



**HAL**  
open science

# A decentralized algorithm for a Mean Field Control problem of Piecewise Deterministic Markov Processes

Adrien Seguret, Thomas Le Corre, Nadia Oudjane

► **To cite this version:**

Adrien Seguret, Thomas Le Corre, Nadia Oudjane. A decentralized algorithm for a Mean Field Control problem of Piecewise Deterministic Markov Processes. 2022. hal-03910622v1

**HAL Id: hal-03910622**

**<https://hal.science/hal-03910622v1>**

Preprint submitted on 22 Dec 2022 (v1), last revised 25 Jan 2024 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A decentralized algorithm for a Mean Field Control problem of Piecewise Deterministic Markov Processes <sup>\*</sup>

Adrien Seguret<sup>1</sup>

Thomas Le Corre<sup>2</sup>

Nadia Oudjane<sup>3</sup>

## Abstract

This paper provides a decentralized approach for the control of a population of  $N$  agents to minimize an aggregate cost. Each agent evolves independently according to a Piecewise Deterministic Markov dynamics controlled via unbounded jumps intensities. The  $N$ -agent high dimensional stochastic control problem is approximated by the limiting mean field control problem. A Lagrangian approach is proposed. Although the mean field control problem is not convex, it is proved to achieve zero duality gap. A stochastic version of the Uzawa algorithm is shown to converge to the primal solution. At each dual iteration of the algorithm, each agent solves its own small dimensional sub problem by means of the Dynamic Programming Principal, while the dual multiplier is updated according to the aggregate response of the agents. Finally, this algorithm is used in a numerical simulation to coordinate the charging of a large fleet of electric vehicles (EVs for short) in order to track a target consumption profile.

## 1 Introduction

This paper focuses on a specific family of stochastic control problems called "large and aggregative". The problem consists in optimizing the strategies of a large number of agents, in a random environment, while the decisions of all the agents interact via the objective function because the latter depends on an "aggregative" term as the sum of the decisions of all the agents. This type of problem occurs naturally in several fields of application such as energy management [16], telecommunications [20], portfolio management [18], robot coordination [4] among others.

To deal with the curse of dimensionality in such high dimensional stochastic control problems, methods based on the value function, have been developed in the convex setting. For instance, Stochastic Dual Dynamic Programming [21] relies on polyhedral approximations of the value function providing good performances for medium sized problems with a number of agents  $N \leq 30$ . In our case, where the number of agents is very large  $N \geq 1000$ , we look more specifically for decentralized solutions, allowing both to respect the privacy of each agent and to reduce the complexity of the problem. In the particular setting, where controls are bounded and strong convexity conditions are satisfied, [25] proposed an iterative stochastic algorithm providing a decentralized solution to such problems. In the same spirit as the Dual Approximate Dynamic Programming [5], this approach relies on a Lagrangian decomposition technique to obtain a decentralized solution, but it takes advantage of a mean field approximation to ensure the validity of the approach when the number of agents is sufficiently large.

This paper aims at extending the decentralized approach developed in [25] to a non-convex framework, with unbounded controls, allowing for more realistic dynamics ruling the evolution of each agent state. In particular, we are interested in the aggregative control of Piecewise Deterministic Markov Processes (PDMPs for short), where the state of each agent is modelled by a PDMP controlled by an unbounded jumps intensity. PDMPs were introduced in [8] as a class of non-diffusion stochastic models, mixing random jumps and deterministic dynamics between jumps. This kind of process is used to model a wide broad of phenomena or situations, such as system reliability and maintenance [10, 28], oil production [29], biology [19, 23] etc... The approach proposed in this paper results in a decentralized algorithm, where at each iteration, each agent is intended to solve its own small dimensional optimal control problem of PDMP. This type of problem is well known and has been studied through dynamic programming

---

<sup>\*</sup>This research benefited from the support of the FMJH Program Gaspard Monge for optimization and operations research and their interactions with data science.

<sup>1</sup>Adrien Seguret is with PSL Research University, Universite Paris-Dauphine, CEREMADE, Place de Lattre de Tassigny, 75016 Paris, France, and with Finance for Energy Market Research Centre (FIME), Paris, France, and with Osiris, EDF R&D, 91120 Palaiseau, France [adrien.seguret@edf.fr](mailto:adrien.seguret@edf.fr)

<sup>2</sup>Thomas Le Corre is with DI ENS, CNRS, PSL University, INRIA Paris [thomas.le-corre@inria.fr](mailto:thomas.le-corre@inria.fr)

<sup>3</sup>Nadia Oudjane is with Osiris, EDF R&D, 91120 Palaiseau, France and with Finance for Energy Market Research Centre (FIME), Paris, France [nadia.oudjane@edf.fr](mailto:nadia.oudjane@edf.fr)

and related equations [1, 7, 9, 15, 27]. Due to the large number of agents, the problem is formulated as a Mean Field Control (MFC for short) problem. Relying on the law of large numbers, the MFC theory consists in controlling a representative agent of the population, where the interaction between the agents is approximated by the interaction of this representative agent with the probability distribution of its state or control [3].

This work is particularly motivated by demand-side management in power systems to help balancing between energy production and consumption. This problem is indeed critical today, due to the increasing share of uncontrollable energies (wind and solar) in the electricity generation mix, which requires to compensate the uncontrollable character of the production by controlling the demand. Controlling the sum of the consumption of each flexible consumer in order to balance the electrical system has already been investigated successfully in the specific framework of Quadratic Kullback-Leibler control problems [2]. In [26] and [17], a mean field assumption is also considered to control the charging of a large fleet of electrical vehicles (EVs), leading to the optimal control of partial differential equation problems.

The novelty of the present paper is to extend the Lagrangian approach proposed in [25], to a particular non-convex setting with unbounded controls by first establishing the existence of a saddle point and then ensuring the convergence of the decentralized algorithm. The first contribution lies in the originality of the proof of a saddle point existence which follows a completely different path than the one developed in [25]. In particular, we make use of regularity results on the solution of the Hamilton Jacobi equation arising in optimal control of PDMPs developed in [24] to show the existence of a saddle point of our Lagrangian problem. The second contribution consists in proving that the Stochastic Uzawa Algorithm proposed in [25] is still providing a converging sequence of controls, in this new setting involving the PDMP dynamics which violates the convexity conditions originally exploited. Finally, an application to the smart charging of an electric vehicles fleet by an aggregator illustrates the performance and the interest of the approach to coordinate the charge of a large number of electric vehicles in order to track a given target power consumption profile for the whole population of electrical vehicles.

The outline of this paper is as follows. In Section 2, we formulate the optimization problem, the assumptions and the main results. Section 3 presents a dual approach of this problem. The Stochastic Uzawa algorithm is presented in Section 4 and is proved to converge in this setting. Section 5 presents simulations of the coordination of power consumption of a large fleet of electrical vehicles.

In the rest of this section we will list some frequently used notation.

**Notation** Let  $\mathcal{X}$  be the state space defined by  $\mathcal{X} := I \times [0, 1]$ , where  $I$  is a finite set, of cardinality  $d \in \mathbb{N}^*$ . The space  $D([0, T], I)$  is the space of càdlàg functions from  $[0, T]$  to  $I$ . The space  $C^1(\mathcal{X}, \mathbb{R})$  denotes the set of real-valued continuously differentiable functions defined on  $\mathcal{X}$ . The set of  $\mathbb{R}_+^d$ -valued Lipschitz continuous functions defined on  $[0, T] \times \mathcal{X}$  is denoted by  $\text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}^d)$ . The space of Borel probability measures on the space  $\mathcal{X}$  is denoted by  $\mathcal{P}(\mathcal{X})$ . We recall the definition of the Wasserstein distance, denoted by  $\mathbb{W}$ , on the space  $\mathcal{P}(\mathcal{X})$ :

$$\mathbb{W}(\mu, \rho) := \sup \left\{ \sum_{i \in I} \int_0^1 \varphi(i, s) (\mu(i, ds) - \rho(i, ds)) \mid \varphi \text{ is } 1 - \text{Lipschitz from } \mathcal{X} \text{ to } \mathbb{R} \right\}.$$

We recall that if a function  $\varphi$  is 1-Lipschitz continuous from  $\mathcal{X}$  to  $\mathbb{R}$ , then  $|\varphi(i, x) - \varphi(j, x)| \leq 1$  for any  $i, j \in I$  and  $x \in [0, 1]$ . The space  $C([0, T], \mathcal{P}(\mathcal{X}))$  is endowed with the distance  $\mathbb{W}_{\|\cdot\|_\infty}$  defined by:

$$\mathbb{W}_{\|\cdot\|_\infty}(m^1, m^2) := \sup_{t \in [0, T]} \mathbb{W}(m^1(t), m^2(t)). \quad (1.1)$$

Let  $\mathbb{H}$  be an Hilbert space and  $F$  be a real valued function defined on  $\mathbb{H}$ . The convex conjugate of  $F$  is denoted by  $F^*$  and is defined for any  $x \in \mathbb{H}$  by  $F^*(x) := \sup_{y \in \mathbb{H}} \langle x, y \rangle_{\mathbb{H}} - F(y)$ .

## 2 Problem formulation and main results

We consider the time interval  $[0, T]$  with  $T > 0$  and a population of  $N$  independent and identically distributed processes, controlled by a central planner via a common feedback control  $\alpha$ . The space of feedbacks, denoted by  $\mathbb{A}$ , is defined by

$$\mathbb{A} := \{ \alpha \in C^0([0, T] \times \mathcal{X}, \mathbb{R}_+^d) : \forall i \in I, \alpha_i(\cdot, i, \cdot) = 0 \}, \quad (2.1)$$

and is endowed with the norm  $\|\cdot\|_\infty$ . The state of an agent, controlled by the jump intensity  $\alpha$ , is given at time  $t$  by  $X_t^\alpha = (Y_t^\alpha, Z_t^\alpha)$  with  $Y_t^\alpha \in I$  and  $Z_t^\alpha \in [0, 1]$ . In the smart charging application in Section 5,  $Y^\alpha$  represents the charging mode of a vehicle, specifying its discrete charging rate (e.g.  $i = 0$  for not charging,  $i = 1$  for charging

and  $i = -1$  for discharging). The continuous variable  $Z^\alpha$  represents the state of charge (SoC for short) of an EV, that is the ratio of energy charged into the battery w.r.t. the total capacity of the battery. We say that the process  $X^\alpha$  is a PDMP( $b, \alpha$ ) controlled by a feedback control  $\alpha$ , if  $X^\alpha = (Y^\alpha, Z^\alpha)$  where  $Y^\alpha$  is a jump process with values in  $I$  switching spontaneously, at jump times  $\{T_k^\alpha\}_{k \in \mathbb{N}}$  given by a Poisson process with intensity  $\alpha$ , while  $Z^\alpha$  follows a deterministic dynamics between two consecutive jumps  $Z_t^\alpha = \phi_{T_k^\alpha, Z_{T_k^\alpha}^\alpha}(Y_{T_k^\alpha}^\alpha, t)$  for  $t \in [T_k^\alpha, T_{k+1}^\alpha)$ . For any  $(\tau, t, j, z) \in [0, T] \times [\tau, T] \times \mathcal{X}$ , the flow  $\phi$  is the unique solution of the ordinary differential equation:

