



HAL
open science

An entropy penalized approach for stochastic control problems. Complete version

Thibaut Bourdais, Nadia Oudjane, Francesco Russo

► **To cite this version:**

Thibaut Bourdais, Nadia Oudjane, Francesco Russo. An entropy penalized approach for stochastic control problems. Complete version. *SIAM Journal on Control and Optimization*, 2026, 64 (1), pp.363-386. <10.1137/25M1741364>. <hal-04193113v3>

HAL Id: hal-04193113

<https://hal.science/hal-04193113v3>

Submitted on 28 Jul 2025

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.



HAL Authorization

An entropy penalized approach for stochastic control problems. Complete version.

THIBAUT BOURDAIS ^{*}, NADIA OUDJANE [†] AND FRANCESCO RUSSO [‡]

July 2025

Abstract

In this paper, we propose an original approach to stochastic control problems. We consider a weak formulation that is written as an optimization (minimization) problem on the space of probability measures. We then introduce a penalized version of this problem obtained by splitting the minimization variables and penalizing the discrepancy between the two variables via an entropy term. We show that the penalized problem provides a good approximation of the original problem when the weight of the entropy penalization term is large enough. Moreover, the penalized problem has the advantage of giving rise to two optimization subproblems that are easy to solve in each of the two optimization variables when the other is fixed. We take advantage of this property to propose an alternating optimization procedure that converges to the infimum of the penalized problem with a rate $O(1/k)$, where k is the number of iterations. The relevance of this approach is illustrated by solving a high-dimensional stochastic control problem aimed at controlling consumption in electrical systems.

Key words and phrases: Stochastic control; optimization; Donsker-Varadhan representation; exponential twist; relative entropy; demand-side management.

2020 AMS-classification: 49M99; 49J99; 60H10; 60J60; 65C05.

1 Introduction

General framework. Stochastic control problems appear in many fields of application such as robotics [41], economics and finance [43]. These problems are, either tackled using the Pontryagin's optimality principle or the dynamic programming principle, which allows the representation of the

^{*}ENSTA Paris, Institut Polytechnique de Paris. Unité de Mathématiques Appliquées (UMA). E-mail: thibaut.bourdais@ensta-paris.fr

[†]EDF R&D, and FiME (Laboratoire de Finance des Marchés de l'Energie (Dauphine, CREST, EDF R&D) www.fime-lab.org). E-mail:nadia.oudjane@edf.fr

[‡]ENSTA Paris, Institut Polytechnique de Paris. Unité de Mathématiques Appliquées (UMA). E-mail:francesco.russo@ensta-paris.fr.

value function via nonlinear Hamilton-Jacobi-Bellman PDEs or Backward Stochastic Differential Equations (BSDEs). The idea of this paper is to propose a radically different approach based on a weak reformulation of the stochastic control problem as an optimization problem on the space of probability measures. We propose an entropic penalization of this optimization problem which suitably approximates the original control problem. We prove the convergence of an alternating optimization procedure to the infimum of the penalized problem: the interest of this procedure is demonstrated in simulation compared with classical techniques relying on dynamic programming. The proof of the convergence of our algorithm relies on geometric arguments rather than classical convex optimization techniques.

Problem formulation. On some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we are interested in a problem of the type

$$J_{strong}^* := \inf_{\nu} \mathbb{E} \left[\int_0^T f(r, X_r^\nu, \nu_r) dr + g(X_T^\nu) \right], \quad (1.1)$$

where ν is a progressively measurable process taking values in some fixed convex domain $\mathbb{U} \subset \mathbb{R}^d$. $X = X^\nu$ will be a controlled diffusion process taking values in \mathbb{R}^d of the form

$$X_t^\nu = x + \int_0^t b(r, X_r^\nu, \nu_r) dr + \int_0^t \sigma(r, X_r^\nu) dW_r. \quad (1.2)$$

Under some mild supplementary assumptions, Problem (1.1) can be reformulated as an optimization program on the space of probability measures, in the form

$$J^* := \inf_{\mathbb{P} \in \mathcal{P}_{\mathbb{U}}} J(\mathbb{P}), \quad \text{with} \quad J(\mathbb{P}) := \mathbb{E}^{\mathbb{P}} \left[\int_0^T f(r, X_r, \nu_r^{\mathbb{P}}) dr + g(X_T) \right], \quad (1.3)$$

$\mathcal{P}_{\mathbb{U}}$ being a set of probability measures, defined in Definition 3.2, such that, under $\mathbb{P} \in \mathcal{P}_{\mathbb{U}}$, the canonical process X is decomposed as

$$X_t = x + \int_0^t b(r, X_r, \nu_r^{\mathbb{P}}) dr + \int_0^t \sigma(r, X_r) dW_r, \quad (1.4)$$

where $\nu^{\mathbb{P}}$ is a progressively measurable process with respect to the canonical filtration \mathcal{F}^X of X taking values in \mathbb{U} and W is some standard Brownian motion. In particular we will have $J_{strong}^* = J^*$. In the sequel, to insist on the path-dependence of ν , we will write $\nu_t = \nu(t, X)$. We refer to Appendix D for the precise link between the different formulations of stochastic control problems (1.1) and (1.3).

One major difficulty in analyzing Problem (1.3) is the lack of convexity of the functional J in (1.3) with respect to \mathbb{P} , even though the literature includes some techniques to transform the original problem into a minimization of a convex functional, see e.g. [2]. For that reason, we cannot rely on classical convex analysis techniques, see e.g. [17], in order to perform related algorithms, see e.g. [6]. As announced above, our method consists in replacing Problem (1.3) with the penalized version

$$\mathcal{J}_\epsilon^* := \inf_{(\mathbb{P}, \mathbb{Q}) \in \mathcal{A}} \mathcal{J}_\epsilon(\mathbb{Q}, \mathbb{P}), \quad \text{with} \quad \mathcal{J}_\epsilon(\mathbb{Q}, \mathbb{P}) := \mathbb{E}^{\mathbb{Q}} \left[\int_0^T f(r, X_r, \nu_r^{\mathbb{P}}) dr + g(X_T) \right] + \frac{1}{\epsilon} H(\mathbb{Q}|\mathbb{P}), \quad (1.5)$$

where \mathcal{A} is a subset of elements $(\mathbb{P}, \mathbb{Q}) \in \mathcal{P}(\Omega)^2$ defined in Definition 3.12, H is the relative entropy, see Definition 2.1, and the penalization parameter $\epsilon > 0$ is intended to vanish to zero, in order to impose $\mathbb{Q} = \mathbb{P}$.

Main contributions. In Theorem 3.15 one shows that the infimum in (1.5) is indeed a minimum $\mathcal{J}_\epsilon^* = \mathcal{J}_\epsilon(\mathbb{Q}_\epsilon^*, \mathbb{P}_\epsilon^*)$, attained on some admissible couple of probability measures $(\mathbb{P}_\epsilon^*, \mathbb{Q}_\epsilon^*) \in \mathcal{A}$. Given one solution $(\mathbb{P}_\epsilon^*, \mathbb{Q}_\epsilon^*)$ of Problem (1.5), Proposition 3.16 shows that \mathbb{P}_ϵ^* is an approximate solution of Problem (1.3), in the sense that the infimum J^* can indeed be approached by $J(\mathbb{P}_\epsilon^*)$ where $\mathbb{P}_\epsilon^* \in \mathcal{P}_\mathbb{U}$ when $\epsilon \rightarrow 0$, and more precisely $J(\mathbb{P}_\epsilon^*) - J^* = O(\epsilon)$. The interest of the penalized Problem (1.5), with respect to the original Problem (1.3), is that the minimization of the functional \mathcal{J}_ϵ , with respect to one variable \mathbb{Q} or \mathbb{P} (the other variable being fixed), can be provided quasi-explicitly. This is the object of Section 5. Indeed, Proposition 5.2 states that the minimization with respect to \mathbb{P} can be reduced to a pointwise minimization, provided that \mathbb{Q} has a Markovian decomposition. In this situation, there exists a function $(t, x) \mapsto u(t, x) \in \mathbb{U}$, such that, for all $(t, x) \in [0, T] \times \mathbb{R}^d$, $u(t, x)$ is independently obtained as the minimum of a strictly convex function and such that the infimum of the minimization $\inf_{\mathbb{P} \in \mathcal{P}_\mathbb{U}} \mathcal{J}_\epsilon(\mathbb{Q}, \mathbb{P})$ is attained by the unique probability measure $\mathbb{P} \in \mathcal{P}_\mathbb{U}$ verifying $\nu_t^\mathbb{P} = u(t, X_t)$. Concerning the minimization with respect to \mathbb{Q} , Proposition 5.3 characterizes the explicit solution of the subproblem. In fact, this is a well-known problem in the area of large deviations, see [16]. It gives rise to a variational representation formula relating log-Laplace transform of the costs and relative entropy, which is linked to a specific case of stochastic optimal control, for which it is possible to linearize the HJB equation by an exponential transform, see [19, 20]. This type of problem is known as *path integral control* and it has been extensively studied with many applications, see [10, 42, 41].

In Section 4 we introduce an alternating minimization procedure (4.1), which consists in solving sequentially each subproblem, in \mathbb{Q} and \mathbb{P} , alternatively. In Theorem 4.6, we prove that the iterated values generated by this procedure converge to the minimum value \mathcal{J}_ϵ^* . We insist again on the fact that \mathcal{J}_ϵ is not jointly convex with respect to (\mathbb{Q}, \mathbb{P}) , so, the proof of Theorem 4.6 relies on geometric arguments developed in [14]. In Section 6, we show the relevance of this algorithm compared to classical dynamic programming techniques, by performing an application to the control of thermostatic loads in power systems.

Link to the literature. Interest in optimization problems on the space of probability measures has increased strongly during the recent years with the Monge-Kantorovitch optimal transport problem, which, for two fixed Borel probability measures on \mathbb{R}^d , ν_1 and ν_2 consists in determining a joint law, whose marginals are precisely ν_1 and ν_2 , minimizing an expected given cost. Benamou and Brenier in [2] propose a dynamical formulation of this problem: it consists in an optimal control problem where the aim is to minimize the integrated kinetic energy of a deterministic dynamical system over a given time horizon, in order to go from the initial law ν_1 to ν_2 as terminal law. In [35], the authors replace the deterministic dynamical system with a diffusion, introducing the so called stochastic mass transportation problem. This consists in controlling the drift of the

diffusion to minimize, over a given finite horizon, a mean integrated cost depending on the drift and the state of the process, while imposing the initial and final distribution of the diffusion. Those authors formulate their problem as an optimization on a space of probability measures, for which they make use of convex duality techniques. In [40], the authors generalize these techniques controlling the volatility as well. Those authors also propose a numerical scheme in order to approximate the dual formulation of their stochastic mass transport problem. In the same spirit as in [35], in this paper, we formulate a stochastic optimal control problem as a minimization on the space of probability measures. However our approach is based, on the one hand, on an entropy correction and, on the other hand, on an alternating procedure.

Similar ideas based on an entropy correction and an alternating procedure were introduced in the context of optimal transport, see e.g. [3, 15]. In [3], the authors are interested in the discrete optimal mass transport problem. To approach that problem, they introduce an entropic regularization which consists in minimizing a relative entropy $H(\gamma|\xi)$ over a subset $\mathcal{K} := \mathcal{K}_1 \cap \mathcal{K}_2$ of joint probability measures on $\mathbb{R}^{d \times d}$, where ξ is a reference probability measure on $\mathbb{R}^{d \times d}$, \mathcal{K}_1 is a subset of $\mathcal{P}(\mathbb{R}^{d \times d})$ with a given first marginal, while \mathcal{K}_2 imposes the second marginal. The solution to this new problem is approximated by a sequence $(\gamma^{(n)})_{n \geq 1}$, where $\gamma^{(n+1)}$ is the entropic projection of $\gamma^{(n)}$ on the set \mathcal{C}_n , where $\mathcal{C}_{2p} := \mathcal{K}_2$ and $\mathcal{C}_{2p+1} := \mathcal{K}_1$ for $p \in \mathbb{N}$. This means $\gamma_n = \arg \min_{\gamma \in \mathcal{C}_n} H(\gamma|\gamma^{(n)})$.

This type of methods and their generalization to continuous states distributions are commonly referred to as Sinkhorn algorithms and are widely used in optimal transport and related fields, such as the Schrödinger Bridge problem, see e.g. [11, 13, 34, 36] for detailed accounts. However, the approach we propose here is resolutely different and differs from these classical methods in two aspects. Firstly, our approach is based on a duplication of the optimization variables, and the entropy correction term we introduce is a *penalty term*, designed to impose equality on the duplicated variables. Furthermore, our alternating procedure aims to sequentially optimize the penalized objective function, in the first and then second variable involved in the entropy penalty, whereas the Sinkhorn alternating projection algorithm is always driven to minimize the cross entropy term with respect to the first variable.

Our reformulation offers both numerical and theoretical advantages. From a numerical point of view, our algorithm relies on two standard optimization sub-problems that are simpler than the original stochastic control problem and that can be tackled by specific numerical schemes. For example, one of the two sub-problems (called exponential twist problem) has an explicit solution, which can be evaluated at each time step by parallel computations of conditional expectations. On the other hand, that sub-problem corresponds to a stochastic control problem with no constraints on the control, and can therefore efficiently be tackled by regression methods [4, 22, 23] or deep learning methods as in [12, 21, 24, 27]. Hence, our algorithm constitutes a complementary approach to existing regression or machine learning techniques developed to solve stochastic control problems.

From a theoretical point of view, the entropy penalization approach offers new perspectives

for reformulating complex stochastic control problems including, for example, constraints on the marginal laws of the controlled process, see e.g. Schrödinger bridge. This is the subject of a the recent paper [8].

The paper is organized as follows. After this Introduction, Section 2 is devoted to the basic definitions and notations. In Section 3 we introduce an entropy penalized optimization problem approaching the original stochastic control problem, see in particular Proposition 3.16. The subsequent Section 4 proposes an alternating minimization procedure to approximate the solution of the entropy penalized problem: in particular Theorem 4.6 establishes that convergence. Section 6 illustrates the interest of the approach on a specific application to demand-side management in power systems. Some new perspectives of our method are sketched in Section 7.

The paper is organized as follows. After this Introduction Section 2 is devoted to the basic definitions and notations. In Section 3 we introduce an entropy penalized optimization problem approaching the original stochastic control problem, see in particular Proposition 3.16. The subsequent Section 4 proposes an alternating minimization procedure to approximate the solution of the entropy penalized problem. In particular Theorem 4.6 establishes that convergence. Section 6 illustrates the interest of the approach on a specific application demand-side management in power systems. Some new perspectives of our method are sketched in Section 7. We conclude the paper with the Appendices, which contain the proof of most the technical intermediate results.

2 Notations and definitions

In this section we introduce the basic notions and notations used throughout this document. In what follows, $T > 0$ will be a fixed time horizon.

- All vectors $x \in \mathbb{R}^d$ are column vectors. Given $x \in \mathbb{R}^d$, $|x|$ will denote its Euclidean norm.
- Given a matrix $A \in \mathbb{R}^{d \times d}$, $\|A\| := \sqrt{\text{Tr}[AA^T]}$ will denote its Frobenius norm.
- Given $\phi \in C^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$, $\partial_t \phi$, $\nabla_x \phi$ and $\nabla_x^2 \phi$ will denote respectively the partial derivative of ϕ with respect to (w.r.t.) $t \in [0, T]$, its gradient and its Hessian matrix w.r.t. $x \in \mathbb{R}^d$.
- Given any bounded function $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}$, we denote by $|\Phi|_\infty$ its supremum.
- \mathbb{U} will denote a closed subset of \mathbb{R}^p for some $p \in \mathbb{N}^*$.
- For any topological spaces E and F , $\mathcal{B}(E)$ will denote the Borel σ -field of E ; $C(E, F)$ (resp. $\mathcal{B}(E, F)$) will denote the linear space of functions from E to F that are continuous (resp. Borel). $\mathcal{P}(E)$ will denote the Borel probability measures on E . Given $\mathbb{P} \in \mathcal{P}(E)$, $\mathbb{E}^{\mathbb{P}}$ will denote the expectation with respect to (w.r.t.) \mathbb{P} .
- Except if differently specified, Ω will denote the space of continuous functions from $[0, T]$ to \mathbb{R}^d . For any $t \in [0, T]$ we denote by $X_t : \omega \in \Omega \mapsto \omega_t$ the coordinate mapping on Ω . We

introduce the σ -field $\mathcal{F} := \sigma(X_r, 0 \leq r \leq T)$. On the measurable space (Ω, \mathcal{F}) , we introduce the *canonical process* $X : \omega \in ([0, T] \times \Omega, \mathcal{B}([0, T]) \otimes \mathcal{F}) \mapsto X_t(\omega) = \omega_t \in (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

We endow (Ω, \mathcal{F}) with the right-continuous filtration $\mathcal{F}_t := \bigcap_{t \leq s \leq T} \sigma(X_r, 0 < r \leq s)$, $t \in [0, T]$.

The filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t))$ will be called the *canonical space* (for the sake of brevity, we denote $(\mathcal{F}_t)_{t \in [0, T]}$ by (\mathcal{F}_t)).

- Given a function $\lambda : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, $p \geq 1$ and a Borel probability \mathbb{Q} on Ω , we will improperly say that $\lambda \in L^q(dt \otimes \mathbb{Q})$ if the map $(t, \omega) \mapsto \lambda(t, X_t(\omega)) \in L^q(dt \otimes \mathbb{Q})$.
- Given a continuous local martingale M , $[M]$ will denote its *quadratic variation*.
- Equality between stochastic processes are in the sense of *indistinguishability*.
- Except if specified otherwise, all properties of processes (e.g. measurability, martingale) are with respect to the canonical filtration $(\mathcal{F}_t)_{t \in [0, T]}$.

Definition 2.1. (*Relative entropy*). Let E be a topological space. Let $\mathbb{P}, \mathbb{Q} \in \mathcal{P}(E)$. The relative entropy $H(\mathbb{Q}|\mathbb{P})$ between the measures \mathbb{P} and \mathbb{Q} is defined by

$$H(\mathbb{Q}|\mathbb{P}) := \begin{cases} \mathbb{E}^{\mathbb{Q}} \left[\log \frac{d\mathbb{Q}}{d\mathbb{P}} \right] & \text{if } \mathbb{Q} \ll \mathbb{P} \\ +\infty & \text{otherwise,} \end{cases} \quad (2.1)$$

with the convention $\log(0/0) = 0$.

Remark 2.2. Let E be a Polish space. The relative entropy $H : \mathcal{P}(E) \times \mathcal{P}(E)$ is non-negative and jointly convex, that is for all $\mathbb{P}_1, \mathbb{P}_2, \mathbb{Q}_1, \mathbb{Q}_2 \in \mathcal{P}(E)$, for all $\lambda \in [0, 1]$, $H(\lambda\mathbb{Q}_1 + (1-\lambda)\mathbb{Q}_2 | \lambda\mathbb{P}_1 + (1-\lambda)\mathbb{P}_2) \leq \lambda H(\mathbb{Q}_1 | \mathbb{P}_1) + (1-\lambda)H(\mathbb{Q}_2 | \mathbb{P}_2)$. Moreover, $(\mathbb{P}, \mathbb{Q}) \mapsto H(\mathbb{Q}|\mathbb{P})$ is lower semicontinuous with respect to the weak-star topology on E^* . We refer to [16] Lemma 1.4.3 for a proof of these properties.

Definition 2.3. (*Minimizing sequence, solution and ϵ -solution*). Let E be a generic set. Let $J : E \mapsto \mathbb{R}$ be a function. Let $J^* := \inf_{x \in E} J(x)$, which can be finite or not.

1. A **minimizing sequence** for J is a sequence $(x_n)_{n \geq 0}$ of elements of E such that $J(x_n) \xrightarrow{n \rightarrow +\infty} J^*$.
2. We will say that $x^* \in E$ is a **solution** to the optimization problem

$$\inf_{x \in E} J(x), \quad (2.2)$$

if $J(x^*) = J^*$. In this case, $J^* = \min_{x \in E} J(x)$.

3. For $\epsilon \geq 0$, we will say that $x^\epsilon \in E$ is an **ϵ -solution** to the optimization Problem (2.2) if $0 \leq J(x^\epsilon) - J^* \leq \epsilon$. We also say that x^ϵ is **ϵ -optimal** for the (optimization) Problem (2.2).

We remark that a 0-solution is a solution of the optimization Problem (2.2).

3 From the stochastic optimal control problem to a penalized optimization problem

In this section we consider a stochastic control problem that we reformulate in terms of an optimization problem on a space of probability measures. Later we propose a penalized version of that problem whose solutions are ε -optimal for the original problem.

3.1 The stochastic optimal control problem

We specify the assumptions and the formulation of the stochastic optimal control Problem (1.3), stated in the Introduction. Let us first consider a drift $b \in \mathcal{B}([0, T] \times \mathbb{R}^d \times \mathbb{U}, \mathbb{R}^d)$ and a diffusion matrix $\sigma \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{R}^{d \times d})$, following the assumptions below.

Hypothesis 3.1. (*Diffusion coefficients*).

1. b is continuous in (t, x, u) .
2. There exists a constant $C_{b,\sigma} > 0$ such that, for all $(t, x) \in [0, T] \times \mathbb{R}^d, u \in \mathbb{U}$,

$$|b(t, x, u)| + \|\sigma(t, x)\| \leq C_{b,\sigma}(1 + |x|). \quad (3.1)$$

3. There exists $c_\sigma > 0$ such that for all $(t, x) \in [0, T] \times \mathbb{R}^d, \xi \in \mathbb{R}^d$,

$$\xi^\top \sigma \sigma^\top(t, x) \xi \geq c_\sigma |\xi|^2. \quad (3.2)$$

4. For all $x \in \mathbb{R}^d$,

$$\lim_{y \rightarrow x} \sup_{0 \leq r \leq T} \|\sigma(r, x) - \sigma(r, y)\| = 0.$$

Let us define the admissible set of probability measures $\mathcal{P}_\mathbb{U}$ for Problem (1.3).

Definition 3.2. Let $\mathcal{P}_\mathbb{U}$ be the set of probability measures on (Ω, \mathcal{F}) such that, for all $\mathbb{P} \in \mathcal{P}_\mathbb{U}$, under \mathbb{P} , the canonical process decomposes as

$$X_t = x + \int_0^t b(r, X_r, \nu_r^\mathbb{P}) dr + M_t^\mathbb{P}, \quad (3.3)$$

where $x \in \mathbb{R}^d$, $M^\mathbb{P}$ is a $(\mathbb{P}, \mathcal{F}_t)$ -local martingale such that $[M^\mathbb{P}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$, $\nu^\mathbb{P}$ is a progressively measurable process with values in \mathbb{U} . If in addition there exists $u^\mathbb{P} \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$ such that $\nu_t^\mathbb{P} = u^\mathbb{P}(t, X_t) dt \otimes d\mathbb{P}$ -a.e, we will denote $\mathbb{P} \in \mathcal{P}_\mathbb{U}^{\text{Markov}}$.

Remark 3.3. The admissible set of probability measures $\mathcal{P}_\mathbb{U}$ for Problem (1.3) imposes an uncontrolled volatility. Indeed the approach developed in the present paper relies on Girsanov's theorem and can not be easily extended to the case of controlled volatility.

Remark 3.4. By classical stochastic calculus arguments, see e.g. Proposition 5.4.6 in [30], we can state the following. If $\mathbb{P} \in \mathcal{P}_{\mathbb{U}}^{Markov}$ in the sense of Definition 3.2, then the following equivalent properties hold.

1. One has

$$X_t = x + \int_0^t b(r, X_r, u^{\mathbb{P}}(r, X_r))dr + M_t^{\mathbb{P}}, \quad (3.4)$$

with $x \in \mathbb{R}^d$, $[M^{\mathbb{P}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r)dr$.

2. \mathbb{P} is solution of the martingale problem (in the sense of Stroock and Varadhan in [39]) associated with the initial condition $(0, x)$ and the operator $\mathcal{L}_{u^{\mathbb{P}}}$ defined, for all bounded functions $\phi \in C_b^{1,2}([0, T] \times \mathbb{R}^d, \mathbb{R})$, $(t, y) \in [0, T] \times \mathbb{R}^d$, by

$$\mathcal{L}_{u^{\mathbb{P}}}\phi(t, y) = \partial_t \phi(t, y) + \langle \nabla_x \phi(t, y), b(t, y, u^{\mathbb{P}}(t, y)) \rangle + \frac{1}{2} Tr[\sigma \sigma^\top(t, y) \nabla_x^2 \phi(t, y)], \quad (3.5)$$

with $\nu^{\mathbb{P}} := u^{\mathbb{P}}(\cdot, X_\cdot)$ in (3.3).

3. \mathbb{P} is a solution (in law) of

$$X_t = x + \int_0^t b(r, X_r, u^{\mathbb{P}}(r, X_r))dr + \int_0^t \sigma(r, X_r) dW_r, \quad (3.6)$$

for some suitable Brownian motion W .

We will often make use of the following proposition.

Proposition 3.5. Assume Hypothesis 3.1 holds. Let $u \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$. There exists a unique probability measure $\mathbb{P}^u \in \mathcal{P}_{\mathbb{U}}^{Markov}$ such that under \mathbb{P}^u the canonical process decomposes as (3.3), with $\nu_t^{\mathbb{P}^u} = u(t, X_t) (= u^{\mathbb{P}^u}(t, X_t))$.

Remark 3.6. In particular, for a given $u : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, the equation

$$X_t = x + \int_0^t b(r, X_r, u(r, X_r))dr + M_t^{\mathbb{P}}, \quad (3.7)$$

where $x \in \mathbb{R}^d$, X is the canonical process, and $M^{\mathbb{P}}$ is a $(\mathbb{P}, \mathcal{F}_t)$ -local martingale such that $[M^{\mathbb{P}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r)dr$, admits a unique solution \mathbb{P} .

Proof of Proposition 3.5. By Theorem 10.1.3 in [39], the martingale problem, associated with the initial condition $(0, x)$ and the operator \mathcal{L}_u defined by (3.5) with $u^{\mathbb{P}} = u$, admits a unique solution \mathbb{P}^u . The result is then a consequence of Remark 3.4. \square

Let then $f \in \mathcal{B}([0, T] \times \mathbb{R}^d \times \mathbb{U}, \mathbb{R})$, $g \in \mathcal{B}(\mathbb{R}^d, \mathbb{R})$, referred to as the *running cost* and the *terminal cost* respectively, and assume that the following holds.

Hypothesis 3.7. (Cost functions).

1. The functions f, g are positive and there exist $C_{f,g} > 0$, $p \geq 1$, such that, for all $(t, x, u) \in [0, T] \times \mathbb{R}^d \times \mathbb{U}$,

$$f(t, x, u) + g(x) \leq C_{f,g}(1 + |x|^p). \quad (3.8)$$

2. f and g are continuous in $(t, x, u) \in [0, T] \times \mathbb{R}^d \times \mathbb{U}$ and $x \in \mathbb{R}^d$ respectively.
3. Let $p \geq 1$ mentioned at item 1. There exist constants $p' > p$ and $C' > 0$ such that $|u|^{p'} \leq C'(1 + f(t, x, u))$, for all $(t, x, u) \in [0, T] \times \mathbb{R}^d \times \mathbb{U}$.

For any $(t, x) \in [0, T] \times \mathbb{R}^d$, we introduce the set

$$K(t, x) := \left\{ (b(t, x, u), z) \mid u \in \mathbb{U}, z \geq f(t, x, u) \right\}. \quad (3.9)$$

Remark 3.8. 1. Item 3. of Hypothesis 3.7 is of course verified if \mathbb{U} is bounded.

