mathematics

MDPI

Article

Robust Statistic Estimation in Constrained Optimal Control Problems of Pollution Accumulation (Part II: Markovian Switchings)

Beatris Adriana Escobedo-Trujillo [1], José Daniel López-Barrientos [2,*], Carmen Geraldi Higuera-Chan [3] and Francisco Alejandro Alaffita-Hernández [4]

[1] Facultad de Ingeniería, Universidad Veracruzana, Coatzacoalcos 96535, Mexico
[2] Facultad de Ciencias Actuariales, Universidad Anáhuac Mexico, Naucalpan de Juárez 52786, Mexico
[3] Departamento de Matemáticas, Universidad de Sonora, Hermosillo 83000, Mexico
[4] Centro de Investigación en Recursos Energéticos y Sustentables, Universidad Veracruzana, Coatzacoalcos 96535, Mexico
* Correspondence: daniel.lopez@anahuac.mx; Tel.: +52-(55)-5627-0210 (ext. 8506)

Abstract: This piece is a follow-up of the research started by the authors on the constrained optimal control problem applied to pollution accumulation. We consider a dynamic system governed by a diffusion process with multiple modes that depends on an unknown parameter. We will study the components of the model and their restrictions and propose a scheme to solve the problem in which it is possible to determine (adaptive) policies that maximize a suitable discounted reward criterion using standard dynamic programming techniques in combination with discrete estimation methods for the unknown parameter. Finally, we develop a numerical example to illustrate our results with a particular case of the method of minimum least square error approximation.

Keywords: consistent estimators; multiple Markovian modes; discounted cost; maximum likelihood estimators; least square errors

MSC: 93E10; 93E20; 93E24; 60J60

Citation: Escobedo-Trujillo, B.A.; López-Barrientos, J.D.; Higuera-Chan, C.G.; Alaffita-Hernández, F.A. Robust Statistic Estimation in Constrained Optimal Control Problems of Pollution Accumulation (Part II: Markovian Switchings). *Mathematics* **2023**, *11*, 1045. https://doi.org/10.3390/math11041045

Academic Editor: David Barilla

Received: 18 January 2023
Revised: 8 February 2023
Accepted: 16 February 2023
Published: 18 February 2023

1. Introduction

Pollution control in large cities is a problem of great interest worldwide, and that is why various organizations are continually seeking strategies to mitigate it. As for scientists, they have begun to analyze models that describe the stock of pollution through ordinary and stochastic differential equations. In particular, optimal control theory has been applied for the optimal management of pollution in economic sciences. This theory considers an economy that consumes some good and, as a by-product of that consumption, generates pollution. The hypotheses in our model are as follows:

- The contamination stock is only gradually dissolved by the environment;
- The growth rate of the pollution is constant or random;
- The flow of pollution is constrained so that it satisfies some mandatory global standards in order to promote sustainable development (see, for instance, ref. [1]).

Social welfare is defined by the net utility from the consumption of some good vis à vis the disutility caused by pollution. Our objective is to find an optimal consumption policy for society. That is, we seek to maximize the difference between the utility function of consumption vs. the disutility caused by the polluting stock (see [2,3]).

This paper represents the second part of a project related to the constrained optimal control problem of pollution accumulation with an unknown parameter. To be in context, we begin by briefly summarizing the results obtained in *Robust statistic estimation in constrained optimal control problems of pollution accumulation (Part I)* (see [4]).

The first part considers the scenery where the dynamic system is given by a diffusion process and depends on an unknown parameter, say θ. First, assuming the parameter θ as

known, an approach with restrictions and one without them is proposed. The respective optimal value functions are V_θ^* and $V_{\theta,\lambda}^*$. Then, estimation techniques for the parameter θ are applied and later combined with the characterizations and results previously analyzed. Roughly speaking, one of the results considers a sequence of estimated parameters (θ_m:m = 1, ...) such that $\theta_m \to \theta$, then the value functions converge (in some sense), that is

$$V_{\theta_n}^* \to V_\theta^* \text{ and } V_{\theta_m,\lambda}^* \to V_{\theta,\lambda}^*.$$

Another result states the existence of optimal policies such that $\pi_{\theta_m} \to \pi_\theta$. Furthermore, the relationship between the value functions V_θ^* and $V_{\theta,\lambda}^*$ is shown.

In this piece, we obtain similar results considering the case where the dynamics of the stock of pollution evolves as a diffusion process with Markovian switching whose drift function, as well as the reward function, depends on the unknown parameter θ. In addition, we impose some natural constraints on the performance index.

To avoid confusion, we try to preserve the notation of Part I in this work. First, a constrained control problem is proposed. Subsequently, assuming θ is the real parameter, we study a standard control problem under the discounted criterion, where it is possible to apply standard techniques and dynamic programming tools to determine optimal policies. Then a (discrete) procedure to estimate the unknown parameter θ is applied in combination with the standard results formerly mentioned to obtain the so-called adaptive policies that maximize a discounted reward criterion with constraints.

The idea is to estimate the parameter θ, and then solve the optimal control problem when such an estimated value is replaced in the problem. In the literature, this approach is known as the *Principle of Estimation and Control*. This problem has been studied in several contexts. For instance, refs. [5–8] and the references therein are about stochastic control systems evolving in discrete time. On the other hand, adaptive optimal control for continuous time is studied in [9–11]. The estimation for diffusion processes using discrete observations has been studied in the works [12–16].

Dynamic optimization has been used to study the problem of pollution accumulation in the past; for example, the papers [17,18] use a linear quadratic model to explain this phenomenon, the article [2] deals with the average payoff in a deterministic framework, while [3,19] extend the former's approach to a stochastic context, and [20] uses a robust stochastic differential game to model the situation. The study [21] is a statistical analysis of the impact of air pollution on public health. In order to develop adaptive policies that are almost surely optimal for the restricted optimization problem under the discounted reward on an infinite horizon with Markovian switchings, we use a statistical estimation approach to determine the unknown parameter θ. These adaptive policies are created by replacing the estimates into optimal stationary controls (that is, the use of the PEC); for more information, see the works of Kurano and Mandl (cf. [7,8]). The statistic estimation method we use for the unknown parameter θ is the so-called *least square estimator* for stochastic differential equations based on many discrete observations. This resembles existing robust estimation techniques, such as the H_∞ method, in the fact that in the applications, the dynamic systems are linear. However, the computational complexity of these techniques is greater. Indeed, with our least square estimator, only the inverse of a matrix must be calculated to obtain the estimator, while there are way more computations to be performed in the other algorithms (see [22,23]). Most risk analysts will not be as familiar with our methods as they are with, for example, the model predictive control, MATLAB's robust control toolbox, or the polynomial chaos expansion method, which have been used in the literature to address similar issues. Since we review a constructive method for robust and adaptive control under deep uncertainty, our findings are similar to those reported in the article [24]. Moreover, our methods also resemble the adaptive moving mesh method for optimal control problems in viscous incompressible fluid used in [25].

This piece can be also be considered an extension of [26–29], who also study adaptive constrained optimal control methods. In fact, ref. [28] studies a constrained optimal control problem, but unlike our case, there, all the parameters are known, while [26] does the same

but in the context of pollution accumulation. The references [27,29] study an unconstrained adaptive optimal control problem. Finally, it is important to highlight the numerical estimation technique that illustrates the results of this article.

The rest of the paper is organized as follows. We present the elements of our model and assumptions in Section 2. Next, Section 3 introduces our optimality criterion and the main results; an interesting numerical example illustrating our results is given in Section 4. We give our conclusions in Section 5, and finally, we included the proof of the important (but rather distracting) Theorem A1 on the convergence of the HJB equation under the topology of relaxed controls in Appendix A.

Notation and Terminology

For vectors $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ and matrices $A = \left(A_{k,p}\right) \in \mathbb{M}_n(\mathbb{R})$, we denote by $|\cdot|$ the Euclidean norm, that is,

$$|x|^2 := \sum_{k=1}^{n} x_k^2 \text{ and } |A|^2 := \text{Tr}(AA^\top) = \sum_{k,p=1}^{n} A_{k,p}^2,$$

where $A^\top$ and $\text{Tr}(\cdot)$ denote the transpose and the trace of matrix, respectively. As an abbreviation, we write ∂_i and ∂_{ij}^2 to refer to $:= \frac{\partial}{\partial x_i}$, and $\frac{\partial^2}{\partial x_i \partial x_j}$, respectively.

Given a Borel set B, we denote by $\mathcal{B}(B)$ its natural σ-algebra. As usual, $\mathcal{C}(\mathcal{O})$, stands for the space of continuous functions whose domain is $\mathcal{O}$ and

$$\mathcal{C}(\mathcal{O} \times E) := \{\nu : \mathcal{O} \times E \to \mathbb{R}^n : \nu(\cdot, i) \in \mathcal{C}(\mathcal{O} \times \mathcal{E}) \text{ for each } i \in E\}.$$

Consequently we denote $\mathcal{C}_b(\mathcal{O} \times E)$ as the subspace of $\mathcal{C}(\mathcal{O} \times E)$ composed by bounded functions. The set $\mathcal{C}^\kappa(\mathcal{O} \times E) := \{\nu : \mathcal{O} \times E \to \mathbb{R}^n : \nu(\cdot, i) \in \mathcal{C}^\kappa(\mathcal{O} \times \mathcal{E}) \text{ for each } i \in E\}$, where $\mathcal{C}^\kappa(\mathcal{O} \times \mathcal{E})$ is the space of all real-valued continuous functions f on the bounded, open and connected subset $\mathcal{O} \subset \mathbb{R}^n$ with continuous derivatives up to order $\kappa \in \mathbb{N}$.

Fix $p \geq 1$ and a measure space $(\Omega, \mathcal{F}, \mu)$, we denote $\mathcal{L}^p(\Omega \times E)$ as the Lebesgue space of functions g on $\Omega \times E$ such that $\int_\Omega |g(x,i)|^p \mu(\mathrm{d}x) < \infty$ for $i \in E$.

Let X and Y be Borel spaces. A stochastic kernel $Q(\cdot|\cdot)$ on X given Y is a function such that $Q(\cdot|y)$ is a probability measure on X for each $y \in Y$ and $Q(B|\cdot)$ is a measurable function on Y for each $B \in \mathcal{B}(X)$.

Finally, the set $\mathcal{P}(B)$ denotes the family of probability measures on B endowed with the topology of weak convergence.

2. Model Formulation and Assumptions

Taking as reference the problem analyzed in Part I, we consider the scenery where the dynamics of the pollution stock is modeled as an n-dimensional controlled stochastic differential equation (SDE) with Markovian switching. Specifically, such a dynamic takes the form

$$\mathrm{d}x(t) = b(x(t), \psi(t), u(t), \theta)\mathrm{d}t + \sigma(x(t), \psi(t))\mathrm{d}W(t), \ (x(0), \psi(0)) = (x_0, \psi_0), \ t \geq 0, \quad (1)$$

where $E = \{1, 2, \dots, N\}$, $b : \mathbb{R}^n \times E \times U \times \Theta \to \mathbb{R}^n$ and $\sigma : \mathbb{R}^n \times E \to \mathbb{R}^{n \times d}$ are given functions, $W(\cdot)$ is an $\mathcal{F}_t$-adapted d-dimensional Wiener process such that $W(t) - W(s)$ and $\mathcal{F}_s$ are pairwise independent, $W(\cdot)$ is independent of $\psi(\cdot)$, and the evolution of the Markov chain ψ has intensity $Q = (q_{ij})_{i,j \in E}$ and transition rule given by

$$\mathbb{P}(\psi(t + \Delta t) = j | \psi(t) = i, (x(s), \psi(s)), s \leq t) = \begin{cases} q_{ij}\Delta t + o(\Delta t), & \text{if } i \neq j, \\ 1 + q_{ii}\Delta t + o(\Delta t), & \text{if } i = j, \end{cases} \quad (2)$$

for $t \geq 0$ and $\sum_{j=1}^{N} q_{ij} = 0$. The compact set $U \subset \mathbb{R}^{n_1}$ is called the control set. In the context of our problem, $u(t)$ is a stochastic process on U such that, at time t, it represents the flow

of consumption which, in turn, is considered bounded to reflect the policies and rules imposed by governments or social entities.

It is important to remark that throughout this work, we assume that θ is an unknown parameter taking values on a compact set $\Theta \subset \mathbb{R}^m$, which is called the parameter set. Note that in the context of pollution problems, θ can be seen as the pollution decay rate.

Now we define the so-called randomized policies, also known as relaxed controls, or just policies.

