optimal Pollution Control with Average Payoff

The application studies the pollution accumulation incurred by the consumption of a certain product, such as gas or petroleum, see [30]. The stock of pollution $x(\cdot)$ is governed by the controlled diffusion process:

$$\mathrm{d}x(t) = [u(t) - \eta(\psi(t))x(t)]\mathrm{d}t + k\mathrm{d}W(t),\ x(0) = x > 0, \tag{68}$$

where $u(t)$ represents the pollution flow generated by an entity due to the consumption of the product, $\eta(\psi(t))$ represents the decay rate of pollution, chosen at each time by nature,

and k is a positive constant. We shall assume that $u(t) \in U = [0, \gamma]$ is bounded and the parameter γ represents the consumption/production restriction. Let $\psi(t)$ be a Markov chain with two states $E = \{1, 2\}$ and a generator Q given by:

$$\begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} = \begin{pmatrix} -\lambda_0 & \lambda_0 \\ \lambda_1 & -\lambda_1 \end{pmatrix}.$$

The reward rate $r : [0, \infty) \times E \times U \to \mathbb{R}$ in this example represents the social welfare and is defined as:

$$r(x, i, u) := F(u) - a(i)x, \ \forall \ (x, i, u) \in [0, \infty) \times E \times U, \tag{69}$$

where $F \in C^2(0, \infty) \cap C(0, \infty)$ and $D = a(i)x \in C([0, \infty) \times E)$ is the social utility of the consumption u and the social disutility of the pollution (x, i), respectively. We assume that the function F in (69) satisfies:

$$\begin{cases} F'(u) > 0, \quad F''(u) < 0, \\ F'(\infty) = F(0) = 0, \quad F'(0+) = F(\infty) = \infty, \end{cases}$$

Clearly, (68) is a liner stochastic differential equation, and satisfies Assumption 1.

Now, we define the Banach space $\mathcal{B}_w(\mathbb{R} \times E)$ and use $w(x, i) := x + i$, $w(x, \Psi) = \sum_{i=1}^{2} \Psi_i w(x.i) = \Psi_1(x + 1) + \Psi_2(x + 2) = x + (1 - \Psi_1)$. Hence, $\lim_{x \to +\infty} w(x, \Psi) = +\infty$ and Assumption 2*i* holds. On the other hand, since the utility function $F(\cdot)$ is continuous on the compact interval $U = [0, \gamma]$, then:

$$|r(x, i, u)| = |F(u) - a(i)x| \leq (\max_{u \in [0, \gamma]} F(u) + \max_{i \in \{1,2\}} a(i))(x + i) = Mw(x, i)$$

where $M := \max_{u \in [0, \gamma]} F(u) + \max_{i \in \{1,2\}} a(i)$; thus, Assumption 3 holds. Note that:

$$\mathcal{L}^{u,\Psi} w(x, \Psi) = u - \eta(i)x - \lambda_0 \Psi_1 + \lambda_1(1 - \Psi_1), \quad \text{for all } x > 0.$$

Thus, taking $q := \max_{i \in E} \eta(i)$ and $p := \max_{u \in [0, \gamma]} u - (\lambda_0 - \lambda_1)\Psi_1$ we obtain:

$$\mathcal{L}^u w(x, \Psi) \leq -pw(x, \Psi) + q \quad \text{for all } x > 0.$$

Therefore, Assumption 2(ii) holds. It can be proven that the process (68) satisfies Assumption 2.6 in [1]; thus, by ([1], Theorem 2.8), $x(t)$ is exponentially ergodic (Assumption 4). In this application, we seek a policy u that maximizes the long-run average welfare $J(x, i, f)$:

$$J(x, i, u) := \liminf_{T \to \infty} \frac{1}{T} \mathbb{E}^u_{x,i} \left[\int_0^T [F(u) - a(i)x] \mathrm{d}t \right].$$

We propose $v(x, \Psi) = v(x) + h(\Psi)$, where $v \in C^2(\mathbb{R} \times E) \cap \mathcal{B}_w(\mathbb{R} \times E)$ and $h \in C^2(S_N)$ as a solution that verify the HJB Equation (27) associated with this pollution control problem. Simple calculations allow us to conclude that the policy on consumption/pollution takes the form:

$$u := f(x, \Psi) = \begin{cases} I(-v'(x)) & \text{if } F'(\gamma) < -v'(x), \\ \gamma & \text{if } F'(\gamma) \geq -v'(x). \end{cases}$$

where $I(-v'(x))$ is the inverse function of derivative F', $f \in \mathbb{F}$.