2. Whenever \mathbb{U} is unbounded, the same hypothesis implies that $|f(t, x, u)| \rightarrow +\infty$ if $|u|$ goes to infinity.
3. Under Hypothesis 3.7, the set $K(t, x)$ is closed. Let indeed $((y_n, z_n))_{n \geq 0}$ be a sequence of elements of $K(t, x)$, which converges toward $(y^*, z^*) \in \mathbb{R}^{d+1}$. Let $(u_n)_{n \geq 0}$ be a sequence of elements of \mathbb{U} such that for all $n \in \mathbb{N}$,

$$y_n = b(t, x, u_n) \quad \text{and} \quad z_n \geq f(t, x, u_n). \quad (3.10)$$

Then, by item 3. of Hypothesis 3.7,

$$\sup_{n \in \mathbb{N}} |u_n|^{p'} \leq C' \left(1 + \sup_{n \in \mathbb{N}} f(t, x, u_n) \right) \leq C' \left(1 + \sup_{n \in \mathbb{N}} z_n \right).$$

Since $(z_n)_{n \in \mathbb{N}}$ converges, it is bounded and the previous inequality implies that $(u_n)_{n \in \mathbb{N}}$ is also bounded. Up to a subsequence, we can thus assume that $(u_n)_{n \in \mathbb{N}}$ converges towards a limit $u^* \in \mathbb{U}$ (recall that \mathbb{U} is closed). Since $b(t, x, \cdot)$ and $f(t, x, \cdot)$ are continuous, letting $n \rightarrow +\infty$ in (3.10), yields $y^* = b(t, x, u^*)$ and $z^* \geq f(t, x, u^*)$. Hence $(y^*, z^*) \in K(t, x)$, and $K(t, x)$ is closed.

We will require the following convexity assumption.

Hypothesis 3.9. (Convex). For all $(t, x) \in [0, T] \times \mathbb{R}^d$, the set $K(t, x)$ is convex.

Remark 3.10. 1. If \mathbb{U} is convex, b is linear w.r.t. to u and f is convex w.r.t. u , then Hypothesis 3.9 holds.

2. Hypothesis 3.9 is a classical convexity assumption when one wants to prove existence of optimal Markovian control to Problem (1.3) in the weak sense, by using compactness arguments, see e.g. [25, 26, 18, 32].

We conclude this section providing a moment estimate, see e.g. Corollary 5.12 of Chapter 2 in [31].

Lemma 3.11. Let b, σ fulfill Hypothesis 3.1 and $q \geq 1$. Then there is a constant $C(q)$, which depends on T and $C_{b, \sigma}$ (and q), such that the following holds.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ be a filtered probability space. Let $\nu : [0, T] \times \Omega \rightarrow \mathbb{U}$ be an (\mathcal{F}_t) -progressively measurable process. Let X be an Itô process on $(\Omega, \mathcal{F}, \mathbb{P})$, which decomposes as

$$X_t = x + \int_0^t b(r, X_r, \nu_r) dr + M_t^{\mathbb{P}},$$

where $M^{\mathbb{P}}$ is a \mathbb{P} -local martingale such that $[M^{\mathbb{P}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$. Then we have

$$\mathbb{E}^{\mathbb{P}} \left[\sup_{0 \leq t \leq T} |X_t|^q \right] \leq C(q).$$

Under Hypotheses 3.1 and 3.7, the function J introduced in (1.3) is well-defined on the set $\mathcal{P}_{\mathbb{U}}$, set out in Definition 3.2. Indeed, by the moment estimate given by Lemma 3.11 one has

$$\mathbb{E}^{\mathbb{P}} \left[\int_0^T f(r, X_r, \nu_r^{\mathbb{P}}) dr + g(X_T) \right] < +\infty,$$

for all $\mathbb{P} \in \mathcal{P}_{\mathbb{U}}$.

3.2 The penalized optimization problem

As mentioned in the Introduction, we reformulate Problem (1.3), by doubling the decision variables and by adding a relative entropy term, in the objective function. The modified penalized Problem is precisely (1.5), where \mathcal{A} is the subset of elements $(\mathbb{P}, \mathbb{Q}) \in \mathcal{P}(\Omega)^2$ defined below.

Definition 3.12. *\mathcal{A} will denote the set of probability measure couples $(\mathbb{P}, \mathbb{Q}) \in \mathcal{P}(\Omega)^2$, such that*

1. $\mathbb{P} \in \mathcal{P}_{\mathbb{U}}$,
2. $H(\mathbb{Q}|\mathbb{P}) < +\infty$.

In the perspective of solving the penalized optimization Problem (1.5), we will introduce in Sections 5.1 and 5.2 two subproblems. The interest of the penalized formulation (1.5) relies on the fact that each of the subproblems, $\inf_{\mathbb{Q} \in \mathcal{P}(\Omega)} \mathcal{J}_\epsilon(\mathbb{Q}, \mathbb{P})$ and $\inf_{\mathbb{P} \in \mathcal{P}_{\mathbb{U}}} \mathcal{J}_\epsilon(\mathbb{Q}, \mathbb{P})$, can be solved by classical techniques described in the literature: those resolutions will constitute the two steps of our alternating minimization algorithm.

The first subproblem, considered in Section 5.2, is a minimization on \mathbb{Q} , the probability \mathbb{P} being fixed, and it is related to a variational representation formula, whose solution is denominated exponential twist, see e.g. [16]. In particular the following result will intervene.

Proposition 3.13. *Let $\varphi : \Omega \rightarrow \mathbb{R}$ be a Borel function and $\mathbb{P} \in \mathcal{P}(\Omega)$. Assume that φ is bounded below. Then*

$$\inf_{\mathbb{Q} \in \mathcal{P}(\Omega)} \mathbb{E}^{\mathbb{Q}}[\varphi(X)] + \frac{1}{\epsilon} H(\mathbb{Q}|\mathbb{P}) = -\frac{1}{\epsilon} \log \mathbb{E}^{\mathbb{P}}[\exp(-\epsilon\varphi(X))]. \quad (3.11)$$

Moreover the problem (3.11) admits a unique solution (minimizer) $\mathbb{Q}^* \in \mathcal{P}(\Omega)$, given by

$$d\mathbb{Q}^* = \frac{\exp(-\epsilon\varphi(X))}{\mathbb{E}^{\mathbb{P}}[\exp(-\epsilon\varphi(X))]} d\mathbb{P}.$$

Proof. The random variable $\varphi(X)$ is bounded below, hence satisfies condition (FE) of [5]. The statement then follows from Proposition 2.5 in [5]. \square

Applying Proposition 3.13 to our framework for $\mathbb{P} \in \mathcal{P}_{\mathbb{U}}$ and $\varphi(X) := \int_0^T f(r, X_r, \nu_r^{\mathbb{P}}) dr + g(X_T)$, we get that, under Hypothesis 3.7, the subproblem $\inf_{\mathbb{Q} \in \mathcal{P}(\Omega)} \mathcal{J}_{\epsilon}(\mathbb{Q}, \mathbb{P})$ admits a unique solution \mathbb{Q}^* given by

$$d\mathbb{Q}^* = \frac{\exp\left(-\epsilon \int_0^T f(r, X_r, \nu_r^{\mathbb{P}}) dr - \epsilon g(X_T)\right)}{\mathbb{E}^{\mathbb{P}}\left[\exp\left(-\epsilon \int_0^T f(r, X_r, \nu_r^{\mathbb{P}}) dr - \epsilon g(X_T)\right)\right]} d\mathbb{P}, \quad (3.12)$$

and that its optimal value is

$$\mathcal{J}_{\epsilon}(\mathbb{Q}^*, \mathbb{P}) = -\frac{1}{\epsilon} \log \mathbb{E}^{\mathbb{P}}\left[\exp\left(-\epsilon \int_0^T f(r, X_r, \nu_r^{\mathbb{P}}) dr - \epsilon g(X_T)\right)\right]. \quad (3.13)$$

This subproblem is further analyzed in Section 5.2. In particular Proposition 5.3 allows to identify \mathbb{Q}^* as the law of a semimartingale with Markovian drift.

Remark 3.14. *Suppose the validity of Hypothesis 3.7. Then $\|d\mathbb{Q}^*/d\mathbb{P}\|_{\infty} < +\infty$.*

Let us discuss now about the second problem, i.e. the subproblem $\inf_{\mathbb{P} \in \mathcal{P}_{\mathbb{U}}} \mathcal{J}_{\epsilon}(\mathbb{Q}, \mathbb{P})$, which will be the object of Section 5.1. This is a minimization on \mathbb{P} , the probability \mathbb{Q} remaining unchanged. The optimal solution of this subproblem is a probability measure \mathbb{P}^u , in the sense of Proposition 3.5, where the function $u \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$, is provided by a pointwise minimization.

The next theorem proves that the penalized Problem (1.5) has a Markovian solution.

Theorem 3.15. *Assume Hypotheses 3.1, 3.7 and 3.9 hold. Then the penalized Problem (1.5) has a solution $(\mathbb{P}_{\epsilon}^*, \mathbb{Q}_{\epsilon}^*) \in \mathcal{A}$, in the sense that $\mathcal{J}_{\epsilon}^* = \mathcal{J}(\mathbb{Q}_{\epsilon}^*, \mathbb{P}_{\epsilon}^*)$. Moreover, under \mathbb{P}_{ϵ}^* , the canonical process is a Markov process and $\nu^{\mathbb{P}_{\epsilon}^*}$, related to \mathbb{P}_{ϵ}^* by Definition 3.2, is such that $\nu_t^{\mathbb{P}_{\epsilon}^*} (= u_{\epsilon}^{\mathbb{P}_{\epsilon}^*}(t, X_t)) = u_{\epsilon}^*(t, X_t)$, for some function $u_{\epsilon}^* \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$ and we also have*

$$d\mathbb{Q}_{\epsilon}^* = \frac{\exp\left(-\epsilon \int_0^T f(r, X_r, u_{\epsilon}^*(r, X_r)) dr - \epsilon g(X_T)\right)}{\mathbb{E}^{\mathbb{Q}_{\epsilon}^*}\left[\exp\left(-\epsilon \int_0^T f(r, X_r, u_{\epsilon}^*(r, X_r)) dr - \epsilon g(X_T)\right)\right]} d\mathbb{P}_{\epsilon}^*. \quad (3.14)$$

The proof of this result relies on several technical lemmas. For the convenience of the reader, it is postponed to Appendix C.

The following proposition justifies the use of the penalized Problem (1.5) to approximately solve the initial stochastic optimal control Problem (1.3). Indeed, the next result states that one can build an approximate solution of Problem (1.3), based on an approximate solution of Problem (1.5)

Proposition 3.16. *We suppose the validity of Hypothesis 3.1 and item 1. of Hypothesis 3.7. Let $\epsilon > 0, \epsilon' \geq 0$ and let $\mathbb{P}_{\epsilon}^{\epsilon'}$ be the first component of an ϵ' -solution of Problem (1.5), in the sense of Definition 2.3, with $E = \mathcal{A}$. We set $Y_{\epsilon}^{\epsilon'} := \int_0^T f(r, X_r, \nu_r^{\epsilon'}) dr + g(X_T)$, where $\nu_r^{\epsilon'}$ corresponds to the $\nu^{\mathbb{P}_{\epsilon}^{\epsilon'}}$, appearing in decomposition (3.3). Then the following holds.*

1. *There is a constant C^* depending only on $C_{b,\sigma}, C_{f,g}, p, d, T$ of Hypothesis 3.7 1. such that $\max(\mathbb{E}[Y_{\epsilon}^{\epsilon'}], \text{Var}^{\mathbb{P}_{\epsilon}^{\epsilon'}}[Y_{\epsilon}^{\epsilon'}]) \leq C^*$, where $\text{Var}^{\mathbb{P}_{\epsilon}^{\epsilon'}}[Y_{\epsilon}^{\epsilon'}]$ denotes the variance of $Y_{\epsilon}^{\epsilon'}$ under $\mathbb{P}_{\epsilon}^{\epsilon'}$.*

2. We have

$$0 \leq J(\mathbb{P}_\epsilon^{\epsilon'}) - J^* \leq \epsilon e^{\epsilon \mathbb{E}[Y_\epsilon^{\epsilon'}]} \text{Var}^{\mathbb{P}_\epsilon^{\epsilon'}} [Y_\epsilon^{\epsilon'}] + \epsilon',$$

where we recall that J and J^* are defined in (1.3).

Remark 3.17. 1. Let $(\mathbb{P}_\epsilon^*, \mathbb{Q}_\epsilon^*)$ be a solution of Problem (1.5) given by Theorem 3.15. Applying item 2. of Proposition 3.16 with $\epsilon' = 0$ implies that \mathbb{P}_ϵ^* is an $\epsilon e^{\epsilon \mathbb{E}[Y_\epsilon^0]} \text{Var}^{\mathbb{P}_\epsilon^*} [Y_\epsilon^0]$ -solution of the original Problem (1.3).

2. By definition of infimum, for $\epsilon' > 0$, the existence of an ϵ' -solution is always guaranteed without any convex assumption on the running cost f w.r.t. the control variable.

3. In the sequel, assuming Hypotheses 3.7 and 3.9 to hold, we will propose an algorithm providing a sequence of ϵ'_n -solutions of the penalized Problem (1.5), where $\epsilon'_n \rightarrow 0$ as $n \rightarrow +\infty$. This will also provide a sequence of $(\epsilon e^{\epsilon \mathbb{E}[Y_\epsilon^{\epsilon'_n}]} \text{Var}^{\mathbb{P}_\epsilon^{\epsilon'_n}} [Y_\epsilon^{\epsilon'_n}] + \epsilon'_n)$ -solutions to the original Problem (1.3) (with a fixed $\epsilon > 0$).

Proof of Proposition 3.16. We first prove item 1. Let $(\mathbb{P}_\epsilon^{\epsilon'}, \mathbb{Q}_\epsilon^{\epsilon'})$ be an ϵ' -solution of Problem (1.5). By Hypothesis 3.7, for all $\epsilon > 0$, on the one hand one has

$$\mathbb{E}[Y_\epsilon^{\epsilon'}] \leq C_{f,g}(T+1) \left(1 + \mathbb{E}^{\mathbb{P}_\epsilon^{\epsilon'}} \left[\sup_{0 \leq t \leq T} |X_t|^p \right] \right),$$

and on the other hand,

$$\text{Var}^{\mathbb{P}_\epsilon^{\epsilon'}} [Y_\epsilon^{\epsilon'}] \leq \mathbb{E}^{\mathbb{P}_\epsilon^{\epsilon'}} [(Y_\epsilon^{\epsilon'})^2] \leq 4C_{f,g}^2(T^2+1) \left(1 + \mathbb{E}^{\mathbb{P}_\epsilon^*} \left[\sup_{0 \leq t \leq T} |X_t|^{2p} \right] \right).$$

Combining these inequalities with Lemma 3.11 implies the existence of a constant C^* , depending only on $C_{b,\sigma}, C_{f,g}, p, d, T$, such that $\max(\mathbb{E}[Y_\epsilon^{\epsilon'}], \text{Var}^{\mathbb{P}_\epsilon^{\epsilon'}} [Y_\epsilon^{\epsilon'}]) \leq C^*$, which is the statement of item 1.

We go on with the proof of item 2. First, a direct application of Lemma F.1 with $\eta = Y_\epsilon^{\epsilon'}$, yields

$$0 \leq \mathbb{E}^{\mathbb{P}_\epsilon^{\epsilon'}} [Y_\epsilon^{\epsilon'}] - \left(-\frac{1}{\epsilon} \log \mathbb{E}^{\mathbb{P}_\epsilon^{\epsilon'}} [\exp(-\epsilon Y_\epsilon^{\epsilon'})] \right) \leq \frac{\epsilon}{2} \text{Var}^{\mathbb{P}_\epsilon^{\epsilon'}} [Y_\epsilon^{\epsilon'}] e^{\epsilon \mathbb{E}^{\mathbb{P}_\epsilon^{\epsilon'}} [Y_\epsilon^{\epsilon'}]}. \quad (3.15)$$

Let then $\tilde{\mathbb{Q}}$ be the solution of $\inf_{\mathbb{Q} \in \mathcal{P}(\Omega)} \mathcal{J}_\epsilon(\mathbb{Q}, \mathbb{P}_\epsilon^{\epsilon'})$, given by (3.12), replacing \mathbb{P} with $\mathbb{P}_\epsilon^{\epsilon'}$. Consequently, by (3.13) $\mathcal{J}_\epsilon(\tilde{\mathbb{Q}}, \mathbb{P}_\epsilon^{\epsilon'}) = -\frac{1}{\epsilon} \log \mathbb{E}^{\mathbb{P}_\epsilon^{\epsilon'}} [\exp(-\epsilon Y_\epsilon^{\epsilon'})]$; replacing the right-hand side of previous expression with $\mathcal{J}_\epsilon(\tilde{\mathbb{Q}}, \mathbb{P}_\epsilon^{\epsilon'})$ in (3.15), we get

$$0 \leq \mathbb{E}^{\mathbb{P}_\epsilon^{\epsilon'}} [Y_\epsilon^{\epsilon'}] - \mathcal{J}_\epsilon(\tilde{\mathbb{Q}}, \mathbb{P}_\epsilon^{\epsilon'}) \leq \epsilon e^{\epsilon \mathbb{E}[Y_\epsilon^{\epsilon'}]} \text{Var}^{\mathbb{P}_\epsilon^{\epsilon'}} [Y_\epsilon^{\epsilon'}]. \quad (3.16)$$

Let $\mathbb{Q}_\epsilon^{\epsilon'}$ be the second component of the ϵ' -solution of Problem (1.5), mentioned in the statement of the current proposition. Observe that $\mathcal{J}_\epsilon(\tilde{\mathbb{Q}}, \mathbb{P}_\epsilon^{\epsilon'}) \leq \mathcal{J}_\epsilon(\mathbb{Q}_\epsilon^{\epsilon'}, \mathbb{P}_\epsilon^{\epsilon'}) \leq \mathcal{J}_\epsilon^* + \epsilon'$. Besides, Problem (1.3) is equivalent to Problem (1.5), under the constraint $\mathbb{Q} = \mathbb{P}$, therefore $\mathcal{J}_\epsilon^* \leq J^*$. Then

$$\mathcal{J}_\epsilon(\tilde{\mathbb{Q}}, \mathbb{P}_\epsilon^{\epsilon'}) - J^* \leq \mathcal{J}_\epsilon^* + \epsilon' - J^* \leq \epsilon'. \quad (3.17)$$

Using (3.16) and (3.17) finally yields

$$0 \leq J(\mathbb{P}_\epsilon^{\epsilon'}) - J^* = \mathbb{E}^{\mathbb{P}_\epsilon^{\epsilon'}} [Y_\epsilon^{\epsilon'}] - \mathcal{J}_\epsilon(\tilde{\mathbb{Q}}, \mathbb{P}_\epsilon^{\epsilon'}) + \mathcal{J}_\epsilon(\tilde{\mathbb{Q}}, \mathbb{P}_\epsilon^{\epsilon'}) - J^* \leq \epsilon e^{\epsilon \mathbb{E}[Y_\epsilon^{\epsilon'}]} \text{Var}^{\mathbb{P}_\epsilon^{\epsilon'}} [Y_\epsilon^{\epsilon'}] + \epsilon'. \quad (3.18)$$

This concludes the proof of item 2. \square

4 Alternating minimization procedure

From now on, ϵ will be implicit in the cost function \mathcal{J}_ϵ to alleviate notations. In this section we will assume Hypotheses 3.1, 3.7 and 3.9. We present an alternating procedure for solving the penalized Problem (1.5). Let $(\mathbb{P}_0, \mathbb{Q}_0) \in \mathcal{A}$. We will define a sequence $(\mathbb{P}_k, \mathbb{Q}_k)_{k \geq 0}$ satisfying the alternating minimization procedure

$$\mathbb{Q}_{k+1} = \arg \min_{\mathbb{Q} \in \mathcal{P}(\Omega)} \mathcal{J}(\mathbb{Q}, \mathbb{P}_k), \quad \mathbb{P}_{k+1} \in \arg \min_{\mathbb{P} \in \mathcal{P}_\mathbb{U}} \mathcal{J}(\mathbb{Q}_{k+1}, \mathbb{P}). \quad (4.1)$$

4.1 Convergence result

The convergence of alternating minimization algorithms has been extensively studied in particular in Euclidean spaces. In general the proof of convergence results requires joint convexity and smoothness properties of the objective function, see [1]. The major difficulty in our case is that the convexity only holds w.r.t \mathbb{Q} (in fact the set $\mathcal{P}_\mathbb{U}$ is not even convex). To prove the convergence we need to rely on techniques which exploit the properties of the entropic penalization. Let us first assume that the initial probability measure $\mathbb{P}_0 \in \mathcal{P}_\mathbb{U}$ is Markovian in the following sense.

Hypothesis 4.1. $\mathbb{P}_0 \in \mathcal{P}_\mathbb{U}^{\text{Markov}}$, see Definition 3.2. In particular, there exists $u^0 (= u^{\mathbb{P}_0}) \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$, such that $\mathbb{P}_0 = \mathbb{P}^{u^0}$.

Let σ^{-1} be the generalized right-inverse of σ , i.e. $\sigma^\top (\sigma \sigma^\top)^{-1}$. For a fixed Borel function $\beta : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$, we set

$$F_\beta : (t, x, u) \in [0, T] \times \mathbb{R}^d \times \mathbb{U} \mapsto f(t, x, u) + \frac{1}{2\epsilon} |\sigma^{-1}(t, x)(\beta(t, x) - b(t, x, u))|^2. \quad (4.2)$$

Given $(t, x) \in [0, T] \times \mathbb{R}^d$, we furthermore introduce the function

$$(y, z) \in \mathbb{R}^d \times \mathbb{R} \mapsto \bar{F}_\beta^{t,x}(y, z) := z + \frac{1}{2\epsilon} |\sigma^{-1}(t, x)(\beta(t, x) - y)|^2. \quad (4.3)$$

We state a lemma which will be used several times in this article and it will be proved in the Appendix E.

Lemma 4.2. Let $(t, x) \in [0, T] \times \mathbb{R}^d$. The following holds.

1. The function $\bar{F}_\beta^{t,x}$, defined by (4.3), restricted to $K(t, x)$, has a unique minimum (y^*, z^*) , which verifies

$$z - z^* + \frac{1}{\epsilon} \langle (\sigma^{-1})^\top \sigma^{-1}(t, x)(y^* - \beta(t, x)), y - y^* \rangle \geq 0 \quad \forall (y, z) \in K(t, x). \quad (4.4)$$

2. (a) Let $u^* \in \mathbb{U}$ such that $y^* = b(t, x, u^*)$ and $z^* \geq f(t, x, u^*)$. Then

$$u^* \in \arg \min_{a \in \mathbb{U}} F_\beta(t, x, a), \quad (4.5)$$

where F_β was defined in (4.2).

(b) Conversely, if $u^* \in \arg \min_{a \in \mathbb{U}} F_\beta(t, x, a)$, then

$$(y^*, z^*) := (b(t, x, u^*), f(t, x, u^*)) \in \arg \min_{(y, z) \in K(t, x)} \bar{F}_\beta^{t, x}(y, z). \quad (4.6)$$

Remark 4.3. Let $\hat{\beta} : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ be a path-dependent function. We extend the definition (4.2) of F_β , by setting

$$\hat{F}_{\hat{\beta}}(t, X, u) := f(t, X_t, u) + \frac{1}{2\epsilon} |\sigma^{-1}(t, X_t)(\hat{\beta}(t, X) - b(t, X_t, u))|^2. \quad (4.7)$$

We remark that, whenever $\hat{u} : [0, T] \times \Omega \rightarrow \mathbb{U}$, $\hat{u}(t, X) = u(t, X_t)$, and $\hat{\beta}(t, X) = \beta(t, X_t)$, we have

$$F_\beta(t, X_t, u(t, X_t)) = \hat{F}_{\hat{\beta}}(t, X, \hat{u}(t, X)).$$

Let \mathbb{P}_0 satisfying Hypothesis 4.1. We set $\mathbb{Q}_0 = \mathbb{P}_0$. We build a sequence $(\mathbb{P}_k, \mathbb{Q}_k)_{k \geq 0}$ of elements of \mathcal{A} , according to the following procedure. Let $k \geq 1$.

- Let

$$d\mathbb{Q}_{k+1} := \frac{\exp\left(-\epsilon \int_0^T f(r, X_r, u^k(r, X_r)) dr - \epsilon g(X_T)\right)}{\mathbb{E}^{\mathbb{P}_k} \left[\exp\left(-\epsilon \int_0^T f(r, X_r, u^k(r, X_r)) dr - \epsilon g(X_T)\right) \right]} d\mathbb{P}_k, \quad (4.8)$$

where $u^k = u^{\mathbb{P}_k}$. By Proposition 5.3 below there exists a measurable function $\beta^{k+1} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that, under \mathbb{Q}_{k+1} , the canonical process decomposes as

$$X_t = x + \int_0^t \beta^{k+1}(r, X_r) dr + M_t^{\mathbb{Q}_{k+1}}, \quad (4.9)$$

where $M^{\mathbb{Q}_{k+1}}$ is a local \mathbb{Q}_{k+1} -martingale such that $[M^{\mathbb{Q}_{k+1}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$.

- By Proposition 5.2 there exists a Borel function $u^{k+1} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{U}$, such that

$$(t, x) \mapsto u^{k+1}(t, x) \in \arg \min_{a \in \mathbb{U}} F_{\beta^{k+1}}(t, x, a), \quad (4.10)$$

where $F_{\beta^{k+1}}$ is given by (4.2). We define $\mathbb{P}_{k+1} := \mathbb{P}^{u^{k+1}}$ according to Proposition 3.5, so that under \mathbb{P}_{k+1} the canonical process decomposes as

$$X_t = x + \int_0^t b(r, X_r, u^{k+1}(r, X_r)) dr + M_t^{\mathbb{P}_{k+1}}, \quad (4.11)$$

where $M^{\mathbb{P}_{k+1}}$ is a local \mathbb{P}_{k+1} -martingale such that $[M^{\mathbb{P}_{k+1}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$. In particular $u^{k+1} = u^{\mathbb{P}_{k+1}}$.

The proof of the lemma below is a direct application of Proposition 5.3 for item 1. Item 2. follows from Propositions 5.2 and 3.5.

Lemma 4.4. *Let $\mathbb{P}_0 = \mathbb{Q}_0 \in \mathcal{P}_{\mathbb{U}}$ satisfying Hypothesis 4.1. Let $(\mathbb{P}_k, \mathbb{Q}_k)_{k \geq 0}$ be given by the recursion (4.8) and just before (4.11). The following holds for $k \geq 0$.*

1. $\mathbb{Q}_{k+1} = \arg \min_{\mathbb{Q} \in \mathcal{P}(\Omega)} \mathcal{J}(\mathbb{Q}, \mathbb{P}_k)$, and

$$\mathcal{J}(\mathbb{Q}_{k+1}, \mathbb{P}_k) = -\frac{1}{\epsilon} \log \mathbb{E}^{\mathbb{P}_k} \left[\exp \left(-\epsilon \int_0^T f(r, X_r, u^k(r, X_r)) dr - \epsilon g(X_T) \right) \right],$$

where $u^k = u^{\mathbb{P}_k}$. Moreover, under \mathbb{Q}_{k+1} the canonical process is a Markov process and $\beta^{k+1} \in L^q(dt \otimes \mathbb{Q}^{k+1})$ for all $1 < q < 2$.

2. $\mathbb{P}_{k+1} \in \arg \min_{\mathbb{P} \in \mathcal{P}_{\mathbb{U}}} \mathcal{J}(\mathbb{Q}_{k+1}, \mathbb{P})$.