Definition 1. *A* policy *is a family $\pi := (\pi_t(\cdot|\cdot,\cdot))_{t\geq 0}$ of stochastic kernels on $\mathcal{B}(U) \times \mathbb{R}^n \times E$ (see Section 1). We denote by Π the set of stationary policies. In particular, a randomized policy is said to be stationary if there is a probability measure $\pi(\cdot|x,i) \in \mathcal{P}(U)$ such that $\pi_t(\cdot|x,i) = \pi(\cdot|x,i)$ for all $t \geq 0$ and $(x,i) \in \mathbb{R}^n \times E$. Let $\mathbb{F}$ be the set of measurable functions f from $\mathbb{R}^n \times E$ to U. We denote the set of stationary Markov policies as $\mathbb{F}_1 := \{f : \mathbb{R}^n \times E \to U : \text{ for each } i \in E, f(\cdot,i) \in \mathbb{F}\}$.*

For each randomized policy $\pi \in \Pi$ and a function whose domain is contained to U, say $v : \mathbb{R}^n \times E \times U \times \Theta \to \mathbb{R}$, we use the abbreviated notation

$$v(x,i,\pi,\theta) := \int_U v(x,i,u,\theta)\pi(\mathrm{d}u|x,i). \tag{3}$$

A suitable adjustment should be made for functions with a different domain.

We endow Π with a topology (see [30]) determined by the convergence criterion defined below (see [31,32], Lemma 3.2 in [30,33].

Definition 2. *A sequence $(\pi_m)_{m\in\mathbb{N}}$ in Π converges to $\pi \in \Pi$ if*

$$\int_{\mathbb{R}^n} g(x,i)h(x,i,\pi_m)\mathrm{d}x \to \int_{\mathbb{R}^n} g(x,i)h(x,i,\pi)\mathrm{d}x.$$

for all $g \in \mathcal{L}^1(\mathbb{R}^n \times E)$, and $h \in \mathcal{C}_b(\mathbb{R}^n \times E \times U)$ (see (3)). Since this mode of convergence was introduced by Warga (cf. [30]), we denote it as $\pi_m \xrightarrow{W} \pi$.

For $\nu(\cdot,\cdot,\theta) \in \mathcal{C}^2(\mathbb{R}^n \times E)$, $u \in U$ and $\theta \in \Theta$, the infinitesimal generator associated with the process $(x(\cdot), \psi(\cdot))$ is

$$\mathbb{L}^{u,\theta}\nu(x,i,\theta) := \sum_{k=1}^{n} b_k(x,i,u,\theta)\partial_k\nu(x,\theta,i) + \frac{1}{2}\sum_{k,\ell=1}^{n} a^{k,\ell}(x,i)\partial^2_{k,\ell}\nu(x,i,\theta) + \sum_{j=1}^{N} q_{ij}\nu(x,i,\theta),$$

where b_k is the k-th component of the drift function b, and $a^{k,\ell}$ is the (k,ℓ) component of the matrix $a(\cdot,\cdot) := \sigma(\cdot,\cdot)\sigma(\cdot,\cdot)^\top$. As in (3), for each policy $\pi \in \Pi$, we write

$$\mathbb{L}^{\pi,\theta}\nu(x,i,\theta) := \int_U \mathbb{L}^{u,\theta}\nu(x,i,\theta)\pi(\mathrm{d}u|x,i).$$

The following set of assumptions and conditions ensures the existence and uniqueness of a strong solution as well as stability of the dynamic system (1) and (2) (see [31,33–35]).

Assumption 1. *(a) The random process (1) belongs to a complete probability space $\left(\Omega, \mathcal{F}, \mathbb{P}^{u,\theta}\right)$. Here, $\{\mathcal{F}_t\}_{t\geq 0}$ is a filtration on $(\Omega,\mathcal{F})$ such that each $\mathcal{F}_t$ is complete relative to $\mathcal{F}$, and $\mathbb{P}^{u,\theta}$ is the law of the state process $x(\cdot)$ given the parameter $\theta \in \Theta$ and the control $u(\cdot)$.*

(b) The drift function $b(\cdot,\cdot,\cdot,\cdot)$ in (1) is continuous and satisfies that for each $R > 0$, there exist non-negative constants $K_\theta(R)$ and $D(R)$ such that, for all $u \in U$, all $|\theta_1|, |\theta_2| \leq R$ and $|x|, |y| \leq R$,

$$|b(x,i,u,\theta) - b(y,i,u,\theta)| \leq K_\theta(R)|x-y|,$$

$$|b(x,i,u,\theta_1) - b(x,i,u,\theta_2)| \leq D(R)|\theta_1 - \theta_2|,$$

Moreover, the function $u \mapsto b(x,i,u,\theta)$ is continuous on U.

(c) *The diffusion coefficient σ satisfies a local Lipschitz condition; that is, for each $R > 0$, there exists a constant $K_1(R) > 0$ such that, for all $|x|, |y|$ less than R,*

$$|\sigma(x,i) - \sigma(y,i)| \leq K_1(R)|x-y|.$$

(d) *A global linear growth condition is satisfied*

$$\sup_{(u,\theta)\in U\times\Theta} |b(x,i,u,\theta)|^2 + |\sigma(x,i)|^2 \leq \tilde{K}(1+|x|^2) \quad \text{for all } x \in \mathbb{R}^n,$$

where $\tilde{K} > 0$ is a constant.

(e) *The matrix $a(x,i) := \sigma(x,i)\sigma(x,i)^\top$ satisfies that, for some constant $K_2 > 0$,*

$$x^\top a(y,i)x \geq K_2|x|^2 \text{ for all } x, y \in \mathbb{R}^n.$$

Remark 1. *(i) Properties such as continuity or Lipschitz continuity given in Assumption 1 are inherited to the drift function $b(x,i,\pi,\theta)$.*

(ii) Under Assumption 1, once a policy $\pi \in \Pi$ and a parameter $\theta \in \Theta$ are fixed, the references [31] and [33] guarantee the existence of a probability space $(\Omega, \mathcal{F}, \mathbb{P}^{\pi,\theta})$ in which there exists a unique process $x^{\pi,\theta}(\cdot)$ with the Markov–Feller property which, in turn, is an almost surely strong solution.

The next hypothesis is known as the Lyapunov stability condition.

Assumption 2. *There exists a function $w \in \mathcal{C}^2(\mathbb{R}^n \times E)$, $w(\cdot,\cdot) \geq 1$ and constants $d \geq \beta > 0$ such that*

(a) $\lim_{|x|\to\infty} w(x,i) = \infty$ *uniformly in $i \in E$.*

(b) $\mathbb{L}^{\pi,\theta}w(x,i) \leq -\beta w(x,i) + d$ *for all $\pi \in \Pi$, $\theta \in \Theta$ and $(x,i) \in \mathbb{R}^n \times E$.*

Assumption 2 essentially asks for a twice-continuously differentiable function to solve the problem at hand. This hypothesis is equivalent to requiring positive-definite matrices in the context of linear matrix inequalities (see [36] and pages 113–135 in [37]). The existence of a function w with the conditions in Assumption 2 implies that the rate functions involved in our model can be unbounded (see Assumption 3). As in the first part, we define next an adequate space for these functions.

Definition 3. *Let v be a function from $\mathbb{R}^n \times E$ to $\mathbf{R}$, we define its w-norm as*

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

Even more, let $\mathcal{B}_w(\mathbb{R}^n \times E)$ be the Banach space of real-valued measurable functions with finite w-norm.

Let r and c be measurable functions from $\mathbb{R}^n \times E \times U \times \Theta$ to $\mathbb{R}$ identified as reward (social welfare) rate and the cost rate, respectively, and let η from $\mathbb{R}^n \times E \times \Theta$ to $\mathbb{R}$ be another measurable function that models the constraint rate. In the context of pollution accumulation, in some situations, such a restriction is due to each country's legal framework, and the cost of cleaning the environment must be bounded for some given quantity.

Assumption 3. *For each $i \in E$ fixed, the payoff rate $r(\cdot,i,\cdot,\cdot)$, the cost rate $c(\cdot,i,\cdot,\cdot)$ and the constraint rate $\eta(\cdot,i,\cdot)$ are continuous on $\mathbb{R}^n \times U \times \Theta$. Moreover, they are locally Lipschitz on $\mathbb{R}^n$, uniformly on E, U and Θ. That is, for each $R > 0$, there are positive constants $K(R)$ and $K_2(R)$ such that for all $|x|, |y| \leq R$,*

$$\begin{aligned}\sup_{(i,u,\theta)\in E\times U\times\Theta}|r(x,i,u,\theta)-r(y,i,u,\theta)|+\sup_{(i,u,\theta)\in E\times U\times\Theta}|c(x,i,u,\theta)-c(y,i,u,\theta)| &\leq K(R)|x-y|,\\ \sup_{(i,\theta)\in E\times\Theta}|\eta(x,i,\theta)-\eta(y,i,\theta)| &\leq K_2(R)|x-y|.\end{aligned}$$

Even more, the rate functions belong to $\mathcal{B}_w(\mathbb{R}^n\times E)$ *and there exists* $M>0$ *such that for all* $(x,i)\in\mathbb{R}^n\times E$,

$$\sup_{(u,\theta)\in U\times\Theta}|\eta(x,i,\theta)|+\sup_{(u,\theta)\in U\times\Theta}|r(x,i,u,\theta)|+\sup_{(u,\theta)\in U\times\Theta}|c(x,i,u,\theta)|\leq Mw(x,i).$$

3. Discounted Optimality Problems and Main Results

Through this section, we establish the contamination problem of our interest in terms of the terminology of optimal control. To this end, we will introduce the functions that evaluate the behavior of the system throughout the process associated with payments, costs, and restrictions.

In order to avoid confusion, we will preserve the notation and the ordering in the presentation of the results from the first part of the project.

3.1. Discounted Optimality Criterion

Definition 4. *Given the initial state* $(x,i)\in\mathbb{R}^n\times E$, *a parameter value* $\theta\in\Theta$ *and a discount rate* $\alpha>0$, *we define the total expected* α*-discounted reward, cost and constraint when the controller uses a policy* π *in* Π *as*

$$\begin{aligned}V(x,i,\pi,r,\theta) &:= \mathbb{E}^{\pi,\theta}_{x,i}\left[\int_0^\infty \mathrm{e}^{-\alpha t}r(x(t),\psi(t),\pi,\theta)\mathrm{d}t\right],\\ V(x,i,\pi,c,\theta) &:= \mathbb{E}^{\pi,\theta}_{x,i}\left[\int_0^\infty \mathrm{e}^{-\alpha t}c(x(t),\psi(t),\pi,\theta)\mathrm{d}t\right] \text{ and}\end{aligned}$$

$$\overline{\eta}(x,i,\pi,\theta):=\alpha\mathbb{E}^{\pi,\theta}_{x,i}\left[\int_0^\infty \mathrm{e}^{-\alpha t}\eta(x(t),\psi(t),\theta)\mathrm{d}t\right].$$

respectively, and $\mathbb{E}^{\pi,\theta}_{x,i}[\cdot]$ *is the expectation of* $\cdot$ *taken with respect to the probability measure* $\mathbb{P}^{\pi,\theta}$ *when* $(x(t),\psi(t))$ *starts at* (x,i).

Proposition 1. *If Assumptions 1–3 hold, the functions* $V(\cdot,\cdot,\pi,r,\theta)$ *and* $V(\cdot,\cdot,\pi,c,\theta)$ *belong to* $\mathcal{B}_w(\mathbb{R}^n\times E)$ *for each* π *in* Π; *in fact, for each* $(x,i)\in\mathbb{R}^n\times E$ *and* $\theta\in\Theta$ *we have*

$$\sup_{\pi\in\Pi}|V(x,i,\pi,r,\theta)|+\sup_{\pi\in\Pi}|V(x,i,\pi,c,\theta)|\leq 2M(\alpha)w(x,i)$$

where $M(\alpha):=M\frac{\alpha+d}{\alpha\beta}$, *the constants* c *and* d *are as in Assumption 2, and* M *is as in Assumption 3 (b).*

Proposition 1 can be obtained directly using the following inequality, which is an application of Dynkin's formula to the function $v(t,x,i):=\mathrm{e}^{\beta t}w(x,i)$, and Assumption 2 (b) yield that, for all $\pi\in\Pi$, $\theta\in\Theta$, $(x,i)\in\mathbb{R}^n\times E$ and $t\geq 0$,

$$\mathbb{E}^{\pi,\theta}_{x,i}[w(x(t),\psi(t))]\leq \mathrm{e}^{-\beta t}w(x,i)+\frac{d}{\beta}\left(1-\mathrm{e}^{-\beta t}\right). \tag{4}$$

Remark 2. *The function* $\overline{\eta}(\cdot,\cdot,\pi,\theta)$ *is in* $\mathcal{B}_w(\mathbb{R}^n\times E)$ *for each* $\pi\in\Pi$. *Moreover, for each* $(x,i)\in\mathbb{R}^n\times E$, *we have*

$$\sup_{\pi\in\Pi}|\overline{\eta}(x,i,\pi,\theta)|\leq\|\eta\|_w\frac{\alpha+d}{\beta}w(x,i).$$

Let $\theta \in \Theta$ be fixed, and again apply Dynkin's formula to the function V (see Theorem 1.45 in p. 48 in [34] or Theorem 1 (iii) in [38]) to yield the following result.