$$\partial_t \phi_{\tau, z}(j, t) = b(j, \phi_{\tau, z}(j, t)) \text{ with } \phi_{\tau, z}(j, \tau) = z, \quad (2.2)$$

where  $b$  is a function given. To be more specific, following the definition of a PDMP given in [11], knowing  $T_k^\alpha$  and  $X_{T_k^\alpha}^\alpha = (Y_{T_k^\alpha}^\alpha, Z_{T_k^\alpha}^\alpha)$ , one obtains  $(T_{k+1}^\alpha, X_{T_{k+1}^\alpha}^\alpha)$  as follows:

$$\begin{cases} T_{k+1, j} := \inf \left\{ t \geq T_k : E_{k+1, j} < \int_{T_k}^t \alpha_j(r, X_{T_k}^\alpha) dr \right\} \\ T_{k+1}^\alpha := \min_{j \in \{1, \dots, d\}} T_{k+1, j} \\ Y_{T_{k+1}^\alpha}^\alpha = \min \{ j \in \{1, \dots, d\} : T_{k+1, j} = T_{k+1}^\alpha \} \\ Z_{T_{k+1}^\alpha}^\alpha := \phi_{Y_{T_{k+1}^\alpha}^\alpha}(Z_{T_k}^\alpha, T_{k+1}^\alpha) \\ X_{T_{k+1}^\alpha}^\alpha = (Y_{T_{k+1}^\alpha}^\alpha, Z_{T_{k+1}^\alpha}^\alpha), \end{cases}$$

where  $\{E_{k, j}\}_{k, j}$  are independent random variables following an exponential distribution of parameter 1, independent of  $X_0^\alpha$ . Indeed, the control process  $\alpha = (\alpha_j)_{j \in I}$  is such that  $\alpha_j(t, x)$  represents the intensity rate, at time  $t \in [0, T]$ , of jumping into mode  $j \in I$  when coming from state  $x \in \mathcal{X}$ . We initialize  $T_0 = 0$  and at time  $t = 0$ , the law of the couple of random variables  $(Y_0^\alpha, Z_0^\alpha)$  is given by  $m^0 \in \mathcal{P}(\mathcal{X})$ . The cost function  $J^N$  for the  $N$ -agents problem is defined for any  $\alpha \in \mathbb{A}$  by:

$$J^N(\alpha) := \mathbb{E} \left[ \int_0^T f \left( t, \frac{1}{N} \sum_{n=1}^N p(t, X_t^{n, \alpha}) \right) dt + \frac{1}{N} \sum_{n=1}^N G(\alpha, X^{n, \alpha}) \right], \quad (2.3)$$

where  $\{X^{n, \alpha}\}_{n \in \{1, \dots, N\}}$  are supposed to be independent PDMP( $b, \alpha$ ) driven by  $d$  independent exponential variables. Indeed, for any  $n \in \{1, \dots, N\}$ ,  $X^{n, \alpha}$  is supposed to be controlled by the common feedback function  $\alpha \in \mathbb{A}$  only depending on the agent state  $X^n$ . Besides, the function  $G$  is common to every agent and defined as the the sum of a running and a terminal cost, such that for any  $x := (y, z) \in D([0, T], I) \times C^0([0, T], [0, 1])$ , by:

$$G(\alpha, x) := \int_0^T \left[ c(t, x_t) + \sum_{j \in I} L(\alpha_j(t, x_t)) \right] dt + g(x_T).$$

From a practical point of view, the function  $f$  represents a coupling cost depending on the aggregate quantity  $\frac{1}{N} \sum_{n=1}^N p(t, X_t^{n, \alpha})$ , and  $G$  the individual cost. In this paper, we focus on the mean field limit control problem, corresponding to the problem with an infinite population. Thus, we introduce the function  $J$  defined, for any  $\alpha \in \mathbb{A}$ , by:

$$J(\alpha) := \int_0^T f(t, \mathbb{E}[p(t, X_t^\alpha)]) dt + \mathbb{E}[G(\alpha, X^\alpha)], \quad (2.4)$$

where  $X^\alpha$  is a PDMP( $b, \alpha$ ). This paper is dedicated to the following problem:

$$\min_{\alpha \in \mathbb{A}} J(\alpha). \quad (2.5)$$

As the cost function (2.4) is nonlinear w.r.t. the expectation term,  $\mathbb{E}[p(t, X_t^\alpha)]$ , via the coupling cost  $f$ , Problem (2.5) goes beyond the scope of optimal control of PDMP. We propose to numerically solve Problem (2.5) by applying the Stochastic Uzawa Algorithm (1), that is detailed in Section 4. This algorithm is introduced in [25] and is a direct application of the stochastic gradient descent algorithm [13]. The main result of this paper, stated in Theorem 2.1, ensures the convergence of the Stochastic Uzawa Algorithm 1 to the solution of Problem (2.5).

Throughout the paper, we assume the following:

### General assumptions

1. The vector field  $b \in C^1(\mathcal{X}, \mathbb{R})$  is assumed to vanish at the boundary:  $b(j, 0) = b(j, 1) = 0$  for any  $j \in I$ .
2. The function  $p \in C^1([0, T] \times \mathcal{X}, \mathbb{R})$  and  $f : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$  is a Carathéodory function being lower semi-continuous (l.s.c. for short), strictly convex and differentiable w.r.t. the second variable. In addition, there exists a constant  $C_f > 0$  such that, for any  $(t, x) \in [0, T] \times \mathbb{R}$ ,

$$\frac{x^2}{2C_f} - C_f \leq f(t, x) \leq C_f \frac{x^2}{2} + C_f. \quad (2.6)$$

3. For any  $i \in I$ , it is assumed that  $c_i \in C^1([0, T] \times [0, 1])$  and  $g_i \in C^1([0, 1])$ .
4. The function  $L : \mathbb{R} \rightarrow \bar{\mathbb{R}}$  is convex and defined by:

$$L(x) := \begin{cases} l(x) & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ +\infty & \text{otherwise,} \end{cases}$$

where  $l \in C^1(\mathbb{R}_+, \mathbb{R}_+)$  is an increasing strongly convex function bounded from above by a quadratic function. More explicitly, there exists  $C > 0$  such that for any  $x \in \mathbb{R}_+$ :

$$\frac{x^2}{C} - C \leq l(x) \leq C(x^2 + 1),$$

where the first inequality is due to the strong convexity of  $l$ . We denote by  $H$  be the convex conjugate of  $L$ . Note that by Assumption 4,  $H$  is non-decreasing, non-negative, and  $H'$  is Lipschitz continuous on  $\mathbb{R}$ .

**Remark 2.1.** *The main role of Assumption 1 is to ensure that the flow defined in (2.2) exists and takes values in  $[0, 1]$ . Assumption 2 ensures that  $\int_0^T f(t, \mathbb{E}[p(t, X_t^\alpha)]) dt$  in (2.4) is well defined for any  $\alpha \in \mathbb{A}$ . The continuity of the map  $\alpha \mapsto \mathbb{E}[G(\alpha, X^\alpha)]$  is an automatic consequence of the properties of  $c$ ,  $g$  and  $L$  given in Assumptions 3 and 4.*

**Remark 2.2.** *The boundary condition on  $b$  may seem restrictive. However, if  $\overline{\text{supp}(m^0)} \subset I \times (0, 1)$  and the vector field  $b$  is such that, for any  $i \in I$ , the sign of  $b(i, \cdot)$  is constant,  $b(i, 0) \geq 0$  and  $b(i, 1) \leq 0$ , then there exists  $\varepsilon > 0$  such that  $\mathbb{P}(Z_t^{n, \alpha} \in (\varepsilon, 1 - \varepsilon)) = 1$  for any  $n \in \{1, \dots, N\}$ , any  $\alpha \in \mathbb{A}$  and any  $t \in [0, T]$ . Therefore, it is possible to approximate such a vector field  $b$  by a smooth function, vanishing at the boundary without modifying the trajectory of  $Z_t^{n, \alpha}$ .*

The main results of the paper are summarized in the following theorem.

**Theorem 2.1.** *Problem 2.5 has a solution. Let  $\{\alpha^k\}_{k \in \mathbb{N}}$  be a sequence in  $\mathbb{A}$  generated by Algorithm 1 (line 5), then the following assertions hold*

- (i) *The sequence  $\{\alpha^k\}_{k \in \mathbb{N}}$  converges a.s. to an element of  $\underset{\alpha \in \mathbb{A}}{\text{argmin}} J(\alpha)$  w.r.t. the norm  $\|\cdot\|_\infty$ .*
- (ii) *The sequence  $\{J(\alpha^k)\}_{k \in \mathbb{N}}$  converges a.s. to  $\min_{\alpha \in \mathbb{A}} J(\alpha)$ .*

This theorem can not be obtained by a direct application of [25], which relies on additional assumptions, that are not verified in the present framework, which involves unbounded controls and PDMP dynamics. The proof of Theorem 2.1 is given in Section 4. Before, we introduce and analyze in the next section a dual problem proved to be equivalent, in some sense, to Problem 2.5.

### 3 Dual approach

In the same vein as [25], a Lagrangian decomposition approach is adopted to obtain a decentralized algorithm. Contrary to [25], the specific assumptions of convexity, that ensure the absence of duality gap and the convergence of the Stochastic Uzawa Algorithm, are not satisfied here. In Sections 3 and 4, we propose new theoretical arguments allowing to demonstrate the validity of the approach proposed in [25] in this specific framework of control of PDMP.

Let  $F$  be defined for any  $v \in L^2(0, T)$  by:

$$F(v) := \int_0^T f(t, v(t)) dt.$$

Owing to the properties of  $f$ , one deduces that function  $F$  is l.s.c., strictly convex, differentiable with a Lipschitz continuous gradient, and it has at least quadratic growth.

**Lemma 3.1.** *The problem*

$$\begin{cases} \min_{\alpha \in \mathbb{A}, v \in L^2(0, T)} \bar{J}(\alpha, v), \\ \bar{J}(\alpha, v) := F(v) + \mathbb{E}[G(\alpha, X^\alpha)], \\ \text{s.t. } \mathbb{E}[p(t, X_t^\alpha)] - v(t) = 0 \text{ a.e. on } [0, T], \end{cases} \quad (3.1)$$

is equivalent to Problem (2.5).