Remark 4.5. *Let $u^0 \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$. We emphasize that the sequence $(u^k)_{k \geq 1}$ of Markovian controls produced by the alternating minimization procedure in (4.8)-(4.10), is independent of the initial law δ_x .*

Indeed the function $\beta = \beta^{k+1}$ appearing at each step, is provided by Proposition 5.3 and it is of the type $\beta(t, x) = b(t, x, u(t, x)) + \lambda(t, x)$, according to the proof. That Proposition 5.3 is a consequence of Corollary 6.12 in [7], which follows from Corollary 6.8 of the same paper. We recall that, the aforementioned function λ was of the form $\frac{\Gamma^v(id)}{v}$, where, v was defined in (5.4), $\phi \mapsto \Gamma^v(\phi)$ was a map introduced in Proposition 5.11 of [7].

The function v , and (taking into account of the step (c) of the proof of that Proposition) Γ^v only depend on the "Regularly Markovian property", (Hypothesis 6.2), which is always fulfilled in our case, see Remark 6.7 of the same paper. That "Regularly Markovian property" is in fact only concerned by the dynamics of \mathbb{P} and not on the initial condition.

Besides, the minimization (5.2) in Proposition 5.2 does not depend on the initial condition provided that β is also independent from it.

The main result of this section is given below.

Theorem 4.6. *Let $\epsilon > 0$ and recall that $\mathcal{J} = \mathcal{J}_\epsilon$ and $\mathcal{J}^* = \mathcal{J}_\epsilon^*$ defined in (1.5), i.e.*

$$\mathcal{J}^* = \inf_{(\mathbb{P}, \mathbb{Q}) \in \mathcal{A}} \mathcal{J}(\mathbb{Q}, \mathbb{P}).$$

Let $\mathbb{P}_0 = \mathbb{Q}_0$ satisfying Hypothesis 4.1. Assume also that Hypotheses 3.1, 3.7 and 3.9 hold. Let $(\mathbb{P}_k, \mathbb{Q}_k)_{k \geq 0}$ be given by the recursion (4.8) and just before (4.11). Then $\mathcal{J}(\mathbb{Q}_k, \mathbb{P}_k) \searrow_{k \rightarrow +\infty} \mathcal{J}^$.*

Moreover, there exists a constant $C > 0$, which only depends on $c_\sigma, C_{b,\sigma}, C_{f,g}, d$ and T (and not on k, ϵ), such that $0 \leq \mathcal{J}(\mathbb{Q}_k, \mathbb{P}_k) - \mathcal{J}^ \leq \frac{C}{k} (1 + \frac{1}{\epsilon})$, for all $k \geq 1$.*

Theorem 4.6 and Proposition 3.16 yield Corollary 4.7 below.

Corollary 4.7. Let $\epsilon > 0$ and J^* as defined in (1.3). Let $\mathbb{P}_0 = \mathbb{Q}_0 \in \mathcal{P}_U$ satisfying Hypothesis 4.1. Let $(\mathbb{P}_k, \mathbb{Q}_k)_{k \geq 0}$ be given by the recursion (4.8) and just before (4.11).

Under the assumptions of Theorem 4.6, there exists a constant $C > 0$, which depends only on $c_\sigma, C_{b,\sigma}, C_{f,g}, d$ and T (and not on k, ϵ) such that for all $k \geq 1$,

$$0 \leq J(\mathbb{P}_k) - J^* \leq \epsilon C + \frac{C}{k} \left(1 + \frac{1}{\epsilon}\right). \quad (4.12)$$

Remark 4.8. We fix $\epsilon > 0$. By Corollary 4.7, approximating J^* with a precision ϵ requires at most $O(1/\epsilon^2)$ iterations of our alternating minimization procedure.

Proof of Corollary 4.7. Let $C_1 > 0$ be the constant appearing in the convergence rate in Theorem 4.6. Let also $C_2 = C^* > 0$ be the constant provided by Proposition 3.16 item 1. We recall that C_1 and C_2 depend only on $c_\sigma, C_{b,\sigma}, C_{f,g}, d$ and T . Let us fix $\epsilon' = \frac{C_1}{k} \left(1 + \frac{1}{\epsilon}\right)$. Theorem 4.6 states that $(\mathbb{P}_k, \mathbb{Q}_k) \in \mathcal{A}$ is an ϵ' -solution of the penalized Problem (1.5). Then, by Proposition 3.16 item 2., we have that

$$0 \leq J(\mathbb{P}_k) - J^* \leq \epsilon e^{\mathbb{E}[Y_k]} \text{Var}^{\mathbb{P}_k}[Y_k] + \frac{C_1}{k} \left(1 + \frac{1}{\epsilon}\right) \leq \epsilon C_2 e^{C_2} + \frac{C_1}{k} \left(1 + \frac{1}{\epsilon}\right), \quad (4.13)$$

where

$$Y_k = \int_0^T f(r, X_r, u^k(r, X_r)) dr + g(X_T),$$

and (4.12) follows from (4.13) setting $C = C_1 \vee C_2 e^{C_2}$. \square

Besides Lemma 4.4, the proof of Theorem 4.6 uses the so called three and four points properties introduced in [14].

Lemma 4.9. (Three points property). We suppose the validity of the hypotheses of Theorem 4.6. For all $\mathbb{Q} \in \mathcal{P}(\Omega)$,

$$\frac{1}{\epsilon} H(\mathbb{Q}|\mathbb{Q}_{k+1}) + \mathcal{J}(\mathbb{Q}_{k+1}, \mathbb{P}_k) \leq \mathcal{J}(\mathbb{Q}, \mathbb{P}_k). \quad (4.14)$$

Proof. We can suppose that $H(\mathbb{Q}|\mathbb{P}_k) < +\infty$, otherwise $\mathcal{J}(\mathbb{Q}, \mathbb{P}_k) = +\infty$ and the inequality holds trivially. Let

$$\varphi : X \mapsto \int_0^T f(r, X_r, u^k(r, X_r)) dr + g(X_T),$$

where u^k (and \mathbb{P}_k) have been defined in (4.11) and just before. By the definition (4.8) we have

$$\frac{d\mathbb{Q}_{k+1}}{d\mathbb{P}_k} = \frac{\exp(-\epsilon\varphi(X))}{\mathbb{E}^{\mathbb{P}_k}[\exp(-\epsilon\varphi(X))]}.$$

Now $d\mathbb{Q}_{k+1}/d\mathbb{P}_k > 0$ implies $\mathbb{Q}_{k+1} \sim \mathbb{P}_k$, hence taking into account $H(\mathbb{Q}|\mathbb{P}_k) < +\infty$, we have that $\mathbb{Q} \ll \mathbb{Q}_{k+1}$, so that, \mathbb{Q} -a.s.,

$$\log \frac{d\mathbb{Q}}{d\mathbb{P}_k} = \log \frac{d\mathbb{Q}}{d\mathbb{Q}_{k+1}} + \log \frac{d\mathbb{Q}_{k+1}}{d\mathbb{P}_k} = \log \frac{d\mathbb{Q}}{d\mathbb{Q}_{k+1}} - \epsilon\varphi(X) - \log \mathbb{E}^{\mathbb{P}_k}[\exp(-\epsilon\varphi(X))].$$

Taking the expectation under \mathbb{Q} in the previous equality and dividing both sides by $\epsilon > 0$, yields

$$\begin{aligned} \frac{1}{\epsilon} H(\mathbb{Q}|\mathbb{Q}_{k+1}) &= \frac{1}{\epsilon} H(\mathbb{Q}|\mathbb{P}_k) + \frac{1}{\epsilon} \log \mathbb{E}^{\mathbb{P}_k} [\exp(-\epsilon\varphi(X))] + \mathbb{E}^{\mathbb{Q}}[\varphi(X)] \\ &= \mathcal{J}(\mathbb{Q}, \mathbb{P}_k) - \mathcal{J}(\mathbb{Q}_{k+1}, \mathbb{P}_k), \end{aligned}$$

where we have used Lemma 4.4 item 1. for the latter equality. \square

Remark 4.10. Whenever $H(\mathbb{Q}|\mathbb{P}_k) < +\infty$, previous proof shows that (4.14) is indeed an equality.

Lemma 4.11. (Four points property). We suppose the validity of the hypotheses of Theorem 4.6. For all $(\mathbb{P}, \mathbb{Q}) \in \mathcal{A}$,

$$\mathcal{J}(\mathbb{Q}, \mathbb{P}_{k+1}) \leq \frac{1}{\epsilon} H(\mathbb{Q}|\mathbb{Q}_{k+1}) + \mathcal{J}(\mathbb{Q}, \mathbb{P}). \quad (4.15)$$

Proof. Let $(\mathbb{P}, \mathbb{Q}) \in \mathcal{A}$. If $H(\mathbb{Q}|\mathbb{Q}_{k+1}) = +\infty$ or $\mathcal{J}(\mathbb{Q}, \mathbb{P}) = +\infty$, the inequality is trivial. We then assume until the end of the proof that $H(\mathbb{Q}|\mathbb{Q}_{k+1}) < +\infty$ and $\mathcal{J}(\mathbb{Q}, \mathbb{P}) < +\infty$.

We first do some preliminary calculations. We recall that, by (4.9), there exists a measurable function $\beta^{k+1} : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that under \mathbb{Q}_{k+1} the canonical process has decomposition

$$X_t = x + \int_0^t \beta^{k+1}(r, X_r) dr + M_t^{\mathbb{Q}_{k+1}},$$

where $M^{\mathbb{Q}_{k+1}}$ is a local martingale under \mathbb{Q}_{k+1} and $[M^{\mathbb{Q}_{k+1}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$. We provide now a useful lower bound for $H(\mathbb{Q}|\mathbb{Q}_{k+1})$. By Lemma A.4 item 1. in the Appendix applied with $\mathbb{P} = \mathbb{Q}_{k+1}$ and the fact that $H(\mathbb{Q}|\mathbb{Q}_{k+1}) < +\infty$, there exists an (\mathcal{F}_t) -progressively measurable process $\alpha = \alpha(\cdot, X)$ such that, under \mathbb{Q} , the canonical process has the decomposition

$$X_t = x + \int_0^t \beta^{k+1}(r, X_r) dr + \int_0^t \sigma \sigma^\top(r, X_r) \alpha(r, X) dr + M_t^{\mathbb{Q}}, \quad (4.16)$$

where $M^{\mathbb{Q}}$ is a local martingale such that $[M^{\mathbb{Q}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$, and

$$H(\mathbb{Q}|\mathbb{Q}_{k+1}) \geq \frac{1}{2} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^\top(r, X_r) \alpha(r, X)|^2 dr \right]. \quad (4.17)$$

We set

$$\hat{\beta}(t, X) := \beta^{k+1}(t, X_t) + \sigma \sigma^\top(t, X_t) \alpha(t, X), \quad (4.18)$$

so that (4.17) can be rewritten

$$H(\mathbb{Q}|\mathbb{Q}_{k+1}) \geq \frac{1}{2} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^{-1}(r, X_r) (\hat{\beta}(r, X) - \beta^{k+1}(r, X_r))|^2 dr \right], \quad (4.19)$$

where we recall that σ^{-1} is the right-inverse of σ .

We proceed now with the proof of the four points property (4.15). Let u^{k+1} and $u^{\mathbb{P}_{k+1}}$ be as in (4.11) and just before so that $u^{k+1} = u^{\mathbb{P}_{k+1}}$. We set

$$\begin{aligned} y_r^{\mathbb{P}} &:= b(r, X_r, \nu^{\mathbb{P}}(r, X)), & y_r^{k+1} &:= b(r, X_r, u^{\mathbb{P}_{k+1}}(r, X_r)), \\ z_r^{\mathbb{P}} &:= f(r, X_r, \nu^{\mathbb{P}}(r, X)), & z_r^{k+1} &:= f(r, X_r, u^{\mathbb{P}_{k+1}}(r, X_r)), \end{aligned} \quad (4.20)$$

where $\nu^{\mathbb{P}}$ (resp. $u^{\mathbb{P}_{k+1}}$) is associated to \mathbb{P} (resp. \mathbb{P}_{k+1}), according to Definition 3.2. Let $\hat{F}_{\hat{\beta}}$ defined in Remark 4.3. Then

$$\begin{aligned} \hat{F}_{\hat{\beta}}(r, X, \nu_r^{\mathbb{P}}) - \hat{F}_{\hat{\beta}}(r, X, u^{\mathbb{P}_{k+1}}(r, X_r)) &= z_r^{\mathbb{P}} - z_r^{k+1} + \frac{1}{2\epsilon} |\sigma^{-1}(r, X_r)(\hat{\beta}(r, X) - y_r^{\mathbb{P}})|^2 \\ &\quad - \frac{1}{2\epsilon} |\sigma^{-1}(r, X_r)(\hat{\beta}(r, X) - y_r^{k+1})|^2. \end{aligned} \quad (4.21)$$

We focus on the last two terms in the previous inequality. We apply the algebraic equality $|a|^2 - |b|^2 = |a - b|^2 + 2\langle a - b, b \rangle$, with $a = \sigma^{-1}(\hat{\beta} - y^{\mathbb{P}})$, $b = \sigma^{-1}(\hat{\beta} - y^{k+1})$, where for conciseness we have omitted the dependencies in (r, X) of all the quantities at hand. So we have

$$\frac{1}{2\epsilon} |\sigma^{-1}(\hat{\beta} - y^{\mathbb{P}})|^2 - \frac{1}{2\epsilon} |\sigma^{-1}(\hat{\beta} - y^{k+1})|^2 = \frac{1}{2\epsilon} |\sigma^{-1}(y^{\mathbb{P}} - y^{k+1})|^2 + \frac{1}{\epsilon} \langle \sigma^{-1}(y^{\mathbb{P}} - y^{k+1}), \sigma^{-1}(y^{k+1} - \hat{\beta}) \rangle.$$

On the other hand,

$$\begin{aligned} \frac{1}{\epsilon} \langle \sigma^{-1}(y^{\mathbb{P}} - y^{k+1}), \sigma^{-1}(y^{k+1} - \hat{\beta}) \rangle &= \frac{1}{\epsilon} \langle \sigma^{-1}(y^{\mathbb{P}} - y^{k+1}), \sigma^{-1}(y^{k+1} - \beta^{k+1}) \rangle \\ &\quad + \frac{1}{\epsilon} \langle \sigma^{-1}(y^{\mathbb{P}} - y^{k+1}), \sigma^{-1}(\beta^{k+1} - \hat{\beta}) \rangle. \end{aligned}$$

Combining what precedes yields

$$\begin{aligned} \frac{1}{2\epsilon} |\sigma^{-1}(\hat{\beta} - y^{\mathbb{P}})|^2 - \frac{1}{2\epsilon} |\sigma^{-1}(\hat{\beta} - y^{k+1})|^2 &= \frac{1}{2\epsilon} |\sigma^{-1}(y^{\mathbb{P}} - y^{k+1})|^2 + \frac{1}{\epsilon} \langle y^{\mathbb{P}} - y^{k+1}, (\sigma^{-1})^\top \sigma^{-1}(y^{k+1} - \beta^{k+1}) \rangle \\ &\quad + \frac{1}{\epsilon} \langle \sigma^{-1}(y^{\mathbb{P}} - y^{k+1}), \sigma^{-1}(\beta^{k+1} - \hat{\beta}) \rangle. \end{aligned}$$

From the inequality (4.21) we then get

$$\begin{aligned} \hat{F}_{\hat{\beta}}(r, X, \nu_r^{\mathbb{P}}) - \hat{F}_{\hat{\beta}}(r, X, u^{k+1}(r, X_r)) &= \frac{1}{2\epsilon} |\sigma^{-1}(r, X_r)(y_r^{\mathbb{P}} - y_r^{k+1})|^2 \\ &\quad + \frac{1}{\epsilon} \langle \sigma^{-1}(r, X_r)(\beta^{k+1}(r, X_r) - \hat{\beta}(r, X)), \sigma^{-1}(r, X_r)(y_r^{\mathbb{P}} - y_r^{k+1}) \rangle \\ &\quad + z_r^{\mathbb{P}} - z_r^{k+1} + \frac{1}{\epsilon} \left\langle (\sigma^{-1})^\top \sigma^{-1}(r, X_r)(y_r^{k+1} - \beta^{k+1}(r, X_r)), y_r^{\mathbb{P}} - y_r^{k+1} \right\rangle. \end{aligned} \quad (4.22)$$

By (4.10) $u^{k+1}(t, x)$ achieves the minimum of $F_{\beta^{k+1}}(t, x, \cdot)$, for all $(t, x) \in [0, T] \times \mathbb{R}^d$, where the application $F_{\beta^{k+1}}$ is the one defined in (4.2). Taking into account (4.10), Lemma 4.2 2(b), applied for any $(t, x) \in [0, T] \times \mathbb{R}^d$ with $u^* = u^{k+1}(t, x)$ and $\beta = \beta^{k+1}$, states that the restriction of the function $\bar{F}_{\beta^{k+1}}^{t, x}$ given by (4.3) to the convex set $K(t, x)$, achieves its minimum at the point $(b(t, x, u^{k+1}(t, x)), f(t, x, u^{k+1}(t, x)))$. Consequently, for the generic probability measure \mathbb{P} , Lemma 4.2 item 1., applied with $\beta = \beta^{k+1}$, $(y^*, z^*) = (y_r^{k+1}, z_r^{k+1})$ and $(y, z) = (y_r^{\mathbb{P}}, z_r^{\mathbb{P}})$, shows that (y_r^{k+1}, z_r^{k+1}) is the unique minimum of $\bar{F}_{\beta^{k+1}}^{t, x}$, so that the term on the third line of the equality (4.22) is non-negative. Then (4.22) yields

$$\begin{aligned} \hat{F}_{\hat{\beta}}(r, X, \nu_r^{\mathbb{P}}) - \hat{F}_{\hat{\beta}}(r, X, u^{k+1}(r, X_r)) dr &\geq \frac{1}{2\epsilon} |\sigma^{-1}(r, X_r)(y_r^{\mathbb{P}} - y_r^{k+1})|^2 \\ &\quad + \frac{1}{\epsilon} \langle \sigma^{-1}(r, X_r)(\beta^{k+1}(r, X_r) - \hat{\beta}(r, X)), \sigma^{-1}(r, X_r)(y_r^{\mathbb{P}} - y_r^{k+1}) \rangle. \end{aligned} \quad (4.23)$$

Next, by the classical inequality $|ab| \leq a^2/2 + b^2/2$ for all $(a, b) \in \mathbb{R}^2$, the right-hand side term in inequality (4.23) gives

$$\begin{aligned} & \frac{1}{\epsilon} \langle \sigma^{-1}(r, X_r)(\beta^{k+1}(r, X_r) - \hat{\beta}(r, X)), \sigma^{-1}(r, X_r)(y_r^{\mathbb{P}} - y_r^{k+1}) \rangle \\ & \geq -\frac{1}{2\epsilon} |\sigma^{-1}(r, X_r)(\hat{\beta}(r, X) - \beta^{k+1}(r, X_r))|^2 \\ & \quad - \frac{1}{2\epsilon} |\sigma^{-1}(r, X_r)(y_r^{\mathbb{P}} - y_r^{k+1})|^2, \end{aligned}$$

and from inequality (4.23) we get

$$\hat{F}_{\hat{\beta}}(r, X, \nu_r^{\mathbb{P}}) + \frac{1}{2\epsilon} |\sigma^{-1}(r, X_r)(\hat{\beta}(r, X) - \beta^{k+1}(r, X_r))|^2 \geq \hat{F}_{\hat{\beta}}(r, X, u^{k+1}(r, X_r)).$$

Integrating previous inequality with respect to $r \in [0, T]$, yields

$$\int_0^T \hat{F}_{\hat{\beta}}(r, X, \nu_r^{\mathbb{P}}) dr + \frac{1}{2\epsilon} \int_0^T |\sigma^{-1}(r, X_r)(\hat{\beta}(r, X) - \beta^{k+1}(r, X_r))|^2 dr \geq \int_0^T \hat{F}_{\hat{\beta}}(r, X, u^{k+1}(r, X_r)) dr. \quad (4.24)$$

Since $H(\mathbb{Q}|\mathbb{P}) < +\infty$, by Definition 3.2, Lemma A.4 item 1. with $\delta(\cdot, X) = b(\cdot, X, \nu^{\mathbb{P}}(\cdot, X))$ (writing $\nu^{\mathbb{P}}(r, X) = \nu_r^{\mathbb{P}}$), states the existence of a predictable process $\tilde{\alpha}$ such that

$$X_t = x + \int_0^t b(r, X_r, \nu_r^{\mathbb{P}}) dr + \int_0^t \sigma \sigma^\top(r, X_r) \tilde{\alpha}(r, X) dr + \tilde{M}_t^{\mathbb{Q}}, \quad (4.25)$$

where $\tilde{M}^{\mathbb{Q}}$ is a $(\mathbb{Q}, \mathcal{F}_t)$ -local martingale with $[\tilde{M}^{\mathbb{Q}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$. By (4.16) and (4.18), under \mathbb{Q} , the canonical process decomposes as

$$X_t = x + \int_0^t \hat{\beta}(r, X) dr + M_t^{\mathbb{Q}}, \quad (4.26)$$

where $M^{\mathbb{Q}}$ is a local martingale verifying $[M^{\mathbb{Q}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$.

Identifying the bounded variation component between (4.25) and decomposition (4.26) (under \mathbb{Q}), yields $\hat{\beta}(r, X) - b(r, X_r, \nu_r^{\mathbb{P}}) = \sigma \sigma^\top(r, X_r) \tilde{\alpha}(r, X)$ and (A.20) in Lemma A.4 item 1. implies that

$$H(\mathbb{Q}|\mathbb{P}) \geq \frac{1}{2} \mathbb{E}^{\mathbb{P}} \left[\int_0^T |\sigma^{-1}(r, X_r)(b(r, X_r, \nu_r^{\mathbb{P}}) - \hat{\beta}(r, X))|^2 dr \right]. \quad (4.27)$$

Then recalling the definition of \mathcal{J} in (1.5), previous inequality (4.27) yields

$$\mathcal{J}(\mathbb{Q}, \mathbb{P}) \geq \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \hat{F}_{\hat{\beta}}(r, X, \nu_r^{\mathbb{P}}) dr + g(X_T) \right]. \quad (4.28)$$

From (4.28) and (4.19) it holds

$$\begin{aligned} \mathcal{J}(\mathbb{Q}, \mathbb{P}) + \frac{1}{\epsilon} H(\mathbb{Q}|\mathbb{Q}_{k+1}) & \geq \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \hat{F}_{\hat{\beta}}(r, X, \nu_r^{\mathbb{P}}) dr + g(X_T) \right] \\ & \quad + \frac{1}{2\epsilon} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^{-1}(r, X_r)(\hat{\beta}(r, X) - \beta^{k+1}(r, X_r))|^2 dr \right], \end{aligned}$$

and by (4.24)

$$\mathcal{J}(\mathbb{Q}, \mathbb{P}) + \frac{1}{\epsilon} H(\mathbb{Q}|\mathbb{Q}_{k+1}) \geq \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \hat{F}_{\hat{\beta}}(r, X, u^{k+1}(r, X_r)) dr + g(X_T) \right]. \quad (4.29)$$

In particular, since $g \geq 0$, we have $\mathbb{E}^{\mathbb{Q}} \left[\int_0^T \hat{F}_{\hat{\beta}}(r, X, u^{k+1}(r, X_r)) dr \right] < +\infty$, hence, recalling the expression (4.7)

$$\mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^{-1}(r, X_r)(b(r, X_r, u^{k+1}(r, X_r)) - \hat{\beta}(r, X))|^2 dr \right] < +\infty.$$

We keep in mind (4.16) and (4.18). By Lemma 4.4 item 2., the decomposition (4.11) is unique in law. Then, by Lemma A.4 item 2. applied to $\mathbb{P} = \mathbb{P}_{k+1} (= \mathbb{P}^{u^{k+1}})$ with $\delta(\cdot, X) = b(\cdot, X, u^{k+1}(\cdot, X))$ and $\gamma = \hat{\beta}$, we have

$$H(\mathbb{Q}|\mathbb{P}_{k+1}) = \frac{1}{2} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^{-1}(r, X_r)(b(r, X_r, u^{k+1}(r, X_r)) - \hat{\beta}(r, X))|^2 dr \right], \quad (4.30)$$

and so

$$\mathcal{J}(\mathbb{Q}, \mathbb{P}_{k+1}) = \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \hat{F}_{\hat{\beta}}(r, X, u^{k+1}(r, X_r)) dr + g(X_T) \right]. \quad (4.31)$$

Finally, combining (4.29) and (4.31), we get

$$\mathcal{J}(\mathbb{Q}, \mathbb{P}) + \frac{1}{\epsilon} H(\mathbb{Q}|\mathbb{Q}_{k+1}) \geq \mathcal{J}(\mathbb{Q}, \mathbb{P}_{k+1}).$$

This concludes the proof. \square

Lemma 4.12. *Let $(\mathbb{P}_\epsilon^*, \mathbb{Q}_\epsilon^*)$ be an optimal solution to Problem (1.5), given by Theorem 3.15, under the assumptions of the aforementioned theorem. Let $(\mathbb{P}_k, \mathbb{Q}_k)_{k \geq 0}$ be given by the recursion (4.8) and just before (4.11).*

There exists a constant $C > 0$, which only depends on $c_\sigma, C_{b,\sigma}, C_{f,g}, d$ and T , such that, for all $k \geq 0$, $\mathcal{J}(\mathbb{Q}_\epsilon^, \mathbb{P}_k) \leq C \left(1 + \frac{1}{\epsilon}\right)$.*

The proof of the result above is postponed to Appendix E for clarity.

Proof of Theorem 4.6. Combining (4.14) in Lemma 4.9 and (4.15) in Lemma 4.11 we get, for all $k \geq 0$, the so called five points property

$$\mathcal{J}(\mathbb{Q}, \mathbb{P}_{k+1}) + \mathcal{J}(\mathbb{Q}_{k+1}, \mathbb{P}_k) \leq \mathcal{J}(\mathbb{Q}, \mathbb{P}_k) + \mathcal{J}(\mathbb{Q}, \mathbb{P}). \quad (4.32)$$

Evaluating (4.32) for (\mathbb{P}, \mathbb{Q}) being the solution $(\mathbb{P}_\epsilon^*, \mathbb{Q}_\epsilon^*)$ of the penalized problem given by Theorem 3.15, we obtain

$$\mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_{k+1}) + \mathcal{J}(\mathbb{Q}_{k+1}, \mathbb{P}_k) \leq \mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_k) + \mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_\epsilon^*), \quad (4.33)$$

and since $\mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_k) < +\infty$, by Lemma 4.12, the previous inequality rewrites

$$\mathcal{J}(\mathbb{Q}_{k+1}, \mathbb{P}_k) - \mathcal{J}^* \leq \mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_k) - \mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_{k+1}), \quad (4.34)$$

where we have used the equality $\mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_\epsilon^*) = \mathcal{J}^*$.