Proposition 2. *Let Assumptions 1, 2 hold, and let v be a measurable function on $\mathbb{R}^n \times E \times U \times \Theta$ satisfying Assumption 3. Then, for $\pi \in \Pi$, the associated expected α-discounted reward $V(\cdot,\cdot,\cdot,\pi,\theta)$ belongs to $\mathcal{W}^{2,p}(\mathbb{R}^n \times E) \cap \mathcal{B}_w(\mathbb{R}^n \times E)$, and satisfies*

$$\alpha V(x,i,\pi,v,\theta) = V(x,i,\pi,v,\theta) + \mathbb{L}^{\pi,\theta}V(x,i,\pi,v,\theta) \text{ for all } (x,i) \in \mathbb{R}^n \times E \text{ and } \theta \in \Theta. \quad (5)$$

Conversely, if some function $\varphi(\cdot,\cdot,\theta)$ in $\mathcal{W}^{2,p}(\mathbb{R}^n \times E) \cap \mathcal{B}_w(\mathbb{R}^n \times E)$ satisfies Equation (5), then

$$\varphi(x,i,\theta) = V(x,i,\pi,v,\theta) \text{ for all } (x,i) \in \mathbb{R}^n \times E \text{ and } \theta \in \Theta. \quad (6)$$

Even more, if relation (5) is an inequality, then (6) holds with the respective inequality.

Consider that $\mathcal{W}^{\ell,p}(\mathbb{R}^n \times E)$ is the Sobolev space of real-valued measurable functions on $\mathbb{R}^n \times E$ whose derivatives up to order $\ell \geq 0$ are in $\mathcal{L}^p(\mathbb{R}^n \times E)$ for $p \geq 1$.

Given the initial conditions $(x,i) \in \mathbb{R}^n \times E$ a parameter $\theta \in \Theta$, and a constraint function η satisfying Assumption 3, we define the set of policies

$$\mathcal{F}_\theta^{x,i} := \Big\{ \pi \in \Pi |\ V(x,i,\pi,c,\theta) \leq \overline{\eta}(x,i,\pi,\theta) \Big\}. \quad (7)$$

We assume, for the moment, that the set defined in (7) is nonempty. Up to this point, we are in a position to formulate the discounted problem with constraints (DPC), which is defined below.

Definition 5. *Given the initial condition $(x,i) \in \mathbb{R}^n \times E$ and the parameter $\theta \in \Theta$ we say that policy $\pi^* \in \Pi$ is optimal for the DPC if $\pi^* \in \mathcal{F}_\theta^{x,i}$ and*

$$V(x,i,\pi^*,r,\theta) = \sup_{\pi \in \mathcal{F}_\theta^{x,i}} V(x,i,\pi,r,\theta).$$

Furthermore, the function $V^*(x,i,r,\theta) := V(x,i,\pi^*,r,\theta)$ is known as the $\alpha-$discount optimal reward for the DPC.

3.2. Unconstrained Discounted Optimality

The objective of this part is to transform the original DPC (presented above) into an unconstrained problem, and thus, to be able to propose results and techniques known in the literature. To this end, we will apply the Lagrange multipliers technique used in [26]. Take $\lambda \leq 0$ and consider the function

$$r^\lambda(x,i,u,\theta) := r(x,i,u,\theta) + \lambda(c(x,i,u,\theta) - \alpha\eta(x,i,\theta)). \quad (8)$$

For our purpose, r^λ represents the new reward rate. Now recalling (3), we write (8) as

$$r^\lambda(x,i,\pi,\theta) := r(x,i,\pi,\theta) + \lambda(c(x,i,\pi,\theta) - \alpha\eta(x,i,\theta)), \quad \pi \in \Pi, \theta \in \Theta.$$

Remark 3. *For each $\alpha > 0$ and $\lambda < 0$, by direct calculations, it is possible to show that $r^\lambda(\cdot,\cdot,\pi,\theta) \in \mathcal{B}_w(\mathbb{R}^n \times E)$ uniformly in $\pi \in \Pi$ and $\theta \in \Theta$. Even more, by Assumption 3, this new reward rate is a Lipschitz function.*

In the same way as in Definition (4), for all $(x,i) \in \mathbb{R}^n \times E$ and $\theta \in \Theta$, we define the function

$$V(x,i,\pi,r^\lambda,\theta) := \mathbb{E}_{x,i}^{\pi,\theta}\left[\int_0^\infty \mathrm{e}^{-\alpha t} r^\lambda(x(t),\psi(t),\pi,\theta)\mathrm{d}t\right].$$

So, the discounted unconstrained problem is defined as follows.

Definition 6 (The adaptive θ-control problem with Markovian switching)**.** *A policy $\pi^* \in \Pi$ is said to be α-discount optimal for the λ-DUP given that θ is the true parameter value, if*

$$V^*(x,i,r^\lambda,\theta) := V(x,i,\pi^*,r^\lambda,\theta) = \sup_{\pi\in\Pi} V(x,i,\pi,r^\lambda,\theta) \tag{9}$$

for all $(x,i) \in \mathbb{R}^n \times E$. The function V^* will be called the value function of the adaptive θ-control problem with Markovian switching.

Let $v : \mathbb{R}^n \times E \times U \times \Theta \to \mathbb{R}$ be a measurable function satisfying the conditions given in Assumption 3. The following result (obtained from [33]) shows that the function $V^*(\cdot,\cdot,v,\theta)$ is the unique solution of (10), and also proves the existence of stationary optimal policies.

Proposition 3. *Suppose that Assumptions 1–3 hold. Then we have the following:*

(i) The α-optimal discount reward $V^*(\cdot,\cdot,v,\theta)$ belongs to $\mathcal{W}^{2,p}(\mathbb{R}^n \times E) \cap \mathcal{B}_w(\mathbb{R}^n \times E)$ and it verifies the discounted reward HJB equation. That is, for all $(x,i) \in \mathbb{R}^n \times E$ and $\theta \in \Theta$,

$$\alpha V^*(x,i,v,\theta) \quad = \quad \sup_{u\in U}\{r(x,i,u,\theta) + \mathbb{L}^{u,\theta}V^*(x,i,v,\theta)\}. \tag{10}$$

Conversely, if a function $\varphi_\theta \in \mathcal{W}^{2,p}(\mathbb{R}^n \times E) \cap \mathcal{B}_w(\mathbb{R}^n \times E)$ satisfies (10), then $\varphi_\theta(x,i) = V^*(x,i,v,\theta)$ for all $(x,i) \in \mathbb{R}^n \times E$.

(ii) There exists a stationary policy $f_\theta^* \in \mathbb{F}$ that maximizes the right-hand side of (10). That is,

$$\alpha V^*(x,i,v,\theta) = r(x,i,f_\theta^*,\theta) + \mathbb{L}^{f_\theta^*,\theta}V^*(x,i,v,\theta) \text{ for all } (x,i) \in \mathbb{R}^n \times E,$$

and f_θ^* is α-discount optimal given that θ is the true parameter value.

Remark 4. *(a) Notice that $V(x,i,\pi,r^\lambda,\theta) = V(x,i,\pi,r,\theta) + \lambda[V(x,i,\pi,c,\theta) - \overline{\eta}(x,i,\pi,\theta)]$, and by Definition 4, $V(x,i,\pi,c,\theta) - \overline{\eta}(x,i,\pi,\theta) = V(x,i,\pi,c-\alpha\eta,\theta)$,*

(b) Remark 3 and Proposition 1 yield that $\sup_{\pi\in\Pi}|V(x,i,\pi,r^\lambda,\theta)| \le M_\alpha^\lambda w(x,i)$, with $M_\alpha^\lambda := N^\lambda \frac{\alpha+d}{\alpha\beta}$ and N^λ is a bound of $\|r^\lambda\|_w$, implying in turn that $V(\cdot,\cdot,\pi,r^\lambda,\theta) \in \mathcal{B}_w(\mathbb{R}^n \times E)$.

(c) If Assumptions 1, 2 and 3 hold, then by Proposition 3.4 in [28], the mappings $\pi \to V(x,i,\pi,v,\theta)$, $\pi \to V(x,i,\pi,c-\alpha\eta,\theta)$ and $\pi \to V(x,i,\pi,r^\lambda,\theta)$ are continuous on Π for each $(x,i) \in \mathbb{R}^n \times E$ and $\theta \in \Theta$.

3.3. Convergence of Value Functions and Estimation Methods

Finally, in this part, we will present one of the main results of this work, which combines optimality and the statistical approximation scheme (in a discrete way) of our unknown parameter. To do this, we define the concept of consistent estimator and the approximation technique that will be used for it.

Definition 7. *A sequence $(\theta_m)_{m\in\mathbb{N}}$ of measurable functions $\theta_m : \Omega \to \Theta$ is said to be a sequence of uniformly strongly consistent (USC) estimators of $\theta \in \Theta$ if, as $m \to \infty$,*

$$\theta_m(\omega) \to \theta \ \ \mathbb{P}^{\pi,\theta} - a.s. \text{ for all } \pi \in \Pi.$$

For ease of notation, we write $\theta_m := \theta_m(\omega) \in \Theta$. Let $v : \mathbb{R}^n \times U \times \Theta \to \mathbb{R}$ be a measurable function satisfying similar conditions as those given in Assumption 3. The following observations and estimation procedure are an adaptation to what was done in the first part and show us that our set of hypotheses and procedures are consistent.

Remark 5. *(a) Let $(\theta_m)_{m\in\mathbb{N}}$ be a sequence of USC estimators of $\theta \in \Theta$ and let $v : \mathbb{R}^n \times E \times \Pi \times \Theta \to \mathbb{R}$ be a function that satisfies the Assumptions 1–3. Theorem 4.5 in [29], guarantees that every sequence $(V(x,i,\pi,v,\theta_m))_{m\in\mathbb{N}}$ converges to $V(x,i,\pi,v,\theta)$, $\mathbb{P}^{\pi,\theta}$ almost surely.*

(b) *Let* $(\pi_m)_{m\in\mathbb{N}}$ *be a sequence in* Π. *Since* Π *is a compact set, there exists a subsequence* $(\pi_{m_k})_{k\in\mathbb{N}} \subset (\pi_m)_{m\in\mathbb{N}}$ *such that* $\pi_{m_k} \xrightarrow{W} \pi \in \Pi$, *and thus, combining Remark 4 (a) and Remark 4 (c), and applying a suitable triangular inequality, it is possible to deduce that for every measurable function* v *satisfying Assumption 3,*

$$V(x,i,\pi_{m_k},v,\theta_{m_k}) \to V(x,i,\pi,v,\theta) \ \ \mathbb{P}^{\pi,\theta}\text{-a.s. as } k\to\infty.$$

(c) *By Proposition 3, and taking into account that* r^λ *in (8), the function* $V^*(\cdot,\cdot,r^\lambda,\theta)$ *verifies (10). In addition, the second part of Proposition 3 ensures the existence of stationary policy* $f_\theta^\lambda \in \mathbb{F}_1$.

(e) *For each* $\lambda \le 0$, $\theta\in\Theta$ *and* $\alpha > 0$, *we define the set*

$$\Pi^{\lambda,\theta} := \Big\{\pi\in\Pi : \alpha V^*(x,i,r^\lambda,\theta) = r^\lambda(x,i,\pi,\theta) + \mathbb{L}^{\pi,\theta}V^*(x,i,r^\lambda,\theta)\ \forall (x,i)\in\mathbb{R}^n\times E\Big\}. \tag{11}$$

Since $\mathbb{F}_1$ *can be seen as an embedding of* Π, *Proposition 3 (ii) guarantees that* $\Pi^{\lambda,\theta}$ *is a nonempty set.*

(f) *As in [4], the set of hypotheses considered in this paper and Lemma 3.15 in [28] ensures that for each* $\theta\in\Theta$ *fixed and any sequence* $(\lambda_m)_{m\in\mathbb{N}}$, *converging to* λ *(with* $\lambda,\lambda_m \le 0$*); if there exists a sequence of policies* $\left(\pi^{\lambda_m,\theta}\right)_{m\in\mathbb{N}} \in \Pi^{\lambda_m,\theta}$ *such that* $\pi^{\lambda_m,\theta} \xrightarrow{W} \pi$, *then* $\pi\in\Pi^{\lambda,\theta}$.