*Proof.* The fact that  $p \in C^1([0, T] \times \mathcal{X}, \mathbb{R})$  implies that the map  $t \mapsto \mathbb{E}[p(t, X_t^\alpha)]$  is in  $L^2(0, T)$ . The equivalence between the two problems then follows from the definitions of  $J$  and  $\bar{J}$ .  $\square$

Let us introduce the Lagrangian  $\mathcal{L} : \mathbb{A} \times L^2(0, T) \times L^2(0, T) \rightarrow \mathbb{R}$ :

$$\mathcal{L}(\alpha, v, \lambda) := \mathcal{L}_1(\alpha, \lambda) + \mathcal{L}_2(v, \lambda), \quad (3.2)$$

where

$$\begin{aligned} \mathcal{L}_1(\alpha, \lambda) &:= G(\alpha, X^\alpha) + \int_0^T \mathbb{E}[p(t, X_t^\alpha)] \lambda(t) dt, \\ \mathcal{L}_2(v, \lambda) &:= F(v) - \int_0^T v(t) \lambda(t) dt, \end{aligned}$$

and the associated dual function  $\mathcal{W} : L^2(0, T) \rightarrow \mathbb{R}$ :

$$\mathcal{W}(\lambda) := \inf_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda) + \inf_{v \in L^2(0, T)} \mathcal{L}_2(v, \lambda). \quad (3.3)$$

The dual problem associated to (3.1) consists of the following maximization problem:

$$\max_{\lambda \in L^2(0, T)} \mathcal{W}(\lambda). \quad (3.4)$$

The following lemma gives the existence of a unique solution of the dual problem (3.4). It is derived from the assumptions on  $F$  and  $G$ .

**Lemma 3.2.** *There exists a unique  $\bar{\lambda} \in L^2(0, T)$  such that  $\bar{\lambda} = \operatorname{argmax}_{\lambda \in L^2(0, T)} \mathcal{W}(\lambda)$ .*

*Proof.* Since  $F$  has a Lipschitz continuous gradient, the function  $\lambda \mapsto F^*(\lambda) := \sup_{v \in L^2(0, T)} \langle v, \lambda \rangle - F(v)$  is strongly convex and l.s.c. [14], and so is the function  $\lambda \mapsto -\inf_{v \in L^2(0, T)} \mathcal{L}_2(v, \lambda)$ . Similarly, for any  $\lambda \in L^2(0, T)$ , the map  $\lambda \mapsto -\inf_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$  is convex and l.s.c.. One deduces that  $-\mathcal{W}$  is strongly convex and l.s.c. on  $L^2(0, T)$ . Thus, problem (3.4) has a unique solution.  $\square$

The below proposition is a key result, that enables to show the convergence of a sequence  $\{\alpha^k\}_k$  generated by Algorithm 4 defined in the next section, to a solution of Problem (2.5).

**Theorem 3.1.** *For any  $\lambda \in L^2(0, T)$ ,  $\operatorname{argmin}_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$  is not empty. In addition, there exists a selection*

$$\lambda \mapsto \alpha[\lambda] \in \operatorname{argmin}_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda), \quad (3.5)$$

such that the map  $L^2(0, T) \ni \lambda \mapsto \alpha[\lambda] \in \mathbb{A}$  is Lipschitz continuous on any bounded subset of  $L^2(0, T)$ . The function  $\lambda \mapsto \alpha[\lambda]$  can be continuously extended to any  $\lambda \in L^2(0, T)$ .

Theorem 3.1 allows to prove that the dual function  $\mathcal{W}$  is Gâteaux differentiable in  $L^2(0, T)$ , that is the main argument of the proof of Theorem 3.2. Theorem 3.2 gives the existence of a saddle point of the Lagrangian function  $\mathcal{L}$ . This result provides a solution of Problem (2.5).

**Theorem 3.2.** *There is no duality gap associated with Problem (3.4), i.e.,*

$$\max_{\lambda \in L^2(0,T)} \mathcal{W}(\lambda) = \min_{\alpha \in \mathbb{A}, v \in L^2(0,T)} \bar{J}(\alpha, v).$$

Besides,  $\operatorname{argmin}_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \bar{\lambda})$  and  $\operatorname{argmin}_{v \in L^2(0,T)} \mathcal{L}_2(v, \bar{\lambda})$  are not empty and, for any  $\bar{\alpha} \in \operatorname{argmin}_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \bar{\lambda})$  and  $\bar{v} \in \operatorname{argmin}_{v \in L^2(0,T)} \mathcal{L}_2(v, \bar{\lambda})$ ,  $((\bar{\alpha}, \bar{v}), \bar{\lambda})$  is a saddle point of the Lagrangian  $\mathcal{L}$  and  $\bar{\alpha}$  is a solution of Problem (2.5).

The main argument for the proof of Theorem 3.2 is that the map  $\lambda \mapsto \mathcal{W}(\lambda)$  is Gâteaux differentiable in  $L^2(0, T)$ . We show that  $\lambda \mapsto \inf_{v \in L^2(0, T)} \mathcal{L}_2(v, \lambda)$  and  $\lambda \mapsto \inf_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$  are both differentiable. The second result is more difficult to prove. Different intermediary results are needed. First, properties of the sub problem w.r.t. the variable  $v$  in the dual problem (3.4) are given in Lemma 3.3. Then, properties of the sub problem w.r.t. the variable  $\alpha$  are proved in Lemma 3.4 in the special case where  $\lambda \in L^\infty(0, T)$ . These properties are extended in Theorem 3.1 to the case where  $\lambda \in L^2(0, T)$ , with the help of Lemmas A.1 in Appendix A and 3.5. Finally, Lemma 3.6 shows that the map  $\lambda \mapsto \inf_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$  is Gâteaux differentiable in  $L^2(0, T)$ . These lemmas are stated and proved below.

### 3.1 Proof of Theorem 3.1

The next Lemma shows that, for any  $\lambda \in L^2(0, T)$ , there exists a unique solution  $v[\lambda]$  of the sub problem and that the map  $\lambda \mapsto \mathcal{L}_2(v[\lambda], \lambda)$  is Gâteaux differentiable.

**Lemma 3.3.**

- (i) For any  $\lambda \in L^2(0, T)$ , there exists a unique  $v[\lambda]$  satisfying  $v[\lambda] = \operatorname{argmin}_{v \in L^2(0, T)} \mathcal{L}_2(v, \lambda)$ .
- (ii) The map  $\lambda \mapsto \min_{v \in L^2(0, T)} \mathcal{L}_2(v, \lambda)$  admits a Gâteaux derivative  $\lambda \mapsto D\mathcal{L}_2(v[\lambda], \lambda)(\cdot)$  in  $L^2(0, T)$ , which satisfies that, for any  $\mu \in L^2(0, T)$ :

$$D\mathcal{L}_2(v[\lambda], \lambda)(\mu) = - \int_0^T v[\lambda](\tau) \mu(\tau) d\tau. \quad (3.6)$$

*Proof.* (i) Since the map  $v \mapsto F(v)$  is l.s.c., strictly convex with at least quadratic growth, one can deduce the existence and uniqueness of  $v[\lambda] := \operatorname{argmin}_{v \in L^2(0, T)} \mathcal{L}_2(v, \lambda)$ .

(ii) Since the function  $F$  is proper, strictly convex and differentiable, classical results from convex analysis give that its convex conjugate  $F^*$  is differentiable [14]. Therefore, the map  $\lambda \mapsto \mathcal{L}_2(v[\lambda], \lambda)$  is differentiable on  $L^2(0, T)$ , with derivative  $\lambda \mapsto v[\lambda]$   $\square$

Regularity results about  $\alpha[\lambda]$ , defined in (3.5), are stated in the next Lemma for any  $\lambda \in L^\infty(0, T)$ .

**Lemma 3.4.** *Let  $\lambda \in L^\infty(0, T)$ . Then, one has:*

- (i) The value function  $u$  associated to the stochastic control problem  $\inf_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$  is the unique solution of the associated Hamilton-Jacobi-Bellman equation (in the sense of [15, Theorem 3.4]) on  $[0, T] \times I \times [0, 1]$ :

$$\begin{aligned} & -\partial_t u(t, i, z) - b(i, z) \partial_z u(t, i, z) - c(t, i, z) - \lambda(t) p(t, i, z) \\ & = \inf_{(a_{i,j})_{j \geq 0}} \sum_{j \in I, j \neq i} L(a_{i,j}) + (u(t, j, z) - u(t, i, z)) a_{i,j}, \end{aligned} \quad (3.7)$$

with the terminal condition:  $u(T, i, z) = g(i, z)$  for any  $(i, z) \in \mathcal{X}$ .

- (ii)  $\alpha[\lambda]$  is the unique solution of the problem  $\inf_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$  and is given, for any  $i, j \in I$  and  $(t, z) \in [0, T] \times [0, 1]$ , by,

$$\alpha[\lambda]_j(t, i, z) = H'(u(t, i, z) - u(t, j, z)).$$

Moreover, the function  $\alpha[\lambda]$  is in  $\operatorname{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d)$ .

- (iii) The function  $L^\infty(0, T) \ni \lambda \mapsto \alpha[\lambda] \in \mathbb{A}$  is Lipschitz continuous, w.r.t. the norm  $\|\cdot\|_2$ , on any bounded subset of  $L^\infty(0, T)$ .

**Remark 3.1.** The proof of Lemma 3.4 essentially relies on [15, Theorem 3.4]. The assumptions needed for the application of this theorem require bounded running cost in the minimization problem  $\min_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$ . Thus, we only consider  $\lambda \in L^\infty(0, T)$  and not in  $L^2(0, T)$  in the statement of Lemma 3.4. The extension of the results of Lemma 3.4 for  $\lambda \in L^2(0, T)$  is stated in Theorem 3.1.

*Proof.* (i) The problem  $\inf_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$  is a stochastic control problem of PDMP. Let  $u$  be the value function of this problem, defined on  $[0, T] \times I \times [0, 1]$  by:

$$\begin{aligned} u(t, i, z) := & \inf_{\alpha \in \mathbb{A}} \int_t^T \sum_{j \in I, j \neq i} \mathbb{E}[H(\alpha_j(\tau, X_\tau^\alpha) \mid X_t^\alpha = (i, z))] d\tau \\ & + \int_t^T \mathbb{E}[c(\tau, X_\tau^\alpha) + \lambda(\tau)p(\tau, X_\tau^\alpha) \mid X_t^\alpha = (i, z)] d\tau \\ & + \mathbb{E}[g(X_T^\alpha) \mid X_t^\alpha = (i, z)] \quad \text{for a.e. } (t, z) \in (0, T) \times (0, 1). \end{aligned}$$

By [15, Theorem 3.4], the value function  $u$  is the unique function satisfying for a.e.  $(t, z, i) \in (0, T) \times I \times (0, 1)$ :

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{u(t, i, z) - u(t+h, i, \phi_{t,z}(i, t+h))}{h} = & c(t, i, z) + p(t, i, z)\lambda(t) \\ & + \inf_{(a_{i,j})_{j \geq 0}} \sum_{j \in I, j \neq i} L(a_{i,j}) + (u(t, j, z) - u(t, i, z))a_{i,j}, \end{aligned}$$

with  $u(T, i, z) = g(i, z)$  on  $\mathcal{X}$ . By [15] and the definition of  $\phi$  in (2.2),  $u$  satisfies for any  $i \in I$  and a.e.  $(t, z) \in [0, T] \times [0, 1]$ :

$$\lim_{h \rightarrow 0} \frac{u(t, i, z) - u(t+h, i, \phi_{t,z}(i, t+h))}{h} = -(\partial_t u(t, i, z) + b(i, z)\partial_z u(t, i, z)).$$

The conclusion follows from the two previous equalities.