Let K be a fixed number of iterations of the algorithm. Summing equation (4.34) between 0 and $K - 1$ and dividing each member of the inequality by K , we get

$$\frac{1}{K} \sum_{k=0}^{K-1} \mathcal{J}(\mathbb{Q}_{k+1}, \mathbb{P}_k) - \mathcal{J}^* \leq \frac{1}{K} (\mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_0) - \mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_K)). \quad (4.35)$$

By construction of the sequence $(\mathbb{P}_k, \mathbb{Q}_k)_{k \geq 0}$, it holds that

$$\mathcal{J}(\mathbb{Q}_{k+1}, \mathbb{P}_k) \geq \mathcal{J}(\mathbb{Q}_{k+1}, \mathbb{P}_{k+1}) \geq \mathcal{J}(\mathbb{Q}_{k+2}, \mathbb{P}_{k+1}) \geq \dots \geq \mathcal{J}(\mathbb{Q}_K, \mathbb{P}_K) \quad (4.36)$$

for all $k \leq K - 1$. Applying (4.36) in (4.35) gives

$$0 \leq \mathcal{J}(\mathbb{Q}_K, \mathbb{P}_K) - \mathcal{J}^* \leq \frac{1}{K} (\mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_0) - \mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_K)) \leq \frac{1}{K} \mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_0). \quad (4.37)$$

Finally, by Lemma 4.12, there exists a constant $C > 0$, which only depends on $c_\sigma, C_{b,\sigma}, C_{f,g}, d$ and T , such that $\mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_0) \leq C(1 + \frac{1}{\epsilon})$ and (4.37) yields

$$0 \leq \mathcal{J}(\mathbb{Q}_K, \mathbb{P}_K) - \mathcal{J}^* \leq \frac{C}{K} \left(1 + \frac{1}{\epsilon}\right).$$

Previous relation proves the convergence of the algorithm and exhibits a convergence rate for fixed ϵ . This concludes the proof. \square

Remark 4.13. *To prove Lemma 4.9 and Lemma 4.11, one can relax the continuity assumption on b in Hypothesis 3.1 and assume instead that $b(t, x, \cdot)$ is continuous for all $(t, x) \in [0, T] \times \mathbb{R}^d$ and $b(\cdot, \cdot, u)$ is measurable for all $u \in \mathbb{U}$. Then \mathcal{J} verifies the so-called five point property (4.32) and Theorem 2 in [14] ensures that $\mathcal{J}(\mathbb{Q}_k, \mathbb{P}_k) \xrightarrow[k \rightarrow +\infty]{} \mathcal{J}^*$. However our proof of Theorem 3.15 strongly relies on the continuity of b in $(t, x, u) \in [0, T] \times \mathbb{R}^d \times \mathbb{U}$, and this stronger regularity allows to exhibit a convergence rate in Theorem 4.6.*

We conclude the section by stating a lemma which is a reformulation in our setting of Proposition 3.9 in [5]. This allows us to estimate the drift β^k in the algorithm via a conditional derivative.

Lemma 4.14. *Assume Hypothesis 3.1 and that a probability \mathbb{P}_0 verifies Hypothesis 4.1. Consider the sequence constructed after Remark 4.3. For almost all $0 \leq t < T$, it holds that*

$$\lim_{h \downarrow 0} \mathbb{E}^{\mathbb{Q}_k} \left[\frac{X_{t+h} - X_t}{h} \mid X_t \right] = \beta^k(t, X_t) \text{ in } L^1(\mathbb{Q}_k). \quad (4.38)$$

Proof. We fix some $1 < p < 2$. By decomposition (4.9), replacing $k + 1$ with k , in order to apply Lemma F.2, it is enough to have $\|\beta^k\|_{L^p(dt \otimes \mathbb{Q}_k)} < +\infty$, which is guaranteed by item 1. of Lemma 4.4. Consequently Lemma F.2 and Remark F.3 yield the result. \square

Remark 4.15. *Our algorithm has the advantage of relying on two standard optimization sub-problems that are simpler than the original stochastic control problem: on the one hand, an exponential twist problem (4.8) and, on the other hand, a convex pointwise optimization problem (4.10). From a numerical point of view, each of the subproblems can be solved numerically by specific approaches. For example, solving the exponential twist problem can be reduced to computing independent conditional expectations on each time step, as shown in Lemma 4.14, hence those computations can be easily parallelized. Our method has a clear advantage with respect to HJB or BSDEs representation of the solution of control problems, which involve nonlinearly nested conditional expectations, because of the backward dynamical programming recursion. In our context the conditional expectation computations can be also efficiently addressed by deep learning methods when the dimension is high, see e.g. [12, 21, 24, 27]. However, in the numerical applications considered in Section 6, we choose to use a simple polynomial regression Monte-Carlo method, since we restrict ourselves to a dimension less than 20. This linearization effect has of course the cost of repeating the procedure along k iterations, allowing a convergence rate $O(\frac{1}{k})$.*

4.2 Entropy penalized Monte-Carlo algorithm

The alternating minimization procedure in Section 4.1 suggests a Monte-Carlo algorithm to approximate a solution to Problem (1.3). In the following, $0 = t_0 \leq t_1 < \dots < t_M = T$, is a regular subdivision of the time interval $[0, T]$ with step Δt , $N \geq 0$ being the number of particles and K the number of descent steps of the algorithm. P_r will denote the set of \mathbb{R}^d -valued polynomials defined on \mathbb{R}^d of degree $\leq r$. Recall that for all $\hat{u} \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$, $\mathbb{P}^{\hat{u}}$ is the probability measure given by Proposition 3.5. The estimation of the drift $\hat{\beta}^k$ in Step 2 of the algorithm below, is performed via regression. This is inspired by (4.38) in Lemma 4.14. The term in the argmin is a weighted Monte-Carlo approximation of the expectation of $\frac{X_{m+1}^n - X_m^n}{\Delta t}$, under the exponential twist of the probability measure $\mathbb{P}^{\hat{u}^{k-1}}$.

Algorithm 1 Entropy penalized Monte-Carlo algorithm

Parameters initialization: $M, N, K \in \mathbb{N}^*$, $r \in \mathbb{N}$, $\Delta t := \frac{T}{M}$, $x \in \mathbb{R}^d$, $\hat{u}^0 \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$.

Simulate: $(X^n)_{1 \leq n \leq N}$, N iid Monte-Carlo path simulations under $\hat{\mathbb{P}}_0 = \mathbb{P}^{\hat{u}^0}$ on the time-grid $(t_m)_{0 \leq m \leq M}$, with $X^n = (X_m^n)_{0 \leq m \leq M}$ and $X_0^n = x$, for all $1 \leq n \leq N$.

for $1 \leq k \leq K$ **do**

Step 1. Compute the weights $(D_n)_{1 \leq n \leq N}$ by

$$D_n = \exp \left(-\epsilon \sum_{m=0}^{M-1} f(t_m, X_m^n, \hat{u}^{k-1}(t_m, X_m^n)) \Delta t - \epsilon g(X_M^n) \right).$$

Step 2. Compute $\hat{\beta}^k = (\hat{\beta}_m^k)_{0 \leq m \leq M-1}$ in (4.9) by the weighted Monte-Carlo approximation of (4.38)

$$\hat{\beta}_m^k \in \arg \min_{\varphi \in \mathcal{P}_r} \frac{1}{\sum_{\ell=1}^N D_\ell} \sum_{n=1}^N D_n \left| \varphi(X_m^n) - \frac{X_{m+1}^n - X_m^n}{\Delta t} \right|^2.$$

Step 3. Simulate new iid Monte-Carlo paths $(X^n)_{1 \leq n \leq N}$ under $\mathbb{P}^{\hat{u}^k}$, where for $0 \leq m \leq M-1$

$$\hat{u}^k(t, x) = \arg \min_{a \in \mathbb{U}} f(t_m, x, a) + \frac{1}{2\epsilon} |\sigma^{-1}(t_m, x)(\hat{\beta}_m^k(x) - b(t_m, x, a))|^2, \quad t \in [t_m, t_{m+1}[. \quad (4.39)$$

end for

return \hat{u}^K

Remark 4.16. The algorithm stores the functions $\hat{\beta}^k$, from which the controls are computed. In our implementation these functions are polynomial regressors (whose coefficients are stored at each time steps) but one could also imagine storing them in the form of neural networks or any other machine learning models. The algorithm actually returns $\hat{\beta}^k$ after k iterations, from which the feedback \hat{u}^k can be evaluated in each point (t, x) by solving the minimization problem (4.39), which defines a measurable function, by Proposition 5.2 below. Thus an optimal feedback control is an output of the algorithm.

An interest of the entropy penalized Monte-Carlo algorithm is that in Lemma 4.14, (4.38) can be independently estimated by regression techniques at each time step t_m , $1 \leq m \leq M$, while in dynamic programming approaches, conditional expectations are recursively computed in time, implying an error accumulation from time $t_M = T$ to t_m . Moreover one can expect that the trajectories simulated under $\mathbb{P}^{\hat{u}^k}$, localize around the optimally controlled trajectories, when the number of iterations k of the algorithm increases to $+\infty$. Hence, the computational effort to estimate the optimal control, focuses on this specific region of the state space, whereas standard regression based Monte-Carlo approaches are blindly exploring the state space, with forward Monte-Carlo simulations of the process.

5 Solving the subproblems

In this section we aim at describing the two subproblems $\inf_{\mathbb{P} \in \mathcal{P}_{\mathbb{U}}} \mathcal{J}(\mathbb{Q}, \mathbb{P})$ and $\inf_{\mathbb{Q} \in \mathcal{P}_{\Omega}} \mathcal{J}(\mathbb{Q}, \mathbb{P})$ appearing in the alternating minimization algorithm proposed in Section 4.

5.1 Pointwise minimization subproblem

Let us first describe the minimization $\inf_{\mathbb{P} \in \mathcal{P}_{\mathbb{U}}} \mathcal{J}(\mathbb{Q}, \mathbb{P})$ where the probability $\mathbb{Q} \in \mathcal{P}(\Omega)$ is fixed and is such that, under \mathbb{Q} , the canonical process is a fixed Itô process. In this section, we assume that Hypotheses 3.1, 3.7 and 3.9 are fulfilled. Let $p \geq 1$ be the real intervening in Hypothesis 3.7 item 1. In the sequel of the present section we also make a specific assumption for a given probability \mathbb{Q} on the canonical space.

Hypothesis 5.1. *There is a Borel function $\beta : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ for which the canonical process X decomposes as*

$$X_t = x + \int_0^t \beta(r, X_r) dr + M_t^{\mathbb{Q}}, \quad (5.1)$$

where $M^{\mathbb{Q}}$ is a local martingale verifying $[M^{\mathbb{Q}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$. Moreover, $\mathbb{E}^{\mathbb{Q}} \left[\sup_{0 \leq r \leq T} |X_r|^p \right] < +\infty$.

For the proposition below we recall that if $u : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{U}$ is a Borel function then $\mathbb{P}^u \in \mathcal{P}_{\mathbb{U}}^{Markov}$ denotes the associated probability measure given by Proposition 3.5.

Proposition 5.2. *There exists a measurable function $(t, x) \mapsto u(t, x) \in \mathbb{U}$ such that*

$$u(t, x) \in \arg \min_{a \in \mathbb{U}} F_{\beta}(t, x, a), \quad (5.2)$$

where F_{β} is given by (4.2), which is well-defined and measurable. Moreover $\mathcal{J}(\mathbb{Q}, \mathbb{P}^u) = \inf_{\mathbb{P} \in \mathcal{P}_{\mathbb{U}}} \mathcal{J}(\mathbb{Q}, \mathbb{P})$.

Proof of Proposition 5.2. We will make use of the function $\bar{F}_{\beta}^{t,x}$ defined by (4.3), defined for all $(t, x) \in [0, T] \times \mathbb{R}^d$. We also keep in mind the definition (3.9) of the convex set $K(t, x)$ where one will consider the restriction of $\bar{F}_{\beta}^{t,x}$.

For all $(t, x) \in [0, T] \times \mathbb{R}^d$ let us consider $(y^*(t, x), z^*(t, x)) \in K(t, x)$ given by Lemma 4.2 item 1. By Theorem B.1 there exists a measurable function $u \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$ such that $y^*(t, x) = b(t, x, u(t, x))$ and $z^*(t, x) \geq f(t, x, u(t, x))$. By Lemma 4.2 item 2.(a),

$$u(t, x) \in \arg \min_{a \in \mathbb{U}} F_{\beta}(t, x, a), \quad \forall (t, x). \quad (5.3)$$

By Proposition 3.5, there is a probability measure \mathbb{P}^u belonging to $\mathcal{P}_{\mathbb{U}}^{Markov}$. Let also $\mathbb{P} \in \mathcal{P}_{\mathbb{U}}$. In particular there exists a progressively measurable process $\nu_r^{\mathbb{P}}$, with values in \mathbb{U} such that under \mathbb{P} the canonical process X has decomposition

$$X_t = x + \int_0^t b(r, X_r, \nu_r^{\mathbb{P}}) dr + M_t^{\mathbb{P}}, \quad t \in [0, T],$$

where $M^{\mathbb{P}}$ is a local martingale verifying $[M^{\mathbb{P}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$. We want to prove that

$$\mathcal{J}(\mathbb{Q}, \mathbb{P}) \geq \mathcal{J}(\mathbb{Q}, \mathbb{P}^u). \quad (5.4)$$

If $\mathcal{J}(\mathbb{Q}, \mathbb{P}) = +\infty$, inequality (5.4) is trivially verified. Assume now that $\mathcal{J}(\mathbb{Q}, \mathbb{P}) < +\infty$. In particular, $H(\mathbb{Q}|\mathbb{P}) < +\infty$ and by Lemma A.4 item 1.(a), there exists a process $\alpha = \alpha(\cdot, X)$ such that under \mathbb{Q} , X decomposes as

$$X_t = x + \int_0^t b(r, X_r, \nu_r^{\mathbb{P}}) dr + \int_0^t \sigma \sigma^\top(r, X_r) \alpha_r dr + \tilde{M}_t^{\mathbb{Q}}, \quad (5.5)$$

where the local martingale $\tilde{M}^{\mathbb{Q}}$ verifies $[\tilde{M}^{\mathbb{Q}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$, and

$$H(\mathbb{Q}|\mathbb{P}) \geq \frac{1}{2} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^\top(r, X_r) \alpha(r, X)|^2 dr \right]. \quad (5.6)$$

Identifying the bounded variation and the local martingale parts in (5.1) and (5.5) yields $\sigma^\top(t, X_t) \alpha(t, X) = \sigma^{-1}(t, X_t) (\beta(t, X_t) - b(t, X_t, \nu_t^{\mathbb{P}})) d\mathbb{Q} \otimes dt$ -a.e. and $\tilde{M}^{\mathbb{Q}} = M^{\mathbb{Q}}$. Replacing in (5.6) we get

$$H(\mathbb{Q}|\mathbb{P}) \geq \frac{1}{2} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^{-1}(r, X_r) (\beta(r, X_r) - b(r, X_r, \nu_r^{\mathbb{P}}))|^2 dr \right],$$

and the previous inequality yields

$$\begin{aligned} \mathcal{J}(\mathbb{Q}, \mathbb{P}) &= \mathbb{E}^{\mathbb{Q}} \left[\int_0^T f(r, X_r, \nu_r^{\mathbb{P}}) dr + g(X_T) \right] + \frac{1}{\epsilon} H(\mathbb{Q}|\mathbb{P}) \\ &\geq \mathbb{E}^{\mathbb{Q}} \left[\int_0^T f(r, X_r, \nu_r^{\mathbb{P}}) dr + g(X_T) + \frac{1}{2\epsilon} \int_0^T |\sigma^{-1}(r, X_r) (\beta(r, X_r) - b(r, X_r, \nu_r^{\mathbb{P}}))|^2 dr \right]. \end{aligned} \quad (5.7)$$

By assumption, $\mathbb{E}^{\mathbb{Q}} \left[\sup_{0 \leq r \leq T} |X_r|^p \right] < +\infty$ and by (3.1) and (3.8) we have

$$\mathbb{E}^{\mathbb{Q}} \left[\int_0^T \left(|b(r, X_r, \nu_r^{\mathbb{P}})| + |f(r, X_r, \nu_r^{\mathbb{P}})| \right) dr \right] < +\infty. \quad (5.8)$$

An application of Fubini's theorem, the tower property and Jensen's inequality for conditional expectation in (5.7) gives

$$\begin{aligned} \mathcal{J}(\mathbb{Q}, \mathbb{P}) &\geq \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \mathbb{E}^{\mathbb{Q}} \left[f(r, X_r, \nu_r^{\mathbb{P}}) \middle| X_r \right] dr + g(X_T) \right. \\ &\quad \left. + \frac{1}{2\epsilon} \int_0^T \left| \sigma^{-1}(r, X_r) \left(\beta(r, X_r) - \mathbb{E}^{\mathbb{Q}} \left[b(r, X_r, \nu_r^{\mathbb{P}}) \middle| X_r \right] \right) \right|^2 dr \right]. \end{aligned} \quad (5.9)$$

Since (5.8) holds, Lemma B.2 applied with $(y_t, z_t) = (b(t, X_t, \nu_t^{\mathbb{P}}), f(t, X_t, \nu_t^{\mathbb{P}}))$ gives the existence of a function $v \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$ such that for almost all $t \in [0, T]$, \mathbb{P} -a.s.

$$\begin{cases} \mathbb{E}^{\mathbb{Q}} \left[b(t, X_t, \nu_t^{\mathbb{P}}) \middle| X_t \right] = b(t, X_t, v(t, X_t)) \\ \mathbb{E}^{\mathbb{Q}} \left[f(t, X_t, \nu_t^{\mathbb{P}}) \middle| X_t \right] \geq f(t, X_t, v(t, X_t)). \end{cases} \quad (5.10)$$

Injecting (5.10) in (5.9) we get

$$\mathcal{J}(\mathbb{Q}, \mathbb{P}) \geq \mathbb{E}^{\mathbb{Q}} \left[\int_0^T f(r, X_r, v(r, X_r)) dr + g(X_T) \right] + \frac{1}{2\epsilon} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^{-1}(r, X_r)(\beta(r, X_r) - b(r, X_r, v(r, X_r)))|^2 dr \right].$$

The previous inequality rewrites

$$\mathcal{J}(\mathbb{Q}, \mathbb{P}) \geq \mathbb{E}^{\mathbb{Q}} \left[\int_0^T F_{\beta}(r, X_r, v(r, X_r)) dr + g(X_T) \right],$$

where we recall that F_{β} was defined in (4.2). By (5.3), for all $t \in [0, T]$ we have

$$F_{\beta}(t, X_t, v(t, X_t)) \geq F_{\beta}(t, X_t, u(t, X_t)) \quad \mathbb{Q}\text{-a.s.},$$

hence

$$\mathcal{J}(\mathbb{Q}, \mathbb{P}) \geq \mathbb{E}^{\mathbb{Q}} \left[\int_0^T F_{\beta}(r, X_r, v(r, X_r)) dr + g(X_T) \right] \geq \mathbb{E}^{\mathbb{Q}} \left[\int_0^T F_{\beta}(r, X_r, u(r, X_r)) dr + g(X_T) \right]. \quad (5.11)$$

In particular,

$$\mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^{-1}(r, X_r)(b(r, X_r, u(t, X_t)) - \beta(r, X_r))|^2 dr \right] < +\infty.$$

By Remark 3.6 the equation (3.7) admits a unique solution. Therefore we can apply item 2. of Lemma A.4 with $\delta(t, X) = b(t, X_t, u(t, X_t))$ and $\gamma(t, X) = \beta(t, X_t)$, and we have

$$H(\mathbb{Q}|\mathbb{P}) = \frac{1}{2} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^{-1}(r, X_r)(b(r, X_r, u(t, X_t)) - \beta(r, X_r))|^2 dr \right],$$

hence

$$\mathbb{E}^{\mathbb{Q}} \left[\int_0^T F_{\beta}(r, X_r, u(r, X_r)) dr + g(X_T) \right] = \mathcal{J}(\mathbb{Q}, \mathbb{P}^u)$$

and previous inequality along with (5.11) yields $\mathcal{J}(\mathbb{Q}, \mathbb{P}) \geq \mathcal{J}(\mathbb{Q}, \mathbb{P}^u)$. \square

5.2 Exponential twist subproblem

In this section we focus on the minimization $\inf_{\mathbb{Q} \in \mathcal{P}(\Omega)} \mathcal{J}(\mathbb{Q}, \mathbb{P})$, $\mathbb{P} \in \mathcal{P}_{\mathbb{U}}^{Markov}$ being the reference probability. Let us denote \mathbb{Q}^* the solution of that problem given by Proposition 3.13.

Proposition 5.3. *Assume that, under \mathbb{P} , the canonical process decomposes as $X_t = x + \int_0^t b(r, X_r, u(r, X_r)) dr + M_t^{\mathbb{P}}$, where $M^{\mathbb{P}}$ is a local martingale such that $[M^{\mathbb{P}}] = \int_0^t \sigma \sigma^{\top}(r, X_r) dr$ and $u \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$. Then there exists $\beta \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{R}^d)$ such that, under \mathbb{Q}^* , the canonical process decomposes as*

$$X_t = x + \int_0^t \beta(r, X_r) dr + M_t^{\mathbb{Q}^*}, \quad (5.12)$$

where $M^{\mathbb{Q}^*}$ is a local martingale such that $[M^{\mathbb{Q}^*}] = \int_0^t \sigma \sigma^{\top}(r, X_r) dr$. Moreover, X is a Markov process under \mathbb{Q}^* and $\beta \in \mathcal{B}([0, T] \times \mathbb{R}^d; \mathbb{R}^d)$ such that $|\beta| \in L^q(dt \otimes d\mathbb{Q}^*)$, for all $1 \leq q < 2$.

Proof. Recall that by Remark 3.4, \mathbb{P} is a solution in law of the SDE

$$dX_t = b(t, X_t, u(t, X_t))dt + \sigma(t, X_t)dW_t, \quad X_0 = x.$$

By Corollary 6.12 in [7] with $\mathbb{Q} = \mathbb{Q}^*$, there is $\lambda \in \mathcal{B}([0, T] \times \mathbb{R}^d; \mathbb{R}^d)$ belonging to $L^q(dt \otimes d\mathbb{Q}^*)$, for all $1 \leq q < 2$, such that (5.12) holds with

$$\beta(t, x) = b(t, x, u(t, x)) + \lambda(t, x).$$

$(t, \omega) \mapsto b(t, X_t(\omega), u(t, X_t(\omega))) \in L^q(dt \otimes d\mathbb{Q}^*)$, taking into account (3.1). The result follows again by Remark 3.4. \square

6 Application to the control of thermostatic loads in power systems

We consider in this section the problem of controlling a large, heterogeneous population of N air-conditioners in order that their overall consumption tracks a given target profile $r = (r_t)_{0 \leq t \leq T}$, on a given time horizon $[0, T]$. This problem was introduced in [28]. Air-conditioners are aggregated in d clusters indexed by $1 \leq i \leq d$, depending on their characteristics. We denote by N_i the number of air-conditioners in the cluster i . Individually, the temperature $X^{i,j}$ in the room with air-conditioner j in cluster i , is assumed to evolve according to the dynamics

$$dX_t^{i,j} = -\theta^i(X_t^{i,j} - x_{out}^i)dt - \kappa^i P_{max}^i u_t^{i,j} dt + \sigma^{i,j} dW_t^{i,j}, \quad X_0^{i,j} = x_0^{i,j}, \quad 1 \leq i \leq d, 1 \leq j \leq N_i, \quad (6.1)$$

where x_{out}^i is the outdoor temperature, θ^i is a positive thermal constant, κ^i is the heat exchange constant and P_{max}^i is the maximal power consumption of an air-conditioner in cluster i . $W^{i,j}$ are independent Brownian motion that represent random temperature fluctuations inside the rooms, such as a window or a door opening. For each cluster, a **local controller** decides at each time step, to turn *ON* or *OFF* some conditioners in the cluster i by setting $u^{i,j} = 1$ or 0, in order to satisfy a **prescribed proportion** of active air-conditioners. We are interested in the global planner problem which consists in computing the prescribed proportion $u^i = \frac{1}{N_i} \sum_{j=1}^{N_i} u^{i,j}$ of air conditioners ON in each cluster in order to track the given target consumption profile $r = (r_t)_{0 \leq t \leq T}$. For each $1 \leq i \leq d$ the average temperature $X^i = \frac{1}{N_i} \sum_{j=1}^{N_i} X^{i,j}$ in the cluster i follows the aggregated dynamics

$$dX_t^i = -\theta^i(X_t^i - x_{out}^i)dt - \kappa^i P_{max}^i u_t^i dt + \sigma^i dW_t^i, \quad X_0^i = x_0^i, \quad (6.2)$$

with

$$W_t^i = \frac{1}{N_i} \sum_{j=1}^{N_i} W_t^{i,j}, \quad \sigma^i = \frac{1}{N_i} \sum_{j=1}^{N_i} \sigma^{i,j} \quad \text{and} \quad x_0^i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_0^{i,j}.$$

We consider the stochastic control Problem (1.3) on the time horizon $[0, T]$ with $\mathbb{U} = [0, 1]^d$ and $T = 2h$. The running cost f is defined for any $(t, x, u) \in [0, T] \times \mathbb{R}^d \times \mathbb{U}$, such that

$$f(t, x, u) := \mu \left(\sum_{i=1}^d \rho_i u_i - r_t \right)^2 + \frac{1}{d} \sum_{i=1}^d \left(\gamma_i (\rho_i u_i)^2 + \eta_i (x_i - x_{max}^i)_+^2 + \eta_i (x_{min}^i - x_i)_+^2 \right), \quad (6.3)$$

where $\rho_i = N_i P_{max}^i / (\sum_{j=1}^d N_j P_{max}^j)$, the first term in the above cost function penalizes the deviation of the overall consumption $\sum_i \rho_i u_t^i$ with respect to the target consumption r_t , γ_i quantifies the penalization for irregular controls in cluster i , while η_i penalizes the exits of the mean temperatures in the cluster i , from a comfort band $[x_{min}^i, x_{max}^i]$. Finally the terminal cost is given by $g(x) = \frac{1}{d} \sum_{i=1}^d |x^i - x_{target}^i|^2$, where x_{target}^i is a target temperature for cluster i . Clearly the cost functions f and g satisfy Hypothesis 3.7. To estimate an optimal policy u^* for this problem, we use Algorithm 1 with a time step $t_{m+1} - t_m = 60s$, for $m = 0, \dots, M$. The parameters of the problem are the same as in [28]. We perform $N_{grid} = 100$ independent runs of the algorithm, providing $(\hat{u}^i)_{1 \leq i \leq N_{grid}}$ estimations of an optimal control on the whole period t_0, t_1, \dots, t_M . For each estimation \hat{u}^i , we simulate $N_{simu} = 1000$ iid trajectories of the process controlled by \hat{u}^i and compute the associated costs $(\mathcal{J}_\ell(\hat{u}^i))_{1 \leq \ell \leq N_{simu}}$. The average cost is finally estimated by $\mathcal{J} = \frac{1}{N_{grid} N_{simu}} \sum_{i=1}^{N_{grid}} \sum_{\ell=1}^{N_{simu}} \mathcal{J}_\ell(\hat{u}^i)$.

To evaluate the performances of our approach, we compare it with the classical regression-based Monte-Carlo technique, relying on the dynamic programming principle in [28]. We underline that we only aim to obtain lower costs compared to the BSDE technique in [28], there are no benchmark costs. The results are reported in Table 1, for dimensions $d = 1, 2, 5, 10, 15, 20$. For both methods, $N = 10^3, 10^4, 5 \times 10^4, 10^5$ particles are used to estimate an optimal policy for each dimension d . For the entropy penalized Monte-Carlo algorithm, we use a penalization parameter $\epsilon = 70$ and $K = 20$ iterations for dimensions $d = 1, 2, 5, 10$ and $\epsilon = 20$ and $K = 60$ iterations for dimensions $d = 15, 20$. Concerning the approximation in Step 1 of the Algorithm 1, we limit ourselves to the set \mathcal{P}_0 of polynomials of degree 0, since the problem is very localized in space. On Table 1 we can observe very good performances, that seem to be weakly sensitive to the dimensions of the problem. On Figure 1, we have reported the cost $\mathcal{J}(\mathbb{Q}_k, \mathbb{P}_k)$ and $\mathcal{J}(\mathbb{P}_k, \mathbb{P}_k) = \mathbb{E}^{\mathbb{P}_k} \left[\int_0^T f(r, X_r, u^k(r, X_r)) dr + g(X_T) \right]$ as a function of the iteration number k , obtained on one run of the algorithm with $d = 20$ and $N = 50000$. These costs are compared to a reference cost obtained with a run of our algorithm with $N = 100000$ particles. As expected, $\mathcal{J}(\mathbb{Q}_k, \mathbb{P}_k)$ is decreasing and converging to a limiting value. It is interesting to notice that $\mathcal{J}(\mathbb{P}_k, \mathbb{P}_k)$ is also decreasing and very close to $\mathcal{J}(\mathbb{Q}_k, \mathbb{P}_k)$. Hence, it seems that the parameter ϵ does not need to be as small to obtain a good approximation of the original control Problem (1.3).