9. Concluding Remarks

Under hypotheses such as uniform ellipticity in Assumption 1*c*, the Lyapunov-like conditions in Assumption 2, and the w-exponential ergodicity in (4) for the average criterion, this work shows the existence of optimal controls for the control problems with discounted and average payoffs, where the dynamic system evolves according to switching diffusion with hidden states. To conclude, we conjecture that the results obtained in this work still hold (with obvious changes) if the hidden Markov chain (ψ) in (1) is replaced with any other diffusion process. Furthermore, these results can be extended to constrained and unconstrained nonzero-sum stochastic differential games with additive structures, which will allow us to model a larger class of practical systems. This will be a topic in future works.

Author Contributions: Conceptualization, B.A.E.-T.; Formal analysis, B.A.E.-T. and J.G.-M.; Investigation, B.A.E.-T., J.G.-M. and G.A.; Methodology, B.A.E.-T., J.G.-M. and J.D.R.-A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Escobedo-Trujillo, B.A.; Hernández-Lerma, O. Overtaking optimality for controlled Markov-modulated diffusions. *J. Optim.* **2011**, *61*, 1405–1426. [CrossRef]
2. Borkar, V.S. The value function in ergodic control of diffusion processes with partial observations. *Stoch. Stoch. Rep.* **1999**, *67*, 255–266. [CrossRef]
3. Borkar, V.S. Dynamic programming for ergodic control with partial observations. *Stoch. Process. Their Appl.* **2003**, *103*, 293–310. [CrossRef]
4. Rieder, U.; Bäuerle, N. Portfolio optimization with unobservable Markov-modulated drift Process. *J. Appl. Probab.* **2005**, 362–378. [CrossRef]
5. Tran, K. Optimal exploitation for hybrid systems of renewable resources under partial observation. *Nonlinear Anal. Hybrid Syst.* **2021**, *40*, 101013. [CrossRef]
6. Tran, K.; Yin, G. Stochastic competitive Lotka–Volterra ecosystems under partial observation: Feedback controls for permanence and extinction. *J. Frankl. Inst.* **2014**, *351*, 4039–4064. [CrossRef]
7. Mao, X.; Yuan, C. *Stochastic Differential Equations with Markovian Switching*; World Scientific Publishing Co.: London, UK, 2006. Available online: https://www.worldscientific.com/doi/pdf/10.1142/p473 (accessed on 20 March 2022). [CrossRef]
8. Yin, G.G.; Zhu, C. Hybrid Switching Diffusions. In *Stochastic Modelling and Applied Probability*; Properties and Applications; Springer: New York, NY, USA, 2010; Volume 63, p. xviii+395. [CrossRef]
9. Yin, G.; Mao, X.; Yuan, C.; Cao, D. Approximation methods for hybrid diffusion systems with state-dependent switching processes: numerical algorithms and existence and uniqueness of solutions. *SIAM J. Math. Anal.* **2009**, *41*, 2335–2352. [CrossRef]
10. Yu, L.; Zhang, Q.; Yin, G. Asset allocation for regime-switching market models under partial observation. *Dynam. Syst. Appl.* **2014**, *23*, 39–61.
11. Ghosh, M.K.; Arapostathis, A.; Marcus, S.I. Optimal control of switching diffusions with application to flexible manufacturing systems. *SIAM J. Control Optim.* **1993**, *31*, 1183–1204. [CrossRef]
12. Ghosh, M.K.; Marcus, S.I.; Arapostathis, A. Controlled switching diffusions as hybrid processes. In Proceedings of the International Hybrid Systems Workshop, New Brunswick, NJ, USA, 22–25 October 1995; Springer: Berlin/Heidelberg, Germany, 1995; pp. 64–75.
13. Zhang, X.; Zhu, Z.; Yuan, C. Asymptotic stability of the time-changed stochastic delay differential equations with Markovian switching. *Open Math.* **2021**, *19*, 614–628. [CrossRef]
14. Zhu, C.; Yin, G. Asymptotic properties of hybrid diffusion systems. *SIAM J. Control Optim.* **2007**, *46*, 1155–1179. [CrossRef]
15. Wonham, W.M. Some applications of stochastic differential equations to optimal nonlinear filtering. *J. SIAM Control Ser. A* **1965**, *2*, 347–369. [CrossRef]
16. Elliott, R.J.; Aggoun, L.; Moore, J.B. *Hidden Markov Models: Estimation and Control*; Springer: Berlin/Heidelberg, Germany, 1995.
17. Cohen, S.N.; Elliott, R.J. *Stochastic Calculus and Applications*, 2nd ed.; Probability and Its Applications; Springer: Cham, Switzerland, 2015; p. xxiii+666. [CrossRef]