(ii) By taking the infimum on the r.h.s. of the equation (3.7) and recalling that  $H$  is the convex conjugate of  $L$ , one obtains that  $u$  satisfies for a.e.  $(t, z, i) \in (0, T) \times I \times (0, 1)$ :

$$-\partial_t u(t, i, z) - b(i, z)\partial_z u(t, i, z) - c_i(t, i, z) - \lambda(t)p(t, i, z) = - \sum_{j \in I, j \neq i} H(u(t, i, z) - u(t, j, z)), \quad (3.8)$$

with  $u(T, i, z) = g(i, z)$  on  $\mathcal{X}$  and where  $x^+ := \max(0, x)$ . Since there exists a unique solution  $u$  to the previous equation, there exists a unique  $\bar{\alpha}_j[\lambda](t, i, z)$ , defined, for any  $(t, z, i, j) \in [0, T] \times [0, 1] \times I \times I$  by:

$$\bar{\alpha}_j(t, i, z) := \operatorname{argmin}_{a \geq 0} L(a) + (u(t, j, z) - u(t, i, z))a = H'(u(t, i, z) - u(t, j, z)). \quad (3.9)$$

By [24], since  $\lambda \in L^\infty$  and  $p \in C^1([0, T] \times \mathcal{X}, \mathbb{R})$ , there exists a unique function  $\bar{u}$  which satisfies for a.e.  $t \in [0, T]$  and for any  $(i, z) \in \mathcal{X}$ :

$$\begin{aligned} \bar{u}(t, i, z) = & \int_t^T \sum_{j \in I} -H(\bar{u}(\tau, i, \phi_{t,z}(i, \tau)) - \bar{u}(\tau, j, \phi_{t,z}(i, \tau))) d\tau \\ & + \int_t^T c(\tau, i, \phi_{t,z}(i, \tau)) + \lambda(\tau)p(\tau, i, \phi_{t,z}(i, \tau)) d\tau + g(i, \phi_{t,z}(i, T)). \end{aligned}$$

Since  $\bar{u}$  is in  $\operatorname{Lip}([0, T] \times \mathcal{X}, \mathbb{R})$ , the set of real valued Lipschitz continuous functions defined on  $[0, T] \times \mathcal{X}$ ,  $\bar{u}$  satisfies, for a.e.  $(t, i, z) \in [0, T] \times \mathcal{X}$ , the equation (3.8). Therefore,  $\bar{u}$  is the value function associated to the stochastic control problem  $\inf_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$ . It comes, from (3.9) and  $u = \bar{u}$ , that  $\alpha[\lambda] \in \operatorname{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d)$  and thus,  $\alpha[\lambda] \in \mathbb{A}$ .

Finally, by [15, Theorem 3.4],  $\alpha[\lambda]$  is in  $\operatorname{argmin}_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$ .

(iii) Let  $k > 0$  and  $\lambda, \mu \in L^\infty(0, T)$  be such that  $\|\lambda\|_2 \leq k$  and  $\|\mu\|_2 \leq k$ . Then, by the same arguments as in the proof of [24, Lemma 4.5], there exists a constant  $K(k, T) > 0$  such that for any  $(t, i, j, z) \in [0, T] \times I \times I \times [0, 1]$ :

$$|\alpha_j[\lambda](t, i, z) - \alpha_j[\mu](t, i, z)| \leq K(k, T)\|p\|_\infty \|\lambda - \mu\|_2.$$

The result follows from taking the supremum over  $[0, T] \times \mathcal{X}$  of the l.h.s. of the previous inequality.  $\square$



Lemma 3.5 enables to show the continuity of the map  $(\alpha, \lambda) \mapsto \mathcal{L}_1(\alpha, \lambda)$ .

**Lemma 3.5.** *The map  $(\alpha, \lambda) \mapsto \mathcal{L}_1(\alpha, \lambda)$  is continuous on  $\text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d) \times L^2(0, T)$ , where  $\text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d) \times L^2(0, T)$  is endowed with the norm  $(\alpha, \lambda) \mapsto \|\alpha\|_\infty + \|\lambda\|_2$ .*

*Proof.* Recalling the expression of  $H$ , one has for any  $(\alpha, \lambda) \in \text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d) \times L^2(0, T)$ :

$$\begin{aligned} \mathcal{L}_1(\alpha, \lambda) &= \int_0^T \int_0^1 \sum_{i \in I} c(t, i, z) m^\alpha(t, i, dz) dt + \int_0^T \int_0^1 \sum_{i \in I} \lambda(t) p(t, i, z) m^\alpha(t, i, dz) dt \\ &\quad + \int_0^T \int_0^1 \sum_{i, j \in I} L(\alpha_j(t, i, z)) m^\alpha(t, i, dz) dt + \int_0^1 \sum_{i \in I} g(i, z) m^\alpha(T, i, dz). \end{aligned}$$

Let  $\{(\alpha^n, \lambda^n)\}_n$  be a sequence in  $\text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d) \times L^2(0, T)$ , converging to  $(\alpha, \lambda)$ . The convergence of  $\{m^{\alpha^n}\}_k$  to  $m^\alpha$ , w.r.t. the distance  $\mathbb{W}_{\|\cdot\|_\infty}$ , is derived from Lemma A.1 in Appendix A. Thus, by the regularity of  $c, g, p$  and  $L$  in Assumptions 2-4, the maps  $t \mapsto \sum_{i \in I} \int_0^1 p(t, i, z) m^{\alpha^n}(t, i, dz)$ ,  $t \mapsto \sum_{i \in I} \int_0^1 c(t, i, z) m^{\alpha^n}(t, i, dz)$ ,  $t \mapsto \sum_{i \in I} \int_0^1 L(\alpha_j^n(t, i, z)) m^{\alpha^n}(t, i, dz)$  converge, w.r.t. the norm  $\|\cdot\|_\infty$ , when  $n$  tends to infinity, respectively to  $t \mapsto \sum_{i \in I} \int_0^1 p(t, i, z) m^\alpha(t, i, dz)$ ,  $t \mapsto \sum_{i \in I} \int_0^1 c(t, i, z) m^\alpha(t, i, dz)$ ,  $t \mapsto \sum_{i \in I} \int_0^1 L(\alpha_j(t, i, z)) m^\alpha(t, i, dz)$ . Similarly  $\sum_{i \in I} \int_0^1 g(i, z) m^{\alpha^n}(T, i, dz)$  converges when  $n$  tends to infinity to  $\sum_{i \in I} \int_0^1 g(i, z) m^\alpha(T, i, dz)$ . The dominated convergence theorem and the convergence of  $\{\lambda^k\}_k$  to  $\lambda$  w.r.t. the norm  $\|\cdot\|_2$  achieves the proof.  $\square$

We are now ready to prove Theorem 3.1. The results obtained in Lemma 3.4 for any  $\lambda \in L^\infty(0, T)$  can now be extended to any  $\lambda \in L^2(0, T)$ .

*Proof of Theorem 3.1.* Let  $\lambda \in L^2(0, T)$ . Since  $L^\infty(0, T)$  is dense in  $L^2(0, T)$ , one can extract a sequence  $\{\lambda^k\}_k$  in  $L^\infty(0, T)$  converging to  $\lambda \in L^2(0, T)$ .

By [24, Remark 4.5], for any  $k \in \mathbb{N}^*$ , the solution  $u$  of (3.7) is such that upper bounds on  $\|u\|_\infty$  and  $\|\partial_s u\|_\infty$  depend on the data of the problem and on  $\|\lambda^k\|_2$ . Since  $\{\lambda^k\}_k$  is converging w.r.t. the norm  $L^2(0, T)$ , the sequence  $\{\|\lambda^k\|_2\}_k$  is uniformly bounded. Thus, by the Lipschitz continuity of the map  $H'$ , guaranteed by Assumption 4, and by equality (3.9), the sequence  $\{\alpha[\lambda^k]\}_k$  is uniformly bounded and uniformly Lipschitz. By the Ascoli-Arzelà Theorem, there exists  $\hat{\alpha} \in \mathbb{A}$  such that  $\{\alpha[\lambda^k]\}_k$  uniformly converges, up to a subsequence, to  $\hat{\alpha} \in \mathbb{A}$  w.r.t. the norm  $\|\cdot\|_\infty$ . In addition, since  $\{\alpha[\lambda^k]\}_k$  is uniformly Lipschitz continuous,  $\hat{\alpha}$  is also Lipschitz continuous on  $[0, T] \times \mathcal{X}$ . Let  $\alpha \in \mathbb{A}$ , by the definition of  $\lambda \mapsto \alpha[\lambda]$ , one has, for any  $k \in \mathbb{N}$ ,

$$0 \leq \mathcal{L}_1(\alpha, \lambda^k) - \mathcal{L}_1(\alpha[\lambda^k], \lambda^k).$$

According to Lemma 3.5,  $(\alpha, \lambda) \mapsto \mathcal{L}_1(\alpha, \lambda)$  is continuous on  $\text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d) \times L^2(0, T)$ . Therefore by taking the limit  $k \rightarrow \infty$  in the previous inequality, one gets:

$$0 \leq \mathcal{L}_1(\alpha, \lambda) - \mathcal{L}_1(\hat{\alpha}, \lambda).$$

Thus,  $\hat{\alpha} \in \text{argmin}_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$ . Defining  $\alpha[\lambda] := \hat{\alpha}$  and using the uniform convergence of  $\{\alpha[\lambda^k]\}_k$  to  $\alpha[\lambda]$ , we deduce that  $\lambda \mapsto \alpha[\lambda]$  is continuous from  $L^2(0, T)$  to  $\mathbb{A}$ . Since  $L^\infty(0, T)$  is dense in  $L^2(0, T)$  and  $\lambda \mapsto \alpha[\lambda]$  is continuous in  $L^2(0, T)$ , the Lipschitz continuity of  $\lambda \mapsto \alpha[\lambda]$  on any bounded subset of  $L^2(0, T)$  is derived from Lemma 3.4.(iii).  $\square$

## 3.2 Proof of Theorem 3.12

Note that rewriting Problem (2.5) w.r.t. to the distribution of the states, one can prove Theorem 3.2 by using a change of variable, as it has been done in [24] for a similar problem. To avoid additional notations, we have preferred to prove Theorem 3.2 by means of Theorem 3.1.

The next lemma states that the map  $\lambda \mapsto \inf_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$  is Gâteaux differentiable in  $L^2(0, T)$ .