Method	$N = 10^3$		$N = 10^4$		$N = 5 \times 10^4$		$N = 10^5$	
	Entropy	BSDE	Entropy	BSDE	Entropy	BSDE	Entropy	BSDE
$d = 1$	7.60($1e^{-6}$)	7.61($6e^{-4}$)	7.59($1e^{-6}$)	7.60($3e^{-4}$)	7.59($1e^{-6}$)	7.60($3e^{-4}$)	7.59($1e^{-6}$)	7.60($3e^{-4}$)
$d = 2$	7.82($2e^{-6}$)	8.24($7e^{-2}$)	7.79($5e^{-7}$)	7.77($1e^{-3}$)	7.78($5e^{-7}$)	7.79($2e^{-4}$)	7.78($5e^{-7}$)	7.78($1e^{-4}$)
$d = 5$	7.34($2e^{-6}$)	14.83(0.64)	7.30($5e^{-7}$)	7.69($6e^{-2}$)	7.30($3e^{-7}$)	7.28($2e^{-3}$)	7.30($3e^{-7}$)	7.27($8e^{-4}$)
$d = 10$	5.96($2e^{-6}$)	28.14(0.64)	5.88($8e^{-7}$)	16.06(0.38)	5.87($5e^{-7}$)	7.96(0.25)	5.87($4e^{-7}$)	6.12(0.08)
$d = 15$	9.15($7e^{-5}$)	37.91(0.60)	8.32($2e^{-5}$)	32.20(0.63)	8.11($5e^{-6}$)	26.69(0.65)	8.08($3e^{-6}$)	22.54(0.56)
$d = 20$	8.80($4e^{-5}$)	34.83(0.45)	7.91($1e^{-5}$)	30.66(0.59)	7.71($3e^{-6}$)	26.21(0.69)	7.68($2e^{-6}$)	23.26(0.59)

Table 1: Simulated costs (within parenthesis, standard deviation) for the relative entropy penalization scheme and a classical BSDE scheme.

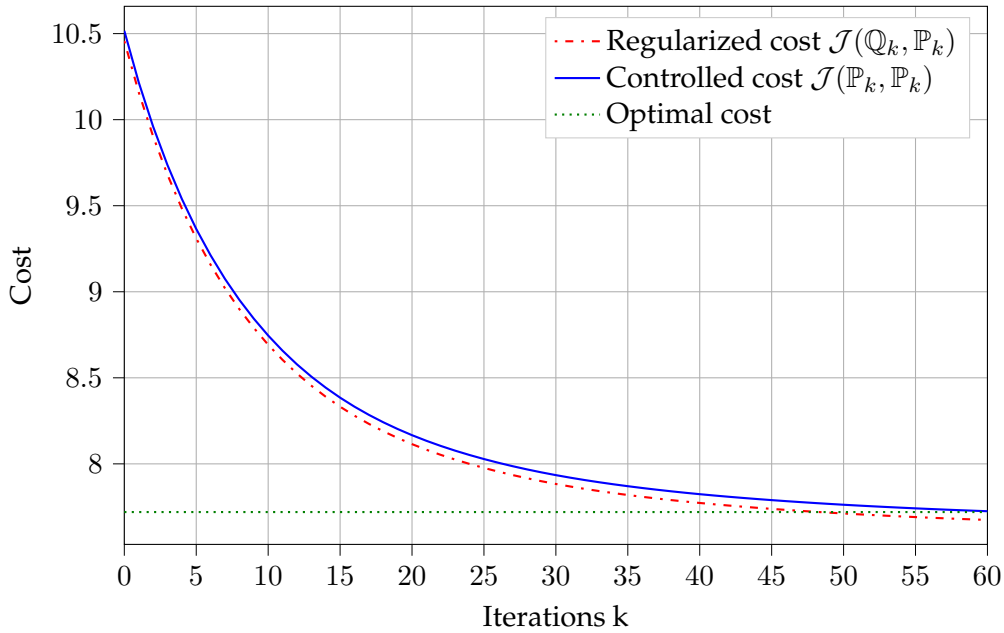


Figure 1: Costs associated with the iterates, generated by the entropy penalized Monte-Carlo algorithm in dimension $d = 20$ with $N = 50000$.

7 Conclusion and perspectives

In this paper we have proposed an original approach to treat stochastic optimal control problems, regarded as optimization programs on the space of probability measures, based on an entropy penalized formulation. In particular this has allowed us to design an alternating minimization procedure to tackle those problems. One additional interest of this entropy penalized formulation is that it can be naturally extended to treat control problems with more complex constraints of the form

$$\inf_{\mathbb{P} \in \mathcal{A} \cap \mathcal{B}} \mathbb{E}^{\mathbb{P}} \left[\int_0^T f(r, X_r, \nu_r^{\mathbb{P}}) dr + g(X_T) \right], \quad (7.1)$$

with a general admissible set of the form $\mathcal{A} \cap \mathcal{B}$ where \mathcal{A} is a convex subset of $\mathcal{P}(\Omega)$ and \mathcal{B} is a subset of $\mathcal{P}(\Omega)$, describing a class of controlled dynamics, which fulfills some technical conditions. A typical example appears when $\mathcal{A} = \{\mathbb{P} \in \mathcal{P}(\Omega) : \mathbb{P}_T = \mu_T\}$, where μ_T is a prescribed (terminal) law and \mathcal{B} imposes an initial law μ_0 . Problem (7.1) then corresponds in this example to a stochastic control problem with prescribed initial and terminal distributions, typically encountered in the fields of martingale optimal transport or Schrödinger Bridge problems. We remark that this formulation covers in particular the one of the present paper setting $\mathcal{A} = \mathcal{P}(\Omega)$ and $\mathcal{B} = \mathcal{P}_{\mathbb{U}}$.

The idea is then to extend the splitting approach of our entropy penalized method, leading us to two simpler subproblems, each one taking into account separately the constraints sets \mathcal{A} and \mathcal{B} . This is the object of a paper in preparation.

Appendices

A Relative entropy related results

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$ be a filtered probability space. Let $\delta = (\delta_t)_{t \in [0, T]}$ (resp. $a = (a_t)_{t \in [0, T]}$) be a progressively measurable process with values in \mathbb{R}^d (resp. in the set of square $d \times d$ non-negative defined symmetric matrices S_d^+). Let X be a continuous process, which decomposes as

$$X_t = x + \int_0^t \delta_r dr + M_t^{\mathbb{P}}, \quad 0 \leq t \leq T, \quad (\text{A.1})$$

where $M^{\mathbb{P}}$ is a continuous $((\mathcal{F}_t), \mathbb{P})$ -local martingale such that $[M^{\mathbb{P}}] = \int_0^\cdot a_r dr$.

The theorem below is the Girsanov's theorem under a finite relative entropy assumption.

Theorem A.1. *Let \mathbb{Q} be a probability measure on (Ω, \mathcal{F}) such that $H(\mathbb{Q}|\mathbb{P}) < +\infty$. Then there exists an \mathbb{R}^d -valued progressively measurable process α such that*

$$\mathbb{E}^{\mathbb{Q}} \left[\int_0^T \alpha_r^\top a_r \alpha_r dr \right] < +\infty, \quad (\text{A.2})$$

and such that, under \mathbb{Q} , the process X is still a continuous semimartingale with decomposition

$$X_t = x + \int_0^t \delta_r dr + \int_0^t a_r \alpha_r dr + M_t^{\mathbb{Q}}, \quad 0 \leq t \leq T, \quad (\text{A.3})$$

where $M^{\mathbb{Q}}$ is a continuous \mathbb{Q} -local martingale and $[M^{\mathbb{Q}}] = \int_0^\cdot a_r dr$. Furthermore,

$$\frac{1}{2} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \alpha_r^\top a_r \alpha_r dr \right] \leq H(\mathbb{Q}|\mathbb{P}). \quad (\text{A.4})$$

Proof. The fact that $H(\mathbb{Q}|\mathbb{P}) < +\infty$ implies in particular that $\mathbb{Q} \ll \mathbb{P}$. Let then $Z_T := d\mathbb{Q}/d\mathbb{P}$ and $(Z_t)_{t \in [0, T]}$ be the càdlàg \mathbb{P} -modification of the martingale $(\mathbb{E}^{\mathbb{P}} [Z_T | \mathcal{F}_t])_{t \in [0, T]}$. By Theorem 3.24, Chapter III in [29], there exists a progressively measurable process α such that decomposition (A.3) holds and

$$\int_0^\cdot \alpha_r^\top a_r \alpha_r dr < +\infty \quad \mathbb{Q}\text{-a.s.} \quad (\text{A.5})$$

as well as

$$[Z, M^{\mathbb{P}}] = \int_0^\cdot a_r \alpha_r Z_{r-} dr, \quad (\text{A.6})$$

with respect to \mathbb{P} , so also with respect to \mathbb{Q} .

Let then $\tau_k := \inf \left\{ t \in [0, T] : \int_0^t \alpha_r^\top a_r \alpha_r dr > k \right\}$, with the convention that $\inf \emptyset = +\infty$. Setting $M^k := \int_0^{\cdot \wedge \tau_k} \alpha_r^\top dM_r^{\mathbb{P}}$ and Z^k the Doléans exponential $\mathcal{E}(M^k)$, we define $d\mathbb{Q}_k := Z_T^k d\mathbb{P}$. By Novikov's criterion (see Proposition 1.15, Chapter VIII in [38]), Z^k is a martingale, therefore \mathbb{Q}_k is a probability measure on (Ω, \mathcal{F}) equivalent to \mathbb{P} since Z_T^k is strictly positive \mathbb{P} -a.s. As $\mathbb{Q} \ll \mathbb{P}$

and $\mathbb{Q}_k \sim \mathbb{P}$, we have $\mathbb{Q} \ll \mathbb{Q}_k$. It follows that \mathbb{P} -a.s., with the notation $\log(0) = -\infty$, and later $0 \log(0) = 0$.

$$\begin{aligned} \log \frac{d\mathbb{Q}}{d\mathbb{P}} &= \log \frac{d\mathbb{Q}}{d\mathbb{Q}_k} + \log \frac{d\mathbb{Q}_k}{d\mathbb{P}} \\ &= \log \frac{d\mathbb{Q}}{d\mathbb{Q}_k} + \log Z_T^k \\ &= \log \frac{d\mathbb{Q}}{d\mathbb{Q}_k} + \int_0^{T \wedge \tau_k} \alpha_r^\top dM_r^\mathbb{P} - \frac{1}{2} \int_0^{T \wedge \tau_k} \alpha_r^\top a_r \alpha_r dr. \end{aligned} \tag{A.7}$$

Previous equality can be of course considered also \mathbb{Q} -a.s. since \mathbb{Q} is "rougher" than \mathbb{P} . Setting $\bar{M} := M^\mathbb{P} - \int_0^\cdot a_r \alpha_r dr$, equality (A.7) rewrites

$$\log \frac{d\mathbb{Q}}{d\mathbb{P}} = \log \frac{d\mathbb{Q}}{d\mathbb{Q}_k} + \int_0^{T \wedge \tau_k} \alpha_r^\top d\bar{M}_r + \frac{1}{2} \int_0^{T \wedge \tau_k} \alpha_r^\top a_r \alpha_r dr \quad \mathbb{Q}\text{-a.s.} \tag{A.8}$$

Taking into account (A.6), Theorem 3.11, Chapter III in [29] states that, \bar{M} is a \mathbb{Q} -local martingale. Since, still with respect to \mathbb{Q} , $[\int_0^\cdot \alpha_r^\top d\bar{M}_r] = \int_0^\cdot \alpha_r^\top a_r \alpha_r dr$, by definition of τ_k , the process $\int_0^{\cdot \wedge \tau_k} \alpha_r^\top d\bar{M}_r$ is a genuine \mathbb{Q} -martingale. Consequently, taking the expectation under \mathbb{Q} in (A.8) gives

$$H(\mathbb{Q}|\mathbb{P}) = H(\mathbb{Q}|\mathbb{Q}_k) + \frac{1}{2} \mathbb{E}^\mathbb{Q} \left[\int_0^{T \wedge \tau_k} \alpha_r^\top a_r \alpha_r dr \right] \geq \frac{1}{2} \mathbb{E}^\mathbb{Q} \left[\int_0^{T \wedge \tau_k} \alpha_r^\top a_r \alpha_r dr \right].$$

Since $\tau_k \xrightarrow[k \rightarrow +\infty]{} +\infty$ increasingly \mathbb{Q} -a.s. by (A.5), a direct application of the monotone convergence theorem then yields

$$H(\mathbb{Q}|\mathbb{P}) \geq \frac{1}{2} \mathbb{E}^\mathbb{Q} \left[\int_0^T \alpha_r^\top a_r \alpha_r dr \right].$$

□

For the following lemma let again X be a process, as at the beginning of the section fulfilling (A.1), this time with $a_t = \sigma \sigma^\top(t, X_t)$. Then by Theorem A.1 there is a progressively measurable process α such that (A.3) holds. For that we have the following estimates.

Lemma A.2. *We suppose the existence of $1 < p < 2$ such that*

$$C_p := \mathbb{E}^\mathbb{P} \left[\int_0^T \|\sigma(r, X_r)\|^{2p/(2-p)} dr \right] < +\infty.$$

Let \mathbb{Q} be a probability measure on (Ω, \mathcal{F}) such that $H(\mathbb{Q}|\mathbb{P}) < +\infty$.

1. If $C_\infty := \|d\mathbb{Q}/d\mathbb{P}\|_\infty < +\infty$, there exists a constant $L > 0$, which depends only on C_p and C_∞ , such that

$$\mathbb{E}^\mathbb{Q} \left[\int_0^T |\sigma^\top(r, X_r) \alpha_r|^p dr \right] \leq L(1 + H(\mathbb{Q}|\mathbb{P})). \tag{A.9}$$

2. Suppose moreover $H(\mathbb{P}|\mathbb{Q}) < +\infty$. Then it holds that

$$\frac{1}{2} \mathbb{E}^\mathbb{P} \left[\int_0^T |\sigma^\top(r, X_r) \alpha_r|^2 dr \right] \leq H(\mathbb{P}|\mathbb{Q}), \tag{A.10}$$

and L can be chosen such that both (A.9) and

$$\mathbb{E}^{\mathbb{P}} \left[\int_0^T |\sigma \sigma^\top(r, X_r) \alpha_r|^p dr \right] \leq L(1 + H(\mathbb{P}|\mathbb{Q})). \quad (\text{A.11})$$

Proof. 1. We recall that $H(\mathbb{Q}|\mathbb{P}) < \infty$. By Hölder's inequality applied on the measure space $([0, T] \times \Omega, \mathcal{B}([0, T]) \otimes \mathcal{F}, dt \otimes d\mathbb{Q})$, it holds that

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma \sigma^\top(r, X_r) \alpha_r|^p dr \right] &\leq \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \|\sigma(r, X_r)\|^p |\sigma^\top(r, X_r) \alpha_r|^p dr \right] \\ &\leq \left(\mathbb{E}^{\mathbb{Q}} \left[\int_0^T \|\sigma(r, X_r)\|^{2p/(2-p)} dr \right] \right)^{1-p/2} \left(\mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^\top(r, X_r) \alpha_r|^2 dr \right] \right)^{p/2}. \end{aligned} \quad (\text{A.12})$$

On the one hand,

$$\mathbb{E}^{\mathbb{Q}} \left[\int_0^T \|\sigma(r, X_r)\|^{2p/(p-2)} dr \right] = \mathbb{E}^{\mathbb{P}} \left[\frac{d\mathbb{Q}}{d\mathbb{P}} \int_0^T \|\sigma(r, X_r)\|^{2p/(p-2)} dr \right] \leq C_\infty C_p. \quad (\text{A.13})$$

On the other hand, by (A.4) in Theorem A.1,

$$\mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^\top(r, X_r) \alpha_r|^2 dr \right] \leq 2H(\mathbb{Q}|\mathbb{P}). \quad (\text{A.14})$$

Combining (A.13) and (A.14) with (A.12), we get

$$\mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma \sigma^\top(r, X_r) \alpha_r|^p dr \right] \leq 2^{p/2} (C_\infty C_p)^{1-p/2} H(\mathbb{Q}|\mathbb{P})^{p/2},$$

and as $p < 2$, using the inequality

$$|a|^q \leq (1 + |a|), \text{ if } q \in]0, 1], \quad (\text{A.15})$$

with $q = 1 - \frac{p}{2}$ and $q = \frac{p}{2}$, we have

$$\mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma \sigma^\top(r, X_r) \alpha_r|^p dr \right] \leq 2(1 + C_\infty C_p)(1 + H(\mathbb{Q}|\mathbb{P})).$$

Setting

$$L := 2(1 + C_p(C_\infty \vee 1)), \quad (\text{A.16})$$

one concludes the proof of item 1.

2. Applying Theorem A.1, we recall the decomposition (A.3), where the local martingale $M^{\mathbb{Q}}$ verifies $[M^{\mathbb{Q}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$, under \mathbb{Q} . As $H(\mathbb{P}|\mathbb{Q}) < +\infty$, interchanging \mathbb{P} and \mathbb{Q} , again Theorem A.1 yields the existence of a progressively measurable process $\tilde{\alpha}$ such that under \mathbb{P} the process X decomposes as

$$X_t = x + \int_0^t \delta_r dr + \int_0^t \sigma \sigma^\top(r, X_r) \alpha_r dr + \int_0^t \sigma \sigma^\top(r, X_r) \tilde{\alpha}_r dr + \tilde{M}_t, \quad (\text{A.17})$$

where \tilde{M} is a \mathbb{P} -local martingale such that $[\tilde{M}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$ and

$$\frac{1}{2} \mathbb{E}^{\mathbb{P}} \left[\int_0^T |\sigma^\top(r, X_r) \tilde{\alpha}_r|^2 dr \right] \leq H(\mathbb{P}|\mathbb{Q}).$$

Identifying the bounded variation and the martingale components of X under \mathbb{P} , in (A.17) and (A.1), we get that $\tilde{M} = M^{\mathbb{P}}$ and $\sigma \sigma^\top(t, X_t) \tilde{\alpha}_t = -\sigma \sigma^\top(t, X_t) \alpha_t dt \otimes d\mathbb{P}$ -a.e. In particular, (A.10) holds. Then, as in the proof of item 1., Hölder's inequality, (A.12) with \mathbb{Q} replaced by \mathbb{P} , and (A.10) yield

$$\begin{aligned} \mathbb{E}^{\mathbb{P}} \left[\int_0^T |\sigma \sigma^\top(r, X_r) \alpha_r|^p dr \right] &\leq \left(\mathbb{E}^{\mathbb{P}} \left[\int_0^T \|\sigma(r, X_r)\|^{2p/(2-p)} dr \right] \right)^{1-p/2} \left(\mathbb{E}^{\mathbb{P}} \left[\int_0^T |\sigma^\top(r, X_r) \alpha_r|^2 dr \right] \right)^{p/2} \\ &\leq 2^{p/2} C_p^{1-p/2} H(\mathbb{P}|\mathbb{Q})^{p/2} \leq 2(1 + C_p)(1 + H(\mathbb{P}|\mathbb{Q})), \end{aligned}$$

where, for the latter inequality we have used again (A.15) with $q = 1 - \frac{p}{2}$ and $q = \frac{p}{2}$ together with (A.10). This finally also implies the result (A.11) with L defined in (A.16). \square

Remark A.3. Let $\tilde{C}_p := \mathbb{E}^{\mathbb{Q}} \left[\int_0^T \|\sigma(r, X_r)\|^{2p/(2-p)} dr \right]$. Item 1. of Lemma A.2 is still valid if one assumes that $\tilde{C}_p < +\infty$ instead of $C_p < +\infty$ and $\|d\mathbb{Q}/d\mathbb{P}\|_\infty$. One only has to replace $C_\infty C_p$ by \tilde{C}_p in the estimates in the proof.

The results of Theorem A.1 can be specified if one considers probability measures on the canonical space $\Omega = C([0, T], \mathbb{R}^d)$. In the following, $\delta, \gamma : [0, T] \times C([0, T], \mathbb{R}^d) \mapsto \mathbb{R}^d$ are progressively measurable functions w.r.t. their corresponding Borel σ -fields. A consequence of Theorem A.1 in this setting is the following.

Lemma A.4. Let $\mathbb{P} \in \mathcal{P}(\Omega)$ such that, under \mathbb{P} the canonical process can be decomposed as

$$X_t = x + \int_0^t \delta(r, X) dr + M_t^{\mathbb{P}}, \quad (\text{A.18})$$

where $M^{\mathbb{P}}$ is a martingale with $[M^{\mathbb{P}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$, where σ verifies item (ii) of Hypothesis 3.1. Let $\mathbb{Q} \in \mathcal{P}(\Omega)$.

1. Assume that $H(\mathbb{Q}|\mathbb{P}) < +\infty$. Then we have the following.

(a) There exists a progressively measurable process α , w.r.t. the natural filtration of X (in particular of the form $\alpha = \alpha(\cdot, X)$) such that, under \mathbb{Q} , X decomposes as

$$X_t = x + \int_0^t \delta(r, X) dr + \int_0^t \sigma \sigma^\top(r, X_r) \alpha(r, X) dr + M_t^{\mathbb{Q}}, \quad (\text{A.19})$$

where $M^{\mathbb{Q}}$ is a martingale with $[M^{\mathbb{Q}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$ and

$$H(\mathbb{Q}|\mathbb{P}) \geq \frac{1}{2} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^\top(r, X_r) \alpha(r, X)|^2 dr \right]. \quad (\text{A.20})$$

(b) If moreover uniqueness in law holds for the SDE (A.18), equality holds in (A.20).

2. Assume that under \mathbb{Q} the canonical process writes

$$X_t = x + \int_0^t \gamma(r, X) dr + M_t^{\mathbb{Q}}, \quad (\text{A.21})$$

where $M^{\mathbb{Q}}$ is a martingale with $[M^{\mathbb{Q}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$ and that uniqueness in law holds for the SDE (A.18). Let σ^{-1} be again the generalized right-inverse of σ . If

$$\mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^{-1}(r, X_r)(\delta(r, X) - \gamma(r, X))|^2 dr \right] < +\infty,$$

then $H(\mathbb{Q}|\mathbb{P}) < +\infty$ and

$$H(\mathbb{Q}|\mathbb{P}) = \frac{1}{2} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^{-1}(r, X_r)(\delta(r, X) - \gamma(r, X))|^2 dr \right]. \quad (\text{A.22})$$

Proof. Part (a) of item 1. of Lemma A.18 is constituted by Theorem A.1 applied to the canonical space equipped with the natural filtration of the canonical process. Item 2. is the object of Lemma 4.4 (iii) in [33].

As far as item 1.(b) is concerned, we apply item 2. with $\gamma(r, X) = \delta(r, X) + \sigma \sigma^\top(r, X_r) \alpha(r, X)$ in (A.21) so that $(\gamma - \delta)(r, X) = \sigma \sigma^\top(r, X_r) \alpha(r, X)$. So $\sigma^{-1}(r, X_r)(\delta - \gamma)(r, X)$ and the equality in (A.20) holds because of (A.22). \square

Remark A.5. By Hypothesis 3.1 on the diffusion coefficient σ , uniqueness in law for the SDE (A.18) holds e.g. if δ is bounded, or if $\delta(r, X) = b(r, X_r, u(r, X_r))$, where b has linear growth in (t, x) independently of u . This follows from Theorem 10.1.3 of [39].

B Measurable selection

The following measurable selection theorem is a direct consequence of Theorem A.9 in [26], setting $y = (t, x)$, $\phi = b$, $i = 1$, $\psi_1 = f$.

Theorem B.1. Suppose the validity of item 1. of Hypothesis 3.1 and item 2. of Hypothesis 3.7. Let $K(t, x)$ be given by (3.9). Let $y \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{R}^d)$ and $z \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{R})$ be two functions such that $(y(t, x), z(t, x)) \in K(t, x)$ for all $(t, x) \in [0, T] \times \mathbb{R}^d$. Then there exists a (measurable) function $u \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$ such that

$$\begin{cases} y(t, x) = b(t, x, u(t, x)) \\ z(t, x) \geq f(t, x, u(t, x)) \end{cases} \text{ for all } (t, x) \in [0, T] \times \mathbb{R}^d.$$

The result below is a simple consequence of Theorem B.1

Lemma B.2. *Let Ω be a Polish space, and let $\mathcal{F} := \mathcal{B}(\Omega)$ be its Borel σ -field. Let $X : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ and $(y, z) : [0, T] \times \Omega \rightarrow \mathbb{R}^{d+1}$ be two processes on (Ω, \mathcal{F}) . Let \mathbb{P} be a probability measure on (Ω, \mathcal{F}) . Assume Hypothesis 3.9 and that $\mathbb{E}^{\mathbb{P}} \left[\int_0^T |y_r| dr \right] < +\infty$ and $\mathbb{E}^{\mathbb{P}} \left[\int_0^T |z_r| dr \right] < +\infty$. Assume moreover that $(y_t, z_t) \in K(t, X_t)$ for almost all $t \in [0, T]$, \mathbb{P} -a.s. Then there exists a function $u \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$ such that for almost all $t \in [0, T]$, \mathbb{P} -a.s.,*

$$\begin{cases} \mathbb{E}^{\mathbb{P}}[y_t|X_t] = b(t, X_t, u(t, X_t)) \\ \mathbb{E}^{\mathbb{P}}[z_t|X_t] \geq f(t, X_t, u(t, X_t)). \end{cases} \quad (\text{B.1})$$

Proof. 1. We set $\Phi_t := (y_t, z_t)$ which belongs a.s. to $K(t, X_t)$. We prove below that, for almost all $t \in [0, T]$,

$$\mathbb{E}^{\mathbb{P}}([\Phi_t|X_t]) \in K(t, X_t) \quad \mathbb{P}\text{-a.s.} \quad (\text{B.2})$$

Indeed, let $t \in [0, T]$ such that $(y_t, z_t) \in K(t, X_t)$ \mathbb{P} -a.s. We set $\mu := \mathcal{L}^{\mathbb{P}}(X_t)$. By Theorem 1.1.6 and Theorem 1.1.8 in [39] there exists a measurable family $(\mathbb{P}_x)_{x \in \mathbb{R}^d}$ of probability measures on (Ω, \mathcal{F}) such that $\mathbb{P}_x(X_t = x) = 1$ for μ -almost all $x \in \mathbb{R}^d$ and $\mathbb{P} = \int_{\mathbb{R}^d} \mathbb{P}_x \mu(dx)$. On the one hand, since $\mathbb{P}(\Phi_t \in K(t, X_t)) = 1$,

$$1 = \mathbb{P}(\Phi_t \in K(t, X_t)) = \int_{\mathbb{R}^d} \mathbb{P}_x(\Phi_t \in K(t, X_t)) \mu(dx) = \int_{\mathbb{R}^d} \mathbb{P}_x(\Phi_t \in K(t, x)) \mu(dx),$$

hence $\mathbb{P}_x(\Phi_t \in K(t, x)) = 1$ for μ -almost all $x \in \mathbb{R}^d$. Consequently, since $K(t, x)$ is a convex closed set, by Theorem 1 in [37],

$$\mathbb{P}_x \left(\mathbb{E}^{\mathbb{P}_x}[\Phi_t|X_t] \in K(t, x) \right) = \mathbb{P}_x \left(\mathbb{E}^{\mathbb{P}_x}[\Phi_t] \in K(t, x) \right) = 1. \quad (\text{B.3})$$

On the other hand, by definition of the conditional expectation, $\mathbb{E}^{\mathbb{P}}[\Phi_t|X_t] = (\mathbb{E}^{\mathbb{P}_x}[\Phi]) \circ X_t$. Consequently,

$$\begin{aligned} \mathbb{P} \left(\mathbb{E}^{\mathbb{P}}[\Phi_t|X_t] \in K(t, X_t) \right) &= \int_{\mathbb{R}^d} \mathbb{P}_x \left(\mathbb{E}^{\mathbb{P}}[\Phi_t|X_t] \in K(t, X_t) \right) \mu(dx) \\ &= \int_{\mathbb{R}^d} \mathbb{P}_x \left(\mathbb{E}^{\mathbb{P}_x}[\Phi_t] \in K(t, x) \right) \mu(dx) \\ &= 1, \end{aligned}$$

where we used (B.3) to conclude. (B.2) is proved.