(g) *Lemma 3.16 in [28] ensures that the mapping* $\lambda \mapsto V^*(x,i,r^\lambda,\theta)$ *is differentiable on* $(-\infty,0)$. *In fact, for each* $\lambda<0$ *and* $\theta\in\Theta$

$$\frac{\partial V^*(x,i,r^\lambda,\theta)}{\partial\lambda} = V(x,i,\pi^\lambda,c,\theta) - \overline{\eta}(x,i,\pi^\lambda,\theta). \tag{12}$$

The unknown parameter θ will be estimated as Pedersen [39] describes. That is, the functions $h_m : \Omega\times\Theta\to\mathbb{R}$, for $m=1,\dots$ will measure how likely the different values of θ are. If for each $\omega\in\Omega$ fixed, the function $h_m(\omega,\theta)$ has a unique maximum point $\theta_m(\omega)\in\Theta$, then θ is estimated by $\theta_m(\omega)$.

Under the assumption that, for $m\in\mathbb{N}$ and $\theta\in\Theta$, $h_m(\cdot,\theta)$ is a measurable function of ω and that it is also twice continuously differentiable in θ for all $\mathbb{P}^{\pi,\theta}-$almost all $\omega\in\Omega$, it is proven that the function $\theta\to h_m(\omega,\theta)$ is continuous and has a unique maximum point $\theta_m(\omega)$ for each $\omega\in\Omega$ fixed. The number $m\in\mathbb{N}$ is the index of a sequence of random experiments on the measurable space $(\Omega,\mathcal{F})$. This method is known as the approximate maximum likelihood estimator.

In our scenery, given a partition of times $\{0=t_0<t_1<t_m:=T\}$ from $[0,T]$, the outcomes of the random experiments will be represented by a sequence $X_T := (x_{t_i} : i=0,\dots,m)$ of a trajectory $x^{u,\theta}(t)$ up to time T on $(\Omega,\mathcal{F}) := (\mathcal{C}([0,\infty)),\mathcal{B}(\mathcal{C}([0,\infty)))$ and the function h_m will be called the least square function (LSE), i.e., $h_m(w,\theta) := LSE(w,\theta)$.

It is evident that $x^{u,\theta}(t)$ in (1) is observed up to a finite time, say T, for which we define

$$LSE(X_T,\theta) := \sum_{i=1}^{m}\left(x_{t_i} - x_{t_{i-1}} - b(x_{t_{i-1}},\psi(t_{i-1}),u_{t_{i-1}},\theta)(t_i-t_{i-1})\right)^2, \tag{13}$$

with the drift function b as in (1). The above function generates the least square estimator until time T with m observations:

$$\theta_{LSE}\equiv\theta_{LSE}(X_T) := \arg\min_{\theta\in\Theta} LSE(X_T,\theta). \tag{14}$$

Remark 6. *The fact that* $x^{u,\theta}(t)$ *in (1) can only be observed in a finite horizon is one of the hypotheses of the so-called model predictive control. However, at least from a theoretical point of view, our version of the PEC makes no such assumption, but still chooses* T *as large as practically possible and thus defines (13) and (14). In this sense, there is a connection between these two perspectives.*

In [12,16,39], the consistency and asymptotic normality of θ_{LSE} are studied. In particular, Shoji (see [16]) shows that the optimization based on the LSE function is equivalent to

the optimization based on the discrete approximate likelihood ratio function in the one-dimensional stochastic differential equation case and with a constant diffusion coefficient considered:

$$MLR(X_T,\theta) := \sum_{i=1}^{m} b(y_{t_{i-1}},u_{t_{i-1}},\theta)\left[\sigma(y_{t_{i-1}})\sigma(y_{t_{i-1}})^\top\right]^{-1}(x_{t_i}-x_{t_{i-1}})$$
$$-\frac{1}{2}\sum_{i=1}^{m}\left\{b(y_{t_{i-1}},u_{t_{i-1}},\theta)^\top\left[\sigma(y_{t_{i-1}})\sigma(y_{t_{i-1}})^T\right]^{-1}\cdot b(y_{t_{i-1}},u_{t_{i-1}},\theta)(t_i-t_{i-1})\right\},$$

with $y_{t_{i-1}} := (x_{t_{i-1}}, \psi(t_{i-1}))$, *and* b and σ as in (1). The MLR function generates the discrete approximate likelihood ratio estimator:

$$\theta_{LR} \equiv \theta_{LR}(X_T) := \arg\max_{\theta\in\Theta} MLR(X_T,\theta).$$

Now, we will establish our main result.

Theorem 1. *Let $(\theta_m)_{m\in\mathbb{N}}$ be a sequence of USC estimators of $\theta\in\Theta$. For each m, let π_m be a α-discount optimal policy. Then there exists a subsequence $(m_k)_k$ of $(m)_m$ and a policy π^* such that $\pi_{m_k} \xrightarrow{W} \pi^*$. Moreover, if Assumptions 1–3 hold, as $k\to\infty$,*

$$V^*(x,i,\theta_{m_k}) \to V^*(x,i,\theta) \;\; \mathbb{P}^{\pi^*,\theta}\text{-a.s. for each } x\in\mathbb{R}^n \text{ and } i\in E,$$

and π^* is α-discount optimal for the θ-control problem $\mathbb{P}^{\pi^*,\theta}$ almost surely.

Proof. Consider a sequence of USC estimators $(\theta_m)_{m\in\mathbb{N}}$ such that $\theta_m\to\theta$ as $m\to\infty$. Let $R>0$, and take the open ball $B_R\times E := \{(x,i)\in\mathbb{R}^n\times E \mid |x|<R,\ i\in E\}$. For $(x,i)\in B_R\times E$, let $(\pi_m)_{m\in\mathbb{N}}\subset\Pi$ be a sequence of α-discounted optimal policies. Since Π is a compact set, there exists a subsequence $(\pi_{m_k})_{k\in\mathbb{N}}\subset(\pi_m)_{m\in\mathbb{N}}$ such that π_{m_k} converges to $\pi^*\in\Pi$ in the topology of relaxed controls given in Definition 2.

Let us first fix an arbitrary $m_k\in\mathbb{N}$. Then, Theorem 6.1 in [33] ensures that the value function $V^*(x,i,\theta_{m_k})$ in (9) is the unique solution of the HJB Equation (10), i.e., it satisfies

$$\alpha V^*(x,i,\theta_{m_k}) \;=\; r(x,i,\pi_{m_k},\theta_{m_k}) + \mathbb{L}^{\pi_{m_k},\theta_{m_k}}V^*(x,i,\theta_{m_k}), \tag{15}$$

and by Theorem 9.11 in [40], there exists a constant C_0 (depending on R) such that, for fixed θ_{m_k} and $p>n$, we have

$$\begin{aligned}|V^*(x,i,\theta_{m_k})|_{\mathcal{W}^{2,p}(B_{2R}\times E)} &\le C_0\left(\left\|V^*(x,i,\theta_{m_k})\right\|_{\mathcal{L}^p(B_{2R}\times E)} + \left\|r(x,i,\pi_{m_k},\theta_{m_k})\right\|_{\mathcal{L}^p(B_{2R}\times E)}\right.\\ &\le (M+M(\alpha))|B_{2R}|^{1/p}\max_{(x,i)\in B_{2R}\times E} w(x,i) < \infty,\end{aligned} \tag{16}$$

where $|\overline{B}_{2R}|$ represents the volume of the closed ball with radius $2R$, and M and $M(\alpha)$ are the constants in Assumption 3 (b) and Proposition 1, respectively.

Now, observe that conditions (a)–(e) of Theorem A1 hold. In fact, for each π_{m_k}, (15) can be written in terms of the operator (A2) as $\mathbb{L}^{\pi_{m_k},\theta}V^*(x,i,\theta_{m_k})=0$ with the functions v_2, λ, ρ equal to zero, $v_1\equiv r$ and $h_{m_k}(x,i)\equiv V^*(x,i,\theta_{m_k})$. So, taking $\xi_{m_k}\equiv 0$, $\lambda\equiv 0$, conditions (a),(c) and (d) hold. In addition, by (16), condition (b) is verified as well.

Then, by Theorem A1, we claim the existence of a function $h(\cdot,\cdot,\theta)\in\mathcal{W}^{2,p}(\mathbb{R}^n\times E)$, together with a subsequence $(m_k : k=1,\dots)$ such that $V^*(\cdot,\cdot,\theta_{m_k})\to h(\cdot,\cdot,\theta)$ uniformly in $B_R\times E$, and pointwise on $\mathbb{R}^n\times E$ as $k\to\infty$ and $\pi_{m_k}\xrightarrow{W}\pi^*$. Furthermore, $h(\cdot,\cdot,\theta)$ satisfies

$$\alpha h(x,i,\theta) = r(x,i,\pi^*) + \mathbb{L}^{\pi^*,\theta}h(x,i,\theta) \;\; \mathbb{P}^{\pi^*,\theta}\text{-a.s.}, \tag{17}$$

with $h(\cdot,\cdot,\theta) \in \mathcal{W}^{2,p}(B_R \times E)$. Since the radius $R > 0$ is arbitrary, we can extend our analysis to all of $(x,i) \in \mathbb{R}^n \times E$.

Thus, as $V^*(x,i,\theta)$ is the unique solution of the HJB equation (17), we can deduce that $h(x,i,\theta)$ coincides with $V^*(x,i,\theta)$. So, by (15) and (17), as $k \to \infty$,

$$V^*(x,i,\theta_{m_k}) \to V^*(x,i,\theta) \ \ \mathbb{P}^{\pi^*,\theta}\text{-a.s., for each } x \in \mathbb{R}^n \text{ and } i \in E.$$

On the other hand, by Proposition 3, for each $i \in E$ and $\theta_{m_k} \in \Theta$ fixed, we have

$$\alpha V^*(x,i,\theta_{m_k}) \quad \geq \quad r(x,i,\pi,\theta_{m_k}) + \mathbb{L}^{\pi,\theta_{m_k}} V^*(x,i,\theta_{m_k}) \ \text{ for all } \pi \in \Pi. \tag{18}$$

Hence, letting $k \to \infty$ and using Theorem A1 from appendix again, we obtain that (18) converges to

$$\alpha V^*(x,i,\theta) \quad \geq \quad r(x,i,\pi,\theta) + \mathbb{L}^{\pi,\theta} V^*(x,i,\theta) \text{ for all } \pi \in \Pi. \tag{19}$$

Thus, by (17) and (19), we obtain

$$\alpha V^*(x,i,\theta) \quad = \quad \sup_{\pi \in \Pi} \{ r(x,i,\pi,\theta) + \mathbb{L}^{\pi,\theta} V^*(x,i,\theta) \}.$$

implying that π^* is α-optimal for the θ-control problem with Markovian switching. □

In the following section, we present a numerical example to illustrate our results. To this end, we implement Algorithm 1. In it, first we introduce the number of iterations in our process and define the variables we need to simulate the dynamic system $x(t)$ and the Markov chain $\psi(t)$. Such simulations are inspired by the algorithm proposed in [41], and allow us to obtain the discrete observations $\{x_k : \ k = 1,2,\dots\}$ needed to feed (13) and (14) and thus approximate the real value of θ.

Algorithm 1: Method of LSE to find θ

Data: Number of iterations to be performed m, for (1) and (2): stepsize dt, arbitrary controllers $u \in U$, drift and diffusion coefficients $b(\cdot,\cdot,\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$; and Q, the generator of the continuous-time Markov chain $\psi(t)$.
Result: Estimation of θ
Simulate $(\tau_i : i = 1,\dots) \subset [0,T]$;
foreach $\tau \in (\tau_i : i = 1,\dots)$ **do**
  Use Q to simulate $\psi(\tau)$;
  Use Euler-Maruyama's method to simulate (1)
end
for $k = 1,\dots,m$ **do**
  $\theta_{LSE} \leftarrow \arg\min_{\theta \in \Theta} LSE(X_t,\theta)$ given by (13)–(14)
end
return θ_{LSE}

Remark 7. *Now we list some limitations of our approach.*

1. *Approximation of the derivative. In our case, we use central differences, but in each application, the approximation type to be used must be analyzed.*
2. *Least squares approximation. The most common restrictions are the amount of data, the regularity of the samples, and the size of the subintervals.*
3. *Euler–Maruyama method. The most common restrictions in this method occur if the differential equation presents stiffness, inappropriate step size, or sudden growth. In our application, the Euler–Maruyama method converges with strong order 1/2 to the true solution. See Theorem 10.2.2 in [42].*

4. Numerical Example

This application complements the one we used in [38]. We represent the stock of pollution as the controlled diffusion process with Markovian switchings of the form

$$\mathrm{d}x(t) = [u(t, \psi(t)) - \theta x(t)]\mathrm{d}t + \sigma \mathrm{d}W(t), \quad x(0) = x > 0, \psi(0) = i, \tag{20}$$

where $\psi(t)$ is a Markov chain with generator

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 \\ \lambda_1 & -\lambda_1 \end{pmatrix}.$$

$\psi(t)$ stands for the perception of society toward the current level of pollution at each time. It takes values from the set $E := \{1, 2\}$. So, if the Markov chain is initially in state $\psi(0) = 1$, then before its first jump from state 1 to state 2 at its first random jump time τ_1, the stock of pollution obeys the following SDE

$$\mathrm{d}x(t) = [u(t, \psi(1)) - \theta x(t)]\mathrm{d}t + \sigma \mathrm{d}W(t), \tag{21}$$

with initial state $x(0) = 0$. At time τ_1, the Markov chain jumps to 2, where it will stay until the next jump, at time τ_2. During the period $[\tau_1, \tau_2]$, the stock of pollution is driven by the SDE

$$\mathrm{d}x(t) = [u(t, \psi(2)) - \theta x(t)]\mathrm{d}t + \sigma \mathrm{d}W(t), \tag{22}$$

with initial value $x(\tau_1)$ at time τ_1, and the stock of pollution switches to (22) from (21). The stock of pollution will continue to alternate between these two states ad infinitum.