**Lemma 3.6.** *The map  $\lambda \mapsto \min_{\alpha \in \mathbb{A}} \mathcal{L}_1(\alpha, \lambda)$  admits a Gâteaux derivative  $\lambda \mapsto D\mathcal{L}_1(\alpha[\lambda], \lambda)(\cdot)$  in  $L^2(0, T)$ , such that, for any  $\mu \in L^2(0, T)$ :*

$$D\mathcal{L}_1(\alpha[\lambda], \lambda)(\mu) = \int_0^T \mathbb{E} \left[ p(\tau, X_\tau^{\alpha[\lambda]}) \right] \mu(\tau) d\tau.$$

*Proof.* For any  $\lambda \in L^2(0, T)$ , we recall that the definition of the Gâteaux derivative  $\lambda \mapsto D\mathcal{L}_1(\alpha[\lambda], \lambda)(\cdot)$  in the direction  $\mu \in L^2(0, T)$ , if it exists, is:

$$D\mathcal{L}_1(\alpha[\lambda], \lambda)(\mu) := \lim_{h \rightarrow 0} \frac{\mathcal{L}_1(\alpha[\lambda + h\mu], \lambda + h\mu) - \mathcal{L}_1(\alpha[\lambda], \lambda)}{h}.$$

For any  $h \in [0, T]$ , we have:

$$\begin{aligned} \int_0^T \mathbb{E} \left[ p(\tau, X_\tau^{\alpha[\lambda + h\mu]}) \right] \mu(\tau) d\tau &= \frac{\mathcal{L}_1(\alpha[\lambda + h\mu], \lambda + h\mu) - \mathcal{L}_1(\alpha[\lambda + h\mu], \lambda)}{h} \\ &\leq \frac{\mathcal{L}_1(\alpha[\lambda + h\mu], \lambda + h\mu) - \mathcal{L}_1(\alpha[\lambda], \lambda)}{h} \\ &\leq \frac{\mathcal{L}_1(\alpha[\lambda], \lambda + h\mu) - \mathcal{L}_1(\alpha[\lambda], \lambda)}{h} \\ &= \int_0^T \mathbb{E} \left[ p(\tau, X_\tau^{\alpha[\lambda]}) \right] \mu(\tau) d\tau. \end{aligned} \tag{3.10}$$

Since  $\lim_{h \rightarrow 0} \lambda + h\mu = \lambda$  in  $L^2(0, T)$ , it turns out from Theorem 3.1 that  $\lim_{h \rightarrow 0} \alpha[\lambda + h\mu] = \alpha[\lambda] \in \text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d)$  w.r.t. the norm  $\|\cdot\|_\infty$ . Since  $p \in C^1([0, T] \times \mathcal{X}, \mathbb{R})$ , one deduces by Lemma A.1 in Appendix A:

$$\begin{aligned} \lim_{h \rightarrow 0} \int_0^T \mathbb{E} \left[ p(\tau, X_\tau^{\alpha[\lambda + h\mu]}) \right] \mu(\tau) d\tau &= \lim_{h \rightarrow 0} \int_0^T \sum_{i \in I} \int_0^1 p(\tau, i, z) m^{\alpha[\lambda + h\mu]}(\tau, i, dz) \mu(\tau) d\tau \\ &= \int_0^T \sum_{i \in I} \int_0^1 p(\tau, i, z) m^{\alpha[\lambda]}(\tau, i, dz) \mu(\tau) d\tau \\ &= \int_0^T \mathbb{E} \left[ p(\tau, X_\tau^{\alpha[\lambda]}) \right] \mu(\tau) d\tau. \end{aligned}$$

The conclusion follows by taking the limit  $h \rightarrow 0$  in (3.10).  $\square$

We now turn to the proof of Theorem 3.2.

*Proof of Theorem 3.2.* According to Lemma 3.3(ii) and Lemma 3.6, one has that  $\mathcal{W}$  is Gâteaux differentiable in  $L^2(0, T)$ . For any  $\lambda \in L^2(0, T)$ , the Gâteaux differential of  $\mathcal{W}$  at  $\lambda$  is denoted by  $D\mathcal{W}(\lambda)(\cdot)$  and is given in the direction  $\mu \in L^2(0, T)$  by:

$$D\mathcal{W}(\lambda)(\mu) = \int_0^T \left( \mathbb{E} \left[ p(\tau, X_\tau^{\alpha[\lambda]}) \right] - v[\lambda](\tau) \right) \mu(\tau) d\tau. \tag{3.11}$$

By the definition of  $\bar{\lambda}$  given in Lemma 3.2, it results that  $0 \in \partial(-\mathcal{W}(\bar{\lambda}))$ . Since  $-\mathcal{W}$  is convex and Gâteaux differentiable, its sub-differential is reduced to a singleton and is equal to  $-D\mathcal{W}$ . One has, for any  $\mu \in L^2(0, T)$ , by the previous equality:

$$\int_0^T \left( \mathbb{E} \left[ p(\tau, X_\tau^{\alpha[\bar{\lambda}]}) \right] - v[\bar{\lambda}](\tau) \right) \mu(\tau) d\tau = \langle \partial(-\mathcal{W}(\bar{\lambda})), \mu \rangle = 0.$$

Therefore, one has for a.e.  $t \in [0, T]$ :

$$\mathbb{E} \left[ p(\tau, X_t^{\alpha[\bar{\lambda}]}) \right] - v[\bar{\lambda}](t) = 0.$$

By the previous equality, one has  $\mathcal{L}(\alpha[\bar{\lambda}], v[\bar{\lambda}], \bar{\lambda}) = \mathcal{L}(\alpha[\bar{\lambda}], v[\bar{\lambda}], \lambda)$  for any  $\lambda \in L^2(0, T)$  and, for any  $\alpha \in \mathbb{A}$  and  $v \in L^2(0, T)$ , satisfying  $\mathbb{E}[p(t, X_t^\alpha)] - v(t) = 0$  a.e on  $[0, T]$ , one has:

$$J(\alpha[\bar{\lambda}]) = \bar{J}(\alpha[\bar{\lambda}], v[\bar{\lambda}]) = \mathcal{L}(\alpha[\bar{\lambda}], v[\bar{\lambda}], \bar{\lambda}) \leq \mathcal{L}(\alpha, v, \bar{\lambda}) = \bar{J}(\alpha, v) = J(\alpha).$$

$\square$

## 4 Stochastic Uzawa Algorithm

While the existence of a unique solution  $\bar{\lambda}$  of the dual problem (3.4) is established in Lemma 3.2, we propose in this section an iterative algorithm converging to  $\bar{\lambda}$ . This algorithm is directly derived from the gradient descent in Hilbert space [13], and has been introduced in [25] to solve dual problems. This algorithm is an adaptation of the Uzawa Algorithm in stochastic optimization settings. The sub gradient of the opposite of the dual function  $\mathcal{W}$  is estimated by Monte Carlo simulations. Convexity assumptions introduced in [25] are not satisfied here.

We consider a sequence  $\{\rho_k\}_k$  satisfying  $\sum_{k=0}^{\infty} \rho_k = \infty$  and  $\sum_{k=0}^{\infty} \rho_k^2 < \infty$  to ensure the convergence. Typically, consider the sequence defined by  $\rho_k = \frac{a}{b+k}$  with  $a > 0$  and  $b > 0$  chosen empirically to accelerate the convergence.

---

### Algorithm 1 Stochastic Uzawa

---

- 1: Initialization  $\lambda^0 \in L^\infty(0, T)$ , set  $\{\rho_k\}$  and  $M \in \mathbb{N}^*$
  - 2:  $k \leftarrow 0$ .
  - 3: **for**  $k = 0, 1, \dots$  **do**
  - 4:    $v^k \leftarrow v[\lambda^k]$  where  $v[\lambda^k]$  is defined in Lemma 3.3.
  - 5:    $\alpha^k \leftarrow \alpha[\lambda^k]$  where  $\alpha[\lambda^k]$  is defined in Lemma 3.4.
  - 6:   Generate  $M$  independent states realizations  $(X^{1,\alpha^k}, \dots, X^{M,\alpha^k})$  independent of previous iterations simulations.
  - 7:    $U^{k+1} \leftarrow v^k - \frac{1}{M} \sum_{j=1}^M p(\cdot, X^{j,\alpha^k})$ .
  - 8:    $\lambda^{k+1} \leftarrow \lambda^k + \rho_k U^{k+1}$ .
- 

The next Lemma shows that it is possible at any iteration  $k$  to compute  $\alpha[\lambda^k]$  defined in line 5 of the Algorithm 1.

**Lemma 4.1.** *For any  $k \in \mathbb{N}$ ,  $\lambda^k \in L^\infty(0, T)$  a.s..*

*Proof.* The proof is done by induction. For  $k = 0$ ,  $\lambda^0$  is initialized as an element of  $L^\infty(0, T)$ . Assume now that  $\lambda^k \in L^\infty(0, T)$  for an integer  $k$ . Since  $p \in C^1([0, T] \times \mathcal{X}, \mathbb{R})$ , one has that a.s.  $\frac{1}{M} \sum_{j=1}^M p(\cdot, X^{j,\alpha^k}) \in L^\infty(0, T)$ . We show now that  $v[\lambda^k] \in L^\infty(0, T)$ . Since  $f$  is l.s.c., it is a normal integrand. By the exchange property [22, Theorem 14.60], one has:

$$\inf_{v \in L^2(0, T)} \int_0^T f(t, v(t)) - v(t)\lambda^k(t) dt = \int_0^T \inf_{v \in \mathbb{R}} f(t, v) - v\lambda^k(t) dt.$$

and for a.e.  $t \in [0, T]$ :

$$v[\lambda^k](t) \in \operatorname{argmin}_{v \in \mathbb{R}} f(t, v) - v\lambda^k(t). \quad (4.1)$$

For any  $t \in [0, T]$ ,  $f^*(t, \cdot)$  denotes the convex conjugate of  $f(t, \cdot)$ . Since for any  $t \in [0, T]$ ,  $f(t, \cdot)$  is strictly convex and by inequality (2.6) has at most quadratic growth,  $f^*(t, \cdot)$  is diferentiable, with linear growth [25, Lemma A.1]. Then, there exists a constant  $C > 0$ , such that for any  $t \in [0, T]$  one has:

$$|\nabla f^*(t, \lambda(t))| \leq C(|\lambda(t)| + 1),$$

where  $\nabla f^*(t, \cdot)$  is the derivative of  $z \mapsto f^*(t, z)$ . By (4.1), one has, for a.e.  $t \in [0, T]$ , that  $v[\lambda^k](t) = \nabla f^*(t, \lambda(t))$ . By previous inequality, one deduces that

$$\|v[\lambda^k]\|_\infty \leq C(\|\lambda^k\|_\infty + 1). \quad (4.2)$$

The conclusion follows from the definition of  $\lambda^{k+1}$  at line 8 of Algorithm 1, previous inequality and the fact that

$$\text{a.s. } \frac{1}{M} \sum_{j=1}^M p(\cdot, X^{j,\alpha^k}) \in L^\infty(0, T). \quad \square$$

The next Lemma provides an estimate on the norm of the gradient of  $\mathcal{W}$ , that is crucial to show the convergence of Algorithm 1.