2. It remains to prove (B.1). Proposition 5.1 in [9] provides two measurable functions \mathcal{Y}, \mathcal{Z} such that for all $t \in [0, T]$, \mathbb{P} -a.s.

$$\left(\mathbb{E}^{\mathbb{P}}[y_t|X_t], \mathbb{E}^{\mathbb{P}}[z_t|X_t] \right) = (\mathcal{Y}(t, X_t), \mathcal{Z}(t, X_t)). \quad (\text{B.4})$$

Let then

$$N := \{(t, x) \in [0, T] \times \mathbb{R}^d \mid (\mathcal{Y}(t, x), \mathcal{Z}(t, x)) \notin K(t, x)\}.$$

The set N is a Borel set, and we now modify the functions \mathcal{Y} and \mathcal{Z} on N and obtain two Borel functions $\hat{\mathcal{Y}}, \hat{\mathcal{Z}}$ defined by

$$\begin{cases} (\hat{\mathcal{Y}}(t, x), \hat{\mathcal{Z}}(t, x)) = (\mathcal{Y}(t, x), \mathcal{Z}(t, x)) & \text{if } (t, x) \notin N \\ (\hat{\mathcal{Y}}(t, x), \hat{\mathcal{Z}}(t, x)) = (b(t, x, u_0), f(t, x, u_0)) & \text{if } (t, x) \in N, \end{cases} \quad (\text{B.5})$$

where $u_0 \in \mathbb{U}$ is fixed. In particular, $(\hat{\mathcal{Y}}(t, x), \hat{\mathcal{Z}}(t, x)) \in K(t, x)$ for all $(t, x) \in [0, T] \times \mathbb{R}^d$. Then by Theorem B.1 there exists a Borel function $u \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$ such that

$$\begin{cases} \hat{\mathcal{Y}}(t, x) = b(t, x, u(t, x)) \\ \hat{\mathcal{Z}}(t, x) \geq f(t, x, u(t, x)) \end{cases} \quad \text{for all } (t, x) \in [0, T] \times \mathbb{R}^d. \quad (\text{B.6})$$

Combining (B.2), (B.4) and (B.6) yields (B.1) for almost all $t \in [0, T]$, \mathbb{P} -a.s. □

C Proof of Theorem 3.15

To simplify the formalism of the proof we will assume that $\epsilon = 1$ and $g = 0$. In the whole section, we can choose $1 \leq p < p'$ as power constants appearing in Hypotheses 3.7. We start by some definitions.

Definition C.1. (Wasserstein space). Let (E, d) be a metric space. We denote $\mathcal{P}^p(E)$ the set of probability measures $\mathbb{P} \in \mathcal{P}(E)$ such that $\int_E (d(x, x_0))^p \mathbb{P}(dx) < +\infty$ for some (and thus for any) $x_0 \in E$. We endow $\mathcal{P}^p(E)$ with the Wasserstein metric

$$d_p(\mathbb{P}, \mathbb{Q}) := \inf \left\{ \int_{E \times E} (d(x, y))^p \rho(dx, dy) : \rho \in \mathcal{P}(E \times E), \rho(\cdot \times E) = \mathbb{P}, \rho(E \times \cdot) = \mathbb{Q} \right\}^{1/p}. \quad (\text{C.1})$$

Definition C.2. (Relaxed controls). We denote \mathcal{V} the set of relaxed controls, that is the set of non-negative measures q on $[0, T] \times \mathbb{U}$ such that we have the following.

1. $q(\cdot \times \mathbb{U})$ is the Lebesgue measure on $[0, T]$, and $q([0, T] \times \cdot)/T$ is a probability measure on $(\mathbb{U}, \mathcal{B}(\mathbb{U}))$.
2. $\int_{[0, T] \times \mathbb{U}} |u|^p q(dr, du) < +\infty$.

The space \mathcal{V} is endowed with the distance $d_{\mathcal{V}}(q_1, q_2) := d_p(q_1/T, q_2/T)$ where d_p is given by (C.1).

Definition C.3. (Extended space). Let $\bar{\Omega} := C([0, T], \mathbb{R}^d) \times \mathcal{V}$ and we denote (X, Λ) its canonical process. The space $\bar{\Omega}$ is endowed with the filtration $(\bar{\mathcal{F}}_t)_{t \in [0, T]}$ defined for all $t \in [0, T]$ by $\bar{\mathcal{F}}_t := \mathcal{F}_t^X \otimes \mathcal{F}_t^\Lambda$ where $\mathcal{F}_t^X := \sigma(X_r, 0 \leq r \leq t)$ and $\mathcal{F}_t^\Lambda := \sigma(\Lambda(A), A \in \mathcal{B}([0, t] \times \mathbb{U}))$. $\bar{\Omega}$ is equipped with the distance $d_{\bar{\Omega}}$ given by $d_{\bar{\Omega}}((x_1, q_1), (x_2, q_2)) := |x_1 - x_2|_\infty + d_{\mathcal{V}}(q_1, q_2)$.

Definition C.4. (Relaxed admissible set). Let $\bar{\mathcal{A}}$ be the subset of $(\mathcal{P}(\bar{\Omega}))^2$ such that $(\bar{\mathbb{P}}, \bar{\mathbb{Q}}) \in \bar{\mathcal{A}}$ if the following holds.

1. $H(\bar{\mathbb{Q}}|\bar{\mathbb{P}}) < +\infty$.

2. Under $\bar{\mathbb{P}}$ the process X decomposes as

$$X_t = x + \int_{[0,t] \times \mathbb{U}} b(r, X_r, u) \Lambda(dr, du) + M_t^{\bar{\mathbb{P}}}, \quad (\text{C.2})$$

where $M^{\bar{\mathbb{P}}}$ is a $(\bar{\mathcal{F}}_t)$ -local martingale verifying $[M^{\bar{\mathbb{P}}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$.

We will denote $\bar{\mathcal{P}}_{\mathbb{U}}$ the set of elements of $\mathcal{P}(\bar{\Omega})$ such that decomposition (C.2) holds.

For $(\bar{\mathbb{P}}, \bar{\mathbb{Q}}) \in \bar{\mathcal{A}}$ we introduce a relaxed problem defined by

$$\bar{\mathcal{J}}^* := \inf_{(\bar{\mathbb{P}}, \bar{\mathbb{Q}}) \in \bar{\mathcal{A}}} \bar{\mathcal{J}}(\bar{\mathbb{Q}}, \bar{\mathbb{P}}) \quad \text{where} \quad \bar{\mathcal{J}}(\bar{\mathbb{Q}}, \bar{\mathbb{P}}) := \mathbb{E}^{\bar{\mathbb{Q}}} \left[\int_{[0,T] \times \mathbb{U}} f(r, X_r, u) \Lambda(dr, du) + g(X_T) \right] + H(\bar{\mathbb{Q}}|\bar{\mathbb{P}}). \quad (\text{C.3})$$

Remark C.5. 1. The notion of relaxed control in Definition C.2 extends the notion of (strict) control $\nu = \nu^{\mathbb{P}}$ as introduced in Definition 3.2. Indeed, a control $\nu : [0, T] \times \Omega \rightarrow \mathbb{U}$ induces a measure on $[0, T] \times \mathbb{U}$ by setting $q^\nu := dt \delta_\nu(du) \in \mathcal{V}$.

2. The set of relaxed controls has two main advantages : it is convex and there exist very convenient tightness criteria to identify its precompact sets using Prokhorov's theorem. This allows to easily prove the existence of a solution to the relaxed Problem (C.3). Under the convexity Hypothesis 3.9, it is then possible to deduce the existence of a solution to the original Problem (1.5).

The strategy of the proof of Theorem 3.15 is the following. We first prove in Proposition C.12 that Problem (C.3) admits a solution $(\bar{\mathbb{P}}^*, \bar{\mathbb{Q}}^*)$ on $\bar{\mathcal{A}}$. We then use Lemma C.13 to compute an optimal solution $(\mathbb{P}^*, \mathbb{Q}^*)$ to the penalized Problem (1.5) derived from $(\bar{\mathbb{P}}^*, \bar{\mathbb{Q}}^*)$. We start by a useful technical result, which is Lemma 3.2 in [32].

Lemma C.6. There exists a \mathcal{F}_t^Λ -predictable process $\bar{\Lambda} : [0, T] \times \mathcal{V} \rightarrow \mathcal{P}(\mathbb{U})$ such that for each $q \in \mathcal{V}$, $\Lambda(q)(dt, du) = dt \bar{\Lambda}_t(q)(du)$.

Based on Lemma C.6, we can now write the canonical process (X, Λ) on $\bar{\Omega}$ as $(X, dt \bar{\Lambda}_t(du))$.

Remark C.7. We list below some facts that will be useful to prove Theorem 3.15.

1. We immediately deduce from Hypothesis 3.1 item 2. and Hypothesis 3.7 item 1. that, for all $t \in [0, T]$,

$$\left| \int_{\mathbb{U}} b(t, X_t, u) \bar{\Lambda}_t(du) \right| \leq \int_{\mathbb{U}} |b(t, X_t, u)| \bar{\Lambda}_t(du) \leq C_{b,\sigma}(1 + |X_t|), \quad (\text{C.4})$$

and

$$\left| \int_{\mathbb{U}} f(t, X_t, u) \bar{\Lambda}_t(du) \right| \leq \int_{\mathbb{U}} |f(t, X_t, u)| \bar{\Lambda}_t(du) \leq C_{f,g}(1 + |X_t|^p). \quad (\text{C.5})$$

2. Let $\bar{\mathbb{P}} \in \bar{\mathcal{P}}_{\mathbb{U}}$. Taking into account decomposition (C.2), (C.4) as well as linear growth of the diffusion coefficient σ in Hypothesis 3.1 item 2., we can apply Lemma 3.11. This yields that for all $q \geq 1$, there exists a constant $C(q)$ which only depends on $C_{b,\sigma}$, T and q such that

$$\mathbb{E}^{\bar{\mathbb{P}}} \left[\left(\sup_{0 \leq t \leq T} |X_t| \right)^q \right] \leq C(q) < +\infty. \quad (\text{C.6})$$

3. Hypothesis 3.1 item 2. implies in particular that

$$|b(t, x, u)| \leq C(1 + |x|^p + |u|^p),$$

for some constant $C > 0$. Since b is continuous in $(t, x, u) \in [0, T] \times \mathbb{R}^d \times \mathbb{U}$ by Hypothesis 3.1 item 1., by Corollary A.5 in [32] applied with $A = \mathbb{U}$, $E = \mathbb{R}^d$ and $\phi = b$, the map

$$(X, \Lambda) \mapsto \int_{[0, t] \times \mathbb{U}} b(r, X_r, u) \Lambda(dr, du),$$

is continuous for $d_{\bar{\Omega}}$. Similarly, Hypothesis 3.7 implies by Corollary A.5 in [32] that the map

$$(X, \Lambda) \mapsto \int_{[0, T] \times \mathbb{U}} f(r, X_r, u) \Lambda(dr, du)$$

is continuous for $d_{\bar{\Omega}}$.

We will need the following simple two technical observations.

Lemma C.8. Let $(\mathbb{P}_n)_{n \geq 1}$ be a sequence of Borel probability measures on a Polish space E that weakly converges towards a probability measure \mathbb{P}_{∞} . Let $\phi : E \rightarrow \mathbb{R}$ be a continuous function. Assume that there exists $\alpha, C > 0$ such that

$$\sup_{n \geq 1} \int_E |\phi(e)|^{1+\alpha} \mathbb{P}_n(de) \leq C. \quad (\text{C.7})$$

Then

$$\int_E \phi(e) \mathbb{P}_n(de) \xrightarrow{n \rightarrow +\infty} \int_E \phi(e) \mathbb{P}_{\infty}(de).$$

Proof. By Skorokhod's representation theorem, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{Q})$, a sequence of random variable $(X_n)_{n \geq 1}$ on Ω and a random variable X such that $\mathcal{L}^{\mathbb{Q}}(X_n) = \mathbb{P}_n$ and $X_n \rightarrow X$ \mathbb{Q} -a.s. Condition (C.7) implies that the sequence $(\phi(X_n))_{n \geq 1}$ is uniformly integrable. Furthermore, by continuity of ϕ , $\phi(X_n) \xrightarrow{n \rightarrow +\infty} \phi(X)$ \mathbb{Q} -a.s. Thus

$$\mathbb{E}^{\mathbb{Q}}[\phi(X_n)] \xrightarrow{n \rightarrow +\infty} \mathbb{E}^{\mathbb{Q}}[\phi(X)]$$

or equivalently

$$\int_E \phi(e) \mathbb{P}_n(de) \xrightarrow{n \rightarrow +\infty} \int_E \phi(e) \mathbb{P}_{\infty}(de).$$

□

Lemma C.9. Let $\bar{\mathbb{P}} \in \bar{\mathcal{P}}_{\mathbb{U}}$. Let $\bar{\mathbb{Q}} \in \mathcal{P}(\bar{\Omega})$ be defined by

$$d\bar{\mathbb{Q}} := \frac{\exp\left(-\int_0^T \int_{\mathbb{U}} f(r, X_r, u) \bar{\Lambda}_r(du) dr\right)}{\mathbb{E}^{\bar{\mathbb{P}}}\left[\exp\left(-\int_0^T \int_{\mathbb{U}} f(r, X_r, u) \bar{\Lambda}_r(du) dr\right)\right]} d\bar{\mathbb{P}}.$$

There exists a constant $C > 0$ only depending on $C_{b,\sigma}, C_{f,g}, T$ and p such that $\|d\bar{\mathbb{Q}}/d\bar{\mathbb{P}}\|_{\infty} \leq C < +\infty$.

Proof. On the one hand, since $f \geq 0$,

$$\exp\left(-\int_0^T \int_{\mathbb{U}} f(r, X_r, u) \bar{\Lambda}_r(du) dr\right) \leq 1. \quad (\text{C.8})$$

On the other hand, from (C.5) and (C.6) in Remark C.7, there exists a constant $C(p)$ which only depends on $C_{b,\sigma}, T$ and p such that

$$\mathbb{E}^{\bar{\mathbb{P}}}\left[\int_0^T \int_{\mathbb{U}} f(r, X_r, u) \bar{\Lambda}_r(du) dr\right] \leq C_{f,g} T (1 + C(p)). \quad (\text{C.9})$$

Then by Jensen's inequality we have

$$\begin{aligned} \mathbb{E}^{\bar{\mathbb{P}}}\left[\exp\left(-\int_0^T \int_{\mathbb{U}} f(r, X_r, u) \bar{\Lambda}_r(du) dr\right)\right] &\geq \exp\left(\mathbb{E}^{\bar{\mathbb{P}}}\left[-\int_0^T \int_{\mathbb{U}} f(r, X_r, u) \bar{\Lambda}_r(du) dr\right]\right) \\ &\geq \exp(-C_{f,g} T (1 + C(p))). \end{aligned} \quad (\text{C.10})$$

Combining (C.8) and (C.10) we get $\|d\bar{\mathbb{Q}}_n/d\bar{\mathbb{P}}_n\|_{\infty} \leq C$ by setting $C := \exp(C_{f,g} T (1 + C(p)))$. \square

We can now start the proof of Theorem 3.15.

Lemma C.10. There exists a minimizing sequence $(\bar{\mathbb{P}}_n, \bar{\mathbb{Q}}_n)_{n \geq 1}$ of $\bar{\mathcal{J}}$ verifying the following.

1. $\sup_{n \geq 1} \left\| \frac{d\bar{\mathbb{Q}}_n}{d\bar{\mathbb{P}}_n} \right\|_{\infty} < +\infty$ and $\sup_{n \geq 1} H(\bar{\mathbb{Q}}_n | \bar{\mathbb{P}}_n) < +\infty$.
2. $(\bar{\mathbb{P}}_n, \bar{\mathbb{Q}}_n)_{n \geq 1}$ is relatively compact in $(\mathcal{P}^p(\bar{\Omega}))^2$.

Proof. In this proof, C denotes a generic non-negative constant. Let $(\bar{\mathbb{P}}_n, \bar{\mathbb{Q}}_n)_{n \geq 1}$ be a minimizing sequence of $\bar{\mathcal{J}}$. Setting

$$d\bar{\mathbb{Q}}_n := \frac{\exp\left(-\int_0^T \int_{\mathbb{U}} f(r, X_r, u) \bar{\Lambda}_r(du) dr\right)}{\mathbb{E}^{\bar{\mathbb{P}}_n}\left[\exp\left(-\int_0^T \int_{\mathbb{U}} f(r, X_r, u) \bar{\Lambda}_r(du) dr\right)\right]} d\bar{\mathbb{P}}_n, \quad (\text{C.11})$$

by Proposition 3.13, $\inf_{\mathbb{Q}} \bar{\mathcal{J}}(\mathbb{Q}, \bar{\mathbb{P}}_n) = \bar{\mathcal{J}}(\bar{\mathbb{Q}}_n, \bar{\mathbb{P}}_n)$ so that $\bar{\mathcal{J}}(\bar{\mathbb{Q}}_n, \bar{\mathbb{P}}_n) \leq \bar{\mathcal{J}}(\bar{\mathbb{Q}}, \bar{\mathbb{P}}_n)$. Hence $(\bar{\mathbb{P}}_n, \bar{\mathbb{Q}}_n)_{n \geq 1}$ is still a minimizing sequence of $\bar{\mathcal{J}}$. Since $\bar{\mathbb{Q}}_n$ is defined by (C.11), $\sup_{n \geq 1} \left\| \frac{d\bar{\mathbb{Q}}_n}{d\bar{\mathbb{P}}_n} \right\|_{\infty} < +\infty$ and $\sup_{n \geq 1} H(\bar{\mathbb{Q}}_n | \bar{\mathbb{P}}_n) < +\infty$ by Lemma C.9. This establishes item 1.

Let us now prove that the sequence $(\bar{\mathbb{P}}_n, \bar{\mathbb{Q}}_n)_{n \geq 1}$ is relatively compact in $(\mathcal{P}^p(\bar{\Omega}))^2$, i.e. item 2. Notice first that since $(\bar{\mathbb{Q}}_n, \bar{\mathbb{P}}_n)_{n \geq 1}$ is a minimizing sequence of $\bar{\mathcal{J}}$, we have

$$\sup_{n \geq 1} \bar{\mathcal{J}}(\bar{\mathbb{Q}}_n, \bar{\mathbb{P}}_n) < +\infty. \quad (\text{C.12})$$

Let then $q' > 1$. Since (C.4) and (3.1) hold, by Problem 3.15, Chapter 5 in [30] applied with $b(t, y) = \int_{\mathbb{U}} b(t, y, u) \bar{\Lambda}_r(du)$, there exists a constant $C > 0$ which only depends on $C_{b,\sigma}, T, q'$ and d such that

$$\mathbb{E}^{\bar{\mathbb{P}}_n} \left[|X_t - X_s|^{2q'} \right] \leq C |t - s|^{q'},$$

hence $(\mathcal{L}^{\bar{\mathbb{P}}_n}(X))_{n \geq 1}$ is a tight sequence by Kolmogorov criteria, see e.g. Problem 4.11, Chapter 2 in [30]. Moreover, by Hypothesis 3.7 item 3. and (C.12),

$$\begin{aligned} \sup_{n \geq 1} \mathbb{E}^{\bar{\mathbb{P}}_n} \left[\int_0^T \int_{\mathbb{U}} |u|^{p'} \bar{\Lambda}_r(du) dr \right] &\leq C' \left(1 + \sup_{n \geq 1} \mathbb{E}^{\bar{\mathbb{P}}_n} \left[\int_0^T \int_{\mathbb{U}} f(r, X_r, u) \bar{\Lambda}_r(du) dr \right] \right) \\ &\leq C' \left(1 + \sup_{n \geq 1} \bar{\mathcal{J}}(\bar{\mathbb{Q}}_n, \bar{\mathbb{P}}_n) \right) < +\infty. \end{aligned} \quad (\text{C.13})$$

Using again (C.6) and by (C.13) we have

$$\sup_{n \geq 1} \mathbb{E}^{\bar{\mathbb{P}}_n} \left[\left(\sup_{0 \leq r \leq T} |X_r| \right)^{p'} + \int_0^T \int_{\mathbb{U}} |u|^{p'} \bar{\Lambda}_r(du) dr \right] < +\infty, \quad (\text{C.14})$$

where we recall that $p' > p \geq 1$ as fixed at the beginning of Appendix C. Since $(\mathcal{L}^{\bar{\mathbb{P}}_n}(X))_{n \geq 1}$ is tight in $\mathcal{P}(\Omega)$ and (C.14) holds, by Proposition B.3 in [32], the sequence $(\bar{\mathbb{P}}_n)_{n \geq 1}$ is relatively compact in $\mathcal{P}^p(\bar{\Omega})$. Now since $\sup_{n \geq 1} \left\| \frac{d\bar{\mathbb{Q}}_n}{d\bar{\mathbb{P}}_n} \right\|_{\infty} < +\infty$ by item 1., $(\mathcal{L}^{\bar{\mathbb{Q}}_n}(X))_{n \geq 1}$ is also tight and (C.14) is also verified replacing $\bar{\mathbb{P}}_n$ by $\bar{\mathbb{Q}}_n$. Hence $(\bar{\mathbb{Q}}_n)_{n \geq 1}$ is also relatively compact in $\mathcal{P}^p(\bar{\Omega})$. This concludes the proof. \square

Lemma C.11. *Let $(\bar{\mathbb{P}}_n, \bar{\mathbb{Q}}_n)_{n \geq 1}$ be a minimizing sequence of $\bar{\mathcal{J}}$ fulfilling items 1. and 2. of Lemma C.10 statement. Any limit point $(\bar{\mathbb{P}}, \bar{\mathbb{Q}})$ of $(\bar{\mathbb{P}}_n, \bar{\mathbb{Q}}_n)_{n \geq 1}$ belongs to $\bar{\mathcal{A}}$.*

Proof. Up to a subsequence, we can assume that the whole sequence $(\bar{\mathbb{P}}_n, \bar{\mathbb{Q}}_n)_{n \geq 1}$ converges in $(\mathcal{P}^p(\bar{\Omega}))^2$ towards $(\bar{\mathbb{P}}, \bar{\mathbb{Q}})$. Let us prove that $(\bar{\mathbb{P}}, \bar{\mathbb{Q}})$ verifies all items of Definition C.4. We first check item 1. We recall that $E := \bar{\Omega}$ is a Polish space. By Remark 2.2, $(\bar{\mathbb{Q}}, \bar{\mathbb{P}}) \mapsto H(\bar{\mathbb{Q}}|\bar{\mathbb{P}})$ is lower semicontinuous with respect to the weak-star convergence on E^* . Since, the convergence in $\mathcal{P}^p(\bar{\Omega})$ implies the weak convergence, we have

$$H(\bar{\mathbb{Q}}|\bar{\mathbb{P}}) \leq \liminf_{n \rightarrow +\infty} H(\bar{\mathbb{Q}}_n|\bar{\mathbb{P}}_n) < +\infty.$$

where we used item 1. of Lemma C.10 to prove the finiteness in previous inequality.

We now verify item 2. of Definition C.4. Let h belonging to the space $C_c^\infty(\mathbb{R}^d)$ of real-valued smooth functions with compact support on \mathbb{R}^d . We set

$$Y := \int_{[0, \cdot] \times \mathbb{U}} b(r, X_r, u) \Lambda(dr, du).$$

By (C.2), under $\bar{\mathbb{P}}_n$ we have $X = x + Y + M^{\bar{\mathbb{P}}_n}$, where $M^{\bar{\mathbb{P}}_n}$ is a $(\bar{\mathcal{F}}_t)$ -local martingale verifying $[M^{\bar{\mathbb{P}}_n}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$. Then by Itô's formula applied to (C.2) under $\bar{\mathbb{P}}_n$, the process

$$N[h] := h(X_\cdot - Y_\cdot) - h(x) - \frac{1}{2} \int_0^\cdot \text{Tr}[\sigma \sigma^\top(r, X_r) \nabla_x^2 h(X_r - Y_r)] dr$$

is a local martingale under $\bar{\mathbb{P}}_n$. Moreover, since h and $\nabla_x^2 h$ are bounded, (3.1) and (C.6) implies that

$$\mathbb{E}^{\bar{\mathbb{P}}_n} \left[\sup_{0 \leq t \leq T} |N[h]_t| \right] \leq 2\|h\|_\infty + T\|\nabla_x^2 h\|_\infty C_{b,\sigma}^2 \left(1 + \mathbb{E}^{\bar{\mathbb{P}}_n} \left[\sup_{0 \leq t \leq T} |X_t|^2 \right] \right) < +\infty,$$

hence $N[h]$ is a genuine $(\bar{\mathbb{P}}_n, \bar{\mathcal{F}}_t)$ -martingale. We then want to prove that $N[h]$ is also a martingale under $\bar{\mathbb{P}}$. Let $0 \leq s < t \leq T$. Let $\psi : C([0, s], \mathbb{R}^d) \times \mathcal{V}_s \rightarrow \mathbb{R}$ be a bounded continuous function, where \mathcal{V}_s is the set of the elements of \mathcal{V} according to Definition C.2 where we have replaced T with s . Then

$$\mathbb{E}^{\bar{\mathbb{P}}_n} [\psi(\mathbf{1}_{[0,s]}X, \mathbf{1}_{[0,s]}\Lambda) N[h]_t] = \mathbb{E}^{\bar{\mathbb{P}}_n} [\psi(\mathbf{1}_{[0,s]}X, \mathbf{1}_{[0,s]}\Lambda) N[h]_s]. \quad (\text{C.15})$$

On the one hand by Remark C.7 item 3., the map

$$(X, \Lambda) \mapsto \int_{[0,t] \times \mathbb{U}} b(r, X_r, u) \Lambda(dr, du)$$

is continuous for $d_{\bar{\Omega}}$, that is $Y = Y(X, \Lambda)$ is continuous for $d_{\bar{\Omega}}$. Since ψ and h are bounded continuous, the function $(X, \Lambda) \mapsto \psi(\mathbf{1}_{[0,s]}X, \mathbf{1}_{[0,s]}\Lambda) (h(X_s - Y_s) - h(x))$ is bounded continuous for $d_{\bar{\Omega}}$ and since $\bar{\mathbb{P}}_n \rightarrow \bar{\mathbb{P}}$ weakly,

$$\mathbb{E}^{\bar{\mathbb{P}}_n} [\psi(\mathbf{1}_{[0,s]}X, \mathbf{1}_{[0,s]}\Lambda) (h(X_s - Y_s) - h(x))] \xrightarrow{n \rightarrow +\infty} \mathbb{E}^{\bar{\mathbb{P}}} [\psi(\mathbf{1}_{[0,s]}X, \mathbf{1}_{[0,s]}\Lambda) (h(X_s - Y_s) - h(x))]. \quad (\text{C.16})$$

On the other hand, since $\nabla_x^2 h$ is bounded, (3.1) yields for all $r \in [0, T]$

$$\left| Tr[\sigma \sigma^\top(r, X_r) \nabla_x^2 h(X_r - Y_r)] \right| \leq 2C_{b,\sigma}^2 \|\nabla_x^2 h\|_\infty (1 + |X_r|^2).$$