18. Yin, G.; Zhang, Q. *Discrete-Time Markov Chains: Two-Time-Scale Methods and Applications*; Stochastic Modelling and Applied Probability; Springer: New York, NY, USA, 2006.
19. Yin, G.G.; Zhu, C. *Hybrid Switching Diffusions: Properties and Applications*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009; Volume 63.
20. Protter, P.E. Stochastic integration and differential equations. In *Stochastic Modelling and Applied Probability*, 2nd ed.; Version 2.1, Corrected Third Printing; Springer: Berlin/Heidelberg, Germany, 2005; Volume 21, p. xiv+419. [CrossRef]
21. Chigansky, P. An ergodic theorem for filtering with applications to stability. *Syst. Control Lett.* **2006**, *55*, 908–917. [CrossRef]
22. Kunita, H. Asymptotic behavior of the nonlinear filtering errors of Markov processes. *J. Multivar. Anal.* **1971**, *1*, 365–393. [CrossRef]
23. Lu X.; Yin, G.; Guo, X. Infinite Horizon Controlled Diffusions with Randomly Varying and State-Dependent Discount Cost Rates. *J. Optim. Theory Appl.* **2017**, *172*, 535–553. [CrossRef]
24. Ghosh, M.K.; Arapostathis, A.; Marcus, S.I. Ergodic control of switching diffusions. *SIAM J. Contr. Optim* **1997**, *35*, 1962–1988. [CrossRef]
25. SchÄl, M. Conditions for optimality and for the limit of n-stage optimal policies to be optimal. *Z. Wahrs. Verw. Gerb.* **1975**, *32*, 179–196. [CrossRef]
26. Ghosh, M.K.; Marcus, S.I. Stochastic differential games with multiple modes. *Stoch. Anal. Appl.* **1998**, *16*, 91–105. [CrossRef]
27. Nguyen, L.H.; Seonghun, P.; Turnip, A.; Hong, K.S. Application of LQR Control Theory to the Design of Modified Skyhook Control Gains for Semi-Active Suspension Systems. In Proceedings of the ICROS-SICE International Joint Conference 2009, Fukuoka, Japan, 18–21 August 2009; pp. 4698–4703.
28. Escobedo-Trujillo, B.; Garrido-Meléndez, J. Stochastic LQR optimal control with white and colored noise: Dynamic programming technique. *Rev. Mex. Ing. QuÍmica* **2021**, *20*, 1111–1127. [CrossRef]
29. Maurya, V.K.; Bhangal, N.S. Optimal Control of Vehicle Active Suspension System. *J. Autom. Control. Eng.* **2018**, *6*, 1111–1127. [CrossRef]
30. Kawaguchi, K.; Morimoto, H. Long-run average welfare in a pollution accumulation model. *J. Econom. Dynam. Control* **2007**, *31*, 703–720. [CrossRef]