**Lemma 4.2.** *There exists  $C > 0$  such that the sequence  $\{U^{k+1}\}_k$  generated by Algorithm 1 (line 7) satisfies, for any  $k \in \mathbb{N}$ ,*

$$\mathbb{E}[\|U^{k+1}\|_2] \leq C(\|\lambda^k\|_2 + 1).$$

*Proof.* By Assumption 2,  $p$  is in  $C^1([0, T] \times \mathcal{X}, \mathbb{R})$  thus, there exists  $C > 0$ , such that, for any  $k \in \mathbb{N}$  and  $j \in \{1, \dots, M\}$ :

$$\mathbb{E}[\|p(\cdot, X^{j, \alpha^k})\|_2] \leq C.$$

The conclusion follows from previous inequality, inequality (4.2) in the proof of Lemma 4.1 and the definition of  $U^{k+1}$  at line 7 of Algorithm 1.  $\square$

The convergence of Algorithm 1 is stated in the next lemma.

**Lemma 4.3.** *The sequence  $\{\lambda^k\}_k$ , generated by Algorithm 1, converges to  $\bar{\lambda}$  a.s. in  $L^2(0, T)$ .*

*Proof.* Since  $\mathcal{W}$  is convex and Gateau differentiable (proof of Theorem 3.2), by Lemma 4.2, the proof is a simple adaptation of the one in [13, Theorem 3.6], as it has been done in [25].  $\square$

Before to give the proof of the main result of this paper, we need the following regularity result on the function  $J$ .

**Lemma 4.4.** *The map  $J$  defined in (2.4) is continuous on  $\text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d)$ .*

*Proof.* By the definition of  $m^\alpha$ , the map  $J$  can be rewritten as:

$$\begin{aligned} J(\alpha) = & \int_0^T f\left(t, \sum_{i \in I} \int_0^1 p(t, i, z) m^\alpha(t, i, dz)\right) dt + \int_0^T \int_0^1 \sum_{i \in I} c(t, i, z) m^\alpha(t, i, dz) dt \\ & + \int_0^T \int_0^1 \sum_{i, j \in I} L(\alpha_j(t, i, z)) m^\alpha(t, i, dz) dt + \int_0^1 \sum_{i \in I} g(i, z) m^\alpha(T, i, dz). \end{aligned}$$

By the regularity of the functions  $p, c, g$  and  $L$  in Assumptions 2-4 and the continuity of the map  $\alpha \mapsto m^\alpha$  on  $\text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d)$  stated in Lemma A.1 in Appendix A, one obtains the continuity of  $J$  over  $\text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d)$  w.r.t. the norm  $\|\cdot\|_\infty$ .  $\square$

We can now prove Theorem 2.1.

*Proof of Theorem 2.1.* This is a direct consequence of Theorem 3.2.

(i) This is a direct consequence of Lemma 4.3 and Theorem 3.1.

(ii) This is a direct consequence of Theorem 2.1.(i) and Lemma 4.4.  $\square$

## 5 Simulations

This section illustrates the results with an example of smart charging. In this use case, the discrete variable taking values in  $I$  represents the charging mode and the continuous variable taking value in  $[0, 1]$  represents the state of charge (SoC) of electrical vehicles (EV) relatively to the maximum energy capacity of the battery.

### 5.1 Definition of the use Case

We consider a large fleet of EVs controlled by a central planner during their charging period  $[0, T]$ . The goal of the central planner is to provide ancillary services to the transmission grid by controlling the aggregate consumption profile of the fleet on the time horizon  $[0, T]$ . More specifically, the central planner aims at making the consumption profile of the fleet to be close to a given profile  $r = (r_t)_{0 \leq t \leq T}$ , supposed to be known on the whole period  $[0, T]$ . In our simulation we compare two situations. In the first case known as V1G, it is assumed that the vehicle batteries can only draw electricity from the grid, then the set of charging modes is  $I = \{0, 1\}$ , where 0 stands for idle mode and 1 for charging. In the second case known as V2G, it is assumed that the batteries can either draw or inject electricity into the grid, then the charging modes are  $I = \{-1, 0, 1\}$ , where  $-1$  corresponds to injection mode. In each situation, each mode  $i \in I$  is characterized by its charging rate  $b(i, \cdot)$  such that

- $i = -1$  corresponds to a V2G mode with  $b(-1, z) = -a$  for  $z \geq 0.25$  and  $b(-1, z) = -az$  for  $z \leq 0.25$ , with  $a = \frac{2}{5}$ .
- $i = 0$  corresponds to a non-charging mode with  $b(0, z) = 0$ ,
- $i = 1$  corresponds to a charging mode with  $b(1, z) = a$  for  $z \leq 0.75$  and  $b(1, z) = a(1 - z)$  for  $z \geq 0.75$ , with  $a = \frac{2}{5}$ ,

The charging rate  $b$  of the batteries has been designed based on industrial data [12]. The state of each EV, controlled by  $\alpha$ , is represented by a PDMP  $X^\alpha = (Y^\alpha, Z^\alpha)$  with  $Y_t^\alpha$ , the charging mode and  $Z_t^\alpha$ , the SoC of the vehicle at time  $t$ . We assume that the power consumption of a vehicle at state  $(i, z) \in \mathcal{X}$  is equal to its instantaneous charging rate  $b(i, z)$ . At time  $t = 0$ ,  $Y_0^\alpha$  equals 0 or 1 with probability 1/2, while  $Z_0^\alpha = \min(1, |z_0| + 0.15)$  where  $z_0 \sim \mathcal{N}(0, 0.2)$ .

To ensure customer satisfaction, the final cost  $g(i, z) := 30 \times (1 - e^{5(z-0.75)})^+$  penalizes vehicles with a final SoC  $z$  lower than 75%. As previously stated, high values of  $\alpha$  are penalized through the cost  $\alpha^2/2$  in order to avoid high frequencies of jumps damaging the batteries. For any  $t \in [0, T]$ ,

$$\begin{cases} c(t, i, z) & = & 0 \\ p(t, i, z) & := & b(i, z) \\ l(a) & := & \frac{a^2}{2} \\ F(v) & := & \kappa \int_0^T (v(t) - r(t))^2 dt \end{cases}$$

with  $\kappa = 1000$  in order to ensure that the overall consumption is close to the profile  $r$ .

The target profile  $r$  is a slight modification of the nominal behaviour of the fleet of EVs (without control), with the same energy consumed over the fixed time horizon  $[0, T]$ . The nominal behaviour corresponds to the situation where the EVs are not required to fit the target profile and seek only to satisfy their own comfort (namely  $F(v) = 0$ ). The idea is to define a realistic target profile so that it is possible for the fleet to follow this profile while satisfying their charging needs. We consider a population of  $N = 10^5$  EVs. The optimal control of Problem (2.5) is computed using Algorithm 1.

## 5.2 Algorithm parameters

For the implementation of the algorithm, the line 5 is computed by discretization of the Hamilton Jacobi-Bellman-Equation III.5 associated to  $\alpha[\lambda]$  defined in III.6 with  $N_t = 1000$  regular points of time and  $N_s = 400$  regular points of SoC. The Lagrangian multipliers  $\lambda^k$  are obtained after  $k = 1000$  iterations and with  $M = 1000$  realizations and are displayed in Figure 4. The initial multiplier is  $\lambda^0 = 0$  and the stepsize sequence is such that  $\rho_k = \min(30, \frac{500}{1+k})$ . For each multiplier  $\lambda$ , the associated optimal strategy  $\alpha[\lambda]$  is computed.

## 5.3 Results

Four periods on Figure 3 are distinguished, depending on whether the target profile is above or below the nominal consumption. The reference profile is displayed in green. EVs are encouraged to consume more than the nominal consumption during periods P1 and P3, and to consume less during periods P2 and P4. First, we observe that both V1G (blue line) and V2G (red line) consumptions are close to the profile  $r$  (green line). The main difference lies in P4, where the possibility for the EVs to perform V2G allows the fleet consumption to stay closer to the target profile. As expected, the profile is better tracked with the V2G mode, due to the additional degree of liberty. On Figure 6 one can observe that V2G is used during periods P2 and P4, when the profile  $r$  is lower than the nominal consumption. Ten trajectories of SoC, generated in the V2G use case, are represented in Figure 1. The green line is an example of a PDMP switching from mode 0 to mode 1 around  $t = 0.2h$ , from mode 1 to mode -1 around  $t = 2.6h$  and then from mode -1 to mode 1 around  $t = 4h$ . The initial and final distribution of the SoC of the fleet are displayed on Figures 7 and 8. While the initial distribution is the same for both scenarios, one can observe that their final distribution are very closed, and that very few EVs have a SoC lower than 0.75% in both cases. Thus, the comfort of each agent is weakly impacted by the V2G mode.

To conclude, the interest of adding V2G is confirmed by the comparison of the optimal cost of each scenario (Figure 2). These optimal costs are decomposed into three parts: the switching cost  $(\frac{1}{2} \int_0^T \sum_{j \neq I} \alpha_j(t, X_t^{i, \alpha})^2 dt)$ , the

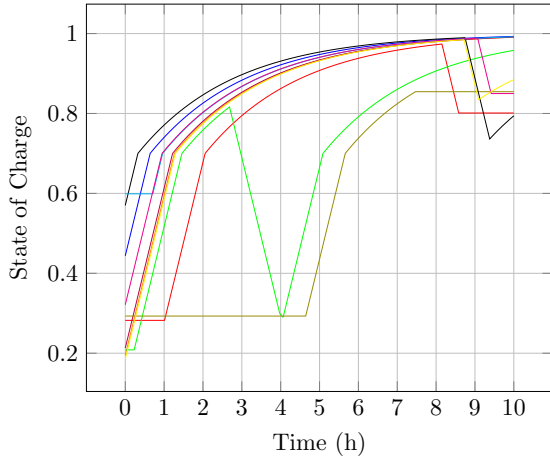


Figure 1: Representation of the SoC of 10 PDMP

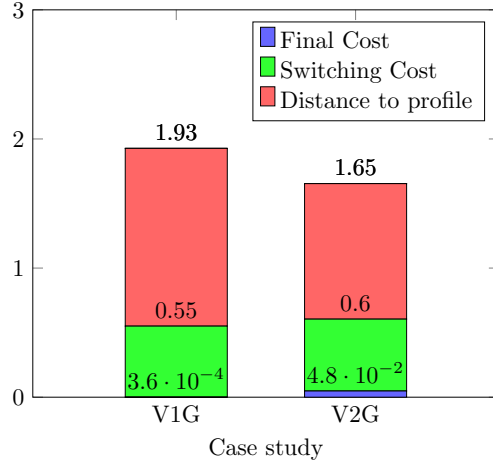


Figure 2: Decomposition of the cost function

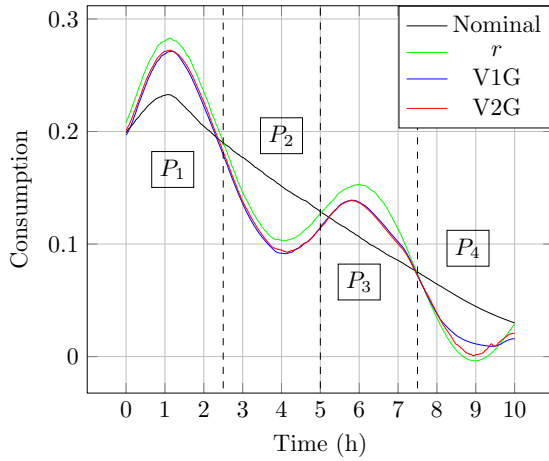


Figure 3: Consumption for V1G and V2G cases compared to the profile and nominal

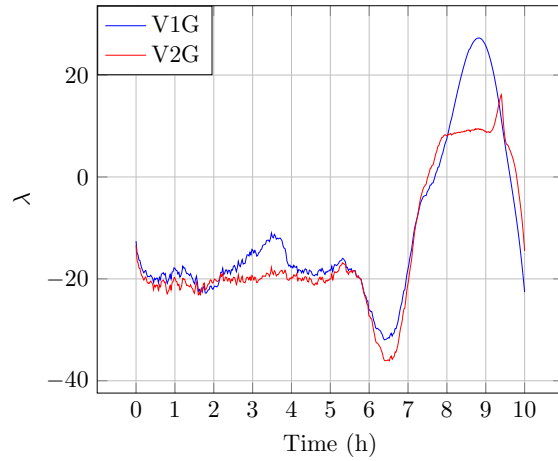


Figure 4: Optimal Lagrangian multiplier  $\lambda$  obtained

final cost ( $g$ ) and the distance to profile ( $F$ ). The low value of the final cost ( $3.6 \times 10^{-4}$  in the V1G use case and  $4.8 \times 10^{-2}$  in the V2G one) shows that, despite the tracking objective, EVs have enough battery at the end of the period, as previously stated. Moreover, this performance is achieved without significant loss in switching costs, thus preserving fleet batteries.