Combining the previous inequality with (C.6) we get that for some $\alpha > 0$,

$$\sup_{n \geq 1} \sup_{r \in [0, T]} \mathbb{E}^{\bar{\mathbb{P}}_n} \left[\left| Tr[\sigma \sigma^\top(r, X_r) \nabla_x^2 h(X_r - Y_r)] \right|^{1+\alpha} \right] < +\infty. \quad (\text{C.17})$$

Hence it holds

$$\sup_{n \geq 1} \mathbb{E}^{\bar{\mathbb{P}}_n} \left[\left| \psi(\mathbf{1}_{[0,s]}X, \mathbf{1}_{[0,s]}\Lambda) \int_0^s Tr[\sigma \sigma^\top(r, X_r) \nabla_x^2 h(X_r - Y_r)] dr \right|^{1+\alpha} \right] < +\infty,$$

and by Lemma C.8 with $E = C([0, s], \mathbb{R}^d) \times \mathcal{V}_s$, we get

$$\begin{aligned} & \mathbb{E}^{\bar{\mathbb{P}}_n} \left[\psi(\mathbf{1}_{[0,s]}X, \mathbf{1}_{[0,s]}\Lambda) \int_0^s Tr[\sigma \sigma^\top(r, X_r) \nabla_x^2 h(X_r - Y_r)] dr \right] \\ & \xrightarrow{n \rightarrow +\infty} \mathbb{E}^{\bar{\mathbb{P}}} \left[\psi(\mathbf{1}_{[0,s]}X, \mathbf{1}_{[0,s]}\Lambda) \int_0^s Tr[\sigma \sigma^\top(r, X_r) \nabla_x^2 h(X_r - Y_r)] dr \right]. \end{aligned} \quad (\text{C.18})$$

Combining (C.16) and (C.18) and letting $n \rightarrow +\infty$ in (C.15) yields

$$\mathbb{E}^{\bar{\mathbb{P}}} [\psi(\mathbf{1}_{[0,s]}X, \mathbf{1}_{[0,s]}\Lambda) N[h]_t] = \mathbb{E}^{\bar{\mathbb{P}}} [\psi(\mathbf{1}_{[0,s]}X, \mathbf{1}_{[0,s]}\Lambda) N[h]_s].$$

Hence the process $N[h]$ is an $((\bar{\mathcal{F}}_t), \bar{\mathbb{P}})$ -martingale for all $h \in C_c^\infty(\mathbb{R}^d)$. By standard stochastic calculus arguments, this implies that under $\bar{\mathbb{P}}$ the process writes $X_t = x + Y_t + M_t^{\bar{\mathbb{P}}}$, where $M^{\bar{\mathbb{P}}}$ is a $(\bar{\mathcal{F}}_t)$ -local martingale verifying $[M^{\bar{\mathbb{P}}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$. Item 2. of Definition C.4 is verified and we conclude that $(\bar{\mathbb{P}}, \bar{\mathbb{Q}}) \in \bar{\mathcal{A}}$. \square

Proposition C.12. *The Problem (C.3) admits a solution $(\bar{\mathbb{P}}^*, \bar{\mathbb{Q}}^*) \in \bar{\mathcal{A}}$, in the sense that $\bar{\mathcal{J}}^* = \bar{\mathcal{J}}(\bar{\mathbb{Q}}^*, \bar{\mathbb{P}}^*)$ which verifies $\|d\bar{\mathbb{Q}}^*/d\bar{\mathbb{P}}^*\|_\infty < +\infty$.*

Proof. Let $(\bar{\mathbb{P}}_n, \bar{\mathbb{Q}}_n)_{n \geq 1}$ be the minimizing sequence given by Lemma C.10 and let $(\bar{\mathbb{P}}, \bar{\mathbb{Q}})$ be any limit point of the sequence $(\bar{\mathbb{P}}_n, \bar{\mathbb{Q}}_n)_{n \geq 1}$. Up to a subsequence we can assume that the whole sequence $(\bar{\mathbb{P}}_n, \bar{\mathbb{Q}}_n)_{n \geq 1}$ converges towards $(\bar{\mathbb{P}}, \bar{\mathbb{Q}})$ in $(\mathcal{P}^p(\bar{\Omega}))^2$. Recall that by Remark C.7 the map

$$(X, \Lambda) \mapsto \int_{[0, T] \times \mathbb{U}} f(r, X_r, u) \Lambda(dr, du)$$

is continuous for $d_{\bar{\Omega}}$. Now by (C.5), we have

$$\left| \int_{[0, T] \times \mathbb{U}} f(r, X_r, u) \Lambda(dr, du) \right| \leq C_{f, g} \left(1 + \sup_{0 \leq r \leq T} |X_r|^p \right),$$

and by (C.6), we deduce that

$$\sup_{n \geq 1} \mathbb{E}^{\bar{\mathbb{P}}_n} \left[\left| \int_{[0, T] \times \mathbb{U}} f(r, X_r, u) \Lambda(dr, du) \right|^{1+\alpha} \right] < +\infty$$

for any $\alpha > 0$. Since $\sup_{n \geq 1} \left\| \frac{d\bar{\mathbb{Q}}_n}{d\bar{\mathbb{P}}_n} \right\|_\infty$ by item 1. of Lemma C.10, it also holds that

$$\sup_{n \geq 1} \mathbb{E}^{\bar{\mathbb{Q}}_n} \left[\left| \int_{[0, T] \times \mathbb{U}} f(r, X_r, u) \Lambda(dr, du) \right|^{1+\alpha} \right] < +\infty.$$

Then by Lemma C.8 applied with $E = \bar{\Omega}$, we have

$$\mathbb{E}^{\bar{\mathbb{Q}}_n} \left[\int_{[0, T] \times \mathbb{U}} f(r, X_r, u) \Lambda(dr, du) \right] \xrightarrow{n \rightarrow +\infty} \mathbb{E}^{\bar{\mathbb{Q}}} \left[\int_{[0, T] \times \mathbb{U}} f(r, X_r, u) \Lambda(dr, du) \right]. \quad (\text{C.19})$$

Again by Remark 2.2, $(\mathbb{Q}, \mathbb{P}) \mapsto H(\mathbb{Q}|\mathbb{P})$ is lower semicontinuous with respect to the weak-star convergence on $\bar{\Omega}$, and we have

$$H(\bar{\mathbb{Q}}|\bar{\mathbb{P}}) \leq \liminf_{n \rightarrow +\infty} H(\bar{\mathbb{Q}}_n|\bar{\mathbb{P}}_n). \quad (\text{C.20})$$

Combining (C.19) and (C.20), we get

$$\bar{\mathcal{J}}^* = \lim_{n \rightarrow +\infty} \bar{\mathcal{J}}(\bar{\mathbb{Q}}_n, \bar{\mathbb{P}}_n) = \liminf_{n \rightarrow +\infty} \bar{\mathcal{J}}(\bar{\mathbb{Q}}_n, \bar{\mathbb{P}}_n) \geq \bar{\mathcal{J}}(\bar{\mathbb{Q}}, \bar{\mathbb{P}}).$$

By Lemma C.11, $(\bar{\mathbb{P}}, \bar{\mathbb{Q}}) \in \bar{\mathcal{A}}$ and we conclude that $(\bar{\mathbb{P}}, \bar{\mathbb{Q}})$ achieves the minimum of $\bar{\mathcal{J}}$. Moreover, we set $\bar{\mathbb{P}}^* := \bar{\mathbb{P}}$ and

$$d\bar{\mathbb{Q}}^* := \frac{\exp\left(-\int_0^T \int_{\mathbb{U}} f(r, X_r, u) \bar{\Lambda}_r(du) dr - g(X_T)\right)}{\mathbb{E}^{\bar{\mathbb{P}}}\left[\exp\left(-\int_0^T \int_{\mathbb{U}} f(r, X_r, u) \bar{\Lambda}_r(du) dr - g(X_T)\right)\right]} d\bar{\mathbb{P}}^*.$$

By Proposition 3.13 we have that $\bar{\mathcal{J}}(\bar{\mathbb{Q}}, \bar{\mathbb{P}}) \geq \bar{\mathcal{J}}(\bar{\mathbb{Q}}^*, \bar{\mathbb{P}}^*)$. Since $(\bar{\mathbb{P}}^*, \bar{\mathbb{Q}}^*) \in \bar{\mathcal{A}}$, $\bar{\mathcal{J}}(\bar{\mathbb{Q}}^*, \bar{\mathbb{P}}^*) = \bar{\mathcal{J}}(\bar{\mathbb{Q}}, \bar{\mathbb{P}})$ and $(\bar{\mathbb{P}}^*, \bar{\mathbb{Q}}^*)$ also achieves the minimum of $\bar{\mathcal{J}}$. Finally $\|d\bar{\mathbb{Q}}^*/d\bar{\mathbb{P}}^*\|_\infty < +\infty$ by Lemma C.9. \square

Lemma C.13. *Let $(\bar{\mathbb{P}}, \bar{\mathbb{Q}}) \in \bar{\mathcal{A}}$ such that $\|d\bar{\mathbb{Q}}/d\bar{\mathbb{P}}\|_\infty < +\infty$. There exists $(\mathbb{P}, \mathbb{Q}) \in \mathcal{A}$ with $\mathbb{P} \in \mathcal{P}_U^{Markov}$ such that $\bar{\mathcal{J}}(\bar{\mathbb{Q}}, \bar{\mathbb{P}}) \geq \mathcal{J}(\mathbb{Q}, \mathbb{P})$.*

Proof. Since $H(\bar{\mathbb{Q}}|\bar{\mathbb{P}}) < +\infty$, by Theorem A.1 applied on the space $\bar{\Omega}$ equipped with the probability measures $\bar{\mathbb{P}}$ and $\bar{\mathbb{Q}}$ with $\delta_r = \int_U b(r, X_r, u)\bar{\Lambda}_r(du)$, $a_r = \sigma\sigma^\top(r, X_r)$, there exists a $(\bar{\mathcal{F}}_t)$ -progressively measurable process $\bar{\alpha}$ such that under $\bar{\mathbb{Q}}$, the canonical process decomposes as

$$X_t = x + \int_0^t \int_U b(r, X_r, u)\bar{\Lambda}_r(du)dr + \int_0^t \sigma\sigma^\top(r, X_r)\bar{\alpha}_r dr + M_t^{\bar{\mathbb{Q}}}, \quad (\text{C.21})$$

where the local martingale $M^{\bar{\mathbb{Q}}}$ verifies $[M^{\bar{\mathbb{Q}}}] = \int_0^\cdot \sigma\sigma^\top(r, X_r)dr$ and

$$H(\bar{\mathbb{Q}}|\bar{\mathbb{P}}) \geq \frac{1}{2}\mathbb{E}^{\bar{\mathbb{Q}}}\left[\int_0^T |\sigma^\top(r, X_r)\bar{\alpha}_r|^2 dr\right]. \quad (\text{C.22})$$

The proof consists in two parts. In the first part we establish some useful estimates related to $\bar{\mathbb{Q}}$ and to the previous decomposition. In the second part we introduce a probability measure $\mathbb{Q} \in \mathcal{P}(\Omega)$ mimicking the time marginals of $\bar{\mathbb{Q}}$ for all $t \in [0, T]$ and another probability measure $\mathbb{P} \in \mathcal{P}_U^{Markov}$ such that $\bar{\mathcal{J}}(\bar{\mathbb{Q}}, \bar{\mathbb{P}}) \geq \mathcal{J}(\mathbb{Q}, \mathbb{P})$.

1. Note first that since $\|d\bar{\mathbb{Q}}/d\bar{\mathbb{P}}\|_\infty < +\infty$, for all $q \geq 1$, by (C.6) in Remark C.7 we have

$$\mathbb{E}^{\bar{\mathbb{Q}}}\left[\left(\sup_{0 \leq r \leq T} |X_r|\right)^q\right] \leq \left\|\frac{d\bar{\mathbb{Q}}}{d\bar{\mathbb{P}}}\right\|_\infty^q \mathbb{E}^{\bar{\mathbb{P}}}\left[\left(\sup_{0 \leq r \leq T} |X_r|\right)^q\right] < +\infty. \quad (\text{C.23})$$

It immediately follows from (C.4) and (C.23) that

$$\mathbb{E}^{\bar{\mathbb{Q}}}\left[\int_0^T \left|\int_U b(r, X_r, u)\bar{\Lambda}_r(du)\right| dr\right] < +\infty, \quad (\text{C.24})$$

and from (C.5) and (C.23) that

$$\mathbb{E}^{\bar{\mathbb{Q}}}\left[\int_0^T \left|\int_U f(r, X_r, u)\bar{\Lambda}_r(du)\right|\right] < +\infty. \quad (\text{C.25})$$

Finally being σ of linear growth because of Hypothesis 3.1 item 2. and (C.6) in Remark C.7, it holds that

$$\mathbb{E}^{\bar{\mathbb{P}}}\left[\int_0^T \|\sigma(r, X_r)\|^q dr\right] < +\infty \quad (\text{C.26})$$

for all $q \geq 1$. Then we can apply Lemma A.2 item 1. which implies that for any $1 < q < 2$

$$\mathbb{E}^{\bar{\mathbb{Q}}}\left[\int_0^T |\sigma\sigma^\top(r, X_r)\bar{\alpha}_r|^q dr\right] < +\infty. \quad (\text{C.27})$$

2. We set $\beta_t := \int_U b(t, X_t, u)\bar{\Lambda}_t(du) + \sigma\sigma^\top(t, X_t)\bar{\alpha}_t$ so that, taking into account (C.21), X decomposes as $X_t = x + \int_0^t \beta_r dr + M_t^{\bar{\mathbb{Q}}}$, where $M_t^{\bar{\mathbb{Q}}}$ is a local martingale such that $[M^{\bar{\mathbb{Q}}}] = \int_0^\cdot \sigma\sigma^\top(r, X_r)dr$ under $\bar{\mathbb{Q}}$ and (C.22) rewrites

$$H(\bar{\mathbb{Q}}|\bar{\mathbb{P}}) \geq \frac{1}{2} \mathbb{E}^{\bar{\mathbb{Q}}} \left[\int_0^T \left| \sigma^{-1}(r, X_r) \left(\beta_r - \int_{\mathbb{U}} b(r, X_r, u) \bar{\Lambda}_r(du) \right) \right|^2 dr \right], \quad (\text{C.28})$$

where σ^{-1} denotes again the generalized right-inverse of σ .

It follows from (C.24), (C.27), and (C.26) together with the assumption $\|d\bar{\mathbb{Q}}/d\bar{\mathbb{P}}\|_\infty < +\infty$, that $\mathbb{E}^{\bar{\mathbb{Q}}} \left[\int_0^T (|\beta_r| + \|\sigma(r, X_r)\|) dr \right] < +\infty$. Then by Corollary 3.7 in [9] there exists a measurable function $\Gamma : [0, T] \times \mathbb{R}^d \mapsto \mathbb{R}^d$ and a probability measure \mathbb{Q} on (Ω, \mathcal{F}) such that the following holds.

- For all $0 \leq t \leq T$,

$$\Gamma(t, X_t) = \mathbb{E}^{\bar{\mathbb{Q}}}[\beta_t | X_t] \quad dt \otimes d\bar{\mathbb{Q}}\text{-a.s.} \quad (\text{C.29})$$

- Under \mathbb{Q} the canonical process can be expressed as $X_t = x + \int_0^t \Gamma(r, X_r) dr + M_t^{\mathbb{Q}}$, where $M^{\mathbb{Q}}$ is a (\mathcal{F}_t) -local martingale with $[M^{\mathbb{Q}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$.
- $\mathcal{L}^{\mathbb{Q}}(X_t) = \mathcal{L}^{\bar{\mathbb{Q}}}(X_t)$, for all $t \in [0, T]$.

Since the estimates (C.24) and (C.25) hold, Lemma B.2 applied with $\Omega = \bar{\Omega}$, $\mathbb{P} = \bar{\mathbb{Q}}$ and

$$(y_t, z_t) = \left(\int_{\mathbb{U}} b(t, X_t, u) \bar{\Lambda}_t(du), \int_{\mathbb{U}} f(t, X_t, u) \bar{\Lambda}_t(du) \right)$$

gives the existence of a measurable function $\bar{u} \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$ such that for almost all $t \in [0, T]$, $\bar{\mathbb{Q}}$ -a.s.,

$$\begin{cases} \mathbb{E}^{\bar{\mathbb{Q}}} \left[\int_{\mathbb{U}} b(t, X_t, u) \bar{\Lambda}_t(du) \middle| X_t \right] = b(t, X_t, \bar{u}(t, X_t)) \\ \mathbb{E}^{\bar{\mathbb{Q}}} \left[\int_{\mathbb{U}} f(t, X_t, u) \bar{\Lambda}_t(du) \middle| X_t \right] \geq f(t, X_t, \bar{u}(t, X_t)). \end{cases} \quad (\text{C.30})$$

(C.30) together with Fubini's theorem then gives

$$\begin{aligned} \mathbb{E}^{\bar{\mathbb{Q}}} \left[\int_0^T \int_{\mathbb{U}} f(r, X_r, u) \bar{\Lambda}_r(du) + g(X_T) \right] &= \mathbb{E}^{\bar{\mathbb{Q}}} \left[\int_0^T f(r, X_r, \bar{u}(r, X_r)) dr + g(X_T) \right] \\ &= \mathbb{E}^{\mathbb{Q}} \left[\int_0^T f(r, X_r, \bar{u}(r, X_r)) dr + g(X_T) \right]. \end{aligned} \quad (\text{C.31})$$

(C.30) together with Fubini's theorem and Jensen's inequality for the conditional expectation applied to (C.28) yields

$$\begin{aligned} H(\bar{\mathbb{Q}}|\bar{\mathbb{P}}) &\geq \frac{1}{2} \int_0^T \mathbb{E}^{\bar{\mathbb{Q}}} \left[\left| \sigma^{-1}(r, X_r) \mathbb{E}^{\bar{\mathbb{Q}}} \left[\left(\beta_r - \int_{\mathbb{U}} b(r, X_r, u) \bar{\Lambda}_r(du) \right) \middle| X_r \right] \right|^2 \right] dr \\ &= \frac{1}{2} \int_0^T \mathbb{E}^{\bar{\mathbb{Q}}} \left[|\sigma^{-1}(r, X_r) (\Gamma(r, X_r) - b(r, X_r, \bar{u}(r, X_r)))|^2 \right] dr, \end{aligned} \quad (\text{C.32})$$

where we used (C.29) and (C.30) in the last equality. Since $\bar{\mathbb{Q}}$ and \mathbb{Q} have the same time marginals, we deduce from (C.32) and Fubini's theorem that

$$\begin{aligned} H(\bar{\mathbb{Q}}|\bar{\mathbb{P}}) &\geq \frac{1}{2} \int_0^T \mathbb{E}^{\bar{\mathbb{Q}}} [|\sigma^{-1}(r, X_r)(\Gamma(r, X_r) - b(r, X_r, \bar{u}(r, X_r)))|^2] dr \\ &= \frac{1}{2} \mathbb{E}^{\bar{\mathbb{Q}}} \left[\int_0^T |\sigma^{-1}(r, X_r)(\Gamma(r, X_r) - b(r, X_r, \bar{u}(r, X_r)))|^2 dr \right]. \end{aligned} \quad (\text{C.33})$$

Finally, let $\mathbb{P} := \mathbb{P}^{\bar{u}} \in \mathcal{P}_{\mathbb{U}}^{\text{Markov}}$ be the unique probability measure given by Proposition 3.5. We recall that, by Remark 3.6, the SDE

$$X_t = x + \int_0^t b(r, X_r, \bar{u}(r, X_r)) dr + M_t^{\mathbb{P}},$$

where $M^{\mathbb{P}}$ is a local martingale with $[M^{\mathbb{P}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$, admits uniqueness in law.

As $H(\bar{\mathbb{Q}}|\bar{\mathbb{P}}) < \infty$, (C.33) and Lemma A.4 2. implies that $H(\mathbb{Q}|\mathbb{P}) < +\infty$ and that

$$H(\mathbb{Q}|\mathbb{P}) = \frac{1}{2} \mathbb{E}^{\mathbb{Q}} \left[\int_0^T |\sigma^{-1}(r, X_r)(\Gamma(r, X_r) - b(r, X_r, \bar{u}(r, X_r)))|^2 dr \right]. \quad (\text{C.34})$$

In particular, $(\mathbb{P}, \mathbb{Q}) \in \mathcal{A}$ and combining (C.31), (C.33) and (C.34) yields $\bar{\mathcal{J}}(\bar{\mathbb{Q}}, \bar{\mathbb{P}}) \geq \mathcal{J}(\mathbb{Q}, \mathbb{P})$. This concludes the proof. □

We are now ready to prove Theorem 3.15.

Proof of Theorem 3.15. Let $(\mathbb{P}, \mathbb{Q}) \in \mathcal{A}$. Let $\bar{\mathbb{P}}$ (resp. $\bar{\mathbb{Q}}$) be the law of $(X, dt\delta_{\nu_t^{\bar{\mathbb{P}}}}(du))$ under \mathbb{P} (resp. \mathbb{Q}). Then $(\bar{\mathbb{P}}, \bar{\mathbb{Q}}) \in (\mathcal{P}(\bar{\Omega}))^2$ and X has clearly the decomposition (C.2) under $\bar{\mathbb{P}}$. Furthermore, one has $d\bar{\mathbb{Q}}/d\bar{\mathbb{P}} = d\mathbb{Q}/d\mathbb{P} \circ \pi_X$, where π_X is the first coordinate projection on $\bar{\Omega}$, and this yields

$$H(\bar{\mathbb{Q}}|\bar{\mathbb{P}}) = \mathbb{E}^{\bar{\mathbb{Q}}} \left[\log \frac{d\bar{\mathbb{Q}}}{d\bar{\mathbb{P}}} \right] = \mathbb{E}^{\bar{\mathbb{Q}}} \left[\log \frac{d\mathbb{Q}}{d\mathbb{P}} \circ \pi_X \right] = \mathbb{E}^{\mathbb{Q}} \left[\log \frac{d\mathbb{Q}}{d\mathbb{P}}(X) \right] = H(\mathbb{Q}|\mathbb{P}).$$

Hence $H(\bar{\mathbb{Q}}|\bar{\mathbb{P}}) < +\infty$, $(\bar{\mathbb{P}}, \bar{\mathbb{Q}}) \in \bar{\mathcal{A}}$ and since

$$\mathbb{E}^{\bar{\mathbb{Q}}} \left[\int_{[0, T] \times \mathbb{U}} f(r, X_r, u) \Lambda(dr, du) + g(X_T) \right] = \mathbb{E}^{\mathbb{Q}} \left[\int_0^T f(r, X_r, \nu_r^{\mathbb{P}}) dr + g(X_T) \right],$$

we get $\bar{\mathcal{J}}(\bar{\mathbb{Q}}, \bar{\mathbb{P}}) = \mathcal{J}(\mathbb{Q}, \mathbb{P})$. Previous computations then show that $\bar{\mathcal{J}}^* \leq \mathcal{J}^*$. Let now $(\bar{\mathbb{P}}^*, \bar{\mathbb{Q}}^*) \in \bar{\mathcal{A}}$ be the solution of (C.3) given by Proposition C.12. In particular, $\bar{\mathcal{J}}(\bar{\mathbb{Q}}^*, \bar{\mathbb{P}}^*) = \bar{\mathcal{J}}^* \leq \mathcal{J}^*$. Let also $(\mathbb{P}^*, \mathbb{Q}^*) \in \mathcal{A}$ be given by Lemma C.13 applied to $(\bar{\mathbb{P}}^*, \bar{\mathbb{Q}}^*)$. We have $\bar{\mathcal{J}}(\bar{\mathbb{Q}}^*, \bar{\mathbb{P}}^*) \geq \mathcal{J}(\mathbb{Q}^*, \mathbb{P}^*)$, hence $\mathcal{J}^* \geq \mathcal{J}(\mathbb{Q}^*, \mathbb{P}^*)$, that is $\mathcal{J}^* = \mathcal{J}(\mathbb{Q}^*, \mathbb{P}^*)$. This implies that $(\mathbb{P}^*, \mathbb{Q}^*)$ is a solution of Problem (1.5). □

D Strong and weak controls

Let $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t \in [0, T]}, \tilde{\mathbb{P}})$ be a filtered probability space endowed with a Brownian motion W . Let \mathcal{V} be the set of $(\tilde{\mathcal{F}}_t)$ -progressively measurable processes ν on $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ taking values in \mathbb{U} such that equation (1.2) has a unique strong solution. We give here some details on the equivalence between a strong formulation of our stochastic optimal control (1.1) formulated on the generic probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$, and our optimization problem (1.3). We have the following result.

Proposition D.1. *Assume Hypotheses 3.7 and 3.1. Recall the definition (1.1) of J_{strong}^* and (1.3) of J^* . Then $J_{strong}^* = J^*$.*

Proof. (i) We first prove that $J_{strong}^* \geq J^*$. Let $(\nu^n)_{n \geq 0}$ be a minimizing sequence of elements of \mathcal{V} for Problem (1.1). For any $n \in \mathbb{N}$, Lemma 3.11 and (3.8) yields

$$\mathbb{E}^{\tilde{\mathbb{P}}} \left[\int_0^T \left(|b(r, X_r^{\nu^n}, \nu_r^n)| + |f(r, X_r^{\nu^n}, \nu_r^n)| + \|\sigma \sigma^\top(r, X_r^{\nu^n})\| \right) dr \right] < +\infty. \quad (\text{D.1})$$

Then by Corollary 3.7 in [9] there exist a measurable function $\Gamma \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{R}^d)$ and a probability measure $\mathbb{P} \in \mathcal{P}(\Omega)$ such that

- For all $0 \leq t \leq T$, $\Gamma(t, X_t) = \mathbb{E}^{\tilde{\mathbb{P}}} [b(t, X_t^{\nu^n}, \nu_t^n) | X_t^{\nu^n}] d\tilde{\mathbb{P}} \otimes dt$ -a.e.
- Under \mathbb{P} the canonical process can be expressed as $X_t = x + \int_0^t \Gamma(r, X_r) dr + M_t^{\mathbb{P}}$, where $M^{\mathbb{P}}$ is a (\mathcal{F}_t) -local martingale with $[M^{\mathbb{P}}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$.
- $\mathcal{L}^{\mathbb{P}}(X_t) = \mathcal{L}^{\tilde{\mathbb{P}}}(X_t)$, $\forall t \in [0, T]$.

Since (D.1) holds, by Lemma B.2 applied with $\Omega = \tilde{\Omega}$, $\mathbb{P} = \tilde{\mathbb{P}}$, $X = X^{\nu^n}$ and $(y_t, z_t) = (b(t, X_t^{\nu^n}, \nu_t^n), f(t, X_t^{\nu^n}, \nu_t^n))$, there exists a function $u^n \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$ such that for almost all $t \in [0, T]$, \mathbb{P} -a.s.,

$$\begin{cases} \mathbb{E}^{\tilde{\mathbb{P}}} [b(t, X_t^{\nu^n}, \nu_t^n) | X_t^{\nu^n}] = b(t, X_t^{\nu^n}, u^n(t, X_t^{\nu^n})) \\ \mathbb{E}^{\tilde{\mathbb{P}}} [f(t, X_t^{\nu^n}, \nu_t^n) | X_t^{\nu^n}] \geq f(t, X_t^{\nu^n}, u^n(t, X_t^{\nu^n})). \end{cases} \quad (\text{D.2})$$

By Fubini's theorem and Jensen's inequality for the conditional expectation, by (D.2) we have

$$\begin{aligned} \mathbb{E}^{\tilde{\mathbb{P}}} \left[\int_0^T f(r, X_r^{\nu^n}, \nu_r^n) dr + g(X_T^{\nu^n}) \right] &\geq \mathbb{E}^{\tilde{\mathbb{P}}} \left[\int_0^T f(r, X_r^{\nu^n}, u^n(r, X_r^{\nu^n})) dr + g(X_T^{\nu^n}) \right] \\ &= \mathbb{E}^{\mathbb{P}} \left[\int_0^T f(r, X_r, u^n(r, X_r)) dr + g(X_T) \right] \\ &\geq \inf_{\mathbb{P} \in \mathcal{P}_{\mathbb{U}}} \mathbb{E}^{\mathbb{P}} \left[\int_0^T f(r, X_r, \nu_r^{\mathbb{P}}) dr + g(X_T) \right], \end{aligned} \quad (\text{D.3})$$

where, for the latter inequality, we have used the fact that $\mathbb{P} \in \mathcal{P}_{\mathbb{U}}$. From (D.3), for all $n \in \mathbb{N}$, we have

$$\mathbb{E}^{\tilde{\mathbb{P}}} \left[\int_0^T f(r, X_r^{\nu^n}, \nu_r^n) dr + g(X_T^{\nu^n}) \right] \geq J^*,$$

and letting $n \rightarrow +\infty$ yields $J_{strong}^* \geq J^*$.