We also consider the pollution flow to be constrained. This means that our controller variable $u(t)$ will be taking values in

$$\begin{array}{ll} [0, \eta] & \text{if } \psi(t) = 1, \text{ or in} \\ [\eta, \gamma] & \text{if } \psi(t) = 2. \end{array}$$

for a constant $0 \leq \eta \leq \gamma$. So, $u(t) := u(t, \psi(t))$. We introduce the reward rate function $r : [0, \infty) \times E \times U \to \mathbb{R}$, that represents the social welfare defined by

$$r(x, i, u) := \sqrt{u} - a(i)x, \text{ for all } (x, i, u) \in [0, \infty) \times E \times U,$$

whereas the cost and constraint rates are

$$\begin{array}{rcl} c(x, i, u) & = & c_1(i)x + c_2(i)u \text{ for all } (x, i, u) \in [0, \infty) \times E \times U, \\ \eta(x, i, \theta) & := & \dfrac{c_1(i)x}{\alpha + \theta} + q, \end{array}$$

where q is a positive constant. Clearly, (20) satisfies Assumption 1. The infinitesimal generator for a function $v \in \mathcal{C}^2(\mathbb{R} \times E)$ is

$$\mathbb{L}^{u,\theta} v(x, i) = [u - \theta x]\frac{\partial v(x, i)}{\partial x} + \frac{1}{2}\sigma^2 \frac{\partial^2 v(x, i)}{\partial x^2} + \sum_{j=0}^{1} q_{ij} v(x, j), \text{ for } x > 0 \text{ and } i \in E.$$

We use $w(x, i) := x^2 i + 1$. It is easy to verify that $\mathbb{L}^{u,\theta} w(x, i) \leq -b_1 w(x, i) + g(x, i, u, \theta)$, with $0 < b_1 < 2\theta - q_{1i}$, where $g(x, i, u, \theta) := b_1 w(x, i) + (2ux - 2\theta x^2 + \sigma^2)i + q_{i1}x^2$.

Take b_1 such that $b_1 - 2\theta + q_{i1} < 0$, and note that for every $(x, i) \in \mathbb{R} \times E$, $(u, \theta) \to g(x, i, u, \theta)$ is continuous on the compact sets U and Θ; therefore, there exists a constant d_1 such that $g(x, i, u, \theta) \leq d_1$ for all $(x, i) \in \mathbb{R} \times E$, $u \in U$ and $\theta \in \Theta$. So, Assumption 2 is satisfied.

In this problem, the payoff rate is $r^{\lambda}(x,i,u) := r(x,i,u) - \lambda c(x,i,u)$ where λ is the Lagrange multiplier, the $\alpha-$discounted expected payoff is

$$V(x,i,\pi,r_1,\theta) := \mathbb{E}_{x,i}^{\pi,\theta}\left[\int_0^{\infty} \mathrm{e}^{-\alpha t} r^{\lambda}(x(t),\psi(t),\pi)\mathrm{d}t\right],$$

and the value function is

$$V^*(x,i,\theta) = \sup_{\pi\in\Pi} V(x,i,\pi,r_1,\theta). \tag{23}$$

In order to find the optimal control and the value function $V^*(x,i,\theta)$ given in (23), we need to solve (10) for each $i \in E$. The HJB equations associated with this example are

$$\begin{aligned}\alpha\varphi(x,1) = \sup_{0\le u\le\eta}\Big\{&\sqrt{u} - a(0)x + \lambda\left[c_1(1)x + c_2(1)u - \alpha\left(\frac{c_1(1)x}{\alpha+\theta} + q\right)\right]\\ &+ (u-\theta x)\frac{\partial\varphi(x,1)}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2\varphi(x,1)}{\partial x^2} + \sum_{j=1}^{2} q_{1j}\varphi(x,j)\Big\} \text{ for all } x>0.\end{aligned} \tag{24}$$

$$\begin{aligned}\alpha\varphi(x,2) = \sup_{0\le u\le\gamma}\Big\{&\sqrt{u} - a(2)x + \lambda\left[c_1(2)x + c_2(2)u - \alpha\left(\frac{c_1(2)x}{\alpha+\theta} + q\right)\right]\\ &+ (u-\theta x)\frac{\partial\varphi(x,2)}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2\varphi(x,2)}{\partial x^2} + \sum_{j=1}^{2} q_{2j}\varphi(x,j)\Big\} \text{ for all } x>0.\end{aligned} \tag{25}$$

Assuming that a solution to (24) and (25) has the form $\varphi(x,i) = k_1(i)x + k_2(i)$ with $k_1, k_2 : E \to \mathbb{R}$ measurable functions, we get $\frac{\partial\varphi(x,i)}{\partial x} = k_1(i)$ and $\frac{\partial^2\varphi(x,i)}{\partial x^2} = 0$. Replacing the derivatives of $\varphi(x,i)$ into (24) and (25), we obtain

$$\begin{aligned} k_1(i) &= \frac{\lambda\theta c_1(i) - (\alpha+\theta)a(i)}{(\alpha+\theta)^2} + \frac{\sum_{j=1}^{2} q_{ij}k_1(j)}{\alpha+\theta},\\ \alpha k_2(1) &= \sup_{0\le u\le\eta}\left(\sqrt{u} - a_{\lambda,\theta,1}u\right) - \lambda\alpha q + \sum_{j=1}^{2} q_{1j}k_2(j), \qquad (26)\\ \alpha k_2(2) &= \sup_{\eta\le u\le\gamma}\left(\sqrt{u} - a_{\lambda,\theta,1}u\right) - \lambda\alpha q + \sum_{j=1}^{2} q_{2j}k_2(j), \qquad (27)\end{aligned}$$

with

$$a_{\lambda,\theta,i} := \frac{(\alpha+\theta)a(i) - \lambda[\theta c_1(i) + (\alpha+\theta)^2 c_2(i)]}{(\alpha+\theta)^2} + \frac{\sum_{j=1}^{2} q_{ij}k_1(j)}{\alpha+\theta} > 0,$$

Notice that the suprema in (26) and (27) are attained at

$$f_{\theta}^{\lambda}(1) = \begin{cases} \frac{1}{4(a_{\lambda,\theta,0})^2} & \text{if } \frac{1}{2\sqrt{\eta}} < a_{\lambda,\theta,0},\\ \eta & \text{if } \frac{1}{2\sqrt{\eta}} \ge a_{\lambda,\theta,0},\end{cases} \tag{28}$$

$$f_{\theta}^{\lambda}(2) = \begin{cases} \frac{1}{4(a_{\lambda,\theta,1})^2} & \text{if } \frac{1}{2\sqrt{\gamma}} < a_{\lambda,\theta,1},\\ \gamma & \text{if } \frac{1}{2\sqrt{\gamma}} \ge a_{\lambda,\theta,1}.\end{cases} \tag{29}$$

Thus, $k_2(\cdot)$ can be written as

$$k_2(i) = \frac{\sqrt{f_\theta^\lambda(i)} - a_{\lambda,\theta,i} f_\theta^\lambda(i)}{\alpha} - \lambda q + \frac{1}{\alpha} \sum_{j=0}^{1} q_{ij} k_2(j).$$

By Proposition 3, the optimal control is (28) and (29), and the value function is $\varphi(x,i)$, i.e.,

$$\begin{aligned} \varphi(x,i) = V^*(x,i,\theta,r) = &\left[\frac{\lambda\theta c_1(i) - (\alpha+\theta)a(i)}{(\alpha+\theta)^2} + \frac{\sum_{j=0}^{1} q_{ij} k_1(j)}{\alpha+\theta} \right] x \\ &+ \frac{\sqrt{f_\theta^\lambda(i)} - a_{\lambda,\theta,i} f_\theta^\lambda(i)}{\alpha} - \lambda q + \frac{1}{\alpha} \sum_{j=0}^{1} q_{ij} k_2(j). \end{aligned} \tag{30}$$

For the numerical experiment, we consider the particular form of (1) given by (20) to test Algorithm 1 with $Q = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$ as the generator of the continuous-time Markov chain $\psi(t)$ embedded within (20). Also, let $x(0) = 0$, $T = 5$, $\mathrm{d}t = 10^{-4}$, $u(t,1) = 0.5$, $u(t,2) = 1.5$, $\sigma = \sqrt{10^{-10}}$, and $\theta = 2.5$ as the true parameter value of the pollution decay rate. These last data allow us to simulate (20) in the interval $[0,5]$, and, for the sake of comparison, it will be considered as the real model (see Figure 1). Based on this information, $m = 50,000$ discrete observations were obtained. Now, we suppose that θ is the unknown parameter and we estimate it by means of the least square function LSE in (13) and (14). Substituting $b(x,i,u,\theta) = u(t,i) - \theta x(t)$ in (13), we obtain the following estimator for each state $i \in E = \{1,2\}$.

$$\theta(i)_{LSE_m} = \frac{\sum_{k=2}^{m-1} u x_k - x_k \mathrm{d}x_k}{\sum_{i=2}^{m-1} x_k^2}, \tag{31}$$

where $\mathrm{d}x_k := \frac{1}{2} \frac{x_{k+1} - x_{k-1}}{t_{k+1} - t_k}$. Given that the dynamic system for $x(t)$ is governed by a stochastic differential equation with Markovian switching, it is not possible to have a single value for θ, but rather a set of values (the number of these values strictly depends on the number of jumps that occur in the interval $[0,T]$), which we will denote as in (31). These approximations allow us to simulate the stochastic differential equation with Markovian switching again with the same jumps. The outputs of the approximate stochastic differential equation with Markovian switching $x^{\theta_m^j}(t)$ and the one with the real value for θ, $x^\theta(t)$ are displayed in Table 1.

To graph the value function $V_\theta^*(x,i) := V^*(x,i,\theta,r)$ given in (30) we take $\eta = 3$, $\gamma = 3$, $a(1) = 1.25$, $a(2) = 2$, $c_1(1) = 100$, $c_1(2) = 150$, $c_2(1) = 10$, $c_1(2) = 1.5$, $\alpha = 0.2$, and $q = 60$, and $\theta(i)_{LSE_m}$ with $m = 10000, 12500, 16667, 25000$ and 5000, see Figure 2.

The symbol $\left(x(t)^{\theta_{LSE_m}}, f_{\theta_{LSE_m}}^\lambda(i), V_{\theta_{LSE_m}}^*(x,i)\right)$ denotes the value estimates of the dynamic system $x(t)$, the optimal control f_θ^λ and of the value function V_θ^* when we take θ_{LSE_m} instead of θ.

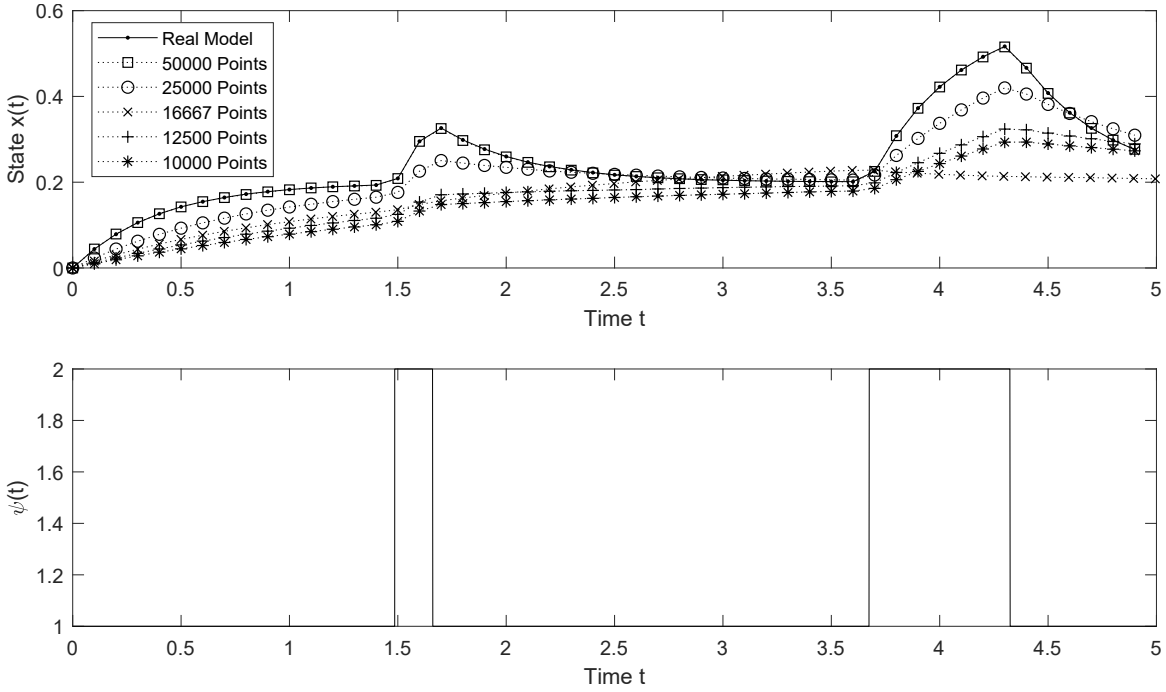

Figure 1. Asymptotic behavior of $x(t)^{\theta_{LSE_m}}$ and $\psi(t)$.