## A Appendix

In this section, we provide some classical results about the distribution  $m^\alpha$  of a process  $X^\alpha$  that is a PDMP( $b, \alpha$ ) (the definition of a PDMP( $b, \alpha$ ) is given in Section 2), when  $\alpha$  is in  $\text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d)$ . We introduce the continuity equation on  $[0, T] \times \mathcal{X}$ :

$$\begin{aligned} \partial_t m(t, i, z) + \partial_z (b(i, z) m(t, i, z)) &= - \sum_{j \neq i} (\alpha_j(t, i, z) m(t, i, z) - \alpha_i(t, j, z) m(t, j, z)) \quad (t, i, z) \in (0, T) \times \mathcal{X}, \\ m_i(0, z) &= m_i^0(z) \quad (i, z) \in \mathcal{X}, \end{aligned} \quad (\text{A.1})$$

where  $m^0 \in \mathcal{P}(\mathcal{X})$  is given. The next definition gives the characterization of weak solutions of (A.1).

**Definition A.1.** A pair  $(m, \alpha)$  satisfies (A.1) in the weak sense if  $t \in [0, T] \mapsto m(t, \cdot) \in \mathcal{P}(\mathcal{X})$  is continuous, for

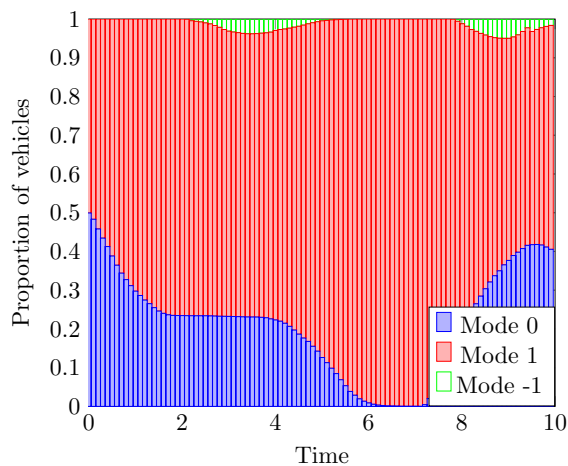
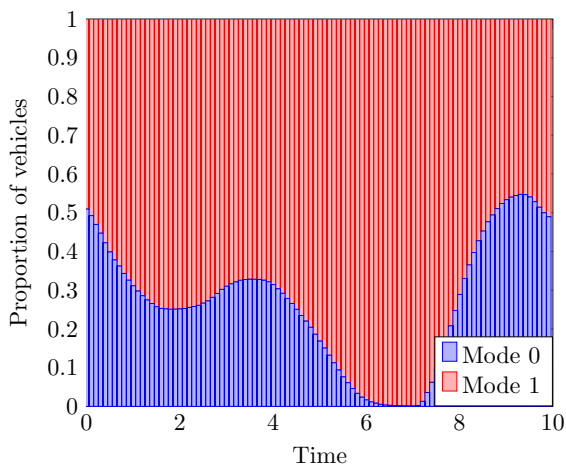


Figure 5: Evolution of the proportion of vehicles for V1G Figure 6: Evolution of the proportion of vehicles for V2G

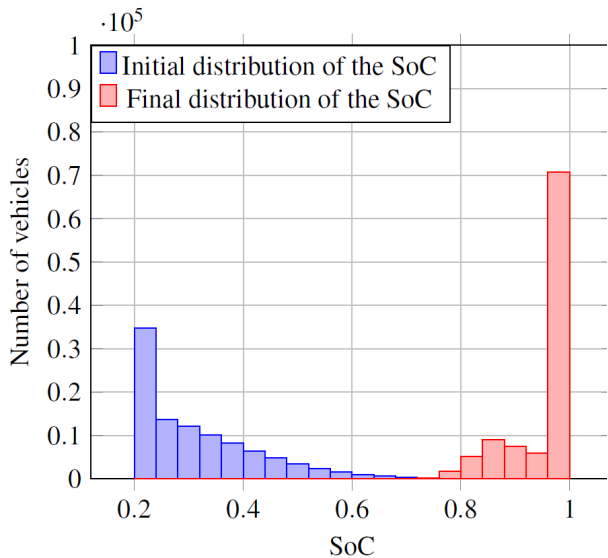


Figure 7: Initial and Final SoC in the V1G case

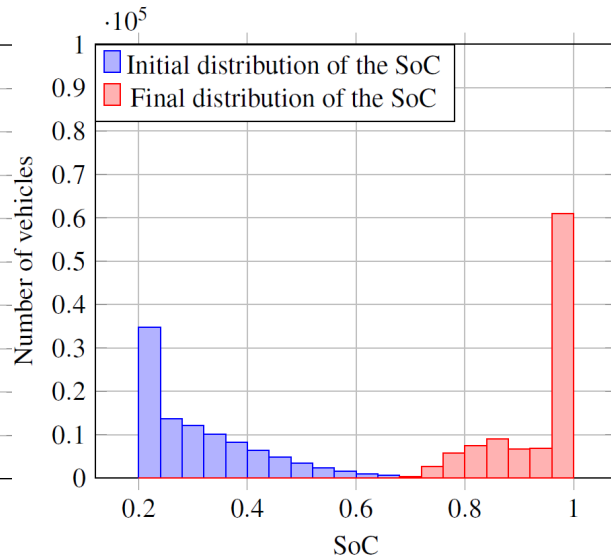


Figure 8: Initial and Final SoC in the V2G case

any  $i, j \in I$  with  $i \neq j$ , it holds that  $\alpha \in L_{m_i}^2([0, T] \times \mathcal{X}, \mathbb{R}^d)$  and for any test function  $\phi \in C_c^\infty([0, T] \times \mathcal{X})$ , we have:

$$\begin{aligned} & \sum_{i \in I} \int_0^1 \phi_i(T, i, z) m(T, i, dz) - \phi(0, i, z) m_i^0(dz) \\ &= \int_0^T \int_0^1 \sum_{i \in I} (\partial_t \phi(t, i, z) + b(i, z) \partial_z \phi(t, i, z)) m(t, i, dz) + \sum_{j \in I, j \neq i} (\phi(t, j, z) - \phi(t, i, z)) \alpha_j(t, i, z) m(t, i, dz) dt. \end{aligned}$$

**Proposition A.1.** *Let  $m^0 \in \mathcal{P}([0, 1] \times I)$ ,  $\alpha \in \text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d)$  and  $X^\alpha$  be a PDMP( $b, \alpha$ ). Then, the distribution  $m^\alpha$  of  $X^\alpha$  is the unique distribution such that  $(m^\alpha, \alpha)$  is a weak solution of (A.1) in the sense of Definition A.1.*

*Proof.* This result is proved in [6] for controls  $\alpha$  that are continuous in space and time independent. The extension of this result to bounded controls that are measurable in time is straightforward.  $\square$

Continuity results of the map  $\alpha \mapsto m^\alpha$  is presented in the next lemma.

**Lemma A.1.** *The map  $\alpha \mapsto m^\alpha$  is continuous from  $\text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d)$  endowed with the norm  $\|\cdot\|_\infty$  to  $C([0, T], \mathcal{P}(\mathcal{X}))$ , endowed with the distance  $\mathbb{W}_{\|\cdot\|_\infty}$ .*

*Proof.* Let  $\psi \in C^1(\mathcal{X})$  and let  $\varphi \in C^1([0, T] \times \mathcal{X})$  be the classical solution of the PDE:

$$\begin{aligned} \partial_\tau \varphi(\tau, i, z) + b(i, z) \partial_s \varphi(\tau, i, z) &= 0 & (\tau, i, z) \in [0, t] \times \mathcal{X} \\ \varphi(t, i, z) &= \psi(i, z) & (i, z) \in \times \mathcal{X}. \end{aligned} \quad (\text{A.2})$$

One has  $\|\varphi\|_\infty = \|\psi\|_\infty$ , and denoting by  $\gamma_\psi$  the Lipschitz constant of  $\psi$ , for any  $t \in [0, T]$   $\varphi(t, \cdot, \cdot)$  is Lipschitz continuous with Lipschitz constant  $\gamma_\psi e^{T\|b'\|_\infty}$ .