(ii) We now prove that $J^* \geq J_{strong}^*$. Let us consider a minimizing sequence $(\mathbb{P}_n)_{n \geq 0}$ of elements of $\mathcal{P}_{\mathbb{U}}$ for Problem (1.3). Notice that, taking into account Lemma 3.11, the estimate (D.1) still holds if we replace $(X^{\nu^n}, \nu^n, \tilde{\mathbb{P}})$ by $(X, \nu^{\mathbb{P}_n}, \mathbb{P}_n)$. Then for all $n \in \mathbb{N}$, again by Corollary 3.7 in [9] together with Lemma B.2 applied with $\mathbb{P} = \mathbb{P}_n$, $(y_t, z_t) = (b(t, X_t, \nu_t^{\mathbb{P}_n}), f(t, X_t, \nu_t^{\mathbb{P}_n}))$, there exist a function $u^n \in \mathcal{B}([0, T] \times \mathbb{R}^d, \mathbb{U})$ and a probability measure $\hat{\mathbb{P}}_n$ on (Ω, \mathcal{F}) such that the following holds.

- For almost all $t \in [0, T]$, \mathbb{P}_n -a.s.

$$\begin{cases} \mathbb{E}^{\mathbb{P}_n} \left[b(t, X_t, \nu_t^{\mathbb{P}_n}) \middle| X_t \right] = b(t, X_t, u^n(t, X_t)) \\ \mathbb{E}^{\mathbb{P}_n} \left[f(t, X_t, \nu_t^{\mathbb{P}_n}) \middle| X_t \right] \geq f(t, X_t, u^n(t, X_t)). \end{cases}$$

- Under $\hat{\mathbb{P}}$ the canonical process decomposes as

$$X_t = x + \int_0^t b(r, X_r, u^n(r, X_r)) dr + M_t^{\hat{\mathbb{P}}_n},$$

where $M^{\hat{\mathbb{P}}_n}$ is an (\mathcal{F}_t) -local martingale such that $[M^{\hat{\mathbb{P}}_n}] = \int_0^\cdot \sigma \sigma(r, X_r) dr$.

- $\mathcal{L}^{\mathbb{P}}(X_t) = \mathcal{L}^{\hat{\mathbb{P}}_n}(X_t)$.

On the one hand, Fubini's theorem and Jensen's inequality for conditional expectation yield

$$\mathbb{E}^{\mathbb{P}_n} \left[\int_0^T f(r, X_r, \nu_r^{\mathbb{P}_n}) dr + g(X_T) \right] \geq \mathbb{E}^{\hat{\mathbb{P}}_n} \left[\int_0^T f(r, X_r, u^n(r, X_r)) dr + g(X_T) \right]. \quad (\text{D.4})$$

On the other hand, Theorem 1.1 in [44] ensures the existence of a unique (strong) solution $X = X^{\nu^n}$ (on the space $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t)_{t \in [0, T]}, \tilde{\mathbb{P}})$) to the SDE

$$dX_t = b(t, X_t, u^n(t, X_t)) dt + \sigma(t, X_t) dW_t, \quad X_0 = x.$$

In particular the process $\nu^n := u^n(\cdot, X^{\hat{u}})$ is an element of \mathcal{V} , and we get by (D.4) that

$$J(\mathbb{P}_n) \geq \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T f(r, X_r^{\nu^n}, \nu_r^n) dr + g(X_T^{\nu^n}) \right] \geq \inf_{\nu \in \mathcal{V}} \mathbb{E}^{\hat{\mathbb{P}}} \left[\int_0^T f(r, X_r^\nu, \nu_r) dr + g(X_T^\nu) \right] = J_{strong}^*.$$

The previous expression gives $J(\mathbb{P}_n) \geq J_{strong}^*$ for all $n \in \mathbb{N}$, and letting $n \rightarrow +\infty$ yields $J^* \geq J_{strong}^*$.

By item (i), we have $J_{strong}^* \geq J^*$, whereas by item (ii), $J^* \geq J_{strong}^*$. Hence $J^* = J_{strong}^*$, and this concludes the proof. \square

E Proofs of two technical lemmata

Proof of Lemma 4.2. 1. The function $\bar{F}_\beta^{t,x}$ is coercive on $K(t, x)$ in the sense of Definition 2.13 in [1]. Since $K(t, x)$ is closed (see Remark 3.8), Theorem 2.14 in [1] gives the existence of a

minimum (y^*, z^*) to $\bar{F}_\beta^{t,x}$ on $K(t, x)$, which is unique since $\bar{F}_\beta^{t,x}$ is strictly convex. Let then $(y, z) \in K(t, x)$. Since $K(t, x)$ is convex, $(\lambda y + (1 - \lambda)y^*, \lambda z + (1 - \lambda)z^*) \in K(t, x)$, for any $\lambda \in]0, 1]$. By definition of (y^*, z^*) we then have

$$\frac{\bar{F}_\beta^{t,x}(\lambda y + (1 - \lambda)y^*, \lambda z + (1 - \lambda)z^*) - \bar{F}_\beta^{t,x}(y^*, z^*)}{\lambda} \geq 0, \quad \text{for all } \lambda \in]0, 1],$$

and since $\bar{F}_\beta^{t,x}$ is of class C^1 on $\mathbb{R}^d \times \mathbb{R}$, letting $\lambda \rightarrow 0$ in the previous inequality yields $\langle \nabla_{(y,z)} \bar{F}_\beta^{t,x}(y^*, z^*), (y, z) \rangle \geq 0$, which rewrites as (4.4).

2. We first observe that $u^* \in \arg \min_{a \in \mathbb{U}} F_\beta(t, x, a)$ is equivalent to

$$F_\beta(t, x, u^*) \leq F_\beta(t, x, a), \quad \forall a \in \mathbb{U} \quad (\text{E.1})$$

and (4.6) is equivalent to

$$\bar{F}_\beta^{t,x}(y, z) \geq \bar{F}_\beta^{t,x}(y^*, z^*), \quad \forall (y, z) \in K(t, x). \quad (\text{E.2})$$

For any $a \in \mathbb{U}$ we set now $(y(a), z(a)) := (b(t, x, a), f(t, x, a))$. Clearly $(y(a), z(a)) \in K(t, x)$ and $(y, z) \in K(t, x)$ if and only if there is $a \in \mathbb{U}$ with $(y, z) = (y(a), z(a))$ and $z \geq z(a)$.

In fact we have

$$\bar{F}_\beta^{t,x}(y(a), z(a)) = F_\beta(t, x, a). \quad (\text{E.3})$$

(a) Let $u^* \in \mathbb{U}$ such that $y^* = y(u^*)$ and $z^* \geq z(u^*)$ and we prove (E.1). By (E.3), for all $a \in \mathbb{U}$, we have

$$\begin{aligned} F_\beta(t, x, a) &= \bar{F}_\beta^{t,x}(y(a), z(a)) \geq \bar{F}_\beta^{t,x}(y^*, z^*) = \bar{F}_\beta^{t,x}(y(u^*), z^*) \\ &\geq \bar{F}_\beta^{t,x}(y(u^*), z(u^*)) = F_\beta(t, x, u^*) \end{aligned}$$

and (E.2) follows.

(b) Let u^* such that (E.1) holds and $a \in \mathbb{U}$ such that $y = y(a), z \geq z(a)$. Then, using again (E.3) we get

$$\begin{aligned} \bar{F}_\beta^{t,x}(y, z) &\geq \bar{F}_\beta^{t,x}(y(a), z(a)) = F_\beta(t, x, a) \\ &\geq F_\beta(t, x, u^*) = \bar{F}_\beta^{t,x}(y(u^*), z(u^*)), \end{aligned}$$

and (4.6) holds. □

Proof of Lemma 4.12. By Remark 3.14 and (3.14) in Theorem 3.15, the quantity $C_\infty := \|d\mathbb{Q}_\epsilon^*/d\mathbb{P}_\epsilon^*\|_\infty$ is finite, so that

$$\mathbb{E}^{\mathbb{Q}_\epsilon^*} \left[\sup_{0 \leq r \leq T} |X_r|^q \right] \leq C_\infty \mathbb{E}^{\mathbb{P}_\epsilon^*} \left[\sup_{0 \leq r \leq T} |X_r|^q \right] \leq C_\infty C(q) < +\infty, \quad (\text{E.4})$$

where $C(q)$ is given by Lemma 3.11. As $H(\mathbb{Q}_\epsilon^*|\mathbb{P}_\epsilon^*) < +\infty$, by Theorem A.1 there exists a progressively measurable process α such that under \mathbb{Q}_ϵ^* the canonical process decomposes as

$$X_t = x + \int_0^t b(r, X_r, u_\epsilon^*(r, X_r))dr + \int_0^t \sigma \sigma^\top(r, X_r) \alpha_r dr + M_t^*, \quad t \in [0, T],$$

where $M^* := M^{\mathbb{Q}_\epsilon^*}$ is a local martingale verifying $[M^*] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$ and u_ϵ^* is the Borel function introduced in Theorem 3.15.

Moreover,

$$H(\mathbb{Q}_\epsilon^*|\mathbb{P}_\epsilon^*) \geq \frac{1}{2} \mathbb{E}^{\mathbb{Q}_\epsilon^*} \left[\int_0^T |\sigma^\top(r, X_r) \alpha_r|^2 dr \right]. \quad (\text{E.5})$$

We set $\beta_t := b(t, X_t, u_\epsilon^*(t, X_t)) + \sigma \sigma^\top(t, X_t) \alpha_t$

Let $k \geq 1$. On the one hand, combining (3.1), (3.2) (E.4) and (E.5) and taking into account (4.9) for $k+1$ replaced with k , we have

$$\begin{aligned} \mathbb{E}^{\mathbb{Q}_\epsilon^*} \left[\int_0^T |\sigma^{-1}(r, X_r) (\beta_r - b(r, X_r, u^k(r, X_r)))|^2 dr \right] &\leq 4 \mathbb{E}^{\mathbb{Q}_\epsilon^*} \left[\int_0^T |\sigma^{-1}(r, X_r) b(r, X_r, u_\epsilon^*(r, X_r))|^2 dr \right] \\ &\quad + 4 \mathbb{E}^{\mathbb{Q}_\epsilon^*} \left[\int_0^T |\sigma^{-1}(r, X_r) b(r, X_r, u^k(r, X_r))|^2 dr \right] \\ &\quad + 4 \mathbb{E}^{\mathbb{Q}_\epsilon^*} \left[\int_0^T |\sigma^\top(r, X_r) \alpha_r|^2 dr \right] \\ &\leq 8c_\sigma C_{b,\sigma}^2 \left(T + \int_0^T \mathbb{E}^{\mathbb{Q}_\epsilon^*} [|X_r|^2] dr \right) + 8H(\mathbb{Q}_\epsilon^*|\mathbb{P}_\epsilon^*) \\ &\leq 8Tc_\sigma C_{b,\sigma}^2 (1 + C_\infty C(2)) + 8H(\mathbb{Q}_\epsilon^*|\mathbb{P}_\epsilon^*). \end{aligned} \quad (\text{E.6})$$

We recall that, by Remark 3.6, the SDE

$$X_t = x + \int_0^t b(r, X_r, u^k(r, X_r))dr + M_t^{\mathbb{P}_k},$$

where $M^{\mathbb{P}_k}$ is a local martingale with $[M^{\mathbb{P}_k}] = \int_0^\cdot \sigma \sigma^\top(r, X_r) dr$, admits uniqueness in law.

The inequality (E.6) implies by Lemma A.4 2. that

$$H(\mathbb{Q}_\epsilon^*|\mathbb{P}_k) = \frac{1}{2} \mathbb{E}^{\mathbb{Q}_\epsilon^*} \left[\int_0^T |\sigma^{-1}(r, X_r) (\beta_r - b(r, X_r, u^k(r, X_r)))|^2 dr \right] < +\infty,$$

hence

$$H(\mathbb{Q}_\epsilon^*|\mathbb{P}_k) \leq 4Tc_\sigma C_{b,\sigma}^2 (1 + C_\infty C(2)) + 4H(\mathbb{Q}_\epsilon^*|\mathbb{P}_\epsilon^*). \quad (\text{E.7})$$

On the other hand, by (3.8) and (E.4),

$$\mathbb{E}^{\mathbb{Q}_\epsilon^*} \left[\int_0^T f(r, X_r, u^k(r, X_r))dr + g(X_T) \right] \leq (T+1)C_{f,g}(1 + C_\infty C(p)). \quad (\text{E.8})$$

Taking into account (1.5) and combining (E.7) and (E.8) yields

$$\begin{aligned} \mathcal{J}(\mathbb{Q}_\epsilon^*, \mathbb{P}_k) &= \mathbb{E}^{\mathbb{Q}_\epsilon^*} \left[\int_0^T f(r, X_r, u^k(r, X_r))dr + g(X_T) \right] + \frac{1}{\epsilon} H(\mathbb{Q}_\epsilon^*|\mathbb{P}_k) \\ &\leq (T+1)C_{f,g}(1 + C_\infty C(p)) + \frac{4Tc_\sigma C_{b,\sigma}^2 (1 + C_\infty C(2))}{\epsilon} + \frac{4}{\epsilon} H(\mathbb{Q}_\epsilon^*|\mathbb{P}_\epsilon^*). \end{aligned} \quad (\text{E.9})$$

Finally, by (3.14) and Jensen's inequality,

$$\begin{aligned}
\frac{1}{\epsilon} H(\mathbb{Q}_\epsilon^* | \mathbb{P}_\epsilon^*) &\leq -\frac{1}{\epsilon} \log \left(\mathbb{E}^{\mathbb{Q}_\epsilon^*} \left[\exp \left(-\epsilon \int_0^T f(r, X_r, u_\epsilon^*(r, X_r)) dr - \epsilon g(X_T) \right) \right] \right) \\
&\leq \mathbb{E}^{\mathbb{Q}_\epsilon^*} \left[\int_0^T f(r, X_r, u_\epsilon^*(r, X_r)) dr + g(X_T) \right] \\
&\leq (T+1) C_{f,g} (1 + C_\infty C(p)),
\end{aligned} \tag{E.10}$$

where we have used (3.8) and (E.4) for the last inequality. Injecting (E.10) in (E.9) yields the desired result by setting $C := 5(T+1)C_{f,g}(1 + C_\infty C(p)) + 4Tc_\sigma C_{b,\sigma}^2(1 + C_\infty C(2))$. \square

F Miscellaneous

We gather in this section two useful technical results. In the following, all the random variables are defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, \mathbb{P})$.

Lemma F.1. *Let η be a square integrable, non-negative random variable. Then for all $\epsilon > 0$,*

$$0 \leq \mathbb{E}[\eta] - \left(-\frac{1}{\epsilon} \log \mathbb{E}[\exp(-\epsilon\eta)] \right) \leq \epsilon \text{Var}[\eta] e^{\epsilon \mathbb{E}[\eta]}.$$

Proof. For all $b \in \mathbb{R}$, it holds by Taylor's formula with integral remainder that

$$e^{-b} = 1 - b + b^2 \int_0^1 (1-t) e^{-tb} dt.$$

A direct application of this formula with $b = \epsilon(\eta - \mathbb{E}[\eta])$ for all $\omega \in \Omega$, yields

$$e^{-\epsilon(\eta - \mathbb{E}[\eta])} = 1 - \epsilon(\eta - \mathbb{E}[\eta]) + \epsilon^2(\eta - \mathbb{E}[\eta])^2 \int_0^1 (1-t) e^{-t\epsilon(\eta - \mathbb{E}[\eta])} dt \leq 1 - \epsilon(\eta - \mathbb{E}[\eta]) + \epsilon^2(\eta - \mathbb{E}[\eta])^2 e^{\epsilon \mathbb{E}[\eta]},$$

taking into account that $\eta \geq 0$.

Taking the expectation in previous inequality we get

$$\mathbb{E} \left[e^{-\epsilon(\eta - \mathbb{E}[\eta])} \right] \leq 1 + \epsilon^2 \text{Var}[\eta] e^{\epsilon \mathbb{E}[\eta]}.$$

Since $\log(1+x) \leq x$ for all $x > -1$, we have

$$\frac{1}{\epsilon} \log \mathbb{E} \left[e^{-\epsilon(\eta - \mathbb{E}[\eta])} \right] \leq \epsilon \text{Var}[\eta] e^{\epsilon \mathbb{E}[\eta]}.$$

Notice that $\mathbb{E}[\eta]$ is a constant, hence $\frac{1}{\epsilon} \log \mathbb{E} \left[e^{-\epsilon(\eta - \mathbb{E}[\eta])} \right] = \mathbb{E}[\eta] - \left(-\frac{1}{\epsilon} \log \mathbb{E} \left[e^{-\epsilon\eta} \right] \right)$. We then have

$$0 \leq \mathbb{E}[\eta] - \left(-\frac{1}{\epsilon} \log \mathbb{E} \left[e^{-\epsilon\eta} \right] \right) \leq \epsilon \text{Var}[\eta] e^{\epsilon \mathbb{E}[\eta]},$$

where the first inequality follows from Jensen's inequality. \square

Lemma F.2. Let $(X_t)_{t \in [0, T]}$ be an (\mathcal{F}_t) -adapted process of the form

$$X_t = x + \int_0^t b_r dr + M_t,$$

where $\mathbb{E} \left[\int_0^T |b_r|^p dr \right] < +\infty$ for some $p > 1$ and where M is a martingale. For Lebesgue almost all $0 \leq t < T$

$$\lim_{h \downarrow 0} \mathbb{E} \left[\frac{X_{t+h} - X_t}{h} \mid \mathcal{F}_t \right] = b_t \text{ in } L^1(\mathbb{P}).$$

Proof. In this proof we extend the process X by continuity after T and b_t by zero for $t > T$. Let $0 < h \leq 1$. Notice first that

$$\mathbb{E} \left[\int_0^T \left| \mathbb{E} \left[\frac{X_{t+h} - X_t}{h} \mid \mathcal{F}_t \right] - b_t \right| dt \right] \leq \mathbb{E} \left[\int_0^T \left| \frac{1}{h} \int_t^{t+h} b_r dr - b_t \right| dt \right],$$

and that for all $\omega \in \Omega$, for almost all $0 \leq t < T$, by Lebesgue differentiation theorem,

$$\frac{1}{h} \int_t^{t+h} b_r(\omega) dr \xrightarrow{h \rightarrow 0} b_t. \quad (\text{F.1})$$

To conclude by a uniform integrability argument w.r.t. $dt \otimes d\mathbb{P}$ we need to prove that

$$\sup_{0 < h \leq 1} \mathbb{E} \left[\int_0^T \left| \frac{1}{h} \int_t^{t+h} b_r dr \right|^p dt \right] < +\infty.$$

Previous expectation, by Hölder inequality, is upper bounded by

$$\mathbb{E} \left[\int_0^T \frac{1}{h} \int_t^{t+h} |b_r|^p dr dt \right] = \mathbb{E} \left[\int_0^T |b_r|^p \frac{1}{h} \int_{(r-h)_+}^r dt dr \right] \leq \mathbb{E} \left[\int_0^T |b_r|^p dr \right] < +\infty,$$

where interchanging the integral inside the expectation is justified by Fubini's theorem. The family $\left(\frac{1}{h} \int_t^{t+h} b_r dr \right)_{0 < h \leq 1}$ is uniformly integrable with respect to $dt \otimes d\mathbb{P}$ and we conclude using the Lebesgue's dominated convergence theorem. \square

Remark F.3. If b_t is a.e. $\sigma(X_t)$ -measurable then the statement of Lemma F.2 still holds replacing the σ -field \mathcal{F}_t with $\sigma(X_t)$. This is an obvious property of the tower property of the conditional expectation.

Acknowledgments

The research of the first named author is supported by a doctoral fellowship PRPhD 2021 of the Région Île-de-France. The research of the second and third named authors was partially supported by the ANR-22-CE40-0015-01 project SDAIM.

References

- [1] A. Beck. *First-order methods in optimization*, volume 25 of *MOS/SIAM Ser. Optim.* Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM); Philadelphia, PA: Mathematical Optimization Society (MOS), 2017.
- [2] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.
- [3] J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comput.*, 37(2):a1111–a1138, 2015.
- [4] C. Bender and T. Moseler. Importance sampling for backward SDEs. *Stochastic Anal. Appl.*, 28(2):226–253, 2010.
- [5] J. Bierkens and H. J. Kappen. Explicit solution of relative entropy weighted control. *Syst. Control Lett.*, 72:36–43, 2014.
- [6] J. F. Bonnans, J. Ch. Gilbert, C. Lemaréchal, and C. A. Sagastizábal. *Numerical optimization. Theoretical and practical aspects. Transl. from the French.* Universitext. 2nd revised. edition, 2006.
- [7] T. Bourdais, N. Oudjane, and F. Russo. A Markovian characterization of the exponential twist of probability measures. *Preprint HAL-04644249*, 2024.
- [8] T. Bourdais, N. Oudjane, and F. Russo. An entropy penalized approach for stochastic optimization with marginal law constraints. complete version. 2025. Preprint hal-04993222.
- [9] G. Brunick and S. Shreve. Mimicking an Itô process by a solution of a stochastic differential equation. *The Annals of Applied Probability*, 23(4):1584–1628, 2013.
- [10] N. Cammardella, A. Bušić, and S. Meyn. Simultaneous allocation and control of distributed energy resources via Kullback-Leibler-Quadratic optimal control. In *2020 American Control Conference (ACC)*, pages 514–520. IEEE, 2020.
- [11] P. Cattiaux and Ch. Léonard. Minimization of the Kullback information of diffusion processes. *Ann. Inst. Henri Poincaré, Probab. Stat.*, 30(1):83–132, 1994.
- [12] Q. Chan-Wai-Nam, J. Mikael, and X. Warin. Machine learning for semi linear PDEs. *J. Sci. Comput.*, 79(3):1667–1712, 2019.
- [13] Y. Chen, T. T. Georgiou, and M. Pavon. On the relation between optimal transport and Schrödinger bridges: a stochastic control viewpoint. *J. Optim. Theory Appl.*, 169(2):671–691, 2016.

- [14] I. Csiszár and G. Tusnády. Information geometry and alternating minimization procedures. Recent results in estimation theory and related topics, Suppl. Issues Stat. Decis. 1, 205-237, 1984.
- [15] M. Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.
- [16] P. Dupuis and R. S. Ellis. *A weak convergence approach to the theory of large deviations*. Wiley Ser. Probab. Stat. Chichester: John Wiley & Sons, 1997.
- [17] I. Ekeland and R. Témam. *Convex analysis and variational problems.*, volume 28 of *Classics Appl. Math.* Philadelphia, PA: Society for Industrial and Applied Mathematics, unabridged, corrected republication of the 1976 English original edition, 1999.
- [18] N. El Karoui, D. Nguyen, and M. Jeanblanc-Picqué. Compactification methods in the control of degenerate diffusions: Existence of an optimal control. *Stochastics*, 20:169–219, 1987.
- [19] W. H. Fleming. Logarithmic transformations and stochastic control. Advances in filtering and optimal stochastic control, Proc. IFIP-WG 7/1 Work. Conf., Cocoyoc/Mex, Lect. Notes Contr. Inf. Sci. 42, 131-141, 1982.
- [20] W. H. Fleming and S. K. Mitter. Optimal control and nonlinear filtering for nondegenerate diffusion processes. *Stochastics*, 8:63–77, 1982.
- [21] M. Germain, H. Pham, and X. Warin. Approximation error analysis of some deep backward schemes for nonlinear PDEs. *SIAM J. Sci. Comput.*, 44(1):a28–a56, 2022.
- [22] E. Gobet and C. Labart. Solving BSDE with adaptive control variate. *SIAM J. Numer. Anal.*, 48(1):257–277, 2010.
- [23] E. Gobet and P. Turkedjiev. Adaptive importance sampling in least-squares Monte Carlo algorithms for backward stochastic differential equations. *Stochastic Processes Appl.*, 127(4):1171–1203, 2017.
- [24] J. Han, A. Jentzen, and W. E. Solving high-dimensional partial differential equations using deep learning. *Proc. Natl. Acad. Sci. USA*, 115(34):8505–8510, 2018.
- [25] U. G. Haussmann. Existence of optimal Markovian controls for degenerate diffusions. Stochastic differential systems, Proc. 3rd Bad Honnef Conf. 1985, Lect. Notes Control Inf. Sci. 78, 171-186, 1986.
- [26] U. G. Haussmann and J. P. Lepeltier. On the existence of optimal controls. *SIAM J. Control Optim.*, 28(4):851–902, 1990.

- [27] C. Huré, H. Pham, and X. Warin. Deep backward schemes for high-dimensional nonlinear PDEs. *Math. Comput.*, 89(324):1547–1579, 2020.
- [28] L. Izydorczyk, N. Oudjane, and F. Russo. A fully backward representation of semilinear PDEs applied to the control of thermostatic loads in power systems. *Monte Carlo Methods and Applications*, 27(4):347–371, 2021.
- [29] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, second edition, 2003.
- [30] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.
- [31] N. V. Krylov. *Controlled diffusion processes*, volume 14 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, 2009. Translated from the 1977 Russian original by A. B. Aries, Reprint of the 1980 edition.
- [32] D. Lacker. Mean field games via controlled martingale problems: existence of Markovian equilibria. *Stochastic Processes Appl.*, 125(7):2856–2894, 2015.
- [33] D. Lacker. Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions. *Probab. Math. Phys.*, 4(2):377–432, 2023.
- [34] Ch. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014.
- [35] T. Mikami and M. Thieullen. Duality theorem for the stochastic optimal control problem. *Stochastic Processes and their Applications*, 116 n.12:1815–1835, 2006.
- [36] M. Pavon, E. G. Tabak, and G. Trigila. The data-driven Schrödinger bridge. *Commun. Pure Appl. Math.*, 74(7):1545–1573, 2021.
- [37] J. Pfanzagl. Convexity and conditional expectations. *Ann. Probab.*, 2:490–494, 1974.
- [38] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.
- [39] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes*. Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1997 edition.
- [40] X. Tan and N. Touzi. Optimal transportation under controlled stochastic dynamics. *The Annals of Probability*, 41(5):3201 – 3240, 2013.

- [41] E. Theodorou, J. Buchli, and S. Schaal. Reinforcement learning of motor skills in high dimensions: A path integral approach. In *2010 IEEE International Conference on Robotics and Automation*, pages 2397–2403, 2010.
- [42] S. Thijssen and H.J. Kappen. Path integral control and state-dependent feedback. *Physical Review E*, 91(3):032104, 2015.
- [43] N. Touzi. *Optimal stochastic control, stochastic target problems, and backward SDE*, volume 29 of *Fields Institute Monographs*. Springer, New York; Fields Institute for Research in Mathematical Sciences, Toronto, ON, 2013. With Chapter 13 by Agnès Tourin.
- [44] X. Zhang. Strong solutions of SDEs with singular drift and Sobolev diffusion coefficients. *Stochastic Processes and their Applications*, 115(11):1805–1818, 2005.