We obtained $m = 50,000$ discrete observations of (20) on $[0,5]$. Given that the Markov chain is known, the vector of jump times $(\tau_1, \tau_2, \tau_3, \tau_4, \tau_5)$ is known as well. The estimator used in each interval $[\tau_k, \tau_{k+1}]$ with $k = 0,1,2,3,4$ and $\tau_0 = 0$ is θ_{LSE_m} given in (31). Figures 1 and 2, together with Table 1 show that, as m increases, the estimator approaches the true parameter value $\theta = 2.5$, and the RMSE between the estimations $\left(x(t)^{\theta_{LSE_m}}, f^{\lambda}_{\theta_{LSE_m}}(i), V^{*}_{\theta_{LSE_m}}(x,i)\right)$ and the actual values of $\left(x(t)^{\theta}, f^{\lambda}_{\theta}(i), V^{*}_{\theta}(x,i)\right)$ decreases, thus implying a good fit.

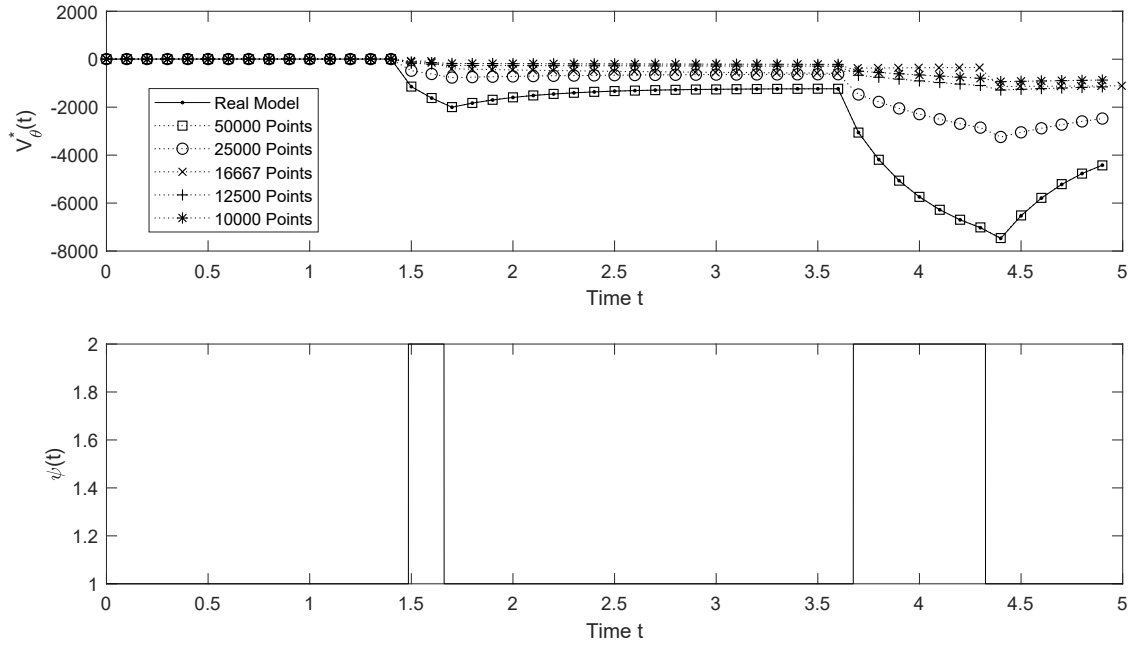

Figure 2. Asymptotic behavior of the optimal reward $V^{*}_{\theta_{LSE_m}}(x(t)^{\theta_{LSE_m}}, i)$ (vertical axis) using the estimator θ_{LSE_m} with m = 10,000, 12,500, 16,667, 25,000 and 5000.

Table 1. Estimated processes using θ_{LSE_m} and the real processes ($\theta = 2.5$).

m	RMSE $(\theta - \theta_{LSE_m})$	RMSE $(x^{\theta} - x^{\theta_{LSE_m}})$	RMSE $(f_{\theta}^{\lambda}(i) - f_{\theta_{LSE_m}}^{\lambda}(i))$	RMSE $(V_{\theta}^{*}(x^{\theta}, i) - V_{\theta_{LSE_m}}^{*}(^{\theta_{LSE_m}}, i))$
50,000	0.00496364	0.00059318	5.72448×10^{-5}	5.09465
25,000	0.00497008	0.0423565	5.72228×10^{-5}	1707.61
16,667	3.23512	0.112074	0.00174892	2648.67
12,500	0.0049845	0.0873181	5.72787×10^{-5}	2518.47
10,000	0.00498072	0.10305	5.72684×10^{-5}	2669.93

Theorem 5.5 in [28] ensures that for a fixed point $z > 0$ such that

$$q < \frac{\eta c_1(i) z}{(\alpha+\theta)^2} + \frac{[\theta c_1(i) + (\alpha+\theta)^2 c_2(i)]\gamma}{\alpha(\alpha+\theta)^2} \text{ for all } i = 1, 2$$

if the inequality $\frac{1}{2\sqrt{\left(\frac{\alpha(\alpha+\theta)^2 q - \alpha\theta c_1(i) z}{\theta c_1(i) + (\alpha+\theta)^2 c_2(i)}\right)}} > \frac{a(i)}{\alpha+\theta}$ holds, then the mapping $\lambda \longmapsto V^*(z, i, \theta, r^{\lambda})$ admits a critical point $\lambda_{z,\theta}^{*} \equiv \lambda_{z,\theta}^{*}(\alpha, z) < 0$ satisfying

$$a_{\lambda_{z,\theta}^{*},\theta}(i) = \frac{(\alpha+\theta)a(i) - \lambda_{z,\theta}^{*}[\theta c_1(i) + (\alpha+\theta)^2 c_2(i)]}{(\alpha+\theta)^2} = \frac{1}{2\sqrt{\left(\frac{\alpha(\alpha+\theta)^2 q - \alpha\theta c_1(i) z}{\theta c_1(i) + (\alpha+\theta)^2 c_2}\right)}}.$$

Therefore, every $\pi^{\lambda_{z,\theta}^{*}} \in \Pi^{\lambda_{z,\theta}^{*}}$ is α-optimal for the DPC and $V(z, i, \pi^{\lambda_{z,i,\theta}^{*}}, c, \theta) = \overline{\eta}(z, i, \pi^{\lambda_{z,\theta}^{*}}, \theta)$; in particular, the α-optimal policy for the DPC is $f_{\theta}^{\lambda_{z,\theta}^{*}} \in \mathbb{F} \cap \Pi^{\lambda_{z,\theta}^{*}}$ of the form

$$f_{\theta}^{\lambda_{z,\theta}^{*}}(i) = \frac{\alpha(\alpha+\theta)^2 q - \alpha\theta c_1(i) z}{\theta c_1(i) + (\alpha+\theta)^2 c_2(i)}, \tag{32}$$

and the α-optimal value for the DPC is given by

$$\begin{aligned} V^*(z, i, \theta, r^{\lambda_{z,\theta}^{*}}) &= V^*(z, \pi^{\lambda_{z,i,\theta}^{*}}, \theta, r) \\ &= -\frac{a(i) z}{\alpha+\theta} + \frac{1}{\alpha}\sqrt{\left(\frac{\alpha(\alpha+\theta)^2 q - \alpha\theta c_1(i) z}{\theta c_{(}i) + (\alpha+\theta)^2 c_2(i)}\right)} - \frac{a(i)}{\alpha+\theta}\left[\frac{(\alpha+\theta)^2 q - \theta c_1(i) z}{\theta c_1(i) + (\alpha+\theta)^2 c_2(i)}\right]. \end{aligned} \tag{33}$$

5. Conclusions

We studied controlled stochastic differential equations with Markovian switching of the form (1), where the drift coefficient depends on an unknown parameter $\theta \in \Theta$.

Two problems were analyzed, each one under a corresponding reward criterion: the discounted unconstrained problems (DUP) and the discounted problem with constraints (DPC) with optimal value functions $V_{\theta}^{*}(x, i, r)$ and $V_{\theta}^{*}(x, i, r^{\lambda})$, respectively. Once a suitable procedure estimation of θ is obtained, it generates a sequence of estimators $(\theta_m)_{m \in \mathbb{N}}$ such that $\theta_m \to \theta$ as $m \to \infty$, and the results obtained guarantee the following:

- For each initial state and parameter θ_m, $V_{\theta_m}^{*} \to V_{\theta}^{*}$ almost surely for both problems.
- For each estimation θ_m and problem (DUP or PDC), there are optimal policies π_{θ_m}.
- There is a subsequence of policies $(\pi_{\theta_{m_k}})_{k \in \mathbb{N}}$ and a policy $\pi_{\theta}^{*} \in \Pi$ such that $\pi_{\theta_{m_k}} \xrightarrow{W} \pi_{\theta}^{*}$, and, moreover, π_{θ}^{*} is optimal for the θ-OCP.

- Similar to the previous point, for the DUP, there is a subsequence of policies and a policy $\pi^*_\theta \in \Pi^{\lambda,\theta}$ such that $\pi^{\lambda_{m_k},\theta_{m_k}} \xrightarrow{W} \pi^*_\theta$, and π^*_θ is optimal for the θ-DUP. Moreover, if $\lambda_{m_k} < 0$ is a critical point of $V^*_{\theta_{m_k}}(x, r^\lambda)$, then π^*_θ is optimal for the θ-DCP.

The numerical part is one of the strengths of this work. Indeed, it aims at solving an estimation problem and a control problem. This task requires knowledge and storage of the optimal policies π_θ for all the values of θ, which may take considerable offline execution time. In addition, we propose and implement an algorithm to approximate θ.

Finally, the idea of modeling the dynamic $x(t)$ as a controlled diffusion process with Markovian switchings allows us to consider extra factors or elements that affect the pollution stock. Such factors could be seen, in particular, as multiple pollution sources. An interesting task or challenge would be to pose this scenario as a multi-objective problem, where both sources of contamination and the stock require to be minimized under certain restrictions. This could be done by adapting and defining a suitable multi-objective linear program (convex program) and guaranteeing the existence of a saddle-point—or Pareto optimal policy—as studied in [43,44]. Another technique, called the multi-objective evolutionary algorithm, combines multi-objective problems with statistical techniques to approximate the Pareto optimal as in [45]. In both cases, it is still necessary to apply extra techniques due to the unknown parameter θ.

Author Contributions: Conceptualization, methodology, and writing/original draft preparation of this research are due to B.A.E.-T., F.A.A.-H. and J.D.L.-B.; software, validation, visualization, and data curation are original of F.A.A.-H.; formal analysis, investigation, writing/review and editing are due to C.G.H.-C.; project administration, funding acquisition are due to J.D.L.-B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universidad Anáhuac México.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to sincerely thank Ekaterina Viktorovna Gromova for her kind invitation to publish this work.