Let  $\alpha, \bar{\alpha} \in \text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d)$ . Since  $X^\alpha$  is a PDMP( $b, \alpha$ ) and  $X^{\bar{\alpha}}$  is a PDMP( $b, \bar{\alpha}$ ), by Proposition A.1,  $(m^\alpha, \alpha)$  and  $(m^{\bar{\alpha}}, \bar{\alpha})$  are both weak solutions of the continuity equation (A.1) on  $[0, T] \times I \times [0, 1]$ . Then, by equality (A.2), one has:

$$\begin{aligned} & \int_0^1 \sum_{i \in I} \psi(i, z) (m^\alpha(t, i, dz) - m^{\bar{\alpha}}(t, i, dz)) \\ &= \sum_{i \in I} \int_0^t \int_0^1 \sum_{j \in I, j \neq i} \varphi(i, \tau, z) \left( -\alpha_j(\tau, i, z) m^\alpha(\tau, i, dz) + \alpha_i(\tau, j, z) m^\alpha(\tau, j, dz) \right) d\tau \\ & \quad + \sum_{i \in I, j \neq i} \int_0^t \int_0^1 \sum_{j \in I} \varphi(i, \tau, z) \left( \bar{\alpha}_j(\tau, i, z) m^{\bar{\alpha}}(\tau, i, dz) - \bar{\alpha}_i(\tau, j, z) m^{\bar{\alpha}}(\tau, j, dz) \right) d\tau \\ &= \sum_{i \in I, j \neq i} \int_0^t \int_0^1 \sum_{j \in I} \varphi(i, \tau, z) \alpha_j(\tau, i, z) \left( m^{\bar{\alpha}}(\tau, i, dz) - m^\alpha(\tau, i, dz) \right) d\tau \\ & \quad + \sum_{i \in I} \int_0^t \int_0^1 \sum_{j \in I, j \neq i} \varphi(i, \tau, z) \alpha_j(\tau, i, z) \left( m^\alpha(\tau, j, dz) - m^{\bar{\alpha}}(\tau, j, dz) \right) d\tau \\ & \quad + \sum_{i \in I} \int_0^t \int_0^1 \sum_{j \in I, j \neq i} \varphi(i, \tau, z) \left( \bar{\alpha}_j(\tau, i, z) - \alpha_j(\tau, i, z) \right) m^{\bar{\alpha}}(\tau, i, dz) d\tau \\ & \quad + \sum_{i \in I} \int_0^t \int_0^1 \sum_{j \in I, j \neq i} \varphi(i, \tau, z) \left( \alpha_i(\tau, j, z) - \bar{\alpha}_i(\tau, j, z) \right) m^{\bar{\alpha}}(\tau, j, dz) d\tau \end{aligned} \quad (\text{A.3})$$

Using that  $\varphi(t, \cdot, \cdot)$  is Lipschitz continuous with Lipschitz constant  $\gamma_\psi e^{T\|b'\|_\infty}$  and that  $\alpha \in \text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d)$ , there exists  $L > 0$  such that:

$$\sum_{i \in I} \int_0^t \int_0^1 \sum_{j \in I, j \neq i} \varphi(i, \tau, z) \alpha_j(\tau, i, z) \left( m^{\bar{\alpha}}(\tau, i, dz) - m^\alpha(\tau, i, dz) \right) d\tau \leq L \gamma_\psi e^{T\|b'\|_\infty} \int_0^T \mathcal{W}(m^\alpha(\tau), m^{\bar{\alpha}}(\tau)) d\tau, \quad (\text{A.4})$$

$$\sum_{i \in I} \int_0^t \int_0^1 \sum_{j \in I, j \neq i} \varphi(i, \tau, z) \alpha_j(\tau, i, z) \left( m^\alpha(\tau, j, dz) - m^{\bar{\alpha}}(\tau, j, dz) \right) d\tau \leq L \gamma_\psi e^{T\|b'\|_\infty} \int_0^T \mathcal{W}(m^\alpha(\tau), m^{\bar{\alpha}}(\tau)) d\tau. \quad (\text{A.5})$$



Since  $\alpha, \bar{\alpha} \in \text{Lip}([0, T] \times \mathcal{X}, \mathbb{R}_+^d)$  and that  $\|\varphi\|_\infty \leq \|\psi\|_\infty$ , one has:

$$\sum_{i \in I} \int_0^t \int_0^1 \sum_{j \in I, j \neq i} \varphi(i, \tau, z) \left( \bar{\alpha}_j(\tau, i, z) - \alpha_j(\tau, i, z) \right) m^{\bar{\alpha}}(\tau, i, dz) d\tau \leq |I| \|\psi\|_\infty \|\alpha - \bar{\alpha}\|_\infty, \quad (\text{A.6})$$

$$\sum_{i \in I} \int_0^t \int_0^1 \sum_{j \in I, j \neq i} \varphi(i, \tau, z) \left( \alpha_i(\tau, j, z) - \bar{\alpha}_i(\tau, j, z) \right) m^{\bar{\alpha}}(\tau, j, dz) d\tau \leq |I| \|\psi\|_\infty \|\alpha - \bar{\alpha}\|_\infty. \quad (\text{A.7})$$

By inequalities (A.4)-(A.7), inequality (A.3) becomes:

$$\int_0^1 \sum_{i \in I} \psi(i, z) (m^\alpha(t, i, dz) - m^{\bar{\alpha}}(t, i, dz)) \leq 2|I| \|\psi\|_\infty \|\alpha - \bar{\alpha}\|_\infty + 2L\gamma_\psi e^{T\|b'\|_\infty} \int_0^t \mathcal{W}(m^\alpha(\tau), m^{\bar{\alpha}}(\tau)) d\tau.$$

Since  $\psi$  is arbitrary and in  $C^1(\mathcal{X})$ , previous inequality becomes, for any  $t \in [0, T]$ ,

$$\mathbb{W}(m^\alpha(t), m^{\bar{\alpha}}(t)) \leq 2|I|^2 \|\psi\|_\infty \|\alpha - \bar{\alpha}\|_\infty + 2Le^{T\|b'\|_\infty} \int_0^t \mathcal{W}(m^\alpha(\tau), m^{\bar{\alpha}}(\tau)) d\tau.$$

By Gronwall inequality, one deduces that there exists a constant  $C > 0$  such that, for any  $t \in [0, T]$ ,

$$\mathbb{W}(m^\alpha(t), m^{\bar{\alpha}}(t)) \leq L \|\alpha - \bar{\alpha}\|_\infty.$$

Thus,  $\mathbb{W}_{\|\cdot\|_\infty}(m^\alpha, m^{\bar{\alpha}}) \leq L \|\alpha - \bar{\alpha}\|_\infty$  and the continuity of  $\alpha \mapsto m^\alpha$  follows.  $\square$

## References

- [1] Nicole Bäuerle and Ulrich Rieder. Optimal control of Piecewise Deterministic Markov Processes with finite time horizon. *Modern trends in controlled stochastic processes: theory and applications*, pages 123–143, 2010.
- [2] Ana Bušić and Sean Meyn. Distributed control of thermostatically controlled loads: Kullback-Leibler optimal control in continuous time. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 7258–7265. IEEE, 2019.
- [3] René Carmona and François Delarue. Forward–backward stochastic differential equations and controlled McKean–Vlasov dynamics. *The Annals of Probability*, 43(5):2647–2700, 2015.
- [4] Guido Carnevale, Andrea Camisa, and Giuseppe Notarstefano. Distributed online aggregative optimization for dynamic multi-robot coordination. *IEEE Transactions on Automatic Control*, 2022.
- [5] Pierre Carpentier, J-Ph Chancelier, Vincent Leclère, and François Pacaud. Stochastic decomposition applied to large-scale hydro valleys management. *European Journal of Operational Research*, 270(3):1086–1098, 2018.
- [6] Christiane Coccozza-Thivent, Robert Eymard, Sophie Mercier, and Michel Roussignol. Characterization of the marginal distributions of Markov processes used in dynamic reliability. *Journal of Applied Mathematics and Stochastic Analysis*, 2006, 2006.
- [7] Oswaldo LV Costa, François Dufour, and Alexei B Piunovskiy. Constrained and unconstrained optimal discounted control of Piecewise Deterministic Markov Processes. *SIAM Journal on Control and Optimization*, 54(3):1444–1474, 2016.
- [8] Mark HA Davis. Piecewise-Deterministic Markov Processes: A general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3):353–376, 1984.
- [9] Benoîte De Saporta, François Dufour, and Alizée Geeraert. Optimal strategies for impulse control of Piecewise Deterministic Markov Processes. *Automatica*, 77:219–229, 2017.
- [10] Benoîte De Saporta and Huilong Zhang. Predictive maintenance for the heated hold-up tank. *Reliability Engineering & System Safety*, 115:82–90, 2013.

- [11] Alain Durmus, Arnaud Guillin, and Pierre Monmarché. Piecewise Deterministic Markov Processes and their invariant measures. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, volume 57, pages 1442–1475. Institut Henri Poincaré, 2021.
- [12] Fastned Support. Qu'est-ce qui détermine la vitesse de recharge? <https://support.fastned.nl/hc/fr/articles/205694717-Qu'est-ce-qui-d%C3%A9termine-la-vitesse-de-recharge->, 2020.
- [13] Caroline Geiersbach and Georg Ch Pflug. Projected stochastic gradients for convex constrained problems in Hilbert spaces. *SIAM Journal on Optimization*, 29(3):2079–2099, 2019.
- [14] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Fundamentals of convex analysis*. Springer Science & Business Media, 2004.
- [15] Yonghui Huang and Xianping Guo. Finite-horizon Piecewise Deterministic Markov decision processes with unbounded transition rates. *Stochastics*, 91(1):67–95, 2019.
- [16] Paulin Jacquot, Olivier Beaude, Pascal Benchimol, Stéphane Gaubert, and Nadia Oudjane. A privacy-preserving disaggregation algorithm for non-intrusive management of flexible energy. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, pages 890–896. IEEE, 2019.
- [17] Caroline Le Floch, Emre Can Kara, and Scott Moura. PDE modeling and control of electric vehicle fleets for ancillary services: A discrete charging case. *IEEE Transactions on Smart Grid*, 9(2):573–581, 2016.
- [18] Man-Fai Leung, Jun Wang, and Duan Li. Decentralized robust portfolio optimization based on cooperative-competitive multiagent systems. *IEEE Transactions on Cybernetics*, 2021.
- [19] Yen Ting Lin and Nicolas E Buchler. Efficient analysis of stochastic gene dynamics in the non-adiabatic regime using Piecewise Deterministic Markov Processes. *Journal of The Royal Society Interface*, 15(138):20170804, 2018.
- [20] Ya-Feng Liu and Yu-Hong Dai. On the complexity of joint subcarrier and power allocation for multi-user OFDMA systems. *IEEE transactions on Signal Processing*, 62(3):583–596, 2013.
- [21] Mario VF Pereira and Leontina MVG Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical programming*, 52(1):359–375, 1991.
- [22] R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009.
- [23] Ryszard Rudnicki and Marta Tyran-Kamińska. Piecewise Deterministic Markov Processes in biological models. In *Semigroups of operators-theory and applications*, pages 235–255. Springer, 2015.
- [24] Adrien Seguret. An optimal control problem for the continuity equation arising in smart charging. arXiv preprint arXiv:2109.12836, October 2022.
- [25] Adrien Seguret, Clémence Alasseur, J Frédéric Bonnans, Antonio De Paola, Nadia Oudjane, and Vincenzo Trovato. Decomposition of high dimensional aggregative stochastic control problems. *arXiv preprint arXiv:2008.09827*, 2020.
- [26] Adrien Séguret, Cheng Wan, and Clémence Alasseur. A mean field control approach for smart charging with aggregate power demand constraints. In *2021 IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, pages 01–05. IEEE, 2021.
- [27] D Vermis. Optimal control of Piecewise Deterministic Markov process. *Stochastics: An International Journal of Probability and Stochastic Processes*, 14(3):165–207, 1985.
- [28] Huilong Zhang, Benoîte De Saporta, François Dufour, and Gilles Deleuze. Dynamic reliability by using simulink and stateflow. *Chemical Engineering Transactions*, 33:529–534, 2013.
- [29] Huilong Zhang, Fares Innal, François Dufour, and Yves Dutuit. Piecewise Deterministic Markov Processes based approach applied to an offshore oil production system. *Reliability Engineering & System Safety*, 126:126–134, 2014.