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

Appendix A. Convergence of the HJB-Equation

Let $v_1, v_2 : \mathcal{O} \times U \times \Theta \times E \to \mathbb{R}$ be two functions with the same properties of the rate functions established in Assumption 3. Furthermore, for every $x \in \mathbb{R}^n$, $k \in E$, $u \in U$, $\alpha > 0$, functions λ and ρ in $\mathcal{B}(\mathcal{O})$, and h in $\mathcal{W}^{2,p}(\bar{\mathcal{O}} \times E)$, let

$$\begin{aligned}\hat{\Psi}(x,k,u,\alpha,\lambda,\rho,\theta;h) :=& v_1(x,k,u,\theta) + \lambda(x)[v_2(x,k,u,\theta) - \rho(x)] \\ &+ \sum_{i=1}^{n} b_i(x,k,u,\theta)\partial_i h(x,k) - \alpha h(x,k),\end{aligned} \tag{A1}$$

where b_i is the i-th component of the drift function b in (1). We also define

$$\mathbb{L}^{u,\theta}h(x,k) := \hat{\Psi}(x,k,u,\alpha,\lambda,\rho,\theta;h) + \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}(x,k)\partial^2_{ij}h(x,k),$$

with a as in Assumption 1 (d). For each $\pi \in \Pi$, we denote

$$\hat{\Psi}(x,k,\pi,\alpha,\lambda,\rho,\theta;h) \quad := \quad \int_U \hat{\Psi}(x,k,u,\alpha,\lambda,\rho,\theta;h)\pi(\mathrm{d}u|x), \quad \text{and}$$

$$\mathbb{L}^{\pi,\theta}h(x,k) \quad := \quad \hat{\Psi}(x,k,\pi,\alpha,\lambda,\rho,\theta;h) + \frac{1}{2}\sum_{i,j=1}^{n} a^{ij}(x,k)\partial_{ij}^{2}h(x,k).$$

The framework we consider requires the interchange of limits, which is an extension of the adaptive case of Theorem 6.1 in [28], Theorem A1 in [26], Theorem 3.4 in [46] and Theorem 5.2 in [47].

Theorem A1. *Let $\mathcal{O}$ be a bounded $\mathcal{C}^2$ domain and suppose that Assumptions 1–3 hold. In addition, assume that there exist sequences $(\lambda_m)_{m\in\mathbb{N}}, (\rho_m)_{m\in\mathbb{N}} \subset \mathcal{B}(\mathcal{O})$, $(\pi_m)_{m\in\mathbb{N}}) \subset \Pi$, $\theta_m \in \Theta$ and $(h_m)_{m\in\mathbb{N}} \equiv (h(\cdot,\cdot,\theta_m))_{m\in\mathbb{N}} \subset \mathcal{W}^{2,p}(\mathcal{O}\times E)$, $(\xi_m)_{m\in\mathbb{N}} \subset \mathcal{L}^p(\mathcal{O}\times E)$, with $p > n$ (n is the dimension of (1)), satisfying the following:*

(a) $\mathbb{L}^{\pi_m,\theta_m}h_m = \xi_m$ *in* $\mathcal{O}\times E$ *for* $m = 1,2,\cdots$.

(b) *There exists a constant* $\tilde{M}_1$ *such that* $\|h_m\|_{\mathcal{W}^{2,p}(\mathcal{O}\times E)} \leq \tilde{M}_1$ *for* $m = 1,2,\cdots$.

(c) ξ_m *converges in* $\mathcal{L}^p(\mathcal{O}\times E)$ *to some function* ξ.

(d) θ_m *converges to some* θ, $\mathbb{P}^{\pi,\theta}$*-a.s.*

(e) ρ_m *converges uniformly to some function* ρ.

(f) $\pi_m \xrightarrow{W} \pi$.

Then, there exists a function $h \in \mathcal{W}^{2,p}(\mathcal{O}\times E)$ and a subsequence $(m_k : k = 1,\dots) \subset \{1,2,\dots\}$ such that $h_{m_k} \to h$ in the norm of $\mathcal{C}^{1,\eta}(\mathcal{O}\times E)$ for $\eta < 1 - \frac{n}{p}$ as $k \to \infty$. Moreover,

$$\mathbb{L}^{\pi,\theta}h = \xi \text{ in } \mathcal{O}\times E \quad \mathbb{P}^{\pi,\theta} - a.s. \tag{A2}$$

Proof. It is known that Sobolev's space $\mathcal{W}^{2,p}(\mathcal{O}\times E)$ is reflexive Theorem 3.5 in [48]. Then, by Theorem 1.17 in [48], for every $\overline{M} \geq 0$, the ball

$$H := \left\{h \in \mathcal{W}^{2,p}(\mathcal{O}\times E) : \|h\|_{\mathcal{W}^{2,p}(\mathcal{O}\times E)} \leq \overline{M}\right\} \tag{A3}$$

is weakly sequentially compact. On the other hand, since $p > n$, by Theorem 6.2 (Part III) in [48], for $0 \leq \eta < 1 - \frac{n}{p}$, the embedding $\mathcal{W}^{2,p}(\mathcal{O}\times E) \hookrightarrow \mathcal{C}^{1,\eta}(\mathcal{O}\times E)$ is compact; hence, it is also continuous, and thus the set H in (A3) is relatively compact in $\mathcal{C}^{1,\eta}(\mathcal{O}\times E)$. This fact ensures the existence of a function $h \in \mathcal{W}^{2,p}(\bar{\mathcal{O}}\times E)$ and a subsequence $\left(h_{m_k}\right)_{k\in\mathbb{N}} \equiv (h_m)_{m\in\mathbb{N}} \subset H$ such that

$$h_m \to h \text{ weakly in } \mathcal{W}^{2,p}(\mathcal{O}\times E) \text{ and strongly in } \mathcal{C}^{1,\eta}(\mathcal{O}\times E). \tag{A4}$$

Now, we show that, as $m \to \infty$,

$$\int_{\mathcal{O}} g(x,k)\hat{\Psi}(x,k,\pi_m,\alpha_m,\lambda_m,\rho_m,\theta_m;h_m)\mathrm{d}x \to \int_{\mathcal{O}} g(x,k)\hat{\Psi}(x,k,\pi,\alpha,\lambda,\rho,\theta;h)\mathrm{d}x, \ \mathbb{P}^{\pi,\theta}\text{-a.s.}, \tag{A5}$$

for all $g \in \mathcal{L}^1(\mathcal{O}\times E)$.

To this end, recall (A1) and note that, given $(x,k) \in \mathcal{O}\times E$, functions $h \in \mathcal{W}^{2,p}(\mathcal{O}\times E)$ and $h_m \in H$, λ_m, λ, ρ_m, $\rho \in \mathcal{B}(\mathcal{O})$, a pair of policies $\pi, \pi_m \in \Pi$, and $\theta_m, \theta \in \Theta, \alpha \geq 0$, the following holds.

$$\begin{aligned}
& \int_{\mathcal{O}} g(x,k)\left|\hat{\Psi}(x,k,\pi_m,\alpha,\lambda_m,\rho_m,\theta_m;h_m) - \hat{\Psi}(x,k,\pi,\alpha,\lambda,\rho,\theta;h)\right|\mathrm{d}x \\
\leq & \int_{\mathcal{O}} g(x,k)|v_1(x,k,\pi_m,\theta_m) - v_1(x,k,\pi,\theta_m)|\mathrm{d}x \\
+ & \int_{\mathcal{O}} g(x,k)|v_1(x,k,\pi,\theta_m) - v_1(x,k,\pi,\theta)|\mathrm{d}x \\
+ & \int_{\mathcal{O}} g(x,k)|\lambda_m(x)v_2(x,k,\pi_m,\theta_m) - \lambda_m(x)v_2(x,k,\pi,\theta_m)|\mathrm{d}x \\
+ & \int_{\mathcal{O}} g(x,k)|\lambda_m(x)v_2(x,k,\pi,\theta_m) - \lambda_m(x)v_2(x,k,\pi,\theta)|\mathrm{d}x
\end{aligned}$$

$$
\begin{aligned}
&+ \int_{\mathcal{O}} g(x,k)|\lambda_m(x)v_2(x,k,\pi,\theta) - \lambda(x)v_2(x,k,\pi,\theta)|\mathrm{d}x \\
&+ \sum_{i=1}^{n}\int_{\mathcal{O}} g(x,k)|\partial_i h_m(x,k)[b_i(x,k,\pi_m,\theta_m) - b_i(x,k,\pi,\theta_m]|\mathrm{d}x \\
&+ \sum_{i=1}^{n}\int_{\mathcal{O}} g(x,k)|\partial_i h_m(x,k)[b_i(x,k,\pi,\theta_m) - b_i(x,k,\pi,\theta)]|\mathrm{d}x \\
&+ \sum_{i=1}^{n}\int_{\mathcal{O}} g(x,k)|b_i(x,k,\pi,\theta)[\partial_i h_m(x,k) - \partial_i h(x,k)]|\mathrm{d}x \\
&+ \int_{\mathcal{O}} g(x,k)|\lambda_m(x)[\rho_m(x) - \rho(x)]\mathrm{d}x| \\
&+ \int_{\mathcal{O}} g(x)|\rho(x)[\lambda_m(x) - \lambda(x)]|\mathrm{d}x \\
&+ \alpha\int_{\mathcal{O}} g(x,k)|h_m(x,k) - h(x,k)|\mathrm{d}x
\end{aligned}
$$

Since the embedding $\mathcal{W}^{2,p}(\mathcal{O}\times E) \hookrightarrow \mathcal{C}^{1,\eta}(\mathcal{O}\times E)$ is continuous, hypothesis (b) together with the definition of the norm $\|\cdot\|_{\mathcal{C}^{1,\eta}(\mathcal{O}\times E)}$ lead to

$$
\max\left\{|h_m|, \max_{1\le i\le n}|\partial_i h_m|\right\} \le \|h_m\|_{\mathcal{C}^{1,\eta}(\mathcal{O}\times E)} \le \bar{M}\|h_m\|_{\mathcal{W}^{2,p}(\mathcal{O}\times E)} \le \bar{M}\tilde{M}_1.
$$

On the other hand, Assumptions 1 and 3, yield that

$$
\sup_{\pi\in\Pi}|b(\cdot,\cdot,\pi,\cdot)| + \sup_{\pi\in\Pi}|v_2(\cdot,\cdot,\pi,\cdot)| \le K(\bar{\mathcal{O}}\times E).
$$

Hence,

$$
\begin{aligned}
&\int_{\mathcal{O}} g(x,k)\left|\hat{\Psi}(x,k,\pi_m,\alpha,\lambda_m,\rho_m,\theta_m;h_m) - \hat{\Psi}(x,k,\pi,\alpha,\lambda,\rho,\theta;h)\right|\mathrm{d}x \\
&\le \int_{\mathcal{O}} g(x,k)|v_1(x,k,\pi_m,\theta_m) - v_1(x,k,\pi,\theta_m)|\mathrm{d}x \\
&\quad + \|g\|_{\mathcal{L}^1(\mathcal{O}\times E)}|v_1(x,k,\pi,\theta_m) - v_1(x,k,\pi,\theta)| \\
&\quad + |\lambda_m|\int_{\mathcal{O}} g(x,k)|v_2(x,k,\pi_m,\theta_m) - v_2(x,k,\pi,\theta_m)|\mathrm{d}x \\
&\quad + |\lambda_m|\|g\|_{\mathcal{L}^1(\mathcal{O}\times E)}|v_2(x,k,\pi,\theta_m) - v_2(x,k,\pi,\theta)| \\
&\quad + K(\bar{\mathcal{O}}\times E)\|\lambda_m - \lambda\|_{\mathcal{B}(\mathcal{O})}\|g\|_{\mathcal{L}^1(\mathcal{O}\times E)} \\
&\quad + \bar{M}\tilde{M}_1 n\max_{1\le i\le n}\int_{\mathcal{O}} g(x,k)|b_i(x,k,\pi_m,\theta_m) - b_i(x,k,\pi,\theta_m)|\mathrm{d}x \\
&\quad + \bar{M}\tilde{M}_1 n\max_{1\le i\le n}\max_{k\in E}\|g\|_{\mathcal{L}^1(\mathcal{O}\times E)}|b_i(x,k,\pi,\theta_m) - b_i(x,k,\pi,\theta)| \\
&\quad + \|h_m - h\|_{\mathcal{C}^{1,\eta}(\mathcal{O}\times E)}2nK(\bar{\mathcal{O}}\times E)\|g\|_{\mathcal{L}^1(\mathcal{O}\times E)} \\
&\quad + |\lambda_m|\|\rho_m - \rho\|_{\mathcal{B}(\mathcal{O})}\|g\|_{\mathcal{L}^1(\mathcal{O}\times E)} \\
&\quad + \|\rho\|_{\mathcal{B}(\mathcal{O})}\|\lambda_m - \lambda\|_{\mathcal{B}(\mathcal{O})}\|g\|_{\mathcal{L}^1(\mathcal{O}\times E)}.
\end{aligned} \tag{A6}
$$

Observe that $v_1(\cdot,\cdot\pi,\theta)$, $v_2(\cdot,\cdot,\pi,\theta)$ and $b_i(\cdot,\cdot,\pi,\theta)$ $i = 1,\cdots,n$ are bounded on $\bar{\mathcal{O}}\times E$, so the weak convergence criterion can be applied. In addition to that, Assumptions 1 (a) and 3 (a) implies that these functions are continuous on Θ. Then, hypotheses (d) to (f), together with (A4), lead to the right-hand side of (A6) going to zero as $m\to\infty$ $\mathbb{P}^{\pi,\theta}$ almost surely, thus proving (A5).

The existence of the constant $K(\bar{\mathcal{O}} \times E)$ used for the analysis in (A6) can be also used to claim that $|\sigma(x,k)| \leq K(\bar{\mathcal{O}} \times E)$, then we can assert that for each g in $\mathcal{L}^{\frac{p}{p-1}}(\mathcal{O} \times E)$,

$$\begin{aligned} \frac{1}{2}\left|\int_{\mathcal{O}} g(x,k)\left[\sum_{i,j=1}^{n} a^{ij}(x,k)\partial^2_{ij}h_m(x,k) - \sum_{i,j=1}^{n} a^{ij}(x,k)\partial^2_{ij}h(x,k)\right]\mathrm{d}x\right| \leq \\ \frac{n^2}{2}\left[K(\bar{\mathcal{O}} \times E)\right]^2 \sum_{i,j=1}^{n}\left|\int_{\mathcal{O}} g(x,k)\left[\partial^2_{ij}h_m(x,k) - \partial^2_{ij}h(x,k)\right]\mathrm{d}x\right| \end{aligned} \tag{A7}$$

Thus the weak convergence of $(h_m : m = 1,2,\dots)$ to h in $\mathcal{W}^{2,p}(\mathcal{O} \times E)$ yields that the right-hand side of (A7) converges to zero almost surely as $m \to \infty$. Notice also that the convergence of (A5) is also valid for all $g \in \mathcal{L}^{\frac{p}{p-1}}(\mathcal{O} \times E)$. The reason is because $\mathcal{L}^{\frac{p}{p-1}}(\mathcal{O}) \times E \subset \mathcal{L}^1(\mathcal{O} \times E)$ (recall the Lebesgue measure on $\mathcal{O}$ is bounded). This last fact together with (A7) and hypothesis (c), yield that for every g in $\mathcal{L}^{\frac{p}{p-1}}(\mathcal{O} \times E)$,

$$\int_{\mathcal{O}} g(x,k)\left[\mathbb{L}^{\pi,\theta}h(x,k) - \xi(x,k)\right]\mathrm{d}x = \lim_{n\to\infty}\int g(x,k)\left[\mathbb{L}^{\pi_m,\theta_m}h_m(x,k) - \xi_m(x,k)\right]\mathrm{d}x = 0$$

$\mathbb{P}^{\pi,\theta}$ almost surely. This fact, along with Theorem 2.10 in [49], implies (A2), i.e., $\mathbb{L}^{\pi,\theta}h = \xi$ $\mathbb{P}^{\pi,\theta}$ almost surely in $\mathcal{O} \times E$. This completes the proof. □

References

1. Vierros, M.K. *Promotion and Strengthening of Sustainable Ocean-Based Economies*; United Nations: New York, NY, USA, 2021.
2. Kawaguchi, K. Optimal Control of Pollution Accumulation with Long-Run Average Welfare. *Environ. Resour. Econ.* **2003**, *26*, 457–468. [CrossRef]
3. Morimoto, H. Optimal Pollution Control with Long-Run Average Criteria. In *Stochastic Control and Mathematical Modeling: Applications in Economics*; Encyclopedia of Mathematics and its Applications; Cambridge University Press: Cambridge, UK, 2010; pp. 237–251. [CrossRef]
4. Escobedo-Trujillo, B.A.; López-Barrientos, J.D.; Higuera-Chan, C.G.; Alaffita-Hernández, F.A. Robust statistic estimation in constrained optimal control problems of pollution accumulation (Part I). *Mathematics* **2023**, *11*, 923. [CrossRef]
5. Hilgert, N.; Minjárez-Sosa, A. Adaptive control of stochastic systems with unknown disturbance distribution: Discounted criteria. *Math. Methods Oper. Res.* **2006**, *63*, 443–460. [CrossRef]
6. Hernández-Lerma, O.; Marcus, S. Technical note: Adaptive control of discounted Markov Decision chains. *J. Optim. Theory Appl.* **1985**, *46*, 227–235. [CrossRef]
7. Kurano, M. Discrete-time markovian decision processes with an unknown parameter-average return criterion. *J. Oper. Res. Soc. Jpn.* **1972**, *15*, 67–76.
8. Mandl, P. Estimation and control in Markov chains. *Adv. Appl. Probab.* **1974**, *6*, 40–60. [CrossRef]
9. Borkar, V.; Ghosh, M. Ergodic Control of Multidimensional Diffusions II: Adaptive Control. *Appl. Math. Optim.* **1990**, *21*, 191–220. [CrossRef]
10. Vrabie, D.; Pastravanu, O.; Abu-Khalaf, M.; Lewis, F. Adaptive optimal control for continuous-time linear systems based on policy iteration. *Automatica* **2009**, *45*, 477–484. [CrossRef]
11. Di Masp, G.; Stettner, L. Bayesian ergodic adaptive control of diffusion processes. *Stochastics Stochastics Rep.* **1997**, *60*, 155–183. [CrossRef]
12. Ralchenko, K. Asymptotic normality of discretized maximum likelihood estimator for drift parameter in homogeneous diffusion model. *Mod. Stochastics Theory Appl.* **2015**, *2*, 17–28. [CrossRef]
13. Duncan, T.; Pasik-Duncan, B.; Stettner, L. Almost self-optimizing strategies for the adaptive control of diffusion processes. *J. Optim. Theory Appl.* **1994**, *81*, 479–507. [CrossRef]
14. Durham, G.; Gallant, A. Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes. *J. Bus. Econ. Stat.* **2002**, *20*, 297–316. [CrossRef]
15. Huzak, M. Estimating a class of diffusions from discrete observations via approximate maximum likelihood method. *Statistics* **2018**, *52*, 239–272. [CrossRef]
16. Shoji, I. A note on asymptotic properties of the estimator derived from the Euler method for diffusion processes at discrete times. *Stat. Probab. Lett.* **1997**, *36*, 153–159. [CrossRef]
17. Athanassoglou, S.; Xepapadeas, A. Pollution control with uncertain stock dynamics: When, and how, to be precautious. *J. Environ. Econ. Manag.* **2012**, *63*, 304–320. [CrossRef]

18. Jiang, K.; You, D.; Li, Z.; Shi, S. A differential game approach to dynamic optimal control strategies for watershed pollution across regional boundaries under eco-compensation criterion. *Ecol. Indic.* **2019**, *105*, 229–241. [CrossRef]
19. Kawaguchi, K.; Morimoto, H. Long-run average welfare in a pollution accumulation model. *J. Econ. Dyn. Control* **2007**, *31*, 703–720. [CrossRef]
20. Jasso-Fuentes, H.; López-Barrientos, J.D. On the use of stochastic differential games against nature to ergodic control problems with unknown parameters. *Int. J. Control* **2015**, *88*, 897–909. [CrossRef]
21. Zhang, Z.; Zhang, G.; Su, B. The spatial impacts of air pollution and socio-economic status on public health: Empirical evidence from China. *Socio-Econ. Plan. Sci.* **2022**, *83*, 101167. [CrossRef]
22. Méndez-Cubillos, X.C.; de Souza, L.C.G. Using of H_∞ Control Method in Attitude Control System of Rigid-Flexible Satellite. *Math. Probl. Eng.* **2009**, 173145. [CrossRef]
23. Shaked, U.; Theodor, Y. H_∞ optimal estimation: A tutorial. In Proceedings of the 31st IEEE Conference on Decision and Control, Tucson, AZ, USA, 16–18 December 1992.
24. Cox, L.A.T., Jr. Confronting Deep Uncertainties in Risk Analysis. *Risk Anal.* **2012**, *32*, 1607–1629. [CrossRef]
25. Lu, J.; Xue, H.; Duan, X. An Adaptive Moving Mesh Method for Solving Optimal Control Problems in Viscous Incompressible Fluid. *Symmetry* **2022**, *14*, 707. [CrossRef]
26. Escobedo-Trujillo, B.A.; López-Barrientos, J.D.; Garrido-Meléndez, J. A Constrained Markovian Diffusion Model for Controlling the Pollution Accumulation. *Mathematics* **2021**, *9*, 1466. [CrossRef]
27. López-Barrientos, J.D.; Jasso-Fuentes, H.; Escobedo-Trujillo, B.A. Discounted robust control for Markov diffusion processes. *Top* **2015**, *23*, 53–76. [CrossRef]
28. Jasso-Fuentes, H.; Escobedo-Trujillo, B.; Mendoza-Pérez, A. The Lagrange and the vanishing discount techniques to controlled diffusions with cost constraints. *J. Math. Anal. Appl.* **2016**, *437*, 999–1035. [CrossRef]
29. Escobedo-Trujillo, B.; Hernández-Lerma, O.; Alaffita-Hernández, F. Adaptive control of diffusion processes with a discounted criterion. *Appl. Math.* **2020**, *47*, 225–253. [CrossRef]
30. Warga, J. *Optimal Control of Differential and Functional Equations*; Academic Press: New York, NY, USA, 1972.
31. Arapostathis, A.; Borkar, V.; Ghosh, M. Ergodic control of diffusion processes. In *Encyclopedia of Mathematics and Its Applications*; Cambridge University Press: Cambridge, UK, 2012; Volume 143.
32. Fleming, W.; Nisio, M. On the stochastic relaxed control for partially observed diffusions. *Nagoya Mathhematical J.* **1984**, *93*, 71–108. [CrossRef]
33. Ghosh, M.; Arapostathis, A.; Marcus, S. Optimal control of switching diffusions with applications to flexible manufacturing systems. *SIAM J. Control Optim.* **1992**, *30*, 1–23.
34. Mao, X.; Yuan, C. *Stochastic Differential Equations with Markovian Switching*; World Scientific Publishing Co.: London, UK, 2006.
35. Jasso-Fuentes, H.; Hernández-Lerma, O. Characterizations of overtaking optimality for controlled diffusion processes. *Appl. Math. Optim.* **2007**, *57*, 349–369. [CrossRef]
36. Dimarogonas, D.V.; Kyriakopoulos, K.J. Lyapunov-like stability of switched stochastic systems. *Proc. 2004 Am. Control Conf.* **2004**, *2*, 1868–1872. [CrossRef]
37. Fathi, M.; Bevrani, H. *Optimization in Electrical Engineering*; Springer International Publishing: Berlin/Heidelberg, Germany, 2019. [CrossRef]
38. Escobedo-Trujillo, B.A.; Higuera-Chan, C.G.; López-Barrientos, J.D. Controlled Switching Diffusions Under Ambiguity: The Average Criterion. *Int. Game Theory Rev.* **2021**, *23*, 2150017. [CrossRef]
39. Pedersen, A.R. Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusions process. *Bernoulli* **1995**, *1*, 257–279. [CrossRef]
40. Gilbarg, D.; Trudinger, N.S. *Elliptic Partial Differential Equations of Second Order*; Springer: Berlin/Heidelberg, Germany, 1998.
41. Yuan, C.; Mao, X. Convergence of the Euler–Maruyama method for stochastic differential equations with Markovian switching. *Math. Comput. Simul.* **2004**, *64*, 223–235. [CrossRef]
42. Kloeden, P.; Platen, E. *Numerical Solutions of Stochastic Differential Equations. Stochastic Modelling and Applied Probability*; Springer: Berlin/Heidelberg, Germany, 1992.
43. Hernández-Lerma, O.; Romera, R. The Scalarization Approach to Multiobjective Markov Control Problems: Why Does It Work? *Appl. Math. Optim.* **2004**, *50*, 279–293. [CrossRef]
44. Jasso-Fuentes, H.; López-Martínez, R.; Minjárez-Sosa, J. Some advances on constrained Markov decision processes in Borel spaces with random state-dependent discount factors. *Optimization* **2022**, 2130699. [CrossRef]
45. Gaspar-Cunha, A.; Covas, J. Robustness in multi-objective optimization using evolutionary algorithms. *Comput. Optim. Appl.* **2008**, *39*, 75–96. [CrossRef]
46. López-Barrientos, J.D. *Basic and Advanced Optimality Criteria for Zero–Sum Stochastic Differential Games*; Centro de Investigación y de Estudios Avanzados del IPN: México, 2012. Available online: www.math.cinvestav.mx/sites/default/files/tesis-daniel-2012.pdf (accessed on 18 January 2023).
47. Alaffita-Hernández, F.A.; Escobedo-Trujillo, B.A.; López-Martínez, R. Constrained stochastic differential games with additive structure: Average and discount payoffs. *J. Dyn. Games* **2018**, *5*, 109–141.

48. Adams, R. *Sobolev Spaces*; Academic Press: New York, NY, USA, 1975.
49. Lieb, E.; Loss, M. *Analysis*; American Mathematical Society: Providence, RI, USA, 2001.